ANALYSIS & PDE

Volume 5 No. 3 2012

mathematical sciences publishers

Analysis & PDE

msp.berkeley.edu/apde

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

The subscription price for 2012 is US $140/year for the electronic version, and $240/year for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA.

Analysis & PDE, at Mathematical Sciences Publishers, Department of Mathematics, 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 Mathematical Sciences Publishers.

PUBLISHED BY
mathematical sciences publishers
http://msp.org/
A NON-PROFIT CORPORATION
Typeset in LaTeX

ANALYSIS AND PDE
Vol. 5, No. 3, 2012
dx.doi.org/10.2140/apde.2012.5.423

ON SOME MICROLOCAL PROPERTIES OF THE RANGE OF A PSEUDODIFFERENTIAL OPERATOR OF PRINCIPAL TYPE

JENS WITTSTEN

We obtain microlocal analogues of results by L. Hörmander about inclusion relations between the ranges of first order differential operators with coefficients in C^∞ that fail to be locally solvable. Using similar techniques, we study the properties of the range of classical pseudodifferential operators of principal type that fail to satisfy condition (Ψ).

1. Introduction

We shall study the properties of the range of a classical pseudodifferential operator $P \in \Psi^m_{\mathrm{cl}}(X)$ that is not locally solvable, where X is a C^∞ manifold of dimension n. Here, classical means that the total symbol of P is an asymptotic sum of homogeneous terms,

$$\sigma_P(x,\xi) = p_m(x,\xi) + p_{m-1}(x,\xi) + \cdots,$$

where p_k is homogeneous of degree k in ξ and p_m denotes the principal symbol of P. When no confusion can occur we will simply refer to σ_P as the symbol of P. We shall restrict our study to operators of principal type, which means that the Hamilton vector field H_{p_m} and the radial vector field are linearly independent when $p_m = 0$. We shall also assume that all operators are properly supported, that is, both projections from the support of the kernel in $X \times X$ to X are proper maps. For such operators, local solvability at a compact set $M \subset X$ means that for every f in a subspace of $C^\infty(X)$ of finite codimension there is a distribution u in X such that

$$Pu = f \tag{1-1}$$

in a neighborhood of M. We can also define microlocal solvability at a set in the cosphere bundle, or equivalently, at a conic set in $T^*(X) \smallsetminus 0$, the cotangent bundle of X with the zero section removed. By a conic set $K \subset T^*(X) \smallsetminus 0$ we mean a set that is conic in the fiber, that is,

$$(x,\xi) \in K \quad \text{implies} \quad (x,\lambda\xi) \in K \quad \text{for all } \lambda > 0.$$

If, in addition, $\pi_x(K)$ is compact in X, where $\pi_x : T^*(X) \to X$ is the projection, then K is said to be compactly based. Thus, we say that P is solvable at the compactly based cone $K \subset T^*(X) \smallsetminus 0$ if there

Research supported in part by the Swedish Research Council.
MSC2000: primary 35S05; secondary 35A07, 58J40, 47G30.
Keywords: pseudodifferential operators, microlocal solvability, principal type, condition (Ψ), inclusion relations, bicharacteristics.

is an integer N such that for every $f \in H^{\mathrm{loc}}_{(N)}(X)$ there exists a $u \in \mathscr{D}'(X)$ with $K \cap WF(Pu - f) = \varnothing$ (see Definition 2.1).

The famous example due to Hans Lewy [1957] of the existence of functions $f \in C^\infty(\mathbb{R}^3)$ such that the equation

$$\partial_{x_1} u + i\partial_{x_2} u - 2i(x_1 + ix_2)\partial_{x_3} u = f$$

does not have any solution $u \in \mathscr{D}'(\Omega)$ in any open nonvoid subset $\Omega \subset \mathbb{R}^3$ contradicted the assumption that partial differential equations with smooth coefficients behave as analytic partial differential equations, for which existence of analytic solutions is guaranteed by the Cauchy–Kovalevsky theorem. This example led to an extension due to Hörmander [1960b; 1960a] in the sense of a necessary condition for a differential equation $P(x, D)u = f$ to have a solution locally for every $f \in C^\infty$. In fact (see [Hörmander 1963, Theorem 6.1.1]), if Ω is an open set in $\mathbb{R}^n$, and P is a differential operator of order m with coefficients in $C^\infty(\Omega)$ such that the differential equation $P(x, D)u = f$ has a solution $u \in \mathscr{D}'(\Omega)$ for every $f \in C_0^\infty(\Omega)$, then $\{p_m, \bar{p}_m\}$ must vanish at every point $(x, \xi) \in \Omega \times \mathbb{R}^n$ for which $p_m(x, \xi) = 0$, where

$$\{a, b\} = \sum_{j=1}^{n} \partial_{\xi_j} a\, \partial_{x_j} b - \partial_{x_j} a\, \partial_{\xi_j} b$$

denotes the Poisson bracket.

In addition to his example, Lewy conjectured that differential operators that fail to have local solutions are essentially uniquely determined by the range. Later Hörmander [1963, Chapter 6.2] proved that if P and Q are two first order differential operators with coefficients in $C^\infty(\Omega)$ and in $C^1(\Omega)$, respectively, such that the equation $P(x, D)u = Q(x, D)f$ has a solution $u \in \mathscr{D}'(\Omega)$ for every $f \in C_0^\infty(\Omega)$, and x is a point in Ω such that

$$p_1(x, \xi) = 0 \quad \text{and} \quad \{p_1, \bar{p}_1\}(x, \xi) \neq 0 \tag{1-2}$$

for some $\xi \in \mathbb{R}^n$, then there is a constant μ such that (at the fixed point x)

$${}^tQ(x, D) = \mu\, {}^tP(x, D),$$

where ${}^tQ$ and ${}^tP$ are the formal adjoints of Q and P. If (1-2) holds for a dense set of points x in Ω and if the coefficients of $p_1(x, D)$ do not vanish simultaneously in Ω, then there is a function $\mu \in C^1(\Omega)$ such that

$$Q(x, D)u = P(x, D)(\mu u). \tag{1-3}$$

Furthermore, for such an operator P and function μ, the equation $P(x, D)u = \mu P(x, D) f$ has a solution $u \in \mathscr{D}'(\Omega)$ for every $f \in C_0^\infty(\Omega)$ if and only if $p_1(x, D)\mu = 0$.

Hörmander also showed that this result extends to operators of higher order in the following way (see [1963, Theorem 6.2.4]). If P is a differential operator of order m with coefficients in $C^\infty(\Omega)$ and μ is a function in $C^m(\Omega)$ such that the equation

$$P(x, D)u = \mu P(x, D) f$$

has a solution $u \in \mathscr{D}'(\Omega)$ for every $f \in C_0^\infty(\Omega)$, then it follows that

$$\sum_{j=1}^{n} \partial_{\xi_j} p_m(x, \xi) \partial_{x_j} \mu(x) = 0$$

for all $x \in \Omega$ and $\xi \in \mathbb{R}^n$ such that

$$\{p_m, \bar{p}_m\}(x, \xi) \neq 0 \quad \text{and} \quad p_m(x, \xi) = 0. \tag{1-4}$$

This means that the derivative of μ must vanish along every bicharacteristic element with initial data (x, ξ) satisfying (1-4), that is, giving rise to nonexistence of solutions.

If P is a pseudodifferential operator such that P is microlocally elliptic near (x_0, ξ_0), then there exists a microlocal inverse, called a parametrix P^{-1} of P, such that in a conic neighborhood of (x_0, ξ_0) we have $PP^{-1} = P^{-1}P =$ Identity modulo smoothing operators. P is then trivially seen to be microlocally solvable near (x_0, ξ_0), and for any pseudodifferential operator Q we can write $Q = PP^{-1}Q + R = PE + R$, where R is a smoothing operator. When the range of Q is microlocally contained in the range of P, we will show the existence of this type of representation for Q in the case when P is a nonsolvable pseudodifferential operator of principal type, although we will have to content ourselves with a weaker statement concerning the Taylor coefficients of the symbol of the operator R (see Theorem 2.19 for the precise formulation of the result). Note that when P is solvable but nonelliptic we cannot hope to obtain such a representation in general; see the remark on page 440.

For pseudodifferential operators of principal type, Hörmander [1985b] proved that local solvability in the sense of (1-1) implies that M has an open neighborhood Y in X where p_m satisfies condition (Ψ), which means that

$$\operatorname{Im} ap_m \text{ does not change sign from } - \text{ to } + \text{ along the oriented bicharacteristics of } \operatorname{Re} ap_m \tag{1-5}$$

over Y for any $0 \neq a \in C^\infty(T^*(Y) \setminus 0)$. The oriented bicharacteristics are the positive flow-outs of the Hamilton vector field $H_{\operatorname{Re} ap_m}$ on $\operatorname{Re} ap_m = 0$. The proof relies on an idea due to Moyer [1978], and uses the fact that condition (1-5) is invariant under multiplication of p_m with nonvanishing factors, and conjugation of P with elliptic Fourier integral operators.

Rather recently Dencker [2006] proved that condition (Ψ) is also sufficient for local and microlocal solvability for operators of principal type. To get local solvability at a point x_0, Dencker assumed the strong form of the nontrapping condition at x_0,

$$p_m = 0 \quad \text{implies} \quad \partial_\xi p_m \neq 0. \tag{1-6}$$

This was the original condition for principal type of Nirenberg and Treves [1970a; 1970b; 1971], which is always obtainable microlocally after a canonical transformation. Thus, we shall study pseudodifferential operators that fail to satisfy condition (Ψ) in place of the condition given by (1-4), and show that such operators are, in analogue with the inclusion relations between the ranges of differential operators that fail to be locally solvable, essentially uniquely determined by the range. However, even though (1-4) is a microlocal condition, we get the mentioned local results for differential operators because of analyticity in ξ of the corresponding symbol. Since this is not generally true for pseudodifferential operators, our results will be inherently microlocal. We will combine the techniques used in [Hörmander 1963] to prove the inclusion relations for differential operators with the approach used in [Hörmander 1985b] to prove the necessity of condition (Ψ) for local solvability of pseudodifferential operators of principal type.

It is possible to extend these results to certain systems of pseudodifferential operators, which will be addressed in a forthcoming joint paper with Nils Dencker.

2. Nonsolvable operators of principal type

Let X be a C^∞ manifold of dimension n. In what follows, C will be taken to be a new constant every time unless stated otherwise. We let $\mathbb{N} = \{0, 1, 2, \dots\}$, and if $\alpha \in \mathbb{N}^n$ is a multiindex $\alpha = (\alpha_1, \dots, \alpha_n)$, we let

$$D_x^\alpha = D_{x_1}^{\alpha_1} \cdots D_{x_n}^{\alpha_n},$$

where $D_{x_j} = -i\partial_{x_j}$. We shall also employ the standard notation $f_{(\alpha)}^{(\beta)}(x,\xi) = \partial_x^\alpha \partial_\xi^\beta f(x,\xi)$ for multi-indices α, β.

In this section we will follow the outline of [Hörmander 1985b, Chapter 26, Section 4]. Recall that the Sobolev space $H_{(s)}(X)$ for $s \in \mathbb{R}$ is a local space, that is, if $\varphi \in C_0^\infty(X)$ and $u \in H_{(s)}(X)$, then $\varphi u \in H_{(s)}(X)$, and the corresponding operator of multiplication is continuous. Thus we can define

$$H_{(s)}^{\mathrm{loc}}(X) = \{u \in \mathscr{D}'(X) : \varphi u \in H_{(s)}(X) \text{ for all } \varphi \in C_0^\infty(X)\}.$$

This is a Fréchet space, and its dual with respect to the inner product on L^2 is

$$H_{(-s)}^{\mathrm{comp}}(X) = H_{(-s)}^{\mathrm{loc}}(X) \cap \mathscr{E}'(X).$$

Definition 2.1. If $K \subset T^*(X) \smallsetminus 0$ is a compactly based cone, we shall say that the range of $Q \in \Psi_{\mathrm{cl}}^m(X)$ is microlocally contained in the range of $P \in \Psi_{\mathrm{cl}}^k(X)$ at K if there exists an integer N such that for every $f \in H_{(N)}^{\mathrm{loc}}(X)$, there exists a $u \in \mathscr{D}'(X)$ with $WF(Pu - Qf) \cap K = \varnothing$.

If $I \in \Psi_{\mathrm{cl}}^0(X)$ is the identity on X, we obtain from Definition 2.1 the definition of microlocal solvability for a pseudodifferential operator (see [Hörmander 1985b, Definition 26.4.3]) by setting $Q = I$. Thus, the range of the identity is microlocally contained in the range of P at K if and only if P is microlocally solvable at K. Note also that if P and Q satisfy Definition 2.1 for some integer N, then due to the inclusion

$$H_{(t)}^{\mathrm{loc}}(X) \subset H_{(s)}^{\mathrm{loc}}(X) \quad \text{if } s < t,$$

the statement also holds for any integer $N' \geq N$. Hence N can always be assumed to be positive. Furthermore, the property is preserved if Q is composed with a properly supported pseudodifferential operator $Q_1 \in \Psi_{\mathrm{cl}}^{m'}(X)$ from the right. Indeed, let g be an arbitrary function in $H_{(N+m')}^{\mathrm{loc}}(X)$. Then $f = Q_1 g \in H_{(N)}^{\mathrm{loc}}(X)$ since the map

$$Q_1 : H_{(s)}^{\mathrm{loc}}(X) \to H_{(s-m')}^{\mathrm{loc}}(X)$$

is continuous for every $s \in \mathbb{R}$, so by Definition 2.1 there exists a $u \in \mathscr{D}'(X)$ with $WF(Pu - Qf) \cap K = \varnothing$. Hence the range of QQ_1 is microlocally contained in the range of P at K with the integer N replaced by $N + m'$.

