ANALYSIS & PDE

Volume 10 No. 4 2017

msp

Analysis & PDE

msp.org/apde

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

The subscription price for 2017 is US $265/year for the electronic version, and $470/year (+$55, 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. 10, No. 4, 2017
dx.doi.org/10.2140/apde.2017.10.757

THE FUGLEDE CONJECTURE HOLDS IN $\mathbb{Z}_p \times \mathbb{Z}_p$

ALEX IOSEVICH, AZITA MAYELI AND JONATHAN PAKIANATHAN

In this paper we study subsets E of $\mathbb{Z}_p^d$ such that any function $f : E \to \mathbb{C}$ can be written as a linear combination of characters orthogonal with respect to E. We shall refer to such sets as spectral. In this context, we prove the Fuglede conjecture in $\mathbb{Z}_p^2$, which says in this context that $E \subset \mathbb{Z}_p^2$ is spectral if and only if E tiles $\mathbb{Z}_p^2$ by translation. Arithmetic properties of the finite field Fourier transform, elementary Galois theory and combinatorial geometric properties of direction sets play the key role in the proof. The proof relies to a significant extent on the analysis of direction sets of Iosevich et al. (*Integers* **11** (2011), art. id. A39) and the tiling results of Haessig et al. (2011).

1. Introduction

Let $E \subset \mathbb{Z}_p^d$, where $\mathbb{Z}_p$, p prime, is the cyclic group of size p and $\mathbb{Z}_p^d$ is the d-dimensional vector space over $\mathbb{Z}_p$. We say that $L^2(E)$ has an orthogonal basis of exponentials (indexed by A) if the following conditions hold:

- (completeness) There exists $A \subset \mathbb{Z}_p^d$ such that for every function $f : E \to \mathbb{C}$ there exist complex numbers $\{c_a\}_{a \in A}$, $A \subset \mathbb{Z}_p^d$, such that

$$f(x) = \sum_{a \in A} \chi(x \cdot a) c_a$$

  for all $x \in E$ where $\chi(u) = e^{2\pi i u/p}$. We shall refer to A as a *spectrum* of E. The expansion above can be applied to functions $f : \mathbb{Z}_p^d \to \mathbb{C}$ by restricting them to E but the equality holds only for $x \in E$.
- (orthogonality) The relation

$$\sum_{x \in E} \chi(x \cdot (a - a')) = 0$$

  holds for every $a, a' \in A$, $a \neq a'$.

  If these conditions hold, we refer to $E \subset \mathbb{Z}_p^d$ as a *spectral* set.

Definition 1.1 (spectral pair). A *spectral pair* (E, A) in $V = \mathbb{Z}_p^d$ is a spectral set E with an orthogonal basis of exponentials indexed by A.

Definition 1.2 (tiling pair). A *tiling pair* (E', A') consists of $E', A' \subset \mathbb{Z}_p^d$ such that every element $v \in V$ can be written uniquely as a sum $v = e' + a'$, $e' \in E'$, $a' \in A'$. Equivalently, (E', A') is a tiling pair if

MSC2010: 05A18, 11P99, 41A10, 42B05, 52C20.
Keywords: exponential bases, Erdős problems, Fuglede conjecture.

$\sum_{a' \in A'} E'(x-a') \equiv 1$ for every $x \in V$. We say that E' *tiles* V by translation if there exists $A' \subset V$ such that (E', V') is a tiling pair. Here and throughout $E(x)$ is the indicator function of E.

The study of the relationship between exponential bases and tiling has its roots in the celebrated Fuglede conjecture in $\mathbb{R}^d$, which says that if $E \subset \mathbb{R}^d$ is of positive Lebesgue measure, then $L^2(E)$ possesses an orthogonal basis of exponentials if and only if E tiles $\mathbb{R}^d$ by translation. Fuglede [1974] proved this conjecture in the case when either the tiling set or the spectrum is a lattice. Katz, Tao and the first author [Iosevich et al. 2003] proved that the Fuglede conjecture holds for convex planar domains.

Terry Tao [2004] disproved the Fuglede conjecture by exhibiting a spectral set in $\mathbb{R}^{12}$ which does not tile. The first step in his argument is the construction of a spectral subset of $\mathbb{Z}_3^5$ of size 6. It is easy to see that this set does not tile because 6 does not divide 3^5. As a by-product, this shows that spectral sets in $\mathbb{Z}_p^d$ do not necessarily tile. See [Kolountzakis and Matolcsi 2006], where the authors also disprove the reverse implication of the Fuglede conjecture. Tao's example raises the natural question of whether and when spectral sets in a variety of settings necessarily tile by translation and vice versa. In this paper we see that the Fuglede conjecture holds in two-dimensional vector spaces over prime fields.

Our main result is the following.

Theorem 1.3. *Let E be a subset of $\mathbb{Z}_p^d$, p an odd prime.*

(i) (*density*) *The space $L^2(E)$ has an orthogonal basis of exponentials indexed by A if and only if $|E| = |A|$ and $\widehat{E}(a-a') = 0$ for all distinct $a, a' \in A$.*

(ii) *If $E \subset \mathbb{Z}_p^d$ is spectral and $|E| > p^{d-1}$ then $E = A = \mathbb{Z}_p^d$.*

(iii) (*divisibility*) *If $E \subset \mathbb{Z}_p^d$ is spectral, then $|E|$ is 1 or a multiple of p.*

(iv) (*Fuglede conjecture in $\mathbb{Z}_p^2$*) *A set $E \subset \mathbb{Z}_p^2$ is a spectral set if and only if E tiles $\mathbb{Z}_p^2$ by translation.*

We note that the Fuglede conjecture holds trivially also in $\mathbb{Z}_p^1$, as a tiling set E must have $|E|$ divide p and thus must be a point or the whole space, and hence is also a spectral set. Conversely, a spectral set E must have size 1 or a multiple of p by the divisibility condition of the theorem above, and so also is either a point or the whole space, and hence is a tiling set. We also note the results of the theorem above also hold for $p = 2$ but we choose to focus on the odd prime case in the rest of the paper. Parts (i)–(iii) extend with no difficulty and indeed imply $|E| \in \{1, 2, 4\}$ if E is either a spectral set or a tiling set. As sets of size 2 are lines, which are both tiling sets and spectral sets, (iv) follows also.

2. Basic properties of spectra

Lemma 2.1. *Suppose that $L^2(E)$ has an orthogonal basis of exponentials and*

$$f : \mathbb{Z}_p^d \to \mathbb{C}.$$

Then the coefficients are given by

$$c_a(f) = |E|^{-1} \sum_{x \in E} \chi(-x \cdot a) f(x).$$

To prove this, observe that if $f(x) = \sum_{a\in A} \chi(x\cdot a)c_a$ for $x \in E$, then

$$\begin{aligned} |E|^{-1} \sum_{x\in E} \chi(-x\cdot a) f(x) &= |E|^{-1} \sum_{x\in E}\sum_{b\in A} \chi(-x\cdot(a-b))c_b(f) \\ &= |E|^{-1} \sum_{b\in A}\sum_{x\in E} \chi(-x\cdot(a-b))c_b(f) = c_a(f) \end{aligned}$$

and the proof is complete.

Lemma 2.2 (delta function). *Suppose that $L^2(E)$ has an orthogonal basis of exponentials with the spectrum A. Let $\delta_0(x) = 1$ if $x = \vec{0}$ and 0 otherwise and suppose $\vec{0} \in E$. Then*

$$\delta_0(x) = |E|^{-1} \sum_{a\in A} \chi(x\cdot a).$$

To prove the lemma, observe that if $f(x) = \delta_0(x)$, then

$$c_a(f) = |E|^{-1} \sum_{x\in E} \chi(-x\cdot a)\delta_0(x) = |E|^{-1}.$$

The conclusion follows from Lemma 2.1.

Lemma 2.3 (Parseval). *Suppose that $L^2(E)$ has an orthogonal basis of exponentials and f is any function on $\mathbb{Z}_p^d$ with values in $\mathbb{C}$. Then*

$$\sum_{a\in A} |c_a(f)|^2 = |E|^{-1} \sum_{x\in E} |f(x)|^2.$$

Lemma 2.4 (density). *Suppose that $L^2(E)$ has an orthogonal basis of exponentials with the spectrum A. Then $|E| = |A|$.*

The set of functions $\{\chi(x\cdot a) : a \in A\}$ is, by completeness, a spanning set for $L^2(E)$ and, by orthogonality, a linearly independent set for $L^2(E)$ and hence is a basis for $L^2(E)$. Thus the cardinality of this set, which is $|A|$, is equal to the dimension of $L^2(E)$, which is $|E|$.

3. Proof of Theorem 1.3

Part (i) of Theorem 1.3 follows easily, as we have seen that if (E, A) is a spectral pair then $|E| = |A|$ and since the orthogonality property can be easily rewritten as $\widehat{E}(a - a') = 0$ for all $a \neq a'$, with $a, a' \in A$. Conversely if (E, A) has the last two properties, it is a spectral pair, as orthogonality implies $\{\chi(x\cdot a) : a \in A\}$ is linearly independent in $L^2(E)$ and $|A| = |E|$ ensures it is a basis and hence that completeness is satisfied.

Definition 3.1 [Iosevich et al. 2011]. We say that two vectors x and x' in $\mathbb{Z}_p^d$ point in the same direction if there exists $t \in \mathbb{F}_q^*$ such that $x' = tx$. Here $\mathbb{F}_q^*$ denotes the multiplicative group of $\mathbb{Z}_p$. Writing this equivalence as $x \sim x'$, we define the set of directions as the quotient

$$\mathcal{D}(\mathbb{Z}_p^d) = \mathbb{Z}_p^d / \sim . \tag{3-1}$$

Similarly, we can define the set of directions determined by $E \subset \mathbb{Z}_p^d$ by

$$\mathcal{D}(E) = E - E/\sim, \tag{3-2}$$

where

$$E - E = \{x - y : x, y \in E\},$$

with the same equivalence relation $\sim$ as in (3-1) above.

The following result, which is one of the two key tools in the proof of our main result, was previously established in [Iosevich et al. 2011].

Theorem 3.2. *A set E does not determine all directions if and only if there is a hyperplane H and $S \subseteq H$ such that E is the graph of a function $f : S \to \mathbb{Z}_p$ over H, which means that relative to some decomposition $\mathbb{Z}_p^d = H \oplus \mathbb{Z}_p$, we have $E = \{(x, f(x)) : x \in S\}$. In particular, if $|E| > p^{d-1}$, every possible direction is determined by E.*

The second main tool in our proof is the following result.

Theorem 3.3 [Haessig et al. 2015, Proposition 3.2]. *Let $E \subset \mathbb{Z}_p^d$. Then $\widehat{E}(m) = 0$ implies that $\widehat{E}(rm) = 0$ for all $r \in \mathbb{Z}_p^*$. Furthermore $\widehat{E}(m) = 0$ for $m \neq 0$ if and only if E is equidistributed on the p hyperplanes $H_t = \{x : x \cdot m = t\}$ for $t \in \mathbb{Z}_p$ in the sense that*

$$\sum_{x \cdot m = t} E(x) = |E \cap H_t|,$$

viewed as a function of t, is constant.

Note this last theorem is a fact about rational-valued functions over prime fields that is *not true* for complex-valued functions in general or over other fields. We give the proof of Theorem 3.3 at the end of the paper for the sake of completeness.

The proof of part (ii) of Theorem 1.3 follows fairly easily from combining Theorems 3.2 and 3.3. Indeed, suppose that $L^2(E)$ has an orthogonal basis of exponentials and $|E| > p^{d-1}$. By Lemma 2.4, $|E| = |A| > p^{d-1}$. By Theorem 3.2, $\mathcal{D}(A) = \mathcal{D}(\mathbb{Z}_p^d)$. Combining this with Theorem 3.3 implies that $\widehat{E}$ vanishes on $\mathbb{Z}_p^d \setminus \vec{0}$. It follows that $E = \mathbb{Z}_p^d$, as claimed.

Part (iii) of Theorem 1.3 is contained in the following result. A spectral pair is called trivial if $(E, A) =$ (point, another point) or $(E, A) = (\mathbb{Z}_p^d, \mathbb{Z}_p^d)$ or $(E, A) = (\varnothing, \varnothing)$. All other spectral pairs are called nontrivial.

Proposition 3.4. *Let p be an odd prime and (E, A) be a nontrivial spectral pair in $\mathbb{Z}_p^d$. Then $|E| = |A| = mp$, where $m \in \{1, 2, 3, \ldots, p^{d-2}\}$.*

To prove Proposition 3.4, let (E, A) be a nontrivial spectral pair in $\mathbb{Z}_p^d$. Then part (i) of Theorem 1.3 shows that $|E| = |A|$ and $\widehat{E}(a - a') = 0$ for distinct $a, a' \in A$. Since the spectral pair (E, A) is nontrivial, $2 \le |E| = |A| \le p^{d-1}$ also. Thus taking two distinct elements $a, a' \in A$ shows that $\widehat{E}(\alpha) = 0$ for $\alpha = a - a' \neq 0$. Thus E is equidistributed on the p parallel hyperplanes

$$H_t = \{x : x \cdot \alpha = t\},$$

$t \in \mathbb{Z}_p$, by Theorem 3.3. Thus if E has $m \geq 1$ elements per hyperplane we have $|E| = |A| = mp$. Then $1 \leq m \leq p^{d-2}$ since $0 < |E| \leq p^{d-1}$. This proves part (iii) of Theorem 1.3.

Observe that if $d = 2$ and (E, A) is a nontrivial spectral pair, then $|E| = |A| = mp$ implies $|E| \geq p$, while $|E| \leq p$ by part (ii) of Theorem 1.3 and so $|E| = |A| = p$. Furthermore, by Theorem 3.2 above, A is a graph of a function $\mathbb{Z}_p \to \mathbb{Z}_p$ since $|A| = p$ and it does not determine all directions. Finally, since E is equidistributed on a family of p parallel lines and $|E| = p$, we see that E is also a graph of a function $\mathbb{Z}_p \to \mathbb{Z}_p$ with respect to some system of axes. The following is an immediate corollary of Proposition 3.4.

Corollary 3.5. *If E is a spectral set in $\mathbb{Z}_p^2$, p an odd prime, then E is either a point, a graph set of order p or the whole space and hence tiles $\mathbb{Z}_p^2$ in all cases.*

This corollary follows from Proposition 3.4 immediately once one notes that any graph set

$$E = \{(x, f(x)) : x \in \mathbb{Z}_p\}$$

for a function f, with respect to some coordinate systems, tiles $\mathbb{Z}_p^2$ using the tiling partner

$$A = \{(0, t) : t \in \mathbb{Z}_p\}.$$

To complete the proof of the Fuglede conjecture in two dimensions over prime fields, which is the content of part (iv) of Theorem 1.3, it remains to show that any tiling set is spectral since we have just shown that any spectral set tiles.

Proposition 3.6 (sets which tile by translation are spectral). *Let p be an odd prime, and let $E \subseteq \mathbb{Z}_p^2$. Suppose that E tiles $\mathbb{Z}_p^2$ by translation. Then E is a spectral set.*

We shall need the following result. We shall prove it at the end of the paper for the sake of completeness.

Theorem 3.7 [Haessig et al. 2015, Theorem 1.7]. *Let E be a set that tiles $\mathbb{Z}_p^2$. Then $|E| = 1$, p or p^2 and E is a graph if $|E| = p$.*

We include a proof of Theorem 3.7 at the end of this paper for completeness.

The cases $|E| = 1$, p^2 are trivially spectral sets so we may reduce to the case that E is a graph, i.e.,

$$E = \{xe_1 + f(x)e_2 : x \in \mathbb{Z}_p\},$$

where e_1, e_2 is a basis for $\mathbb{Z}_p^2$ and $f : \mathbb{Z}_p \to \mathbb{Z}_p$ is a function. By changing the function if necessary we can assume e_2 is orthogonal to e_1 as long as $e_1 \cdot e_1 \neq 0$, i.e., e_1 does not generate an isotropic line. This is always the case if $p \equiv 3 \bmod 4$. In the case when $p \equiv 1 \bmod 4$, it is possible that e_1 generates one of the two isotropic lines

$$\{(t, it) : t \in \mathbb{Z}_p\},$$

where i is one of the two distinct solutions of the equation $x^2+1=0$. The reason this case needs to be treated separately is that $(t_1, it_1)\cdot(t_2, it_2)=0$ for all $t_1, t_2 \in \mathbb{Z}_p$. To deal with this, we note that the other solution of the equation $x^2+1=0$ is given by $-i$ and we take e_2 to be on the other isotropic line in the plane, given by

$$\{(t, -it) : t \in \mathbb{Z}_p\},$$

with e_2 normalized so that $e_1 \cdot e_2 = 1$.

There are two situations to consider.

Case 1: e_1 and e_2 are orthogonal. Then we will take $A = \{xe_1 : x \in \mathbb{Z}_p\}$. To show that (E, A) is a spectral pair, we need only show that the set $\{\chi(ae_1 \cdot x) : a \in \mathbb{Z}_p\}$ is orthogonal in $L^2(E)$. By Theorem 3.3 this happens if and only if $\widehat{E}((a-a')e_1) = 0$ for all distinct $a, a' \in \mathbb{Z}_p$, which happens if and only if E equidistributes on the p parallel lines normal to e_1, i.e., on the p parallel lines of constant e_1-coordinate in the (e_1, e_2)-grid. This is clearly the case as E is a graph over the e_1-coordinate and hence has exactly one element on each of these parallel lines, so this case is proven.

Case 2: e_1 and e_2 generate the two isotropic lines in $\mathbb{Z}_p^2$, $p = 1 \bmod 4$. In this case $e_1 \cdot e_2 \neq 0$ but $e_1 \cdot e_1 = e_2 \cdot e_2 = 0$. Since E is equidistributed along the p parallel lines of constant e_1-coordinate, it is easy to see that these are the same family of lines as $H_t = \{x : x \cdot e_2 = t\}$, $t \in \mathbb{Z}_p$. Thus in this case using $A = \{ae_2 : a \in \mathbb{Z}_p\}$ we find that $\widehat{E}((a-a')e_2) = 0$ for distinct $a, a' \in \mathbb{Z}_p$ and so (E, A) is a spectral pair. Thus E is still spectral in this case and the theorem is proven in all cases.

4. Proof of Theorem 3.3

We include the proof of Theorem 3.3 for the sake of completeness. We have

$$\widehat{E}(m) = p^{-d} \sum_{x \in \mathbb{Z}_p^d} \chi(-x \cdot m) E(x) = 0$$

for some $m \neq (0, \ldots, 0)$. Let $\xi = \chi(-1) = e^{-2\pi i/p}$. Note that ξ is a primitive p-th root of unity. It follows that

$$0 = \sum_{x \in \mathbb{Z}_p^d} \xi^{x \cdot m} E(x) = \sum_{t \in \mathbb{Z}_p} \xi^t \sum_{x \cdot m = t} E(x).$$

Let

$$n(t) = \sum_{x \cdot m = t} E(x) \in \mathbb{Q},$$

so

$$\sum_{t \in \mathbb{Z}_p} \xi^t n(t) = 0.$$

This means that ξ is a root of the rational polynomial

$$P(u) = \sum_{t=0}^{p-1} n(t) u^t.$$

The minimal polynomial of ξ, over $\mathbb{Q}$, is

$$Q(u) = 1 + u + \cdots + u^{p-1},$$

so by elementary Galois theory, $P(u)$ is a constant multiple of $Q(u)$ since ξ is a root of the rational polynomial P and Q is the minimal polynomial of ξ. It follows that the coefficients of $n(t)$ are independent of t. This proves the second assertion of Theorem 3.3, namely that E is equidistributed on the hyperplanes $H_t = \{x \in \mathbb{Z}_p^d : x \cdot m = t\}$.

Let us now prove that if $\widehat{E}(m) = 0$ for some $m \neq (0, \ldots, 0)$, then $\widehat{E}(rm) = 0$ for all $r \neq 0$. We have

$$\sum_{x \in \mathbb{Z}_p^d} \chi(-x \cdot rm) E(x) = \sum_{t \in \mathbb{Z}_p} \xi^t \sum_{x \cdot rm = t} E(x) = \sum_{t \in \mathbb{Z}_p} \xi^t \sum_{x \cdot m = tr^{-1}} E(x) = \sum_{t \in \mathbb{Z}_p} \xi^t n(r^{-1} t).$$

For a fixed r, it follows from above that $n(r^{-1}t)$ is independent of t. Therefore

$$\sum_{t \in \mathbb{Z}_p} \xi^t n(r^{-1} t) = \sum_{t \in \mathbb{Z}_p} \xi^t n(t) = 0$$

and the proof of the claim follows. This completes the proof of Theorem 3.3.

Note the proof above generalizes to rational-valued functions but not to real- or complex-valued functions. The reason is that over $\mathbb{R}$ or $\mathbb{C}$, a polynomial that ξ is a root of need not be a multiple of $1 + x + x^2 + \cdots + x^{p-1}$; for example, $P(x) = x - \xi$ or $P(x) = (x - \xi)(x - \bar{\xi}) = x^2 - 2\cos(2\pi/p) + 1$.

5. Proof of Theorem 3.7

Let A denote the set that tiles E. Note that $|E||A| = p^2$, so $|E| = 1$, p or p^2. If $|E| = 1$ then E is a point and we are done. If $|E| = p^2$ then E is the whole plane and we are done, so without loss of generality let $|E| = p$.

If $\widehat{E}(m)$ never vanishes then E is a point and we are done. On the other hand if $\widehat{E}(m) = 0$ for some nonzero m, then it vanishes on L, the line passing through the origin and $m \neq \vec{0}$. Thus if we set $L^\perp$ to be the line through the origin, perpendicular to m, we see that

$$\widehat{L^\perp}(s) \widehat{E}(s) = 0$$

for all nonzero s. This is because by a straightforward calculation

$$\widehat{L^\perp}(s) = q^{-(d-1)} L(s).$$

Since $|L^\perp| = p = |E|$ we then see that E tiles $\mathbb{F}_p^2$ by $L^\perp$.

Since $\widehat{E}(m) = 0$ for some nonzero vector m, we see that E is equidistributed on the set of p lines $H_t = \{x : x \cdot m = t\}$, $t \in \mathbb{F}_p$. Since $|E| = p$ this means there is exactly one point of E on each of these lines.

We will now choose a coordinate system in which E will be represented as a graph of a function. The coordinate system will either be an orthogonal system or an isotropic system depending on the nature of the vector m. There are two cases to consider.

Case 1: $m \cdot m \neq 0$: We may set $e_1 = m$ and e_2 a vector orthogonal to m. Now $\{e_1, e_2\}$ is an orthogonal basis because e_2 does not lie on the line through m, as this line is not isotropic. If we take a general vector $hx = x_1 e_1 + x_2 e_2$ we see that $hx \cdot m = x_1(m \cdot m)$ and so the lines H_t, $t \in \mathbb{F}_p$, are the same as the lines of constant x_1-coordinate with respect to this orthogonal basis $\{e_1, e_2\}$. Thus there is a unique value of x_2 for any given value of x_1 so that $x_1 e_1 + x_2 e_2 \in E$. Thus $E = \{x_1 e_1 + f(x_1) e_2 : x_1 \in \mathbb{Z}_p\} = \text{Graph}(f)$ for some function $f : \mathbb{F}_p \to \mathbb{Z}_p$.

Case 2: $m \cdot m = 0$: We may set $e_1 = m$. In this case any vector orthogonal to e_1 lies on the line generated by e_1 and so cannot be part of a basis with e_1. Instead we select e_2 off the line generated by e_1 and scale it so that $e_1 \cdot e_2 = 1$. Then by subtracting a suitable multiple of e_1 from e_2 one can also ensure $e_2 \cdot e_2 = 0$.

Thus $\{e_1, e_2\}$ is a basis consisting of two linearly independent isotropic vectors. With respect to this basis, the dot product is represented by the matrix

$$\mathbb{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

which exhibits the plane as the hyperbolic plane. This case can only occur when $p = 1 \bmod 4$.

Note when we express a general vector $x = x_1 e_1 + x_2 e_2$ with respect to this basis we have $x \cdot m = x_2$; thus the lines $\{H_t : t \in \mathbb{Z}_p\}$ are the same as the lines of constant x_2-coordinate with respect to this basis and E has a unique point on each of these lines. Thus $E = \{f(x_2)e_1 + x_2 e_2 : x_2 \in \mathbb{Z}_p\} = \mathrm{Graph}(f)$ is a graph with respect to this isotropic coordinate system.

Finally we note any function $f : \mathbb{Z}_p \to \mathbb{F}_p$ is given by a polynomial of degree at most $p-1$, explicitly expressed in the form

$$f(x) = \sum_{k \in \mathbb{Z}_p} f(k) \frac{\prod_{j \neq k}(x - j)}{\prod_{j \neq k}(k - j)}.$$

References

[Fuglede 1974] B. Fuglede, "Commuting self-adjoint partial differential operators and a group theoretic problem", *J. Functional Analysis* **16** (1974), 101–121. MR Zbl

[Haessig et al. 2015] C. D. Haessig, A. Iosevich, J. Pakianathan, S. Robins, and L. Vaicunas, "Tiling, circle packing and exponential sums over finite fields", preprint, 2015. arXiv

[Iosevich et al. 2003] A. Iosevich, N. Katz, and T. Tao, "The Fuglede spectral conjecture holds for convex planar domains", *Math. Res. Lett.* **10**:5-6 (2003), 559–569. MR Zbl

[Iosevich et al. 2011] A. Iosevich, H. Morgan, and J. Pakianathan, "On directions determined by subsets of vector spaces over finite fields", *Integers* **11** (2011), art. id. A39. MR Zbl

[Kolountzakis and Matolcsi 2006] M. N. Kolountzakis and M. Matolcsi, "Tiles with no spectra", *Forum Math.* **18**:3 (2006), 519–528. MR Zbl

[Tao 2004] T. Tao, "Fuglede's conjecture is false in 5 and higher dimensions", *Math. Res. Lett.* **11**:2-3 (2004), 251–258. MR Zbl

Received 3 Dec 2015. Revised 15 Dec 2015. Accepted 11 Mar 2016.

ALEX IOSEVICH: iosevich@math.rochester.edu
Department of Mathematics, University of Rochester, Rochester, NY 14627, United States

AZITA MAYELI: amayeli@qcc.cuny.edu
Department of Mathematics and Computer Science, Queensborough Community College, 222-05 56th Ave., Bayside, NY 11364, United States

JONATHAN PAKIANATHAN: jonpak@math.rochester.edu
Department of Mathematics, University of Rochester, Rochester, NY 14627, United States

ANALYSIS AND PDE
Vol. 10, No. 4, 2017
dx.doi.org/10.2140/apde.2017.10.765

DISTORTED PLANE WAVES IN CHAOTIC SCATTERING

MAXIME INGREMEAU

We provide a precise description of distorted plane waves for semiclassical Schrödinger operators under the assumption that the classical trapped set is hyperbolic and that a certain topological pressure (a quantity defined using thermodynamical formalism) is negative. Distorted plane waves are generalized eigenfunctions of the Schrödinger operator which differ from free plane waves, $e^{i\langle x,\xi\rangle/h}$, by an outgoing term. Under our assumptions we show that they can be written as a convergent sum of Lagrangian states. That provides a description of their semiclassical defect measures in the spirit of quantum ergodicity and extends results of Guillarmou and Naud obtained for hyperbolic quotients to our setting.

1. Introduction

In this paper, we will consider on $\mathbb{R}^d$ a semiclassical Hamiltonian of the form

$$P_h = -h^2\Delta + V(x), \quad V \in C_c^\infty(\mathbb{R}^d).$$

We will study the "distorted plane waves", or "scattering states" associated to P_h. They are a family of functions $E_h^\xi \in C^\infty(\mathbb{R}^d)$ with parameter $\xi \in \mathbb{S}^d$ (the direction of propagation of the incoming wave) which are generalized eigenfunctions of P_h; that is to say, they satisfy the differential equation

$$(P_h - 1)E_h^\xi = 0, \tag{1}$$

but are not in $L^2(\mathbb{R}^d)$ (since P_h has no embedded eigenvalues in $\mathbb{R}^+$).

These distorted plane waves resemble the actual plane waves, in the sense that we may write

$$E_h^\xi(x) = e^{\frac{i}{h}x\cdot\xi} + E_{\text{out}}^\xi, \tag{2}$$

where E_{out} is outgoing in the sense that it satisfies the Sommerfeld radiation condition:

$$\lim_{|x|\to\infty} |x|^{(d-1)/2}\left(\frac{\partial}{\partial|x|} - \frac{i}{h}\right)E_{\text{out}}^\xi(x) = 0. \tag{3}$$

One can show (see, for instance, [Melrose 1995, §2; Dyatlov and Zworski 2017, §4]) that for any $\xi \in \mathbb{S}^{d-1}$ and $h > 0$, there exists a unique function E_h^ξ satisfying conditions (1), (2) and (3).

Condition (3) may be equivalently stated by asking that E_{out}^ξ is the image of a function in $C_c^\infty(\mathbb{R}^d)$ by the outgoing resolvent $(P_h - (1+i0)^2)^{-1}$, or by asking that E_{out}^ξ be of the form

$$E_{\text{out}}^\xi(x) = e^{i|x|/h}|x|^{-(d-1)/2}\left(a_h^\xi(\omega) + O\left(\frac{1}{|x|}\right)\right),$$

MSC2010: 35P20, 35P25, 81Q50.
Keywords: scattering theory, quantum chaos, semiclassical measures, distorted plane waves.

where $\omega = x/|x|$. The function $a_h(\xi,\omega) := a_h^{\xi}(\omega)$ is called the *scattering amplitude*, and is the integral kernel of the scattering matrix minus identity. The scattering amplitude, and hence the distorted plane waves, are central objects in scattering theory.

The aim of this paper is to discuss the behaviour of distorted plane waves in the semiclassical limit $h \to 0$. Distorted plane waves can be seen as an analogue, on manifolds of infinite volume, of the eigenfunctions of a Schrödinger operator on a compact manifold. It is therefore natural to ask questions similar to those in the compact case: what can be said about the semiclassical measures of distorted plane waves, about the behaviour of their L^p norms as $h \to 0$, and about their nodal sets and nodal domains?

The answer to these questions will depend in a drastic way on the properties of the underlying *classical dynamics*. Let us define the classical Hamiltonian by

$$p(x,\xi) = |\xi|^2 + V(x),$$

and the layer of energy 1 as

$$\mathcal{E} = \{\rho \in T^*\mathbb{R}^d : p(\rho) = 1\}.$$

Note that this is a noncompact set, but its intersection with any fibre T_x^*X is compact.

We also denote, for each $t \in \mathbb{R}$, the Hamiltonian flow generated by p by $\Phi^t : T^*\mathbb{R}^d \to T^*\mathbb{R}^d$. For $\rho \in \mathcal{E}$, we will say that $\rho \in \Gamma^{\pm}$ if $\{\Phi^t(\rho) : \pm t \leq 0\}$ is a bounded subset of $T^*\mathbb{R}^d$; that is to say, ρ does not "go to infinity", respectively in the past or in the future. The sets $\Gamma^{\pm}$ are called respectively the *outgoing* and *incoming* tails (at energy 1).

The *trapped set* is defined as

$$K := \Gamma^+ \cap \Gamma^-. \tag{4}$$

It is a flow-invariant set, and it is compact, because V is compactly supported.

If the trapped set is empty, then we can easily describe the distorted plane waves in the semiclassical limit. Namely, one can show (see [Dyatlov and Guillarmou 2014, §5.1]) that E_h^{ξ} is a *Lagrangian* (*WKB*) *state*. Furthermore, for any $\chi \in C_c^{\infty}(\mathbb{R}^d)$, the norm $\|\chi E_h^{\xi}\|_{L^2}$ is bounded independently of h.

However, if the trapped set is nonempty, the distorted plane waves may not be bounded uniformly in L^2_{loc} as $h \to 0$. Actually, $\|\chi E_h^{\xi}\|_{L^2}$ could grow exponentially fast as $h \to 0$. If we want this quantity to remain bounded uniformly in h, we must therefore make some additional assumptions on the classical dynamics. Let us now detail these assumptions.

Hypotheses on the classical dynamics.

- *Hyperbolicity assumption*: In the sequel, we will suppose that the potential V is such that the trapped set contains no fixed point, and is a *hyperbolic set*. We refer to Section 2.1.2 for the definition of a hyperbolic set. The potential in Figure 1 is an example of such a potential.

- *Topological pressure assumption*: For our result on distorted plane waves to hold, we must also make the assumption (Hypothesis 46) that the topological pressure associated to half the logarithm of the unstable Jacobian of the flow on K is negative. The definition of the topological pressure will be recalled in Section 3.4. Hypothesis 46 roughly says that the system is "very open". One should note that in dimension 2, this condition is equivalent to the fact that the Hausdorff dimension of K is strictly smaller

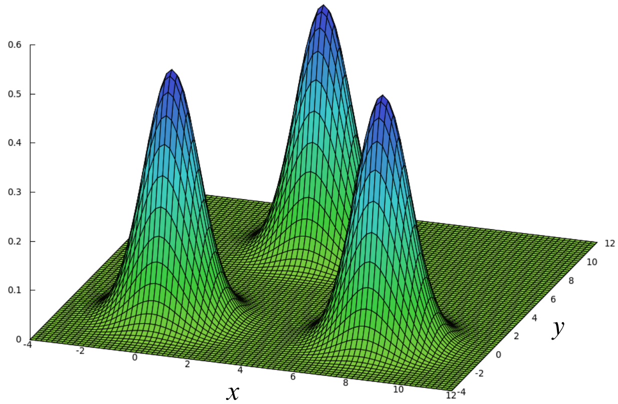


Figure 1. An example of a potential on $(\mathbb{R}^2, g_{\text{flat}})$ such that the dynamics is hyperbolic on the trapped set in some energy range. (See [Sjöstrand 1990, Appendix C] for details.)

than 2. In the three-bumps potential of Figure 1, this condition is satisfied if the three bumps are far enough from each other, but it is not satisfied if the bumps are close to each other.

• *Transversality assumption*: Our last assumption does not concern directly the classical dynamics, but the Lagrangian manifold[1]

$$\Lambda_\xi := \{(x, \xi) : x \in \mathbb{R}^d\}. \tag{5}$$

Note that the plane wave $e^{\frac{i}{h} x\cdot\xi}$ is a Lagrangian state associated with the Lagrangian manifold Λ_ξ.

We need to make a *transversality assumption* on Λ_ξ. This assumption roughly says that the direction ξ defining Λ_ξ is such that the incoming tail Γ^- and Λ_ξ intersect transversally. We postpone the precise statement of this assumption to Hypothesis 16 in Section 2.1.4. This assumption is probably generic in ξ, although we don't know how to prove it. In [Ingremeau 2017], we show that it is always satisfied for every ξ, when we consider geometric scattering on a manifold of nonpositive curvature.

Statement of the results. In Theorem 47, we will give a precise description of E_h^ξ as a sum of WKB states, under the assumptions above. Since the precise statement of the theorem is a bit technical, we postpone it to Section 3.5, and only state two important consequences of this result.

The first one is a bound analogous to what we would get in the nontrapping case.

Theorem 1. *Suppose that Hypothesis 10 on hyperbolicity holds, that Hypothesis 46 concerning the topological pressure is satisfied, and that $\xi \in \mathbb{S}^{d-1}$ is such that Λ_ξ satisfies Hypothesis 16 of transversality.*

Let $\chi \in C_c^\infty(X)$. Then there exists a constant $C_{\xi,\chi}$ independent of h such that for any $h > 0$, we have

$$\|\chi E_h^\xi\|_{L^2} \leq C_{\xi,\chi}. \tag{6}$$

Remark 2. The bound (6) could not be obtained directly from resolvent estimates. Indeed, as we will see in Section 3.3.2, the term E_{out} in (2) can be written as the outgoing resolvent $(P_h - (1 + i0)^2)^{-1}$

[1]By a Lagrangian manifold, we mean a d-dimensional submanifold of a $2d$-dimensional symplectic manifold, on which the symplectic form vanishes. We will allow Lagrangian manifolds to have boundaries, and to be disconnected.

applied to a term which is compactly supported, and whose L^2 norm is $O(h)$. Therefore, we have a priori that $\|\chi E_h^\xi\|_{L^2} \leq O(h)\|\chi(P_h-(1+i0)^2)^{-1}\chi\|_{L^2\to L^2}$, as least if the support of χ is large enough. But under Hypotheses 10 and 46, it is known from [Nonnenmacher and Zworski 2009] (see Theorem 45) that

$$\left\|\chi(P_h-(1+i0)^2)^{-1}\chi\right\|_{L^2\to L^2} \leq C\frac{|\log h|}{h},$$

and such estimates are sharp in the presence of trapping (see [Bony et al. 2010]). Such a priori estimates would therefore only give $\|\chi E_h^\xi\|_{L^2} \leq C|\log h|$.

Our next result concerns the *semiclassical measure* of E_h^ξ. Consider on $T^*\mathbb{R}^d$ the measure μ_0^ξ given by

$$\mathrm{d}\mu_0^\xi(x,v) = \mathrm{d}x\,\delta_{v=\xi}.$$

The measure μ_0^ξ is the semiclassical measure associated to $e^{\frac{i}{h}x\cdot\xi}$, in the sense that for any $\psi\in C_c^\infty(T^*\mathbb{R}^d)$ and any $\chi\in C_c^\infty(\mathbb{R}^d)$, we have

$$\lim_{h\to 0}\left\langle \mathrm{Op}_h(\psi)\chi e^{\frac{i}{h}x\cdot\xi}, \chi e^{\frac{i}{h}x\cdot\xi}\right\rangle = \int_{T^*\mathbb{R}^d}\chi^2(x)\psi(x,v)\,\mathrm{d}\mu_0^\xi(x,v).$$

For the definition and properties of the Weyl quantization Op_h, we refer the reader to Section 3.1.1.

We then define a measure μ^ξ on $T^*\mathbb{R}^d$ by

$$\int_{T^*\mathbb{R}^d} a\,\mathrm{d}\mu^\xi := \lim_{t\to\infty}\int_{T^*\mathbb{R}^d} a\circ\Phi^t\,\mathrm{d}\mu_0^\xi$$

for any $a\in C_c^0(T^*\mathbb{R}^d)$.

We will show in Section 6.3 that this limit exists under our above assumptions. Actually, the proof will not use Hypothesis 46 that the topological pressure of half the unstable jacobian is negative, but the much weaker assumption that the topological pressure of the unstable jacobian is negative.

The following theorem tells us that, under our hypotheses, μ^ξ is the semiclassical measure associated to E_h^ξ, and it gives us a precise description of μ^ξ close to the trapped set.

Theorem 3. *Suppose that Hypothesis 10 on hyperbolicity holds, that Hypothesis 46 concerning the topological pressure is satisfied, and that $\xi\in\mathbb{S}^{d-1}$ is such that Λ_ξ satisfies Hypothesis 16 of transversality.*

Then for any $\psi\in C_c^\infty(T^*\mathbb{R}^d)$ and any $\chi\in C_c^\infty(\mathbb{R}^d)$, we have

$$\langle \mathrm{Op}_h(\psi)\chi E_h^\xi, \chi E_h^\xi\rangle = \int_{T^*\mathbb{R}^d}\psi(x,v)\,\mathrm{d}\mu^\xi(x,v) + O(h^c).$$

Furthermore, for any $\rho\in K$, there exists a small neighbourhood $U_\rho\subset T^*\mathbb{R}^d$ of ρ, and a local change of symplectic coordinates $\kappa_\rho: U_\rho\to T^*\mathbb{R}^d$ with $\kappa_\rho(\rho)=0$ such that the following holds. There exists a constant $c>0$ and two sequences of functions $f_n,\phi_n\in C_c^\infty(\mathbb{R}^d)$ for $n\in\mathbb{N}$ such that for any $(y,\eta)\in\kappa_\rho(U_\rho)$, we have

$$\mathrm{d}\mu^\xi(\kappa_\rho^{-1}(y,\eta)) = \sum_{n=0}^\infty f_n(y)\delta_{\{\eta=\partial\phi_n(y)\}}dy,$$

and where the functions f_n *satisfy*

$$\sum_{n=0}^{\infty} \|f_n\|_{C^0} < \infty. \tag{7}$$

Remark 4. Theorem 3 tells us that the distorted plane waves E_h^ξ have a unique semiclassical measure. This result is therefore analogous to the quantum unique ergodicity conjecture for eigenfunctions of the Laplace–Beltrami operator on manifolds of negative curvature. However, on compact manifolds of negative curvature, the semiclassical measure we expect is the Liouville measure. Here, the semiclassical measure given by Theorem 3 is very different from the Liouville measure, since, close to the trapped set, it is concentrated on a countable union of Lagrangian submanifolds of T^*X. There is therefore a deep difference between compact and noncompact manifolds concerning the semiclassical measure of eigenfunctions, a fact which was already noted in [Guillarmou and Naud 2014].

Idea of proof. Theorems 1 and 3 will be deduced from a precise description of the distorted plane waves E_h^ξ microlocally near the trapped set. In Theorem 47, we will show that, microlocally near the trapped set, E_h^ξ can be written as a convergent sum of WKB states. Let us now explain how this result is obtained.

By definition, the distorted plane waves E_h^ξ are generalized eigenfunctions of the operator P_h. Therefore, if we write $U(t) = e^{-\frac{i}{h}P_h}$ for the Schrödinger propagator associated to P_h, we would like to write formally that $U(t)E_h^\xi = e^{-\frac{it}{h}}E_h^\xi$. Of course, this expression can only be formal, since $E_h^\xi \notin L^2$, but we will give it a precise meaning by truncating it by some cut-off functions.

By equation (2), E_h^ξ may be decomposed into two terms, which we will write as E_h^0 and E_h^1 in the sequel. E_h^0 is a Lagrangian state associated to the Lagrangian manifold Λ_ξ, while E_h^1 is the image of a smooth compactly supported function by the resolvent $(P_h - (1+i0)^2)^{-1}$.

Using some resolvent estimates and hyperbolic dispersion estimates, we will show in the sequel that, for any compactly supported function χ, we have $\lim_{t\to\infty} \|\chi U(t)E_h^1\| = 0$.

Therefore, in order to describe E_h^ξ, we only have to study $U(t)E_h^0$ for some very long times. Since E_h^0 is a Lagrangian state, its evolution can be described using the WKB method. To do this, we will have to understand the classical evolution of the Lagrangian manifold Λ_ξ for large times. We will show that for any $t > 0$, the restriction of $\Phi^t(\Lambda_\xi)$ to a region close to the trapped set consists of finitely many Lagrangian manifolds, most of which are very close to the "outgoing tail" of the trapped set (see Theorem 17 for more details).

Relation to other works. The study of the high frequency behaviour of eigenfunctions of Schrödinger operators, and of their semiclassical measures, in the case where the associated classical dynamics has a chaotic behaviour, has a long story. It goes back to the classical works [Shnirelman 1974; Zelditch 1987; Colin de Verdière 1985] dealing with quantum ergodicity on compact manifolds.

Analogous results on manifolds of infinite volume are much more recent. In [Dyatlov and Guillarmou 2014], the authors studied the semiclassical measures associated to distorted plane waves in a very general framework, with very mild assumptions on the classical dynamics. The counterpart of this generality is that the authors have to average on directions ξ and on an energy interval of size h to be able to define the semiclassical measure of distorted plane waves. Their result can be seen as a form of quantum ergodicity result on noncompact manifolds, although no "ergodicity" assumption is made.

In [Guillarmou and Naud 2014], the authors considered the case where $X = \Gamma\backslash\mathbb{H}^d$ is a manifold of infinite volume, with sectional curvature constant equal to -1 (convex cocompact hyperbolic manifold), and with the assumption that the Hausdorff dimension of the limit set of Γ is smaller than $(d-1)/2$. In this setting, distorted plane waves are often called *Eisenstein series*. The authors prove that there is a unique semiclassical measure for the Eisenstein series with a given incoming direction, and they give a very explicit formula for it. This result can hence be seen as a quantum unique ergodicity result in infinite volume.

Our result is a generalization of those of [Guillarmou and Naud 2014]. Indeed, we also obtain a unique semiclassical measure for the distorted plane waves with a given incoming direction. Our assumption on the topological pressure is a natural generalization of the assumption on the Hausdorff dimension of the limit set of Γ to the case of nonconstant curvature. As in [Guillarmou and Naud 2014], the main ingredient of the proof is a decomposition of the distorted plane waves as a sum of WKB states. Although our description of the distorted plane waves and of their semiclassical measure is slightly less explicit than that of [Guillarmou and Naud 2014], our methods are much more versatile, since they rely on the properties of the Hamiltonian flow close to the trapped set, instead of relying on the global quotient structure.

In [Dyatlov 2012], the author was able to obtain semiclassical convergence of distorted plane waves on manifolds of finite volume (with cusps), by working at complex energies; see also [Bonthonneau 2014] for more precise results. The main argument of [Dyatlov 2012], [Bonthonneau 2014] and [Dyatlov and Guillarmou 2014], which is to describe the distorted plane waves as plane waves propagated during a long time by the Schrödinger flow, is the starting point of our proof. However, the reason for the convergence in the long-time limit is very different in the papers above than in the present paper.

Many of the tools used in this paper were inspired by [Nonnenmacher and Zworski 2009]. We will use the notations and methods of this paper a lot.

Most of the results of the present paper can be made more precise if we suppose that we work on a manifold of nonpositive sectional curvature, without a potential. This has been studied in [Ingremeau 2017], where the author is able to show, by using the methods developed in the present paper, that distorted plane waves are bounded in L^∞_{loc} independently of h, and to give sharp bounds on the Hausdorff measure of nodal sets of the real part of distorted plane waves restricted to a compact set.

Organization of the paper. In Section 2, we will state and prove a result concerning the propagation by the Hamiltonian flow of Lagrangian manifolds similar to Λ_ξ near the trapped set, under general assumptions. In Section 3, we will state Theorem 47, which is our main theorem, giving a description of distorted plane waves as a sum of WKB states. We will deduce Theorem 1 as an easy corollary. In Section 4, we will recall various tools which were introduced in [Nonnenmacher and Zworski 2009], and which will play a role in the proof of Theorem 47. We shall then prove Theorem 47 in Section 5. Section 6 will be devoted to the proof of the Theorem 3.

The main reason why we want to state Theorem 47 for generalized eigenfunctions that are more general than distorted plane waves on $\mathbb{R}^d$ is that our results do also apply if the manifold is hyperbolic near infinity (which allows us to recover some of the results of [Guillarmou and Naud 2014]), as is shown in [Ingremeau 2017, Appendix B]. Our results do probably also apply if the manifold is asymptotically hyperbolic; this shall be pursued elsewhere.

2. Propagation of Lagrangian manifolds

2.1. *General assumptions for propagation of Lagrangian manifolds.* Let (X, g) be a noncompact complete Riemannian manifold of dimension d, and let $V : X \to \mathbb{R}$ be a smooth compactly supported potential.

We denote by $p(x, \xi) = p(\rho) : T^*X \to \mathbb{R}$, $p(x, \xi) = \|\xi\|^2 + V(x)$, the classical Hamiltonian.

For each $t \in \mathbb{R}$, we denote by $\Phi^t : T^*X \to T^*X$ the Hamiltonian flow at time t for the Hamiltonian p.

Given any smooth function $f : X \to \mathbb{R}$, it may be lifted to a function $f : T^*X \to \mathbb{R}$, which we denote by the same letter. We may then define $\dot{f}, \ddot{f} \in C^\infty(T^*X)$ to be the derivatives of f with respect to the Hamiltonian flow:

$$\dot{f}(x, \xi) := \frac{df(\Phi^t(x, \xi))}{dt}\bigg|_{t=0}, \quad \ddot{f}(x, \xi) := \frac{d^2 f(\Phi^t(x, \xi))}{dt^2}\bigg|_{t=0}.$$

2.1.1. *Hypotheses near infinity.* We suppose the following conditions are fulfilled.

Hypothesis 5 (structure of X near infinity). *We suppose the manifold (X, g) is such that the following holds*:

(1) *There exists a compactification $\overline{X}$ of X, that is, a compact manifold with boundaries $\overline{X}$ such that X is diffeomorphic to the interior of $\overline{X}$. The boundary $\partial\overline{X}$ is called the boundary at infinity.*

(2) *There exists a boundary-defining function b on X, that is, a smooth function $b : \overline{X} \to [0, \infty)$ such that $b > 0$ on X, and b vanishes to first order on $\partial\overline{X}$.*

(3) *There exists a constant $\epsilon_0 > 0$ such that for any point $(x, \xi) \in \mathcal{E}$,*

$$\textit{if } b(x, \xi) \le \epsilon_0 \textit{ and } \dot{b}(x, \xi) = 0 \textit{ then } \ddot{b}(x, \xi) < 0.$$

Note that, although part (3) of the hypothesis makes reference to the Hamiltonian flow, it is only an assumption on the manifold (X, g) and not on the potential V, because V is assumed to be compactly supported.

Example 6. $\mathbb{R}^d$ fulfills Hypothesis 5 by taking the boundary-defining function $b(x) = (1 + |x|^2)^{-1/2}$. We then have $\overline{X} \equiv B(0, 1)$.

Example 7. The Poincaré space $\mathbb{H}^d$ also fulfills Hypothesis 5. Indeed, in the ball model $B_0(1) = \{x \in \mathbb{R}^d : |x| < 1\}$, where $|\cdot|$ denotes the Euclidean norm, $\mathbb{H}^d$ compactifies to the closed unit ball, and the boundary-defining function $b(x) = 2(1 - |x|)/(1 + |x|)$ fulfills conditions (2) and (3).

We will write $X_0 := \{x \in X : b(x) \ge \epsilon_0/2\}$. By possibly taking ϵ_0 smaller, we can assume that $\mathrm{supp}(V) \subset \{x \in X; b(x) > \epsilon_0\}$. We will call X_0 the *interaction region*. We will also write

$$W_0 := T^*(X \backslash X_0) = \{\rho \in T^*X : b(\rho) < \epsilon_0/2\}, \quad \mathcal{W}_0 = W_0 \cap \mathcal{E}. \tag{8}$$

By possibly taking ϵ_0 even smaller, we may ask that

$$\forall \rho \in \mathcal{W}_0, \quad b(\Phi^1(\rho)) < \epsilon_0. \tag{9}$$

Definition 8. If $\rho = (x, \xi) \in \mathcal{E}$, we say that ρ escapes directly in the forward direction, denoted $\rho \in \mathcal{DE}_+$, if $b(x) < \epsilon_0/2$ and $\dot{b}(x, \xi) \leq 0$.

If $\rho = (x, \xi) \in \mathcal{E}$, we say that ρ escapes directly in the backward direction, denoted $\rho \in \mathcal{DE}_-$, if $b(x) < \epsilon_0/2$ and $\dot{b}(x, \xi) \geq 0$.

Note that we have

$$\mathcal{W}_0 = \mathcal{DE}_- \cup \mathcal{DE}_+.$$

Part (3) of Hypothesis 5 implies the following *geodesic convexity* result, which reflects the fact that once a trajectory has left the interaction region, it cannot come back to it.

Lemma 9. *For any* $t \geq 0$, *we have*

$$\Phi^t(\mathcal{E} \cap T^*X_0) \cap \mathcal{DE}_- = \varnothing.$$

Proof. Suppose that there exists a $\rho \in \Phi^t(\mathcal{E} \cap T^*X_0) \cap \mathcal{DE}_-$ for some $t \geq 0$. Then there exists $\rho' \in \mathcal{E} \cap T^*X_0$ such that $\rho = \Phi^t(\rho')$. Let us consider $f(s) := b(\Phi^s(\rho'))$. We have $f(0) > \epsilon_0/2$, $f(t) < \epsilon_0/2$ and $f'(t) \geq 0$ by hypothesis. This is impossible, because by Hypothesis 5, point (3), whenever $f(s) \leq \epsilon_0$ and $f'(s) = 0$, we have $f''(s) < 0$. □

2.1.2. *Hyperbolicity.* Recall that the *trapped set* was defined in (4). In the sequel, we will always suppose that the trapped set is a *hyperbolic set*, as follows.

Hypothesis 10 (hyperbolicity of the trapped set). *We assume that* K *is a hyperbolic set for the flow* $\Phi^t_{|\mathcal{E}}$. *That is to say, there exists a metric* g_{ad} *on a neighbourhood of* K *included in* $\mathcal{E}$, *and* $\lambda > 0$, *such that the following holds. For each* $\rho \in K$, *there is a decomposition*

$$T_\rho \mathcal{E} = \mathbb{R} H_p(\rho) \oplus E^+_\rho \oplus E^-_\rho$$

such that

$$\|d\Phi^t_\rho(v)\|_{g_{\mathrm{ad}}} \leq e^{-\lambda|t|} \|v\|_{g_{\mathrm{ad}}} \quad \textit{for all } v \in E^{\mp}_\rho,\ \pm t \geq 0.$$

We will call $E^\pm$ the *unstable* (resp. *stable*) subspaces at the point ρ.

We may extend g_{ad} to a metric on the whole energy layer, so that outside of the interaction region, it coincides with the metric on T^*X induced from the Riemannian metric on X. From now on, d will denote the Riemannian distance associated to this metric on $\mathcal{E}$.

Let us recall a few properties of hyperbolic dynamics (see [Katok and Hasselblatt 1995, Chapter 6] for the proofs of the statements).

(i) The hyperbolic set is *structurally stable*, in the following sense. For $\boldsymbol{E} > 0$, define the layer of energy $\boldsymbol{E}$ as

$$\mathcal{E}_{\boldsymbol{E}} := \{\rho \in T^*X : p(\rho) = \boldsymbol{E}\}, \tag{10}$$

and the trapped set at energy $\boldsymbol{E}$ as

$$K_{\boldsymbol{E}} := \{\rho \in \mathcal{E}_{\boldsymbol{E}} : \Phi^t(\rho) \text{ remains in a compact set for all } t \in \mathbb{R}\}. \tag{11}$$

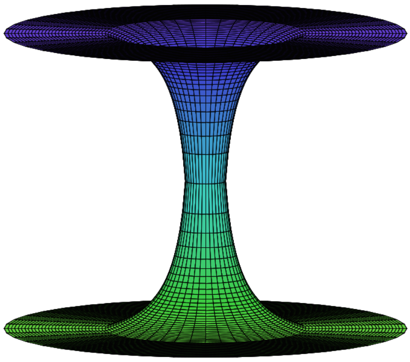


Figure 2. A surface which has negative curvature close to the trapped set of the geodesic flow, and which is isometric to two copies of $\mathbb{R}^2 \backslash B(0, R_0)$ outside of a compact set. It satisfies Hypothesis 10 near the trapped set and Hypothesis 5 at infinity.

If K is a hyperbolic set for $\Phi^t_{|\boldsymbol{E}}$, then

$$\exists \delta > 0,\ \forall \boldsymbol{E} \in (1-\delta, 1+\delta), \quad K_{\boldsymbol{E}} \text{ is a hyperbolic set for } \Phi^t_{|\mathcal{E}_{\boldsymbol{E}}}. \tag{12}$$

(ii) $d\Phi^t_\rho(E^\pm_\rho) = E^\pm_{\Phi^t(\rho)}$.

(iii) $K \ni \rho \mapsto E^\pm_\rho \subset T_\rho(\mathcal{E})$ is Hölder-continuous.

(iv) Any $\rho \in K$ admits local strongly (un)stable manifolds $W^\pm_{\text{loc}}(\rho)$ tangent to $E^\pm_\rho$, defined by

$$W^\pm_{\text{loc}}(\rho) = \big\{\rho' \in \mathcal{E} : d(\Phi^t(\rho), \Phi^t(\rho')) < \epsilon \text{ for all } \pm t \leq 0 \text{ and } \lim_{t \to \mp\infty} d(\Phi^t(\rho'), \Phi^t(\rho)) = 0\big\},$$

where $\epsilon > 0$ is some small number.

We call

$$E^{+0}_\rho := E^+_\rho \oplus \mathbb{R} H_p(\rho), \quad E^{-0}_\rho := E^-_\rho \oplus \mathbb{R} H_p(\rho)$$

the *weak unstable* and *weak stable* subspaces at the point ρ respectively.

2.1.3. *Adapted coordinates.* Let us now describe the construction of a local system of coordinates which is adapted to the stable and unstable directions near a point. In the sequel, these coordinates will be considered as fixed, and used to state Theorem 17.

Lemma 11. *Let $\rho \in K$. There exists an adapted system of symplectic coordinates (y^ρ, η^ρ) on a neighbourhood of ρ in T^*X such that the following holds*:

(i) $\rho \equiv (0, 0)$.

(ii) $E^+_\rho = \text{span}\{(\partial/\partial y^\rho_i)(\rho) : i = 2, \ldots, d\}$.

(iii) $E^-_\rho = \text{span}\{(\partial/\partial \eta^\rho_i)(\rho) : i = 2, \ldots, d\}$.

(iv) $\eta^\rho_1 = p - 1$ *is the energy coordinate.*

(v) $\big\langle (\partial/\partial y^\rho_i)(\rho), (\partial/\partial y^\rho_j)(\rho) \big\rangle_{g_{\text{ad}}(\rho)} = \delta_{i,j},\ i, j = 2, \ldots, d.$

Proof. We may identify a neighbourhood of $\rho \in T^*X$ with a neighbourhood of $(0,0) \in T^*\mathbb{R}^d$. Let us take $e_1^\rho = H_p(\rho)$, and complete it into a basis $(e_1^\rho, \dots, e_d^\rho)$ of E_ρ^{+0} such that $\langle e_i^\rho, e_j^\rho \rangle_{g_{\mathrm{ad}}(\rho)} = 1$ for $2 \le i, j \le d$.

Since $E^{\pm 0}$ are Lagrangian subspaces (which follows from the hyperbolicity assumption), it is then possible to find vectors $(f_1^\rho, \dots, f_d^\rho)$ such that $E_\rho^- = \operatorname{span}\{f_2^\rho, \dots, f_d^\rho\}$ and such that $\omega(f_j^\rho, e_k^\rho) = \delta_{j,k}$ for any $1 \le j, k \le d$. In particular, we have $\omega(f_1^\rho, e_1^\rho) = dp(f_1) = 1$.

From Darboux's theorem, there exists a nonlinear symplectic chart $(y^{\mathrm{flat}}, \eta^{\mathrm{flat}})$ near the origin such that $\eta_1^{\mathrm{flat}} = p - 1$. There also exists a linear symplectic transformation A such that the coordinates $(y, \eta) = A(y^{\mathrm{flat}}, \eta^{\mathrm{flat}})$ satisfy $\eta_1 = \eta_1^{\mathrm{flat}}$ as well as

$$\eta_1 = p - 1, \quad \frac{\partial}{\partial y_j}(0,0) = e_j \text{ and } \frac{\partial}{\partial \eta_j}(0,0) = f_j, \quad j = 1, \dots, d. \qquad \square$$

We will often write

$$\boldsymbol{y}^\rho := (y_2^\rho, \dots, y_d^\rho) \quad \text{and } \boldsymbol{\eta}^\rho := (\eta_2^\rho, \dots, \eta_d^\rho). \tag{13}$$

For any $\epsilon > 0$, write $D_\epsilon = \{u \in \mathbb{R}^{d-1} : |u| < \epsilon\}$. We define the following polydisk centred at ρ:

$$U^\rho(\epsilon) \equiv \big\{(y^\rho, \eta^\rho) : |y_1^\rho| < \epsilon, |\eta_1^\rho| < \delta,\ \boldsymbol{y}^\rho \in D_\epsilon,\ \boldsymbol{\eta} \in D_\epsilon\big\}, \tag{14}$$

where δ comes from (12).

We also define *unstable Lagrangian manifolds*, which are needed in the statement of Theorem 17.

Definition 12. Let $\Lambda \subset \mathcal{E}$ be an isoenergetic Lagrangian manifold (not necessarily connected) included in a small neighbourhood W of a point $\rho \in K$, and let $\gamma > 0$. We will say that Λ is a *γ-unstable Lagrangian manifold* (or that Λ is in the γ-unstable cone) in the coordinates (y^ρ, η^ρ) if it can be written in the form

$$\Lambda = \{(y^\rho; 0, F(y^\rho)) : y^\rho \in D\},$$

where $D \subset \mathbb{R}^d$ is an open subset with finitely many connected components, and with piecewise smooth boundary, and $F : \mathbb{R}^d \to \mathbb{R}^{d-1}$ is a smooth function with $\|dF\|_{C^0} \le \gamma$.

Note that, since F is defined on $\mathbb{R}^d$, a γ-unstable manifold may always be seen as a submanifold of a *connected γ-unstable Lagrangian manifold.*

Let us also note that, since Λ is isoenergetic and is Lagrangian, an immediate computation shows that F does not depend on y_1^ρ, so that Λ can actually be put in the form

$$\Lambda = \{(y^\rho; 0, f(\boldsymbol{y}^\rho)) : y^\rho \in D\},$$

where $f : \mathbb{R}^{d-1} \to \mathbb{R}^{d-1}$ is a smooth function with $\|df\|_{C^0} \le \gamma$.

2.1.4. *Hypotheses on the incoming Lagrangian manifold.* Let us consider an isoenergetic Lagrangian manifold $\mathcal{L}_0 \subset \mathcal{E}$ of the form

$$\mathcal{L}_0 := \{(x, \phi(x)) : x \in X_1\},$$

where X_1 is a closed subset of $X \backslash X_0$ with finitely many connected components and piecewise smooth boundary, and $\phi : x \to \phi(x),\ X_2 \mapsto T_x^*X$, is a smooth covector field defined on some neighbourhood X_2 of X_1.

We make the following additional hypothesis on $\mathcal{L}_0$:

Hypothesis 13 (invariance hypothesis). *We suppose that $\mathcal{L}_0$ satisfies the following invariance hypotheses*:

$$\forall t \geq 0, \quad \Phi^t(\mathcal{L}_0) \cap \mathcal{DE}_- = \mathcal{L}_0 \cap \mathcal{DE}_-. \tag{15}$$

Example 14. Given a $\xi \in \mathbb{R}^d$ with $|\xi|^2 = 1$, the Lagrangian manifold Λ_ξ defined in the Introduction fulfills Hypothesis 13.

Example 15. Suppose that $(X \backslash X_0, g) \cong (\mathbb{R}^d \backslash B(0, R), g_{\mathrm{Eucl}})$ for some $R > 0$. Then the incoming spherical Lagrangian, defined by

$$\Lambda_{\mathrm{sph}} := \left\{ \left(x, -\frac{x}{|x|} \right) : |x| > R \right\},$$

fulfills Hypothesis 13.

We also make the following transversality assumption on the Lagrangian manifold $\mathcal{L}_0$. It roughly says that $\mathcal{L}_0$ intersects the stable manifold transversally.

Hypothesis 16 (transversality hypothesis). *We suppose that $\mathcal{L}_0$ is such that, for any $\rho \in K$, for any $\rho' \in \mathcal{L}_0$, for any $t \geq 0$, we have*

$$\Phi^t(\rho') \in W^-_{\mathrm{loc}}(\rho) \implies W^-_{\mathrm{loc}}(\rho) \text{ and } \Phi^t(\mathcal{L}_0) \text{ intersect transversally at } \Phi^t(\rho'),$$

that is to say

$$T_{\Phi^t(\rho')}\mathcal{L}_0 \oplus T_{\Phi^t(\rho')} W^-_{\mathrm{loc}}(\rho) = T_{\Phi^t(\rho')}\mathcal{E}. \tag{16}$$

Note that (16) is equivalent to $T_{\Phi^t(\rho')}\mathcal{L}_0 \cap T_{\Phi^t(\rho')} W^-_{\mathrm{loc}}(\rho) = \{0\}$.

On $X = \mathbb{R}^d$, Hypothesis 16 is likely to hold for almost every $\xi \in \mathbb{S}^{d-1}$, at least for a generic V. In [Ingremeau 2017], the author shows that this hypothesis is satisfied for every ξ on manifolds of nonpositive curvature which have several Euclidean ends (like the one in Figure 2), when there is no potential.

2.2. *Statement of the result.* Let us now state the main result of this section, which describes the "truncated evolution" of Lagrangian manifolds.

Truncated Lagrangians. Let $(W_a)_{a\in A}$ be a finite family of open sets in T^*X. Let $N \in \mathbb{N}$, and let $\alpha = \alpha_0, \alpha_1 \cdots \alpha_{N-1} \in A^N$. Let Λ be a Lagrangian manifold in T^*X. We define the sequence of (possibly empty) Lagrangian manifolds $(\Phi^k_\alpha(\Lambda))_{0\leq k\leq N}$ by recurrence by

$$\Phi^0_\alpha(\Lambda) = \Lambda \cap W_{\alpha_0}, \quad \Phi^{k+1}_\alpha(\Lambda) = W_{\alpha_{k+1}} \cap \Phi^1(\Phi^k_\alpha(\Lambda)).$$

In the sequel, we will consider families with indices in $A = A_1 \sqcup A_2 \sqcup \{0\}$. For any $\alpha \in A^N$ such that $\alpha_{N-1} \neq 0$, we will define

$$\tau(\alpha) := \max\{1 \leq i \leq N-1 : \alpha_i = 0\} \tag{17}$$

if there exists $1 \leq i \leq N-1$ with $\alpha_i = 0$, and $\tau(\alpha) = 0$ otherwise.

Theorem 17. *Suppose that the manifold X satisfies Hypothesis 5 at infinity, that the Hamiltonian flow (Φ^t) satisfies Hypothesis 10, and that the Lagrangian manifold $\mathcal{L}_0$ satisfies Hypothesis 13 on invariance as well as Hypothesis 16 of transversality.*

Fix $\gamma_{\text{uns}} > 0$ small enough. There exists $\varepsilon_0 > 0$ such that the following holds. Let $(W_a)_{a\in A_1}$ be any open cover of K in T^*X of diameter $< \varepsilon_0$ such that there exist points $\rho_a \in W_a \cap K$, and such that the adapted coordinates (y^a, η^a) centred on ρ^a are well defined on W_a for every $a \in A_1$. Then we may complete this cover into $(W_a)_{a\in A}$ an open cover of $\mathcal{E}$ in T^*X where $A = A_1 \sqcup A_2 \sqcup \{0\}$ (with W_0 defined as in (8)) such that the following holds.

There exists $N_{\text{uns}} \in \mathbb{N}$ such that for all $N \in \mathbb{N}$, for all $\alpha \in A^N$ and all $a \in A_1$, we have $W_a \cap \Phi^N_\alpha(\mathcal{L}_0)$ is either empty, or is a Lagrangian manifold in some unstable cone in the coordinates (y^a, η^a).

Furthermore, if $N - \tau(\alpha) \geq N_{\text{uns}}$, then $W_a \cap \Phi^N_\alpha(\mathcal{L}_0)$ is a γ_{uns}-unstable Lagrangian manifold in the coordinates (y^a, η^a).

Remark 18. For a sequence $\alpha \in A^N$, $N - \tau(\alpha)$ corresponds to the time spent in the interaction region. Our last statement therefore says that if a part of $\mathcal{L}_0$ stays in the interaction region for long enough when propagated, then its tangents will form a small angle with the unstable direction at ρ^a.

Remark 19. The constant ε_0 and the sets $(W_a)_{a\in A_2}$ depend on the Lagrangian manifold $\mathcal{L}_0$. If we take a whole family of Lagrangian manifolds $(\mathcal{L}_z)_{z\in Z}$ satisfying Hypotheses 13 and 16, then we will need some additional conditions on the whole family to be able to find a common choice of ε_0 and $(W_a)_{a\in A_2}$ independent of $z \in Z$. An example of such a condition will be provided by equations (36) and (37). Note that these equations are automatically satisfied if Z is finite.

2.3. *Proof of Theorem 17.*

Proof. From now on, we will fix a $\gamma_{\text{uns}} > 0$.

Let $\rho_0 \in K$, and consider the system of adapted coordinates in a neighbourhood of ρ_0 constructed in Section 2.1.3. Recall that the set $U^{\rho_0}(\epsilon)$ was defined in (14). We define a *Poincaré section* by

$$\Sigma^{\rho_0} = \Sigma^{\rho_0}(\epsilon) = \{(y^{\rho_0}, \eta^{\rho_0}) \in U^{\rho_0}(\epsilon) : y_1^{\rho_0} = \eta_1^{\rho_0} = 0\}.$$

Note that the spaces $E^{\pm}_{\rho_0}$ are tangent to Σ^{ρ_0}, and that the coordinates $(\boldsymbol{y}^{\rho_0}, \boldsymbol{\eta}^{\rho_0})$ introduced in (13) form a symplectic chart on Σ^{ρ_0}.

Actually, we will often need a nonsymplectic system of coordinates built from the coordinates (y^ρ, η^ρ).

Before building this nonsymplectic system of coordinates, let us explain why it is a crucial ingredient of our argument. The main tool in the proof of Theorem 17 is the so-called "inclination lemma", which roughly says that a Lagrangian manifold which intersects the stable manifold transversally will get more and more unstable when propagated in the future. This is a very easy result in the case of linear hyperbolic diffeomorphisms, but we must add some quantifiers in the case of nonlinear dynamics to make it rigorous. Namely, one can say, as in [Nonnenmacher and Zworski 2009, Proposition 5.1], that given a $\gamma > 0$, there exists $\epsilon_\gamma > 0$ such that if Λ is a γ-unstable Lagrangian manifold included in some $U^\rho(\epsilon_\gamma)$, then for any ρ', $\Phi^1(\Lambda) \cap U^{\rho'}(\epsilon_\gamma)$ is still γ-unstable.

However, we may not use this result directly for the following reason. The smaller we take ϵ, the longer the points of the Lagrangian manifold $\mathcal{L}_0$ may spend in the part of the interaction region which is not affected by the hyperbolic dynamics before entering in some $U^\rho(\epsilon)$ for some $\rho \in K$. Yet the longer they spend in this "intermediate" region, the more stable the Lagrangian manifold may a priori become. To avoid such a circular reasoning, we should introduce another system of coordinates, in which the description of the propagation of the Lagrangian manifolds in the intermediate region is easier.

2.3.1. *Alternative coordinates.* In this section we will describe a system of "alternative", or "twisted" coordinates built from the one we introduced in Section 2.1.3, but which may differ slightly from them.

Given a $\rho \in K$, we introduce a system of smooth coordinates $(\tilde{y}^\rho, \tilde{\eta}^\rho)$ as follows.

On Σ^ρ, these coordinates are such that

$$W^{0+}_{\mathrm{loc}}(\rho) \cap \Sigma^\rho \equiv \{(\tilde{\boldsymbol{y}}^\rho, 0) : \tilde{\boldsymbol{y}}^\rho \in D_\epsilon\}, \quad W^{0-}_{\mathrm{loc}}(\rho) \cap \Sigma^\rho \equiv \{(0, \tilde{\boldsymbol{\eta}}^\rho) : \tilde{\boldsymbol{\eta}}^\rho \in D_\epsilon\},$$

and if we denote by L_ρ the map

$$L_\rho : (\boldsymbol{y}^\rho, \boldsymbol{\eta}^\rho) \mapsto (\tilde{\boldsymbol{y}}^\rho, \tilde{\boldsymbol{\eta}}^\rho) \tag{18}$$

defined in a neighbourhood of $(0, 0)$, we have

$$dL_\rho(0, 0) = \mathrm{Id}_{\mathbb{R}^{2d-2}}. \tag{19}$$

Now, if $\hat{\rho}$ has straight coordinates $(y^\rho(\hat{\rho}), \eta^\rho(\hat{\rho}))$, we let $\hat{\rho}' \in \Sigma^\rho$ be the point with straight coordinates $(0, \boldsymbol{y}^\rho(\hat{\rho}), 0, \boldsymbol{\eta}^\rho(\hat{\rho}))$. We then define the twisted coordinates of $\hat{\rho}$ by

$$\tilde{y}^\rho_1(\hat{\rho}) = y^\rho_1(\hat{\rho}), \quad \tilde{\eta}^\rho_1(\hat{\rho}) = \eta^\rho_1(\hat{\rho}), \quad \tilde{\boldsymbol{y}}^\rho(\hat{\rho}) = \tilde{\boldsymbol{y}}^\rho(\hat{\rho}'), \quad \tilde{\boldsymbol{\eta}}^\rho(\hat{\rho}) = \tilde{\boldsymbol{\eta}}^\rho(\hat{\rho}').$$

Note that this system of coordinates doesn't have to be symplectic.

We have

$$\frac{\partial \boldsymbol{y}^\rho_j}{\partial \tilde{y}^\rho_1} = \frac{\partial \boldsymbol{\eta}^\rho_j}{\partial \tilde{y}^\rho_1} = 0 \quad \text{for } j = 1, \dots, d-1, \qquad \frac{\partial y^\rho_1}{\partial \tilde{y}^\rho_1} = 1. \tag{20}$$

Given a $\rho \in K$, and $\epsilon, \epsilon' > 0$, we define

$$\widetilde{U}^\rho(\epsilon, \epsilon') \equiv \big\{(\tilde{y}^\rho, \tilde{\eta}^\rho) : |\tilde{y}^\rho_1| < \epsilon,\ |\tilde{\eta}^\rho_1| < \delta,\ \tilde{\boldsymbol{y}}^\rho \in D_{\epsilon'},\ \tilde{\boldsymbol{\eta}}^\rho \in D_\epsilon\big\}, \tag{21}$$

where δ is an energy interval on which the dynamics remains uniformly hyperbolic.

Finally, the Poincaré section in the alternative coordinates is represented as

$$\widetilde{\Sigma}^\rho(\epsilon, \epsilon') := \big\{(\tilde{y}^\rho, \tilde{\eta}^\rho) \in \widetilde{U}^\rho(\epsilon, \epsilon') : \tilde{y}^\rho_1 = \tilde{\eta}^\rho_1 = 0\big\}.$$

In the sequel, we will be working most of the time in a situation where $\epsilon' \ll \epsilon$ (that is, with sets much thinner in the unstable direction than in the stable direction).

The main reason why we needed to introduce alternative coordinates is that they give a simpler expression for the *Poincaré map* (see Remark 20). Let us now define this map.

2.3.2. *The Poincaré map.* Let $\rho_0 \in K$, and let $\epsilon > 0$ be small enough so that the twisted coordinates around ρ_0 and $\Phi^1(\rho_0)$ are well defined in some neighbourhoods $\widetilde{U}^{\rho_0}(\epsilon,\epsilon)$ and $\widetilde{U}^{\Phi^1(\rho_0)}(\epsilon,\epsilon)$. The *Poincaré map* κ_{ρ_0} is defined, for $\rho \in \widetilde{\Sigma}^{\rho_0}(\epsilon)$ near ρ_0, by taking the intersection of the trajectory $(\Phi^s(\rho))_{|s-1|\leq\epsilon}$ with the section $\widetilde{\Sigma}^{\Phi^1(\rho_0)}$ (this intersection consists of at most one point). In the sequel, we will sometimes omit the reference to ρ_0 and simply write the Poincaré map κ.

The map κ_{ρ_0} need not be symplectic, since it is defined in the twisted coordinates which need not be symplectic. However, if we had defined the Poincaré map in the straight coordinates, it would have been automatically symplectic. The linearisations of the two systems of coordinates are identical at ρ_0 by (19). Therefore, by using the hyperbolicity assumption, we see that the differential of κ at ρ_0 takes the form

$$d\kappa(\rho_0) \equiv \begin{pmatrix} A & 0 \\ 0 & {}^tA^{-1} \end{pmatrix},$$

and there exists

$$\nu = e^{-\lambda} < 1 \tag{22}$$

such that the matrix A satisfies

$$\|A^{-1}\| \leq \nu, \tag{23}$$

where $\|\cdot\|$ corresponds to the matrix norm. Hence, the Poincaré map κ_{ρ_0} takes the form

$$\kappa_{\rho_0}(\tilde{\boldsymbol{y}}^{\rho_0},\tilde{\boldsymbol{\eta}}^{\rho_0}) = \big(A\tilde{\boldsymbol{y}}^{\rho_0} + \tilde{\alpha}(\tilde{\boldsymbol{y}}^{\rho_0},\tilde{\boldsymbol{\eta}}^{\rho_0}), {}^tA^{-1}\tilde{\boldsymbol{\eta}}^{\rho_0} + \tilde{\beta}(\tilde{\boldsymbol{y}}^{\rho_0},\tilde{\boldsymbol{\eta}}^{\rho_0})\big), \tag{24}$$

and the functions $\tilde{\alpha}$ and $\tilde{\beta}$ satisfy

$$\tilde{\alpha}(0,\tilde{\boldsymbol{\eta}}^{\rho_0}) = \tilde{\beta}(\tilde{\boldsymbol{y}}^{\rho_0},0) \equiv 0 \quad \text{and} \quad d\tilde{\alpha}(0,0) = d\tilde{\beta}(0,0) = 0. \tag{25}$$

We therefore have

$$\|\tilde{\alpha}\|_{C^1(V)} \leq C_0\epsilon, \quad \|\tilde{\beta}\|_{C^1(V)} \leq C_0\epsilon \tag{26}$$

for some constant C_0, since κ is uniformly C^2.

Remark 20. Equation (25) is the main reason why we needed to introduce alternative coordinates, and will play a key role in the proof of Lemma 31. If we had defined the Poincaré map in the straight coordinates, we wouldn't have had $\alpha(0,\boldsymbol{\eta}^{\rho_0}) = 0$ or $\beta(\boldsymbol{y}^{\rho_0},0) = 0$.

Remark 21. By compactness of the trapped set, the constants C_0 and ν may be chosen independent of the point ρ_0. We may also find a $\mathcal{C} > 1$ such that, independently of ρ_0 and ρ_1 in K, we have

$$\|A\| \leq \mathcal{C}. \tag{27}$$

Finally, by possibly taking C_0 larger, we may assume that all the second derivatives of the map L_ρ defined in (18) are bounded by C_0 independently on $\rho \in K$.

2.3.3. *Changes of coordinates and Lagrangian manifolds.* Let us describe how a Lagrangian manifold is affected when we go from twisted coordinates to straight coordinates centred at the same point.

Lemma 22. *Suppose that a Lagrangian manifold $\Lambda \subset \widetilde{U}^\rho(\epsilon,\epsilon)$ may be written in the twisted coordinates centred on $\rho \in K$ as $\Lambda = \{(\tilde{y}_1^\rho, \tilde{y}^\rho; 0, \widetilde{F}(\tilde{y}^\rho)) : \tilde{y}^\rho \in_\rho\}$, where $_\rho \subset \mathbb{R}^d$ is a small open set, and with $\|d\widetilde{F}\|_{C^0} \leq \gamma$. Suppose furthermore that*

$$C_0\epsilon\gamma < 1.$$

Then, in the straight coordinates, Λ *may be written as*

$$\Lambda = \{(y_1^\rho, \boldsymbol{y}^\rho; 0, f(\boldsymbol{y}^\rho)) : \boldsymbol{y}^\rho \in D_\rho\},$$

with $\|df\|_{C^0} \leq \gamma(1 - C_0\gamma\epsilon)^{-1}(1 + 2C_0\epsilon)$.

Proof. To lighten the notations, we will not write the indices ρ.

Points on Λ are parametrized by the coordinate $\tilde{y}$. We may hence see their straight coordinates u, s as functions of $\tilde{y}$.

By equations (19), (20) and Remark 21, we have

$$\frac{\partial y}{\partial \tilde{y}} = \frac{\partial y}{\partial \tilde{y}} + \frac{\partial y}{\partial \tilde{\eta}} \frac{\partial \widetilde{F}(\tilde{y})}{\partial \tilde{y}} = I + R$$

with $\|R\| \leq C_0\gamma\epsilon < 1$.

Therefore, on Λ, we know $\tilde{y} \mapsto y$ is invertible. We may hence write $\boldsymbol{\eta}$ as a function of y, and we have

$$\frac{\partial \boldsymbol{\eta}}{\partial y} = \frac{\partial \tilde{y}}{\partial y}\left[\frac{\partial \boldsymbol{\eta}}{\partial \tilde{y}} + \frac{\widetilde{F}(\tilde{y})}{\partial \tilde{y}} \frac{\partial \boldsymbol{\eta}}{\partial \tilde{\boldsymbol{\eta}}}\right] = (I + R)^{-1}(\gamma(I + R')),$$

with $\|R'\| \leq 2C_0\epsilon$. Hence $\|\partial\boldsymbol{\eta}/\partial y\| \leq \gamma(1 - C_0\gamma\epsilon)^{-1}(1 + 2C_0\epsilon)$.

That $\boldsymbol{\eta}$ is actually independent of y_1 comes from the fact that Λ is an isoenergetic Lagrangian manifold, and that we are working in symplectic coordinates. □

Let us now describe the change between two systems of twisted coordinates. Let $\rho, \rho' \in K$. If they are close enough to each other, the map $L : (\tilde{y}^\rho, \tilde{\eta}^\rho) \mapsto (\tilde{y}^{\rho'}, \tilde{\eta}^{\rho'})$ is well defined on a set containing both ρ and ρ', of diameter $d(\rho, \rho')$.

Combining the fact that the (un)stable subspaces $E_\rho^\pm$ are Hölder continuous with respect to $\rho \in K^\delta$ with some Hölder exponent $\mathsf{p} > 0$, and point (v) of Lemma 11, we get

$$dL_{(0,0)} = \mathsf{L} + R_{\rho,\rho'}, \tag{28}$$

where

$$\|R_{\rho,\rho'}\| \leq C d^{\mathsf{p}}(\rho, \rho') \quad \text{for some } \mathsf{p} > 0, \tag{29}$$

and where L is of the form

$$\mathsf{L} = \begin{pmatrix} U_y & 0 \\ 0 & \mathsf{L}_\eta \end{pmatrix}$$

for some unitary matrix U_y. Here, L_η might not be unitary, but it is invertible, and by compactness of K, $\|\mathsf{L}_\eta\|^{-1}$ may be bounded independently on ρ.

Now, by compactness, the second derivatives of L may be bounded independently of ρ and ρ'. Therefore, for any ρ'' in a neighbourhood of ρ, we have

$$dL_{\rho''} = dL_{(0,0)} + R_{\rho''}, \tag{30}$$

with $R_{\rho''} \leq C' d(\rho, \rho'')$ and C' independent of ρ'.

By possibly enlarging C_0, we may assume that $\|\mathsf{L}_\eta\|^{-1} \leq C_0$. We may also assume that $C_0/2$ is larger than the constants C and C' appearing in the bounds on $R_{\rho,\rho'}$ and $R_{\rho''}$.

We will use the previous remarks in the form of the following lemma, which describes the effect of a change of twisted coordinates on a Lagrangian manifold.

Lemma 23. *Let $\rho, \rho' \in K$ be such that $d(\rho, \rho') < \epsilon$, and let Λ be a Lagrangian manifold which may be written in the twisted coordinates centred on ρ as $\Lambda = \{(\tilde{y}_1^\rho, \boldsymbol{\tilde{y}}^\rho; 0, \widetilde{F}^\rho(\tilde{y}^\rho)) : \tilde{y}^\rho \in \varrho\}$, where $\varrho \subset \mathbb{R}^d$ is a small open set, and with $\|d\widetilde{F}^\rho\|_{C^0} \leq \gamma < 1/(4C_0\epsilon^{\mathsf{p}})$.*

Then, $\Lambda \cap \widetilde{U}^{\rho'}(\epsilon, \epsilon)$ may be written in the coordinates centred at ρ' as

$$\Lambda \cap \widetilde{U}^{\rho'}(\epsilon, \epsilon) = \{(\tilde{y}_1^{\rho'}, \boldsymbol{\tilde{y}}^{\rho'}; 0, \widetilde{F}^{\rho'}(\tilde{y}^{\rho'})) : \tilde{y}^{\rho'} \in \varrho'\},$$

where $\varrho' \subset \mathbb{R}^d$ is a small open set, and with

$$\|d\widetilde{F}^{\rho'}\|_{C^0} \leq (\gamma(1 + C_0\epsilon^{\mathsf{p}}) + C_0\epsilon^{\mathsf{p}})(1 - 2\gamma C_0\epsilon^{\mathsf{p}})^{-1} < \infty.$$

Proof. Consider points on Λ. By assumption, their $\tilde{\eta}^\rho$-coordinate is a function of their $\tilde{y}^\rho$-coordinate. Therefore, using the map L, their coordinates $(\tilde{y}^{\rho'}, \tilde{\eta}^{\rho'})$ may be seen as functions of $\tilde{y}^\rho$.

Let us denote by L_y and L_η the two components of L. By definition, we have

$$\tilde{y}^{\rho'} = L_y(\tilde{y}^\rho, \tilde{\eta}^\rho) = L_y(\tilde{y}^\rho, \widetilde{F}^\rho(\tilde{y}^\rho)),$$

where $\widetilde{F}^\rho(\tilde{y}^\rho)$ satisfies $\|\partial\widetilde{F}^\rho(\tilde{y}^\rho)/\partial\tilde{y}^\rho\| \leq \gamma$. Therefore, we have

$$\frac{\partial\tilde{y}^{\rho'}}{\partial\tilde{y}^\rho} = \frac{\partial L_y}{\partial\tilde{y}^\rho} + \frac{\partial\widetilde{F}^\rho(\tilde{y}^\rho)}{\partial\tilde{y}^\rho}\frac{\partial L_y}{\partial\tilde{\eta}^\rho} = U + \widetilde{R},$$

where U is unitary.

By equations (28) and (30), we have $\|\widetilde{R}\| \leq 2\gamma C_0\epsilon^{\mathsf{p}} < 1$ by assumption. Therefore, $\tilde{y}^\rho \mapsto \tilde{y}^{\rho'}$ is invertible, and we have $\|\partial\tilde{y}^\rho/\partial\tilde{y}^{\rho'}\| \leq (1 - 2\gamma C_0\epsilon^{\mathsf{p}})^{-1}$. We may see $\tilde{\eta}^{\rho'}$ as a function of $\tilde{y}^{\rho'}$, and we have

$$\left\|\frac{\partial\tilde{\eta}^{\rho'}}{\partial\tilde{y}^{\rho'}}\right\| = \left\|\frac{\partial\tilde{y}^\rho}{\partial\tilde{y}^{\rho'}}\frac{\partial\tilde{\eta}^{\rho'}}{\partial\tilde{y}^\rho} + \frac{\partial\tilde{y}^\rho}{\partial\tilde{y}^{\rho'}}\frac{\partial\tilde{\eta}^\rho}{\partial\tilde{y}^\rho}\frac{\partial\tilde{\eta}^{\rho'}}{\partial\tilde{\eta}^\rho}\right\| \leq (1 - 2\gamma C_0\epsilon^{\mathsf{p}})^{-1}(C_0\epsilon^{\mathsf{p}} + \gamma(1 + C_0\epsilon^{\mathsf{p}})),$$

and the lemma follows. □

2.3.4. *Propagation for bounded times.* Let us fix a $\nu_1 \in (\nu, 1)$, where ν was defined in (22). Recall that p was defined in (29) as the Hölder exponent of the stable and unstable directions. From now on, we fix an $\epsilon > 0$ small enough so that

$$\frac{\nu + C_0\epsilon^{\mathsf{p}}}{\nu^{-1} - C_0\epsilon^{\mathsf{p}}} < \nu_1, \quad \text{and} \quad \frac{C_0\epsilon^{\mathsf{p}}}{\nu^{-1} - 2C_0\epsilon^{\mathsf{p}}} < \frac{\gamma_{\mathrm{uns}}(1 - \nu_1)}{8}, \tag{31}$$

$$\left(1 - \frac{(1 + \nu_1)\gamma_{\mathrm{uns}}}{1 + 2C_0\epsilon^{\mathsf{p}})}C_0\epsilon^{\mathsf{p}}\right)^{-1}\left(\gamma_{\mathrm{uns}}\frac{(1 + \nu_1)(1 + C_0\epsilon^{\mathsf{p}})}{2 + 4C_0\epsilon^{\mathsf{p}}} + C_0\epsilon^{\mathsf{p}}\right) < \frac{\gamma_{\mathrm{uns}}}{1 + 2C_0\epsilon^{\mathsf{p}}}. \tag{32}$$

This is possible because $(1 + \nu_1)/2 < 1$. We also ask that $C_0\epsilon^{\mathsf{p}} < 1/2$. Note that, although condition (32) looks horrible, it is designed to work well with Lemma 23.

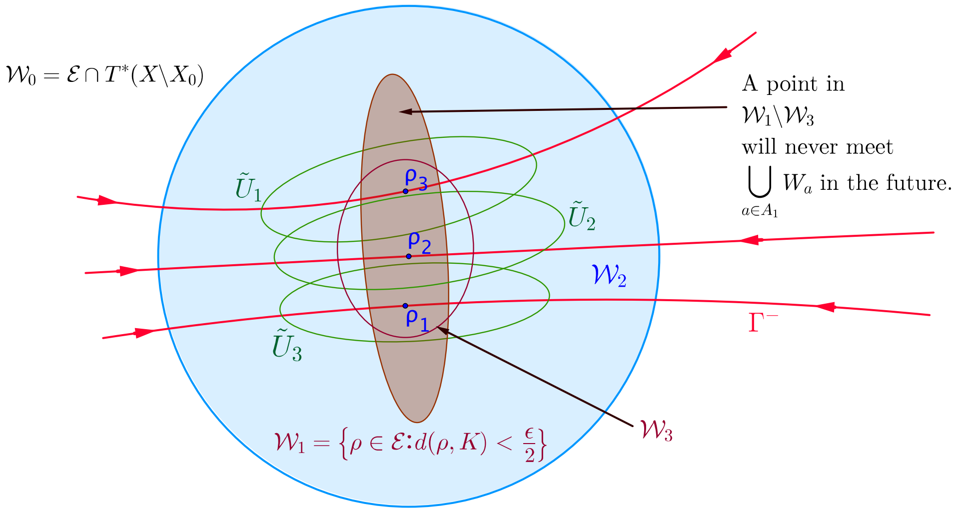


Figure 3. A representation of some of the different sets we introduce in the proof of Theorem 17, intersected with a Poincaré section.

Let us introduce a first decomposition of the energy layer. Recall that we defined $\mathcal{W}_0$ in (8) as the external part of the energy layer. We define $\mathcal{W}_1 := \{\rho \in \mathcal{E} \backslash \mathcal{W}_0 : d(\rho, K) < \epsilon/2\}$ for the part of the energy layer close to the trapped set, and $\mathcal{W}_2 := \{\rho \in \mathcal{E} \backslash \mathcal{W}_0 : d(\rho, K) \geq \epsilon/2\}$ for the intermediate region. See Figure 3 for a representation of these different sets. Note that we will later introduce a finer open cover of the energy layer, using the sets W_a appearing in the statement of the theorem.

The following lemma tells us that the set $\mathcal{W}_2$ is a transient set, that is to say, points spend only a finite time inside it.

Lemma 24. *There exists $N_\epsilon \in \mathbb{N}$ an integer which depends on ϵ such that for all $\rho \in \mathcal{W}_2$, we have either $\Phi^{N_\epsilon}(\rho) \in \mathcal{W}_0$ or $\Phi^{-N_\epsilon}(\rho) \in \mathcal{W}_0$.*

Proof. This result comes from the uniform transversality of the stable and unstable manifolds (which is a direct consequence of the compactness of K).

It gives us the existence of a $d_1(\epsilon) > 0$ such that, for all $\rho \in \mathcal{W}_2 \cup \mathcal{W}_1$,

$$d(\rho, \Gamma^+) + d(\rho, \Gamma^-) \leq 2d_1 \quad \Longrightarrow \quad d(\rho, K) \leq \epsilon/2.$$

We may therefore write

$$\mathcal{W}_2 = \{\rho \in \mathcal{W}_2 : d(\rho, \Gamma^-) > d_1\} \cup \{\rho \in \mathcal{W}_2 : d(\rho, \Gamma^-) > d_1\}.$$

A point in the first set will leave the interaction region in finite time in the future, while a point in the second set will leave it in finite time in the past. By compactness, we can find a uniform N_ϵ such as the one in the statement of the lemma. $\square$

The following lemma is a consequence of the transversality assumption we made. It tells us that when we propagate $\mathcal{L}_0$ during a finite time N and restrict it to a small set $\widetilde{U}^\rho(\epsilon, \varrho)$ close to the trapped set, we obtain a finite union of Lagrangian manifolds in the alternative coordinates. Here, the size ϱ of the set in the unstable direction depends on N, but its size ϵ in the stable direction does not.

Lemma 25. *Let $N \in \mathbb{N}$. There exists $\mathcal{N}_N \in \mathbb{N}$, $\tilde{\varrho}_N > 0$ and $\tilde{\gamma}_N > 0$ such that for all $0 < \varrho \leq \tilde{\varrho}_N$, for all $\rho \in K$, and for all $1 \leq t \leq N$, the set $\Phi^t(\mathcal{L}_0) \cap \widetilde{U}^\rho(\epsilon, \varrho)$ can be written in the coordinates $(\tilde{y}^\rho, \tilde{\eta}^\rho)$ as the union of at most $\mathcal{N}_N$ disjoint Lagrangian manifolds, which are all $\tilde{\gamma}_N$-unstable*:

$$\Phi^t(\mathcal{L}_0) \cap \widetilde{U}^\rho(\epsilon, \varrho) \equiv \bigcup_{l=0}^{l(\varrho)} \hat{\Lambda}_l,$$

with $l(\varrho) \leq \mathcal{N}_N$ and

$$\hat{\Lambda}_l = \{(\tilde{y}_1^\rho, \boldsymbol{\tilde{y}}^\rho; 0, f^l(\tilde{y}^\rho)) : \boldsymbol{\tilde{y}}^\rho \in D_\varrho\}$$

for some smooth functions f^l with $\|df^l(\tilde{y}^\rho)\|_{C^0(D_\epsilon)} \leq \tilde{\gamma}_N$.

Proof. Let us consider a $1 \leq t \leq N$. First of all, since Φ^t is a symplectomorphism, it sends Lagrangian manifolds to Lagrangian manifolds. The restriction of a Lagrangian manifold to a region of phase space is a union of Lagrangian manifolds.

We now have to prove that, if we take ϱ small enough, these Lagrangian manifolds are all $\tilde{\gamma}_N$ unstable for some $\tilde{\gamma}_N > 0$ which is independent of ρ.

Let $\rho \in K$. By hypothesis, $W^-_{\text{loc}}(\rho)$ and $\Phi^t(\mathcal{L}_0)$ are transverse when they intersect.

Therefore, in a small neighbourhood of the stable manifold $\{\boldsymbol{\tilde{y}}^\rho = 0\}$, each connected component of $\Phi^t(\mathcal{L}_0)$ may be projected smoothly on the twisted unstable manifold $\{\boldsymbol{\tilde{\eta}}^\rho = 0\}$. That is to say, there exists a $\varrho > 0$ and a $\gamma > 0$ such that each connected component of $\Phi^t(\mathcal{L}_0) \cap \widetilde{U}^\rho(\epsilon, \varrho)$ is γ-unstable in the twisted coordinates around ρ for some $\gamma > 0$.

Now, since the changes of coordinates between twisted coordinates are continuous, we may use the compactness of K to find uniform constants $\varrho > 0$ and $\gamma > 0$ such that each connected component of $\Phi^t(\mathcal{L}_0) \cap \widetilde{U}^\rho(\epsilon, \varrho)$ is γ-unstable in the twisted coordinates around ρ, independently of $\rho \in K$ and $1 \leq t \leq N$.

By compactness of $\widetilde{U}^\rho(\epsilon, \varrho)$, the number of Lagrangian manifolds making up $\Phi^t(\mathcal{L}_0) \cap \widetilde{U}^\rho(\epsilon, \varrho)$ is finite. ☐

Applying this lemma to $N = N_\epsilon + 2$, we define the following constants, which we shall need later in the proof (recall that γ_{uns} has been fixed):

$$(\gamma_0, \varrho_0) := (\tilde{\gamma}_{N_\epsilon+2}, \tilde{\varrho}_{N_\epsilon+2}), \tag{33}$$

$$N_1 := \left\lfloor \frac{\log(\gamma_{\text{uns}}/4\gamma_0)}{\log((1+\nu_1)/2)} \right\rfloor + 1, \quad N_{\text{uns}} := N_1 + N_\epsilon + 2, \tag{34}$$

$$\varrho_1 := \min\Big(\frac{\epsilon}{2\gamma_0}, \varrho_0\Big), \quad \varrho_2 := \min\big((\mathcal{C} + C_0\epsilon^{\mathsf{P}})^{-N_{\text{uns}}}\varrho_1, \tilde{\delta}_{N_{\text{uns}}}\big), \tag{35}$$

where $\mathcal{C}$ comes from Remark 21, and C_0 comes from equation (26).

Remark 26. As explained in Lemma 24, N_ϵ is the maximal time spent by a trajectory in the intermediate region $\mathcal{W}_2$. The time N_1 will be the time necessary to incline a γ_0-unstable Lagrangian manifold to a γ_{uns}-unstable Lagrangian manifold, as explained in Proposition 30. As for the constant ϱ_2, it has been chosen small enough so that at each step of the aforementioned propagation during a time N_1, the Lagrangian manifolds we consider are contained in a single coordinate chart, as explained in Proposition 30.

Remark 27. The constant ε_0 in Theorem 17 will depend only on γ_0 and ϱ_0. Therefore, the proof of Lemma 25 tells us that if we consider a whole family of Lagrangian manifolds $(\mathcal{L}_z)_{z\in Z}$ satisfying Hypotheses 13 and 16, we will be able to find an $\varepsilon_0 > 0$ uniform in $z \in Z$ provided we have the following uniform transversality condition:

$$\forall t \in N,\ \forall \rho \in K,\ \exists \delta, \gamma > 0 \text{ such that } \forall z \in Z,\ \Phi^t(\mathcal{L}_z) \cap \widetilde{U}^{\rho}(\epsilon, \delta) \text{ is } \gamma\text{-unstable}. \tag{36}$$

Lemma 28. *There exists a neighbourhood $\mathcal{W}_3$ of $\Gamma^- \cap \mathcal{W}_1$ in $\mathcal{E}$, a finite set of points $(\rho_i)_{i\in I} \subset K$ and $0 < \epsilon_1 < \varrho_1$, such that the following holds:*

(i) *The sets $(\widetilde{U}_i)_{i\in I} := (\widetilde{U}^{\rho_i}(\epsilon, \varrho_2))_{i\in I}$ form an open cover of a neighbourhood of $\mathcal{W}_3$.*

(ii) $\rho \in [\mathcal{W}_1 \backslash \mathcal{W}_3] \cup \{\rho' \in \mathcal{W}_2 : d(\rho', \Gamma^-) > d_1\} \implies \forall t \geq 0,\ d(\Phi^t(\rho)), K) \geq \epsilon_1$.

(iii) *For any open set W of diameter $< \epsilon_1$ included in $\mathcal{W}_3$, there exists an $i \in I$ such that $W \subset \widetilde{U}_i$.*

Proof. The sets $(\widetilde{U}^{\rho}(\epsilon, \varrho_2))_{\rho\in K}$ form an open cover of a neighbourhood of $(\Gamma^- \cap \mathcal{W}_1)$. Let us denote by $\mathcal{W}_3$ such a neighbourhood.

By compactness, we may extract from it a finite open cover $(\widetilde{U}_i)_{i\in I} := (\widetilde{U}^{\rho_i}(\epsilon, \varrho_2))_{i\in I}$, which still satisfies (i).

Since $\mathcal{W}_3$ is a neighbourhood of $\Gamma^- \cap \mathcal{W}_1$, there exists a constant $\varrho_2' > 0$ such that the following holds:

$$\forall \rho \in \mathcal{W}_1 \backslash \mathcal{W}_3, \quad d(\rho, \Gamma^-) > \varrho_2'.$$

Therefore, there exists $0 < \epsilon_1 < \min(\varrho_1, \epsilon)$ such that

$$\rho \in [\mathcal{W}_1 \backslash \mathcal{W}_3] \cup \{\rho' \in \mathcal{W}_2 : d(\rho', \Gamma^-) \geq d_1\} \implies \forall t \geq 0,\ d(\Phi^t(\rho)), K) > \epsilon_1,$$

which is (ii). Finally, since the set $\widetilde{U}_i$ are open, we may shrink ϵ_1 so that (iii) is satisfied. □

Remark 29. The constant ε_0 appearing in Theorem 17 will be smaller than ϵ_1 (see Lemma 33); therefore each of the sets $(W_a)_{a\in A_1}$ will be contained in some $\widetilde{U}_i$. Furthermore, we will have $W_a \subset \{\rho \in \mathcal{E} : d(\rho, K) < \varepsilon_0\}$. Hence, a point $\rho \in [\mathcal{W}_1 \backslash \mathcal{W}_3] \cup \{\rho' \in \mathcal{W}_2 : d(\rho', \Gamma^-) \geq d_1\}$ will not be contained in any of the sets $(W_a)_{a\in A_1}$ when propagated in the future.

Lemma 25 tells us that $\Phi^{N_\epsilon}(\mathcal{L}_0) \cap \widetilde{U}_i$ consists of finitely many γ_0-unstable Lagrangian manifolds. Our aim will now be to take a Lagrangian manifold included in a $\widetilde{U}_{i_1}$, to propagate it during some time $N \geq N_1$, then to restrict it to a $\widetilde{U}_{i_2}$ for $i_1, i_2 \in I$. The remaining part of the Lagrangian, which is in $\mathcal{W}_1 \backslash \mathcal{W}_3$, will not meet the sets $(W_a)_{a\in A_1}$ when propagated in the future, as explained in Remark 29.

2.3.5. *Propagation in the sets $\widetilde{U}_i$.* For $N \in \mathbb{N}$ and $\iota = (i_0 i_1 \cdots i_{N-1}) \in I^N$, we define

$$\Phi_\iota(\Lambda) := \Phi^1\big(\widetilde{U}_{i_{N-1}} \cap \Phi^1(\cdots \Phi^1(\widetilde{U}_{i_0} \cap \Lambda) \cdots)\big).$$

The propagation of Lagrangian manifolds in the sets $\widetilde{U}_i$ is described in the following proposition, which is the cornerstone of the proof of Theorem 17. Recall that γ_{uns} was chosen arbitrarily at the beginning of the proof, and that N_1 was defined in (34).

Proposition 30. *Let $N \geq N_1$, $\iota = (i_0 i_1 \cdots i_{N-1}) \in I^N$ and $i \in I$. Let $\Lambda^0 \subset \widetilde{U}_{i_0}$ be an isoenergetic Lagrangian manifold which is γ_0-unstable in the twisted coordinates centred on ρ_{i_0}. Then $\widetilde{U}_i \cap \Phi_\iota(\Lambda)$ is a Lagrangian manifold contained in $\widetilde{U}_i$, and it is $(\gamma_{\mathrm{uns}}/(1+2C_0\epsilon^{\mathrm{p}})^2)$-unstable in the twisted coordinates centred on ρ_i.*

Proof. The first part of the proof consists in understanding how $\Phi^n(\Lambda^0)$ behaves for $n \leq N_1$, in the twisted coordinates centred on ρ_{i_0}. This is the content of the following lemma, which is an adaptation to our context of the "inclination lemma". (See [Katok and Hasselblatt 1995, Theorem 6.2.8]; see also [Nonnenmacher and Zworski 2009, Proposition 5.1] for a statement closer to our context and notation.)

Lemma 31. *$\Phi^{N_1}(\Lambda^0)$ is a Lagrangian manifold, which can be written in the chart $(\tilde{y}^{\Phi^{N_1}(\rho_{i_0})}, \tilde{\eta}^{\Phi^{N_1}(\rho_{i_0})})$ in the form*

$$\Phi^{N_1}(\Lambda^0) \equiv \big\{(\tilde{y}_1^{\Phi^{N_1}(\rho_{i_0})}, \tilde{\boldsymbol{y}}^{\Phi^{N_1}(\rho_{i_0})}; 0, f^{N_1}(\tilde{\boldsymbol{y}}^{\Phi^{N_1}(\rho_{i_0})})) : \tilde{\boldsymbol{y}}^{\Phi^{N_1}(\rho_{i_0})} \in D_{N_1}\big\},$$

with $D_{N_1} \subset B(0, \varrho_1)$ and $\|df^{N_1}\|_{C^0(D_k)} \leq (1+\nu_1)\gamma_{\mathrm{uns}}/4$.

Note that $\Phi^{N_1}(\Lambda^0)$ is a priori not contained in a single set $\widetilde{U}_i$, but the lemma states that it is contained in the set $\widetilde{U}^{\Phi^{N_1}(\rho_{i_0})}(\epsilon, \varrho_1)$, where the twisted coordinates are well defined.

Proof. By assumption, Λ^0 may be put in the form

$$\Lambda^0 \equiv \big\{(\tilde{y}_1^{\rho_{i_0}}, \tilde{\boldsymbol{y}}^{\rho_{i_0}}; 0, f^0(\tilde{\boldsymbol{y}}^{\rho_{i_0}})) : |\tilde{\boldsymbol{y}}^{\rho_{i_0}}| < \varrho_2\big\}, \quad \text{with } \|df^0(\tilde{\boldsymbol{y}}^{\rho_{i_0}})\|_{C^0} \leq \gamma_0.$$

We will consider restrictions of the Lagrangian manifolds at intermediate times to the Poincaré sections centred at $\Phi^k(\rho_{i_0})$:

$$\Lambda^k_{\mathrm{sec}} := \Phi^k(\Lambda^0) \cap \Sigma^{\Phi^k(\rho_{i_0})}(\epsilon, \varrho_0).$$

We have $\Lambda^{k+1}_{\mathrm{sec}} = \kappa^k(\Lambda^k_{\mathrm{sec}})$, where $\kappa^k := \kappa_{\Phi^k(\rho_{i_0}), \Phi^{k+1}(\rho_{i_0})}$ is of the form (24). From equation (24) and the definition of $\mathcal{C}$, we see that the maximal rate of expansion in the unstable direction is bounded by $(\mathcal{C} + C_0\epsilon^{\mathrm{p}})$. Therefore, the definition of ϱ_2 implies that for any $k \leq N_1$, the projection of Λ^k_{sec} on the unstable direction is supported in $B(0, \varrho_1)$.

To lighten the notations, we will write $\tilde{\boldsymbol{y}}^k$ and $\tilde{\boldsymbol{\eta}}^k$ instead of $\tilde{\boldsymbol{y}}^{\Phi^k(\rho_{i_0})}$ and $\tilde{\boldsymbol{\eta}}^{\Phi^k(\rho_{i_0})}$.

Let $k \geq 0$, and suppose we may write

$$\Lambda^k_{\mathrm{sec}} \equiv \{(\tilde{\boldsymbol{y}}^k, f^k(\tilde{\boldsymbol{y}}^k)) : \tilde{\boldsymbol{y}}^k \in D_k\},$$

where $D_k \subset B(0, \varrho_1)$, and $\|df^k\|_{C^0} \leq \gamma_k$ for some $0 < \gamma_k \leq \gamma_0$.

Note that the key point in the following computations is that, since we have chosen "alternative" coordinates, we have $|\partial_{\boldsymbol{\eta}} \tilde{\alpha}^k(\tilde{\boldsymbol{y}}^k, \tilde{\boldsymbol{\eta}}^k)| \leq C_0 \tilde{\boldsymbol{y}}^k \leq C_0 \varrho_1$.

The projection of $\Phi^1_{|\Lambda^k_{\mathrm{sec}}}$ on the horizontal subspace is given by

$$\tilde{\boldsymbol{y}}^k \mapsto \tilde{\boldsymbol{y}}^{k+1} = \pi\Phi^1(\tilde{\boldsymbol{y}}^k, f^k(\tilde{\boldsymbol{y}}^k)) = A_k \tilde{\boldsymbol{y}}^k + \tilde{\alpha}^k(\tilde{\boldsymbol{y}}_k, f^k(\tilde{\boldsymbol{y}}_k)),$$

where for each k, we have A_k is a matrix as in (23).

By differentiating, we obtain

$$\frac{\partial \tilde{\boldsymbol{y}}^{k+1}}{\partial \tilde{\boldsymbol{y}}^k} = A_k + \frac{\partial \tilde{\alpha}^k}{\partial \tilde{\boldsymbol{y}}^k} + \frac{\partial \tilde{\alpha}^k}{\partial \tilde{\boldsymbol{\eta}}^k}\frac{\partial f_k}{\partial \tilde{\boldsymbol{y}}^k} = A_k + r_k,$$

where r_k has entries bounded by $C_0 \varrho_1 \gamma_0 \leq C_0 \epsilon$.

Therefore, the map is invertible, and $\tilde{\boldsymbol{y}}^{k+1} \mapsto \tilde{\boldsymbol{y}}^k$ is contracting. This implies that $\Lambda_{\mathrm{sec}}^{k+1}$ can be represented as a graph

$$\Lambda_{\mathrm{sec}}^{k+1} \equiv \{(\tilde{\boldsymbol{y}}^{k+1}, f^{k+1}(\tilde{\boldsymbol{y}}^{k+1})) : \tilde{\boldsymbol{y}}^{k+1} \in D_{k+1}\},$$

with

$$f^{k+1}(\tilde{\boldsymbol{y}}^{k+1}) = {}^t A_k^{-1} f^k(\tilde{\boldsymbol{y}}^k) + \tilde{\beta}_k(\tilde{\boldsymbol{y}}^k, f^k(\tilde{\boldsymbol{y}}^k)).$$

Differentiating with respect to $\tilde{\boldsymbol{y}}^{k+1}$, we get

$$\frac{\partial f^{k+1}}{\partial \tilde{\boldsymbol{y}}^{k+1}} = \left(\frac{\partial \tilde{\boldsymbol{y}}^k}{\partial \tilde{\boldsymbol{y}}^{k+1}}\right)\left[\left({}^t A_k^{-1} + \partial_{\boldsymbol{\eta}} \tilde{\beta}^k(\tilde{\boldsymbol{y}}^k, f^k(\tilde{\boldsymbol{y}}^k))\right)\frac{\partial f^k}{\partial \tilde{\boldsymbol{y}}^k}(\tilde{\boldsymbol{y}}^k) + \partial_{\tilde{\boldsymbol{y}}} \tilde{\beta}^k(\tilde{\boldsymbol{y}}^k, f^k(\tilde{\boldsymbol{y}}^k))\right].$$

Therefore, we have

$$\begin{aligned}\left\|\frac{\partial f^{k+1}}{\partial \tilde{\boldsymbol{y}}^{k+1}}\right\| &\leq \frac{\|{}^t A_k^{-1}\|\gamma_k + |\partial_{\tilde{\boldsymbol{y}}} \tilde{\beta}| + |\partial_{\tilde{\boldsymbol{\eta}}} \tilde{\beta}|\gamma_k}{\nu^{-1} - |\partial_{\tilde{\boldsymbol{y}}} \tilde{\alpha}| - |\partial_{\tilde{\boldsymbol{\eta}}} \tilde{\beta}|\gamma_k} \leq \frac{\gamma_k \nu + C_0 \epsilon^{\mathsf{p}}(1+\gamma_k)}{\nu^{-1} - 2C_0 \epsilon^{\mathsf{p}}} \\ &\leq \nu_1 \gamma_k + \frac{(1-\nu_1)\gamma_{\mathrm{uns}}}{8} = \gamma_k\left(\nu_1 + \frac{\gamma_{\mathrm{uns}}(1-\nu_1)}{8\gamma_k}\right),\end{aligned}$$

where the last inequality comes from (31). First of all, the fact that this slope is bounded uniformly on $\Lambda_{\mathrm{sec}}^{k+1}$ implies that $\Lambda_{\mathrm{sec}}^{k+1}$ can indeed be written in the form

$$\Lambda_{\mathrm{sec}}^{k+1} \equiv \{(\tilde{\boldsymbol{y}}^{k+1}, f^{k+1}(\tilde{\boldsymbol{y}}^{k+1})) : \tilde{\boldsymbol{y}}^{k+1} \in D_{k+1}\},$$

where $D_{k+1} \subset B(0, \varrho_1)$, and $\|df^{k+1}\|_{C^0} \leq \gamma_{k+1}$, where

$$\gamma_{k+1} \leq \gamma_k\left(\nu_1 + \frac{\gamma_{\mathrm{uns}}(1-\nu_1)}{8\gamma_k}\right).$$

Now, if $\gamma_k > \gamma_{\mathrm{uns}}/4$, then

$$\nu_1 + \frac{\gamma_{\mathrm{uns}}(1-\nu_1)}{8\gamma_k} < \frac{1+\nu_1}{2} < 1,$$

so that γ_k decreases exponentially fast, while if $\gamma_k \leq (1+\nu_1)\gamma_{\mathrm{uns}}/4$, then $\gamma_{k+1} < (1+\nu_1)\gamma_{\mathrm{uns}}/4$.

The time N_1 has been chosen large enough so that $\gamma_{N_1} < (1+\nu_1)\gamma_{\mathrm{uns}}/4$, which concludes the proof of the lemma. □

After times $N > N_1$, the Lagrangian manifold may not be included in $\widetilde{U}^{\Phi^N(\rho_{i_0})}(\epsilon, \varrho_1)$. Therefore, we may have to use a change of coordinates. By Lemma 31, at time N_1, our Lagrangian manifold $\Phi^{N_1}(\Lambda^0)$ is included in $\widetilde{U}^{\Phi^{N_1}(\rho_{i_0})}(\epsilon, \varrho_1)$ and is $((1+\nu_1)\gamma_{\mathrm{uns}}/4)$-unstable.

We want to study $\widetilde{U}_j \cap \Phi^{N_1}(\Lambda^0)$ for $j \in I$ in the coordinates centred at ρ_j, and to apply the computations made in the proof of Lemma 31 again. Let us see how all this works.

If, for some $j \in I$, we have $\widetilde{U}_j \cap \Phi^{N_1}(\Lambda^0) \neq \varnothing$, then $d(\Phi^{N_1}(\rho_{i_0}), \rho_j) < \epsilon$. Now, by applying Lemma 23 as well as equation (32), we obtain that $\Phi^{N_1}(\Lambda^0) \cap \widetilde{U}_j$ is $(\gamma_{\text{uns}}/2)$-unstable in the twisted coordinates centred at ρ_j.

We may continue this argument of changing coordinates and propagating to any time $N \geq N_1$: we always obtain a single Lagrangian manifold which is $((1+\nu_1)\gamma_{\text{uns}}/4)$-unstable. This concludes the proof of Proposition 30, because we assumed that $C_0\epsilon^{\mathrm{p}} < 1/2$. □

Remark 32. In [Nonnenmacher and Zworski 2009, Proposition 5.1], the authors prove using the chain rule that for each $\ell \in \mathbb{N}$, there exists a constant $\mathcal{C}_\ell$ large enough such that the following holds. If $i_1, i_2 \in I$ and if $\Lambda \subset \widetilde{U}_{i_1}$ is a Lagrangian manifold in some unstable cone, generated by a function f in the coordinates $(\tilde{y}^{\rho_{i_1}}, \tilde{\eta}^{\rho_{i_1}})$ with $\|f\|_{C^\ell} \leq \mathcal{C}_\ell$, then $\Phi^1(\Lambda) \cap \widetilde{U}_{i_2}$ is a union of finitely many Lagrangian manifolds, all of which are in some unstable cone in the coordinates $(\tilde{y}^{\rho_{i_2}}, \tilde{\eta}^{\rho_{i_2}})$, and are generated by functions with a C^ℓ norm smaller than $\mathcal{C}_\ell$.

In particular, this shows that on the Lagrangian manifold $\Phi_\iota^N(\Lambda)$ described in Proposition 30, the function $s^{\rho_i}(y^{\rho_i})$ has a C^ℓ norm smaller than $\mathcal{C}_\ell$, where $\mathcal{C}_\ell$ is a constant independent of N.

2.3.6. *Properties of the sets $(W_a)_{a\in A_1}$.* The following lemma is an adaptation of Lemma 25 to the "straight coordinates". Note that the main reason why we want to use these straight coordinates is because they are symplectic, which will play a crucial role in the proof of Theorem 47.

Lemma 33. *There exists $\varepsilon_0 < \epsilon_1$ such that, if $(W_a)_{a\in A_1}$ is an adapted cover of K of diameter ε_0 such that for each $a \in A_1$, we have $W_a \cap \mathcal{W}_0 = \varnothing$, and there exists a point $\rho_a \in W_a \cap K \neq \varnothing$, then there exist $\mathcal{N}_{N_{\text{uns}}} \in \mathbb{N}$ and γ' such that the following holds.*

For each $a \in A_1$, for each $1 \leq N \leq N_{\text{uns}}$, the set $\Phi^N(\mathcal{L}_0) \cap W_a$ consists of at most $\mathcal{N}_{N_{\text{uns}}}$ Lagrangian manifolds, all of which are γ'-unstable in the straight coordinates centred on ρ_a.

Proof. Let us choose $\varepsilon_0 > 0$ small enough so that $C_0\varepsilon_0\tilde{\gamma}_{N_{\text{uns}}} < 1$ and such that each set of diameter smaller than ε_0 and which intersects K is contained in some $\widetilde{U}^\rho(\epsilon,\delta)$, with $\delta < \tilde{\delta}_{N_{\text{uns}}}$. By applying Lemma 25, we know that there exists $\mathcal{N}_{N_{\text{uns}}} \in \mathbb{N}$, $\tilde{\delta}_{N_{\text{uns}}} > 0$ and $\tilde{\gamma}_{N_{\text{uns}}} > 0$ such that for all $0 < \delta \leq \tilde{\delta}_{N_{\text{uns}}}$, for all $\rho \in K$ and for all $1 \leq N \leq N_{\text{uns}}$, the set $\Phi^N(\mathcal{L}_0) \cap \widetilde{U}^\rho(\epsilon,\delta)$ can be written in the coordinates $(\tilde{y}^\rho, \tilde{\eta}^\rho)$ as the union of at most $\mathcal{N}_{N_{\text{uns}}}$ Lagrangian manifolds, which are all $\tilde{\gamma}_{N_{\text{uns}}}$-unstable. This gives us the statement in the twisted coordinates. To go to the straight coordinates, we may simply use Lemma 22 thanks to the assumption made on ε_0. □

For any $a \in A_1$, $1 \leq k \leq N_{\text{uns}}$, the set $W_a \cap \Phi^k(\mathcal{L}_0)$ consists of finitely many Lagrangian manifolds. Let us define $\mathrm{d}_{a,k}$ as the minimal distance (with respect to the distance d) between the Lagrangian manifolds which make up $W_a \cap \Phi^k(\mathcal{L}_0)$, with the convention that this quantity is equal to $+\infty$ if $W_a \cap \Phi^k(\mathcal{L}_0)$ consists of a single Lagrangian manifold or is empty. We then set

$$\mathrm{d} := \min\Big(\varepsilon_0, \min_{\substack{a\in A_1\\ 1\leq k\leq N_{\text{uns}}}} \{\mathrm{d}_{a,k}\}\Big) > 0.$$

Remark 34. If we consider a whole family of Lagrangian manifolds $(\mathcal{L}_z)_{z\in Z}$ satisfying Hypotheses 13 and 16, we will be able to apply Theorem 17 to them with sets $(W_a)_{a\in A_2}$ independent of $z\in Z$ provided the constant d is well-defined, that is to say, provided we have

$$\inf_{\substack{a\in A_1, z\in Z\\ 1\leq k\leq N_{\text{uns}}}} \{\mathsf{d}^z_{a,k}\} > 0, \tag{37}$$

where $\mathsf{d}^z_{a,k}$ is the minimal distance between the Lagrangian manifolds which make up $W_a\cap\Phi^k(\mathcal{L}_z)$, with the convention that this quantity is equal to $+\infty$ if $W_a\cap\Phi^k(\mathcal{L}_z)$ consists of a single Lagrangian manifold or is empty.

The flow (Φ^t) is C^1 with respect to time, and hence Lipschitz on $[0, N_{\text{uns}}]$. Therefore, there exists a constant $C>0$ such that for all $t\in[0,N_{\text{uns}}]$ and for all $\rho_1,\rho_2\in\mathcal{E}$, we have

$$d(\Phi^t(\rho_1),\Phi^t(\rho_2))\leq C d(\rho_1,\rho_2).$$

We take

$$\varepsilon_2 := \mathsf{d}/C.$$

We now complete $(W_a)_{a\in A_1}$ to cover the whole energy layer.

2.3.7. *Construction and properties of the sets* $(W_a)_{a\in A_2}$. Recall that $W_0=T^*(X\backslash X_0)$, and that b is the boundary-defining function introduced in Hypothesis 5.

We build the sets $(W_a)_{a\in A_2}$ so that, if we set $A=A_1\cup A_2\cup\{0\}$, the following holds:

- Each of the sets $(W_a)_{a\in A_2}$ has a diameter smaller than ε_2.
- For each $a\in A_2$, we have $d(W_a,K)>\varepsilon_2/2$.
- $(W_a)_{a\in A}$ is an open cover of $\mathcal{E}$.

Our next lemma is the first brick of the proof of the uniqueness of the Lagrangian manifold making up $\Phi^N_\alpha(\mathcal{L}_0)$. It relies on the fact that the sets $(W_a)_{a\in A_2}$ have been built small enough.

Lemma 35. *Let $k\leq N_{\text{uns}}$, $\alpha\in A^k$, and $a\in A_1$. Then the set $W_a\cap\Phi^k_\alpha(\mathcal{L}_0)$ is empty or consists of a single Lagrangian manifold.*

Proof. Let us suppose that $\Phi^k(\mathcal{L}_0)\cap W_a$ is nonempty. We have seen in Lemma 33 that it consists of finitely many Lagrangian manifolds, with a distance between them larger than d. Therefore, for any $1\leq k'\leq k$, the sets $\Phi^{-k'}(\Phi^k(\mathcal{L}_0)\cap W_a)$ consist of Lagrangian manifolds which are at a distance larger than ε_2 from each other. Because of the assumption (9) we made, we have $\alpha_{k'}\in A_2$ for some $k'\leq k$. Since the sets $(W_a)_{a\in A_2}$ have a diameter smaller than ε_2, they separate the Lagrangian manifolds which make up $\Phi^{-k'}(\Phi^k(\mathcal{L}_0)\cap W_a)$. We deduce from this the lemma. ☐

2.3.8. *Structure of the admissible sequences.* We will now state two of lemmas which put some constraints on the sequences $\alpha\in A^N$, with $\alpha_N\in A_1$ such that $\Phi^N_\alpha(\mathcal{L}_0)\neq\varnothing$.

The first of these lemmas tell us that we may restrict ourselves to sequences such that $\alpha_k\neq 0$ for $k\geq 1$.

Lemma 36. *Let $N \in \mathbb{N}$, and let $\alpha \in A^N$, and $a \in A_1$. Suppose that $\alpha_k = 0$ for some $1 \leq k \leq N-1$, and that $W_a \cap \Phi_\alpha^N(\mathcal{L}_0) \neq \varnothing$. Then*

$$W_a \cap \Phi_\alpha^N(\mathcal{L}_0) \subset \Phi^{N-k}_{\alpha_{k+1}\cdots\alpha_{N-1}}(\mathcal{L}_0).$$

Proof. By hypothesis, $\Phi^k_{\alpha_1\cdots\alpha_k}(\mathcal{L}_0) \subset \mathcal{W}_0$, and it intersects $\mathcal{W}_1$ in the future. We have $\mathcal{W}_0 = \mathcal{DE}_- \cup \mathcal{DE}_+$, and a point in $\mathcal{DE}_+$ cannot intersect $\mathcal{W}_1$ in the future. Therefore, the points in $\Phi^k_{\alpha_1\cdots\alpha_k}(\mathcal{L}_0)$ which intersect $\mathcal{W}_1$ in the future are all in $\mathcal{DE}_-$. But by Lemma 9, the point in $\mathcal{DE}_-$ can only have preimages in W_0. Therefore, we have

$$W_a \cap \Phi_\alpha^N(\mathcal{L}_0) \subset W_a \cap \Phi^N_{0\cdots 0\alpha_{k+1}\cdots\alpha_{N-1}}(\mathcal{L}_0) \subset \Phi^{N-k}_{\alpha_{k+1}\cdots\alpha_{N-1}}(\mathcal{L}_0),$$

where the second inclusion comes from Hypothesis 13. □

Let us now take advantage of Remark 29 to show that, from time $k \geq N_\epsilon + 2$, all the interesting dynamics takes place in $\mathcal{W}_3$.

Lemma 37. *Let $N \geq N_\epsilon + 2$, and $\alpha \in A^N$ with $\alpha_i \neq 0$ for $i \geq 1$.*

Let $N_\epsilon + 2 \leq k \leq N$, and $\rho \in \Phi^k_{\alpha_1\cdots\alpha_k}(\mathcal{L}_0)$ be such that $\Phi^{N-k}(\rho) \in W_a$ for some $a \in A_1$. Then $\rho \in \mathcal{W}_3$.

Proof. If $\rho \in \mathcal{W}_1$, then the result follows from Remark 29. We must therefore check that we cannot have $\rho \in \mathcal{W}_2 \cup \mathcal{W}_0$. First of all, note that Lemma 9 implies that we cannot have $\rho \in \mathcal{W}_0$. This lemma also implies that for each $a' \in A_1 \cup A_2$, we have

$$\Phi^1(W_{a'} \backslash W_0) \cap \mathcal{DE}_- = \varnothing. \tag{38}$$

Suppose now that $\rho \in \mathcal{W}_2$. Since $k \geq N_\epsilon + 2$, and $\alpha_i \neq 0$ for $i \geq 1$, we have $\Phi^{-N_\epsilon - 1}(\rho) \in W_{a'}$ for some $a' \in A_1 \cup A_2$. Therefore, by equation (38), we have $\Phi^{-N_\epsilon}(\rho) \notin \mathcal{W}_0$.

By the proof of Lemma 24, this would imply that $d(\rho, \Gamma^-) \geq d_1$. By Remark 29, this implies that we cannot have $\Phi^{N-k}(\rho) \in W_a$ for some $a \in A_1$, a contradiction. □

2.3.9. *End of the proof of Theorem 17.* Let $N \geq 0$, $\alpha \in A^N$ and $a \in A_1$. If $N \leq N_{\text{uns}}$, the result of Theorem 17 is a consequence of Lemmas 33 and 35.

Consider now $N \geq N_{\text{uns}} > N_\epsilon + 2$. We will assume that $W_a \cap \Phi_\alpha^N(\mathcal{L}_0) \neq \varnothing$. Thanks to Lemma 36 and to Hypothesis 13, we may assume that $\alpha_i \neq 0$ for all $i \geq 1$.

From Lemma 37, we deduce that

$$W_a \cap \Phi_\alpha^N(\mathcal{L}_0) \subset \bigcup_{\substack{\iota \in I^{N-N_\epsilon-1} \\ \iota_{N-N_\epsilon} = i_\alpha}} \Phi_\iota(\Phi^{N_\epsilon+2}_{\alpha_1\cdots\alpha_{N_\epsilon+2}}(\mathcal{L}_0)), \tag{39}$$

where $i_\alpha \in I$ is such that $W_{\alpha_N} \subset \widetilde{U}_{i_\alpha}$.

Let us define

$$\Lambda_k := \{\rho \in \Phi_\alpha^k(\mathcal{L}_0) : \forall k' \geq 0,\ \Phi^{k'}(\rho) \in W_{\alpha_{k+k'}}\}.$$

By Lemma 37, for each $k \geq N_\epsilon + 2$, we have $\Lambda_k \subset \mathcal{W}_3 \cap W_{\alpha_k}$. Therefore, by Lemma 28(iii), there exists a $i_k \in I$ such that $\Lambda_k \subset \widetilde{U}_{i_k}$, and we obtain that

$$W_a \cap \Phi_\alpha^N(\mathcal{L}_0) \subset \Phi^{N-N_\epsilon-2}_{i_{N_\epsilon+2}\cdots i_N}(\Phi^{N_\epsilon+2}_{\alpha_1\cdots\alpha_{N_\epsilon+2}}(\mathcal{L}_0)).$$

We know from Lemmas 25 and 35 that $\Phi^{N_\epsilon+2}_{\alpha_1\cdots\alpha_{N_\epsilon+2}}(\mathcal{L}_0)$ consists of a single Lagrangian manifold, which is γ_0-unstable in the coordinates centred on any point of K. Applying Proposition 30, we know that the right-hand side of (39) is a Lagrangian manifold which is $(\gamma_{\text{uns}}/(1+2C_a\epsilon^{\mathsf{p}})^2)$-unstable in the twisted coordinates centred on ρ_{i_α}.

We first apply Lemma 23 to write this Lagrangian manifold in the twisted coordinates centred on ρ_a. Thanks to equation (32), it is $(\gamma_{\text{uns}}/(1+2C_a\epsilon^{\mathsf{p}}))$-unstable. We then use Lemma 22 to write this Lagrangian manifold in the straight coordinates centred on ρ_{α_N}, and we deduce that it is γ_{uns}-unstable. This concludes the proof of Theorem 17. □

Remark 38. Therefore, in the coordinates (y^a, η^a), the set $W_a \cap \Phi^N_\alpha(\mathcal{L}_0)$ may be put in the form

$$W_a \cap \Phi^N_\alpha(\mathcal{L}_0) \equiv \{(y_1^a, \boldsymbol{y}^a; 0, f_{N,\alpha,a}(\boldsymbol{y}^a)) : \boldsymbol{y}^a \in D_{N,\alpha,a}\}$$

for some open set $D_{N,\alpha,a} \subset \mathbb{R}^d$.

Remark 32 tells us that for any $\ell \in \mathbb{N}$, the functions $f_{N,\alpha,a}$ have C^ℓ norms which are bounded independently of N, α and a.

3. Generalized eigenfunctions

We shall state our results about generalized eigenfunctions under rather general assumptions. We shall then explain why these assumptions hold in the case of distorted plane waves on manifolds which are Euclidean near infinity.

In the sequel, we will consider a Riemannian manifold (X, g) with a real-valued potential $V \in C_c^\infty(X)$, and define the Schrödinger operator

$$P_h = -h^2\Delta_g - c_0 h^2 + V(x).$$

Here $c_0 > 0$ is a constant, which will be 0 in the case of Euclidean-near-infinity manifolds (see Section 3.3 for the definition of such manifolds).

Before stating our assumptions, let us recall a few definitions and facts from semiclassical analysis.