The property given by Definition 2.1 is also preserved under composition of both P and Q with a properly supported pseudodifferential operator from the left. This follows immediately from the fact that properly supported pseudodifferential operators are microlocal, that is,

$$WF(Au) \subset WF(u) \cap WF(A) \quad \text{for } u \in \mathscr{D}'(X).$$

Remark. In Definition 2.1 we may always assume that $f \in H_{(N)}^{\mathrm{comp}}(X)$ and $u \in \mathscr{E}'(X)$ when considering a fixed cone K. In fact, assume

$$Qf = Pu + g,$$

where $f \in H_{(N)}^{\mathrm{loc}}(X)$ and $u, g \in \mathscr{D}'(X)$ with $WF(g) \cap K = \varnothing$, and let $Y \Subset X$ satisfy $K \subset T^*(Y) \smallsetminus 0$. (We write $Y \Subset X$ when $\overline{Y}$ is compact and contained in X.) Since P and Q are properly supported we can find $Z_1, Z_2 \subset X$ such that $Pv = 0$ in Y if $v = 0$ in Z_1, and $Qv = 0$ in Y if $v = 0$ in Z_2. We may of course assume that $Y \Subset Z_j$ for $j = 1, 2$. Fix $\phi_j \in C_0^\infty(X)$ with $\phi_j = 1$ on Z_j. Then we have $Pu = P(\phi_1 u)$ and $Qf = Q(\phi_2 f)$ in Y, so

$$\varnothing = WF(Qf - Pu) \cap K = WF(Q(\phi_2 f) - P(\phi_1 u)) \cap K$$

where $\phi_1 u$ and $\phi_2 f$ have compact support. Hence we may assume that $u \in \mathscr{E}'(X)$ and $f \in H_{(N)}^{\mathrm{comp}}(X) = H_{(N)}^{\mathrm{loc}}(X) \cap \mathscr{E}'(X)$ to begin with. Note that this also implies $g = Qf - Pu \in \mathscr{E}'(X)$ since P and Q are properly supported.

The following easy example will prove useful when discussing inclusion relations between the ranges of solvable but nonelliptic operators.

Example 2.2. If $X \subset \mathbb{R}^n$ is open, and $K \subset T^*(X) \smallsetminus 0$ is a compactly based cone, then the range of $D_1 = -i\partial/\partial x_1$ is microlocally contained in the range of D_2 at K. In fact, this is trivially true since both operators are surjective $\mathscr{D}'(X) \to \mathscr{D}'(X)/C^\infty(X)$. To see that for example D_1 is surjective, we note that by the remark on page 427 it suffices to show that there exists a number $N \in \mathbb{Z}$ such that the equation $D_1 u = f$ has a solution $u \in \mathscr{D}'(X)$ for every $f \in H_{(N)}^{\mathrm{comp}}(X) = H_{(N)}^{\mathrm{loc}}(X) \cap \mathscr{E}'(X)$. By [Hörmander 1983b, Theorem 10.3.1] this is satisfied for every $N \in \mathbb{Z}$ if $u \in H_{(N+1)}^{\mathrm{loc}}(X)$ is given by $E * f$, where E is the regular fundamental solution of D_1.

Just as the microlocal solvability of a pseudodifferential operator P gives an a priori estimate for the adjoint P^*, we have the following result for operators satisfying Definition 2.1.

Lemma 2.3. *Let $K \subset T^*(X) \smallsetminus 0$ be a compactly based cone. Let $Q \in \Psi_{\mathrm{cl}}^m(X)$ and $P \in \Psi_{\mathrm{cl}}^k(X)$ be properly supported pseudodifferential operators such that the range of Q is microlocally contained in the range of P at K. If $Y \Subset X$ satisfies $K \subset T^*(Y)$ and if N is the integer in Definition 2.1, then for every positive integer κ we can find a constant C, a positive integer ν and a properly supported pseudodifferential* *operator A with $WF(A) \cap K = \varnothing$ such that*

$$\|Q^* v\|_{(-N)} \le C(\|P^* v\|_{(\nu)} + \|v\|_{(-N-\kappa-n)} + \|Av\|_{(0)}) \tag{2-1}$$

for all $v \in C_0^\infty(Y)$.

Since (2-1) holds for any κ, it is actually superfluous to include the dimension n in the norm $\|v\|_{(-N-\kappa-n)}$. However, for our purposes, it turns out that this is the most convenient formulation.

Proof. We shall essentially adapt the proof of [Hörmander 1985b, Lemma 26.4.5]. Let $\|\cdot\|_{(s)}$ denote a norm in $H_{(s)}^{\mathrm{comp}}(X)$ that defines the topology in $H_{(s)}^{c}(M) = H_{(s)}^{\mathrm{loc}}(X) \cap \mathscr{E}'(M)$ for every compact set $M \subset X$. (The reason we change notation from $H_{(s)}^{\mathrm{comp}}(M)$ to $H_{(s)}^{c}(M)$ when M is compact is to signify that $H_{(s)}^{c}(M)$ is a Hilbert space for each fixed compact set M.) Let $Y \Subset Z \Subset X$, and take $\chi \in C_0^\infty(X)$ with $\operatorname{supp}\chi = \bar{Z}$ to be a real-valued cutoff function identically equal to 1 in a neighborhood of Y. Then $\chi Q f \in H_{(N-m)}^{c}(\bar{Z})$ for all $f \in H_{(N)}^{\mathrm{comp}}(X)$ since Q is properly supported, and we claim that for fixed $f \in H_{(N)}^{\mathrm{comp}}(X)$ we have for some C, ν and A as in the statement of the lemma

$$|(\chi Q f, v)| \le C(\|P^* v\|_{(\nu)} + \|v\|_{(-N-\kappa-n)} + \|Av\|_{(0)}) \tag{2-2}$$

for all $v \in C_0^\infty(Y)$. Indeed, by hypothesis and the remark on page 427 we can find u and $\tilde{g}$ in $\mathscr{E}'(X)$ with $WF(\tilde{g}) \cap K = \varnothing$ such that

$$\chi Q f = Q f - (1-\chi) Q f = P u + \tilde{g} - (1-\chi) Q f.$$

Since $K \subset T^*(Y)$ and $\chi \equiv 1$ near Y we get $WF((1-\chi)Qf) \cap K = \varnothing$, so $\chi Q f = P u + g$ for some $g \in \mathscr{E}'(X)$ with $WF(g) \cap K = \varnothing$. Thus

$$(\chi Q f, v) = (u, P^* v) + (g, v) \quad \text{for } v \in C_0^\infty(Y).$$

Now choose properly supported pseudodifferential operators B_1 and B_2 of order 0 with $I = B_1 + B_2$ and $WF(B_1) \cap WF(g) = \varnothing$ and $WF(B_2) \cap K = \varnothing$, which is possible since $WF(g) \cap K = \varnothing$. Since $g \in \mathscr{E}'(X)$ and $B_1 : \mathscr{E}'(X) \to \mathscr{E}'(X)$ is continuous and microlocal we get $B_1 g \in C_0^\infty(X)$, so $(B_1 g, v)$ can be estimated by $C\|v\|_{(-N-\kappa-n)}$. Also, $g \in H_{(-\mu)}^{\mathrm{loc}}(X)$ for some $\mu > 0$ so if B is properly supported and elliptic of order μ, and $B' \in \Psi_{\mathrm{cl}}^{-\mu}(X)$ is a properly supported parametrix of B, then

$$B_2^* v = B' B B_2^* v + L B_2^* v, \tag{2-3}$$

where $L \in \Psi^{-\infty}(X)$ and both B' and L are continuous $H_{(s)}^{\mathrm{comp}}(X) \to H_{(s+\mu)}^{\mathrm{comp}}(X)$. Hence

$$|(B_2 g, v)| \le C\|B_2^* v\|_{(\mu)} \le C(\|B B_2^* v\|_{(0)} + \|B_2^* v\|_{(0)}),$$

and if we apply the identity (2-3) to $\|B_2^* v\|_{(0)}, \|B_2^* v\|_{(-\mu)}, \ldots$ sufficiently many times, and then recall that B_2^* is properly supported and of order 0, we obtain

$$|(B_2 g, v)| \le C(\|B B_2^* v\|_{(0)} + \|v\|_{(-N-\kappa-n)}).$$

Since we chose B to be properly supported this gives (2-2) with $A = B B_2^*$.

For fixed κ, suppose V is the space $C_0^\infty(Y)$ equipped with the topology defined by the seminorms $\|v\|_{(-N-\kappa-n)}$, $\|P^*v\|_{(\nu)}$ for $\nu = 1, 2, \dots$, and $\|Av\|_{(0)}$, where A is a properly supported pseudodifferential operator with $K \cap WF(A) = \varnothing$. It suffices to use a countable sequence $A_1, A_2, \dots$, where A_ν is noncharacteristic of order ν in a set that increases to $(T^*(X) \setminus 0) \setminus K$ as $\nu \to \infty$. Thus V is a metrizable space. The sesquilinear form $(\chi Qf, v)$ in the product of the Hilbert space $H^c_{(N-m)}(\overline{Z})$ and the metrizable space V is obviously continuous in χQf for fixed v, and by (2-2) it is also continuous in v for fixed f. Hence it is continuous, which means that for some ν and C

$$|(\chi Qf, v)| \le C\|Qf\|_{(N-m)}(\|P^*v\|_{(\nu)} + \|v\|_{(-N-\kappa-n)} + \|Av\|_{(0)})$$

for all $f \in H^{\mathrm{comp}}_{(N)}(X)$ and $v \in C_0^\infty(Y)$. Now Q is continuous from $H^{\mathrm{comp}}_{(N)}(X)$ to $H^{\mathrm{comp}}_{(N-m)}(X)$ so we have $\|Qf\|_{(N-m)} \le C\|f\|_{(N)}$. Since $\chi \equiv 1$ near Y and $(\chi Q)^* = Q^*\chi$, this yields the estimate

$$|(f, Q^*v)| \le C\|f\|_{(N)}(\|P^*v\|_{(\nu)} + \|v\|_{(-N-\kappa-n)} + \|Av\|_{(0)}). \tag{2-4}$$

For $v \in C_0^\infty(Y)$ and Q^* properly supported we have $Q^*v \in C_0^\infty(X)$, and therefore also $Q^*v \in H^{\mathrm{loc}}_{(-N)}(X)$. Viewing Q^*v as a functional on $H^{\mathrm{comp}}_{(N)}(X)$, the dual of $H^{\mathrm{loc}}_{(-N)}(X)$ with respect to the standard inner product on L^2, we obtain (2-1) after taking the supremum over all $f \in H^{\mathrm{comp}}_{(N)}(X)$ with $\|f\|_{(N)} = 1$. □

We will need the following analogue of [Hörmander 1985b, Proposition 26.4.4]. Recall that $\mathcal{H} : T^*(Y) \setminus 0 \to T^*(X) \setminus 0$ is a canonical transformation if and only if its graph $C_{\mathcal{H}}$ in the product $(T^*(X) \setminus 0) \times (T^*(Y) \setminus 0)$ is Lagrangian with respect to the difference $\sigma_X - \sigma_Y$ of the symplectic forms of $T^*(X)$ and $T^*(Y)$ lifted to $T^*(X) \times T^*(Y) = T^*(X \times Y)$. This differs in sign from the symplectic form $\sigma_X + \sigma_Y$ of $T^*(X \times Y)$ so it is the *twisted graph*

$$C'_{\mathcal{H}} = \{(x, \xi, y, -\eta) : (x, \xi, y, \eta) \in C_{\mathcal{H}}\},$$

which is Lagrangian with respect to the standard symplectic structure in $T^*(X \times Y)$.

Proposition 2.4. *Let $K \subset T^*(X) \setminus 0$ and $K' \subset T^*(Y) \setminus 0$ be compactly based cones and let χ be a homogeneous symplectomorphism from a conic neighborhood of K' to one of K such that $\chi(K') = K$. Let $A \in I^{m'}(X \times Y, \Gamma')$ and $B \in I^{m''}(Y \times X, (\Gamma^{-1})')$, where Γ is the graph of χ, and assume that A and B are properly supported and noncharacteristic at the restriction of the graphs of χ and χ^{-1} to K' and to K respectively, while $WF'(A)$ and $WF'(B)$ are contained in small conic neighborhoods. Then the range of the pseudodifferential operator Q in X is microlocally contained in the range of the pseudodifferential operator P in X at K if and only if the range of the pseudodifferential operator BQA in Y is microlocally contained in the range of the pseudodifferential operator BPA in Y at K'.*

Proof. Choose $A_1 \in I^{-m''}(X \times Y, \Gamma')$ and $B_1 \in I^{-m'}(Y \times X, (\Gamma^{-1})')$ properly supported such that

$$K' \cap WF(BA_1 - I) = \varnothing, \qquad K \cap WF(A_1B - I) = \varnothing,$$
$$K' \cap WF(B_1A - I) = \varnothing, \qquad K \cap WF(AB_1 - I) = \varnothing.$$

Assume that the range of Q is microlocally contained in the range of P at K and choose N as in Definition 2.1. Let $g \in H^{\mathrm{loc}}_{(N+m')}(Y)$ and set $f = Ag \in H^{\mathrm{loc}}_{(N)}(X)$. Then we can find $u \in \mathcal{D}'(X)$ such that

$K \cap WF(Pu - Qf) = \varnothing$. Let $v = B_1 u \in \mathscr{D}'(Y)$. Then

$$WF(Av - u) = WF((AB_1 - I)u)$$

does not meet K, so $K \cap WF(PAv - Qf) = \varnothing$. Recalling that $f = Ag$ this implies

$$K' \cap WF(BPAv - BQAg) = \varnothing,$$

so the range of BQA is microlocally contained in the range of BPA at K'. Conversely, if the range of BQA is microlocally contained in the range of BPA at K', it follows that the range of A_1BQAB_1 is microlocally contained in the range of A_1BPAB_1 at K. Since

$$K \cap WF(A_1BPAB_1u - A_1BQAB_1f) = K \cap WF(Pu - Qf),$$

this means that the range of Q is microlocally contained in the range of P at K, which proves the proposition. □

Before we can state our main theorem, we need to study the geometric situation that occurs when p fails to satisfy condition (Ψ). Recall that by [Hörmander 1985b, Theorem 26.4.12] we may always assume that the nonvanishing factor in condition (1-5) is a homogeneous function. We begin with a lemma concerning a reduction of the general case.

Lemma 2.5. *Let p and q be homogeneous smooth functions on $T^*(X)\smallsetminus 0$, and let $t \mapsto \gamma(t)$, for $a \le t \le b$, be a bicharacteristic interval of* $\operatorname{Re} qp$ *such that $q(\gamma(t)) \neq 0$ for $a \le t \le b$. If*

$$\operatorname{Im} qp(\gamma(a)) < 0 < \operatorname{Im} qp(\gamma(b)), \tag{2-5}$$

then there exists a proper subinterval $[a', b'] \subset [a, b]$, possibly reduced to a point, such that

(i) $\operatorname{Im} qp(\gamma(t)) = 0$ *for $a' \le t \le b'$,*

(ii) *for every $\varepsilon > 0$ there exist $a' - \varepsilon < s_- < a'$ and $b' < s_+ < b' + \varepsilon$ such that* $\operatorname{Im} qp(\gamma(s_-)) < 0 < \operatorname{Im} qp(\gamma(s_+))$.

If $\gamma(t)$ is defined for $a \le t \le b$ we shall in the sequel say that $\operatorname{Im} qp$ changes sign from $-$ to $+$ on γ if (2-5) holds. If $\gamma|_{[a',b']}$ is the restriction of γ to $[a', b']$ and (i) and (ii) hold we shall say that $\operatorname{Im} qp$ *strongly* changes sign from $-$ to $+$ on $\gamma|_{[a',b']}$.

Proof. It suffices to regard the case that $q = 1$, $X = \mathbb{R}^n$, p is homogeneous of degree 1 with $\operatorname{Re} p = \xi_1$, and the bicharacteristic of $\operatorname{Re} p$ is given by

$$a \le x_1 \le b, \quad x' = (x_2, \dots, x_n) = 0, \quad \xi = \varepsilon_n. \tag{2-6}$$

Here $\varepsilon_n = (0, \dots, 0, 1) \in \mathbb{R}^n$, and we shall in what follows write ξ^0 in place of ε_n'. The proof of this fact is taken from [Hörmander 1985b, page 97] and is given here for the purpose of reference later, in particular in connection with Definition 2.11 below.

Choose a pseudodifferential operator Q with principal symbol q. If we let $P_1 = QP$, then the principal symbol of P_1 is $p_1 = qp$ so $\operatorname{Im} p_1$ changes sign from $-$ to $+$ on the bicharacteristic γ of $\operatorname{Re} p_1$. Now choose Q_1 to be of order $1 -$ degree P_1 with positive, homogeneous principal symbol. If p_2 is the principal symbol of $P_2 = Q_1 P_1$, it follows that Re p_1 and Re p_2 have the same bicharacteristics, including orientation, and since p_2 is homogeneous of degree 1 these can be considered to be curves on the cosphere bundle $S^*(X)$. Moreover, Im p_1 and Im p_2 have the same sign, so Im p_2 changes sign from $-$ to $+$ along $\gamma \subset S^*(X)$. If γ is a closed curve on $S^*(X)$ we can pick an arc that is not closed where the sign change still occurs. If we assume this to be done, then [Hörmander 1985b, Proposition 26.1.6] states that there exists a C^∞ homogeneous canonical transformation χ from an open conic neighborhood of (2-6) to one of γ such that $\chi(x_1, 0, \varepsilon_n) = \gamma(x_1)$ and $\chi^*(\operatorname{Re} p_2) = \xi_1$. Since the Hamilton field is symplectically invariant it follows that the equations of a bicharacteristic are invariant under the action of canonical transformations, that is, $\tilde{\gamma}$ is a bicharacteristic of $\chi^*(\operatorname{Re} p_2)$ if and only if $\chi(\tilde{\gamma})$ is a bicharacteristic of Re p_2. This proves the claim.

In accordance with the notation in [Hörmander 1985b, page 97], let $(x', \xi') = (0, \xi^0)$ and consider

$$L(0, \xi^0) = \inf\{t - s : a < s < t < b,\ \operatorname{Im} p(s, 0, \varepsilon_n) < 0 < \operatorname{Im} p(t, 0, \varepsilon_n)\}.$$

For every small $\delta > 0$ there exist s_δ and t_δ such that $a < s_\delta < t_\delta < b$, $\operatorname{Im} p(s_\delta, 0, \varepsilon_n) < 0 < \operatorname{Im} p(t_\delta, 0, \varepsilon_n)$ and $t_\delta - s_\delta < L(0, \xi^0) + \delta$. Choose a sequence $\delta_j \to 0$ such that the limits $a' = \lim s_{\delta_j}$ and $b' = \lim t_{\delta_j}$ exist. Then $b' - a' = L(0, \xi^0)$ and in view of (2-5) we have $a < a' \le b' < b$ by continuity. Moreover, $\operatorname{Im} p(t, 0, \varepsilon_n) = 0$ for $a' \le t \le b'$. This is clear if $a' = b'$. If on the other hand $\operatorname{Im} p(t, 0, \varepsilon_n)$ is, say, strictly positive for some $a' < t < b'$, then $L(0, \xi^0) \le t - s_{\delta_j} \to t - a' < b' - a'$, a contradiction. Thus (i) holds.

To prove (ii), let $\varepsilon > 0$. After possibly reducing to a subsequence we may assume that the sequences $\{s_{\delta_j}\}$ and $\{t_{\delta_j}\}$ given above are monotone increasing and decreasing, respectively. It then follows by (i) that $s_{\delta_j} < a' \le b' < t_{\delta_j}$ for all j. Since $s_{\delta_j} \to a'$ and $t_{\delta_j} \to b'$ we can choose j so that $a' - \varepsilon < s_{\delta_j} < a'$ and $b' < t_{\delta_j} < b' + \varepsilon$. By construction we have $\operatorname{Im} p(s_{\delta_j}, 0, \varepsilon_n) < 0 < \operatorname{Im} p(t_{\delta_j}, 0, \varepsilon_n)$. This completes the proof. □

Although it will not be needed here, we note that if $[a', b']$ is the interval given by Lemma 2.5 and $a' < b'$, then in addition to (i) and (ii) we also have

(iii) there exists a $\delta > 0$ such that $\operatorname{Im} qp(\gamma(s)) \le 0 \le \operatorname{Im} qp(\gamma(t))$ for all $a' - \delta < s < a'$ and $b' < t < b' + \delta$.

Indeed, the infimum $L(0, \xi^0) = b' - a'$ would otherwise satisfy $L(0, \xi^0) < \delta$ for every δ in view of (ii), which is a contradiction when $a' < b'$.

We next recall the definition of a one-dimensional bicharacteristic.

Definition 2.6. A one-dimensional bicharacteristic of the pseudodifferential operator with homogeneous principal symbol p is a C^1 map $\gamma : I \to T^*(X) \setminus 0$, where I is an interval on $\mathbb{R}$, such that

(i) $p(\gamma(t)) = 0$ for $t \in I$,

(ii) $0 \ne \gamma'(t) = c(t) H_p(\gamma(t))$ if $t \in I$

for some continuous function $c : I \to \mathbb{C}$.

Let P be an operator of principal type on a C^∞ manifold X with principal symbol p, and suppose p fails to satisfy condition (Ψ) in X. By (1-5) there is a function q in $C^\infty(T^*(X) \setminus 0)$ such that $\operatorname{Im} qp$

changes sign from $-$ to $+$ on a bicharacteristic γ of $\operatorname{Re} qp$, where $q \neq 0$. As can be seen in [Hörmander 1985b, pages 96–97], we can then find a compact one-dimensional bicharacteristic interval $\Gamma \subset \gamma$ or a characteristic point $\Gamma \in \gamma$ such that the sign change occurs on bicharacteristics of $\operatorname{Re} qp$ arbitrarily close to Γ. What we mean by this will be clear from the following discussion, although we will not use this terminology in the sequel. By the proof of Lemma 2.5 it suffices to regard the case that $q = 1$, $X = \mathbb{R}^n$, p is homogeneous of degree 1 with $\operatorname{Re} p = \xi_1$, and the bicharacteristic of $\operatorname{Re} p$ is given by (2-6).

We shall now study a slightly more general situation is some detail. If $\gamma = I \times \{w_0\}$, where $I = [a, b]$, we shall by $|\gamma|$ denote the usual arc length in $\mathbb{R}^{2n}$, so that $|\gamma| = b - a$. Furthermore, we will assume that all curves are bicharacteristics of $\operatorname{Re} p = \xi_1$, that is, $w_0 = (x', 0, \xi') \in \mathbb{R}^{2n-1}$. We owe parts of this exposition to Nils Dencker.

Lemma 2.7. *Assume that* $\operatorname{Im} p$ *strongly changes sign from* $-$ *to* $+$ *on* $\gamma = [a, b] \times \{w_0\}$. *Then for any* $\delta > 0$ *there exist* $\varepsilon > 0$, $a - \delta < s_- < a$ *and* $b < s_+ < b + \delta$ *such that* $\pm \operatorname{Im} p(s_\pm, w) > 0$ *for any* $|w - w_0| < \varepsilon$.

Proof. Since $t \mapsto \operatorname{Im} p(t, w_0)$ strongly changes sign on $[a, b]$ we can find $s_\pm$ satisfying the conditions so that $\pm \operatorname{Im} p(s_\pm, w_0) > 0$. By continuity we can find $\varepsilon_\pm > 0$ so that $\pm \operatorname{Im} p(s_\pm, w) > 0$ for any $|w - w_0| < \varepsilon_\pm$. The lemma now follows if we take $\varepsilon = \min(\varepsilon_-, \varepsilon_+)$. □

Definition 2.8. Let $\gamma = [a, b] \times \{w_0\}$ and $\gamma_j = [a_j, b_j] \times \{w_j\}$. If $\liminf_{j\to\infty} a_j \geq a$, $\limsup_{j\to\infty} b_j \leq b$ and $\lim_{j\to\infty} w_j = w_0$, then we shall write $\gamma_j \dashrightarrow \gamma$ as $j \to \infty$. If in addition $\lim_{j\to\infty} a_j = a$ and $\lim_{j\to\infty} b_j = b$ then we shall write $\gamma_j \to \gamma$ as $j \to \infty$.

Definition 2.9. If γ is a bicharacteristic of $\operatorname{Re} p = \xi_1$ and there exists a sequence $\{\gamma_j\}$ of bicharacteristics of $\operatorname{Re} p$ such that $\operatorname{Im} p$ strongly changes sign from $-$ to $+$ on γ_j for all j and $\gamma_j \dashrightarrow \gamma$ as $j \to \infty$, we set

$$L_p(\gamma) = \inf_{\{\gamma_j\}} \{\liminf_{j\to\infty} |\gamma_j| : \gamma_j \dashrightarrow \gamma \text{ as } j \to \infty\}, \tag{2-7}$$

where the infimum is taken over all such sequences. We shall write $L_p(\gamma) \geq 0$ to signify the existence of such a sequence $\{\gamma_j\}$.