3.1. *Refresher on semiclassical analysis.*

3.1.1. *Pseudodifferential calculus.* We shall use the class $S^{\text{comp}}(T^*X)$ of symbols $a \in C_c^\infty(T^*X)$, which may depend on h, but whose seminorms and supports are all bounded independently of h. We will sometimes write $S^{\text{comp}}(X)$ for the set of symbols in $S^{\text{comp}}(T^*X)$ which depend only on the base variable. If U is an open subset of T^*X, we will denote by $S^{\text{comp}}(U)$ the set of functions in $S^{\text{comp}}(T^*X)$ whose support is contained in U.

Definition 39. Let $a \in S^{\mathrm{comp}}(T^*Y)$. We will say that a is a *classical symbol* if there exists a sequence of symbols $a_k \in S^{\mathrm{comp}}(T^*Y)$ such that for any $n \in \mathbb{N}$,

$$a - \sum_{k=0}^{n} h^k a_k \in h^{n+1} S^{\mathrm{comp}}(T^*Y).$$

We will then write

$$a^0(x,\xi) := \lim_{h \to 0} a(x,\xi;h)$$

for the *principal symbol* of a.

We associate to $S^{\mathrm{comp}}(T^*X)$ the class of pseudodifferential operators $\Psi_h^{\mathrm{comp}}(X)$, through a surjective quantization map

$$\mathrm{Op}_h : S^{\mathrm{comp}}(T^*X) \to \Psi_h^{\mathrm{comp}}(X).$$

This quantization map is defined using coordinate charts, and the standard Weyl quantization on $\mathbb{R}^d$. It is therefore not intrinsic. However, the principal symbol map

$$\sigma_h : \Psi_h^{\mathrm{comp}}(X) \to S^{\mathrm{comp}}(T^*X)/hS^{\mathrm{comp}}(T^*X)$$

is intrinsic, and we have

$$\sigma_h(A \circ B) = \sigma_h(A)\sigma_h(B)$$

and

$$\sigma_h \circ \mathrm{Op} : S^{\mathrm{comp}}(T^*X) \to S^{\mathrm{comp}}(T^*X)/hS^{\mathrm{comp}}(T^*X)$$

is the natural projection map.

For more details on all these maps and their construction, we refer the reader to [Zworski 2012, Chapter 14].

For $a \in S^{\mathrm{comp}}(T^*X)$, we say its essential support is equal to a given compact $K \Subset T^*X$, denoted by

$$\operatorname{ess\,supp}_h a = K \Subset T^*X,$$

if and only if, for all $\chi \in S(T^*X)$,

$$\operatorname{supp} \chi \subset (T^*X \backslash K) \quad \Longrightarrow \quad \chi a \in h^\infty S(T^*X).$$

For $A \in \Psi_h^{\mathrm{comp}}(X)$, $A = \mathrm{Op}_h(a)$, we define the wave front set of A as

$$\mathrm{WF}_h(A) = \operatorname{ess\,supp}_h a,$$

noting that this definition does not depend on the choice of the quantisation. When K is a compact subset of T^*X and $\mathrm{WF}_h(A) \subset K$, we will sometimes say that A is *microsupported* inside K.

Let us now state a lemma which is a consequence of Egorov theorem [Zworski 2012, Theorem 11.1]. Recall that $U(t)$ is the Schrödinger propagator $U(t) = e^{-itP_h/h}$.

Lemma 40. *Let $A, B \in \Psi_h^{\mathrm{comp}}(X)$, and suppose that $\Phi^t(\mathrm{WF}_h(A)) \cap \mathrm{WF}_h(B) = \varnothing$. Then we have*

$$AU(t)B = O_{L^2 \to L^2}(h^\infty).$$

If U, V are bounded open subsets of T^*X, and if $T, T' : L^2(X) \to L^2(X)$ are bounded operators, we shall say that $T \equiv T'$ *microlocally* near $U \times V$ if there exist bounded open sets $\widetilde{U} \supset \overline{U}$ and $\widetilde{V} \supset \overline{V}$ such that for any $A, B \in \Psi_h^{\mathrm{comp}}(X)$ with $\mathrm{WF}(A) \subset \widetilde{U}$ and $\mathrm{WF}(B) \subset \widetilde{V}$, we have

$$A(T - T')B = O_{L^2 \to L^2}(h^\infty)$$

Tempered distributions. Let $u = (u(h))$ be an h-dependent family of distributions in $\mathcal{D}'(X)$. We say it is *h-tempered* if for any bounded open set $U \subset X$, there exists $C > 0$ and $N \in \mathbb{N}$ such that

$$\|u(h)\|_{H_h^{-N}(U)} \leq C h^{-N},$$

where $\|\cdot\|_{H_h^{-N}(U)}$ is the semiclassical Sobolev norm.

For a tempered distribution $u = (u(h))$, we say that a point $\rho \in T^*X$ does not lie in the wave front set $\mathrm{WF}(u)$ if there exists a neighbourhood V of ρ in T^*X such that for any $A \in \Psi_h^{\mathrm{comp}}(X)$ with $\mathrm{WF}(a) \subset V$, we have $Au = O(h^\infty)$.

3.1.2. *Lagrangian distributions and Fourier integral operators.*

Phase functions. Let $\phi(x, \theta)$ be a smooth real-valued function on some open subset U_ϕ of $X \times \mathbb{R}^L$ for some $L \in \mathbb{N}$. We call x the *base variable* and θ the *oscillatory variable*. We say that ϕ is a *nondegenerate phase function* if the differentials $d(\partial_{\theta_1}\phi) \cdots d(\partial_{\theta_L}\phi)$ are linearly independent on the *critical set*

$$C_\phi := \{(x, \theta) : \partial_\theta \phi = 0\} \subset U_\phi.$$

In this case

$$\Lambda_\phi := \{(x, \partial_x \phi(x, \theta)) : (x, \theta) \in C_\phi\} \subset T^*X$$

is an immersed Lagrangian manifold. By shrinking the domain of ϕ, we can make it an embedded Lagrangian manifold. We say that ϕ *generates* Λ_ϕ.

Lagrangian distributions. Given a phase function ϕ and a symbol $a \in S^{\mathrm{comp}}(U_\phi)$, consider the h-dependent family of functions

$$u(x; h) = h^{-L/2} \int_{\mathbb{R}^L} e^{i\phi(x,\theta)/h} a(x, \theta; h)\, d\theta. \tag{40}$$

We call $u = (u(h))$ a *Lagrangian distribution*, (or a *Lagrangian state*) generated by ϕ. By the method of nonstationary phase, if $\operatorname{supp} a$ is contained in some h-independent compact set $K \subset U_\phi$, then

$$\mathrm{WF}_h(u) \subset \big\{(x, \partial_x \phi(x, \theta)) : (x, \theta) \in C_\phi \cap K\big\} \subset \Lambda_\phi.$$

Definition 41. Let $\Lambda \subset T^*X$ be an embedded Lagrangian submanifold. We say that an h-dependent family of functions $u(x; h) \in C_c^\infty(X)$ is a (compactly supported and compactly microlocalized) *Lagrangian distribution associated to* Λ, if it can be written as a sum of finitely many functions of the form (40), for different phase functions ϕ parametrizing open subsets of Λ, plus an $O(h^\infty)$ remainder. We will denote by $I^{\mathrm{comp}}(\Lambda)$ the space of all such functions.

Fourier integral operators. Let X, X' be two manifolds of the same dimension d, and let κ be a symplectomorphism from an open subset of T^*X to an open subset of T^*X'. Consider the Lagrangian

$$\Lambda_\kappa = \{(x', -\nu'; x, \nu) : \kappa(x, \nu) = (x', \nu')\} \subset T^*X' \times T^*X = T^*(X' \times X).$$

A compactly supported operator $U : \mathcal{D}'(X) \to C_c^\infty(X')$ is called a (semiclassical) *Fourier integral operator* associated to κ if its Schwartz kernel $K_U(x', x)$ lies in $h^{-d/2} I^{\mathrm{comp}}(\Lambda_\kappa)$. We write $U \in I^{\mathrm{comp}}(\kappa)$. The $h^{-d/2}$ factor is explained as follows: the normalization for Lagrangian distributions is chosen so that $\|u\|_{L^2} \sim 1$, while the normalization for Fourier integral operators is chosen so that $\|U\|_{L^2(X) \to L^2(X')} \sim 1$.

Note that if $\kappa \circ \kappa'$ is well defined, and if $U \in I^{\mathrm{comp}}(\kappa)$ and $U' \in I^{\mathrm{comp}}(\kappa')$, then $U \circ U' \in I^{\mathrm{comp}}(\kappa \circ \kappa')$.

If $U \in I^{\mathrm{comp}}(\kappa)$ and $O \subset T^*X$ is an open bounded set, we shall say that U is *microlocally unitary* near O if $U^*U \equiv I_{L^2(X) \to L^2(X)}$ microlocally near $O \times \kappa(O)$.

3.1.3. *Local properties of Fourier integral operators.* In this section we shall see that, if we work locally, we may describe many Fourier integral operators without the help of oscillatory coordinates. In particular, following [Nonnenmacher and Zworski 2009, §4.1], we will recall the effect of a Fourier integral operator on a Lagrangian distribution which has no caustics. We will recall in Section 4.2 how this formalism may be applied to the study of the Schrödinger propagator.

Let $\kappa : T^*\mathbb{R}^d \to T^*\mathbb{R}^d$ be a local symplectic diffeomorphism. By performing phase-space translations, we may assume that κ is defined in a neighbourhood of $(0,0)$ and that $\kappa(0,0) = (0,0)$.

Without loss of generality, we can find linear Lagrangian subspaces, $\Gamma_j, \Gamma_j^\perp \subset T^*\mathbb{R}^d$, $j = 0, 1$, with the following properties:

- $\Gamma_j^\perp$ is transversal to Γ_j.
- If π_j (resp. $\pi_j^\perp$) is the projection $T^*\mathbb{R}^d \to \Gamma_j$ along $\Gamma_j^\perp$ (resp. the projection $T^*\mathbb{R}^d \to \Gamma_j^\perp$ along Γ_j), then, for some neighbourhood U of ρ_0, the map
$$\kappa(U) \times U \to \Gamma_1 \times \Gamma_0^\perp, \quad (\kappa(\rho), \rho) \mapsto \pi_1(\kappa(\rho)) \times \pi_0^\perp,$$
is a local diffeomorphism from the graph of $\kappa|_U$ to a neighbourhood of the origin in $\Gamma_1 \times \Gamma_0^\perp$.

Let A_j, $j = 0, 1$ be linear symplectic transformations with the properties

$$A_j(\Gamma_j) = \{(x, 0)\} \subset T^*\mathbb{R}^d \quad \text{and} \quad A_j(\Gamma_j^\perp) = \{(0, \xi)\} \subset T^*\mathbb{R}^d,$$

and let M_j be *metaplectic quantizations* of the A_j as defined in [Dimassi and Sjöstrand 1999, Appendix to Chapter 7]. Then the rotated diffeomorphism

$$\tilde{\kappa} := A_1 \circ \kappa \circ A_0^{-1}$$

is such that the projection from the graph of $\tilde{\kappa}$

$$T^*\mathbb{R}^d \times T^*\mathbb{R}^d \to \in \mathbb{R}^d \times \mathbb{R}^d, \quad (x^1, \xi^1; x^0, \xi^0) \mapsto (x^1, \xi^0), \quad (x^1, \xi^1) = \tilde{\kappa}(x^0, \xi^0), \tag{41}$$

is a diffeomorphism near the origin. It then follows that there exists a unique function $\tilde{\psi} \in C^\infty(\mathbb{R}^d \times \mathbb{R}^d)$ such that for (x^1, ξ^0) near $(0,0)$,

$$\tilde{\kappa}(\tilde{\psi}'_\xi(x^1, \xi^0), \xi^0) = (x^1, \tilde{\psi}'_x(x^1, \xi^0)), \quad \det \tilde{\psi}''_{x\xi} \neq 0 \text{ and } \tilde{\psi}(0,0) = 0.$$

The function $\tilde{\psi}$ is said to *generate* the transformation $\tilde{\kappa}$ near $(0,0)$.

Note that if $\widetilde{T} \in I^{\text{comp}}(\tilde{\kappa})$, then

$$T := M_1^{-1} \circ \widetilde{T} \circ M_0 \in I^{\text{comp}}(\kappa). \tag{42}$$

Thanks to assumption (41), a Fourier integral operator $\widetilde{T} \in I^{\text{comp}}(\tilde{\kappa})$ may then be written in the form

$$\widetilde{T}u(x^1) := \frac{1}{(2\pi h)^d} \iint_{\mathbb{R}^{2n}} e^{i(\tilde{\psi}(x^1,\xi^0)-\langle x^0,\xi^0\rangle)/h} \alpha(x^1,\xi^0;h)u(x^0)\,dx^0\,d\xi^0, \tag{43}$$

with $\alpha \in S^{\text{comp}}(\mathbb{R}^{2d})$.

Now, let us state a lemma which was proven in [Nonnenmacher and Zworski 2009, Lemma 4.1], and which describes the effect of a Fourier integral operator of the form (43) on a Lagrangian distribution which projects on the base manifold without caustics.

Lemma 42. *Consider a Lagrangian $\Lambda_0 = \{(x_0, \phi_0'(x_0)) : x \in \Omega_0\}$, $\phi_0 \in C_b^\infty(\Omega_0)$, contained in a small neighbourhood $V \subset T^*\mathbb{R}^d$ such that κ is generated by ψ near V. We assume that*

$$\kappa(\Lambda_0) = \Lambda_1 = \{(x, \phi_1'(x)) : x \in \Omega_1\}, \quad \phi_1 \in C_b^\infty(\Omega_1).$$

Then, for any symbol $a \in S^{\text{comp}}(\Omega_0)$, the application of a Fourier integral operator T of the form (43) to the Lagrangian state

$$a(x)e^{i\phi_0(x)/h}$$

associated with Λ_0 can be expanded, for any $L > 0$, into

$$T(ae^{i\phi_0/h})(x) = e^{i\phi_1(x)/h}\Big(\sum_{j=0}^{L-1} b_j(x)h^j + h^L r_L(x,h)\Big),$$

where $b_j \in S^{\text{comp}}$, and for any $\ell \in \mathbb{N}$, we have

$$\|b_j\|_{C^\ell(\Omega_1)} \le C_{\ell,j}\|a\|_{C^{\ell+2j}(\Omega_0)}, \quad 0 \le j \le L-1,$$
$$\|r_L(\cdot,h)\|_{C^\ell(\Omega_1)} \le C_{\ell,L}\|a\|_{C^{\ell+2L+n}(\Omega_0)}.$$

The constants $C_{\ell,j}$ depend only on κ, α and $\sup_{\Omega_0} |\partial^\beta \phi_0|$ for $0 < |\beta| \le 2\ell + j$.

3.2. *Assumptions on the generalized eigenfunctions.* We consider generalized eigenfunctions of P_h at energy 1, that is to say, a family of smooth functions $E_h \in C^\infty(X)$ indexed by $h \in (0,1]$ which satisfy

$$(P_h - 1)E_h = 0.$$

We will furthermore assume that these generalized eigenfunctions may be decomposed as follows.

Hypothesis 43. *We suppose that E_h can be put in the form*

$$E_h = E_h^0 + E_h^1, \tag{44}$$

where E_h^0 is a tempered distribution which is a Lagrangian state associated to a Lagrangian manifold which satisfies Hypothesis 13 of invariance, as well as Hypothesis 16 of transversality, and where E_h^1 is a tempered distribution such that for each $\rho \in \mathrm{WF}_h(E_h^1)$, we have $\rho \in \mathcal{E}$.

Furthermore, we suppose that E_h^1 *is* ***outgoing*** *in the sense that there exists* $\epsilon_2 > 0$ *such that for all* $\chi, \chi' \in C_c^\infty$ *such that* $\chi \equiv 1$ *on* $\{x \in X : b(x) \geq \epsilon_2\}$*, there exists* $T_\chi > 0$ *such that for all* $t \geq T_\chi$*, we have*

$$\Phi^t\big(\mathrm{WF}((1-\chi)\chi' E_h^1)\big) \cap \mathrm{spt}(\chi) = \varnothing. \tag{45}$$

The most natural example of such generalized eigenfunctions is given by distorted plane waves, which we are now going to define. Note that they depend on a parameter $\xi \in \partial\overline{X}$, so that they actually form a whole family of generalized eigenfunctions.

It is also possible to define generalized eigenfunctions which satisfy Hypothesis 43 on manifolds which are hyperbolic near infinity. This is done in [Ingremeau 2017, Appendix B]; the construction mainly follows [Dyatlov and Guillarmou 2014, §6], but some work has to be done to check that E_h^1 is a tempered distribution.

3.3. *Distorted plane waves on Euclidean-near-infinity manifolds.*

Definition 44. We say that X is *Euclidean near infinity* if there exists a compact set $X_0 \subset X$ and a $R_0 > 0$ such that $X \backslash X_0$ has finitely many connected components, which we denote by $X_1, \dots, X_l$, such that for each $1 \leq i \leq l$, we have (X_i, g) is isometric to $(\mathbb{R}^d \backslash B(0, R_0), g_{\mathrm{Eucl}})$.

The surface in Figure 2 is an example of a Euclidean-near-infinity manifold. We may assume that $\operatorname{supp} V \subset X_0$. Also, any Euclidean-near-infinity manifold fulfills Hypothesis 5. Indeed, we may take a boundary-defining function b such that $b(x) = (1+|x|^2)^{-1/2}$ if $x \in X_i$ which we identify with $\mathbb{R}^d \backslash B(0, R_0)$.

To define distorted plane waves, we will simply give a definition of each of the two terms which compose them as in (44).

3.3.1. *Definition of* E_h^0. By definition of a Euclidean-near-infinity manifold, we have

$$X = X_0 \sqcup \left(\bigsqcup_{i=1}^N X_i\right)$$

with X_0 compact, and for each $1 \leq i \leq N$, there exists an isometric isomorphism

$$x_i : X_i \to \mathbb{R}^d \backslash B(0, R_0), \tag{46}$$

equipped with the Euclidean metric g_0.

The boundary of $\overline{X}$ may then be identified with a union of spheres:

$$\partial X \cong \bigsqcup_{i=1}^N S_i,$$

with $S_i \cong \mathbb{S}^n$.

Let $\xi \in \partial\overline{X}$. We have $\xi \in S_i$ for some $1 \leq i \leq m$. Take a smooth function $\tilde{\chi} : X \to [0, 1]$ which vanishes outside of X_i, and which is equal to 1 in a neighbourhood of S_i.

We define the incoming wave E_h^0 by $E_h^0(\xi, \cdot) : X \to \mathbb{C}$ by

$$E_h^0(\xi, x) = \begin{cases} \tilde{\chi}(x) e^{\frac{i}{h} x_i(x)\cdot\xi} & \text{if } x \in X_i, \\ 0 & \text{otherwise.} \end{cases}$$

If we write $\mathcal{L}_0$ for the Lagrangian submanifold (with boundaries) $X_i \times \{\xi\} \subset T^*X$, then E_h^0 is a Lagrangian distribution associated to $\mathcal{L}_0$, which satisfies Hypothesis 13 of invariance.

3.3.2. *Definition of the distorted plane waves.* Let us set

$$F_h := -[P_h, \tilde{\chi}]E_h^0(\xi).$$

Note that we have $F_h \in S^{\mathrm{comp}}(X)$.

Recall that the outgoing resolvent $R_h(1)$ is defined as $R_h(1) := \lim_{\epsilon\to 0^+}(P_h - (1+i\epsilon)^2)^{-1}$, the limit being taken in the topology of bounded operators from $L^2_{\mathrm{comp}}(X)$ to $L^2_{\mathrm{loc}}(X)$.

We shall use the following resolvent estimate, which was proved in [Nonnenmacher and Zworski 2009].

Theorem 45 (resolvent estimates for Euclidean-near-infinity manifolds). *Let X be a Euclidean-near-infinity manifold such that Hypothesis 10 on hyperbolicity and Hypothesis 46 on topological pressure hold. Then for any $\chi \in C_c^\infty(X)$, there exists $C > 0$ such that for all $0 < h < h_0$, we have*

$$\|\chi R_h(1)\chi\|_{L^2(X)\to L^2(X)} \le C\frac{\log(1/h)}{h}. \tag{47}$$

We define

$$E_h^1 := R_h(1)F_h,$$

which is a tempered distribution thanks to Theorem 45.

We then define the distorted plane wave as

$$E_h^\xi := E_h^0 + E_h^1.$$

To check the outgoing assumption on E_h^1, we must explain why there exists $\epsilon_2 > 0$ such that for all $\chi, \chi' \in C_c^\infty$ with $\chi \equiv 1$ on $\{x \in X : b(x) \ge \epsilon_2\}$, there exists $T_\chi > 0$ such that for all $t > T_\chi$, we have

$$\Phi^t\big(\mathrm{WF}((1-\chi)\chi' E_h^1)\big) \cap \mathrm{spt}(\chi) = \varnothing. \tag{48}$$

From [Dyatlov and Guillarmou 2014, §6.2], we know that for any $\rho \in \mathrm{WF}_h(E_h^1)$, we have $\rho \in \mathcal{E}$, and either $\rho \in \Gamma^+$ or there exists a $t > 0$ such that $\Phi^{-t}(\rho) = (x,\xi)$ where $x \in \mathrm{spt}(\partial\tilde{\chi})$, where $\tilde{\chi}$ is as in Section 3.3.1.

We may take $\epsilon_2 < \epsilon_0$ small enough so that $\mathrm{spt}(\tilde{\chi}) \subset \{x \in X : b(x) > \epsilon_2\}$. Suppose that $\rho = (x,\xi)$ is such that $x \in \mathrm{spt}(1-\chi)$ and $\pi_X(\Phi^t(\rho)) \in \mathrm{spt}(\chi)$. Then, by geodesic convexity, $(x,-\xi) \in \mathcal{DE}_+$. Therefore, since $\mathrm{spt}(\tilde{\chi}) \subset \{x \in X : b(x) > \epsilon_2\}$ and $\mathrm{spt}(1-\chi) \subset \{x \in X : b(x) < \epsilon_2\}$ and since b decreases in the future along the trajectory of $(x,-\xi)$, it is impossible that there exists $t > 0$ such that $\Phi^{-t}(\rho) = (x,\xi)$ where $x \in \mathrm{spt}(\partial\tilde{\chi})$. Therefore, if $\rho \in \Phi^t\big(\mathrm{WF}((1-\chi)\chi' E_h^1)\big) \cap \mathrm{spt}(\chi)$, we must have $\rho \in \mathcal{DE}_+$.

On the other hand, if $\rho \in \mathcal{DE}_+$, then (48) is always satisfied as long as T_χ is large enough so that $\Phi^{T_\chi}\big(\mathcal{DE}_+ \cap T^*(\mathrm{spt}(1-\chi))\big) \cap T^*\mathrm{spt}(\chi) = \varnothing$. This shows that E_h^1 is outgoing.

Finally, one readily checks that we have, in the sense of PDEs,

$$(P_h - 1)E_h^\xi = 0.$$

We will sometimes simply write E_h instead of E_h^ξ, to avoid cumbersome notations.

The definition of E_h seems to depend on the choices of the cut-off functions we made. Actually, the distorted plane waves can be defined in a much more intrinsic fashion, using the structure of the resolvent at infinity. We don't want to enter into the details here (see [Dyatlov and Guillarmou 2014, §6; Melrose 1995, Chapter 2]).

3.4. *Topological pressure.* We shall now give a definition of topological pressure, so as to formulate Hypothesis 46. Recall that the distance d was defined in Section 2.1.2, and that it was associated to the adapted metric. We say that a set $\mathcal{S} \subset K$ is (ϵ, t)-separated if for $\rho_1, \rho_2 \in \mathcal{S}$, $\rho_1 \neq \rho_2$, we have $d(\Phi^{t'}(\rho_1), \Phi^{t'}(\rho_2)) > \epsilon$ for some $0 \leq t \leq t'$. (Such a set is necessarily finite.)

The metric g_{ad} induces a volume form Ω on any d-dimensional subspace of $T(T^*\mathbb{R}^d)$. Using this volume form, we will define the unstable Jacobian on K. For any $\rho \in K$, the determinant map

$$\Lambda^n d\Phi^t(\rho)|_{E_\rho^{+0}} : \Lambda^n E_\rho^{+0} \to \Lambda^n E_{\Phi^t(\rho)}^{+0}$$

can be identified with the real number

$$\det(d\Phi^t(\rho)|_{E_\rho^{+0}}) := \frac{\Omega_{\Phi^t(\rho)}(d\Phi^t v_1 \wedge d\Phi^t v_2 \wedge \cdots \wedge d\Phi^t v_n)}{\Omega_\rho(v_1 \wedge v_2 \wedge \cdots \wedge v_n)},$$

where $(v_1, \dots, v_n)$ can be any basis of E_ρ^{+0}. This number defines the unstable Jacobian:

$$\exp \lambda_t^+(\rho) := \det(d\Phi^t(\rho)|_{E_\rho^{+0}}). \tag{49}$$

From there, we take

$$Z_t(\epsilon, s) := \sup_{\mathcal{S}} \sum_{\rho \in \mathcal{S}} \exp(-s\lambda_t^+(\rho)),$$

where the supremum is taken over all (ϵ, t)-separated sets. The pressure is then defined as

$$\mathcal{P}(s) := \lim_{\epsilon \to 0} \limsup_{t \to \infty} \frac{1}{t} \log Z_t(\epsilon, s).$$

This quantity is actually independent of the volume form Ω and of the metric chosen: after taking logarithms, a change in Ω or in the metric will produce a term $O(1)/t$, which is not relevant in the $t \to \infty$ limit.

Hypothesis 46. *We assume the following inequality on the topological pressure associated with Φ^t on K:*

$$\mathcal{P}\left(\tfrac{1}{2}\right) < 0. \tag{50}$$

We will give an equivalent definition of topological pressure in Section 4.1, better suited to our purpose.

3.5. *Statement of the results concerning distorted plane waves.* We may now formulate our main result.

Theorem 47. *Suppose that the manifold X satisfies Hypothesis 5 at infinity, and that the Hamiltonian flow (Φ^t) satisfies Hypothesis 10 on hyperbolicity and Hypothesis 46 concerning the topological pressure. Let E_h be a generalized eigenfunction of the form described in Hypothesis 43, where E_h^0 is associated to a Lagrangian manifold $\mathcal{L}_0$ which satisfies the invariance Hypothesis 13 as well as the transversality Hypothesis 16.*

Then there exists a finite set of points $(\rho_b)_{b\in B_1}\subset K$ *and a family* $(\Pi_b)_{b\in B_1}$ *of operators in* $\Psi_h^{\mathrm{comp}}(X)$ *microsupported in a small neighbourhood of* ρ_b *such that* $\sum_{b\in B_1}\Pi_b = I$ *microlocally on a neighbourhood of* K *in* T^*X *such that the following holds.*

Let $\mathcal{U}_b : L^2(X)\to L^2(\mathbb{R}^d)$ *be a Fourier integral operator quantizing the symplectic change of local coordinates* $\kappa_b : (x,\xi)\mapsto (y^{\rho_b},\eta^{\rho_b})$, *and which is microlocally unitary on the microsupport of* Π_b.

For any $r>0$, *there exists* $M_r>0$ *such that we have*

$$\mathcal{U}_b\Pi_b E_h(y^{\rho_b}) = \sum_{n=0}^{\lfloor M_{r,\ell}|\log h|\rfloor}\sum_{\beta\in\widetilde{\mathcal{B}}_n} e^{i\phi_{n,\beta,b}(y^{\rho_b})/h}a_{n,\beta,b}(y^{\rho_b};h)+R_r, \tag{51}$$

where $a_{n,\beta,b}\in S^{\mathrm{comp}}(\mathbb{R}^d)$ *are classical symbols, and each* $\phi_{n,\beta,b}$ *is a smooth function independent of* h, *and defined in a neighbourhood of the support of* $a_{n,\beta,b}$. *The set* $\widetilde{\mathcal{B}}_n$ *will be defined in* (85). *Its cardinal behaves like some exponential of* n.

We have the following estimate on the remainder

$$\|R_r\|_{L^2} = O(h^r).$$

For any $\ell\in\mathbb{N}$, $\epsilon>0$, *there exists* $C_{\ell,\epsilon}$ *such that for all* $n\geq 0$ *and for all* $h\in(0,h_0]$, *we have*

$$\sum_{\beta\in\widetilde{\mathcal{B}}_n}\|a_{n,\beta,b}\|_{C^\ell}\leq C_{\ell,\epsilon}e^{n(\mathcal{P}(1/2)+\epsilon)}. \tag{52}$$

Remark 48. This theorem can be considered as a quantum analogue of Theorem 17. Indeed, as we explained in Section 1, we will prove it by describing the evolution of the Schrödinger flow of Lagrangian states, while Theorem 17 described the evolution by the Hamiltonian flow of associated Lagrangian manifolds. Actually, the sets containing the microsupports of the operators $(\Pi_b)_{b\in B_1}$ will be built from the sets $(W_a)_{a\in A_1}$ constructed in Theorem 17, as explained in Section 4.1.

Remark 49. The remainder R_r is compactly microlocalised, since the other two terms in the decomposition (51) are compactly microlocalised. Therefore, for any $\ell\in\mathbb{N}$, by possibly taking M_r larger, we may ask that

$$\|R_r\|_{C^\ell} = O(h^r).$$

Theorem 47 may be used to identify the semiclassical measures associated to our generalized eigenfunctions, as in Theorem 3. We shall do this only microlocally close to the trapped set, since the expression for the semiclassical measure on the whole manifold may become very complicated.

Let us denote by π_b the principal symbol of the operators Π_b introduced in the statement of Theorem 47. The following corollary is a more precise version of (the second part of) Theorem 3.

Corollary 50. *There exists a constant* $0<c\leq 1$ *and functions* $e_{n,\beta,b}$ *for* $n\in\mathbb{N}$, $\beta\in\widetilde{\mathcal{B}}_n$ *and* $b\in B_1$ *such that for any* $a\in C_c^\infty(T^*X)$ *and for any* $\chi\in C_c^\infty(X)$, *we have*

$$\left\langle \mathrm{Op}_h(\pi_b^2 a)\chi E_h,\chi E_h\right\rangle = \int_{T^*X} a(x,v)\,\mathrm{d}\mu_{b,\chi}(x,v)+O(h^c),$$

with

$$d\mu_{b,\chi}(\kappa_b^{-1}(y^{\rho_b},\eta^{\rho_b})) = \sum_{n=0}^{\infty}\sum_{\beta\in\widetilde{\mathcal{B}}_n} e_{n,\beta,b}(y^{\rho_b})\delta_{\{\eta^{\rho_b}=\partial\phi_{j,n}(y^{\rho_b})\}}dy^{\rho_b}.$$

The functions $e_{n,\beta,b}$ satisfy an exponential decay estimate as in (52).

The functions $e_{n,\beta,b}$ will be closely related to $a^0_{n,\beta,b}(y^{\rho_b})$, the principal symbol of $a_{n,\beta,b}(y^{\rho_b})$ appearing in (51). Actually, $e_{n,\beta,b}(y^{\rho_b})$ will either be the square of the modulus of $a^0_{n,\beta,b}(y^{\rho_b})$, or the square of the modulus of the sum of a finite number of $a^0_{n,\beta,b}(y^{\rho_b})$ for different values of n and β. These different terms will come from the fact that a point may belong to $\Phi^{n,t_0}_\beta(\mathcal{L}_0)$ for several values of n,β.

3.6. *Strategy of proof.* To study the asymptotic behaviour of the distorted plane wave as h goes to zero, we would like to write that $\widetilde{U}(t)E_h = E_h$, where $\widetilde{U}(t) := e^{it/h}U(t)$. However, this equation can only be formal, because $E_h \notin L^2(X)$. Instead, we use [Dyatlov and Guillarmou 2014, Lemma 3.10]:

Lemma 51. *Let $\chi\in C_c^\infty(X)$. Take $t\in\mathbb{R}$, and a cut-off function $\chi_t\in C_c^\infty(X)$ supported in the interior of a compact set K_t, such that*

$$d_g(\operatorname{supp}\chi, \operatorname{supp}(1-\chi_t)) > 2|t|,$$

where d_g denotes the Riemannian distance on M. Then, for any $\xi\in\mathbb{S}^d$, we have

$$\chi E_h = \chi\widetilde{U}(t)\chi_t E_h + O(h^\infty\|E_h\|_{L^2(K_t)}). \tag{53}$$

Since E_h is a tempered distribution by assumption, we have, for any $t>0$ and $\chi\in C_c^\infty(X)$,

$$\|\chi E_h - \chi\widetilde{U}(t)\chi_t E_h\|_{L^2} = O(h^\infty),$$

where χ_t is as in Lemma 51.

We may then iterate this equation as follows: we write that $\chi_t = \chi + \chi_t(1-\chi)$, and obtain

$$\chi E_h = \chi\widetilde{U}(t)((1-\chi)\chi_t)E_h + \chi\widetilde{U}(t)\chi\widetilde{U}(t)\chi_t E_h + O(h^\infty).$$

We may iterate this method to times $Nt \le Mt|\log h|$ for any given $M>0$. We obtain

$$\chi E_h = (\chi\widetilde{U}(t))^N\chi_t E_h + \sum_{k=1}^{N}(\chi\widetilde{U}(t))^k(1-\chi)\chi_t E_h + O(h^\infty). \tag{54}$$

Now, choose $\chi\in C_c^\infty(X)$ as in Hypothesis 43, and take $t>T_\chi$.

Lemma 52. *Let $t>T_\chi$, $M>0$, and $\chi\in C_c^\infty(X)$ be such that $\chi\equiv 1$ on $\{x\in X : b(b)>\epsilon_2\}$, where $\epsilon_2<\epsilon_0$ is as in Hypothesis 43. For any $k\le M|\log h|$, we have*

$$\|(\chi\widetilde{U}(t))^k(1-\chi)\chi_t E_h^1\|_{L^2} = O(h^\infty).$$

Proof. We only have to prove that $\|(\chi\widetilde{U}(t))(1-\chi)\chi_t E_h^1\|_{L^2} = O(h^\infty)$. This is a consequence of (45) in Hypothesis 43. ∎

Therefore, we have for any $\chi \in C_c^\infty(X)$ as in Lemma 52,

$$\chi E_h = (\chi \widetilde{U}(t))^N \chi_t E_h^0 + (\chi \widetilde{U}(t))^N \chi_t E_h^1 + \sum_{k=1}^{N} (\chi \widetilde{U}(t))^k (1-\chi)\chi_t E_h^0 + O(h^\infty). \tag{55}$$

Let us now introduce tools from [Nonnenmacher and Zworski 2009] to analyse these terms in more detail.

4. Tools for the proofs of Theorem 47

4.1. ***Another definition of topological pressure.*** Recall that $\mathcal{E}_{\boldsymbol{E}}$ and $K_{\boldsymbol{E}}$ were defined in (10) and (11) respectively. For any $\delta > 0$ small enough so that (12) holds, we define

$$\mathcal{E}^\delta := \bigcup_{|\boldsymbol{E}-1|<\delta} \mathcal{E}_{\boldsymbol{E}}, \quad K^\delta := \bigcup_{|\boldsymbol{E}-1|<\delta} K_{\boldsymbol{E}}.$$

Let $\mathcal{W} = (W_a)_{a\in A_1}$ be a finite open cover of $K^{\delta/2}$ such that the W_a are all strictly included in $\mathcal{E}^\delta$ and of diameter $< \varepsilon_0$, where ε_0 comes from Theorem 17. For any $T \in \mathbb{N}^*$, define $W(T) := (W_\alpha)_{\alpha \in A_1^T}$ by

$$W_\alpha := \bigcap_{k=0}^{T-1} \Phi^{-k}(W_{a_k}),$$

where $\alpha = a_0, \dots, a_{T-1}$. Let $\mathcal{A}'_T$ be the set of $\alpha \in A_1^T$ such that $W_\alpha \cap K^\delta \neq \varnothing$. If $V \subset \mathcal{E}^\delta$, $V \cap K^{\delta/2} \neq \varnothing$, define

$$S_T(V) := -\inf_{\rho\in V\cap K^{\delta/2}} \lambda_T^+(\rho), \quad \text{with } \lambda_T^+ \text{ as in (49)},$$

$$Z_T(\mathcal{W}, s) := \inf\Big\{ \sum_{\alpha\in\mathcal{A}_T} \exp\{sS_T(W_\alpha)\} : \mathcal{A}_T \subset \mathcal{A}'_T,\ K^{\delta/2} \subset \bigcup_{\alpha\in\mathcal{A}_T} W_\alpha \Big\},$$

$$\mathcal{P}^\delta(s) := \lim_{\operatorname{diam}\mathcal{W}\to 0} \lim_{T\to\infty} \frac{1}{T} \log Z_T(\mathcal{W}, s).$$

The topological pressure is then

$$\mathcal{P}(s) = \lim_{\delta\to 0} \mathcal{P}^\delta(s). \tag{56}$$

Recall that we assumed that

$$\mathcal{P}\big(\tfrac{1}{2}\big) < 0.$$

Let us fix $\epsilon_0 > 0$ so that $\mathcal{P}\big(\tfrac{1}{2}\big) + 2\epsilon_0 < 0$. Then there exists $t_0 > 0$, and $\widehat{\mathcal{W}}$ an open cover of K^δ with $\operatorname{diam}(\widehat{\mathcal{W}}) < \varepsilon_0$ such that

$$\Big| \frac{1}{t_0} \log Z_{t_0}(\widehat{\mathcal{W}}, s) - \mathcal{P}^\delta(s) \Big| \le \epsilon_0. \tag{57}$$

We can find $\mathcal{A}_{t_0}$ so that $\{W_\alpha : \alpha \in \mathcal{A}_{t_0}\}$ is an open cover of K^δ in $\mathcal{E}^\delta$ and such that

$$\sum_{\alpha\in\mathcal{A}_{t_0}} \exp\{sS_{t_0}(W_\alpha)\} \le \exp\{t_0(\mathcal{P}^\delta(s) + \epsilon_0)\}.$$

Therefore, if we take δ small enough, and if we rename $\{W_\alpha : \alpha \in \mathcal{A}_{t_0}\}$ as $\{V_b : b \in B_1\}$, we have

$$\sum_{b\in B_1} \exp\{\tfrac{1}{2} S_{t_0}(V_b)\} \le \exp\{t_0(\mathcal{P}(\tfrac{1}{2}) + 2\epsilon_0)\}. \tag{58}$$

By taking t_0 large enough, we can assume that $\log(1+\epsilon_0) + t_0(\mathcal{P}(1/2)+\epsilon_0) < 0$.

A new open cover of $\mathcal{E}$. By hypothesis, the diameter of $\widehat{\mathcal{W}}$ in (57) is smaller than ε_0, so that we may apply Theorem 17 to it. We complete it into an open cover $(W_a)_{a\in A}$ as in Theorem 17, and if $\alpha \in A^N$ for some $N \ge 0$, we define as previously $W_\alpha := \bigcap_{k=0}^{N-1} \Phi^{-k}(W_{a_k})$.

Let us rewrite as $(V_b)_{b\in B_2}$ the sets $(W_\alpha)_{\alpha\in A^{t_0}}$, where $\alpha \in A^{t_0}\backslash\mathcal{A}_{t_0}$ such that $\alpha_k \neq 0$ for some $0 \le k \le t_0 - 1$. We will also write V_0 for the set $W_{0,0,\dots,0}$.

If we write $B = B_1 \sqcup B_2 \sqcup \{0\}$, the sets $(V_b)_{b\in B}$ form an open cover of $\mathcal{E}$ in T^*X.

Actually, by compactness of the interaction region, we may find a $\delta > \delta' > 0$ small enough so that (12) holds and such that, by replacing V_0 by $V_0 \cap \mathcal{E}^\delta$, the sets $(V_b)_{b\in B}$ form an open cover of $\mathcal{E}^{\delta'}$ included in $\mathcal{E}^\delta$.

If $\beta = b_0 \cdots b_{N-1} \in B^N$ for some $N \in \mathbb{N}$, and if Λ is a Lagrangian manifold, we will define for each $0 \le k \le N-1$ the set $\Phi_\beta^{k,t_0}(\Lambda)$ by

$$\Phi_\beta^{0,t_0}(\Lambda) = \Lambda \cap V_{b_0},$$
$$\Phi_\beta^{k,t_0}(\Lambda) = \Phi^{t_0}(V_{b_k} \cap \Phi_\beta^{k-1,t_0}(\Lambda)) \quad \text{for } 1 \le k \le N-1.$$

By the definition of the sets $b \in B$, we have $\Phi_\beta^{N,t_0}(\Lambda) = \Phi_{\alpha_\beta}^{Nt_0}(\Lambda)$, where $\alpha_\beta \in A^{Nt_0}$ is the concatenation of all the sequences which make up the b_k, $0 \le k \le N-1$.

Therefore, once we have fixed a point $\rho^b \in K \cap V_b$ for each $b \in B_1$, we have the following analogue of Theorem 17.

Corollary 53. *If there exists $N'_{\text{uns}} \in \mathbb{N}$ such that for all $N \in \mathbb{N}$, for all $\beta \in B^N$ and all $b \in B_1$, then $V_b \cap \Phi_\beta^{N,t_0}(\mathcal{L}_0)$ is either empty, or is a Lagrangian manifold in some unstable cone in the coordinates $(y^{\rho_b}, \eta^{\rho_b})$.*

Furthermore, if $N - \tau(\beta) \ge N'_{\text{uns}}$, then $V_b \cap \Phi_\beta^N(\mathcal{L}_0)$ is a γ_{uns}-unstable Lagrangian manifold in the coordinates $(y^{\rho_b}, \eta^{\rho_b})$.

Remark 54 (new definition of the sets $(V_b)_{b\in B_1}$). The sets $(V_b)_{b\in B_1}$ form an open cover of K. By compactness, they form an open cover of $\{\rho \in \mathcal{E} : d(\rho, K) \le \epsilon_3\}$ for some $\epsilon_3 > 0$. Hence, if for each $b \in B_2$ we replace each V_b by $V_b \cap \{\rho \in \mathcal{E} : d(\rho, K) > \epsilon_3/2\}$, which we still denote by V_b, the sets $(V_b)_{b\in B}$ still form an open cover of $\mathcal{E}$, and the conclusions of Corollary 53 do still apply.

By adapting the proof of Lemma 24, we see that by possibly enlarging N'_{uns}, we may suppose that for all $b \in B_2$ and for all $\rho \in V_b$, we have $\Phi^{N'_{\text{uns}}t_0}(\rho) \in V_0\backslash\big(\bigcup_{b\in B_2} V_b\big)$ or $\Phi^{-N'_{\text{uns}}t_0}(\rho) \in V_0\backslash\big(\bigcup_{b\in B_2} V_b\big)$.

Note also that thanks to Lemma 9, for any $b \in B_1 \cup B_2$ and for any $k \ge 1$, we have $\Phi^{kt_0}(\mathcal{L}_0 \cap V_b)) \cap \mathcal{W}_0 \cap \mathcal{DE}_- = \varnothing$.

Remark 55. In [Nonnenmacher and Zworski 2009, Proposition 5.2] the authors proved the following statement. There exists a $\gamma_1 > 0$ such that the following holds. Let $b, b' \in B_1$, and let Λ be a Lagrangian manifold contained in V_b, γ-unstable in the coordinates $(y^{\rho_b}, \eta^{\rho_b})$ for some $\gamma \le \gamma_1$. Then $\Phi^{t_0}(\Lambda) \cap V_{b'}$ is also a Lagrangian manifold which is γ-unstable in the coordinates $(y^{\rho_{b'}}, \eta^{\rho_{b'}})$.

Furthermore, the map $y^{\rho_b} \mapsto y^{\rho_{b'}}$ obtained by projecting $\Phi^{t_0}|_\Lambda$ onto the planes $\{(y^{\rho_b}, \eta^{\rho_b}) : \eta^{\rho_b} = 0\}$ and $\{(y^{\rho_{b'}}, \eta^{\rho_{b'}}) : \eta^{\rho_{b'}} = 0\}$ satisfies the following estimate on its domain of definition:

$$\det\left(\frac{\partial y^{\rho_{b'}}}{\partial y^{\rho_b}}\right) = (1 + O(\epsilon^{\mathrm{p}}))e^{\lambda^+_{t_0}(\rho_b)},$$

where $\lambda^+_{t_0}(\rho_b)$ is the unstable Jacobian of ρ_b, defined in (49).

In the sequel, we will always suppose that $\gamma_{\mathrm{uns}} < \gamma_1$.

For each $b \in B_1$, we will denote by $\mathcal{U}_b$ a Fourier integral operator quantizing the local change of symplectic coordinates $(x, \xi) \mapsto (y^{\rho_b}, \eta^{\rho_b})$.

4.2. ***The Schrödinger propagator as a Fourier integral operator.*** Let us explain how the formalism of Section 3.1.3 may be used to describe the Schrödinger propagator $U(t)$ acting on $L^2(X)$. We shall state a lemma proven in [Nonnenmacher and Zworski 2009, Lemma 4.2]. Recall that for $0 < \delta < 1$, we defined $\mathcal{E}^\delta$ as $\bigcup_{|\boldsymbol{E}-1|<\delta} \mathcal{E}_{\boldsymbol{E}}$.

Lemma 56. *Let $V_0 \Subset \mathcal{E}^\delta$, $V_1 \subset \Phi^t(V_0)$ for some $t > 0$. Take some $\rho_0 \in V_0 \cap \mathcal{E}$ and set $\rho_1 = \Phi^t(\rho_0) \in V_1$. Let $f_j : \pi(V_j) \to \mathbb{R}^d$, $j = 0, 1$ be local coordinates such that $f_0(\pi(\rho_0)) = f_1(\pi(\rho_1)) = 0 \in \mathbb{R}^d$. They induce on V_0 and V_1 the symplectic coordinates*

$$F_j(x, \xi) := \big(f_j(x), (df_j(x)^t)^{-1}\xi - \xi^{(j)}\big), \quad j = 0, 1,$$

where $\xi^{(j)} \in \mathbb{R}^d$ is fixed by the condition $F_j(\rho_j) = (0, 0)$. Then the operator on $L^2(\mathbb{R}^d)$,

$$T(t) := e^{-i\langle x, \xi^{(1)}\rangle/h}(f_1^{-1})^* U(t)(f_0)^* e^{i\langle x, \xi^{(0)}\rangle/h},$$

is of the form (42) for some choice of the A_j microlocally near $(0, 0) \times (0, 0)$.

4.3. ***Iterations of Fourier integral operators.*** We recall here the main results from [Nonnenmacher and Zworski 2009, §4] concerning the iterations of semiclassical Fourier integral operators in $T^*\mathbb{R}^d$.

Let $V \subset T^*\mathbb{R}^d$ be an open neighbourhood of 0, and take a sequence of symplectomorphisms $(\kappa_i)_{i=1,\dots,N}$ from V to $T^*\mathbb{R}^d$ such that for all $i \in \{1, \dots, N\}$, we have $\kappa_i(0) \in V$, and the projection

$$(x_1, \xi_1; x_0, \xi_0) \mapsto (x_1, \xi_0), \quad \text{where } (x_1, \xi_1) = \kappa(x_0, \xi_0),$$

is a diffeomorphism close to the origin. We consider Fourier integral operators (T_i) which quantise κ_i and which are microlocally unitary near an open set $U \times U$, where $U \Subset V$, which contains the origin. Let $\Omega \subset \mathbb{R}^d$ be an open set such that $U \Subset T^*\Omega$, and, for all i, we have $\kappa_i(U) \Subset T^*\Omega$. For each i, we take a smooth cut-off function $\chi_i \in C_c^\infty(U; [0, 1])$, and let

$$S_i := \mathrm{Op}_h(\chi_i) \circ T_i. \tag{59}$$

Let us consider a family of Lagrangian manifolds $\Lambda_k = \{(x, \phi'_k(x)) : x \in \Omega\} \subset T^*\mathbb{R}^d,\ k = 0, \dots, N$, such that

$$|\partial^\alpha \phi_k| \le C_\alpha, \quad 0 \le k \le N,\ \alpha \in \mathbb{N}^d. \tag{60}$$

We assume that there exists a sequence of integers $(i_k \in \{1, \dots, J\})_{k=1,\dots,N}$ such that

$$\kappa_{i_{k+1}}(\Lambda_k \cap U) \subset \Lambda_{k+1}, \quad k = 0, \dots, N-1.$$

We define g_k by

$$g_k(x) = \pi \circ \kappa_{i_k}^{-1}(x, \phi'_k(x)).$$

That is to say, $\kappa_{i_k}^{-1}(x, \phi'_k(x)) = (g_k(x), \phi'_{k-1}(g_k(x)))$.

We will say that a point $x \in \Omega$ is N-admissible if we can define recursively a sequence by $x^N = x$, and, for $k = N, \dots, 1$, we have $x^{k-1} = g_k(x^k)$. This procedure is possible if, for any k, we have x^k is in the domain of definition of g_k.

Let us assume that, for any admissible sequence $(x^N \cdots x^0)$, the Jacobian matrices are uniformly bounded from above:

$$\left\| \frac{\partial x^k}{\partial x^l} \right\| = \left\| \frac{\partial (g_{k+1} \circ g_{k+2} \circ \cdots \circ g_l)}{\partial x^l}(x^l) \right\| \le C_D, \quad 0 \le k < l \le N,$$

where C_D is independent of N. This assumption roughly says that the maps g_k are (weakly) contracting.

We will also use the notation

$$D_k := \sup_{x \in \Omega} |\det dg_k(x)|^{1/2}, \quad J_k := \prod_{k'=1}^{k} D_{k'},$$

and assume that the D_k are uniformly bounded: $1/C_D \le D_k \le C_D$.

The following result can be found in [Nonnenmacher and Zworski 2009, Proposition 4.1].

Proposition 57. *We use the above definitions and assumptions, and take N arbitrarily large, possibly varying with h. Take any $a \in S^{\mathrm{comp}}$ and consider the Lagrangian state $u = a e^{i\phi_0/h}$ associated with the Lagrangian Λ_0. Then we may write*

$$(S_{i_N} \circ \cdots \circ S_{i_1})(a e^{i\phi_0/h})(x) = e^{i\phi_N(x)/h} \left(\sum_{j=0}^{L-1} h^j a_j^N(x) + h^L R_L^N(x, h) \right),$$

where each $a_j^N \in C_c^\infty(\Omega)$ depends on h only through N, and $R_L^N \in C^\infty((0,1]_h, \mathcal{S}(\mathbb{R}^d))$. If $x^N \in \Omega$ is N-admissible, and defines a sequence $(x^k),\ k = N, \dots, 1$, then

$$|a_0^N(x^N)| = \left(\prod_{k=1}^{N} \chi_{i_k}(x^k, \phi'_k(x^k)) |\det dg_k(x^k)|^{1/2} \right) |a(x^0)|;$$

otherwise $a_j^N(x^N) = 0,\ j = 0, \dots, L-1$. We also have the bounds

$$\|a_j^N\|_{C^\ell(\Omega)} \le C_{j,\ell} J_N (N+1)^{\ell+3j} \|a\|_{C^{\ell+2j}(\Omega)}, \quad j = 0, \dots, L-1,\ \ell \in \mathbb{N}, \tag{61}$$

$$\|R_L^N\|_{L^2(\mathbb{R}^d)} \leq C_L \|a\|_{C^{2L+d}(\Omega)} (1+C_0 h)^N \sum_{k=1}^{N} J_k k^{3L+d}, \tag{62}$$

$$\|R_L^N\|_{C^\ell(\mathbb{R}^d)} \leq C_{L,l} h^{-d/2-\ell} \|a\|_{C^{2L+d}(\Omega)} (1+C_0 h)^N \sum_{k=1}^{N} J_k k^{3L+d}. \tag{63}$$

The constants $C_{j,\ell}$, C_0 and C_L depend on the constants in (60) and on the operators $\{S_j\}_{j=1}^J$.

We shall mainly be using this proposition in the case where for all k, we have $D_k \leq \nu < 1$. In this case, the estimates (61), (62) and (63) imply that for any $\ell \in \mathbb{N}$, there exists C_ℓ independent of N such that for any $N \in \mathbb{N}$, we have

$$\|a^N\|_{C^\ell} \leq \|a_0^N\|_{C^\ell}(1+C_\ell h). \tag{64}$$

4.4. *Microlocal partition.* We take a partition of unity $\sum_{b\in B} \pi_b$ such that

$$\sum_{b\in B} \pi_b(x) \equiv 1 \quad \text{for all } x \in \mathcal{E}^{\delta'},$$

and $\mathrm{supp}(\pi_b) \subset V_b \subset \mathcal{E}^\delta$ for all $b \in B$.

For $b \in B_1 \cup B_2$, we set $\Pi_b := \mathrm{Op}_h(\pi_b)$. We have

$$\mathrm{WF}_h(\Pi_b) \subset V_b \cap \mathcal{E}^\delta \quad \text{and} \quad \Pi_b = \Pi_b^*.$$

We then set

$$\Pi_0 := \mathrm{Id} - \sum_{b\in B_1\cup B_2} \Pi_b.$$

We can decompose the propagator at time t_0 into

$$\widetilde{U}(t_0) = \sum_{b\in B} \widetilde{U}_b, \quad \text{where } \widetilde{U}_b := \Pi_b e^{it_0/h} U(t_0).$$

The propagator at time Nt_0 may then be decomposed as

$$\widetilde{U}(Nt_0) = \sum_{\beta\in B^N} \widetilde{U}_\beta, \tag{65}$$

where $\widetilde{U}_\beta := \widetilde{U}_{\beta_{N-1}} \circ \cdots \circ \widetilde{U}_{\beta_0}$.

4.5. *Hyperbolic dispersion estimates.* We will use the following hyperbolic dispersion estimate, coming from [Nonnenmacher and Zworski 2009, Proposition 6.3], the proof of which can be found in Section 7 of that paper.

Lemma 58 (hyperbolic dispersion estimate). *Let $M > 0$ be fixed. There exists an $h_0 > 0$ and a $C > 0$ such that for any $0 < h < h_0$, for any $N < M \log(1/h)$ and for any $\beta \in B_1^N$, we have*

$$\|\widetilde{U}_\beta\|_{L^2\to L^2} \leq C h^{-d/2} (1+\epsilon_0)^N \prod_{j=1}^{N} \exp\left[\tfrac{1}{2} S_{t_0}(V_{\beta_j})\right]. \tag{66}$$

5. Proof of Theorem 47

Proof. Having introduced these different tools, we may now come back to the proof of Theorem 47.

5.1. *Decomposition of χE_h.* Let $\chi \in C_c^\infty(X)$ be as in Lemma 52. We may suppose $T_\chi \le t_0$. Then, by equation (55), we have

$$\chi E_h = (\chi \tilde{U}(t_0))^N \chi_{t_0} E_h + \sum_{k=1}^{N} (\chi \tilde{U}(t_0))^k (1-\chi)\chi_{t_0} E_h^0 + O(h^\infty), \tag{67}$$

where the cut-off function $\chi_{t_0} \in C_c^\infty(X)$ is such that

$$d_X(\operatorname{supp}\chi, \operatorname{supp}(1-\chi_{t_0})) > 2|t_0|,$$

where d_X denotes the Riemannian distance on X.

We shall require the following lemma. The proof of (i) is the same as that of Lemma 24, while the proof of (ii) essentially follows from point (3) of Hypothesis 5.

Lemma 59. (i) *There exists $N_\chi \in \mathbb{N}$ such that for any $N \in \mathbb{N}$ if $\rho \in \operatorname{supp}(\chi_{t_0})$ and $\Phi^N(\rho) \in \operatorname{supp}(\chi)$, then for any $N_\chi \le k \le N - N_\chi$, we have $\Phi^{kt_0}(\rho) \in V_b$ for some $b \in B_1 \cup B_2$.*

(ii) *If $\rho \in \mathcal{E}$ is such that $\Phi^{kt_0}(\rho) \in V_0$ for some $k \in \mathbb{N}$, but $\Phi^{(k+1)t_0}(\rho) \in V_b$ for some $b \in B_1 \cup B_2$, then $\Phi^{k'}(\rho)$ is in $\mathcal{DE}_-$ (and hence in V_0) for any $k' \le k$.*

From Lemma 59, we deduce that for any $k \ge 2N_\chi + 2$, we have

$$(\chi \tilde{U}(t_0))^k = \sum_{l=0}^{N_\chi+1} (\chi \tilde{U}(t_0))^{N_\chi+1} \Big(\sum_{\beta \in B^{k-2N_\chi-2+l}} \tilde{U}_\beta \Big) (\chi \tilde{U}_0)^{N_\chi - l} + O_{L^2\to L^2}(h^\infty). \tag{68}$$

For any $N \in \mathbb{N}\backslash\{0\}$, define $\mathcal{B}_N \subset (B_1 \cup B_2)^N$ by

$$\mathcal{B}_N := \begin{cases} (B_1\cup B_2)^N & \text{if } N \le 2N'_{\text{uns}} + 2, \\ (B_1\cup B_2)^{N'_{\text{uns}}+1} B_1^{N-2N'_{\text{uns}}-2} (B_1\cup B_2)^{N'_{\text{uns}}+1} & \text{otherwise.} \end{cases} \tag{69}$$

Lemma 60. *For any $N \ge 2N'_{\text{uns}} + 2$ and for any $\beta \in (B_1\cup B_2)^N \backslash \mathcal{B}_N$, we have*

$$\|\tilde{U}_\beta\|_{L^2\to L^2} = O(h^\infty).$$

Proof. Let $\beta \in (B_1\cup B_2)^N\backslash\mathcal{B}_N$. Then there exists $N'_{\text{uns}} + 2 \le k \le N - N'_{\text{uns}} + 2$ such that $\beta_k \in B_2$. Recall from Remark 54 that N'_{uns} is such that for any $\rho \in V_{\beta_k}$, we have $\Phi^{N'_{\text{uns}}t_0}(\rho) \in V_0\backslash\big(\bigcup_{b\in B_2} V_b\big)$ or $\Phi^{-N'_{\text{uns}}t_0}(\rho) \in V_0\backslash\big(\bigcup_{b\in B_2} V_b\big)$. The result then follows from Lemma 40. ☐

Equation (68) may then be rewritten as

$$(\chi \tilde{U}(t_0))^k = \sum_{l=0}^{N_\chi+1} (\chi \tilde{U}(t_0))^{N_\chi+1} \Big(\sum_{\beta \in \mathcal{B}_{k-2N_\chi-2+l}} \tilde{U}_\beta \Big) (\chi \tilde{U}_0)^{N_\chi - l} + O_{L^2\to L^2}(h^\infty). \tag{70}$$

By summing over k and reordering the terms, we get, for any $K > 2N_\chi + 3N'_{\text{uns}} + 4$,

$$\sum_{k=0}^{K} (\chi \tilde{U}(t_0))^k = \sum_{n=1}^{\mathfrak{N}} \sum_{l=0}^{N_\chi+1} (\chi \tilde{U}(t_0))^{N_\chi+1} \Big(\sum_{\beta \in \mathcal{B}_{n+RN'_{\text{uns}}+2}} \tilde{U}_\beta \Big) (\chi \tilde{U}_0)^l$$
$$- \sum_{n=K-2N_\chi-2}^{\mathfrak{N}} \sum_{l=0}^{\mathfrak{N}-n} (\chi \tilde{U}(t_0))^{N_\chi+1} \Big(\sum_{\beta \in \mathcal{B}_{n+3N'_{\text{uns}}+2}} \tilde{U}_\beta \Big) (\chi \tilde{U}_0)^l$$
$$+ \sum_{l=0}^{K-\mathfrak{N}-1} (\chi \tilde{U}(t_0))^l + O_{L^2 \to L^2}(h^\infty), \tag{71}$$

where $\mathfrak{N} = K - 3N'_{\text{uns}} - N_\chi - 4$.

Let us note that from Lemma 42 and Hypothesis 13, for each $0 \le l \le N_\chi$, there exists $\chi_l \in S^{\text{comp}}(X)$ such that

$$(\chi \tilde{U}_0)^{N_\chi - l} (1-\chi) \chi_{t_0} E_h^0 = \chi_l E_h^0 + O(h^\infty). \tag{72}$$

Let us introduce the notation

$$\bar{\chi} := \sum_{l=0}^{N_\chi+1} \chi_l. \tag{73}$$

Thanks to equation (71), we can study the different terms in equation (67). The first term in the right-hand side of (67) may be bounded by the following lemma.

Lemma 61. *Let $r > 0$. We may find a constant $M_r \ge 0$ such that for any $M > M_r$ and for any $M_r|\log h| \le N \le M|\log h|$, we have*

$$\|(\chi \tilde{U}(t_0))^N \chi_{t_0} E_h\|_{L^2} = O(h^r).$$

Proof. We use (70), Lemma 58 and the topological pressure assumption to obtain

$$\begin{aligned}
\|(\chi \tilde{U}(t_0))^N \chi_{t_0} E_h\|_{L^2} &\le C \Big\| \sum_{\beta \in \mathcal{B}_{N-2N_\chi-2}} \tilde{U}_\beta \chi_{t_0} E_h \Big\| + O(h^\infty) \\
&\le C \sum_{\beta \in B_1^{N-2N_{\text{uns}}-2N_\chi-4}} \|\tilde{U}_\beta \chi_{t_0} E_h\| \\
&\le C h^{-d/2} (1+\epsilon_0)^N \sum_{\beta \in B_1^{N-2N_{\text{uns}}-2N_\chi-4}} \prod_{j=1}^{N-2N_{\text{uns}}-2} \exp\big[\tfrac{1}{2} S_{t_0}(V_{\beta_j})\big] \|\chi_{t_0} E_h\| \\
&\le C h^{-d/2} (1+\epsilon_0)^N \Big(\sum_{b \in B_1} \exp\big[\tfrac{1}{2} S_{t_0}(V_b)\big] \Big)^N \|\chi_{t_0} E_h\| \\
&\le C h^{-d/2} (1+\epsilon_0)^N \exp\big\{ N t_0 \big(\mathcal{P}\big(\tfrac{1}{2}\big) + 2N\epsilon_0 \big) \big\} \|\chi_{t_0} E_h\|.
\end{aligned}$$

By assumption, E_h is a tempered distribution, so that $\|\chi_{t_0} E_h\|_{L^2} \leq C/h^{r''}$. Therefore

$$\|(\chi \widetilde{U}(t_0))^N \chi_{t_0} E_h\|_{L^2} \leq C h^{-r''-d/2-\epsilon} \exp\{N t_0(\mathcal{P}(\tfrac{1}{2}) + 2N\epsilon_0)\}$$

for some small ϵ. The lemma follows by taking M_r large enough. $\square$

Using Lemma 61, and equation (71), we may rewrite equation (67) as

$$\begin{aligned}\chi E_h = &\sum_{n=1}^{M_r|\log h|} \sum_{l=0}^{N_\chi+1} (\chi \widetilde{U}(t_0))^{N_\chi+1} \Big(\sum_{\beta\in\mathcal{B}_{n+3N'_{\mathrm{uns}}+2}} \widetilde{U}_\beta\Big)(\chi\widetilde{U}_0)^l (1-\chi)\chi_{t_0} E_h^0 \\ &- \sum_{n=M_r|\log h|-N_\chi}^{M_r|\log h|} \sum_{l=0}^{M_r|\log h|-N_\chi-2-n} (\chi \widetilde{U}(t_0))^{N_\chi+1} \Big(\sum_{\beta\in\mathcal{B}_{n+3N'_{\mathrm{uns}}+2}} \widetilde{U}_\beta\Big)(\chi\widetilde{U}_0)^l (1-\chi)\chi_{t_0} E_h^0 \\ &+ \sum_{l=0}^{3N'_{\mathrm{uns}}+N_\chi+3} (\chi \widetilde{U}(t_0))^l (1-\chi)\chi_{t_0} E_h^0 + O_{L^2}(h^r).\end{aligned}$$

The second term may be bounded by $O(h^r)$ thanks to Lemma 61. By using equations (72) and (73), we get

$$\chi E_h = \sum_{n=1}^{M_r|\log h|} (\chi \widetilde{U}(t_0))^{N_\chi+1} \Big(\sum_{\beta\in\mathcal{B}_{n+3N'_{\mathrm{uns}}+2}} \widetilde{U}_\beta\Big) \bar{\chi} E_h^0 + \sum_{l=0}^{3N'_{\mathrm{uns}}+N_\chi+3} (\chi \widetilde{U}(t_0))^l (1-\chi)\chi_{t_0} E_h^0 + O_{L^2}(h^r). \tag{74}$$

5.2. *Evolution of the WKB states.*

5.2.1. *Construction of $\widetilde{\mathcal{B}}_0$.* From now on, we fix $b \in B_1$ and $r > 1$. We may write

$$\mathcal{U}_b \Pi_b \sum_{l=0}^{3N'_{\mathrm{uns}}+N_\chi+3} (\chi \widetilde{U}(t_0))^l (1-\chi)\chi_{t_0} E_h^0 = \sum_{l=0}^{N_\chi+3N'_{\mathrm{uns}}+3} \sum_{\beta\in B^l} \mathcal{U}_b \Pi_b U_\beta^\chi (1-\chi)\chi_{t_0} E_h^0, \tag{75}$$

where we have used the notation

$$U_\beta^\chi = \chi \widetilde{U}_{\beta_l} \chi \cdots \chi \widetilde{U}_{\beta_0}. \tag{76}$$

Note that each of the $\mathcal{U}_b \Pi_b U_\beta^\chi$ is a Fourier integral operator from $L^2(X)$ to $L^2(\mathbb{R}^d)$. Thanks to Corollary 53, we may use Lemma 42 to describe the action of each of these Fourier integral operators on the Lagrangian state $(1-\chi)\chi_{t_0} E_h^0$. If we denote by $\widetilde{\mathcal{B}}_0$ the set $\bigcup_{l=0}^{N_\chi+3N'_{\mathrm{uns}}+3} B^l$, we may write

$$\mathcal{U}_b \Pi_b \sum_{l=0}^{N_\chi+3N'_{\mathrm{uns}}+3} (\chi \widetilde{U}(t_0))^l (1-\chi)\chi_{t_0} E_h^0 = \sum_{\beta\in\widetilde{\mathcal{B}}_0} e_{0,\beta,b}, \tag{77}$$

where $e_{0,\beta,b}(y^b) = e^{\phi_{0,\beta,b}(y^{\rho_b})/h} a_{0,\beta,b}(y^{\rho_b}; h)$, with $a_{0,\beta,b}$ and $\phi_{0,\beta,b}$ as in the statement of Theorem 47.

Let us now consider the other terms on the right-hand side of equation (74), which will be indexed by $\widetilde{\mathcal{B}}_n$, $n \geq 1$.

5.2.2. *Evolution in the intermediate region.* Let $n \geq 1$ and $\beta \in \mathcal{B}_{n+3N'_{\mathrm{uns}}+2}$. By the definition of $\mathcal{B}_{n+3N'_{\mathrm{uns}}+2}$, for $N'_{\mathrm{uns}}+1 \leq i \leq n+2N'_{\mathrm{uns}}+1$, we have $\beta_i \in B_1$.

According to Theorem 17, $\Phi_\beta^{2N'_{\mathrm{uns}}+1,t_0}(\mathcal{L}_0)$ consists of a single Lagrangian manifold, which is γ_{uns}-unstable in the symplectic coordinates in $V_{\beta_{2N'_{\mathrm{uns}}+1}}$.

Thus, we may say that $\widetilde{U}_{\beta_0\cdots\beta_{2N'_{\mathrm{uns}}+1}}(\bar{\chi}E_h^0)$ is a Lagrangian state associated to the Lagrangian manifold $\Phi_\beta^{2N'_{\mathrm{uns}}+1,t_0}(\mathcal{L}_0)$. Thanks to Lemma 56, we may use Lemma 42 to write

$$\big(\mathcal{U}_{\beta_{2N'_{\mathrm{uns}}+1}}\Pi_{\beta_{2N'_{\mathrm{uns}}+1}}\widetilde{U}_{\beta_0\cdots\beta_{2N'_{\mathrm{uns}}+1}}(\bar{\chi}E_h^0)\big)(y^{\rho_{\beta_{2N'_{\mathrm{uns}}+1}}}) = a(y^{\rho_{\beta_{2N'_{\mathrm{uns}}+1}}};h)e^{i\phi(y^{\rho_{\beta_{2N'_{\mathrm{uns}}+1}}})/h}$$

for some $a \in S^{\mathrm{comp}}(\mathbb{R}^d)$.

5.2.3. *Propagation of Lagrangian states close to the trapped set.* To lighten the notations, let us write $\hat{n} := n+2N'_{\mathrm{uns}}+1$.

For each $2N'_{\mathrm{uns}}+1 \leq k \leq \hat{n}$, we write

$$T_{\beta_{k'+1},\beta_{k'}} := \mathcal{U}_{\beta_{k'+1}}\widetilde{U}_{\beta_{k'+1}}\mathcal{U}^*_{\beta'_k}.$$

Now $T_{\beta_{k'+1},\beta_{k'}}$ is an operator quantising the map $\kappa_{\beta_{k'},\beta_{k'+1}}$ obtained by expressing Φ^{t_0} in the coordinates $(y^{\beta_{k'}},\eta^{\beta_{k'}}) \mapsto (y^{\beta_{k'+1}},\eta^{\beta_{k'+1}})$. It is of the form (59).

We will write

$$T_\beta^{2N'_{\mathrm{uns}}+1,\hat{n}} := T_{\beta_{\hat{n}},\beta_{\hat{n}}}\circ\cdots\circ T_{\beta_{2N'_{\mathrm{uns}}+2},\beta_{2N'_{\mathrm{uns}}+1}}.$$

Thanks to Remark 55, we may apply Proposition 57 to describe the action of $T_\beta^{2N'_{\mathrm{uns}}+1,\hat{n}}$ on the Lagrangian state $\mathcal{U}_{\beta_{2N'_{\mathrm{uns}}}+1}\widetilde{U}_{\beta_0\cdots\beta_{2N'_{\mathrm{uns}}+1}}(\bar{\chi}E_h^0)$. Note that

$$T_\beta^{2N'_{\mathrm{uns}}+1,\hat{n}}\mathcal{U}_{\beta_{2N_{\mathrm{uns}}}+1}\widetilde{U}_{\beta_0\cdots\beta_{2N_{\mathrm{uns}}+1}} = \mathcal{U}_{\beta_{\hat{n}}}\widetilde{U}_{\beta_0\cdots\beta_{\hat{n}}}.$$

We obtain that $\mathcal{U}_{\beta_{\hat{n}+1}}\Pi_{\beta_{\hat{n}+1}}\widetilde{U}_{\beta_0\cdots\beta_{\hat{n}}}(\bar{\chi}E_h^0) = e_{\hat{n},\beta}$, with

$$e_{\hat{n},\beta}(y) = a^{\hat{n},\beta}(y)e^{i\phi_{\hat{n},\beta}(y)/h}, \quad y \in \mathbb{R}^d. \tag{78}$$

In the notation of Section 4.3, by Remark 55 that for any $N'_{\mathrm{uns}}+1 \leq k' \leq \hat{n}$, we have

$$D_{k'} = S_T(V_{\beta_{k'}})(1+O(\epsilon^{\mathrm{p}})) < 1.$$

We therefore set

$$J_{\beta_{N'_{\mathrm{uns}}+1}\cdots\beta_{\hat{n}}} := \prod_{k'=N_{\mathrm{uns}}+1}^{\hat{n}}\big(S_{t_0}(V_{\beta_{k'}})(1+O(\epsilon^{\mathrm{p}}))\big). \tag{79}$$

Thanks to equation (61) in Proposition 57 and equation (64), we obtain for any $\ell \in \mathbb{N}$,

$$\|a^{\hat{n},\beta}\|_{C^\ell} \leq (1+C_\ell h)C'_\ell J_{\beta_{N'_{\mathrm{uns}}+1}\cdots\beta_{\hat{n}}}(\hat{n}+1)^\ell \tag{80}$$

for some constants C_ℓ, C'_ℓ.

5.2.4. *End of the propagation.* Using equation (74) and the results of the previous subsection, we have

$$
\begin{aligned}
\chi E_h = &\sum_{n=1}^{M_r|\log h|} (\chi \tilde{U}(t_0))^{N_\chi+1}\Big(\sum_{\beta\in\mathcal{B}_{n+3N'_{\text{uns}}+2}} \tilde{U}_{\beta_{\hat{n}}\cdots\beta_n}\mathcal{U}^*_{\beta_{\hat{n}}} e_{\hat{n},\beta}\Big)\\
&+\sum_{l=0}^{N_\chi+3N'_{\text{uns}}+3} (\chi\tilde{U}(t_0))^l(1-\chi)\chi_{t_0}E_h^0 + O_{L^2}(h^r),
\end{aligned}
\tag{81}
$$

with

$$
\mathcal{U}_b\Pi_b \sum_{l=0}^{N_\chi+3N'_{\text{uns}}+3} (\chi\tilde{U}(t_0))^l(1-\chi)\chi_{t_0}E_h^0 = \sum_{\beta\in\tilde{\mathcal{B}}_0} e_{0,\beta,b}.
$$

To finish the proof, we have to apply $\mathcal{U}_b\Pi_b(\chi\tilde{U}(t_0))^{N_\chi+1}\tilde{U}_{\beta_{\hat{n}}\cdots\beta_n}\mathcal{U}^*_{\beta_{\hat{n}}}$ to $e_{\hat{n},\beta}$.

To do this, one should once again decompose the propagator, and study

$$
\sum_{\beta'\in B^{N_\chi+1}} \mathcal{U}_b\Pi_b U^\chi_{\beta'}\tilde{U}_{\beta_{\hat{n}}\cdots\beta_n}\mathcal{U}^*_{\beta_{\hat{n}}} e_{\hat{n},\beta}, \tag{82}
$$

with $U^\chi_{\beta'}$ as in (76). To analyse each of the terms on the right-hand side of (82), we use once again Lemma 42 (the lemma may be applied, thanks to Theorem 17 and to Lemma 56).

We obtain that

$$
\mathcal{U}_b\Pi_b U^\chi_{\beta'}\tilde{U}_{\beta_{\hat{n}}\cdots\beta_n}\mathcal{U}^*_{\beta_{\hat{n}}} e_{\hat{n},\beta}(y) = a^{n,\beta,\beta'}(y)e^{i\phi_{n,\beta,\beta'}(y)/h}, \quad y\in\mathbb{R}^d, \tag{83}
$$

and thanks to equation (80), we get

$$
\|a^{n,\beta,\beta'}\|_{C^\ell} \le (1+C_\ell h)C'_\ell J_{\beta_{N'_{\text{uns}}+1}\cdots\beta_{\hat{n}}}(\hat{n}+1)^\ell \tag{84}
$$

for some constants C_ℓ, C'_ℓ.

For any $n\ge 1$, we write

$$
\tilde{\mathcal{B}}_n = \mathcal{B}_{n+3N'_{\text{uns}}+2}\times B^{N_\chi+1}. \tag{85}
$$

As announced, the cardinal of $\tilde{\mathcal{B}}_n$ grows exponentially with n. If $\beta=(\beta',\beta'')\in\tilde{\mathcal{B}}_n$ with $\beta\in\mathcal{B}_{n+2N'_{\text{uns}}+1}$, we define

$$
a_{n,\beta,b} = a^{N_{n+2N_\chi+2,l},\beta,\beta'}, \quad \phi_{n,\beta,b} = \phi_{N_{n+2N_\chi+2,l},\beta,\beta'}.
$$

With these notations, combining (81) with (83) gives us the decomposition (51).

The key point to obtaining estimate (52) is to notice that for any $N\ge N'_{\text{uns}}+1$, we have, thanks to (58),

$$
\begin{aligned}
\sum_{\beta_{N'_{\text{uns}}+1}\cdots\beta_{\hat{N}}\in B_1^{N-N'_{\text{uns}}-1}} J_{\beta_{N'_{\text{uns}}+1}\cdots\beta_{\hat{N}}} &= \Big(\sum_{b\in B_1} S_{t_0}(V_b)(1+O(\epsilon^{\mathsf{p}}))\Big)^{N-N'_{\text{uns}}-1}\\
&\le \exp\big[(N-N'_{\text{uns}}-1)\big(t_0\mathcal{P}\big(\tfrac{1}{2}\big)(1+O(\epsilon^{\mathsf{p}}))\big)\big].
\end{aligned}
\tag{86}
$$

By applying (86) for $N=N_{n+2N_\chi+2,l}$, and combining it with (84), we get (52). $\square$

Note that, although the statement of Theorem 47 describes the generalized eigenfunctions E_h only very close to the trapped set, (81) can be used to describe E_h in any compact set, though in a less explicit way.

Using the estimate (52) as well as the fact that $\|\chi \widetilde{U}(t_0)\|_{L^2\to L^2} \le 1$ and $\|\mathcal{U}_b\|_{L^2\to L^2} \le 1$, we deduce Theorem 1.

6. Semiclassical measures

The main ingredient in the proof of Corollary 50 is nonstationary phase. Let us recall the estimate we will use, and which can be proven by integrating by parts.

Let $a, \phi \in S^{\mathrm{comp}}(X)$. We consider the oscillatory integral

$$I_h(a,\phi) := \int_X a(x) e^{\frac{i\phi(x,h)}{h}} \,\mathrm{d}x.$$

Proposition 62. *Let $\epsilon > 0$. Suppose that there exists $C > 0$ such that for all $x \in \mathrm{spt}(a)$ and for all $0 < h < h_0$, we have $|\partial\phi(x,h)| \ge C h^{1/2-\epsilon}$. Then*

$$I_h(a,\phi) = O(h^\infty).$$

We shall only give a sketch of proof here, and refer to [Hörmander 1983, §7.7] for more details.

Sketch of proof. To prove this result, we simply integrate by parts, noting that

$$I_h(a,\phi) = \frac{h}{i} \int_X \frac{a}{|\partial\phi|^2} \partial\phi \cdot \partial(e^{\frac{i\phi(x,h)}{h}}) \,\mathrm{d}x.$$

Hence, when we integrate by parts, the worst term in the integrand will involve second derivatives of ϕ times $h/|\partial\phi|^2$, and will therefore be a $O(h^{2\epsilon})$ by assumption. By integrating by parts more times, we will gain a factor $h^{2\epsilon}$ every time, so that $I_h(a,\phi)$ is actually a $O(h^\infty)$. ☐

Note that the sketch of proof above tells us that, if we could say that when $\partial\phi(x,h)$ is small, then the higher derivatives of ϕ are small as well, i.e., if we had

$$\forall k \ge 2,\ \exists C_k \quad \text{such that} \quad |\partial^k\phi(x,h)| \le C_k |\partial\phi(x,h)|,$$

then we would have $I_h(a,\phi) = O(h^\infty)$ provided $|\partial\phi(x,h)| \ge C h^{1-\epsilon}$. However, it is not clear that we can estimate the higher derivatives of the phase functions which appear in this section.

6.1. ***Distance between the Lagrangian manifolds.*** To take advantage of Proposition 62, we need a lower bound on the distance between the Lagrangian manifolds which make up $\Phi^{n,t_0}(\mathcal{L}_0) \cap V_b$. To prove such a lower bound, let us first state an elementary topological lemma.

Lemma 63. *There exists $c_0 > 0$ such that for any $\rho, \rho' \in T^*X_0 \cap \mathcal{E}$ such that $d(\rho,\rho') < c_0$, there exists $b \in B$ such that $\rho, \rho' \in V_b$.*

Proof. Suppose for contradiction that for any $\epsilon > 0$, there exists $\rho_\epsilon, \rho'_\epsilon$ such that $d(\rho_\epsilon, \rho'_\epsilon) < \epsilon$ and such that for all $b \in B$ such that $\rho_\epsilon \in V_b$, we have $y_\epsilon \notin V_b$. By compactness of $T^*X_0 \cap \mathcal{E}$, we may suppose that ρ_ϵ converges to some ρ. We then have $\rho'_\epsilon \to x$, and if $b \in B$ is such that $\rho \in V_b$, then $\rho_\epsilon, \rho'_\epsilon \in V_b$ for ϵ small enough, a contradiction. ☐

We may now state our lower bound on the distance between the Lagrangian leaves which make up $\Phi^{n,t_0}(\mathcal{L}_0)\cap V_b$.

Let $N\in\mathbb{N}$, $\beta\in B^N$ and $b\in B_1$. The set $\Phi^{n,t_0}_{\beta}(\mathcal{L}_0)\cap V_b$ may be written in the form $\{(y^{\rho_b}, \partial\tilde{\phi}_{n,\beta,b}(y^{\rho_b})\}$ for some smooth function $\tilde{\phi}_{N,\beta,b}$.

For any $\beta\in B^N$, $\beta'\in B^{N'}$, let us define

$$\sigma(\beta,\beta') := \max(N-\tau(\beta), N'-\tau(\beta')),$$

with $\tau(\beta)$ as in (17).

Proposition 64. *There exist constants $C_1', C_2'>0$ such that for any $N, N'\in\mathbb{N}$, for any $\beta\in B^N$, $\beta'\in B^{N'}$, for any $b\in B_1$ and for any y^{ρ_b}, we have either $\partial\tilde{\phi}_{N,\beta,b}(y^{\rho_b}) = \partial\tilde{\phi}_{N',\beta',b}(y^{\rho_b})$ or*

$$|\partial\tilde{\phi}_{N,\beta,b}(y^{\rho_b}) - \partial\tilde{\phi}_{N',\beta',b}(y^{\rho_b})| \geq C_1' e^{C_2'\sigma(\beta,\beta')}.$$

Proof. Since $T^*X_0\cap\mathcal{E}$ is compact, we may find a constant $C>0$ such that for any $\rho,\rho'\in\mathcal{E}\cap T^*X_0$,

$$d(\Phi^t(\rho),\Phi^t(\rho'))\leq e^{Ct} d(\rho,\rho'), \tag{87}$$

where d is the distance on the energy layer which we introduced in Section 2.1.2.

Let $b\in B_1$, and $y^{\rho_b}\in D_{\beta,b}\cap D_{\beta',b}$ be such that

$$\partial\tilde{\phi}_{N,\beta,b}(y^{\rho_b}) \neq \partial\tilde{\phi}_{N',\beta',b}(y^{\rho_b}).$$

Let us denote by ρ the point $(y^{\rho_b}, \partial\tilde{\phi}_{N,\beta,b}(y^{\rho_b}))$ and by ρ' the point $(y^{\rho_b}, \partial\tilde{\phi}_{N',\beta',b}(y^{\rho_b}))$.

We claim that there exists $0\leq k\leq\sigma(\beta,\beta')$ such that for each $b'\in B$, if $\Phi^{-kt_0}(\rho)\in V_{b'}$, then $\Phi^{-kt_0}(\rho')\notin V_b$. Indeed, if no such k existed, then for each k, there would exist $b_k\in B$ such that $\Phi^{-kt_0}(\rho)\in V_{b_k}$ and $\Phi^{-kt_0}(\rho')\in V_{b_k}$ for each $0\leq k\leq\sigma(\beta,\beta')$. We would then have $\rho\in\Phi^{\max(N,N'),t_0}_{\beta''}(\mathcal{L}_0)$ and $\rho'\in\Phi^{\max(N,N'),t_0}_{\beta''}(\mathcal{L}_0)$ for some sequence β'' built by possibly adding some 0's in front of the sequences β and β'. This would contradict the statement of Corollary 53.

Thanks to Lemma 63, we deduce from this that there exists $0\leq k\leq\sigma(\beta,\beta')$ such that

$$d(\Phi^{-kt_0}(\rho),\Phi^{-kt_0}(\rho'))\geq c_0.$$

Combining this fact with equation (87), we get

$$d(\rho,\rho')\geq c_0 e^{-C\sigma(\beta,\beta')}$$

Since all metrics are equivalent on a compact set, we may compare $d(\rho,\rho')$ with $|\partial\tilde{\phi}_{N,\beta,b}(y^{\rho_b}) - \partial\tilde{\phi}_{N',\beta',b}(y^{\rho_b})|$ and we deduce from this the proposition. ☐

Using the definition of $\widetilde{\mathcal{B}}_n$, we deduce the following result about the functions $\phi_{n,\beta,b}$ in the statement of Theorem 47.

Corollary 65. *There exist constants $C_1', C_2'>0$ such that for any $n, n'\in\mathbb{N}$, for any $\beta\in\widetilde{\mathcal{B}}_n$, $\beta'\in\widetilde{\mathcal{B}}_{n'}$, for any $b\in B_1$ and for any y^{ρ_b}, we have either $\partial\phi_{n,\beta,b}(y^{\rho_b}) = \partial\phi_{n',\beta',b}(y^{\rho_b})$ or*

$$|\partial\phi_{n,\beta,b}(y^{\rho_b}) - \partial\phi_{n',\beta',b}(y^{\rho_b})| \geq C_1' e^{C_2'\min(n,n')}.$$

6.2. ***Proof of Corollary 50.*** We shall now prove Corollary 50, which we recall.

Corollary 50. *There exists a constant $0 < c \leq 1$ and functions $e_{n,\beta,b}$ for $n \in \mathbb{N}$, $\beta \in \widetilde{\mathcal{B}}_n$ and $b \in B_1$ such that for any $a \in C_c^\infty(T^*X)$ and for any $\chi \in C_c^\infty(X)$, we have*

$$\langle \mathrm{Op}_h(\pi_b^2 a)\chi E_h, \chi E_h\rangle = \int_{T^*X} a(x,v)\,\mathrm{d}\mu_{b,\chi}(x,v) + O(h^c),$$

with

$$\mathrm{d}\mu_{b,\chi}(\kappa_b^{-1}(y^{\rho_b},\eta^{\rho_b})) = \sum_{n=0}^{\infty}\sum_{\beta\in\widetilde{\mathcal{B}}_n} e_{n,\beta,b}(y^{\rho_b})\delta_{\{\eta^{\rho_b}=\partial\phi_{j,n}(y^{\rho_b})\}}dy^{\rho_b},$$

The functions $e_{n,\beta,b}$ satisfy the estimate (52).

Proof. Take any small $\epsilon > 0$, and set

$$M := \frac{1}{2C_2'} - \epsilon, \quad c := (M-\epsilon)\mathcal{P}\left(\tfrac{1}{2}\right) = \frac{\mathcal{P}\left(\frac{1}{2}\right)}{2C_2'} - \epsilon,$$

where C_2' comes from Corollary 65.

Let $a \in C_c^\infty(T^*X)$, $\chi \in C_c^\infty(X)$ and $b \in B_1$. Using the fact that $\mathrm{Op}_h(\mathrm{ab}) = \mathrm{Op}_h(\mathrm{a})\mathrm{Op}_h(\mathrm{b}) + O_{L^2\to L^2}(h)$ for any $\mathrm{a}, \mathrm{b} \in S^{\mathrm{comp}}(X)$, the self-adjointness of Π_b, and the unitarity of $\mathcal{U}_b$ on the microsupport of Π_b, we see that we have

$$\begin{aligned}\langle \mathrm{Op}_h(\pi_b^2 a)\chi E_h, \chi E_h\rangle_{L^2(X)} &= \langle \mathrm{Op}_h(a)\Pi_b\chi E_h, \Pi_b\chi E_h\rangle_{L^2(X)} + O(h)\\ &= \langle \mathcal{U}_b\mathrm{Op}_h(a)\mathcal{U}_b^*\mathcal{U}_b\Pi_b E_h, \mathcal{U}_b\Pi_b\chi E_h\rangle_{L^2(X)} + O(h).\end{aligned}$$

Now, using Egorov's theorem ([Zworski 2012, Theorem 11.1]), we know that

$$\mathcal{U}_b\mathrm{Op}_h(a)\mathcal{U}_b^*\mathcal{U}_b\Pi_b = \mathrm{Op}_h(a_b)\mathcal{U}_b\Pi_b + O_{L^2(X)\to L^2(\mathbb{R}^d)}(h^\infty),$$

where $a_b = a\circ\kappa_b + O_{L^2}(h)$. Using decomposition (51), we have

$$\langle \mathrm{Op}_h(\pi_b^2 a)\chi E_h, \chi E_h\rangle_{L^2(X)} = \sum_{n=0}^{\lfloor M_c|\log h|\rfloor}\sum_{\beta\in\widetilde{\mathcal{B}}_n}\left\langle \mathrm{Op}_h(a_b)[e^{i\phi_{n,\beta,b}/h}a_{n,\beta,b}], \sum_{n'=0}^{\lfloor M_c|\log h|\rfloor}\sum_{\beta'\in\widetilde{\mathcal{B}}_{n'}} e^{i\phi_{n',\beta',b}/h}a_{n',\beta',b}\right\rangle + O(h^c). \tag{88}$$

But thanks to estimate (52),

$$\sum_{n=0}^{\lfloor M_c|\log h|\rfloor}\sum_{\beta\in\widetilde{\mathcal{B}}_n} e^{i\phi_{n,\beta,b}/h}a_{n,\beta,b} = \sum_{n=0}^{\lfloor M|\log h|\rfloor}\sum_{\beta\in\widetilde{\mathcal{B}}_n} e^{i\phi_{n,\beta,b}/h}a_{n,\beta,b} + O_{L^2}(h^c),$$

so that

$$\langle \mathrm{Op}_h(\pi_b^2 a)\chi E_h, \chi E_h\rangle_{L^2(X)} = \sum_{n=0}^{\lfloor M|\log h|\rfloor}\sum_{\beta\in\widetilde{\mathcal{B}}_n}\left\langle \mathrm{Op}_h(a_b)[e^{i\phi_{n,\beta,b}/h}a_{n,\beta,b}], \sum_{n'=0}^{\lfloor M|\log h|\rfloor}\sum_{\beta'\in\widetilde{\mathcal{B}}_{n'}} e^{i\phi_{n',\beta',b}/h}a_{n',\beta',b}\right\rangle + O(h^c). \tag{89}$$

We now want to fix a $n \leq M|\log h|$ and a $\beta \in \widetilde{\mathcal{B}}_n$, and to analyse the behaviour of

$$\Big\langle \mathrm{Op}_h(a_b)[e^{i\phi_{n,\beta,b}/h}a_{n,\beta,b}], \sum_{n'=0}^{\lfloor M|\log h|\rfloor}\sum_{\beta'\in\widetilde{\mathcal{B}}_n} e^{i\phi_{n',\beta',b}/h}a_{n',\beta',b}\Big\rangle.$$

Let us define $Y_{n',\beta'} = \{y^{\rho_b} \in \mathrm{spt}(\phi_{n,\beta,b}) \cap \mathrm{spt}(\phi_{n',\beta',b}) : \partial\phi_{n',\beta',b}(y^{\rho_b}) = \partial\phi_{n,\beta,b}(y^{\rho_b})\}$. We have

$$\begin{aligned}\langle \mathrm{Op}_h(a_b)&[e^{i\phi_{n,\beta,b}/h}a_{n,\beta,b}], e^{i\phi_{n',\beta',b}/h}a_{n',\beta',b}\rangle \\ &= \int_{Y_{n',\beta'}} \big(\mathrm{Op}_h(a_b)[e^{i\phi_{n,\beta,b}/h}a_{n,\beta,b}]\big)(y^{\rho_b})e^{i\phi_{n',\beta',b}(y^{\rho_b})/h}a_{n',\beta',b}(y^{\rho_b};h)\,dy^{\rho_b} \\ &\quad + \int_{\mathbb{R}^d\backslash Y_{n',\beta'}} \big(\mathrm{Op}_h(a_b)[e^{i\phi_{n,\beta,b}/h}a_{n,\beta,b}]\big)(y^{\rho_b})e^{i\phi_{n',\beta',b}(y^{\rho_b})/h}a_{n',\beta',b}(y^{\rho_b};h)\,dy^{\rho_b}. \end{aligned} \tag{90}$$

Recall that the integrals are well defined, because the phase functions are well defined in a neighbourhood of the functions $a_{n,\beta,b}$.

The second term on the right-hand side of (90) is a $O(h^\infty)$. Indeed, the image of a Lagrangian state by a pseudodifferential operator is still a Lagrangian state with the same phase. Therefore, we are computing scalar products between Lagrangian states with respective phases $\phi_{n,\beta,b}$ and $\phi_{n',\beta',b}$.

Now, by the choice of M, and by Corollary 65, we know that for each $y^{\rho_b} \in \mathbb{R}^d\backslash Y_{n',\beta'}$ we have $|\partial\phi_{n,\beta,b}(y^{\rho_b}) - \partial\phi_{n',\beta',b}(y^{\rho_b})| \geq Ch^{1/2+\epsilon}$ for some $C, \epsilon > 0$. Hence by Proposition 62, we deduce that the second term on the right-hand side of (90) is a $O(h^\infty)$.

We should now try to understand the properties of the set $Y_{n',\beta'}$.

First of all, $Y_{n',\beta'}$ is an open set. Indeed, if $y^{\rho_b} \in Y_{n',\beta'}$, then the point $\rho = (y^{\rho_b}, \partial\phi_{n,\beta,b}(y^{\rho_b}))$ (in the coordinates centred at ρ_b) belongs to $\Phi_\beta^{n,t_0}(\mathcal{L}_0)$ as well as to $\Phi_{\beta'}^{n',t_0}(\mathcal{L}_0)$ in the notation of Proposition 64. Suppose for simplicity that $n = n'$ (the general case works the same). Then the condition $y^{\rho_b} \in Y_{n',\beta'}$ simply means that $\Phi^{n-k}(\rho)$ was both in V_{β_k} and in $V_{\beta'_k}$ at each intermediate time k. This is clearly an open condition.

On the other hand, by continuity of the phase functions, $Y_{n',\beta'}$ is a closed set. Therefore, $Y_{n',\beta'}$ consists of a certain number of connected components of the support of $\phi_{n',\beta'}$.

We know that the support of $a_{n',\beta',b}$ is included in the domain of definition of $\phi_{n',\beta',b}$. Therefore, some of the connected components of $\mathrm{spt}(a_{n',\beta',b})$ may be included in $Y_{n',\beta'}$, while others are included in $\mathbb{R}^d\backslash Y_{n',\beta'}$, but none of them may intersect both sets. Therefore, if we set $a_{n',\beta',b}^{n,\beta}(y^{\rho_b}) = a_{n',\beta',b}(y^{\rho_b})$ if $y^{\rho_b} \in Y_{n',\beta'}$ and equal to 0 otherwise, then $a_{n',\beta',b}^{n,\beta} \in S$, and we have

$$\begin{aligned}\Big\langle \mathrm{Op}_h(a_b)&[e^{i\phi_{n,\beta,b}/h}a_{n,\beta,b}], \sum_{n'=0}^{\lfloor M|\log h|\rfloor}\sum_{\beta'\in\mathcal{B}_n} e^{i\phi_{n',\beta',b}/h}a_{n',\beta',b}\Big\rangle \\ &= \int_{\mathbb{R}^d} \big(\mathrm{Op}_h(a_b)[e^{i\phi_{n,\beta,b}/h}a_{n,\beta,b}]\big)(y^{\rho_b})e^{-i\phi_{n,\beta,b}(y^{\rho_b})/h}\Big(\sum_{n'=0}^{\lfloor M|\log h|\rfloor}\sum_{\beta'\in\mathcal{B}_n} a_{n',\beta',b}^{n,\beta}\Big)(y^{\rho_b})\,\mathrm{d}y^{\rho_b}. \end{aligned} \tag{91}$$

Let us write

$$\tilde{a}_{n,\beta,b} := \sum_{n'=0}^{\lfloor M|\log h|\rfloor} \sum_{\beta'\in\mathcal{B}_n} a_{n',\beta',b}^{n,\beta}.$$

So $\tilde{a}_{n,\beta,b}(y^{\rho_b})$ is the sum of all the symbols in the expansion (51) having phase $\phi_{n,\beta,b}(y^{\rho_b})$. We see by the estimate (52) that $\tilde{a}_{n,\beta,b}$ satisfies (52) itself, and that

$$\Big\langle \mathrm{Op}_h(a_b)[e^{i\phi_{n,\beta,b}/h}a_{n,\beta,b}], \sum_{n'=0}^{\lfloor M|\log h|\rfloor} \sum_{\beta'\in\mathcal{B}_n} e^{i\phi_{n',\beta',b}/h}a_{n',\beta',b}\Big\rangle$$
$$= \int_{\mathbb{R}^d} \big(\mathrm{Op}_h(a_b)[e^{i\phi_{n,\beta,b}/h}a_{n,\beta,b}]\big)(y^{\rho_b})e^{-i\phi_{n,\beta,b}(y^{\rho_b})/h}\tilde{a}_{n,\beta,b}(y^{\rho_b})\,\mathrm{d}y^{\rho_b} + O(h^\infty).$$

We may then compute this expression using stationary phase, just as to compute the semiclassical measure of a Lagrangian state (see [Zworski 2012, §5.1]). We obtain

$$\Big\langle \mathrm{Op}_h(a_b)[e^{i\phi_{n,\beta,b}/h}a_{n,\beta,b}], \sum_{n'=0}^{\lfloor M|\log h|\rfloor} \sum_{\beta'\in\mathcal{B}_n} e^{i\phi_{n',\beta',b}/h}a_{n',\beta',b}\Big\rangle = \int_{\mathbb{R}^{2n}} a_b\,\mathrm{d}\mu_{n,\beta,b},$$

where

$$d\mu_{n,\beta,b} = a_{n,\beta,b}(y^{\rho_b})\overline{\tilde{a}_{n,\beta,b}(y^{\rho_b})}\delta_{\{\eta^{\rho_b}=\partial\phi_{n,\beta,b}(y^{\rho_b})\}}\mathrm{d}y^{\rho_b}.$$

Summing over all n,β and using equation (89), we obtain indeed that

$$\langle \mathrm{Op}_h(\pi_b^2 a)E_h, E_h\rangle = \int_{T^*X} a(x,\xi)\,\mathrm{d}\mu_{b,\chi}(x,\xi) + O(h^c),$$

with $(\kappa_b)^*\mu_{b,\chi} = \sum_{n=0}^\infty \sum_{\beta\in\mathcal{B}_n}\mu_{n,\beta,b}$; that is to say

$$\mathrm{d}\mu_{b,\chi}(\kappa_b^{-1}(y^{\rho_b},\eta^{\rho_b})) = \sum_{n=0}^\infty \sum_{\beta\in\mathcal{B}_n} e_{n,\beta,b}(y^{\rho_b})\delta_{\{\eta^{\rho_b}=\partial\phi_{n,\beta,b}(y^{\rho_b})\}}dy^{\rho_b},$$

where $e_{n,\beta,b}(y^{\rho_b}) := \lim_{h\to 0}(a_{n,\beta,b}\overline{\tilde{a}_{n,\beta,b}})(y^{\rho_b})$. This concludes the proof of Corollary 50. ☐

6.3. ***Construction of the measure μ^ξ.*** In the Introduction we defined the measure μ^ξ by

$$\int_{T^*\mathbb{R}^d} a\,\mathrm{d}\mu^\xi := \lim_{t\to\infty}\int_{T^*\mathbb{R}^d} a\circ\Phi^t\,\mathrm{d}\mu_0^\xi$$

for any $a\in C_c^0(T^*\mathbb{R}^d)$. We will now give a sketch of the proof of why the hyperbolicity and transversality hypotheses, along with the assumption that $\mathcal{P}(1)<0$, imply that the above limit exists.

Note that the assumption $\mathcal{P}(1)<0$ is really less restrictive than $\mathcal{P}\big(\frac{1}{2}\big)<0$. For instance, if we assume that the flow (Φ^t) is *axiom A*, that is to say, that the periodic orbits are dense in K, then [Bowen 1975, §4.C] guarantees us that $\mathcal{P}(1)<0$.

Note that, if a is nonnegative, then $t\mapsto\int_{T^*\mathbb{R}^d} a\circ\Phi^t\,\mathrm{d}\mu_0^\xi$ is nondecreasing, so that we only have to show that this quantity is bounded.

If μ is a measure, we define $\Phi^t_*\mu$ by

$$\int_{T^*\mathbb{R}^d} a \,\mathrm{d}(\Phi^t_*\mu) := \int_{T^*\mathbb{R}^d} a \circ \Phi^t \,\mathrm{d}\mu.$$

If $\pi \in C^\infty(T^*\mathbb{R}^d; [0,1])$ we define the measure $\pi\mu$ by

$$\int_{T^*\mathbb{R}^d} a \,\mathrm{d}(\pi\mu) := \int_{T^*X} a\pi \,\mathrm{d}\mu.$$

Remark 66. Note that if μ is the semiclassical measure associated to a Lagrangian state ϕ_h, then $\pi\mu$ is the semiclassical measure associated to $\sqrt{pi}\phi_h$, and, by Egorov's theorem, $\Phi^t_*\mu$ is the semiclassical measure associated to $U(t)\phi_h$.

We shall use the functions π_b from Section 4.4. If $\beta \in B^n$, we set

$$\Phi_\beta\mu := \pi_{\beta_n}\Phi^{t_0}_*(\cdots\pi_{\beta_2}\Phi^{t_0}_*(\pi_{\beta_1}\Phi^{t_0}_*\mu)).$$

Let ϕ_h be a Lagrangian state associated to a Lagrangian manifold which is γ-unstable in the coordinates (y^ρ, η^ρ), and let μ be the semiclassical measure associated to ϕ_h. The propagation $U_\beta\phi_h$ can be described using the methods of Section 4.3 along with the results of Section 2. In particular, we obtain, like in [Nonnenmacher and Zworski 2009, (7.12)], that we may find $C, \epsilon > 0$ such that for all $N \in \mathbb{N}$ and all $\beta \in B_1^N$, we have

$$\|U_\beta\phi_h\|_{L^2} \le C(1 + C\epsilon)^N \prod_{j=1}^{N} \exp\bigl[\tfrac{1}{2}S_{t_0}(V_{\beta_j})\bigr].$$

We may deduce from this the following bound for the measure $\Phi_\beta\mu$. Note that this could also be deduced directly from the transport equations for measures, without using Schrödinger propagators and Egorov's theorem.

For any $a \in C^0_c(T^*X)$, if $\beta \in B_1^N$, we have that

$$\langle \Phi_\beta\mu, a\rangle \le C(a)(1 + C\epsilon)^N \prod_{j=1}^{N} \exp[S_{t_0}(V_{\beta_j})].$$

By possibly taking the sets V_b smaller, we may ensure, just like in Section 4.1, that

$$\sum_{b \in B_1} \exp\{S_{t_0}(V_b)\} \le \exp\{t_0(\mathcal{P}(1) + \epsilon)\}.$$

Therefore, we obtain that

$$\sum_{\beta \in B_1^N} \langle \Phi_\beta\mu, a\rangle \le C(a)\exp[-Nt_0(\mathcal{P}(1) - \epsilon)]. \tag{92}$$

If we assume that the flow (Φ^t) is *axiom A*, that is to say, that the periodic orbits are dense in K, then [Bowen 1975, §4.C] guarantees us that $\mathcal{P}(1) < 0$.

Now, we have that

$$\Phi_*^{N t_0} \mu^\xi = \sum_{\beta \in \widetilde{\mathcal{B}}^N} \Phi_\beta \mu^\xi,$$

and we may use (92) along with the assumption that $\mathcal{P}(1) < 0$ to show that, if a is nonnegative, $t \mapsto \int_{T^*\mathbb{R}^d} a \circ \Phi^t \, d\mu_0^\xi$ is bounded.

Showing that μ^ξ is the semiclassical measure associated to E_h follows from [Dyatlov and Guillarmou 2014, §5.1] (which relies on Egorov's theorem), along with estimate (47).

Acknowledgements

The author is partially supported by the Agence Nationale de la Recherche project GeRaSic (ANR-13-BS01-0007-01).

The author would like to thank Stéphane Nonnenmacher for suggesting this project, as well as for his advice during the redaction of this paper. He also thanks the anonymous referee for suggesting several clarifications and improvements in the paper.

References

[Bonthonneau 2014] Y. Bonthonneau, "Long time quantum evolution of observables", preprint, 2014. arXiv

[Bony et al. 2010] J.-F. Bony, N. Burq, and T. Ramond, "Minoration de la résolvante dans le cas captif", *C. R. Math. Acad. Sci. Paris* **348**:23–24 (2010), 1279–1282. MR Zbl

[Bowen 1975] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lecture Notes in Mathematics **470**, Springer, 1975. MR Zbl

[Dimassi and Sjöstrand 1999] M. Dimassi and J. Sjöstrand, *Spectral asymptotics in the semi-classical limit*, London Mathematical Society Lecture Note Series **268**, Cambridge University Press, 1999. MR Zbl

[Dyatlov 2012] S. Dyatlov, "Microlocal limits of Eisenstein functions away from the unitarity axis", *J. Spectr. Theory* **2**:2 (2012), 181–202. MR Zbl

[Dyatlov and Guillarmou 2014] S. Dyatlov and C. Guillarmou, "Microlocal limits of plane waves and Eisenstein functions", *Ann. Sci. Éc. Norm. Supér.* (4) **47**:2 (2014), 371–448. MR Zbl

[Dyatlov and Zworski 2017] S. Dyatlov and M. Zworski, "Mathematical theory of scattering resonances", book in progress, 2017, available at http://math.mit.edu/~dyatlov/res/res_20170323.pdf.

[Guillarmou and Naud 2014] C. Guillarmou and F. Naud, "Equidistribution of Eisenstein series on convex co-compact hyperbolic manifolds", *Amer. J. Math.* **136**:2 (2014), 445–479. MR Zbl

[Hörmander 1983] L. Hörmander, *The analysis of linear partial differential operators, I: Distribution theory and Fourier analysis*, Grundlehren der Mathematischen Wissenschaften **256**, Springer, 1983. MR Zbl

[Ingremeau 2017] M. Ingremeau, "Distorted plane waves on manifolds of nonpositive curvature", *Comm. Math. Phys.* **350**:2 (2017), 845–891. MR Zbl

[Katok and Hasselblatt 1995] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Encyclopedia of Mathematics and its Applications **54**, Cambridge University Press, 1995. MR Zbl

[Melrose 1995] R. B. Melrose, *Geometric scattering theory*, Cambridge University Press, 1995. MR Zbl

[Nonnenmacher and Zworski 2009] S. Nonnenmacher and M. Zworski, "Quantum decay rates in chaotic scattering", *Acta Math.* **203**:2 (2009), 149–233. MR Zbl

[Shnirelman 1974] A. I. Shnirelman, "Ergodic properties of eigenfunctions", *Uspehi Mat. Nauk* **29**:6(180) (1974), 181–182. In Russian. MR Zbl

[Sjöstrand 1990] J. Sjöstrand, "Geometric bounds on the density of resonances for semiclassical problems", *Duke Math. J.* **60**:1 (1990), 1–57. MR Zbl

[Colin de Verdière 1985] Y. Colin de Verdière, "Ergodicité et fonctions propres du laplacien", *Comm. Math. Phys.* **102**:3 (1985), 497–502. MR Zbl

[Zelditch 1987] S. Zelditch, "Uniform distribution of eigenfunctions on compact hyperbolic surfaces", *Duke Math. J.* **55**:4 (1987), 919–941. MR Zbl

[Zworski 2012] M. Zworski, *Semiclassical analysis*, Graduate Studies in Mathematics **138**, American Mathematical Society, Providence, RI, 2012. MR Zbl

Received 7 Jan 2016. Revised 19 Nov 2016. Accepted 7 Mar 2017.

MAXIME INGREMEAU: maxime.ingremeau@math.u-psud.fr
Université Paris-Sud, 91400 Orsay, France

ANALYSIS AND PDE
Vol. 10, No. 4, 2017
dx.doi.org/10.2140/apde.2017.10.817

msp

A FOURIER RESTRICTION THEOREM FOR A TWO-DIMENSIONAL SURFACE OF FINITE TYPE

STEFAN BUSCHENHENKE, DETLEF MÜLLER AND ANA VARGAS

The problem of $L^q(\mathbb{R}^3) \to L^2(S)$ Fourier restriction estimates for smooth hypersurfaces S of finite type in $\mathbb{R}^3$ is by now very well understood for a large class of hypersurfaces, including all analytic ones. In this article, we take up the study of more general $L^q(\mathbb{R}^3) \to L^r(S)$ Fourier restriction estimates, by studying a prototypical model class of two-dimensional surfaces for which the Gaussian curvature degenerates in one-dimensional subsets. We obtain sharp restriction theorems in the range given by Tao in 2003 in his work on paraboloids. For high-order degeneracies this covers the full range, closing the restriction problem in Lebesgue spaces for those surfaces. A surprising new feature appears, in contrast with the nonvanishing curvature case: there is an extra necessary condition. Our approach is based on an adaptation of the bilinear method. A careful study of the dependence of the bilinear estimates on the curvature and size of the support is required.

1. Introduction

Let S be a smooth hypersurface in $\mathbb{R}^n$ with surface measure $d\sigma_S$. The Fourier restriction problem for S, proposed by E. M. Stein in the seventies, asks for the range of exponents q and r for which the estimate

$$\left(\int_S |\hat{f}|^r \, d\sigma_S\right)^{\frac{1}{r}} \le C \|f\|_{L^q(\mathbb{R}^n)} \tag{1-1}$$

holds true for every $f \in \mathcal{S}(\mathbb{R}^n)$, with a constant C independent of f. There was a lot of activity on this problem in the seventies and early eighties. The sharp range in dimension $n = 2$ for curves

Buschenhenke was partially supported by the ERC grant 277778. Buschenhenke and Müller were partially supported by the DFG grant MU 761/ 11-1. Vargas was partially supported by MICINN/MINECO grants MTM2010-16518 and MTM2013–40945 (Spain).
MSC2010: 42B10.
Keywords: Fourier restriction, finite type, multilinear, bilinear.

with nonvanishing curvature was determined through work by C. Fefferman [1970], E. M. Stein and A. Zygmund [1974]. In higher dimensions, the sharp $L^q - L^2$ result for hypersurfaces with nonvanishing Gaussian curvature was obtained by Stein [1986] and P. A. Tomas [1975] (see also [Strichartz 1977]). Some more general classes of surfaces were treated by A. Greenleaf [1981].

The question about the general $L^q - L^r$ restriction estimates is nevertheless still wide open. Fundamental progress has been made since the nineties, with contributions by many. Major new ideas were introduced in particular by J. Bourgain [1991; 1995b] and T. Wolff [1995], which led to important further steps towards an understanding of the case of nonvanishing Gaussian curvature. These ideas and methods were further developed by A. Moyua, A. Vargas, L. Vega and T. Tao [Moyua et al. 1996; 1999; Tao et al. 1998], who established the so-called bilinear approach (which had been anticipated in the work of C. Fefferman [1970] and had implicitly been present in the work of Bourgain [1995a]) for hypersurfaces with nonvanishing Gaussian curvature for which all principal curvatures have the same sign. The same method was applied to the light cone by Tao and Vargas [2000a; 2000b]. The climax of the application of that bilinear method to these types of surfaces is due to Tao [2001b] (for principal curvatures of the same sign), and Wolff [2001] and Tao [2001a] (for the light cone). In particular, in these last two papers the sharp linear restriction estimates for the light cone in $\mathbb{R}^4$ were obtained.

For the case of nonvanishing curvature but principal curvatures of different signs, analogous results in $\mathbb{R}^3$ were proved by S. Lee [2006] and Vargas [2005]. Results for the light cone were previously obtained in $\mathbb{R}^3$ by B. Barceló [1985], who also considered more general cones [Barceló 1986]. These results were improved to sharp theorems by S. Buschenhenke [2015]. The bilinear approach also produced results for hypersurfaces with $k \leq n-2$ nonvanishing principal curvatures [Lee and Vargas 2010].

More recently, J. Bourgain and L. Guth [2011] made further important progress on the case of nonvanishing curvature by making use also of multilinear restriction estimates due to J. Bennett, A. Carbery and T. Tao [Bennett et al. 2006].

On the other hand, general finite-type surfaces in $\mathbb{R}^3$ (without assumptions on the curvature) have been considered in work by I. Ikromov, M. Kempe and D. Müller [Ikromov et al. 2010; Ikromov and Müller 2011; 2012; 2014], and the sharp range of Stein–Tomas-type $L^q - L^2$ restriction estimates has been determined for a large class of smooth, finite-type hypersurfaces, including all analytic hypersurfaces.

It is our aim in this work to take up the latter branch of development by considering a certain model class of hypersurfaces in dimension three with varying curvature and study more general $L^q - L^r$ restriction estimates. Our approach will again be based on the bilinear method.[1] In our model class, the degeneracy of the curvature will take place along one-dimensional subvarieties. For analytic hypersurfaces whose Gaussian curvature does not vanish identically, this kind of behavior is typical, even though in our model class the zero varieties will still be linear (or the union of two linear subsets). Even though our model class would seem to be among the simplest possible surfaces of such behavior, we will see that they require a very intricate study. We hope that this work will give some insight also for future research on more general types of hypersurfaces.

[1] When preparing this article, the multilinear approach seemed still not sufficiently developed for our needs, since estimates with sharp dependence on the transversality were lacking. For recent progress on this issue, we refer to [Ramos 2016].

Independently of our work, a result for rotationally invariant surfaces with degeneracy of the curvature at a single point has been obtained recently by B. Stovall [2015].

1A. *Outline of the problem: the adjoint setting.* We start with a description of the surfaces that we want to study. We will consider surfaces that are graphs of smooth functions defined on $Q =]0,1[\times]0,1[$,

$$\Gamma = \operatorname{graph}(\phi) = \{(\xi, \phi(\xi)) : \xi \in Q\}.$$

The surface Γ is equipped with the surface measure, σ_Γ. It will be more convenient to use duality and work in the adjoint setting. The adjoint restriction operator is given by

$$\mathcal{R}^* f(x) = \widehat{f\, d\sigma_\Gamma}(x) = \int_\Gamma f(\xi) e^{-ix\cdot\xi}\, d\sigma_\Gamma(\xi), \tag{1-2}$$

where $f \in L^s(\Gamma, \sigma_\Gamma)$. The restriction problem is therefore equivalent to the question of finding the appropriate range of exponents for which the estimate

$$\|\mathcal{R}^* f\|_{L^p(\mathbb{R}^3)} \le C \|f\|_{L^s(\Gamma, d\sigma_\Gamma)}$$

holds with a constant C independent of the function $f \in L^s(\Gamma, d\sigma_s)$. We shall require the following properties of the functions ϕ:

Let $m_1, m_2 \in \mathbb{R}$, $m_1, m_2 \ge 2$. We say that a function ϕ is of *normalized type* (m_1, m_2) if there exist $\phi_{(1)}, \phi_{(2)} \in C^\infty(]0,1[, \mathbb{R})$ and $a, b > 0$ such that

$$\phi(\xi_1, \xi_2) = \phi_{(1)}(\xi_1) + \phi_{(2)}(\xi_2) \tag{1-3}$$

on $]0,1[\times]0,1[$, where the derivatives of the $\phi_{(i)}$ satisfy

$$\phi''_{(i)}(\xi_i) \sim \xi_i^{m_i-2}, \tag{1-4}$$

$$|\phi^{(k)}_{(i)}(\xi_i)| \lesssim \xi_i^{m_i-k} \quad \text{for } k \ge 3. \tag{1-5}$$

The constants hidden in these estimates are assumed to be admissible in the sense that they only depend on m_1, m_2 and the order of the derivative, but not explicitly on the $\phi_{(i)}$.

Note we have restricted ourselves to the open square Q which does not contain the origin in order to allow also for noninteger values of m_1 and m_2.

One would of course expect that small perturbations of such functions ϕ, depending on both ξ_1 and ξ_2, should lead to hypersurfaces sharing the same restriction estimates as our model class above. However, such perturbation terms are not covered by our proof. It seems that the treatment of these more general situations would require even more intricate arguments, which will have to take the underlying geometry of the surface into account. We plan to study these questions in the future.

The prototypical example of a normalized function of type (m_1, m_2) is of course $\phi(\xi) = \xi_1^{m_1} + \xi_2^{m_2}$. For m_1 and m_2 integer, others arise simply as follows:

Remarks 1.1. (i) Let $\varepsilon > 0$ and $\varphi \in C^\infty(]-\varepsilon, \varepsilon[, \mathbb{R})$ be of *finite type* $m \in \mathbb{N}$ in 0, i.e.,

$$\varphi(0) = \varphi'(0) = \cdots = \varphi^{(m-1)}(0) = 0 \ne \varphi^{(m)}(0).$$

Assume $\varphi^{(m)}(0) > 0$. Then there exist $\varepsilon' \in (0, \varepsilon)$ such that

$$\varphi^{(k)}(t) \sim t^{m-k}$$

for all $0 \le k \le m$, $|t| < \varepsilon'$.

(ii) Further let $\phi(\xi) = \phi_{(1)}(\xi_1) + \phi_{(2)}(\xi_2)$, $|\xi| \le \varepsilon$, where $\phi_{(i)} \in C^\infty(]-\varepsilon, \varepsilon[, \mathbb{R})$ is of finite type m_i in 0 with $\phi_{(i)}^{(m_i)}(0) > 0$. Then there exists an $\bar{\varepsilon} > 0$ such that $y \mapsto \phi(\bar{\varepsilon} y)$ is of normalized finite type (m_1, m_2).

Proof. (i) Since φ has a zero of order m at the origin, we can find some $\varepsilon_0 > 0$, a smooth function $\chi_0 :]-\varepsilon_0, \varepsilon_0[\to]0, \infty[$ and a sign $\sigma = \pm 1$ such that

$$\varphi(t) = \sigma t^m \chi_0(t)$$

for all $|t| < \varepsilon_0$. It is then easy to see that this implies $\varphi^{(k)}(t) \sim t^{m-k}$.

(ii) Choose $\bar{\varepsilon} > 0$ such that for both $i = 1, 2$, $0 \le k \le m_i$ and all $0 \le t \le \bar{\varepsilon}$,

$$\phi_{(i)}^{(k)}(t) \sim t^{(m_i - k)}.$$

Then for all $0 \le s \le 1$,

$$\frac{d^k}{ds^k} \phi_{(i)}(\bar{\varepsilon} s) \sim \bar{\varepsilon}^k (\bar{\varepsilon} s)^{(m_i - k)} = \bar{\varepsilon}^{m_i} s^{(m_i - k)}. \qquad \square$$

In order to formulate our main theorem, adapting Varchenko's notion of height to our setting, we introduce the height h of the surface by

$$\frac{1}{h} = \frac{1}{m_1} + \frac{1}{m_2}.$$

Let us also put $\bar{m} = m_1 \vee m_2 = \max\{m_1, m_2\}$ and $m = m_1 \wedge m_2 = \min\{m_1, m_2\}$.

Theorem 1.2. *Let $p > \max\{\frac{10}{3}, h+1\}$, $1/s' \ge (h+1)/p$ and $1/s + (2\bar{m}+1)/p < (\bar{m}+2)/2$. Then $\mathcal{R}^*$ is bounded from $L^{s,t}(\Gamma, d\sigma_\Gamma)$ to $L^{p,t}(\mathbb{R}^3)$ for every $1 \le t \le \infty$.*

If in addition $s \le p$ or $1/s' > (h+1)/p$, then $\mathcal{R}^*$ is even bounded from $L^s(\Gamma, d\sigma_\Gamma)$ to $L^p(\mathbb{R}^3)$.

Remarks 1.3. (i) Notice that the "critical line" $1/s' = (h+1)/p$ and the line $1/s + (2\bar{m}+1)/p = (\bar{m}+2)/2$ in the $(1/s, 1/p)$-plane intersect at the point $(1/s_0, 1/p_0)$ given by

$$\frac{1}{s_0} = \frac{3\bar{m} + m - m\bar{m}}{4\bar{m} + 2m}, \qquad \frac{1}{p_0} = \frac{\bar{m} + m}{4\bar{m} + 2m}. \tag{1-6}$$

This shows in particular that the point $(1/s_0, 1/p_0)$ lies strictly above (if $m > 2$) or on the bisectrix $1/s = 1/p$ (if $m = 2$).

The condition $1/s' \ge (h+1)/p$ in the theorem is necessary and in fact dictated by homogeneity (Knapp box examples).

(ii) By (i), the condition

$$\frac{\bar{m}+2}{2} > \frac{2\bar{m}+1}{p} + \frac{1}{s} \tag{1-7}$$

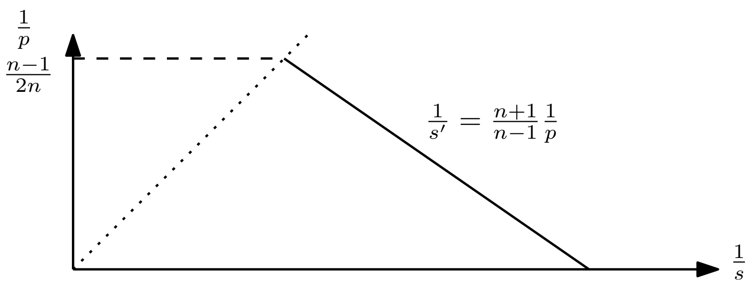


Figure 1. Conjectured range of p and s for nonvanishing Gaussian curvature.

only plays a role above the bisectrix. It is necessary too when $p < s$, hence, in view of (i), if $m > 2$. If $m = 2$, it is necessary with the possible exception of the case where

$$s_0 = p_0 = \frac{4(\bar{m}+1)}{\bar{m}+2},$$

for which we do not have an argument. Our proof in Section 1C will reflect the fact that for $m_j > 2$, the behavior of the operator must be worse than for the case $m_j = 2$.

(iii) From the first condition in the theorem, we see that $p \geq h+1$ is also necessary. Moreover, we shall show in Section 1C that strong-type estimates are not possible unless $s \leq p$ or $1/s' > h + 1/p$. The condition $p > \frac{10}{3}$ is due to the use of the bilinear method, as this exponent gives the sharp bilinear result for the paraboloid, and it is surely not sharp. Nevertheless, when $h > \frac{7}{3}$, we obtain the sharp result.

A new phenomenon appears in these surfaces. In the case of nonvanishing Gaussian curvature, it is conjectured that the sharp range is given by the homogeneity condition $1/s' \geq (h+1)/p$ (with $h = 2/(n-1)$, hence $h+1 = (n+1)/(n-1)$), and a second condition, $p > 2n/(n-1)$, due to the decay rate of the Fourier transform of the surface measure. A similar result is conjectured for the light cone (cf. Figure 1). In contrast to this, we show in our theorem that for the class of surfaces Γ under consideration a third condition appears, namely (1-7).

Let us briefly discuss the different situations that may arise in Theorem 1.2, depending on the choice of m_1 and m_2:

First observe that $1/p_0$ in (1-6) is above the critical threshold $1/p_c = \frac{3}{10}$, if $\bar{m} \leq 2m$. In this case, the new condition

$$\frac{1}{s} + \frac{2\bar{m}+1}{p} = \frac{\bar{m}+2}{2}$$

will not show up in our theorem. So for $\bar{m} \leq 2m$, we are in the situation of either Figure 2 $\big($if $h \leq \frac{7}{3}$, i.e., $h+1 \leq \frac{10}{3}\big)$ or of Figure 3 $\big($if $h > \frac{7}{3}\big)$. Notice that in the last case our theorem is sharp.

It might also be interesting to compare p_0 not only with the condition $p > \frac{10}{3}$, which is due to the bilinear method, but with the conjectured range $p > 3$. We always have $p_0 \geq 3$, while we have $p_0 = 3$ only if $m_1 = m_2$; i.e., a reasonable conjecture is that the new condition (1-7) should always appear for inhomogeneous surfaces with $m_1 \neq m_2$. In the case $\bar{m} > 2m$, our new condition might be visible.

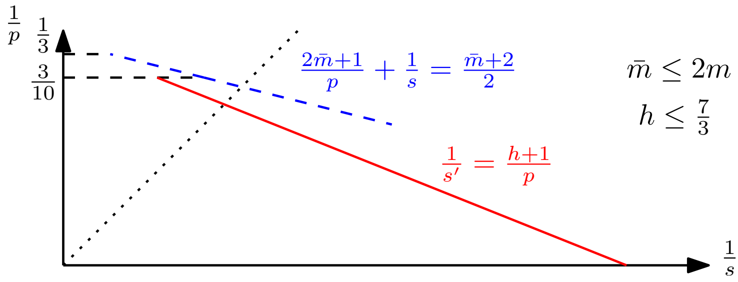


Figure 2. Range of p and s in Theorem 1.2.

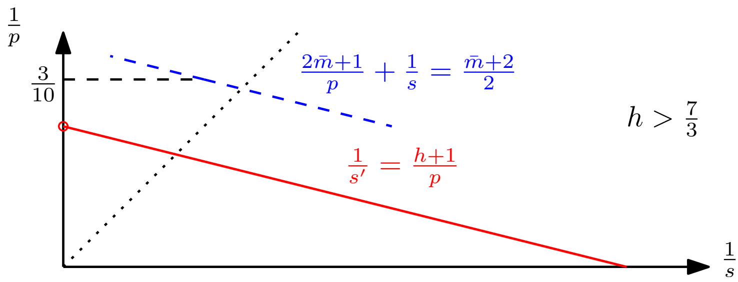


Figure 3. Range of p and s in Theorem 1.2.

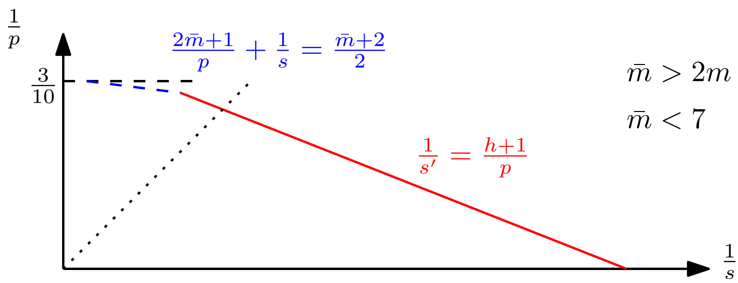


Figure 4. Range of p and s in Theorem 1.2.

Observe next that the line

$$\frac{1}{s} + \frac{2\bar{m}+1}{p} = \frac{\bar{m}+2}{2}$$

intersects the $(1/p)$-axis where

$$p = p_1 = \frac{4\bar{m}+2}{\bar{m}+2}.$$

Thus there are two subcases:

For $\bar{m} < 7$ we have $p_1 < \frac{10}{3}$, corresponding to Figure 4, and our new condition appears.

For $\bar{m} \ge 7$ we may either have $p_0 \ge p_1 \ge h+1$ (which is equivalent to $\bar{m}m \ge 3\bar{m} + m$) and thus Figure 5 applies, or $p_0 < p_1 < h+1$ (which is equivalent to $\bar{m}m < 3\bar{m} + m$), and we are in the situation of Figure 3; here again the new condition becomes relevant. Observe that in the two last mentioned cases, i.e., for $\bar{m} \ge 7$, our theorem is always sharp.

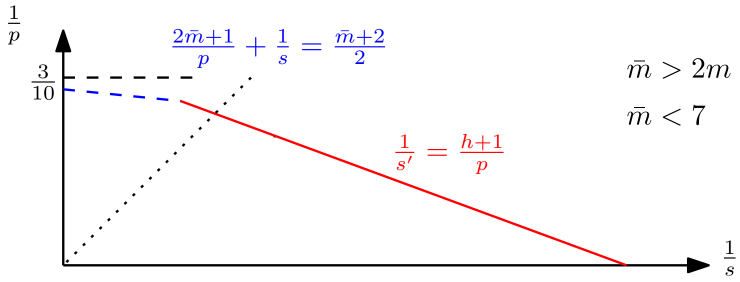


Figure 5. Range of p and s in Theorem 1.2.

Further observe that the appearance of a third condition, besides the classical ones, is natural: Fix $m_1 = 2$ and let $\bar{m} = m_2 \to \infty$. Then the contact order in the second coordinate direction degenerates. Hence, we would expect to find the same p-range as for a two-dimensional cylinder, which agrees with the range for a parabola in the plane, namely $p > 4$ (see [Fefferman 1970; Zygmund 1974]). Since $h \to 2$ as $\bar{m} = m_2 \to \infty$, the condition $p > \max\{\frac{10}{3}, h+1\}$ becomes $p > \frac{10}{3}$ in the limit, which would lead to a larger range than expected. However, the new extra condition

$$\frac{1}{s} + \frac{2\bar{m}+1}{p} < \frac{\bar{m}+2}{2}$$

becomes $p > 4$ for $\bar{m} \to \infty$, as is to be expected.

The restriction problem for the graph of functions $\phi(x) = \xi_1^{m_1} + \xi_2^{m_2}$ (and related surfaces) was studied by E. Ferreyra and M. Urciuolo [2009], however by simpler methods, which led to weaker results than ours. In their approach, they made use of the invariance of this surface under suitable nonisotropic dilations as well as of the one-dimensional results for curves. This allowed them to obtain some results for $p > 4$, in the region below the homogeneity line, i.e., for $1/s' > (h+1)/p$. Our results are stronger in two ways: they include the critical line and, more importantly, when $h < 3$, we obtain a larger range for p.

As for the points on the critical line in the range $p > 4$, let us indicate that these points can in fact also be obtained by means of a simple summation argument involving Lorentz spaces and real interpolation. This can be achieved by means of a summation trick going back to ideas by Bourgain [1985] (see for instance [Tao et al. 1998; Lee 2003]). Details are given in Section A1 of this article.