Remark. The definition of $L_p(\gamma)$ corresponds to what is denoted by L_0 in [Hörmander 1985b, page 97], when $\gamma = [a, b] \times \{w_0\}$ is given by (2-6) and

$$\operatorname{Im} p(a, w_0) < 0 < \operatorname{Im} p(b, w_0). \tag{2-8}$$

To prove this claim, we begin by showing that $L_p(\gamma) \leq L_0$, after having properly defined L_0. To this end, let $\tilde{\gamma} = [\tilde{a}, \tilde{b}] \times \{\tilde{w}\}$ be a bicharacteristic of $\operatorname{Re} p$ such that $\operatorname{Im} p$ changes sign on $\tilde{\gamma}$. For w close to w_0 we set

$$\mathscr{L}_p(\tilde{\gamma}, w) = \inf\{t - s : \tilde{a} < s < t < \tilde{b},\ \operatorname{Im} p(\tilde{a}, w) < 0 < \operatorname{Im} p(\tilde{b}, w)\}.$$

(Using the notation in [Hörmander 1985b, page 97] we would have $\mathscr{L}_p(\gamma, w) = L(x', \xi')$ if $w = (x', 0, \xi')$.) Then

$$L_0 = \liminf_{w\to w_0} \mathscr{L}_p(\gamma, w).$$

By an adaptation of the arguments in [Hörmander 1985b, page 97] it follows from the definition of L_0 that we can find a sequence $\{\gamma_j\}$ of bicharacteristics of $\operatorname{Re} p$ with $\gamma_j = [a_j, b_j] \times \{w_j\}$ such that

$$\operatorname{Im} p(a_j, w_j) < 0 < \operatorname{Im} p(b_j, w_j) \quad \text{for all } j,$$

where $\lim w_j = w_0$ and the limits $a_0 = \lim a_j$ and $b_0 = \lim b_j$ exist, belong to the interval (a, b) and satisfy $b_0 - a_0 = L_0$. If we for each j apply Lemma 2.5 to γ_j we obtain a sequence of bicharacteristics $\Gamma_j \subset \gamma_j$ of $\operatorname{Re} p$ such that $\operatorname{Im} p$ strongly changes sign from $-$ to $+$ on Γ_j, where $|\Gamma_j| = \mathscr{L}_p(\gamma_j, w_j) < |\gamma_j|$. Clearly $\Gamma_j \dashrightarrow \gamma$ as $j \to \infty$. Since $a < a_j \le b_j < b$ if j is sufficiently large it follows that for such j we have $\mathscr{L}_p(\gamma, w_j) \le \mathscr{L}_p(\gamma_j, w_j)$ by definition. This implies

$$\begin{aligned} L_0 = \liminf_{w \to w_0} \mathscr{L}_p(\gamma, w) &\le \liminf_{j\to\infty} \mathscr{L}_p(\gamma, w_j) \\ &\le \liminf_{j\to\infty} |\Gamma_j| \le \limsup_{j\to\infty} |\Gamma_j| \le \lim_{j\to\infty} |\gamma_j| = L_0, \end{aligned} \tag{2-9}$$

so $|\Gamma_j| \to L_0$ as $j \to \infty$. Thus $L_p(\gamma) \le L_0$.

For the reversed inequality, suppose $\{\tilde\gamma_j\}$ is any sequence satisfying the properties of Definition 2.9, with $\tilde\gamma_j = [\tilde a_j, \tilde b_j] \times \{\tilde w_j\}$. By assumption we have $\operatorname{Im} p(\tilde a_j, \tilde w_j) = \operatorname{Im} p(\tilde b_j, \tilde w_j) = 0$ for all j, which together with (2-8) and a continuity argument implies the existence of a positive integer j_0 such that

$$a < \tilde a_j \le \tilde b_j < b \quad \text{for all } j \ge j_0.$$

If $\tilde\gamma_{j,\delta} = [\tilde a_j - \delta, \tilde b_j + \delta] \times \{\tilde w_j\}$, this means that for small $\delta > 0$ and sufficiently large j we have

$$\mathscr{L}_p(\gamma, \tilde w_j) \le \mathscr{L}_p(\tilde\gamma_{j,\delta}, \tilde w_j).$$

Since $\operatorname{Im} p$ strongly changes sign from $-$ to $+$ on $\tilde\gamma_j$, the infimum in the right side exists for every $\delta > 0$ and is bounded from above by $\tilde b_j - \tilde a_j + 2\delta$. Taking the limit as $\delta \to 0$ yields $\mathscr{L}_p(\gamma, \tilde w_j) \le |\tilde\gamma_j|$. Since $\tilde w_j \to w_0$ as $j \to \infty$ the definition of L_0 now gives

$$L_0 \le \liminf_{j\to\infty} \mathscr{L}_p(\gamma, \tilde w_j) \le \liminf_{j\to\infty} |\tilde\gamma_j|, \tag{2-10}$$

and since the sequence $\{\tilde\gamma_j\}$ was arbitrary, we obtain $L_0 \le L_p(\gamma)$ by Definition 2.9. This proves the claim.

When no confusion can occur we will omit the dependence on p in Definition 2.9. We note that if $L_p(\gamma)$ exists, then $L_p(\gamma) \le |\gamma|$ by definition. Also, if $\operatorname{Im} p$ strongly changes sign from $-$ to $+$ on γ then Lemma 2.7 implies that the conditions of Definition 2.9 are satisfied. This proves the first part of the following result.

Corollary 2.10. *Let $\gamma = [a, b] \times \{w_0\}$ be a bicharacteristic of* $\operatorname{Re} p = \xi_1$. *If* $\operatorname{Im} p$ *strongly changes sign from $-$ to $+$ on γ, then $0 \le L_p(\gamma) \le |\gamma|$. Moreover, for every $\delta, \varepsilon > 0$ there exists a bicharacteristic $\tilde\gamma = \tilde\gamma_{\delta,\varepsilon}$ of* $\operatorname{Re} p$ *with*

$$\tilde\gamma = [\tilde a, \tilde b] \times \{\tilde w\}, \quad a - \varepsilon < \tilde a \le \tilde b < b + \varepsilon, \quad |\tilde w - w_0| < \varepsilon,$$

such that $\operatorname{Im} p$ *strongly changes sign from $-$ to $+$ on $\tilde\gamma$ and $|\tilde\gamma| < L_p(\gamma) + \delta$.*

Proof. The existence of the sequence $\{\Gamma_j\}$ in the preceding remark can after some adjustments be used to prove the second part of Corollary 2.10, but we prefer the following direct proof.

Given $\delta > 0$ we can by Definition 2.9 find a sequence $\gamma_j = [a_j, b_j] \times \{w_j\}$ of bicharacteristics of $\operatorname{Re} p$ such that $\gamma_j \dashrightarrow \gamma$ as $j \to \infty$, $\operatorname{Im} p$ strongly changes sign from $-$ to $+$ on γ_j and $\liminf_{j\to\infty} |\gamma_j| < L(\gamma) + \delta$. After reducing to a subsequence we may assume $|\gamma_j| < L(\gamma) + \delta$ for all j. We have $\liminf_{j\to\infty} a_j \geq a$, so for every ε there exists a $j_1(\varepsilon)$ such that $a_j > a - \varepsilon$ for all $j \geq j_1$. Similarly there exists a $j_2(\varepsilon)$ such that $b_j < b + \varepsilon$ for all $j \geq j_2$. Also, $w_j \to w_0$ as $j \to \infty$, so there exists a $j_3(\varepsilon)$ such that $|w_j - w_0| < \varepsilon$ for all $j \geq j_3$. Hence we can take $\tilde{\gamma} = \gamma_{j_0}$, where $j_0 = \max(j_1, j_2, j_3)$. □

Consider now the general case when $\operatorname{Im} qp$ changes sign from $-$ to $+$ on a bicharacteristic $\gamma \subset T^*(X) \smallsetminus 0$ of $\operatorname{Re} qp$, where $q \neq 0$, that is, (2-5) holds. In view of the proof of Lemma 2.5 we can by means of (2-7) define a minimality property of a subset of the curve γ in the following sense.

Definition 2.11. Let $I \subset \mathbb{R}$ be a compact interval possibly reduced to a point and let $\tilde{\gamma} : I \to T^*(X) \smallsetminus 0$ be a characteristic point or a compact one-dimensional bicharacteristic interval of the homogeneous function $p \in C^\infty(T^*(X) \smallsetminus 0)$. Suppose that there exists a function $q \in C^\infty(T^*(X) \smallsetminus 0)$ and a C^∞ homogeneous canonical transformation χ from an open conic neighborhood V of

$$\Gamma = \{(x_1, 0, \varepsilon_n) : x_1 \in I\} \subset T^*(\mathbb{R}^n)$$

to an open conic neighborhood $\chi(V) \subset T^*(X) \smallsetminus 0$ of $\tilde{\gamma}(I)$ such that

(i) $\chi(x_1, 0, \varepsilon_n) = \tilde{\gamma}(x_1)$ and $\operatorname{Re} \chi^*(qp) = \xi_1$ in V,

(ii) $L_{\chi^*(qp)}(\Gamma) = |\Gamma|$.

Then we say that $\tilde{\gamma}(I)$ is a minimal characteristic point or a minimal bicharacteristic interval if $|I| = 0$ or $|I| > 0$, respectively.

The definition of the arclength is of course dependent of the choice of Riemannian metric on $T^*(\mathbb{R}^n)$. However, since we are only using the arclength to compare curves where one is contained within the other and both are parametrizable through condition (i), the results here and Definition 2.11 in particular are independent of the chosen metric. By choosing a Riemannian metric on $T^*(X)$, one could therefore define the minimality property given by Definition 2.11 through the corresponding arclength in $T^*(X)$ directly, although there, the notion of convergence of curves is somewhat trickier. We shall not pursue this any further.

Note that condition (i) implies that $q \neq 0$ and $\operatorname{Re} H_{qp} \neq 0$ on $\tilde{\gamma}$, and that by definition, a minimal bicharacteristic interval is a compact one-dimensional bicharacteristic interval. Moreover, if $\operatorname{Im} qp$ changes sign from $-$ to $+$ on a bicharacteristic $\gamma \subset T^*(X) \smallsetminus 0$ of $\operatorname{Re} qp$, where $q \neq 0$, then we can always find a minimal characteristic point $\tilde{\gamma} \in \gamma$ or a minimal bicharacteristic interval $\tilde{\gamma} \subset \gamma$. In view of the proof of Lemma 2.5, this follows from the conclusion of the extensive remark beginning on page 432 together with (2-9). The following proposition shows that this continues to hold even when the assumption (2-5) is relaxed in the sense of Definition 2.9. We will state this result only in the (very weak) generality needed here.

Proposition 2.12. *Let $\gamma = [a, b] \times \{w_0\}$ be a bicharacteristic of* $\operatorname{Re} p = \xi_1$, *and assume that $L(\gamma) \geq 0$. Then there exists a minimal characteristic point $\Gamma \in \gamma$ of p or a minimal bicharacteristic interval $\Gamma \subset \gamma$ of p of length $L(\gamma)$ if $L(\gamma) = 0$ or $L(\gamma) > 0$, respectively. If $\Gamma = [a_0, b_0] \times \{w_0\}$ and $a_0 < b_0$, that is, $L(\gamma) > 0$, then*

$$\operatorname{Im} p^{(\beta)}_{(\alpha)}(t, w_0) = 0 \tag{2-11}$$

for all α, β with $\beta_1 = 0$ if $a_0 \leq t \leq b_0$. Conversely, if γ is a minimal characteristic point or a minimal bicharacteristic interval then $L(\gamma) = |\gamma|$.

Lemma 2.13. *Let γ and γ_j for $j \geq 1$ be bicharacteristics of* $\operatorname{Re} p = \xi_1$, *and assume that* $\operatorname{Im} p$ *strongly changes sign from $-$ to $+$ on γ_j for each j. If $\gamma_j \dashrightarrow \gamma$ as $j \to \infty$ then $L(\gamma) \leq \liminf_{j\to\infty} L(\gamma_j)$.*

Proof. Let $\gamma_j = [a_j, b_j] \times \{w_j\}$ and $\gamma = [a, b] \times \{w_0\}$. Since $\operatorname{Im} p$ strongly changes sign from $-$ to $+$ on γ_j we can by Corollary 2.10 for each j find a bicharacteristic $\tilde{\gamma}_j = [\tilde{a}_j, \tilde{b}_j] \times \{\tilde{w}_j\}$ of $\operatorname{Re} p$ with

$$a_j - 1/j < \tilde{a}_j \leq \tilde{b}_j < b_j + 1/j \quad \text{and} \quad |\tilde{w}_j - w_j| < 1/j,$$

such that $\operatorname{Im} p$ strongly changes sign from $-$ to $+$ on $\tilde{\gamma}_j$ and $|\tilde{\gamma}_j| < L(\gamma_j) + 1/j$. Now $|\tilde{w}_j - w_0| \leq |\tilde{w}_j - w_j| + |w_j - w_0|$, and since $\liminf_{j\to\infty} \tilde{a}_j \geq \liminf_{j\to\infty}(a_j - 1/j) \geq a$ and correspondingly for $\tilde{b}_j$, we find that $\tilde{\gamma}_j \dashrightarrow \gamma$ as $j \to \infty$. Thus

$$L(\gamma) \leq \liminf_{j\to\infty} |\tilde{\gamma}_j| \leq \liminf_{j\to\infty} (L(\gamma_j) + 1/j). \qquad \square$$

Proof of Proposition 2.12. We may without loss of generality assume that $w_0 = (0, \varepsilon_n) \in \mathbb{R}^{2n-1}$. The last statement is then an immediate consequence of Definition 2.11. To prove the theorem it then also suffices to show that we can find a characteristic point $\Gamma \in \gamma$ of p, or a compact one-dimensional bicharacteristic interval $\Gamma \subset \gamma$ of p of length $L(\gamma)$, with the property that in any neighborhood of Γ there is a bicharacteristic of $\operatorname{Re} p$ where $\operatorname{Im} p$ strongly changes sign from $-$ to $+$. This is done by adapting the arguments in [Hörmander 1985b, page 97], which also yields (2-11).

For small $\delta > 0$ we can find $\varepsilon(\delta)$ with $0 < \varepsilon < \delta$ such that $L(\tilde{\gamma}) > L(\gamma) - \delta/2$ for any bicharacteristic $\tilde{\gamma} = [\tilde{a}, \tilde{b}] \times \{\tilde{w}\}$ with $a - \varepsilon < \tilde{a} \leq \tilde{b} < b + \varepsilon$ and $|\tilde{w} - w_0| < \varepsilon$ such that $\operatorname{Im} p$ strongly changes sign from $-$ to $+$ on $\tilde{\gamma}$. Indeed, otherwise there would exist a $\delta > 0$ such that for each (sufficiently large) k we can find a bicharacteristic $\gamma_k = [a_k, b_k] \times \{w_k\}$ with $a - 1/k < a_k \leq b_k < b + 1/k$ and $|w_k - w_0| < 1/k$ such that $\operatorname{Im} p$ strongly changes sign from $-$ to $+$ on γ_k and $L(\gamma_k) \leq L(\gamma) - \delta/2$. This implies that $\gamma_k \dashrightarrow \gamma$ as $k \to \infty$, so by Lemma 2.13 we obtain

$$L(\gamma) \leq \liminf_{k\to\infty} L(\gamma_k) \leq L(\gamma) - \delta/2,$$

a contradiction. Since $L(\gamma) \geq 0$ we have by Corollary 2.10 for some $|w_\delta - w_0| < \varepsilon$ and $a - \varepsilon < a_\delta \leq b_\delta < b + \varepsilon$ with $w_\delta = (x'_\delta, 0, \xi'_\delta)$ that $\operatorname{Im} p$ strongly changes sign from $-$ to $+$ on the bicharacteristic $\gamma_\delta = [a_\delta, b_\delta] \times \{w_\delta\}$, and $|\gamma_\delta| < L(\gamma) + \delta/4$. Thus,

$$L(\gamma) - \delta/2 < |\gamma_\delta| < L(\gamma) + \delta/4. \tag{2-12}$$

We claim that $\operatorname{Im} p$ and all derivatives with respect to x' and ξ' must vanish at (t, w_δ) if $a_\delta+\delta<t<b_\delta-\delta$. Indeed, by Lemma 2.7 we can find a $\rho>0$, $a_\delta-\delta/4<s_-<a_\delta$ and $b_\delta<s_+<b_\delta+\delta/4$ such that

$$\operatorname{Im} p(s_-, w)<0<\operatorname{Im} p(s_+, w) \quad \text{for all } |w-w_\delta|<\rho.$$

If $\operatorname{Im} p$ and all derivatives with respect to x' and ξ' do not vanish at (t, w_δ) if $a_\delta+\delta<t<b_\delta-\delta$, then we can choose $w=(x', 0, \xi')$ so that $|w-w_\delta|<\rho$, $|w-w_0|<\varepsilon$ and $\operatorname{Im} p(t, w)\neq 0$ for some $a_\delta+\delta<t<b_\delta-\delta$. It follows that the required sign change of $\operatorname{Im} p(x_1, w)$ must occur on one of the intervals (s_-, t) and (t, s_+), which are shorter than $L(\gamma)-\delta/2$. This contradiction proves the claim.

Now choose a sequence $\delta_j\to 0$ as $j\to\infty$ such that $\lim a_{\delta_j}$ and $\lim b_{\delta_j}$ exist. If we denote these limits by a_0 and b_0, respectively, then $L(\gamma)=b_0-a_0$ by (2-12), and (2-11) holds if $a_0<b_0$. In particular, if $a_0<b_0$ then

$$H_p(\gamma(t))=(1+i\partial \operatorname{Im} p(\gamma(t))/\partial\xi_1)\gamma'(t) \quad \text{for } a_0\le t\le b_0,$$

so if $\Gamma=\{(t, w_0): t\in[a_0, b_0]\}$ then Γ is a compact one-dimensional bicharacteristic interval of p with the function c in Definition 2.6 given by

$$c(t)=(1+i\partial \operatorname{Im} p(\Gamma(t))/\partial\xi_1)^{-1}. \qquad \square$$

Proposition 2.12 allows us to make some additional comments on the implications of Definition 2.11. With the notation in the definition, we note that condition (ii) implies that there exists a sequence $\{\Gamma_j\}$ of bicharacteristics of $\operatorname{Re}\chi^*(qp)$ on which $\operatorname{Im}\chi^*(qp)$ strongly changes sign from $-$ to $+$, such that $\Gamma_j\to\Gamma$ as $j\to\infty$. By our choice of terminology, the sequence $\{\Gamma_j\}$ may simply be a sequence of points when $L(\Gamma)=0$. Conversely, if $\{\Gamma_j\}$ is a point sequence then $L(\Gamma)=0$. Also note that if $\tilde\gamma(I)$ is minimal, and condition (i) in Definition 2.11 is satisfied for some other choice of maps q' and χ', then condition (ii) also holds for q' and χ'; in other words,

$$L_{\chi^*(qp)}(\Gamma)=|\Gamma|=L_{(\chi')^*(q'p)}(\Gamma).$$

This follows by an application of Proposition 2.12 together with [Hörmander 1985b, Lemma 26.4.10]. It is then also clear that $\tilde\gamma(I)$ is a minimal characteristic point or a minimal bicharacteristic interval of the homogeneous function $p\in C^\infty(T^*(X)\setminus 0)$ if and only if $\Gamma(I)$ is a minimal characteristic point or a minimal bicharacteristic interval of $\chi^*(qp)\in C^\infty(T^*(\mathbb{R}^n)\setminus 0)$ for any maps q and χ satisfying condition (i) in Definition 2.11.

The proof of [Hörmander 1985b, Theorem 26.4.7] stating that condition (Ψ) is necessary for local solvability relies on the imaginary part of the principal symbol satisfying (2-11). By Proposition 2.12, it is clear that (2-11) holds on a minimal bicharacteristic interval Γ in the case $q=1$ and $\operatorname{Re} p=\xi_1$. However, we shall require that we can find bicharacteristics arbitrarily close to Γ for which the following stronger result is applicable, at least if $\operatorname{Im} p$ does not depend on ξ_1 as is the case for the standard normal form. This will be made precise below.

Proposition 2.14. *Let* $p = \xi_1 + i \operatorname{Im} p$. *Assume that* $\operatorname{Im} p$ *strongly changes sign from* $-$ *to* $+$ *on* $\gamma = [a, b] \times \{w\}$ *and that* $L(\gamma) \geq |\gamma| - \varrho$ *for some* $0 < \varrho < |\gamma|/2$. *If* $\operatorname{Im} p$ *does not depend on* ξ_1 *then for any* $\kappa > \varrho$ *we find that* $\operatorname{Im} p$ *vanishes identically in a neighborhood of* $I_\kappa \times \{w\}$, *where* $I_\kappa = [a+\kappa, b-\kappa]$.

The statement would of course be void if the hypotheses hold only for $\varrho \geq |\gamma|/2$, for then $I_\kappa = \varnothing$.

Proof. If the statement is false, there exists a $\kappa > 0$ such that $\operatorname{Im} p \not\equiv 0$ near $I_\kappa \times \{w\}$. Thus there exists a sequence $(s_j, w_j) \dashrightarrow I_\kappa \times \{w\}$ such that $\operatorname{Im} p(s_j, w_j) \neq 0$ for all j. Since $\operatorname{Im} p$ does not depend on ξ_1 we can choose w_j to have ξ_1 coordinate equal to zero for all j, so that (s_j, w_j) is contained in a bicharacteristic of $\operatorname{Re} p$. We may choose a subsequence so that for some $s \in I_\kappa$ we have $|s_j - s| \to 0$ and $|w_j - w| \to 0$ monotonically, and either $\operatorname{Im} p(s_j, w_j) > 0$ or $-\operatorname{Im} p(s_j, w_j) > 0$ for all j. We shall consider the case with positive sign, the negative case works similarly.

Choose $\delta < (\kappa - \varrho)/3$ and use Lemma 2.7. We find that there exists $a - \delta < s_- < a$ and $\varepsilon > 0$ such that $\operatorname{Im} p(s_-, v) < 0$ for any $|v - w| < \varepsilon$. Choose $k > 0$ so that $|s_j - s| < \delta$ and $|w_j - w| < \varepsilon$ when $j > k$. Then $t \mapsto \operatorname{Im} p(t, w_j)$ changes sign from $-$ to $+$ on $I_j = [s_-, s_j]$, which has length

$$|I_j| = s_j - s_- \leq |s_j - s| + s - a + a - s_- < |\gamma| - \kappa + 2\delta < |\gamma| - \varrho - \delta.$$

If we for each j apply Lemma 2.5 to $I_j \times \{w_j\}$ and let $j \to \infty$ we obtain a contradiction to the hypothesis $L(\gamma) \geq |\gamma| - \varrho$. ☐

One could state Proposition 2.14 without the condition that the imaginary part is independent of ξ_1. The invariant statement would then be that the restriction of the imaginary part to the characteristic set of the real part vanishes in a neighborhood of γ.

The fact that Proposition 2.14 assumes that $\operatorname{Im} p$ strongly changes sign from $-$ to $+$ on γ means that the conditions are not in general satisfied when γ is a minimal bicharacteristic interval. As mentioned above, we will instead show that arbitrarily close to a minimal bicharacteristic interval one can always find bicharacteristics for which Proposition 2.14 is applicable. Before we state the results we introduce a helpful definition together with some (perhaps contrived but illustrative) examples.

Definition 2.15. A minimal bicharacteristic interval $\Gamma = [a_0, b_0] \times \{w_0\} \subset T^*(\mathbb{R}^n) \smallsetminus 0$ of the homogeneous function $p = \xi_1 + i \operatorname{Im} p$ of degree 1 is said to be ϱ-minimal if there exists a $\varrho \geq 0$ such that $\operatorname{Im} p$ vanishes in a neighborhood of $[a_0 + \kappa, b_0 - \kappa] \times \{w_0\}$ for any $\kappa > \varrho$.

By a 0-minimal bicharacteristic interval Γ we thus mean a minimal bicharacteristic interval such that the imaginary part vanishes in a neighborhood of any proper closed subset of Γ. Note that this does not hold for minimal bicharacteristic intervals in general.

Example 2.16. Let $f \in C^\infty(\mathbb{R})$ be given by

$$f(t) = \begin{cases} -e^{-1/t^2} & \text{if } t < 0, \\ 0 & \text{if } 0 \leq t \leq 2, \\ e^{-1/(t-2)^2} & \text{if } t > 2 \end{cases} \tag{2-13}$$

and let $\phi \in C^\infty(\mathbb{R})$ be a smooth cutoff function with $\operatorname{supp}\phi = [0,2]$ such that $\phi > 0$ on $(0,2)$. If $\xi = (\xi_1, \xi')$ then

$$p_1(x,\xi) = \xi_1 + i|\xi'|(f(x_1) + x_2\phi(x_1))$$

is homogeneous of degree 1. If we write $x = (x_1, x_2, x'')$ then for any fixed $(x'', \xi') \in \mathbb{R}^{n-2} \times \mathbb{R}^{n-1}$ with $\xi' \neq 0$ we find that $\{(x_1, x_2, x'', 0, \xi') : x_1 = a, x_2 = c\}$ is a minimal characteristic point of p_1 if $c \geq 0$ and $a = 0$ or if $c \leq 0$ and $a = 2$. Note that if $\xi' \neq 0$ then $\operatorname{Im} p_1$ changes sign from $-$ to $+$ on the bicharacteristic $\gamma(x_1) = \{(x_1, 0, x'', 0, \xi')\}$ of $\operatorname{Re} p_1$, but that none of the points $\{\gamma(x_1) : 0 < x_1 < 2\}$ are minimal characteristic points.[1] On the other hand, if f is given by (2-13) let

$$h(x,\xi') = \begin{cases} |\xi'|f(x_1-1)e^{1/x_2} & \text{if } x_2 < 0, \\ 0 & \text{if } x_2 = 0, \\ |\xi'|f(x_1)e^{-1/x_2} & \text{if } x_2 > 0 \end{cases}$$

be the imaginary part of $p_2(x,\xi)$. If $\operatorname{Re} p_2 = \xi_1$ then p_2 is homogeneous of degree 1 and

$$\Gamma_c = \{(x_1, x_2, x'', 0, \xi') : x_2 = c, x_1 \in I_c\}$$

is a minimal bicharacteristic interval of p_2 for any $(x'', \xi') \in \mathbb{R}^{n-2} \times \mathbb{R}^{n-1}$ with $\xi' \neq 0$ if $c \geq 0$ and $I_c = [0,2]$ or if $c \leq 0$ and $I_c = [1,3]$. Moreover, if $c \lessgtr 0$ then Γ_c is a 0-minimal bicharacteristic interval. However, there is no $\varrho > 0$ such that the minimal bicharacteristic interval $\Gamma = \{(x_1, 0, x'', 0, \xi') : 0 \leq x_1 \leq 2\}$ is ϱ-minimal. The same holds for the minimal bicharacteristic interval $\tilde{\Gamma} = \{(x_1, 0, x'', 0, \xi') : 1 \leq x_1 \leq 3\}$. Figure 1 shows a cross-section of the characteristic sets of $\operatorname{Im} p_1$ and $\operatorname{Im} p_2$.