1B. ***Passage from surface to Lebesgue measure.*** We will always consider hypersurfaces $S = \{(\eta, \phi(\eta)) : \eta \in U\}$ which are the graphs of functions ϕ that are smooth on an open bounded subset $U \subset \mathbb{R}^d$ and continuous on the closure of U. The adjoint of the Fourier restriction operator associated to S is then given by

$$\mathcal{R}^* f(x,t) = \widehat{f\, d\sigma_S}(x,t) = \int_S f(\xi) e^{-i(x,t)\cdot\xi}\, d\sigma_S(\xi), \quad (x,t) \in \mathbb{R}^d \times \mathbb{R} = \mathbb{R}^{d+1},$$

where $d\sigma_S = (1 + |\nabla\phi(\eta)|^2)^{\frac{1}{2}}\, d\eta$ denotes the Riemannian surface measure of S. Here, $f : S \to \mathbb{C}$ is a function on S, but we shall often identify it with the corresponding function $\tilde{f} : U \to \mathbb{C}$, given by

$\tilde{f}(\eta) = f(\eta, \phi(\eta))$. Correspondingly, we define

$$R^*_{\mathbb{R}^d} g(x,t) := \widehat{g\,d\nu}(x,t) := \int_U g(\eta) e^{-i(x\eta + t\phi(\eta))}\, d\eta, \quad (x,t) \in \mathbb{R}^d \times \mathbb{R} = \mathbb{R}^{d+1},$$

for every function $g \in L^1(U)$ on U. We shall occasionally address $d\nu = d\nu_S$ as the "Lebesgue measure" on S, in contrast with the surface measure $d\sigma = d\sigma_S$. Moreover, to emphasize which surface S is meant, we shall occasionally also write $R^*_{\mathbb{R}^d} = R^*_{S,\mathbb{R}^d}$. Observe that if there is a constant A such that

$$|\nabla\phi(\eta)| \le A, \quad \eta \in U \tag{1-8}$$

(this applies for instance to our class of hypersurfaces Γ, since we assume $m_i \ge 2$), then the Lebesgue measure $d\nu$ and the surface measure $d\sigma$ are comparable, up to some positive multiplicative constants depending only on A. Moreover, since

$$R^* f = R^*_{\mathbb{R}^d}\big(\tilde{f}(1 + |\nabla\phi(\eta)|^2)^{\frac{1}{2}}\big), \tag{1-9}$$

the L^q-norms $\|\tilde{f}\|_{L^q(d\eta)}$ and $\|f\|_{L^q(d\sigma_S)} = \|\tilde{f}(1+|\nabla\phi(\eta)|^2)^{\frac{1}{2}q}\|_{L^q(d\eta)}$ of $\tilde{f}$ and of f are comparable too. Throughout the article, we shall therefore apply the following.

Convention 1.4. Whenever $|\nabla\phi| \lesssim 1$, with some slight abuse of notation, we shall denote the function f on S and the corresponding function $\tilde{f}$ on U by the same symbol f, and write $R^*_{\mathbb{R}^d} f$ in place of $R^*_{\mathbb{R}^d} \tilde{f}$.

In view of these observations, we shall in the sequel mainly work with the operator $R^*_{\mathbb{R}^d}$ associated to the hypersurface Γ, in place of R^*.

1C. *Necessary conditions.* The condition $p > h+1$ is in some sense the weakest one. Indeed, the second condition already implies $p \ge h+1$, and even $p > h+1$ when $s < \infty$. Thus the condition $p > h+1$ only plays a role when the critical line $1/s' = (h+1)/p$ intersects the axis $1/s = 0$ at a point where $p > p_c = \frac{10}{3}$ (see Figure 3).

However, the condition $p > h+1$ is necessary as well (although some kind of weak-type estimate might hold true at the endpoint). This can be shown by analyzing the oscillatory integral defined by $R^*_{\mathbb{R}^d} 1$ (see [Sogge 1987] for similar arguments). For the sake of simplicity, we shall do this only for the model case $\phi(\xi) = \xi_1^{m_1} + \xi_2^{m_2}$ (the more general case can be treated by similar, but technically more involved arguments).

Lemma 1.5. *Assume $m \ge 2$.*

(i) *If $1 \ll \mu \ll \lambda \ll \mu^m$, then*

$$\left|\int_0^\delta e^{i(\mu\xi - \lambda(\xi^m + \mathcal{O}(\xi^{m+1})))}\, d\xi\right| \ge C_\delta \mu^{-\frac{m-2}{2m-2}} \lambda^{-\frac{1}{2m-2}},$$

provided $\delta > 0$ is sufficiently small.

(ii) *If $1 \ll \mu^m \ll \lambda$, $0 \le \alpha < 1$ and $0 \le \beta < 1$, then*

$$\left|\int_0^1 e^{i(\mu\xi - \lambda\xi^m)} \xi^{-\alpha} |\log(\xi/2)|^{-\beta}\, d\xi\right| \gtrsim \lambda^{\frac{\alpha-1}{m}} (\log\lambda)^{-\beta}.$$

Proof. (i) Apply the transformation $\xi \mapsto (\mu/\lambda)^{\frac{1}{m-1}}\xi$ to obtain

$$\int_0^\delta e^{i(\mu\xi-\lambda\xi^m+\mathcal{O}(\xi^{m+1}))}\,d\xi=\int_0^{\delta(\frac{\lambda}{\mu})^{1/(m-1)}}\Big(\frac{\mu}{\lambda}\Big)^{\frac{1}{m-1}}e^{i(\frac{\mu^m}{\lambda})^{1/(m-1)}\phi(\xi)}\,d\xi=\int_0^1+\int_1^{\delta(\frac{\lambda}{\mu})^{1/(m-1)}},$$

where $\phi(\xi)=\xi-\xi^m+\mathcal{O}((\mu/\lambda)^{\frac{1}{m-1}}\xi^{m+1})$, $(\mu/\lambda)^{\frac{1}{m-1}}\ll 1$ and $(\mu^m/\lambda)^{\frac{1}{m-1}}\gg 1$. The phase function $\phi(\xi)$ has a unique critical point at $\xi_0=\xi_0(\mu/\lambda)$ in $[0,1]$ lying very close to $m^{-\frac{1}{m-1}}$. Applying well-known asymptotic expansions for oscillatory integrals with nondegenerate critical points (see, e.g., [Stein 1993]), we find that

$$\Big|\int_0^1\Big|\gtrsim\Big(\frac{\mu}{\lambda}\Big)^{\frac{1}{m-1}}\Big(\frac{\mu^m}{\lambda}\Big)^{-\frac{1}{2}\frac{1}{m-1}}.$$

Moreover, integrating by parts in the second integral leads to

$$\Big|\int_1^{\delta(\frac{\lambda}{\mu})^{1/(m-1)}}\Big|\lesssim C_1\Big(\frac{\mu}{\lambda}\Big)^{\frac{1}{m-1}}\Big(\frac{\mu^m}{\lambda}\Big)^{-\frac{1}{m-1}},$$

provided δ is sufficiently small. These estimates imply

$$\Big|\int_0^1\Big(\frac{\mu}{\lambda}\Big)^{\frac{1}{m-1}}e^{i(\mu\xi-\lambda\xi^m+\mathcal{O}(\xi^{m+1}))}\,d\xi\Big|\gtrsim\Big(\frac{\mu}{\lambda}\Big)^{\frac{1}{m-1}}\Big(\frac{\mu^m}{\lambda}\Big)^{-\frac{1}{2}\frac{1}{m-1}}=\mu^{-\frac{m-2}{2m-2}}\lambda^{-\frac{1}{2m-2}}.$$

(ii) Apply the change of variables $\xi\mapsto\lambda^{-\frac{1}{m}}\xi$ to obtain

$$\begin{aligned}&\Big|\int_0^1 e^{i(\mu\xi-\lambda\xi^m)}\xi^{-\alpha}|\log(\xi/2)|^{-\beta}\,d\xi\Big|\\&\qquad=\lambda^{\frac{\alpha-1}{m}}|\log(\lambda^{-\frac{1}{m}})|^{-\beta}\Big|\int_0^{\lambda^{1/m}}e^{i(\mu\lambda^{-1/m}\xi-\xi^m)}\xi^{-\alpha}\Big(1+\frac{|\log(\xi/2)|}{\frac{1}{m}\log\lambda}\Big)^{-\beta}d\xi\Big|\\&\qquad\gtrsim\lambda^{\frac{\alpha-1}{m}}(\log\lambda)^{-\beta}.\end{aligned}$$

Notice here that $\lambda^{\frac{1}{m}}\gg 1$ and $\mu\lambda^{-\frac{1}{m}}\ll 1$, and that, as $\lambda\to\infty$, the last oscillatory integral tends to $\int_0^\infty e^{-i\xi^m}\xi^{-\alpha}\,d\xi\neq 0$ (which is easily seen). □

Part (ii) of the lemma implies

$$\Big|\int_0^1 e^{-i(x_1\xi-x_3\xi^{m_1})}\,d\xi\Big|\gtrsim x_3^{-\frac{1}{m_1}}$$

for $1\le x_3<\infty$, $1\ll x_1^{m_1}\ll x_3$, and since $R^*_{\mathbb{R}^d}1=\widehat{d\nu}$, we find that

$$\begin{aligned}\|R^*_{\mathbb{R}^d}1\|_p^p&\ge\int_1^\infty\int_{1\ll x_2\ll x_3^{1/m_2}}\int_{1\ll x_1\ll x_3^{1/m_1}}|R^*_{\mathbb{R}^d}1(x_1,x_2,-x_3)|^p\,dx_1\,dx_2\,dx_3\\&\gtrsim\int_1^\infty\int_{1\ll x_2\ll x_3^{1/m_2}}dx_2\int_{1\ll x_1\ll x_3^{1/m_1}}dx_1 x_3^{-p(\frac{1}{m_1}+\frac{1}{m_2})}\,dx_3\\&\gtrsim\int_1^\infty x_3^{(1-p)(\frac{1}{m_1}+\frac{1}{m_2})}\,dx_3=\int_1^\infty x_3^{-\frac{p-1}{h}}\,dx_3.\end{aligned}$$

If the adjoint Fourier restriction operator is bounded, the integral has to be finite; thus necessarily $(p-1)/h > 1$, i.e., $p > h+1$.

Next, to see that the condition (1-7), i.e.,

$$\frac{\bar{m}+2}{2} > \frac{2\bar{m}+1}{p} + \frac{1}{s},$$

is necessary in Theorem 1.2, we consider the subsurface

$$\Gamma_0 = \{(\xi, \phi(\xi)) : \xi \in [0,1] \times [\tfrac{1}{2}, \tfrac{1}{2}+\delta]\},$$

where $\delta > 0$ is assumed to be sufficiently small. On this subsurface, the principal curvature in the ξ_2-direction is bounded from below. This means that, after applying a suitable affine transformation of coordinates, the restriction problem for the surface Γ_0 is equivalent to the one for the surface

$$\Gamma_{m_1,2} = \{(\xi_1, \xi_2, \xi_1^{m_1} + c(\xi_2^2 + \mathcal{O}(\xi_2^3))) : (\xi_1, \xi_2) \in [0,1] \times [0,\delta]\},$$

where $c > 0$.

As stated in Remarks 1.3, the condition (1-7) only plays a role above the bisectrix $1/s = 1/p$. So, assume $p < s$ (as explained, this excludes only the case where $m = 2$ and $s = p = p_0$). Then we may choose $\beta < 1$ such that $\beta s > 1 > \beta p$. Assume $\mathcal{R}^*_\Gamma$ is bounded from $L^s(\Gamma)$ to $L^p(\mathbb{R}^3)$; i.e., $\mathcal{R}^*_{\Gamma_{m_1,2}}$ is bounded from $L^s(\Gamma_{m_1,2})$ to $L^p(\mathbb{R}^3)$. Passing again from the surface measure $d\sigma$ to the "Lebesgue measure" $d\nu$ on $\Gamma_{m_1,2}$, define $f(\xi_1, \xi_2) = \xi_1^{-1/s} \log(\xi_1/2)^{-\beta} \in L^s(\Gamma_{m_1,2}, d\nu)$. Then

$$\begin{aligned} |\widehat{f\,d\nu}(x_1, x_2, -t)| &= \left| \int_{[0,1]\times[0,\delta]} e^{i(x_1\xi_1 + x_2\xi_2 - t(\xi_1^{m_1} + c(\xi_2^2 + \mathcal{O}(\xi_2^3))))} \xi_1^{-\frac{1}{s}} \log(\xi_1/2)^{-\beta} \, d(\xi_1, \xi_2) \right| \\ &= \left| \int_0^1 e^{i(x_1\xi_1 - t\xi_1^{m_1})} \xi_1^{-\frac{1}{s}} \log(\xi_1/2)^{-\beta} \, d\xi_1 \right| \left| \int_0^\delta e^{i(x_2\xi_2 - tc(\xi_2^2 + \mathcal{O}(\xi_2^3)))} \, d\xi_2 \right|. \end{aligned}$$

We estimate the first integral by means of Lemma 1.5(ii), and for the second one we use Lemma 1.5(i) (with $m = 2$), which leads to

$$\begin{aligned} \infty > \|f\|_s^p \gtrsim \|\widehat{f\,d\nu}\|_p^p &\gtrsim \int_N^\infty \int_{t^{1/2}}^t \int_1^{t^{1/m_1}} t^{-\frac{p}{s' m_1}} t^{-\frac{p}{2}} (\log t)^{-\beta p} \, dx_1 \, dx_2 \, dt \\ &\approx \int_N^\infty t^{1-\frac{p}{2}} t^{\frac{1}{m_1} - \frac{p}{s' m_1}} (\log t)^{-\beta p} \, dt, \end{aligned}$$

provided N is chosen sufficiently large. This implies that necessarily

$$1 - \frac{p}{2} + \frac{1}{m_1} - \frac{p}{s' m_1} < -1,$$

which is equivalent to

$$\frac{m_1+2}{2} > \frac{2m_1+1}{p} + \frac{1}{s}.$$

Interchanging the roles of ξ_1 and ξ_2, we obtain the same inequality for m_2 and hence for $\bar{m} = m_1 \vee m_2$, and we arrive at (1-7).

Let us finally prove that on the critical line $1/s' = (h+1)/p$ one cannot have strong-type estimates above the bisectrix $1/s = 1/p$, i.e., for $s > p$. In this regime, we find some $1 > r > 0$ such that $1/s < r < 1/p$. Let

$$f(\xi) = \xi_2^{-\frac{m_2}{sh}} |\log(\xi_2/2)|^{-r} \chi_{\{\xi_1^{m_1} \le \xi_2^{m_2}\}}(\xi).$$

It is easy to check that $f \in L^s(\Gamma)$ since $1 < rs$. Now assume $1 \ll x_j^{m_j} \ll t$ for $j = 1, 2$; more precisely choose $N \gg 1$ and assume $N^2 \le N x_j^{m_j} \le t$ for $j = 1, 2$. Then

$$(R^*_{\mathbb{R}^d} f)(x_1, x_2, -t) = \int_0^1 e^{-i(x_2\xi_2 - t\xi_2^{m_2})} \xi_2^{-\frac{m_2}{sh}} |\log(\xi_2/2)|^{-r} \int_0^{\xi_2^{m_2/m_1}} e^{-i(x_1\xi_1 - t\xi_1^{m_1})}\, d\xi_1\, d\xi_2.$$

Since $x_1^{m_1} \ll t$ is equivalent to $(\xi_2^{m_2/m_1} x_1)^{m_1} \ll t\xi_2^{m_2}$, Lemma 1.5(ii) gives

$$\left| \int_0^{\xi_2^{m_2/m_1}} e^{-i(x_1\xi_1 - t\xi_1^{m_1})}\, d\xi_1 \right| = \xi_2^{\frac{m_2}{m_1}} \left| \int_0^1 e^{-i(\xi_2^{m_2/m_1} x_1 \eta - t\xi_2^{m_2}\eta^{m_1})}\, d\eta \right| \gtrsim t^{-\frac{1}{m_1}}.$$

Applying Lemma 1.5 once more, we obtain

$$\left|(R^*_{\mathbb{R}^d} f)(x_1, x_2, -t)\right| \gtrsim t^{-\frac{1}{m_1}} t^{\frac{1}{sh} - \frac{1}{m_2}} \log^{-r}(t/2) = t^{-\frac{1}{p}(1+\frac{1}{h})} \log^{-r}(t/2),$$

where we made use of $1/s' = (h+1)/p$. Thus we get

$$\begin{aligned} \|R^*_{\mathbb{R}^d} f\|_p^p &\gtrsim \int_{N^2}^{\infty} \int_{N^{1/m_2}}^{(t/N)^{1/m_2}} \int_{N^{1/m_1}}^{(t/N)^{1/m_1}} t^{-1-\frac{1}{h}} \log^{-rp}(t/2)\, dx_1\, dx_2\, dt \\ &\approx \int_{N^2}^{\infty} t^{-1} \log^{-rp}(t/2)\, dt = \infty, \end{aligned}$$

since $rp < 1$.

Let us finish this subsection by adding a few more observations and remarks.

(a) First, observe that Γ_0 is a subset of

$$\Gamma_1 = \{(\xi, \phi(\xi)) \in \Gamma : |\xi| \sim 1\}.$$

(b) One can use the dilations $(\xi_1, \xi_2) \mapsto (r^{1/m_1}\xi_1, r^{1/m_2}\xi_2)$, $r > 0$, in order to decompose $Q = [0, 1] \times [0, 1]$ into "dyadic annuli" which, after rescaling, reduces the restriction problem in many situations to the one for Γ_1 (this kind of approach is used extensively in [Ikromov et al. 2010; Ikromov and Müller 2011], as well as in [Ferreyra and Urciuolo 2009]).

Indeed, on the one hand, any restriction estimate on Γ clearly implies the same estimate also for the subsurface Γ_1. On the other hand, the estimates for the dyadic pieces sum up below the sharp critical line (this is the approach in [Ferreyra and Urciuolo 2009]), i.e., when $1/s' > (h+1)/p$. Moreover, in many situations one may apply Bourgain's summation trick in a similar way to that described in Section A1 in order to establish weak-type estimates also when $(1/s, 1/p)$ lies on the critical line, i.e., when $1/s' = (h+1)/p$. However, we shall not pursue this approach here, since it would not give too much of a simplification for us and since our approach (outlined in the next subsection) seems to lead to an even somewhat sharper

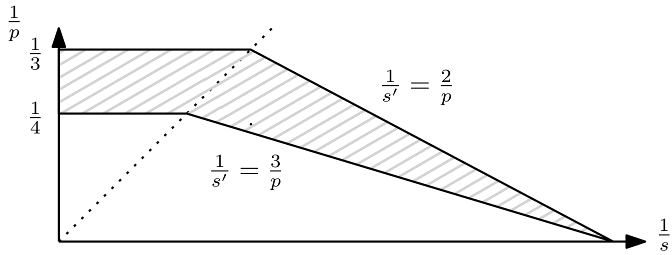


Figure 6. Region on which (1-10) is valid.

result. Moreover, it seems useful and more systematic to understand bilinear restriction estimates for quite general pairs of pieces of our surfaces Γ, and not only the ones which would arise from Γ_1.

(c) On Γ_1, one of the two principal curvatures may vanish, but not both. Notice also that by dividing Γ_1 into a finite number of pieces lying in sufficiently small angular sectors and applying a suitable affine transformation to each of them, we may reduce to surfaces of the form

$$\Gamma_{m,2} = \{(\xi_1, \xi_2, \psi_m(\xi_1) + \xi_2^2 + \mathcal{O}(\xi_2^3)) : \xi_1, \xi_2 \in [0, 1]\},$$

where $\psi_m(\xi_1) \sim \xi_1^m$ as before, with $m = m_1$ or $m = m_2$ (see also our previous discussion of necessary conditions). Applying then a further dyadic decomposition in ξ_1, we see that we may essentially reduce to subsurfaces on which $\xi_1 \sim \varepsilon$, with $\varepsilon > 0$ a small dyadic number. Note that on these we have nonvanishing Gaussian curvature, but the lower bounds of the curvature depend on $\varepsilon > 0$. A rescaling then leads to surfaces of the form

$$P_T = \bigl\{\bigl(\xi_1, \xi_2, \xi_1^2 + \xi_2^2 + \mathcal{O}(\xi_1^3 + T^{-1}\xi_2^3)\bigr) : \xi_1 \in [0, 1], \xi_2 \in [0, T]\bigr\},$$

with $T = \varepsilon^{-m/2} \gg 1$. A prototype of such a situation would be the part of the standard paraboloid lying above a very long-stretched rectangle. Although Fourier restriction estimates for the paraboloid have been studied extensively, the authors are not aware of any results that would give the right control on the dependence on the parameter $T \gg 1$. Indeed, one can prove that the lower bound

$$\|R_T^*\|_{L^s(P_T)\to L^p(\mathbb{R}^3)} \gtrsim T^{(\frac{1}{p}-\frac{1}{s})_+} \tag{1-10}$$

for the adjoint restriction operator $R_T^* = R_{P_T,\mathbb{R}^2}^*$ associated to Lebesgue measure on P_T holds true for all s and p for which $(1/s, 1/p)$ lies within the shaded region in Figure 6, and a reasonable conjecture is that also the reverse inequality essentially holds true, maybe up to an extra factor T^δ, i.e., that

$$\|R_T^*\|_{L^s(P_T)\to L^p(\mathbb{R}^3)} \le C_\delta T^{\delta+(\frac{1}{p}-\frac{1}{s})_+} \tag{1-11}$$

for every $\delta > 0$.

We give some hints as to why (1-10) holds true and why the inverse inequality (with δ-loss) seems a reasonable conjecture. Let $d\nu_T$ denote the "Lebesgue measure" on P_T. Then by Lemma 1.5,

$$|\widehat{d\nu_T}(x_1, x_2, t)| \gtrsim t^{-\frac{1}{2}} \int_0^T e^{i(x_2\xi_2 + t[\xi_2^2 + \mathcal{O}(T^{-1}\xi_2^3)])}\, d\xi_2 = Tt^{-\frac{1}{2}} \int_0^1 e^{i(x_2T\eta + tT^2[\eta^2 + \mathcal{O}(\eta^3)])}\, d\eta \gtrsim t^{-1},$$

provided $x_1 \ll t$ and $x_2 \ll Tt$ (we may arrange matters in the preceding reductions so that the error term $\mathcal{O}(\eta^3)$ is small compared to η^2). Hence, since we assume $p > 3$,

$$\|\widehat{d\nu_T}\|_{L^p(\mathbb{R}^3)} \gtrsim T^{\frac{1}{p}}.$$

Obviously $\|1\|_{L^s(P_T, d\nu_T)} = T^{1/s}$, so we see that

$$\|R_T^*\|_{L^s(P_T)\to L^p(\mathbb{R}^3)} \gtrsim T^{\left(\frac{1}{p}-\frac{1}{s}\right)}.$$

Restricting P_T to the region where $\xi_2 \le 1$, we see that also $\|R_T^*\|_{L^s(P_T)\to L^p(\mathbb{R}^3)} \gtrsim 1$, and combining these two lower bounds gives (1-10).

On the other hand, from Remark 4, (2.4) in [Ferreyra and Urciuolo 2009] we easily obtain by an obvious rescaling argument that for $1/s' = 3/p$ and $p > 4$ (hence $1/p < 1/s$), we have

$$\|R_T^*\|_{L^s(P_T, d\nu_T)\to L^p(\mathbb{R}^3)} \le C,$$

uniformly in T. It is conjectured that for the entire paraboloid $\mathcal{P} = \{(\xi_1, \xi_2, \xi_1^2 + \xi_2^2) : (\xi_1, \xi_2) \in \mathbb{R}^2\}$, the adjoint restriction operator $R^*_{\mathcal{P},\mathbb{R}^2}$ is bounded for $1/s' = 2/p$ and $p > 3$ (hence $1/p < 1/s$). It would be reasonable to expect the same kind of behavior for suitable perturbations of the paraboloid, and subsets of those, such as P_T (maybe with an extra factor T^δ for any $\delta > 0$). By complex interpolation, the previous estimate in combination with the latter conjectural estimate would lead to

$$\|R_T^*\|_{L^s(P_T, d\nu_T)\to L^p(\mathbb{R}^3)} \le C_\delta T^\delta$$

for every $\delta > 0$, provided that $1/p < 1/s$ and $2/p < 1/s' < 3/p$. In combination with a trivial application of Hölder's inequality this leads to the conjecture (1-11),

$$\|R_T^*\|_{L^s(P_T)\to L^p(\mathbb{R}^3)} \le C_\delta T^{\delta + \left(\frac{1}{p}-\frac{1}{s}\right)_+}$$

for every $\delta > 0$, provided $(1/s, 1/p)$ lies within the shaded region in Figure 6.

1D. ***The strategy of the approach.*** We will study certain bilinear operators. For a suitable pair of subsurfaces $S_1, S_2 \subset S$ (we will be more specific on this point later), we seek to establish bilinear estimates

$$\|R^*_{\mathbb{R}^2} f_1 R^*_{\mathbb{R}^2} f_2\|_{L^p(\mathbb{R}^3)} \le C_p C(S_1, S_2)\, \|f_1\|_{L^2(S_1)} \|f_2\|_{L^2(S_2)},$$

for functions f_1, f_2 supported in S_1 and S_2, respectively.

For hypersurfaces with nonvanishing Gaussian curvature and principal curvatures of the same sign, the sharp estimates of this type, under the appropriate transversality assumption, appeared in [Tao 2003b] (after previous partial results in [Tao et al. 1998; Tao and Vargas 2000a]). For the light cone in any dimension, the analogous results were established in [Wolff 2001; Tao 2001a] (improving on earlier results in [Bourgain 1995a; Tao and Vargas 2000a]). For the case of principal curvatures of different sign, or with a smaller number of nonvanishing principal curvatures, sharp bilinear results are also known [Lee 2006; Vargas 2005; Lee and Vargas 2010].

What is crucial for us is to know how the constant $C(S_1, S_2)$ explicitly depends on the pair of surfaces S_1 and S_2, in order to be able to sum all the bilinear estimates that we obtain for pairs of pieces of our given surface, to pass to a linear estimate. Classically, this is done by proving a bilinear estimate for one "generic" class of subsurfaces. For instance, if S is the paraboloid, then other pairs of subsurfaces can be reduced to it by means suitable affine transformations and homogeneous rescalings. However, general surfaces do not come with such a kind of self-similarity under these transformations, and it is one of the features of this article that we establish new, very precise bilinear estimates.

The bounds on the constant $C(S_1, S_2)$ that we establish will depend on the size of the domains and local principal curvatures of the subsurfaces, and we shall have to keep track of these during the whole proof. In this sense, many of the lemmas are generalized, quantitative versions of well-known results from classical bilinear theory.

The pairs of subsurfaces that we would like to discuss are pieces of the surface sitting over two dyadic rectangles and satisfying certain separation or "transversality" assumptions. However, such a rectangle might touch one of the axes, where some principle curvature is vanishing. In this case we will decompose dyadically a second time. But even on these smaller sets, we do not have the correct "transversality" conditions; we first have to find a proper rescaling such that the scaled subsurfaces allow us to run the bilinear machinery.

The following section will begin with the bilinear argument to provide us with a very general bilinear result for sufficiently "good" pairs of surfaces. In the subsequent section, we construct a suitable scaling in order to apply this general result to our situation. After rescaling and several additional arguments, we pass to a global bilinear estimate and finally proceed to the linear estimate.

A few more remarks on the notion will be useful: as mentioned before, it is very important to know precisely how the constants depend on the specific choice of subsurfaces. Moreover, there will appear other constants, depending possibly on m_1, m_2, p, q, or other quantities, but not explicitly on the choice of subsurfaces. We will not keep track of such types of constants, since it would even set a false focus and distract the reader. Instead we will simply use the symbol $\lesssim$ for an inequality involving one of these constants of minor importance. To be more precise on this, later we introduce a family of pairs of subsurfaces $\mathcal{S}_0$. Then for quantities $A, B : \mathcal{S}_0 \to \mathbb{R}$ the inequality $A \lesssim B$ means there exists a constant $C > 0$ such that $A(S_1, S_2) \leq CB(S_1, S_2)$ uniformly for all $(S_1, S_2) \in \mathcal{S}_0$.

Moreover, we will also use the notation $A \sim B$ if $A \lesssim B$ and $B \lesssim A$. We will even use this notation for vectors, meaning their entries are comparable in each coordinate. Similarly, we write $A \ll B$ if there exists a constant $c > 0$ such that $A(S_1, S_2) \leq cB(S_1, S_2)$ for all $(S_1, S_2) \in \mathcal{S}_0$ and c is "small enough" for our purposes. This notion of being "sufficiently small" will in general depend on the situation and further constants, but the choice will be uniform in the sense that it will work for all pairs of subsurfaces in the class $\mathcal{S}_0$.

The inner product of two vectors x, y will usually be denoted by xy or $x \cdot y$, and occasionally also by $\langle x, y \rangle$.

2. General bilinear theory

2A. *Wave packet decomposition.* We begin with what is basically a well-known result, although we need a more quantitative version (cf. [Tao 2003a; Lee 2006]).

Lemma 2.1. *Let $U \subset \mathbb{R}^d$ be an open and bounded subset, and let $\phi \in C^\infty(U, \mathbb{R})$. We assume there exist constants $\kappa > 0$ and $D \le 1/\kappa$ such that $\|\partial^\alpha \phi\|_\infty \le A_\alpha \kappa D^{2-|\alpha|}$ for all $\alpha \in \mathbb{N}^d$ with $|\alpha| \ge 2$. Then for every $R \ge 1$ there exists a wave packet decomposition adapted to ϕ with tubes of radius $R/D = R'$ and length $R^2/(D^2\kappa) = (R')^2/\kappa$, where we have put $R = R'D$.*

More precisely, consider the index sets $\mathcal{Y} = R'\mathbb{Z}^d$ and $\mathcal{V} = (R')^{-1}\mathbb{Z}^d \cap U$, and define for $w = (y, v) \in \mathcal{Y} \times \mathcal{V} = \mathcal{W}$ the tube

$$T_w = \left\{(x,t) \in \mathbb{R}^d \times \mathbb{R} : |t| \le \frac{(R')^2}{\kappa},\ |x - y + t\nabla\phi(v)| \le R'\right\}. \tag{2-1}$$

Then, given any function $f \in L^2(U)$, there exist functions (wave packets) $\{p_w\}_{w\in\mathcal{W}}$ and coefficients $c_w \in \mathbb{C}$ such that $R^*_{\mathbb{R}^d} f$ can be decomposed as

$$R^*_{\mathbb{R}^d} f(x,t) = \sum_{w\in\mathcal{W}} c_w p_w(x,t)$$

for every $t \in \mathbb{R}$ with $|t| \le (R')^2/\kappa$, in such a way that the following hold true:

(P1) $p_w = R^*_{\mathbb{R}^d}(\mathfrak{F}^{-1}_{\mathbb{R}^d}(p_w(\cdot, 0)))$.

(P2) $\operatorname{supp} \mathfrak{F}_{\mathbb{R}^{d+1}} p_w \subset B((v, \phi(v)), 2/R')$.

(P3) *p_w is essentially supported in T_w; i.e.,*

$$|p_w(x,t)| \le C_N (R')^{-\frac{d}{2}} \left(1 + \frac{|x - y + t\nabla\phi(v)|}{R'}\right)^{-N}$$

for every $N \in \mathbb{N}$. In particular, $\|p_w(\cdot, t)\|_2 \lesssim 1$.

(P4) *For all $W \subset \mathcal{W}$, we have $\left\|\sum_{w\in W} p_w(\cdot, t)\right\|_2 \lesssim |W|^{\frac{1}{2}}$.*

(P5) $\|c\|_{\ell^2} \lesssim \|f\|_{L^2}$.

Moreover, the constants arising explicitly (such as the C_N) or implicitly in these estimates can be chosen to depend only on the constants A_α but no further on the function ϕ, and also not on the other quantities R, D and κ (such constants will be called admissible).

Remarks 2.2. (i) Notice that no bound is required on $\nabla\phi$ at this stage; however, such bounds will become important later (for instance in (iii)).

(ii) Denote by $N(v)$ the normal vector at $(v, \phi(v))$ to the graph of ϕ which is given by $N(v) = (-\nabla\phi(v), 1)$. Since $(R')^2/\kappa \ge R'$, we may thus rewrite

$$T_w = (y, 0) + \left\{tN(v) : |t| \le \frac{(R')^2}{\kappa}\right\} + \mathcal{O}(R').$$

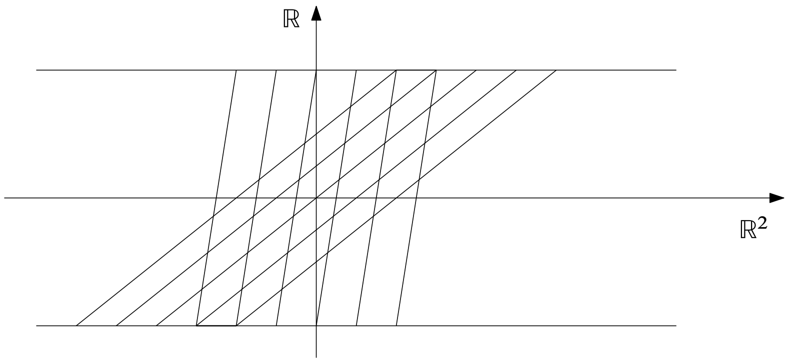


Figure 7. The tubes T_w fill a horizontal strip.

Moreover,

$$|x - y + t\nabla\phi(v)| = |(x,t) - (y,0) - tN(v)| \geq \operatorname{dist}((x,t), T_w).$$

It is then easily seen that (P3) can be rewritten as

$$|p_w(z)| \leq C_N (R')^{-\frac{d}{2}} \left(1 + \frac{\operatorname{dist}(z, T_w)}{R'}\right)^{-N}$$

for all $z \in \mathbb{R}^{d+1}$ with $|\langle z, e_{d+1}\rangle| \leq (R')^2/\kappa$, where e_{d+1} denotes the last vector of the canonical basis of $\mathbb{R}^{d+1}$. This justifies the statement that "p_w is essentially supported in T_w".

(iii) Notice further that we can reparametrize the wave packets by lifting $\mathcal{V}$ to $\widetilde{\mathcal{V}} = \{(v, \phi(v)) : v \in \mathcal{V}\} \subset S$. If we now assume $\|\nabla\phi\| \lesssim 1$, then we have $|(v,\phi(v)) - (v', \phi(v'))| \sim |v - v'|$, and thus $\widetilde{\mathcal{V}}$ becomes an $(R')^{-1}$-net in S. Finally, we shall identify a parameter $y \in \mathbb{R}^d$ with the point $(y, 0)$ in the hyperplane $\mathbb{R}^d \times \{0\}$.

Proof of Lemma 2.1. We will basically follow the proof by Lee [2006]; the only new feature consists in elaborating the precise role of the constant κ.

Let $\psi, \hat{\eta} \in C_0^\infty(B(0,1))$ be chosen in such a way that for $\eta_y(x) = \eta((x-y)/R')$, $\psi_v(\xi) = \psi(R'(\xi - v))$ we have $\sum_{v\in\mathcal{V}} \psi_v = 1$ on U and $\sum_{y\in\mathcal{Y}} \eta_y = 1$. We also choose a slightly bigger function $\tilde{\psi} \in C_0^\infty(B(0,3))$ such that $\tilde{\psi} = 1$ on $B(0,2) \supset \operatorname{supp}\psi + \operatorname{supp}\hat{\eta}$, and put $\tilde{\psi}_v(\xi) = \tilde{\psi}(R'(\xi - v))$. Then the functions

$$F_{(y,v)} = \mathfrak{F}^{-1}_{\mathbb{R}^d}(\widehat{\psi_v f\, \eta_y}) = (\psi_v f) * \check{\eta}_y, \quad y \in \mathcal{Y},\ v \in \mathcal{V},$$

are essentially well localized in both position and momentum/frequency space. Define $q_w = R^*_{\mathbb{R}^d}(F_w)$, $w = (y, v) \in \mathcal{W}$; up to a certain factor c_w, which will be determined later, these are already the announced wave packets, i.e., $q_w = c_w p_w$.

Since $f = \sum_{w\in W} F_w$, we then have the decomposition $R^*_{\mathbb{R}^d} f = \sum_{w \in W} q_w$. Let us concentrate on property (P3) — the other properties are then rather easy to establish. Since $\operatorname{supp} F_{(y,v)} \subset B(v, 2/R')$, we have, for every $w = (y, v) \in \mathcal{W}$,

$$\begin{aligned} q_w(x,t) &= \int e^{-i(x\xi+t\phi(\xi))} F_w(\xi)\,d\xi = \int e^{-i(x\xi+t\phi(\xi))} F_w(\xi)\tilde{\psi}_v(\xi)\,d\xi \\ &= (2\pi)^{-d} \iint e^{-i\left((x-z)\xi+t\phi(\xi)\right)}\tilde{\psi}_v(\xi)\,d\xi\,\widehat{F}_w(z)\,dz \\ &= (2\pi)^{-d}(R')^{-d} \iint e^{-i\left((x-z)(\frac{\xi}{R'}+v)+t\phi(\frac{\xi}{R'}+v)\right)}\tilde{\psi}(\xi)\,d\xi\,\widehat{F}_w(z)\,dz \\ &= (2\pi)^{-d}(R')^{-d} \int K(x-z,t)\widehat{F}_w(z)\,dz, \end{aligned}$$

with the kernel

$$K(x,t) = \int e^{i\left(x(\frac{\xi}{R'}+v)+t\phi(\frac{\xi}{R'}+v)\right)}\tilde{\psi}(\xi)\,d\xi.$$

We claim that

$$|K(x,t)| \lesssim \left(1 + \frac{|x+t\nabla\phi(v)|}{R'}\right)^{-N} \tag{2-2}$$

for every $N \in \mathbb{N}$. To this end, we shall estimate the oscillatory integral

$$K_\lambda = \int e^{i\lambda\Phi(\xi)}\tilde{\psi}(\xi)\,d\xi,$$

with phase

$$\Phi(\xi) = \frac{x\left(\frac{\xi}{R'}+v\right)+t\phi\left(\frac{\xi}{R'}+v\right)}{1+(R')^{-1}|x+t\nabla\phi(v)|},$$

where we put $\lambda = 1 + (R')^{-1}|x+t\nabla\phi(v)|$. In order to prove (2-2), we may assume $|x+t\nabla\phi(v)| \gg R'$. Then integrations by parts will lead to $|K_\lambda| \lesssim \lambda^{-N}$ for all $N \in \mathbb{N}$, hence to (2-2), provided we can show that

$$|\nabla\Phi(\xi)| \sim 1 \quad \text{for all } \xi, \tag{2-3}$$

$$\|\partial^\alpha\Phi\|_\infty \lesssim 1 \quad \text{for all } \alpha \geq 2, \tag{2-4}$$

and that the constants in these estimates are admissible. But,

$$\frac{|t|\left|\nabla\phi\left(\frac{\xi}{R'}+v\right)-\nabla\phi(v)\right|}{|x+t\nabla\phi(v)|} \ll \frac{|t|\left|\nabla\phi\left(\frac{\xi}{R'}+v\right)-\nabla\phi(v)\right|}{R'} \lesssim \frac{|t|}{(R')^2}\|\phi''\|_\infty \leq \frac{1}{\kappa}\|\phi''\|_\infty \leq 1$$

for every $\xi \in \operatorname{supp}\tilde{\psi}$, hence

$$|t|\left|\nabla\phi\left(\frac{\xi}{R'}+v\right)-\nabla\phi(v)\right| \ll |x+t\nabla\phi(v)|.$$

Thus

$$|\nabla\Phi(\xi)| = \frac{\left|x+t\nabla\phi\left(\frac{\xi}{R'}+v\right)\right|}{R'+|x+t\nabla\phi(v)|} = \frac{\left|x+t\nabla\phi(v)-t\left[\nabla\phi(v)-\nabla\phi\left(\frac{\xi}{R'}+v\right)\right]\right|}{R'+|x+t\nabla\phi(v)|} \sim \frac{|x+t\nabla\phi(v)|}{R'+|x+t\nabla\phi(v)|} \sim 1,$$

which verifies (2-3). And, for $|\alpha| \geq 2$ we have

$$|\partial^\alpha \Phi(\xi)| \leq \left| t(R')^{-|\alpha|}(\partial^\alpha \phi)\left(\frac{\xi}{R'} + v\right)\right| \lesssim \frac{(R')^2}{\kappa}(R')^{-|\alpha|}\kappa D^{2-|\alpha|} \leq (DR')^{2-|\alpha|} = R^{2-|\alpha|} \leq 1,$$

which gives (2-4). It is easily checked that the constants in these estimates can be chosen to be admissible. Following the proof in [Lee 2006], we conclude that

$$\begin{aligned} |q_w(x,t)| &\lesssim (R')^{-d} \int \big|K(x-z-y,t)\widehat{F}_w(z+y)\big|\,dz \\ &= (R')^{-d} \int \left|K(x-z-y,t)\eta\left(\frac{z}{R'}\right)\widehat{\psi_v f}(z+y)\right|dz \\ &\lesssim \left(1+\frac{|x-y+t\nabla\phi(v)|}{R'}\right)^{-N} M(\widehat{\psi_v f})(y), \end{aligned}$$

where M denotes the Hardy–Littlewood maximal operator. Thus, we obtain (P3) by choosing $c_w = c_{(y,v)} = (R')^{d/2} M(\widehat{\psi_v f})(y)$.

Properties (P1) and (P2) follow from the definition of the wave packets. From (P2) and (P3) we can deduce (P4). For (P5), we refer to [Lee 2006]. □

In view of our previous remarks, it is easy to restate Lemma 2.1 in a more coordinate-free way. For any given hyperplane $H = n^\perp \subset \mathbb{R}^{d+1}$, with n a unit vector (so that $\mathbb{R}^{d+1} = H + \mathbb{R}n$), define the partial Fourier (co)transform

$$\mathfrak{F}_H^{-1} f(\xi + tn) = \int_H f(x+tn)e^{ix\cdot\xi}\,dx, \quad \xi \in H,\ t \in \mathbb{R}.$$

Moreover, if $U \subset H$ is open and bounded, and if $\phi_H \in C^\infty(U,\mathbb{R})$ is given, then consider the smooth hypersurface $S = \{\eta + \phi_H(\eta)n : \eta \in U\} \subset \mathbb{R}^{d+1}$, and define the corresponding Fourier extension operator

$$R_H^* f(x+tn) = \int_U f(\eta)e^{-i(x\eta + t\phi_H(\eta))}\,d\eta = \int_U f(\eta)e^{-i\langle x+tn,\eta+\phi_H(\eta)n\rangle}\,d\eta$$

for $(x,t) \in H \times \mathbb{R}$ and $f \in L^2(U)$. Notice that $R^*_{\mathbb{R}^d}$ corresponds to the special case $H = \mathbb{R}^d \times \{0\}$, and thus by means of a suitable rotation, mapping e_{d+1} to n, we immediately obtain the following.

Corollary 2.3 (wave packet decomposition). *Let $U \subset H$ be an open and bounded subset, and let $\phi_H \in C^\infty(U,\mathbb{R})$. We assume that there are constants $\kappa > 0$ and $D \leq 1/\kappa$ such that $\|\phi_H^{(l)}\|_\infty \leq A_l \kappa D^{2-l}$ for every $l \in \mathbb{N}$ with $l \geq 2$, where $\phi_H^{(l)}$ denotes the total derivative of ϕ_H of order l, and in addition that $\|\phi'\|_\infty \leq A$. Then for every $R \geq 1$ there exists a wave packet decomposition adapted to S and the decomposition of $\mathbb{R}^{d+1}$ into $\mathbb{R}^{d+1} = H + \mathbb{R}n$, with tubes of radius $R/D = R'$ and length $R^2/(D^2\kappa) = (R')^2/\kappa$, where $R = R'D$.*

More precisely, there exists an R'-lattice $\mathcal{Y}$ in H and an $(R')^{-1}$-net $\mathcal{V}$ in S such that the following hold true: if we denote by $\mathcal{W}$ the index set $\mathcal{W} = \mathcal{Y} \times \mathcal{V}$ and associate to $w = (y,v) \in \mathcal{Y} \times \mathcal{V} = \mathcal{W}$ the tube-like set

$$T_w = y + \left\{tN(v) : |t| \leq \frac{(R')^2}{\kappa}\right\} + B(0,R'), \tag{2-5}$$

then for every given function $f \in L^2(U)$ there exist functions (wave packets) $\{p_w\}_{w\in\mathcal{W}}$ and coefficients $c_w \in \mathbb{C}$ such that for every $x = x' + tn \in \mathbb{R}^{d+1}$ with $|t| \le (R')^2/\kappa$ and $x' \in H$, we may decompose $R_H^* f(x)$ as

$$R_H^* f(x) = \sum_{w\in\mathcal{W}} c_w p_w(x),$$

in such a way that the following hold true:

(P1′) $p_w = R_H^*(\mathfrak{F}_H^{-1}(p_w|_H))$.

(P2′) $\operatorname{supp}\mathfrak{F}_{\mathbb{R}^{d+1}} p_w \subset B(v,(R')^{-1})$ *and* $\operatorname{supp}\mathfrak{F}_H(p_w(\cdot + tn)) \subset B(v', \mathcal{O}((R')^{-1}))$,*where v' denotes the orthogonal projection of $v \in S$ to H.*

(P3′) *p_w is essentially supported in T_w; i.e.,*

$$|p_w(x)| \le C_N (R')^{-1}\left(1 + \frac{\operatorname{dist}(x,T_w)}{R'}\right)^{-N}.$$

(P4′) *For all $W \subset \mathcal{W}$, we have* $\|\sum_{w\in W} p_w(\cdot + tn)\|_{L^2(H)} \lesssim |W|^{\frac{1}{2}}$.

(P5′) $\|c\|_{\ell^2} \lesssim \|f\|_{L^2}$.

Moreover, the constants arising in these estimates can be chosen to depend only on the constants A_l and A, but no further on the function ϕ_H, and also not on the other quantities R, D and κ (such constants will be called admissible).

Notice that, unlike as in Lemma 2.1, we may here choose an $(R')^{-1}$-net in S in place of an R' -lattice in H for the parameter set $\mathcal{V}$, because of our assumed bound on ϕ_H'.

It will become important that under suitable additional assumptions on the position of a given hyperplane H, we may reparametrize a given smooth hypersurface $S = \{(\xi, \phi(\xi)) : \xi \in U\}$ (where U is an open subset of $\mathbb{R}^d$) also of the form

$$S = \{\eta + \phi_H(\eta)n : \eta \in U_H\},$$

where U_H is an open subset of H and $\phi_H \in C^\infty(U_H, \mathbb{R})$.

Lemma 2.4 (reparametrization). *Let $H_1 = n_1^\perp$ and $H_2 = n_2^\perp$ be two hyperplanes in $\mathbb{R}^{d+1}$, where n_1 and n_2 are given unit vectors. Let $K = H_1 \cap H_2$, and choose unit vectors h_1, h_2 orthogonal to K such that $H_1 = K + \mathbb{R}h_1$ and $H_2 = K + \mathbb{R}h_2$. Let $U_1 \subset H_1$ be an open bounded subset such that for every $x' \in K$, the section $U_1^{x'} = \{u \in \mathbb{R} : x' + uh_1 \subset U_1\}$ is an (open) interval, and let $\phi_1 \in C^\infty(U_1, \mathbb{R})$ satisfy the assumptions of Corollary 2.3. Setting $B = \kappa D^2$ and $r = D^{-1}$, an equivalent way to state this is that there are constants $B, r > 0$ such that $Br \le 1$, $\|\phi_1'\|_\infty \le A$ and $\|\phi_1^{(l)}\|_\infty \le A_l B r^l$ for every $l \in \mathbb{N}$ with $l \ge 2$. Denote by S the hypersurface*

$$S = \{\eta + \phi_1(\eta)n_1 : \eta \in U_1\} \subset \mathbb{R}^{d+1},$$

and again by $v \mapsto N(v)$ the corresponding unit normal field on S.

Assume furthermore that the vector n_2 is transversal to S; i.e., $|\langle n_2, N(v)\rangle| \geq a > 0$ for all $v \in S$. Then there exist an open bounded subset $U_2 \subset H_2$ such that for every $x' \in K$, the section $U_2^{x'} = \{s \in \mathbb{R} : x' + sh_2 \in U_2\}$ is an interval, and a function $\phi_2 \in C^\infty(U_2, \mathbb{R})$ so that we may rewrite

$$S = \{\xi + \phi_2(\xi)n_2 : \xi \in U_2\}. \tag{2-6}$$

Moreover, the derivatives of ϕ_2 satisfy estimates of the same form as those of ϕ_1, up to multiplicative constants which are admissible, i.e., which depend only on the constants A_l, A and a.

Finally, given any $f_1 \in L^2(U_1)$, there exists a unique function $f_2 \in L^2(U_2)$ such that

$$R^*_{H_1} f_1 = R^*_{H_2} f_2, \tag{2-7}$$

and $\|f_1\|_2 \sim \|f_2\|_2$, where the constants in these estimates are admissible.

Proof. Assume that (2-6) holds true. Then, given any point $\eta + \phi_H(\eta)n_1 \in S$, with $\eta = x' + uh_1 \in U_1$, $x' \in K$, we find some $\xi = x' + sh_2 \in U_2$ such that

$$x' + uh_1 + \phi_1(x' + uh_1)n_1 = x' + sh_2 + \phi_2(x' + sh_2)n_2, \tag{2-8}$$

which shows that necessarily

$$s = \langle uh_1 + x' + \phi_1(x' + uh_1)n_1, h_2\rangle. \tag{2-9}$$

Let us therefore define the mapping $G : U_1 \to H_2$ by

$$G(x' + uh_1) = x' + \langle uh_1 + x' + \phi_1(x' + uh_1)n_1, h_2\rangle h_2.$$

Moreover, fixing an orthonormal basis $E_1, \dots, E_{d-1}$ of K and extending this by the vector h_1 or h_2 in order to obtain bases of H_1 and H_2 respectively and working in the corresponding coordinates, we may assume without loss of generality that U_1 is an open subset of $\mathbb{R}^{d-1} \times \mathbb{R}$, since $\dim K = d - 1$, and that G is a mapping $G : U_1 \to \mathbb{R}^{d-1} \times \mathbb{R}$, given by

$$G(x', u) = (x', g(x', u)),$$

where

$$g(x', u) = \langle x' + uh_1 + \phi_1(x', u)n_1, h_2\rangle.$$

To show that G is a diffeomorphism onto its image $U_2 = G(U_1)$, observe that

$$\partial_u G(x', u) = (0, \partial_u g(x', u)) = \big(0, \langle h_1 + \partial_u \phi_1(x', u)n_1, h_2\rangle\big).$$

On the other hand, the vector

$$N_0 = -\partial_u \phi_1(x', u)h_1 - \sum_{j=1}^{k} \partial_{x_j} \phi(x', u)E_j + n_1$$

is normal to S at the point $x' + uh_1 + \phi_1(x' + uh_1)n_1$ $\big($here $x' = \sum_{j=1}^{d-1} x_j E_j\big)$, and $|N_0| \sim 1$. Thus, our transversality assumption implies

$$|\langle -\partial_u \phi_1(x', u)h_1 + n_1, n_2\rangle| \gtrsim a > 0. \tag{2-10}$$

But, $\{h_j, n_j\}$ forms an orthonormal basis of $K^\perp$ for $j = 1, 2$, and thus, rotating all these vectors by an angle of $\pi/2$, we see that (2-10) is equivalent to $|\langle \partial_u \phi_1(x', u) n_1 + h_1, h_2\rangle| \gtrsim a > 0$, so that

$$|\partial_u g(x', u)| \gtrsim a > 0.$$

Given the special form of G, this also implies

$$|\det G'(x', u)| = |\partial_u g(x', u)| \gtrsim a > 0.$$

Consequently, for x' fixed, the mapping $u \mapsto g(x', u)$ is a diffeomorphism from the interval $U_1^{x'}$ onto an open interval $U_2^{x'}$, and thus G is bijective onto its image U_2, in fact even a diffeomorphism, and U_2 fibers into the intervals $U_2^{x'}$. Indeed, the inverse mapping $F = G^{-1} : U_2 \to U_1$ of G is also of the form

$$F(x', s) = (x', f(x', s)),$$

where

$$g(x', f(x', s)) = s. \tag{2-11}$$

In combination with (2-8) this shows that (2-6) holds indeed true, with

$$\phi_2(x', s) = f(x', s)\langle h_1, n_2\rangle + \phi_1(F(x', s))\langle n_1, n_2\rangle. \tag{2-12}$$

Moreover, if $f_1 \in L^2(U_1)$, then, by (2-8) and a change of coordinates,

$$\begin{aligned} R^*_{H_1} f_1(y) &= \iint_{U_1} f_1(x', u) e^{-i\langle y, x' + u h_1 + \phi_1(x', u) n_1\rangle}\, dx'du \\ &= \iint_{U_2} f_1(F(x', s)) |\det F'(x', s)|\, e^{-i\langle y, x' + s h_2 + \phi_2(x', s) n_2\rangle}\, dx'ds, \end{aligned}$$

so that (2-7) holds true, with

$$f_2(x' + s h_2) = f_1(x' + f(x', s) h_1)\, |\det F'(x', s)|. \tag{2-13}$$

Our estimates for derivatives of F show that $|\det F'(x', s)| \sim 1$, with admissible constants, so that in particular $\|f_1\|_2 \sim \|f_2\|_2$.

What remains is the control of the derivatives of ϕ_2. This somewhat technical part of the proof will be based on Faà di Bruno's theorem and is deferred until the Appendix (see Section A2). $\square$

We shall from now on restrict ourselves to dimension $d = 2$. The following lemma will deal with the separation of tubes along certain types of curves, for a special class of 2-hypersurfaces. It will later be applied to intersection curves of two hypersurfaces.

Lemma 2.5 (tube-separation along the intersection curve). *Let $\mathcal{Y}$, $\mathcal{V}$, $\mathcal{W}$, R, T_w be as in Corollary 2.3. Moreover assume $\phi \in C^\infty(U, \mathbb{R})$, $U \subset \mathbb{R}^2$, such that $\partial_i^2 \phi(x) \sim \kappa_i$ for all $x \in U$, $i = 1, 2$, and $\partial_1\partial_2\phi = 0$. Define $\kappa = \kappa_1 \vee \kappa_2$. Let $\gamma = (\gamma_1, \gamma_2)$ be a curve in U with $|\dot\gamma_i| \sim 1$ for $i = 1, 2$. Then for all pairs of points $v_1, v_2 \in \operatorname{im}(\gamma) + \mathcal{O}((R')^{-1})$ such that $v_1 - v_2 = j/R'$, where $j \in \mathbb{Z}^2$ and $|j| \gg 1$, the following separation condition holds true (again with constants in these estimates which are admissible in the* *obvious sense*):

$$|\nabla\phi(v_1)-\nabla\phi(v_2)|\sim|j|\frac{R'}{(R')^2/\kappa}.$$

Proof. Choose t_1, t_2 such that $v_i=\gamma(t_i)+\mathcal{O}((R')^{-1})$. Then

$$|\nabla\phi(v_i)-\nabla\phi(\gamma(t_i))|\le\|\phi''|_U\|_\infty|v_i-\gamma(t_i)|\lesssim\frac{\kappa}{R'}. \tag{2-14}$$

Therefore

$$\begin{aligned}\frac{|j|}{R'}&=|v_1-v_2|=|\gamma(t_1)-\gamma(t_2)|+\mathcal{O}((R')^{-1})\\&\sim|\dot\gamma_1||t_1-t_2|+|\dot\gamma_2||t_1-t_2|+\mathcal{O}((R')^{-1})\sim|t_1-t_2|+\mathcal{O}((R')^{-1}),\end{aligned}$$

and since $|j|\gg 1$, we see that $|t_1-t_2|\sim|j|/R'$. By our assumptions on ϕ and (2-14), we thus see that there exist s_1 and s_2 lying between t_1 and t_2 such that

$$\begin{aligned}|\nabla\phi(v_1)-\nabla\phi(v_2)|&\ge|\nabla\phi(\gamma(t_1))-\nabla\phi(\gamma(t_2))|-\kappa\mathcal{O}((R')^{-1})\\&\sim\big(|\partial_1^2\phi(\gamma(s_1))\dot\gamma(s_1)|+|\partial_2^2\phi(\gamma(s_2))\dot\gamma(s_2)|\big)\,|t_1-t_2|-\kappa\mathcal{O}((R')^{-1})\\&\sim(\kappa_1+\kappa_2)\frac{|j|}{R'}+\kappa\mathcal{O}((R')^{-1})\sim|j|\frac{\kappa}{R'},\end{aligned}$$

where we used again that $|j|\gg 1$. □

2B. ***A bilinear estimate for normalized hypersurfaces.*** In this section, we shall work under the following:

General Assumptions. Let $\phi\in C^\infty(\mathbb{R}^2)$ such that $\partial_1\partial_2\phi\equiv 0$, and let

$$S_j=\{(\eta,\phi(\eta):\eta\in U_j\},\quad U_j=r^{(j)}+[0,d_1^{(j)}]\times[0,d_2^{(j)}],\quad j=1,2,$$

where $r^{(j)}\in\mathbb{R}^2$ and $d_1^{(j)},d_2^{(j)}>0$. We assume the principal curvature of S_j in the direction of η_1 is comparable to $\kappa_1^{(j)}>0$, and in the direction of η_2 to $\kappa_2^{(j)}>0$, up to some fixed multiplicative constants. We then put for $j=1,2$,

$$\begin{gathered}\kappa^{(j)}=\kappa_1^{(j)}\vee\kappa_2^{(j)},\quad\bar\kappa_i=\kappa_i^{(1)}\vee\kappa_i^{(2)},\quad\bar\kappa=\bar\kappa_1\vee\bar\kappa_2=\kappa^{(1)}\vee\kappa^{(2)},\\ \bar d_i=d_i^{(1)}\vee d_i^{(2)},\quad D=\min_{i,j}d_i^{(j)}.\end{gathered} \tag{2-15}$$

The vector field $N=(-\nabla\phi,1)$ is normal to S_1 and S_2, and thus $N_0=N/|N|$ is a unit normal field to these hypersurfaces. We make the following additional assumptions:

(i) For all $i,j=1,2$ and all $\eta\in U_j$, we have

$$|\partial_i\phi(\eta)-\partial_i\phi(r^{(j)})|\lesssim\kappa_i^{(j)}d_i^{(j)}\quad\text{and}\quad\bar\kappa_i\bar d_i\lesssim 1 \tag{2-16}$$

(notice that the first inequality follows already from our earlier assumptions).

(ii) For all $\eta\in U_1\cup U_2$ and for all $\alpha\in\mathbb{N}^2$, $|\alpha|\ge 2$, we have $|\partial^\alpha\phi(\eta)|\lesssim\bar\kappa D^{2-|\alpha|}$.

(iii) For $i=1,2$, i.e., with respect to both variables, the following separation condition holds true:

$$|\partial_i\phi(\eta^1)-\partial_i\phi(\eta^2)|\sim 1\quad\text{for all }\eta^j\in U_j,\ j=1,2. \tag{2-17}$$

The set of all pairs (S_1, S_2) of hypersurfaces satisfying these properties will be denoted by $\mathcal{S}_0$ (note that it does depend on the constants hidden by the symbols $\lesssim$ and $\sim$).

The main goal of this chapter will be to establish a local, bilinear Fourier extension estimate on suitable cuboids adapted to the wave packets.

Theorem 2.6. *Assume $\frac{5}{3} \le p \le 2$. Let us choose $r \in \mathbb{R}^2$ such that $r = r^{(j)}$ if $\kappa^{(j)} = \kappa^{(1)} \wedge \kappa^{(2)}$. Then for every $\alpha > 0$ there exist constants $C_\alpha, \gamma_\alpha > 0$ such that for every pair $S = (S_1, S_2) \in \mathcal{S}_0$, every parameter $R \ge 1$ and all functions $f_j \in L^2(S_j)$, $j = 1, 2$, we have*

$$\| R^*_{\mathbb{R}^2} f_1 \, R^*_{\mathbb{R}^2} f_2 \|_{L^p(Q^0_{S_1,S_2}(R))} \le C_\alpha R^\alpha (\kappa^{(1)}\kappa^{(2)})^{\frac{1}{2}-\frac{1}{p}} D^{3-\frac{5}{p}} \log^{\gamma_\alpha}(C_0(S)) \, \|f_1\|_2 \|f_2\|_2, \tag{2-18}$$

where

$$Q^0_{S_1,S_2}(R) = \left\{ x \in \mathbb{R}^3 : |x_i + \partial_i \phi(r) x_3| \le \frac{R^2}{D^2 \bar{\kappa}}, i = 1, 2, \ |x_3| \le \frac{R^2}{D^2(\kappa^{(1)} \wedge \kappa^{(2)})} \right\}, \tag{2-19}$$

with

$$C_0(S) = \frac{\bar{d}_1^2 \bar{d}_2^2}{D^4} (D[\kappa^{(1)} \wedge \kappa^{(2)}])^{-\frac{1}{p}} (D\kappa^{(1)} D\kappa^{(2)})^{-\frac{1}{2}}. \tag{2-20}$$

Notice that $C_0(S) \gtrsim 1$.

Remark 2.7. If $\kappa^{(1)} = \kappa^{(2)} = \bar{\kappa}$, then r is not well defined. But in this case the two sets

$$Q^0_{S_1,S_2}(R; j) = \left\{ x \in \mathbb{R}^3 : |x_i + \partial_i \phi(r^{(j)}) x_3| \le \frac{R^2}{D^2 \bar{\kappa}}, i = 1, 2, \ |x_3| \le \frac{R^2}{D^2 \bar{\kappa}} \right\}, \quad j = 1, 2,$$

essentially coincide. Indeed, since $|\nabla\phi(r^{(1)}) - \nabla\phi(r^{(2)})| \sim 1$ (due to the transversality assumption (iii)), an easy geometric consideration shows that

$$a Q^0_{S_1,S_2}(R; 1) \subset Q^0_{S_1,S_2}(R; 2) \subset b Q^0_{S_1,S_2}(R; 1)$$

for some constants a, b which do not depend on R and the class $\mathcal{S}_0$ from which $S = (S_1, S_2)$ is taken.

By applying a suitable affine transformation whose linear part fixes the points of $\mathbb{R}^2 \times \{0\}$, if necessary, we may assume without loss of generality that $r = 0$ and $\nabla\phi(r) = 0$. Notice that conditions (i)–(iii) and the conclusion of the theorem are invariant under such affine transformations.

In fact, we shall then prove estimate (2-18) in the theorem on the even larger cuboid

$$Q_{S_1,S_2}(R) = \left\{ x \in \mathbb{R}^3 : |x_{i_0}| \le \frac{R^2}{D^2 \bar{\kappa}}, \ \|x\|_\infty \le \frac{R^2}{D^2(\kappa^{(1)} \wedge \kappa^{(2)})} \right\} \tag{2-21}$$

for an appropriate choice of the coordinate direction x_{i_0}, $i_0 \in \{1, 2\}$, in which the cuboid has smaller side length. Later we shall need to combine different cuboids which may possibly have their smaller side lengths in different directions. Then it will become necessary to restrict to their intersection, which leads to (2-19).

Indeed, we shall see that there will be two directions in which the side length of the cuboids are dictated by the length of the wave packets, and one remaining third direction for which we shall have more freedom in choosing the side length.

Observe also that $\bar{\kappa}_i \bar{d}_i \lesssim 1$, and thus we may even assume without loss of generality that

$$\bar{\kappa}_i \bar{d}_i \ll 1 \quad \text{for all } i = 1, 2 \tag{2-22}$$

simply by decomposing S_1 and S_2 into a finite number of subsets for which the side lengths of corresponding rectangles U_j are sufficiently small fractions of the given $d_i^{(j)}$.

For $\eta^j \in U_j$ define

$$\phi_1(\eta) = \phi(\eta - \eta^2) + \phi(\eta^2), \quad \eta \in \eta^2 + U_1,$$
$$\phi_2(\eta) = \phi(\eta - \eta^1) + \phi(\eta^1), \quad \eta \in \eta^1 + U_2.$$

The set $((\eta^2, \phi(\eta^2)) + S_1) \cap ((\eta^1, \phi(\eta^1)) + S_2) = \mathrm{graph}(\phi_1) \cap \mathrm{graph}(\phi_2)$ will be called an *intersection curve* of S_1 and S_2. It agrees with the graph of ϕ_1 (or ϕ_2) restricted to the set where $\psi = \phi_1 - \phi_2 = 0$. On this set, the normal field $N_j(\eta) = (-\nabla\phi_j(\eta), 1)$ forms the conical set

$$\Gamma_j = \{sN_j(\eta) : s \in \mathbb{R}, \psi(\eta) = 0\}.$$

In the sequel, we shall use the abbreviation $j+1 \bmod 2 = 2$, if $j = 1$, and $j+1 \bmod 2 = 1$, if $j = 2$.

Lemma 2.8. *Let* $(S_1, S_2) \in \mathcal{S}_0$. *Assume* $\nabla\phi(r) = 0$ *for some* $r \in S_1 \cup S_2$ *and* $\bar{\kappa}_i \bar{d}_i \ll 1$. *Then the following hold true*:

(a) $D\kappa_i^{(j)} \ll 1$ *for all* $i, j = 1, 2$.

(b) $|\nabla\phi(\eta)| \lesssim 1$ *for all* $(\eta, \phi(\eta)) \in S_1 \cup S_2$.

(c) *The unit normal fields on* S_1 *and* S_2 *are transversal*; *i.e.*,

$$|N_0(\eta^1) - N_0(\eta^2)| \sim 1 \quad \textit{for all } (\eta^j, \phi(\eta^j)) \in S_j. \tag{2-23}$$

(d) N_j *and* $\Gamma_{j+1 \bmod 2}$ *are transversal for* $j = 1, 2$ *and for any choice of intersection curve of* S_1 *and* S_2.

(e) *If* γ *is a parametrization by the arclength* t *of the projection of an intersection curve of* S_1 *and* S_2 *to the first two coordinates* $\eta \in \mathbb{R}^2$, *then* $|\dot\gamma_1| \sim 1 \sim |\dot\gamma_2|$.

Proof. We shall denote by $\eta = \tilde{x} \in \mathbb{R}^2$ the projection of a point in $x \in \mathbb{R}^3$ to its first two coordinates. Part (a) is clear since $D = \min_{i,j=1,2} d_i^{(j)}$. To prove (b), notice that for any $x, x' \in S_1 \cup S_2$ we have $|\nabla\phi(\tilde{x}) - \nabla\phi(\tilde{x}')| \lesssim 1$: if x and x' belong to different hypersurface S_j, we apply condition (iii) on page 838, and if x and x' are in the same hypersurface S_j, we use condition (a). Thus we have $|\nabla\phi(\tilde{x})| = |\nabla\phi(\tilde{x}) - \nabla\phi(r)| \lesssim 1$ for all $x \in S_1 \cup S_2$.

This gives $|N(\tilde{x})| = \sqrt{1 + |\nabla\phi(\tilde{x})|^2} \sim 1$ for all $x \in S_1 \cup S_2$, which already implies the transversality of the normal fields:

$$|N_0(\eta^1) - N_0(\eta^2)| \sim |N(\eta^1) - N(\eta^2)| = |\nabla\phi(\eta^1) - \nabla\phi(\eta^2)| \sim 1$$

for all $(\eta^j, \phi(\eta^j)) \in S_j$, $j = 1, 2$.

We shall prove (e) first, since (e) will be needed for the proof of (d). It suffices to prove that $|\partial_i \psi(\eta)| \sim 1$ for all η such that $\eta - \eta^j \in U_{j+1 \bmod 2}$, $\eta^j \in U_j$, since the tangent to the curve γ at any point $\gamma(t)$ is orthogonal to $\nabla\psi(\gamma(t))$. But, in view of (2-17),

$$|\partial_i\psi(\eta)| = |\partial_i\phi(\eta-\eta^2) - \partial_i\phi(\eta-\eta^1)| \sim 1.$$

For (d), since the claim is symmetric in $j \in \{1,2\}$, it suffices to show that N_1 and Γ_2 are transversal. Since we have

$$|N_1(\eta) - N_1(\eta')| = |\nabla\phi_1(\eta) - \nabla\phi_1(\eta')| \lesssim \kappa_1^{(1)} d_1^{(1)} + \kappa_2^{(1)} d_2^{(1)} \ll 1$$

for all $\eta, \eta' \in U_1 + \eta^2$, whereas $|N_1(\eta)| \sim 1$ for all $\eta \in U_1 + \eta^2$, it is even enough to show that $N_1(\eta)$ and the tangent space $T_{N_2(\eta)}\Gamma_2$ of Γ_2 at the point $N_2(\eta)$ are transversal. Since γ is a parametrization by arclength of the zero set of ψ, the tangent space of Γ_2 at the point $N_2(\eta)$ for $\eta = \gamma(t)$ is spanned by $N_2(\eta)$ and $(-D^2\phi_2(\eta)\,\dot\gamma(t), 0)$, where $D^2\phi_2$ denotes the Hessian matrix of ϕ_2. But, recalling that we assume $\partial_1\partial_2\phi \equiv 0$, we see that the vectors $N_2(\eta)$ and $(1/\kappa^{(2)})(-D^2\phi_2(\eta)\,\dot\gamma(t), 0)$ form an "almost" orthonormal frame for the tangent space $T_{N_2(\eta)}\Gamma_2$, and thus the transversality can be checked by estimating the volume V of the parallelepiped spanned by $N_1(\eta)$ and these two vectors, which is given by

$$V = \begin{vmatrix} -\partial_1\phi_1(\eta) & -\partial_2\phi_1(\eta) & 1 \\ -\partial_1\phi_2(\eta) & -\partial_2\phi_2(\eta) & 1 \\ \frac{1}{\kappa^{(2)}}\partial_1^2\phi_2(\eta)\dot\gamma_1(t) & \frac{1}{\kappa^{(2)}}\partial_1^2\phi_2(\eta)\dot\gamma_2(t) & 0 \end{vmatrix} = \frac{1}{\kappa^{(2)}}\big|-\partial_1^2\phi_2(\eta)\dot\gamma_1(t)\partial_2\psi(\eta) + \partial_2^2\phi_2(\eta)\dot\gamma_2(t)\partial_1\psi(\eta)\big|.$$

Since $\psi \circ \gamma = 0$ by definition, we have $\partial_1\psi(\eta)\dot\gamma_1(t) + \partial_2\psi(\eta)\dot\gamma_2(t)$; hence

$$\partial_2\psi(\eta) = -\partial_1\psi(\eta)\frac{\dot\gamma_1(t)}{\dot\gamma_2(t)}.$$

Thus

$$V = \frac{|\partial_1\psi(\eta)|}{\kappa^{(2)}|\dot\gamma_2(t)|}\big(\partial_1^2\phi_2(\eta)\dot\gamma_1^2(t) + \partial_2^2\phi_2(\eta)\dot\gamma_2^2(t)\big) \sim \big|\partial_1\phi(\eta-\eta^2) - \partial_1\phi(\eta-\eta^1)\big|\frac{\kappa_1^{(2)} + \kappa_2^{(2)}}{\kappa^{(2)}} \sim 1. \qquad \square$$

We now come to the introduction of the wave packets that we shall use in the proof of Theorem 2.6. Let us assume without loss of generality that

$$\kappa^{(1)} \le \kappa^{(2)}, \tag{2-24}$$

i.e., $r = r^{(1)}$ and $\nabla\phi(r^{(1)}) = 0$.

Next, since S_1 is horizontal at $(r^{(1)}, \phi(r^{(1)}))$, we may use the wave packet decomposition from Corollary 2.3, with normal n_1 and hyperplane H_1 given by

$$n_1 = (0,0,1) \quad \text{and} \quad H_1 = \mathbb{R}^2 \times \{0\}$$

in order to obtain the decomposition

$$R^*_{\mathbb{R}^2} f_1 = R^*_{H_1} f_1 = \sum_{w_1 \in \mathcal{W}_1} c_{w_1} p_{w_1}, \quad w_1 \in \mathcal{W}_1, \tag{2-25}$$

into wave packets $p_{w_1}, w_1 \in \mathcal{W}_1$ of length $(R')^2/\kappa^{(1)}$, directly by means of Lemma 2.1. By $T_{w_1}, w_1 \in \mathcal{W}_1$, we denote the associated set of tubes. Recall that this decomposition is valid on the set $P_1 = \mathbb{R}^2 \times [-(R')^2/\kappa^{(1)}, (R')^2/\kappa^{(1)}]$.

Let us next turn to S_2 and $R^*_{\mathbb{R}^2} f_2$. If we would keep the same coordinate system for S_2, we would have to truncate even further in x_3-direction, since $(R')^2/\kappa^{(2)} \leq (R')^2/\kappa^{(1)}$. However, by (2-17) we have for $\eta \in U_2$ and both $i = 1$ and $i = 2$ that

$$|\langle e_i, N(\eta)\rangle| = |\partial_i \phi(\eta)| = |\partial_i \phi(\eta) - \partial \phi(r^{(1)})| \sim 1.$$

This means that we may apply Lemma 2.4 to S_2 in order to reparametrize S_2 by an open subset (denoted again by U_2) of the hyperplane $H_2 = n_2^\perp$ given by

$$n_2 = e_{i_0} \quad \text{and} \quad H_2 = \{n_2\}^\perp = \{e_{i_0}\}^\perp.$$

We may thus replace the function f_2 by a function (also denoted by f_2) on U_2 of comparable L^2-norm, and replace $R^*_{\mathbb{R}^2} f_2$ by $R^*_{H_2} f_2$ in the subsequent arguments.

Next, applying Corollary 2.3, now with $H = H_2$, for $i_0 = 1$, as well as for $i_0 = 2$, we may decompose $R^*_{H_2} f_2$ as

$$R^*_{H_2} f_2 = \sum_{w_2 \in \mathcal{W}_2} c_{w_2} p_{w_2}, \quad w_2 \in \mathcal{W}_2, \tag{2-26}$$

on the set

$$P_2 = \left\{ x \in \mathbb{R}^3 : |\langle x, n_2\rangle| \leq \frac{(R')^2}{\kappa^{(2)}} \right\} = \left\{ x \in \mathbb{R}^3 : |x_{i_0}| \leq \frac{(R')^2}{\kappa^{(2)}} \right\}$$

by means of wave packets of length $(R')^2/\kappa^{(2)}$. The associated set of tubes is denoted by $T_{w_2}, w_2 \in \mathcal{W}_2$.

In order to decide how to choose i_0, we observe that for $\eta \in U_1$, our definitions (2-15) in combination with the estimates (2-16) and (2-22) show that

$$|\partial_i \phi(\eta) - \partial_i \phi(r^{(1)})| \lesssim \kappa_i^{(1)} d_i^{(1)} \leq \frac{\kappa_i^{(1)}}{\bar{\kappa}_i} \bar{\kappa}_i \bar{d}_i \ll \frac{\kappa_i^{(1)}}{\bar{\kappa}_i}.$$

Notice that the wave packets associated to S_1 are roughly pointing in the direction of $N(r^{(1)}) = (0, 0, 1)$. More precisely, if we project a wave packet pointing in the direction of $N(\eta)$, $\eta \in U_1$, to the coordinate x_i, $i = 1, 2$, then by the previous estimates we see that we obtain an interval of length comparable to

$$\left|\left\langle e_i, \frac{(R')^2}{\kappa^{(1)}} N(\eta)\right\rangle\right| = \frac{(R')^2}{\kappa^{(1)}} |\partial_i \phi(\eta)| = \frac{(R')^2}{\kappa^{(1)}} |\partial_i \phi(\eta) - \partial_i \phi(r^{(1)})| \ll \frac{(R')^2}{\bar{\kappa}} \frac{\bar{\kappa}}{\kappa^{(1)}} \frac{\kappa_i^{(1)}}{\bar{\kappa}_i}. \tag{2-27}$$

Let us therefore choose i_0 so that

$$\frac{\kappa_{i_0}^{(1)}}{\bar{\kappa}_{i_0}} = \frac{\kappa_1^{(1)}}{\bar{\kappa}_1} \wedge \frac{\kappa_2^{(1)}}{\bar{\kappa}_2}.$$

Then

$$\bar{\kappa} \frac{\kappa_{i_0}^{(1)}}{\bar{\kappa}_{i_0}} = (\bar{\kappa}_1 \vee \bar{\kappa}_2) \left(\frac{\kappa_1^{(1)}}{\bar{\kappa}_1} \wedge \frac{\kappa_2^{(1)}}{\bar{\kappa}_2} \right) \leq \kappa_1^{(1)} \vee \kappa_2^{(1)} = \kappa^{(1)},$$

and thus by (2-27) and (2-24)

$$\left|\left\langle e_{i_0}, \frac{(R')^2}{\kappa^{(1)}} N(\eta)\right\rangle\right| \ll \frac{(R')^2}{\bar{\kappa}} = \frac{(R')^2}{\kappa^{(2)}}.$$

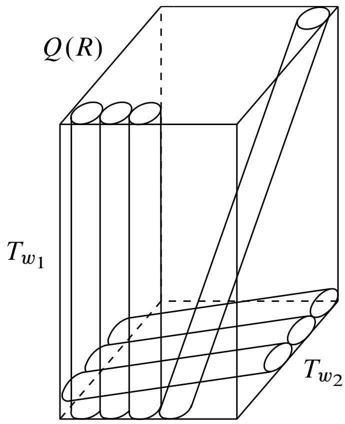


Figure 8. The wave packets filling the cuboid $Q_{S_1,S_2}(R)$.

This means that the geometry fits well: the wave packets associated to S_1 do not turn too much into the direction of x_{i_0}; projected to this coordinate, their length is smaller than the length of the wave packets associated to S_2, which are essentially pointing in the direction of the i_0-th coordinate axis (see Figure 8).

However, for the remaining coordinate direction x_i, $i \in \{1,2\} \setminus \{i_0\}$, we cannot guarantee such a behavior. But notice that by (2-24),

$$\begin{aligned} P_1 \cap P_2 &= \left(\mathbb{R}^2 \times \left[-\frac{(R')^2}{\kappa^{(1)}}, \frac{(R')^2}{\kappa^{(1)}}\right]\right) \cap \left\{x \in \mathbb{R}^3 : |x_{i_0}| \le \frac{(R')^2}{\kappa^{(2)}}\right\} \\ &= \left\{(x_1,x_2) \in \mathbb{R}^2 : |x_{i_0}| \le \frac{(R')^2}{\kappa^{(2)}}\right\} \times \left[-\frac{(R')^2}{\kappa^{(1)}}, \frac{(R')^2}{\kappa^{(1)}}\right] \\ &\supset \left\{x \in \mathbb{R}^3 : |x_{i_0}| \le \frac{(R')^2}{\bar{\kappa}},\ \|x\|_\infty \le \frac{(R')^2}{\kappa^{(1)} \wedge \kappa^{(2)}}\right\} = Q_{S_1,S_2}(R); \end{aligned}$$

i.e., on the cuboid $Q_{S_1,S_2}(R)$ we may apply our development into wave packets to the wave packets associated to the hypersurface S_1, as well as those associated to S_2.

For every $\alpha > 0$, let us denote by $E(\alpha)$ the following statement:

> There exist constants $C_\alpha > 0$ and $\gamma_\alpha > 0$ such that for all pairs $S = (S_1, S_2) \in \mathcal{S}_0$, all $R \ge 1$ and all $f_j \in L^2(U_j)$, $j = 1,2$, (which we may also regard as functions on S_j) the following estimate holds true:
>
> $$\begin{aligned} &\|R^*_{H_1} f_1 R^*_{H_2} f_2\|_{L^p(Q_{S_1,S_2}(R))} \\ &\qquad \le C_\alpha R^\alpha \log^{\gamma_\alpha}(1+R) (\kappa^{(1)}\kappa^{(2)})^{\frac{1}{2}-\frac{1}{p}} D^{3-\frac{5}{p}} \log^{\gamma_\alpha}(C_0(S)) \|f_1\|_2 \|f_2\|_2. \end{aligned} \tag{2-28}$$
>
> Here, $C_0(S)$ denotes the constant defined in Theorem 2.6.

Our goal will be to show that $E(\alpha)$ holds true for every $\alpha > 0$, which would prove Theorem 2.6. To this end, we shall apply the method of induction on scales.

Observe that the intersection of two of the transversal tubes $T_{w_1}, w_1 \in \mathcal{W}_1$, and $T_{w_2}, w_2 \in \mathcal{W}_2$, will always be contained in a cube of side length $\mathcal{O}(R')$. Let us therefore decompose $\mathbb{R}^3$ by means of a grid of side length R' into cubes q of the same side length, and let $\{q\}_{q\in\mathcal{Q}}$ be a family of such cubes covering $Q_{S_1,S_2}(R)$. By c_q we shall denote the center of the cube q. Choose $\chi \in \mathcal{S}(\mathbb{R}^3)$ with supp $\hat{\chi} \subset B(0,1)$ and $\hat{\chi}(0) = 1/(2\pi)^n$, and put $\chi_q(x) = \chi(x - c_q/(R'))$. Poisson's summation formula then implies $\sum \chi_q = 1$ on $\mathbb{R}^3$, so that in particular we may assume $\sum_{q\in\mathcal{Q}} \chi_q = 1$ on $Q_{S_1,S_2}(R)$.

Notice that our approach slightly differs from the standard usage of induction on scales, where χ_q is chosen to be the characteristic function of q, and not a smoothened version of it. The price we shall have to pay is that some arguments will become a bit more technical, but the compact Fourier support of the functions χ_q will become crucial later.

For a given index set $W_j \subset \mathcal{W}_j$, $j = 1, 2$, of wave packets (see (2-25), (2-26)), we denote by

$$W_j(q) = \{w_j \in W_j : T_{w_j} \cap R^\delta q \neq \varnothing\}$$

the collection of all the tubes of type j passing through (a slightly thickened) cube q. Here, $\delta > 0$ is a small parameter which will be fixed later, and $R^\delta q$ denotes the dilate of q by the factor R^δ having the same center c_q as q.

Let us denote by $\mathcal{N}$ the set $\mathcal{N} = \{2^n : n \in \mathbb{N}\} \cup \{0\}$. In order to count the magnitude of the number of wave packets W_j passing through a given cube q, we introduce the sets

$$Q^\mu = \{q : |W_j(q)| \sim \mu_j,\, j = 1, 2\}, \quad \mu = (\mu_1, \mu_2) \in \mathcal{N}^2.$$

Obviously the Q^μ form a partition of the family of all cubes $q \in \mathcal{Q}$. For $w_j \in W_j$, we further introduce the set of all cubes in Q^μ close to T_{w_j}:

$$Q^\mu(w_j) = \{q \in Q^\mu : T_{w_j} \cap R^\delta q \neq \varnothing\}.$$

Finally, we determine the number of such cubes by means of the sets

$$W_j^{\lambda_j,\mu} = \{w_j \in W_j : |Q^\mu(w_j)| \sim \lambda_j\}, \quad \lambda_j, \mu_1, \mu_2 \in \mathcal{N}.$$

For every fixed μ, the family $\{W_j^{\lambda_j,\mu}\}_{\lambda_j\in\mathcal{N}}$ forms a partition of W_j.

We are now in a position to reduce the statement $E(\alpha)$ to a formulation in terms of wave packets.

2C. ***Reduction to a wave packet formulation.*** Following basically a standard pigeonholing argument in combination with (P5), the estimate in $E(\alpha)$ can easily be reduced to a bilinear estimate for sums of wave packets (modulo an increase of the exponent γ_α by 5). It is in this reduction that some power of the logarithmic factor $\log(C_0(S))$ will appear, and we shall have to be a bit more precise than usual in order to identify $C_0(S)$ as the expression given by (2-20).

Lemma 2.9. *Let $\alpha > 0$. Assume there are constants $C_\alpha, \gamma_\alpha > 0$ such that for all $(S_1, S_2) \in \mathcal{S}_0$ (parametrized by the open subsets $U_j \subset H_j$) the following estimate is satisfied*:

Given any two families of wave packets $\{p_{w_1}\}_{w_1\in\mathcal{W}_1}$ and $\{p_{w_2}\}_{w_2\in\mathcal{W}_2}$ associated to S_1 and S_2 respectively, as in the wave packet decomposition Corollary 2.3, where the p_{w_j}, $j = 1, 2$, satisfy *uniformly the estimates in (P2)–(P5), for all $R \geq 1$, all $\lambda_j, \mu_j \in \mathcal{N}$ and all subsets $W_j \subset \mathcal{W}_j$, $j = 1, 2$, we have (with admissible constants)*

$$\Big\| \prod_{j=1,2} \sum_{w_j \in W_j^{\lambda_j,\mu}} p_{w_j} \sum_{q \in Q^\mu} \chi_q \Big\|_{L^p(Q_{S_1,S_2}(R))} \leq C_\alpha R^\alpha \log^{\gamma_\alpha}(1+R)(\kappa^{(1)}\kappa^{(2)})^{\frac{1}{2}-\frac{1}{p}} D^{3-\frac{5}{p}} \log^{\gamma_\alpha}(C_0(S))|W_1|^{\frac{1}{2}}|W_2|^{\frac{1}{2}}. \tag{2-29}$$

Then $E(\alpha)$ holds true.

Proof. In order to show $E(\alpha)$, we may assume without loss of generality that $\|f_j\|_2 = 1$, $j = 1, 2$. Let us use the abbreviation $C_0(S) = C_0$.

First observe that for fixed q and v_j, the number of y_j such that the tube $T_{(y_j,v_j)}$ passes through $R^\delta q$ is bounded by $R^{c\delta}$, whereas the total number of $v_j \in V_j$ is bounded by

$$|V_j| \sim (R')^2|U_j| \leq R^2 \frac{\bar{d}_1\bar{d}_2}{D^2}. \tag{2-30}$$

Thus we have

$$|W_j(q)| \leq R^{2+c\delta}\frac{\bar{d}_1\bar{d}_2}{D^2} \leq R^{2+c}\frac{\bar{d}_1^2\bar{d}_2^2}{D^4}(D[\kappa^{(1)} \wedge \kappa^{(2)}])^{-\frac{1}{p}}(D\kappa^{(1)}D\kappa^{(2)})^{-\frac{1}{2}} = R^{c'}C_0,$$

where we have used property (a) of Lemma 2.8. Consequently $Q^\mu = \varnothing$, if $\mu_j \gg R^{c'}C_0$ for some j. Similarly, the number of cubes q of side length R' such that $R^\delta q$ intersects with a tube T_{w_j} of length $(R')^2/\kappa^{(j)}$ is bounded by $R^{c\delta}R'/\kappa^{(j)} = R^{1+c\delta}/D\kappa^{(j)}$. Since $D \leq \bar{d}_1, \bar{d}_2$, this implies

$$|Q^\mu(w_j)| \leq \frac{R^{1+c\delta}}{D\kappa^{(j)}} \leq R^{c'}\frac{\bar{d}_1^4\bar{d}_2^4}{D^8}(D[\kappa^{(1)} \wedge \kappa^{(2)}])^{-\frac{2}{p}}(D\kappa^{(1)}D\kappa^{(2)})^{-1} = R^{c'}C_0^2,$$

and thus $W_j^{\lambda_j,\mu} = \varnothing$, if $\lambda_j \gg R^{c'}C_0^2$. For $C \geq 0$ let us put $\mathcal{N}(C) = \{\nu \in \mathcal{N} : \nu \lesssim C\}$. Since $C_0 \gtrsim 1$, we then see that

$$\mathcal{Q} = \bigcup_{\mu_1,\mu_2 \in \mathcal{N}(R^{c'}C_0^2)} Q^\mu,$$

and for every fixed μ,

$$W_j = \bigcup_{\lambda_j \in \mathcal{N}(R^{c'}C_0^2)} W_j^{\lambda_j,\mu}.$$

These decompositions in combination with our assumed estimate (2-29) imply

$$\begin{aligned}\Big\| \prod_{j=1,2} \sum_{w_j \in W_j} p_{w_j} \Big\|_{L^p(Q_{S_1,S_2}(R))} &\leq \sum_{\lambda_1,\lambda_2,\mu_1,\mu_2 \in \mathcal{N}(R^{c'}C_0^2)} \Big\| \prod_{j=1,2} \sum_{w_j \in W_j^{\lambda_j,\mu}} p_{w_j} \sum_{q \in Q^\mu} \chi_q \Big\|_{L^p(Q_{S_1,S_2}(R))} \\ &\leq C_\alpha R^\alpha \log^4(R^{c'}C_0^2) \log^{\gamma_\alpha}(1+R)(\kappa^{(1)}\kappa^{(2)})^{\frac{1}{2}-\frac{1}{p}} D^{3-\frac{5}{p}} \log^{\gamma_\alpha}(C_0)|W_1|^{\frac{1}{2}}|W_2|^{\frac{1}{2}}\end{aligned}$$

for every $W_j \subset \mathcal{W}_j$, $j = 1, 2$; hence

$$\Big\| \prod_{j=1,2} \sum_{w_j \in W_j} p_{w_j} \Big\|_{L^p(Q_{S_1,S_2}(R))} \leq C_\alpha R^\alpha \log^{\gamma_\alpha+4}(1+R)(\kappa^{(1)}\kappa^{(2)})^{\frac{1}{2}-\frac{1}{p}} D^{3-\frac{5}{p}} \log^{\gamma_\alpha+4}(C_0)|W_1|^{\frac{1}{2}}|W_2|^{\frac{1}{2}}. \tag{2-31}$$

Recall next that $R^* f_j = \sum_{w_j \in \mathcal{W}_j} c_{w_j} p_{w_j}$. We introduce the subsets $W_j^k = \{w_j \in \mathcal{W}_j : |c_{w_j}| \sim 2^{-k}\}$, which allow us to partition $\mathcal{W}_j$ into $\bigcup_{k \in \mathbb{N}} W_j^k$. We fix some k_0, whose precise value will be determined later. Then

$$\Big\| \sum_{k>k_0} \sum_{w_1 \in W_1^k} \sum_{w_2 \in \mathcal{W}_2} c_{w_1} p_{w_1} c_{w_2} p_{w_2} \Big\|_{L^p(Q_{S_1,S_2}(R))} \leq |Q_{S_1,S_2}(R)|^{\frac{1}{p}} \sum_{k>k_0} \Big\| \sum_{w_1 \in W_1^k} \sum_{w_2 \in \mathcal{W}_2} c_{w_1} p_{w_1} c_{w_2} p_{w_2} \Big\|_\infty.$$

The wave packets p_{w_j} are well separated with respect to the parameter y_j, and by (P4), their L^∞- norm is of order $\mathcal{O}((R')^{-1})$. Moreover, by (2-30) the number of v_j's is bounded by $R^2 \bar{d}_1 \bar{d}_2 / D^2$. Furthermore, $|c_{w_1}| \lesssim 2^{-k}$ for every $w_1 \in W_1^k$, and by (P5) we have $|c_{w_2}| \leq \|\{c_{w_2}\}_{w_2 \in \mathcal{W}_2}\|_{\ell^2} \lesssim \|f_2\|_2 = 1$. Combining all this information, we may estimate

$$\Big\| \sum_{k>k_0} \sum_{w_1 \in W_1^k} \sum_{w_2 \in \mathcal{W}_2} c_{w_1} p_{w_1} c_{w_2} p_{w_2} \Big\|_{L^p(Q_{S_1,S_2}(R))} \lesssim \Big(\frac{(R')^6}{[\kappa^{(1)} \wedge \kappa^{(2)}]\kappa^{(1)}\kappa^{(2)}} \Big)^{\frac{1}{p}} \frac{\bar{d}_1^2 \bar{d}_2^2}{D^4} R^4 (R')^{-2} \sum_{k>k_0} 2^{-k}$$
$$\sim R^{\frac{6}{p}+2} C_0 D^{3-\frac{5}{p}} (\kappa^{(1)}\kappa^{(2)})^{\frac{1}{2}-\frac{1}{p}} 2^{-k_0}.$$

If we now choose $k_0 = \log_2 C_0 + \log R^{\frac{6}{p}+2}$, then we obtain

$$\Big\| \sum_{k>k_0} \sum_{w_1 \in W_1^k} \sum_{w_2 \in \mathcal{W}_2} c_{w_1} p_{w_1} c_{w_2} p_{w_2} \Big\|_{L^p(Q_{S_1,S_2}(R))} \lesssim D^{3-\frac{5}{p}} (\kappa^{(1)}\kappa^{(2)})^{\frac{1}{2}-\frac{1}{p}}. \tag{2-32}$$

In a similar way we also get

$$\Big\| \sum_{k_1 \leq k_0} \sum_{w_1 \in W_1^{k_1}} \sum_{k_2 > k_0} \sum_{w_2 \in W_2^{k_2}} c_{w_1} p_{w_1} c_{w_2} p_{w_2} \Big\|_{L^p(Q_{S_1,S_2}(R))} \lesssim D^{3-\frac{5}{p}} (\kappa^{(1)}\kappa^{(2)})^{\frac{1}{2}-\frac{1}{p}}. \tag{2-33}$$

The remaining terms can simply be estimated by

$$\Big\| \sum_{k_1,k_2=1}^{k_0} \sum_{w_1 \in W_1^{k_1}} \sum_{w_2 \in \mathcal{W}_2^{k_2}} c_{w_1} p_{w_1} c_{w_2} p_{w_2} \Big\|_{L^p(Q_{S_1,S_2}(R))} \lesssim \sum_{k_1,k_2=1}^{k_0} 2^{-k_1-k_2} \Big\| \sum_{w_1 \in W_1^{k_1}} \sum_{w_2 \in \mathcal{W}_2^{k_2}} c_{w_1} 2^{k_1} p_{w_1} c_{w_2} 2^{k_2} p_{w_2} \Big\|_{L^p(Q_{S_1,S_2}(R))}.$$

Since $|c_{w_j}2^{k_j}|\sim 1$ for $w_j\in W_j^{k_j}$, it is appropriate to apply (2-31) to the modified wave packets $\tilde p_{w_j}=c_{w_j}2^{k_j}p_{w_j}$:

$$\Bigl\|\sum_{k_1,k_2=1}^{k_0}\sum_{w_1\in W_1^{k_1}}\sum_{w_2\in \mathcal{W}_2^{k_2}} c_{w_1}p_{w_1}c_{w_2}p_{w_2}\Bigr\|_{L^p(Q_{S_1,S_2}(R))} \le C_\alpha R^\alpha \log^{\gamma_\alpha+4}(1+R)(\kappa^{(1)}\kappa^{(2)})^{\frac12-\frac1p}D^{3-\frac5p}\log^{\gamma_\alpha+4}(C_0)\sum_{k_1,k_2=1}^{k_0}2^{-k_1-k_2}|W_1^{k_1}|^{\frac12}|W_2^{k_2}|^{\frac12}.$$

But observe that by (P5),

$$\sum_{k_1,k_2=1}^{k_0}2^{-k_1-k_2}|W_1^{k_1}|^{\frac12}|W_2^{k_2}|^{\frac12}\le k_0\Bigl(\sum_{k_1=1}^{k_0}|W_1^{k_1}|2^{-2k_1}\sum_{k_2=1}^{k_0}|W_2^{k_2}|2^{-2k_2}\Bigr)^{\frac12} \lesssim k_0\Bigl(\sum_{k_1=1}^{k_0}\sum_{w_1\in W_1^{k_1}}|c_{w_1}|^2\sum_{k_2=1}^{k_0}\sum_{w_2\in W_2^{k_2}}|c_{w_2}|^2\Bigr)^{\frac12}\lesssim k_0\|f_1\|_2\|f_2\|_2=k_0,$$

and thus

$$\Bigl\|\sum_{k_1,k_2=1}^{k_0}\sum_{w_1\in W_1^{k_1}}\sum_{w_2\in W_1^{k_2}} c_{w_1}p_{w_1}c_{w_2}p_{w_2}\Bigr\|_{L^p(Q_{S_1,S_2}(R))} \lesssim C_\alpha R^\alpha \log^{\gamma_\alpha+5}(1+R)(\kappa^{(1)}\kappa^{(2)})^{\frac12-\frac1p}D^{3-\frac5p}\log^{\gamma_\alpha+5}(C_0). \tag{2-34}$$

Combining (2-32)–(2-34), we find that

$$\|R^*_{H_1}f_1R^*_{H_2}f_2\|_{L^p(Q_{S_1,S_2}(R))}=\Bigl\|\prod_{j=1,2}\sum_{w_j\in\mathcal{W}_j}c_{w_j}p_{w_j}\Bigr\|_{L^p(Q_{S_1,S_2}(R))} \lesssim C_\alpha R^\alpha \log^{\gamma_\alpha+5}(1+R)(\kappa^{(1)}\kappa^{(2)})^{\frac12-\frac1p}D^{3-\frac5p}\log^{\gamma_\alpha+5}(C_0),$$

which verifies $E(\alpha)$. □

2D. *Bilinear estimates for sums of wave packets.* Let $v_j\in\mathcal{V}_j$, $j=1,2$, and define the ($\mathcal{O}(1/R')$ thickened) "intersection" of the transversal hypersurfaces S_1 and S_2 by

$$\Pi_{v_1,v_2}=(v_1+S_2)\cap(v_2+S_1)+\mathcal{O}((R')^{-1}).$$

For any subset $W_j\subset\mathcal{W}_j$, let

$$W_j^{\Pi_{v_1,v_2}}=\{w_j'\in W_j: v_j'+v_{j+1}\in\Pi_{v_1,v_2}\}$$

(where $j+1$ is to be interpreted mod 2 as before, i.e., we will use the shorthand notation $j+1=j+1 \bmod 2$ in the sequel whenever $j+1$ appears as an index), and denote by

$$V_j=\{v_j'\in\mathcal{V}_j:(y_j',v_j')\in W_j\ \text{ for some } y_j'\in\mathcal{Y}\}$$

the $\mathcal{V}$-projection of W_j. Further let

$$\begin{aligned} V_j^{\Pi_{v_1,v_2}} &= \{v_j' \in \mathcal{V}_j : (y_j', v_j') \in W_j^{\Pi_{v_1,v_2}} \text{ for some } y_j' \in \mathcal{Y}_j\} \\ &= \{v_j' \in \mathcal{V}_j : \text{ there is some } y_j' \in \mathcal{Y}_j \text{ s.t. } (y_j', v_j') \in W_j \text{ and } v_j' + v_{j+1} \in \Pi_{v_1,v_2}\}. \end{aligned}$$

Lemma 2.10. *Let $W_j \subset \mathcal{W}_j$, $j = 1, 2$. Then*

$$\Big\| \sum_{w_1 \in W_1} \sum_{w_2 \in W_2} p_{w_1} p_{w_2} \Big\|_{L^1(Q_{S_1,S_2}(R))} \le \frac{(R')^2}{\sqrt{\kappa^{(1)}\kappa^{(2)}}} |W_1|^{\frac{1}{2}} |W_2|^{\frac{1}{2}}, \tag{2-35}$$

$$\Big\| \sum_{w_1 \in W_1} \sum_{w_2 \in W_2} p_{w_1} p_{w_2} \Big\|_{L^2(Q_{S_1,S_2}(R))} \lesssim (R')^{-\frac{1}{2}} \min_j \sup_{v_1,v_2} |V_j^{\Pi_{v_1,v_2}}|^{\frac{1}{2}} |W_1|^{\frac{1}{2}} |W_2|^{\frac{1}{2}}. \tag{2-36}$$

Proof. We shall closely follow the arguments in [Lee and Vargas 2010], in particular the proof of Lemma 2.2, with only slight modifications.

The first estimate is easy. Using Hölder's inequality, we see that

$$\begin{aligned} \Big\| \sum_{w_1 \in W_1} \sum_{w_2 \in W_2} p_{w_1} p_{w_2} \Big\|_{L^1(Q_{S_1,S_2}(R))} &\le \prod_{j=1,2} \Big\| \sum_{w_j \in W_j} p_{w_j} \Big\|_{L^2(Q_{S_1,S_2}(R))} \\ &\le \prod_{j=1,2} \bigg(\int_{-(R')^2/\kappa^{(j)}}^{(R')^2/\kappa^{(j)}} \Big\| \sum_{w_j \in W_j} p_{w_j}(\cdot + t n_j) \Big\|_{L^2(H_j)}^2 dt \bigg)^{\frac{1}{2}} \\ &\lesssim \prod_{j=1,2} \frac{R'}{\sqrt{\kappa^{(j)}}} |W_j|^{\frac{1}{2}}, \end{aligned}$$

where we have used (P4) in the last estimate. The second one is more involved. We write

$$\Big\| \sum_{w_1 \in W_1} \sum_{w_2 \in W_2} p_{w_1} p_{w_2} \Big\|_{L^2(Q_{S_1,S_2}(R))}^2 = \sum_{w_1 \in W_1} \sum_{w_2 \in W_2} \sum_{v_1' \in V_1} \sum_{v_2' \in V_2} \Big\langle p_{w_1} \sum_{y_2' \in Y_2(v_2')} p_{w_2'}, \, p_{w_2} \sum_{y_1' \in Y_1(v_1')} p_{w_1'} \Big\rangle,$$

where $Y_j(v_j') = \{y \in \mathcal{Y}_j : (y, v_j') \in W_j\}$ (recall that V_j is $\mathcal{V}$-projection of W_j). Since for $j = 1, 2$ the Fourier transform of $\sum_{y_{j+1}' \in Y_{j+1}(v_{j+1}')} p_{w_{j+1}'} p_{w_j}$ is supported in a ball of radius $\mathcal{O}((R')^{-1})$ centered at $v_{j+1}' + v_j$, we may assume that the intersection of these two balls is nonempty, and thus

$$v_1' + v_2 = v_2' + v_1 + \mathcal{O}((R')^{-1}).$$

Especially

$$v_{j+1}' + v_j \in \Pi_{v_1,v_2} \quad \text{and} \quad v_j' \in V_j^{\Pi_{v_1,v_2}}, \quad j = 1, 2.$$

This implies

$$\begin{aligned} &\Big\| \sum_{w_1 \in W_1} \sum_{w_2 \in W_2} p_{w_1} p_{w_2} \Big\|_{L^2(Q_{S_1,S_2}(R))}^2 \\ &\qquad \le \sum_{w_1 \in W_1} \sum_{w_2 \in W_2} \sum_{v_1' \in V_1^{\Pi_{v_1,v_2}}} \sum_{v_2'} \int_{\mathbb{R}^3} |p_{w_1} p_{w_2}| \, dx \Big\| \sum_{y_2' \in Y(v_2')} p_{w_2'} \Big\|_\infty \Big\| \sum_{y_1' \in Y(v_1')} p_{w_1'} \Big\|_\infty, \end{aligned}$$

where $v_2' = v_1' + v_2 - v_1 + \mathcal{O}((R')^{-1})$ in the rightmost sum. Observe that there are at most $\mathcal{O}(1)$ possible choices for v_2' such that

$$v_2' = v_1' + v_2 - v_1 + \mathcal{O}((R')^{-1}).$$

Since the wave packets p_{w_j} are essentially supported in the tubes T_{w_j}, which are well separated with respect to the parameter y, the sum in y_j' can be replaced by the supremum, up to some multiplicative constant. Since T_{w_1} and T_{w_2} satisfy the transversality condition (2-23), $p_{w_1}p_{w_2}$ decays rapidly away from the intersection $T_{w_1} \cap T_{w_2}$; i.e.,

$$\int_{\mathbb{R}^3} |p_{w_1}p_{w_2}|\,dx \lesssim \int_{\mathbb{R}^3} (R')^{-2}\Big(1+\frac{|x|}{R'}\Big)^{-N} dx = R'\int_{\mathbb{R}^3}(1+|x|)^{-N}\,dx \sim R'.$$

We thus obtain

$$\begin{aligned}\Big\|\sum_{\substack{w_1\in W_1\\ w_2\in W_2}} p_{w_1}p_{w_2}\Big\|^2_{L^2(Q_{S_1,S_2}(R))} &\lesssim R'|W_1||W_2|\sup_{v_1,v_2}|V_1^{\Pi_{v_1,v_2}}|\prod_{j=1,2}\sup_{w_j'\in W_j}\|p_{w_j'}\|_\infty\\ &\lesssim (R')^{-1}|W_1||W_2|\sup_{v_1,v_2}|V_1^{\Pi_{v_1,v_2}}|. \qquad (2\text{-}37)\end{aligned}$$

Repeating the same computation with the roles of v_1' and v_2' interchanged gives (2-36). ☐

2E. *Basis of the induction-on-scales argument.* In order to start our induction on scales, we need to establish a base case estimate which will respect the form of our estimate (2-29). This will require a somewhat more sophisticated approach than what is done usually, based on the following.

Lemma 2.11. *Let $V_j \subset \mathcal{V}_j$. Then $\min_j \sup_{v_1\in\mathcal{V}_1, v_2\in\mathcal{V}_2} |V_j^{\Pi_{v_1,v_2}}| \lesssim R$.*

Proof. Define the graph mapping $\Phi : U_1 \cup U_2 \to S_1 \cup S_2$, $\Phi(x) = (x, \phi(x))$. If $v_j' = \Phi(x_j') \in V_j^{\Pi_{v_1,v_2}}$, then $v_j' + v_{j+1} \in \Pi_{v_1,v_2}$, and for $x_{j+1} = \Phi^{-1}(v_{j+1})$ we have $x_j' + x_{j+1} \in \gamma(I) + \mathcal{O}((R')^{-1})$, where $\gamma : I \to [U_1+x_2]\cap[U_2+x_1] \subset \mathbb{R}^2$ is a parametrization by arclength of the projection to the (x_1,x_2)-space of the intersection curve Π_{v_1,v_2}. Recall from Lemma 2.8(e) that our assumptions imply that then γ will be close to a diagonal, i.e., $|\dot\gamma_i| \sim 1$, $i = 1, 2$.

For all $t, t' \in I$, we have $\gamma(t), \gamma(t') \in [U_1 + x_2] \cap [U_2 + x_1]$; hence

$$\min_j d_i^{(j)} \ge |\gamma_i(t) - \gamma_i(t')| \ge \min_{t''\in I} |\dot\gamma_i(t'')|\,|t-t'| \sim |t-t'|.$$

This implies $|I| = \sup_{t,t'\in I}|t-t'| \lesssim \min_{i,j} d_i^{(j)} = D$; hence $L(\gamma) \lesssim D$, and thus

$$\begin{aligned}|V_j^{\Pi_{v_1,v_2}}| &\sim |\Phi^{-1}(V_j^{\Pi_{v_1,v_2}})|\\ &\le \big|\{x_j' \in \Phi^{-1}(\mathcal{V}_j) : x_j' \in \gamma(I) - x_{j+1} + \mathcal{O}((R')^{-1})\}\big|\\ &\lesssim L(\gamma)/((R')^{-1})\\ &\lesssim DR' = R,\end{aligned}$$

since $\Phi^{-1}(\mathcal{V}_j)$ is an $(R')^{-1}$-grid in U_j. ☐

Corollary 2.12. *$E(1)$ holds true, provided $\frac{4}{3} \le p \le 2$.*

Proof. Due to Lemma 2.9, it is enough to show the corresponding estimate for wave packets (2-29) with $\alpha = 1$. But, estimating $|V_j^{\Pi_{v_1,v_2}}|$ on the right-hand side of (2-36) in Lemma 2.10 by means of Lemma 2.11, we obtain

$$\Big\| \prod_{j=1,2} \sum_{w_j \in W_j^{\lambda_j,\mu}} p_{w_j} \sum_{q \in Q^\mu} \chi_q \Big\|_{L^2(Q_{S_1,S_2}(R))} \lesssim (R')^{-\frac{1}{2}} R^{\frac{1}{2}} |W_1|^{\frac{1}{2}} |W_1|^{\frac{1}{2}}.$$

Interpolating this with the corresponding L^1-estimate that we obtain from (2-35), we arrive at

$$\begin{aligned}\Big\| \prod_{j=1,2} \sum_{w_j \in W_j^{\lambda_j,\mu}} p_{w_j} \sum_{q \in Q^\mu} \chi_q \Big\|_{L^p(Q_{S_1,S_2}(R))} &\lesssim (\kappa^{(1)}\kappa^{(2)})^{\frac{1}{2}-\frac{1}{p}} (R')^{\frac{5}{p}-3} R^{1-\frac{1}{p}} |W_1|^{\frac{1}{2}} |W_1|^{\frac{1}{2}} \\ &\le (\kappa^{(1)}\kappa^{(2)})^{\frac{1}{2}-\frac{1}{p}} D^{3-\frac{5}{p}} R |W_1|^{\frac{1}{2}} |W_1|^{\frac{1}{2}},\end{aligned}$$

provided $\frac{4}{3} \le p \le 2$. □

2F. *Further decompositions.* In a next step, by some slight modification of the usual approach, we introduce a further decomposition of the cuboid $Q_{S_1,S_2}(R)$ defined in (2-21) into smaller cuboids b whose dimensions are those of $Q_{S_1,S_2}(R)$ shrunk by a factor $R^{-2\delta}$; i.e., all of the b's will be translates of $Q_{S_1,S_2}(R^{1-\delta})$. Here, $\delta > 0$ is a sufficiently small parameter to be chosen later. Since

$$\frac{(R')^2}{\bar\kappa} R^{-2\delta} = \frac{R^{1-2\delta} R'}{D\bar\kappa} \ge R^{1-2\delta} R',$$

the smallest side length of b is still much larger than the side length $R^\delta R'$ of the thickened cubes $R^\delta q$ introduced at the end of Section 2B. Observe further that the number of cuboids b into which $Q_{S_1,S_2}(R)$ will be decomposed is of the order $R^{c\delta}$.[2]

If $\mu \in \mathcal{N}^2$ is a fixed pair of dyadic numbers, and if $w_j \in W_j$, then we assign to w_j a cuboid $b(w_j)$ in such a way that $b(w_j)$ contains a maximal number of q's from $Q^\mu(w_j)$ among all the cuboids b. We say that $b \sim w_j$ if b is contained in $10b(w_j)$ (the cuboid having the same center as $b(w_j)$ but scaled by a factor of 10). Notice that if $b \not\sim w_j$, then this does not necessarily mean that there are only few cubes $q \in Q^\mu(w_j)$ contained in b (since the cuboid $b(w_j)$ may not be unique), but it does imply that there are many cubes q lying "away" from b. To be more precise, if $b \not\sim w_j$, then

$$\big|\{q \in Q^\mu(w_j) : q \cap 5b = \varnothing\}\big| \ge \big|\{q \in Q^\mu(w_j) : q \subset b(w_j)\}\big| \gtrsim R^{-c\delta} |Q^\mu(w_j)|, \tag{2-38}$$

since only $\mathcal{O}(R^{2\delta})$ cuboids b meet T_{w_j}.

For a fixed b, we can decompose any given set $W_j \subset \mathcal{W}_j$ into $W_j^{\not\sim b} = \{w_j \in W_j : b \not\sim w_j\}$ and $W_j^{\sim b} = \{w_j \in W_j : b \sim w_j\}$. Thus we have

$$\Big\| \prod_{j=1,2} \sum_{w_j \in W_j^{\lambda_j,\mu}} p_{w_j} \sum_{q \in Q^\mu} \chi_q \Big\|_{L^p(Q_{S_1,S_2}(R))} \le \sum_b \Big\| \prod_{j=1,2} \sum_{w_j \in W_j^{\lambda_j,\mu}} p_{w_j} \sum_{q \in Q^\mu} \chi_q \Big\|_{L^p(b)} = \mathrm{I} + \mathrm{II} + \mathrm{III}, \tag{2-39}$$

[2]Here and in the subsequent considerations, c will denote some constant which is independent of R and S_1, S_2, but whose precise value may vary from line to line.

where

$$\mathrm{I}=\sum_b\Big\|\prod_{j=1,2}\sum_{w_j\in W_j^{\lambda_j,\mu,\sim b}}p_{w_j}\sum_{q\in Q^\mu}\chi_q\Big\|_{L^p(b)},$$

$$\mathrm{II}=\sum_b\Big\|\sum_{w_1\in W_1^{\lambda_1,\mu,\not\sim b}}p_{w_1}\sum_{w_2\in W_2^{\lambda_2,\mu}}p_{w_2}\sum_{q\in Q^\mu}\chi_q\Big\|_{L^p(b)},$$

$$\mathrm{III}=\sum_b\Big\|\sum_{w_1\in W_1^{\lambda_1,\mu,\sim b}}p_{w_1}\sum_{w_2\in W_2^{\lambda_2,\mu,\not\sim b}}p_{w_2}\sum_{q\in Q^\mu}\chi_q\Big\|_{L^p(b)}.$$

As usual in the bilinear approach, part I, which comprises the terms of highest density of wave packets over the cuboids b, will be handled by means of an inductive argument. The treatment of part II (and analogously of part III) will be based on a combination of geometric and combinatorial arguments. It is only here that the very choice of the $b(w_j)$ will become crucial.

Lemma 2.13. *Let $\alpha>0$, and assume that $E(\alpha)$ holds true. Then*

$$I\le C_\alpha R^{\alpha(1-\delta)}\log^{\gamma_\alpha}(1+R)(\kappa^1\kappa^2)^{\frac12-\frac1p}D^{3-\frac5p}\log^{\gamma_\alpha}(C_0(S))|W_1|^{\frac12}|W_2|^{\frac12}.$$

Proof. To shorten notation, write $C_1=C_\alpha(\kappa^1\kappa^2)^{\frac12-\frac1p}D^{3-\frac5p}\log^{\gamma_\alpha}(C_0(S))$. Recall the reproducing formula (P1) in Corollary 2.3: $p_{w_j}=R^*_{H_j}(\mathfrak{F}_{H_j}(p_{w_j}|_{H_j}))$. Since every cuboid b is a translate of $Q_{S_1,S_2}(R^{1-\delta})$, and since a translation of $R^*_{H_j}g$ corresponds to a modulation of the function g, we see that $E(\alpha)$ implies

$$\begin{aligned}
I&=\sum_b\Big\|\prod_{j=1,2}\sum_{w_j\in W_j^{\lambda_j,\mu,\sim b}}p_{w_j}\sum_{q\in Q^\mu}\chi_q\Big\|_{L^p(b)}\\
&\le\sum_b\Big\|\prod_{j=1,2}R^*_{H_j}\Big(\sum_{w_j\in W_j^{\lambda_j,\mu,\sim b}}\mathfrak{F}_{H_j}(p_{w_j}|_{H_j})\Big)\Big\|_{L^p(b)}\\
&\le C_1(R^{1-\delta})^\alpha\log^{\gamma_\alpha}(1+R^{1-\delta})\sum_b\prod_{j=1,2}\Big\|\sum_{w_j\in W_j^{\lambda_j,\mu,\sim b}}\mathfrak{F}_{H_j}(p_{w_j}|_{H_j})\Big\|_{L^2(H_j)}\\
&\le C_1R^{\alpha(1-\delta)}\log^{\gamma_\alpha}(1+R)\sum_b\prod_{j=1,2}|W_j^{\lambda_j,\mu,\sim b}|^{\frac12}.
\end{aligned}$$

In the last estimate, we have made use of property (P4). Moreover, using Hölder's inequality, we obtain

$$\sum_b\prod_{j=1,2}|W_j^{\lambda_j,\mu,\sim b}|^{\frac12}\le\prod_{j=1,2}\Big(\sum_b|W_j^{\lambda_j,\mu,\sim b}|\Big)^{\frac12},$$

where, due to Fubini's theorem (for sums),

$$\sum_b|W_j^{\lambda_j,\mu,\sim b}|=\sum_b|\{w_j\in W_j^{\lambda_j,\mu}:w_j\sim b\}|=\sum_{w_j\in W_j^{\lambda_j,\mu}}|\{b:\ b\sim w_j\}|\lesssim|W_j|.$$

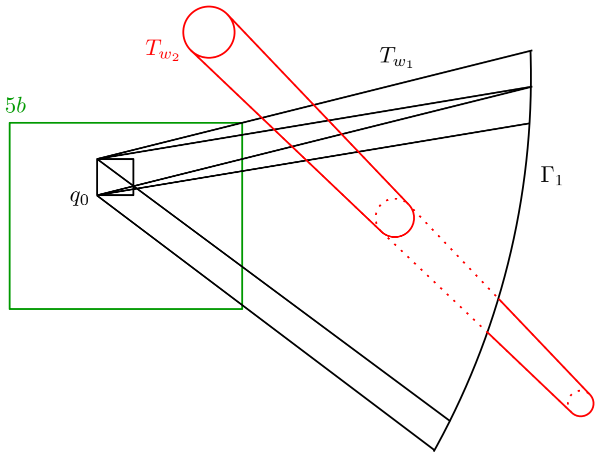


Figure 9. The geometry in Lemma 2.14.

In combination, these estimates yield

$$I \leq C_1 R^{\alpha(1-\delta)} \log^{\gamma_\alpha}(1+R) \prod_{j=1,2} |W_j|^{\frac{1}{2}}. \qquad \square$$

2G. ***The geometric argument.*** We next turn to the estimation of II and III. A crucial tool will be the following lemma, which is a variation of Lemma 2.3 in [Lee and Vargas 2010].

Lemma 2.14. *Let $\lambda_j, \mu_j \in \mathcal{N}$, $W_j \subset \mathcal{W}_j$, $v_j \in \mathcal{V}_j$, $j = 1, 2$, and let b and q_0 be cuboids from our collections such that $q_0 \cap 2b \neq \varnothing$. If we define $W_j^{\lambda_j,\mu,\not\sim b}(q_0) = W_j^{\lambda_j,\mu,\not\sim b} \cap W_j(q_0)$, then*

(i) $\lambda_1\mu_2 \, \big|[W_1^{\lambda_1,\mu,\not\sim b}(q_0)]^{\Pi_{v_1,v_2}}\big| \lesssim R^{c\delta}|W_2|$,

(ii) $\lambda_2\mu_1 \, \big|[W_2^{\lambda_2,\mu,\not\sim b}(q_0)]^{\Pi_{v_1,v_2}}\big| \lesssim R^{c\delta}|W_1|$.

Proof. We only show (i), since the proof of (ii) is analogous. Set

$$\Gamma_1 = \bigcup \{T_{w_1} : w_1 \in [W_1^{\lambda_1,\mu,\not\sim b}(q_0)]^{\Pi_{v_1,v_2}}\} \setminus 5b, \quad Q^\mu_{\Gamma_1} = \{q \in Q^\mu : R^\delta q \cap \Gamma_1 \neq \varnothing\}.$$

Since we have seen in Lemma 2.8(d) that T_{w_2} is transversal to Γ_1, we have

$$|Q^\mu_{\Gamma_1} \cap Q^\mu(w_2)| \lesssim R^{c\delta}. \tag{2-40}$$

Due to the separation of the tube directions, the sets $T_{w_1} \setminus 5b$ do not overlap too much. To be more precise, we claim that for all cubes $q \in Q^\mu_{\Gamma_1}$,

$$\big|\{w_1 \in [W_1^{\lambda_1,\mu,\not\sim b}(q_0)]^{\Pi_{v_1,v_2}} : R^\delta q \cap T_{w_1} \setminus 5b \neq \varnothing\}\big| \lesssim R^{c\delta}. \tag{2-41}$$

Indeed, let $w_1, w_1' \in [W_1^{\lambda_1,\mu,\not\sim b}(q_0)]^{\Pi_{v_1,v_2}}$ and $x \in R^\delta q \cap T_{w_1} \setminus 5b$, $x' \in R^\delta q \cap T_{w_1'} \setminus 5b$. The definition of $W_1(q_0)$ means that we can find $x_0 \in R^\delta q_0 \cap T_{w_1}$ and $x_0' \in R^\delta q_0 \cap T_{w_1'}$; then we may write

$$x = x_0 + |x - x_0| N(v_1) + \mathcal{O}(R') \quad \text{and} \quad x' = x_0' + |x' - x_0'| N_0(v_1') + \mathcal{O}(R'). \tag{2-42}$$

Furthermore we have

$$\big||x-x_0|-|x'-x_0'|\big| \le |x-x'|+|x_0-x_0'| = \mathcal{O}(R^{c\delta}R'). \tag{2-43}$$

Since T_{w_1} has length $(R')^2/\kappa^{(1)}$, so that the length of b in the direction of T_{w_1} is at least $R^{-2\delta}(R')^2/\kappa^{(1)}$, and since $x_0 \in R^\delta q_0 \subset 4b$ but $x \notin 5b$, we conclude that

$$R^{-2\delta}\frac{(R')^2}{\kappa^{(1)}} \le |x-x_0|. \tag{2-44}$$

Applying Lemma 2.5, and consecutively making use of the estimates (2-44), (2-43), (2-42) and again (2-43), we obtain

$$\begin{aligned} R'|v_1-v_1'| &\lesssim \frac{R'}{\kappa^{(1)}}|N(v_1)-N(v_1')| \\ &\lesssim R^{2\delta}(R')^{-1}|x-x_0|\,|N_0(v_1)-N_0(v_1')| \\ &\lesssim R^{2\delta}(R')^{-1}\big||x-x_0|N_0(v_1)-|x'-x_0'|N_0(v_1')\big| + \mathcal{O}(R^{c\delta}) \\ &\lesssim R^{2\delta}(R')^{-1}\big(|x-x'|+|x_0'-x_0|\big)+\mathcal{O}(R^{c\delta}) = \mathcal{O}(R^{c\delta}). \end{aligned}$$

Recall also that the direction of a tube T_{w_1} with $w_1=(y_1,v_1)$ depends only on v_1, and thus the set of all these directions corresponding to the set

$$\big\{w_1 \in [W_1^{\lambda_1,\mu,\not\sim b}(q_0)]^{\Pi_{v_1,v_2}} : R^\delta q \cap T_{w_1}\setminus 5b\big\}$$

has cardinality $\mathcal{O}(R^{c\delta})$. But, for a fixed direction v_1, the number of parameters y_1 such that the tube $T_{(y_1,v_1)}$ passes through $R^{c\delta}q_0$ is bounded by $\mathcal{O}(R^{c\delta})$ anyway, and thus (2-41) holds true.

Recall next from (2-38) that for $w_1 \not\sim b$ we have $R^{-c\delta}|Q^\mu(w_1)| \lesssim |\{q \in Q^\mu(w_1) : q \cap 5b = \varnothing\}|$. Since for $w_1 \in W_1^{\lambda_1,\mu}$ we have $|Q^\mu(w_1)| \sim \lambda_1$, we may thus estimate

$$\begin{aligned} R^{-c\delta}\lambda_1\big|[W_1^{\lambda_1,\mu,\not\sim b}(q_0)]^{\Pi_{v_1,v_2}}\big| &\lesssim R^{-c\delta}\sum\big|Q^\mu(w_1)\big| \\ &\lesssim \sum\big|\{q\in Q^\mu(w_1) : q\cap 5b=\varnothing\}\big| \\ &\le \sum\big|\{q\in Q^\mu : R^\delta q\cap T_{w_1}\neq\varnothing,\ R^\delta q\cap 5b=\varnothing\}\big| \\ &\le \sum\big|\{q\in Q^\mu : R^\delta q\cap (T_{w_1}\setminus 5b)\neq\varnothing\}\big| \\ &= \sum_{q\in Q^\mu}\big|\big\{w_1\in[W_1^{\lambda_1,\mu,\not\sim b}(q_0)]^{\Pi_{v_1,v_2}} : R^\delta q\cap(T_{w_1}\setminus 5b)\neq\varnothing\big\}\big| \\ &= R^{c\delta}|Q^\mu_{\Gamma_1}|, \end{aligned}$$

where sums are taken over $w_1 \in [W_1^{\lambda_1,\mu,\not\sim b}(q_0)]^{\Pi_{v_1,v_2}}$ unless otherwise indicated, and where we have used (2-41) in the last estimate. But, by (2-40), we also have

$$\mu_2|Q^\mu_{\Gamma_1}| = \sum_{q\in Q^\mu_{\Gamma_1}}|W_2(q)| \le \sum_{w_2\in W_2}|Q^\mu_{\Gamma_1}\cap Q^\mu(w_2)| \lesssim R^{c\delta}|W_2|,$$

and combining this with the previous estimate we arrive at the desired estimate in (i). □

Lemma 2.15. *Let $0<\delta<\frac{1}{4}$. Then*

$$\mathrm{II}=\sum_b \Big\| \sum_{w_1\in W_1^{\lambda_1,\mu,\not\sim b}} p_{w_1} \sum_{w_2\in W_2^{\lambda_2,\mu}} p_{w_2} \sum_{q\in Q^\mu} \chi_q \Big\|_{L^p(b)} \le C_\alpha R^{c\delta} (\kappa^1\kappa^2)^{\frac12-\frac1p} D^{3-\frac5p} |W_1|^{\frac12} |W_2|^{\frac12} \tag{2-45}$$

and

$$\mathrm{III}=\sum_b \Big\| \sum_{w_1\in W_1^{\lambda_1,\mu,\sim b}} p_{w_1} \sum_{w_2\in W_2^{\lambda_2,\mu,\not\sim b}} p_{w_2} \sum_{q\in Q^\mu} \chi_q \Big\|_{L^p(b)} \le C_\alpha R^{c\delta} (\kappa^1\kappa^2)^{\frac12-\frac1p} D^{3-\frac5p} |W_1|^{\frac12} |W_2|^{\frac12}. \tag{2-46}$$

Proof. We will only prove the first inequality; the proof of second one works in a similar way. Since the number of b's over which we sum in (2-45) is of the order $R^{c\delta}$, it is enough to show that for every fixed b

$$\Big\| \sum_{w_1\in W_1^{\lambda_1,\mu,\not\sim b}} p_{w_1} \sum_{w_2\in W_2^{\lambda_2,\mu}} p_{w_2} \sum_{q\in Q^\mu} \chi_q \Big\|_{L^p(b)} \le C_\alpha R^{c\delta} (\kappa^1\kappa^2)^{\frac12-\frac1p} D^{3-\frac5p} |W_1|^{\frac12} |W_2|^{\frac12}. \tag{2-47}$$

For $p=1$, we apply (2-35) from Lemma 2.10:

$$\begin{aligned}\Big\| \sum_{w_1\in W_1^{\lambda_1,\mu,\not\sim b}} p_{w_1} \sum_{w_2\in W_2^{\lambda_2,\mu}} p_{w_2} \sum_{q\in Q^\mu} \chi_q \Big\|_{L^1(b)} &\lesssim \Big\| \sum_{w_1\in W_1^{\lambda_1,\mu,\not\sim b}} p_{w_1} \sum_{w_2\in W_2^{\lambda_2,\mu}} p_{w_2} \Big\|_{L^1(Q_{S_1,S_2}(R))} \\ &\le \frac{(R')^2}{\sqrt{\kappa^{(1)}\kappa^{(2)}}} |W_1|^{\frac12} |W_1|^{\frac12}.\end{aligned}$$

For $p=2$, we claim that

$$\Big\| \sum_{w_1\in W_1^{\lambda_1,\mu,\not\sim b}} p_{w_1} \sum_{w_2\in W_2^{\lambda_2,\mu}} p_{w_2} \sum_{q\in Q^\mu} \chi_q \Big\|^2_{L^2(b)} \lesssim C_\alpha R^{c\delta} (R')^{-1} |W_1||W_2|. \tag{2-48}$$

The desired inequality (2-45) will then follow by means of interpolation with the previous L^1-estimate — notice here that $R^{5/p-3}\le 1$ since $\frac53\le p$.

To prove (2-48), recall that the side lengths of b are of the form

$$\left(\frac{(R')^2}{\kappa^{(j)}}\right) R^{-2\delta} = \frac{R'}{D\kappa^{(j)}} R^{1-2\delta} \ge R' R^{1-2\delta}, \quad j\in\{1,2\}.$$

If $q\cap 2b=\varnothing$, then for $x\in b$ we have $|x-c_q|\ge \inf_{y\notin 2b}|x-y| = d(x,(2b)^c)\ge R'R^{1-2\delta}$. Therefore for every $x\in b$,

$$\begin{aligned}\Big| \sum_{\substack{q\in Q^\mu \\ q\cap 2b=\varnothing}} \chi_q(x) \Big| &\le C_N \sum_{\substack{l\in\mathbb{N} \\ 2^l\ge R^{1-2\delta}}} \sum_{\substack{q\in Q^\mu \\ |x-c_q|\sim R'2^l}} \left(1+\frac{|x-c_q|}{R'}\right)^{-N-2} \lesssim C_N \sum_{\substack{l\in\mathbb{N} \\ 2^l\ge R^{1-2\delta}}} \big|\{q: |x-c_q|\sim R'2^l\}\big| 2^{-(N+2)l} \\ &\sim C_N \sum_{\substack{l\in\mathbb{N} \\ 2^l\ge R^{1-2\delta}}} 2^{-Nl} \sim C_N R^{-(1-2\delta)N} = C_{\delta,N'} R^{-N'}.\end{aligned} \tag{2-49}$$

The last step requires that $\delta < \frac{1}{2}$. Choosing N sufficiently large, we see that by Lemma 2.10 and Lemma 2.11,

$$\begin{aligned}\Bigg\| \sum_{w_1 \in W_1^{\lambda_1,\mu,\not\sim b}} p_{w_1} \sum_{w_2 \in W_2^{\lambda_2,\mu}} p_{w_2} \sum_{\substack{q \in Q^\mu \\ q \cap 2b = \varnothing}} \chi_q \Bigg\|_{L^2(b)}^2 &\lesssim \Bigg\| \sum_{w_1 \in W_1^{\lambda_1,\mu,\not\sim b}} p_{w_1} \sum_{w_2 \in W_2^{\lambda_2,\mu}} p_{w_2} \Bigg\|_{L^2}^2 \Bigg\| \sum_{\substack{q \in Q^\mu \\ q \cap 2b = \varnothing}} \chi_q \Bigg\|_{L^\infty(b)}^2 \\ &\lesssim C_{\delta,N'} (R')^{-1} |W_1||W_2| \min_j \sup_{v_1,v_2} |V_j^{\Pi_{v_1,v_2}}| R^{-2N'} \\ &\lesssim C_{\delta,N'} (R')^{-1} |W_1||W_2| R^{1-2N'} \\ &\lesssim C'_{\delta,N''} (R')^{-1} |W_1||W_2| R^{-N''}.\end{aligned}$$

Thus it is enough to consider the sum over the set $Q_b^\mu = \{q \in Q^\mu : q \cap 2b \neq \varnothing\}$. For fixed w_1, w_2, we split this set into the subsets $Q_b^\mu(w_1,w_2) = Q_b^\mu \cap Q^\mu(w_1) \cap Q^\mu(w_2)$ and $Q_b^\mu \cap Q^\mu(w_1) \setminus Q^\mu(w_2)$ and

$$Q_b^\mu \setminus Q^\mu(w_1) = \big(Q_b^\mu \cap Q^\mu(w_2) \setminus Q^\mu(w_1)\big) \cup \big(Q_b^\mu \setminus (Q^\mu(w_2) \cap Q^\mu(w_1))\big).$$

Except for the first set, the contributions by the other subsets can be treated in the same way, since they are all special cases of the following situation:

Let $Q_0 = Q_0(w_1,w_2) \subset Q_b^\mu$ such that there exists an $j \in \{1,2\}$ with $R^\delta q \cap T_{w_j} = \varnothing$ for all $q \in Q_0$. Then

$$\Bigg\| \sum_{w_1 \in W_1^{\lambda_1,\mu,\not\sim b}} p_{w_1} \sum_{w_2 \in W_2^{\lambda_2,\mu}} p_{w_2} \sum_{q \in Q_0} \chi_q \Bigg\|_{L^2(b)}^2 \lesssim C_\alpha R^{c\delta} (R')^{-1} |W_1||W_2|. \tag{2-50}$$

Notice that the right-hand side is just what we need for (2-48).