Lemma 2.17. *Let $p = \xi_1 + i \operatorname{Im} p$, and assume that $L(\gamma) > 0$ and that* $\operatorname{Im} p$ *does not depend on ξ_1. Then one can find $\tilde{\gamma}_j \subset \gamma_j \dashrightarrow \gamma$ such that $|\tilde{\gamma}_j| \to L(\gamma)$,* $\operatorname{Im} p$ *strongly changes sign from $-$ to $+$ on γ_j and* $\operatorname{Im} p$ *vanishes in a neighborhood of $\tilde{\gamma}_j$.*

[1] If the factor x_2 in $\operatorname{Im} p_1$ is raised to the power 3 for example, then it turns out that $\{\gamma(x_1) : 0 < x_1 < 2\}$ is a one-dimensional bicharacteristic interval of p_1, and not only a bicharacteristic of the real part. It is obviously not minimal though, nor does it contain any minimal characteristic points.

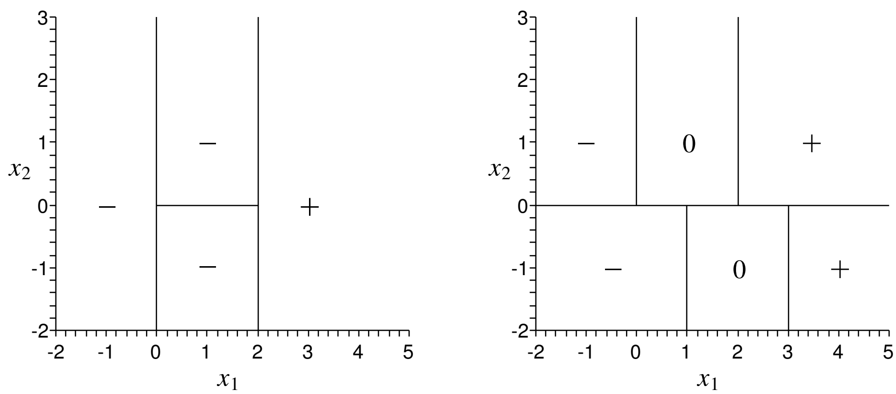

Figure 1. Cross-sections of the characteristic sets of $\operatorname{Im} p_1$ and $\operatorname{Im} p_2$, respectively.

Note that the conditions imply that $\tilde{\gamma}_j \dashrightarrow \gamma$ as $j \to \infty$.

Proof. Choose $\gamma_j \dashrightarrow \gamma$ when $j \to \infty$ as in the proof of Proposition 2.12, so that $\operatorname{Im} p$ strongly changes sign from $-$ to $+$ on γ_j and $L(\gamma) = \lim_{j\to\infty} |\gamma_j|$. By Lemma 2.13 and Corollary 2.10 we have

$$L(\gamma) \le \liminf_{j\to\infty} L(\gamma_j) \le \liminf_{j\to\infty} |\gamma_j| = L(\gamma).$$

Thus we can for every $\varepsilon > 0$ choose j so that $|L(\gamma) - |\gamma_j|| < \varepsilon$ and $|L(\gamma_j) - |\gamma_j|| < \varepsilon$. If we choose $\varepsilon < L(\gamma)/5$ then

$$2\varepsilon < (L(\gamma) - \varepsilon)/2 < |\gamma_j|/2.$$

Hence, if $\gamma_j = [a_j, b_j] \times w_j$ then by using Proposition 2.14 on γ_j we find that $\operatorname{Im} p$ vanishes identically in a neighborhood of $\tilde{\gamma}_j = [a_j + 2\varepsilon, b_j - 2\varepsilon] \times \{w_j\}$. Now choose a sequence $\varepsilon_k \to 0$ as $k \to \infty$. Then $\tilde{\gamma}_{j(k)} \subset \gamma_{j(k)}$ and assuming as we may that $j(k) > j(k')$ if $k > k'$ we obtain $|\tilde{\gamma}_{j(k)}| \to L(\gamma)$ as $k \to \infty$, which completes the proof. □

If $\Gamma \subset \gamma$ is a minimal bicharacteristic interval in $T^*(\mathbb{R}^n) \setminus 0$ of the homogeneous function $p = \xi_1 + i \operatorname{Im} p$ of degree 1, where the imaginary part is independent of ξ_1, then by Definition 2.11 and Proposition 2.12 we have $0 < |\Gamma| = L(\Gamma)$. By the proof of Lemma 2.17 there exists a sequence $\gamma_j \to \Gamma$ of bicharacteristics of $\operatorname{Re} p$ such that $\operatorname{Im} p$ strongly changes sign from $-$ to $+$ on γ_j and vanishes identically in a neighborhood of a subinterval $\tilde{\gamma}_j \subset \gamma_j$. Moreover, $\tilde{\gamma}_j \to \Gamma$ as $j \to \infty$. By Lemma 2.13 we have $L(\gamma_j) > 0$ for sufficiently large j, so according to Proposition 2.12 we can for each such j find a minimal bicharacteristic interval $\Gamma_j \subset \gamma_j$. We have $\gamma_j \to \Gamma$ as $j \to \infty$ and since

$$|\Gamma| = L(\gamma) \le \liminf_{j\to\infty} L(\gamma_j) = \liminf_{j\to\infty} |\Gamma_j| \le \limsup_{j\to\infty} |\Gamma_j| \le \lim_{j\to\infty} |\gamma_j| = |\Gamma|,$$

it follows that $\Gamma_j \to \Gamma$ as $j \to \infty$. Since also $\tilde{\gamma}_j \subset \gamma_j$ and $\tilde{\gamma}_j \to \Gamma$ as $j \to \infty$, the intersection $\tilde{\gamma}_j \cap \Gamma_j$ must be nonempty for large j. For such j it follows that $\tilde{\gamma}_j$ must be a proper subinterval of Γ_j, for if not, this would contradict the fact that Γ_j is a minimal bicharacteristic interval. Hence we can find a sequence $\{\varrho_j\}$ of positive numbers with $\varrho_j \to 0$ as $j \to \infty$, such that Γ_j is a ϱ_j-minimal bicharacteristic interval. We have thus proved the following theorem, which concludes our study of the bicharacteristics.

Theorem 2.18. *If Γ is a minimal bicharacteristic interval in $T^*(\mathbb{R}^n) \setminus 0$ of the homogeneous function $p = \xi_1 + i \operatorname{Im} p$ of degree 1, where the imaginary part is independent of ξ_1, then there exists a sequence $\{\Gamma_j\}$ of ϱ_j-minimal bicharacteristic intervals of p such that $\Gamma_j \to \Gamma$ and $\varrho_j \to 0$ as $j \to \infty$.*

We can now state our main theorem, which yields necessary conditions for inclusion relations between the ranges of operators that fail to be microlocally solvable.

Theorem 2.19. *Let $K \subset T^*(X) \setminus 0$ be a compactly based cone. Let $P \in \Psi^k_{\mathrm{cl}}(X)$ and $Q \in \Psi^{k'}_{\mathrm{cl}}(X)$ be properly supported pseudodifferential operators such that the range of Q is microlocally contained in the range of P at K, where P is an operator of principal type in a conic neighborhood of K. Let p_k be the homogeneous principal symbol of P, and let $I = [a_0, b_0] \subset \mathbb{R}$ be a compact interval possibly reduced to a point. Suppose that K contains a conic neighborhood of $\gamma(I)$, where $\gamma : I \to T^*(X) \setminus 0$ is either*

(a) *a minimal characteristic point of p_k, or*

(b) *a minimal bicharacteristic interval of p_k with injective regular projection in $S^*(X)$.*

Then there exists a pseudodifferential operator $E \in \Psi_{\mathrm{cl}}^{k'-k}(X)$ such that the terms in the asymptotic sum of the symbol of $Q - PE$ have vanishing Taylor coefficients at $\gamma(I)$.

The hypotheses of Theorem 2.19 imply that P is not solvable at the cone K. Indeed, solvability at $K \subset T^*(X) \smallsetminus 0$ implies solvability at any smaller closed cone, and in view of Definition 2.11 it follows by [Hörmander 1985b, Theorem 26.4.7′] together with [Hörmander 1985b, Proposition 26.4.4] that P is not solvable at the cone generated by $\gamma(I)$. Conversely, suppose that P is an operator of principal type that is not microlocally solvable in any neighborhood of a point $(x_0, \xi_0) \in T^*(X) \smallsetminus 0$. Then the principal symbol p_k fails to satisfy condition (1-5) in every neighborhood of (x_0, ξ_0) by [Dencker 2006, Theorem 1.1]. In view of the alternative version of condition (1-5) given by [Hörmander 1985b, Theorem 26.4.12], it is then easy to see using [Hörmander 1985a, Theorem 21.3.6] and [Hörmander 1985b, Lemma 26.4.10] that (x_0, ξ_0) is a minimal characteristic point of p_k, so Theorem 2.19 applies there.

We also mention that if P is of principal type and γ is a minimal bicharacteristic interval of the principal symbol p_k contained in a curve along which p_k fails to satisfy condition (1-5), then γ has injective regular projection in $S^*(X)$ by the proof of [Hörmander 1985b, Theorem 26.4.12].

Remark. As pointed out in the introduction, we cannot hope to obtain a result such as Theorem 2.19 for solvable nonelliptic operators in general. Indeed, Example 2.2 shows that if $X \subset \mathbb{R}^n$ is open, and $K \subset T^*(X) \smallsetminus 0$ is a compactly based cone, then the range of D_2 is microlocally contained in the range of D_1 at K. If there were to exist a pseudodifferential operator $e(x, D) \in \Psi_{\mathrm{cl}}^0(X)$ such that all the terms in the symbol of $R(x, D) = D_2 - D_1 \circ e(x, D)$ have vanishing Taylor coefficients at a point $(x_0, \xi_0) \in K$ contained in a bicharacteristic of the principal symbol $\sigma(D_1) = \xi_1$ of D_1, then in particular this would hold for the principal symbol

$$\sigma(R)(x, \xi) = \xi_2 - \xi_1 e_0(x, \xi),$$

if e_0 denotes the principal symbol of $e(x, D)$. However, taking the ξ_2 derivative of the equation above and evaluating at (x_0, ξ_0) then immediately yields the contradiction $0 = 1$ since (x_0, ξ_0) belongs to the hypersurface $\xi_1 = 0$.

In the proof of the theorem we may assume that P and Q are operators of order 1. In fact, the discussion following Definition 2.1 shows that if the conditions of Theorem 2.19 hold and $Q_1 \in \Psi_{\mathrm{cl}}^{k-k'}(X)$ and $Q_2 \in \Psi_{\mathrm{cl}}^{1-k}(X)$ are properly supported, then the range of $Q_2 Q Q_1 \in \Psi_{\mathrm{cl}}^1(X)$ is microlocally contained in the range of $Q_2 P \in \Psi_{\mathrm{cl}}^1(X)$ at K. If the theorem holds for operators of the same order k then there exists an operator $E \in \Psi_{\mathrm{cl}}^0(X)$ such that all the terms in the asymptotic expansion of the symbol of $QQ_1 - PE$ have vanishing Taylor coefficients at $\gamma(I)$. If we choose Q_1 to be elliptic, then we can find a parametrix Q_1^{-1} of Q_1 such that

$$Q - PEQ_1^{-1} \equiv (QQ_1 - PE) \circ Q_1^{-1} \mod \Psi^{-\infty}(X)$$

has symbol

$$\sigma_{A\circ Q_1^{-1}}(x,\xi) \sim \sum \partial_\xi^\alpha \sigma_A(x,\xi) D_x^\alpha \sigma_{Q_1^{-1}}(x,\xi)/\alpha! \tag{2-14}$$

with $A = QQ_1 - PE$. Clearly, all the terms in the asymptotic expansion of the symbol of $Q - PEQ_1^{-1}$ then have vanishing Taylor coefficients at $\gamma(I)$, and $E_1 = EQ_1^{-1} \in \Psi_{\mathrm{cl}}^{k'-k}(X)$, so the theorem holds with E replaced by E_1. If the theorem holds for operators of order 1 we can choose Q_2 elliptic and use the same argument to show that if all the terms in the asymptotic expansion of the symbol of $Q_2QQ_1 - Q_2PE$ have vanishing Taylor coefficients at $\gamma(I)$, then the same holds for

$$Q - PEQ_1^{-1} \equiv Q_2^{-1} \circ (Q_2QQ_1 - Q_2PE) \circ Q_1^{-1} \mod \Psi^{-\infty}(X),$$

where Q_2^{-1} is a parametrix of Q_2. Here we use the fact that if $\gamma(I)$ is a minimal characteristic point or a minimal bicharacteristic interval of the principal symbol of P, then this also holds for the principal symbol of Q_2P by Definition 2.11.

For pseudodifferential operators, the property that all terms in the asymptotic expansion of the total symbol have vanishing Taylor coefficients is preserved under conjugation with Fourier integral operators associated with a canonical transformation (see Lemma A.1). Thus we will be able to prove Theorem 2.19 by local arguments and an application of Proposition 2.4.

Let $\gamma : I \to T^*(X)\smallsetminus 0$, with $I = [a_0, b_0] \subset \mathbb{R}$, be the map given by Theorem 2.19. By using [Hörmander 1985a, Theorem 21.3.6] or [Hörmander 1985b, Theorem 26.4.13], when γ is a characteristic point or a one-dimensional bicharacteristic, respectively, we can find a C^∞ canonical transformation χ from a conic neighborhood of $\Gamma = \{(x, \varepsilon_n) : x_1 \in I,\ x' = 0\}$ in $T^*(\mathbb{R}^n)\smallsetminus 0$ to a conic neighborhood of $\gamma(I)$ in $T^*(X)\smallsetminus 0$ and a C^∞ homogeneous function b of degree 0 with no zero on $\gamma(I)$ such that $\chi(x_1, 0, \varepsilon_n) = \gamma(x_1)$ for $x_1 \in I$ and

$$\chi^*(bp_1) = \xi_1 + if(x, \xi'), \tag{2-15}$$

where f is real-valued, homogeneous of degree 1 and independent of ξ_1. Thus, by the hypotheses of Theorem 2.19 one can in any neighborhood of Γ find an interval in the x_1 direction where f changes sign from $-$ to $+$ for increasing x_1. Also, if I is an interval then f vanishes of infinite order on Γ by (2-11), and by Theorem 2.18 there exists a sequence $\{\Gamma_j\}$ of ϱ_j-minimal bicharacteristics of $\chi^*(bp_1)$ such that $\varrho_j \to 0$ and $\Gamma_j \to \Gamma$ as $j \to \infty$.

The existence of the canonical transformation χ together with Proposition 2.4 implies that we can find Fourier integral operators A and B such that the range of BQA is microlocally contained in the range of BPA at a cone K' containing Γ, where the principal symbol of BPA is given by (2-15). In view of Lemma A.1 we may therefore reduce the proof to the case that $P, Q \in \Psi_{\mathrm{cl}}^1(\mathbb{R}^n)$ and the principal symbol p of P is given by (2-15). In accordance with the notation in Proposition 2.4 we will assume that the range of Q is microlocally contained in the range of P at a cone K containing Γ, thus renaming K' to K. If

$$\sigma_Q = q_1 + q_0 + \cdots$$

is the asymptotic sum of homogeneous terms of the symbol of Q, we can then use the Malgrange preparation theorem (see [Hörmander 1983a, Theorem 7.5.6]) to find $e_0, r_1 \in C^\infty$ near Γ such that

$$q_1(x,\xi) = (\xi_1 + if(x,\xi'))e_0(x,\xi) + r_1(x,\xi'),$$

where r_1 is independent of ξ_1. Restricting to $|\xi| = 1$ and extending by homogeneity we can make e_0 and r_1 homogeneous of degree 0 and 1, respectively. The term of degree 1 in the symbol of $Q - P \circ e_0(x, D)$ is $r_1(x,\xi')$. Again, by Malgrange's preparation theorem we can find $e_{-1}, r_0 \in C^\infty$ near Γ such that

$$q_0(x,\xi) - \sigma_0(P \circ e_0(x,D))(x,\xi) = (\xi_1 + if(x,\xi'))e_{-1}(x,\xi) + r_0(x,\xi'),$$

where e_{-1} and r_0 are homogeneous of degree -1 and 0, respectively, and r_0 is independent of ξ_1. The term of degree 0 in the symbol of

$$Q - P \circ e_0(x,D) - P \circ e_{-1}(x,D)$$

is $r_0(x,\xi')$. Repetition of the argument allows us to write

$$Q = P \circ E + R(x, D_{x'}) \tag{2-16}$$

where $\sigma_R(x,\xi') = r_1(x,\xi') + r_0(x,\xi') + \cdots$ is an asymptotic sum of homogeneous terms, all independent of ξ_1. Thus $R(x, D_{x'})$ is a pseudodifferential operator in the $n-1$ variables x' depending on x_1 as a parameter. Furthermore, the range of $R(x, D_{x'})$ is microlocally contained in the range of P at K. Indeed, suppose N is the integer given by Definition 2.1. If $g \in H^{\mathrm{loc}}_{(N)}(\mathbb{R}^n)$, then $Rg = PEg - Qg = Pv - Qg$ for some $v \in \mathscr{D}'(\mathbb{R}^n)$, and there exists a $u \in \mathscr{D}'(\mathbb{R}^n)$ such that

$$K \cap WF(Qg - Pu) = \varnothing.$$

Hence, $WF(P(v-u) - Rg)$ does not meet K, so the range of R is microlocally contained in the range of P at K. We claim that under the assumptions of Theorem 2.19, this implies that all terms in the asymptotic sum of the symbol of the operator $R(x, D_{x'})$ in (2-16) have vanishing Taylor coefficients at Γ, thus proving Theorem 2.19. The proof of this claim will be based on the two theorems stated below. As we have seen, the principal symbol p of P may be assumed to have the normal form given by (2-15). By means of Theorem 2.20 below, we shall also use the fact that an even simpler normal form exists near a point where $p = 0$ and $\{\operatorname{Re} p, \operatorname{Im} p\} \neq 0$. To prove these two theorems, we will use techniques that actually require the lower order terms of P to be independent of ξ_1 near Γ. However, we claim that this may always be assumed. In fact, Malgrange's preparation theorem implies that

$$p_0(x,\xi) = a(x,\xi)(\xi_1 + if(x,\xi')) + b(x,\xi')$$

where a is homogeneous of degree -1 and b homogeneous of degree 0, as demonstrated in the construction of the operators E and R above. The term of degree 0 in the symbol of $(I - a(x,D))P$ is equal to $b(x,\xi')$. Repetition of the argument implies that there exists a classical operator $\tilde{a}(x,D)$ of order -1 such that $(I - \tilde{a}(x,D))P$ has principal symbol $\xi_1 + if(x,\xi')$ and all lower order terms are independent of ξ_1. The microlocal property of pseudodifferential operators immediately implies that the range of $(I - \tilde{a}(x, D))Q$ is microlocally contained in the range of $(I - \tilde{a}(x, D))P$ at K. Hence, if there are operators E and R with

$$R = (I - \tilde{a}(x, D))Q - (I - \tilde{a}(x, D))PE$$

such that all terms in the asymptotic expansion of the symbol of R have vanishing Taylor coefficients at Γ, then this also holds for the symbol of $Q - PE \equiv (I - \tilde{a}(x, D))^{-1}R \bmod \Psi^{-\infty}$, since this property is preserved under composition with elliptic pseudodifferential operators by (2-14).

Theorem 2.20. *Suppose that in a conic neighborhood Ω of*

$$\Gamma' = \{(0, \varepsilon_n)\} \subset T^*(\mathbb{R}^n) \setminus 0$$

P has the form $P = D_1 + ix_1 D_n$ and the symbol of $R(x, D_{x'})$ is given by the asymptotic sum

$$\sigma_R = \sum_{j=0}^{\infty} r_{1-j}(x, \xi'),$$

with r_{1-j} homogeneous of degree $1 - j$ and independent of ξ_1. If there exists a compactly based cone $K \subset T^*(\mathbb{R}^n) \setminus 0$ containing Ω such that the range of R is microlocally contained in the range of P at K, then all the terms in the asymptotic sum of the symbol of R have vanishing Taylor coefficients at Γ'.

Theorem 2.21. *Suppose that in a conic neighborhood Ω of*

$$\Gamma' = \{(x_1, x', 0, \xi') : a \le x_1 \le b\} \subset T^*(\mathbb{R}^n) \setminus 0$$

the principal symbol of P has the form

$$p(x, \xi) = \xi_1 + if(x, \xi'),$$

where f is real-valued and homogeneous of degree 1, and suppose that if $b > a$ then f vanishes of infinite order on Γ' and there exists a $\varrho \ge 0$ such that for any $\varepsilon > \varrho$ one can find a neighborhood of

$$\Gamma'_\varepsilon = \{(x_1, x', 0, \xi') : a + \varepsilon \le x_1 \le b - \varepsilon\}, \tag{2-17}$$

where f vanishes identically. Suppose also that

$$f(x, \xi') = 0 \quad \textit{implies} \quad \partial f(x, \xi')/\partial x_1 \le 0 \tag{2-18}$$

in Ω and that in any neighborhood of Γ' one can find an interval in the x_1 direction where f changes sign from $-$ to $+$ for increasing x_1. Furthermore, suppose that in Ω the symbol of $R(x, D_{x'})$ is given by the asymptotic sum

$$\sigma_R = \sum_{j=0}^{\infty} r_{1-j}(x, \xi')$$

with r_{1-j} homogeneous of degree $1 - j$ and independent of ξ_1. If the lower order terms $p_0, p_{-1}, \ldots$ in the symbol of P are independent of ξ_1 near Γ', and there exists a compactly based cone $K \subset T^*(\mathbb{R}^n) \setminus 0$ containing Ω such that the range of R is microlocally contained in the range of P at K, then all the *terms in the asymptotic sum of the symbol of R have vanishing Taylor coefficients on Γ'_ϱ if $a < b$, and at Γ' if $a = b$.*

Assuming these results for the moment, we can now show how Theorem 2.19 follows.

End of proof of Theorem 2.19. Recall that

$$\Gamma = \{(x_1, 0, \varepsilon_n) : a_0 \le x_1 \le b_0\} \subset T^*(\mathbb{R}^n) \setminus 0.$$

By what we have shown, it suffices to regard the case $Q = PE + R$, where we may assume that the conditions of Theorem 2.21 are all satisfied in a conic neighborhood Ω of Γ, with the exception of (2-18) and the condition concerning the existence of a neighborhood of (2-17) in which f vanishes identically when $a_0 < b_0$. We consider three cases.

(i) Γ is an interval. We then claim that condition (2-18) imposes no restriction. Indeed, if there is no neighborhood of Γ in which (2-18) holds, then there exists a sequence $\{\gamma_j\} = \{(t_j, x'_j, 0, \xi'_j)\}$ such that $a_0 \le \liminf t_j \le \limsup t_j \le b_0$, $(x'_j, \xi'_j) \to (0, \xi^0) \in \mathbb{R}^{2n-2}$,

$$f(t_j, x'_j, \xi'_j) = 0 \quad \text{and} \quad \partial f(t_j, x'_j, \xi'_j)/\partial x_1 > 0 \tag{2-19}$$

for each j. By (2-19) we can choose a sequence $0 < \delta_j \to 0$ such that

$$f(t_j - \delta_j, x'_j, \xi'_j) < 0 < f(t_j + \delta_j, x'_j, \xi'_j).$$

In view of Definition 2.9 we must therefore have $L(\Gamma) = 0$. Since Γ is minimal, this implies that $|\Gamma| = 0$, so $\gamma_j \to \Gamma$. Thus, if there is no neighborhood of Γ in which (2-18) holds, then Γ is a point, and we will in this case use the existence of the sequence $\{\gamma_j\}$ satisfying (2-19) to reduce the proof of Theorem 2.19 to Theorem 2.20, as demonstrated in case (iii) below. In the present case however, Γ is assumed to be an interval, so there exists a neighborhood $\mathcal{U}$ of Γ in which (2-18) holds. We may assume that $\mathcal{U} \subset \Omega$ and since f is homogeneous of degree 1 we may also assume that $\mathcal{U}$ is conic.