For the proof of (2-50), assume without loss of generality that $j = 1$. Let $q \in Q_0$. Then $T_{w_1} \cap R^\delta q = \varnothing$, and for all $x \in (R^\delta/2)q$ we have

$$(R^\delta/2)R' \le \operatorname{dist}(x, (R^\delta q)^c) \le \operatorname{dist}(x, T_{w_1}).$$

Thus for every $x \in Q_{S_1,S_2}(R)$, we have $\operatorname{dist}(x, T_{w_1}) \ge (R^\delta/2)R'$ or $x \notin (R^\delta/2)q$. In the first case, we have

$$|p_{w_1}(x)| \le C_N (R')^{-1} \left(1 + \frac{\operatorname{dist}(x, T_{w_1})}{R'}\right)^{-2N} \le C'_N (R')^{-1} R^{-\delta N} \left(1 + \frac{\operatorname{dist}(x, T_{w_1})}{R'}\right)^{-N}. \tag{2-51}$$

One the other hand, in the second case, where $x \notin (R^\delta/2)q$, we have $(R^\delta/2)R' \le |x - c_q|$. Using the rapid decay of the Schwartz function ϕ we then see that

$$|\chi_q(x)| = \left|\chi\left(\frac{x - c_q}{R'}\right)\right| \le C_N \left(\frac{|x - c_q|}{R'}\right)^{-N} \le C'_N R^{-\delta N}. \tag{2-52}$$

Applying an argument similar to the one used in (2-49), we even obtain

$$\left|\sum_{q \in Q_0} \chi_q(x)\right| \le C''_N R^{-\delta N}$$

for all $x \notin (R^\delta/2)q$. To summarize, we obtain that for every $x \in Q_{S_1,S_2}(R)$,

$$\Big| p_{w_1} \sum_{q \in Q_0(w_1,w_2)} \chi_q \Big|(x) \le C(N,\delta)(R')^{-1} R^{-\delta N} \Big(1 + \frac{\operatorname{dist}(x, T_{w_1})}{R'}\Big)^{-N}. \tag{2-53}$$

This means that the expression $p_{w_1} \sum_{q \in Q_0(w_1,w_2)} \chi_q$ cannot only be estimated in the same way as the original wave packet p_{w_1}, but we even obtain an improved estimate because of an additional factor $R^{-\delta N}$. If we replace p_{w_1} by p_{w_2} on the left-hand side, we obtain in a similar way just the standard wave packet estimate

$$\Big| p_{w_2} \sum_{q \in Q_0(w_1,w_2)} \chi_q \Big|(x) \lesssim \|p_{w_2}\|_\infty \lesssim (R')^{-1} \Big(1 + \frac{\operatorname{dist}(x, T_{w_2})}{R'}\Big)^{-N}, \tag{2-54}$$

without an additional factor.

We can now finish the proof of (2-50), basically by following the ideas of the proof of the estimate (2-36) in Lemma 2.10. The crucial argument was the fact that the Fourier transform of $p_{w'_{j+1}} p_{w_j}$ is supported in $v'_{j+1} + v_j + \mathcal{O}((R')^{-1})$. Since $\operatorname{supp} \hat{\chi}_q = \operatorname{supp} \hat{\chi}(R' \cdot) \subset B(0, (R')^{-1})$, the Fourier support of $p_{w_1} p_{w_2} \sum_{q \in Q_0(w_1,w_2)} \chi_q$ remains essentially the same. It is at this point that we need that the functions χ_q have compact Fourier support. The modified wave packets $p_{w_i} \sum_{q \in Q_0(w_1,w_2)} \chi_q$ are still well separated with respect to the parameter y_i, for fixed direction v_i, thanks to (2-53) and (2-54). Thus the argument from Lemma 2.10 applies, and by the analogue of (2-37) we obtain

$$\begin{aligned}
&\Big\| \sum_{\substack{w_1 \in W_1^{\lambda_1,\mu,\not\sim b} \\ w_2 \in W_2^{\lambda_2,\mu}}} p_{w_1} p_{w_2} \sum_{q \in Q_0(w_1,w_2)} \chi_q \Big\|_{L^2(b)}^2 \\
&\quad \lesssim R' |W_1||W_2| \min_j \sup_{v_1,v_2} |V_j^{\Pi_{v_1,v_2}}| \sup_{\substack{w_1 \in W_1 \\ w'_2 \in W_2}} \Big\| p_{w'_2} \sum_{q \in Q_0(w_1,w'_2)} \chi_q \Big\|_\infty \sup_{\substack{w'_1 \in W_1 \\ w_2 \in W_2}} \Big\| p_{w'_1} \sum_{q \in Q_0(w'_1,w_2)} \chi_q \Big\|_\infty \\
&\quad \lesssim C_{\delta,N'} (R')^{-1} |W_1||W_2| \min_j \sup_{v_1,v_2} |V_j^{\Pi_{v_1,v_2}}| R^{-N'} \lesssim C_{\delta,N'} (R')^{-1} R^{1-N'} |W_1||W_2|.
\end{aligned}$$

In the second inequality, we have made use of (2-53) and (2-54), and the last one is based on Lemma 2.11. This concludes the proof of (2-50).

What remains to be controlled are the contributions by the cubes q from $Q_b^\mu(w_1, w_2)$. Notice that the kernel $K(q,q') = \chi_q(x)\chi_{q'}(x)$ satisfies Schur's test condition

$$\sup_q \sum_{q'} \chi_q(x)\chi_{q'}(x) \lesssim \sum_{q'} \chi_{q'}(x) \lesssim 1,$$

with a constant not depending on x. Let us put

$$f_q = \sum p_{w_1} p_{w_2},$$

where the sum is taken over $w_1 \in W_1^{\lambda_1,\mu,\not\sim b}(q)$ and $w_2 \in W_2^{\lambda_2,\mu}(q)$. Observe that for $w_1 \in W_1^{\lambda_1,\mu,\not\sim b}$ and $w_2 \in W_2^{\lambda_2,\mu}$, we have $q \in Q_b^\mu(w_1, w_2)$ if and only if $q \in Q_b^\mu$ and $w_1 \in W_1^{\lambda_1,\mu,\not\sim b}(q)$ and $w_2 \in W_2^{\lambda_2,\mu}(q)$.

Then we see that we may estimate

$$\begin{aligned}\Big\| \sum_{\substack{w_1 \in W_1^{\lambda_1,\mu,\not\sim b}\\ w_2 \in W_2^{\lambda_2,\mu}}} p_{w_1} p_{w_2} \sum_{q \in Q_b^\mu(w_1,w_2)} \chi_q \Big\|_{L^2(b)}^2 &= \Big\| \sum_{q \in Q_b^\mu} \chi_q f_q \Big\|_{L^2(b)}^2 \\ &= \int_b \Big| \sum_{q,q' \in Q_b^\mu} \chi_q \chi_{q'} f_q f_{q'} \Big| \, dx = \int_b \Big| \sum_{q,q' \in Q_b^\mu} K(q,q') f_q f_{q'} \Big| \, dx \\ &\lesssim \int_b \sum_{q \in Q_b^\mu} |f_q|^2 \, dx = \sum_{\substack{q \in Q^\mu\\ q \cap 2b \neq \varnothing}} \|f_q\|_{L^2(b)}^2.\end{aligned}$$

Invoking also Lemma 2.10 and Lemma 2.14(i), we thus obtain

$$\begin{aligned}&\Big\| \sum_{\substack{w_1 \in W_1^{\lambda_1,\mu,\not\sim b}\\ w_2 \in W_2^{\lambda_2,\mu}}} p_{w_1} p_{w_2} \sum_{q \in Q_b^\mu(w_1,w_2)} \chi_q \Big\|_{L^2(b)}^2 \\ &\quad\lesssim \sum_{\substack{q \in Q^\mu\\ q \cap 2b \neq \varnothing}} (R')^{-1} |W_1^{\lambda_1,\mu,\not\sim b}(q)| \, |W_2^{\lambda_2,\mu}(q)| \sup_{v_1,v_2} \big| [W_1^{\lambda_1,\mu,\not\sim b}(q)]^{\Pi_{v_1,v_2}} \big| \\ &\quad\lesssim R^{c\delta} (R')^{-1} \sum_{\substack{q \in Q^\mu\\ q \cap 2b \neq \varnothing}} |W_1^{\lambda_1,\mu}(q)| \, |W_2(q)| \frac{|W_2|}{\lambda_1 \mu_2} \\ &\quad\lesssim R^{c\delta} (R')^{-1} \sum_{w_1 \in W_1^{\lambda_1,\mu}} |Q^\mu(w_1)| \frac{|W_2|}{\lambda_1} \lesssim R^{c\delta} (R')^{-1} |W_1| \, |W_2|.\end{aligned}$$

This completes the proof of estimate (2-45), and hence of Lemma 2.15. □

2H. *Induction on scales.* We can now easily complete the proof of Theorem 2.6 by following standard arguments.

Corollary 2.16. *There exist constants $c, \delta_0 > 0$ such that $c\delta_0 > 1$ and such that the following holds true*:

Whenever $\alpha > 0$ is such that $E(\alpha)$ holds true, then $E(\max\{\alpha(1-\delta), c\delta\})$ holds true for every δ such that $0 < \delta < \delta_0$.

Proof. Let us put $\delta_0 = \frac{1}{4}$. Then the previous Lemmas 2.13 and 2.15 imply

$$\begin{aligned}&\Big\| \prod_{j=1,2} \sum_{w_j \in W_j^{\lambda_j,\mu}} p_{w_j} \sum_{q \in Q^\mu} \chi_q \Big\|_{L^p(Q_{S_1,S_2}(R))} \\ &\quad\leq \mathrm{I} + \mathrm{II} + \mathrm{III} \\ &\quad\lesssim \big(C_\alpha R^{\alpha(1-\delta)} \log^{\gamma_\alpha}(1+R) + C_\delta R^{c\delta}\big) (\kappa^1 \kappa^2)^{\frac{1}{2}-\frac{1}{p}} D^{3-\frac{5}{p}} \log^{\gamma_\alpha}(C_0) |W_1|^{\frac{1}{2}} |W_2|^{\frac{1}{2}} \\ &\quad\lesssim C_{\alpha,\delta} R^{\alpha(1-\delta)\vee c\delta} \log^{\gamma_\alpha}(1+R) (\kappa^1 \kappa^2)^{\frac{1}{2}-\frac{1}{p}} D^{3-\frac{5}{p}} \log^{\gamma_\alpha}(C_0) |W_1|^{\frac{1}{2}} |W_2|^{\frac{1}{2}}\end{aligned}$$

whenever $\delta < \delta_0$, where $C_0 = C_0(S) \gtrsim 1$ is defined in (2-20). By Lemma 2.9, this estimate implies $E(\alpha(1-\delta) \vee c\delta)$. Finally, by simply increasing the constant c, if necessary, we may also ensure $c\delta_0 > 1$. $\square$

Corollary 2.17. *$E(\alpha)$ holds true for every $\alpha > 0$.*

This completes also the proof of Theorem 2.6.

Proof. Define inductively the sequence $\alpha_0 = 1$, $\alpha_{j+1} = c\alpha_j/(c+\alpha_j)$, which is decreasing and converges to 0. It therefore suffices to prove that $E(\alpha_j)$ is valid for every $j \in \mathbb{N}$. But, by Corollary 2.12, $E(\alpha_0) = E(1)$ does hold true. Moreover, Corollary 2.16 shows that $E(\alpha_j)$ implies $E(\alpha_{j+1})$, for if we choose $\delta = \alpha_j/(c+\alpha_j)$, then $\delta < 1/c < \delta_0$ and $\alpha_j(1-\delta) = c\delta = \alpha_j c/(c+\alpha_j) = \alpha_{j+1}$, and thus we may conclude by induction. $\square$

3. Scaling

For the proof of our main theorem, we shall have to perform a kind of Whitney-type decomposition of $S \times S$ into pairs of patches of hypersurfaces (S_1, S_2) and prove very precise bilinear restriction estimates for those. In order to reduce these estimates to Section 2B, we shall need to rescale simultaneously the hypersurfaces S_1, S_2 for each such pair (S_1, S_2) in a suitable way. To this end, we shall denote here and in the sequel by $R^*_{S_1,S_2}$ the bilinear Fourier extension operator

$$R^*_{S_1,S_2}(f_1, f_2) = R^*_{\mathbb{R}^2} f_1 \cdot R^*_{\mathbb{R}^2} f_2, \quad f_1 \in L^2(U_1),\ f_2 \in L^2(U_2)$$

associated to any pair of hypersurfaces (S_1, S_2) given as the graphs $S_j = \{(\xi, \phi_j(\xi)) : \xi \in U_j\}$, $j = 1, 2$.

The following trivial lemma comprises the effect of the type of rescaling that we shall need.

Lemma 3.1. *Let $S_j = \{(\xi, \phi(\xi) : \xi \in U_j\}$, where again $U_j \subset \mathbb{R}^d$ is open and bounded for $j = 1, 2$. Let $A \in \mathrm{GL}(d, \mathbb{R})$, $a > 0$, put $\phi^s(\eta) = (1/a)\phi(A\eta)$, and let*

$$S_j^s = \{(\eta, \phi^s(\eta)) : \eta \in U_j^s\}, \quad U_j^s = A^{-1}(U_j), \quad j = 1, 2.$$

For any measurable subset $Q^s \subset \mathbb{R}^{d+1}$, we set $Q = \{x : ({}^tAx', ax_{d+1}) \in Q^s\}$. Assume the following estimate holds true:

$$\|R^*_{S_1^s,S_2^s}(g_1, g_2)\|_{L^p(Q^s)} \le C_s \|g_1\|_2 \|g_2\|_2 \quad \text{for all } g_j \in L^2(U_j^s).$$

Then

$$\|R^*_{S_1,S_2}(f_1, f_2)\|_{L^p(Q)} \le C_s |\det A|^{\frac{1}{p'}} a^{-\frac{1}{p}} \|f_1\|_2 \|f_2\|_2 \quad \text{for all } f_j \in L^2(U_j).$$

We now return to our model hypersurface (see (1-3), (1-4) and (1-5)), which is the graph of

$$\phi(\xi_1, \xi_2) = \phi_{(1)}(\xi_1) + \phi_{(2)}(\xi_2)$$

on $]0, 1[\times]0, 1[$, where the derivatives of the $\phi_{(i)}$ satisfy

$$\phi''_{(i)}(\xi_i) \sim \xi_i^{m_i - 2},$$
$$|\phi^{(k)}_{(i)}(\xi_i)| \lesssim \xi_i^{m_i - k} \quad \text{for } k \ge 3,$$

and where $m_1, m_2 \in \mathbb{R}$ are such that $m_i \ge 2$.

We shall apply the preceding lemma to pairs $S_1 = S$ and $S_2 = \widetilde{S}$ of patches of this hypersurface on which the following assumptions are met:

General Assumptions. Let $S = \{(\xi, \phi(\xi)) : \xi \in U\}$ and $\widetilde{S} = \{(\xi, \phi(\xi)) : \xi \in \widetilde{U}\}$, where $U = r + [0, d_1] \times [0, d_2]$ and $\widetilde{U} = \tilde{r} + [0, \tilde{d}_1] \times [0, \tilde{d}_2]$, with $r = (r_1, r_2)$ and $\tilde{r} = (\tilde{r}_1, \tilde{r}_2)$.

We assume that for $i = 1, 2$ we have $r_i \geq d_i$ and $\tilde{r}_i \geq \tilde{d}_i$, so that the principal curvature $\phi''_{(i)}$ of S with respect to ξ_i is comparable to $\kappa_i = r_i^{m_i-2}$, and that of $\widetilde{S}$ is comparable to $\tilde{\kappa}_i = \tilde{r}_i^{m_i-2}$. We put

$$\begin{gathered} \bar{d}_i = d_i \vee \tilde{d}_i, \quad \bar{r}_i = r_i \vee \tilde{r}_i, \quad \Delta r_i = r_i - \tilde{r}_i, \\ \kappa = \kappa_1 \vee \kappa_2, \quad \tilde{\kappa} = \tilde{\kappa}_1 \vee \tilde{\kappa}_2, \bar{\kappa}_i = \kappa_i \vee \tilde{\kappa}_i, \quad \bar{\kappa} = \kappa \vee \tilde{\kappa} = \bar{\kappa}_1 \vee \bar{\kappa}_2. \end{gathered} \tag{3-1}$$

In addition, *we assume that for each direction ξ_1 and ξ_2 the rectangle U or $\widetilde{U}$ respectively on which the corresponding principal curvature is bigger (which means that its projection to the ξ_i-axis is the one further to the right) has also bigger length in this direction.* This is easily seen to be equivalent to

$$(\kappa_i d_i) \vee (\tilde{\kappa}_i \tilde{d}_i) = \bar{\kappa}_i \bar{d}_i. \tag{3-2}$$

Last, but not least, we assume the rectangles U and $\widetilde{U}$ are separated with respect to both variables ξ_i, $i = 1, 2$, in the following sense:

$$\operatorname{dist}_{\xi_i}(U, \widetilde{U}) = \inf\bigl\{|\xi_i - \tilde{\xi}_i| : \xi \in U,\ \tilde{\xi} \in \widetilde{U}\bigr\} \sim |\Delta r_i| \sim \bar{d}_i. \tag{3-3}$$

Given these assumptions, we shall introduce a rescaling as follows: we put

$$a_1 = \bar{\kappa}_2 \bar{d}_2, \quad a_2 = \bar{\kappa}_1 \bar{d}_1, \tag{3-4}$$

and

$$\phi^s(\eta) = \frac{1}{a} \phi(A\eta) = \frac{1}{a_1 a_2} \phi(a_1 \eta_1, a_2 \eta_2). \tag{3-5}$$

The quantities that arise after this scaling will be denoted by a superscript s; i.e.,

$$\begin{gathered} r_i^s = \frac{r_i}{a_i}, \quad d_i^s = \frac{d_i}{a_i}, \quad \kappa_i^s = \frac{1}{a_1 a_2} a_i^2 \kappa_i = \frac{a_i}{a_{i+1 \bmod 2}} \kappa_i, \\ D^s = \min\{d_1^s, d_2^s, \tilde{d}_1^s, \tilde{d}_2^s\}, \quad U^s = r^s + [0, d_1^s] \times [0, d_2^s], \end{gathered}$$

with corresponding expressions for $\tilde{r}^s$, $\tilde{d}_i^s$, $\tilde{\kappa}_i^s$ and $\widetilde{U}^s$. For later use, recall also the normal field N on $S \cup \widetilde{S}$ defined by $N(\xi, \phi(\xi)) = (-\nabla\phi(\xi), 1)$ and the corresponding unit normal field $N_0 = N/|N|$. After scaling, the corresponding normal fields on $S^s \cup \widetilde{S}^s$ will be denoted by N^s and N_0^s. With our choice of scaling, the following lemma holds true:

Lemma 3.2 (scaling). (i) *For $i = 1, 2$ and all $\eta \in U^s$ and $\tilde{\eta} \in \widetilde{U}^s$ we have*

$$|\partial_i \phi^s(\eta) - \partial_i \phi^s(r^s)| \lesssim \kappa_i^s d_i^s \lesssim 1 \quad \textit{and} \quad |\partial_i \phi^s(\tilde{\eta}) - \partial_i \phi^s(\tilde{r}^s)| \lesssim \tilde{\kappa}_i^s \tilde{d}_i^s \lesssim 1.$$

Moreover, $\bar{\kappa}_i^s \bar{d}_i^s = 1$.

(ii) *For every $|\alpha| \geq 2$ and all $\eta \in U^s$ and $\tilde{\eta} \in \widetilde{U}^s$,*

$$|\partial^\alpha \phi^s(\eta)| \lesssim \kappa^s |d_1^s \wedge d_2^s|^{2-|\alpha|} \quad \textit{and} \quad |\partial^\alpha \phi^s(\tilde{\eta})| \lesssim \tilde{\kappa}^s |\tilde{d}_1^s \wedge \tilde{d}_2^s|^{2-|\alpha|}.$$

(iii) *For $i = 1, 2$, i.e., with respect to both variables, the separation condition*

$$|\partial_i \phi^s(\eta) - \partial_i \phi^s(\tilde{\eta})| \sim 1 \quad \textit{for all } \eta \in S,\ \tilde{\eta} \in \widetilde{S}$$

holds true.

In particular, the rescaled pair of hypersurfaces $(S^s, \widetilde{S}^s)$ satisfies the general assumptions (i)–(iii) introduced before Theorem 2.6.

Proof. Observe first that

$$\bar{d}_i^s = \frac{\bar{d}_i}{a_i}, \quad \bar{\kappa}_i^s = \frac{1}{a_1 a_2} a_i^2 \bar{\kappa}_i,$$

and thus, by the definition of a_i, we see that $\bar{\kappa}_i^s \bar{d}_i^s = 1$.

Next, in order to prove (i), observe that for $\eta \in U^s$,

$$|\partial_i \phi^s(\eta) - \partial_i \phi^s(r^s)| \leq \sup_{\eta' \in U} |\partial_i^2 \phi^s(\eta')| |\eta_i - r_i^s| \lesssim \kappa_i^s d_i^s,$$

with $\kappa_i^s d_i^s \leq \bar{\kappa}_i^s \bar{d}_i^s = 1$.

As for (ii), notice that also $\partial_1 \partial_2 \phi^s \equiv 0$. In the unscaled situation, we have for $k \geq 2$ and every $\xi \in U$,

$$|\partial_i^k \phi(\xi)| \lesssim \xi_i^{m_i - k} \sim \partial_i^2 \phi(\xi) \xi_i^{2-k} \sim \kappa_i \xi_i^{2-k}.$$

Thus, for $\eta \in U^s$, we find that

$$|\partial_i^k \phi^s(\eta)| = \frac{1}{a_1 a_2} a_i^k |\partial_i^k \phi(A\eta)| \lesssim \frac{1}{a_1 a_2} a_i^k \kappa_i (a_i \eta_i)^{2-k} = \frac{a_i^2}{a_1 a_2} \kappa_i \eta_i^{2-k} = \kappa_i^s \eta_i^{2-k}.$$

On the other hand, for $\eta \in U^s$ we have

$$\eta_i \geq r_i^s = \frac{r_i}{a_i} \geq \frac{d_i}{a_i} = d_i^s \geq d_1^s \wedge d_2^s,$$

and thus we conclude that

$$|\partial_i^k \phi^s(\eta)| \lesssim \kappa^s (d_1^s \wedge d_2^s)^{2-k}, \quad k \geq 2.$$

In the same way, we obtain the corresponding result for $\eta \in \widetilde{U}^s$. These estimates imply (ii).

Finally, in order to prove (iii), let $\xi = (\xi_1, \xi_2) \in U$ and $\tilde{\xi} = (\tilde{\xi} - 1, \tilde{\xi}_2) \in \widetilde{U}$. Then, by (3-3), we see that $|\xi_i - \tilde{\xi}_i| \sim \bar{d}_i$. Moreover, if for instance $r_i < \tilde{r}_i$ (the other case can be treated analogously), then by (3-3) we even have $r_i + d_i + c\bar{d}_i \leq \tilde{r}_i$ for some admissible constant $c > 0$ such that $c < 1$. But then $\kappa_i \lesssim |\phi''_{(i)}(t)| \leq \tilde{\kappa}_i$ for every t in between ξ_i and $\tilde{\xi}_i$, and moreover $\phi''_{(i)}(t) \sim \tilde{\kappa}_i = \bar{\kappa}_1$ on the subinterval $[\tilde{r}_i - c\bar{d}_i/4, \tilde{r}_i]$, and thus

$$|\partial_i \phi(\xi) - \partial_i \phi(\tilde{\xi})| = \left| \int_{\xi_i}^{\tilde{\xi}_i} \phi''_{(i)}(t)\, dt \right| \sim \bar{\kappa}_i \bar{d}_i = a_{i+1 \bmod 2};$$

hence

$$|\partial_i\phi^s(\eta)-\partial_i\phi^s(\tilde\eta)| = \frac{|\partial_i\phi(A\eta)-\partial_i\phi(A\tilde\eta)|}{a_{i+1 \bmod 2}} \sim 1. \tag{3-6}$$

This completes the proof. ☐

In view of Lemma 3.2, we may now apply Theorem 2.6 to the rescaled phase function ϕ^s. According to (2-19), the scaled cuboids are given by

$$Q^0_{S^s,\tilde S^s}(R)=\left\{x\in\mathbb{R}^3:\ |x_i+\partial_i\phi^s(r_0^s)x_3|\le \frac{R^2}{(D^s)^2\bar\kappa^s}, i=1,2,\ |x_3|\le \frac{R^2}{(D^s)^2(\kappa^s\wedge\tilde\kappa^s)}\right\},$$

with $r_0^s=r^s$ if $\kappa^s=\kappa^s\wedge\tilde\kappa^s$, and $r_0^s=\tilde r^s$ if $\tilde\kappa^s=\kappa^s\wedge\tilde\kappa^s$. Thus, if $\frac53\le p\le 2$, then for every $\alpha>0$ we obtain the following estimate, valid for every $R\ge 1$:

$$\|R^*_{S^s,\tilde S^s}\|_{L^2\times L^2\to L^p(Q^s_{S^s,\tilde S^s}(R))}\le (\kappa^s\tilde\kappa^s)^{\frac12-\frac1p}(D^s)^{3-\frac5p}\log^{\gamma_\alpha}(C_0^s)C_\alpha R^\alpha,$$

with (compare to (2-20))

$$C_0^s=\frac{\bar d_1^{s\,2}\bar d_2^{s\,2}}{(D^s)^4}(D^s[\kappa^s\wedge\tilde\kappa^s])^{-\frac1p}(D^s\kappa^sD^s\tilde\kappa^s)^{-\frac12}.$$

Recall here that $R^*_{S,\tilde S}(f_1,f_2)=R^*_{\mathbb{R}^2}f_1\cdot R^*_{\mathbb{R}^2}f_2$, if $f_1\in L^2(U)$, $f_2\in L^2(\widetilde U)$. Scaling back by means of Lemma 3.1, we obtain

$$\begin{aligned}\|R^*_{S,\tilde S}\|_{L^2\times L^2\to L^p(Q_{S,\tilde S}(R))}&\le (a_1a_2)^{1-\frac2p}(\kappa^s\tilde\kappa^s)^{\frac12-\frac1p}(D^s)^{3-\frac5p}\log^{\gamma_\alpha}(C_0^s)C_\alpha R^\alpha\\ &=(a_1a_2\kappa^s\cdot a_1a_2\tilde\kappa^s)^{\frac12-\frac1p}(D^s)^{3-\frac5p}\log^{\gamma_\alpha}(C_0^s)C_\alpha R^\alpha,\end{aligned} \tag{3-7}$$

where

$$\begin{aligned}Q_{S,\tilde S}(R)&=\left\{x\in\mathbb{R}^3:\ |a_ix_i+\partial_i\phi^s(r_0^s)a_1a_2x_3|\le\frac{R^2}{(D^s)^2\bar\kappa^s},\ i=1,2,\ |a_1a_2x_3|\le\frac{R^2}{(D^s)^2\kappa^s\wedge\tilde\kappa^s}\right\}\\ &=\left\{x\in\mathbb{R}^3:\ |x_i+\partial_i\phi(r_0)x_3|\le\frac{R^2}{a_i(D^s)^2\bar\kappa^s},\ i=1,2,\ |x_3|\le\frac{R^2}{a_1a_2(D^s)^2\kappa^s\wedge\tilde\kappa^s}\right\}.\end{aligned}$$

But, by (3-4), we have

$$\bar\kappa^s=\bar\kappa_1^s\vee\bar\kappa_2^s=\frac{a_1}{a_2}\bar\kappa_1\vee\frac{a_2}{a_1}\bar\kappa_2=\frac{a_1}{\bar d_1}\vee\frac{a_2}{\bar d_2}=\frac{\bar\kappa_2\bar d_2^2\vee\bar\kappa_1\bar d_1^2}{\bar d_1\bar d_2} \tag{3-8}$$

and

$$D^s=\min\{d_1^s,d_2^s,\tilde d_1^s,\tilde d_2^s\}\le\min\left\{\frac{\bar d_1}{a_1},\frac{\bar d_2}{a_2}\right\}=(\bar\kappa^s)^{-1};$$

hence

$$a_i(D^s)^2\bar\kappa^s\le a_iD^s\le\bar d_i,\quad i=1,2,$$

and also

$$a_1a_2(D^s)^2(\kappa^s\wedge\tilde\kappa^s)\le D^sa_1a_2\le a_2\bar d_1\wedge a_1\bar d_2=\bar\kappa_1\bar d_1^2\wedge\bar\kappa_2\bar d_2^2.$$

These estimates imply that
$$Q_{S,\tilde{S}}(R) \supset Q^1_{S,\tilde{S}}(R), \tag{3-9}$$
if we put
$$Q^1_{S,\tilde{S}}(R) = \left\{ x \in \mathbb{R}^3 : \; |x_i + \partial_i \phi(r_0) x_3| \le \frac{R^2}{\bar{d}_i},\; i = 1, 2,\; |x_3| \le \frac{R^2}{\bar{\kappa}_1 \bar{d}_1^2 \wedge \bar{\kappa}_2 \bar{d}_2^2} \right\}.$$
Moreover, by (3-4) we have
$$d_i^s = \frac{d_i}{a_i} = \frac{\bar{\kappa}_i \bar{d}_i}{a_1 a_2} d_i,$$
and
$$\min\{d_i^s, \tilde{d}_i^s\} = \frac{\bar{\kappa}_i}{a_1 a_2} \bar{d}_i \min\{d_i, \tilde{d}_i\} = \frac{\bar{\kappa}_i d_i \tilde{d}_i}{a_1 a_2}.$$
Furthermore,
$$a_1 a_2 \kappa^s \sim a_1 a_2 \left(\frac{a_2}{a_1} \kappa_2 + \frac{a_1}{a_2} \kappa_1 \right) = (\bar{\kappa}_1^2 \bar{d}_1^2 \kappa_2 + \bar{\kappa}_2^2 \bar{d}_2^2 \kappa_1) = \bar{\kappa}_1 \bar{\kappa}_2 \left(\bar{\kappa}_1 \bar{d}_1^2 \frac{\kappa_2}{\bar{\kappa}_2} + \bar{\kappa}_2 \bar{d}_2^2 \frac{\kappa_1}{\bar{\kappa}_1} \right). \tag{3-10}$$
Thus the product of the first two factors on the right-hand side of (3-7) can be rewritten as
$$\begin{aligned}
&(a_1 a_2 \kappa^s \cdot a_1 a_2 \tilde{\kappa}^s)^{\frac{1}{2} - \frac{1}{p}} (D^s)^{3 - \frac{5}{p}} \\
&\quad = (a_1 a_2)^{\frac{5}{p} - 3} (\bar{\kappa}_1 \bar{\kappa}_2)^{1 - \frac{2}{p}} \left(\bar{\kappa}_1 \bar{d}_1^2 \frac{\kappa_2}{\bar{\kappa}_2} + \bar{\kappa}_2 \bar{d}_2^2 \frac{\kappa_1}{\bar{\kappa}_1} \right)^{\frac{1}{2} - \frac{1}{p}} \left(\bar{\kappa}_1 \bar{d}_1^2 \frac{\tilde{\kappa}_2}{\bar{\kappa}_2} + \bar{\kappa}_2 \bar{d}_2^2 \frac{\tilde{\kappa}_1}{\bar{\kappa}_1} \right)^{\frac{1}{2} - \frac{1}{p}} \min_i (\bar{\kappa}_i d_i \tilde{d}_i)^{3 - \frac{5}{p}} \\
&\quad = (\bar{\kappa}_1 \bar{d}_1 \bar{\kappa}_2 \bar{d}_2)^{\frac{5}{p} - 3} (\bar{\kappa}_1 \bar{\kappa}_2)^{1 - \frac{2}{p}} \left(\bar{\kappa}_1 \bar{d}_1^2 \frac{\kappa_2}{\bar{\kappa}_2} + \bar{\kappa}_2 \bar{d}_2^2 \frac{\kappa_1}{\bar{\kappa}_1} \right)^{\frac{1}{2} - \frac{1}{p}} \left(\bar{\kappa}_1 \bar{d}_1^2 \frac{\tilde{\kappa}_2}{\bar{\kappa}_2} + \bar{\kappa}_2 \bar{d}_2^2 \frac{\tilde{\kappa}_1}{\bar{\kappa}_1} \right)^{\frac{1}{2} - \frac{1}{p}} \min_i (\bar{\kappa}_i d_i \tilde{d}_i)^{3 - \frac{5}{p}} \\
&\quad = (\bar{\kappa}_1 \bar{\kappa}_2)^{\frac{3}{p} - 2} (\bar{d}_1 \bar{d}_2)^{\frac{5}{p} - 3} \left(\bar{\kappa}_1 \bar{d}_1^2 \frac{\kappa_2}{\bar{\kappa}_2} + \bar{\kappa}_2 \bar{d}_2^2 \frac{\kappa_1}{\bar{\kappa}_1} \right)^{\frac{1}{2} - \frac{1}{p}} \left(\bar{\kappa}_1 \bar{d}_1^2 \frac{\tilde{\kappa}_2}{\bar{\kappa}_2} + \bar{\kappa}_2 \bar{d}_2^2 \frac{\tilde{\kappa}_1}{\bar{\kappa}_1} \right)^{\frac{1}{2} - \frac{1}{p}} \min_i (\bar{\kappa}_i d_i \tilde{d}_i)^{3 - \frac{5}{p}}.
\end{aligned}$$

For $a, b \in (0, \infty)$ write
$$q(a, b) = \frac{a \vee b}{a \wedge b} = \frac{a}{b} \vee \frac{b}{a} \ge 1.$$
A lower bound for D^s is
$$D^s = \frac{d_1 \wedge \tilde{d}_1}{a_1} \wedge \frac{d_2 \wedge \tilde{d}_2}{a_2} \ge \left(\frac{d_1 \wedge \tilde{d}_1}{\bar{d}_1} \wedge \frac{d_2 \wedge \tilde{d}_2}{\bar{d}_2} \right) \left(\frac{\bar{d}_1}{a_1} \wedge \frac{\bar{d}_2}{a_2} \right) \ge \frac{1}{q(d_1, \tilde{d}_1) q(d_2, \tilde{d}_2)} \frac{1}{\bar{\kappa}^s}, \tag{3-11}$$
where we have used (3-8) in the last inequality. And, from formula (3-10) we can deduce
$$\kappa^s \gtrsim \frac{\bar{\kappa}_1 \bar{\kappa}_2}{a_1 a_2} (\bar{\kappa}_1 \bar{d}_1^2 \vee \bar{\kappa}_2 \bar{d}_2^2) \frac{\kappa_1}{\bar{\kappa}_1} \frac{\kappa_2}{\bar{\kappa}_2} \ge \frac{\bar{\kappa}_1 \bar{d}_1^2 \vee \bar{\kappa}_2 \bar{d}_2^2}{\bar{d}_1 \bar{d}_2} \frac{1}{q(\kappa_1, \bar{\kappa}_1) q(\kappa_2, \bar{\kappa}_2)} = \frac{\bar{\kappa}^s}{q(\kappa_1, \bar{\kappa}_1) q(\kappa_2, \bar{\kappa}_2)}, \tag{3-12}$$
where we have again applied (3-8) in the last step. Combining (3-11) and (3-12), we obtain
$$(D^s \kappa^s)^{-1} \lesssim \prod_{i=1,2} q(\kappa_i, \tilde{\kappa}_i) q(d_i, \tilde{d}_i), \tag{3-13}$$

and then by symmetry also

$$(D^s\tilde{\kappa}^s)^{-1} \lesssim \prod_{i=1,2} q(\kappa_i,\tilde{\kappa}_i)q(d_i,\tilde{d}_i).$$

We may now estimate the constant C_0^s in the following way, using (3-13) in the first inequality, (3-11) in the second one and (3-8) in the third one (being generous in the exponents, since C_0^s appears only logarithmically):

$$\begin{aligned}
C_0^s &= \frac{\bar{d}_1^{s\,2}\bar{d}_2^{s\,2}}{(D^s)^4}(D^s[\kappa^s\wedge\tilde{\kappa}^s])^{-\frac{1}{p}}(D^s\kappa^s D^s\tilde{\kappa}^s)^{-\frac{1}{2}} \le \frac{\bar{d}_1^{s\,2}\bar{d}_2^{s\,2}}{(D^s)^4}\Big(\prod_{i=1,2} q(\kappa_i,\tilde{\kappa}_i)q(d_i,\tilde{d}_i)\Big)^{\frac{1}{p}+1} \\
&\le \Big(\prod_{i=1,2} q(\kappa_i,\tilde{\kappa}_i)q(d_i,\tilde{d}_i)\Big)^{\frac{1}{p}+5}(\bar{d}_1^s\bar{d}_2^s)^2(\bar{\kappa^s})^4 \\
&\le \Big(\prod_{i=1,2} q(\kappa_i,\tilde{\kappa}_i)q(d_i,\tilde{d}_i)\Big)^{\frac{1}{p}+5}\Big(\frac{\bar{d}_1\bar{d}_2}{a_1a_2}\Big)^2\Big(\frac{\bar{\kappa}_1\bar{d}_1^2\vee\bar{\kappa}_2\bar{d}_2^2}{\bar{d}_1\bar{d}_2}\Big)^4 \\
&= \Big(\prod_{i=1,2} q(\kappa_i,\tilde{\kappa}_i)q(d_i,\tilde{d}_i)\Big)^{\frac{1}{p}+5}\Big(\frac{(\bar{\kappa}_1\bar{d}_1^2\vee\bar{\kappa}_2\bar{d}_2^2)^2}{\bar{\kappa}_1\bar{d}_1^2\bar{\kappa}_2\bar{d}_2^2}\Big)^2 \\
&= \Big(\prod_{i=1,2} q(\kappa_i,\tilde{\kappa}_i)q(d_i,\tilde{d}_i)\Big)^{\frac{1}{p}+5} q(\bar{\kappa}_1\bar{d}_1^2,\bar{\kappa}_2\bar{d}_2^2)^2.
\end{aligned}$$

Combining all these estimates, we finally arrive at the following.

Corollary 3.3. *Let $\frac{5}{3}\le p\le 2$. For every $\alpha>0$ there exist $C_\alpha,\gamma_\alpha>0$ such that, for every pair of patches of hypersurfaces S and $\widetilde{S}$ as described in our general assumptions at the beginning of this section and every $R>0$, we have*

$$\begin{aligned}
\|R^*_{S,\widetilde{S}}\|_{L^2(S)\times L^2(\widetilde{S})\to L^p(Q^1_{S,\widetilde{S}}(R))} &\le C_\alpha R^\alpha(\bar{\kappa}_1\bar{\kappa}_2)^{\frac{3}{p}-2}(\bar{d}_1\bar{d}_2)^{\frac{5}{p}-3}\min_i(\bar{\kappa}_i d_i\tilde{d}_i)^{3-\frac{5}{p}} \\
&\quad\times\Big(\bar{\kappa}_1\bar{d}_1^2\frac{\kappa_2}{\bar{\kappa}_2}\vee\bar{\kappa}_2\bar{d}_2^2\frac{\kappa_1}{\bar{\kappa}_1}\Big)^{\frac{1}{2}-\frac{1}{p}}\Big(\bar{\kappa}_1\bar{d}_1^2\frac{\tilde{\kappa}_2}{\bar{\kappa}_2}\vee\bar{\kappa}_2\bar{d}_2^2\frac{\tilde{\kappa}_1}{\bar{\kappa}_1}\Big)^{\frac{1}{2}-\frac{1}{p}} \\
&\quad\times\Big[1+\log^{\gamma_\alpha}\Big(q(\bar{\kappa}_1\bar{d}_1^2,\bar{\kappa}_2\bar{d}_2^2)\prod_{i=1,2} q(d_i,\tilde{d}_i)q(\kappa_i,\tilde{\kappa}_i)\Big)\Big], \qquad (3\text{-}14)
\end{aligned}$$

where, in correspondence with our Convention 1.4, we have put $R^*_{S,\widetilde{S}}(f_1,f_2)=R^*_{\mathbb{R}^2}f_1\cdot R^*_{\mathbb{R}^2}f_2$, $f_1\in L^2(S)$, $f_2\in L^2(\widetilde{S})$.

4. Globalization and ε-removal

4A. ***General results.*** The next task will be to extend our inequalities (3-14) from the cuboids $Q^1_{S,\widetilde{S}}(R)$ to the whole space, and to get rid of the factor R^α. There is a certain amount of "globalization" or "ε-removal" technique available for this purpose, in particular Lemma 2.4 by Tao and Vargas [2000a], which in return follows ideas from [Bourgain 1995b]. We shall need to adapt those techniques to our setting, in which it will be important to understand more precisely how the corresponding estimates will depend on the parameters κ_j and d_j, $j = 1, 2$.

To this end, let us consider two hypersurfaces S_1 and S_2 in $\mathbb{R}^{d+1}$, defined as graphs $S_j = \{(x, \phi_j(x)) : x \in U_j\}$, and assume there is a constant A such that

$$|\nabla\phi_j(x)| \le A \tag{4-1}$$

for all $x \in U_j$, $j = 1, 2$. We will consider the measures ν_j defined on S_j by

$$\int_{S_j} g\, d\nu_j = \int_{U_j} f(x, \phi_j(x))\, dx.$$

Note that, under the assumption (4-1), these measures are equivalent to the surface measures on S_1 and S_2. We write again

$$R^*_{S_1,S_2}(f_1, f_2) = R^*_{\mathbb{R}^d} f_1\, R^*_{\mathbb{R}^d} f_2.$$

Denote by $B(0, R) = \{x \in \mathbb{R}^{d+1} : |x| \le R\}$ the ball of radius R. Our main result in this section is the following.

Lemma 4.1. *Let $C_1, C_2, \alpha, s > 0$, $R_0 \ge 1$, $1 \le p_0 < p \le \infty$, and let S_1, S_2 be hypersurfaces with ν_1, ν_2, respectively, satisfying (4-1), and let μ be a positive Borel measure on $\mathbb{R}^{d+1}$. Assume that for all $R \ge R_0$ and all $f_j \in L^2(S_j, \nu_j)$, $j = 1, 2$,*

(i) $\|R^*_{S_1,S_2}(f_1, f_2)\|_{L^{p_0}(B(0,R),\mu)} \le C_1 R^\alpha \|f_1\|_{L^2(S_1,\nu_1)} \|f_2\|_{L^2(S_2,\nu_2)}$,

(ii) $|\widehat{d\nu_j}(x)| \le C_2(1 + |x|)^{-s}$ *for all* $x \in \mathbb{R}^{d+1}$,

and that $(1 + 2\alpha/s)/p < 1/p_0$. Then

$$\|R^*_{S_1,S_2}(f_1, f_2)\|_{L^p(\mathbb{R}^{d+1},\mu)} \le C' \|f_1\|_{L^2(S_1,\nu_1)} \|f_2\|_{L^2(S_2,\nu_2)} \tag{4-2}$$

for all $f_j \in L^2(S_j, \nu_j)$, $j = 1, 2$, where C' only depends on $C_1, C_2, R_0, \alpha, s, p, p_0$.

Proof. We shall follow the proof of Lemma 2.4 in [Tao and Vargas 2000a] and only briefly sketch the main arguments, indicating those changes in the proof that will be needed in our setting. The main difference with [Tao and Vargas 2000a] is that instead of a Stein–Tomas-type estimate, we will use the trivial bound

$$\|R^*_{\mathbb{R}^d} f_j\|_{L^\infty(\mathbb{R}^{d+1},\mu)} \le \|f_j\|_{L^1(\nu_j)} \le \|f_j\|^{\frac12}_{L^2(S_j,\nu_j)} |\widehat{d\nu_j}(0)|^{\frac12} \le C_2^{\frac12} \|f_j\|^{\frac12}_{L^2(S_j,\nu_j)}, \tag{4-3}$$

where we have used our hypothesis (ii).

By (4-3) and interpolation, it then suffices to prove a weak-type estimate of the form

$$\mu(E_\lambda) \lesssim \lambda^{-p}, \quad \lambda > 0, \tag{4-4}$$

assuming $\|f_j\|_{L^2(\nu_j)} = 1$, $j = 1, 2$. Here, $E_\lambda = \{\mathrm{Re}(R^*_{\mathbb{R}^d} f_1\, R^*_{\mathbb{R}^d} f_2) > \lambda\}$. Given $\lambda > 0$, let us abbreviate $E = E_\lambda$. We may also assume $\mu(E) \gtrsim 1$. Chebyshev's inequality implies

$$\lambda\mu(E) \lesssim \|\chi_E R^*_{\mathbb{R}^d} f_1\, R^*_{\mathbb{R}^d} f_2\|_{L^1(\mu)},$$

and thus it suffices to show

$$\|\chi_E R^*_{\mathbb{R}^d} g_1 \, R^*_{\mathbb{R}^d} g_2\|_{L^1(\mu)} \lesssim \mu(E)^{\frac{1}{p'}} \|g_1\|_{L^2(\nu_1)} \|g_2\|_{L^2(\nu_1)} \tag{4-5}$$

for arbitrary L^2-functions g_1 and g_2 (which are completely independent of f_1 and f_2).

To this end, fix g_2 with $\|g_2\|_{L^2(\nu_2)} \sim 1$, and define $T = T_{E,g_2}$ as the linear operator

$$Tg_1 = \chi_E R^*_{\mathbb{R}^d} g_1 \, R^*_{\mathbb{R}^d} g_2.$$

Then, (4-5) is equivalent to the inequality

$$\|Tg_1\|_{L^1(\mu)} \lesssim \mu(E)^{\frac{1}{p'}} \|g_1\|_{L^2(\nu_1)}.$$

By duality, it suffices to show

$$\|T^* F\|_{L^2(d\nu_1)} \lesssim \mu(E)^{\frac{1}{p'}} \|F\|_{L^\infty(\mu)},$$

where T^* is (essentially) the adjoint operator

$$T^* F = \mathcal{F}^{-1}(\chi_E R^*_{\mathbb{R}^d} g_2 F \mu),$$

and $\mathcal{F}^{-1}$ is the inverse Fourier transform. We may assume $\|F\|_{L^\infty}(\mu) \lesssim 1$.

By squaring this and applying Plancherel's theorem, we reduce ourselves to showing

$$|\langle \widetilde{F}\, d\mu * \widehat{d\nu_1}, \widetilde{F}\, d\mu\rangle| \lesssim \mu(E)^{\frac{2}{p'}}, \tag{4-6}$$

where $\widetilde{F} = \chi_E (R^*_{\mathbb{R}^d} g_2) F$. Note that the hypotheses on F and g_2 and inequality (4-3) imply

$$\|\widetilde{F}\|_{L^1(\mu)} = \|\chi_E (R^*_{\mathbb{R}^d} g_2) F\|_{L^1(\mu)} \le \|\chi_E\|_{L^1(\mu)} \|R^*_{\mathbb{R}^d} g_2\|_{L^\infty(\mu)} \|F\|_{L^\infty(\mu)} \lesssim \mu(E). \tag{4-7}$$

From this point on, we follow the proof of [Tao and Vargas 2000a] with the obvious changes. Let $R > 1$ be a quantity to be chosen later. Let ϕ be a bump function which equals 1 for $|x| \lesssim 1$ and vanishes for $|x| \gg 1$, and write $d\nu_1 = d\nu_1^R + d\nu_{1\,R}$, where

$$\widehat{d\nu_{1\,R}}(x) = \phi\left(\frac{x}{R}\right) \widehat{d\nu_1}(x). \tag{4-8}$$

From hypothesis (ii) we have

$$\|\widehat{d\nu_1^R}\|_\infty \lesssim R^{-s},$$

and so by (4-7) we have

$$|\langle \widetilde{F}\, d\mu * \widehat{d\nu_1^R}, \widetilde{F}\, d\mu\rangle| \lesssim R^{-s} \mu(E)^2.$$

We now choose R to be

$$R = \mu(E)^{\frac{2}{sp}}, \tag{4-9}$$

so that the contribution of $d\nu_1^R$ to (4-6) is acceptable. Thus (4-6) reduces to

$$|\langle \widetilde{F}\, d\mu * \widehat{d\nu_{1\,R}}, \widetilde{F}\, d\mu\rangle| \lesssim \mu(E)^{\frac{2}{p'}}.$$

Following the arguments in [Tao and Vargas 2000a] and skipping details, we may then reduce the problem to proving

$$\|\chi_E \hat{\tilde{g}}_1 \hat{\tilde{g}}_2\|_{L^1(\mu)} \lesssim R^{-\frac{1}{2}} R^{-\frac{1}{2}} \mu(E)^{\frac{1}{p'}} \|\tilde{g}_1\|_2 \|\tilde{g}_2\|_2,$$

where $\tilde{g}_i$ is an arbitrary function on the $1/R$ neighborhood of $S_{i,R}$ for $i = 1, 2$. By Hölder's inequality it suffices to show

$$\|\hat{\tilde{g}}_1 \hat{\tilde{g}}_2\|_{L^{p_0}(\mu)} \lesssim \mu(E)^{-\frac{1}{p_0'}} R^{-\frac{1}{2}} R^{-\frac{1}{2}} \mu(E)^{\frac{1}{p'}} \|\tilde{g}_1\|_2 \|\tilde{g}_2\|_2. \tag{4-10}$$

Moreover, using the first hypothesis of the lemma, we obtain

$$\|\hat{\tilde{g}}_1 \hat{\tilde{g}}_2\|_{L^{p_0}(\mu)} \lesssim R^{\alpha-1} \|\hat{\tilde{g}}_1\|_2 \|\hat{\tilde{g}}_2\|_2.$$

Comparing this with (4-10), we see that we will be done if

$$R^{\alpha} \lesssim \mu(E)^{-\frac{1}{p_0'}} \mu(E)^{\frac{1}{p'}} = \mu(E)^{\frac{1}{p_0} - \frac{1}{p}}.$$

But this follows from (4-9) and the assumption $(1 + 2\alpha/s)/p < 1/p_0$. □

4B. ***Application to the setting of Section 3.*** Let us now come back to the situation described by our General Assumptions in Lemma 3.2; i.e., we are interested in pairs of surfaces $S = \operatorname{graph}(\phi|_U)$, $U = r + [0, d_1] \times [0, d_2]$, with principal curvatures on S comparable to $\kappa_i = r_i^{m_i - 2}$, $r_i \geq d_i$, and $\widetilde{S} = \operatorname{graph}(\phi|_{\widetilde{U}})$, with corresponding quantities $\tilde{r}_i$, $\tilde{d}_i$, $\tilde{\kappa}_i$, $\tilde{\kappa}$.

Recall also the notation defined in (1-3), (3-1), and assume the conditions (3-2) and (3-3) are satisfied.

We consider the measure ν_S supported on S given by

$$\int_S f \, d\nu_S := \int_U f(x_1, x_2, \phi(x_1, x_2)) \, dx_1 \, dx_2,$$

and define $\nu_{\widetilde{S}}$ on $\widetilde{S}$ analogously.

4B1. *Decay of the Fourier transform.*

Lemma 4.2. *Let $s = 1/(m_1 \vee m_2)$. For any $r^0 \in U \cup \widetilde{U}$ we then have the uniform estimate for $x \in \mathbb{R}^3$*

$$\begin{aligned} &|\widehat{d\nu_S}(x)| + |\widehat{d\nu_{\widetilde{S}}}(x)| \\ &\quad \leq C_s \bar{d}_1 \bar{d}_2 \big(1 + |\bar{d}_1(x_1 + \partial_1\phi(r^0)x_3)| + |\bar{d}_2(x_2 + \partial_2\phi(r^0)x_3)| + |(\bar{\kappa}_1 \bar{d}_1^2 \vee \bar{\kappa}_2 \bar{d}_2^2)x_3|\big)^{-s}. \end{aligned} \tag{4-11}$$

Proof. We only consider $\nu = \nu_S$, since the proof for $\nu_{\widetilde{S}}$ is analogous. Recall that ϕ splits into $\phi(x) = \phi_{(1)}(x_1) + \phi_{(2)}(x_2)$, so that

$$|\widehat{d\nu}(x)| = \left| \int_{r_1}^{r_1+d_1} e^{-i(x_1\xi_1 + x_3\phi_{(1)}(\xi_1))} \, d\xi_1 \int_{r_2}^{r_2+d_2} e^{-i(x_2\xi_2 + x_3\phi_{(2)}(\xi_2))} \, d\xi_2 \right|.$$

Next, for $i \in \{1,2\}$, we have

$$\begin{aligned} I_i &= \left| \int_{r_i}^{r_i+d_i} e^{-i(x_i\xi_i + x_3\phi_{(i)}(\xi_i))}\, d\xi_i \right| = \left| \int_0^{d_i} e^{-i(x_i(r_i+y_i)+x_3\phi_{(i)}(r_i+y_i))}\, dy_i \right| \\ &= \left| \int_0^{d_i} e^{-i\left((x_i+\phi'_{(i)}(r_i)x_3)y_i + x_3(\phi_{(i)}(r_i+y_i)-\phi_{(i)}(r_i)-\phi'_{(i)}(r_i)y_i)\right)}\, dy_i \right| \\ &= d_i \left| \int_0^1 e^{-i((x_i+\phi'_{(i)}(r_i)x_3)d_iy_i + x_3\kappa_i d_i^2 \Psi_i(d_iy_i))}\, dy_i \right|, \end{aligned}$$

where $\Psi_i(y_i) = \left(\phi_{(i)}(r_i+d_1y_i) - \phi_{(i)}(r_i) - \phi'_{(i)}(r_i)d_iy_i\right)/(\kappa_i d_i^2)$, so that in particular

$$\left| \frac{d}{dy_i}\Psi_i(y_i) \right| = \left| \frac{\phi'_{(i)}(r_i+d_iy_i) - \phi'_{(i)}(r_i)}{\kappa_i d_i^2} d_i \right| \lesssim \frac{\kappa_i d_i}{\kappa_i d_i^2} d_i \sim 1, \qquad \frac{d^2}{dy_i^2}\Psi_i(y_i) = \frac{\phi''_{(i)}(r_i+d_iy_i)}{\kappa_i d_i^2} d_i^2 \sim 1.$$

Therefore, by either applying van der Corput's lemma of order 2, or by integrating by parts (if $|d_i(x_i+\phi'_{(i)}(r_i)x_3)| \gg |\kappa_i d_i^2 x_3|$,) we obtain

$$I_i \lesssim d_i(1 + |d_i(x_i+\phi'_{(i)}(r_i)x_3)| + |\kappa_i d_i^2 x_3|)^{-\frac{1}{2}}. \tag{4-12}$$

We next claim that the distortion $d_i/\bar{d}_i$ in the side lengths is bounded by the distortion in the size of the space variable r_i, i.e.,

$$\frac{d_i}{\bar{d}_i} \lesssim \frac{r_i}{\bar{r}_i}. \tag{4-13}$$

If $r_i \sim \bar{r}_i$, the statement is obvious, so assume $r_i \ll \bar{r}_i$. Then $\tilde{r}_i = \bar{r}_i$, and furthermore by our assumptions we have $d_i \le r_i$ and $\bar{r}_i \sim |r_i - \tilde{r}_i| \lesssim \bar{d}_i$ (compare to the separation condition (3-3)). Thus (4-13) follows also in this case. As $\kappa_i = r_i^{m_i-2}$, we conclude from (4-13) that

$$\frac{\kappa_i d_i^2}{\bar{\kappa}_i \bar{d}_i^2} \gtrsim \left(\frac{d_i}{\bar{d}_i}\right)^{m_i}. \tag{4-14}$$

In combination, the estimates (4-13) and (4-14) imply

$$1 + |d_i(x_i+\phi'_{(i)}(r_i)x_3)| + |\kappa_i d_i^2 x_3| \gtrsim \left(\frac{d_i}{\bar{d}_i}\right)^{m_i} \left(1 + |\bar{d}_i(x_i+\phi'_{(i)}(r_i)x_3)| + |\bar{\kappa}_i \bar{d}_i^2 x_3|\right).$$

Since we may replace the exponent $-\frac{1}{2}$ in the right-hand side of (4-12) by $-1/m_i$, we now see that we may estimate

$$I_i \lesssim \bar{d}_i(1 + |\bar{d}_i(x_i+\phi'_{(i)}(r_i)x_3)| + |\bar{\kappa}_i \bar{d}_i^2 x_3|)^{-\frac{1}{m_i}}. \tag{4-15}$$

Finally, in order to pass from the point r to an arbitrary point $r^0 \in U \cup \widetilde{U}$ in these estimates, observe that by (3-3) we have $|r_i - r_i^0| \le |r_i - \tilde{r}_i| + \bar{d}_i \sim \bar{d}_i$, and hence

$$\bar{d}_i |\phi'_{(i)}(r_i) - \phi'_{(i)}(r_i^0)| \le \bar{\kappa}_i |r_i - r_i^0| \bar{d}_i \lesssim \bar{\kappa}_i \bar{d}_i^2,$$

since $|\phi''_{(i)}| \lesssim \bar\kappa_i$ on $[r_i, r_i + d_i] \cup [\tilde r_i, \tilde r_i + \tilde d_i]$. Therefore (4-15) implies that also

$$I_i \lesssim \bar d_i (1 + |\bar d_i (x_i + \partial_i \phi(r^0) x_3)| + |\bar\kappa_i \bar d_i^2 x_3|)^{-\frac{1}{m_i}}.$$

The estimate (4-11) is now immediate. □

4B2. *Linear change of variables and verification of the assumptions of Lemma 4.1.* In view of Lemma 4.2, let us fix $r^0 \in U \cup \widetilde U$, and define the linear transformation $T = T_{S,\widetilde S}$ of $\mathbb{R}^3$ by

$$T(x) = \big(\bar d_1 (x_1 + \partial_1 \phi(r^0) x_3), \bar d_2 (x_2 + \partial_2 \phi(r^0) x_3), (\bar\kappa_1 \bar d_1^2 \vee \bar\kappa_2 \bar d_2^2) x_3\big).$$

Then estimate (4-11) reads

$$|\widehat{d\nu_S}(x)| + |\widehat{d\nu_{\widetilde S}}(x)| \le C_s \bar d_1 \bar d_2 (1 + |T(x)|)^{-s}.$$

Therefore, in order to apply Lemma 4.1, we will consider the rescaled surfaces

$$S_1 = (T^t)^{-1} S \quad \text{and} \quad S_2 = (T^t)^{-1} \widetilde S. \tag{4-16}$$

Then we find that

$$\begin{aligned} S_1 &= \{(T^t)^{-1}(x_1, x_2, \phi_{(1)}(x_1) + \phi_{(2)}(x_2)) : (x_1, x_2) \in U\} \\ &= \left\{\left(\frac{x_1}{\bar d_1}, \frac{x_2}{\bar d_2}, \frac{1}{\bar\kappa_1 \bar d_1^2 \vee \bar\kappa_2 \bar d_2^2}\big(-\partial_1\phi(r_0) x_1 - \partial_2 \phi(r_0) x_2 + \phi_{(1)}(x_1) + \phi_{(2)}(x_2)\big)\right) : (x_1, x_2) \in U\right\} \\ &= \{(y_1, y_2, \psi(y_1, y_2)) : (y_1, y_2) \in U_1\}, \end{aligned}$$

where $U_1 = \{(y_1, y_2) = (x_1/\bar d_1, x_2/\bar d_2) : (x_1, x_2) \in U\}$ is a square of side length ≤ 1 and

$$\psi(y_1, y_2) = \frac{1}{\bar\kappa_1 \bar d_1^2 \vee \bar\kappa_2 \bar d_2^2}\big(-\bar d_1 \partial_1 \phi(r_0) y_1 - \bar d_2 \partial_2 \phi(r_0) y_2 + \phi_{(1)}(\bar d_1 y_1) + \phi_{(2)}(\bar d_2 y_2)\big).$$

We have a similar expression for S_2.

In S_1 we consider the measure $d\nu_1$ defined by

$$\int_{S_1} g\, d\nu_1 = \frac{1}{\bar d_1 \bar d_2} \int_S g((T^t)^{-1} x)\, d\nu_S(x).$$

By our definition of dν and ψ, this may be rewritten as

$$\begin{aligned} \int_{S_1} g\, d\nu_1 &= \frac{1}{\bar d_1 \bar d_2} \int_U g\big((T^t)^{-1}(x_1, x_2, \phi(x_1, x_2))\big)\, dx_1\, dx_2 \\ &= \frac{1}{\bar d_1 \bar d_2} \int_U g\left(\frac{x_1}{\bar d_1}, \frac{x_2}{\bar d_2}, \psi\left(\frac{x_1}{\bar d_1}, \frac{x_2}{\bar d_2}\right)\right) dx_1\, dx_2 = \int_{U_1} g(y_1, y_2, \psi(y_1, y_2))\, dy_1\, dy_2. \end{aligned}$$

Moreover, we have

$$\widehat{g\, d\nu_1}(\xi) = \frac{1}{\bar d_1 \bar d_2}\, \widehat{\big(g \circ (T^t)^{-1} d\nu_S\big)}\, (T^{-1}\xi), \tag{4-17}$$

and therefore

$$|\widehat{d\nu_1}(x)| \le C_s(1+|x|)^{-s}.$$

We have a similar estimate for $\widehat{d\nu_2}$. Thus, the hypothesis (ii) in Lemma 4.1 is satisfied. To check that condition (4-1) is satisfied for S_1 and S_2 too, we compute

$$\left|\frac{\partial\psi}{\partial y_1}\right| = \frac{1}{\bar\kappa_1\bar d_1^2 \vee \bar\kappa_2\bar d_2^2}\left|-\bar d_1\partial_1\phi(r_0)+\bar d_1\phi'_{(1)}(\bar d_1 y_1)\right|.$$

Writing $r_0 = (\bar d_1 y_{1,0}, \bar d_2 y_{2,0})$, we see that

$$\begin{aligned}\left|\frac{\partial\psi}{\partial y_1}\right| &= \frac{1}{\bar\kappa_1\bar d_1^2 \vee \bar\kappa_2\bar d_2^2}\left|-\bar d_1\phi'_{(1)}(\bar d_1 y_{1,0})+\bar d_1\phi'_{(1)}(\bar d_1 y_1)\right| \\ &\sim \frac{1}{\bar\kappa_1\bar d_1^2 \vee \bar\kappa_2\bar d_2^2}|\bar d_1^2(y_1-y_{1,0})\phi''_{(1)}| \le \frac{\bar\kappa_1\bar d_1^2}{\bar\kappa_1\bar d_1^2 \vee \bar\kappa_2\bar d_2^2}|y_1-y_{1,0}| \le C_{m_1,m_2},\end{aligned}$$

and in a similar way we find that the derivative with respect to y_2 is bounded. Hence, hypothesis (4-1) is satisfied for ψ in place of ϕ.

What remains to be checked is condition (i) in Lemma 4.1. Observe first that our local bilinear estimate for S and $\tilde S$ in Corollary 3.3 is restricted to cuboids (see (3-9))

$$Q^1(R) = Q^1_{S,\tilde S}(R) = \left\{x\in\mathbb{R}^3 : |x_i+\partial_i\phi(r^0)x_3| \le \frac{R}{\bar d_i}, i=1,2,\ |x_3| \le \frac{R}{\bar\kappa_1\bar d_1^2 \wedge \bar\kappa_2\bar d_2^2}\right\}, \tag{4-18}$$

where r^0 is either[3] r or $\tilde r$. Obviously $T^{-1}(B(0,R)) = \{x\in\mathbb{R}^3 : |Tx|\le R\} \subset Q^1(R)$.

Define

$$\begin{aligned}A &= (\bar\kappa_1\bar\kappa_2)^{-2}(\bar d_1\bar d_2)^{-3}\min_i(\bar\kappa_i d_i\tilde d_i)^3\left(\bar\kappa_1\bar d_1^2\frac{\kappa_2}{\bar\kappa_2}\vee\bar\kappa_2\bar d_2^2\frac{\kappa_1}{\bar\kappa_1}\right)^{\frac12}\left(\bar\kappa_1\bar d_1^2\frac{\tilde\kappa_2}{\bar\kappa_2}\vee\bar\kappa_2\bar d_2^2\frac{\tilde\kappa_1}{\bar\kappa_1}\right)^{\frac12}(1+\log^{\gamma_\alpha} Q),\\ B &= (\bar\kappa_1\bar\kappa_2)^3(\bar d_1\bar d_2)^5\min_i(\bar\kappa_i d_i\tilde d_i)^{-5}\left(\bar\kappa_1\bar d_1^2\frac{\kappa_2}{\bar\kappa_2}\vee\bar\kappa_2\bar d_2^2\frac{\kappa_1}{\bar\kappa_1}\right)^{-1}\left(\bar\kappa_1\bar d_1^2\frac{\tilde\kappa_2}{\bar\kappa_2}\vee\bar\kappa_2\bar d_2^2\frac{\tilde\kappa_1}{\bar\kappa_1}\right)^{-1},\end{aligned} \tag{4-19}$$

where

$$Q = Q(S,\tilde S) = q(\bar\kappa_1\bar d_1^2, \bar\kappa_2\bar d_2^2)\prod_{i=1,2} q(d_i,\tilde d_i)q(\kappa_i,\tilde\kappa_i)$$

and $q(a,b) = (a\vee b)/a\wedge b \ge 1$ are defined to be the maximal quotient of a and b. In some sense Q is a "degeneracy quotient" that measures how much (for instance) quantities $d_i, \tilde d_i$ differ from their maximum $\bar d_i$.

Then the estimate (3-14) in Corollary 3.3, valid for $\frac53 \le p \le 2$, can be rewritten in terms of these quantities as

$$Q^1(R)\|R^*_{S,\tilde S}\|_{L^2(S)\times L^2(\tilde S)\to L^p(Q^1(R))} \le C_\alpha R^\alpha A B^{\frac1p}. \tag{4-20}$$

[3] Recall that we have some algorithm for how to choose r^0, but this will not be relevant here.

Now, in order to check hypothesis (i) in Lemma 4.1, let us choose for μ the measure on $\mathbb{R}^3$ given by

$$d\mu = \widetilde{B}^{-1}d\xi, \quad \text{where } \widetilde{B} = |\det T|\left(\frac{A}{\bar{d}_1\bar{d}_2}\right)^{p_0}B,$$

and where $d\xi$ denotes the Lebesgue measure. Notice also that (4-17) implies that, for any measurable set $E\subset\mathbb{R}^3$ and any exponent p, we have

$$\|R^*_{S_1,S_2}(f_1,f_2)\|_{L^p(E,\mu)} = \frac{A^{-\frac{p_0}{p}}B^{-\frac{1}{p}}}{(\bar{d}_1\bar{d}_2)^{2-\frac{p_0}{p}}}\big\|R^*_{S,\widetilde{S}}\big(f_1\circ(T^t)^{-1}, f_2\circ(T^t)^{-1}\big)\big\|_{L^p(T^{-1}(E),d\xi)}. \tag{4-21}$$

In particular, we obtain

$$\begin{aligned}\|R^*_{S_1,S_2}(f_1,f_2)\|_{L^{p_0}(B(0,R),\mu)} &= \frac{A^{-1}B^{-\frac{1}{p_0}}}{\bar{d}_1\bar{d}_2}\big\|R^*_{S,\widetilde{S}}\big(f_1\circ(T^t)^{-1}, f_2\circ(T^t)^{-1}\big)\big\|_{L^{p_0}(T^{-1}(B(0,R)),d\xi)}\\ &\le \frac{A^{-1}B^{-\frac{1}{p_0}}}{\bar{d}_1\bar{d}_2}\big\|R^*_{S,\widetilde{S}}\big(f_1\circ(T^t)^{-1}, f_2\circ(T^t)^{-1}\big)\big\|_{L^{p_0}(Q^1(R),d\xi)}.\end{aligned}$$

Invoking (4-20), we thus see that for $\frac{5}{3}\le p_0\le 2$ and every $\alpha>0$,

$$\begin{aligned}\|R^*_{S_1,S_2}(f_1,f_2)\|_{L^{p_0}(B(0,R),\mu)} &\le \frac{1}{\bar{d}_1\bar{d}_2}C_\alpha R^\alpha\|f_1\circ(T^t)^{-1}\|_{L^2(d\nu_S)}\|f_2\circ(T^t)^{-1}\|_{L^2(d\nu_{\widetilde{S}})}\\ &= C_\alpha R^\alpha\|f_1\|_{L^2(d\nu_1)}\|f_2\|_{L^2(d\nu_2)},\end{aligned}$$

which shows that hypothesis (i) in the Lemma 4.1 is satisfied. Applying this lemma and using again identity (4-21) and the definitions of μ, ν_1 and ν_2, we find that for any g_1 and g_2 supported in S and $\widetilde{S}$, respectively, and any p satisfying the assumptions of Lemma 4.1, we have

$$\|R^*_{S,\widetilde{S}}(g_1,g_2)\|_{L^p(d\xi)} \le C(\bar{d}_1\bar{d}_2)^{1-\frac{p_0}{p}}A^{\frac{p_0}{p}}B^{\frac{1}{p}}\|g_1\|_{L^2(d\nu_1)}\|g_2\|_{L^2(d\nu_2)}. \tag{4-22}$$

Finally, putting $\varepsilon = 1-p_0/p$, and recalling that we may choose α in Lemma 4.1 as small as we wish, then by applying Hölder's inequality in order to replace the L^2-norms on the right-hand side of (4-22) by the L^q-norms, we arrive at the following global estimate:

Theorem 4.3. *Let $\frac{5}{3}<p\le 2$, $q\ge 2$, $\varepsilon>0$. Then there exist constants $C=C_{p,\varepsilon}$ and $\gamma=\gamma_{p,\varepsilon}>0$ such that*

$$\begin{aligned}\|R^*_{S,\widetilde{S}}\|_{L^q(S)\times L^q(\widetilde{S})\to L^p(\mathbb{R}^n)} \le{}& C(\bar{\kappa}_1\bar{\kappa}_2)^{\frac{3}{p}-2+2\varepsilon}(\bar{d}_1\bar{d}_2)^{\frac{5}{p}-3+4\varepsilon}\\ &\times(d_1d_2\tilde{d}_1\tilde{d}_2)^{\frac{1}{2}-\frac{1}{q}}(1+\log^\gamma Q)\min_i(\bar{\kappa}_i d_i\tilde{d}_i)^{3-3\varepsilon-\frac{5}{p}}\\ &\times\left(\bar{\kappa}_1\bar{d}_1^2\frac{\kappa_2}{\bar{\kappa}_2}\vee\bar{\kappa}_2\bar{d}_2^2\frac{\kappa_1}{\bar{\kappa}_1}\right)^{\frac{1-\varepsilon}{2}-\frac{1}{p}}\left(\bar{\kappa}_1\bar{d}_1^2\frac{\tilde{\kappa}_2}{\bar{\kappa}_2}\vee\bar{\kappa}_2\bar{d}_2^2\frac{\tilde{\kappa}_1}{\bar{\kappa}_1}\right)^{\frac{1-\varepsilon}{2}-\frac{1}{p}}\end{aligned} \tag{4-23}$$

uniformly in S and $\widetilde{S}$, where $Q=q(\bar{\kappa}_1\bar{d}_1^2,\bar{\kappa}_2\bar{d}_2^2)\prod_{i=1,2}q\big(d_i,\tilde{d}_i\big)q(\kappa_i,\tilde{\kappa}_i)$ and $q(a,b)=(a\vee b)/a\wedge b$.

5. Dyadic summation

Recall that our hypersurface of interest is the graph of a smooth function $\phi(x_1, x_2) = \phi_{(1)}(x_1) + \phi_{(2)}(x_2)$ defined over the square $]0, 1[\times]0, 1[$. We assume ϕ to be extended continuously to the closed square $Q = [0, 1] \times [0, 1]$ (this extension will in the end not really play any role, but it will be more convenient to work with a closed square). By means of a kind of Whitney decomposition of the direct product $Q \times Q$ near the "diagonal", following some standard procedure in the bilinear approach, we can decompose $Q \times Q$ into products of congruent rectangles U and $\widetilde{U}$ of dyadic side lengths, which are "well-separated neighbors" in some sense. The next step will therefore consist in establishing bilinear estimates for pairs of subhypersurfaces supported over such pairs of neighboring rectangles. Notice that if one of these rectangles meets one of the coordinate axes, then the principal curvature in at least one coordinate direction will no longer be of a certain size, but will indeed go down to zero within this rectangle. We then perform an additional dyadic decomposition of this rectangle in order to achieve that both principal curvatures will be of a certain size on each of the dyadic subrectangles (see Figure 10). To these we can then apply our estimates from Theorem 4.3. Thus, in this section we shall work under the following:

General Assumptions. For $k_i, \tilde{k}_i, j_i \in \mathbb{N}$,

$$U = [k_1 2^{-j_1}, (k_1 + 1)2^{-j_1}] \times [k_2 2^{-j_2}, (k_2 + 1)2^{-j_2}],$$
$$\widetilde{U} = [\tilde{k}_1 2^{-j_1}, (\tilde{k}_1 + 1)2^{-j_1}] \times [\tilde{k}_2 2^{-\tilde{j}_2}, (\tilde{k}_2 + 1)2^{-\tilde{j}_2}],$$

are two congruent closed bidyadic rectangles in $[0, 1] \times [0, 1]$ whose side length and distance between them in the x_i-direction is equal to $\rho_i = 2^{-j_i}$, both for $i = 1$ and $i = 2$.

By $\varkappa_i$ we denote the maximum value of the principal curvature in the x_i-direction of both $S = \operatorname{graph}(\phi|_U)$ and $\widetilde{S} = \operatorname{graph}(\phi|_{\widetilde{U}})$.

Theorem 5.1. *Let $\frac{5}{3} < p < 2$, $q \geq 2$, $\varepsilon > 0$, and assume $(m_1 \vee m_2 + 3)\left(\frac{1}{p} - \frac{1}{2}\right) < \frac{1}{q'}$. Then we have*

$$\|R^*_{S,\widetilde{S}}\|_{L^q(S)\times L^q(\widetilde{S})\to L^p(\mathbb{R}^3)} \leq C_{p,q,\varepsilon}(\rho_1\rho_2)^{\frac{2}{q'}-\frac{1}{p}}(\varkappa_1\rho_1^2 \vee \varkappa_2\rho_2^2)^{\frac{1}{p}-1+\varepsilon}(\varkappa_1\rho_1^2 \wedge \varkappa_2\rho_2^2)^{1-\frac{2}{p}-\varepsilon}. \quad (5\text{-}1)$$

Proof. If U does not intersect with the x_i-axis, then the principal curvature in x_i-direction on U is indeed comparable to $\varkappa_i$. Otherwise we decompose U further into sets with (roughly) constant principal curvatures in order to apply the previous results. More precisely, to each dyadic interval $I = [k2^{-j}, (k + 1)2^{-j}]$, $k, j \in \mathbb{N}$, we associate a family of subsets $\{I(l)\}_{l\in\mathcal{N}_0}$ with $\bigcup_{l\in\mathcal{N}_0} I(l) = I$, according to the following two alternatives:

(i) If $k > 0$, then choose $\mathcal{N}_0 = \{0\}$ and $I(0) = I$.

(ii) If $k = 0$, then choose $\mathcal{N}_0 = \mathbb{N} = \{1, 2, 3, \dots\}$ and $I(l) = [2^{-l}(k + 1)2^{-j}, 2^{1-l}(k + 1)2^{-j}]$.

If we write $U = I_1 \times I_2$, then denote by $\{I_i(l_i)\}_{l_i\in\mathcal{N}_i}$ their associated family, and let $U(l) = I(l_1) \times I(l_2)$, $l = (l_1, l_2) \in \mathcal{N} = \mathcal{N}_1 \times \mathcal{N}_2$ and $S(l) = \operatorname{graph}(\phi|_{U(l)})$. Define $\widetilde{\mathcal{N}}$, $\widetilde{U}(l)$ and $\widetilde{S}(l)$, $l \in \widetilde{\mathcal{N}}$, in an analogous manner. Other relevant quantities are the principal curvatures on $U(l)$, i.e.,

$$\kappa_i(l_i) := 2^{-l_i(m_i-2)}\varkappa_i, \quad (5\text{-}2)$$

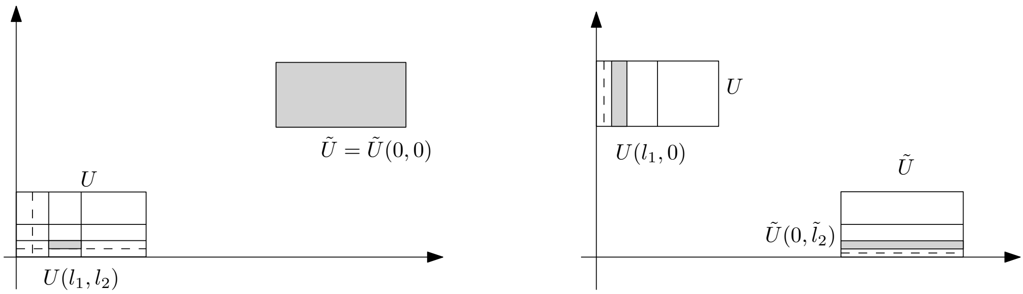


Figure 10. Two possibilities for the decomposition into subboxes.

and the side lengths of $U(l)$

$$d_i(l_i) = 2^{-l_i}\rho_i. \tag{5-3}$$

A simple but crucial observation is that since I_i and $\widetilde{I}_i$ are separated for both $i=1$ and $i=2$, we have $\mathcal{N}_i = \{0\}$ or $\widetilde{\mathcal{N}}_i = \{0\}$ (see Figure 10). Hence $l_i = 0$ or $\tilde l_i = 0$ for each pair $(l_i, \tilde l_i) \in \mathcal{N}_i \times \widetilde{\mathcal{N}}_i$, and thus

$$\bar\kappa_i(l_i,\tilde l_i) := \max\{\kappa_i(l_i), \tilde\kappa_i(\tilde l_i)\} = \max\{2^{-l_i(m_i-2)}, 2^{-\tilde l_i(m_i-2)}\}\varkappa_i = \varkappa_i, \tag{5-4}$$

$$\bar d_i(l_i,\tilde l_i) := \max\{d_i(l_i), \tilde d_i(\tilde l_i)\} = \max\{2^{-l_i}, 2^{-\tilde l_i}\}\rho_i = \rho_i. \tag{5-5}$$

We conclude that

$$\frac{\kappa_i(l_i)}{\bar\kappa_i(l_i,\tilde l_i)} = 2^{-l_i(m_i-2)}, \tag{5-6}$$

$$\frac{\tilde\kappa_i(l_i)}{\bar\kappa_i(l_i,\tilde l_i)} = 2^{-\tilde l_i(m_i-2)}\,\frac{d_i(l_i)}{\bar d_i} = 2^{-l_i}, \tag{5-7}$$

$$d_i(l_i)\tilde d_i(\tilde l_i) = 2^{-l_i-\tilde l_i}\rho_i^2 = 2^{-l_i \vee \tilde l_i}\rho_i^2. \tag{5-8}$$

Hence

$$Q = q(\bar\kappa_1\bar d_1^2, \bar\kappa_2\bar d_2^2)\prod_{i=1,2} q\big(d_i(l_i), \tilde d_i(\tilde l_i)\big)\, q\big(\kappa_i(l_i), \tilde\kappa_i(\tilde l_i)\big) \le \frac{\bar\kappa_1\bar d_1^2 \vee \bar\kappa_2\bar d_2^2}{\bar\kappa_1\bar d_1^2 \wedge \bar\kappa_2\bar d_2^2}\, 2^{m_1(l_1+\tilde l_1)+m_2(l_2+\tilde l_2)}. \tag{5-9}$$

Thus, if we apply inequality (4-23) from Theorem 4.3 to the pairs of hypersurfaces $S(l), \widetilde{S}(\tilde l)$ and estimate by means of (5-4)–(5-9), then we get

$$\begin{aligned}
&\|R^*_{S,\widetilde S}\|_{L^q\times L^q\to L^p} \\
&\quad\le \sum_{l\in\mathcal{N},\tilde l\in\widetilde{\mathcal{N}}} \|R^*_{S(l),\widetilde S(\tilde l)}\|_{L^q\times L^q\to L^p} \\
&\quad\lesssim (\varkappa_1\rho_1^2\varkappa_2\rho_2^2)^{\frac{3}{p}-2+2\varepsilon}(\rho_1\rho_2)^{\frac{2}{q'}-\frac{1}{p}} \log^\gamma\Big(\frac{\varkappa_1\rho_1^2}{\varkappa_2\rho_2^2} + \frac{\varkappa_2\rho_2^2}{\varkappa_1\rho_1^2}\Big) \\
&\qquad\times\Big(\sum_{l\in\mathcal{N},\tilde l\in\widetilde{\mathcal{N}}} [1+l_1+\tilde l_1+l_2+\tilde l_2]^\gamma \big(\varkappa_1\rho_1^2 2^{-l_1-\tilde l_1}\wedge \varkappa_2\rho_2^2 2^{-l_2-\tilde l_2}\big)^{3-3\varepsilon-\frac{5}{p}}\, 2^{-(l_1+\tilde l_1+l_2+\tilde l_2)(\frac12-\frac1q)}\Big) \\
&\qquad\times\big(\varkappa_1\rho_1^2 2^{-l_2(m_2-2)} \vee \varkappa_2\rho_2^2 2^{-l_1(m_1-2)}\big)^{\frac{1-\varepsilon}{2}-\frac1p}\big(\varkappa_1\rho_1^2 2^{-\tilde l_2(m_2-2)} \vee \varkappa_2\rho_2^2 2^{-\tilde l_1(m_1-2)}\big)^{\frac{1-\varepsilon}{2}-\frac1p}.
\end{aligned}$$

We claim

$$\begin{aligned}\sum_{l\in\mathcal{N},\tilde{l}\in\widetilde{\mathcal{N}}} &[1+l_1+\tilde{l}_1+l_2+\tilde{l}_2]^\gamma \big(\varkappa_1\rho_1^2 2^{-l_1-\tilde{l}_1}\wedge\varkappa_2\rho_2^2 2^{-l_2-\tilde{l}_2}\big)^{3-3\varepsilon-\frac{5}{p}} 2^{-(l_1+\tilde{l}_1+l_2+\tilde{l}_2)(\frac{1}{2}-\frac{1}{q})}\\ &\times\big(\varkappa_1\rho_1^2 2^{-l_2(m_2-2)}\vee\varkappa_2\rho_2^2 2^{-l_1(m_1-2)}\big)^{\frac{1-\varepsilon}{2}-\frac{1}{p}}\big(\varkappa_1\rho_1^2 2^{-\tilde{l}_2(m_2-2)}\vee\varkappa_2\rho_2^2 2^{-\tilde{l}_1(m_1-2)}\big)^{\frac{1-\varepsilon}{2}-\frac{1}{p}}\\ &\lesssim(\varkappa_1\rho_1^2\wedge\varkappa_2\rho_2^2)^{3-3\varepsilon-\frac{5}{p}}(\varkappa_1\rho_1^2\vee\varkappa_2\rho_2^2)^{1-\varepsilon-\frac{2}{p}}.\end{aligned}\tag{5-10}$$

Taking this for granted, we would arrive at estimate (5-1):

$$\begin{aligned}&\|R^*_{S,\widetilde{S}}\|_{L^2\times L^2\to L^p(Q_{S,\widetilde{S}}(R))}\\ &\quad\lesssim(\varkappa_1\rho_1^2\varkappa_2\rho_2^2)^{\frac{3}{p}-2+2\varepsilon}(\rho_1\rho_2)^{\frac{2}{q'}-\frac{1}{p}}(\varkappa_1\rho_1^2\vee\varkappa_2\rho_2^2)^{1-\varepsilon-\frac{2}{p}}(\varkappa_1\rho_1^2\wedge\varkappa_2\rho_2^2)^{3-3\varepsilon-\frac{5}{p}}\log^\gamma\Big(\frac{\varkappa_1\rho_1^2}{\varkappa_2\rho_2^2}+\frac{\varkappa_2\rho_2^2}{\varkappa_1\rho_1^2}\Big)\\ &\quad=(\rho_1\rho_2)^{\frac{2}{q'}-\frac{1}{p}}(\varkappa_1\rho_1^2\vee\varkappa_2\rho_2^2)^{\frac{1}{p}-1+\varepsilon}(\varkappa_1\rho_1^2\wedge\varkappa_2\rho_2^2)^{1-\frac{2}{p}-\varepsilon}\log^\gamma\Big(\frac{\varkappa_1\rho_1^2}{\varkappa_2\rho_2^2}+\frac{\varkappa_2\rho_2^2}{\varkappa_1\rho_1^2}\Big).\end{aligned}$$

We are thus left with the estimation of the dyadic sum in (5-10). Let

$$\mu=\frac{1}{p}-\frac{1-\varepsilon}{2}>0,\qquad \nu=3-3\varepsilon-\frac{5}{p}>0,\qquad \omega=\frac{1}{2}-\frac{1}{q}>0,\qquad c_i=m_i-2.$$

Then $c_i\mu<\nu+\omega$ is equivalent to $m_i\big(\frac{1}{p}-\frac{1}{2}\big)+\mathcal{O}(\varepsilon)<\frac{1}{q'}$. This is satisfied since by our assumptions in the theorem we have $m_i\big(\frac{1}{p}-\frac{1}{2}\big)<\frac{1}{q'}$, and we can choose ε arbitrarily small.

Estimate (5-10) will then be an easy consequence of the next lemma. Indeed, recalling our earlier observation that for each pair $(l_i,\tilde{l}_i)\in\mathcal{N}_i\times\widetilde{\mathcal{N}}_i$ one of the entries l_i or $\tilde{l}_i$ must be zero, we see that we have to sum over at most two of the parameters l_1, l_2, $\tilde{l}_1$, $\tilde{l}_2$.

Thus, there are four possibilities: if exactly two of the parameters are nonzero, then there are two distinct cases: either these parameters belong to the same surface (i.e., $l_1=l_2=0$ or $\tilde{l}_1=\tilde{l}_2=0$), which correspond to the left picture in Figure 10, or the nonzero parameters belong to two different surfaces, as in the "over cross" situation shown in the picture on the right hand side of Figure 10. The remaining two possibilities are firstly that only one parameter l_1, l_2, $\tilde{l}_1$, $\tilde{l}_2$ is nonzero, which happens if only one of the rectangles $U,\widetilde{U}$ touches only one of the axes, and secondly the situation where both rectangles are located away from the axes. In this last situation, we have indeed no further decomposition and only one term to sum.

The first two of the aforementioned possibilities can be dealt with directly by the next lemma. But, notice that the corresponding sums of course dominate the sums over fewer parameters (or even none), which allows to also handle the remaining two possibilities. □

Lemma 5.2. *Let $\mu,\omega\ge 0$, $\nu>0$, $n,c_1,c_2\ge 0$ such that $(c_1\vee c_2)\mu<\nu+\omega$, and let $a,b\in\mathbb{R}_+$. Then*

$$\begin{aligned}\sum_{l_1,l_2\in\mathbb{N}}&(1+l_1+l_2)^n 2^{-(l_1+l_2)\omega}(a2^{-l_2c_2}\vee b)^{-\mu}(a\vee b2^{-l_1c_1})^{-\mu}(a2^{-l_1}\wedge b2^{-l_2})^\nu\\ &\le\sum_{l_1,l_2\in\mathbb{N}}(1+l_1+l_2)^n 2^{-(l_1+l_2)\omega}(a\vee b)^{-\mu}(a2^{-l_2c_2}\vee b2^{-l_1c_1})^{-\mu}(a2^{-l_1}\wedge b2^{-l_2})^\nu\\ &\lesssim(a\vee b)^{-2\mu}(a\wedge b)^\nu.\end{aligned}$$

In the last estimate, the constant hidden by the symbol $\lesssim$ will depend only on the exponent n.

We remark that the bound in this lemma is essentially sharp, as one can immediately see by looking at the term with $l_1 = 0 = l_2$. Notice that the proof is easier when $\omega > 0$.

Proof. To prove the first inequality, observe that $a2^{-l_2c_2} \vee b2^{-l_1c_1}$ is bounded by $a2^{-l_2c_2} \vee b$ as well as by $a \vee b2^{-l_1c_1}$, and hence by the minimum of these expressions. Therefore we have

$$\begin{aligned}(a2^{-l_2c_2} \vee b) \wedge (a \vee b2^{-l_1c_1}) &\geq a2^{-l_2c_2} \vee b2^{-l_1c_1},\\ (a2^{-l_2c_2} \vee b) \vee (a \vee b2^{-l_1c_1}) &= a \vee b;\end{aligned}$$

hence

$$(a2^{-l_2c_2} \vee b)(a \vee b2^{-l_1c_1}) \geq (a \vee b)(a2^{-l_2c_2} \vee b2^{-l_1c_1}).$$

Using the symmetry in this estimate, it suffices to estimate

$$S = a^{\nu} \sum_{\substack{l_1,l_2\in\mathbb{N}\\ a2^{-l_1}\leq b2^{-l_2}}} l_1^n l_2^n 2^{-(l_1+l_2)\omega}(a2^{-l_2c_2} \vee b2^{-l_1c_1})^{-\mu}2^{-l_1\nu}.$$

On the one hand, we have

$$\begin{aligned}S &\leq a^{\nu}b^{-\mu}\sum_{l_1} l_1^n 2^{l_1(c_1\mu-\nu-\omega)} \sum_{l_2:a2^{-l_1}\leq b2^{-l_2}} l_2^n 2^{-l_2\omega}\\ &\leq a^{\nu}b^{-\mu}\log^{n+1}\Big(\frac{a}{b}+\frac{b}{a}\Big)\sum_{l_1} l_1^{2n+1}2^{l_1(c_1\mu-\nu-\omega)} \lesssim a^{\nu}b^{-\mu}\log^{n+1}\Big(\frac{a}{b}+\frac{b}{a}\Big).\end{aligned}$$

In the case $\omega > 0$, we might get along even without the log-term. On the other hand,

$$\begin{aligned}S &\leq a^{\nu-\mu}\sum_{l_2} l_2^n 2^{l_2(c_2\mu-\omega)} \sum_{l_1:a2^{-l_1}\leq b2^{-l_2}} l_1^n 2^{-l_1(\nu+\omega)}\\ &\leq a^{\nu-\mu}\sum_{l_2} l_2^n 2^{l_2(c_2\mu-\omega)} \sum_{l_1:a2^{-l_1}\leq b2^{-l_2}} l_1^n 2^{-l_1\nu}\\ &\sim a^{\nu-\mu}\log^{n}\Big(\frac{a}{b}+\frac{b}{a}\Big)\Big(\frac{b}{a}\Big)^{\nu}\sum_{l_2} l_2^{2n}2^{l_2(c_2\mu-\nu-\omega)} \sim a^{-\mu}b^{\nu}\log^{n}\Big(\frac{a}{b}+\frac{b}{a}\Big).\end{aligned}$$

Combining these two estimates, we obtain

$$\frac{S}{\log^{n+1}(\frac{a}{b}+\frac{b}{a})} \lesssim a^{-\mu}b^{\nu} \wedge a^{\nu}b^{-\mu} = (a\vee b)^{-\mu}(a\wedge b)^{\nu}. \qquad \square$$

6. Passage from bilinear to linear estimates

Recall that $\bar{m} = m_1 \vee m_2$, $m = m_1 \wedge m_2$ and $1/h = 1/m_1 + 1/m_2$. The first step to prove our main theorem, Theorem 1.2, is the following Lorentz space estimate for the adjoint restriction operator R^* associated to $\Gamma = \mathrm{graph}(\phi)$.

Theorem 6.1. *Let $p_0 = 1 + \bar{m}/(\bar{m}+m)$, $2p > \max\{\frac{10}{3}, 2p_0, h+1\}$ and $1/s' \geq (h+1)/(2p)$. Then R^* is bounded from $L^{s,t}(\Gamma, d\nu)$ to $L^{2p,t}(\mathbb{R}^3)$ for any $1 \leq t \leq \infty$.*

Proof. We begin by observing that we may assume

$$\frac{h+1}{p} > 1. \tag{6-1}$$

Indeed, if $2p \geq 2(h+1)$, then we have the Stein–Tomas-type result that R^* is bounded from $L^2(\Gamma, d\nu)$ to $L^{2p}(\mathbb{R}^3)$ (see [Ikromov et al. 2010; Ikromov and Müller 2011]). Interpolating this with the trivial estimate from $L^1(\Gamma, d\nu)$ to $L^\infty(\mathbb{R}^3)$ and applying Hölder's inequality on Γ, we see that the situation where $(h+1)/p \leq 1$ is settled in Theorem 6.1.

In the remaining cases, interpolation theory for Lorentz spaces (see, e.g., [Grafakos 2008]) shows that it suffices to prove the restricted weak-type estimate

$$\|\widehat{\chi_\Omega\, d\nu}\|_{2p} \lesssim |\Omega|^{\frac{1}{s}} \tag{6-2}$$

for any measurable set $\Omega \subset Q = [0,1] \times [0,1]$.

To this end we perform the kind of Whitney decomposition mentioned in Section 5 of $Q \times Q = \bigcup_j \bigcup_{k \approx \tilde{k}} \tau_{jk} \times \tau_{j\tilde{k}}$ into "well-separated neighboring rectangles" τ_{jk} and $\tau_{j\tilde{k}}$, where

$$\tau_{jk} = [(k_1-1)2^{-j_1}, k_1 2^{-j_1}] \times [(k_2-1)2^{-j_2}, k_2 2^{-j_2}],$$

and where $k \approx \tilde{k}$ means that $2 \leq |k_i - \tilde{k}_i| \leq C$, $i = 1, 2$ (see [Lee 2006; Vargas 2005]). Then we may estimate

$$\|\widehat{\chi_\Omega\, d\nu}\|_{2p}^2 = \|\widehat{\chi_\Omega\, d\nu}\,\widehat{\chi_\Omega\, d\nu}\|_p \leq \sum_j \biggl(\sum_{k\sim\tilde{k}} \|\mathfrak{F}(\chi_{\Omega\cap\tau_{jk}}\, d\nu)\mathfrak{F}(\chi_{\Omega\cap\tau_{j\tilde{k}}}\, d\nu)\|_p^{p^*}\biggr)^{\frac{1}{p^*}},$$

where

$$p^* = \min\{p, p'\}, \tag{6-3}$$

with $1/p + 1/p' = 1$. The last step can be obtained by interpolation between the case $p = 2$, where one may apply Plancherel's theorem, and the cases $p = 1$ and $p = \infty$, which are simply treated by means of the triangle inequality (see Lemma 6.3 in [Tao and Vargas 2000a]). We claim

$$(\bar{m}+3)\Bigl(\frac{1}{p} - \frac{1}{2}\Bigr) < \frac{h+1}{2p}. \tag{6-4}$$

Case 1: $\bar{m} \leq 2m$. Then $\bar{m} \leq 3h$ and

$$(\bar{m}+3)\Bigl(\frac{1}{p} - \frac{1}{2}\Bigr) - \frac{h+1}{2p} \leq 3(h+1)\Bigl(\frac{1}{p} - \frac{1}{2}\Bigr) - \frac{h+1}{2p} = (h+1)\Bigl(\frac{5}{2p} - \frac{3}{2}\Bigr) < 0$$

according to our assumptions.

Case 2: $\bar{m} > 2m$. Here,

$$h+1 = \frac{\bar{m}m + \bar{m} + m}{\bar{m}+m} > \frac{\bar{m}m + 3m}{\bar{m}+m} = (\bar{m}+3)\frac{m}{\bar{m}+m},$$

and thus

$$(\bar m+3)\Big(\frac{1}{p}-\frac{1}{2}\Big)-\frac{h+1}{2p}<(\bar m+3)\Big(\frac{1}{p}-\frac{1}{2}-\frac{m}{\bar m+m}\frac{1}{2p}\Big)$$
$$=(\bar m+3)\Big(\frac{1}{2p}\Big(1+\frac{\bar m}{\bar m+m}\Big)-\frac{1}{2}\Big)<0,$$

because of our assumption $2p>2p_0$.

In both cases, these estimates show that we may choose $q\ge 2$ such that

$$(\bar m+3)\Big(\frac{1}{p}-\frac{1}{2}\Big)<\frac{1}{q'}<\frac{h+1}{2p} \tag{6-5}$$

(recall here (6-1), which allows to choose $q\ge 2$).

The first inequality allows us to apply Theorem 5.1 to the pair of hypersurfaces

$$S_{jk}=\{(\xi,\phi(\xi)):\xi\in\tau_{jk}\}\quad\text{and}\quad S_{j\tilde k}=\{(\xi,\phi(\xi)):\xi\in\tau_{j\tilde k}\},$$

with

$$\rho_i=2^{-j_i},\quad \varkappa_i\sim(k_i2^{-j_i})^{m_i-2}\sim(\tilde k_i2^{-j_i})^{m_i-2},\quad \varkappa_i\rho_i^2\sim k_i^{m_i-2}2^{-j_im_i}. \tag{6-6}$$

Without loss of generality, we may assume

$$k\in I:=\{k:k_1^{m_1-2}2^{-j_1m_1}\ge k_2^{m_2-2}2^{-j_2m_2}\}, \tag{6-7}$$

i.e., $\varkappa_1\rho_1^2\ge\varkappa_2\rho_2^2$. Thus, by Theorem 5.1,

$$\begin{aligned}\|R^*_{S_{jk},S_{j\tilde k}}\|_{L^q(S_{jk})\times L^q(S_{j\tilde k})\to L^p(\mathbb{R}^3)}&\\ &\lesssim(\rho_1\rho_2)^{\frac{2}{q'}-\frac{1}{p}}(\varkappa_1\rho_1^2\vee\varkappa_2\rho_2^2)^{-\frac{1}{p}}\bigg(\frac{\varkappa_1\rho_1^2\vee\varkappa_2\rho_2^2}{\varkappa_1\rho_1^2\wedge\varkappa_2\rho_2^2}\bigg)^{\frac{2}{p}-1+\varepsilon}\\ &=2^{-(j_1+j_2)(\frac{2}{q'}-\frac{1}{p})}k_1^{-(m_1-2)\frac{1}{p}}2^{j_1m_1\frac{1}{p}}\bigg(\frac{k_1^{m_1-2}}{k_2^{m_2-2}}2^{-(j_1m_1-j_2m_2)}\bigg)^{\frac{2}{p}-1+\varepsilon}=A_j\cdot B_{k,j}^{\frac{1}{p}},\end{aligned}$$

if we define

$$A_j=2^{-(j_1+j_2)(\frac{2}{q'}-\frac{1}{p})}2^{j_1m_1\frac{1}{p}}2^{-(j_1m_1-j_2m_2)\frac{2}{p}-1+\varepsilon},$$
$$B_{j,\tilde k}\sim B_{k,j}=k_1^{-(m_1-2)}\bigg(\frac{k_1^{m_1-2}}{k_2^{m_2-2}}\bigg)^{2-p+\varepsilon p}.$$

Since $|\{\tilde k:k\sim\tilde k\}|\lesssim 1$ for fixed k, we conclude that

$$\begin{aligned}\|\widehat{\chi_\Omega\,d\nu}\|_{2p}^2&\lesssim\sum_j A_j\bigg(\sum_{k\sim\tilde k}\big(B_{k,j}^{\frac{1}{p}}|\Omega\cap\tau_{jk}|^{\frac{1}{q}}|\Omega\cap\tau_{j\tilde k}|^{\frac{1}{q}}\big)^{p^*}\bigg)^{\frac{1}{p^*}}\\ &\lesssim\sum_j A_j\bigg(\sum_k B_{k,j}^{\frac{p^*}{p}}|\Omega\cap\tau_{jk}|^{\frac{2p^*}{q}}\bigg)^{\frac{1}{p^*}}.\end{aligned}$$

Therefore we are reduced to showing

$$\sum_j A_j \Big(\sum_k B_{k,j}^{\frac{p^*}{p}} |\Omega\cap\tau_{jk}|^{\frac{2p^*}{q}}\Big)^{\frac{1}{p^*}} \lesssim |\Omega|^{\frac{2}{s}}. \tag{6-8}$$

6A. ***Further reduction.*** We have the decomposition

$$\frac{2p^*}{q} = \frac{\alpha}{r^*} + \frac{1}{r^{*\prime}},$$

where $r* \in [1,\infty]$ will be determined later, and introduce $r = r^* p^*/p$. Applying Hölder's inequality to the summation in k, with Hölder exponent $r^* \geq 1$, we get

$$\begin{aligned}\Big(\sum_{k\in I} B_{k,j}^{\frac{p^*}{p}} |\Omega\cap\tau_{jk}|^{\frac{2p^*}{q}}\Big)^{\frac{1}{p^*}} &\leq \Big(\sum_{k\in I} B_{k,j}^{\frac{p^*r^*}{p}} |\Omega\cap\tau_{jk}|^{\alpha}\Big)^{\frac{1}{p^*r^*}} \Big(\sum_{k\in I} |\Omega\cap\tau_{jk}|\Big)^{\frac{1}{p^*r^{*\prime}}} \\ &\leq \Big(\sum_{k\in I} B_{k,j}^{r}\Big)^{\frac{1}{pr}} \min\{|\Omega|, 2^{-j_1-j_2}\}^{\frac{\alpha}{p^*r^*}} |\Omega|^{\frac{1}{p^*}(1-\frac{1}{r^*})} \\ &= \Big(\sum_{k\in I} B_{k,j}^{r}\Big)^{\frac{1}{pr}} \min\{|\Omega|, 2^{-j_1-j_2}\}^{\frac{2}{q}-\frac{1}{p^*}+\frac{1}{pr}} |\Omega|^{\frac{1}{p^*}-\frac{1}{pr}}.\end{aligned}$$

Moreover we have $|\Omega| \leq |Q| = 1$, as well as $1/s' \geq h + 1/(2p)$, i.e., $2-(h+1)/p \geq 2/s$. Therefore $|\Omega|^{2-(h+1)/p} \leq |\Omega|^{2/s}$, and thus in order to prove (6-8), it suffices to show that

$$|\Omega|^{2-\frac{h+1}{p}-\frac{1}{p^*}+\frac{1}{pr}} \gtrsim \sum_j A_j \min\{|\Omega|, 2^{-j_1-j_2}\}^{\frac{2}{q}-\frac{1}{p^*}+\frac{1}{pr}} \Big(\sum_{k\in I} B_{k,j}^{r}\Big)^{\frac{1}{rp}},$$

i.e., that

$$\begin{aligned}|\Omega|^{2-\frac{h+1}{p}-\frac{1}{p^*}+\frac{1}{pr}} \gtrsim \sum_j 2^{-(j_1m_1-j_2m_2)\frac{2}{p}-1+\varepsilon} 2^{-(j_1+j_2)(\frac{2}{q'}-\frac{1}{p})} 2^{j_1m_1\frac{1}{p}} \min\{|\Omega|, 2^{-j_1-j_2}\}^{\frac{2}{q}-\frac{1}{p^*}+\frac{1}{pr}} \\ \times \Big(\sum_{k\in I} k_1^{(m_1-2)(1-p+\varepsilon p)r} k_2^{(m_2-2)(p-2-\varepsilon p)r}\Big)^{\frac{1}{pr}}.\end{aligned}$$

We apply the change of variables $l = j_1 + j_2 \in \mathbb{N}$, $l' = j_1m_1 - j_2m_2 \in \mathbb{Z}$, such that

$$j_1 = \frac{m_2 l + l'}{m_1+m_2}.$$

Then the exponent in j_1, j_2 becomes

$$\begin{aligned}(j_1m_1-j_2m_2)\Big(1-\frac{2}{p}-\varepsilon\Big)+(j_1+j_2)\Big(\frac{1}{p}-\frac{2}{q'}\Big)+j_1m_1\frac{1}{p} &= l'\Big(1-\frac{2}{p}-\varepsilon\Big)+l\Big(\frac{1}{p}-\frac{2}{q'}\Big)+\frac{m_1m_2l+m_1l'}{m_1+m_2}\frac{1}{p} \\ &= \frac{l'}{p}\Big(p-\varepsilon p-\frac{m_1+2m_2}{m_1+m_2}\Big)+l\Big(\frac{h+1}{p}-\frac{2}{q'}\Big).\end{aligned}$$

The summation over $k \in I_{l'} = \{k_1^{m_1-2} \geq k_2^{m_2-2} 2^{l'}\}$ is independent of l, and thus we have finally reduced the proof of (6-8) to proving the following two decoupled estimates:[4]

$$\sum_{l'=-\infty}^{\infty} 2^{\frac{l'}{p}(p-\varepsilon p-\frac{m_1+2m_2}{m_1+m_2})} \Bigl(\sum_{k\in I_{l'}} k_1^{(m_1-2)(1-p+\varepsilon p)r} k_2^{(m_2-2)(p-2-\varepsilon p)r} \Bigr)^{\frac{1}{pr}} < \infty \tag{6-9}$$

and

$$\sum_{l=0}^{\infty} 2^{l(\frac{h+1}{p}-\frac{2}{q'})} \min\{|\Omega|, 2^{-l}\}^{\frac{2}{q}-\frac{1}{p^*}+\frac{1}{pr}} \lesssim |\Omega|^{2-\frac{h+1}{p}-\frac{1}{p^*}+\frac{1}{pr}}. \tag{6-10}$$

6B. ***The case $m > 2$.*** In this case we have both $m_1 > 2$ and $m_2 > 2$.

6B1. *Summation in k.* We compare the sum over k in (6-9) with an integral. We claim

$$\int_{\substack{k_1,k_2\geq 1\\ k_1^{m_1-2}\geq k_2^{m_2-2}2^{l'}}} k_1^{(m_1-2)(1-p+\varepsilon p)r} k_2^{(m_2-2)(p-\varepsilon p-2)r}\, dk \lesssim \begin{cases} 2^{|l'|(\frac{1}{m_1-2}+(1-p+\varepsilon p)r)}, & l' \geq 0,\\ |l'| 2^{|l'|(\frac{1}{m_2-2}+(p-\varepsilon p-2)r)_+}, & l' < 0, \end{cases} \tag{6-11}$$

provided that

$$a := \frac{1}{m_1-2} + (1-p+\varepsilon p)r < 0 \tag{6-12}$$

and

$$a+b = \frac{1}{m_1-2} + \frac{1}{m_2-2} - r < 0, \tag{6-13}$$

where

$$b := \frac{1}{m_2-2} + (p-\varepsilon p-2)r \in \mathbb{R}. \tag{6-14}$$

For the moment, we will simply assume these conditions hold true. We shall collect several further conditions on the exponent r and verify at the end of this section that we can indeed find an r such that all these conditions are satisfied.

By means of the coordinate transformation $s = k_1^{m_1-2}$, $t = k_2^{m_2-2}$ (i.e., $dk \sim s^{\frac{1}{m_1-2}-1} t^{\frac{1}{m_2-2}-1} d(s,t)$), (6-11) simplifies to showing

$$J(a,b) = \iint_{\substack{s,t\geq 1\\ s\geq t2^{l'}}} s^a t^b \frac{ds}{s}\frac{dt}{t} \lesssim \begin{cases} 2^{|l'|a}, & l' \geq 0,\\ 2^{|l'|b_+}, & l' < 0, \end{cases} \tag{6-15}$$

provided that $a < 0$, $a+b < 0$. Here we have set $b_+ = b \vee 0$. Changing t' to $s2^{-l'}/t$, the set of integration for the t-variable $\{t : t \geq 1, s2^{-l'}/t \geq 1\}$ transforms into $\{t' : s2^{-l'}/t' \geq 1, t' \geq 1\}$, and thus, since we

[4] Technically, we only have to sum over the smaller set $l' \in m_1\mathbb{N} - m_2\mathbb{N}$.

assume $a+b<0$,

$$J(a,b)=\iint\limits_{\substack{s,t'\geq 1\\ s\geq t'2^{l'}}} s^a\Big(\frac{s}{t'}2^{-l'}\Big)^b\frac{ds}{s}\frac{dt'}{t'}=2^{-l'b}\int_{t'=1}^{\infty}t'^{-b}\int_{s=1\vee t'2^{l'}}^{\infty}s^{a+b}\frac{ds}{s}\frac{dt'}{t'}$$
$$=2^{-l'b}\int_1^\infty(1\vee t'2^{l'})^{a+b}t'^{-b}\frac{dt'}{t'}.$$

If $l'\geq 0$, then clearly $1\vee t'2^{l'}=t'2^{l'}$, and since $a<0$, we get

$$J(a,b)=2^{l'a}\int_1^\infty t'^a\frac{dt'}{t'}\sim 2^{l'a}.$$

And, if $l'<0$, then we can split it into

$$J(a,b)=2^{-l'b}\int_1^{2^{|l'|}}t'^{-b}\frac{dt'}{t'}+2^{l'a}\int_{2^{|l'|}}^\infty t'^a\frac{dt'}{t'}=2^{-l'b}\frac{1-2^{l'b}}{b}+\int_1^\infty u^a\frac{du}{u}$$
$$\lesssim|l'|(2^{|l'|b}+1)\sim|l'|2^{|l'|b_+}$$

(notice that the additional factor $|l'|$ arises in fact only when $b=0$). This proves (6-15).

6B2. *Summation in l'.* In order to apply (6-11) to (6-9), we split the sum in (6-9) into summation over $l'\geq 0$ and summation over $l'<0$. In the first case $l'\geq 0$, we obtain

$$\sum_{l'\geq 0}2^{\frac{l'}{p}(p-\varepsilon p-\frac{m_1+2m_2}{m_1+m_2})}\Big(\sum_{k\in I}k_1^{(m_1-2)(1-p+\varepsilon p)r}k_2^{(m_2-2)(p-\varepsilon p-2)r}\Big)^{\frac{1}{rp}}$$
$$\lesssim\sum_{l'\geq 0}2^{\frac{l'}{p}(p-\varepsilon p-1-\frac{m_2}{m_1+m_2})}2^{l'(\frac{1}{pr}\frac{1}{m_1-2}+\frac{1-p+\varepsilon p}{p})}$$
$$=\sum_{l'\geq 0}2^{\frac{l'}{p}(\frac{1}{r}\frac{1}{m_1-2}-\frac{m_2}{m_1+m_2})}. \tag{6-16}$$

The sum is finite provided

$$\frac{1}{r}<\frac{m_2(m_1-2)}{m_1+m_2}, \tag{6-17}$$

which gives yet another condition for our collection.

In the second case $l'<0$, we have

$$\sum_{l'<0}2^{\frac{l'}{p}(p-\varepsilon p-\frac{m_1+2m_2}{m_1+m_2})}\Big(\sum_{k\in I}k_1^{(m_1-2)(1-p+\varepsilon p)r}k_2^{(m_2-2)(p-\varepsilon p-2)r}\Big)^{\frac{1}{rp}}$$
$$\lesssim\sum_{l'<0}2^{\frac{l'}{p}(p-\varepsilon p-\frac{m_1+2m_2}{m_1+m_2})}|l'|2^{|l'|(\frac{1}{pr}\frac{1}{m_2-2}+\frac{p-\varepsilon p-2}{p})_+}$$
$$=\sum_{l'<0}|l'|2^{\frac{l'}{p}(p-\varepsilon p-\frac{m_1+2m_2}{m_1+m_2}-(\frac{1}{r}\frac{1}{m_2-2}+p-\varepsilon p-2)_+)}. \tag{6-18}$$

Notice that for sufficiently small $\varepsilon > 0$ we have $p - \varepsilon p > p_0 = 1 + \bar{m}/(\bar{m}+m) \geq 1 + m_2/(m_1+m_2)$, and therefore

$$p - \varepsilon p - \frac{m_1 + 2m_2}{m_1 + m_2} > 0. \tag{6-19}$$

Thus the last sum in (6-18) converges in the case where

$$\frac{b}{r} = \frac{1}{r}\frac{1}{m_2 - 2} + p - \varepsilon p - 2 \leq 0.$$

This shows that we only need to discuss the case where $b > 0$, in which we need that

$$0 < p - \varepsilon p - \frac{m_1 + 2m_2}{m_1 + m_2} - \frac{1}{r}\frac{1}{m_2 - 2} - p + \varepsilon p + 2 = \frac{m_1}{m_1 + m_2} - \frac{1}{r}\frac{1}{m_2 - 2},$$

which is equivalent to

$$\frac{1}{r} < \frac{m_1(m_2 - 2)}{m_1 + m_2}. \tag{6-20}$$

Notice that this is of the same form as (6-17), only with the roles of m_1 and m_2 interchanged.

6B3. *Summation in l.* Recall that we want to show estimate (6-10), i.e.,

$$\sum_{l=0}^{\infty} 2^{l\left(\frac{h+1}{p} - \frac{2}{q'}\right)} \min\{2^{-l}, |\Omega|\}^{\frac{2}{q} - \frac{1}{p^*} + \frac{1}{pr}} \lesssim |\Omega|^{2 - \frac{h+1}{p} - \frac{1}{p^*} + \frac{1}{pr}}.$$

We claim it is sufficient to show that for $\mu > 0$ and $\nu - \mu > 0$,

$$\int_0^\infty e^{x\mu} \min\{e^{-x}, A\}^\nu \, dx \lesssim A^{\nu - \mu}. \tag{6-21}$$

Indeed, given (6-21), we apply it with $A = |\Omega|$, $\mu = (h+1)/p - 2/q'$ and $\nu = 2/q - 1/p^* + 1/(pr)$. Due to the choice of q in (6-5), we have $\mu > 0$. Moreover we want

$$0 < \nu - \mu = 2 - \frac{1}{p^*} + \frac{1}{pr} - \frac{h+1}{p} = \frac{1}{p}\left(2p - h - 1 - \frac{p}{p*} + \frac{1}{r}\right).$$

Notice that if $p \leq 2$, then $p/p* = 1$, but if $p > 2$, then $p/p* = p(1 - 1/p) = p - 1$. Thus $p/p* = 1 + (p-2)_+$ for all $1 \leq p \leq \infty$, i.e., the condition which is required here is

$$\frac{1}{r} > h + 2 - 2p + (p-2)_+. \tag{6-22}$$

In order to verify (6-21), observe that

$$\int_0^\infty e^{x\mu} \min\{e^{-x}, A\}^\nu \, dx = \int_{\ln A}^\infty e^{y\mu} A^{-\mu} \min\{e^{-y}A, A\}^\nu \, dy = A^{\nu-\mu} \int_{\ln A}^\infty e^{y\mu} \min\{e^{-y}, 1\}^\nu \, dy.$$

The last integral can be estimated by

$$\int_{\ln A}^\infty e^{y\mu} \min\{e^{-y}, 1\}^\nu \, dy \leq \int_{-\infty}^0 e^{y\mu} \, dy + \int_0^\infty e^{-y(\nu-\mu)} \, dy,$$

which is convergent since $\mu > 0$ and $\nu - \mu > 0$.

It still remains to be checked whether there exists some $1 \le r^* < \infty$ (for $m > 2$) for which r satisfies the conditions (6-12), (6-13), (6-17), (6-20) and (6-22).

This task will be accomplished in Lemma 6.2. First, we discuss the situation where $m = 2$.

6C. *The case m = 2.* We will just give some hints for how to modify the previous proof for this situation. In this case, $r = \infty$ turns out to be an appropriate choice, and the inequality that we need to start the argument with here reads

$$\Big(\sum_{k\in I} B_{k,j}^{\frac{p^*}{p}} |\Omega \cap \tau_{jk}|^{\frac{2p^*}{q}}\Big)^{\frac{1}{p^*}} \le (\sup_{k\in I} B_{k,j})^{\frac{1}{p}} \min\{|\Omega|, 2^{-j_1-j_2}\}^{\frac{2}{q}-\frac{1}{p^*}} |\Omega|^{\frac{1}{p^*}}.$$

This is very easy to prove, provided $2p^*/q \ge 1$ (notice that this condition corresponds to our previous decomposition of $2p^*/q$ when $r = \infty$). To see that indeed $2p^*/q \ge 1$, recall from (6-5) that $1/q' < (h+1)/(2p)$. Then, it is enough to check that $2p^*(1-(h+1)/(2p)) > 1$, i.e., $h+1-2p+p/p^* < 0$. The last condition is equivalent to $h+2-2p+(p-2)_+ < 0$. However, this is what we shall indeed verify in the proof of Lemma 6.2 (compare to estimate (6-30) when $m = 2$).

Observe next that we may rewrite the integral in (6-11) in terms of the L^r-norm as

$$\big\|(k_1^{(m_1-2)(1-p+\varepsilon p)} k_2^{(m_2-2)(p-\varepsilon p-2)})_{k\in I_{l'}}\big\|_r \lesssim \begin{cases} 2^{|l'|(\frac{1}{r}\frac{1}{m_1-2}+1-p+\varepsilon p)}, & l' \ge 0, \\ 2^{|l'|(\frac{1}{r}\frac{1}{m_2-2}+p-\varepsilon p-2)_+}, & l' < 0, \end{cases}$$

provided the conditions (6-12) and (6-13) hold true, i.e., that

$$\frac{1}{m_1-2} + (1-p+\varepsilon p)r < 0 \quad \text{and} \quad \frac{1}{m_1-2} + \frac{1}{m_2-2} - r < 0.$$

This gives rise to the conjecture that (for $r = \infty$) we should have

$$\sup_{k\in I} k_1^{(m_1-2)(1-p+\varepsilon p)} k_2^{(m_2-2)(p-\varepsilon p-2)} \le \sup_{s\ge t2^{l'}} s^{1-p+\varepsilon p} t^{p-\varepsilon p-2} \lesssim \begin{cases} 2^{|l'|(1-p+\varepsilon p)}, & l' \ge 0, \\ 2^{|l'|(p-\varepsilon p-2)_+}, & l' < 0, \end{cases} \tag{6-23}$$

which would suffice in this case. But notice that the conditions (6-12) and (6-13) are formally fulfilled for $r = \infty$, and it is then easy to check that (6-23) indeed holds true, even in the case $m = 2$.

6C1. *Summation in l'.* The summation in l' becomes simpler here. We split again into the sums over $l' \ge 0$ and $l' < 0$, and obtain for the first half of the sum in (6-16)

$$\sum_{l'\ge 0} 2^{\frac{l'}{p}(p-\varepsilon p-\frac{m_1+2m_2}{m_1+m_2})} 2^{l'\frac{1-p+\varepsilon p}{p}} = \sum_{l'\ge 0} 2^{-\frac{l'}{p}\frac{m_2}{m_1+m_2}} < \infty.$$

The second part of the sum becomes (compare to (6-18))

$$\sum_{l'<0} 2^{\frac{l'}{p}(p-\varepsilon p-\frac{m_1+2m_2}{m_1+m_2}-(p-\varepsilon p-2)_+)}.$$

We already know from (6-19) that $p-\varepsilon p-(m_1+2m_2)/(m_1+m_2)>0$. Thus the sum converges if $p-\varepsilon p\le 2$. For $p-\varepsilon p>2$, notice that

$$p-\varepsilon p-\frac{m_1+2m_2}{m_1+m_2}-(p-\varepsilon p-2)_+=\frac{m_1}{m_1+m_2}>0,$$

and thus the sum is finite.

6C2. *Summation in l.* It remains to show that

$$\sum_{l=0}^{\infty}2^{l(\frac{h+1}{p}-\frac{1}{q'})}\min\{|\Omega|,2^{-l}\}^{\frac{1}{q}-\frac{1}{p^*}}\lesssim|\Omega|^{2-\frac{h+1}{p}-\frac{1}{p^*}},$$

which is the special case $r=\infty$ of (6-10). We saw that this holds true provided (6-22) is valid, i.e., if $1/r>h+2-2p+(p-2)_+$.

However, if $m=2$, then

$$2p>p_0=\frac{2\bar m}{\bar m+2}+2=h+2.$$

Thus for the case $p\le 2$ we have $h+2-2p+(p-2)_+=h+2-2p<0$. For the case $p>2$ notice that

$$h+2-2p+(p-2)_+=h-p=\frac{2\bar m}{\bar m+2}-p<2-p<0. \qquad \square$$

6D. ***Final considerations.*** We finally verify that there is indeed always some r for which all necessary conditions (6-12), (6-13), (6-17), (6-20) and (6-22) are satisfied in the case $m>2$. Recall that

$$\frac{2p^*}{q}=\frac{\alpha}{r^*}+\frac{1}{r^{*\prime}},$$

and notice that it will suffice to verify the following equivalent inequalities:

$$\frac{1}{r}<(m_1-2)(p-1), \tag{6-24}$$

$$\frac{1}{r}<\frac{(m_1-2)(m_2-2)}{m_1+m_2-4}, \tag{6-25}$$

$$\frac{1}{r}<\frac{m_2(m_1-2)}{m_1+m_2}, \tag{6-26}$$

$$\frac{1}{r}<\frac{m_1(m_2-2)}{m_1+m_2}, \tag{6-27}$$

$$\frac{1}{r}>h+2-2p+(p-2)_+. \tag{6-28}$$

Lemma 6.2. *Assume $m>2$ and $2p>\max\{2p_0,h+1\}$, where we recall that $p_0=1+\bar m/(\bar m+m)$. Define*

$$J=]0,1+(p-2)_+]\cap\Big]h+2-2p+(p-2)_+,\frac{\bar m(m-2)}{\bar m+m}\Big[.$$

Then $J \neq \varnothing$, and for every $1/r \in J$ we have

$$r^* = \frac{rp}{p^*} \geq 1, \quad \text{and} \quad \alpha = r^*\left(\frac{2p^*}{q} - \frac{1}{r^{*\prime}}\right) > 0, \tag{6-29}$$

and moreover the inequalities (6-24), (6-25), (6-26), (6-27) and (6-28) are valid.

Proof. First of all, we will show that $J \neq \varnothing$. We need to see that

$$h + 2 - 2p + (p-2)_+ < \frac{\bar{m}(m-2)}{\bar{m}+m} = h - \frac{2\bar{m}}{\bar{m}+m}, \tag{6-30}$$

i.e., that $2p_0 = 2 + 2\bar{m}/(\bar{m}+m) < 2p - (p-2)_+$. For the case $p \leq 2$, this holds true since $2p > 2p_0$. If $p > 2$, observe that

$$h + 2 - 2p + (p-2)_+ = h - p < h - 2 \leq h - \frac{2\bar{m}}{\bar{m}+m} = \frac{\bar{m}(m-2)}{\bar{m}+m}.$$

Thus both intervals used for the definition of J are not empty, but we still have to check that their intersection is not trivial. Since we assume $2p > h+1$, we have

$$h + 2 - 2p + (p-2)_+ < 1 + (p-2)_+.$$

And, for $m > 2$, we also have $0 < \bar{m}(m-2)/(\bar{m}+m)$, which shows that $J \neq \varnothing$.

Next, if $1/r \in J$, then in particular $1/r \leq 1 + (p-2)_+ = p/p^*$, and thus $r^* = rp/p^* \geq 1$. To prove (6-29), observe that due to our choice of q in (6-5) we have $1/q > 1 - (h+1)/(2p)$, and thus it suffices to prove that

$$2p^*\left(1 - \frac{h+1}{2p}\right) > \frac{1}{r^{*\prime}} = 1 - \frac{p^*}{rp}.$$

This inequality is equivalent to

$$\frac{1}{r} > \frac{p}{p*} + h + 1 - 2p = h + 2 - 2p + (p-2)_+,$$

and thus is satisfied.

Considering the remaining conditions listed before the statement of the lemma, notice that (6-28) is immediate by the definition of J. Furthermore we have

$$\frac{1}{r} < \frac{\bar{m}(m-2)}{\bar{m}+m} = \frac{m_1m_2 - 2\bar{m}}{m_1+m_2} \leq \frac{m_1m_2 - 2m_i}{m_1+m_2}$$

for both $i = 1, 2$, which gives (6-26) and (6-27). To obtain (6-24), we estimate

$$\frac{1}{r} < \frac{\bar{m}(m-2)}{\bar{m}+m} \leq \frac{\bar{m}(m_1-2)}{\bar{m}+m} = (p_0-1)(m_1-2) < (p - \varepsilon p - 1)(m_1-2).$$

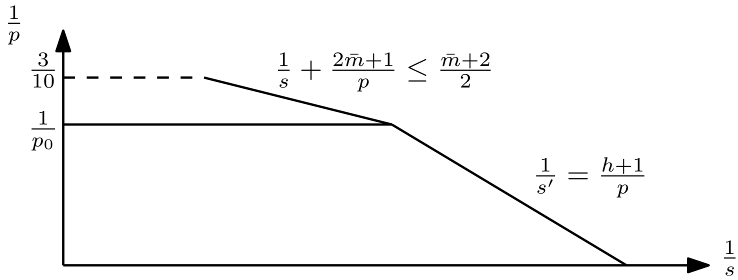


Figure 11. Range of p and q in Theorem 1.2.

Finally, observe that we have the following equivalences:

$$\begin{aligned} \frac{\bar{m}(m-2)}{\bar{m}+m} \leq \frac{(m_1-2)(m_2-2)}{m_1+m_2-4} \quad &\iff \quad \frac{\bar{m}}{\bar{m}+m} \leq \frac{\bar{m}-2}{\bar{m}+m-4} \\ &\iff \quad \bar{m}(\bar{m}+m)-4\bar{m} \leq \bar{m}(\bar{m}+m)-2(\bar{m}+m) \\ &\iff \quad m \leq \bar{m}. \end{aligned}$$

Hence (6-25) holds true as well. ☐

6E. ***Finishing the proof.*** We can now conclude the proof of our main result, Theorem 1.2:

Corollary 6.3. *Let $2p > \max\{\frac{10}{3}, h+1\}$, $1/s' \geq (h+1)/(2p)$ and $1/s+(2\bar{m}+1)/(2p) < (\bar{m}+2)/2$. Then R^* is bounded from $L^{s,t}(\Gamma)$ to $L^{2p,t}(\mathbb{R}^3)$ for every $1 \leq t \leq \infty$. If moreover $s \leq 2p$ or $1/s' > (h+1)/(2p)$, then R^* is bounded from $L^s(\Gamma)$ to $L^{2p}(\mathbb{R}^3)$.*

Proof. The crucial observation is that the intersection point of the two lines

$$\frac{1}{s'} = \frac{h+1}{2p} \quad \text{and} \quad \frac{2\bar{m}+1}{2p} + \frac{1}{s} = \frac{\bar{m}+2}{2}$$

has the p-coordinate $p = \hat{p}_0 = 1 + \bar{m}/(\bar{m}+m)$ (comparing with (1-6), notice that $\hat{p}_0 = p_0/2$). So, what remains is to establish estimates for R^* for the missing points $(1/s, 1/p)$ lying within the sectorial region defined by the conditions $(2\bar{m}+1)/(2p) + 1/s < (\bar{m}+2)/2$ and $1/p > 1/\hat{p}_0$ (the region above the horizontal threshold line $1/p = 1/p_0$ from Theorem 6.1 (see Figure 11).

Notice also that if $m \geq \bar{m}/2$, then $\hat{p}_0 \leq \frac{5}{3}$, i.e., $p_0 \leq \frac{10}{3}$, and hence the condition $1/s+(2\bar{m}+1)/(2p) < (\bar{m}+2)/2$ becomes redundant.

Moreover, the condition $1/s + (2\bar{m}+1)/(2p) < (\bar{m}+2)/2$ does only depend on $\bar{m}$, and not on m, whereas the condition $1/s' = (h+1)/(2p)$ depends on the height h, i.e., on both m_1 and m_2.

This leads to the following *heuristic idea*: Assume we fix $\bar{m}$ and consider a family of surfaces $\Gamma_{\bar{m},m^\sharp}$ corresponding to exponents $m_1 = \bar{m}$ and $m_2 = m^\sharp$ for different exponents $m < m^\sharp$ such that $\Gamma_{\bar{m},m} = \Gamma$ (think for instance of the graph of $x_1^{m^\sharp} + x_2^{\bar{m}}$ for $m^\sharp \neq m$). Let us then compare the restriction estimates that we have so far for the surface $\Gamma = \Gamma_{\bar{m},m}$ with the ones for the hypersurfaces $\Gamma_{\bar{m},m^\sharp}$. Denote by h and $h^\sharp$ the heights of these hypersurfaces. Then $h < h^\sharp$, so that the critical line $1/s' = (h^\sharp+1)/(2p)$ lies below the critical line $1/s' = (h+1)/(2p)$ for Γ, but its intersection point with the corresponding

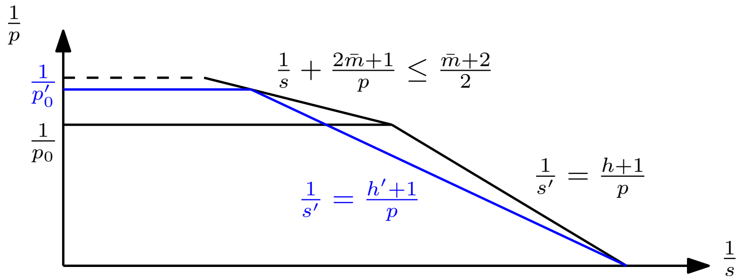


Figure 12. Variation of the minimal exponent.

horizontal threshold line $1/p = 1/p_0^\sharp$, where $p_0^\sharp = 2 + 2\bar{m}/(\bar{m} + m^\sharp) < p_0$, lies above the previous intersection point (see Figure 12).

This suggests that for our theorem, it should essentially be sufficient to "increase" $m^\sharp$ until $\bar{m} = 2m^\sharp$, because then we would have $p_0^\sharp = 2 + 2\bar{m}/(\bar{m} + m^\sharp) = \frac{10}{3}$. In other words, for any point $(1/s, 1/(2p))$ fulfilling the assumptions of Theorem 1.2, we would find an $m^\sharp \in]m, \bar{m}/2]$ such that $(1/s, 1/(2p))$ satisfies the requirements of Theorem 6.1 corresponding to the surface $\Gamma_{\bar{m},m'}$. Thus we would obtain the restriction estimate for the surface $\Gamma_{\bar{m},m^\sharp}$ at the point $(1/s, 1/(2p))$. However, since this surface has "less curvature" than $\Gamma_{\bar{m},m}$, as $m^\sharp > m$, the corresponding restriction inequality should hold true for $\Gamma_{\bar{m},m} = \Gamma$ as well.

To turn these heuristics into a solid proof, we just need to check that the bound for the bilinear operator that we obtained in Theorem 5.1 is increasing in m. Recall that for subsurfaces $S, \tilde{S} \subset S_{\bar{m},m}$ under the assumptions of the aforementioned theorem we obtained the bound

$$\|R^*_{S,\tilde{S}}\|_{L^s(S)\times L^s(\tilde{S})\to L^p(\mathbb{R}^3)} \lesssim C_{\bar{m},m} := (\rho_1\rho_2)^{\frac{2}{s'}-\frac{1}{p}} (\varkappa_1\rho_1^2 \vee \varkappa_2\rho_2^2)^{\frac{1}{p}-1+\varepsilon} (\varkappa_1\rho_1^2 \wedge \varkappa_2\rho_2^2)^{1-\frac{2}{p}-\varepsilon},$$

which we apply to $\rho_i = 2^{-j_i}$ and $\varkappa_i = (k_i 2^{-j_i})^{m_i-2}$ (see (6-6)). If we denote by $\rho_i^\sharp, \varkappa_i^\sharp$ the corresponding quantities associated to the exponents $\bar{m}$ and $m^\sharp$, then clearly $\rho_i^\sharp = \rho_i$ and $\varkappa_i^\sharp \le \varkappa_i$. Since we seek to extend the range of validity of Theorem 6.1, we may assume that $2p \le p_0 < 4$, and moreover that $2p \ge p_0^\sharp > 2$. Then we have $1/p - 1 + \varepsilon < 0$ and $1 - 2/p - \varepsilon < 0$ for sufficient small $\varepsilon > 0$, and hence

$$C_{\bar{m},m} \le (\rho_1^\sharp\rho_2^\sharp)^{\frac{2}{s'}-\frac{1}{p}} \big(\varkappa_1^\sharp(\rho_1^\sharp)^2 \vee \varkappa_2^\sharp(\rho_2^\sharp)^2\big)^{\frac{1}{p}-1+\varepsilon} \big(\varkappa_1^\sharp(\rho_1^\sharp)^2 \wedge \varkappa_2^\sharp(\rho_2^\sharp)^2\big)^{1-\frac{2}{p}-\varepsilon} = C_{\bar{m},m^\sharp}.$$

Proceeding with the latter estimate from here on as before in our proof of Theorem 6.1, but working now with $m^\sharp$ in place of m, we arrive at the statement of Corollary 6.3. $\square$

Appendix

A1. ***A short argument to improve [Ferreyra and Urciuolo 2009] to the critical line.*** We consider the set $A_0 = \{x \in \mathbb{R}^2 : \frac{1}{2} < |x| \le 1\}$ and define $H = 2\bar{m}/(2 + \bar{m})$. Note that $H < h$. Ferreyra and Urciuolo proved that for every p for which $p > 4$ and $1/s' > (H + 1)/p$, there is a constant $C_{p,s} > 0$ such that, for every function f_0 with supp $f_0 \subset A$, we have

$$\|R^*_{\mathbb{R}^2} f_0\|_p \le C_{p,s}\|f_0\|_s.$$

Rescaling this, we obtain

$$\|R^*_{\mathbb{R}^2} f_j\|_p \le C_{p,s} 2^{\frac{j}{h}(-\frac{1}{s'}+\frac{h+1}{p})}\|f_j\|_s \tag{A-1}$$

for every function f_j such that

$$\operatorname{supp} f_j \subset \left\{(x_1,x_2) : 2^{-\frac{j+1}{m_1}} \le x_1 \le 2^{-\frac{j}{m_1}},\ 2^{-\frac{j+1}{m_2}} \le x_2 \le 2^{-\frac{j}{m_2}}\right\}$$

and the same range of p, s.

Given a function f supported in the unit ball of $\mathbb{R}^2$, we decompose $f = \sum_{j=0}^{\infty} f_j$, where the functions f_j have supports as above. Then,

$$\left|\left\{x : |R^*_{\mathbb{R}^2} f(x)| > \lambda\right\}\right| \le \left|\left\{x : \left|\sum_{j=J}^{\infty} R^*_{\mathbb{R}^2} f_j(x)\right| > \frac{\lambda}{2}\right\}\right| + \left|\left\{x : \left|\sum_{1}^{j=J} R^*_{\mathbb{R}^2} f_j(x)\right| > \frac{\lambda}{2}\right\}\right|,$$

for some J to be chosen appropriately. Using Chebyshev's inequality, the last expression can be bounded above by

$$\left(\frac{2}{\lambda}\right)^{p_1}\left\|\sum_{j=J}^{\infty} R^*_{\mathbb{R}^2} f_j\right\|^{p_1}_{L^{p_1}} + \left(\frac{2}{\lambda}\right)^{p_2}\left\|\sum_{j=1}^{J} R^*_{\mathbb{R}^2} f_j\right\|^{p_2}_{L^{p_2}}.$$

Let us choose exponents $p_1 > p > p_2$ such that $1/s' = (h+1)/p$ and $(h+1)/p_2 > 1/s' > (h+1)/p_1 > (H+1)/p$. We use the triangle inequality and (A-1) and sum the resulting geometric series, obtaining the inequality

$$\left|\left\{x : |R^*_{\mathbb{R}^2} f(x)| > \lambda\right\}\right| \lesssim \left(\frac{2}{\lambda}\right)^{p_1} 2^{\frac{J}{h}(-\frac{p_1}{s'}+h+1)}\|f\|^{p_1}_{L^s} + \left(\frac{2}{\lambda}\right)^{p_2} 2^{\frac{J}{h}(-\frac{p_2}{s'}+h+1)}\|f\|^{p_2}_{L^s}.$$

By choosing J such that $2^J = \left(\|f\|_{L^s}/\lambda\right)^{hs'}$, we then arrive at the weak-type estimate

$$\left|\left\{x\ :\ |R^*_{\mathbb{R}^2} f(x)| > \lambda\right\}\right| \lesssim \left(\frac{\|f\|_{L^s}}{\lambda}\right)^{(h+1)s'} = \left(\frac{\|f\|_{L^s}}{\lambda}\right)^{p}.$$

From this, by interpolation with the trivial bound $\|R^*_{\mathbb{R}^2}\|_{L^1\to L^\infty} \le 1$, we obtain the desired strong-type estimate.

A2. ***Faà di Bruno's theorem and completion of the proof of Lemma 2.4.*** The formula of Faà di Bruno is a chain rule for higher-order derivatives of the composition of two functions. This is well known for functions in one real variable. However, we need a version for several variables.

Lemma A.1 (formula of Faà di Bruno). *Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$, and let $g = (g^1, \dots, g^m) \in C^\infty(U, V)$ and $f \in C^\infty(V, \mathbb{R}^l)$. For $\alpha \in \mathbb{N}^n$, we put $A_\alpha = \{\gamma \in \mathbb{N}^n : 1 \le |\gamma| \le |\alpha|\}$. Then $f \circ g$ is smooth, and for* *every $\alpha \in \mathbb{N}^n$ we have*

$$\partial^\alpha (f \circ g) = \alpha! \sum_{1 \le |\beta| \le |\alpha|} (\partial^\beta f) \circ g \sum_k \prod_{j=1}^m \prod_{\gamma \in A_\alpha} \left(\frac{\partial^\gamma g^j}{\gamma!} \right)^{k_\gamma^j},$$

where the sum in k is over all mappings $k : \{1, \dots, m\} \times A_\alpha \to \mathbb{N}$, $(j, \gamma) \mapsto k_\gamma^j$, such that

$$\sum_{\gamma \in A_\alpha} k_\gamma^j = \beta_j \tag{A-1}$$

for all $j = 1, \dots, m$ and

$$\sum_{j=1}^m \sum_{\gamma \in A_\alpha} k_\gamma^j \gamma = \alpha. \tag{A-2}$$

Proof. The elegant short proof in [Spindler 2005] for the one-dimensional case can easily be adapted to the higher-dimensional situation. ☐

We now come back to the proof of Lemma 2.4 and establish the still-missing estimates for the derivatives of the function ϕ_2 (given explicitly by (2-12)). Notice that these estimates cannot simply be obtained by means of a scaling argument, since the first-order derivatives are assumed to exhibit a different behavior than the higher-order derivatives.

We shall not really make use of formula (2-12), but rather proceed as follows: denoting by $e_1, \dots, e_{d+1}$ the canonical basis of $\mathbb{R}^{d+1}$, after applying a suitable orthogonal transformation to $\mathbb{R}^{d+1}$ we may and shall assume $n_1 = (0, \dots, 0, 1) = e_{d+1}$, and $E_1 = e_1, \dots, E_{d-1} = e_{d-1}$ and $e_d = h_1$ (recall here from the first part of the proof of Lemma 2.4 that $E_1, \dots, E_{d-1}$ is an orthonormal basis of $K = H_1 \cap H_2$). Then we may regard U_1 as a subset of $\mathbb{R}^d$, and we consider the function

$$H(\eta, \tau) = \tau - \phi_1(\eta), \quad \eta \in U_1,\ \tau \in \mathbb{R},$$

whose set of zeros agrees exactly with S. Observe first that the derivatives of H satisfy almost the same kind of estimates as ϕ_1:

$$\|H'\|_\infty \le \sqrt{A^2 + 1}, \qquad \|H^{(l)}\|_\infty \le A_l B r^l \quad \text{for every } l \ge 2. \tag{A-3}$$

Let $\psi(\xi) = \xi + \phi_2(\xi) n_2$, $\xi \in U_2$, be the parametrization of S induced by ϕ_2. Moreover, we introduce coordinates on U_2 by writing $\xi = \xi_1 E_1 + \dots + \xi_{d-1} E_{d-1} + \xi_d h_2$. Then obviously

$$H(\psi(\xi)) = 0 \quad \text{for all } \xi \in U_2. \tag{A-4}$$

Furthermore,

$$\frac{\partial \psi}{\partial \xi_j} = E_j + \frac{\partial \phi_2}{\partial \xi_j} n_2, \quad j = 1, \dots, d-1, \qquad \frac{\partial \psi}{\partial \xi_d} = h_2 + \frac{\partial \phi_2}{\partial \xi_d} n_2, \tag{A-5}$$

and

$$\partial_\xi^\alpha \psi = \partial_\xi^\alpha \phi_2 \, n_2 \quad \text{for all } \alpha \in \mathbb{N}^d,\ |\alpha| \ge 2. \tag{A-6}$$

From (A-4) and (A-5) we obtain that for $j = 1, \dots, d$,

$$\frac{\partial \phi_2}{\partial \xi_j}(\xi) = -\frac{\langle (\nabla H)(\psi(\xi)), \tilde{e}_j \rangle}{\langle (\nabla H)(\psi(\xi)), n_2 \rangle}, \tag{A-7}$$

if we put $\tilde{e}_j = E_j = e_j$, if $j = 1, \dots, d-1$ and $\tilde{e}_d = h_2$. Notice also that our transversality condition $|\langle n_2, N(v)\rangle| \geq a > 0$ for all $v \in S$ implies $|\langle (\nabla H)(\psi(\xi)), n_2\rangle| \geq a$. Thus (A-7) implies

$$\left| \frac{\partial \phi_2}{\partial \xi_j}(\xi) \right| \leq \frac{A+1}{a}. \tag{A-8}$$

It remains to show that

$$\|\partial^\alpha \phi_2\|_\infty = \|\partial^\alpha \psi\|_\infty \leq \widetilde{A}_l B r^{|\alpha|} \tag{A-9}$$

for every $|\alpha| \geq 2$, where we have used the abbreviation $\partial = \partial_\xi$. By induction, we may assume that for every $\gamma \in \mathbb{N}^d$ with $2 \leq |\gamma| < |\alpha|$ inequality (A-9) holds true.[5] Applying the partial derivative of order α to (A-4) yields

$$\partial^\alpha (H \circ \psi) = 0.$$

We apply the formula of Faà di Bruno (Lemma A.1). First, we discuss the summands in Faà di Bruno's formula with $|\beta| = 1$, say $\beta = e_{j_0}$ for some $j_0 = 1, \dots, m$. How many k's are there for which $\sum_{\gamma \in A_\alpha} k_\gamma^j = \beta_j = \delta_{j j_0}$ and $\sum_{j=1}^m \sum_{\gamma \in A_\alpha} k_\gamma^j \gamma = \alpha$? By the first condition, there exists a γ_0 such that $k_{\gamma_0}^{j_0} = 1$ and $k_\gamma^j = 0$ for $j \neq j_0$ or $\gamma \neq \gamma_0$. But then the second condition implies $\gamma_0 = \alpha$. Thus we obtain

$$\begin{aligned}\sum_{|\beta|=1} (\partial^\beta H) \circ \psi \sum_k \prod_{j=1}^m \prod_{\gamma \in A_\alpha} \left(\frac{\partial^\gamma \psi^j}{\gamma!} \right)^{k_\gamma^j} &= \sum_{j_0=1}^m (\partial_{j_0} H) \circ \psi \left(\frac{\partial^\alpha \psi^{j_0}}{\alpha!} \right)^{k_\alpha^{j_0}} \\ &= \frac{1}{\alpha!} \langle (\nabla H) \circ \psi, \partial^\alpha \psi \rangle = \frac{\partial^\alpha \phi_2}{\alpha!} \langle (\nabla H) \circ \psi, n_2 \rangle,\end{aligned}$$

where we have used (A-6) once more. This implies

$$|\partial^\alpha \phi_2| \leq \frac{\alpha!}{a} \left| \sum_{|\beta|=2}^{|\alpha|} (\partial^\beta H) \circ \psi \sum_k \prod_{j=1}^m \prod_{\gamma \in A_\alpha} \left(\frac{\partial^\gamma \psi^j}{\gamma!} \right)^{k_\gamma^j} \right|,$$

where the sum in k is over all mappings $k : \{1, \dots, m\} \times A_\alpha \to \mathbb{N}$, $(j, \gamma) \mapsto k_\gamma^j$ such that $\sum_{\gamma \in A_\alpha} k_\gamma^j = \beta_j$ for all $j = 1, \dots, m$ and $\sum_{j=1}^m \sum_{\gamma \in A_\alpha} k_\gamma^j \gamma = \alpha$. Observe that for all k appearing in the above sum, we have $k_\alpha^j = 0$ for all $j = 1, \dots, m$:

For, otherwise there would be some j_0 such that $k_\alpha^{j_0} = 1$ and $k_\gamma^j = 0$ if $\gamma \neq \alpha$ or $j \neq j_0$, a contradiction to $2 \leq |\beta| = \sum_{j,\gamma} k_\gamma^j$.

Thus, if $k_\gamma^j \neq 0$ for an exponent in the above sum, then we have $|\gamma| < |\alpha|$, and therefore our induction hypothesis implies the following:

[5] At the start of the induction with $|\alpha| = 2$, the range of such γ's is empty.

If $|\gamma| \ge 2$, then we may estimate $|\partial^\gamma \psi^j| \le A_{|\gamma|} B r^{|\gamma|}$. And, if $|\gamma| = 1$, then in view of (A-5) and (A-8), we may estimate $|\partial^\gamma \psi^j| \le 1 + (A+1)/a \lesssim 1$. Making also use of (A-3), we then arrive at the estimation

$$|\partial^\alpha \phi_2| \lesssim \sum_{|\beta|=2}^{|\alpha|} B r^{|\beta|} \sum_k \prod_{j=1}^{m} \prod_{|\gamma|=2}^{|\alpha|} [B r^{|\gamma|}]^{k_\gamma^j}$$
$$\le \sum_{|\beta|=2}^{|\alpha|} \sum_k B^{1+\sum_j \sum_{|\gamma|=2}^{|\alpha|} k_\gamma^j} r^{|\beta|+\sum_j \sum_{|\gamma|=2}^{|\alpha|} k_\gamma^j |\gamma|}.$$

Notice that we have

$$|\beta| = \sum_j \beta_j = \sum_j \sum_{|\gamma|=1}^{|\alpha|} k_\gamma^j = \sum_j \sum_{|\gamma|=2}^{|\alpha|} k_\gamma^j + \sum_j \sum_{|\gamma|=1} k_\gamma^j |\gamma|,$$

and thus

$$B^{1+\sum_j \sum_{|\gamma|=2}^{|\alpha|} k_\gamma^j} r^{|\beta|+\sum_j \sum_{|\gamma|=2}^{|\alpha|} k_\gamma^j |\gamma|} = B r^{\sum_j \sum_{|\gamma|=1}^{|\alpha|} k_\gamma^j |\gamma|} (Br)^{\sum_j \sum_{|\gamma|=2}^{|\alpha|} k_\gamma^j} \le B r^{|\alpha|},$$

where we have made use of our assumption $Br \le 1$. This proves also (A-9).

Acknowledgments

The authors wish to thank the referee for several helpful comments and suggestions. Moreover, they wish to express their gratitude to the Hausdorff Research Institute for Mathematics (HIM) in Bonn and its staff for its hospitality and for providing a wonderful working atmosphere.

References

[Barceló 1985] B. Barceló Taberner, "On the restriction of the Fourier transform to a conical surface", *Trans. Amer. Math. Soc.* **292**:1 (1985), 321–333. MR Zbl

[Barceló 1986] B. Barceló, "The restriction of the Fourier transform to some curves and surfaces", *Studia Math.* **84**:1 (1986), 39–69. 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 1985] J. Bourgain, "Estimations de certaines fonctions maximales", *C. R. Acad. Sci. Paris Sér. I Math.* **301**:10 (1985), 499–502. MR

[Bourgain 1991] J. Bourgain, "Besicovitch type maximal operators and applications to Fourier analysis", *Geom. Funct. Anal.* **1**:2 (1991), 147–187. MR Zbl

[Bourgain 1995a] 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 1995b] J. Bourgain, "Some new estimates on oscillatory integrals", pp. 83–112 in *Essays on Fourier analysis in honor of Elias M. Stein* (Princeton, NJ, 1991), edited by C. Fefferman et al., Princeton Math. Ser. **42**, Princeton Univ. Press, 1995. 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

[Buschenhenke 2015] S. Buschenhenke, "A sharp L^p-L^q-Fourier restriction theorem for a conical surface of finite type", *Math. Z.* **280**:1 (2015), 367–399. MR Zbl

[Fefferman 1970] C. Fefferman, "Inequalities for strongly singular convolution operators", *Acta Math.* **124**:1 (1970), 9–36. MR Zbl

[Ferreyra and Urciuolo 2009] E. Ferreyra and M. Urciuolo, "Fourier restriction estimates to mixed homogeneous surfaces", *J. Inequal. Pure Appl. Math.* **10**:2 (2009), art. id. 35, 11 pp. MR Zbl

[Grafakos 2008] L. Grafakos, *Classical Fourier analysis*, 2nd ed., Graduate Texts in Mathematics **249**, Springer, 2008. MR Zbl

[Greenleaf 1981] A. Greenleaf, "Principal curvature and harmonic analysis", *Indiana Univ. Math. J.* **30**:4 (1981), 519–537. MR Zbl

[Ikomov and Müller 2012] I. A. Ikomov and D. Müller, "$L^p - L^2$ Fourier restriction for hypersurfaces in $\mathbb{R}^3$, I", preprint, 2012. arXiv

[Ikomov and Müller 2014] I. A. Ikomov and D. Müller, "$L^p - L^2$ Fourier restriction for hypersurfaces in $\mathbb{R}^3$, II", preprint, 2014. arXiv

[Ikromov and Müller 2011] I. A. Ikromov and D. Müller, "Uniform estimates for the Fourier transform of surface carried measures in $\mathbb{R}^3$ and an application to Fourier restriction", *J. Fourier Anal. Appl.* **17**:6 (2011), 1292–1332. MR Zbl

[Ikromov et al. 2010] I. A. Ikromov, M. Kempe, and D. Müller, "Estimates for maximal functions associated with hypersurfaces in $\mathbb{R}^3$ and related problems of harmonic analysis", *Acta Math.* **204**:2 (2010), 151–271. 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

[Moyua et al. 1996] A. Moyua, A. Vargas, and L. Vega, "Schrödinger maximal function and restriction properties of the Fourier transform", *Internat. Math. Res. Notices* **1996**:16 (1996), 793–815. MR Zbl

[Moyua et al. 1999] A. Moyua, A. Vargas, and L. Vega, "Restriction theorems and maximal operators related to oscillatory integrals in $\mathbb{R}^3$", *Duke Math. J.* **96**:3 (1999), 547–574. MR Zbl

[Ramos 2016] J. Ramos, "The trilinear restriction estimate with sharp dependence on the transversality", preprint, 2016. arXiv

[Sogge 1987] C. D. Sogge, "A sharp restriction theorem for degenerate curves in $\mathbb{R}^2$", *Amer. J. Math.* **109**:2 (1987), 223–228. MR Zbl

[Spindler 2005] K. Spindler, "A short proof of the formula of Faà di Bruno", *Elem. Math.* **60**:1 (2005), 33–35. MR Zbl

[Stein 1986] E. M. Stein, "Oscillatory integrals in Fourier analysis", pp. 307–355 in *Beijing lectures in harmonic analysis* (Beijing, 1984), edited by E. M. Stein, Ann. of Math. Stud. **112**, Princeton Univ. Press, 1986. 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

[Stovall 2015] B. Stovall, "Linear and bilinear restriction to certain rotationally symmetric hypersurfaces", preprint, 2015.

[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

[Tao 2001a] 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 2001b] T. Tao, "From rotating needles to stability of waves: emerging connections between combinatorics, analysis, and PDE", *Notices Amer. Math. Soc.* **48**:3 (2001), 294–303. MR Zbl

[Tao 2003a] T. Tao, "Recent progress on the restriction conjecture", preprint, 2003. arXiv

[Tao 2003b] T. Tao, "A sharp bilinear restrictions estimate for paraboloids", *Geom. Funct. Anal.* **13**:6 (2003), 1359–1384. 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

[Tao et al. 1998] T. Tao, A. Vargas, and L. Vega, "A bilinear approach to the restriction and Kakeya conjectures", *J. Amer. Math. Soc.* **11**:4 (1998), 967–1000. MR Zbl

[Tomas 1975] P. A. Tomas, "A restriction theorem for the Fourier transform", *Bull. Amer. Math. Soc.* **81**:2 (1975), 477–478. MR Zbl

[Vargas 2005] A. Vargas, "Restriction theorems for a surface with negative curvature", *Math. Z.* **249**:1 (2005), 97–111. MR Zbl

[Wolff 1995] T. Wolff, "An improved bound for Kakeya type maximal functions", *Rev. Mat. Iberoamericana* **11**:3 (1995), 651–674. MR Zbl

[Wolff 2001] T. Wolff, "A sharp bilinear cone restriction estimate", *Ann. of Math.* (2) **153**:3 (2001), 661–698. MR Zbl

[Zygmund 1974] A. Zygmund, "On Fourier coefficients and transforms of functions of two variables", *Studia Math.* **50**:2 (1974), 189–201. MR Zbl

Received 3 Feb 2016. Revised 2 Sep 2016. Accepted 22 Jan 2017.

STEFAN BUSCHENHENKE: buschenhenke@math.uni-kiel.de
Mathematisches Seminar, Christian-Albrechts-Universität Kiel, Ludewig-Meyn Str. 4, D-24098 Kiel, Germany

DETLEF MÜLLER: mueller@math.uni-kiel.de
Mathematisches Seminar, Christian-Albrechts-Universität Kiel, Ludewig-Meyn Str. 4, D-24098 Kiel, Germany

ANA VARGAS: ana.vargas@uam.es
Department of Mathematics, Autonomous University of Madrid, 28049 Madrid, Spain

ANALYSIS AND PDE
Vol. 10, No. 4, 2017
dx.doi.org/10.2140/apde.2017.10.893

ON THE 3-DIMENSIONAL WATER WAVES SYSTEM ABOVE A FLAT BOTTOM

XUECHENG WANG

As a starting point for studying the long-time behavior of the 3-dimensional water waves system in the flat bottom setting, we try to improve the understanding of the Dirichlet–Neumann operator in this set-up. As an application, we study the 3-dimensional gravity waves system and derive a new energy estimate of L^2–L^∞ type, which has good structure in the L^∞-type space. This has been used in our Ph.D. thesis (2016) to prove the global regularity of the 3-dimensional gravity waves system for suitably small initial data.

1. Introduction

1A. ***The full water waves system above a flat bottom.*** We are interested in the long-time behavior of the 3-dimensional water waves system for suitably small initial data in the flat-bottom setting.

The water waves system describes the evolution of an inviscid incompressible fluid with constant density (e.g., water) inside a time-dependent region $\Omega(t)$, which has a free interface $\Gamma(t)$ and a fixed flat bottom Σ. Above the domain $\Omega(t)$, there is a vacuum.

Without loss of generality, we normalize the depth of $\Omega(t)$ to be 1. In the Eulerian coordinate system, we can represent the domain $\Omega(t)$, the interface $\Gamma(t)$ and the bottom Σ as follows:

$$\Omega(t) := \{(x, y) : x \in \mathbb{R}^2,\ -1 \le y \le h(t, x)\},$$
$$\Gamma(t) := \{(x, y) : x \in \mathbb{R}^2,\ y = h(t, x)\}, \quad \Sigma := \{(x, y) : x \in \mathbb{R}^2,\ y = -1\}.$$

We remark that, for the case we are considering, the size of $h(t, \cdot)$ will be small for all time.

We assume that the velocity field is irrotational. The evolution of fluid is subject to the gravity effect or the surface tension effect. We can describe the evolution of fluid by the Euler equation as

$$\begin{cases} \partial_t u + u \cdot \nabla u = -\nabla p - g(0, 0, 1), \\ \nabla \cdot u = 0, \quad \nabla \times u = 0, \quad u(0) = u_0, \end{cases} \tag{1-1}$$

MSC2010: 35Q35, 76B15.
Keywords: 3-dimensional water waves, finite depth, flat bottom, new energy estimate.

where g denotes the constant of the gravity effect.

Moreover, we have the boundary conditions

$$\begin{cases} u\cdot\vec{\boldsymbol{n}}=0 & \text{on } \Sigma,\\ P=\sigma H(h) & \text{on } \Gamma(t),\\ \partial_t+u\cdot\nabla \text{tangents to } \bigcup_t\Gamma(t) & \text{on } \Gamma(t),\end{cases} \tag{1-2}$$

where σ denotes the surface tension coefficient and $H(h)$ denotes the mean curvature of the interface, which is given by

$$H(h)=\nabla\cdot\left(\frac{\nabla h}{\sqrt{1+|\nabla h|^2}}\right).$$

The first boundary condition in (1-2) means that the fluid cannot go through the fixed bottom. The second boundary condition in (1-2) comes from the Young–Laplace equation for the pressure. The third boundary condition in (1-2) represents the kinematic boundary condition, which says that the free interface moves with the normal component of the velocity.

Recall that the velocity field is irrotational. Hence, we can represent it in terms of a velocity potential ϕ. We use ψ to denote the restriction of the velocity potential to the boundary $\Gamma(t)$, i.e., $\psi(t,x):=\phi(t,x,h(t,x))$. From the incompressible condition and the boundary conditions, we can derive the following Laplace equation with two boundary conditions, Neumann-type on the bottom and Dirichlet-type on the interface:

$$(\Delta_x+\partial_y^2)\phi=0,\qquad \frac{\partial\phi}{\partial\vec{\boldsymbol{n}}}\bigg|_{\Sigma}=0,\qquad \phi|_{\Gamma(t)}=\psi. \tag{1-3}$$

Hence, we can reduce (e.g., see [Zakharov 1968]) the motion of fluid inside the water region $\Omega(t)$ to the evolution of the height h and the restricted velocity potential ψ on the interface $\Gamma(t)$:

$$\begin{cases} \partial_t h=G(h)\psi,\\ \partial_t\psi=-gh+\sigma H(h)-\frac{1}{2}|\nabla\psi|^2+\dfrac{(G(h)\psi+\nabla h\cdot\nabla\psi)^2}{2(1+|\nabla h|^2)},\end{cases} \tag{1-4}$$

where $G(h)\psi=\sqrt{1+|\nabla h|^2}\mathcal{N}(h)\psi$ and $\mathcal{N}(h)\psi$ is the Dirichlet–Neumann operator on the interface.

The system (1-4) has the conservation law

$$\mathcal{H}\big(h(t),\psi(t)\big):=\left[\int \tfrac{1}{2}\psi(t)G(h(t))\psi(t)+\tfrac{1}{2}g|h(t)|^2+\frac{\sigma|\nabla h(t)|^2}{1+\sqrt{1+|\nabla h(t)|^2}}\right]=\mathcal{H}\big(h(0),\psi(0)\big).$$

Intuitively speaking, after diagonalizing the system (1-4), we find ourselves dealing with the following type of quasilinear dispersive equation:

$$(\partial_t+i\tilde{\Lambda})u=\mathcal{N}(u,\nabla u),\quad \tilde{\Lambda}=\sqrt{|\nabla|\tanh(|\nabla|)(g+\sigma|\nabla|^2)},\quad u=h+i\tilde{\Lambda}^{-1}|\nabla|\tanh|\nabla|\psi, \tag{1-5}$$

$$u:\mathbb{R}_t\times\mathbb{R}_x^2\to\mathbb{C}. \tag{1-6}$$

Readers can temporarily take (1-5) for granted. It will be much clearer after we obtain the linear term of the Dirichlet–Neumann operator, which is $|\nabla|\tanh|\nabla|\psi$, in Section 3.

1B. ***Motivation and the main result of this paper.*** Note that the best decay rate that one can expect for a 2-dimensional dispersive equation is $1/t$, which is critical in establishing the global regularity for small initial data.

For a 2-dimensional nonlinear dispersive equation, generally speaking, it is crucial to know what the quadratic terms are when studying the long-time behavior of the solution. Unfortunately, to the best of our knowledge, there is no previous work that addresses this issue for the water waves system in the flat-bottom setting. It motivated us to study the problem in this paper.

Identifying the quadratic terms requires a more careful analysis of the Dirichlet–Neumann operator in the flat-bottom setting. Note that the water waves system in the Eulerian coordinate formulation (1-4) is dimensionless. Since we don't want to limit our scope to the 3-dimensional setting, in this paper, we will identify structures inside the Dirichlet–Neumann operator as much as we can.

We summarize and explain several important properties of the Dirichlet–Neumann operator here to help readers understand the discussion of it in this paper. These properties will play important roles in the study of the long-time behavior of the water waves system.

(i) Unlike the infinite-depth setting, in the flat-bottom setting, we do not have the null structure in the low-frequency part. More precisely, if the frequencies of two inputs are 1 and 0 respectively, then the size of the symbol is 1 (flat-bottom setting) instead of 0 (infinite-depth setting).

We remark that the principal symbol of the Dirichlet–Neumann operator in the flat-bottom setting is still the same as in the infinite-depth setting. Intuitively speaking, the high-frequency parts of the Dirichlet–Neumann operator in the two settings are almost the same.

(ii) We give the explicit formula for the quadratic terms of the Dirichlet–Neumann operator, which provides the first step in studying the long-time behavior of (1-5).

(iii) We formulate the cubic and higher-order terms of the Dirichlet–Neumann operator in a fixed-point-type formulation, which provides a good way to control the cubic and higher-order terms over time.

As a starting point and also as an example, we study a specific setting of the water waves system (1-4), which is the gravity water waves system. More precisely, we consider the gravity effect and neglect the surface tension effect. After normalizing the gravity effect constant g to be 1, the system (1-4) is reduced to

$$\begin{cases} \partial_t h = G(h)\psi, \\ \partial_t \psi = -h - \frac{1}{2}|\nabla\psi|^2 + \dfrac{(G(h)\psi + \nabla h\cdot\nabla\psi)^2}{2(1+|\nabla h|^2)}. \end{cases} \tag{1-7}$$

Correspondingly, the diagonalized equation (1-5) is reduced to the quasilinear dispersive equation

$$(\partial_t + i\Lambda)u = \mathcal{N}(u,\nabla u), \quad \Lambda = \sqrt{|\nabla|\tanh|\nabla|}, \quad u = h + i\Lambda\psi. \tag{1-8}$$

For the water waves system in the flat-bottom setting, a typical issue is that the phases are highly degenerate at the low-frequency part. For example, we consider a phase associated with a quadratic term of (1-8),

$$\Lambda(|\xi|) - \Lambda(|\xi-\eta|) + \Lambda(|\eta|) \approx (|\xi| - |\xi-\eta| + |\eta|) - \tfrac{1}{6}(|\xi|^3 - |\xi-\eta|^3 + |\eta|^3), \quad |\eta| \le |\xi| \sim |\xi-\eta| \ll 1.$$

When ξ and $-\eta$ are in the same direction, the above phase is of size $|\xi|^2|\eta|$, which is highly degenerate. Because of this issue, generally speaking, there is no hope to prove the sharp $1/t$ decay rate of the nonlinear solution over time. As a result, a rough energy estimate is not sufficient to control the growth of energy in the long run. However, it turns out that there is a relatively simple way to control the growth of energy. It relies on two observations about the system (1-7):

(i) We can derive a new energy estimate of L^2-L^∞ type after carefully analyzing the structures inside the quadratic terms in (1-7). The input inside the quadratic terms is, roughly speaking, not put in L^∞ but rather in a weaker L^∞-type space, which has derivatives in front. See (1-12).

(ii) The low-frequency parts of the derivatives compensate for the decay rate of the solution of (1-7). We can prove that the solution with some derivatives in front decays sharply, despite the fact that the solution itself may not have the sharp decay rate. The proof of this fact involves a very delicate Fourier analysis. Interested readers are referred to [Wang 2016] for more details.

Before stating our main result, we define the function spaces

$$\|f\|_{\widetilde{W}^\gamma} := \sum_{k\geq 0, k\in\mathbb{Z}} 2^{\gamma k}\|P_k f\|_{L^\infty} + \|P_{\leq 0} f\|_{L^\infty}, \tag{1-9}$$

$$\|f\|_{\widehat{W}^{\gamma,\alpha}} := \sum_{k\in\mathbb{Z}} (2^{\alpha k} + 2^{\gamma k})\|P_k f\|_{L^\infty}, \quad 0\leq\alpha\leq\gamma, \qquad \|f\|_{\widehat{W}^\gamma} := \|f\|_{\widehat{W}^{\gamma,0}}. \tag{1-10}$$

Theorem 1.1. *Let $0<\delta<c$, $\alpha\in(0,1]$, and $N_0\geq 6$, where c is some sufficiently small constant. If the initial data $(h_0,\Lambda\psi_0)\in H^{N_0+1/2}(\mathbb{R}^2)\times H^{N_0}(\mathbb{R}^2)$ satisfies the smallness condition*

$$\|(h_0,\Lambda\psi_0)\|_{\widetilde{W}^4}\leq\delta, \tag{1-11}$$

then there exists $T>0$ such that the system (1-7) has the unique solution

$$(h,\Lambda\psi)\in C^0\big([0,T]; H^{N_0}(\mathbb{R}^2)\times H^{N_0}(\mathbb{R}^2)\big).$$

Moreover, we have a new type of energy estimate in the time interval of existence:

$$\frac{d}{dt}E_{N_0}(t) \lesssim_{N_0} \big[\|(h,\Lambda\psi)(t)\|_{\widehat{W}^{4,\alpha}} + \|(h,\Lambda\psi)(t)\|^2_{\widehat{W}^4}\big]E_{N_0}(t), \tag{1-12}$$

where the energy $E_{N_0}(t)$ is defined in (5-3). The size of energy is comparable to $\|(h,\Lambda\psi)(t)\|^2_{H^{N_0}}$.

Remark 1.2. Note that smallness condition is not assumed in [Alazard, Burq and Zuily 2011; 2014a; 2014b; Lannes 2005] to derive the local wellposedness. For the purpose of obtaining a global solution, we impose the smallness condition (1-11) to derive our desired estimate (1-12), which is the first step to obtaining global existence for small initial data.

In [Wang 2016], based on the results we obtained in this paper, we show that the solution of the system (1-7) exists globally and scatters to a linear solution. We will study the long-time behavior of the water waves system (1-4) in other settings in the future. For example, do we still have global solutions if only the surface tension is effective or both the gravity and the surface tension are effective? We expect that the results we obtained in this paper will be very helpful to the future study of the water waves system in the flat-bottom setting.

1C. ***Previous results.*** To be concise, we mainly discuss work on the local behavior of the water waves system in this subsection. For a more detailed discussion on the long-time behavior, please refer to the introduction of [Wang 2016].

Starting with [Nalimov 1974] and [Yosihara 1982], there has been a considerable amount of work on the local theory of the water waves system. In the framework of Sobolev spaces and without smallness assumptions on the initial data, the local wellposedness was first obtained by Wu [1997; 1999] for the gravity waves system. The local wellposedness was also obtained when the surface tension is effective by Beyer and Günther [1998]. Later, different methods were developed and many important results were obtained to improve our understanding of the local behavior of the water waves system. Among them, we mention [Christodoulou and Lindblad 2000; Ambrose and Masmoudi 2005; Lannes 2005; Shatah and Zeng 2008; Coutand and Shkoller 2007; Alazard, Burq and Zuily 2011; 2014a; 2014b].

Roughly speaking, the local existence for the water waves system (1-4) holds even when the initial interface has an unbounded curvature and the bottom is very rough. A fixed-length separation between the interface and the bottom is sufficient. See [Alazard, Burq and Zuily 2011; 2014a; 2014b; Lannes 2005] for more details and more precise descriptions.

1D. ***Main ideas and the outline of this paper.*** To prove our main theorem, we have to pay attention to both the low- and high-frequency parts.

For the high-frequency part, due to the quasilinear nature of the gravity waves system (1-7), we have to get around the difficulty of losing one derivative. Thanks to [Lannes 2005; Alazard and Métivier 2009; Alazard, Burq and Zuily 2011; 2014a; 2014b], we can utilize the paralinearization method to get around the potential loss of one derivative. However, for their purposes, only the high-frequency part has been carefully studied in their works. In this paper, we will do the paralinearization process and pay special attention to the low-frequency part at the same time.

For the low-frequency part, more careful estimates of the Dirichlet–Neumann operator are essential since it is not straightforward to see the fact that we can gain α derivatives for input in $\widehat{W}^{4,\alpha}$. For example, for the quadratic term $\nabla h \cdot \nabla \psi$ of the Dirichlet–Neumann operator, it is problematic to gain α derivatives when ψ has smaller frequency because the total number of derivatives of ψ in (1-12) is $1+\alpha$ in the low-frequency part when the input ψ of the quadratic terms is in L^∞.

To conclude the argument, we will use the hidden structure inside the system (1-7) for different scenarios. Without describing too many details, we give two examples as follows to explain the main ideas:

(i) When ψ has a smaller frequency inside $\nabla h \cdot \nabla \psi$, we can use the hidden symmetry to move one derivative from ∇h to $\nabla \psi$ during the energy estimate; hence we have two derivatives in total for ψ.

(ii) For some terms, e.g., the good remainder term of the paralinearization process, we can lower their regularities to L^2. Hence, we can put $\nabla \psi$ in L^2 and put ∇h in L^∞; as a result the desired estimate (1-12) also holds for this case.

Outline: In Section 2, we introduce notation and give a quick summary of paradifferential calculus. In Section 3, we study various properties of the Dirichlet–Neumann operator. In Section 4, we use the paralinearization method to show the good structures inside the system (1-7), which help us find good substitution variables. In Section 5, we prove the new energy estimate (1-12) by using the symmetries inside the equations satisfied by the good substitution variables. In the Appendix, we calculate explicitly the quadratic terms of good remainder terms. This is intended to help readers understand the fact that we can gain α derivatives in (1-12) for an input of quadratic terms, which lies in the L^∞-type space.

2. Preliminaries

2A. *Notation.* For any two numbers A and B, we use $A \lesssim B$ and $B \gtrsim A$ to denote $A \le CB$, where C is an absolute constant. We use $A \lesssim_\epsilon B$ to denote $A \le C_\epsilon B$, where the constant C_ϵ depends on ϵ. For an integer $k \in \mathbb{Z}$, we use k_+ to denote $\max\{k, 0\}$ and use k_- to denote $\min\{k, 0\}$.

Throughout this paper, we will abuse the notation of Λ. When there is no lower script associated with Λ, we let $\Lambda := \sqrt{\tanh(|\nabla|)|\nabla|}$, which is the linear operator associated with the system (1-8). For $p \in \mathbb{N}_+$, we use $\Lambda_p(\mathcal{N})$ to denote the p-th order terms of a nonlinearity $\mathcal{N}$ when a Taylor expansion for the nonlinearity $\mathcal{N}$ is available. For example, $\Lambda_2[\mathcal{N}]$ denotes the quadratic term of $\mathcal{N}$. We also use $\Lambda_{\ge p}[\mathcal{N}]$ to denote the p-th and higher-order terms. More precisely, $\Lambda_{\ge p}[\mathcal{N}] := \sum_{q \ge p} \Lambda_q[\mathcal{N}]$. In this paper, the Taylor expansion and $\Lambda_p[\cdot]$ are in terms of h and ψ when there is no special annotation.

We fix an even smooth function $\tilde{\psi} : \mathbb{R} \to [0, 1]$, which is supported in $\left[-\frac{3}{2}, \frac{3}{2}\right]$ and is equal to 1 in $\left[-\frac{5}{4}, \frac{5}{4}\right]$. For any $k \in \mathbb{Z}$, define

$$\psi_k(x) := \tilde{\psi}(x/2^k) - \tilde{\psi}(x/2^{k-1}), \quad \psi_{\le k}(x) := \tilde{\psi}(x/2^k), \quad \psi_{\ge k}(x) := 1 - \psi_{\le k-1}(x).$$

Denote the projection operators P_k, $P_{\le k}$ and $P_{\ge k}$ by the Fourier multipliers ψ_k, $\psi_{\le k}$ and $\psi_{\ge k}$ respectively. For a well-defined function f, we will also use the notation f_k to abbreviate $P_k f$.

The Fourier transform is defined as

$$\mathcal{F}(f)(\xi) = \int_{\mathbb{R}^2} e^{-ix\cdot\xi} f(x)\, dx.$$

For two well-defined functions f and g and a bilinear form $Q(f, g)$, we will use the convention that the symbol $q(\cdot\,,\cdot)$ of $Q(\cdot\,,\cdot)$ is defined in the following sense throughout this paper:

$$\mathcal{F}[Q(f, g)](\xi) = \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \hat{f}(\xi-\eta)\hat{g}(\eta) q(\xi-\eta, \eta)\, d\eta. \tag{2-1}$$

Meanwhile, for a trilinear form $C(f, g, h)$, its symbol $c(\cdot\,,\cdot\,,\cdot)$ is defined in the following sense:

$$\mathcal{F}[C(f, g, h)](\xi) = \frac{1}{16\pi^4} \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \hat{f}(\xi-\eta)\hat{g}(\eta-\sigma)\hat{h}(\sigma) c(\xi-\eta, \eta-\sigma, \sigma)\, d\eta\, d\sigma.$$

2B. *Multilinear estimate.* We define a class of symbols with an associated norm as

$$\mathcal{S}^\infty := \left\{m : \mathbb{R}^4 \text{ or } \mathbb{R}^6 \to \mathbb{C},\ m \text{ is continuous and } \|\mathcal{F}^{-1}(m)\|_{L^1} < \infty\right\},$$

$$\|m\|_{\mathcal{S}^\infty} := \|\mathcal{F}^{-1}(m)\|_{L^1}, \quad \|m(\xi, \eta)\|_{\mathcal{S}^\infty_{k,k_1,k_2}} := \left\|m(\xi, \eta)\psi_k(\xi)\psi_{k_1}(\xi-\eta)\psi_{k_2}(\eta)\right\|_{\mathcal{S}^\infty},$$

$$\|m(\xi, \eta, \sigma)\|_{\mathcal{S}^\infty_{k,k_1,k_2,k_3}} := \left\|m(\xi, \eta, \sigma)\psi_k(\xi)\psi_{k_1}(\xi-\eta)\psi_{k_2}(\eta-\sigma)\psi_{k_3}(\sigma)\right\|_{\mathcal{S}^\infty}.$$

Lemma 2.1. *Assume that $m, m' \in \mathcal{S}^\infty$ and $p, q, r, s \in [1, \infty]$. Then the estimates*

$$\|m \cdot m'\|_{\mathcal{S}^\infty} \lesssim \|m\|_{\mathcal{S}^\infty} \|m'\|_{\mathcal{S}^\infty}, \tag{2-2}$$

$$\left\| \mathcal{F}^{-1}\left[\int_{\mathbb{R}^2} m(\xi,\eta) \hat{f}(\xi-\eta)\hat{g}(\eta)\, d\eta\right]\right\|_{L^p} \lesssim \|m\|_{\mathcal{S}^\infty} \|f\|_{L^q} \|g\|_{L^r} \quad \text{if } \tfrac{1}{p} = \tfrac{1}{q} + \tfrac{1}{r}, \tag{2-3}$$

$$\left\| \mathcal{F}^{-1}\left[\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} m'(\xi,\eta,\sigma) \hat{f}(\xi-\eta)\hat{h}(\sigma)\hat{g}(\eta-\sigma)\, d\eta\, d\sigma\right]\right\|_{L^p} \lesssim \|m'\|_{\mathcal{S}^\infty} \|f\|_{L^q} \|g\|_{L^r} \|h\|_{L^s} \tag{2-4}$$

hold for well-defined functions $f(x)$, $g(x)$, and $h(x)$, where $\frac{1}{p} = \frac{1}{q} + \frac{1}{r} + \frac{1}{s}$.

To estimate the $\mathcal{S}^\infty_{k,k_1,k_2}$ norm and the $\mathcal{S}^\infty_{k,k_1,k_2,k_3}$ norm of symbols, we repeatedly use the following:

Lemma 2.2. *For $i \in \{1,2,3\}$, if $f : \mathbb{R}^{2i} \to \mathbb{C}$ is a smooth function and $k_1, \ldots, k_i \in \mathbb{Z}$, then we have the estimate*

$$\left\| \int_{\mathbb{R}^{2i}} f(\xi_1,\ldots,\xi_i) \prod_{j=1}^{i} e^{i x_j \cdot \xi_j} \psi_{k_j}(\xi_j)\, d\xi_1 \cdots d\xi_i \right\|_{L^1_{x_1,\ldots,x_i}} \lesssim \sum_{m=0}^{i+1} \sum_{j=1}^{i} 2^{m k_j} \|\partial_{\xi_j}^m f\|_{L^\infty}. \tag{2-5}$$

Proof. The cases when $i = 1, 3$ can be estimated in the same way as the case when $i = 2$. We only do the case $i = 2$ in detail here. Through scaling, it is sufficient to prove the above estimate for the case when $k_1 = k_2 = 0$. From Plancherel's theorem, we have the two estimates

$$\left\| \int_{\mathbb{R}^{2i}} f(\xi_1,\xi_2) e^{i(x_1\cdot\xi_1 + x_2\cdot\xi_2)} \psi_0(\xi_1)\psi_0(\xi_2)\, d\xi_1\, d\xi_2 \right\|_{L^2_{x_1,x_2}} \lesssim \|f(\xi_1,\xi_2)\|_{L^\infty_{\xi_1,\xi_2}},$$

$$\left\| (|x_1|+|x_2|)^3 \int_{\mathbb{R}^{2i}} f(\xi_1,\xi_2) e^{i(x_1\cdot\xi_1 + x_2\cdot\xi_2)} \psi_0(\xi_1)\psi_0(\xi_2)\, d\xi_1\, d\xi_2 \right\|_{L^2_{x_1,x_2}} \lesssim \sum_{m=0}^{3} \left[\|\partial_{\xi_1}^m f\|_{L^\infty} + \|\partial_{\xi_2}^m f\|_{L^\infty}\right],$$

which are sufficient to finish the proof of (2-5). □

2C. *Paradifferential calculus.* In this subsection, we discuss some necessary background material from paradifferential calculus. For more details and related topics, please refer to [Métivier 2008].

Definition 2.3. Given $\rho \in \mathbb{N}_+$, $\rho \geq 0$ and $m \in \mathbb{R}$, we use $\Gamma^m_\rho(\mathbb{R}^2)$ to denote the space of locally bounded functions $a(x,\xi)$ on $\mathbb{R}^2 \times (\mathbb{R}^2/\{0\})$, which are C^∞ with respect to ξ for $\xi \neq 0$. Moreover, they satisfy the estimate

$$\forall |\xi| \geq \tfrac{1}{2}, \quad \|\partial_\xi^\alpha a(\cdot,\xi)\|_{W^{\rho,\infty}} \lesssim_\alpha (1+|\xi|)^{m-|\alpha|}, \quad \alpha \in \mathbb{N}^2,$$

where $W^{\rho,\infty}$ is the usual Sobolev space. Note that $W^{\rho,\infty}$ contains the spaces $\widetilde{W}^\rho$ and $\widehat{W}^{\rho,\alpha}$, which are defined in (1-9) and (1-10), as subspaces.

Remark 2.4. In the above definitions, ρ is not necessarily an integer, but the integer case is sufficient for our purposes.

Definition 2.5. (i) We use $\dot{\Gamma}^m_\rho(\mathbb{R}^2)$ to denote the subspace of $\Gamma^m_\rho(\mathbb{R}^2)$ which consists of symbols that are homogeneous of degree m in ξ.

(ii) If $a = \sum_{0 \leq j < \rho} a^{(m-j)}$, where $a^{(m-j)} \in \dot{\Gamma}^{m-j}_{\rho-j}(\mathbb{R}^2)$, then we say $a^{(m)}$ is the principal symbol of a.

(iii) An operator T is said to be of order m, where $m \in \mathbb{R}$, if for all $\mu \in \mathbb{R}$, it is bounded from $H^{\mu}(\mathbb{R}^2)$ to $H^{\mu-m}(\mathbb{R}^2)$. We use S^m to denote the set of all operators of order m.

For a symbol $a \in \Gamma_\rho^m$, we can define its norm as

$$M_\rho^m(a) := \sup_{|\alpha|\le 2+\rho} \sup_{|\xi|\ge \frac{1}{2}} \left\| (1+|\xi|)^{|\alpha|-m} \partial_\xi^\alpha a(\cdot,\xi) \right\|_{W^{\rho,\infty}}.$$

For $a, f \in L^2$ and a pseudodifferential operator $\tilde{a}(x,\xi)$, we define the operators $T_a f$ and $T_{\tilde{a}} f$ as

$$T_a f = \mathcal{F}^{-1}\left[\int_{\mathbb{R}} \hat{a}(\xi-\eta)\theta(\xi-\eta,\eta)\hat{f}(\eta)\,d\eta\right], \quad T_{\tilde{a}} f = \mathcal{F}^{-1}\left[\int_{\mathbb{R}} \mathcal{F}_x(\tilde{a})(\xi-\eta,\eta)\theta(\xi-\eta)\hat{f}(\eta)\,d\eta\right], \tag{2-6}$$

where the cut-off function $\theta(\xi-\eta,\eta)$ is given by

$$\theta(\xi-\eta,\eta) = \begin{cases} 1 & \text{when } |\xi-\eta| \le 2^{-10}|\eta| \text{ and } |\eta| \ge 1, \\ 0 & \text{when } |\xi-\eta| \ge 2^{10}|\eta| \text{ or } |\eta| \le 1. \end{cases}$$

For two well-defined functions a and b, we have the paraproduct decomposition

$$ab = T_a b + T_b a + \mathcal{R}(a,b), \tag{2-7}$$

where $\mathcal{R}(a,b)$ contains those terms in which a and b have comparable size of frequencies or the frequency of the output is less than 1.

We have the following composition lemma for paradifferential operators. It can be found, for example, in [Alazard, Burq and Zuily 2011; Métivier 2008].

Lemma 2.6. *Let $m \in \mathbb{R}$ and $\rho > 0$. If given symbols $a \in \Gamma_\rho^m(\mathbb{R}^d)$ and $b \in \Gamma_\rho^{m'}(\mathbb{R}^d)$ we define*

$$a \sharp b = \sum_{|\alpha|<\rho} \frac{1}{i^{|\alpha|}\alpha!} \partial_\xi^\alpha a\, \partial_x^\alpha b,$$

then for all $\mu \in \mathbb{R}$, there exists a constant K such that

$$\|T_a T_b - T_{a\sharp b}\|_{H^\mu \to H^{\mu-m-m'+\rho}} \le K M_\rho^m(a) M_\rho^{m'}(b). \tag{2-8}$$

Remark 2.7. It may be too early to give this remark here. However, we think that it is a good idea to keep the following simple observation in mind, which will be very helpful to see the equivalence relations later on. The simple observation is that if the symbols a and b all depend on ∇h instead of h, then the rough estimate (2-8) is sufficient to gain one derivative in the low-frequency part.

Lemma 2.8. *Let $m \in \mathbb{R}$, $\rho > 0$ and $a \in \Gamma_\rho^m(\mathbb{R}^d)$. If we use $(T_a)^*$ to denote the adjoint operator of T_a and use $\bar{a}$ to denote the complex conjugate of a, then $(T_a)^* - T_{a^*}$ is of order $m-\rho$, where*

$$a^* = \sum_{|\alpha|<\rho} \frac{1}{i^{|\alpha|}\alpha!} \partial_\xi^\alpha \partial_x^\alpha \bar{a}.$$

Moreover, the norm of the operator $(T_a)^* - T_{a^*}$ is bounded by $M_\rho^m(a)$.

Proof. See [Alazard, Burq and Zuily 2011, Theorem 3.10]. □

Remark 2.9. In most applications of Lemma 2.8, we have $m \le 1$. If we let $\rho = 1$ in the above lemma, then it is easy to see $a^* = \bar{a}$. If, moreover, a is real, then $a^* = \bar{a} = a$.

3. Dirichlet–Neumann operator

The main goal of this section is to study various properties of the Dirichlet–Neumann operator, which provide a foundation for carrying out the processes of paralinearization and symmetrization in Section 4 and obtaining the new energy estimate (1-12) in Section 5. The study of the Dirichlet–Neumann operator is mainly reduced to a study of the velocity potential inside the water region $\Omega(t)$.

Recall the smallness condition (1-11) of the initial data. From the local wellposedness result of the gravity waves system (1-7), we know that there exists a positive time T such that the estimate

$$\sup_{t\in[0,T]} \|(h,\Lambda\psi)(t)\|_{\widetilde{W}^4} \le 2\delta \tag{3-1}$$

holds, which means that the L^∞ norm of solution remains small in the time interval $[0,T]$. Throughout the rest of this paper, we restrict ourselves to the time interval $[0,T]$.

3A. ***Type I formulation of the Laplace equation (1-3).*** In this subsection, we reduce the Laplace equation (1-3) to a favorable formulation so that we can solve it and identify the fixed-point-type structure inside the Laplace equation, which further enables us to estimate the Dirichlet–Neumann operator.

We do a change of variables and map the water region $\Omega(t)$ to the strip $\mathcal{S} := \mathbb{R}^2\times[-1,0]$ using

$$(x,y)\to(x,z),\qquad z:=\frac{y-h(t,x)}{h(t,x)+1}.$$

Very naturally, the inverse transformation is given by

$$y = h+(h+1)z.$$

Define the velocity potential in the (x,z)-coordinate system as $\varphi(x,z):=\phi(x,h+(h+1)z)$. From direct computations, we have the identities

$$\phi(x,y)=\varphi\Big(x,\frac{y-h}{1+h}\Big),\qquad \partial_y\phi=\frac{\partial_z\varphi}{1+h},\qquad \partial_y^2\phi=\frac{\partial_z^2\varphi}{(1+h)^2}, \tag{3-2}$$

$$\partial_{x_i}\phi=\partial_{x_i}\varphi+\partial_z\varphi\bigg[\frac{-\partial_{x_i}h}{1+h}-\frac{(y-h)\partial_{x_i}h}{(1+h)^2}\bigg]=\partial_{x_i}\varphi-\frac{(y+1)\partial_{x_i}h}{(1+h)^2}\partial_z\varphi, \tag{3-3}$$

$$\partial_{x_i}^2\phi=\partial_{x_i}^2\varphi-2\frac{(y+1)\partial_{x_i}h}{(1+h)^2}\partial_z\partial_{x_i}\varphi+\bigg[\frac{-(y+1)\partial_{x_i}^2h}{(1+h)^2}+2\frac{(y+1)(\partial_{x_i}h)^2}{(1+h)^3}\bigg]\partial_z\varphi+\frac{(y+1)^2(\partial_{x_i}h)^2}{(1+h)^4}\partial_z^2\varphi.$$

From the above identities and (1-3), it is easy to derive the equation

$$(\Delta_x+\partial_y^2)\phi=0 \quad\Longrightarrow\quad P_{x,z}\varphi:=\big[\Delta_x+\tilde a\partial_z^2+\tilde b\cdot\nabla\partial_z+\tilde c\partial_z\big]\varphi=0, \tag{3-4}$$

where

$$\tilde a=\frac{(y+1)^2|\nabla h|^2}{(1+h)^4}+\frac{1}{(1+h)^2}=\frac{1+(z+1)^2|\nabla h|^2}{(1+h)^2}, \tag{3-5}$$

$$\tilde b=-2\frac{(y+1)\nabla h}{(1+h)^2}=\frac{-2(z+1)\nabla h}{1+h},\qquad \tilde c=\frac{-(z+1)\Delta_x h}{(1+h)}+2\frac{(z+1)|\nabla h|^2}{(1+h)^2}. \tag{3-6}$$

To sum up, we can reduce the Laplace equation (1-3) with two boundary conditions in terms of φ as follows:

$$P_{x,z}\varphi = 0, \quad \varphi|_{z=0} = \psi, \quad \partial_z\varphi|_{z=-1} = 0, \quad (x,z) \in \mathbb{R}^2 \times [-1,0]. \tag{3-7}$$

3B. ***Type II formulation of the Laplace equation* (1-3).** In this subsection, we reduce the Laplace equation (1-3) into another favorable formulation, which will be used to do the paralinearization of the Dirichlet–Neumann operator in Section 4A.

We remark that we don't use the type I formulation (3-7) to do the paralinearization process because the coefficients $\tilde{a}, \tilde{b}, \tilde{c}$ in (3-5) and (3-6) are very complicated, which complicates the paralinearization process and prevents us from seeing clearly the principal symbol of the Dirichlet–Neumann operator.

Recall the smallness condition (3-1). Since the height of interface is very small, we know that there exists a curve parallel to the interface $\Gamma(t)$ with depth $\frac{1}{2}$ inside $\Omega(t)$. More precisely, we have

$$\Omega_1(t) := \left\{(x,y) : x \in \mathbb{R}^2,\ h(t,x) - \tfrac{1}{2} \le y \le h(t,x)\right\}, \quad \Omega_1(t) \subset \Omega(t).$$

Define

$$\Omega_2(t) := \left\{(x,y) : x \in \mathbb{R}^2, h(t,x) - \tfrac{1}{4} \le y \le h(t,x)\right\}, \quad \Omega_2(t) \subset \Omega_1(t) \subset \Omega(t), \tag{3-8}$$

$$\tilde{\phi}(x,y) := \chi(y - h(t,x))\phi(x,y), \quad (x,y) \in \Omega_1(t), \quad \chi(z) = 1 \text{ if } z \ge -\tfrac{1}{4}, \quad \operatorname{supp}(\chi) \subset \left[-\tfrac{1}{2}, 0\right], \tag{3-9}$$

where $\chi(x)$ is a fixed Schwartz function.

Recall the Laplace equation (1-3). From (3-9), it is easy to derive the identities

$$\begin{aligned} &\Delta_{x,y}\tilde{\phi} = \tilde{g} := \Delta_{x,y}[\chi\phi] - \chi\Delta_{x,y}\phi, \quad (x,y) \in \Omega_1(t), \\ &\tilde{\phi}(x,y) = \phi(x,y), \quad \tilde{g}(x,y) = 0, \quad (x,y) \in \Omega_2(t). \end{aligned} \tag{3-10}$$

We can map the water region $\Omega_1(t)$ to the strip $\mathcal{S}' := \mathbb{R}^2 \times \left[-\frac{1}{2}, 0\right]$ by changing the coordinate system using

$$(x,y) \to (x,w), \quad w := y - h(t,x).$$

Define the velocity potential in the (x,w)-coordinate system as $\Phi(x,w) := \tilde{\phi}(x, \omega + h(t,x))$. Hence $\tilde{\phi}(x,y) = \Phi(x, y - h(t,x))$. From (3-10), it is easy to verify that the equality

$$\mathcal{P}_{x,w}\Phi := \left[\Delta_x + a'\partial_w^2 + b' \cdot \nabla\partial_w + c'\partial_w\right]\Phi = g'(x,w) := \tilde{g}(x, \omega + h(t,x)) \tag{3-11}$$

holds, where

$$a' = 1 + |\nabla h|^2, \quad b' = -2\nabla h, \quad c' = -\Delta h. \tag{3-12}$$

Remark 3.1. From (3-5), (3-6), and (3-12), it is easy to see that the coefficients in (3-11) satisfied by Φ are much easier and more favorable than the coefficients in (3-4) satisfied by φ. However, the formulation satisfied by Φ in (3-11) cannot be used as the starting point because we don't know the estimates of Φ in the first place.

From the above definitions, the following identities hold inside the water region $\Omega_2(t)$, see (3-8), and the corresponding regions in the new coordinate systems:

$$\Phi(x,w)=\varphi\Big(x,\frac{w}{1+h}\Big),\quad \varphi(x,z)=\Phi(x,(1+h)z),\quad (x,w)\in\mathbb{R}^2\times\big[-\tfrac{1}{4},0\big],$$
$$\partial_{x_i}\Phi=\partial_{x_i}\varphi-\frac{w\partial_z\varphi\partial_{x_i}h}{(1+h)^2},\quad \partial_w\Phi=\frac{\partial_z\varphi}{1+h}. \tag{3-13}$$

From (3-2) and (3-13), the Dirichlet–Neumann operator $G(h)\psi$ in terms of φ and Φ and the quadratic terms of $G(h)\psi$ are given by

$$G(h)\psi=[-\nabla h\cdot\nabla\phi+\partial_y\phi]|_{y=h}=\frac{1+|\nabla h|^2}{1+h}\partial_z\varphi|_{z=0}-\nabla\psi\cdot\nabla h, \tag{3-14}$$
$$G(h)\psi=(1+|\nabla h|^2)\partial_w\Phi|_{w=0}-\nabla h\cdot\nabla\psi, \tag{3-15}$$
$$\Lambda_2[G(h)\psi]=\Lambda_2[\partial_z\varphi|_{z=0}]-\Lambda_1[\partial_z\varphi|_{z=0}]h-\nabla\psi\cdot\nabla h. \tag{3-16}$$

3C. *A fixed-point-type formulation for the Dirichlet–Neumann operator.* In this subsection, our main goal is to obtain basic estimates for the Dirichlet–Neumann operator with special attention to the low-frequency part, which will further help us to obtain a new energy estimate.

To this end, we study the reduced Laplace equation (3-7) and formulate $\nabla_{x,z}\varphi$ into a fixed-point-type formulation, which enables us to use a fixed-point-type argument.

After moving all nonlinear terms to the right-hand, we can rewrite (3-7) as

$$\partial_z^2\varphi+\Delta_x\varphi=(\partial_z-|\nabla|)(\partial_z+|\nabla|)\varphi=g(z):=(1-\tilde{a})\partial_z^2\varphi-\tilde{b}\cdot\nabla\partial_z\varphi-\tilde{c}\partial_z\varphi. \tag{3-17}$$

Now, we will solve $\varphi(z)$ from (3-17) by treating $g(z)$ in (3-17) as a given nonlinearity. Define $\tilde{h}(x,z):=(\partial_z-|\nabla|)\varphi$. Very naturally, we have

$$\begin{cases}(\partial_z+|\nabla|)\tilde{h}=g,\\ (\partial_z-|\nabla|)\varphi=\tilde{h},\quad \varphi|_{z=0}=\psi,\ \partial_z\varphi|_{z=-1}=0.\end{cases} \tag{3-18}$$

We can solve the above system of equations with $\tilde{h}(-1)$ to be determined:

$$\tilde{h}(z)=e^{-z|\nabla|}\tilde{h}(-1)+\int_{-1}^{z}e^{-(z-z')|\nabla|}g(z')\,dz', \tag{3-19}$$

$$\begin{aligned}\varphi(z)&=e^{z|\nabla|}\varphi(0)+\int_0^z e^{(z-z')|\nabla|}\tilde{h}(z')\,dz'\\
&=e^{z|\nabla|}\psi-\int_z^0 e^{(z-z')|\nabla|}\Big[e^{-z'|\nabla|}\tilde{h}(-1)+\int_{-1}^{z'}e^{-(z'-s)|\nabla|}g(s)\,ds\Big]\,dz'\\
&=e^{z|\nabla|}\psi-\tfrac{1}{2}|\nabla|^{-1}[e^{-z|\nabla|}-e^{z|\nabla|}]\tilde{h}(-1)-\int_{-1}^{z}\int_z^0 e^{(z+s-2z')|\nabla|}g(s)\,dz'\,ds\\
&\qquad\qquad-\int_z^0\int_s^0 e^{(z+s-2z')|\nabla|}g(s)\,dz'\,ds\\
&=e^{z|\nabla|}\psi-\tfrac{1}{2}|\nabla|^{-1}[e^{-z|\nabla|}-e^{z|\nabla|}]\tilde{h}(-1)+\frac{1}{2}\int_{-1}^{0}|\nabla|^{-1}e^{(z+s)|\nabla|}g(s)\,ds\\
&\qquad\qquad-\frac{1}{2}\int_{-1}^{0}|\nabla|^{-1}e^{-|z-s||\nabla|}g(s)\,ds.\end{aligned} \tag{3-20}$$

The unknown $\tilde{h}(-1)$ is determined by the Neumann-type boundary condition $\partial_z\varphi|_{z=-1}=0$. We calculate $\partial_z\varphi$ from the formula (3-20) and have the equality

$$\partial_z\varphi=|\nabla|e^{z|\nabla|}\psi+\tfrac{1}{2}[e^{z|\nabla|}+e^{-z|\nabla|}]\tilde{h}(-1)+\frac{1}{2}\int_{-1}^{0}e^{(z+s)|\nabla|}g(s)\,ds-\frac{1}{2}\int_{-1}^{0}e^{-|z-s||\nabla|}\operatorname{sign}(s-z)g(s)\,ds.$$

After evaluating the above equality at the point $z=-1$, we have

$$\tilde{h}(-1)=-\frac{2|\nabla|e^{-|\nabla|}\psi}{e^{-|\nabla|}+e^{|\nabla|}}-\int_{-1}^{0}\frac{e^{(s-1)|\nabla|}-e^{-(s+1)|\nabla|}}{e^{-|\nabla|}+e^{|\nabla|}}g(s)\,ds, \tag{3-21}$$

which further gives us

$$\begin{aligned}\partial_z\varphi=\frac{e^{(z+1)|\nabla|}-e^{-(z+1)|\nabla|}}{e^{-|\nabla|}+e^{|\nabla|}}|\nabla|\psi-\frac{1}{2}\frac{e^{z|\nabla|}+e^{-z|\nabla|}}{e^{-|\nabla|}+e^{|\nabla|}}\int_{-1}^{0}[e^{(s-1)|\nabla|}-e^{-(s+1)|\nabla|}]g(s)\,ds\\+\frac{1}{2}\int_{-1}^{0}e^{(z+s)|\nabla|}g(s)ds-\frac{1}{2}\int_{-1}^{0}e^{-|z-s||\nabla|}\operatorname{sign}(s-z)g(s)\,ds.\end{aligned} \tag{3-22}$$

Moreover, we can reduce (3-20):

$$\begin{aligned}\varphi(z)=\left[\frac{e^{-(z+1)|\nabla|}+e^{(z+1)|\nabla|}}{e^{-|\nabla|}+e^{|\nabla|}}\right]\psi+\frac{1}{2}|\nabla|^{-1}\frac{e^{-z|\nabla|}-e^{z|\nabla|}}{e^{-|\nabla|}+e^{|\nabla|}}\int_{-1}^{0}[e^{(s-1)|\nabla|}-e^{-(s+1)|\nabla|}]g(s)\,ds\\+\frac{1}{2}\int_{-1}^{0}|\nabla|^{-1}e^{(z+s)|\nabla|}g(s)\,ds-\frac{1}{2}\int_{-1}^{0}|\nabla|^{-1}e^{-|z-s||\nabla|}g(s)\,ds.\end{aligned} \tag{3-23}$$

However, we cannot use the formulation (3-23) to estimate the velocity potential and the Dirichlet–Neumann operator because $g(z)$ actually depends on the velocity potential $\varphi(z)$; see (3-17).

To get around this issue, we observe that there exists a fixed-point-type structure inside $g(z)$. Recall (3-17), (3-5), and (3-6). Note that

$$g=\partial_z\left[\frac{2h+h^2-(z+1)^2|\nabla h|^2}{(1+h)^2}\partial_z\varphi+\frac{2(z+1)\nabla h\cdot\nabla\varphi}{1+h}\right]-\frac{2\nabla h\cdot\nabla\varphi}{1+h}+\frac{(z+1)\Delta h}{1+h}\partial_z\varphi,$$

and

$$\frac{(z+1)\Delta h}{1+h}\partial_z\varphi=\nabla\cdot\left[\frac{(z+1)\nabla h\partial_z\varphi}{1+h}\right]+\frac{(z+1)|\nabla h|^2\partial_z\varphi}{(1+h)^2}-\partial_z\left[\frac{(z+1)\nabla h\cdot\nabla\varphi}{1+h}\right]+\frac{\nabla h\cdot\nabla\varphi}{1+h}.$$

Hence, we can decompose the nonlinearity $g(z)$ into three parts:

$$g(z)=\partial_z g_1(z)+g_2(z)+\nabla\cdot g_3(z), \tag{3-24}$$

where

$$g_1(z)=\frac{2h+h^2-(z+1)^2|\nabla h|^2}{(1+h)^2}\partial_z\varphi+\frac{(z+1)\nabla h\cdot\nabla\varphi}{1+h},\quad g_1(-1)=0, \tag{3-25}$$

$$g_2(z)=\frac{(z+1)|\nabla h|^2\partial_z\varphi}{(1+h)^2}-\frac{\nabla h\cdot\nabla\varphi}{1+h},\quad g_3(z)=\frac{(z+1)\nabla h\partial_z\varphi}{1+h}. \tag{3-26}$$

To simplify the notation, we define

$$\tilde{h}_1 := \frac{2h+h^2}{(1+h)^2}, \quad \tilde{h}_2 := \frac{|\nabla h|^2}{(1+h)^2}, \quad \tilde{h}_3 := \frac{\nabla h}{1+h}. \tag{3-27}$$

As a result, we have

$$g_1(z) = \tilde{h}_1 \partial_z \varphi - (z+1)^2 \tilde{h}_2 \partial_z \varphi + (z+1)\tilde{h}_3 \cdot \nabla \varphi, \tag{3-28}$$

$$g_2(z) = (z+1)\tilde{h}_2 \partial_z \varphi - \tilde{h}_3 \cdot \nabla \varphi, \quad g_3(z) = (z+1)\tilde{h}_3 \partial_z \varphi. \tag{3-29}$$

Note that $g_1(z)$, $g_2(z)$, and $g_3(z)$ are all linear with respect to $\nabla_{x,z}\varphi(z)$.

After decomposing $g(s)$ in (3-23) into three parts, $\partial_s g_1$, g_2 and $\nabla \cdot g_3$, we integrate by parts in s to move the derivative ∂_s in front of $\partial_s g_1$. As a result, we have

$$\begin{aligned}
\varphi(z) = {} & \left[\frac{e^{-(z+1)|\nabla|} + e^{(z+1)|\nabla|}}{e^{-|\nabla|} + e^{|\nabla|}} \right] \psi \\
& + \frac{1}{2} |\nabla|^{-1} \frac{e^{-z|\nabla|} - e^{z|\nabla|}}{e^{-|\nabla|} + e^{|\nabla|}} \int_{-1}^{0} \left[e^{(s-1)|\nabla|}(g_2 + \nabla \cdot g_3 - |\nabla| g_1) - e^{-(s+1)|\nabla|}(g_2 + \nabla \cdot g_3 + |\nabla| g_1) \right] ds \\
& + \frac{1}{2} \int_{-1}^{0} |\nabla|^{-1} e^{(z+s)|\nabla|} [g_2 + \nabla \cdot g_3 - |\nabla| g_1] \, ds \\
& - \frac{1}{2} \int_{-1}^{0} |\nabla|^{-1} e^{-|z-s||\nabla|} [g_2 + \nabla \cdot g_3 - \operatorname{sign}(z-s)|\nabla| g_1] \, ds.
\end{aligned} \tag{3-30}$$

Now, we know that the nonlinearity in (3-30) is linear with respect to $\nabla_{x,z}\varphi$.

To see the fixed-point-type structure of $\nabla_{x,z}\varphi$, we take the derivative $\nabla_{x,z}$ on both sides of (3-30). As a result, we derive a fixed-point-type formulation for $\nabla_{x,z}\varphi$:

$$\begin{aligned}
\nabla_{x,z}\varphi = {} & \left[\left[\frac{e^{-(z+1)|\nabla|} + e^{(z+1)|\nabla|}}{e^{-|\nabla|} + e^{|\nabla|}} \right] \nabla \psi, \frac{e^{(z+1)|\nabla|} - e^{-(z+1)|\nabla|}}{e^{-|\nabla|} + e^{|\nabla|}} |\nabla| \psi \right] \\
& + \frac{1}{2} \left[\frac{\nabla}{|\nabla|} \frac{e^{-z|\nabla|} - e^{z|\nabla|}}{e^{-|\nabla|} + e^{|\nabla|}} \int_{-1}^{0} \left[e^{(s-1)|\nabla|}(g_2 + \nabla \cdot g_3 - |\nabla| g_1) - e^{-(s+1)|\nabla|}(g_2 + \nabla \cdot g_3 + |\nabla| g_1) \right] ds, \right. \\
& \qquad \left. - \frac{e^{z|\nabla|} + e^{-z|\nabla|}}{e^{-|\nabla|} + e^{|\nabla|}} \int_{-1}^{0} \left[e^{(s-1)|\nabla|}(g_2 + \nabla \cdot g_3 - |\nabla| g_1) - e^{-(s+1)|\nabla|}(g_2 + \nabla \cdot g_3 + |\nabla| g_1) \right] ds \right] \\
& + \frac{1}{2} \left[\int_{-1}^{0} \frac{\nabla}{|\nabla|} e^{(z+s)|\nabla|} [g_2 + \nabla \cdot g_3 - |\nabla| g_1] \, ds - \int_{-1}^{0} \frac{\nabla}{|\nabla|} e^{-|z-s||\nabla|} [g_2 + \nabla \cdot g_3 - \operatorname{sign}(z-s)|\nabla| g_1] \, ds, \right. \\
& \qquad \left. \int_{-1}^{0} e^{(z+s)|\nabla|} [g_2 + \nabla \cdot g_3 - |\nabla| g_1] \, ds - \int_{-1}^{0} e^{-|z-s||\nabla|} [\operatorname{sign}(s-z)(g_2 + \nabla \cdot g_3) + |\nabla| g_1] \, ds \right] \\
& + [\mathbf{0}, g_1(z)].
\end{aligned} \tag{3-31}$$

To simplify the notation, we define operators

$$K_1(z,s) := \frac{1}{2}\left[\frac{\nabla}{|\nabla|}\frac{e^{-z|\nabla|}-e^{z|\nabla|}}{e^{-|\nabla|}+e^{|\nabla|}}e^{(s-1)|\nabla|}+\frac{\nabla}{|\nabla|}e^{(z+s)|\nabla|}, -\frac{e^{z|\nabla|}+e^{-z|\nabla|}}{e^{-|\nabla|}+e^{|\nabla|}}e^{(s-1)|\nabla|}+e^{(z+s)|\nabla|}\right], \tag{3-32}$$

$$K_2(z,s) := \frac{1}{2}\left[\frac{\nabla}{|\nabla|}\frac{e^{-z|\nabla|}-e^{z|\nabla|}}{e^{-|\nabla|}+e^{|\nabla|}}e^{-(s+1)|\nabla|}, -\frac{e^{z|\nabla|}+e^{-z|\nabla|}}{e^{-|\nabla|}+e^{|\nabla|}}e^{-(s+1)|\nabla|}\right], \tag{3-33}$$

$$K_3(z,s) := \frac{1}{2}\left[\frac{\nabla}{|\nabla|}e^{-|z-s||\nabla|}, e^{-|z-s||\nabla|}\operatorname{sign}(s-z)\right]. \tag{3-34}$$

With the above operators, we can rewrite (3-31) as

$$\begin{aligned}\nabla_{x,z}\varphi = &\left[\left[\frac{e^{-(z+1)|\nabla|}+e^{(z+1)|\nabla|}}{e^{-|\nabla|}+e^{|\nabla|}}\right]\nabla\psi, \frac{e^{(z+1)|\nabla|}-e^{-(z+1)|\nabla|}}{e^{-|\nabla|}+e^{|\nabla|}}|\nabla|\psi\right] + [\mathbf{0}, g_1(z)] \\ &+ \int_{-1}^{0}[K_1(z,s)-K_2(z,s)-K_3(z,s)](g_2(s)+\nabla\cdot g_3(s))\,ds \\ &\quad + \int_{-1}^{0} K_3(z,s)|\nabla|\operatorname{sign}(z-s)g_1(s) - |\nabla|[K_1(z,s)+K_2(z,s)]g_1(s)\,ds.\end{aligned} \tag{3-35}$$

To make sure that we can conclude the fixed-point-type argument, we need to estimate the operators $K_i(z,s)$ so that the issue of losing derivatives does not exist. More precisely, the following lemma holds.

Lemma 3.2. *For $k,\gamma \ge 0$, we have the estimates*

$$\sum_{i=1}^{3}\left\|\int_{-1}^{0}K_i(z,s)\nabla g(s)\,ds\right\|_{L_z^\infty H^k} + \left\|\int_{-1}^{0}K_i(z,s)g(s)\,ds\right\|_{L_z^\infty H^k} \lesssim \|g(z)\|_{L_z^\infty H^k}, \tag{3-36}$$

$$\sum_{i=1}^{3}\left\|\int_{-1}^{0}K_i(z,s)\nabla g(s)\,ds\right\|_{L_z^\infty \widetilde{W}^\gamma} + \left\|\int_{-1}^{0}[K_1(z,s)-K_2(z,s)-K_3(z,s)]g(s)\,ds\right\|_{L_z^\infty \widetilde{W}^\gamma} \lesssim \|g(z)\|_{L_z^\infty \widetilde{W}^\gamma}. \tag{3-37}$$

Proof. We first prove the desired estimate (3-36). Recall (3-32), (3-33), and (3-34). From Lemma 2.2, we have

$$\sup_{z,s\in[-1,0]}\left\|\mathcal{F}^{-1}\left[\mathcal{F}\big([K_1(z,s)-K_2(z,s)-K_3(z,s)-[0,(1-\operatorname{sign}(s-z))/2]]\big)(\xi)\psi_{k_1}(\xi)\right]\right\|_{L^1} \lesssim 2^{k_{1,-}}, \tag{3-38}$$

$$\sup_{z,s\in[-1,0]}\sum_{i=1}^{3}\left\|\mathcal{F}^{-1}\left[\mathcal{F}\big(K_i(z,s)\big)(\xi)\psi_{k_1}(\xi)\right]\right\|_{L^1} \lesssim 1. \tag{3-39}$$

We will use above estimates for the case when $k_1 < 0$. However, when $k_1 \ge 0$, we cannot use the estimate (3-39) directly to estimate the left-hand side of (3-36); otherwise we lose one derivative. An important observation is that the integration with respect to s actually compensates for the loss.

For any fixed $k \ge 0$, $k \in \mathbb{Z}$, we have the following formulation in terms of the kernel:

$$\int_{-1}^{0}K_i(z,s)\nabla P_k[g(s)]\,ds = \int_{-1}^{0}\int_{\mathbb{R}^2}K_{i;k}(z,s,y)g(s,x-y)\,dy\,ds, \tag{3-40}$$

where

$$K_{i;k}(z,s,y) = \int_{\mathbb{R}^2} e^{iy\cdot\xi}\mathcal{F}(K_1(z,s))(\xi)\psi_k(\xi)\xi\, d\xi. \tag{3-41}$$

After integration by parts in ξ many times, we have the pointwise estimate

$$|K_{i;k}(z,s,y)| \lesssim 2^{3k}(1+2^k|y|+2^k|z-s|)^{-10} \tag{3-42}$$

for $i\in\{1,2,3\}$, which further implies that the kernel $K_{i;k}(z,s,y)$ belongs to $L^1_{s,y}$ for fixed z. Therefore, from (3-39) and (3-42), we have the estimate

$$\begin{aligned}&\big|\text{the left-hand side of (3-36)}\big|^2\\ &\quad\lesssim \sum_{k_1\le 0}\|P_{k_1}[g(z)]\|^2_{L^\infty_z L^2} + \sum_{i=1}^{3}\sum_{k_1\ge 0} 2^{2kk_1}\|K_{i;k_1}(z,s)\|^2_{L^1_{s,y}}\|P_{k_1}[g(z)]\|^2_{L^\infty_z L^2} \lesssim \|g(z)\|^2_{L^\infty_z H^k},\end{aligned}$$

hence finishing the proof of (3-36). Very similarly, from (3-38), (3-39), and (3-42), our desired estimate (3-37) follows in the same way. □

From (3-35) and estimates in Lemma 3.2, now it is clear that we can estimate $\nabla_{x,z}\varphi$ by using a fixed-point-type argument.

However, if we do it naively, then the resulting estimate will not tell the difference between $\nabla_x\varphi$ and $\partial_z\varphi$. To capture the fact that $\partial_z\varphi$ actually has two derivatives at the low-frequency part, while $\nabla_x\varphi$ only has one derivative, we decompose $\nabla_{x,z}\varphi$ as

$$\nabla_{x,z}\varphi = \Lambda_1[\nabla_{x,z}\varphi] + \Lambda_{\ge 2}[\nabla_{x,z}\varphi]. \tag{3-43}$$

From (3-35), it is easy to see that $\Lambda_1[\nabla_{x,z}\varphi]$ is given by

$$\Lambda_1[\nabla_{x,z}\varphi] = \left[\left[\frac{e^{-(z+1)|\nabla|}+e^{(z+1)|\nabla|}}{e^{-|\nabla|}+e^{|\nabla|}}\right]\nabla\psi,\ \frac{e^{(z+1)|\nabla|}-e^{-(z+1)|\nabla|}}{e^{-|\nabla|}+e^{|\nabla|}}|\nabla|\psi\right]. \tag{3-44}$$

From (3-44), it is easy to see that $\Lambda_1[\partial_z\varphi]$ has two derivatives at the low-frequency part. Now, the goal is reduced to estimating $\Lambda_{\ge 2}[\nabla_{x,z}\varphi]$, which is done again by a fixed-point-type argument.

Recall (3-35). To identify the fixed-point-type structure inside $\Lambda_{\ge 2}[\nabla_{x,z}\varphi]$, it is sufficient to reformulate $\Lambda_{\ge 2}[g_i(z)],\ i\in\{1,2,3\}$.

Recall (3-28) and (3-29). After using the decomposition (3-43) for $\nabla_{x,z}\varphi$ in $g_i(z),\ i\in\{1,2,3\}$, we have the decomposition of $\Lambda_{\ge 2}[g_i(z)],\ i\in\{1,2,3\}$,

$$\begin{aligned}\Lambda_{\ge 2}[g_1(z)] &= \tilde{h}_1\Lambda_{\ge 2}[\partial_z\varphi] - (z+1)^2\tilde{h}_2\Lambda_{\ge 2}[\partial_z\varphi] + (z+1)\tilde{h}_3\cdot\Lambda_{\ge 2}[\nabla\varphi]\\ &\qquad + \tilde{h}_1\Lambda_1[\partial_z\varphi] - (z+1)^2\tilde{h}_2\Lambda_1[\partial_z\varphi] + (z+1)\tilde{h}_3\cdot\Lambda_1[\nabla\varphi],\end{aligned} \tag{3-45}$$

$$\Lambda_{\ge 2}[g_2(z)] = (z+1)\tilde{h}_2\Lambda_{\ge 2}[\partial_z\varphi] - \tilde{h}_3\cdot\Lambda_{\ge 2}[\nabla\varphi] + (z+1)\tilde{h}_2\Lambda_1[\partial_z\varphi] - \tilde{h}_3\cdot\Lambda_1[\nabla\varphi], \tag{3-46}$$

$$\Lambda_{\ge 2}[g_3(z)] = (z+1)\tilde{h}_3\Lambda_{\ge 2}[\partial_z\varphi] + (z+1)\tilde{h}_3\Lambda_1[\partial_z\varphi]. \tag{3-47}$$

From (3-45), (3-46), and (3-47), now it is easy to see that there exists a fixed-point-type structure for $\Lambda_{\ge 2}[\nabla_{x,z}\varphi]$ in $\Lambda_{\ge 2}[g_i(z)],\ i\in\{1,2,3\}$. From the standard fixed-point-type argument and the estimates in Lemma 3.2, we obtain basic estimates for $\Lambda_{\geq 2}[\nabla_{x,z}\varphi]$, which further give us more precise estimates for $\nabla_{x,z}\varphi$ from (3-43).

More precisely, our main results in this subsection are summarized as follows,

Lemma 3.3. *For $\gamma', k' \geq 1$, $0 < \delta \ll 1$, $\alpha \in (0,1]$, if $h \in \widetilde{W}^{\gamma'} \cap H^{k'}$ satisfies the smallness assumption*

$$\|h\|_{\widetilde{W}^{\gamma'}} < \delta, \tag{3-48}$$

then the following L^2-type estimate and L^∞-type estimate for the velocity potential φ hold:

$$\|\nabla_{x,z}\varphi\|_{L^\infty_z H^k} \lesssim \|\nabla\psi\|_{H^k} + \|h\|_{H^{k+1}}\|\nabla\psi\|_{\widetilde{W}^0}, \tag{3-49}$$

$$\|\nabla_x\varphi\|_{L^\infty_z \widetilde{W}^\gamma} \lesssim \|\nabla\psi\|_{\widetilde{W}^\gamma}, \quad \|\partial_z\varphi\|_{L^\infty_z \widetilde{W}^\gamma} \lesssim \|\nabla\psi\|_{\widehat{W}^{\gamma,\alpha}} + \|h\|_{\widetilde{W}^{\gamma+1}}\|\nabla\psi\|_{\widetilde{W}^\gamma}, \tag{3-50}$$

$$\|\Lambda_{\geq 2}[\nabla_{x,z}\varphi]\|_{L^\infty_z \widetilde{W}^\gamma} \lesssim \|\nabla\psi\|_{\widetilde{W}^\gamma}\|h\|_{\widetilde{W}^{\gamma+1}}, \tag{3-51}$$

$$\|\Lambda_{\geq 2}[\nabla_{x,z}\varphi]\|_{L^\infty_z H^k} \lesssim \|h\|_{\widetilde{W}^1}\||\nabla|\psi\|_{H^k} + \|\nabla\psi\|_{\widetilde{W}^0}\|h\|_{H^{k+1}}, \tag{3-52}$$

where $k \leq k'-1$ and $1 \leq \gamma \leq \gamma'-1$. In the above estimates, the range of z for the L^∞_z norm is $[-1,0]$.

Proof. We first estimate $\Lambda_{\geq 2}[\nabla_{x,z}\varphi]$. Recall (3-35), (3-45), (3-46) and (3-47). From estimate (3-37) in Lemma 3.2, we have

$$\begin{aligned}\|\Lambda_{\geq 2}[\nabla_{x,z}\varphi]\|_{L^\infty_z \widetilde{W}^\gamma} &\lesssim \|\Lambda_{\geq 2}[(g_1(z), g_2(z), g_3(z))]\|_{L^\infty_z \widetilde{W}^\gamma} \\ &\lesssim \|h\|_{\widetilde{W}^{\gamma+1}}\|\Lambda_{\geq 2}[\nabla_{x,z}\varphi]\|_{L^\infty_z \widetilde{W}^\gamma} + \|h\|_{\widetilde{W}^{\gamma+1}}\|\nabla\psi\|_{\widetilde{W}^\gamma}.\end{aligned}$$

Hence, by the smallness condition (3-48),

$$\|\Lambda_{\geq 2}[\nabla_{x,z}\varphi]\|_{L^\infty_z \widetilde{W}^\gamma} \lesssim \|h\|_{\widetilde{W}^{\gamma+1}}\|\nabla\psi\|_{\widetilde{W}^\gamma}. \tag{3-53}$$

Very similarly, from estimate (3-36) in Lemma 3.2, we have

$$\begin{aligned}&\|\Lambda_{\geq 2}[\nabla_{x,z}\varphi]\|_{L^\infty_z H^k} \\ &\lesssim \|\Lambda_{\geq 2}[(g_1(z), g_2(z), g_3(z))]\|_{L^\infty_z H^k} \\ &\lesssim \|h\|_{\widetilde{W}^1}\|\Lambda_{\geq 2}[\nabla_{x,z}\varphi]\|_{L^\infty_z H^k} + \|h\|_{H^{k+1}}\|\Lambda_{\geq 2}[\nabla_{x,z}\varphi]\|_{L^\infty_z \widetilde{W}^0} + \|h\|_{H^{k+1}}\|\nabla\psi\|_{\widetilde{W}^0} + \|\nabla\psi\|_{H^k}\|h\|_{\widetilde{W}^1} \\ &\lesssim \|h\|_{\widetilde{W}^1}\|\Lambda_{\geq 2}[\nabla_{x,z}\varphi]\|_{L^\infty_z H^k} + \|h\|_{H^{k+1}}\|\nabla\psi\|_{\widetilde{W}^0}(1+\|h\|_{\widetilde{W}^1}) + \|\nabla\psi\|_{H^k}\|h\|_{\widetilde{W}^1}.\end{aligned}$$

Again, by the smallness assumption (3-48), we conclude

$$\|\Lambda_{\geq 2}[\nabla_{x,z}\varphi]\|_{L^\infty_z H^k} \lesssim \|h\|_{H^{k+1}}\|\nabla\psi\|_{\widetilde{W}^0} + \|\nabla\psi\|_{H^k}\|h\|_{\widetilde{W}^1}. \tag{3-54}$$

From estimates (3-53) and (3-54) and the explicit formulas of $\Lambda_1[\nabla_{x,z}\varphi]$ in (3-44), we have

$$\begin{aligned}&\|\nabla_{x,z}\varphi\|_{L^\infty H^k} \lesssim \|\nabla\psi\|_{H^k} + \|h\|_{H^{k+1}}\|\nabla\psi\|_{\widetilde{W}^0}, \quad \|\nabla_x\varphi\|_{L^\infty_z \widetilde{W}^\gamma} \lesssim \|\nabla\psi\|_{\widetilde{W}^\gamma}, \\ &\|\partial_z\varphi\|_{L^\infty_z \widetilde{W}^\gamma} \lesssim \|\Lambda^2\psi\|_{\widetilde{W}^\gamma} + \|h\|_{\widetilde{W}^{\gamma+1}}\|\nabla\psi\|_{\widetilde{W}^\gamma} \lesssim \|\nabla\psi\|_{\widehat{W}^{\gamma,\alpha}} + \|h\|_{\widetilde{W}^{\gamma+1}}\|\nabla\psi\|_{\widetilde{W}^\gamma}.\end{aligned}$$

□

3D. *The quadratic terms of the Dirichlet–Neumann operator.* The content of this subsection is not related to the proof of our main theorem. However, it is crucial to the study of the long-time behavior of the water waves system in the flat-bottom setting.

Generally speaking, the main enemies of the global existence for a 2-dimensional dispersive equation are the quadratic terms. The first step is to know exactly what the enemies are. Surprisingly, as a byproduct of the fixed-point-type formulation (3-35), we can calculate explicitly the quadratic terms of the Dirichlet–Neumann operator.

More precisely, the main result of this subsection is stated as follows:

Lemma 3.4. *In terms of h and ψ, the quadratic terms of the Dirichlet–Neumann operator are*

$$\Lambda_2[G(h)\psi] = -\nabla\cdot(h\nabla\psi) - |\nabla|\tanh|\nabla|(h|\nabla|\tanh|\nabla|\psi). \tag{3-55}$$

Remark 3.5. Before we proceed to prove the above lemma, we compare the main difference between the flat-bottom setting, which is less studied, and the infinite depth setting, which is recently well-studied. In the infinite-depth setting, the quadratic terms of the Dirichlet–Neumann operator are

$$\text{(infinite-depth setting)}\quad \Lambda_2[G(h)\psi] = -\nabla\cdot(h\nabla\psi) - |\nabla|(h|\nabla|\psi). \tag{3-56}$$

If the frequency η of ψ is of size 1 and the frequency $\xi-\eta$ of h is of size 0, from (3-55) and (3-56), it is easy to check the size of the symbol of quadratic terms:

$$\text{(flat-bottom setting)}\quad \xi\cdot\eta - |\xi|\,|\eta|\tanh|\xi|\tanh|\eta| = \frac{4|\xi|^2}{(e^{|\xi|}+e^{-|\xi|})^2}\sim 1,$$

$$\text{(infinite-depth setting)}\quad -|\xi|\,|\eta| + \xi\cdot\eta = 0.$$

That is to say, unlike the infinite-depth setting, we do not have the null structure at the low-frequency part in the flat-bottom setting. As a result, we expect a much stronger nonlinear effect from the quadratic terms, which makes the global regularity problem in the flat-bottom setting more delicate and more difficult than the infinite-depth setting.

Proof of Lemma 3.4. Recall (3-14) and (3-44). We have

$$\Lambda_2[G(h)\psi] = \Lambda_2[\partial_z\varphi|_{z=0}] - h|\nabla|\tanh|\nabla|\psi - \nabla h\cdot\nabla\psi. \tag{3-57}$$

Hence, the problem is reduced to calculating explicitly the quadratic terms of $\partial_z\varphi|_{z=0}$. Recalling (3-35),

$$\begin{aligned}
\Lambda_2[\partial_z\varphi|_{z=0}] &= -\frac{1}{e^{|\nabla|}+e^{-|\nabla|}}\int_{-1}^{0}[e^{(s-1)|\nabla|}-e^{-(s+1)|\nabla|}][\Lambda_2[g_2+\nabla\cdot g_3]]\,ds\\
&\quad+\frac{1}{e^{|\nabla|}+e^{-|\nabla|}}\int_{-1}^{0}[e^{(s-1)|\nabla|}+e^{-(s+1)|\nabla|}]|\nabla|[\Lambda_2[g_1]]\,ds\\
&\quad+\int_{-1}^{0}e^{s|\nabla|}\Lambda_2[g_2+\nabla\cdot g_3-|\nabla|g_1]\,ds+\Lambda_2[g_1(0)]\\
&=\int_{-1}^{0}\frac{e^{(s+1)|\nabla|}+e^{-(s+1)|\nabla|}}{e^{|\nabla|}+e^{-|\nabla|}}\Lambda_2[g_2+\nabla\cdot g_3]\,ds\\
&\quad-\int_{-1}^{0}\frac{e^{(s+1)|\nabla|}-e^{-(s+1)|\nabla|}}{e^{|\nabla|}+e^{-|\nabla|}}|\nabla|\Lambda_2[g_1(s)]\,ds+\Lambda_2[g_1(0)].
\end{aligned} \tag{3-58}$$

From (3-25), (3-26), and (3-44), it is easy to derive the equalities

$$
\begin{aligned}
\Lambda_2[g_1(s)] &= 2h\frac{e^{(s+1)|\nabla|}-e^{-(s+1)|\nabla|}}{e^{|\nabla|}+e^{-|\nabla|}}|\nabla|\psi+(s+1)\nabla h\cdot\nabla\frac{e^{(s+1)|\nabla|}+e^{-(s+1)|\nabla|}}{e^{|\nabla|}+e^{-|\nabla|}}\psi,\\
\Lambda_2[g_1(0)] &= 2h|\nabla|\tanh|\nabla|\psi+\nabla h\cdot\nabla\psi,\\
\Lambda_2[g_2(s)] &= -\nabla h\cdot\nabla\frac{e^{(s+1)|\nabla|}+e^{-(s+1)|\nabla|}}{e^{|\nabla|}+e^{-|\nabla|}}\psi,\\
\Lambda_2[g_3(s)] &= (s+1)\nabla h\frac{e^{(s+1)|\nabla|}-e^{-(s+1)|\nabla|}}{e^{|\nabla|}+e^{-|\nabla|}}|\nabla|\psi.
\end{aligned}
\tag{3-59}
$$

After plugging in the above explicit formula of $\Lambda_2[g_i(z)]$, $i\in\{1,2,3\}$, the goal is to calculate explicitly the symbols of the two integrals in (3-58). Define

$$
Q_1(h,\psi):=\int_{-1}^{0}\frac{e^{(s+1)|\nabla|}+e^{-(s+1)|\nabla|}}{e^{|\nabla|}+e^{-|\nabla|}}\Lambda_2[g_2+\nabla\cdot g_3]\,ds=Q_{1,1}(h,\psi)+Q_{1,2}(h,\psi),\tag{3-60}
$$

$$
Q_2(h,\psi):=-\int_{-1}^{0}\frac{e^{(s+1)|\nabla|}-e^{-(s+1)|\nabla|}}{e^{|\nabla|}+e^{-|\nabla|}}|\nabla|\Lambda_2[g_1]\,ds=Q_{2,1}(h,\psi)+Q_{2,2}(h,\psi),\tag{3-61}
$$

where

$$
\begin{aligned}
Q_{1,1}(h,\psi)&=\int_{-1}^{0}\frac{e^{(s+1)|\nabla|}+e^{-(s+1)|\nabla|}}{e^{|\nabla|}+e^{-|\nabla|}}\left[\nabla\cdot[(s+1)\nabla h\frac{e^{(s+1)|\nabla|}-e^{-(s+1)|\nabla|}}{e^{|\nabla|}+e^{-|\nabla|}}|\nabla|\psi\right]ds,\\
Q_{1,2}(h,\psi)&=\int_{-1}^{0}\frac{e^{(s+1)|\nabla|}+e^{-(s+1)|\nabla|}}{e^{|\nabla|}+e^{-|\nabla|}}\left[-\nabla h\cdot\nabla\frac{e^{(s+1)|\nabla|}+e^{-(s+1)|\nabla|}}{e^{|\nabla|}+e^{-|\nabla|}}\psi\right]ds,\\
Q_{2,1}(h,\psi)&=-\int_{-1}^{0}\frac{e^{(s+1)|\nabla|}-e^{-(s+1)|\nabla|}}{e^{|\nabla|}+e^{-|\nabla|}}|\nabla|\left[2h\frac{e^{(s+1)|\nabla|}-e^{-(s+1)|\nabla|}}{e^{|\nabla|}+e^{-|\nabla|}}|\nabla|\psi\right]ds,\\
Q_{2,2}(h,\psi)&=-\int_{-1}^{0}\frac{e^{(s+1)|\nabla|}-e^{-(s+1)|\nabla|}}{e^{|\nabla|}+e^{-|\nabla|}}|\nabla|\left[(s+1)\nabla h\cdot\nabla\frac{e^{(s+1)|\nabla|}+e^{-(s+1)|\nabla|}}{e^{|\nabla|}+e^{-|\nabla|}}\psi\right]ds.
\end{aligned}
$$

The symbol $q_{i,j}(\xi-\eta,\eta)$ of the bilinear operator $Q_{i,j}(h,\psi)$, $i,j\in\{1,2\}$, is given by

$$
\begin{aligned}
q_{1,1}(\xi-\eta,\eta)&=\frac{-\xi\cdot(\xi-\eta)|\eta|}{(e^{|\xi|}+e^{-|\xi|})(e^{|\eta|}+e^{-|\eta|})}\int_{-1}^{0}(s+1)[e^{(s+1)|\xi|}+e^{-(s+1)|\xi|}][e^{(s+1)|\eta|}-e^{-(s+1)|\eta|}]\,ds\\
&=\frac{-\xi\cdot(\xi-\eta)|\eta|}{(e^{|\xi|}+e^{-|\xi|})(e^{|\eta|}+e^{-|\eta|})}\biggl[\frac{(|\xi|+|\eta|-1)e^{|\xi|+|\eta|}-(-|\xi|-|\eta|-1)e^{-|\xi|-|\eta|}}{(|\xi|+|\eta|)^2}\\
&\qquad+\frac{(|\eta|-|\xi|-1)e^{|\eta|-|\xi|}-(|\xi|-|\eta|-1)e^{|\xi|-|\eta|}}{(|\xi|-|\eta|)^2}\biggr],
\end{aligned}
\tag{3-62}
$$

$$
\begin{aligned}
q_{1,2}(\xi-\eta,\eta)&=\frac{(\xi-\eta)\cdot\eta}{(e^{|\xi|}+e^{-|\xi|})(e^{|\eta|}+e^{-|\eta|})}\int_{-1}^{0}[e^{(s+1)|\xi|}+e^{-(s+1)|\xi|}][e^{(s+1)|\eta|}+e^{-(s+1)|\eta|}]\,ds\\
&=\frac{(\xi-\eta)\cdot\eta}{(e^{|\xi|}+e^{-|\xi|})(e^{|\eta|}+e^{-|\eta|})}\left[\frac{e^{|\xi|+|\eta|}-e^{-|\xi|-|\eta|}}{|\xi|+|\eta|}+\frac{e^{|\xi|-|\eta|}-e^{|\eta|-|\xi|}}{|\xi|-|\eta|}\right],
\end{aligned}
\tag{3-63}
$$

$$\begin{aligned} q_{2,1}(\xi-\eta,\eta) &= \frac{-2|\xi|\,|\eta|}{(e^{|\xi|}+e^{-|\xi|})(e^{|\eta|}+e^{-|\eta|})}\int_{-1}^{0}[e^{(s+1)|\xi|}-e^{-(s+1)|\xi|}][e^{(s+1)|\eta|}-e^{-(s+1)|\eta|}]\,ds \\ &= \frac{-2|\xi|\,|\eta|}{(e^{|\xi|}+e^{-|\xi|})(e^{|\eta|}+e^{-|\eta|})}\left[\frac{e^{|\xi|+|\eta|}-e^{-|\xi|-|\eta|}}{|\xi|+|\eta|}-\frac{e^{|\xi|-|\eta|}-e^{|\eta|-|\xi|}}{|\xi|-|\eta|}\right], \end{aligned} \tag{3-64}$$

$$\begin{aligned} q_{2,2}(\xi-\eta,\eta) &= \frac{|\xi|(\xi-\eta)\cdot\eta}{(e^{|\xi|}+e^{-|\xi|})(e^{|\eta|}+e^{-|\eta|})}\int_{-1}^{0}(s+1)[e^{(s+1)|\xi|}-e^{-(s+1)|\xi|}][e^{(s+1)|\eta|}+e^{-(s+1)|\eta|}]\,ds \\ &= \frac{|\xi|(\xi-\eta)\cdot\eta}{(e^{|\xi|}+e^{-|\xi|})(e^{|\eta|}+e^{-|\eta|})}\left[\frac{(|\xi|+|\eta|-1)e^{|\xi|+|\eta|}-(-|\xi|-|\eta|-1)e^{-|\xi|-|\eta|}}{(|\xi|+|\eta|)^2}\right. \\ &\qquad\left.-\frac{(|\eta|-|\xi|-1)e^{|\eta|-|\xi|}-(|\xi|-|\eta|-1)e^{|\xi|-|\eta|}}{(|\xi|-|\eta|)^2}\right]. \end{aligned} \tag{3-65}$$

In the above computations, we have used the simple fact

$$\int_{-1}^{0}(s+1)e^{(s+1)a}\,ds=\frac{1+(a-1)e^{a}}{a^2}.$$

From (3-57)–(3-61), we have

$$\Lambda_2[G(h)\psi]:=\widetilde{Q}(h,\psi)=Q_1(h,\psi)+Q_2(h,\psi)+h|\nabla|\tanh|\nabla|\psi.$$

Therefore, the symbol $\tilde{q}(\xi-\eta,\eta)$ of $\widetilde{Q}(h,\psi)$ is given by

$$\tilde{q}(\xi-\eta,\eta)=\sum_{i,j=1,2}q_{i,j}(\xi-\eta,\eta)+\frac{e^{|\eta|}-e^{-|\eta|}}{e^{|\eta|}+e^{-|\eta|}}|\eta|.$$

Although the above formulae look complicated, actually there are cancellations inside. Note that

$$\begin{aligned} &q_{1,2}(\xi-\eta,\eta)+q_{2,1}(\xi-\eta,\eta)+\frac{e^{|\eta|}-e^{-|\eta|}}{e^{|\eta|}+e^{-|\eta|}}|\eta| \\ &\quad=\frac{\xi\cdot\eta}{(e^{|\xi|}+e^{-|\xi|})(e^{|\eta|}+e^{-|\eta|})}\left[\frac{e^{|\xi|+|\eta|}-e^{-|\xi|-|\eta|}}{|\xi|+|\eta|}+\frac{e^{|\xi|-|\eta|}-e^{|\eta|-|\xi|}}{|\xi|-|\eta|}\right] \\ &\qquad-\frac{|\xi|\,|\eta|}{(e^{|\xi|}+e^{-|\xi|})(e^{|\eta|}+e^{-|\eta|})}\left[\frac{[e^{|\xi|+|\eta|}-e^{-|\xi|-|\eta|}]}{|\xi|+|\eta|}-\frac{e^{|\xi|-|\eta|}-e^{|\eta|-|\xi|}}{|\xi|-|\eta|}\right], \end{aligned} \tag{3-66}$$

$$\begin{aligned} &q_{1,1}(\xi-\eta,\eta)+q_{2,2}(\xi-\eta,\eta) \\ &\quad=\frac{(-|\xi|\,|\eta|+\xi\cdot\eta)}{(e^{|\xi|}+e^{-|\xi|})(e^{|\eta|}+e^{-|\eta|})}\frac{(|\xi|+|\eta|-1)e^{|\xi|+|\eta|}-(-|\xi|-|\eta|-1)e^{-|\xi|-|\eta|}}{|\xi|+|\eta|} \\ &\qquad-\frac{(\xi|\,|\eta|+\xi\cdot\eta)}{(e^{|\xi|}+e^{-|\xi|})(e^{|\eta|}+e^{-|\eta|})}\frac{(|\eta|-|\xi|-1)e^{|\eta|-|\xi|}-(|\xi|-|\eta|-1)e^{|\xi|-|\eta|}}{|\xi|-|\eta|}. \end{aligned} \tag{3-67}$$

From (3-66) and (3-67), now it is easy to verify

$$\tilde{q}(\xi-\eta,\eta)=\xi\cdot\eta-|\xi|\,|\eta|\tanh|\xi|\tanh|\eta|. \tag{3-68}$$

Hence our desired equality (3-55) holds. □

Lemma 3.6. *For $k_1, k_2, k \in \mathbb{Z}$, the following estimate holds for the symbol of the quadratic terms for the Dirichlet–Neumann operator*:

$$\|\tilde{q}(\xi-\eta,\eta)\|_{\mathcal{S}^\infty_{k,k_1,k_2}} \lesssim 2^{k+k_2}. \tag{3-69}$$

Proof. From (3-68) and the estimate in Lemma 2.2, it is straightforward to derive the above estimate. □

3E. ***A fixed-point-type formulation for $\Lambda_{\geq 3}[\nabla_{x,z}\varphi]$.*** As in the previous subsection, the content of this subsection is not related to the proof of the main theorem but is related to the future study of the long-time behavior of the water waves system in different settings.

Although, intuitively speaking, the quadratic terms are the leading terms for the dispersive equation (1-8) in 2 dimensions, we also have to control the cubic and higher-order remainder terms to see that their effects are indeed small over time. In this subsection, our goal is to formulate $\Lambda_{\geq 3}[\nabla_{x,z}\varphi]$ into a fixed-point-type formulation, which provides a good way to estimate the cubic and higher-order remainder terms.

Recall the fixed-point-type formulation of $\nabla_{x,z}\varphi$ in (3-35), we truncate it at the cubic-and-higher level and get

$$\begin{aligned}\Lambda_{\geq 3}[\nabla_{x,z}\varphi] = [\mathbf{0}, \Lambda_{\geq 3}[g_1(z)]] + \int_{-1}^{0} [K_1(z,s) - K_2(z,s) - K_3(z,s)]\big(\Lambda_{\geq 3}[g_2(s)] + \nabla\cdot\Lambda_{\geq 3}[g_3(s)]\big)\, ds \\ + \int_{-1}^{0} K_3(z,s)|\nabla|\operatorname{sign}(z-s)\Lambda_{\geq 3}[g_1(s)] - |\nabla|[K_1(z,s)+K_2(z,s)]\Lambda_{\geq 3}[g_1(s)]\, ds.\end{aligned} \tag{3-70}$$

Recall (3-28) and (3-29). Similar to the decomposition we did in (3-45)–(3-47), we can separate $\Lambda_{\geq 3}[g_i(z)]$, $i \in \{1,2,3\}$, into two parts: (i) one of them contains $\Lambda_{\geq 3}[\nabla_{x,z}\varphi]$, which involves the fixed-point structure; (ii) the other part does not depend on $\Lambda_{\geq 3}[\nabla_{x,z}\varphi]$, and hence can be estimated directly.

More precisely, we decompose $\Lambda_{\geq 3}[g_i(z)]$, $i \in \{1,2,3\}$, as follows:

$$\begin{aligned}\Lambda_{\geq 3}[g_1(z)] = \tilde{h}_1\Lambda_{\geq 3}[\partial_z\varphi] - (z+1)^2\tilde{h}_2\Lambda_{\geq 3}[\partial_z\varphi] + (z+1)\tilde{h}_3\cdot\Lambda_{\geq 3}[\nabla\varphi] \\ + \sum_{i=1,2}\Lambda_{\geq 3-i}[\tilde{h}_1]\Lambda_i[\partial_z\varphi] - (z+1)^2\Lambda_{\geq 3-i}[\tilde{h}_2]\Lambda_i[\partial_z\varphi] + (z+1)\Lambda_{\geq 3-i}[\tilde{h}_3]\cdot\Lambda_i[\nabla\varphi],\end{aligned} \tag{3-71}$$

$$\begin{aligned}\Lambda_{\geq 3}[g_2(z)] = (z+1)\tilde{h}_2\Lambda_{\geq 3}[\partial_z\varphi] - \tilde{h}_3\cdot\Lambda_{\geq 3}[\nabla\varphi] \\ + \sum_{i=1,2}(z+1)\Lambda_{\geq 3-i}[\tilde{h}_2]\Lambda_i[\partial_z\varphi] - \Lambda_{\geq 3-i}[\tilde{h}_3]\cdot\Lambda_i[\nabla\varphi],\end{aligned} \tag{3-72}$$

$$\Lambda_{\geq 3}[g_3(z)] = (z+1)\tilde{h}_3\Lambda_{\geq 3}[\partial_z\varphi] + \sum_{i=1,2}(z+1)\Lambda_{\geq 3-i}[\tilde{h}_3]\Lambda_i[\partial_z\varphi]. \tag{3-73}$$

From (3-27), it is easy to verify that

$$\Lambda_{\geq 2}[\tilde{h}_1] = h^2 - (2h+h^2)\tilde{h}_1, \quad \Lambda_{\geq 2}[\tilde{h}_2] = \tilde{h}_2, \quad \Lambda_{\geq 2}[\tilde{h}_3] = -h\tilde{h}_3. \tag{3-74}$$

We can summarize the above decomposition in the following lemma.

Lemma 3.7. *We have*

$$\Lambda_{\geq 3}[\nabla_{x,z}\varphi(z)] = \sum_{i=1}^{3} C_z^i(h,\psi,\tilde{h}_i) + h\widetilde{C}_z^i(h,\psi,\tilde{h}_i) + T_z^i(\tilde{h}_i, \Lambda_{\geq 3}[\nabla_{x,z}\varphi]),$$

where C_z^i and $\widetilde{C}_z^i$ are some trilinear operators and T_z^i is some bilinear operator. Assume that the corresponding symbols are $c_z^i(\cdot,\cdot,\cdot)$, $\tilde{c}_z^i(\cdot,\cdot,\cdot)$, and $t_z^i(\cdot,\cdot)$ respectively. Then we have the estimates

$$\sup_{z\in[-1,0]} \sum_{i=1}^{3} \|c_z^i(\xi_1,\xi_2,\xi_3)\|_{\mathcal{S}^\infty_{k,k_1,k_2,k_3}} + \|\tilde{c}_z(\xi_1,\xi_2,\xi_3)\|_{\mathcal{S}^\infty_{k,k_1,k_2,k_3}} \lesssim 2^{3\max\{k_1,k_2,k_3\}_+}, \tag{3-75}$$

$$\sup_{z\in[-1,0]} \sum_{i=1}^{3} \|t_z^i(\xi_1,\xi_2)\|_{\mathcal{S}^\infty_{k,k_1,k_2}} \lesssim 2^{3\max\{k_1,k_2\}_+}. \tag{3-76}$$

Proof. The proof is straightforward. From Lemma 2.2, our desired estimates (3-75) and (3-76) can be derived by checking the symbol of each term inside the equations (3-70), (3-71), (3-72) and (3-73). Note that there are at most three derivatives in total. □

4. Paralinearization and symmetrization of the system

Since the gravity waves system (1-7) is quasilinear and lacks symmetric structures inside, we cannot use this system directly to do the energy estimate because of the difficulty of losing one derivative.

To identify the hidden symmetries inside the gravity waves system (1-7) and get around the issue of losing derivatives, we use the method of paralinearization and symmetrization which was introduced and studied in [Alazard and Métivier 2009; Alazard, Burq and Zuily 2011; 2014a; 2014b]. Interested readers may refer to those works for more details. Here, we only briefly discuss this method to help readers understand how this method works and get a sense of what they will read about in this section.

For a fully nonlinear term, it is very hard to tell which part actually loses derivatives and which part does not lose derivatives, which is clearly very important to get around the issue of losing derivatives. With the help of the paralinearization process, we can identify the part that actually loses derivatives, which is the real issue. In Section 4A, we will do the paralinearization process for the nonlinearity of the equation satisfied by the height h, which is the Dirichlet–Neumann operator. In Section 4B, we will do the paralinearization process for the nonlinearity of the equation satisfied by the velocity potential ψ.

Knowing which part loses derivatives is certainly very helpful, but it does not imply that we can get around the issue of losing derivatives because the original system lacks good symmetric structures. With the help of the symmetrization process, in Section 4C, we identify good substitution variables so that the system of equations satisfied by the good substitution variables has the requisite symmetries. Moreover, the good substitution variables have size of energy comparable to that of the original variables. Therefore, instead of doing the energy estimate for the original variables, we do an energy estimate for the good substitution variables.

4A. ***Paralinearization of the Dirichlet–Neumann operator.*** In this subsection, our main goal is to identify which part of the Dirichlet–Neumann operator actually loses derivatives by using the paralinearization method. In the meantime, we also pay attention to the low-frequency part for the purpose of proving our new energy estimate (1-12).

More precisely, the goal of this subsection is to prove the following proposition.

Proposition 4.1. *Let $k \geq 6$, $\alpha \in (0,1]$. Assume that $(h, \Lambda\psi) \in H^k$ and h satisfies the smallness condition (3-48). Then we have*

$$G(h)\psi = T_\lambda(\psi - T_B h) - T_V \cdot \nabla h + F(h)\psi = \Lambda^2(\psi - T_B h) + T_{\lambda-|\xi|}(\psi - T_B h) - T_V \cdot \nabla h + \widetilde{F}(h)\psi, \tag{4-1}$$

where

$$\lambda := \sqrt{(1+|\nabla h|^2)|\xi|^2 - (\nabla h \cdot \xi)^2}, \tag{4-2}$$

$$B \overset{\text{abbr}}{=} B(h)\psi = \frac{G(h)\psi + \nabla h \cdot \nabla \psi}{1+|\nabla h|^2}, \quad V \overset{\text{abbr}}{=} V(h)\psi = \nabla\psi - B\nabla h. \tag{4-3}$$

The good remainder terms $F(h)\psi$ and $\widetilde{F}(h)\psi$ do not lose derivatives and satisfy the estimate

$$\|\Lambda_{\geq 2}[F(h)\psi]\|_{H^k} + \|\widetilde{F}(h)\psi\|_{H^k} \lesssim_k \big[\|(h,\Lambda\psi)\|_{\widehat{W}^{4,\alpha}} + \|(h,\Lambda\psi)\|^2_{\widehat{W}^4}\big]\big[\|h\|_{H^k} + \|\nabla\psi\|_{H^{k-1}}\big]. \tag{4-4}$$

Remark 4.2. We remark that, unlike the infinite-depth setting, the good remainder term $F(h)\psi$ in (4-1) actually contains a linear term, which is $[\tanh(|\nabla|) - 1]|\nabla|\psi \in H^\infty$.

For simplicity, we define the following equivalence relation. For two well-defined nonlinearities A and B, which are nonlinear with respect to h and ψ, we say

$$A \approx B \iff A - B \text{ is a good error term in the sense of (4-5)},$$

$$\|\text{good error term}\|_{H^k} \lesssim_k \big[\|(h,\Lambda\psi)\|_{\widehat{W}^{4,\alpha}} + \|(h,\Lambda\psi)\|^2_{\widehat{W}^4}\big]\big[\|h\|_{H^k} + \|\nabla\psi\|_{H^{k-1}}\big], \quad \alpha \in (0,1],\ k \geq 0. \tag{4-5}$$

Recall (3-15). Note that, essentially speaking, the only fully nonlinear term inside the Dirichlet–Neumann operator $G(h)\psi$ is $\partial_w\Phi|_{w=0}$. So the task is reduced to identifying which part of $\partial_w\Phi$ actually loses a derivative.

To this end, we will show that there exists a pseudodifferential operator $A(x,\xi)$ such that $\partial_w\Phi - T_A(\Phi - T_{\partial_w\Phi}h)$ actually does not lose derivatives, where $\Phi - T_{\partial_w\Phi}h$ is the so-called good unknown variable. This step is very nontrivial and technical. Unfortunately, to the best of our knowledge, there is no physical intuitive explanation available. It relies heavily on the study of good structures for the Laplace equation (3-11). We do this step in detail in the following subsubsection.

4A1. *Paralinearization of the Laplace equation (3-11).* Recall due to (3-11) and the fact that $g'(\cdot, w) = 0$ when $w \in \left[-\frac{1}{4}, 0\right]$, we have

$$\begin{aligned} &[\Delta_x + a'\partial_w^2 + b' \cdot \nabla\partial_w + c'\partial_w]\Phi = 0, \\ &a' = 1 + |\nabla h|^2 \approx 1 + 2T_{\nabla h} \cdot \nabla h, \quad b' = -2\nabla h, \quad c' = -\Delta h. \end{aligned} \tag{4-6}$$

We remark that w is also restricted inside $\left[-\frac{1}{4}, 0\right]$ in the rest of this paper.

Before proceeding to the paralinearization process for (4-6), we need some necessary estimates of Φ. Essentially speaking, under a certain smallness condition, the size of Φ is comparable to φ. Note that we already have necessary estimates of φ; see Lemma 3.3. More precisely, from the definition of Φ (see Section 3B) and estimates of φ in Lemma 3.3, the following lemma holds.

Lemma 4.3. *Under the smallness estimate (3-1), we have the following estimates for $k \geq 1, \gamma \leq 3$:*

$$\sup_{w\in[-1/4,0]} \|\nabla_{x,w}\Phi\|_{H^k} \lesssim \|\nabla\psi\|_{H^k} + \|h\|_{H^{k+1}}\|\nabla\psi\|_{\widetilde{W}^1}, \tag{4-7}$$

$$\sup_{w\in[-1/4,0]} \|\nabla_x\Phi\|_{\widetilde{W}^\gamma} \lesssim \|\nabla\psi\|_{\widetilde{W}^\gamma}, \qquad \sup_{w\in[-1/4,0]} \|\partial_w\Phi\|_{\widetilde{W}^\gamma} \lesssim \|\nabla\psi\|_{\widehat{W}^{\gamma,\alpha}} + \|h\|_{\widetilde{W}^{\gamma+1}}\|\nabla\psi\|_{\widetilde{W}^\gamma}, \tag{4-8}$$

$$\sup_{w\in[-1/4,0]} \|\Lambda_{\geq 2}[\nabla_{x,w}\Phi]\|_{L^2} \lesssim \big[\|(h,\Lambda\psi)\|_{\widehat{W}^{2,\alpha}} + \|(h,\Lambda\psi)\|^2_{\widehat{W}^2}\big]\big[\|h\|_{H^1} + \|\nabla\psi\|_{L^2}\big]. \tag{4-9}$$

Proof. This is postponed to the end of this subsection for the purpose of improving the presentation. □

After paralinearizing (4-6), we have

$$P\Phi + 2T_{\partial_w^2\Phi}T_{\nabla h}\cdot\nabla h - 2T_{\nabla\partial_w\Phi}\cdot\nabla h - T_{\partial_w\Phi}\Delta h \approx 0, \tag{4-10}$$

where

$$P := [\Delta + T_{a'}\partial_w^2 + T_{b'}\cdot\nabla\partial_w + T_{c'}\partial_w]. \tag{4-11}$$

To see why the equivalence relation (4-10) holds, we mention that we can always put ∇h in L^∞ and put $\partial_w\Phi$ and $\partial_w^2\Phi$ in L^2.

Define $W := \Phi - T_{\partial_w\Phi}h$. As in [Alazard, Burq and Zuily 2011], we claim that $PW \approx 0$ when $w \in \big[-\frac{1}{4}, 0\big]$. After using (2-7) and the composition in Lemma 2.6, the following equivalence relations hold:

$$
\begin{aligned}
PW \approx 0 \iff\ & P[T_{\partial_w\Phi}h] + 2T_{\partial_w^2\Phi}T_{\nabla h}\cdot\nabla h - 2T_{\nabla\partial_w\Phi}\cdot\nabla h - T_{\partial_w\Phi}\Delta h \approx 0 && \text{(4-12)}\\
\iff\ & \big[T_{a'}T_{\partial_w^3\Phi}h + T_{b'}\cdot\nabla T_{\partial_w^2\Phi}h + T_{c'}T_{\partial_w^2\Phi}h + 2T_{\nabla\partial_w\Phi}\cdot\nabla h\big] \\
& \qquad + \big[2T_{\partial_w^2\Phi}T_{\nabla h}\cdot\nabla h - 2T_{\nabla\partial_w\Phi}\cdot\nabla h\big] \approx 0 && \text{(4-13)}\\
\iff\ & \big[T_{b'}\cdot T_{\partial_w^2\Phi}\nabla h + 2T_{\nabla\partial_w\Phi}\cdot\nabla h\big] + \big[2T_{\partial_w^2\Phi}T_{\nabla h}\cdot\nabla h - 2T_{\nabla\partial_w\Phi}\cdot\nabla h\big] \approx 0 && \text{(4-14)}\\
\iff\ & T_{[b'\partial_w^2\Phi + 2\partial_w^2\Phi\nabla h]}\cdot\nabla h \approx 0 && \text{(4-15)}\\
\iff\ & 0 \approx 0, \quad \text{as } b' = -2\nabla h. && \text{(4-16)}
\end{aligned}
$$

Obviously, (4-16) holds. Hence, we can reverse the directions of all arrows back to conclude $PW \approx 0$.

Although tedious, it is not difficult to verify that all $\approx$ equivalence relations hold in all the above equations. As a typical example, we give a detailed proof of (4-13) here. To prove (4-13), it is sufficient to estimate $T_{\Delta\partial_w\Phi}h$. From the estimate (4-8) in Lemma 4.3, we have

$$\sup_{w\in[-1/4,0]} \|T_{\Delta\partial_w\Phi}h\|_{H^k} \lesssim \sup_{w\in[-1/4,0]} \|h\|_{H^k}\|\partial_w\Phi\|_{\widetilde{W}^2} \lesssim [\|\Lambda\psi\|_{\widehat{W}^{4,\alpha}} + \|(h,\Lambda\psi)\|^2_{\widehat{W}^4}]\|h\|_{H^k}.$$

Hence, the equivalence relation (4-13) holds. All other equivalence relations can be obtained very similarly.

The next step is to decompose the equation $PW \approx 0$ into a forward evolution equation and a backward evolution equation. As a result, from Lemma 4.6, we can show that $\partial_w W - T_A W$ actually does not lose derivatives. Note that $\partial_w W - T_A W \approx \partial_w \Phi - T_A(\Phi - T_{\partial_w \Phi} h)$. Hence, our desired result is obtained.

More precisely, we have the following lemma.

Lemma 4.4. *There exist two symbols $a = a(x,\xi)$ and $A(x,\xi)$ with*

$$a = a^{(1)} + a^{(0)}, \quad A = A^{(1)} + A^{(0)},$$

where

$$\begin{aligned} a^{(1)}(x,\xi) &= \frac{1}{1+|\nabla h|^2}\big(i\nabla h\cdot\xi - \sqrt{(1+|\nabla h|^2)|\xi|^2 - (\nabla h\cdot\xi)^2}\big),\\ A^{(1)}(x,\xi) &= \frac{1}{1+|\nabla h|^2}\big(i\nabla h\cdot\xi + \sqrt{(1+|\nabla h|^2)|\xi|^2 - (\nabla h\cdot\xi)^2}\big),\\ a^{(0)}(x,\xi) &= \frac{1}{A^{(1)} - a^{(1)}}\Big(i\,\partial_\xi a^{(1)}\cdot\partial_x A^{(1)} - \frac{\Delta h\ a^{(1)}}{1+|\nabla h|^2}\Big),\\ A^{(0)}(x,\xi) &= \frac{1}{a^{(1)} - A^{(1)}}\Big(i\,\partial_\xi a^{(1)}\cdot\partial_x A^{(1)} - \frac{\Delta h\ A^{(1)}}{1+|\nabla h|^2}\Big), \end{aligned} \tag{4-17}$$

such that

$$P = T_{a'}(\partial_w - T_a)(\partial_w - T_A) + R_0 + R_1\partial_w, \quad a'(a+A) = ib'\cdot\xi + c', \tag{4-18}$$

$$a'\Big[a^{(1)}A^{(1)} + \frac{1}{i}\partial_\xi a^{(1)}\cdot\partial_x A^{(1)} + a^{(1)}A^{(0)} + a^{(0)}A^{(1)}\Big] = a'(a\sharp A) = -|\xi|^2, \tag{4-19}$$

where

$$R_0 = T_{a'}T_aT_A - \Delta, \quad R_1 = -T_{a'}T_{a+A} + T_{b'}\cdot\nabla + T_{c'}. \tag{4-20}$$

Moreover, the following estimate holds for good error operators R_0 and R_1:

$$\|R_0 f\|_{H^k} + \|R_1 f\|_{H^{k+1}} \lesssim \|\nabla h\|_{\widetilde{W}^3}\|f\|_{H^k}. \tag{4-21}$$

Proof. Most parts of above lemma are cited directly from [Alazard, Burq and Zuily 2011, Lemma 3.18]. Given the a priori decomposition (4-18), from (4-11), we can calculate explicitly the formulae of R_0 and R_1, which are given in (4-20). Note that as a' doesn't depend on ξ, from (4-18)–(4-20), we have the identities

$$\begin{aligned} R_1 &= -T_{a'}T_{a+A} + T_{a'(a+A)} = -T_{a'}T_{a+A} + T_{(a'\sharp(a+A))},\\ R_0 &= T_{a'}[T_aT_A - T_{a\sharp A}] + T_{a'}T_{a\sharp A} - T_{a'(a\sharp A)} = T_{a'}[T_aT_A - T_{a\sharp A}] + T_{(a'-1)}T_{a\sharp A} - T_{(a'-1)\sharp(a\sharp A)}. \end{aligned}$$

From explicit formulations of a', a and A, we can see that $a', a'-1 \in \Gamma_2^0(\mathbb{R}^2)$ $a, A, a+A \in \Gamma_2^1(\mathbb{R}^2)$ and $a\sharp A \in \Gamma_2^2(\mathbb{R}^2)$. The following estimates on their symbolic bounds hold:

$$M_2^2(a\sharp A) + M_2^0(a') \lesssim 1, \quad M_2^1(a) + M_2^1(A) + M_2^1(a+A) \lesssim \|\nabla h\|_{\widetilde{W}^3}, \quad M_2^0(a'-1) \lesssim \|\nabla h\|^2_{\widetilde{W}^3}.$$

From estimate (2-8) in Lemma 2.6, we have

$$\|R_1 f\|_{H^{k+1}} \lesssim M_2^0(a')M_2^1(a+A)\|f\|_{H^k} \lesssim \|\nabla h\|_{\widetilde{W}^3}\|f\|_{H^k},$$
$$\|R_0 f\|_{H^k} \lesssim \big[M_2^0(a')M_2^1(a)M_2^1(A) + M_2^0(a'-1)M_2^2(a\sharp A)\big]\|f\|_{H^k} \lesssim \|\nabla h\|_{\widetilde{W}^3}^2\|f\|_{H^k}.$$

Hence finishing the proof of (4-21). □

In the following lemma, we prove that $\partial_w W - T_A W$ doesn't lose a derivative.

Lemma 4.5. *Let $A(x,\xi)$ be as defined in Lemma 4.4. For $k \geq 1$, we have the estimate*

$$\big\|\Lambda_{\geq 2}[(\partial_w W - T_A W)]|_{w=0}\big\|_{H^k} \lesssim_k \big[\|(h,\Lambda\psi)\|_{\widehat{W}^{4,\alpha}} + \|(h,\Lambda\psi)\|_{\widehat{W}^4}^2\big]\big[\|h\|_{H^k} + \|\nabla\psi\|_{H^{k-1}}\big]. \tag{4-22}$$

Proof. Recalling the decomposition of the operator P in (4-18) and the fact that $PW \approx 0$, we have

$$T_{a'}(\partial_w - T_a)(\partial_w - T_A)W \approx -R_0 W - R_1\partial_w W,$$

which further gives us

$$(\partial_w - T_a)(\partial_w - T_A)W \approx \tilde{g},$$

where

$$\tilde{g} = T_{a'^{-1}}[-R_0 W - R_1\partial_w W] + [I - T_{a'^{-1}}T_{a'}](\partial_w - T_a)(\partial_w - T_A)W,$$
$$= T_{a'^{-1}}[-R_0 W - R_1\partial_w W] + \big[I - T_1 + T_{(a'-1)(a'^{-1}-1)} - T_{(a'-1)}T_{(a'^{-1}-1)}\big](\partial_w - T_a)(\partial_w - T_A)W.$$

From the estimate (4-21) in Lemma 4.4, and the fact that $T_{(a'-1)(a'^{-1}-1)} - T_{(a'-1)}T_{(a'^{-1}-1)}$ is of order -2, we have

$$\sup_{w\in[-1/4,0]} \|\Lambda_{\geq 2}[\tilde{g}(w)]\|_{H^k} \lesssim \|\nabla h\|_{\widetilde{W}^3}\big[\|P_{\geq 1/2}[W]\|_{H^k} + \|\partial_w W\|_{H^{k-1}}\big]$$
$$\lesssim \big[\|(h,\Lambda\psi)\|_{\widehat{W}^{4,\alpha}} + \|(h,\Lambda\psi)\|_{\widehat{W}^4}^2\big]\big[\|h\|_{H^k} + \|\nabla\psi\|_{H^{k-1}}\big].$$

Note that $[\partial_w^2 + \Delta]\Lambda_1[W(w)] = [\partial_w^2 + \Delta]\Lambda_1[\Phi(w)] = 0$ when $w \in \big[-\frac{1}{4}, 0\big]$; see (4-6). It is easy to see that we have the equivalence relation

$$(\partial_w - T_a)\Lambda_{\geq 2}[(\partial_w - T_A)W] + \Lambda_{\geq 2}\big[(\partial_w - T_a)\Lambda_1[(\partial_w - T_A)W]\big] \approx \Lambda_{\geq 2}[\tilde{g}]. \tag{4-23}$$

Note that

$$\Lambda_1[\partial_w\Phi] = \Lambda_1[\partial_z\varphi(w/(1+h))] = \frac{e^{(w+1)|\nabla|} - e^{-(w+1)|\nabla|}}{e^{-|\nabla|} + e^{|\nabla|}}|\nabla|\psi,$$
$$\Lambda_1[\Phi] = \Lambda_1[\varphi(w/(1+h))] = \frac{e^{-(w+1)|\nabla|} + e^{(w+1)|\nabla|}}{e^{-|\nabla|} + e^{|\nabla|}}\psi.$$

It is easy to verify that

$$\Lambda_1[(\partial_w - T_A)W] = \Lambda_1[\partial_w\Phi - T_{|\xi|}\Phi] \in H^\infty,$$
$$\big\|\Lambda_{\geq 2}\big[(\partial_w - T_a)\Lambda_1[(\partial_w - T_A)W]\big]\big\|_{H^k} \lesssim \|\nabla h\|_{\widetilde{W}^3}\|\nabla\psi\|_{H^{k-1}}.$$

Therefore, from (4-23), we have

$$(\partial_w - T_a)\Lambda_{\geq 2}[(\partial_w - T_A)W] \approx \Lambda_{\geq 2}[\tilde{g}].$$

We reformulate the above equation as

$$(\partial_w + T_{-a})\Lambda_{\geq 2}[(\partial_w - T_A)W] = \Lambda_{\geq 2}[\tilde{g}] + \hat{g},$$

where

$$\hat{g} = \text{error term from} \approx \text{equivalence relation.}$$

Recalling the precise formula of a in Lemma 4.4, we know that $-a$ satisfies the assumption in Lemma 4.6. We can first choose a series of constants $\{\tau_i\}_{i=1}^{k}$ such that $\tau_{i+1} = 4\tau_i$ and $\tau_k \geq -\frac{1}{5}$ and then keep iterating the estimate (4-25). As a result, we have the estimate

$$\begin{aligned}
&\big\|\Lambda_{\geq 2}[(\partial_w - T_A)W|_{w=0}]\big\|_{H^k} \\
&\qquad \lesssim_k \sup_{w\in[-1/5,0]} \big[\|\Lambda_{\geq 2}[(\partial_w - T_A)W(w,\cdot)]\|_{L^2} + \|\tilde{g}(w)\|_{H^{k-1+\epsilon}} + \|\hat{g}(w)\|_{H^{k-1+\epsilon}}\big] \\
&\qquad \lesssim_k \big[\|(h,\Lambda\psi)\|_{\widehat{W}^{4,\alpha}} + \|(h,\Lambda\psi)\|_{\widehat{W}^4}^2\big]\big[\|h\|_{H^k} + \|\nabla\psi\|_{H^{k-1}}\big],
\end{aligned} \tag{4-24}$$

which concludes the proof. ☐

Lemma 4.6. *Let $a \in \Gamma_2^1(\mathbb{R}^2)$ and suppose it satisfies the assumption* $\mathrm{Re}[a(x,\xi)] \geq c|\xi|$ *for some positive constant c. If u solves the equation*

$$(\partial_w + T_a)u(w,\cdot) = g(w,\cdot),$$

then we know that the following estimate holds for any fixed and sufficiently small constant τ, and arbitrarily small constant $\epsilon > 0$:

$$\sup_{w\in[\tau,0]} \|u(w)\|_{H^k} \lesssim M_2^1(a)\frac{1+|\tau|}{|\tau|}\Big[\sup_{z\in[4\tau,0]} \|u(w)\|_{H^{k-2(1-\epsilon)}} + \sup_{z\in[4\tau,0]} \|g(z)\|_{H^{\mu-(1-\epsilon)}}\Big]. \tag{4-25}$$

Proof. A detailed proof can be found in [Alazard and Delort 2015] by combining Lemma 2.2.7 and the proof of Lemma 2.2.8. ☐

4A2. *Paralinearization of the Dirichlet–Neumann operator.* In this subsubsection, we use the result we obtained in the last subsubsection, which is the fact that $\partial_w W - T_A W$ doesn't lose derivatives, to identify which part of the Dirichlet–Neumann operator loses derivatives.

Recall (3-15). For the reader's convenience, we rewrite it as

$$G(h)\psi = \big((1+|\nabla h|^2)\partial_w\Phi - \nabla h\cdot\nabla\Phi\big)\big|_{w=0}.$$

Define

$$\underline{V} := \nabla\Phi - \partial_w\Phi\nabla h, \quad \underline{V}\big|_{w=0} = V.$$

Now we let w be inside the range $\big[-\frac{1}{4}, 0\big]$ instead of being restricted to the boundary. By using (2-7) and Lemma 2.6, we have the paralinearization result

$$\begin{aligned}
+|\nabla h|^2)\partial_w\Phi - \nabla h\cdot\nabla\Phi &\approx T_{1+|\nabla h|^2}\partial_w\Phi + 2T_{\partial_w\Phi}T_{\nabla h}\cdot\nabla h - T_{\nabla h}\cdot\nabla\Phi - T_{\nabla\Phi}\cdot\nabla h \\
&\approx T_{1+|\nabla h|^2}\partial_w\Phi + T_{2\nabla h\partial_w\Phi - \nabla\Phi}\cdot\nabla h - T_{\nabla h}\cdot\nabla\Phi
\end{aligned}$$

$$
\begin{aligned}
&= T_{1+|\nabla h|^2}\partial_w(W+T_{\partial_w\Phi}h)+T_{2\nabla h\partial_w\Phi-\nabla\Phi}\cdot\nabla h-T_{\nabla h}\cdot\nabla(W+T_{\partial_w\Phi}h)\\
&\approx T_{1+|\nabla h|^2}\partial_w W+T_{2\nabla h\partial_w\Phi-\nabla\Phi}\cdot\nabla h-T_{\nabla h}\cdot\nabla W-T_{\nabla h\partial_w\Phi}\cdot\nabla h\\
&= T_{1+|\nabla h|^2}\partial_w W-T_{\nabla h}\cdot\nabla W-T_{\underline{V}}\cdot\nabla h\\
&= T_{1+|\nabla h|^2}[\partial_w W-T_A W]+[T_{1+|\nabla h|^2}T_A-T_{\nabla h}\cdot\nabla]W-T_{\underline{V}}\cdot\nabla h\\
&= T_\lambda W-T_{\underline{V}}\cdot\nabla h+T_{1+|\nabla h|^2}[\partial_w W-T_A W]+R_2W,
\end{aligned}
\tag{4-26}
$$

where

$$R_2:=[T_{1+|\nabla h|^2}T_A-T_{(1+|\nabla h|^2)A^{(1)}}],$$

and where λ is given in (4-2). In (4-26), we used the identity

$$\lambda=(1+|\nabla h|^2)A^{(1)}-i\xi\cdot\nabla h,$$

where $A^{(1)}$ is given in (4-17). Note that

$$R_2=T_{(1+|\nabla h|^2)}T_{A^{(0)}}+\big[T_{(1+|\nabla h|^2)}T_{A^{(1)}}-T_{(1+|\nabla h|^2)A^{(1)}}\big]=T_{a'}T_{A^0}+\big[T_{(a'-1)}T_{A^{(1)}}-T_{(a'-1)\sharp A^{(1)}}\big].$$

Now, it is easy to see that R_2 is an operator of order 0 with an upper bound given by $\|\nabla h\|_{\widetilde{W}^3}$. Hence, we have the estimate

$$
\begin{aligned}
\|R_2W|_{w=0}\|_{H^k} &= \big\|R_2[P_{\geq 1/2}[W]|_{w=0}]\big\|_{H^k}\\
&\lesssim \|\nabla h\|_{\widetilde{W}^3}\big\|P_{\geq 1/2}[\psi-T_{B(h)\psi}h]\big\|_{H^k}\\
&\lesssim \big[\|(h,\Lambda\psi)\|_{\widehat{W}^{4,\alpha}}+\|(h,\Lambda\psi)\|^2_{\widehat{W}^4}\big]\big[\|h\|_{H^k}+\|\nabla\psi\|_{H^{k-1}}\big].
\end{aligned}
\tag{4-27}
$$

Combining (4-26), (4-27), and the estimate (4-22) in Lemma 4.5, it's easy to see that Proposition 4.1 holds.

Now, we give the postponed proof of Lemma 4.3.

Proof of Lemma 4.3. For fixed $w\in\big[-\frac14,0\big]$, it's easy to see $\Phi(w)=\varphi(w/(1+h(x)))$ and that we have the identity

$$\nabla_{x,w}\Phi=\nabla_{x,z}\varphi(w/(1+h))+\left[\frac{-w\nabla h\partial_z\varphi(w/(1+h))}{(1+h)^2},\frac{-h\partial_z\varphi(w/(1+h))}{1+h}\right].\tag{4-28}$$

Therefore, we know that the leading term of $\nabla_{x,w}\Phi(w)$ is $\nabla_{x,z}\varphi(w/(1+h))$. Under the smallness estimate (3-1), to estimate $\nabla_{x,w}\Phi$, it is sufficient to estimate $\nabla_{x,z}\varphi(w/(1+h))$.

Recall that according to the fixed-point-type formulation of $\nabla_{x,z}\varphi$ in (3-35), we study the linear term on the right-hand side of (3-35) first. Define

$$
\begin{aligned}
p_\pm(w,x,\xi) &:= \pm\frac{e^{-(w+1)|\xi|}(e^{h(x)w|\xi|/(1+h(x))}-1)}{e^{-|\xi|}+e^{|\xi|}}+\frac{e^{(w+1)|\xi|}(e^{-h(x)w|\xi|/(1+h(x))}-1)}{e^{-|\xi|}+e^{|\xi|}}\\
&= \sum_{n\geq 1}\frac{1}{n!}\left[\pm\frac{e^{-(w+1)|\xi|}}{e^{|\xi|}+e^{-|\xi|}}(w|\xi|)^n\left(\frac{h(x)}{1+h(x)}\right)^n+\frac{e^{|\xi|}}{e^{|\xi|}+e^{-|\xi|}}e^{w|\xi|}(w|\xi|)^n\left(\frac{h(x)}{1+h(x)}\right)^n\right]\\
&= \sum_{n\geq 1}\frac{1}{n!}[\pm f_n^1(w,\xi)g_n(x)+f_n^2(w,\xi)g_n(x)],
\end{aligned}
$$

where

$$f_n^1(w,\xi) := \frac{e^{-(w+1)|\xi|}}{e^{|\xi|}+e^{-|\xi|}}(w|\xi|)^n, \quad f_n^2(w,\xi) := \frac{e^{|\xi|}}{e^{|\xi|}+e^{-|\xi|}}e^{w|\xi|}(w|\xi|)^n, \quad g_n(x) := \left(\frac{h(x)}{1+h(x)}\right)^n.$$

It is easy to verify that

$$\Lambda_1[\nabla_x\varphi](x, w/(1+h(x))) = \Lambda_1[\nabla_x\varphi](x,w) + P_1\psi(x,w),$$
$$\Lambda_1[\partial_z\varphi](x, w/(1+h(x))) = \Lambda_1[\partial_z\varphi](x,w) + P_2\psi(x,w),$$

where

$$P_1\psi(x,w) := \frac{1}{4\pi^2}\int_{\mathbb{R}^2} e^{ix\cdot\xi}\hat{\psi}(\xi) i\xi p_+(w,x,\xi)\,d\xi, \quad P_2\psi(x,w) := \frac{1}{4\pi^2}\int_{\mathbb{R}^2} e^{ix\cdot\xi}\hat{\psi}(\xi)|\xi| p_-(w,x,\xi)\,d\xi.$$

We will show that, under the smallness estimate (3-1), the size of $\Lambda_1[\nabla_{x,z}\varphi](x, w/(1+h(x)))$ is almost same as the size of $\Lambda_1[\nabla_{x,z}\varphi](x,w)$. For $k \in \mathbb{Z}$, we define

$$\begin{aligned} p_{\pm,k}(w,x,\xi) &:= \frac{1}{4\pi^2}\int_{\mathbb{R}^2} e^{ix\cdot\sigma}\mathcal{F}_x(p)(w,\sigma,\xi)\psi_k(\sigma)\,d\sigma = \frac{1}{4\pi^2}\sum_{n\geq 1}\frac{1}{n!}\big[\pm f_n^1(w,\xi) + f_n^2(w,\xi)\big]P_k[g_n](x), \\ P_{1,k}\psi(x,w) &:= \int_{\mathbb{R}^2} e^{ix\cdot\xi}\hat{\psi}(\xi) i\xi p_{+,k}(w,x,\xi)\,d\xi, \\ P_{2,k}\psi(x,w) &:= \int_{\mathbb{R}^2} e^{ix\cdot\xi}\hat{\psi}(\xi)|\xi| p_{-,k}(w,x,\xi)\,d\xi. \end{aligned} \tag{4-29}$$

Since $P_2\psi$ can be treated in the same way as $P_1\psi$, we only estimate $P_1\psi$ in detail here. We have the decomposition

$$P_1\psi = \sum_{k_1,k_2\in\mathbb{Z}} P_{1,k_1}\psi_{k_2} = \mathrm{I} + \mathrm{II}, \quad \mathrm{I} = \sum_{k_2\leq k_1} P_{1,k_2}\psi_{k_1}, \quad \mathrm{II} = \sum_{k_1\leq k_2} P_{1,k_2}\psi_{k_1}.$$

From the bilinear estimate of L^2–L^∞ type (2-3) in Lemma 2.1, it is easy to see that we have the following estimates

$$\|\mathrm{I}\|_{H^k} \lesssim \left(\sum_{k_1} 2^{2k_1+2kk_{1,+}}\|P_{k_1}\psi\|_{L^2}^2\left(\sum_{n\geq 1}\frac{1}{n!}\|P_{\leq k_1}g_n\|_{L^\infty}\right)^2\right)^{\frac{1}{2}} \lesssim \|\nabla\psi\|_{H^k}\|h\|_{\widetilde{W}^1},$$

$$\|\mathrm{II}\|_{H^s} \lesssim \sum_{n\geq 1}\frac{1}{n!}\sum_{k_2\leq k_1} 2^{k_2+kk_{1,+}}\|P_{k_1}g_n\|_{L^2}\|P_{k_2}\psi\|_{L^\infty} \lesssim \|h\|_{H^{s+1}}\|\nabla\psi\|_{\widetilde{W}^1},$$

$$\|\mathrm{I}\|_{\widetilde{W}^\gamma} \lesssim \|h\|_{\widetilde{W}^\gamma}\|\nabla\psi\|_{\widetilde{W}^\gamma}, \quad \|\mathrm{II}\|_{\widetilde{W}^\gamma} \lesssim \|h\|_{\widetilde{W}^\gamma}\|\nabla\psi\|_{\widetilde{W}^\gamma}.$$

Therefore

$$\begin{aligned} \big\|\Lambda_1[\nabla_{x,z}\varphi]\big(x, w/(1+h(x))\big)\big\|_{H^k} &\lesssim \|\nabla\psi\|_{H^k} + \|h\|_{H^{k+1}}\|\nabla\psi\|_{\widetilde{W}^1}, \\ \big\|\Lambda_1[\nabla_x\varphi]\big(x, w/(1+h(x))\big)\big\|_{\widetilde{W}^\gamma} &\lesssim \|\nabla\psi\|_{\widetilde{W}^\gamma} + \|h\|_{\widetilde{W}^\gamma}\|\nabla\psi\|_{\widetilde{W}^\gamma} \lesssim \|\nabla\psi\|_{\widetilde{W}^\gamma}, \\ \big\|\Lambda_1[\partial_z\varphi]\big(x, w/(1+h(x))\big)\big\|_{\widetilde{W}^\gamma} &\lesssim \|\Lambda^2\psi\|_{\widetilde{W}^\gamma} + \|h\|_{\widetilde{W}^\gamma}\|\nabla\psi\|_{\widetilde{W}^\gamma}. \end{aligned}$$

From the above estimates and (4-28), we conclude

$$\|\Lambda_1[\nabla_x\Phi]\|_{\widetilde{W}^\gamma} \lesssim \|\nabla\psi\|_{\widetilde{W}^\gamma}, \quad \|\Lambda_1[\partial_w\Phi]\|_{\widetilde{W}^\gamma} \lesssim \|\Lambda^2\psi\|_{\widetilde{W}^\gamma} + \|h\|_{\widetilde{W}^\gamma}\|\nabla\psi\|_{\widetilde{W}^\gamma}, \tag{4-30}$$

$$\begin{aligned}\|\Lambda_1[\nabla_{x,w}\Phi]\|_{H^k} &\lesssim \big\|\Lambda_1[\nabla_{x,z}\varphi]\big(x, w/(1+h)\big)\big\|_{H^k} + \big\|\Lambda_1[\nabla_{x,z}\varphi]\big(x, w/(1+h)\big)\big\|_{\widetilde{W}^0}\|h\|_{H^{k+1}} \\ &\lesssim \|\nabla\psi\|_{H^k} + \|h\|_{H^{k+1}}\|\nabla\psi\|_{\widetilde{W}^1}.\end{aligned} \tag{4-31}$$

Following a similar procedure, we can handle the integral part in (3-35) in the same way. Similar to what we did in the proof of Lemma 3.2, we use the size of the symbol directly when $|\xi| \le 1$ and estimate the associated kernel when $|\xi| \ge 1$. As a result, we have the estimates

$$\begin{aligned}&\big\|\text{all terms in the right-hand side of (3-35) except for the linear part } \big(x, w/(1+h(x))\big)\big\|_{H^k} \\ &\quad\lesssim \sum_{i=1}^{3} \|g_i(z,\cdot)\|_{L^\infty_z H^k} + \|h\|_{H^{k+1}}\|g_i(z,\cdot)\|_{L^\infty_z\widetilde{W}^0} \lesssim \|\nabla\psi\|_{\widetilde{W}^0}\|h\|_{H^{k+1}} + \|\nabla\psi\|_{H^k}\|h\|_{\widetilde{W}^0} \\ &\quad\lesssim \|\nabla\psi\|_{\widetilde{W}^0}\|h\|_{H^{k+1}} + \|\nabla\psi\|_{H^k}\|h\|_{\widetilde{W}^0}.\end{aligned} \tag{4-32}$$

$$\begin{aligned}&\big\|\text{all terms in the right-hand side of (3-35) except for the linear part } \big(x, w/(1+h(x))\big)\big\|_{\widetilde{W}^\gamma} \\ &\qquad\lesssim \sum_{i=1,2,3} \|g_i(z,\cdot)\|_{L^\infty_z\widetilde{W}^\gamma} \lesssim \|h\|_{\widetilde{W}^{\gamma+1}}\|\nabla\psi\|_{\widetilde{W}^\gamma}.\end{aligned} \tag{4-33}$$

From (4-28), (4-30)–(4-33), now it's easy to see that estimates (4-7) and (4-8) hold.

Now, we proceed to prove (4-9). From (4-28) and the same procedure as above, we have the estimate

$$\begin{aligned}\|\Lambda_{\ge 2}[\nabla_{x,w}\Phi]\|_{L^2} &\lesssim \big\|\Lambda_{\ge 2}[\nabla_{x,z}\varphi]\big(x, w/(1+h(x))\big)\big\|_{L^2} + \|h\|_{H^1}\big\|\partial_z\varphi\big(x, w/(1+h(x))\big)\big\|_{\widetilde{W}^0} \\ &\lesssim \|h\|_{H^1}\big[\|\Lambda\psi\|_{\widehat{W}^{2,\alpha}} + \|h\|_{\widetilde{W}^1}\|\nabla\psi\|_{\widetilde{W}^1}\big] + \sum_{i=1}^{3}\|g_i\|_{L^\infty_z L^2}.\end{aligned} \tag{4-34}$$

Recall (3-25) and (3-26). Note that $\nabla\varphi$ appears together with ∇h inside the quadratic terms of $g_i(z)$, $i \in \{1,2,3\}$. When estimating the $L^\infty_z L^2$ norm of $g_i(z)$, $i \in \{1,2,3\}$, we always put $\nabla\varphi$ in L^2 and put $\partial_z\varphi$ in L^∞. As a result, the following estimate holds, i.e., our desired estimate (4-9) holds:

$$\begin{aligned}\text{(4-34)} &\lesssim \|h\|_{H^1}\|\Lambda\psi\|_{\widehat{W}^{2,\alpha}} + \big[\|(h,\Lambda\psi)\|_{\widehat{W}^{2,\alpha}} + \|(h,\Lambda\psi)\|_{\widehat{W}^2}\big](\|h\|_{H^1} + \|\nabla\psi\|_{L^2}) \\ &\lesssim \big[\|(h,\Lambda\psi)\|_{\widehat{W}^{2,\alpha}} + \|(h,\Lambda\psi)\|_{\widehat{W}^2}\big](\|h\|_{H^1} + \|\nabla\psi\|_{L^2}).\end{aligned}$$

□

4B. ***Paralinearization of the equation satisfied by the velocity potential.*** In this subsection, our main goal is to do the paralinearization process for the nonlinearity of the equation satisfied by ψ in (1-7), which shows which part of the nonlinearity actually loses derivatives.

More precisely, the main result of this subsection is stated in the following proposition,

Proposition 4.7. *We have the paralinearization*

$$\tfrac{1}{2}|\nabla\psi|^2 - \frac{(\nabla h\cdot\nabla\psi + G(h)\psi)^2}{2(1+|\nabla h|^2)} \approx T_V\cdot\nabla[\psi - T_{B(h)\psi}h] - T_B G(h)\psi \tag{4-35}$$

for the nonlinearity of the equation satisfied by ψ.

Proof. Recall that $V=\nabla\psi-\nabla hB$. From (2-7) and the composition Lemma 2.6, we have

$$
\begin{aligned}
\tfrac12|\nabla\psi|^2-\frac{(\nabla h\cdot\nabla\psi+G(h)\psi)^2}{2(1+|\nabla h|^2)}&\\
&=\tfrac12|\nabla\psi|^2-\tfrac12(1+|\nabla h|^2)B^2\\
&=\tfrac12|V|^2+V\cdot\nabla hB+\tfrac12|\nabla h|^2B^2-\tfrac12(1+|\nabla h|^2)B^2=\tfrac12|V|^2+V\cdot\nabla hB-\tfrac12B^2\\
&\approx T_V\cdot V+T_B(V\cdot\nabla h)+T_{V\cdot\nabla h}B-T_BB=T_V\cdot V+T_{V\cdot\nabla h}B-T_BG(h)\psi\\
&=T_V\cdot\nabla\psi-T_V\cdot(\nabla hB)+T_{V\cdot\nabla h}B-T_BG(h)\psi\\
&\approx T_V\cdot\nabla\psi-T_VT_B\cdot\nabla h-[T_V\cdot T_{\nabla h}-T_{V\cdot\nabla h}]B-T_BG(h)\psi\\
&\approx T_V\cdot[\nabla\psi-T_B\cdot\nabla h]-T_BG(h)\psi\approx T_V\cdot\nabla[\psi-T_{B(h)\psi}h]-T_BG(h)\psi.
\end{aligned}
$$
□

4C. *Symmetrization of the full system.* Based on the paralinearization results we obtained in previous subsections, in this subsection, we will find out the good substitution variables by doing the symmetrization process such that the resulting system has the requisite symmetric structures inside.

Define $\omega=\psi-T_{B(h)\psi}h$, which is the so-called good unknown variable. After combining the good decomposition (4-1) in Proposition 4.1 and the good decomposition (4-35) in Proposition 4.7, we reduce the system of equations satisfied by h and ψ to the system of equations satisfied by h and ω,

$$
\begin{cases}\partial_t h=\Lambda^2\omega+T_{\lambda-|\xi|}\omega-T_V\cdot\nabla h+\widetilde{F}(h)\psi,\\ \partial_t\omega=-T_ah-T_V\cdot\nabla\omega+f',\end{cases}\tag{4-36}
$$

where

$$
a:=1+\partial_tB+V\cdot\nabla B,
$$

which is the so-called Taylor coefficient, and f' is a good error term in the sense of estimate (4-5).

However, the system (4-36) cannot be used to do the energy estimate. When using the system (4-36) to do the energy estimate, one might find that the term

$$
\int_{\mathbb{R}^2}\partial_x^Nh\,\partial_x^N[T_{\lambda-|\xi|}\omega]+\partial_x^N\Lambda\omega\partial_x^N\Lambda[-T_ah],\quad\text{where } N \text{ is the prescribed top derivative level,}\tag{4-37}
$$

loses one derivative and cannot be simply treated.

To get around this difficulty, we will symmetrize the system (4-36) by following the same procedures in [Alazard, Burq and Zuily 2014a]. Define

$$
U_1:=h+T_\alpha h,\quad U_2:=\Lambda[\omega+T_\beta\omega]\approx[T_{\sqrt{\lambda}-|\xi|^{1/2}}\omega]+\Lambda\omega,\quad U:=U_1+iU_2,\tag{4-38}
$$

where

$$
\alpha:=\sqrt{a}-1,\quad \omega:=\psi-T_{B(h)\psi}h,\quad \beta:=\sqrt{\lambda/|\xi|}-1=\sqrt[4]{(1+|\nabla h|^2)-(\nabla h\cdot\xi/|\xi|)^2}-1.\tag{4-39}
$$

Note that

$$
\Lambda_1[a]=\Lambda_1[\partial_t\Lambda^2\psi]=-\Lambda^2h,\quad \Lambda_1[\partial_ta]=-\Lambda^4\psi,\quad \Lambda_1[\partial_t\alpha]=-\tfrac12\Lambda^4\psi,\quad \Lambda_1[\alpha]=-\tfrac12\Lambda^2h.
$$

We take the estimates of the Taylor coefficient in Lemma 4.8 for granted first. Then it is easy to see that the following estimate and equivalence relations hold:

$$\|(U_1, U_2)-(h, \Lambda\psi)\|_{H^k} \lesssim \|(h, \Lambda\psi)\|_{\widetilde{W}^3}\|(h, \Lambda\psi)\|_{H^k}, \tag{4-40}$$

$$\begin{aligned}
\partial_t U_1 &\approx \Lambda^2\omega + [T_{\lambda-|\xi|}\omega] - T_V\cdot\nabla h + T_{\partial_t\alpha}h + T_\alpha[\partial_t h] \\
&\approx [\Lambda^2 - T_{|\xi|}]\omega + T_\lambda\omega - T_V\cdot\nabla U_1 + T_V\cdot T_\alpha\nabla h + T_\alpha\big[T_\lambda\omega - T_V\cdot\nabla h\big] \\
&\approx [\Lambda^2 - T_{|\xi|}]\Lambda^{-1}U_2 + T_{\lambda\sqrt{a}}\omega - T_V\cdot\nabla U_1 \approx \Lambda U_2 + [T_{\sqrt{\lambda}\alpha}T_{\sqrt{\lambda}}\omega] - T_V\cdot\nabla U_1 \\
&\approx \Lambda U_2 + [T_{\sqrt{\lambda}\alpha}U_2] - T_V\cdot\nabla U_1,
\end{aligned} \tag{4-41}$$

$$\begin{aligned}
\partial_t U_2 &\approx \Lambda(1+T_\beta)[-T_a h - T_V\cdot\nabla\omega] + \Lambda T_{\partial_t\beta}\omega \approx -\Lambda U_1 - [T_{\sqrt{\lambda}}T_\alpha U_1] - T_V\cdot\nabla T_{\sqrt{\lambda}}\omega + T_{\partial_t\beta}\Lambda\omega \\
&\approx -\Lambda U_1 - [T_{\sqrt{\lambda}\alpha}U_1] - T_V\cdot\nabla U_2 + T_{\partial_t\beta}U_2.
\end{aligned} \tag{4-42}$$

Hence, the problematic terms in (4-37) become the terms (modulo good error terms)

$$\int_{\mathbb{R}^2} \partial_x^N U_1 \partial_x^N [T_{\sqrt{\lambda}\alpha}U_2] - \partial_x^N U_2 \partial_x^N [T_{\sqrt{\lambda}\alpha}U_1], \quad \text{where } N \text{ is the prescribed top derivative level.} \tag{4-43}$$

Therefore, we can move derivatives in (4-43) around so that these cubic terms do not lose derivatives. See (5-5) for more details.

4D. *Estimates of the Taylor coefficient.* The main goal of this subsection is to obtain some basic estimates for the Taylor coefficient, which are necessary for the energy estimate.

Lemma 4.8. *Under the smallness estimate (3-1), for $\gamma \le 3$, $\gamma_1 \le 2$ we have the estimates*

$$\begin{aligned}
\|a-1\|_{H^k} &\lesssim \|\nabla h\|_{H^k} + \|(h, \nabla\psi)\|_{H^k}\|(h, \nabla\psi)\|_{\widetilde{W}^2}, \\
\|a-1\|_{\widetilde{W}^\gamma} &\lesssim \|\nabla h\|_{\widetilde{W}^\gamma} + \big[\|\nabla\psi\|_{\widetilde{W}^{\gamma+1}} + \|\nabla h\|_{\widetilde{W}^\gamma}\big]^2, \\
\|\partial_t a\|_{H^k} &\lesssim \|\nabla\psi\|_{H^{k+1}} + \|(h, \nabla\psi)\|_{\widetilde{W}^3}\|(h, \nabla\psi)\|_{H^{k+1}}, \\
\|\partial_t a\|_{\widetilde{W}^{\gamma_1}} &\lesssim \|\Lambda^2\psi\|_{\widetilde{W}^{\gamma_1+1}} + \|(\nabla h, \nabla\psi)\|^2_{\widetilde{W}^{\gamma_1+1}}.
\end{aligned}$$

Proof. Recall (4-3), (1-7), and

$$a = 1 + V\cdot\nabla B + \partial_t B, \quad B = \partial_z\varphi/(1+h)|_{z=0}.$$

To estimate a and $\partial_t a$, it is sufficient to estimate $\partial_z\partial_t\varphi$ and $\partial_z\partial_t^2\varphi$. From the fixed-point-type formulation of $\nabla_{x,z}\varphi$ in (3-35), we can derive the equality

$$\begin{aligned}
\nabla_{x,z}\partial_t\varphi = &\left[\left[\frac{e^{-(z+1)|\nabla|}+e^{(z+1)|\nabla|}}{e^{-|\nabla|}+e^{|\nabla|}}\right]\nabla\partial_t\psi, \frac{e^{(z+1)|\nabla|}-e^{-(z+1)|\nabla|}}{e^{-|\nabla|}+e^{|\nabla|}}|\nabla|\partial_t\psi\right] + [\mathbf{0}, \partial_t g_1(z)] + \\
&+ \int_{-1}^0 [K_1(z,s) - K_2(z,s) - K_3(z,s)](\partial_t g_2(s) + \nabla\cdot\partial_t g_3(s))\,ds \\
&\quad + \int_{-1}^0 K_3(z,s)|\nabla|\operatorname{sign}(z-s)\partial_t g_1(s) - |\nabla|[K_1(z,s)+K_2(z,s)]\partial_t g_1(s)\,ds.
\end{aligned} \tag{4-44}$$

Following the same fixed-point-type argument that we used in the proof of Lemma 3.3, we can derive the estimates

$$
\begin{aligned}
&\|\nabla_{x,z}\partial_t\varphi\|_{L^\infty_z\widetilde{W}^\gamma}\\
&\qquad\lesssim \|\nabla h\|_{\widetilde{W}^\gamma}+\|\partial_t h\|_{\widetilde{W}^{\gamma+1}}\|\nabla_{x,z}\varphi\|_{L^\infty_z\widetilde{W}^\gamma}\lesssim \|\nabla h\|_{\widetilde{W}^\gamma}+\big[\|\nabla\psi\|_{\widetilde{W}^{\gamma+1}}+\|\nabla h\|_{\widetilde{W}^\gamma}\big]^2, \qquad (4\text{-}45)\\
&\|\nabla_{x,z}\partial_t\varphi\|_{L^\infty_z H^k}\\
&\qquad\lesssim \|\nabla\partial_t\psi\|_{H^k}+\|h\|_{H^{k+1}}\|\nabla_{x,z}\partial_t\varphi\|_{\widetilde{W}^1}+\|\partial_t h\|_{H^{k+1}}\|\nabla_{x,z}\varphi\|_{\widetilde{W}^1}+\|\partial_t h\|_{\widetilde{W}^1}\|\nabla_{x,z}\varphi\|_{L^\infty_z H^k}\\
&\qquad\lesssim \|\nabla h\|_{H^k}+\|(h,\nabla\psi)\|_{H^k}\|(h,\nabla\psi)\|_{\widetilde{W}^2}.
\end{aligned}
$$

We can take another time derivative at both sides of (4-44) to derive a fixed-point-type formulation for $\nabla_{x,z}\partial_t^2\varphi$. Following a similar argument, we can derive the estimate

$$\|\nabla_{x,z}\partial_t^2\varphi\|_{L^\infty_z\widetilde{W}^\gamma}\lesssim \|\partial_t^2\psi\|_{\widetilde{W}^{\gamma+1}}+\|\partial_t h\|_{\widetilde{W}^{\gamma+1}}\|\partial_t\nabla_{x,z}\varphi\|_{L^\infty_z\widetilde{W}^\gamma}+\|\partial_t^2 h\|_{\widetilde{W}^{\gamma+1}}\|\nabla_{x,z}\varphi\|_{L^\infty_z\widetilde{W}^\gamma}.$$

Recalling the system of equations satisfied by h and ψ in (1-7), we have

$$
\begin{aligned}
\partial_t^2 h &= \partial_t G(h)\psi = \partial_t[(1+|\nabla h|^2)B-\nabla h\cdot\nabla\psi],\\
\partial_t^2\psi &= -\partial_t h+\nabla\psi\cdot\nabla\partial_t\psi+(1+|\nabla h|^2)B\partial_t B+\nabla h\cdot\nabla\partial_t h B^2.
\end{aligned}
$$

Hence,

$$\|\partial_t^2\psi\|_{\widetilde{W}^{\gamma+1}}\lesssim \|\Lambda^2\psi\|_{\widetilde{W}^{\gamma+1}}+\big[\|\nabla h\|_{\widetilde{W}^{\gamma+1}}+\|\nabla\psi\|_{\widetilde{W}^{\gamma+1}}\big]^2.$$

Combining the above estimate, (4-45) and (3-50) in Lemma 3.3, we have

$$\|\nabla_{x,z}\partial_t^2\varphi\|_{L^\infty_z\widetilde{W}^\gamma}\lesssim \|\Lambda^2\psi\|_{\widetilde{W}^{\gamma+1}}+\big[\|\nabla h\|_{\widetilde{W}^{\gamma+1}}+\|\nabla\psi\|_{\widetilde{W}^{\gamma+1}}\big]^2.$$

Following the same argument, we derive the L^2-type estimate of $\partial_t^2\varphi$,

$$
\begin{aligned}
\|\nabla_{x,z}\partial_t^2\varphi\|_{L^\infty_z H^k}\lesssim{}& \|\partial_t^2\psi\|_{H^{k+1}}+\|\partial_t h\|_{H^{k+1}}\|\nabla_{x,z}\partial_t\varphi\|_{L^\infty_z\widetilde{W}^1}+\|\partial_t h\|_{\widetilde{W}^1}\|\nabla_{x,z}\partial_t\varphi\|_{L^\infty_z H^k}\\
&+\|\partial_t^2 h\|_{H^k}\|\nabla_{x,z}\varphi\|_{L^\infty_z\widetilde{W}^1}+\|\partial_t^2 h\|_{\widetilde{W}^1}\|\nabla_{x,z}\varphi\|_{L^\infty_z H^k}\\
&+\|h\|_{H^{k+1}}\|\partial_t^2\nabla_{x,z}\varphi\|_{L^\infty_z\widetilde{W}^1}+\|h\|_{\widetilde{W}^1}\|\nabla_{x,z}\partial_t^2\varphi\|_{L^\infty_z H^k},
\end{aligned}
$$

which further gives us the estimate

$$
\begin{aligned}
\|\nabla_{x,z}\partial_t^2\varphi\|_{L^\infty_z H^k}&\lesssim \|\Lambda^2\psi\|_{H^{k+1}}+\|\nabla h\|_{\widetilde{W}^2}\|\Lambda^2\psi\|_{H^{k+1}}+\|\nabla\psi\|_{\widetilde{W}^2}\|h\|_{H^{k+1}}\\
&\lesssim \|\nabla\psi\|_{H^{k+1}}+\|(h,\nabla\psi)\|_{\widetilde{W}^3}\|(h,\nabla\psi)\|_{H^{k+1}}.
\end{aligned}
$$

Therefore, our desired estimates of the Taylor coefficient hold. □

5. Energy estimate

The goal in this section is to prove our main result, Theorem 1.1. Since the energy of (U_1,U_2) is comparable with the energy of $(h,\Lambda\psi)$, see(4-40), it is sufficient to estimate the energy of (U_1,U_2). Let

N_0 be the prescribed top regularity level. From (4-41) and (4-42), we know that the system of equations satisfied by (U_1, U_2) is given by

$$\begin{cases} \partial_t U_1 - \Lambda U_2 = T_{\sqrt{\lambda}\alpha} U_2 - T_V . \nabla U_1 + \mathcal{R}_1, \\ \partial_t U_2 + \Lambda U_1 = -T_{\sqrt{\lambda}\alpha} U_1 - T_V \cdot \nabla U_2 + \mathcal{R}_2. \end{cases} \tag{5-1}$$

The precise formulations of good remainder terms $\mathcal{R}_1$ and $\mathcal{R}_2$ are not so important in the energy estimate. From (4-41) and (4-42), we know that they are good error terms, i.e.,

$$\|\mathcal{R}_1\|_{H^{N_0}} + \|\mathcal{R}_2\|_{H^{N_0}} \lesssim_{N_0} \big[\|(h, \Lambda\psi)\|_{\widehat{W}^{4,\alpha}} + \|(h, \Lambda\psi)\|^2_{\widehat{W}^4}\big] \|(\eta, \Lambda\psi)\|_{H^{N_0}}. \tag{5-2}$$

Define the energy of U_1 and U_2 as

$$E_{N_0}(t) := \frac{1}{2}\Big[\|U_1\|_{L^2} + \|U_2\|_{L^2} + \sum_{\substack{k+j=N_0 \\ 0\le k, j \in \mathbb{Z}}} \|\partial_1^k \partial_2^j U_1\|^2_{L^2} + \|\partial_1^k \partial_2^j U_2\|^2_{L^2}\Big]. \tag{5-3}$$

From (5-1), we have

$$\begin{aligned} \Big|\frac{d}{dt} E_{N_0}(t)\Big| &\lesssim \|(U_1, U_2)\|_{H^{N_0}} \|(\mathcal{R}_1, \mathcal{R}_2)\|_{H^{N_0}} \\ &\quad + \sum_{\substack{k+j=N_0 \\ 0\le k, j \in \mathbb{Z}}} \Big| \int_{\mathbb{R}^2} \big[\partial_1^k \partial_2^j U_1 \partial_1^k \partial_2^j [-T_V \cdot \nabla U_1] + \partial_1^k \partial_2^j U_2 \partial_1^k \partial_2^j [-T_V \cdot \nabla U_2]\big]\, dx \Big| \\ &\quad + \Big| \int_{\mathbb{R}^2} \partial_1^k \partial_2^j U_1 \partial_1^k \partial_2^j ([T_{\sqrt{\lambda}\alpha} U_2]) - \partial_1^k \partial_2^j U_2 \partial_1^k \partial_2^j ([T_{\sqrt{\lambda}\alpha} U_1]) \Big| \\ &\lesssim_{N_0} \big[\|\nabla V\|_{\widetilde{W}^1} + \|(h, \Lambda\psi)\|_{\widehat{W}^{4,\alpha}} + \|(h, \Lambda\psi)\|^2_{\widehat{W}^4}\big] \|(U_1, U_2)\|^2_{H^{N_0}} + \mathcal{E}_{N_0} \\ &\lesssim_{N_0} \big[\|(h, \Lambda\psi)\|_{\widehat{W}^{4,\alpha}} + \|(h, \Lambda\psi)\|^2_{\widehat{W}^4}\big] \|(U_1, U_2)\|^2_{H^{N_0}} + \mathcal{E}_{N_0}, \end{aligned} \tag{5-4}$$

where

$$\begin{aligned} \mathcal{E}_{N_0} &= \sum_{\substack{k+j=N_0 \\ 0\le k, j \in \mathbb{Z}}} \Big| \int_{\mathbb{R}^2} \big[\partial_1^k \partial_2^j U_1 T_{\sqrt{\lambda}\alpha} \partial_1^k \partial_2^j U_2 - \partial_1^k \partial_2^j U_2 T_{\sqrt{\lambda}\alpha} \partial_1^k \partial_2^j U_1\big] \Big| \\ &= \sum_{\substack{k+j=N_0 \\ 0\le k, j \in \mathbb{Z}}} \Big| \int_{\mathbb{R}^2} \big[\partial_1^k \partial_2^j U_1 \big(T_{\sqrt{\lambda}\alpha} - (T_{\sqrt{\lambda}\alpha})^*\big) \partial_1^k \partial_2^j U_2\big] \Big|. \end{aligned} \tag{5-5}$$

Recall that $\sqrt{\lambda}\alpha \in M_1^{1/2}$ is a symbol of order $\frac{1}{2}$. Note that it is real. Hence, from Lemma 2.8, we know that $(\sqrt{\lambda}\alpha)^*$ is $\sqrt{\lambda}\alpha$, and that the operator $(T_{\sqrt{\lambda}\alpha})^* - T_{\sqrt{\lambda}\alpha}$ is of order $-\frac{1}{2}$. As a result, the estimate

$$\mathcal{E}_{N_0} \lesssim M_1^{1/2}(\sqrt{\lambda}\alpha) \|(U_1, U_2)\|^2_{H^{N_0}} \lesssim \|\nabla h\|_{\widetilde{W}^2} \|(U_1, U_2)\|^2_{H^{N_0}} \tag{5-6}$$

holds. Combining the above estimate with (5-4) and (4-40), we have

$$\Big|\frac{d}{dt} E_{N_0}(t)\Big| \lesssim_{N_0} \big[\|(h, \Lambda\psi)\|_{\widehat{W}^{4,\alpha}} + \|(h, \Lambda\psi)\|^2_{\widehat{W}^4}\big] \|(h, \Lambda\psi)\|^2_{H^{N_0}}.$$

Appendix: Quadratic terms of the good remainders

In this section, we calculate explicitly the quadratic terms of the good reminder terms $\mathcal{R}_1$ and $\mathcal{R}_2$ to help readers understand the fact that we can gain one derivative in the new energy estimate (1-12) for the inputs of quadratic terms, which are put in the L^∞-type space. Recall (4-38) and (4-39). We have

$$\Lambda_1[B]=\Lambda^2\psi,\quad \Lambda_1[a]=\Lambda_1[\partial_t B]=-\Lambda^2 h,\quad \Lambda_1[\alpha]=-\tfrac{1}{2}\Lambda^2 h,\quad \Lambda_1[\beta]=0.$$

Recall (5-1). By using the above definitions, we can reduce the equations satisfied by U_1 and U_2 to the equations

$$\begin{cases}\partial_t h-\Lambda^2\psi=\widetilde{Q}_1(h,\psi)+\Lambda_2[\mathcal{R}_1](h,\psi)+\text{cubic and higher}_1,\\ \partial_t\Lambda\psi+\Lambda h=\widetilde{Q}_2(h,\psi)+\Lambda_2[\mathcal{R}_2](h,\psi)+\text{cubic and higher}_2,\end{cases}\tag{5-7}$$

satisfied by h and ψ, where

$$\widetilde{Q}_1(h,\psi)=-\Lambda^2(T_{\Lambda^2\psi}h)+\tfrac{1}{2}T_{\Lambda^2h}\Lambda^2\psi+\tfrac{1}{2}T_{\Lambda^4\psi}h-\tfrac{1}{2}(T_{\Lambda^2h}|\nabla|^{1/2}\Lambda\psi)-T_{\nabla\psi}\cdot\nabla h,$$
$$\widetilde{Q}_2(h,\psi)=\Lambda(T_{\Lambda^2\psi}\Lambda^2\psi-T_{\Lambda^2h}h)+\Lambda\big(\tfrac{1}{2}T_{\Lambda^2h}h\big)+\tfrac{1}{2}(T_{\Lambda^2h}|\nabla|^{1/2}h)-T_{\nabla\psi}\cdot\nabla\Lambda\psi.$$

Recall (1-7) and (3-55) in Lemma 3.4. We have

$$\begin{aligned}\Lambda_2[\mathcal{R}_1](h,\psi)&=\Lambda_2[G(h)\psi]-\widetilde{Q}_1(h,\psi)\\&=-\nabla\cdot(T_h\nabla\psi)-\nabla\cdot\mathcal{R}(h,\nabla\psi)-T_{\Delta^2\psi}h-\Lambda^2(T_h\Lambda^2\psi)\\&\qquad-\Lambda^2\mathcal{R}(h,\Lambda^2\psi)+\tfrac{1}{2}T_{\Lambda^2h}\Lambda(\Lambda-|\nabla|^{1/2})\psi-\tfrac{1}{2}T_{\Lambda^4\psi}h,\end{aligned}\tag{5-8}$$

$$\begin{aligned}\Lambda_2[\mathcal{R}_2](h,\psi)&=\Lambda\big[-\tfrac{1}{2}|\nabla\psi|^2+\tfrac{1}{2}|\Lambda^2\psi|^2\big]-\widetilde{Q}_2(h,\psi)\\&=\Lambda(-T_{\nabla\psi}\cdot\nabla\psi)+T_{\nabla\psi}\cdot\nabla\Lambda\psi\\&\qquad+\tfrac{1}{2}\big(-\Lambda(T_{\Lambda^2h}h)+T_{\Lambda^2h}|\nabla|^{1/2}h\big)-\tfrac{1}{2}\Lambda\mathcal{R}(\nabla\psi,\nabla\psi)+\tfrac{1}{2}\Lambda\mathcal{R}(\Lambda^2\psi,\Lambda^2\psi).\end{aligned}\tag{5-9}$$

Note that

$$\Lambda-|\nabla|^{1/2}=|\nabla|^{1/2}(\sqrt{\tanh|\nabla|}-1)=\frac{-2e^{-|\nabla|}|\nabla|^{1/2}}{(\sqrt{\tanh|\nabla|}+1)(e^{|\nabla|}+e^{-|\nabla|})}.\tag{5-10}$$

Now, it is easy to see that $\Lambda_2[\mathcal{R}_2](h,\psi)$ and $\Lambda_2[\mathcal{R}_2](h,\psi)$ do not lose derivatives. It remains to check that we can gain one derivative in the L^∞-type space. By (5-8) and (5-9), it is sufficient to check the term

$$-\nabla\cdot\big(T_h\nabla\psi\big)-\Lambda^2(T_h\Lambda^2\psi).\tag{5-11}$$

The corresponding symbol for the above quadratic terms is

$$\big(\xi\cdot\eta-|\xi||\eta|\tanh|\xi|\tanh|\eta|\big)\theta(\xi-\eta,\eta),\quad |\xi-\eta|\ll|\xi|\sim|\eta|.$$

We decompose this symbol into two parts:

$$p_1(\xi-\eta,\eta)=\xi\cdot\eta-|\xi||\eta|=-\tfrac{1}{2}|\xi-\eta|^2+\tfrac{1}{2}|\xi|^2+\tfrac{1}{2}|\eta|^2-|\xi||\eta|=-\tfrac{1}{2}|\xi-\eta|^2+\tfrac{1}{2}(|\xi|-|\eta|)^2,$$
$$p_2(\xi-\eta,\eta)=|\xi||\eta|(1-\tanh|\xi|\tanh|\eta|).$$

Now, it is clear that the first part of (5-11), which is determined by $p_1(\xi-\eta,\eta)$, does not lose derivatives and gains two derivatives for h. For the second part of (5-11), which is determined by $p_2(\xi-\eta,\eta)$, we can lower its regularity to L^2. Hence, we can place ψ in L^∞ and h in L^2. As a result, we always gain one derivative for inputs of quadratic terms that are in L^∞.

Acknowledgements

I would like to thank my Ph.D. advisor Alexandru Ionescu for many helpful discussions and suggestions. The first version of this paper was done when I was visiting Fudan University and BICMR, Peking University. I would like to thank Zhen Lei and BICMR for their warm hospitality during the visits.

References

[Alazard and Delort 2015] T. Alazard and J.-M. Delort, *Sobolev estimates for two dimensional gravity water waves*, Astérisque **374**, Société Mathématique de France, Paris, 2015. MR Zbl

[Alazard and Métivier 2009] T. Alazard and G. Métivier, "Paralinearization of the Dirichlet to Neumann operator, and regularity of three-dimensional water waves", *Comm. Partial Differential Equations* **34**:10–12 (2009), 1632–1704. 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 2014a] T. Alazard, N. Burq, and C. Zuily, "On the Cauchy problem for gravity water waves", *Invent. Math.* **198**:1 (2014), 71–163. MR Zbl

[Alazard, Burq and Zuily 2014b] T. Alazard, N. Burq, and C. Zuily, "Strichartz estimates and the Cauchy problem for the gravity water waves equations", preprint, 2014. arXiv

[Ambrose and Masmoudi 2005] D. M. Ambrose and N. Masmoudi, "The zero surface tension limit of two-dimensional water waves", *Comm. Pure Appl. Math.* **58**:10 (2005), 1287–1315. MR Zbl

[Beyer and Günther 1998] K. Beyer and M. Günther, "On the Cauchy problem for a capillary drop, I: Irrotational motion", *Math. Methods Appl. Sci.* **21**:12 (1998), 1149–1183. 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

[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

[Lannes 2005] D. Lannes, "Well-posedness of the water-waves equations", *J. Amer. Math. Soc.* **18**:3 (2005), 605–654. MR Zbl

[Métivier 2008] G. Métivier, *Para-differential calculus and applications to the Cauchy problem for nonlinear systems*, Centro di Ricerca Matematica Ennio De Giorgi (CRM) Series **5**, Edizioni della Normale, Pisa, 2008. MR Zbl

[Nalimov 1974] V. I. Nalimov, "The Cauchy–Poisson problem", *Dinamika Splošn. Sredy* **18** (1974), 104–210. In Russian. MR

[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

[Wang 2016] X. Wang, *Global solutions for the gravity water waves system: infinite depth setting and flat bottom setting*, Ph.D. thesis, Princeton University, 2016, available at http://search.proquest.com/docview/1845307995. MR

[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 1999] S. Wu, "Well-posedness in Sobolev spaces of the full water wave problem in 3-D", *J. Amer. Math. Soc.* **12**:2 (1999), 445–495. MR Zbl

[Yosihara 1982] H. Yosihara, "Gravity waves on the free surface of an incompressible perfect fluid of finite depth", *Publ. Res. Inst. Math. Sci.* **18**:1 (1982), 49–96. MR Zbl

[Zakharov 1968] V. E. Zakharov, "Stability of periodic waves of finite amplitude on the surface of a deep fluid", *J. Appl. Mech. Tech. Phys.* **9**:2 (1968), 190–194.

Received 18 Feb 2016. Revised 16 Jan 2017. Accepted 25 Feb 2017.

XUECHENG WANG: xuecheng@math.princeton.edu
Mathematics Department, Princeton University, Princeton, NJ 08544, United States

ANALYSIS AND PDE
Vol. 10, No. 4, 2017
dx.doi.org/10.2140/apde.2017.10.929

msp

IMPROVING BECKNER'S BOUND VIA HERMITE FUNCTIONS

PAATA IVANISVILI AND ALEXANDER VOLBERG

We obtain an improvement of the Beckner inequality $\|f\|_2^2 - \|f\|_p^2 \le (2-p)\|\nabla f\|_2^2$ valid for $p \in [1,2]$ and the Gaussian measure. Our improvement is essential for the intermediate case $p \in (1,2)$, and moreover, we find the natural extension of the inequality for any real p.

1. Introduction

1.1. *The history of the problem.* The Poincaré inequality [Nash 1958] for the standard Gaussian measure

$$d\gamma_n = \frac{e^{-|x|^2/2}}{\sqrt{(2\pi)^n}}\, dx$$

states that

$$\int_{\mathbb{R}^n} f^2\, d\gamma_n - \left(\int_{\mathbb{R}^n} f\, d\gamma_n\right)^2 \le \int_{\mathbb{R}^n} |\nabla f|^2\, d\gamma_n \tag{1}$$

for any smooth bounded function $f : \mathbb{R}^n \to \mathbb{R}$. Later William Beckner [1989] generalized (1) for any real power p, $1 \le p \le 2$, as follows:

$$\int_{\mathbb{R}^n} f^p\, d\gamma_n - \left(\int_{\mathbb{R}^n} f\, d\gamma_n\right)^p \le \frac{p(p-1)}{2} \int_{\mathbb{R}^n} f^{p-2} |\nabla f|^2\, d\gamma_n \tag{2}$$

for any smooth bounded $f : \mathbb{R}^n \to (0,\infty)$. We caution the reader that in [Beckner 1989], inequality (2) was formulated in a slightly different but equivalent form (see Theorem 1, inequality (3) in that paper). It should be also mentioned that in the case $p = 2$, inequality (2) does coincide with (1) for all $f \ge 0$ but it does not imply the Poincaré inequality for the functions taking the negative values, especially when $\int_{\mathbb{R}^n} f\, d\gamma_n = 0$. If $p \to 1+$ then (2) provides us with log-Sobolev inequality (see [Beckner 1989]). In general, the constant $p(p-1)/2$ is sharp in the right-hand side of (2), as can be seen for $n = 1$ on the test functions $f(x) = e^{\varepsilon x}$ by sending $\varepsilon \to 0$.

Later Beckner's inequality (2) was studied by many mathematicians for different measures, in different settings and for different spaces as well. We refer the reader to [Arnold et al. 2007; Da Pelo et al. 2016;

Volberg is partially supported by the NSF grant DMS-1600065 and by the Hausdorff Institute for Mathematics, Bonn, Germany. This paper is also based upon work supported by the National Science Foundation under Grant No. DMS-1440140 while the authors were in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Spring 2017 semester.
MSC2010: primary 42B37, 52A40, 35K55, 42C05, 60G15; secondary 33C15, 46G12.
Keywords: Poincaré inequality, log-Sobolev inequality, Sobolev inequality, Beckner inequality, Gaussian measure, log-concave measures, semigroups, Hermite polynomials, Hermite differential equation, confluent hypergeometric functions, Turán's inequality, error term in Jensen's inequality, phi-entropy, phi-Sobolev, F-Sobolev, phi-divergence, information theory, backwards heat, Monge–Amperè with drift, exterior differential systems.

Barthe et al. 2006; 2007; Barthe and Roberto 2003; Bobkov and Tetali 2003; Bobkov and Götze 1999; Boucheron et al. 2005; Chafaï 2004; Wang 2005; Latała and Oleszkiewicz 2000; Kolesnikov 2007].

An analysis done in [Ivanisvili and Volberg 2015c] indicates that the right-hand side of (2) can be improved. In the present paper we address this issue: what is the precise estimate of the difference given in the left-hand side of (2), and can the requirement $p \in [1,2]$ be avoided by slightly changing the right-hand side of (2)?

We give complete answers to these questions. For example, if $p = \frac{3}{2}$ we will obtain an improvement in Beckner's inequality (2):

$$\int_{\mathbb{R}^n} f^{3/2}\,d\gamma_n - \left(\int_{\mathbb{R}^n} f\,d\gamma_n\right)^{3/2} \le \int_{\mathbb{R}^n} \left(f^{3/2} - \frac{1}{\sqrt{2}}(2f - \sqrt{f^2+|\nabla f|^2})\sqrt{f + \sqrt{f^2+|\nabla f|^2}}\right) d\gamma_n. \tag{3}$$

The left-hand side of (3) coincides with the left-hand side of (2) for $p = \frac{3}{2}$, but the right-hand side of (3) is strictly smaller than the right-hand side in (2). Indeed, notice that we have the *pointwise* inequality

$$x^{3/2} - \frac{1}{\sqrt{2}}(2x - \sqrt{x^2+y^2})\sqrt{x+\sqrt{x^2+y^2}} \le \tfrac{3}{8}x^{-1/2}y^2 \quad \text{for all } x, y \ge 0, \tag{4}$$

which follows from the homogeneity, i.e., take $x = 1$, and the rest is a direct computation which follows by introducing a new variable

$$u := \sqrt{1+\sqrt{1+y^2}}.$$

As one can see, the improvement of Beckner's inequality (2) is essential. Indeed, if $y \to \infty$ then the right-hand side of (4) increases as y^2 whereas the left-hand side of (4) increases as $y^{3/2}$. Also notice that if $x \to 0$ then the difference of both sides of (4) tends to infinity. The only place where the quantities in (4) are comparable is when $y/x \to 0$.

1.2. ***Main results.*** Let k be a real parameter. Let $H_k(x)$ be the Hermite function which satisfies the Hermite differential equation

$$H_k'' - xH_k' + kH_k = 0, \qquad x \in \mathbb{R}, \tag{5}$$

and which grows relatively slowly, that is, $H_k(x) = x^k + o(x^k)$ as $x \to +\infty$. If k is a nonnegative integer then H_k is the *probabilists'* Hermite polynomial of degree k with the leading coefficient 1; for example, $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, etc. In general, for arbitrary $k \in \mathbb{R}$ one should think that H_k is the analytic extension of the Hermite polynomials in k (existence and many other properties will be mentioned in Section 2).

For $k \in \mathbb{R}$, let R_k be the rightmost zero of $H_k(x)$ (see Lemma 7). If $k \le 0$ then we set $R_k = -\infty$. Define $F_k(x)$ as

$$F_k\left(\left|\frac{H_k'(q)}{H_k(q)}\right|\right) = \frac{H_{k+1}(q)}{H_k^{1+1/k}(q)} \quad \text{for } q \in (R_k, \infty). \tag{6}$$

We will see in the next section that $F_k \in C^2([0,\infty))$ is well defined and $F_k(0) = 1$. Moreover, if $k > -1$ then F_k will be a decreasing concave function, and if $k < -1$ then F_k will be an increasing convex function.

One may observe that

$$F_1(y) = 1 - y^2, \quad F_2(y) = \frac{1}{\sqrt{2}}(2 - \sqrt{1+y^2})\sqrt{1+\sqrt{1+y^2}}.$$

If $k = 0$ then definition (6) should be understood in the limiting sense as

$$F_{\exp}(H_{-1}(q)) = q \exp\left(\alpha - \int_1^q H_{-1}(s)\, ds\right) \quad \text{for all } q \in \mathbb{R}, \tag{7}$$

where

$$\alpha = \int_1^\infty \left(H_{-1}(s) - \frac{1}{s}\right) ds \approx -0.266\ldots. \tag{8}$$

Theorem 1. *For any $p \in \mathbb{R} \setminus [0, 1]$ and any smooth bounded $f \geq 0$ with $\int_{\mathbb{R}^n} f^p \, d\gamma_n < \infty$ we have*

$$\int_{\mathbb{R}^n} f^p F_{1/(p-1)}\left(\frac{|\nabla f|}{f}\right) d\gamma_n \leq \left(\int_{\mathbb{R}^n} f \, d\gamma_n\right)^p. \tag{9}$$

The inequality is reversed if $p \in (0, 1)$.

Proposition 2. *We have*

$$1 - \frac{p(p-1)}{2} t^2 \leq F_{1/(p-1)}(t) \quad \textit{for all } t \geq 0, \ p \in (1, 2]. \tag{10}$$

It remains to notice that estimate (10) applied to (9) immediately gives (2).

The improvement will be essential when $t \to \infty$. For example, it will become clear in the next section that as $t \to \infty$ we have

$$F_{1/(p-1)}(t) \sim -t^p \left(H'_{1/(p-1)}(R_{1/(p-1)})\right)^{1-p} \quad \text{for } p > 1. \tag{11}$$

Another immediate application of Theorem 1 is the following corollary.

Corollary 3. *For any $p \in (1, 2]$ and any smooth bounded $f \geq 0$ we have*

$$\int_{\mathbb{R}^n} f^p \, d\gamma_n - \left(\int_{\mathbb{R}^n} f \, d\gamma_n\right)^p \leq \left(H'_{1/(p-1)}(R_{1/(p-1)})\right)^{1-p} \int_{\mathbb{R}^n} |\nabla f|^p \, d\gamma_n. \tag{12}$$

Estimate (12) will follow by showing that, for any $y \geq 0$, the map

$$x \to x^p - x^p F_{1/(1-p)}\left(\frac{y}{x}\right) \tag{13}$$

is decreasing for $x > 0$, and the limit $x \to 0$ gives (12) by (11).

Appearance of the roots of Hermite functions in (12) seems quite unexpected, especially when these estimates are obtained on the Hamming cube. For example, in [Ivanisvili and Volberg 2016] we were able to extend (12) to the Hamming cube but for a particular power $p = \frac{3}{2}$:

$$\mathbb{E} f^{3/2} - (\mathbb{E} f)^{3/2} \leq \frac{1}{\sqrt{2}} \mathbb{E} |\nabla f|^{3/2}, \quad f : \{-1, 1\}^n \to \mathbb{R}_+. \tag{14}$$

We refer the reader to that paper for the notations, and we notice that the result announced there is a counterpart of (9) for $p = \frac{3}{2}$ on the Hamming cube, where the identity $x^{3/2} F_2(y/x) = \Re(x + iy)^{3/2}$

was used. Next, let $A \subset \{-1,1\}^n$, and let $w_A(x)$ denote the number of edges containing x between the set A and its complement. Clearly $w_A(x)$ lives on the *boundary* of the set A: $w_A(x) = 4|\nabla \mathbb{1}_A|^2$. If A has cardinality 2^{n-1} then the classical edge-isoperimetric inequality [Harper 1966] states that $\sum_{x\in\{-1,1\}^n} w_A(x) \geq 2^n$. On the other hand, taking $f = \mathbb{1}_A$ in (14) gives

$$\sum_{x\in\{-1,1\}^n} w_A(x)^{3/4} \geq (2-\sqrt{2})2^n,$$

which is a new edge-isoperimetric inequality and does not follow from the classical one.

Theorem 1 generates several inequalities. If $p \to 1+$ then (9) gives the log-Sobolev inequality. If $p = 2$ then (9) provides us with the Poincaré inequality. If $p \to \pm\infty$ then we obtain a new Sobolev inequality:

Corollary 4. *For any smooth bounded f we have*

$$\int_{\mathbb{R}^n} \exp(f)\, F_{\exp}(|\nabla f|)\, d\gamma_n \leq \exp\left(\int_{\mathbb{R}^n} f\, d\gamma_n\right),$$

where $F_{\exp}$ is defined in (7).

Finally if $p \to 0$ we obtain a new "negative log-Sobolev" inequality:

Corollary 5. *For any smooth bounded $f \geq 0$ with $\int_{\mathbb{R}^n} \ln f\, d\gamma_n > -\infty$ we have*

$$\int_{\mathbb{R}^n} -\ln f\, d\gamma_n + \ln\left(\int_{\mathbb{R}^n} f\, d\gamma_n\right) \leq \int_{\mathbb{R}^n} -F_{-\ln}\left(\frac{|\nabla f|}{f}\right) d\gamma_n,$$

where $F_{-\ln}(t)$ is defined as

$$F_{-\ln}\left(\frac{H_{-2}(x)}{H_{-1}(x)}\right) = \int_1^x H_{-1}(s)\, ds - c + \ln H_{-1}(x), \qquad x \in \mathbb{R}.$$

All these estimates extend to uniformly log-concave probability measures in the following sense (for the proof see Section 3).

Corollary 6. *Let $d\mu = e^{-U}\, dx$ be a probability measure, where* $\operatorname{Hess} U \geq R \cdot \mathrm{Id}$ *for some $R > 0$. For any $p \in \mathbb{R}\setminus[0,1]$ and any smooth bounded $f \geq 0$ with $\int_{\mathbb{R}^n} f^p\, d\mu < \infty$ we have*

$$\int_{\mathbb{R}^n} f^p F_{1/(p-1)}\left(\frac{|\nabla f|}{f\sqrt{R}}\right) d\mu \leq \left(\int_{\mathbb{R}^n} f\, d\mu\right)^p. \tag{15}$$

The inequality is reversed if $p \in (0,1)$.

The limiting cases of (15) when $p \to \pm\infty$ and $p \to 0$ should be understood in the sense of functions $F_{\exp}$ and $F_{-\ln}$ as in Corollary 4 and Corollary 5.

To summarize, the current paper provides us with estimates of Φ-entropy (see [Chafaï 2004])

$$\mathbf{Ent}^{\Phi}_{\gamma_n}(f) := \int_{\mathbb{R}^n} \Phi(f)\, d\gamma_n - \Phi\left(\int_{\mathbb{R}^n} f\, d\gamma_n\right)$$

for the following examples:

- $\Phi(x) = x^p$ for $p \in \mathbb{R} \setminus [0, 1]$ using Theorem 1.
- $\Phi(x) = -x^p$ for $p \in (0, 1)$ using Theorem 1.
- $\Phi(x) = e^x$ using Corollary 4, or by taking $p \to \pm\infty$ in Theorem 1.
- $\Phi(x) = -\ln x$ using Corollary 5, or by taking $p \to 0$ in Theorem 1.
- $\Phi(x) = x \ln x$ by taking $p \to 1$ in Theorem 1.

2. The proof of the theorem

The proof of the theorem amounts to checking that the real-valued function

$$M(x, y) = x^p F_k\left(\frac{y}{x}\right), \quad k = \frac{1}{1-p}, \tag{16}$$

defined on $[\varepsilon, \infty) \times [0, \infty)$ for any $\varepsilon > 0$, obeys a necessary smoothness condition, has a boundary condition $M(x, 0) = x^p$ and satisfies the partial differential inequality

$$\begin{pmatrix} M_{xx} + M_y/y & M_{xy} \\ M_{xy} & M_{yy} \end{pmatrix} \leq 0, \tag{17}$$

with reversed inequality in (17) if $p \in (0, 1)$. Then by Theorem 1 in [Ivanisvili and Volberg 2015c] we obtain that

$$\int_{\mathbb{R}^n} f^p F_k\left(\frac{|\nabla f|}{f}\right) d\gamma_n = \int_{\mathbb{R}^n} M(f, |\nabla f|)\, d\gamma_n \leq M\left(\int_{\mathbb{R}^n} f\, d\gamma_n, 0\right) = \left(\int_{\mathbb{R}^n} f\, d\gamma_n\right)^p$$

for any smooth bounded $f \geq \varepsilon$, which is the statement of the theorem we want to prove (except we need to justify the passage to the limit $\varepsilon \to 0$ and this will be done later). Notice that the inequality is reversed if $p \in (0, 1)$; indeed, in this case we should work with $-M(x, y)$ instead of $M(x, y)$.

Next we will need some tools regarding the Hermite functions H_k.

2.1. *Properties of Hermite functions.*

H_k can be defined (see [Hayman and Ortiz 1975]) by

$$H_k(x) = -\frac{2^{-k/2} \sin(\pi k)\, \Gamma(k+1)}{2\pi} \sum_{n=0}^{\infty} \frac{\Gamma((n-k)/2)}{n!} (-x\sqrt{2})^n, \tag{18}$$

or in terms of the confluent hypergeometric functions (see [Durand 1975]) by

$$H_k(x) = \sqrt{\frac{2^k}{\pi}} \left[\cos\left(\frac{\pi k}{2}\right) \Gamma\left(\frac{k+1}{2}\right) {}_1F_1\left(-\frac{k}{2}, \frac{1}{2}; \frac{x^2}{2}\right) + t\sqrt{2} \sin\left(\frac{\pi k}{2}\right) \Gamma\left(\frac{k}{2}+1\right) {}_1F_1\left(\frac{1-k}{2}, \frac{3}{2}; \frac{x^2}{2}\right)\right]. \tag{19}$$

If k is a nonnegative integer then one should understand (18) and (19) in the limiting sense. Notice the recurrence properties

$$H_k'(x) = k H_{k-1}(x), \tag{20}$$

$$H_{k+1}(x) = x H_k(x) - H_k'(x). \tag{21}$$

These properties follow from (18) and the fact that $\Gamma(z+1) = z\Gamma(z)$.

We also notice that

$$H_k(x) := e^{x^2/4} D_k(x), \tag{22}$$

where $D_k(x)$ is the *parabolic cylinder function*; i.e., it is the solution of the equation

$$D_k'' + \left(k + \frac{1}{2} - \frac{x^2}{4}\right) D_k = 0.$$

Since $H_k(x)$ is an entire function in x and k (see [Temme 2015] for the parabolic cylinder function), sometimes it will be convenient to write $H(x,k)$ instead of $H_k(x)$. The precise asymptotic for $x \to +\infty$, $x > 0$ and any $k \in \mathbb{R}$ is given by

$$H_k(x) \sim x^k \cdot \sum_{n=0}^{\infty} (-1)^n \frac{(-k)_{2n}}{n!\,(2x^2)^n}. \tag{23}$$

Here $(a)_n = 1$ if $n = 0$, and $(a)_n = a(a+1)\cdots(a+n-1)$ if $n > 0$. When $x \to -\infty$ we have

$$H_k(x) \sim |x|^k \cos(k\pi) \sum_{n=0}^{\infty} (-1)^n \frac{(-k)_{2n}}{n!\,(2x^2)^n} + \frac{\sqrt{2\pi}}{\Gamma(-k)} |x|^{-k-1} e^{x^2/2} \sum_{n=0}^{\infty} \frac{(1+k)_{2n}}{n!\,(2x^2)^n}. \tag{24}$$

We refer the reader to [Temme 2015; Olver et al. 2010]. For instance, for (23) we can use the asymptotic formula (12.9.1) in [Olver et al. 2010] for the parabolic cylinder function. To verify (24) we can express $H_k(-x)$ as a linear combination of two parabolic cylinder functions but with argument x instead of $-x$, see (12.2.15) in [Olver et al. 2010], and then we can use (12.9.1) and (12.9.2) in the same paper.

Next we will need the result of Elbert and Muldoon [1999] which describes the behavior of the real zeros of $H_k(x)$ for any real k.

Lemma 7. *For $k \le 0$, the function $H_k(x)$ has no real zeros, and it is positive on the real axis. For $n < k \le n+1$, $n = 0, 1, \ldots,$ the function $H_k(x)$ has $n+1$ real zeros. Each zero is increasing function of k on its interval of definition.*

The proof of the lemma is Theorem 3.1 in [Elbert and Muldoon 1999]. It is explained in the paper that as k passes through each nonnegative integer n a new leftmost zero appears at $-\infty$, while the rightmost zero passes through the largest zero of $H_k(x)$. They also include more precise information about the asymptotic behavior of the zeros as $k \to \infty$.

Further we will need Turán's inequality for H_k for any real k.

Lemma 8. *We have Turán's inequality*:

$$H_k^2(x) - H_{k-1}(x) H_{k+1}(x) > 0 \quad \textit{for all } k \in \mathbb{R},\ x \ge L_k, \tag{25}$$

where L_k denotes the leftmost zero of H_k. If $k \le 0$ then $L_k = -\infty$.

The lemma is known as Turán's inequality when k is a nonnegative integer. Unfortunately we could not find the reference in the case when k is different from a positive integer; therefore we decided to include the proof of the lemma.

The following is borrowed from [Madhava Rao and Thiruvenkatachar 1949].

Proof. Take $f(x) = e^{-x^2/2}\big(H_k^2(x) - H_{k-1}(x)H_{k+1}(x)\big)$. Asymptotic formulas (23) and (24) imply that

$$\begin{aligned} \lim_{x\to+\infty} f(x) &= 0 \quad \text{for all } k \in \mathbb{R}, \\ f(x) &\sim \sqrt{2\pi}|x| > 0 \quad \text{for } x \to -\infty,\ k = 0, \\ f(x) &\sim \frac{2\pi e^{x^2/2}}{\Gamma(-k)\Gamma(-k+1)}|x|^{-2k-2} \quad \text{for } x \to -\infty,\ k \notin \{0\} \cup \mathbb{N}. \end{aligned} \tag{26}$$

On the other hand, notice that by (20) and (21) we have

$$f'(x) = -e^{-x^2/2} H_k H_{k-1}. \tag{27}$$

If $k \le 0$ then by Lemma 7 we have $f' < 0$, and because of the conditions $f(-\infty) = +\infty$ and $f(\infty) = 0$ we obtain that $f > 0$ on $\mathbb{R}$. To verify the statement for $k > 0$ we notice that

$$f''(x) = e^{-x^2/2}(H_k^2 - kH_{k-1}^2). \tag{28}$$

Now we notice that if $H_k(c) = 0$ then $H_{k-1}(c) \neq 0$. Indeed, assume to the contrary that $H_{k-1}(c) = 0$. Then by (20) we have $H_k'(c) = 0$ and by (5) we obtain $H_k''(c) = 0$, and again taking derivatives in (20) we obtain that $H_{k-2}(c) = 0$. Repeating this process we obtain that $H_{k-N}(c) = 0$ for any large integer $N > 0$. But this contradicts Lemma 7.

Thus by (27) and (28) we obtain that c is a local minimum of f if and only if $H_{k-1}(c) = 0$. Then $f(c) = e^{-x^2/2}H_k^2(c) > 0$. Finally we obtain that $f : [L_k, \infty) \to \mathbb{R}$ is positive on its local minimum points, $f(\infty) = 0$ and $f(L_k) > 0$ (because H_{k-1} and H_{k+1} have opposite signs at zeros of H_k by (21)). Therefore $f > 0$ on $[L_k, \infty) \to \mathbb{R}$ and the lemma is proved. □

Remark 9. If $k \in \mathbb{N}$ then H_k is the probabilists' Hermite polynomial of degree k, so $f(x)$ will be even and inequality (25) will hold for all $x \in \mathbb{R}$, which confirms the classical Turán's inequality. However, if $k > 0$ but $k \notin \mathbb{N}$ then (25) fails when $x \to -\infty$; see (26).

Finally the next corollary together with Lemma 7 implies that

$$\left|\frac{H_k'}{H_k}\right| = \operatorname{sign}(k)\frac{H_k'(q)}{H_k(q)}$$

is positive and decreasing for $q \in (R_k, \infty)$ and $k \in \mathbb{R} \setminus \{0\}$.

Corollary 10. *For any $x \ge L_k$ and any $k \in \mathbb{R} \setminus \{0\}$ we have*

$$\operatorname{sign}[(H_k')^2 - H_k H_k''] = \operatorname{sign}(k).$$

Proof. The proof follows from Lemma 8 and the identity

$$k(H_k^2 - H_{k-1}H_{k+1}) = (H_k')^2 - H_k H_k'' \tag{29}$$

by (5), (20) and (21). □

2.2. ***Checking the partial differential inequality.*** Let $p = 1 + 1/k$. Further we assume $k \neq 0, -1$. Define $F = F_k$ as in the Introduction:

$$F(t) = \frac{H_{k+1}(q)}{H_k^{1+1/k}(q)}, \quad \text{where } \left|\frac{H_k'(q)}{H_k(q)}\right| = t, \; q \in (R_k, \infty), \; t \in (0, \infty). \tag{30}$$

Notice that by Corollary 10, the function

$$\left|\frac{H_k'(q)}{H_k(q)}\right| = \operatorname{sign}(k)\frac{H_k'(q)}{H_k(q)}$$

is positive decreasing in q for $q \in (R_k, \infty)$; moreover, by (23) and (20) we have $H_k'(q)/H_k(q) \sim k/q$ when $q \to +\infty$. From the same asymptotic formulas it follows that when $t \to 0+$ we have

$$F(t) = 1 - \frac{p(p-1)}{2} t^2 + O(t^4).$$

Therefore F is a well-defined function and $F \in C^2([0, \infty))$.

Take a positive $\varepsilon > 0$ and define $M(x, y)$ as in (16):

$$M(x, y) := x^p F\left(\frac{y}{x}\right) \quad \text{for } y \geq 0, \; x > \varepsilon > 0. \tag{31}$$

Clearly $M(x, \sqrt{y}) \in C^2([\varepsilon, \infty) \times [0, \infty))$. By Theorem 1 in [Ivanisvili and Volberg 2015c] we have the inequality

$$\int_{\mathbb{R}^n} M(f, |\nabla f|)\, d\gamma_n \leq M\left(\int_{\mathbb{R}^n} f\, d\gamma_n, 0\right) \tag{32}$$

for all smooth bounded $f \geq \varepsilon$ if (17) holds. In terms of F (see (31)) condition (17) takes the form

$$tFF''p(p-1) + F'F'' - t(p-1)^2(F')^2 \geq 0, \quad \text{i.e., the determinant of (17) is nonnegative,} \tag{33}$$

$$F''(t+t^3) + F'(2t^2 + 1 - 2pt^2) + Fp(p-1)t \leq 0, \quad \text{i.e., the trace of (17) is nonpositive,} \tag{34}$$

where $t = y/x$ is the argument of F. In fact we will show that we have equality in (33) instead of inequality; therefore the sign of (17) will depend on the sign of the trace in (34). We will see that inequality (34) will be reversed for $p \in (0, 1)$.

From (30), (29), (20) and (21) we obtain

$$F'(t) = -\frac{k+1}{|k|}\frac{1}{H_k^{1/k}}, \tag{35}$$

$$F''(t) = \frac{F'}{|k|}\frac{H_k H_{k-1}}{H_k^2 - H_{k+1}H_{k-1}}, \tag{36}$$

$$F(t) = -\frac{|k|}{k+1}\frac{H_{k+1}}{H_k} F'. \tag{37}$$

If we plug (36) and (37) into (33) we obtain that the left-hand side of (33) is zero. If we plug (36) and (37) into (34) we obtain

$$\text{left-hand side of (34)} = \left[\frac{(kH_{k-1}^2 - H_k^2 + H_{k-1}H_{k+1})^2 + H_{k-1}^2 H_k^2}{H_k^2(H_k^2 - H_{k+1}H_{k-1})}\right] F'.$$

Thus the sign of left-hand side of (34) coincides with the sign of F', which coincides with $\operatorname{sign}(-(k+1))$. The condition $p \in \mathbb{R} \setminus [0, 1]$ implies that $k > -1$ and therefore (17) holds. The condition $p \in (0, 1)$ implies that $k < -1$ and therefore the inequality in (17) is reversed.

Thus we have obtained (32) for smooth bounded functions $f \geq \varepsilon$. Next we claim that for an arbitrary smooth bounded $f \geq 0$ with $\int_{\mathbb{R}^n} f^p \, d\gamma_n < \infty$, we can apply the inequality to $f_\varepsilon := f + \varepsilon$ and send ε to 0 in (9). Indeed, it follows from (6) and (23) that as $t \to \infty$ we have

$$F(t) \sim \begin{cases} -t^{1+1/k}(H_k'(R_k))^{-1/k} & \text{for } k > 0, \\ \operatorname{sign}(-1-k)\left(\dfrac{e^{t^2/2}\sqrt{2\pi}}{t|\Gamma(-1-k)|}\right)^{-1/k} |1+k|^{1+1/k} & \text{for } k < 0,\ k \neq -1. \end{cases} \tag{38}$$

Thus for $p > 1$, that is, $k > 0$, the claim about the limit follows from the estimate $|F(t)| \leq C_1 + C_2 t^p$ together with the Lebesgue dominated convergence theorem.

If $p < 0$, that is, $k \in (-1, 0)$, we rewrite (9) in a standard way as

$$\int_{\mathbb{R}^n} f_\varepsilon^p \, d\gamma_n - \left(\int_{\mathbb{R}^n} f_\varepsilon \, d\gamma_n\right)^p \leq \int_{\mathbb{R}^n} f_\varepsilon^p \left(1 - F\left(\frac{|\nabla f|}{f_\varepsilon}\right)\right) d\gamma_n. \tag{39}$$

Since f is bounded, $f \geq 0$ and $\int_{\mathbb{R}^n} f^p \, d\gamma_n < \infty$, there is no issue with the left-hand side of (39) when $\varepsilon \to 0$. For the right-hand side of (39) we notice that the function $x^p(1 - F(y/x))$ is nonnegative and decreasing in x. Then the claim follows from the monotone convergence theorem. The nonnegativity follows from the observation that $F(0) = 1$ and $F' < 0$ (see (35) where we have $k > -1$). The monotonicity follows from (6), (35), (20) and the straightforward computations

$$\frac{\partial}{\partial x}(x^p(1 - F(y/x))) = x^{p-1}(p - pF(t) + tF'(t)) = x^{p-1} p \left[1 - \frac{q}{H_k^{1/k}(q)}\right], \tag{40}$$

where

$$|k| \frac{H_{k-1}(q)}{H_k(q)} = t = \frac{y}{x}$$

and $q \in (R_k, \infty)$. The last expression in (40) is negative because

$$1 \geq F(t) = \frac{H_{k+1}}{H_k^{1+1/k}} = \frac{qH_k - kH_{k-1}}{H_k^{1+1/k}} > \frac{q}{H_k^{1/k}}.$$

Finally if $p \in (0, 1)$, that is, $k < -1$, we have the opposite inequality in (39). In this case the situation is absolutely the same as for $k \in (-1, 0)$ except now we should consider the function $x^p(F(y/x) - 1)$, which is nonnegative and decreasing in x; see (40). This finishes the proof of the theorem.

Now let us show Proposition 2. Since $F(0) = 1$, it is enough to show a stronger inequality, namely $F' + p(p-1)t \geq 0$. From (35) and the fact that $k > 1$ since $p \in (1, 2)$, we obtain that it is enough to show the inequality

$$-\frac{p}{H_k^{1/k}} + p(p-1)\frac{H_k'}{H_k} \geq 0 \quad \text{for all } k \geq 1,\ q \in (R_k, \infty).$$

Using (20) and $p = 1 + 1/k$ we notice that the inequality can be rewritten as $1 \geq H_k(q)/H_{k-1}^{k/(k-1)}(q)$ for all $q \in (R_k, \infty)$. To verify the last inequality recall that $F(0) = 1$ and $F'(t) < 0$. Therefore $F(t) \leq 1$. We also recall the definition of $F(t)$; see (30). It follows that $1 \geq F = H_{k+1}/H_k^{1+1/k}$ for all $k > 0$. The last inequality is the same as

$$1 \geq \frac{H_k(q)}{H_{k-1}^{k/(k-1)}(q)} \quad \text{for all } q \in (R_k, \infty),\ k > 1. \tag{41}$$

This finishes the proof of Proposition 2.

To verify Corollary 3 we only need to prove the monotonicity of the map (13) for $p \in (1, 2]$, that is, $k \geq 1$, and the rest will follow from (38). If $k = 1$ there is nothing to prove; therefore we assume $k > 1$. By (40) it is enough to show that $L(q) := H_k^{1/k}(q) - q \leq 0$ for $q \in (R_k, \infty)$. The growth condition (24) on H_k implies that $\lim_{q\to\infty} L(q) = 0$. If $L'(q) \geq 0$ then we are done. Using (20) we notice that $L'(q) \geq 0$ is equivalent to (41), which was already proved.

2.3. ***Proof of Corollaries 4 and 5.*** Notice that as $y \to 0$ we have

$$F_{\exp}(y) = 1 - \frac{y^2}{2} + O(y^4) \quad \text{and} \quad F_{-\ln}(y) = -\frac{y^2}{2} + O(y^4).$$

One can check that

$$M_{\exp}(x, y) := e^x F_{\exp}(y), \quad M_{\exp}(x, 0) = e^x, \quad M_{\exp}(x, \sqrt{y}) \in C^2(\mathbb{R} \times \mathbb{R}_+),$$
$$M_{-\ln}(x, y) := -\ln(x) + F_{-\ln}\left(\frac{y}{x}\right), \quad M_{-\ln}(x, 0) = -\ln x, \quad x > 0,$$

and $M_{-\ln}(x, \sqrt{y}) \in C^2([\varepsilon, \infty) \times \mathbb{R}^+)$ for any $\varepsilon > 0$. By straightforward computations we notice that if we set $\psi(q) = \alpha - \int_1^q H_{-1}(s)\,ds$ then using the identity $1 = qH_{-1}(q) + H_{-2}(q)$ we obtain

$$F_{\exp}(H_{-1}) = qe^{\psi}, \quad F'_{\exp}(H_{-1}) = -e^{\psi} \quad \text{and} \quad F''_{\exp}(H_{-1}) = -\frac{H_{-1}}{H_{-2}}.$$

Similarly we compute that

$$F'_{-\ln}\left(\frac{H_{-2}}{H_{-1}}\right) = -H_{-1} \quad \text{and} \quad F''_{-\ln}\left(\frac{H_{-2}}{H_{-1}}\right) = -\frac{H_{-2}H_{-1}^2}{H_{-1}^2 - H_{-2}}.$$

Next one notices that $M_{\exp}$ and $M_{-\ln}$ satisfy (17) (in fact the determinant of (17) is zero). Then by Theorem 1 in [Ivanisvili and Volberg 2015c] we obtain the corollaries. The passage to the limit for $M_{-\ln}(x, y)$ when $\varepsilon \to 0$ follows from the monotone convergence theorem. Indeed, we notice that $-F_{-\ln}(y/x) \geq 0$ is decreasing in x. We apply Corollary 5 to $f_\varepsilon = f + \varepsilon$ and send $\varepsilon \to 0$.

2.3.1. *How we guessed the functions $M_{\exp}$ and $M_{-\ln}$.* One may ask how to find the functions $M_{\exp}$ and $M_{-\ln}$. To find $M_{\exp}$ we should apply (9) to functions $f = e^{g/p}$, where g is some fixed function. Then (9) takes the form

$$\int_{\mathbb{R}^n} e^g F_{1/(p-1)}\left(\frac{|\nabla g|}{p}\right) d\gamma_n \le \left(\int_{\mathbb{R}^n} e^{g/p}\, d\gamma_n\right)^p. \tag{42}$$

Now we take $p \to \infty$. The right-hand side of (42) tends to $\exp(\int_{\mathbb{R}^n} g\, d\gamma_n)$. For the left-hand side of (42) we should compute the limit

$$F_{\exp}(t) := \lim_{p\to\infty} F_{1/(p-1)}\left(\frac{t}{p}\right) = \lim_{p\to\infty} F_{1/(p-1)}\left(\frac{t}{p-1}\right) = \lim_{k\to 0+} F_k(tk).$$

In fact all equalities can be justified by direct calculations using the fact that $H_k(x) = H(x,k)$ is the entire function of x and k; see [Temme 2015] for the parabolic cylinder function and formula (22).

It is clear that $F_{\exp}(0) = 1$. Next if we take $k \to 0+$ in (6) we obtain

$$\lim_{k\to 0+} F_k\left(\left|\frac{H_k'}{H_k}\right|\right) = \lim_{k\to 0+} F_k\left(k\frac{H_{k-1}}{H_k}\right) = \lim_{k\to 0+} F_k\left(k\frac{H_{-1}}{H_0}\right) = F_{\exp}(H_{-1}).$$

On the other hand, for the right-hand side of (6) we have

$$\lim_{k\to 0+} \frac{H_{k+1}(q)}{H_k^{1+1/k}} = q \lim_{k\to 0+} H_k^{-1/k}.$$

Here we have used $H_0(q) = 1$ and $H_1(q) = q$. Thus it remains to find $\lim_{k\to 0+} H_k^{-1/k}$. If we take the derivative in k of (20) we obtain $H_{xk}(x,k) = H(x,k-1) + kH_k(x,k)$ (here subindices denote partial derivatives). Now taking $k = 0$ we obtain $H_{xk}(x,0) = H(x,-1)$. Thus $H_k(x,0)$ is an antiderivative of $H(x,-1) = H_{-1}$. So

$$\lim_{k\to 0+} H_k^{-1/k} = \lim_{k\to 0+} \exp\left(-\frac{1}{k}\ln\bigl(1 + kH_k(x,0) + o(k)\bigr)\right) = \exp\left(-\int H_{-1}(s)\, ds\right).$$

Finally we obtain

$$F_{\exp}(H_{-1}(q)) = q\exp\left(C - \int_1^q H_{-1}\right). \tag{43}$$

In order to satisfy the condition $F_{\exp}(0) = 1$, the constant c must be chosen as $C = \int_1^\infty (H_{-1} - 1/s)\, ds$; indeed send $q \to \infty$ in (43). This gives Corollary 4. It is worth mentioning that we have also obtained

$$H_k(x,0) = \int_1^x H_{-1}(s)\, ds - \alpha;$$

see (8).

To find $M_{-\ln}$, let $F(x,k) := F_k(x)$. Let $F_k(x,k)$ denote the partial derivative in k of $F(x,k)$. If we send $p \to 0,\ p < 0$ in (9) and compare the terms of order p we obtain

$$\int_{\mathbb{R}^n} \left(\ln f - F_k\left(\frac{|\nabla f|}{f}, -1\right)\right) d\gamma_n \ge \ln\left(\int_{\mathbb{R}^n} f\, d\gamma_n\right).$$

It remains to find the function $F_k(x,-1)$. Let us equate terms of order $(k+1)$ as $k\to -1$, $k<-1$ in

$$F\left(\frac{|H_x(x,k)|}{H(x,k)},k\right)=\frac{H(x,k+1)}{H(x,k)^{1+1/k}}.$$

Straightforward computation shows that

$$F_k\left(\frac{H_{-2}(x)}{H_{-1}(x)},-1\right)=H_k(x,0)+\ln H_{-1}(x)=\int_1^x H_{-1}(s)\,ds-\alpha+\ln H_{-1}(x),$$

where

$$\alpha=\int_1^\infty\left(H_{-1}(s)-\frac{1}{s}\right)ds.$$

3. Concluding remarks

The reader may wonder how we guessed the choice (16). Of course it was not a random guess. Function (16) is the best possible in the sense that the determinant of (17) is identically zero:

$$\begin{gathered}M_{yy}\left(M_{xx}+\frac{M_y}{y}\right)-M_{xy}^2=0,\\ M(x,0)=x^p\quad\text{for } x\ge 0.\end{gathered}\tag{44}$$

Initially this was the way we started looking for $M(x,y)$ as the solution of the Monge–Ampère equation with a drift (44). By a proper change of variables, the equation reduces to the backwards heat equation (see [Ivanisvili and Volberg 2015c] for more details where the connection with the theory of exterior differential systems of R. Bryant et al. [1991] was exploited)

$$u_{xx}+u_t=0,\tag{45}$$

$$u(x,0)=Cx^{p/(p-1)}\quad\text{for } x\ge 0.\tag{46}$$

One can notice that the Hermite polynomials do satisfy (45) and (46) when $p/(p-1)$ is a positive integer. In general, one should invoke Hermite functions and this is the reason for the appearance of these functions in our theorem.

Another possibility is to assume that $M(x,y)$ should be homogeneous of degree p, which forces M to have the form (31) for some F. Next setting $h=F/F'$ and further by a subtle change of variables, one obtains Hermite differential equation (5).

Nevertheless, for the formal proof of Theorem 1 we do not need to go through the details. We have $M(x,y)$ defined by (16) and we just need to check that it satisfies the desired properties.

That $M(x,y)$ satisfies (17) makes it possible to extend Theorem 1 in a semigroup setting for uniformly log-concave probability measures. Indeed, let $d\mu=e^{-U}\,dx$, where $\operatorname{Hess} U\ge R\cdot\mathrm{Id}$, $R>0$. Let $L=\Delta-\nabla U\cdot\nabla$ and $P_t=e^{tL}$ be the semigroup with generator L; see [Ivanisvili and Volberg 2015c; Bakry et al. 2014].

Corollary 11. *For any $p\in\mathbb{R}\setminus[0,1]$ and any smooth bounded $f\ge 0$ with $\int_{\mathbb{R}^n} f^p\,d\mu<\infty$ we have*

$$P_t\left[f^pF_{1/(p-1)}\left(\frac{|\nabla f|}{f\sqrt{R}}\right)\right]\le (P_tf)^pF_{1/(p-1)}\left(\frac{|\nabla P_tf|}{P_tf\sqrt{R}}\right).$$

The inequality is reversed if $p\in(0,1)$.

Proof. Notice that $\widetilde{M}(x, y) = M(x, y/\sqrt{R})$ satisfies (17). Now it remains to use inequality (2.3) from [Ivanisvili and Volberg 2015c]. □

By taking $t \to \infty$ and using the fact that $|\nabla P_t f| \le e^{-tR} P_t |\nabla f|$, we obtain Corollary 6.

Finally we would like to mention that having characterization (17) of functional inequalities (32) makes our approach to problem (9) systematic. Very similar *local estimates* happen to rule some *global inequalities*. We refer the reader to our recent papers on this subject [Ivanisvili and Volberg 2015a;2015b; Ivanisvili 2016].

Acknowledgements

We are very grateful to Robert Bryant who suggested a change of variables in (33), and an anonymous referee for valuable suggestions.

References

[Arnold et al. 2007] A. Arnold, J.-P. Bartier, and J. Dolbeault, "Interpolation between logarithmic Sobolev and Poincaré inequalities", *Commun. Math. Sci.* **5**:4 (2007), 971–979. MR Zbl

[Bakry et al. 2014] D. Bakry, I. Gentil, and M. Ledoux, *Analysis and geometry of Markov diffusion operators*, Grundlehren der Mathematischen Wissenschaften **348**, Springer, 2014. MR Zbl

[Barthe and Roberto 2003] F. Barthe and C. Roberto, "Sobolev inequalities for probability measures on the real line", *Studia Math.* **159**:3 (2003), 481–497. MR Zbl

[Barthe et al. 2006] F. Barthe, P. Cattiaux, and C. Roberto, "Interpolated inequalities between exponential and Gaussian, Orlicz hypercontractivity and isoperimetry", *Rev. Mat. Iberoam.* **22**:3 (2006), 993–1067. MR Zbl

[Barthe et al. 2007] F. Barthe, P. Cattiaux, and C. Roberto, "Isoperimetry between exponential and Gaussian", *Electron. J. Probab.* **12**:44 (2007), 1212–1237. MR Zbl

[Beckner 1989] W. Beckner, "A generalized Poincaré inequality for Gaussian measures", *Proc. Amer. Math. Soc.* **105**:2 (1989), 397–400. MR Zbl

[Bobkov and Götze 1999] S. G. Bobkov and F. Götze, "Exponential integrability and transportation cost related to logarithmic Sobolev inequalities", *J. Funct. Anal.* **163**:1 (1999), 1–28. MR Zbl

[Bobkov and Tetali 2003] S. Bobkov and P. Tetali, "Modified log-Sobolev inequalities, mixing and hypercontractivity", pp. 287–296 in *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing* (San Diego, CA, 2003), ACM, New York, 2003. MR Zbl

[Boucheron et al. 2005] S. Boucheron, O. Bousquet, G. Lugosi, and P. Massart, "Moment inequalities for functions of independent random variables", *Ann. Probab.* **33**:2 (2005), 514–560. MR Zbl

[Bryant et al. 1991] R. L. Bryant, S. S. Chern, R. B. Gardner, H. L. Goldschmidt, and P. A. Griffiths, *Exterior differential systems*, Mathematical Sciences Research Institute Publications **18**, Springer, 1991. MR Zbl

[Chafaï 2004] D. Chafaï, "Entropies, convexity, and functional inequalities: on Φ-entropies and Φ-Sobolev inequalities", *J. Math. Kyoto Univ.* **44**:2 (2004), 325–363. MR Zbl

[Da Pelo et al. 2016] P. Da Pelo, A. Lanconelli, and A. I. Stan, "An extension of the Beckner's type Poincaré inequality to convolution measures on abstract Wiener spaces", *Stoch. Anal. Appl.* **34**:1 (2016), 47–64. MR Zbl

[Durand 1975] L. Durand, "Nicholson-type integrals for products of Gegenbauer functions and related topics", pp. 353–374 in *Theory and application of special functions* (Madison, WI, 1975), edited by R. A. Askey, Academic Press, New York, 1975. MR Zbl

[Elbert and Muldoon 1999] A. Elbert and M. E. Muldoon, "Inequalities and monotonicity properties for zeros of Hermite functions", *Proc. Roy. Soc. Edinburgh Sect. A* **129**:1 (1999), 57–75. MR Zbl

[Harper 1966] L. H. Harper, "Optimal numberings and isoperimetric problems on graphs", *J. Combinatorial Theory* **1** (1966), 385–393. MR Zbl

[Hayman and Ortiz 1975] W. K. Hayman and E. L. Ortiz, "An upper bound for the largest zero of Hermite's function with applications to subharmonic functions", *Proc. Roy. Soc. Edinburgh Sect. A* **75**:3 (1975), 182–197. MR Zbl

[Ivanisvili 2016] P. Ivanisvili, "Boundary value problem and the Ehrhard inequality", preprint, 2016. arXiv

[Ivanisvili and Volberg 2015a] P. Ivanisvili and A. Volberg, "Bellman partial differential equation and the hill property for classical isoperimetric problems", preprint, 2015. arXiv

[Ivanisvili and Volberg 2015b] P. Ivanisvili and A. Volberg, "Hessian of Bellman functions and uniqueness of the Brascamp–Lieb inequality", *J. Lond. Math. Soc.* (2) **92**:3 (2015), 657–674. MR Zbl

[Ivanisvili and Volberg 2015c] P. Ivanisvili and A. Volberg, "Isoperimetric functional inequalities via the maximum principle: the exterior differential systems approach", preprint, 2015. arXiv

[Ivanisvili and Volberg 2016] P. Ivanisvili and A. Volberg, "Poincaré inequality 3/2 on the Hamming cube", preprint, 2016. arXiv

[Kolesnikov 2007] A. V. Kolesnikov, "Modified log-Sobolev inequalities and isoperimetry", *Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl.* **18**:2 (2007), 179–208. MR Zbl

[Latała and Oleszkiewicz 2000] R. Latała and K. Oleszkiewicz, "Between Sobolev and Poincaré", pp. 147–168 in *Geometric aspects of functional analysis*, Lecture Notes in Math. **1745**, Springer, 2000. MR Zbl

[Madhava Rao and Thiruvenkatachar 1949] B. S. Madhava Rao and V. R. Thiruvenkatachar, "On an inequality concerning orthogonal polynomials", *Proc. Indian Acad. Sci., Sect. A.* **29** (1949), 391–393. MR Zbl

[Nash 1958] J. Nash, "Continuity of solutions of parabolic and elliptic equations", *Amer. J. Math.* **80** (1958), 931–954. MR Zbl

[Olver et al. 2010] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark (editors), *NIST handbook of mathematical functions*, Cambridge University Press, 2010. MR Zbl

[Temme 2015] N. M. Temme, *Asymptotic methods for integrals*, Series in Analysis **6**, World Scientific Publishing, Hackensack, NJ, 2015. MR Zbl

[Wang 2005] F.-Y. Wang, "A generalization of Poincaré and log-Sobolev inequalities", *Potential Anal.* **22**:1 (2005), 1–15. MR Zbl

Received 28 Jun 2016. Revised 30 Jan 2017. Accepted 18 Mar 2017.

PAATA IVANISVILI: ivanishvili.paata@gmail.com
Department of Mathematics, Kent State University, Kent, OH 44240, United States

ALEXANDER VOLBERG: volberg@math.msu.edu
Department of Mathematics, Michigan State University, East Lansing, MI 48824, United States

ANALYSIS AND PDE
Vol. 10, No. 4, 2017
dx.doi.org/10.2140/apde.2017.10.943

msp

POSITIVITY FOR FOURTH-ORDER SEMILINEAR PROBLEMS RELATED TO THE KIRCHHOFF–LOVE FUNCTIONAL

GIULIO ROMANI

We study the ground states of the following generalization of the Kirchhoff–Love functional,

$$J_\sigma(u) = \int_\Omega \frac{(\Delta u)^2}{2} - (1-\sigma)\int_\Omega \det(\nabla^2 u) - \int_\Omega F(x,u),$$

where Ω is a bounded convex domain in $\mathbb{R}^2$ with $C^{1,1}$ boundary and the nonlinearities involved are of sublinear type or superlinear with power growth. These critical points correspond to least-energy weak solutions to a fourth-order semilinear boundary value problem with Steklov boundary conditions depending on σ. Positivity of ground states is proved with different techniques according to the range of the parameter $\sigma \in \mathbb{R}$ and we also provide a convergence analysis for the ground states with respect to σ. Further results concerning positive radial solutions are established when the domain is a ball.

1. Introduction

The energy of a thin hinged plate under the action of a vertical external force of density f can be computed by the Kirchhoff–Love functional

$$I_\sigma(u) = \int_\Omega \frac{(\Delta u)^2}{2} - (1-\sigma)\int_\Omega \det(\nabla^2 u) - \int_\Omega f u,$$

where the bounded domain $\Omega \subset \mathbb{R}^2$ describes the shape of the plate and u its deflection from the original unloaded position. Since the plate is supposed to be hinged, the natural space in which to consider our problem is $H^2(\Omega) \cap H_0^1(\Omega)$. The coefficient σ, called the *Poisson ratio*, depends on the material and measures its transverse expansion (resp. contraction), according to its positive (resp. negative) sign, when subjected to an external compressing force. Due to some thermodynamic considerations in elasticity theory, the physical relevant interval for σ is $\left(-1, \frac{1}{2}\right)$. A detailed derivation of the model can be found in [Ventsel and Krauthammer 2001], while a mathematical analysis concerning the positivity-preserving property for I_σ has been carried out by Parini and Stylianou [2009]. Besides a further extension of their results, here we are interested in a direct generalization of the Kirchhoff–Love functional, namely when

This work has been carried out thanks to the support of the A*MIDEX grant (no. ANR-11-IDEX-0001-02) funded by the French Government "Investissements d'Avenir" program.
MSC2010: 35G30, 49J40.
Keywords: biharmonic operator, positivity-preserving property, semilinear problem, positive least-energy solutions, Nehari manifold.

the density f may depend also on the deflection of the plate itself:

$$J_\sigma(u) = \int_\Omega \frac{(\Delta u)^2}{2} - (1-\sigma)\int_\Omega \det(\nabla^2 u) - \int_\Omega F(x,u), \tag{1-1}$$

where $F(x,s) = \int_0^s f(x,t)\,dt$, and furthermore we let $\sigma \in \mathbb{R}$. We are mainly interested in a power-type nonlinearity, namely

$$F(x,u) = \frac{g(x)|u|^{p+1}}{p+1}, \quad \text{where } g \in L^1(\Omega) \text{ and } g > 0 \text{ in } \Omega. \tag{1-2}$$

In particular we look for existence and positivity of those critical points which have the lowest energy, referred to in the literature as *ground states*.

If the boundary is sufficiently smooth, searching for critical points of J_σ with the nonlinearity (1-2) is equivalent to finding weak solutions of the fourth-order semilinear boundary problem

$$\begin{cases} \Delta^2 u = g(x)|u|^{p-1}u & \text{in } \Omega, \\ u = \Delta u - (1-\sigma)\kappa u_n = 0 & \text{on } \partial\Omega, \end{cases} \tag{1-3}$$

where u_n stands for the normal derivative of u on $\partial\Omega$ and κ is the signed curvature of the boundary (positive on convex parts). These kinds of mixed boundary conditions are usually called *Steklov* from their first appearance in [Stekloff 1902] and they are an intermediate situation between Navier boundary conditions (when $\sigma = 1$) and Dirichlet boundary conditions ($u = u_n = 0$, seen as the limit case as $\sigma \to +\infty$).

Although fourth-order (or more generally, higher-order) problems have garnered attention even from the first decade of the 20th century, most of the literature deals with the Navier case, where the maximum principle still holds, or with Dirichlet boundary conditions, where Green's function arguments are available. Conversely, problems like (1-3) have been intensively studied only in the last decade, focusing on the associated boundary eigenvalue problems (see [Ferrero et al. 2005; Bucur et al. 2009]), the positivity-preserving property of the solution operator (see [Gazzola and Sweers 2008]) and some semilinear problems (see, for instance, [Berchio and Gazzola 2011; Berchio et al. 2006; 2007]).

This paper is a contribution to the study of semilinear subcritical biharmonic Steklov problems in low dimension. Here, we mainly focus on a nonlinearity of power-type as in [Berchio et al. 2007], where the critical exponent in high dimensions is considered and the domain is a ball. On the other hand, although some related subcritical problems have already appeared in [Berchio et al. 2006], we consider a slightly different kind of nonlinearity; we let σ be lying not only in the physical relevant interval, and the techniques involved are different.

Besides the existence of ground states for J_σ, we mainly investigate their positivity. The question is quite challenging since, like most fourth-order problems, one has to face the lack of a maximum principle. Moreover, we will show that positivity is strongly related to the parameter σ and different techniques are needed to cover different regions in which σ lies: the superharmonic method, some convergence arguments and the dual cones decomposition.

The main results contained in this paper may be summarized as follows:

Theorem 1.1 (existence, positivity). *Let $\Omega \subset \mathbb{R}^2$ be a bounded convex domain with $\partial\Omega$ of class $C^{1,1}$ and let $f(x,s) = g(x)|s|^{p-1}s$, with $p \in (0,1) \cup (1,+\infty)$ and $g \in L^1(\Omega)$, $g > 0$ a.e. in Ω. Then there exist $\sigma^* \leq -1$ and $\sigma_1 > 1$ (depending on Ω and possibly infinite) such that the functional J_σ has no positive critical points if $\sigma \leq \sigma^*$, while it admits (at least) a positive ground state if $\sigma \in (\sigma^*, \sigma_1)$.*

Theorem 1.2 (convergence). *Under the previous assumptions for Ω and f, let $(u_k)_{k\in\mathbb{N}}$ be a sequence of ground states for the respective sequence of functionals $(J_{\sigma_k})_{k\in\mathbb{N}}$. Up to a subsequence,*

(i) *if $\sigma_k \searrow \sigma^*$, then $u_k \to 0$ in $H^2(\Omega)$ in the case $p > 1$, while $u_k \to +\infty$ in $L^\infty(\Omega)$ if $p \in (0,1)$;*

(ii) *if $\sigma_k \to 1$, then $u_k \to \bar{u}$ in $W^{2,q}(\Omega)$ for every $q > 2$, where $\bar{u}$ is a ground state for the Navier problem;*

(iii) *if $\sigma_k \to +\infty$, then $u_k \to U$ in $H^2(\Omega)$, where U is a ground state for the Dirichlet problem.*

Notice that Theorem 1.1 might also be seen as an extension to the semilinear setting of the main positivity results established by Gazzola and Sweers [2008, Theorem 4.1] for the linear case.

Finally, we want to stress our attempt to impose only the strictly necessary assumptions on the domain in order to obtain our results and to have a well-defined second boundary condition in (1-3).

The paper is organized as follows: after a few preliminary results (Section 2), we establish existence (Section 3) and positivity (Section 4) of ground states of J_σ when σ belongs to the range $(-1,1]$ (which contains the relevant physical interval) both for f sublinear and superlinear; the latter is due to an argument based on the Nehari manifold. Except for the last section, the rest of the paper is devoted to proving Theorem 1.1 in the cases $\sigma \leq -1$ (Section 5) and $\sigma > 1$ (Section 6). While the first situation is quite easy to handle, the positivity in the second is more delicate and requires different tools. In this context and also for this purpose, Theorem 1.2 will be established. Finally, Section 7 provides a further investigation in the case Ω is the unit ball, concerning generic positive radially symmetric solutions.

2. Notation and preliminary results

Throughout the paper, $\nabla^2 u$ stands for the Hessian matrix of u and the derivatives are denoted by subscripts $(u_x, u_{xy}, \dots)$. Moreover, n and τ will be the exterior normal and the tangent vector, and u_n and u_τ the normal and the tangential derivative of u. We say that u is *superharmonic* in Ω when $-\Delta u \geq 0$ in Ω and $u = 0$ on $\partial\Omega$; u is *strictly superharmonic* when we have in addition that $-\Delta u \not\equiv 0$.

Let $N \geq 2$. We say $\Omega \subset \mathbb{R}^N$ is a *domain* when it is open and connected; moreover, Ω has *a boundary of class* $C^{k,1}$ if $\partial\Omega$ can be described in local coordinates by a C^k function with Lipschitz continuous k-th derivatives. Finally, Ω satisfies a *uniform external ball condition* if there exists $R > 0$ such that for all $x \in \partial\Omega$ there exists a ball B_R of radius R such that $x \in \partial B_R$ and $B_R \subset \mathbb{R}^N \setminus \overline{\Omega}$.

The topological dual of a normed space X is denoted by X^*; for $q \in [1,+\infty]$, the $L^q(\Omega)$ norm is denoted by $\|\cdot\|_q$ and we have

$$\||\nabla^k \cdot|\|_q := \Big(\sum_{|\alpha|=k} \|D^\alpha \cdot\|_q^q\Big)^{\frac{1}{q}},$$

where α is a multi-index.

Let us also recall that, if $\Omega \subset \mathbb{R}^2$ is a bounded domain with Lipschitz boundary, by Sobolev embeddings (see [Adams and Fournier 2003, Theorem 4.12, Part II]), $H^2(\Omega) \hookrightarrow C^{0,\lambda}(\overline{\Omega})$ for any $\lambda \in (0,1)$; thus we have continuous embedding in $L^q(\Omega)$ for every $q \in [1,\infty]$.

Finally, we present some very useful facts about equivalence of norms in $H^2(\Omega) \cap H_0^1(\Omega)$. The quoted results have been already obtained by Nazarov, Stylianou and Sweers in [Nazarov et al. 2012]; we will include the proof of the second equivalence in order to have a self-contained exposition.

Lemma 2.1. *Let $\Omega \subset \mathbb{R}^N$ bounded with a Lipschitz boundary and $\sigma \in (-1,1]$.*

(i) $\big\| |\nabla^2 \cdot| \big\|_2$ *and* $\|\cdot\|_{H^2(\Omega)}$ *are equivalent norms on* $H^2(\Omega) \cap H_0^1(\Omega)$.

(ii) *If $\sigma = 1$, assume additionally that Ω satisfies a uniform external ball condition. Then*

$$\|u\|_{H_\sigma(\Omega)} := \left(\int_\Omega (\Delta u)^2 - 2(1-\sigma) \int_\Omega \det(\nabla^2 u) \right)^{\frac{1}{2}} \tag{2-1}$$

defines a norm on $H^2(\Omega) \cap H_0^1(\Omega)$ equivalent to the standard norm.

Proof. We prove here only (ii) and we refer to [Nazarov et al. 2012, Corollary 5.4] for a proof of (i). Firstly

$$\begin{aligned} \|u\|^2_{H_\sigma(\Omega)} &= \int_\Omega u_{xx}^2 + u_{yy}^2 + 2u_{xy}^2 + 2\sigma(u_{xx}u_{yy} - u_{xy}^2) \\ &\le \big\| |\nabla^2 u| \big\|_2^2 + 2|\sigma| \left(\frac{u_{xx}^2 + u_{yy}^2}{2} + u_{xy}^2 \right) = (1+|\sigma|) \big\| |\nabla^2 u| \big\|_2^2. \end{aligned}$$

Moreover, if $\sigma \in (-1,1)$, one has

$$\begin{aligned} \|u\|^2_{H_\sigma(\Omega)} &= \int_\Omega u_{xx}^2 + u_{yy}^2 + 2(1-\sigma)u_{xy}^2 + 2\sigma u_{xx}u_{yy} \\ &\ge \int_\Omega u_{xx}^2 + u_{yy}^2 + 2(1-\sigma)u_{xy}^2 - |\sigma|(u_{xx}^2 + u_{yy}^2) \ge (1-|\sigma|) \big\| |\nabla^2 u| \big\|_2^2. \end{aligned} \tag{2-2}$$

The proof is completed by applying (i) and noticing that the map

$$(u,v)_{H_\sigma} \mapsto \int_\Omega \Delta u \, \Delta v - (1-\sigma) \int_\Omega u_{xx}v_{yy} + u_{yy}v_{xx} - 2u_{xy}v_{xy}$$

defines a scalar product on $H^2(\Omega) \cap H_0^1(\Omega)$ for every $\sigma \in (-1,1)$ by the inequality (2-2). In the special case $\sigma = 1$, one has $\|u\|_{H_1(\Omega)} = \|\Delta u\|_2$, which is an equivalent norm on $H^2(\Omega) \cap H_0^1(\Omega)$ provided the external ball condition is satisfied (see [Adolfsson 1992]). □

In the following, $C_0 = C_0(\Omega)$ and $C_A = C_A(\Omega)$ indicate the smallest positive constants such that

$$\|u\|^2_{H^2(\Omega)} := \|u\|_2^2 + \big\| |\nabla u| \big\|_2^2 + \big\| |\nabla^2 u| \big\|_2^2 \le C_0 \big\| |\nabla^2 u| \big\|_2^2 \tag{2-3}$$

and

$$\|u\|^2_{H^2(\Omega)} \le C_A \|\Delta u\|_2^2 \tag{2-4}$$

for every $u \in H^2(\Omega) \cap H_0^1(\Omega)$.

3. Existence of ground states

In this section we investigate the existence of critical points of the generalized Kirchhoff–Love functional $J_\sigma : H^2(\Omega) \cap H_0^1(\Omega) \to \mathbb{R}$ defined in (1-1) in the physical relevant interval $\sigma \in (-1, 1]$. Hereafter, we assume Ω to be a bounded domain in $\mathbb{R}^2$. Concerning the nonlinearity, the functional J_σ is well defined once we impose $F(\cdot, s) \in L^1(\Omega)$ and $F(x, \cdot) \in C^1(\mathbb{R})$ (and thus there exists $f(x, \cdot)$ continuous such that $F(x,s) = \int_0^s f(x,t)\,dt$) and a power-type growth control on F, namely the existence of $a, b \in L^1(\Omega)$ such that $|F(x,s)| \le a(x) + b(x)|s|^q$ for some $q > 0$. With these assumptions on F, it is a standard fact to prove that J_σ is a C^1 functional with Fréchet derivative

$$J'_\sigma(u)[v] = \int_\Omega \Delta u\, \Delta v - (1-\sigma)\int_\Omega (u_{xx}v_{yy} + u_{yy}v_{xx} - 2u_{xy}v_{xy}) - \int_\Omega f(x,u)v.$$

Notice that, if Ω satisfies the assumptions of Lemma 2.1, we can rewrite the functional as

$$J_\sigma(u) = \tfrac{1}{2}\|u\|^2_{H_\sigma(\Omega)} - \int_\Omega F(x,u).$$

Our aim is to investigate the *ground states* of the functional J_σ, i.e., the critical points on which the functional assumes the lowest value. In fact, besides the interest from a physical point of view, we are able to characterize them variationally and thus to apply a larger number of analytical tools.

Since the geometry of the functional plays an important role, from now on we have to distinguish between the sublinear case, that is, when the density f has at most a slow linear growth in the real variable (as it will be specified in the following), and the superlinear case, the opposite one. In fact, we will see that in the first case J_σ behaves similarly to the linear Kirchhoff–Love functional studied in [Parini and Stylianou 2009] since it is coercive and ground states are global minima, while, in the second case, J_σ has a mountain pass geometry and the ground states are saddle points. Moreover, although in the sequel we will be mainly interested in the power-type nonlinearity as in (1-2), in the sublinear case we can easily generalize our analysis to a larger class of nonlinearities, as specified in Proposition 3.1.

We exclude from our analysis the case of general linear growths for the nonlinearity, for instance $f(x,u) = \lambda g(x)u$, since (1-3) becomes an eigenvalue problem and can be investigated with standard techniques (see also [Berchio et al. 2006, Theorem 4]).

Sublinear case.

Proposition 3.1. *With the assumptions for σ and Ω as in Lemma 2.1, let $p \in (0,2)$ and suppose*

$$|F(x,s)| \le d(x) + c(x)|s|^p + \tfrac{1}{2}(1-|\sigma|)C_0^{-1}s^2, \tag{H}$$

where $c, d \in L^1(\Omega)$. Then the functional J_σ is weakly lower semicontinuous and coercive; hence there exists a global minimizer of J_σ in $H^2(\Omega) \cap H_0^1(\Omega)$.

Proof. Let $(u_k)_{k\in\mathbb{N}} \subset H^2(\Omega) \cap H_0^1(\Omega) \ni u$ be such that $u_k \rightharpoonup u$ weakly in $H^2(\Omega)$; since it is bounded in $H^2(\Omega)$ and consequently in $L^\infty(\Omega)$, one has

$$|F(x,u_k)| \le d(x) + c(x)M^p + \tfrac{1}{2}(1-|\sigma|)C_0^{-1}M^2$$

for some $M > 0$, which is integrable over Ω. Moreover, by the compactness of the embedding $H^2(\Omega) \hookrightarrow L^p(\Omega)$, there exists a subsequence $(u_{k_j})_{j\in\mathbb{N}}$ such that $u_{k_j} \to u$ in $L^p(\Omega)$ for a suitable $p \geq 1$, so $F(x, u_{k_j}(x)) \to F(x, u(x))$ a.e. in Ω by continuity of $F(x, \cdot)$. Hence, by the dominated convergence theorem, we have $\int_\Omega F(x, u_{k_j}) \to \int_\Omega F(x, u)$. This, together with the weakly lower semicontinuity of the norm, implies the same property for J_σ. If $\sigma \in (-1, 1)$, by (2-3)

$$\begin{aligned} J_\sigma(u) &\geq \tfrac{1}{2}(1-|\sigma|)\big\||\nabla^2 u|\big\|_2^2 - \|d\|_1 - C^p\|c\|_1\|u\|^p_{H^2(\Omega)} - \tfrac{1}{2}(1-|\sigma|)C_0^{-1}\|u\|_2^2 \\ &\geq \tfrac{1}{2}(1-|\sigma|)C_0^{-1}\big\||\nabla^2 u|\big\|_2^2 - \|c\|_1 C^p C_0^{\frac{p}{2}}\big\||\nabla^2 u|\big\|_2^p - \|d\|_1; \end{aligned}$$

by (i) of Lemma 2.1, we deduce that $J_\sigma(u) \to +\infty$ as $\|u\|_{H^2(\Omega)} \to +\infty$, since $p \in (0, 2)$. Easier computations provide a similar estimate to conclude the proof also if $\sigma = 1$. $\square$

Remark 3.2 (model case). As an application of Proposition 3.1, we may consider the following kind of sublinearity:

$$F(x, u) = g(x)|u|^{p+1} + d(x)u,$$

where $p \in (0, 1)$ and $d, g \in L^1(\Omega)$. In this case the functional is coercive and verifies (H). Notice also that if $g = 0$ we retrieve the linear Kirchhoff–Love functional considered in [Parini and Stylianou 2009].

Superlinear case. This case is more involved and we have to restrict to the nonlinearity (1-2) with $p > 1$:

$$J_\sigma(u) := \int_\Omega \frac{(\Delta u)^2}{2} - (1-\sigma)\int_\Omega \det(\nabla^2 u) - \int_\Omega \frac{g(x)|u|^{p+1}}{p+1}. \tag{3-1}$$

Here the functional is not coercive anymore: in fact, fixing any $u \in H^2(\Omega) \cap H_0^1(\Omega) \setminus \{0\}$, we have $J_\sigma(tu) \to -\infty$ as $t \to +\infty$. Following some arguments of [Castro et al. 1997; Grumiau and Parini 2008], we will make use of the method of the Nehari manifold to infer the existence of a (nontrivial) critical point. After some preliminary results, we will show that in our manifold the infimum of J_σ is attained and then, using a deformation lemma, we will prove it is a critical point for J_σ in $H^2(\Omega) \cap H_0^1(\Omega)$.

Let us define the *Nehari manifold* of J_σ as the set

$$\mathcal{N}_\sigma := \big\{u \in (H^2(\Omega) \cap H_0^1(\Omega)) \setminus \{0\} \;\big|\; J'_\sigma(u)[u] = 0\big\},$$

which clearly contains all nontrivial critical points of J_σ. First of all, notice that $u \in \mathcal{N}_\sigma$ if and only if

$$\int_\Omega (\Delta u)^2 - 2(1-\sigma)\int_\Omega \det(\nabla^2 u) = \int_\Omega g(x)|u|^{p+1},$$

so one has two equivalent formulations for J_σ restricted on $\mathcal{N}_\sigma$,

$$J_{\sigma|\mathcal{N}_\sigma}(u) = \left(\frac{1}{2} - \frac{1}{p+1}\right)\int_\Omega g(x)|u|^{p+1} = \left(\frac{1}{2} - \frac{1}{p+1}\right)\left(\int_\Omega (\Delta u)^2 - 2(1-\sigma)\int_\Omega \det(\nabla^2 u)\right), \tag{3-2}$$

which implies $J_{\sigma|\mathcal{N}_\sigma}(u) > 0$ for every $u \in \mathcal{N}_\sigma$.

A crucial step will be to study what happens on the half-lines of $H^2(\Omega) \cap H_0^1(\Omega)$:

Lemma 3.3. *Let $u \in H^2(\Omega) \cap H_0^1(\Omega) \setminus \{0\}$ and the half-line r_u be defined as $r_u := \{tu \mid t > 0\}$. The intersection between r_u and $\mathcal{N}_\sigma$ consists of a unique point $t^*(u)u$, where*

$$t^*(u) := \left(\frac{\int_\Omega (\Delta u)^2 - 2(1-\sigma)\int_\Omega \det(\nabla^2 u)}{\int_\Omega g(x)|u|^{p+1}}\right)^{\frac{1}{p-1}}. \tag{3-3}$$

Moreover $J_\sigma(t^*(u)u) = \max_{t>0} J_\sigma(tu)$.

Proof. For $t > 0$ and a fixed $u \in H^2(\Omega) \cap H_0^1(\Omega) \setminus \{0\}$, we have $tu \in \mathcal{N}_\sigma$ if and only if

$$t^2\left[\int_\Omega (\Delta u)^2 - 2(1-\sigma)\int_\Omega \det(\nabla^2 u)\right] = t^{p+1}\int_\Omega g(x)|u|^{p+1},$$

from which we deduce $t = t^*(u)$. Moreover, define

$$\eta(t) := J_\sigma(tu) = \frac{t^2}{2}\left[\int_\Omega (\Delta u)^2 - 2(1-\sigma)\int_\Omega \det(\nabla^2 u)\right] - \frac{t^{p+1}}{p+1}\int_\Omega g(x)|u|^{p+1}.$$

If we look for $\bar{t} > 0$ such that $\eta'(\bar{t}) = 0$, we find again that $\bar{t} = t^*(u)$ and, since $\eta'(t)(t - t^*(u)) < 0$ for $t \neq t^*(u)$, we have that $t^*(u)u$ is the unique global maximum in the half-line r_u. □

Lemma 3.4. *The Nehari manifold is bounded away from* 0; *i.e.*, $0 \notin \overline{\mathcal{N}_\sigma}$.

Proof. Suppose first that $\sigma \in (-1, 1)$ and let $u \in H^2(\Omega) \cap H_0^1(\Omega) \setminus \{0\}$. By Lemmas 2.1 and 3.3 and the embedding $H^2(\Omega) \hookrightarrow L^\infty(\Omega)$, the following chain of inequalities holds:

$$\begin{aligned}(1+|\sigma|)\,\|t^*(u)u\|^2_{H^2(\Omega)} &\geq \|t^*(u)u\|^2_{H_\sigma(\Omega)} = (t^*(u))^{p+1}\int_\Omega g(x)|u|^{p+1}\\ &\geq (C_0^{-1}(1-|\sigma|))^{\frac{p+1}{p-1}}\frac{\|u\|^{\frac{2(p+1)}{p-1}}_{H^2(\Omega)}}{\left(\int_\Omega g(x)|u|^{p+1}\right)^{\frac{2}{p-1}}}\\ &\geq C(\Omega,p,\sigma)\frac{\|u\|^{\frac{2(p+1)}{p-1}}_{H^2(\Omega)}}{\left(\|g\|_1\,\|u\|^{p+1}_{H^2(\Omega)}\right)^{\frac{2}{p-1}}} = \frac{C(\Omega,p,\sigma)}{\|g\|_1^{\frac{2}{p-1}}}.\end{aligned}$$

If $\sigma = 1$, one can deduce the same result using the equivalent norm on $H^2(\Omega) \cap H_0^1(\Omega)$ given by $\|\Delta\cdot\|_2$. In both cases, there exists a uniform bound from below for the $H^2(\Omega)$ norm of the elements in the Nehari manifold and thus 0 cannot be a cluster point for $\mathcal{N}_\sigma$. □

Proposition 3.5. *There exists $u \in \mathcal{N}_\sigma$ such that $J_\sigma(u) = \inf_{v\in\mathcal{N}_\sigma} J_\sigma(v) =: c$.*

Proof. As already noticed, $c \geq 0$, since it attains positive values on $\mathcal{N}_\sigma$. Let now $(u_k)_{k\in\mathbb{N}} \subset \mathcal{N}_\sigma$ be a minimizing sequence for J_σ: we claim that $(u_k)_{k\in\mathbb{N}}$ is bounded in $H^2(\Omega)$ norm. In fact, if $\sigma \in (-1, 1)$, there exists a constant $C > 0$ such that, for every $k \in \mathbb{N}$,

$$C \geq J_\sigma(u_k) = \left(\frac{1}{2} - \frac{1}{p+1}\right)\|u_k\|^2_{H_\sigma(\Omega)} \geq \left(\frac{1}{2} - \frac{1}{p+1}\right)(1-|\sigma|)C_0^{-1}\|u_k\|^2_{H^2(\Omega)},$$

while (2-4) provides the right estimate in the case $\sigma=1$. Hence, there exists a subsequence $(u_{k_j})_{j\in\mathbb{N}}\subset\mathcal{N}_\sigma$ and $u\in H^2(\Omega)\cap H_0^1(\Omega)\setminus\{0\}$ such that $u_{k_j}\rightharpoonup u$ weakly in $H^2(\Omega)$ (and so weakly in $(H^2(\Omega)\cap H_0^1(\Omega),\|\cdot\|_{H\sigma})$ by Lemma 2.1) and strongly in $L^\infty(\Omega)$, by compact embedding. Consider now $t^*=t^*(u)$ such that $t^*u\in\mathcal{N}_\sigma$: by weak semicontinuity of the norm,

$$\begin{aligned} c=\inf_{v\in\mathcal{N}_\sigma} J_\sigma(v)\le J(t^*u)&=(t^*)^2\biggl[\int_\Omega\frac{(\Delta u)^2}{2}-(1-\sigma)\int_\Omega\det(\nabla^2u)\biggr]-(t^*)^{p+1}\int_\Omega\frac{g(x)|u|^{p+1}}{p+1}\\ &\le\liminf_{j\to+\infty}\biggl((t^*)^2\biggl[\int_\Omega\frac{(\Delta u_{k_j})^2}{2}-(1-\sigma)\int_\Omega\det(\nabla^2u_{k_j})\biggr]-(t^*)^{p+1}\int_\Omega\frac{g(x)|u_{k_j}|^{p+1}}{p+1}\biggr)\\ &=\liminf_{j\to+\infty}J_\sigma(t^*u_{k_j})\le\liminf_{j\to+\infty}J_\sigma(u_{k_j})=c, \end{aligned}\tag{3-4}$$

where the last inequality holds because the supremum of J_σ in each half-line $\{tu_{k_j}\mid t>0\}$ is achieved exactly in u_{k_j} by Lemma 3.3. Hence, the infimum of J_σ on $\mathcal{N}_\sigma$ is attained on t^*u. □

In the proof of Proposition 3.5 something weird happened: we took a minimizing sequence, which converges to an element u and we proved that there exists $\alpha=t^*(u)\in\mathbb{R}$ such that αu is the minimum point of our functional J_σ. One expects that the minimum is u itself and not a dilation of it. Indeed, one may prove that $t^*=1$. In fact, with the same notation as in that proof, from (3-4) we deduce $J_\sigma(u_{k_j})\to c=J_\sigma(t^*u)$ by construction and $t^*u\in\mathcal{N}_\sigma$, so

$$J_\sigma(u_{k_j})\to\Bigl(\frac12-\frac{1}{p+1}\Bigr)\int_\Omega g(x)|t^*u|^{p+1}.$$

Moreover, we took the sequence to be in the Nehari manifold itself, so

$$J_\sigma(u_{k_j})=\Bigl(\frac12-\frac{1}{p+1}\Bigr)\int_\Omega g(x)|u_{k_j}|^{p+1},$$

and we have that $u_{k_j}\to u$ strongly in $L^\infty(\Omega)$; thus

$$J_\sigma(u_{k_j})\to\Bigl(\frac12-\frac{1}{p+1}\Bigr)\int_\Omega g(x)|u|^{p+1}.$$

By the uniqueness of the limit, we must have $t^*=1$, so $u\in\mathcal{N}_\sigma$.

Theorem 3.6. *The minimum u of J_σ in $\mathcal{N}_\sigma$ is a critical point for J_σ in $H^2(\Omega)\cap H_0^1(\Omega)$.*

Proof. Suppose by contradiction that u is not a critical point. Since the functional is C^1, there exists a ball centered in u and $\varepsilon>0$ such that, for all $v\in B$,

$$c-\varepsilon\le J_\sigma(v)\le c+\varepsilon,\quad \|J'_\sigma(v)\|_{(H^2(\Omega)\cap H_0^1(\Omega))^*}\ge\tfrac12\varepsilon,$$

where $c=J_\sigma(u)=\inf_{v\in\mathcal{N}_\sigma}J_\sigma(v)$. Notice that on the half-line r_u, the point u is the global maximum, so $J_\sigma(v)<c$ for each $v\in B\cap r_u$, $v\ne u$.

If we define $a=c-\varepsilon$, $b=c+\varepsilon$, $\delta=8$, $S=\overline{B_r(u)}$ and $S_0=\overline{H^2(\Omega)\cap H_0^1(\Omega)\setminus B'}$, where $r>0$ such that $B_r(u)\Subset B'\Subset B$, applying [Gasiński and Papageorgiou 2006, Proposition 5.1.25], there exists a locally Lipschitz homotopy of homeomorphisms Γ_t on $H^2(\Omega)\cap H_0^1(\Omega)$ such that

(i) $t \mapsto J_\sigma(\Gamma(t, v))$ is decreasing in $B_r(u)$ and, in general, nonincreasing;

(ii) $J_\sigma(\Gamma(t, v)) = v$ for $v \in S_0$ and $t \in [0, 1]$, and so also for all $v \in \partial B$.

From (i) we deduce that $J_\sigma(\Gamma(t, v)) < c$ for every $v \in B \cap r_u$ and $t \neq 0$. Moreover, define the map $\psi : B \cap r_u \to \mathbb{R}$ such that

$$\psi(v) := J'_\sigma(\Gamma(1, v))[\Gamma(1, v)]$$

and consider $v \in \partial B \cap r_u$, so there exists $\alpha \neq 1$ such that $v = \alpha u$: we know from (ii) that $\Gamma(1, v) = v$ and, by Lemma 3.3, $J'_\sigma(\alpha u)[\alpha u] > 0$ if $\alpha \in (0, 1)$ and $J'_\sigma(\alpha u)[\alpha u] < 0$ if $\alpha \in (1, +\infty)$, so $\psi(v)(v-u) < 0$ on $\partial B \cap r_u$. As a result, since one can think of ψ as a continuous map from $[x_1, x_2] \to \mathbb{R}$, where x_1 and x_2 correspond to the intersections between the half-line r_u and the ball B, and since $\psi(x_1) > 0$ and $\psi(x_2) < 0$, there exists a zero of ψ in (x_1, x_2); i.e., there exists $\bar{v} \in B \cap r_u$ such that $J'_\sigma(\Gamma(1, \bar{v}))[\Gamma(1, \bar{v})] = 0$.

Setting $w := \Gamma(1, \bar{v})$, we have $w \in \mathcal{N}_\sigma$ and $J_\sigma(w) = J_\sigma(\Gamma(1, \bar{v})) < c = \inf_{v \in \mathcal{N}_\sigma} J_\sigma(v)$, a contradiction. □

So far, we proved the existence of a ground state for J_σ. Actually, one can say more about the existence of general critical points by means of the Krasnoselskiĭ genus theory (see [Ambrosetti and Malchiodi 2007, Section 10.2]). In fact, since our framework is subcritical, it is quite standard to prove the Palais–Smale condition for J_σ by a compact embedding of $H^2(\Omega)$ in every Lebesgue space. Moreover, our functional is C^1, even and bounded from below on the unit sphere of $H^2(\Omega) \cap H^1_0(\Omega)$: indeed, if $\|u\|_{H_\sigma(\Omega)} = 1$, then $\|u\|_\infty < C$ for some $C > 0$, so

$$J_\sigma(u) = \frac{1}{2} - \int_\Omega \frac{g(x)|u|^{p+1}}{p+1} \geq \frac{1}{2} - \frac{C^{p+1}\|g\|_1}{p+1} > -\infty.$$

Hence, by [Ambrosetti and Malchiodi 2007, Proposition 10.8], one can ensure the existence of an infinite number of couples of critical points. The same argument may also be applied for the general sublinear case, provided $F(x, s) = F(x, -s)$ for every $s \in \mathbb{R}$.

4. An identity and the positivity of ground states in convex domains

The aim of this section is to prove positivity for the ground states found in the previous section. Notice that the problematic term in J_σ is the one involving the determinant of the Hessian matrix. In order to overcome this difficulty, we need to rewrite it in an equivalent way, transforming it into a boundary term which can be handled in order to prove the desired positivity. Nevertheless, since the signed curvature of the boundary will be involved, we need to impose some regularity on $\partial\Omega$. We will basically deduce the same statement as Lemma 2.5(ii) of [Parini and Stylianou 2009], but we will extend it to a larger class of domains.

A crucial identity. Our goal is to generalize the following result:

Theorem 4.1 [Parini and Stylianou 2009, Lemma 2.5]. *Let Ω be a bounded domain in $\mathbb{R}^2$ with $C^{2,1}$ boundary, and let κ be its signed curvature. Then for all $u \in H^2(\Omega)$ and for every $\varphi \in H^3(\Omega)$, defining $K(u) := \int_\Omega \det(\nabla^2 u)\, dx$, we have*

$$\langle K'(u), \varphi \rangle = \int_{\partial\Omega} (\kappa \varphi_n u_n + \varphi_{\tau\tau} u_n - \varphi_{\tau n} u_\tau). \qquad (\mathrm{F_{PS}})$$

Hence, for all $u \in H^2(\Omega) \cap H_0^1(\Omega)$,

$$K(u) = \frac{1}{2}\int_{\partial\Omega} \kappa u_n^2. \tag{F}$$

Going into the details of its proof, one can actually realize that the strong regularity assumption on the boundary was needed only to derive (F) from (F_{PS}) because the authors used the density of $H^3(\Omega) \cap H_0^1(\Omega)$ in $H^2(\Omega) \cap H_0^1(\Omega)$, which strongly relied on the fact that $\partial\Omega \in C^{2,1}$ (see [Parini and Stylianou 2009, Lemma 2.3]). Nevertheless, (F_{PS}) requires only that all the elements therein are well defined. Hence, our starting point is the following:

Corollary 4.2. *Let $\Omega \subset \mathbb{R}^2$ be a bounded domain of class $C^{1,1}$. Then for every $v \in C^\infty(\overline{\Omega})$,*

$$K(v) = \tfrac{1}{2}\langle K'(v), v\rangle = \frac{1}{2}\int_{\partial\Omega} (\kappa v_n^2 - (v_{n\tau} + v_{\tau n})v_\tau). \tag{$F_{PS}2$}$$

Proof. One only has to notice that if $\partial\Omega \in C^{1,1}$, then κ is in $L^\infty(\partial\Omega)$ and

$$\int_{\partial\Omega} (v_{n\tau}v_\tau + v_n v_{\tau\tau}) = \int_{\partial\Omega} (v_n v_\tau)_\tau = 0,$$

as $\partial\Omega$ is a closed curve and by the definition of the tangential derivative (i.e., as $(d/ds)u(\gamma(s))$, where γ is the parametrization of the curve $\partial\Omega$ in the arch parameter s). $\square$

Our strategy consists of two steps: using ($F_{PS}2$), we will firstly prove that (F) holds also for every $v \in C_0^{1,1}(\overline{\Omega}) := \{u \in C^{1,1}(\overline{\Omega}) \mid u_{|\partial\Omega} = 0\}$; then, by a density result, we will transfer (F) from $C_0^{1,1}(\overline{\Omega})$ to $H^2(\Omega) \cap H_0^1(\Omega)$. We will make use of the following lemma, which makes a well-known result more precise:

Lemma 4.3. *Let $\Omega \subset \mathbb{R}^N$ be a bounded domain of class C^1 and $u \in C^{1,1}(\overline{\Omega})$. Then there exists a sequence $(u_k)_{k\in\mathbb{N}} \in C^\infty(\overline{\Omega})$ such that $u_k \to u$ in $H^2(\Omega)$ and $\|u_k\|_{W^{2,\infty}(\Omega)} \le C\|u\|_{W^{2,\infty}(\Omega)}$ for some positive constant C.*

Proof. First of all notice that $C^{1,1}(\overline{\Omega})$ can be equivalently seen as $W^{2,\infty}(\Omega)$, which is a subset of $H^2(\Omega)$ since Ω is a bounded domain; moreover the fact that $C^\infty(\overline{\Omega})$ is dense in $H^2(\Omega)$ in $H^2(\Omega)$ norm if $\partial\Omega$ is of class C^1 is a standard fact (see [Evans 2010, Section 5.3.3, Theorem 3]), so the only statement to be verified is the $W^{2,\infty}(\Omega)$ estimate. Since the main tool in the proof of the $H^2(\Omega)$ convergence is the local approximation, which is achieved by mollification, we only have to prove that the same inequality holds there. So, let $v \in L^\infty(\Omega)$, $\varepsilon > 0$ and consider

$$v_\varepsilon(x) := (\eta_\varepsilon * v)(x) = \int_{B_\varepsilon(0)} \eta_\varepsilon(y)v(x-y)\,dy,$$

where η_ε is the standard mollifier in $\mathbb{R}^N$, that is $\eta_\varepsilon := \varepsilon^{-n}\eta(x/\varepsilon)$ and

$$\eta(x) = \widetilde{C} e^{\frac{1}{|x|^2-1}} \chi_{B_1(0)}(x),$$

where $\widetilde{C} > 0$ such that $\int_{B_1(0)} \eta(z)\,dz = 1$. So v_ε is in $\Omega_\varepsilon := \{x \in \Omega \mid d(x, \partial\Omega) > \varepsilon\}$ and we know that $v_\varepsilon \in C^\infty(\Omega_\varepsilon)$ and η_ε is such that $\int_{B_\varepsilon(0)} \eta_\varepsilon(z)\,dz = 1$.

We claim that $\|v_\varepsilon\|_{L^\infty(\Omega_\varepsilon)} \le \|v\|_{L^\infty(\Omega)}$. In fact,

$$\|v_\varepsilon\|_{L^\infty(\Omega_\varepsilon)} \le \sup_{x\in\Omega_\varepsilon} \int_{B_\varepsilon(0)} |\eta_\varepsilon(z)||v(x-z)|\,dz \le \|v\|_{L^\infty(\Omega)} \int_{B_\varepsilon(0)} |\eta_\varepsilon(z)|\,dz = \|v\|_{L^\infty(\Omega)}.$$

Also for derivatives of v the same inequality holds, because for any admissible multi-index α we know that $D^\alpha(v_\varepsilon) = (D^\alpha(v))_\varepsilon$ (see [Gilbarg and Trudinger 1998, Lemma 7.3]).

At this point, following the aforementioned proof of [Evans 2010], it is easy to derive the desired result. ☐

Proposition 4.4. *Let $\Omega \subset \mathbb{R}^2$ be a bounded domain of class $C^{1,1}$. Then, for all $u \in C_0^{1,1}(\overline{\Omega})$,*

$$\int_\Omega \det(\nabla^2 u) = \frac{1}{2}\int_{\partial\Omega} \kappa u_n^2.$$

Proof. Applying Lemma 4.3, let $(u_k)_{k\in\mathbb{N}} \subset C^\infty(\overline{\Omega})$ be a sequence converging to u in $H^2(\Omega)$, whose norms in $W^{2,\infty}$ are controlled by the $W^{2,\infty}$ norm of u. By Corollary 4.2, the following identity holds:

$$K(u_k) = \frac{1}{2}\int_{\partial\Omega} \big[\kappa(u_k)_n^2 - ((u_k)_{n\tau} + (u_k)_{\tau n})(u_k)_\tau\big]. \tag{4-1}$$

By the convergence in $H^2(\Omega)$, one clearly has $K(u_k) \to K(u)$; moreover, since $\kappa \in L^\infty(\partial\Omega)$ and using the trace theorem, one can deduce also that

$$\int_{\partial\Omega} \kappa(u_k)_n^2 \to \int_{\partial\Omega} \kappa u_n^2.$$

Finally we have to consider the terms in which tangential derivatives are involved. Similarly to the normal derivative, one has $(u_k)_\tau \to u_\tau$ in $L^2(\partial\Omega)$, so $(u_k)_\tau \to 0$ in $L^2(\partial\Omega)$, since $u_{|\partial\Omega} = 0$. Furthermore,

$$(u_k)_{n\tau} = \nabla(u_k)_n \cdot \tau = \nabla(\nabla u_k \cdot n)\cdot\tau = (\nabla^2 u_k \cdot n + \nabla u_k \cdot \nabla n)\cdot\tau$$

and (see [Sperb 1981, Chapter 4])

$$(u_k)_{\tau n} = \sum_{i,j=1}^{2} \frac{\partial^2 u_k}{\partial x_i\,\partial x_j}\tau_i n_j$$

and one can infer that $(u_k)_{n\tau}$ and $(u_k)_{\tau n}$ are uniformly bounded in $L^2(\partial\Omega)$. In fact, since the u_k are C^∞ functions and using Lemma 4.3,

$$\begin{aligned}\|(u_k)_{n\tau}\|_{L^2(\partial\Omega)} &\le |\partial\Omega|^{\frac12}\|(u_k)_{n\tau}\|_{L^\infty(\partial\Omega)} \le |\partial\Omega|^{\frac12}\big(\big\||\nabla^2 u_k\cdot n|\big\|_{L^\infty(\partial\Omega)} + \big\||\nabla u_k\cdot\nabla n|\big\|_{L^\infty(\partial\Omega)}\big)\\ &\le 2|\partial\Omega|^{\frac12}\|n\|_{W^{1,\infty}(\partial\Omega)}\|u_k\|_{W^{2,\infty}(\Omega)} \le C(\Omega)\|u\|_{W^{2,\infty}(\Omega)}\end{aligned}$$

and similarly for $(u_k)_{\tau n}$. Consequently,

$$\int_{\partial\Omega} \big((u_k)_{n\tau} + (u_k)_{\tau n}\big)(u_k)_\tau \to 0.$$ ☐

In order to extend (F) to the space $H^2(\Omega)\cap H_0^1(\Omega)$, we need a density result (Lemma 4.6 below) which is taken from [Stylianou 2010, Theorem 2.2.4] and which can be adapted to our context: in fact, it concerns C^2 functions and diffeomorphisms but, with a little care, one can obtain the same result also in the class $C^{1,1}$.

Definition 4.5 [Adams and Fournier 2003, 3.40, p. 77]. Let Φ be a one-to-one transformation of a domain $\Omega \subset \mathbb{R}^N$ onto a domain $G \subset \mathbb{R}^N$ having inverse $\Psi := \Phi^{-1}$. We say that Φ is a $C^{1,1}$ diffeomorphism if, writing $\Phi = (\Phi_1, \dots, \Phi_N)$ and $\Psi = (\Psi_1, \dots, \Psi_N)$, then $\Phi_i \in C^{1,1}(\overline{\Omega})$ and $\Psi_i \in C^{1,1}(\overline{G})$ for every $i \in \{1, \dots, N\}$.

Lemma 4.6. *Let $\Omega \subset \mathbb{R}^N$ be bounded and open such that for every $x \in \partial\Omega$ there exists a $j \in \{0, \dots, N-1\}$, $\varepsilon > 0$ and a $C^{1,1}$-diffeomorphism $\Phi : \mathbb{R}^N \to \mathbb{R}^N$ such that the following hold*:

- $\Phi(x) = 0$.
- $\Phi(B_\varepsilon(x) \cap \Omega) \subset S_j := \{x = (x_1, \dots, x_N) \in \Omega \mid x_i > 0\,, \forall i > j\}$.
- $\Phi(B_\varepsilon(x) \cap \partial\Omega) \subset \partial S_j$.

Then
$$\overline{C_0^{1,1}(\overline{\Omega})}^{\|\cdot\|_{H^2(\Omega)}} = H^2(\Omega) \cap H_0^1(\Omega).$$

Theorem 4.7. *Let $\Omega \subset \mathbb{R}^2$ be a bounded domain of class $C^{1,1}$. Then, for all $u \in H^2(\Omega) \cap H_0^1(\Omega)$,*
$$\int_\Omega \det(\nabla^2 u) = \frac{1}{2} \int_{\partial\Omega} \kappa u_n^2. \tag{F}$$

Proof. Let $u \in H^2(\Omega) \cap H_0^1(\Omega)$; since the assumptions on the boundary are clearly fulfilled if $\partial\Omega$ is of class $C^{1,1}$, applying Lemma 4.6 we get an approximating sequence $(u_k)_{k\in\mathbb{N}} \subset C_0^{1,1}(\overline{\Omega})$ converging in $H^2(\Omega)$ to u. With the same steps as in the proof of Proposition 4.4, by the $H^2(\Omega)$ convergence, we have both $K(u_k) \to K(u)$ and $\int_{\partial\Omega} \kappa (u_k)_n^2 \to \int_{\partial\Omega} \kappa u_n^2$ and one concludes by the uniqueness of the limit. □

From the functional to the PDE. As already briefly mentioned in the Introduction, if the boundary is smooth enough ($\partial\Omega$ of class $C^{4,\alpha}$ for $\alpha > 0$), standard elliptic regularity results apply and one can integrate by parts the Euler–Lagrange equation from J_σ to see that critical points satisfy (1-3). On the other hand, assuming only that the boundary is of class $C^{1,1}$, the signed curvature is well defined in $L^\infty(\Omega)$ and we can have a weak formulation of problem (1-3). More precisely, in this case, by a *weak solution* of (1-3) here we mean a function $u \in H^2(\Omega) \cap H_0^1(\Omega)$ which satisfies
$$\int_\Omega \Delta u\, \Delta\varphi - (1-\sigma)\int_{\partial\Omega} \kappa u_n \varphi_n = \int_\Omega g(x)|u|^{p-1}u\varphi \quad \text{for all } \varphi \in H^2(\Omega) \cap H_0^1(\Omega). \tag{4-2}$$
Consequently, we can equivalently say "ground states of J_σ" or "ground state solutions for (1-3)". For a proof of the equivalence of the two problems, we refer to [Gazzola and Sweers 2008].

Positivity of ground states in convex domains. Assuming $\partial\Omega$ is of class $C^{1,1}$, Theorem 4.7 enables us to rewrite the functional J_σ in a more convenient way: in fact, we deduce that for every $u \in H^2(\Omega) \cap H_0^1(\Omega)$,
$$J_\sigma(u) = \int_\Omega \frac{(\Delta u)^2}{2} - \frac{1-\sigma}{2}\int_{\partial\Omega} \kappa u_n^2 - \int_\Omega F(x,u), \tag{4-3}$$
where we recall that $F(x,s) = \int_0^s f(x,t)\,dt$.

With this formulation, now we are able to establish the positivity of ground states of the functional J_σ in convex domains with boundary of class $C^{1,1}$ if the density function $f(x,u)$ is nonnegative, both in the sublinear and superlinear case. We will make use of the method of the superharmonic function, which is quite a standard tool when dealing with fourth-order problems and which has already been successfully used, for instance, in [Berchio et al. 2006; Gazzola and Sweers 2008; Nazarov et al. 2012], and whose core is contained in the following lemma:

Lemma 4.8. *Let $\Omega \subset \mathbb{R}^N$ be a bounded convex domain; fix $u \in H^2(\Omega) \cap H_0^1(\Omega)$ and define $\tilde{u}$ as the unique solution in $H_0^1(\Omega)$ of the Poisson problem*

$$\begin{cases} -\Delta \tilde{u} = |\Delta u| & \text{in } \Omega, \\ \tilde{u} = 0 & \text{on } \partial\Omega. \end{cases} \tag{4-4}$$

Then $\tilde{u} \in H^2(\Omega) \cap H_0^1(\Omega)$ and either $\tilde{u} > |u|$ in Ω and $\tilde{u}_n^2 \ge u_n^2$ on $\partial\Omega$ or $\tilde{u} = u$ in Ω.

Proof. Since Ω is convex by assumption, it satisfies in particular a uniform external ball condition and thus, by [Adolfsson 1992], we infer $\tilde{u} \in H^2(\Omega)$. Suppose $\tilde{u} \not\equiv u$. Since in particular $-\Delta\tilde{u} \ge \Delta u$ holds, by the maximum principle for strong solutions (see [Gilbarg and Trudinger 1998, Theorem 9.6]), one has $\tilde{u} > -u$ in Ω and so $\tilde{u}_n \le -u_n$. Similarly, $-\Delta\tilde{u} \ge -\Delta u$ implies also $\tilde{u} > u$ and $\tilde{u}_n \le u_n$ and so, combining them, we have the result. □

Proposition 4.9 (sublinear case). *Let $\Omega \subset \mathbb{R}^2$ be a bounded convex domain with $\partial\Omega$ of class $C^{1,1}$ and $\sigma \in (-1, 1]$. In addition to the assumption (H), suppose also that $f \ge 0$ and is positive for a subset of positive measure. If $u \in H^2(\Omega) \cap H_0^1(\Omega)$ is a nontrivial minimizer of J_σ, then u is strictly superharmonic in Ω, and thus positive.*

Proof. Firstly notice that $\kappa \ge 0$ a.e. on $\partial\Omega$ by the convexity of Ω. From u, define its superharmonic function $\tilde{u}$ as in Lemma 4.8. Supposing $\tilde{u} \not\equiv u$, by that result we can infer

$$J_\sigma(\tilde{u}) = \int_\Omega \frac{(\Delta\tilde{u})^2}{2} - \frac{1-\sigma}{2}\int_{\partial\Omega} \kappa\tilde{u}_n^2 - \int_\Omega F(x,\tilde{u}) \le \int_\Omega \frac{(\Delta u)^2}{2} - \frac{1-\sigma}{2}\int_{\partial\Omega} \kappa u_n^2 - \int_\Omega F(x,\tilde{u}). \tag{4-5}$$

Nevertheless, since $(\partial F/\partial s) = f \ge 0$, we have also that $F(x,u) < F(x,\tilde{u})$, and thus $J_\sigma(\tilde{u}) < J_\sigma(u)$, which leads to a contradiction. Hence necessarily $\tilde{u}$ coincides with u, so $-\Delta u = -\Delta\tilde{u} = |\Delta u| \ge 0$. As $u = 0$ on $\partial\Omega$ and $u \not\equiv 0$, we deduce $u > 0$ in Ω. □

It is clear that, when $f(x,0) \ne 0$, by Proposition 3.1, we always find a *nontrivial* global minimizer, which is positive by Proposition 4.9. For homogeneous nonlinearities this is not true in general. Anyway, for our model $f(x,s) = g(x)|s|^{p-1}s$, if we restrict our attention to the Nehari set, we easily see

$$J_\sigma(u) = \left(\frac{1}{2} - \frac{1}{p+1}\right)\|u\|_{H_\sigma} < 0$$

for every $u \ne 0$. So it is clear that in the minimization process we do not fall on 0. The same argument holds for more general nonlinearities $f(x,u)$, provided

$$\tfrac{1}{2} f(x,u)u - F(x,u) < 0 \quad \text{for all } u \in H^2(\Omega) \cap H_0^1(\Omega).$$

For instance this holds when $f(x,s) = g(x)|s|^{p-1}s + h(x)|s|^{q-1}s$ for $g, h > 0,\ p, q \in (0,1)$.

Remark 4.10. We stress here that, as a direct consequence of Proposition 4.9, we have obtained the positivity-preserving property also in the case of f not depending on u, i.e., for the linear Kirchhoff–Love functional I_σ (see also Remark 3.2). This generalizes the corresponding result by Parini and Stylianou [2009, Theorem 3.1] for bounded convex domains assuming only $C^{1,1}$ regularity on the boundary.

In our sublinear model case $f(x,s) = g(x)|s|^{p-1}s,\ p \in (0,1)$, something more may be deduced: in fact, Lemma 3.3 still applies and, with the same steps as in the proof of Lemma 3.4, (reversing the inequalities since now $p-1<0$), one ends up with

$$\|u\|_{H^2(\Omega)} \le \left(\frac{\|g\|_1 C(\Omega)}{(1-|\sigma|)C_0^{-1}}\right)^{\frac{1}{1-p}} \qquad \text{for all } u \in \mathcal{N}_\sigma.$$

As a result, we can state the following:

Proposition 4.11. *Let Ω be a bounded Lipschitz domain in $\mathbb{R}^2$ and let $g \in L^1(\Omega)$ be positive a.e. in Ω. For every $\sigma \in (-1,1)$ fixed, all critical points of J_σ with $f(x,s) = g(x)|s|^{p-1}s$ and $p \in (0,1)$ are uniformly bounded in $H^2(\Omega)$.*

Notice that by continuous embedding $H^2(\Omega) \hookrightarrow L^\infty(\Omega)$, one may also infer an a priori L^∞ bound for all critical points of J_σ. The estimate becomes also uniform with respect to σ if we restrict $\sigma \in I \Subset (-1,1)$.

Concerning the superlinear case with the nonlinearity (1-2), we obtain the same positivity result with the same assumptions on Ω and σ:

Proposition 4.12 (superlinear case). *Let $\Omega \subset \mathbb{R}^2$ be a bounded convex domain with $\partial\Omega$ of class $C^{1,1}$ and $\sigma \in (-1,1]$. Moreover suppose $f(x,u) = g(x)|u|^{p-1}u$, where $p>1$ and $g \in L^1(\Omega)$ positive a.e. in Ω. Then the ground states of the functional J_σ are positive in Ω.*

Proof. Suppose, by contradiction, that there exists $u \in \mathcal{N}_\sigma$ such that $J_\sigma(u) = \inf\{J_\sigma(v) \mid v \in \mathcal{N}_\sigma\}$ and u is not positive. With the same spirit of the proof of Proposition 4.9, consider the superharmonic function $\tilde{u}$ associated to u and suppose they are not the same. This time the inequality (4-5) is not sufficient to have a contradiction since we do not know whether $\tilde{u} \in \mathcal{N}_\sigma$. Nevertheless, by Lemma 3.3, there exists $t^* := t^*(\tilde{u}) \in \mathbb{R}^+$ such that $t^*\tilde{u} \in \mathcal{N}_\sigma$. Then,

$$\begin{aligned} J_\sigma(t^*\tilde{u}) &= (t^*)^2\left[\int_\Omega \frac{(\Delta\tilde{u})^2}{2} - \frac{1-\sigma}{2}\int_{\partial\Omega}\kappa\tilde{u}_n^2\right] - (t^*)^{p+1}\int_\Omega \frac{g(x)|\tilde{u}|^{p+1}}{p+1} \\ &< (t^*)^2\left[\int_\Omega \frac{(\Delta u)^2}{2} - \frac{1-\sigma}{2}\int_{\partial\Omega}\kappa u_n^2\right] - (t^*)^{p+1}\int_\Omega \frac{g(x)|u|^{p+1}}{p+1} = J_\sigma(t^*u) \le J_\sigma(u), \end{aligned} \tag{4-6}$$

which is again a contradiction. Notice that the last inequality holds since, by Lemma 3.3, J_σ restricted to every half-line attains its maximum on the Nehari manifold. Thus necessarily $\tilde{u}$ coincides with u, which implies that u is strictly superharmonic and thus positive. □

Remark 4.13. Notice that in the proofs of Propositions 4.9 and 4.12, if σ lies in the interval $(-1,1]$, the assumption Ω convex was necessary to have the good inequality for the second term of J_σ; on the other hand, if $\sigma>1$ we do not have anymore the right sign and we cannot conclude the argument.

5. Beyond the physical bounds: $\sigma \leq -1$

So far, we studied the existence of critical points of the functional J_σ with the assumption $\sigma \in (-1, 1]$, we described in a variational way the geometry of the ground states and we finally established their positivity. The aim of this section is to study what happens to the ground states of J_σ if we let the parameter be in the whole $\mathbb{R}$. Again, we are especially interested in studying their positivity.

Since the study is rather different if $\sigma \leq -1$ or $\sigma > 1$, we divide the subject into two sections. In both, we will always assume that $\Omega \subset \mathbb{R}^2$ is a bounded convex domain of class $C^{1,1}$ so that Theorem 4.7 holds. Moreover, as it seems more interesting from a mathematical point of view, we mainly focus on the superlinear case $f(x, u) = g(x)|u|^{p-1}u$ with $p > 1$, pointing out, if needed, the necessary adaptation for the sublinear power $p \in (0, 1)$.

A Steklov eigenvalue problem. Let us begin by recalling some known facts about the eigenvalue problem associated to equation (1-3) (see [Gazzola and Sweers 2008] or, for the case $\kappa = 1$, [Bucur et al. 2009; Berchio et al. 2006]):

$$\begin{cases} \Delta^2 u = 0 & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \\ \Delta u = d\kappa u_n & \text{on } \partial\Omega. \end{cases} \tag{5-1}$$

We define a *Steklov eigenvalue* to be a real value d such that (1-3) admits a nontrivial weak solution, named a *Steklov eigenfunction*, i.e., $u \in H^2(\Omega) \cap H_0^1(\Omega)$, $u \neq 0$, such that for all $\varphi \in H^2(\Omega) \cap H_0^1(\Omega)$,

$$\int_\Omega \Delta u \, \Delta\varphi - d \int_{\partial\Omega} \kappa u_n \varphi_n = 0. \tag{5-2}$$

First of all, d must be positive. In fact, if u is a Steklov eigenfunction, taking $u = \varphi$ in (5-2),

$$d \int_{\partial\Omega} \kappa (u_n)^2 = \int_\Omega (\Delta u)^2 > 0,$$

since $\|\Delta \cdot\|_2$ is a norm in $H^2(\Omega) \cap H_0^1(\Omega)$. As $\kappa \geq 0$, we have both $d > 0$ and $\int_{\partial\Omega} \kappa u_n^2 > 0$. As a complementary result, in order to show nontrivial solutions of (5-1), without loss of generality, we can restrict to the subset

$$\mathcal{H} = \left\{ u \in H^2(\Omega) \cap H_0^1(\Omega) \;\middle|\; \int_{\partial\Omega} \kappa (u_n)^2 \neq 0 \right\}.$$

Definition 5.1. We denote by $\tilde{\delta}_1(\Omega)$ the first Steklov eigenvalue for problem (5-1):

$$\tilde{\delta}_1(\Omega) := \inf_{\mathcal{H}\setminus\{0\}} \frac{\|\Delta u\|_2^2}{\int_{\partial\Omega} \kappa u_n^2}.$$

Proposition 5.2. *The first Steklov eigenvalue is attained, is positive and there exists a unique (up to a multiplicative constant) corresponding Steklov eigenfunction, which is positive in Ω.*

Proof. We refer to [Gazzola and Sweers 2008, Lemma 4.4], just noticing that the continuity of the curvature assumed therein was not necessary to obtain this result. ☐

A nonexistence and an existence result. From Proposition 5.2, it is easy to deduce a nonexistence result for positive solutions if σ is negative enough:

Proposition 5.3. *If $\sigma \le \sigma^* := 1 - \tilde{\delta}_1(\Omega)$, there is no nonnegative nontrivial solution for the Steklov boundary problem (1-3).*

Proof. Let u be a nonnegative solution for (1-3) and $\Phi_1 > 0$ be the first Steklov eigenfunction; we use Φ_1 as a test function in (4-2):

$$\int_\Omega \Delta u\, \Delta\Phi_1 - (1-\sigma)\int_{\partial\Omega} \kappa u_n(\Phi_1)_n = \int_\Omega g(x)u^p\Phi_1.$$

Then, interpreting u this time as a test function in (5-2), we have

$$\int_\Omega \Delta u\, \Delta\Phi_1 = \tilde{\delta}_1(\Omega)\int_{\partial\Omega} \kappa(\Phi_1)_n u_n.$$

Combining the two equalities,

$$(\tilde{\delta}_1(\Omega) - (1-\sigma))\int_{\partial\Omega} \kappa(\Phi_1)_n u_n = \int_\Omega g(x)u^p\Phi_1 > 0.$$

Again by positivity of u and Φ_1, we have $u_n \le 0$ and $(\Phi_1)_n \le 0$ so, as $\kappa \ge 0$, we finally end up with $\tilde{\delta}_1(\Omega) - 1 + \sigma > 0$, which is exactly what we wanted. ☐

Remark 5.4. We already proved that our problem (1-3) admits positive solutions whenever $\sigma \in (-1, 1]$ with the same assumptions on Ω. Hence, we infer that, $\tilde{\delta}_1(\Omega) \ge 2$ and we have equality if $\Omega = B_1(0)$ (see [Berchio et al. 2006, Proposition 12]). This result was already proved for C^2 bounded convex domains of $\mathbb{R}^2$ by Parini and Stylianou [2009, Remark 3.3], using Fichera's duality principle.

The next step is to investigate what happens if $\sigma \in (\sigma^*, -1]$ in case this interval is nonempty. We will show that the existence and the positivity results found for $\sigma \in (-1, 1]$ can be extended for this case. In fact, the only restriction we have to overcome is the fact that here Lemma 2.1 is not the right way to prove that the first two terms in the functional J_σ define indeed a norm on $H^2(\Omega) \cap H_0^1(\Omega)$.

Lemma 5.5. *For every $\sigma > \sigma^*$, the map*

$$u \mapsto \left[\int_\Omega (\Delta u)^2 - (1-\sigma)\int_{\partial\Omega} \kappa(u_n)^2\right]^{\frac{1}{2}} := \|u\|_{H_\sigma}$$

is a norm in $H^2(\Omega) \cap H_0^1(\Omega)$ equivalent to the standard norm.

Proof. By the definition of $\tilde{\delta}_1(\Omega)$ as an inf, we have $\|\Delta u\|_2^2 \ge \tilde{\delta}_1(\Omega) \int_{\partial\Omega} \kappa u_n^2$ for each $u \in H^2(\Omega) \cap H_0^1(\Omega)$ and so, if $d > 0$ (which corresponds to $\sigma < 1$),

$$\int_\Omega (\Delta u)^2 \ge \int_\Omega (\Delta u)^2 - d\int_{\partial\Omega} \kappa u_n^2 \ge \int_\Omega (\Delta u)^2\left(1 - \frac{d}{\tilde{\delta}_1(\Omega)}\right). \tag{5-3}$$

On the other hand, if $d < 0$ (so that $\sigma > 1$),

$$\int_\Omega (\Delta u)^2 \le \int_\Omega (\Delta u)^2 + |d|\int_{\partial\Omega} \kappa u_n^2 \le \int_\Omega (\Delta u)^2\left(1 + \frac{|d|}{\tilde{\delta}_1(\Omega)}\right).$$

As a result, we have to impose that $d < \tilde{\delta}_1(\Omega)$ to have the positivity of the constant in the first estimate, while no restriction occurs in the second. The proof is completed noticing that the map

$$(u, v)_{H_\sigma} \mapsto \int_\Omega \Delta u\, \Delta v - d \int_{\partial\Omega} \kappa u_n v_n$$

defines a scalar product on $H^2(\Omega) \cap H_0^1(\Omega)$ by inequality (5-3) for all $d < \tilde{\delta}_1(\Omega)$. □

Proposition 5.6. *Let $\Omega \subset \mathbb{R}^2$ be a bounded convex domain with boundary $C^{1,1}$ and suppose $\sigma \in (\sigma^*, -1]$; then the functional J_σ admits a positive ground state.*

Proof. It is sufficient to notice that Lemma 3.4 holds for these values of σ if we replace Lemma 2.1 by Lemma 5.5, while all the other propositions that led to the existence and the positivity of ground states are not affected by this change. □

Remark 5.7 (sublinear case). Both Propositions 5.3 and 5.6 hold in the case of a function $f(x, u)$ which verifies the assumption (H) (modifying in a suitable way the constant in front of the quadratic term) and $f \geq 0$, $f \not\equiv 0$.

Approaching σ^*. As we know now the existence of positive ground state solutions for $\sigma \in (\sigma^*, 1]$ and that there are no positive solutions if $\sigma \leq \sigma^*$, a natural question that arises is determining the behaviour of a sequence $(u_k)_{k\in\mathbb{N}}$, each of them being a ground state for the respective functional J_{σ_k}, as $\sigma_k \searrow \sigma^*$. We will find an antipodal result for $f(x, u) = g(x)|u|^{p-1}u$ as $p \in (1, +\infty)$ or $p \in (0, 1)$.

The following proof is an adaptation of [Berchio et al. 2007, Theorem 1], which covers the critical case $f(x, u) = |u|^{2^*-2}u$ when the dimension is $N \geq 5$. Moreover, the authors considered a slightly different notion of solution, that is, the minimizers of the Rayleigh quotient associated to the boundary value problem:

$$R_\sigma(u) := \frac{\|\Delta u\|_2^2 - (1-\sigma)\int_{\partial\Omega} \kappa u_n^2}{\left(\int_\Omega g(x)|u|^{p+1}\right)^{\frac{2}{p+1}}} \tag{5-4}$$

Anyway, it is a standard fact to prove that every ground state of J_σ is also a minimizer of R_σ, while the converse is also true, up to a multiplication by a constant.

Theorem 5.8. *Let Ω be as in Proposition 5.6 and $\sigma_k \searrow \sigma^*$ as $k \to +\infty$. If $p \in (0, 1)$, then $\|u_k\|_\infty \to +\infty$, while, if $p > 1$, then $\|u_k\|_{H^2(\Omega)} \to 0$.*

Proof. Let $p > 0$, $p \neq 1$; by the remark above, each ground state u_k is such that

$$R_{\sigma_k}(u_k) = \inf_{0 \neq u \in H^2(\Omega)\cap H_0^1(\Omega)} R_{\sigma_k}(u) := \Sigma_{\sigma_k} \geq 0.$$

By Proposition 5.2, there exists a positive first Steklov eigenfunction Φ_1; since we have $\|\Delta\Phi_1\|_2^2 = (1-\sigma^*)\int_{\partial\Omega} \kappa(\Phi_1)_n^2$, we have

$$0 \leq \Sigma_{\sigma_k} \leq R_{\sigma_k}(\Phi_1) = (\sigma_k - \sigma^*) \frac{\int_{\partial\Omega} \kappa(\Phi_1)_n^2}{\left(\int_\Omega g(x)|\Phi_1|^{p+1}\right)^{\frac{2}{p+1}}} \to 0$$

as $k \to +\infty$. Moreover, since u_k is a ground state for J_{σ_k},

$$\|\Delta u_k\|_2^2 - (1-\sigma_k)\int_{\partial\Omega} \kappa (u_k)_n^2 = \int_\Omega g(x)|u_k|^{p+1}$$

and, since $R_{\sigma_k}(u_k) = \Sigma_{\sigma_k}$, we deduce

$$\left(\int_\Omega g(x)|u_k|^{p+1}\right)^{\frac{p-1}{p+1}} = \Sigma_{\sigma_k} \to 0.$$

Hence, if $p>1$, then $\int_\Omega g(x)|u_k|^{p+1} \to 0$; otherwise, if $p \in (0,1)$, then $\int_\Omega g(x)|u_k|^{p+1} \to +\infty$, which implies, by the Hölder inequality as $g \in L^1(\Omega)$, that $\|u_k\|_\infty \to +\infty$.

We have now to prove that, if $p>1$, this convergence to 0 is actually in the natural norm $H^2(\Omega)$. By Lemma 5.5, $\|\cdot\|_{H_{\sigma_k}}$ is a norm in $H^2(\Omega) \cap H_0^1(\Omega)$ for every k, so we are able to decompose in that norm the Hilbert space as $H^2(\Omega) \cap H_0^1(\Omega) = \operatorname{span}(\Phi_1) \oplus [\operatorname{span}(\Phi_1)]^\perp$. Thus, for every k there exist a unique $\alpha_k \in \mathbb{R}$ and $\psi_k \in [\operatorname{span}(\Phi_1)]^\perp$ such that $u_k = \alpha_k \Phi_1 + \psi_k$.

Hence, for k large enough,

$$\begin{aligned} o(1) \ge \int_\Omega g(x)|u_k|^{p+1} &= \|\Delta u_k\|_2^2 - (1-\sigma_k)\int_{\partial\Omega} \kappa (u_k)_n^2 = (u_k, u_k)_{H_{\sigma_k}} \\ &= \alpha_k^2 (\Phi_1, \Phi_1)_{H_{\sigma_k}} + (\psi_k, \psi_k)_{H_{\sigma_k}}. \end{aligned} \tag{5-5}$$

First of all,

$$(\Phi_1, \Phi_1)_{H_{\sigma_k}} = \|\Delta \Phi_1\|_2^2 - (1-\sigma_k)\int_{\partial\Omega} \kappa (\Phi_1)_n^2 = (\sigma_k - \sigma^*)\int_{\partial\Omega} \kappa (\Phi_1)_n^2. \tag{5-6}$$

Moreover, denoting by $\tilde{\delta}_2(\Omega)$ the second eigenvalue of the Steklov problem, i.e.,

$$\tilde{\delta}_2(\Omega) = \inf_{\operatorname{span}(\Phi_1)^\perp \setminus \{0\}} \frac{\|\Delta v\|_2^2}{\int_{\partial\Omega} \kappa v_n^2},$$

and defining $\sigma^{**} := 1 - \tilde{\delta}_2(\Omega)$, we get

$$\|\Delta \psi_k\|_2^2 \ge (1-\sigma^{**})\int_{\partial\Omega} \kappa (\psi_k)_n^2,$$

from which

$$(\psi_k, \psi_k)_{H_{\sigma_k}} = \|\Delta\psi_k\|_2^2 - (1-\sigma_k)\int_{\partial\Omega} \kappa(\psi_k)_n^2 \ge \|\Delta\psi_k\|_2^2 - \frac{1-\sigma_k}{1-\sigma^{**}}\|\Delta\psi_k\|_2^2 = \frac{\sigma_k - \sigma^{**}}{1-\sigma^{**}}\|\Delta\psi_k\|_2^2. \tag{5-7}$$

As a result, combining (5-5) with (5-6) and (5-7), we get

$$o(1) \ge \int_\Omega g(x)|u_k|^{p+1} = \alpha_k^2(\sigma_k - \sigma^*)\int_{\partial\Omega} \kappa(\Phi_1)_n^2 + \frac{\sigma_k - \sigma^{**}}{1-\sigma^{**}}\|\Delta\psi_k\|_2^2.$$

Since we proved in Proposition 5.2 that the first Steklov eigenfunction is simple, we have $\sigma^{**} < \sigma^*$ and, recalling that $\sigma_k > \sigma^*$ by assumption, necessarily $\|\Delta\psi_k\|_2 \to 0$. Hence,

$$\begin{aligned} \int_\Omega g(x)|\alpha_k\Phi_1|^{p+1} \le \int_\Omega g(x)\big[|u_k| + |\psi_k|\big]^{p+1} &\le 2^p \int_\Omega g(x)\big[|u_k|^{p+1} + |\psi_k|^{p+1}\big] \\ &\le 2^p \int_\Omega g(x)|u_k|^{p+1} + C^{p+1}(\Omega)\|g\|_1 \|\psi_k\|_{H^2(\Omega)} \to 0. \end{aligned}$$

As a result, $\alpha_k \to 0$ and we finally obtain

$$\|u_k\|_{H^2(\Omega)} \le |\alpha_k| \, \|\Phi_1\|_{H^2(\Omega)} + \|\psi_k\|_{H^2(\Omega)} \to 0. \qquad \square$$

If we read carefully the proof of Theorem 5.8, we notice that the fact that each u_k is a ground state for J_σ was necessary only to deduce that $\int_\Omega g(x)|u_k|^{p+1} \to 0$, while to prove the convergence to 0 in $H^2(\Omega)$ norm it was only sufficient that each u_k is a critical point (actually, an element of the Nehari manifold $\mathcal{N}_{\sigma_k}$, since the only step of the proof involved is (5-5)). Consequently, we can directly state the following lemma, which will be useful when we will look at the radial case in Section 7:

Lemma 5.9. *Let $(u_k)_{k\in\mathbb{N}}$ be a sequence of critical points of J_{σ_k} in the superlinear case such that $\int_\Omega g(x)|u_k|^{p+1} \to 0$ as $\sigma_k \searrow \sigma^*$. Then $\|u_k\|_{H^2(\Omega)} \to 0$.*

6. Beyond the physical bounds: $\sigma > 1$

As briefly announced at the beginning of the previous section, here we want to investigate the behaviour of the ground states of J_σ when $\sigma > 1$. We assume again hereafter that $\Omega \subset \mathbb{R}^2$ is a bounded convex domain with $C^{1,1}$ boundary and (1-2) concerning the nonlinearity. As a consequence, the extension of the existence result is straightforward: in fact, in this case, by Lemma 5.5, $\|\cdot\|_{H_\sigma(\Omega)}$ is still a norm on $H^2(\Omega) \cap H^1_0(\Omega)$ and we can repeat the usual steps. Notice also that it is equivalent by these assumptions on Ω to consider critical points of J_σ as far as weak solutions of the semilinear problem (1-3).

The extension of positivity in this case seems not to be obvious, as already noticed in Remark 4.13. We will provide here two different proofs (which will produce two slightly different results); the first one relies on the study of the convergence of ground states as $\sigma \to 1$, which in the limit yields the Navier case, while the second is based on the method of dual cones by Moreau, connecting our semilinear problem with the linear one. We point out that the convergence result might be also of independent interest.

In the following, we will always consider the exponent of the nonlinearity (1-2) to be $p > 1$, but similar results can be proved also in the sublinear framework (see Remarks 6.6 and 6.22).

Convergence of ground states of J_σ to ground states of J_{NAV} as $\sigma \to 1$. In this section, $(u_k)_{k\in\mathbb{N}}$ will always denote a sequence of ground states solutions of the Steklov problems

$$\begin{cases} \Delta^2 u = g(x)|u|^{p-1}u & \text{in } \Omega, \\ u = \Delta u - (1-\sigma_k)\kappa u_n = 0 & \text{on } \partial\Omega \end{cases} \tag{6-1}$$

for a sequence $(\sigma_k)_{k\in\mathbb{N}}$ converging to 1. Moreover, in order to underline the peculiarity of the problem when $\sigma = 1$, we set $J_{\mathrm{NAV}} := J_1$, whose critical points are the weak solution of the following Navier problem:

$$\begin{cases} \Delta^2 u = g(x)|u|^{p-1}u & \text{in } \Omega, \\ u = \Delta u = 0 & \text{on } \partial\Omega, \end{cases} \tag{6-2}$$

Finally, $\bar{u}$ will always denote a ground state of J_{NAV}. Our main result is to prove the convergence $u_k \to \bar{u}$ in the natural norm, i.e., in $H^2(\Omega)$, as $\sigma_k \to 1$, no matter if σ_k is less or greater than 1. First of all, a weaker result is enough:

Lemma 6.1. *Let $(u_k)_{k\in\mathbb{N}}$ and $\bar u$ be as specified above. If $u_k \rightharpoonup \bar u$ weakly in $H^2(\Omega)$, then (up to a subsequence) $u_k \to \bar u$ strongly in $H^2(\Omega)$ as $\sigma_k \to 1$.*

Proof. As $u_k \rightharpoonup \bar u$ weakly in $H^2(\Omega)$, there exists $M > 0$ such that $\|u_k\|^2_{H^2(\Omega)} \le M$. Moreover, for each $k \in \mathbb{N}$, we know u_k is a solution of (6-1) and $\bar u$ of the Navier problem (6-2); thus, for every test function $\varphi \in H^2(\Omega) \cap H^1_0(\Omega)$,

$$\begin{aligned} \int_\Omega \Delta u_k\, \Delta\varphi - (1-\sigma_k)\int_{\partial\Omega} \kappa (u_k)_n \varphi_n &= \int_\Omega g(x)|u_k|^{p-1}u_k\varphi, \\ \int_\Omega \Delta\bar u\, \Delta\varphi &= \int_\Omega g(x)|\bar u|^{p-1}\bar u\varphi. \end{aligned} \tag{6-3}$$

Hence

$$\begin{aligned} &C_A^{-1}\|u_k-\bar u\|^2_{H^2(\Omega)} \\ &\quad\le \|\Delta u_k - \Delta \bar u\|_2^2 = \int_\Omega \Delta u_k\, \Delta(u_k-\bar u) - \int_\Omega \Delta\bar u\, \Delta(u_k-\bar u) \\ &\quad= (1-\sigma_k)\int_{\partial\Omega}\kappa (u_k)_n(u_k-\bar u)_n + \left[\int_\Omega g(x)|u_k|^{p-1}u_k(u_k-\bar u) - \int_\Omega g(x)|\bar u|^{p-1}\bar u(u_k-\bar u)\right]. \end{aligned}$$

For the first term,

$$\begin{aligned} \left|(1-\sigma_k)\int_{\partial\Omega}\kappa(u_k)_n(u_k-\bar u)_n\right| &\le |1-\sigma_k|C_T^2\|\kappa\|_{L^\infty(\partial\Omega)}\|u_k\|_{H^2(\Omega)}\|u_k-\bar u\|_{H^2(\Omega)} \\ &\le |1-\sigma_k|C_T^2\|\kappa\|_{L^\infty(\partial\Omega)}M(M+\|\bar u\|_{H^2(\Omega)}) \to 0, \end{aligned}$$

where C_T is the constant in the trace theorem. Concerning the second, it is enough to invoke the dominated convergence theorem as we have pointwise convergence and since

$$\big|g(x)\big(|u_k|^{p-1}u_k - |\bar u|^{p-1}\bar u\big)(u_k-\bar u)\big| \le |g(x)|\big[C(\Omega)^p M^p + |\bar u|^p\big]\big[C(\Omega)M + \bar u\big] \in L^1(\Omega),$$

where $C(\Omega)$ is the constant in the embedding $H^2(\Omega) \hookrightarrow L^\infty(\Omega)$. □

Remark 6.2. This result holds not only for ground states, but for generic solutions; i.e., if $(u_k)_{k\in\mathbb{N}}$ is a sequence of weak solutions of the Steklov problem (6-1) and $\bar u$ a weak solution of the Navier problem (6-2) and we know that $u_k \rightharpoonup \bar u$ weakly in $H^2(\Omega)$, then, up to a subsequence, it converges strongly too.

A crucial observation is that the Nehari manifolds are nested with respect to the parameter σ:

Lemma 6.3. *Let $\sigma_1 < \sigma_2$ and fix $u \in H^2(\Omega)\cap H^1_0(\Omega)\setminus\{0\}$. Then*

$$t^*_{\sigma_1}(u) \le t^*_{\sigma_2}(u).$$

Proof. In fact, $-(1-\sigma_1) < -(1-\sigma_2)$ and so

$$t^*_{\sigma_1}(u) = \left(\frac{\int_\Omega(\Delta u)^2 - (1-\sigma_1)\int_{\partial\Omega}\kappa u_n^2}{\int_\Omega g(x)|u|^{p+1}}\right)^{\frac{1}{p-1}} \le \left(\frac{\int_\Omega(\Delta u)^2 - (1-\sigma_2)\int_{\partial\Omega}\kappa u_n^2}{\int_\Omega g(x)|u|^{p+1}}\right)^{\frac{1}{p-1}} = t^*_{\sigma_2}(u). \quad □$$

Notice that if $u \in H^2_0(\Omega)$ then one has the equality; if we suppose moreover that $\kappa > 0$ a.e., we deduce also the converse.

Proposition 6.4. *The sequence of ground states $(u_k)_{k\in\mathbb{N}}$ is bounded in $H^2(\Omega)$.*

Proof. Set $k_{\max}$ such that $\sigma_{k_{\max}} = \max\{(\sigma_k)_{k\in\mathbb{N}}, 1\}$ and so $u_{k_{\max}}$ is a ground state for $J_{\sigma_{k_{\max}}}$ (with the convention that if $\sigma_{k_{\max}} = 1$, then $u_{k_{\max}}$ is a ground state for J_{NAV}).

Defining $w_k := t^*_{\sigma_k}(u_{k_{\max}})u_{k_{\max}} \in \mathcal{N}_{\sigma_k}$, that is, the "projection" of $u_{k_{\max}}$ on the Nehari manifold $\mathcal{N}_{\sigma_k}$ along its half-line, one has

$$\int_\Omega g(x)|u_k|^{p+1} \le \int_\Omega g(x)|w_k|^{p+1} \le \int_\Omega g(x)|u_{k_{\max}}|^{p+1}. \tag{6-4}$$

Indeed, the first inequality comes from the fact that u_k is a ground state of J_{σ_k}, which has the equivalent formulation (3-2); the second is obtained by Lemma 6.3 since

$$\begin{aligned}\int_\Omega g(x)|w_k|^{p+1} &= (t^*_{\sigma_k}(u_{k_{\max}}))^{p+1}\int_\Omega g(x)|u_{k_{\max}}|^{p+1}\\ &\le (t^*_{\sigma_{k_{\max}}}(u_{k_{\max}}))^{p+1}\int_\Omega g(x)|u_{k_{\max}}|^{p+1} = \int_\Omega g(x)|u_{k_{\max}}|^{p+1}.\end{aligned}$$

Furthermore, for a generic $\sigma > 0$ (and here we can assume it without loss of generality),

$$\int_\Omega (\Delta u)^2 - (1-\sigma)\int_{\partial\Omega} \kappa u_n^2 \ge \min\{\sigma, 1\} C_A(\Omega)\|u\|^2_{H^2(\Omega)}. \tag{6-5}$$

In fact, if $\sigma \in [1, +\infty)$ the proof is straightforward since $-(1-\sigma) \ge 0$; otherwise, if $\sigma \in (0, 1)$,

$$\begin{aligned}\int_\Omega (\Delta u)^2 - (1-\sigma)\int_{\partial\Omega}\kappa u_n^2 &= \int_\Omega (\Delta u)^2 + 2(1-\sigma)\int_\Omega (-\det(\nabla^2 u))\\ &= \int_\Omega \bigl[u_{xx}^2 + u_{yy}^2 + 2\sigma u_{xx}u_{yy} + 2(1-\sigma)u_{xy}^2\bigr] \ge \sigma\int_\Omega(\Delta u)^2 + 2(1-\sigma)\int_\Omega u_{xy}^2\\ &\ge \sigma\int_\Omega (\Delta u)^2 \ge \sigma C_A^{-1}(\Omega)\|u\|^2_{H^2(\Omega)}.\end{aligned}$$

As a result, combining (6-4) with (6-5), we get

$$\|u_k\|^2_{H^2(\Omega)} \le \frac{C_A(\Omega)}{\min\{\sigma_k, 1\}}\int_\Omega g(x)|u_k|^{p+1} \le \frac{C_A(\Omega)}{\min\{\sigma_k, 1\}}\int_\Omega g(x)|u_{k_{\max}}|^{p+1},$$

which is the estimate we needed. □

As a direct consequence of Proposition 6.4, the sequence $(u_k)_{k\in\mathbb{N}}$, up to a subsequence, is weakly convergent to some $u_\infty \in H^2(\Omega)\cap H^1_0(\Omega)$ with strong convergence in $L^\infty(\Omega)$. It is also easy to see that u_∞ is a weak solution of the Navier problem (6-2): it is enough to apply to (6-3) the weak convergence in $H^2(\Omega)$, the strong convergence in $L^2(\partial\Omega)$ of the normal derivatives and the dominated convergence theorem. As a consequence, by Lemma 6.1, the convergence $u_k \to u_\infty$ is strong in $H^2(\Omega)$.

Theorem 6.5. *Let $\sigma_k \to 1$ and Ω be a bounded convex domain in $\mathbb{R}^2$ with boundary of class $C^{1,1}$. Then the sequence $(u_k)_{k\in\mathbb{N}}$ of ground state solutions for the Steklov problems (6-1) admits a subsequence $(u_{k_j})_{j\in\mathbb{N}}$ which converges in $H^2(\Omega)$ to u_∞, which is a ground state for the Navier problem (6-2), and thus strictly superharmonic.*

Proof. Clearly, as u_∞ is weak solution of (6-2), we have $J_{\mathrm{NAV}}(u_\infty) \geq \inf_{\mathcal{N}_{\mathrm{NAV}}} J_{\mathrm{NAV}}$. Now we have to prove the converse inequality. Firstly, we have $J_{\mathrm{NAV}}(u_\infty) \leq \liminf_{k\to+\infty} J_{\sigma_k}(u_k)$. Indeed,

$$\begin{aligned}\liminf_{k\to+\infty} J_{\sigma_k}(u_k) &= \liminf_{k\to+\infty} \int_\Omega \frac{(\Delta u_k)^2}{2} - \lim_{k\to+\infty} \frac{1-\sigma_k}{2} \int_{\partial\Omega} \kappa (u_k)_n^2 - \lim_{k\to+\infty} \int_\Omega \frac{g(x)|u_k|^{p+1}}{p+1} \\ &\geq \int_\Omega \frac{(\Delta u_\infty)^2}{2} - \int_\Omega \frac{g(x)|u_\infty|^{p+1}}{p+1} = J_{\mathrm{NAV}}(u_\infty),\end{aligned}$$

having used the compactness of the map $\partial_n : H^1(\Omega)^2 \to L^2(\Omega)$ and the dominated convergence theorem. Moreover, if we suppose $\sigma_k < 1$ for k large enough, by Lemma 6.3 (with a similar argument to that in (6-4)), for all $k \in \mathbb{N}$ we have

$$J_{\sigma_k}(u_k) = \left(\frac{1}{2} - \frac{1}{p+1}\right) \int_\Omega g(x)|u_k|^{p+1} \leq \left(\frac{1}{2} - \frac{1}{p+1}\right) \int_\Omega g(x)|u_\infty|^{p+1} = J_{\mathrm{NAV}}(u_\infty), \tag{6-6}$$

so in this case we are done. If otherwise $\sigma_k > 1$ for a infinite number of indices, (6-6) does not hold. In this case, without loss of generality, we can assume that $\sigma_k \searrow 1$. By the existence theorems in Section 3, we know that there exists a ground state $\bar{u} \in H^2(\Omega) \cap H_0^1(\Omega)$ for J_{NAV} and we define $\bar{u}_k := t^*_{\sigma_k}(\bar{u})\bar{u}$ to be the "projection" on the Nehari manifold $\mathcal{N}_{\sigma_k}$. Then $\|\bar{u}_k - \bar{u}\|_{H^2(\Omega)} = |1 - t^*_{\sigma_k}(\bar{u})| \|\bar{u}\|_{H^2(\Omega)}$ with

$$1 - (t^*_{\sigma_k}(\bar{u}))^{p-1} \overset{[\bar{u}\in\mathcal{N}_{\mathrm{NAV}}]}{=} (t^*_{\mathrm{NAV}}(\bar{u}))^{p-1} - (t^*_{\sigma_k}(\bar{u}))^{p-1} = 2(1-\sigma_k)\frac{\int_\Omega \det(\nabla^2 \bar{u})}{\int_\Omega g(x)|\bar{u}|^{p+1}} \to 0,$$

so $\bar{u}_k \to \bar{u}$ in $H^2(\Omega)$, which implies

$$\int_\Omega g(x)|\bar{u}_k|^{p+1} \to \int_\Omega g(x)|\bar{u}|^{p+1}. \tag{6-7}$$

Nevertheless, since u_k is a ground state of J_{σ_k},

$$\int_\Omega g(x)|\bar{u}_k|^{p+1} \overset{[\bar{u}_k\in\mathcal{N}_{\sigma_k}]}{=} \left(\frac{1}{2} - \frac{1}{p+1}\right) J_{\sigma_k}(\bar{u}_k) \geq \left(\frac{1}{2} - \frac{1}{p+1}\right) J_{\sigma_k}(u_k) \overset{[u_k\in\mathcal{N}_{\sigma_k}]}{=} \int_\Omega g(x)|u_k|^{p+1}; \tag{6-8}$$

furthermore, since we assumed $\sigma_k > 1$ and by Lemma 6.3,

$$\begin{aligned}\int_\Omega g(x)|u_k|^{p+1} &\geq \int_\Omega g(x)|t^*_{\mathrm{NAV}}(u_k)u_k|^{p+1} = \left(\frac{1}{2} - \frac{1}{p+1}\right) J_{\mathrm{NAV}}(t^*_{\mathrm{NAV}}(u_k)u_k) \\ &\geq \left(\frac{1}{2} - \frac{1}{p+1}\right) J_{\mathrm{NAV}}(\bar{u}) = \int_\Omega g(x)|\bar{u}|^{p+1}.\end{aligned} \tag{6-9}$$

Combining (6-7), (6-8) and (6-9), we find that

$$\int_\Omega g(x)|u_k|^{p+1} \to \int_\Omega g(x)|\bar{u}|^{p+1}, \tag{6-10}$$

from which $J_{\sigma_k}(u_k) \to J_{\mathrm{NAV}}(\bar{u})$, which completes our equality.

To conclude, notice that we have already obtained in the proof of Proposition 4.12 that ground states of the Navier equation (6-2) are strictly superharmonic. $\square$

Remark 6.6. The same analysis may be adapted also for the sublinear case $p \in (0, 1)$, paying attention to some minor changes: for instance, Lemma 6.3 holds with the reverse inequality, but this compensates for the fact that this time the coefficient $\frac{1}{2} - \frac{1}{p}$ in the equivalent formulation of J_σ is negative.

Regularity of solutions and $W^{2,q}$ convergence of ground states. The convergence result of the previous section will be used to derive positivity of ground states when σ lies in a right neighborhood of 1. Nevertheless, we will need a $C^{0,1}$ convergence to be able to control the normal derivatives on the boundary; thus we have to upgrade our convergence to a stronger norm. The first step will be to investigate, for a fixed $\sigma > \sigma^*$, the regularity of solutions of (1-3) and (6-2) with just a slightly more regular boundary (actually, we will have to impose that $\partial\Omega$ is of class C^2). This will be obtained by means of the following lemma by Gazzola, Grunau and Sweers, which follows from a result by Agmon, Douglis and Nirenberg [Agmon et al. 1959, Theorem 15.3′, p. 707]:

Lemma 6.7 [Gazzola et al. 2010, Corollary 2.23]. *Let $q > 1$ and take an integer $m \geq 4$. Assume that $\partial\Omega \in C^m$ and $a \in C^{m-2}$. Then there exists $C = C(m, q, a, \Omega) > 0$ such that*

$$\|u\|_{W^{m,q}(\Omega)} \leq C\big(\|u\|_q + \|\Delta^2 u\|_{W^{m-4,q}(\Omega)} + \|u\|_{W^{m-1/q,q}(\partial\Omega)} + \|\Delta u - a u_n\|_{W^{m-2-1/q,q}(\partial\Omega)}\big)$$

for every $u \in W^{m,q}(\Omega)$. The same statement holds for any $m \geq 2$ provided the norms on the right-hand side are suitably interpreted.

Hence we have to define $\Delta^2 u$ as a distribution in $W^{-2,q}(\Omega)$, i.e., acting on functions in $W_0^{2,q'}(\Omega)$. Let $u \in H^2(\Omega) \cap H_0^1(\Omega)$ be a weak solution of (1-3); we define the linear functional over $H^2(\Omega)$

$$\Delta^2 u : H^2(\Omega) \ni \varphi \mapsto \langle \Delta^2 u, \varphi \rangle := \int_\Omega \Delta u \, \Delta\varphi,$$

which is well defined and continuous. If we let

$$u_g^p : \varphi \mapsto \langle u_g^p, \varphi \rangle := \int_\Omega g(x)|u|^{p-1} u\varphi,$$

it is clearly well defined and continuous on $W_0^{2,q'}(\Omega)$ and, by the weak formulation of the PDE, on the subset $H_0^2(\Omega)$ it acts identically as $\Delta^2 u$. As a result, we define

$$\Delta^2 u : W_0^{2,q'}(\Omega) \ni \varphi \mapsto \langle \Delta^2 u, \varphi \rangle := \int_\Omega g(x)|u|^{p-1} u\varphi. \tag{6-11}$$

Proposition 6.8. *If $\partial\Omega \in C^2$, for every $\sigma > \sigma^*$ the weak solutions of Steklov and Navier problems (6-1) and (6-2) lie in $W^{2,q}(\Omega)$ for every $q > 2$.*

Proof. Let $u \in H^2(\Omega) \cap H_0^1(\Omega)$ be a weak solution of (1-3). Applying Lemma 6.7 with $m = 2$ and $a = (1-\sigma)\kappa \in C^0(\partial\Omega)$ ($a = 0$ for the Navier case), we find

$$\|u\|_{W^{2,q}(\Omega)} \leq C(q, \sigma, \Omega)\big(\|u\|_q + \|\Delta^2 u\|_{W^{-2,q}(\Omega)}\big),$$

which is well defined in view of (6-11). Since

$$\|\Delta^2 u\|_{W^{-2,q}(\Omega)} = \sup_{0 \neq \varphi \in W_0^{2,q'}(\Omega)} \frac{\left|\int_\Omega g(x)|u|^{p-1} u\varphi\right|}{\|\varphi\|_{W_0^{2,q'}(\Omega)}} \leq C(p, q, \Omega)\|g\|_1 \|u\|_{H^2(\Omega)}^p, \tag{6-12}$$

we finally deduce from (6-12) that

$$\|u\|_{W^{2,q}(\Omega)} \le C(q,\sigma,\Omega)\big(\|u\|_q + C(p,q,\Omega)\|g\|_1 \|u\|^p_{H^2(\Omega)}\big) < +\infty. \qquad \square$$

We stress that we did not use either the fact that u is a ground state solution, or its positivity: the above result holds true for every weak solution of Steklov and Navier problems.

Let us now recall the following interpolation result:

Lemma 6.9 (interpolation of fractional Sobolev spaces, [Brezis and Mironescu 2001, Corollary 2]). *For $0 \le s_1 < s_2 < +\infty$, $1 < p_1, p_2 < +\infty$, for every s, p such that $s = \theta s_1 + (1-\theta)s_2$ and $\frac{1}{p} = \frac{\theta}{p_1} + \frac{1-\theta}{p_2}$, we have*

$$\|f\|_{W^{s,p}(\mathbb{R}^N)} \le C\|f\|^{\theta}_{W^{s_1,p_1}(\mathbb{R}^N)}\|f\|^{1-\theta}_{W^{s_2,p_2}(\mathbb{R}^N)}.$$

Proposition 6.10. *Let Ω be of class C^2 and $(u_k)_{k\in\mathbb{N}}$ be a sequence of weak solutions for the Steklov problems (6-1) converging in $H^2(\Omega)$ to $\bar u$, a weak solution for the Navier problem (6-2). Then the convergence is in $W^{2,q}(\Omega)$ for every $q \ge 2$.*

Proof. Let $q \ge 2$ and apply the regularity estimate of Lemma 6.7 to $u_k - \bar u$ with $m = 2$, $a = 0$:

$$\|u_k - \bar u\|_{W^{2,q}(\Omega)} \le C(q,\Omega)\big(\|u_k-\bar u\|_q + \|\Delta^2 u_k - \Delta^2\bar u\|_{W^{-2,q}(\Omega)} + |1-\sigma_k|\,\|\kappa(u_k)_n\|_{W^{-1/q,q}(\partial\Omega)}\big), \tag{6-13}$$

using that on $\partial\Omega$ we have $\Delta(u_k-\bar u) - a(u_k-\bar u)_n = \Delta u_k - \Delta\bar u = (1-\sigma_k)\kappa(u_k)_n$.

By (6-11) and the dominated convergence theorem,

$$\|\Delta^2 u_k - \Delta^2\bar u\|_{W^{-2,q}(\Omega)} = \sup_{0\ne\varphi\in W^{2,q'}_0(\Omega)} \frac{\big|\int_\Omega g(x)|u_k|^{p-1}u_k\varphi - \int_\Omega g(x)|\bar u|^{p-1}\bar u\varphi\big|}{\|\varphi\|_{W^{2,q'}_0(\Omega)}} \to 0,$$

similarly to (6-12). We need now to prove that $(\kappa(u_k)_n)_{k\in\mathbb{N}}$ is bounded in $W^{-\frac{1}{q},q}(\partial\Omega)$. Notice that if we provide a uniform bound in $L^q(\partial\Omega)$, then we are done. In fact $W^{-\frac{1}{q},q}(\partial\Omega) := W^{\frac{1}{q},q'}(\partial\Omega)^*$ and $W^{\frac{1}{q},q'}(\partial\Omega) \hookrightarrow L^{q'}(\partial\Omega)$, so we directly infer $W^{-\frac{1}{q},q}(\partial\Omega) \hookleftarrow L^q(\partial\Omega)$.

Moreover, it is known that, with our assumptions on $\partial\Omega$, the normal trace of functions in $W^{s,p}(\Omega)$ lies in $L^p(\partial\Omega)$, provided $s > 1 + \frac{1}{p}$ (for this and some further sharper results, see [Marschall 1987, Theorem 2]). Hence,

$$\begin{aligned}\|\kappa(u_k)_n\|_{W^{-1/q,q}(\partial\Omega)} &\le C(q,\Omega)\|\kappa(u_k)_n\|_{L^q(\partial\Omega)} \le C(q,\Omega)\|\kappa\|_{L^\infty(\partial\Omega)}\|(u_k)_n\|_{L^q(\partial\Omega)}\\ &\le C(q,\Omega,s)\|\kappa\|_{L^\infty(\partial\Omega)}\|u_k\|_{W^{s,q}(\Omega)}\end{aligned} \tag{6-14}$$

for some $s > 1 + \frac{1}{q}$. Thus, we need to find an appropriate fractional Sobolev space in which $H^2(\Omega)$ is embedded. We claim that $H^2(\Omega) \hookrightarrow W^{1+\frac{3}{2q},q}(\Omega)$. Actually, it is enough to prove that $H^1(\Omega) := W^{1,2}(\Omega) \hookrightarrow W^{\frac{3}{2q},q}(\Omega)$ by the definition of $W^{s,p}(\Omega)$ for $s > 1$. So, let $u \in W^{1,2}(\Omega)$; by the Stein total extension theorem [Adams and Fournier 2003, Theorem 5.24] there exists $U \in W^{1,2}(\mathbb{R}^2)$ such that $U_{|\Omega} = u$ a.e. and $\|U\|_{W^{1,2}(\mathbb{R}^2)} \le C\|u\|_{W^{1,2}(\Omega)}$ for some positive constant independent of u. Applying the interpolation result Lemma 6.9 to U with $\theta = \frac{3}{2q}$ and the Sobolev embedding $W^{1,2}(\mathbb{R}^2) \hookrightarrow L^{4q-6}(\mathbb{R}^2)$

since $4q-6 \geq 2$,

$$\|U\|_{W^{3/2q,q}(\mathbb{R}^2)} \leq C\|U\|_{W^{1,2}(\mathbb{R}^2)}^{\frac{3}{2q}} \|U\|_{L^{4q-6}(\mathbb{R}^2)}^{1-\frac{3}{2q}} \leq C_1\|U\|_{W^{1,2}(\mathbb{R}^2)}.$$

Hence,

$$\|u\|_{W^{3/(2q),q}(\Omega)} = \|U\|_{W^{3/(2q),q}(\Omega)} \leq \|U\|_{W^{3/(2q),q}(\mathbb{R}^2)} \leq C_1\|U\|_{W^{1,2}(\mathbb{R}^2)} \leq C_2\|u\|_{W^{1,2}(\Omega)}.$$

As a result, noticing that $s = 1+\frac{3}{2q} > 1+\frac{1}{q}$, we can continue (6-14), obtaining

$$\|\kappa(u_k)_n\|_{W^{-1/q,q}(\partial\Omega)} \leq C(q,\Omega)\|\kappa\|_{L^\infty(\partial\Omega)}\|u_k\|_{W^{1+3/(2q),q}(\Omega)} \leq \widetilde{C}(q,\Omega)\|\kappa\|_{L^\infty(\partial\Omega)}\|u_k\|_{H^2(\Omega)},$$

which is uniformly bounded in k. Combining estimate (6-13) with the ones above for the second and the third terms of (6-13), we finally end up with the strong convergence in $W^{2,q}(\Omega)$. $\square$

Extending positivity, part 1: A convergence argument. Let us start by noticing that, by Morrey's embeddings, the convergence in $W^{2,q}(\Omega)$ for every $q \geq 2$ of Proposition 6.10 implies the convergence in $C^{1,\alpha}(\overline{\Omega})$ for every $\alpha < 1$, thus in particular in $C^1(\overline{\Omega})$. This will be the main ingredient in the next proof.

Proposition 6.11. *Let $\Omega \subset \mathbb{R}^2$ be a bounded convex domain of class C^2 and $(\sigma_k)_{k\in\mathbb{N}}$ be a sequence of parameters such that $\sigma_k \searrow 1$ and $(u_k)_{k\in\mathbb{N}}$ be a sequence of ground states for the functional J_{σ_k}. Then there exists a subsequence $(u_{k_j})_{j\in\mathbb{N}}$ and $j_0 \in \mathbb{N}$ such that $u_{k_j} > 0$ in Ω for every $j \geq j_0$.*

Proof. By Propositions 6.8 and 6.10 and by the previous observation, we know that, up to a subsequence, $u_k \to \bar{u}$ in $C^1(\overline{\Omega})$ for some $\bar{u}$, a ground state for J_{NAV}.

Since Ω has a C^2 boundary, the interior sphere condition holds and one can extend the outer normal vector n in a small neighborhood $\omega_0 \subset \Omega$ of $\partial\Omega$ and thus define here $\bar{u}_n := \nabla\bar{u}\cdot n$ (see [Sperb 1981, Chapter 4]). Moreover, since $\bar{u}$ is strictly superharmonic, the normal derivative $\bar{u}_n$ is negative on $\partial\Omega$ and, by compactness of $\partial\Omega$ and continuity of $\bar{u}_n$, there exists $\alpha > 0$ such that

$$\bar{u}_{n_{|\partial\Omega}} \leq -\alpha < 0.$$

Hence, again by continuity, there exists a second neighborhood $\omega \subset \omega_0$ of $\partial\Omega$ such that

$$\bar{u}_{n_{|\omega}} \leq -\tfrac{2}{3}\alpha < 0.$$

Take now $\varepsilon_1 = \frac{1}{3}\alpha$: by the $C^1(\overline{\Omega})$ convergence, there exists $k_1 \in \mathbb{N}$ such that for every $k \geq k_1$ and $x \in \omega$,

$$|(u_k)_n(x)| \geq |\bar{u}_n(x)| - |(u_k)_n(x) - \bar{u}_n(x)| > \tfrac{2}{3}\alpha - \big\||n|\big\|_{L^\infty(\omega)}\big\||\nabla u_k - \nabla\bar{u}|\big\|_{L^\infty(\Omega)} > \tfrac{2}{3}\alpha - \varepsilon_1 > \tfrac{1}{3}\alpha.$$

By the interior sphere condition, the map $\omega \to \partial\Omega,\ x \mapsto x_0$, such that $d(x,x_0) = \inf\{d(x,y) \mid y \in \partial\Omega\}$ is well defined and the vector $x - x_0$ has the same direction as $n(x)$ and $n(x_0)$. Hence by the Lagrange theorem and recalling that $u_{k_{|\partial\Omega}} = 0$, for $x \in \omega$,

$$|u_k(x)| = |u_k(x) - u_k(x_0)| \geq \min_{y\in[x_0,x]} |(u_k)_n(y)|\,|x - x_0| > \tfrac{1}{3}\alpha|x - x_0| > 0. \tag{6-15}$$

Moreover, notice that by compactness of $\Omega_0 := \Omega \setminus \omega$, the remaining part of Ω, we have

$$\bar{u}_{|\Omega_0} \geq \min_{\Omega_0} \bar{u} := m > 0$$

and so by the uniform convergence it is easy to deduce that, for k large enough, $u_k(x) > \frac{1}{2}m$ for every $x \in \Omega_0$. The result follows by combining this with (6-15). □

Theorem 6.12. *Let $\Omega \subset \mathbb{R}^2$ be a bounded convex domain of class C^2; then there exists $\sigma_1 > 1$ such that for every $\sigma \in (1, \sigma_1)$ the ground states of J_σ are positive in Ω.*

Proof. By contradiction, suppose that such σ_1 does not exist. Hence we would be able to find a sequence $(\sigma_k) \searrow 1$ such that for each of them there exists a ground state u_k for J_{σ_k} which is not positive. This would contradict Proposition 6.11. □

Remark 6.13. As we are dealing with continuous functions, since $H^2(\Omega) \hookrightarrow C^0(\bar{\Omega})$, we are interested in the strict positivity *everywhere in* Ω and not only a.e. in Ω. Theorem 6.12 gives a positive answer for this question: in fact, as $\bar{u} \in H^2(\Omega) = W^{2,N}(\Omega)$ is strictly superharmonic, by the strong maximum principle for strong solutions [Gilbarg and Trudinger 1998, Theorem 9.6], we deduce that it cannot achieve its minimum on the interior of Ω; thus $\bar{u}(x) > 0$ *for every* $x \in \Omega$. By the C^1 convergence we deduce the same strict inequality for u_σ, with $\sigma \in (1, \sigma_1)$.

Extending positivity, part 2: Moreau dual cones decomposition. Our aim is to investigate a further extension of the positivity result found in Theorem 6.12, possibly for the whole range $\sigma \in (1, +\infty)$. It seems natural if we think of the following fact: similarly to what we already obtained for the Navier problem, one can prove the convergence in $H^2(\Omega)$, as $\sigma \to +\infty$, of a sequence of ground states of J_σ to a least-energy solution of the Dirichlet problem

$$\begin{cases} \Delta^2 u = g(x)|u|^{p-1}u & \text{in } \Omega, \\ u = u_n = 0 & \text{on } \partial\Omega, \end{cases} \tag{6-16}$$

at least when κ is positive a.e. on $\partial\Omega$. Since we already know that in some cases the ground states of (6-16) are positive (for instance if Ω is a ball, see [Ferrero et al. 2007]), we expect to be able to completely extend positivity for such domains.

After a brief explanation of the convergence just mentioned above, we will apply Moreau's method of dual cones to infer the intervals of positivity for the semilinear problem. At the end, one may also compare the resulting analysis with the respective one for the linear problem with the same boundary conditions, due to Gazzola and Sweers [2008].

The Dirichlet problem. The argument is similar to what we used in the subsection on page 961 for the convergence to the Navier problem, but now we have to pay attention to the fact that in this case the two functional spaces are different ($H^2(\Omega) \cap H_0^1(\Omega)$ for the Steklov problem and $H_0^2(\Omega)$ for the Dirichlet). We are not giving here the details of the proof of the existence of ground states of (6-16), as it can be obtained as for the Steklov framework by the Nehari method of Section 3. In the following, we assume Ω to be a bounded convex domain in $\mathbb{R}^2$ with boundary of class $C^{1,1}$ and $\sigma > 1$. We suppose also that the curvature κ is positive a.e, that is $\partial\Omega$ has parts that are not flat. Moreover, as usual, u_k will always denote a ground state for J_{σ_k} and $\bar{u}$ a ground state for $J_{\text{DIR}} : H_0^2(\Omega) \to \mathbb{R}$ defined as

$$J_{\text{DIR}}(u) = \frac{1}{2}\int_\Omega (\Delta u)^2 - \frac{1}{p+1}\int_\Omega g(x)|u|^{p+1},$$

whose critical points are weak solutions of (6-16). Moreover, as in the Steklov case, we define the Nehari manifold for J_{DIR}:

$$\mathcal{N}_{\mathrm{DIR}} := \{u \in H_0^2(\Omega)\setminus\{0\} \mid J'_{\mathrm{DIR}}(u)[u] = 0\}.$$

First of all, notice that, by the definition of J_σ, for each σ,

$$J_\sigma|_{H_0^2(\Omega)} = J_{\mathrm{DIR}}, \tag{6-17}$$

so $\mathcal{N}_\sigma$ restricted to the subspace $H_0^2(\Omega)$ coincides with $\mathcal{N}_{\mathrm{DIR}}$.

Theorem 6.14. *Let $\sigma_k \to +\infty$ and Ω be a bounded convex domain in $\mathbb{R}^2$ with boundary of class $C^{1,1}$. Assume also that the curvature k is positive a.e on $\partial\Omega$. Then the sequence $(u_k)_{k\in\mathbb{N}}$ of ground states of $(J_{\sigma_k})_{k\in\mathbb{N}}$ admits a subsequence $(u_{k_j})_{j\in\mathbb{N}}$ convergent in $H^2(\Omega)$ to $\bar{u}$, which is a ground state for the Dirichlet problem (6-16).*

Proof. We follow the same steps as in the subsection on page 961 to deduce Theorem 6.5. Firstly, we prove that $(u_k)_{k\in\mathbb{N}}$ is bounded in $H^2(\Omega)$. Indeed, fix $\bar{w} \in H_0^2(\Omega)$, a ground state for the Dirichlet problem (6-16). Then

$$\begin{aligned}\|\Delta u_k\|_2^2 &\le \int_\Omega (\Delta u_k)^2 - (1-\sigma_k)\int_{\partial\Omega} \kappa (u_k)_n^2 = \int_\Omega g(x)|u_k|^{p+1} = \inf_{v\in\mathcal{N}_{\sigma_k}} \int_\Omega g(x)|v|^{p+1} \\ &\le \inf_{v\in\mathcal{N}_{\sigma_k}\cap H_0^2(\Omega)} \int_\Omega g(x)|v|^{p+1} = \int_\Omega g(x)|\bar{w}|^{p+1}.\end{aligned} \tag{6-18}$$

Hence, there exists $\bar{u} \in H^2(\Omega)\cap H_0^1(\Omega)$ such that, up to a subsequence, $u_k \rightharpoonup \bar{u}$ weakly in $H^2(\Omega)$. Moreover, (6-18) implies that

$$0 \le (\sigma_k - 1)\int_{\partial\Omega} \kappa (u_k)_n^2 \le \int_\Omega g(x)|u_k|^{p+1} \le C(\Omega, p)\|g\|_1 \|u_k\|_{H^2(\Omega)}^{p+1} \le D(\Omega, p, g)$$

and, taking into account that $\sigma_k \to +\infty$, we deduce that

$$\int_{\partial\Omega} \kappa (u_k)_n^2 \to 0.$$

Furthermore, by the compactness of the map $\partial_n : H^2(\Omega) \to L^2(\partial\Omega)$, we have also that

$$\int_{\partial\Omega} \kappa (u_k)_n^2 \to \int_{\partial\Omega} \kappa \bar{u}_n^2.$$

Hence, combining the two and recalling that we assumed $\kappa > 0$ on $\partial\Omega$, we deduce that $\bar{u}_n \equiv 0$ on $\partial\Omega$ and thus $\bar{u} \in H_0^2(\Omega)$.

Finally, testing the weak formulation of problem (6-1) with $\varphi \in H_0^2(\Omega)$ and passing to the limit as $k \to +\infty$, we deduce that

$$\int_\Omega \Delta\bar{u}\,\Delta\varphi = \int_\Omega g(x)|\bar{u}|^{p-1}\bar{u}\varphi,$$

so $\bar{u}$ is a solution of the Dirichlet problem (6-16) and, similarly to Lemma 6.1, we can prove that the convergence is strong in $H^2(\Omega)$. It remains to prove that $\bar{u}$ is actually a ground state for J_{DIR}. Let $\bar{w} \in H_0^2(\Omega)$ be a ground state solution of J_{DIR}. Then, by (6-17),

$$m = J_{\mathrm{DIR}}(\bar{w}) = J_{\sigma_k}(t^*_{\sigma_k}(\bar{w})\bar{w}) \geq \inf_{\mathcal{N}_{\sigma_k} \cap H_0^2(\Omega)} J_{\sigma_k} \geq \inf_{\mathcal{N}_{\sigma_k}} J_{\sigma_k} = J_{\sigma_k}(u_k);$$

hence we deduce that $m \geq \liminf_{k\to+\infty} J_{\sigma_k}(u_k)$. Moreover, by strong convergence,

$$J_{\mathrm{DIR}}(\bar{u}) = \left(\frac{1}{2} - \frac{1}{p+1}\right) \int_\Omega g(x)|\bar{u}|^{p+1} = \lim_{k\to+\infty} \left(\frac{1}{2} - \frac{1}{p+1}\right) \int_\Omega g(x)|u_k|^{p+1} = \lim_{k\to+\infty} J_{\sigma_k}(u_k).$$

Finally, since $\bar{u}$ is a solution of the Dirichlet problem (6-16), we have $\bar{u} \in \mathcal{N}_{\mathrm{DIR}}$, so

$$m \leq J_{\mathrm{DIR}}(\bar{u}) \leq \liminf_{k\to+\infty} J_{\sigma_k}(u_k) \leq m. \qquad \square$$

Moreau dual cones decomposition. So far, we have proved the existence of ground states for the Dirichlet problem (6-16) and the convergence result as $\sigma \to +\infty$. Proving positivity of ground states of (6-16) is quite a hard subject, since it strongly relies on the geometry of the domain, even in the linear case, where $f(x,u) = f(x)$: we refer to [Sweers 2001] for a short survey. Anyway, there are some cases in which it holds: for instance, the Dirichlet problem in the ball has been studied in [Ferrero et al. 2007], which covers the case where $g \equiv 1$, but whose arguments hold also in the general situation.

Our strategy is mainly inspired by this last work and it was firstly applied to fourth-order problems by Gazzola and Grunau [2001]. Briefly, we use Moreau decomposition in dual cones (for the original paper, see [Moreau 1962]) to obtain from a supposed sign-changing ground state solution u, a function w of one sign and in the same space with a strictly lower energy level, leading to a contradiction. In our case, in order to apply this machinery, we have to impose that the associated linear problem is positivity preserving: this will be the connection between the two problems.

Definition 6.15. Let $\Omega \subset \mathbb{R}^2$ be a bounded domain of class $C^{1,1}$ and fix $\sigma \in \mathbb{R}$. The linear Steklov boundary problem

$$\begin{cases} \Delta^2 u = f & \text{in } \Omega, \\ u = \Delta u - (1-\sigma)\kappa u_n = 0 & \text{on } \partial\Omega \end{cases} \tag{6-19}$$

is *positivity preserving* in Ω if there exists a unique solution $u \in H^2(\Omega) \cap H_0^1(\Omega)$ and $f \geq 0$ implies $u \geq 0$, and this holds for each $f \in L^2(\Omega)$. We shorten this by saying that "Ω is a [PPP$_\sigma$] domain for (6-19)".

Definition 6.16. Let H be a Hilbert space with scalar product $(\cdot,\cdot)_H$ and $K \subset H$ be a nonempty closed convex cone. Its *dual cone* K^* is defined as

$$K^* := \{w \in H \mid (w,v)_H \leq 0 \text{ for all } v \in K\}.$$

Theorem 6.17 (Moreau dual cone decomposition, [Gazzola et al. 2010, Theorem 3.4]). *Let H be a Hilbert space with scalar product $(\cdot,\cdot)_H$ and K and K^* as before. Then for every $u \in H$, there exists a unique couple $(u_1,u_2) \in K \times K^*$ such that $u = u_1 + u_2$ and $(u_1,u_2)_H = 0$.*

Our aim is to apply this result with $(H, \|\cdot\|_H) = (H^2(\Omega) \cap H_0^1(\Omega); \|\cdot\|_{H_\sigma})$, where $\|\cdot\|_{H_\sigma}$ is the norm (2-1), and $K := \{v \in H \mid v \geq 0\}$, the cone of nonnegative functions, looking for a decomposition of each element of the space in positive and negative "parts". Hence we need a characterization of the dual cone K^*:

Lemma 6.18. *If Ω is a [PPP$_\sigma$] domain for (6-19) for a fixed $\sigma \in \mathbb{R}$, then $K^* \subseteq \{w \in H \mid w < 0$ a.e.$\} \cup \{0\}$.*

Proof. We adapt here the proof of [Gazzola et al. 2010, Proposition 3.6]. Let $\varphi \in C_c^\infty(\Omega)$, $\varphi \geq 0$ and let $v_\varphi \in H^2(\Omega) \cap H_0^1(\Omega)$ be the unique weak solution of the linear problem

$$\begin{cases} \Delta^2 v_\varphi = \varphi & \text{in } \Omega, \\ v_\varphi = \Delta v_\varphi - (1-\sigma)\kappa (v_\varphi)_n = 0 & \text{on } \partial\Omega; \end{cases}$$

that is, for every test function $w \in H^2(\Omega) \cap H_0^1(\Omega)$, we have

$$(v_\varphi, w)_{H_\sigma} := \int_\Omega \Delta v_\varphi \, \Delta w - (1-\sigma) \int_{\partial\Omega} \kappa (v_\varphi)_n w_n = \int_\Omega \varphi w.$$

Hence, suppose $w = u \in K^*$: as Ω is a [PPP$_\sigma$] domain and $\varphi \geq 0$, we deduce that $v_\varphi \geq 0$, so $v_\varphi \in K$ and thus $(v_\varphi, u)_{H_\sigma} \leq 0$. As a result, we have obtained that for every $\varphi \in C_c^\infty(\Omega)$, $\varphi \geq 0$, we have $\int_\Omega \varphi u \leq 0$, which implies that $u \leq 0$ a.e. in Ω.

Moreover, let us suppose that the null set of u, namely $N := \{x \in \Omega \mid u(x) = 0\}$, has positive measure, consider $\psi := \chi_N \neq 0$ and let v_0 be the unique solution of the linear Navier problem

$$\begin{cases} \Delta^2 v_0 = \psi & \text{in } \Omega, \\ v_0 = \Delta v_0 = 0 & \text{on } \partial\Omega. \end{cases} \tag{6-20}$$

Then v_0 is strictly superharmonic by the maximum principle; thus $v_0 > 0$ and, by the Hopf Lemma, $(v_0)_n < 0$. As a result, for any function $v \in H^2(\Omega) \cap H_0^1(\Omega)$ one can produce two positive constants α, β such that $v + \alpha v_0 \geq 0$ and $v - \beta v_0 \leq 0$. Moreover we claim that $(u, v_0)_{H_\sigma} \geq 0$. In fact, as v_0 is the weak solution of (6-20) and by the definition of ψ,

$$\int_\Omega \Delta u \, \Delta v_0 = \int_\Omega u \psi = \int_N u = 0.$$

Thus, since $\sigma > 1$, $\kappa \geq 0$, $u_n \leq 0$ as $u \geq 0$, and $(v_0)_n < 0$,

$$(u, v_0)_{H_\sigma} := \int_\Omega \Delta u \, \Delta v_0 - (1-\sigma) \int_{\partial\Omega} \kappa u_n (v_0)_n \geq 0.$$

As a result, recalling that $u \in K^*$, $v + \alpha v_0 \in K$ and $v - \beta v_0 \in (-K)$, we have the chain of inequalities

$$0 \geq (u, v + \alpha v_0)_{H_\sigma} = (u, v)_{H_\sigma} + \alpha (u, v_0)_{H_\sigma} \geq (u, v)_{H_\sigma} \geq (u, v)_{H_\sigma} - \beta (u, v_0)_{H_\sigma} = (u, v - \beta v_0)_{H_\sigma} \geq 0,$$

which implies that $(u, v)_{H_\sigma} = 0$, and this holds for all $v \in H^2(\Omega) \cap H_0^1(\Omega)$. Hence this is true also for v defined as the unique solution of the Steklov problem

$$\begin{cases} \Delta^2 v = u & \text{in } \Omega, \\ v = \Delta v - (1-\sigma)\kappa v_n = 0 & \text{on } \partial\Omega, \end{cases} \tag{6-21}$$

and, using u as a test function, we deduce that

$$0 = (u, v)_{H_\sigma} = \int_\Omega u^2 = \|u\|_2^2,$$

which implies $u = 0$ a.e. $\square$

Proposition 6.19. *Let $\sigma > 1$ and suppose Ω is a $[PPP_\sigma]$ domain for (6-19). Then the ground states of J_σ are a.e. strictly of only one sign.*

Proof. Let $u \in H^2(\Omega) \cap H_0^1(\Omega)$ be such a ground state and suppose by contradiction that u is sign-changing. Denoting as before the cone of nonnegative functions by K, by Moreau decomposition there exists a unique couple $(u_1, u_2) \in K \times K^*$ such that $u = u_1 + u_2$ and $(u_1, u_2)_{H_\sigma} = 0$. Hence we know that $u_1 \geq 0$ and, by Lemma 6.18, $u_2 < 0$. Moreover, u is supposed to change sign, so $u_1 \neq 0$.

Defining $w := u_1 - u_2 \in H^2(\Omega) \cap H_0^1(\Omega)$, we have $w > |u|$. Indeed,

$$w = u_1 - u_2 > u_1 + u_2 = u, \quad w = u_1 - u_2 > -u_1 - u_2 = -u.$$

Consequently, $\int_\Omega g(x)|w|^{p+1} > \int_\Omega g(x)|u|^{p+1}$ and, since the decomposition is orthogonal under that norm, $\|w\|_{H_\sigma}^2 = \|u_1\|_{H_\sigma}^2 + \|u_2\|_{H_\sigma}^2 = \|u\|_{H_\sigma}^2$. Moreover, by Lemma 3.3, there exists $t^* := t^*(w) \in (0, +\infty)$ such that $w^* := t^*(w)w \in \mathcal{N}_\sigma$. Hence we deduce

$$\begin{aligned} J_\sigma(w^*) &= \frac{(t^*)^2}{2}\|w\|_{H_\sigma}^2 - \frac{(t^*)^{p+1}}{p+1}\int_\Omega g(x)|w|^{p+1} \\ &< \frac{(t^*)^2}{2}\|u\|_{H_\sigma}^2 - \frac{(t^*)^{p+1}}{p+1}\int_\Omega g(x)|u|^{p+1} = J_\sigma(t^*(w)u) \leq J_\sigma(u), \end{aligned}$$

since u is the maximum of J_σ on the half-line $\{tu \mid t \in (0, +\infty)\}$ by Lemma 3.3; thus we have a contradiction again, since u was the infimum of J_σ on the Nehari manifold $\mathcal{N}_\sigma$. Hence we infer that $u \geq 0$.

Finally, as u is a critical point of J_σ, we have for each a positive test function $\varphi \in H^2(\Omega) \cap H_0^1(\Omega)$,

$$(u, \varphi)_{H_\sigma} = \int_\Omega \Delta u\, \Delta\varphi - (1-\sigma)\int_{\partial\Omega} \kappa u_n \varphi_n = \int_\Omega g(x) u^p \varphi \geq 0,$$

which implies that $-u \in K^*$. Applying now Lemma 6.18, we get $-u < 0$, that is, $u > 0$. $\square$

As a consequence, the problem of proving positivity of ground state is led back to a problem of positivity preserving for the linear problem, which was already tackled and solved by Gazzola and Sweers [2008].

Theorem 6.20. *Let $\sigma > 1$ and $\Omega \subset \mathbb{R}^2$ be a bounded convex domain with $\partial\Omega$ of class C^2. There exists $\tilde{\delta}_c(\Omega) \in (1, +\infty]$ such that if $\sigma \in (1, \tilde{\delta}_c(\Omega))$, the ground states of the functional J_σ are a.e. strictly of only one sign.*

Proof. We follow the notation of [Gazzola and Sweers 2008]. Choosing $\beta = \kappa$ in Theorem 4.1(iii) of that paper, we infer the existence of $\delta_{c,\kappa}(\Omega) \in [-\infty, 0)$ such that if $(1-\sigma)\kappa \geq \delta_{c,\kappa}(\Omega)\kappa$, then the positivity preserving for problem (6-19) holds in Ω. Hence, defining $\tilde{\delta}_c(\Omega) := 1 + |\delta_{c,\kappa}(\Omega)|$, we can apply Proposition 6.19, provided $\sigma < \tilde{\delta}_c(\Omega)$. $\square$

Comparing Theorems 6.20 and 6.12, one may argue that we have nothing more than what we already knew: in both we obtain the existence of $\sigma_1 = \sigma_1(\Omega) > 1$ such that for all $\sigma \in (1, \sigma_1)$ the ground state solutions of problem (1-3) are positive. Nevertheless, in Theorem 6.20 we get further precise information about how the interval of positivity depends on the domain, relating it strongly with the positivity-preserving property. This fact is striking in the case of the disk and allows us to finally answer the question which opened the section.

Corollary 6.21. *Let $B \subset \mathbb{R}^2$ be a disk and let $\sigma > 1$. Then the ground states of the functional J_σ are a.e. strictly of only one sign.*

Proof. It is enough to notice that here $\kappa = 1$ and applying [Gazzola and Sweers 2008, Theorem 2.9] one can deduce $\delta_{c,\kappa}(B) = -\infty$, which implies $\tilde{\delta}_c(B) = +\infty$. □

One should finally notice that here the positivity found by the dual cones method is up to a subset of the domain with zero Lebesgue measure, so *almost everywhere* in Ω. This is the price we have to pay to extend the positivity beyond the parameter σ_1 found in Theorem 6.12 (see also Remark 6.13).

Remark 6.22. Again, up to some easy modifications in the proofs, both the convergence in Theorem 6.14 and the positivity result in Theorem 6.20 hold also in the sublinear case $p \in (0, 1)$.

7. Radial case

This section is devoted to some further investigations when the domain is a disk in $\mathbb{R}^2$ and the function g is radial, regarding existence, positivity and some qualitative properties of radially symmetric solutions. Moreover, we establish the counterpart of the convergence results of Sections 5 and 6, but for general radial positive solutions.

For simplicity, we focus on the problem

$$\begin{cases} \Delta^2 u = g(x)|u|^{p-1}u & \text{in } B, \\ u = \Delta u - (1-\sigma)u_n = 0 & \text{on } \partial B, \end{cases} \tag{7-1}$$

where $B := B_1(0) \subset \mathbb{R}^2$, $g = g(|x|)$ lies in $L^1(B)$ and it is strictly positive inside B. Moreover, we let $\sigma \in \mathbb{R}$ and $p \in (0, 1) \cup (1, +\infty)$ to cover both the sublinear and the superlinear case. Notice that the curvature does not appear in the mixed boundary condition since $\kappa(B) \equiv 1$.

Positive radially decreasing solutions and global bounds. First of all, by Proposition 5.3, our analysis concerns only the range $\sigma > -1$: in fact, if $\Omega = B$, one has $\sigma^* = -1$, since the first Steklov eigenvalue $\tilde{\delta}_1(B)$ is 2 (see [Berchio et al. 2006, Proposition 12]).

Retracing exactly the same steps of Sections 3 and 4, it is quite easy to obtain the existence of a positive radial solution. In fact, confining ourselves to the closed subspace of radial functions

$$H_{\mathrm{rad}}(B) := \{u \in H^2(B) \cap H_0^1(B) \mid u(x) = u(|x|) \text{ for all } x \in B\} = \mathrm{Fix}_{O(2)}(H^2(B) \cap H_0^1(B)),$$

we deduce the existence of a critical point of J_σ restricted to $H_{\mathrm{rad}}(B)$. Then it is enough to notice that J_σ is invariant under the action of $O(2)$ and to apply the principle of symmetric criticality due to Palais (see [Willem 1996, Theorem 1.28]), retrieving that these points are critical for J_σ also with respect to the whole space.

Finally, if we restrict to the interval $(-1,1]$, the positivity of such critical points is proved as in Propositions 4.9 and 4.12, realizing that the superharmonic function of a radially symmetric function is radial too (see (4-4)). On the other hand, if $\sigma > 1$, one can apply the dual cone decomposition to the Hilbert space $H_{\mathrm{rad}}(B)$ and argue as in Lemma 6.18 and Proposition 6.19, taking into account that B is a $[\mathrm{PPP}_\sigma]$ domain for every $\sigma > -1$. Summarizing, we have shown the following:

Proposition 7.1. *Let $p \in (0,1)\cup(1,+\infty)$, $g = g(|x|) \in L^1(B)$, $g > 0$. If $\sigma \le -1$, there is no positive nonnegative nontrivial solution for (7-1), while, if $\sigma > -1$, there exists at least a positive radial solution, which is strictly superharmonic whenever $\sigma \in (-1,1]$.*

Now, we want to prove some qualitative properties of radial positive solutions of (7-1). The first result concerns the radial behaviour, while the second the uniform boundedness in $L^\infty(B)$. Before proving these results, one should notice that such solutions are strong solutions, namely in $W^{4,q}(B)$, provided $g \in L^q(B)$ for some $q > 2$ and also classical assuming in addition that $g \in W^{1,q}(B)$ for some $q > 2$. This is a straightforward application of Lemma 6.7 combined with Morrey's embeddings.

Lemma 7.2. *Let $B := B_R(0)$ be the ball of radius R in $\mathbb{R}^2$ centered in 0, $q > 2$ and $\tilde{h} \in W^{2,q}(B)$ be radial. Defining $h : [0,R] \to \mathbb{R}$ to be its restriction to the radial variable, for all $t \in [0,R]$ the following equality holds*:

$$t h'(t) = \int_0^t s\,\Delta h(s)\,ds. \tag{7-2}$$

Proof. If h is of class C^2, it comes directly from integration by parts and from the radial representation of the laplacian as

$$\Delta\tilde{h}(x) = h''(|x|) + \frac{1}{|x|}h'(|x|).$$

Otherwise, let $(\tilde{f}_k)_{k\in\mathbb{N}} \subset C^\infty(\bar{B})$ be such that $\tilde{f}_k \to \tilde{h}$ in $W^{2,q}(B)$, so in $C^1(\bar{B})$. Since $\tilde{h}$ is radial, we claim that it is possible to choose each $\tilde{f}_k$ to be radial and we denote its restriction to the radial variable as f_k. If so, for every $k \in \mathbb{N}$, we have

$$t f_k'(t) = \int_0^t s\,\Delta f_k(s)\,ds.$$

As a result, as $k \to +\infty$,

$$\left|\int_0^t s(\Delta f_k(s) - \Delta h(s))\,ds\right| = \frac{1}{2\pi}\|\Delta\tilde{f}_k - \Delta\tilde{h}\|_{L^1(B_t(0))} \le C(q)\|\tilde{f}_k - \tilde{h}\|_{W^{2,q}(B)} \to 0.$$

The result is proved by the convergence in $C^1(\bar{B})$ and the uniqueness of the limit. Now we have to justify our previous claim. Since $\tilde{h} \in W^{2,q}(B)$, we have

$$\sum_{i,\alpha}\int_B \left|\frac{\partial^\alpha\tilde{h}}{\partial i^\alpha}(x,y)\right|^q dx\,dy < +\infty,$$

where $i \in \{x, y\}$ and α is a multi-index of length $0 \le |\alpha| \le 2$. Since each $(\partial^\alpha \tilde{h})/(\partial i^\alpha)$ is radial, this is equivalent to saying that $h \in W^{2,q}([0, R], r)$, that is, the weighted Sobolev space with weight r. Hence, by [Kufner 1985, Theorem 7.4] ($M = \{0\}$, $\varepsilon = 1$ in notation therein), there exists a sequence $(f_k)_{k\in\mathbb{N}} \subset C^\infty([0, R])$ such that $f_k \to h$ in $W^{2,q}([0, R], r)$, that is

$$\sum_{i,\alpha} \int_0^R r \left| \frac{\partial^\alpha h}{\partial i^\alpha}(r) - f_k(r) \right|^q dr \to 0.$$

Hence, defining $F_k(x) := f_k(|x|)$, each $F_k \in C^\infty(\bar{B})$ is radial and

$$\|\tilde{h} - F_k\|_{W^{2,q}(B)} = \sum_{i,\alpha} \int_B \left| \frac{\partial^\alpha \tilde{h}}{\partial i^\alpha}(x, y) - F_k(x, y) \right|^q dx\, dy = 2\pi \sum_{i,\alpha} \int_0^R r \left| \frac{\partial^\alpha h}{\partial i^\alpha}(r) - f_k(r) \right|^q dr \to 0,$$

and the claim is proved. □

Proposition 7.3 (radial decay). *Assume $g \in L^q(B)$ for some $q > 2$, g is radial and $g > 0$, and let $u \not\equiv 0$ be a nonnegative radial solution of (7-1) with $\sigma \in (-1, 1]$ and $p \in (0, 1) \cup (1, +\infty)$. Then u is strictly radially decreasing; thus $u > 0$ in B.*

Proof. By the assumption on g, we infer that u is a strong solution; thus $w := \Delta u \in W^{2,q}(B)$. Since $\Delta w = \Delta^2 u = g(|x|)u^p \ge 0$ in $[0, 1]$, applying Lemma 7.2, we have $w' > 0$ in $(0, 1]$. Hence Δu is strictly increasing in $(0, 1]$. Moreover, since u is nonnegative and $u(1) = 0$, we have $u'(1) \le 0$; hence, using the second boundary condition, $\Delta u(1) = (1-\sigma)u'(1) \le 0$. Since Δu is strictly increasing in $(0, 1]$, we deduce that $\Delta u < 0$ in $[0, 1)$, and finally, applying again Lemma 7.2, $u' < 0$ in $(0, 1]$. □

In the next result we find a uniform upper bound for positive radial solutions of (7-1), which may be seen as a superlinear counterpart of Proposition 4.11. We will make use of a blow up method which goes back to Gidas and Spruck [1981], and which was adapted to the polyharmonic case by Reichel and Weth [2009; 2010]. Briefly, our argument will be the following: supposing the existence of a sequence of positive radial solutions with diverging L^∞ norm, we rescale each of them in order to have another sequence of functions with the same L^∞ norm, satisfying the same equation in nested domains which tend to occupy the whole $\mathbb{R}^2$. Then we show that, up to a subsequence, it converges uniformly on compact subsets to a continuous nonnegative but nontrivial function. This turns out to be a solution of the same equation on $\mathbb{R}^2$, which is a contradiction with the following Liouville-type result by Wei and Xu, with $N = 2$ and $m = 2$:

Lemma 7.4 [Wei and Xu 1999, Theorem 1.4]. *Let $m \in \mathbb{N}$ and assume that $p > 1$ if $N \le 2m$ and $1 < p \le (N + 2m)/(N - 2m)$ if $N > 2m$. If u is a classical nonnegative solution of*

$$(-\Delta)^m u = u^p \quad \text{in } \mathbb{R}^N,$$

then $u \equiv 0$.

In our proof we will make also use of the following local regularity estimate, which is a particular case of a more general result by Reichel and Weth:

Lemma 7.5 [Reichel and Weth 2009, Corollary 6]. *Let $\Omega = B_R(0) \subset \mathbb{R}^N$, $m \in \mathbb{N}$, $h \in L^p(\Omega)$ for some $p \in (1, +\infty)$ and suppose $u \in W^{2m,p}(\Omega)$ satisfies*

$$(-\Delta)^m u = h \quad \text{in } \Omega.$$

Then there exists a constant $C = C(R, N, p, m)$ such that for any $\delta \in (0, 1)$,

$$\|u\|_{W^{2m,p}(B_{\delta R}(0))} \le \frac{C}{(1-\delta)^{2m}} \big(\|h\|_{L^p(B_R(0))} + \|u\|_{L^p(B_R(0))}\big).$$

Proposition 7.6. *Let $\sigma \in (-1, 1]$. Let $g \in L^q(B)$ for some $q > 2$, g be radial and $g > 0$. Suppose also that g is continuous in 0. Then, there exists $C > 0$ independent of σ such that $\|u\|_\infty \le C$ for every u radial positive solution of (7-1).*

Proof. By contradiction, suppose there exists a sequence $(v_k)_{k\in\mathbb{N}}$ of radial positive solutions such that $\|v_k\|_\infty \nearrow +\infty$. According to Proposition 7.3, each v_k is radially decreasing, so $v_k(0) = \|v_k\|_\infty \nearrow +\infty$. For each $k \ge 1$, define

$$u_k(x) = \lambda_k^{\frac{4}{p-1}} v_k(\lambda_k x),$$

where $\lambda_k \in \mathbb{R}^+$ are such that $\lambda_k^{\frac{4}{p-1}} = 1/v_k(0)$. With this choice, each u_k satisfies

$$\begin{cases} \Delta^2 u_k = g(|\lambda_k x|) u_k^p & \text{in } B_{\frac{1}{\lambda_k}}(0), \\ u_k = \Delta u_k - (1-\sigma)\lambda_k (u_k)_n = 0 & \text{on } \partial B_{\frac{1}{\lambda_k}}(0), \end{cases} \tag{7-3}$$

is in $W^{4,q}(B_{\frac{1}{\lambda_k}}(0))$, radially decreasing and

$$\|u_k\|_{L^\infty(B_{1/\lambda_k}(0))} = u_k(0) = \lambda_k^{\frac{4}{p-1}} v_k(0) = 1. \tag{7-4}$$

We claim that the sequence $(u_k)_{k\in\mathbb{N}}$ is uniformly bounded on compact sets of $\mathbb{R}^2$ in $W^{4,q}$ norm. In fact, let $K \subset \mathbb{R}^2$ be compact; then there exists $\rho > 0$ such that $B_\rho(0) \supset K$ and, for k large enough, each u_k is in K since $B_{\frac{1}{\lambda_k}}(0) \supset B_{2\rho}(0)$ definitively. For such k, by (7-4) and applying Lemma 7.5 with $\Omega = B_{2\rho}(0)$, $m = N = 2$ and $\delta = \frac{1}{2}$,

$$\begin{aligned} \|u_k\|_{W^{4,q}(K)} \le \|u_k\|_{W^{4,q}(B_\rho(0))} &\le \frac{C(\rho, q)}{1/2^4} (\|\Delta^2 u_k\|_{L^q(B_{2\rho}(0))} + \|u_k\|_{L^q(B_{2\rho}(0))}) \\ &\le 16 C(\rho, q)(\|g(|\lambda_k \cdot|)\|_{L^q(B_{2\rho}(0))} + |B_{2\rho}(0)|^{\frac{1}{q}}). \end{aligned} \tag{7-5}$$

Moreover, fixing $\varepsilon > 0$ and supposing k large enough,

$$\|g(|\lambda_k \cdot|)\|_{L^q(B_{2\rho}(0))} = (4\pi\rho^2)^{\frac{1}{q}} \left(\frac{1}{|B_{2\rho\lambda_k}(0)|} \int_{B_{2\rho\lambda_k}(0)} |g(y)|^q \, dy \right)^{\frac{1}{q}} \le (4\pi\rho^2)^{\frac{1}{q}} g(0) + \varepsilon, \tag{7-6}$$

where the last inequality follows from the Lebesgue differentiation theorem. Hence, combining (7-5) with (7-6), we infer $\|u_k\|_{W^{4,q}(K)} \le C(p, q, K, g)$, which is uniform on k. Incidentally, notice that this constant does not depend on σ. Hence we find $u \in W^{4,q}(K)$ such that, up to subsequences, $u_k \to u$ in $C^3(K)$, where $u \in C^3(\mathbb{R}^2)$, $u \geq 0$ and $u(0) = 1$ by (7-4) and satisfying

$$\Delta^2 u = g(0)u^p \quad \text{in } \mathbb{R}^2.$$

So, by a bootstrap method, we deduce that u is also a classical solution. Finally, setting for all $x \in \mathbb{R}^2$ $w(x) := u(bx)$ with $b := g(0)^{-\frac{1}{4}}$, one has w is a nonnegative solution of

$$\Delta^2 w = w^p \quad \text{in } \mathbb{R}^2,$$

with $w(0) = u(0) = 1$, which contradicts Lemma 7.4. □

Convergence results. We want to investigate what happens at the endpoints of the interval $(-1, 1]$ in which σ lies, by means of the last results. More precisely, our aim is to examine if any result similar to Theorems 5.8 and 6.5 can be found assuming $(u_k)_{k\in\mathbb{N}}$ to be a sequence of positive radial solutions of (7-1) with $\sigma = \sigma_k$ but without imposing any "minimizing" requirement. Unless otherwise stated, we assume $g \equiv 1$ and $p > 1$.

Let us start with the behaviour for $\sigma \to 1$, where the main ideas are taken from the same result for ground states. Notice that we know everything for the Navier problem in the ball: in fact, Dalmasso [1995] proved that there exists a unique positive solution, which is radially symmetric and radially decreasing thanks to a result by Troy [1981].

Proposition 7.7. *Let $(u_k)_{k\in\mathbb{N}}$ be a sequence of positive radial solutions of (7-1) with $\sigma_k \nearrow 1$. Then $u_k \to \bar{u}$ in $H^2(B)$, where $\bar{u}$ is the unique positive solution of the Navier problem.*

Proof. We firstly claim that such a sequence is bounded in $H^2(B)$. Indeed, by Proposition 7.6,

$$\|u_k\|^2_{H^2(B)} \leq C_0\|\Delta u_k\|_2^2 \leq C_0\left(1 - \frac{1-\sigma_k}{2}\right)^{-1}\|u_k\|^2_{H_{\sigma_k}} = \frac{2C_0}{1+\sigma_k}\|u_k\|^{p+1}_{p+1} \leq 2\pi C_0 C^{p+1}.$$

Hence, we can extract a subsequence $(u_{k_j})_{j\in\mathbb{N}}$ such that there exists $\bar{v} \in H^2(B) \cap H^1_0(B)$ such that $u_{k_j} \rightharpoonup \bar{v}$ weakly in $H^2(B)$. By Lemma 6.1, together with Remark 6.2, one can infer that this subsequence is actually strongly convergent in $H^2(B)$ and then that $\bar{v}$ is a weak solution of the Navier problem (thus classical by regularity theory). Moreover, since the convergence is pointwise, we immediately deduce that $\bar{v}$ is nonnegative, radially symmetric and radially nonincreasing. Nevertheless, by Proposition 7.3, $\bar{v}$ is actually strictly decreasing and positive in B, so it coincides with the unique positive solution $\bar{u}$ of the Navier problem. By the uniqueness of the limit and applying Urysohn subsequence principle, we retrieve the convergence of the whole sequence $(u_k)_{k\in\mathbb{N}}$ from which we started. □

Let us now investigate the case $\sigma \to -1$. As already noticed in Lemma 5.9, it is enough to understand the behaviour of the $L^{p+1}(B)$ norm of a sequence of solutions to infer the convergence in $H^2(B)$ norm. Since the proof of Theorem 5.8 strongly relies on the fact that it deals with ground states, we need a different technique. The first step is a Pohozaev-type identity by Mitidieri [1993]: it will allow us to prove an inequality involving $L^p(B)$ and $L^{p+1}(B)$ norms which, combined with the uniform bound of Proposition 7.6, will lead us to the convergence result.

Lemma 7.8 [Mitidieri 1993, Proposition 2.2]. *Let Ω be a smooth domain and $u \in C^4(\overline{\Omega})$. The following identity holds*:

$$\int_\Omega (\Delta^2 u)x\cdot\nabla u - \frac{N}{2}\int_\Omega (\Delta u)^2 - (N-2)\int_\Omega \nabla\Delta u\cdot\nabla u$$
$$= -\frac{1}{2}\int_{\partial\Omega}(\Delta u)^2 x\cdot n + \int_{\partial\Omega}\big((\Delta u)_n(x\cdot\nabla u) + u_n(x\cdot\nabla\Delta u) - \nabla\Delta u\cdot\nabla u(x\cdot n)\big).$$

Corollary 7.9. *Suppose u is a positive solution for problem (7-1) with $g\equiv 1$. Then the following identity holds*:

$$\int_{\partial B_R}\Big((\Delta u)_n + (1-\sigma)\Big(1-\frac{1-\sigma}{2}\Big)u_n\Big)u_n = -\Big(1+\frac{2}{p+1}\Big)\int_{B_R} u^{p+1}. \tag{7-7}$$

Proof. By similar computations as in the proof contained in [Berchio et al. 2007, Section 6], from Lemma 7.8 one infers

$$\Big(\frac{N-4}{2} - \frac{N}{p+1}\Big)\int_\Omega u^{p+1} = \int_{\partial\Omega}\big(x\cdot\nabla\Delta u + \tfrac{1}{2}N(1-\sigma)\kappa u_n - \tfrac{1}{2}(1-\sigma)^2\kappa^2 u_n(x\cdot n)\big)u_n. \tag{7-8}$$

If $N=2$ and $\Omega = B$, we have $x=n$ and $\kappa=1$, so $x\cdot\nabla\Delta u = (\Delta u)_n$ and (7-7) follows. □

The next result follows from some ideas of Berchio and Gazzola: we give here a sketch, while we refer to [Berchio and Gazzola 2011, Proposition 4], for a more detailed proof.

Lemma 7.10. *Let $\sigma\in(-1,1)$ and u be a positive radial solution of problem (7-1) with $g\equiv 1$. Then the following estimate holds*:

$$\|u\|_{p+1}^{p+1} \ge \tfrac{3}{64}\big(1-\tfrac{3}{64}(1-\sigma)\big)\frac{1}{\pi(1+\sigma)}\frac{p+1}{p+3}\|\Delta^2 u\|_1^2. \tag{7-9}$$

Proof. By radial symmetry, (7-7) reduces to

$$2(\Delta u)'(1)u'(1) + (1-\sigma)(1+\sigma)(u'(1))^2 = -\frac{p+3}{p+1}\frac{1}{\pi}\int_B u^{p+1}. \tag{7-10}$$

Moreover, by the divergence theorem we have

$$u'(1) = \frac{1}{2\pi}\int_B \Delta u \quad\text{and}\quad (\Delta u)'(1) = \frac{1}{2\pi}\int_B \Delta^2 u,$$

so, taking the first Steklov eigenfunction $w(x) = \frac{1}{4}(1-|x|^2)$ and after some elementary computations, one gets

$$\Big(\int_B \Delta^2 u - (1-\sigma)\int_B w\,\Delta^2 u\Big)\int_B w\,\Delta^2 u = \frac{p+3}{p+1}(1+\sigma)\pi\int_B u^{p+1}. \tag{7-11}$$

Noticing that $0\le w\le \frac{1}{4}$, we have

$$\frac{3}{64}\int_B \Delta^2 u \le \int_B w\,\Delta^2 u \le \frac{1}{4}\int_B \Delta^2 u.$$

Hence, defining now $d := (1-\sigma)$, $s := \int_B w\,\Delta^2 u$ and $A := \int_B \Delta^2 u$, the left-hand side of (7-11) becomes

$$As - ds^2, \quad \text{with } s\in\big[\tfrac{3}{64}A; \tfrac{1}{4}A\big].$$

Since $d > 0$, we know $\psi : s \mapsto As - ds^2$ is a concave function, so it attains its minimum on the extremal values of the interval: in this case, with $0 < d < 2$, one has

$$\psi(s) \geq \tfrac{3}{64}\big(1 - \tfrac{3}{64}d\big)A^2.$$

Combining this with (7-11), one finds the desired estimate (7-9). □

Theorem 7.11. *Let $\sigma_k \searrow -1$ and $(u_k)_{k\in\mathbb{N}}$ be a sequence of positive radial functions, each of them a solution of the problem (7-1) with $g \equiv 1$ and $\sigma = \sigma_k$. Then, $u_k \to 0$ in $H^2(B)$.*

Proof. By Lemma 5.9, it is enough to prove the convergence in $L^{p+1}(B)$ norm. Since every solution of (7-1) is smooth, we have $\|\Delta^2 u_k\|_1 = \|u_k\|_p^p$. Moreover, by the uniform L^∞ estimate found in Proposition 7.6, we know that there exists a constant $C > 0$ not depending on σ_k, such that

$$\|u_k\|_{p+1}^{p+1} \leq \|u_k\|_\infty^{p+1}|B| \leq \pi C^{p+1}.$$

As a result, using the estimate provided by Lemma 7.10, one has

$$\frac{1+\sigma_k}{1-\frac{3}{64}(1-\sigma_k)} \geq \frac{p+1}{p+3}\,\frac{3}{64\pi^2 C^{p+1}}\|u_k\|_p^{2p}, \tag{7-12}$$

so, letting $\sigma_k \to -1$ we deduce $\|u_k\|_p \to 0$. This, together with the $L^\infty(B)$ estimate of Proposition 7.6, gives us the convergence in $L^{p+1}(B)$ and so the desired result. □

8. Open problems

We end our paper with some unsolved questions that would complete the present investigation.

- *If Ω is a ball, are the ground states of J_σ radially symmetric?*

In fact, we deduced the existence of ground states and radial solutions which are indeed ground states among all possible radial solutions; both of them are positive and have the same behaviour when $\sigma \to -1$ and $\sigma \to 1$. But no standard techniques such as the Talenti symmetrization principle seem to apply (except for the Navier case) to prove that these classes of functions are indeed the same.

- *Are the radial positive solutions radially decreasing if $\sigma > 1$?*

Indeed, the radial decay property proved in Proposition 7.3 does not apply in this setting and, by now, we cannot extend Proposition 7.6 for these values of σ.

Moreover, in the spirit of [Dalmasso 1995] and [Ferrero et al. 2007]:

- *Can we say something about the uniqueness of (at least) the positive radially symmetric ground state of J_σ for some values of σ?*

Finally, all the techniques developed in Section 3 strongly relied on the assumptions we made on the boundary, that is, $\partial\Omega$ of class $C^{1,1}$, in order to have $\kappa \in L^\infty(\partial\Omega)$. In particular, Theorem 4.7 allowed us to rewrite in an appropriate way our functional. Also the convexity played a crucial role in proving the positivity: see in particular Propositions 4.9, 4.12 and 5.6 as well as Theorem 6.20.

- *May we deduce the positivity of ground states of J_σ when the domain Ω is not convex anymore or with less regularity on the boundary*?

Since in the Navier case their positivity is always assured simply by the maximum principle, we expect that, even without the convexity assumption, it continues to hold whenever σ belongs to a neighborhood of 1 which may depend on "how far" the domain is from being convex.

Concerning the regularity of the boundary, if we consider the particular case of a convex polygon $\mathcal{P}$, it is known that ground states of J_σ are positive for every σ: in fact, the superharmonic method applies easily once we have $\int_{\mathcal{P}} \det(\nabla^2 u) = 0$ thanks to a result by Grisvard [1992, Lemma 2.2.2]. We believe that positivity for ground states of J_σ still holds imposing, for instance, only Lipschitz regularity for $\partial\Omega$.

Acknowledgements

The author wishes to express his gratitude to Enea Parini for his active interest in the publication of this paper and to François Hamel for many valuable and stimulating conversations. The author wants also to thank the anonymous referee for the careful reading of the manuscript and helpful comments and suggestions.

References

[Adams and Fournier 2003] R. A. Adams and J. J. F. Fournier, *Sobolev spaces*, 2nd ed., Pure and Applied Mathematics (Amsterdam) **140**, Elsevier/Academic Press, Amsterdam, 2003. MR Zbl ↑946, ↑954, ↑966

[Adolfsson 1992] V. Adolfsson, "L^2-integrability of second-order derivatives for Poisson's equation in nonsmooth domains", *Math. Scand.* **70**:1 (1992), 146–160. MR Zbl ↑946, ↑955

[Agmon et al. 1959] S. Agmon, A. Douglis, and L. Nirenberg, "Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions, I", *Comm. Pure Appl. Math.* **12** (1959), 623–727. MR Zbl ↑965

[Ambrosetti and Malchiodi 2007] A. Ambrosetti and A. Malchiodi, *Nonlinear analysis and semilinear elliptic problems*, Cambridge Studies in Advanced Mathematics **104**, Cambridge University Press, 2007. MR Zbl ↑951

[Berchio and Gazzola 2011] E. Berchio and F. Gazzola, "Positive solutions to a linearly perturbed critical growth biharmonic problem", *Discrete Contin. Dyn. Syst. Ser. S* **4**:4 (2011), 809–823. MR Zbl ↑944, ↑978

[Berchio et al. 2006] E. Berchio, F. Gazzola, and E. Mitidieri, "Positivity preserving property for a class of biharmonic elliptic problems", *J. Differential Equations* **229**:1 (2006), 1–23. MR Zbl ↑944, ↑947, ↑955, ↑957, ↑958, ↑973

[Berchio et al. 2007] E. Berchio, F. Gazzola, and T. Weth, "Critical growth biharmonic elliptic problems under Steklov-type boundary conditions", *Adv. Differential Equations* **12**:4 (2007), 381–406. MR Zbl ↑944, ↑959, ↑978

[Brezis and Mironescu 2001] H. Brezis and P. Mironescu, "Gagliardo–Nirenberg, composition and products in fractional Sobolev spaces", *J. Evol. Equ.* **1**:4 (2001), 387–404. MR Zbl ↑966

[Bucur et al. 2009] D. Bucur, A. Ferrero, and F. Gazzola, "On the first eigenvalue of a fourth order Steklov problem", *Calc. Var. Partial Differential Equations* **35**:1 (2009), 103–131. MR Zbl ↑944, ↑957

[Castro et al. 1997] A. Castro, J. Cossio, and J. M. Neuberger, "A sign-changing solution for a superlinear Dirichlet problem", *Rocky Mountain J. Math.* **27**:4 (1997), 1041–1053. MR Zbl ↑948

[Dalmasso 1995] R. Dalmasso, "Uniqueness theorems for some fourth-order elliptic equations", *Proc. Amer. Math. Soc.* **123**:4 (1995), 1177–1183. MR Zbl ↑977, ↑979

[Evans 2010] L. C. Evans, *Partial differential equations*, 2nd ed., Graduate Studies in Mathematics **19**, American Mathematical Society, Providence, RI, 2010. MR Zbl ↑952, ↑953

[Ferrero et al. 2005] A. Ferrero, F. Gazzola, and T. Weth, "On a fourth order Steklov eigenvalue problem", *Analysis* (*Munich*) **25**:4 (2005), 315–332. MR Zbl ↑944

[Ferrero et al. 2007] A. Ferrero, F. Gazzola, and T. Weth, "Positivity, symmetry and uniqueness for minimizers of second-order Sobolev inequalities", *Ann. Mat. Pura Appl.* (4) **186**:4 (2007), 565–578. MR Zbl ↑968, ↑970, ↑979

[Gasiński and Papageorgiou 2006] L. Gasiński and N. S. Papageorgiou, *Nonlinear analysis*, Series in Mathematical Analysis and Applications **9**, Chapman & Hall/CRC, Boca Raton, FL, 2006. MR Zbl ↑950

[Gazzola and Grunau 2001] F. Gazzola and H.-C. Grunau, "Critical dimensions and higher order Sobolev inequalities with remainder terms", *NoDEA Nonlinear Differential Equations Appl.* **8**:1 (2001), 35–44. MR Zbl ↑970

[Gazzola and Sweers 2008] F. Gazzola and G. Sweers, "On positivity for the biharmonic operator under Steklov boundary conditions", *Arch. Ration. Mech. Anal.* **188**:3 (2008), 399–427. MR Zbl ↑944, ↑945, ↑954, ↑955, ↑957, ↑968, ↑972, ↑973

[Gazzola et al. 2010] F. Gazzola, H.-C. Grunau, and G. Sweers, *Polyharmonic boundary value problems: positivity preserving and nonlinear higher order elliptic equations in bounded domains*, Lecture Notes in Mathematics **1991**, Springer, 2010. MR Zbl ↑965, ↑970, ↑971

[Gidas and Spruck 1981] B. Gidas and J. Spruck, "A priori bounds for positive solutions of nonlinear elliptic equations", *Comm. Partial Differential Equations* **6**:8 (1981), 883–901. MR Zbl ↑975

[Gilbarg and Trudinger 1998] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, revised 2nd ed., Grundlehren der Mathematischen Wissenschaften **224**, Springer, 1998. MR Zbl ↑953, ↑955, ↑968

[Grisvard 1992] P. Grisvard, *Singularities in boundary value problems*, Recherches en Mathématiques Appliquées **22**, Springer, 1992. MR Zbl ↑980

[Grumiau and Parini 2008] C. Grumiau and E. Parini, "On the asymptotics of solutions of the Lane–Emden problem for the p-Laplacian", *Arch. Math.* (*Basel*) **91**:4 (2008), 354–365. MR Zbl ↑948

[Kufner 1985] A. Kufner, *Weighted Sobolev spaces*, John Wiley & Sons, New York, 1985. MR Zbl ↑975

[Marschall 1987] J. Marschall, "The trace of Sobolev–Slobodeckij spaces on Lipschitz domains", *Manuscripta Math.* **58**:1–2 (1987), 47–65. MR Zbl ↑966

[Mitidieri 1993] E. Mitidieri, "A Rellich type identity and applications", *Comm. Partial Differential Equations* **18**:1–2 (1993), 125–151. MR Zbl ↑977, ↑978

[Moreau 1962] J.-J. Moreau, "Décomposition orthogonale d'un espace hilbertien selon deux cônes mutuellement polaires", *C. R. Acad. Sci. Paris* **255** (1962), 238–240. MR Zbl ↑970

[Nazarov et al. 2012] S. A. Nazarov, A. Stylianou, and G. Sweers, "Hinged and supported plates with corners", *Z. Angew. Math. Phys.* **63**:5 (2012), 929–960. MR Zbl ↑946, ↑955

[Parini and Stylianou 2009] E. Parini and A. Stylianou, "On the positivity preserving property of hinged plates", *SIAM J. Math. Anal.* **41**:5 (2009), 2031–2037. MR Zbl ↑943, ↑947, ↑948, ↑951, ↑952, ↑956, ↑958

[Reichel and Weth 2009] W. Reichel and T. Weth, "A priori bounds and a Liouville theorem on a half-space for higher-order elliptic Dirichlet problems", *Math. Z.* **261**:4 (2009), 805–827. MR Zbl ↑975, ↑976

[Reichel and Weth 2010] W. Reichel and T. Weth, "Existence of solutions to nonlinear, subcritical higher order elliptic Dirichlet problems", *J. Differential Equations* **248**:7 (2010), 1866–1878. MR Zbl ↑975

[Sperb 1981] R. P. Sperb, *Maximum principles and their applications*, Mathematics in Science and Engineering **157**, Academic Press, New York, 1981. MR Zbl ↑953, ↑967

[Stekloff 1902] W. Stekloff, "Sur les problèmes fondamentaux de la physique mathématique", *Ann. Sci. École Norm. Sup.* (3) **19** (1902), 191–259. MR Zbl ↑944

[Stylianou 2010] A. Stylianou, *Comparison and sign preserving properties of bilaplace boundary value problems in domains with corners*, Ph.D. thesis, Universität zu Köln, 2010, available at http://www.mi.uni-koeln.de/~astylian/StylianouDissertation.pdf. Zbl ↑953

[Sweers 2001] G. Sweers, "When is the first eigenfunction for the clamped plate equation of fixed sign?", pp. 285–296 in *Proceedings of the USA-Chile Workshop on Nonlinear Analysis* (Viña del Mar-Valparaiso, Chile, 2000), edited by R. Manasevich and P. Rabinowitz, Electron. J. Differ. Equ. Conf. **6**, Southwest Texas State Univ., San Marcos, TX, 2001. MR Zbl ↑970

[Troy 1981] W. C. Troy, "Symmetry properties in systems of semilinear elliptic equations", *J. Differential Equations* **42**:3 (1981), 400–413. MR Zbl ↑977

[Ventsel and Krauthammer 2001] E. Ventsel and T. Krauthammer, *Thin plates and shells: theory: analysis, and applications*, Dekker, New York, 2001. ↑943

[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 ↑975

[Willem 1996] M. Willem, *Minimax theorems*, Progress in Nonlinear Differential Equations and their Applications **24**, Birkhäuser, Boston, 1996. MR Zbl ↑974

Received 29 Jun 2016. Revised 6 Feb 2017. Accepted 7 Mar 2017.

GIULIO ROMANI: giulio.romani@univ-amu.fr
Aix Marseille Univ, CNRS, Centrale Marseille, Institute de Mathématique de Marseille (I2M), 13453 Marseille, France

ANALYSIS AND PDE
Vol. 10, No. 4, 2017
dx.doi.org/10.2140/apde.2017.10.983

GEOMETRIC CONTROL CONDITION FOR THE WAVE EQUATION WITH A TIME-DEPENDENT OBSERVATION DOMAIN

JÉRÔME LE ROUSSEAU, GILLES LEBEAU, PEPPINO TERPOLILLI AND EMMANUEL TRÉLAT

We characterize the observability property (and, by duality, the controllability and the stabilization) of the wave equation on a Riemannian manifold Ω, with or without boundary, where the observation (or control) domain is time-varying. We provide a condition ensuring observability, in terms of propagating bicharacteristics. This condition extends the well-known geometric control condition established for fixed observation domains.

As one of the consequences, we prove that it is always possible to find a time-dependent observation domain of arbitrarily small measure for which the observability property holds. From a practical point of view, this means that it is possible to reconstruct the solutions of the wave equation with only few sensors (in the Lebesgue measure sense), at the price of moving the sensors in the domain in an adequate way.

We provide several illustrating examples, in which the observation domain is the rigid displacement in Ω of a fixed domain, with speed v, showing that the observability property depends both on v and on the wave speed. Despite the apparent simplicity of some of our examples, the observability property can depend on nontrivial arithmetic considerations.

1. Introduction and main result

1A. ***Framework.*** Studies of the stabilization and the controllability for the wave equation go back to the works of D. L. Russell [1971a; 1971b]. The work of J.-L. Lions [1988a] was very important in the formalization of many controllability questions. In the case of a manifold without boundary Ω, the pioneering work of J. Rauch and M. Taylor [1974] related the question of fast stabilization, that is, exhibiting an exponential decay of the energy, to a geometric condition connecting the damping

G. Lebeau acknowledges the support of the European Research Council, ERC-2012-ADG, project number 320845: Semi Classical Analysis of Partial Differential Equations. E. Trélat acknowledges the support by FA9550-14-1-0214 of the EOARD-AFOSR.
MSC2010: 35L05, 93B07, 93C20.
Keywords: wave equation, geometric control condition, time-dependent observation domain.

region $\omega \subset \Omega$ and the rays of geometrical optics, resulting in the now celebrated geometric control condition (in short, GCC). The damped wave equation takes the form

$$\partial_t^2 u - \Delta u + \chi_\omega \, \partial_t u = 0.$$

Using that the energy of the solution to a hyperbolic equation is largely carried along the rays, if one assumes that any ray will have reached the region ω where the operator is dissipative in a finite time, one can prove that the energy decays exponentially in time, with an additional unique continuation argument that allows one to handle the low-frequency part of the energy. The work [Rauch and Taylor 1974] only treated the case of a manifold Ω without boundary, leaving open the case of manifolds with boundary until the work of C. Bardos, G. Lebeau, and J. Rauch [Bardos et al. 1992]. The understanding of the propagation of singularities in the presence of the boundary $\partial\Omega$, after the seminal work of R. Melrose and J. Sjöstrand [1978; 1982], was a key element in the proof of [Bardos et al. 1992], providing a generalized notion of rays, taking reflections at the boundary into account as well as glancing and gliding phenomena. The geometric condition for ω, now an open subset of $\overline{\Omega}$, is then the requirement that every generalized ray should meet the damping region ω in a finite time. The resulting stabilization estimate then takes the form

$$E_0(u(t)) \leqslant C e^{-C't} E_0(u(0)),$$

where E_0 is the energy

$$E_0(u(t)) = \|u(t)\|^2_{H^1(\Omega)} + \|\partial_t u(t)\|^2_{L^2(\Omega)}.$$

Note that, if an open set ω does not fulfill the geometric control condition, then only a logarithmic type of energy decay can be achieved in general [Lebeau 1996; Lebeau and Robbiano 1995; Burq 1998].[1]

The question of exact controllability relies on the same line of arguments as for the exponential stabilization. By exact controllability in time $T > 0$, for the control wave equation

$$\partial_t^2 u - \Delta u = \chi_\omega(x) f,$$

one means, given an arbitrary initial state (u_0, u_1) and an arbitrary final state (u_0^F, u_1^F), the ability to find f such that $(u_{t=T}, \partial_t u_{t=T}) = (u_0^F, u_1^F)$ starting from $(u_{t=0}, \partial_t u_{t=0}) = (u_0, u_1)$. If the energy level is $(u(t), \partial_t u(t)) \in H^1(\Omega) \oplus L^2(\Omega)$, it is natural to seek $f \in L^2((0, T) \times \Omega)$. Boundary conditions can be of Dirichlet or Neumann types.

In fact, as is well known, both exponential stabilization and exact controllability of the wave equation in a domain Ω, with a damping or a control only acting in an open region ω of $\overline{\Omega}$, are equivalent to an observability estimate for a free wave. For such a wave, the energy is constant with respect to time. The observability inequality takes the following form: for some constant $C > 0$ and some $T > 0$, we have

$$E_0(u) \leqslant C \int_0^T \|\partial_t u\|^2_{L^2(\omega)} \, dt. \tag{1}$$

For the issue of exact controllability, the time $T > 0$ in this inequality is then the control time (horizon). If the open set ω fulfills the geometric control condition, then the results of [Rauch and Taylor 1974;

[1] In fact, intermediate decay rates have been established in particular geometrical settings; see for instance [Schenck 2011] for almost exponential decay, and [Burq and Hitrik 2007; Phung 2007; Anantharaman and Léautaud 2014] for polynomial decay.

Bardos et al. 1992] show that the infimum of all possible such times T coincides with the infimum of all possible times in the geometric control condition. Note however that there are cases in which the geometric control condition does not hold, and yet the observability inequality (1) is valid: the case Ω is a sphere and ω is a half-sphere is a typical example.

A glance at inequality (1) shows that observability is in fact to be understood as occurring in a space-time domain, here $(0, T) \times \omega$. It is then natural to wonder if observability can hold if it is replaced by some other open subset of $(0, T) \times \overline{\Omega}$. This is the subject of the present article.

The motivation for such a study can be seen as fairly theoretical. However, in practical issues, in different industrial contexts, for nondestructive testing, safety applications, as well as tomography techniques used for imaging bodies (human or not), this question becomes quite relevant. In fact, the industrial framework of seismic exploration was the original motivation for this work. In the different fields we mentioned, data are collected to be exploited in an interpretation step which involves the solution of some inverse problem. The point is that the device used to collect data does not fit well with the usual geometric condition which is crucial to obtain an observability result. In some cases it appears of great interest to be able to tackle situations where the observation set is time-dependent. In others, the reduction of data volume may be sought, while preserving the data quality. One may also face a situation in which all sensors cannot be active at the same time.

The example of seismic data acquisition can help the reader get a grasp on the industrial need to better design data acquisition procedures. In the case of a towed marine seismic data acquisition campaign, a typical setup consists in six parallel streamers with length 6000 m, separated by a distance of 100 m, floating at a depth of 8 m. The basic receiving elements are pressure-sensitive hydrophones composed of piezoelectric ceramic crystal devices that are placed some 20 to 50 m apart along each streamer. A source (a carefully designed air gun array) is shot every 25 m while the boat moves. The seismic data experiment lasts around 8 s. One understands with this description that a huge amount (terabytes) of data is recorded during one such acquisition campaign above an area of interest beneath the sea floor. Of course, the velocity of the ship and of the streamers is very small as compared to that of the seismic waves (1500 m/s in water and up to 5000 m/s for examples in salt bodies that are typical in the North Sea or in the Gulf of Mexico). Yet, however small it may be, one can question its impact on the quality of the data. One could also want not to use all receivers at a single time but rather to design a dynamic (software-based) array of receivers during the time of the seismic experiment. The reader will of course realize that the mathematical results we present here are very far from solving this problem. They however give some leads on what important theoretical issues can be.

An inspection of the proof of [Bardos et al. 1992] shows that it uses the invariance of the observation cylinder $(0, T) \times \omega$ with respect to time in a crucial way. Hence, the method, if not modified, cannot be applied to a general open subset of $(0, T) \times \overline{\Omega}$. One of the contributions of the present work is to remedy this issue. In fact, this is done by a significant simplification of the argument of [Bardos et al. 1992], yielding a less technical aspect in one of the steps of the proof. Eventually, the result that we obtain is in fact faithful to the intuition one may have. The proper geometric condition to impose on an open subset Q of $(0, T) \times \overline{\Omega}$, for an observability condition of the form

$$E_0(u) \leqslant C \iint_Q |\partial_t u|^2 \, dt \, dx \tag{2}$$

to hold is the following: for any generalized ray $t \mapsto x(t)$ initiated at time $t = 0$, there should be a time $0 < t_1 < T$ such that the ray is located in $Q \cap \{t = t_1\}$, that is, $(t_1, x(t_1)) \in Q$. This naturally generalizes the usual geometric control condition in the case where Q is the cylinder $(0, T) \times \omega$.

One of the interesting consequences of our analysis lies in the following fact: if the geometric condition holds for a time-dependent domain Q, a thinner domain for which the condition holds as well can be simply obtained by picking a neighborhood of the boundary of Q. This can be viewed as a step towards the reduction of the amount of data collected in the practical applications mentioned above.

We complete our analysis with a set of examples in *very simple* geometrical situations. Some of these examples show that even if "many" rays are missed by a static domain, a moving version of this domain can capture in finite time all rays, even with a very slow motion. However other examples show situations in which "very few" rays are missed, and a slow motion of the observation set allows one to capture these rays, yet implying that other rays remain away from the moving observation region for any positive time. Those examples may become hard to analyse because of the complexity of the Hamiltonian dynamics that governs the rays. Yet, they illustrate that naive strategies can fail to achieve the fulfillment of the geometric control condition. Those examples show that further study would be of interest, with a study of the increase or decrease of the minimal control time as an observation set is moved around. Some examples show that this minimal control time may not be continuous with respect to the dynamics we impose on a moving control region.

1B. ***Setting.*** Let (M, g) be a smooth d-dimensional Riemannian manifold, with $d \geqslant 1$. Let Ω be an open bounded connected subset of M, with a smooth boundary if $\partial\Omega \neq \varnothing$. We consider the wave equation

$$\partial_t^2 u - \triangle_g u = 0 \tag{3}$$

in $\mathbb{R} \times \Omega$. Here, $\triangle_g$ denotes the Laplace–Beltrami operator on M, associated with the metric g on M. If the boundary $\partial\Omega$ of Ω is nonempty, then we consider boundary conditions of the form

$$Bu = 0 \quad \text{on } \mathbb{R} \times \partial\Omega, \tag{4}$$

where the operator B is either

- the Dirichlet trace operator, $Bu = u_{|\partial\Omega}$;
- or the Neumann trace operator, $Bu = \partial_n u_{|\partial\Omega}$, where ∂_n is the outward normal derivative along $\partial\Omega$.

Our study encompasses the case where $\partial\Omega = \varnothing$: in this case, Ω is a compact connected d-dimensional Riemannian manifold. Measurable sets are considered with respect to the Riemannian measure dx_g (if M is the usual Euclidean space $\mathbb{R}^n$ then dx_g is the usual Lebesgue measure).

In the case of a manifold without boundary or in the case of homogeneous Neumann boundary conditions, the Laplace–Beltrami operator is not invertible on $L^2(\Omega)$ but is invertible in

$$L^2_0(\Omega) = \left\{ u \in L^2(\Omega) \;\middle|\; \int_\Omega u(x) \, dx_g = 0 \right\}.$$

In what follows, we set $X = L^2_0(\Omega)$ in the boundaryless case or in the Neumann case, and $X = L^2(\Omega)$ in the Dirichlet case (in both cases, the norm on X is the usual L^2-norm). We denote by $A = -\triangle_g$ the Laplace operator defined on X with domain

$$D(A) = \{u \in X \mid Au \in X \text{ and } Bu = 0\}$$

with one of the above boundary conditions whenever $\partial\Omega \neq \varnothing$.

Note that A is a selfadjoint positive operator. In the case of Dirichlet boundary conditions, $X = L^2(\Omega)$ and we have

$$D(A) = H^2(\Omega) \cap H^1_0(\Omega), \quad D(A^{1/2}) = H^1_0(\Omega) \quad \text{and} \quad D(A^{1/2})' = H^{-1}(\Omega),$$

where the dual is considered with respect to the pivot space X. For Neumann boundary conditions, $X = L^2_0(X)$ and we have

$$D(A) = \{u \in H^2(\Omega) \cap L^2_0(X) \mid (\partial u/\partial n)_{|\partial\Omega} = 0\} \quad \text{and} \quad D(A^{1/2}) = H^1(\Omega) \cap L^2_0(X).$$

The Hilbert spaces $D(A)$, $D(A^{1/2})$, and $D(A^{1/2})'$ are respectively endowed with the norms $\|u\|_{D(A)} = \|Au\|_{L^2(\Omega)}$, $\|u\|_{D(A^{1/2})} = \|A^{1/2}u\|_{L^2(\Omega)}$ and $\|u\|_{D(A^{1/2})'} = \|A^{-1/2}u\|_{L^2(\Omega)}$.

For all $(u^0, u^1) \in D(A^{1/2}) \times X$ (resp. $X \times D(A^{1/2})'$), there exists a unique solution $u \in \mathscr{C}^0(\mathbb{R}; D(A^{1/2})) \cap \mathscr{C}^1(\mathbb{R}; X)$ (resp. $u \in \mathscr{C}^0(\mathbb{R}; X) \cap \mathscr{C}^1(\mathbb{R}; D(A^{1/2})'))$ of (3)–(4) such that $u_{|t=0} = u^0$ and $\partial_t u_{|t=0} = u^1$. In both cases, such solutions of (3)–(4) are to be understood in a weak sense.

Remark 1.1. In (3), we consider the classical d'Alembert wave operator $\Box_g = \partial_t^2 - \triangle_g$. In fact, the results of the present article remain valid for the more general wave operators of the form

$$P = \partial_t^2 - \sum_{i,j} a_{ij}(x)\partial_{x_i}\partial_{x_j} + \text{lower-order terms},$$

where $(a_{ij}(x))$ is a smooth real-valued symmetric positive definite matrix, and where the lower-order terms are smooth and do not depend on t. We insist on the fact that our approach is limited to operators with time-independent coefficients as in [Bardos et al. 1992].

1C. *Observability.* Let Q be an open subset of $\mathbb{R} \times \overline{\Omega}$. We denote by χ_Q the characteristic function of Q, defined by $\chi_Q(t,x) = 1$ if $(t,x) \in Q$ and $\chi_Q(t,x) = 0$ otherwise. We set

$$\omega(t) = \{x \in \Omega \mid (t,x) \in Q\},$$

so that $Q = \{(t,x) \in \mathbb{R} \times \Omega \mid t \in \mathbb{R},\ x \in \omega(t)\}$. Let $T > 0$ be arbitrary. We say that (3)–(4) is observable on Q in time T if there exists $C > 0$ such that

$$C\|(u^0,u^1)\|^2_{D(A^{1/2})\times X} \leqslant \|\chi_Q\, \partial_t u\|^2_{L^2((0,T)\times\Omega)} = \int_0^T \int_{\omega(t)} |\partial_t u(t,x)|^2\, dx_g\, dt \tag{5}$$

for all $(u^0, u^1) \in D(A^{1/2}) \times X$, where u is the solution of (3)–(4) with initial conditions $u_{|t=0} = u^0$ and $\partial_t u_{|t=0} = u^1$. One refers to (5) as to an *observability inequality*.

The observability inequality (5) is stated for initial conditions $(u^0, u^1) \in D(A^{1/2}) \times X$. Other energy spaces can be used. An important example is the following proposition.

Proposition 1.2. *The observability inequality (5) is equivalent to having $C > 0$ such that*

$$C\|(u^0, u^1)\|^2_{X\times D(A^{1/2})'} \leqslant \|\chi_Q u\|^2_{L^2((0,T)\times\Omega)} = \int_0^T \int_{\omega(t)} |u(t,x)|^2 \, dx_g \, dt \tag{6}$$

for all $(u^0, u^1) \in X \times D(A^{1/2})'$, where u is the solution of (3)–(4) with initial conditions $u_{|t=0} = u^0$ and $\partial_t u_{|t=0} = u^1$.

This proposition is proven in Section 2B.

In the existing literature, the observation is most often made on cylindrical domains $Q = (0, T) \times \omega$ for some given $T > 0$, meaning that $\omega(t) = \omega$. In such a case where the observation domain ω is stationary, it is known that, within the class of smooth domains Ω, the observability property holds if the pair (ω, T) satisfies the *geometric control condition* (in short, GCC) in Ω (see [Bardos et al. 1992; Burq and Gérard 1997]). Roughly speaking, it says that every geodesic propagating in Ω at unit speed, and reflecting at the boundary according to the classical laws of geometrical optics, so-called generalized geodesics, should meet the open set ω within time T.

In the present article, our goal is to extend the GCC to *time-dependent observation domains*. For a precise statement of the GCC, we first recall the definition of generalized geodesics and bicharacteristics.

1C1. *The generalized bicharacteristic flow of R. Melrose and J. Sjöstrand.* First, we define generalized bicharacteristics in the interior of Ω. There, they coincide with the classical notion of bicharacteristics. Second, we define generalized bicharacteristics in the neighborhood of the boundary.

Symbols and bicharacteristics in the interior. The principal symbol of $-\triangle_g$ coincides with the cometric g^* defined by

$$g^*_x(\xi, \xi) = \max_{v \in T_x M \setminus\{0\}} \frac{\langle \xi, v\rangle^2}{g_x(v, v)}$$

for every $x \in M$ and every $\xi \in T^*_x M$. In local coordinates, we denote by $g_{ij}(x)$ the Riemannian metric g at point x; that is, $g(v, \tilde{v})(x) = g_{ij}(x) v^i(x) \tilde{v}^j(x)$, for $v, \tilde{v} \in TM$, that is, two vector fields, and by $g^{ij}(x)$ the cometric g^* at x, that is $g^*(\omega, \tilde{\omega})(x) = g^{ij}(x)\omega_i(x)\tilde{\omega}_j(x)$ for $\omega, \tilde{\omega} \in T^*M$, that is, two 1-forms. In local coordinates, the Laplace–Beltrami reads

$$-\triangle_g = -\mathrm{g}(x)^{-1/2}\, \partial_i (\mathrm{g}(x)^{1/2} g^{ij}(x) \partial_j).$$

In $\mathbb{R} \times M$, the principal symbol of the wave operator $\partial_t^2 - \triangle_g$ is then $p(t, x, \tau, \xi) = -\tau^2 + g^*_x(\xi, \xi)$. In $T^*(\mathbb{R} \times M)$, the Hamiltonian vector field H_p associated with p is given by $H_p f = \{p, f\}$ for $f \in \mathscr{C}^1(T^*(\mathbb{R}\times M))$. In local coordinates, H_p reads

$$H_p = \partial_\tau p \partial_t + \nabla_\xi p \,\nabla_x - \nabla_x p\, \nabla_x = -2\tau\partial_t + 2g^{jk}(x)\xi_k \partial_{x_j} - \partial_{x_j} g^{ik}(x) \xi_i \xi_k \partial_{\xi_j},$$

with the usual Einstein summation convention. Along the integral curves of H_p, the value of p is constant as $H_p p = 0$. Thus, the characteristic set $\mathrm{Char}(p) = \{p = 0\}$ is invariant under the flow of H_p. In $T^*(\mathbb{R} \times M)$, bicharacteristics are defined as the maximal integral curves of H_p that lay in $\mathrm{Char}(p)$. The projections of the bicharacteristics onto M, using the variable t as a parameter, coincide with the geodesics on M associated with the metric g travelled at speed 1.

We set $Y = \mathbb{R} \times \overline{\Omega}$. We denote by $\mathrm{Char}_Y(p)$ the characteristic set of p above Y, given by

$$\mathrm{Char}_Y(p) = \big\{\rho=(t,x,\tau,\xi) \in T^*(\mathbb{R} \times M) \setminus 0 \;\big|\; x \in \overline{\Omega} \text{ and } p(\rho)=0\big\}.$$

Coordinates and Hamiltonian vector fields near and at the boundary. Close to the boundary $\mathbb{R} \times \partial\Omega$, using normal geodesic coordinates (x', x_d), the principal symbol of the Laplace–Beltrami operator reads $\xi_d^2 + \ell(x, \xi')$. Set $y = (t, x)$, $y' = (t, x')$ and $y_n = x_d$. Here $n = d + 1$. In these coordinates, the principal symbol of the wave operator takes the form $p(y', y_n, \eta', \eta_n) = \eta_n^2 + r(y, \eta')$, where r is a smooth y_n-family of tangential (differential) symbols, and the boundary $\mathbb{R} \times \partial\Omega$ is locally parametrized by y' and given by $\{y_n = 0\}$. The open set $\mathbb{R} \times \Omega$ is locally given by $\{y_n > 0\}$.

The variables $\eta = (\eta', \eta_n)$ are the cotangent variables associated with $y = (y', y_n)$. We set

$$\partial T^*Y = \{\rho=(y,\eta) \in T^*(\mathbb{R} \times M) \mid y_n=0\}$$

as the boundary of $T^*Y = \{\rho=(y,\eta) \in T^*(\mathbb{R} \times M) \mid y \in Y\}$. In those local coordinates, the associated Hamiltonian vector field H_p is given by

$$H_p = \nabla_{\eta'} r \nabla_{y'} + 2\eta_n \partial_{y_n} - \nabla_y r \nabla_\eta.$$

We denote by r_0 the trace of r on ∂T^*Y, that is, $r_0(y', \eta') = r(y', y_n=0, \eta')$. We then introduce the Hamiltonian vector field above the submanifold $\{y_n = 0\}$:

$$H_{r_0} = \nabla_{\eta'} r_0 \nabla_{y'} - \nabla_{y'} r_0 \nabla_{\eta'}.$$

The compressed cotangent bundle. On Y, for points $y = (y', y_n)$ near the boundary, we define the vector fiber bundle ${}^bTY = \bigcup_{y\in Y} {}^bT_yY$, generated by the vector fields $\partial_{y'}$ and $y_n\partial_{y_n}$, in the local coordinates introduced above. We then have the natural map

$$j : T^*Y \to {}^bT^*Y = \bigcup_{y\in Y} ({}^bT_yY)^*, \quad (y;\eta) \mapsto (y; \eta', y_n\eta_n),$$

expressed here in local coordinates for simplicity. In particular:

- If $y \in \mathbb{R} \times \Omega$ then ${}^bT_y^*Y = j(T_y^*Y)$ is isomorphic to $T_y^*Y = T_y^*(\mathbb{R} \times M)$.
- If $y \in \mathbb{R} \times \partial\Omega$ then ${}^bT_y^*Y = j(T_y^*Y)$ is isomorphic to $T_y^*(\mathbb{R} \times \partial\Omega)$.

The bundle ${}^bT^*Y$ is called the *compressed cotangent bundle*, and we see that it allows one to patch together $T_y^*(\mathbb{R} \times M)$ in the interior of Ω and $T_y^*(\mathbb{R} \times \partial\Omega)$ at the boundary in a smooth manner, despite the discrepancy in their dimensions.

Decomposition of the characteristic set at the boundary. We set $\Sigma = j(\mathrm{Char}_Y(p)) \subset {}^bT^*Y$ and

$$\Sigma_0 = \Sigma_{|y_n=0} \subset \partial\, {}^bT^*Y = {}^bT^*Y_{|y_n=0} \simeq T^*(\mathbb{R} \times \partial\Omega).$$

Using local coordinates (for convenience here), we then define $G \subset \Sigma_0$ by $r(y, \eta') = r_0(y', \eta') = 0$ as the glancing set and $H = \Sigma_0 \setminus G$ as the hyperbolic set. Hence, if $\rho = (y', y_n=0, \eta') \in \Sigma_0$ then

$$\rho \in H \iff r_0(y', \eta') < 0, \quad \rho \in G \iff r_0(y', \eta') = 0.$$

The set of points $(y', y_n = 0, \eta') \in {}^bT^*Y_{|y_n=0}$ such that $r_0(y', \eta') > 0$ is referred to as the elliptic set E. We also set $\hat{\Sigma} = \Sigma \cup E = \Sigma \cup {}^bT^*Y_{|y_n=0}$ and we set the cosphere quotient space $S^*\hat{\Sigma}$ to be $\hat{\Sigma}/(0, +\infty)$.

The glancing set is itself written as $G = G^2 \supset G^3 \supset \cdots \supset G^\infty$, with $\rho = (y', y_n = 0, \eta)$ in G^{k+2} if and only if

$$\eta_n = r_0(\rho) = 0, \qquad H^j_{r_0} r_1(\rho) = 0, \quad 0 \leqslant j < k,$$

where $r_1(\rho) = \partial_{y_n} r(y', y_n = 0, \eta')$. Finally, we write $G^2 \setminus G^3$, the set of glancing points of order exactly 2, as the union of the diffractive set G^2_d and of the gliding set G^2_g; that is, $G^2 \setminus G^3 = G^2_d \cup G^2_g$, with

$$\rho \in G^2_d \text{ (resp. } G^2_g) \iff \rho \in G^2 \setminus G^3 \text{ and } r_1(\rho) < 0 \text{ (resp. } > 0).$$

Similarly, for $\ell \geqslant 2$, we write $G^{2\ell} \setminus G^{2\ell+1}$, the set of glancing points of order exactly $k = 2\ell$, as the union of the diffractive set $G^{2\ell}_d$ and the gliding set $G^{2\ell}_g$; that is, $G^{2\ell} \setminus G^{2\ell+1} = G^{2\ell}_d \cup G^{2\ell}_g$, with

$$\rho \in G^{2\ell}_d \text{ (resp. } G^{2\ell}_g) \iff \rho \in G^{2\ell} \setminus G^{2\ell+1} \text{ and } H^{2\ell-2}_{r_0} r_1(\rho) < 0 \text{ (resp. } > 0).$$

We shall call points in $G_d = \bigcup_{\ell \geqslant 1} G^{2\ell}_d$ diffractive.[2]

Observe that a bicharacteristic going through a point of G^k projects onto a geodesic on M that has a contact of order k with $\mathbb{R} \times \partial\Omega$.

Generalized bicharacteristics. What we introduced above now allows us to give a precise definition of generalized bicharacteristics above Y.

Definition 1.3. A generalized bicharacteristic of p is a differentiable map

$$\mathbb{R} \setminus B \ni s \mapsto \gamma(s) \in (\mathrm{Char}_Y(p) \setminus \partial T^*Y) \cup j^{-1}(G),$$

where B is a subset of $\mathbb{R}$ made of isolated points, that satisfies the following properties:

(i) $\gamma'(s) = H_p(\gamma(s))$ if either $\gamma(s) \in \mathrm{Char}_Y(p) \setminus \partial T^*Y$ or $\gamma(s) \in j^{-1}(G_d)$.

(ii) $\gamma'(s) = H_{r_0}(\gamma(s))$ if $\gamma(s) \in j^{-1}(G \setminus G_d)$.

(iii) If $s_0 \in B$, there exists $\delta > 0$ such that $\gamma(s) \in \mathrm{Char}_Y(p) \setminus \partial T^*Y$ for $s \in (s_0 - \delta, s_0) \cup (s_0, s_0 + \delta)$. Moreover, the limits $\rho^\pm = (y^\pm, \eta^\pm) = \lim_{s \to s_0^\pm} \gamma(s)$ exist and $y_n^- = y_n^+ = 0$; i.e., $\rho^\pm \in \mathrm{Char}_Y(p) \cap \partial T^*Y$, $y^{-\prime} = y^{+\prime}$, $\eta^{-\prime} = \eta^{+\prime}$, and $\eta_n^- = -\eta_n^+$. That is, ρ^+ and ρ^- lay in the same hyperbolic fiber above a point in ${}^bT^*Y_{|y_n=0}$: $j(\rho^+) = j(\rho^-) \in H$.

Point (i) describes the generalized bicharacteristic in the interior, that is, in $T^*(\mathbb{R} \times \Omega)$, and at diffractive points, where it coincides with part of a classical bicharacteristic as defined above. Point (ii) describes the behavior in $G \setminus G_d$, thus explaining that a generalized bicharacteristic can enter or leave the boundary ∂T^*Y or locally remain in it. Point (iii) describes reflections when the boundary ∂T^*Y is reached transversally by a classical bicharacteristic, that is, at a point of the hyperbolic set. While $s \mapsto \xi(s)$

[2] In the sense of Taylor and Melrose the terminology diffractive only applies to G^2_d. Here, we chose to extend it to $\bigcup_{\ell \geqslant 1} G^{2\ell}_d$ as we shall refer to nondiffractive points, that is, points in the complement of G_d, in Section 4. In the literature nondiffractive points are defined this way but diffractive points are often defined according to Taylor and Melrose. Then, the set of nondiffractive points and the set of diffractive points are not complements of one another; a source of confusion.

exhibits a jump at such a point, $s \mapsto t(s)$ and $s \mapsto x(s)$ can both be extended by continuity there. We shall thus proceed with this extension. Above, for clarity we chose to state point (iii) in local coordinates near the hyperbolic point. The generalized bicharacteristics are however defined as a geometrical object, independent of the choice of coordinates.

Definition 1.4. Compressed generalized bicharacteristics are the image under the map j of the generalized bicharacteristics defined above.

If ${}^b\gamma = j(\gamma)$ is such a compressed generalized bicharacteristic, then ${}^b\gamma : \mathbb{R} \to {}^bT^*Y \setminus E$ is a continuous map (if one introduces the proper natural topology on ${}^bT^*Y$).

Using t as parameter, generalized geodesics for Ω, travelled at speed 1, are then the projection on M of the (compressed) generalized bicharacteristics. Generalized geodesics remain in $\overline{\Omega}$. We shall call a "ray" this projection following the terminology of geometrical optics.

An important result is then the following.

Proposition 1.5. *A (compressed) generalized bicharacteristic with no point in G^∞ is uniquely determined by any one of its points.*

We refer to [Melrose and Sjöstrand 1978] for a proof of this result and for many more details on generalized bicharacteristics (see also [Hörmander 1985, Section 24.3]).

1C2. *A time-dependent geometric control condition.* With the notion of compressed generalized bicharacteristic recalled in Section 1C1, we can state the geometric condition adapted to a time-dependent control domain.

Definition 1.6. Let Q be an open subset of $\mathbb{R} \times \overline{\Omega}$, and let $T > 0$. We say that (Q, T) satisfies the *time-dependent geometric control condition* (in short, t-GCC) if every generalized bicharacteristic ${}^b\gamma : \mathbb{R} \to {}^bT^*Y \setminus E$, $s \mapsto (t(s), x(s), \tau(s), \xi(s))$, is such that there exists $s \in \mathbb{R}$ such that $t(s) \in (0, T)$ and $(t(s), x(s)) \in Q$. We say that Q satisfies the t-GCC if there exists $T > 0$ such that (Q, T) satisfies the t-GCC.

The control time $T_0(Q, \Omega)$ is defined by

$$T_0(Q, \Omega) = \inf\{T > 0 \mid (Q, T) \text{ satisfies the } t\text{-GCC}\},$$

with the agreement that $T_0(Q, \Omega) = +\infty$ if Q does not satisfy the t-GCC.

The t-GCC property of Definition 1.6 is a time-dependent version of the usual GCC.

Remark 1.7. Several remarks are in order.

(1) The t-GCC assumption implies that the set $\mathcal{O} = \bigcup_{t\in(0,T)} \omega(t)$ is a control domain that satisfies the usual GCC for a time $T > T_0(Q, \Omega)$.

(2) It is interesting to note that the control time $T_0(Q, \Omega)$ is not a continuous function of the domains for any reasonable topology (see Remark 3.1 below).

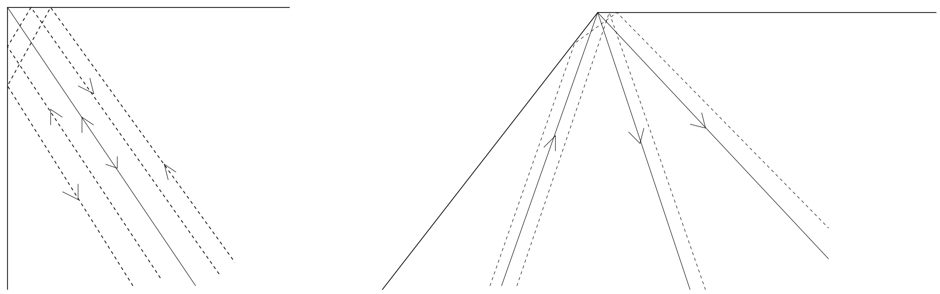


Figure 1. Reflections of generalized geodesics at some corner: a right angle corner (left) and a general corner (right).

(3) Observe that if (Q, T) satisfies the t-GCC, a similar geometric condition may not occur if the time interval $(0, T)$ is replaced by $(t_0, t_0 + T)$. As the set Q is not a cylinder in general, by nature the t-GCC is not invariant under time translation.

(4) Note that Q cannot be chosen as an open set of $\mathbb{R} \times \Omega$ instead of $\mathbb{R} \times \overline{\Omega}$. Consider indeed the case of a disk: if Q is an open set of $\mathbb{R} \times \Omega$ then the ray that glides along the boundary never enters Q. This coincides with the so-called whispering gallery phenomenon.

1D. *Main result.*

Theorem 1.8. *Let Q be an open subset of $\mathbb{R} \times \overline{\Omega}$ that satisfies the t-GCC. Let $T > T_0(Q, \Omega)$. If $\partial\Omega \neq \varnothing$, we assume moreover that no generalized bicharacteristic has a contact of infinite order with $(0, T) \times \partial\Omega$, that is, $G^\infty = \varnothing$. Then, the observability inequality (5) holds.*

Theorem 1.8 is proven in Section 2A. By Proposition 1.2 we have the following result.

Theorem 1.8′. *Under the same assumptions as Theorem 1.8 the observability inequality (6) holds.*

Remark 1.9. (1) In the case where $\partial\Omega \neq \varnothing$, the assumption of the absence of any ray having a contact of infinite order with the boundary is classical (see [Bardos et al. 1992]). Note that this assumption is not useful if M and $\partial\Omega$ are analytic. This assumption is used in a crucial way in the proof of the theorem to ensure uniqueness of the generalized bicharacteristic flow, as stated in Proposition 1.5.

(2) We have assumed here that, if $\partial\Omega \neq \varnothing$, then $\partial\Omega$ is smooth. The case where $\partial\Omega$ is not smooth is open. Even the case where $\partial\Omega$ is piecewise analytic is open. The problem is that, in that case, the generalized bicharacteristic flow is not well defined since there is no uniqueness of a bicharacteristic passing over a point. This fact is illustrated in Figure 1 (right) where a ray reflecting at some angle can split into two rays. However, it clearly follows from our proof that Theorem 1.8 is still valid if the domain Ω is such that this uniqueness property holds (like in the case of a rectangle). In general, we conjecture that the conclusion of Theorem 1.8 holds true if *all* generalized bicharacteristics meet T^*Q within time T. This would require however extending the classical theory of propagation of singularities. Proving this fact is beyond of the scope of the present article. We may however assert here, in the present context, that the result of Theorem 1.8 is valid in any d-dimensional orthotope.

Remark 1.10. In the case of a 1-dimensional wave equation with Dirichlet boundary conditions, the corresponding statement of Theorem 1.8 is proven in [Castro et al. 2014] by means of the d'Alembert formula. The proof we provide in Section 2A is general and follows [Bardos et al. 1992; Burq and Gérard 1997]. In fact, as already mentioned in Section 1A, a key step of the approach of these papers is simplified here, and more precisely, it consists of several steps. Firstly, a weaker version of the observability inequality is proven; in the present article, this is done in Lemma 2.1. Secondly, the so-called *set of invisible solutions*, defined by (16), is shown to be reduced to zero; in the present article, this is done in Lemma 2.3. Thirdly, the observability inequality is proven to hold by means of the result of the two previous steps. Our simplification with respect to [Bardos et al. 1992; Burq and Gérard 1997] lies in the second step. The argument is much shorter than the original one and, in the present analysis of a time-varying observation region, it turns out to be crucial, as the more classical argument of [Bardos et al. 1992; Burq and Gérard 1997] cannot be applied.

1E. *Consequences.*

1E1. *Controllability.* By the usual duality argument (the Hilbert uniqueness method, see [Lions 1988a; 1988b]), we have the following equivalent result for the control of the wave equation with a time-dependent control domain, based on the observability inequality (6) that follows from Theorem 1.8′.

Theorem 1.8″. *Let Q be an open subset of $\mathbb{R}\times\overline{\Omega}$ that satisfies the t-GCC. Let $T > T_0(Q,\Omega)$. If $\partial\Omega\neq\varnothing$ we assume moreover that no generalized bicharacteristic has a contact of infinite order with $(0,T)\times\partial\Omega$, that is, $G^\infty=\varnothing$. Setting $\omega(t)=\{x\in\Omega \mid (t,x)\in Q\}$, we consider the wave equation with internal control*

$$\partial_t^2 u-\triangle_g u=\chi_Q f \tag{7}$$

in $(0,T)\times\Omega$, with Dirichlet or Neumann boundary conditions (4) whenever $\partial\Omega\neq\varnothing$, and with $f\in L^2((0,T)\times\Omega)$. Then, the controlled equation (7) is exactly controllable in the space $D(A^{1/2})\times X$, meaning that, for all (u^0,u^1) and (v^0,v^1) in $D(A^{1/2})\times X$, there exists $f\in L^2((0,T)\times\Omega)$ such that the corresponding solution of (7), with $(u_{|t=0},\partial_t u_{|t=0})=(u^0,u^1)$, satisfies $(u_{|t=T},\partial_t u_{|t=T})=(v^0,v^1)$.

Remark 1.11. In the above result the control operator is $f\mapsto\chi_Q f$. We could choose instead a control operator $f\mapsto b(t,x)f$ with b smooth and such that $Q=\{(t,x)\in\mathbb{R}\times\overline{\Omega}\mid b(t,x)>0\}$. Then with the same t-GCC we also have exact controllability in this case. The equivalent observability inequality is then

$$C\|(u^0,u^1)\|^2_{X\times D(A^{1/2})'}\leqslant\|bu\|^2_{L^2((0,T)\times\Omega)}=\int_0^T\int_\Omega|b(t,x)u(t,x)|^2\,dx_g\,dt.$$

1E2. *Observability with few sensors.* We give another interesting consequence of Theorem 1.8, in connection with the very definition of the t-GCC property, which can be particularly relevant in view of practical applications.

Corollary 1.12. *Let $Q\subset\mathbb{R}\times\overline{\Omega}$ be an open subset with Lipschitz boundary and let $T>0$ be such that (Q,T) satisfies the t-GCC. Then, every open subset $\mathcal{V}$ of $[0,T]\times\overline{\Omega}$ (for the topology induced by $\mathbb{R}\times M$), containing $\partial(Q\cap([0,T]\times\overline{\Omega}))$ is such that $(\mathcal{V},T)$ satisfies the t-GCC and, consequently, observability holds for such an open subset.*

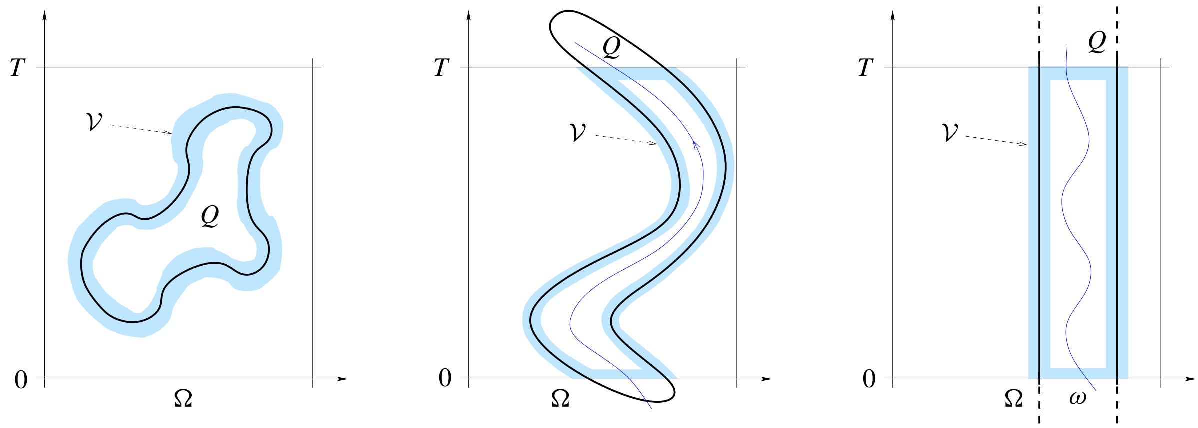


Figure 2. Neighborhoods $\mathcal{V}$ of $\partial(Q \cap ([0,T] \times \overline{\Omega}))$ in $[0,T] \times \overline{\Omega}$ (for the induced topology). In middle and right pictures, a potential bicharacteristic that remains in the interior of Q is represented.

Proof. Let $\mathbb{R} \ni s \mapsto {}^b\gamma(s)$ be a compressed generalized bicharacteristic with $t = t(s)$. As (Q,T) satisfies the t-GCC, there exists $t_1 \in (0,T)$ and $s_1 \in \mathbb{R}$ such that $t_1 = t(s_1)$ and ${}^b\gamma(s_1) \in j(T^*(Q))$. Now, there are two cases:

Case 1: There exists $s_2 \in \mathbb{R}$ such that $t_2 = t(s_2) \in (0,T)$ and ${}^b\gamma(s_2) \notin j(T^*(Q))$. The continuity of $s \mapsto {}^b\gamma(s)$ into ${}^bT^*Y \setminus E$, in particular of its projection on $\mathbb{R} \times \overline{\Omega}$, then allows one to conclude that there exists $s_3 \in \mathbb{R}$ such that $t_3 = t(s_3) \in (0,T)$ and ${}^b\gamma(s_3) \in j(T^*(\mathcal{V}))$.

Case 2: For all $s \in \mathbb{R}$ such that $t = t(s) \in (0,T)$ we have ${}^b\gamma(s) \in j(T^*(Q))$. Such a bicharacteristic is illustrated in Figure 2 (middle and right). Then $s \mapsto {}^b\gamma(s)$ enters $j(T^*(\mathcal{W}))$ for any neighborhood $\mathcal{W}$ of $\{T\} \times \omega(T)$ (or $\{0\} \times \omega(0)$). Thus, there exists $s_2 \in \mathbb{R}$ such that $t_2 = t(s_2) \in (0,T)$ and ${}^b\gamma(s_2) \in j(T^*(\mathcal{V}))$. ☐

Remark 1.13. (1) The main interest of Corollary 1.12 is that it allows one to take the open set $\mathcal{V}$ "as small as possible", provided that it contains the boundary of $Q \cap ([0,T] \times \overline{\Omega})$ (see Figure 2). As a practical consequence, only few sensors are needed to ensure the observability property, or, by duality, the controllability property, thus reducing the cost of an experiment.

Somehow, with an internal control we have $d+1$ degrees of freedom for the control of d variables, and this explains intuitively why the choice of a "thin" open set $\mathcal{V}$ is possible. The above corollary roughly states that control is still feasible by using only d degrees of freedom. In terms of the control domain, this means that we only need a control domain that is any open neighborhood of a set of Hausdorff dimension d.

(2) Observe that the proof of Corollary 1.12 and Figure 2 (middle and right) shows in fact that it suffices to choose $\mathcal{V}$ as the union of a neighborhood of $\partial Q \cap (0,T) \times \overline{\Omega}$ and a neighborhood of $Q \cap \{t = 0\}$ (or $Q \cap \{t = T\}$).

(3) Note that, if an open subset ω of $\overline{\Omega}$ satisfies the usual GCC, then a small neighborhood of $\partial\omega$ does not satisfy necessarily the GCC. In contrast, when considering time-space control domains (i.e., subsets of

$\mathbb{R} \times \overline{\Omega}$), the situation is different. For instance, if ω satisfies the GCC, then any open subset of $[0, T] \times \overline{\Omega}$ containing $\partial([0, T] \times \omega) = ([0, T] \times \partial\omega) \cup (\{0\} \times \omega) \cup (\{T\} \times \omega)$ satisfies the t-GCC; see Figure 2 (right).

(4) Note that $T(Q, \Omega) \leqslant \liminf T(\mathcal{V}, \Omega)$ as $\mathcal{V}$ shrinks to $\partial(Q \cap ([0, T] \times \overline{\Omega}))$, and that equality may fail as there may exist some bicharacteristics propagating inside Q and not reaching $\mathcal{V}$ for $t \in (t_1, t_2)$ for with $0 < t_1 < t_2 < T$; see Figure 2 (middle and right).

1E3. *Stabilization.* Theorem 1.8 has the following consequence for wave equations with a damping localized on a domain Q that is time-periodic.

Corollary 1.14. *Let Q be an open subset of $\mathbb{R} \times \overline{\Omega}$, satisfying the t-GCC. Let $T > T_0(Q, \Omega)$. If $\partial\Omega \neq \varnothing$, we assume moreover that no generalized bicharacteristic has a contact of infinite order with $(0, T) \times \partial\Omega$, that is, $G^\infty = \varnothing$. Setting $\omega(t) = \{x \in \Omega \mid (t, x) \in Q\}$, we assume that ω is T-periodic, that is, $\omega(t + T) = \omega(t)$, for almost every $t \in (0, T)$. We consider the wave equation with a local internal damping term*

$$\partial_t^2 u - \triangle_g u + \chi_\omega \, \partial_t u = 0 \tag{8}$$

in $(0, T) \times \Omega$, with Dirichlet or Neumann boundary conditions (4) whenever $\partial\Omega \neq \varnothing$. Then, there exists $\mu \geqslant 0$ and $\nu > 0$ such that any solution of (7), with $(u(0), \partial_t u(0)) \in D(A^{1/2}) \times X$, satisfies

$$E_0(u)(t) \leqslant \mu E_0(u)(0) e^{-\nu t},$$

where we have set $E_0(u)(t) = \frac{1}{2}\big(\|A^{1/2}u(t)\|^2_{L^2(\Omega)} + \|\partial_t u(t)\|^2_{L^2(\Omega)}\big)$.

Corollary 1.14 is proven in Section 2C.

2. Proofs

2A. ***Proof of Theorem 1.8.*** The proof follows the classical chain of arguments developed in [Bardos et al. 1992; Burq and Gérard 1997], with yet a simplification of one of the key steps, as already pointed out in Remark 1.10. This simplification is a key element here. The original proof scheme would not allow one to conclude in the case of a time-dependent control domain.

For a solution u of (3)–(4), with $u_{|t=0} = u^0 \in D(A^{1/2})$ and $\partial_t u_{|t=0} = u^1 \in X$, we use the natural energy

$$E_0(u)(t) = \tfrac{1}{2}\big(\|u(t)\|^2_{D(A^{1/2})} + \|\partial_t u(t)\|^2_X\big). \tag{9}$$

In the proof, we use the fact that this energy remains constant as time evolves, that is,

$$E_0(u)(t) = E_0(u)(0) = \tfrac{1}{2}\big(\|u^0\|^2_{D(A^{1/2})} + \|u^1\|^2_X\big) = \tfrac{1}{2}\|(u^0, u^1)\|^2_{D(A^{1/2}) \times X}, \tag{10}$$

and we shall simply write $E_0(u)$ at places.

We first achieve a weak version of the observability inequality.

Lemma 2.1. *There exists $C > 0$ such that*

$$C\|(u^0, u^1)\|^2_{D(A^{1/2}) \times X} \leqslant \|\chi_Q \, \partial_t u\|^2_{L^2((0,T) \times \Omega)} + \|(u^0, u^1)\|^2_{X \times D(A^{1/2})'} \tag{11}$$

for all $(u^0, u^1) \in D(A^{1/2}) \times X$, where u is the corresponding solution of (3)–(4) with $u_{|t=0} = u^0$ and $\partial_t u_{|t=0} = u^1$.

Remark 2.2. With respect to the desired observability inequality (5), the inequality (11) exhibits a penalization term, $\|(u^0, u^1)\|_{X\times D(A^{1/2})'}$, on the right-hand side. Note that the Sobolev spaces under consideration have no importance, and instead of $X \times D(A^{1/2})'$ we could as well have chosen $H^{1/2-s}(\Omega)\times H^{-1/2-s}(\Omega)$ for any $s > 0$. The key point lies in the fact that the space $D(A^{1/2})\times X$ is compactly embedded into $X \times D(A^{1/2})'$.

Proof. We prove the result by contradiction. We assume that there exists a sequence $(u_n^0, u_n^1)_{n\in\mathbb{N}}$ in $D(A^{1/2}) \times X$ such that

$$\|(u_n^0, u_n^1)\|_{D(A^{1/2})\times X} = 1 \quad \forall n \in \mathbb{N}, \tag{12}$$

$$\|(u_n^0, u_n^1)\|_{X\times D(A^{1/2})'} \to 0 \quad \text{as } n \to +\infty, \tag{13}$$

$$\|\chi_Q\, \partial_t u_n\|_{L^2((0,T)\times\Omega)} \to 0 \quad \text{as } n \to +\infty, \tag{14}$$

where u_n is the solution of (3)–(4) satisfying $u_{n|t=0} = u_n^0$ and $\partial_t u_{n|t=0} = u_n^1$. From (12), the sequence $(u_n^0, u_n^1)_{n\in\mathbb{N}}$ is bounded in $D(A^{1/2}) \times X$, and using (13) the only possible closure point for the weak topology of $D(A^{1/2}) \times X$ is $(0, 0)$. Therefore, the sequence $(u_n^0, u_n^1)_{n\in\mathbb{N}}$ converges to $(0, 0)$ for the weak topology of $D(A^{1/2})\times X$. By continuity of the flow with respect to initial data, it follows that the sequence $(u_n)_{n\in\mathbb{N}}$ of corresponding solutions converges to 0 for the weak topology of $H^1((0, T)\times\Omega)$; in particular, it is bounded.

Up to a subsequence (still denoted $(u_n)_{n\in\mathbb{N}}$ in what follows), according to Proposition A.1 in the Appendix, there exists a microlocal defect measure μ on the cosphere quotient space $S^*\hat{\Sigma}$ introduced in Section 1C such that

$$(Ru_n, u_n) \to \langle\mu, \kappa(R)\rangle \quad \text{as } n \to +\infty \tag{15}$$

for every $R \in \Psi^0(Y)$ with $\kappa(R)$ to be understood as a continuous function on $S^*\hat{\Sigma}$.

It follows from (14) that μ vanishes in $j(T^*Q)\cap S^*\hat{\Sigma}$. As is well known, the measure μ is invariant[3] under the compressed generalized bicharacteristic flow [Lebeau 1996; Burq and Lebeau 2001]. The definition of this flow is recalled in Section 1C1.

The t-GCC assumption for Q then implies that μ vanishes identically (see [Bardos et al. 1992; Burq and Gérard 1997]). This precisely means that $(u_n)_{n\in\mathbb{N}}$ strongly converges to 0 in $H^1((0, T)\times\Omega)$.

Now, we let $0 < t_1 < t_2 < T$. The above strong convergence implies that

$$\int_{t_1}^{t_2} E_0(u_n)(t)\, dt \to 0 \quad \text{as } n \to +\infty,$$

[3]The theorem of propagation for measures is proven in [Lebeau 1996] for a damped wave equation with Dirichlet boundary condition. We claim that the same proof applies with Neumann boundary condition. First, using the notations and result numbers in that paper, we still have $\mu_\partial = 0$ in the Neumann case, where μ_∂ is defined by (A.18). Then, the proof of Theorem A.1 about the propagation of the measure still applies in the Neumann case: First, one can assume that the tangential operator Q_0 that appears in (A.28) satisfies $\partial_x Q_0|_{x=0} = 0$, to assure that $Q_0 u$ still satisfies the Neumann boundary condition. Then inequality (A.29) holds true as well, since the theorem of propagation of Melrose and Sjöstrand [1978; 1982] holds true in the Neumann case. Finally, estimate (A.33) of [Lebeau 1996] remains valid since the energy estimate holds true in the Neumann case.

with the energy E_0 defined in (9). As this energy is preserved (with respect to time t), this implies that

$$E_0(u_n)(0) = \tfrac{1}{2}\big(\|u_n^0\|^2_{D(A^{1/2})} + \|u_n^1\|^2_X\big) \to 0 \quad \text{as } n \to +\infty,$$

yielding a contradiction. □

We define the set of *invisible solutions* as

$$N_T = \big\{ v \in H^1((0,T)\times\Omega) \;\big|\; v \text{ is a solution of (3)–(4)}, \\ \text{with } v_{|t=0} \in D(A^{1/2}),\ \partial_t v_{|t=0} \in X \text{ and } \chi_Q\, \partial_t v = 0 \big\}, \quad (16)$$

equipped with the norm $\|v\|^2_{N_T} = \|v_{|t=0}\|^2_{D(A^{1/2})} + \|\partial_t v_{|t=0}\|^2_X$. Clearly, N_T is closed.

Lemma 2.3. *We have $N_T = \{0\}$.*

In other words, due to the t-GCC assumption, there is no nontrivial invisible solution.

Proof. First, the t-GCC assumption combined with the propagation of singularities along the generalized bicharacteristic flow (see [Hörmander 1985, Theorem 24.5.3]) implies that all elements of N_T are smooth functions on $(0,T)\times\Omega$, up to the boundary. In particular, if $v \in N_T$ then $\partial_t v \in N_T$.

Second, we remark that, since the operator $\partial_t^2 - \triangle_g$ is time-independent (as well as the boundary condition), the space N_T is invariant under the action of the operator ∂_t.

Third, applying the weak observability inequality of Lemma 2.1 gives

$$C\|v\|^2_{N_T} = C\|(v_{|t=0}, \partial_t v_{|t=0})\|_{D(A^{1/2})\times X} \leqslant \|(v_{|t=0}, \partial_t v_{|t=0})\|_{X\times D(A^{1/2})'}$$

for every $v \in N_T$. Since $D(A^{1/2}) \times (D(A^{1/2}))'$ is compactly embedded into $X \times D(A^{1/2})'$, this implies that the unit ball of N_T is compact and thus N_T is finite-dimensional.

We are now in a position to prove the lemma. The proof goes by contradiction. Let us assume that $N_T \neq \{0\}$. The operator $\partial_t : N_T \to N_T$ has at least one (complex) eigenvalue λ, associated with an eigenfunction $v \in N_T \setminus \{0\}$. Since $\partial_t v = \lambda v$, it follows that $v(t,x) = e^{\lambda t} w(x)$, and since $(\partial_t^2 - \triangle_g)v = 0$ we obtain $(\lambda^2 - \triangle_g)w = 0$. Note that $\lambda \neq 0$ (in the Neumann case, we have $w \in L^2_0(\Omega)$). Now, take any $t \in (0,T)$ such that $\omega(t) = \{x \in \Omega \mid (t,x) \in Q\} \neq \varnothing$. Since $\chi_Q\, \partial_t v = 0$ and thus $\chi_Q v = 0$, it follows that $w = 0$ on the open set $\omega(t)$. By elliptic unique continuation we then infer that $w = 0$ on the whole Ω, and hence $v = 0$. This is a contradiction. □

Let us finally derive the observability inequality (5). To this aim, the compact term on the right-hand side of (11) must be removed. We argue again by contradiction, assuming that there exists a sequence $(u_n^0, u_n^1)_{n\in\mathbb{N}}$ in $D(A^{1/2}) \times X$ such that

$$\|(u_n^0, u_n^1)\|_{D(A^{1/2})\times X} = 1 \qquad \forall n \in \mathbb{N}, \tag{17}$$

$$\|\chi_Q\, \partial_t u_n\|_{L^2((0,T)\times\Omega)} \to 0 \quad \text{as } n \to +\infty, \tag{18}$$

where u_n is the solution of (3)–(4) such that $u_{n\,|t=0} = u_n^0$ and $\partial_t u_{n|t=0} = u_n^1$. From (17), the sequence $(u_n^0, u_n^1)_{n\in\mathbb{N}}$ is bounded in $D(A^{1/2}) \times X$, and therefore, extracting if necessary a subsequence, it converges to some $(u^0, u^1) \in D(A^{1/2}) \times X$ for the weak topology. Let u be the solution of (3)–(4) such that $u_{|t=0} = u^0$ and $\partial_t u_{|t=0} = u^1$. Then, $\chi_Q\, \partial_t u_n \to \chi_Q\, \partial_t u$ weakly in $L^2((0,T)\times\Omega)$ implying

$$\|\chi_Q\, \partial_t u\|_{L^2((0,T)\times\Omega)} \leqslant \liminf_{n\to+\infty} \|\chi_Q\, \partial_t u_n\|_{L^2((0,T)\times\Omega)} = 0,$$

and hence $u \in N_T$. It follows from Lemma 2.3 that $u = 0$. In particular, we have then $(u^0, u^1) = (0, 0)$ and hence $(u_n^0, u_n^1)_{n\in\mathbb{N}}$ converges to $(0, 0)$ for the weak topology of $D(A^{1/2}) \times X$, and thus, by compact embedding, for the strong topology of $X \times D(A^{1/2})'$. Applying the weak observability inequality (11) raises a contradiction. This concludes the proof of Theorem 1.8. □

2B. ***Proof of Proposition 1.2.*** First, we assume that the observability inequality (5) holds. Let v be a solution of (3)–(4), with initial conditions $(v^0, v^1) \in X \times D(A^{1/2})'$. We set $u = \int_0^t v(s)\, ds - A^{-1}v^1$. Then $\partial_t u = v$ and we have $u_{|t=0} = u^0 = -A^{-1}v^1 \in D(A^{1/2})$, and $\partial_t u_{|t=0} = u^1 = v^0 \in X$. Moreover, we have

$$\partial_t^2 u(t) = \partial_t v(t) = \int_0^t \partial_t^2 v(s)\, ds + \partial_t v(0) = -\int_0^t Av(s)\, ds + v^1 = -Au(t).$$

Since $v = \partial_t u$ and $\|(u^0, u^1)\|_{D(A^{1/2})\times X} = \|(A^{-1}v^1, v^0)\|_{D(A^{1/2})\times X} = \|(v^0, v^1)\|_{X\times D(A^{1/2})'}$, applying the observability inequality (5) to u, we obtain (6).

Second, we assume that the observability inequality (6) holds. Let u be a solution of (3)–(4), with initial conditions $(u^0, u^1) \in D(A^{1/2}) \times X$. We set $v = \partial_t u$. Then v is clearly a solution of (3)–(4), with $v_{|t=0} = v^0 = u^1 \in X$ and

$$\partial_t v_{|t=0} = v^1 = \partial_t^2 u_{|t=0} = -Au_{|t=0} = -Au^0 \in D(A^{1/2})'.$$

Since

$$\|(v^0, v^1)\|_{X\times D(A^{1/2})'} = \|(u^1, Au^0)\|_{X\times D(A^{1/2})'} = \|(u^0, u^1)\|_{D(A^{1/2})\times X},$$

applying the observability inequality (6) to $v = \partial_t u$, we obtain (5). □

2C. ***Proof of Corollary 1.14.*** It is proven in [Haraux 1989] that a second-order linear equation with (bounded) damping has the exponential energy decay property if and only if the corresponding conservative linear equation is observable. The extension to our framework is straightforward. We however give a proof of Corollary 1.14 for completeness.

By Theorem 1.8, there exists $C_0 > 0$ such that

$$C_0\big(\|A^{1/2}\phi_{|t=0}\|^2_{L^2(\Omega)} + \|\partial_t\phi_{|t=0}\|^2_{L^2(\Omega)}\big) \leqslant \int_0^S \|\partial_t\phi\|^2_{L^2(\omega(t))}\, dt, \quad \text{where } S = \ell T,\ \ell \in \mathbb{N}^*, \tag{19}$$

for ϕ solution of $\partial_t^2\phi = \triangle_g\phi$, with $(\phi_{|t=0}, \partial_t\phi_{|t=0}) \in D(A^{1/2}) \times X$.

Let now u be a solution of (8) with $(u_{|t=0}, \partial_t u_{|t=0}) \in D(A^{1/2}) \times X$. Let us prove that we have an exponential decay for its energy. We consider ϕ as above with the initial conditions $\phi_{|t=0} = u_{|t=0}$ and $\partial_t\phi_{|t=0} = \partial_t u_{|t=0}$. Then, setting $\theta = u - \phi$, we have

$$\partial_t^2\theta - \triangle_g\theta = \chi_\omega\, \partial_t u, \quad \theta_{|t=0} = 0, \ \ \partial_t\theta_{|t=0} = 0. \tag{20}$$

Observe that $\partial_t u \in L^2(\mathbb{R}\times\Omega)$ yielding $\theta \in \mathscr{C}^0(\mathbb{R}; D(A^{1/2})) \cap \mathscr{C}^1(\mathbb{R}; X)$.

Replacing the right-hand side of (20) by f in $H^1(\mathbb{R}\times\Omega)$, we have $\theta\in\mathscr{C}^0(\mathbb{R};\,D(A))\cap\mathscr{C}^1(\mathbb{R};\,D(A^{1/2}))$. Recalling the definition of E_0 in the statement of Corollary 1.14, we find

$$\frac{d}{dt}E_0(\theta)(t)=\langle\partial_t\theta(t),\,A\theta(t)+\partial_t^2\theta(t)\rangle_{L^2(\Omega)}=\langle\partial_t\theta(t),\,f\rangle_{L^2(\Omega)}.$$

Continuity with respect to f and a density argument then yield $\frac{d}{dt}E_0(\theta)(t)=\langle\partial_t\theta(t),\,\chi_\omega\,\partial_t u\rangle_{L^2(\Omega)}$. With two integrations with respect to $t\in(0,S)$, using that $E_0(\theta)(0)=0$, we obtain, by the Fubini theorem,

$$\int_0^S E_0(\theta)(t)\,dt=\int_0^S(S-t)\int_{\omega(t)}\partial_t\theta(t,x)\,\partial_t u(t,x)\,dx_g\,dt.$$

With the Young inequality, we have

$$\int_0^S E_0(\theta)(t)\,dt\leqslant S^2\int_0^S\|\partial_t u\|^2_{L^2(\omega(t))}\,dt+\frac{1}{4}\int_0^S\|\partial_t\theta\|^2_{L^2(\Omega)}\,dt.$$

With the definition of $E_0(\theta)(t)$, we then infer that

$$\int_0^S\|\partial_t\theta\|^2_{L^2(\Omega)}\,dt\leqslant 4S^2\int_0^S\|\partial_t u\|^2_{L^2(\omega(t))}\,dt. \tag{21}$$

Now, since $\phi=u-\theta$, we have $\|\partial_t\phi\|^2_{L^2(\omega(t))}\leqslant 2\|\partial_t u\|^2_{L^2(\omega(t))}+2\|\partial_t\theta\|^2_{L^2(\Omega)}$, yielding, using (21),

$$\int_0^S\|\partial_t\phi\|^2_{L^2(\omega(t))}\,dt\leqslant(2+8S^2)\int_0^S\|\partial_t u\|^2_{L^2(\omega(t))}\,dt. \tag{22}$$

Arguing as above, we have $\frac{d}{dt}E_0(u)(t)=-\|\partial_t u(t,x)\|^2_{L^2(\omega(t))}$. Using this property, inequalities (19) and (22), and the fact that $\phi_{|t=0}=u_{|t=0}$ and $\partial_t\phi_{|t=0}=\partial_t u_{|t=0}$, we deduce that

$$C_0E_0(u)(0)=C_0E_0(\phi)(0)\leqslant(2+8S^2)\big(E_0(u)(0)-E_0(u)(S)\big),$$

or rather $E_0(u)(S)\leqslant(1-C_0/(2+8S^2))E_0(u)(0)$. For S chosen sufficiently large, that is, for $\ell\in\mathbb{N}^*$ chosen sufficiently large, we thus have $E_0(u)(S)\leqslant\alpha E_0(u)(0)$ with $0<\alpha<1$.

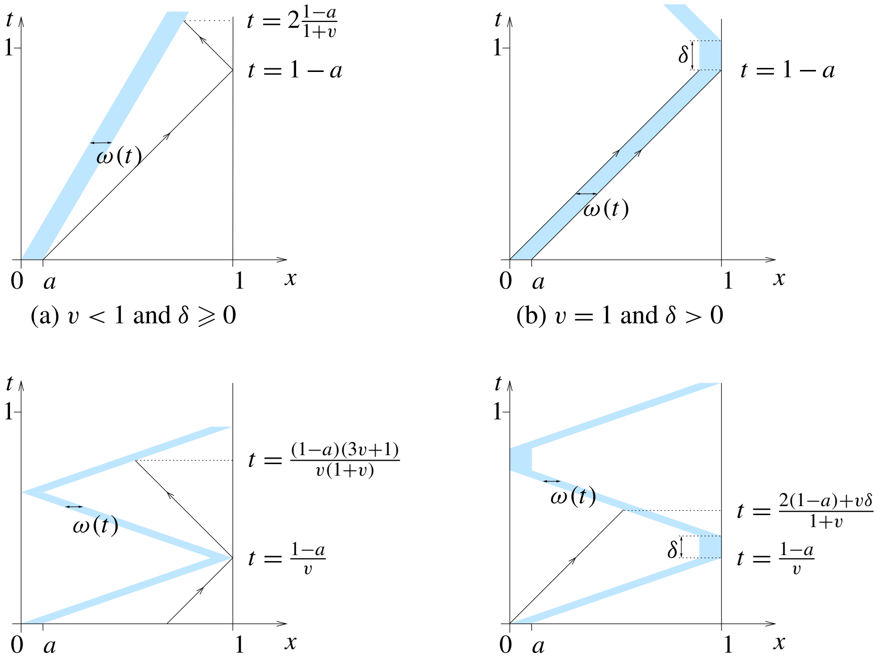
Since ω is T-periodic and thus S-periodic, the above reasoning can be done on any interval $(kS,(k+1)S)$, yielding $E_0(u)((k+1)S)\leqslant\alpha E_0(u)(kS)$ for every $k\in\mathbb{N}$. Hence, we obtain $E_0(u)(kS)\leqslant\alpha^kE_0(u)(0)$.

For every $t\in[kS,(k+1)S)$, noting that $k=[t/S]>t/S-1$, and that $\log(\alpha)<0$, it follows that $\alpha^k<(1/\alpha)\exp(\ln(\alpha)/St)$ and hence $E_0(u)(t)\leqslant E_0(u)(S)\leqslant\mu\exp(-\nu t)E_0(u)(0)$ for some positive constants μ and ν that are independent of u. □

3. Some examples and counterexamples

In the forthcoming examples, we shall consider several geometries in which the observation (or control) domain $\omega(t)=\{x\in\Omega\mid(t,x)\in Q\}$ is the rigid displacement in Ω of a fixed domain, with velocity v. Then the resulting observability property depends on the value of v with respect to the wave speed.

In all our examples, in the presence of a boundary we shall consider Dirichlet boundary conditions. In that case, generalized bicharacteristics behave as described in Section 1C1. We recall that, if parametrized by time t, the projections of the generalized bicharacteristics on the base manifold travel at speed 1.



(a) $v < 1$ and $\delta \geqslant 0$ (b) $v = 1$ and $\delta > 0$

(c) $v \geqslant 1$ and $\delta = 0$ (d) $v > 1$ and $\delta > 0$

Figure 3. Time-varying domains in dimension 1.

3A. ***In dimension 1.*** We consider $M = \mathbb{R}$ (Euclidean) and $\Omega = (0, 1)$. The rays have a speed equal to 1. We set $I = (0, a)$ for some fixed $a \in (0, 1)$, and we assume that the control domain $\omega(t)$ is equal to the translation of the interval I with fixed speed $v > 0$. We have, then, $\omega(t) = (vt, vt + a)$ as long as $t \in (0, (1-a)/v)$. When $\omega(t)$ touches the boundary, we assume that it is "reflected" after a time-delay $\delta \geqslant 0$ according to the following rule: if $(1-a)/v \leqslant t \leqslant (1-a)/v + \delta$ then $\omega(t) = (1-a, 1)$. For larger times $t \geqslant (1-a)/v + \delta$ (and before the second reflection), the set $\omega(t)$ moves in the opposite direction with the same speed (see Figure 3).

This simple example is of interest as it exhibits that the control time depends on the value of the velocity v with respect to the wave speed (which is equal to 1 here). We denote by $T_0(v, a, \delta)$ the control time. With simple computations (see also Figure 3), we establish that

$$T_0(v, a, \delta) = \begin{cases} 2(1-a)/(1+v) & \text{if } 0 \leqslant v < 1 \text{ and } \delta \geqslant 0, \\ 1-a & \text{if } v = 1 \text{ and } \delta > 0, \\ (1-a)(3v+1)/(v(1+v)) & \text{if } \ v \geqslant 1 \text{ and } \delta = 0, \\ (2(1-a)+v\delta)(1+v) & \text{if } v > 1 \text{ and } \delta > 0. \end{cases}$$

Remark 3.1. Note that the control time $T_0(v, a, \delta)$ is discontinuous in v and δ. The control time is not continuous with respect to the domain Q as already mentioned in Remark 1.7.

3B. ***A moving domain on a sphere.*** Let $M = \Omega = S^2$, the unit sphere of $\mathbb{R}^3$, be endowed with the metric induced by the Euclidean metric of $\mathbb{R}^3$. Let us consider spherical coordinates (θ, φ) on M, in which

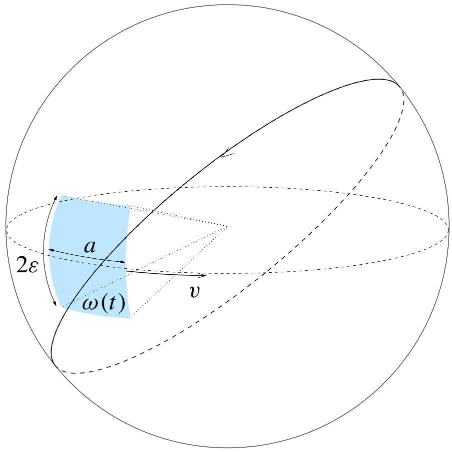


Figure 4. A time-varying domain on the unit sphere and a typical ray (great circle).

$\varphi = 0$ represents the horizontal plane (latitude zero), and θ is the angle describing the longitude along the equator. Let $a \in (0, 2\pi)$ and $\varepsilon \in (0, \pi/2)$ be arbitrary. For $v > 0$, we set

$$\omega(t) = \{(\theta, \varphi) \mid |\varphi| < \varepsilon,\ \ vt < \theta < vt + a\}$$

for every $t \in \mathbb{R}$. The set $\omega(t)$ is a spherical square drawn on the unit sphere, with angular length equal to 2ε in latitude, and a in longitude, and moving along the equator with speed equal to v (see Figure 4). We denote by $T_0(v, a, \varepsilon)$ the control time as defined in Section 1C2.

For this example, an important fact is the following: every (geodesic) ray on the sphere propagates at speed 1 along a great circle, with half-period π. We thus have a simple description of all possible rays. Note that, as the radius is 1, the speed coincides with the angular speed.

Proposition 3.2. *Let $a \in (0, 2\pi)$ and $\varepsilon \in (0, 1)$ be arbitrary. Then $T_0(v, a, \varepsilon) < +\infty$ except for a finite number of critical speeds $v > 0$. Moreover*:

- *$T_0(v, a, \varepsilon) \sim (\pi - a)/v$ as $v \to 0$.*
- *If $v > v_1 = (2\pi - a + 2\varepsilon)/(2\varepsilon)$ then $T_0(v, a, \varepsilon) < \infty$. If $v \to +\infty$ then $T_0(v, a, \varepsilon) \to \pi - 2\varepsilon$.*

Besides, if $v \in \mathbb{Q}$, then there exist $a_0 > 0$ and $\varepsilon_0 > 0$ such that $T_0(v, a, \varepsilon) = +\infty$ for every $a \in (0, a_0)$ and every $\varepsilon \in (0, \varepsilon_0)$.

Obtaining an analytic expression of T_0 as a function of (v, a, ε) seems to be very difficult.

Note that the asymptotics above still make sense if either $a > 0$ or $\varepsilon > 0$ are small. This shows that we can realize the observability property with a subset of arbitrary small Lebesgue measure (compare with Corollary 1.12).

Proof of Proposition 3.2. First, we observe the following. Consider a ray propagating along the equator (with angular speed 1), in the same direction as $\omega(t)$. If $v = 1$, depending on its initial condition, this ray either never enters $\omega(t)$ or remains in it for all time. Hence, $T_0(v, a, \varepsilon) = +\infty$. In contrast, if $v \neq 1$, then, such a ray enters $\omega(t)$ for a time $0 \leqslant t < 2\pi/|v - 1|$, as $2\pi/|v - 1| > (2\pi - a)/|v - 1|$.

Second, we treat the cases v large and v small, and we compute the asymptotics of $T_0(v, a, \varepsilon)$ (in the argument, both a and ε are kept fixed).

Case $\boldsymbol{v}$ small: If $2\pi v < a$, then every ray goes full circle in a time shorter than that it takes for the domain to travel the distance a. It is then clear that every ray will have met $\omega(t)$ as soon as $\omega(t)$ has travelled halfway way along the equator (up to the thickness of $\omega(t)$ and the travel time of the ray itself). In other words, we then have $(\pi - a)/v \leqslant T_0(v, a, \varepsilon) \leqslant (\pi - a)/v + 2\pi$ for $v < v_0 = a/(2\pi)$.

Case $\boldsymbol{v}$ large: If v grows to infinity, then the situation becomes intuitively as if we have a static control domain forming a strip of constant width $\varepsilon > 0$ around the equator. For such a strip, the control time is $\pi - 2\varepsilon$. More precisely, let us assume $(2\pi - a + 2\varepsilon)/v < 2\varepsilon$. Every ray entering the region $\{|\varphi| < \varepsilon\}$ spends a time at least equal to 2ε in this region. At worst, the control domain will have to travel the distance $2\pi - a + 2\varepsilon$ to "catch" this ray (going full circle and more than the longitudinal distance travelled by the ray itself). The condition $v > v_1 = (2\pi - a + 2\varepsilon)/(2\varepsilon)$ thus implies that all rays enter the moving open domain $\omega(t)$ within time $\pi - 2\varepsilon + 2(\pi + \varepsilon)/v$. Hence, $\pi - 2\varepsilon \leqslant T_0(v, a, \varepsilon) \leqslant \pi - 2\varepsilon + 2(\pi + \varepsilon)/v$.

Third, we consider the case $v_0 \leqslant v \leqslant v_1$. To get some intuition, we consider, in a first step, that a and ε are both very small, and thus consider $\omega(t)$ as point moving along the equator. According to the first observation we made above, let us consider a ray propagating along a great circle that is transversal to the equator. It meets the equator at times $t_k = t_0 + k\pi,\ k \in \mathbb{Z}$, for some t_0. If v is irrational then the set of positions of the "points" $\omega(t_k)$, given by $x(t_k) = \cos(vt_k)$ and $y(t) = \sin(vt_k)$ in the plane (x, y) containing the equator, is dense in the equator. Adding some thickness to $\omega(t)$, that is, having $a > 0$ and $\varepsilon > 0$, we find that every ray encounters the moving open set $\omega(t)$ in a finite time if v is irrational. By a compactness argument we then obtain $T_0(v, a, \varepsilon) < \infty$ if v is irrational.

Fourth, considering again that $a > 0$ and $\varepsilon > 0$ are both very small, we shall now see that there do exist rays, transversal to the equator, that never meet the moving "point" $\omega(t)$ whenever $v \in \mathbb{Q}$. Writing $v = p/q$ with p and q positive integers, the set of points reached by $(\cos(vt_k), \sin(vt_k))$ at times $t_k = t_0 + k\pi$, with $k \in \mathbb{Z}$, is finite. The following lemma yields a more precise statement.

Lemma 3.3. *Let p and q be two coprime integers. We have*

$$\{kp\pi/q \bmod 2\pi \mid k = 1, \dots, 2q\} = \begin{cases} \{k\pi/q \mid k = 1, \dots, 2q\} & \text{if } p \text{ is odd}, \\ \{2k\pi/q \mid k = 1, \dots, q\} & \text{if } p \text{ is even (and } q \text{ odd)}. \end{cases}$$

Thus, if $v = p/q$, with p and q coprime integers, the points $(\cos(vt_k), \sin(vt_k))$ form the vertices of a regular polygon in the disk. There are exactly $2q$ (resp. q) such vertices if p is odd (resp. even). In this situation, it is always possible to find a ray transversal to the equator that never meets this set of vertices. Now, this phenomenon persists in the case $a > 0$ and $\varepsilon > 0$ if both are chosen sufficiently small. We have thus proven that, given $v \in \mathbb{Q} \cap [v_0, v_1]$, there exist $0 < a_0 < 2\pi$ and $0 < \varepsilon_0 < \pi/2$ such that $T_0(v, a, \varepsilon) = +\infty$ for all $a \in (0, a_0)$ and $\varepsilon \in (0, \varepsilon_0)$. Note also that if $a > 2\pi/q$ and $\varepsilon > 0$ then every ray meets $\omega(t)$ in some finite time. By a compactness argument we then obtain $T_0(v, a, \varepsilon) < \infty$. From that last observation, we infer that, given $a > 0$ and $\varepsilon > 0$ fixed, the set of rational velocities $v \in (v_0, v_1) \cap \mathbb{Q}$ for which $T_0(v, a, \varepsilon) = +\infty$ is finite. $\square$

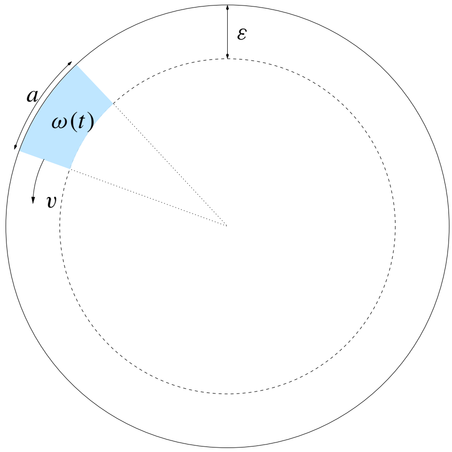


Figure 5. A time-varying domain on the unit disk.

Proof of Lemma 3.3. We note that $\{kp\pi/q \mod 2\pi \mid k = 1, \dots, 2q\} = \{kp\pi/q \mod 2\pi \mid k \in \mathbb{Z}\}$. It thus suffices to prove the following two statements:

(1) For p even: $\forall k' \in \{1, \dots, q\}$, $\exists k \in \mathbb{Z}$, $\exists m \in \mathbb{Z}$ such that $2k' = kp + 2mq$.

(2) For p odd: $\forall k' \in \{1, \dots, 2q\}$, $\exists k \in \mathbb{Z}$, $\exists m \in \mathbb{Z}$ such that $k' = kp + 2mq$.

Since p and q are coprime, there exists $(a, b) \in \mathbb{Z}^2$ such that $ap + bq = 1$. Moreover, if (a, b) is a solution of that diophantine equation, then all other solutions are given by $(a + qn, b - pn)$, with $n \in \mathbb{Z}$. Multiplying by $2k'$, we infer that $2k' = 2k'ap + 2k'bq$, and the first statement above follows. For the second statement, we note that, if p is odd, then, changing b into $b - pn$ if necessary, we may assume that b is even, say $b = 2b'$. Then, multiplying by k', we infer that $k' = k'ap + 2k'b'q$, and the second statement follows. □

Before moving on to the next example, we stress again that the peculiarity of the present example (unit sphere) is that all rays are periodic, with the same period 2π. The study of other Zoll manifolds would be of interest. The situation turns out to be drastically different in the case of a disk, due to "secular effects" implying a precession phenomenon, as we are now going to describe.

3C. *A moving domain near the boundary of the unit disk.* Let $M = \mathbb{R}^2$ (Euclidean) and let $\Omega = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$ be the unit disk. Let $a \in (0, 2\pi)$ and $\varepsilon \in (0, 1)$ be arbitrary. We set, in polar coordinates,

$$\omega(t) = \left\{(r, \theta) \in [0, 1] \times \mathbb{R} \mid 1 - \varepsilon < r < 1,\ vt < \theta < vt + a\right\}$$

for every $t \in \mathbb{R}$. The time-dependent set $\omega(t)$ moves at constant angular speed v, anticlockwise, along the boundary of the disk (see Figure 5). Its radial length is ε and its angular length is a.

Proposition 3.4. *The following properties hold*:

(1) *Let $a \in (0, 2\pi)$ and $\varepsilon \in (0, 1)$ be arbitrary. We have $T_0(v, a, \varepsilon) < +\infty$ for every $v > v_0 = (2\pi + 2\varepsilon - a)/(2\varepsilon)$, and we have $T_0(v, a, \varepsilon) \sim 2 - 2\varepsilon$ as $v \to +\infty$.*

(2) *If there exists* $n \in \mathbb{N} \setminus \{0, 1\}$ *such that* $v \sin(\pi/n) \in \pi\mathbb{Q}$, *then there exist* $a_0 \in (0, 2\pi)$ *and* $\varepsilon_0 \in (0, 1)$ *such that* $T_0(v, a, \varepsilon) = +\infty$ *for all* $a \in (0, a_0)$ *and* $\varepsilon \in (0, \varepsilon_0)$.

(3) *For every* $v \geqslant 1$, *for every* $a \in (0, 2\pi)$, *there exists* $\varepsilon_0 > 0$ *such that* $T_0(v, a, \varepsilon) = +\infty$ *for every* $\varepsilon \in (0, \varepsilon_0)$.

(4) *For every* $v \geqslant 0$ *and* $a \in (0, \pi)$, *there exists* $\varepsilon_0 \in (0, 1)$ *such that* $T_0(v, a, \varepsilon) = +\infty$ *for every* $\varepsilon \in (0, \varepsilon_0)$.

As for the case of the sphere presented in Section 3B, obtaining an analytic expression of T_0 as a function of (v, a, ε) seems a difficult task.

The fact that $T_0(v, a, \varepsilon) = +\infty$ provided that $a > 0$ and $\varepsilon > 0$ are chosen sufficiently small is in strong contrast with the case of the sphere. This is due to the fact that, in the disk, the structure of the rays is much more complex: there are large families of periodic and almost-periodic rays. The ray drawn in Figure 7 produces some sort of "secular effect", itself implying a precession whose speed can be tuned to coincide with the speed v of $\omega(t)$, provided $v \geqslant 1$. We shall use this property in the proof.

We stress that for the third property we do not need to assume that a is small. Actually, a is any element of $(0, 2\pi)$. If a is close to 2π, then $\omega(t)$ is almost a ring located at the boundary, moving with angular speed v, with a "hole". As this hole moves around with speed v there is a ray that periodically hits the boundary and reflects from it exactly at the hole position.

Proof of Proposition 3.4. For $a \in (0, 2\pi)$ and $\varepsilon \in (0, 1)$ fixed, if v is very large, then the situation gets close to that of a static control domain which is a ring of width ε, located at the boundary of the disk. For this static domain the control time is $2 - 2\varepsilon$. In fact, all rays enter the region $\Omega_\varepsilon = \{1 - \varepsilon < r < 1\}$ and the shortest time spent there is 2ε. During such time the angular distance travelled by the ray is less than 2ε. Hence, if $(2\pi + 2\varepsilon - a)/v < 2\varepsilon$, one knows for sure that the ray will be "caught" by the moving open set $\omega(t)$ before it leaves Ω_ε. Thus, for $v > v_0 = (2\pi + 2\varepsilon - a)/(2\varepsilon)$ all rays enter $\omega(t)$ in finite time. Moreover, we have $2 - 2\varepsilon \leqslant T_0(v, a, \varepsilon) \leqslant 2 - 2\varepsilon + (2\pi + 2\varepsilon)/v$. This yields the announced asymptotics for $T_0(v, a, \varepsilon)$.

Let us now investigate the three cases where $T_0(v, a, \varepsilon) = +\infty$ as stated in the proposition. For the sake of intuition, it is simpler to first assume that $\varepsilon > 0$ and $a > 0$ are very small, and hence, that $\omega(t)$ is close to being a point moving along the boundary of the disk, given by $(\cos(vt), \sin(vt))$.

Let us then consider, as illustrated in Figure 6, *periodic* rays propagating "anticlockwise" in the disk (with speed equal to 1), and reflecting at the boundary of the disk according to Section 1C1, that is, according to geometrical optics. The trajectory of such rays forms a regular polygon with vertices at the boundary of the unit disk. Let $n \geqslant 2$ be the number of vertices. For $n = 2$, the ray travels along a diameter of the disk, and passes through the origin; it is 4-periodic. For $n = 3$, the trajectory of the ray forms an equilateral triangle centered at the origin; etc. The length of an edge of such a regular polygon with $n \geqslant 2$ vertices is equal to $2\sin(\pi/n)$. This means that there exists $t_0 \in \mathbb{R}$ such that this ray reaches the boundary at times $t_k = t_0 + 2k\sin(\pi/n)$. Hence, if $2v\sin(\pi/n) = (2p\pi)/q$ with p and q positive integers, then the moving point $(\cos(vt), \sin(vt))$, taken at times t_k ranges over a finite number of points of $\partial\Omega$. Therefore, there exists a periodic ray with n vertices never meeting $\omega(t)$. This property remains

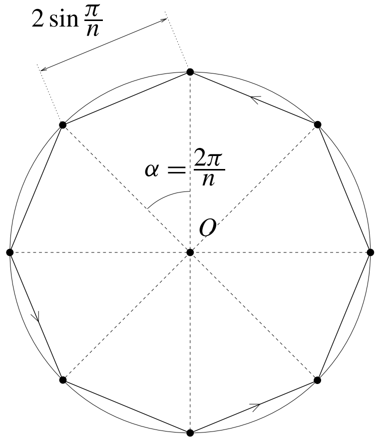


Figure 6. A periodic ray yielding a regular polygon.

clearly true for values of $a > 0$ and of $\varepsilon > 0$ chosen sufficiently small. This show the second statement of the proposition.

Let us now consider a ray propagating in the disk, as drawn in Figure 7, and reflecting at the boundary at consecutive points P_k, $k \in \mathbb{N}$. Denote by O the center of the disk, and by α the oriented angle $\widehat{P_0 O P_1}$. If $0 < \alpha < \pi$, then the ray appears to be going “anticlockwise” as in Figure 7 (left); if $\alpha = \pi$ then the ray bounces back and forth on a diameter of the disk; if $\pi < \alpha < 2\pi$, then the ray appears to be going “clockwise” as in Figure 7 (right).

In any case, the distance P_0P_1, and more generally P_kP_{k+1}, is equal to $2\sin(\alpha/2)$. Since the speed of the ray is equal to 1, the ray starting from P_0 at time $t = 0$ reaches the point P_1 at time $2\sin(\alpha/2)$, the point

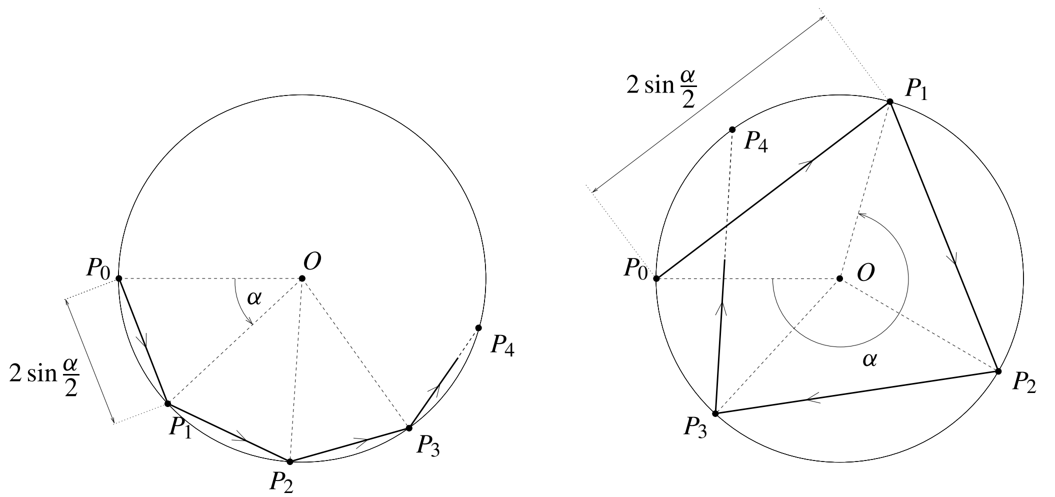


Figure 7. Rays propagating “anticlockwise”“, $0 < \alpha \leqslant \pi$, (left) and “clockwise”“, $\pi \leqslant \alpha < 2\pi$, (right).

P_k at time $t_k = 2k\sin(\alpha/2)$, etc. Let $t \mapsto P(t)$ be the curve propagating *anticlockwise* along the unit circle, with constant angular speed, passing exactly through the points P_k at time t_k. Its angular speed is then

$$w_P(\alpha) = \frac{\alpha}{2\sin(\alpha/2)},$$

and we call it the *precession speed*. This is the speed at which the discrete points P_k propagate "anticlockwise" along the unit circle. Now, if the set $\omega(t)$ has the angular speed $v = w_P(\alpha)$ (for some $\alpha \in (0, 2\pi)$), then there exists rays, as in Figure 7, that never meet $\omega(t)$, if $a \in (0, 2\pi)$, provided that $\varepsilon > 0$ is chosen sufficiently small. Since the function $\alpha \mapsto w_P(\alpha)$ is monotone increasing from $(0, 2\pi)$ to $(1, \infty)$, it follows that, for every $v \in (1, \infty)$, there exists $\alpha \in (0, 2\pi)$ such that $v = w_P(\alpha)$, and therefore $T_0(v, a, \varepsilon) = +\infty$ provided that ε is chosen sufficiently small. For $v = 1$, there exists a gliding ray that never meets $\omega(t)$. This can be seen as the limiting case $\alpha \to 0$, as rays can be concentrated near the gliding ray. We thus have proven the third statement of the proposition.

Now, still working with the configurations drawn in Figure 7 (right), for $\alpha \in (\pi, 2\pi)$, let $P(t)$ be the curve propagating *anticlockwise* along the unit circle, with constant angular speed, passing successively through P_0 at time 0, through P_2 at time $4\sin(\alpha/2)$, and P_{2k} at time $2k\sin(\alpha/2)$. Its angular speed is

$$w_P(\alpha) = \frac{2\alpha - 2\pi}{4\sin(\alpha/2)} = \frac{\alpha - \pi}{2\sin(\alpha/2)}.$$

The function $\alpha \mapsto w_P(\alpha)$ is monotone increasing from $(\pi, 2\pi)$ to $(0, +\infty)$. If the set $\omega(t)$, with $a \in (0, \pi)$ and $\varepsilon > 0$ small, is initially (at time 0) located between the points P_0 and P_0' its diametrically opposite point, and if $v = w_P(\alpha)$, then the ray drawn in Figure 7 (right) never meets $\omega(t)$. This is illustrated in Figure 8. We have thus proven the last statement of the proposition. □

Remark 3.5. (1) It is interesting to note that, even for domains such that a is close to 2π, the t-GCC property fails if $v > 1$ and if ε is chosen too small. This example is striking, because in that case, if the control domain were static, then it would satisfy the usual GCC (this is true as soon as $a > \pi$). This example shows that, when considering a control domain satisfying the GCC, then, when making it move, the t-GCC property may fail. However, this example is a domain moving faster than the actual wave speed. This is rather nonphysical.

(2) For the fourth property of the previous proposition we obtain a moving open set $\omega(t)$ with an "angular measure" that is less than π, that is, $0 < a < \pi$ (see Figure 8). In fact, if one allows for $\omega(t)$ to be not connected, but rather the union of two connected components, for any velocity $v \in (0, +\infty)$ we can exhibit moving sets whose "angular measure" is a close as one wants to 2π and yet the t-GCC does not hold. This is illustrated in Figure 9.

(3) If v cannot be chosen as large as desired (for physical reasons), Proposition 3.4 states that the t-GCC does not hold true if $a > 0$ and $\varepsilon > 0$ are too small. As shown in the proof above, this lack of observability is due to a secular effect caused by geodesics whose trace at the boundary produces a pattern that itself varies in time, with a precession speed that can be tuned to match that of the control domain. In fact, a precession speed can be obtained as slow as one wants if α is chosen such that $\pi < \alpha < \pi + \delta$ with $\delta > 0$ small. This is illustrated in Figure 10 (left). If one is close to regular polygons, as illustrated in

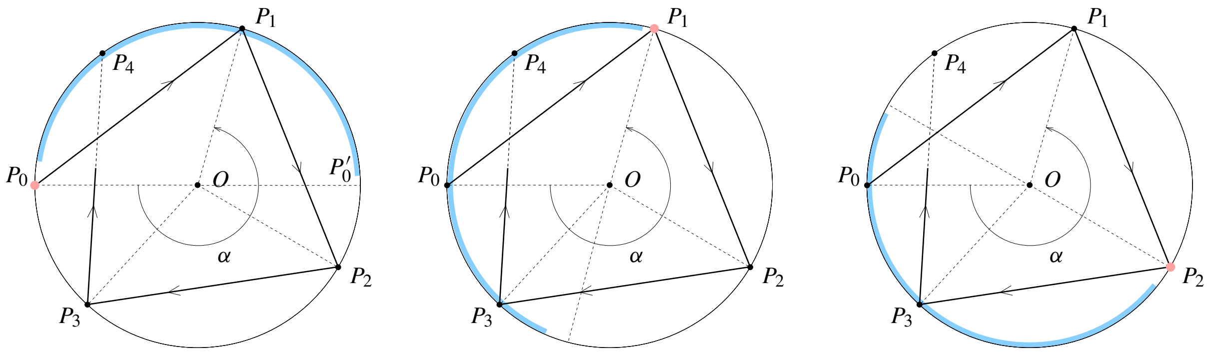


Figure 8. Illustration of Proposition 3.4(4) with $t = 0$ (left), $t = 2\sin(\alpha/2)$ (middle), and $t = 4\sin(\alpha/2)$ (right).

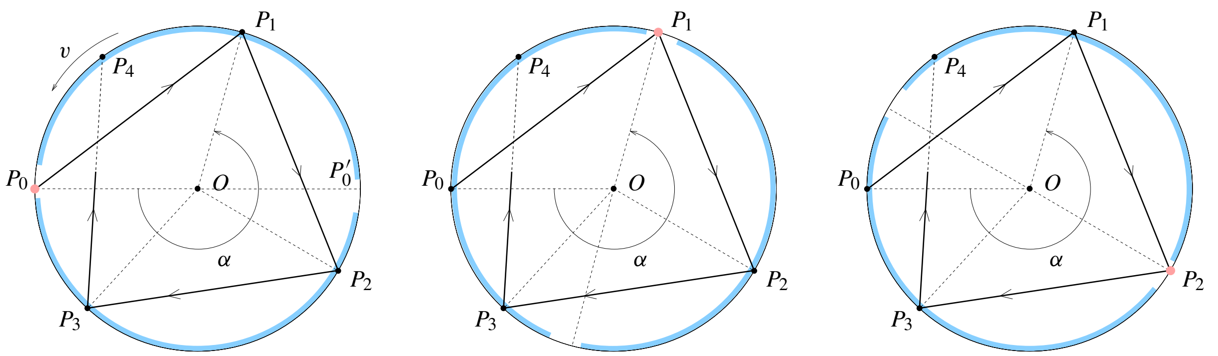


Figure 9. Case of a disconnected moving open set $\omega(t)$ moving anticlockwise, not satisfying the t-GCC with yet a very large "angular" measure with $t = 0$ (left), $t = 2\sin(\alpha/2)$ (middle), and $t = 4\sin(\alpha/2)$ (right).

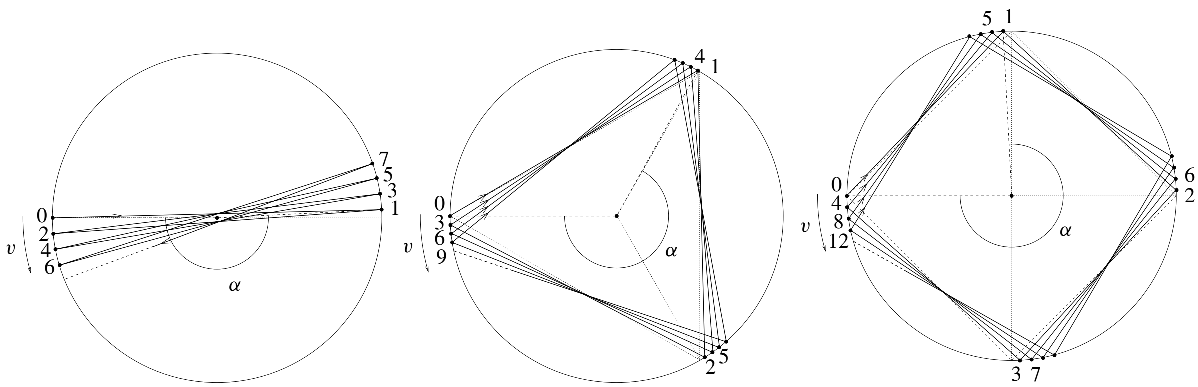


Figure 10. Cases of slow precession speeds with $0 < \alpha - 2\pi(n-1)/n < \delta$, with δ small and $n \geqslant 2$, that is, with a trajectory "close" to that of a periodic ray that forms a regular polygon (see Figure 6): $n = 2$ and $\pi < \alpha < \pi + \delta$ (left), $n = 3$ and $4\pi/3 < \alpha < 4\pi/3 + \delta$ (middle), $n = 4$ and $3\pi/2 < \alpha < 3\pi/2 + \delta$ (right). Above P_n is denoted by n.

Figure 10, one obtains a precession pattern that can be used to deduce other families of examples of moving domains $\omega(t)$ with multiple connected components with velocities $v > 0$ that do not satisfy the t-GCC.

(4) Proposition 3.4 has been established for the domain drawn in Figure 5, sliding anticlockwise along the boundary with a constant angular speed. Other situations can be of interest: we could allow the domain to move with a nonconstant angular speed. For instance, we could allow the domain to move anticlockwise within a certain horizon of time, and then clockwise. This would certainly improve the observability property. A situation that can be much more interesting in view of practical issues is to let the angular speed of the control domain evolve according to $v(t) = v + \beta \sin(\gamma t)$ (with $\beta > 0$ small), that is, with a speed oscillating around a constant value v. We expect that such a configuration, with an appropriate choice of coefficients, will yield the observability property to be more robust, by avoiding the situation described in Proposition 3.4(2) (nonobservability for a dense set of speeds).

We now state a positive result. For $a \in (0, 2\pi)$ and $\varepsilon \in (0, 1)$, assume now that the domain $\omega(t)$ is given by $\omega(t) = \{(r, \theta) \in [0, 1] \times \mathbb{R} \mid 1 - \varepsilon < r < 1,\ \theta_0(t) < \theta < a + \theta_0(t)\}$, with

$$\theta_0(t) = \begin{cases} 0 & \text{if } 0 \leqslant t < t_0, \\ v(t - t_0) & \text{if } t_0 \leqslant t \end{cases}$$

for some $t_0 > 0$ and $v > 0$. We set $Q = \bigcup_{t \geqslant 0} \omega(t)$. In this configuration, at first the domain is still, and then one lets it move.

Proposition 3.6. *If* $4\pi/5 < a < \pi$, $t_0 > 2\pi$, *then there exists* $0 < v < 1$ *such that* $T_0(Q, \Omega) < \infty$.

Remark 3.7. (1) The important aspect of this result lies in the following facts. First, if at rest, the observability set does not satisfy the geometric control condition; hence, its motion is crucial for the t-GCC to hold. Second, the motion is performed at a velocity v that is less than that of the wave speed; we thus have a physically meaningful example.

(2) The result is not optimal as we do not exploit the thickness ε of the domain $\omega(t)$ in the proof.

(3) It would be interesting to further study this "stop-and-go" strategy and see how small the value of $a > 0$ can be chosen.

Proof of Proposition 3.6. Let $0 < t < t_0$; then $\omega(t) = \omega(0)$ is still. First, we consider the ray associated with $0 \leqslant \alpha < a$, or symmetrically $2\pi - a < \alpha \leqslant 2\pi$. Then the movement of the successive points P_k, $k \in \mathbb{N}$, is anticlockwise, or clockwise, respectively; see Figure 7. Depending on the case considered we denote $\beta = \alpha$ or $\beta = 2\pi - \alpha$. The above condition thus reads better as $0 \leqslant \beta < a$. In both cases, the (unsigned) angular distance between two points is precisely β. The successive points P_k, $k \in \mathbb{N}$, thus end up meeting $\omega(0)$ in finite time. (The case $\beta = 0$ coincides with a gliding ray that has angular speed 1; it thus meets $\omega(0)$ in finite time.) Let us consider $\beta \neq 0$. The maximal number of steps it takes for any ray associated with β to enter $\omega(0)$ is then $\lfloor (2\pi - a/\beta) \rfloor + 1$ yielding a maximal time $T_0(\alpha) = 2\sin(\beta/2)(\lfloor (2\pi - a/\beta) \rfloor + 1)$, as the time lapse between two points P_k is $2\sin(\beta/2)$. Here, the notation $\lfloor \cdot \rfloor$ stands for the usual floor function. We thus need $T_0 > \max(2\pi - a, \sup_{0<\beta<a} T_0(\alpha))$. The value $2\pi - a$ accounts for the gliding rays ($\beta = 0$). Here, we give a crude upper bound for $T_0(\alpha)$,

observing that

$$T_0(\alpha) \leqslant \frac{\sin(\beta/2)}{\beta/2}(2\pi - a + \beta) \leqslant 2\pi - a + \beta < 2\pi.$$

We thus see that if we choose $t_0 > 2\pi$ then all the rays associated with the angle $0 \leqslant \beta < a$ enter $\omega(0)$ for $0 \leqslant t < t_0$.

Second, we consider $t \geqslant t_0$ and we are left only with the rays that are associated with $a \leqslant \alpha \leqslant 2\pi - a$. For these rays we consider the two sequences of points $(P_{2k})_k$ and $(P_{2k+1})_k$. The time lapse for a ray to go from one point to the consecutive point in these two sequences is $4\sin(\alpha/2)$ and this is associated with the (signed) angle $2(\alpha - \pi)$. In fact, as $a > \pi/2$, if $a \leqslant \alpha \leqslant \pi$, then both sequences move clockwise and if $\pi \leqslant \alpha \leqslant 2\pi - a$, then both sequences move anticlockwise.

If $v > 0$ is the angular speed of $\omega(t)$ for $t \geqslant t_0$ then we require $2(\alpha - \pi) < 4\sin(\alpha/2)v < a + 2(\alpha - \pi)$, that is,

$$\frac{\alpha - \pi}{2\sin(\alpha/2)} < v < \frac{a + 2(\alpha - \pi)}{4\sin(\alpha/2)} \quad \text{for } a \leqslant \alpha \leqslant 2\pi - a. \tag{23}$$

Observe that $a + 2(\alpha - \pi) > 0$ as $a > \pi/2$. The left inequality in (23) is necessary as it implies that the anticlockwise moving open set $\omega(t)$ will be faster than the two sequences given above, a necessary condition to be able to catch points in those sequences. In particular, this necessary condition is clearly filled if the sequences move clockwise, that is, if $a \leqslant \alpha \leqslant \pi$. The right inequality in (23) expresses that $\omega(t)$ will not turn too fast and then miss the discrete sequences of points. In fact, during a time interval of length $4\sin(\alpha/2)$ the relative angular displacement of the sequence and the moving set $\omega(t)$ is $\ell = 4\sin(\alpha/2)v - 2(\alpha - \pi)$ and with (23) we have $0 < \ell < a$. This expresses that the sequence points cannot be missed.

As both bounds of (23) are increasing functions for $\alpha \in (a, 2\pi - a)$, we obtain the sufficient condition

$$\frac{\pi - a}{2\sin(a/2)} < v < \frac{3a - 2\pi}{4\sin(a/2)}.$$

We see that it can be satisfied if $a > 4\pi/5$. Observing that

$$a \mapsto h(a) = \frac{3a - 2\pi}{4\sin(a/2)}$$

increases on $(0, \pi]$ and $0 < h(\pi) = \pi/4 < 1$ we see that the found admissible velocities are such that $0 < v < 1$. □

3D. ***A moving domain in a square.*** Let $M = \mathbb{R}^2$ (Euclidean) and $\Omega = (0, 1)^2$ be the unit square. We recall that, as discussed in Remark 1.9(2), the statement of Theorem 1.8 is still valid in the square, because the generalized bicharacteristic flow is well defined.

Let $a \in (0, 1)$. We consider the fixed domain $\tilde{\omega}_0 = (-a, a)^2$, a square centered at the origin $(0, 0)$, and we set $\omega_0 = \tilde{\omega}_0 \cap \Omega$ (see Figure 11). Since there are periodic rays, bouncing back and forth between opposite sides of the square, that remain away from ω_0, the GCC does not hold true for ω_0, and the wave equation cannot be observed from the domain ω_0 in the sense of (5).

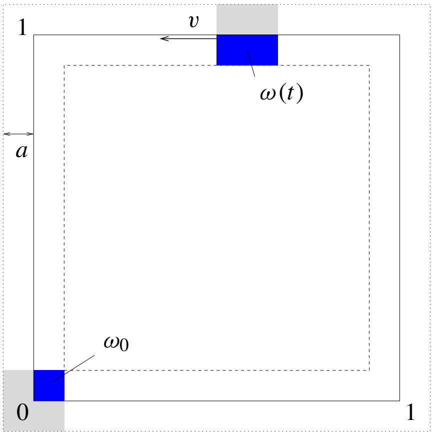


Figure 11. A time-varying domain in the square $(0,1)^2$.

Now, for a given $T>0$, consider a continuous path $t\in[0,T]\mapsto(x(t),y(t))$ in the closed square $[0,1]^2$, with $(x(0),y(0))=(0,0)$. We set

$$\tilde{\omega}(t)=\big(x(t)-a,x(t)+a\big)\times\big(y(t)-a,y(t)+a\big)\quad\text{and}\quad\omega(t)=\tilde{\omega}(t)\cap\Omega.$$

To avoid the occurrence of periodic rays, as described above, that never meet $\omega(t)$, a necessary condition for the t-GCC to hold true is

$$[a,1-a]\subset x([0,T])\quad\text{and}\quad[a,1-a]\subset y([0,T]).$$

Let us assume that the point $(x(t),y(t))$ moves precisely along the boundary of the square $[0,1]^2$, anticlockwise, and with a constant speed v. The path $t\mapsto(x(t),y(t))$ only exhibits singularities when reaching a corner of the square $[0,1]^2$, where the direction of the speed is discontinuous (see Figure 11).

We denote by $T_0(v,a)$ the control time.

Proposition 3.8. *We have the following two results*:

(1) *Let* $a\in\left(0,\frac{1}{2}\right)$ *be arbitrary. For* $v>v_0=(2-a)/a$ *we have* $T_0(v,a)<+\infty$, *and moreover*, $T_0(v,a)\sim\max(\sqrt{2}(1-2a),0)$ *as* $v\to+\infty$.

(2) *If* $v\in\bigcup_{(p,q)\in\mathbb{N}}\sqrt{p^2+q^2}\,\mathbb{Q}$, *there exists* $a_0>0$ *such that* $T_0(v,a)=+\infty$ *if* $a\in(0,a_0)$.

Proof of Proposition 3.8. The argument for the first property is the same as that developed in the other examples. For $v>0$ large, the situation becomes intuitively as if we have a static control domain that forms a a-thick strip along the boundary of the square. For this case, the geometric control time is $\sqrt{2}(1-2a)$ if $a\in\left(0,\frac{1}{2}\right)$ and 0 otherwise. More precisely, if a ray enters this strip, it remains in it at least for a time $2a$. During such time, it travels at most a lateral distance equal to $2a$ (wave speed is 1). If during that time the control domain goes all around and travels also the additional $2a$ distance, we can be sure that this ray will be "caught" by the moving domain. For $v>v_0=(4+2a)/2a=(2-a)/a$, the control time can thus be estimated by $\max(\sqrt{2}(1-2a),0)\leqslant T_0(v,a)\leqslant\max(\sqrt{2}(1-2a),0)+(4+2a)/v$. Hence, the announced asymptotics.

For the second property, as for the other examples, considering a small at first, and thus the set $\omega(t)$ to be a simple point running along the boundary, greatly helps intuition. We start by considering 2-periodic rays that bounce back and forth between two opposite sides of the square. They reflect at boundaries at times $t_k = t_0 + k,\ k \in \mathbb{Z}$, for some $t_0 \in \mathbb{R}$. If $v = p/q$ is rational, then the positions of the "moving point" $\omega(t)$ at times t_k range over a finite number of points. One can thus easily identify 2-periodic rays that never meet that moving point. This property remains true if $a > 0$ is chosen sufficiently small.

Let us consider more general periodic rays. All rays propagating in the square can be described as follows. Let $(x_0, y_0) \in [0, 1]^2$ be arbitrary. Let us consider a ray $t \mapsto (x(t), y(t))$ starting from (x_0, y_0) at time $t = t_0$, with a slope $\tan(\alpha) \in \bar{\mathbb{R}}$ for some $\alpha \in (-\pi, \pi]$. Setting $c = \cos\alpha$ and $s = \sin\alpha$, we define $\tilde{x}(t) = x_0 + (t - t_0)c$ and $\tilde{y}(t) = y_0 + (t - t_0)s$. Then, for times $t \in \mathbb{R}$ such that $|t - t_0|$ is small (possibly only if $t \leqslant t_0$ or $t \geqslant t_0$), the ray is given by $t \mapsto (\tilde{x}(t), \tilde{y}(t))$. Introducing $\hat{x}(t), \hat{y}(t) \in [0, 2)$ such that $\hat{x}(t) = \tilde{x}(t) \bmod 2$ and $\hat{y}(t) = \tilde{y}(t) \bmod 2$ it can be seen, by "developing the square" by means of plane symmetries, that the ray is given by

$$x(t) = \begin{cases} \hat{x}(t) & \text{if } 0 \leqslant \hat{x}(t) \leqslant 1, \\ 2 - \hat{x}(t) & \text{if } 1 < \hat{x}(t) < 2, \end{cases} \qquad y(t) = \begin{cases} \hat{y}(t) & \text{if } 0 \leqslant \hat{y}(t) \leqslant 1, \\ 2 - \hat{y}(t) & \text{if } 1 < \hat{y}(t) < 2. \end{cases} \tag{24}$$

A ray is periodic if and only if $\tan\alpha = p/q \in \mathbb{Q} \cup \{+\infty, -\infty\}$, with p and q relatively prime integers (including the case $q = 0$). The period is equal to $2\sqrt{p^2 + q^2}$. In this case, we have $c = q/\sqrt{p^2 + q^2}$ and $s = p/\sqrt{p^2 + q^2}$. Such a ray reflects from the boundary of the square at times t in the union of the following (possibly empty) subsets of $\mathbb{R}$:

$$A_x = t_0 + \{(k - x_0)/c \mid k \in \mathbb{Z}\}, \quad A_y = t_0 + \{(k - y_0)/s \mid k \in \mathbb{Z}\}.$$

The set $M = M(x_0, y_0, p, q)$ of associated points where this ray meets the boundary is then finite and independent of t_0.

If now $v = r\sqrt{p^2 + q^2}$ with $r \in \mathbb{Q}^+$, at times $t \in A_x \cup A_y$, the "moving point" $\omega(t)$ ranges over a finite set $L = L(x_0, y_0, t_0, p, q, r)$ of points on the boundary of the square, as the accumulated distance travelled along the boundary of the square $(0, 1)^2$ is of the form $d_k = t_0 v + (k - x_0)r(p^2 + q^2)/q$ or $d'_k = t_0 v + (k - y_0)r(p^2 + q^2)/p$ and simply needs to be considered modulo 4. Adjusting the value of the time t_0, we can enforce $M \cap L = \varnothing$. Hence, the associated ray never meets the moving point $\omega(t)$. Finally, as the number of points is finite, this property remains true if $a > 0$ is chosen sufficiently small. □

Remark 3.9. If $\tan\alpha \in \mathbb{R} \setminus \mathbb{Q}$, then the set of points at which the corresponding ray reflects at the boundary $\partial\Omega$ is dense in $\partial\Omega$. In fact at such a point, using the parametrization given in the proof above, we have either $\tilde{x}(t) = x_0 + (t - t_0)\cos\alpha \in \mathbb{Z}$, or $\tilde{y}(t) = y_0 + (t - t_0)\sin\alpha \in \mathbb{Z}$. For instance, if $\tilde{x}(t) = 2k \in \mathbb{Z}$, that is, $x(t) = 0$, meaning that we consider point on the left-hand side of the square, then the corresponding $\tilde{y}(t)$ satisfies $\tilde{y}(t) = (y_0 - x_0 \tan\alpha) + 2k \tan\alpha$. Using the fact that the set that $\{2k\beta \bmod 2 \mid k \in \mathbb{Z}\}$ is dense in $[0, 2]$ if and only if $\beta \in \mathbb{R} \setminus \mathbb{Q}$, we conclude that $\hat{y}(t) = \tilde{y}(t) \bmod 2$ is dense in $[0, 2]$ and thus $y(t)$ is dense in $[0, 1]$ considering (24). From this density, we deduce that the analysis in the case $\tan\alpha \in \mathbb{R} \setminus \mathbb{Q}$ may be quite intricate.

3E. ***An open question.*** Let Ω be a domain of $\mathbb{R}^d$. Let ω_0 be a small disk in Ω, of center $x_0 \in \Omega$ and of radius $\varepsilon > 0$. Let $v > 0$ arbitrary. Given a path $t \mapsto x(t)$ in Ω, we define $\omega(t) = B(x(t), \varepsilon)$ (an open ball), with $x(0) = x_0$. We say that the path $x(\cdot)$ is admissible if $\omega(t) \subset \Omega$ for every time t. We raise the following question:

> Do there exist $T > 0$ and an admissible $\mathscr{C}^1$ path $t \mapsto x(t)$ in Ω, with speed less than or equal to v, such that (Q, T) satisfies the t-GCC? $(\star)$

Here, we have set $Q = \{(t, x) \in \mathbb{R} \times \Omega \mid t \in \mathbb{R},\ x \in \omega(t)\}$. Of course, many variants are possible: the observation set is not necessarily a ball, its velocity may be constant or not. The examples of Sections 3A–3D have shown that addressing the question $(\star)$ is far from obvious. Assumptions on the domain Ω could be made; for instance, one may wonder whether ergodicity of Ω may help or not.

Note that, above in $(\star)$, we restrict the speed of $t \mapsto x(t)$. In fact, using arguments as in the beginnings of the proofs of Propositions 3.4 and 3.8, if $t \mapsto x(t)$ is periodic and if $\bigcup_{t\in\mathbb{R}} B(x(t), \varepsilon)$ satisfies the “static” geometric control condition, then for a sufficiently large speed of the moving point, the set Q satisfies the t-GCC. An estimate of the minimal speed can be derived from the inner diameter of $\bigcup_{t\in\mathbb{R}} B(x(t), \varepsilon)$.

4. Boundary observability and control

In this section, we briefly extend our main results to the case of boundary observability. We consider the framework of Section 1B, and we assume that $\partial\Omega$ is not empty. We still consider the wave equation (3), and we restrict ourselves, for simplicity, to Dirichlet conditions along the boundary.

Let R be an open subset of $\mathbb{R} \times \partial\Omega$. We set

$$\Gamma(t) = \{x \in \partial\Omega \mid (t, x) \in R\}.$$

We say that the Dirichlet wave equation is observable on R in time T if there exists $C > 0$ such that

$$C\big\|\big(u(0,\cdot), u_t(0,\cdot)\big)\big\|^2_{H^1_0(\Omega)\times L^2(\Omega)} \leqslant \int_0^T \int_{\Gamma(t)} \left|\frac{\partial u}{\partial n}(t, x)\right|^2 d\mathcal{H}^{n-1}\, dt \tag{25}$$

for all solutions of (3) with Dirichlet boundary conditions. Here, $\mathcal{H}^{n-1}$ is the $(n-1)$-Hausdorff measure.

In the static case (that is, if $\Gamma(t) \equiv \Gamma$ does not depend on t), the observability property holds true as soon as (Γ, T) satisfies the following GCC (see [Bardos et al. 1992; Burq and Gérard 1997]):

> *Let $T > 0$. The open set $\Gamma \subset \partial\Omega$ satisfies the geometric control condition if the projection onto M of every (compressed) generalized bicharacteristic meets Γ at a time $t \in (0, T)$ at a nondiffractive point.*

Recall the definition of nondiffractive points given in Section 1C. The t-GCC is then defined similarly to Definition 1.6.

Definition 4.1. Let R be an open subset of $\mathbb{R} \times \partial\Omega$, and let $T > 0$. We say that (R, T) satisfies the *time-dependent geometric control condition* (in short, t-GCC) if every compressed generalized bicharacteristic ${}^b\gamma : \mathbb{R} \to {}^bT^*Y \setminus E$, $s \mapsto (t(s), x(s), \tau(s), \xi(s))$, is such that there exists $s \in \mathbb{R}$ such that $t(s) \in (0, T)$ and $(t(s), x(s)) \in R$ and ${}^b\gamma(s) \in H \cup G \setminus G_d$ (a hyperbolic point or a glancing yet nondiffractive point).

We say that R satisfies the t-GCC if there exists $T > 0$ such that (R, T) satisfies the t-GCC.

The control time $T_0(R, \Omega)$ is defined by

$$T_0(R, \Omega) = \inf\{T > 0 \mid (R, T) \text{ satisfies the } t\text{-GCC}\},$$

with the agreement that $T_0(R, \Omega) = +\infty$ if R does not satisfy the t-GCC.

Theorem 4.2. *Let R be an open subset of $\mathbb{R} \times \partial\Omega$ that satisfies the t-GCC. Let $T > T_0(R, \Omega)$. We assume moreover that no (compressed) generalized bicharacteristic has a contact of infinite order with $(0, T) \times \partial\Omega$, that is, $G^\infty = \varnothing$. Then, the observability inequality (25) holds.*

Proof. The proof is similar to the proof of Theorem 1.8 done in Section 2A. We just point out that the set of invisible solutions is defined by

$$N_T = \left\{ u \in H^1((0, T) \times \Omega) \;\middle|\; u \text{ is a Dirichlet solution of (3) and } \chi_R \frac{\partial u}{\partial n} = 0 \right\}.$$

If $u \in N_T$ and $\rho \in T^*((0, T) \times \Omega)$, we wish to prove that u is smooth at ρ. Let ${}^b\gamma(s) = (x(s), t(s), \xi(s), \tau(s))$ be the compressed bicharacteristic that originates from ρ (at $s = 0$). There exists $s_0 \in \mathbb{R}$ such that $t(s_0) \in (0, T)$ and $(t_0, x_0) = (t(s_0), x(s_0)) \in R$. To fix ideas, let us assume that $s_0 \geqslant 0$ and let us set ${}^b\gamma_0 = {}^b\gamma(s) \in T^*\partial Y$. Because of the t-GCC, we may assume that ${}^b\gamma_0$ is a nondiffractive point. Note that the case $s_0 \leqslant 0$ can be treated similarly.

Let now V be an open neighborhood of (t_0, x_0) in $\mathbb{R} \times M$ such that $V_\partial = V \cap (\mathbb{R} \times \partial\Omega) \subset R$. In V, we extend the function $u(t, x)$ by zero outside $\mathbb{R} \times \Omega$ and denote this extension by $\underline{u}$. Since $u_{|V_\partial} = \partial_n u_{|V_\partial} = 0$ we observe that $\underline{u}$ solves $P\underline{u} = 0$ in V. As γ_0 is nondiffractive, the natural bicharacteristic associated with γ_0 has points outside Ω in any neighborhood of γ_0. By propagation of singularities for $\underline{u}$ we thus find that $\underline{u}$ is smooth at γ_0. Then, by propagation of singularities along the compressed generalized bicharacteristic flow (see [Melrose and Sjöstrand 1978; 1982; Hörmander 1985]), we find that u is smooth at ρ. Having u smooth in $(0, T) \times \Omega$, we see that if $u \in N_T$ then $\partial_t u \in N_T$. The rest of the proof follows. $\square$

Remark 4.3. By duality, we have, as well, a boundary controllability result, as in Theorem 1.8$''$.

Remark 4.4. It is interesting to analyze, in this context of boundary observability, the examples of the disk and of the square studied in Section 3.

- For the disk (see Section 3C): we set

$$\Gamma(t) = \left\{ (r, \theta) \in [0, 1] \times \mathbb{R} \;\middle|\; r = 1,\ vt < \theta < vt + a \right\}.$$

This is the limit case of the case of Section 3C, with $\varepsilon = 0$. With respect to Proposition 3.4, it is not true anymore that $T_0(v, a) < +\infty$ if v is chosen sufficiently large.

The other items of Proposition 3.4, providing sufficient conditions such that $T(v, a) = +\infty$, are still valid.

- For the square (see Section 3D), $\omega(t)$ is the translation of the segment $(0, 2a)$ along the boundary, moving anticlockwise and with constant speed v.

With respect to Proposition 3.8, it is not true anymore that $T_0(v, a) < +\infty$ if v is large enough. The second item of Proposition 3.8, providing a sufficient condition such that $T(v, a) = +\infty$, is still valid.

We stress that, in Propositions 3.4 and 3.8, the fact that $T_0 < +\infty$ for v large enough was due to the fact that the width of the observation domain is positive. This remark shows that observability is even more difficult to realize with moving observation domains located at the boundary.

Appendix: A class of test operators near the boundary

We denote by $\Psi^m(Y)$ the space of operators of the form $R = R_{\text{int}} + R_{\tan}$ where:

- R_{int} is a classical pseudodifferential operator of order m with compact support in $\mathbb{R} \times \Omega$, that is, satisfying $R_{\text{int}} = \varphi R_{\text{int}} \varphi$ for some $\varphi \in \mathscr{C}_c^\infty(\mathbb{R} \times \Omega)$;
- $R_{\tan}$ is a classical tangential pseudodifferential operator of order m. In the local normal geodesic coordinates introduced in Section 1C1, such an operator only acts in the y'-variables.

If $\sigma(R_{\text{int}})$ and $\sigma(R_{\tan})$ denote the homogeneous principal symbols of R_{int} and $R_{\tan}$ respectively, we observe that their restriction to $\text{Char}_Y(p) \cup T^*(\mathbb{R} \times \partial\Omega)$ is well defined. Then, by means of the map $j : T^*Y \to {}^bT^*Y$, the function $\sigma(R_{\text{int}})_{|\text{Char}_Y(P)} + \sigma(R_{\tan})_{|\text{Char}_Y(P) \cup T^*(\mathbb{R}\times\partial\Omega)}$ yields a *continuous* map on $\hat{\Sigma} = j(\text{Char}_Y(p)) \cup E$, that we denote by $\kappa(R)$. Its homogeneity yields a continuous function on $S^*\hat{\Sigma} = \hat{\Sigma}/(0, +\infty)$.

We then have the following proposition (see [Lebeau 1996; Burq and Lebeau 2001]).

Proposition A.1. *Let $(u_n)_{n\in\mathbb{N}}$ be a bounded sequence of $H^1(\mathbb{R} \times \Omega)$ that satisfies $(\partial_t^2 - \triangle_g)u_n = 0$ and weakly converges to* 0. *Then, there exist a subsequence $(u_{\varphi(n)})_{n\in\mathbb{N}}$ and a positive measure μ on $S^*\hat{\Sigma}$ such that*

$$(Ru_{\varphi(n)}, u_{\varphi(n)})_{H^1(\Omega)} \to \langle \mu, \kappa(R)\rangle \quad \textit{as } n \to +\infty, \ \textit{where } R \in \Psi^0(Y).$$

This extends results in the interior introduced in the seminal works [Gérard 1991; Tartar 1990].

References

[Anantharaman and Léautaud 2014] N. Anantharaman and M. Léautaud, "Sharp polynomial decay rates for the damped wave equation on the torus", *Anal. PDE* **7**:1 (2014), 159–214. MR Zbl

[Bardos et al. 1992] C. Bardos, G. Lebeau, and J. Rauch, "Sharp sufficient conditions for the observation, control, and stabilization of waves from the boundary", *SIAM J. Control Optim.* **30**:5 (1992), 1024–1065. MR Zbl

[Burq 1998] N. Burq, "Décroissance de l'énergie locale de l'équation des ondes pour le problème extérieur et absence de résonance au voisinage du réel", *Acta Math.* **180**:1 (1998), 1–29. MR Zbl

[Burq and Gérard 1997] N. Burq and P. Gérard, "Condition nécessaire et suffisante pour la contrôlabilité exacte des ondes", *C. R. Acad. Sci. Paris Sér. I Math.* **325**:7 (1997), 749–752. MR Zbl

[Burq and Hitrik 2007] N. Burq and M. Hitrik, "Energy decay for damped wave equations on partially rectangular domains", *Math. Res. Lett.* **14**:1 (2007), 35–47. MR Zbl

[Burq and Lebeau 2001] N. Burq and G. Lebeau, "Mesures de défaut de compacité, application au système de Lamé", *Ann. Sci. École Norm. Sup.* (4) **34**:6 (2001), 817–870. MR Zbl

[Castro et al. 2014] C. Castro, N. Cîndea, and A. Münch, "Controllability of the linear one-dimensional wave equation with inner moving forces", *SIAM J. Control Optim.* **52**:6 (2014), 4027–4056. MR

[Gérard 1991] P. Gérard, "Microlocal defect measures", *Comm. Partial Differential Equations* **16**:11 (1991), 1761–1794. MR Zbl

[Haraux 1989] A. Haraux, "Une remarque sur la stabilisation de certains systèmes du deuxième ordre en temps", *Portugal. Math.* **46**:3 (1989), 245–258. MR Zbl

[Hörmander 1985] L. Hörmander, *The analysis of linear partial differential operators, III: Pseudodifferential operators*, Grundlehren der Mathematischen Wissenschaften **274**, Springer, 1985. MR Zbl

[Lebeau 1996] G. Lebeau, "Équation des ondes amorties", pp. 73–109 in *Algebraic and geometric methods in mathematical physics* (Kaciveli, 1993), edited by A. Boutet de Monvel and V. Marchenko, Math. Phys. Stud. **19**, Kluwer Acad. Publ., Dordrecht, 1996. MR Zbl

[Lebeau and Robbiano 1995] G. Lebeau and L. Robbiano, "Contrôle exact de l'équation de la chaleur", *Comm. Partial Differential Equations* **20**:1-2 (1995), 335–356. MR Zbl

[Lions 1988a] J.-L. Lions, *Contrôlabilité exacte, perturbations et stabilisation de systèmes distribués, Tome 1: Contrôlabilité exacte*, Recherches en Mathématiques Appliquées **8**, Masson, Paris, 1988. MR Zbl

[Lions 1988b] J.-L. Lions, "Exact controllability, stabilization and perturbations for distributed systems", *SIAM Rev.* **30**:1 (1988), 1–68. MR Zbl

[Melrose and Sjöstrand 1978] R. B. Melrose and J. Sjöstrand, "Singularities of boundary value problems, I", *Comm. Pure Appl. Math.* **31**:5 (1978), 593–617. MR Zbl

[Melrose and Sjöstrand 1982] R. B. Melrose and J. Sjöstrand, "Singularities of boundary value problems, II", *Comm. Pure Appl. Math.* **35**:2 (1982), 129–168. MR Zbl

[Phung 2007] K. D. Phung, "Polynomial decay rate for the dissipative wave equation", *J. Differential Equations* **240**:1 (2007), 92–124. MR Zbl

[Rauch and Taylor 1974] J. Rauch and M. Taylor, "Exponential decay of solutions to hyperbolic equations in bounded domains", *Indiana Univ. Math. J.* **24** (1974), 79–86. MR Zbl

[Russell 1971a] D. L. Russell, "Boundary value control of the higher-dimensional wave equation", *SIAM J. Control* **9** (1971), 29–42. MR Zbl

[Russell 1971b] D. L. Russell, "Boundary value control theory of the higher-dimensional wave equation, II", *SIAM J. Control* **9** (1971), 401–419. MR Zbl

[Schenck 2011] E. Schenck, "Exponential stabilization without geometric control", *Math. Res. Lett.* **18**:2 (2011), 379–388. MR Zbl

[Tartar 1990] L. Tartar, "H-measures, a new approach for studying homogenisation, oscillations and concentration effects in partial differential equations", *Proc. Roy. Soc. Edinburgh Sect. A* **115**:3-4 (1990), 193–230. MR Zbl

Received 13 Jul 2016. Revised 25 Jan 2017. Accepted 7 Mar 2017.

JÉRÔME LE ROUSSEAU: jlr@math.univ-paris13.fr
Université Paris-Nord, CNRS UMR 7339, Laboratoire d'analyse géométrie et applications (LAGA), Institut universitaire de France, 99 Avenue Jean Baptiste Clément, 93430 Villetaneuse France

GILLES LEBEAU: gilles.lebeau@unice.fr
Université de Nice Sophia-Antipolis, CNRS UMR 7351, Laboratoire J.-A. Dieudonné, Parc Valrose, 06108 Nice France

PEPPINO TERPOLILLI: peppino.terpolilli@total.com
Total, Centre scientifique et technique Jean-Feger, Avenue Larribau, 64000 Pau, France

EMMANUEL TRÉLAT: emmanuel.trelat@upmc.fr
Sorbonne Universités, UPMC Univ Paris 06, CNRS UMR 7598, Laboratoire Jacques-Louis Lions, Institut universitaire de France, 75005 Paris, France

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 10 No. 4 2017