By Theorem 2.18, there exists a sequence $\{\Gamma_j\}$ of ϱ_j-minimal bicharacteristic intervals such that $\varrho_j \to 0$ and $\Gamma_j \to \Gamma$ as $j \to \infty$. For sufficiently large j we have $\Gamma_j \subset \mathcal{U}$. Hence, if

$$\Gamma_j = \{(x_1, x'_j, 0, \xi'_j) : a_j \le x_1 \le b_j\}$$

then all the terms in the asymptotic sum of the symbol of R have vanishing Taylor coefficients on

$$\Gamma_{\varrho_j} = \{(x_1, x'_j, 0, \xi'_j) : a_j + \varrho_j \le x_1 \le b_j - \varrho_j\}$$

by Theorem 2.21. Since $\Gamma_{\varrho_j} \to \Gamma$ as $j \to \infty$, and all the terms in the asymptotic sum of the symbol of R are smooth functions, it follows that all the terms in the asymptotic sum of the symbol of R have vanishing Taylor coefficients on Γ. This proves Theorem 2.19 in this case.

(ii) Γ is a point and condition (2-18) holds. Then all the terms in the asymptotic sum of the symbol of R have vanishing Taylor coefficients on Γ by Theorem 2.21, so Theorem 2.19 follows.

(iii) Γ is a point and (2-18) is false. Let $\{\gamma_j\}$ be the sequence satisfying (2-19). Then $\{\operatorname{Re} p, \operatorname{Im} p\}(\gamma_j) > 0$ and $p(\gamma_j) = 0$ for each j since $\gamma_j = (t_j, x'_j, 0, \xi'_j)$. For fixed j we may assume that $\gamma_j = (0, \eta)$ and use

[Hörmander 1985a, Theorem 21.3.3] to find a canonical transformation χ together with Fourier integral operators A, B, A_1 and B_1 as in Proposition 2.4 such that $\chi(0, \varepsilon_n) = \gamma_j$, and $BPA = D_1 + ix_1 D_n$ in a conic neighborhood Ω of $\{(0, \varepsilon_n)\}$. Repetition of the arguments above allows us to write

$$BQA = BPAE + R(x, D_{x'}), \tag{2-20}$$

where the range of R is microlocally contained in the range of BPA at some compactly based cone K' containing Ω with $\chi(K') = K$. As before, E and R have classical symbols. Then all the terms in the asymptotic expansion of the symbol of R have vanishing Taylor coefficients at $\{(0, \varepsilon_n)\}$ by Theorem 2.20, and therefore all the terms in the asymptotic expansion of the symbol of $A_1 R B_1$ have vanishing Taylor coefficients at γ_j by Lemma A.1 in the appendix. Since the Fourier integral operators are chosen so that

$$K \cap WF(A_1 B - I) = \varnothing \quad \text{and} \quad K \cap WF(AB_1 - I) = \varnothing,$$

we have

$$\varnothing = K \cap WF(A_1 BQAB_1 - A_1 BPAEB_1 - A_1 RB_1) = K \cap WF(Q - PAEB_1 - A_1 RB_1)$$

in view of (2-20). Hence, all the terms in the asymptotic expansion of the symbol of

$$Q - PE_1 = A_1 RB_1 + S, \quad \text{where } \ WF(S) \cap K = \varnothing, \tag{2-21}$$

have vanishing Taylor coefficients at γ_j if $E_1 = AEB_1$. (Strictly speaking, the change of base variables $\gamma_j \mapsto (0, \eta)$ should be represented in (2-21) by conjugation of a linear transformation $\kappa : \mathbb{R}^n \to \mathbb{R}^n$, but this could be integrated in the Fourier integral operators A_1 and B_1 so it has been left out since it will not affect the arguments below.) It is clear that $E_1 \in \Psi^0_{\mathrm{cl}}(\mathbb{R}^n)$.

We have now shown that for each j there exists an operator $E_j \in \Psi^0_{\mathrm{cl}}(\mathbb{R}^n)$ such that all the terms in the asymptotic expansion of the symbol of $Q - PE_j$ have vanishing Taylor coefficients at γ_j. To construct the operator E in Theorem 2.19, we do the following. For each j, denote the symbol of E_j by

$$e^j(x, \xi) \sim \sum_{l=0}^{\infty} e^j_{-l}(x, \xi)$$

where $e^j_0(x, \xi)$ is the principal part, and $e^j_{-l}(x, \xi)$ is homogeneous of degree $-l$. If q is the principal symbol of Q, then by Proposition A.3 there exists a function $e_0 \in C^\infty(T^*(\mathbb{R}^n) \smallsetminus 0)$, homogeneous of degree 0, such that $q - pe_0$ has vanishing Taylor coefficients at Γ.

This argument can be repeated for lower order terms. Indeed, if $\sigma_Q = q + q_0 + \cdots$, then the term of degree 0 in σ_{Q-PE_j} is

$$\sigma_0(Q - PE_j) = \tilde{q}_j - pe^j_{-1},$$

where (see (2-25) below)

$$\tilde{q}_j(x, \xi) = q_0(x, \xi) - p_0(x, \xi) e^j_0(x, \xi) - \sum_k \partial_{\xi_k} p(x, \xi) D_{x_k} e^j_0(x, \xi).$$

We can write

$$p(x,\xi)e^j_{-1}(x,\xi) = p(x,\xi/|\xi|)e^j_{-1}(x,\xi/|\xi|),$$

so that $\tilde{q}_j(x,\xi)$, $p(x,\xi/|\xi|)$ and $e^j_{-1}(x,\xi/|\xi|)$ are all homogeneous of degree 0. Since

$$\partial_x^\alpha \partial_\xi^\beta e_0(\Gamma) = \lim_{j\to\infty} \partial_x^\alpha \partial_\xi^\beta e_0^j(\gamma_j)$$

it follows by Proposition A.3 that there is a function $g \in C^\infty(T^*(\mathbb{R}^n) \setminus 0)$, homogeneous of degree 0, such that

$$q_0(x,\xi) - p_0(x,\xi)e_0(x,\xi) - \sum_k \partial_{\xi_k} p(x,\xi) D_{x_k} e_0(x,\xi) - p(x,\xi/|\xi|)g(x,\xi)$$

has vanishing Taylor coefficients at Γ. Putting $e_{-1}(x,\xi) = |\xi|^{-1} g(x,\xi)$ we find that

$$\partial_x^\alpha \partial_\xi^\beta e_{-1}(\Gamma) = \lim_{j\to\infty} \partial_x^\alpha \partial_\xi^\beta e^j_{-1}(\gamma_j),$$

and that

$$\sigma_0(Q - P\circ e_0(x,D) - P\circ e_{-1}(x,D))$$

has vanishing Taylor coefficients at Γ. Continuing this way we successively obtain functions $e_m(x,\xi) \in C^\infty(T^*(\mathbb{R}^n) \setminus 0)$, homogeneous of degree m for $m \le 0$, such that

$$\sigma_Q - \Big(\sum_{m=0}^{M} e_{-m}\Big)\sigma_P \quad \operatorname{mod} S_{\mathrm{cl}}^{-M}$$

has vanishing Taylor coefficients at Γ. If we let E have symbol

$$\sigma_E(x,\xi) \sim \sum_{m=0}^{\infty} (1-\phi(\xi))e_{-m}(x,\xi)$$

with $\phi \in C_0^\infty$ equal to 1 for ξ close to 0, then $E \in \Psi^0_{\mathrm{cl}}(\mathbb{R}^n)$ and all terms in the asymptotic expansion of the symbol of $Q - PE$ have vanishing Taylor coefficients at Γ. We have proved Theorem 2.19. □

Remark. Instead of reducing to the study of the normal form $P = D_{x_1} + ix_1 D_{x_n}$ when condition (2-18) does not hold, as in case (iii) above, one could show that the terms in the asymptotic expansion of the operator R given by (2-16) has vanishing Taylor coefficients at every point in the sequence $\{\gamma_j\}$ satisfying (2-19) using techniques very similar to those used to prove Theorem 2.21. Theorem 2.19 would then follow by continuity, but the proof of the analogue of Theorem 2.20 would be more involved. In particular, we would have to construct a phase function w solving the eikonal equation

$$\partial w/\partial x_1 - if(x, \partial w/\partial x') = 0$$

approximately instead of explicitly (confer the proofs of Theorems 2.21 and 2.20, respectively). For fixed j this could be accomplished by adapting the approach in [Hörmander 1963; Hörmander 1966] (for a brief discussion, see [Hörmander 1981, p. 83]), where one has $f = 0$ and $\partial f/\partial x_1 > 0$ at $(0,\xi^0)$ instead of at γ_j.

We shall now show how our results relate to the ones referred to in the introduction, beginning with (1-3). There, it sufficed to have the coefficients of P and Q in C^∞ and C^1, respectively. However, in order for Theorem 2.19 to qualify, we must require both P and Q to have smooth coefficients. On the other hand, we shall only require the equation $Pu = Qf$ to be microlocally solvable (at an appropriate cone K) as given by Definition 2.1. Note that if P is a first order differential operator on an open set $\Omega \subset \mathbb{R}^n$, such that the principal symbol p of P satisfies condition (1-4) at a point $(x, \xi) \in T^*(\Omega) \smallsetminus 0$, then either $\{\operatorname{Re} p, \operatorname{Im} p\} > 0$ at (x, ξ), or $\{\operatorname{Re} p, \operatorname{Im} p\} > 0$ at $(x, -\xi)$. (The order of the operator is not important; the statement is still true for a differential operator of order m, since the Poisson bracket is then homogeneous of order $2m - 1$.) Assuming the former, this implies that (x, ξ) satisfies condition (a) in Theorem 2.19 by an application of [Hörmander 1985a, Theorem 21.3.3] and Lemma 2.7. In order to keep the formulation of the following result as simple as possible, we will assume that there exists a compactly based cone $K \subset T^*(\Omega) \smallsetminus 0$ with nonempty interior such that K contains the appropriate point $(x, \pm\xi)$, and such that the equation $Pu = Qf$ is microlocally solvable at K. This is clearly the case if the equation $Pu = Qf$ is locally solvable in Ω in the weak sense suggested by (1-1).

Corollary 2.22. *Let $\Omega \subset \mathbb{R}^n$ be open, and let $P(x, D)$ and $Q(x, D)$ be two first order differential operators with coefficients in $C^\infty(\Omega)$. Let p be the principal symbol of P, and let x_0 be a point in Ω such that*

$$p(x_0, \xi_0) = 0 \quad \textit{and} \quad \{\operatorname{Re} p, \operatorname{Im} p\}(x_0, \xi_0) > 0 \tag{2-22}$$

for some $\xi_0 \in \mathbb{R}^n$. If $K \subset T^*(\Omega) \smallsetminus 0$ is a compactly based cone containing (x_0, ξ_0) such that the range of Q is microlocally contained in the range of P at K, then there exists a constant μ such that (at the fixed point x_0)

$$Q^*(x_0, D) = \mu P^*(x_0, D), \tag{2-23}$$

where Q^* and P^* are the adjoints of Q and P.

Proof. By (2-22), $P \in \Psi^1_{\mathrm{cl}}(\Omega)$ is an operator of principal type microlocally near (x_0, ξ_0). P and Q therefore satisfy the hypotheses of Theorem 2.19, and in view of the discussion above regarding the point (x, ξ) we find that there exists an operator $E \in \Psi^0_{\mathrm{cl}}(\Omega)$ such that all the terms in the asymptotic expansion of the symbol of $Q - PE$ have vanishing Taylor coefficients at (x_0, ξ_0). By the discussion following (3-7) on page 452 below, it follows that the same must hold for the adjoint $Q^* - E^*P^*$. If we let Q^* and P^* have symbols $\sigma_{Q^*}(x, \xi) = q_1(x, \xi) + q_0(x)$ and $\sigma_{P^*}(x, \xi) = p_1(x, \xi) + p_0(x)$, then E^*P^* has principal symbol $e_0 p_1$ if $\sigma_{E^*} = e_0 + e_{-1} + \dots$ denotes the symbol of E^*. Hence

$$\partial q_1(x_0, \xi_0)/\partial \xi_k = e_0(x_0, \xi_0)\partial p_1(x_0, \xi_0)/\partial \xi_k \quad \text{for } 1 \le k \le n$$

and $p_1(x_0, \xi_0) = \overline{p(x_0, \xi_0)} = 0$. Since q_1 and p_1 are polynomials in ξ of degree 1, this means that at the fixed point x_0 we have $q_1(x_0, \xi) = \mu p_1(x_0, \xi)$ for $\xi \in \mathbb{R}^n$, where the constant μ is given by the value of e_0 at (x_0, ξ_0). Moreover,

$$0 = \partial_{\xi_j}\partial_{\xi_k} q_1(x_0, \xi_0) = \partial_{\xi_j} e_0(x_0, \xi_0)\partial_{\xi_k} p_1(x_0, \xi_0) + \partial_{\xi_k} e_0(x_0, \xi_0)\partial_{\xi_j} p_1(x_0, \xi_0). \tag{2-24}$$

By assumption, the coefficients of $p(x, D)$ do not vanish simultaneously, so the same is true for $p_1(x, D)$. Hence $\partial_{\xi_j} p_1(x_0, \xi_0) \neq 0$ for some j. Assuming this holds for $j = 1$, we find by choosing $j = k = 1$ in (2-24) that $\partial_{\xi_1} e_0(x_0, \xi_0) = 0$. But this immediately yields

$$\partial_{\xi_k} e_0(x_0, \xi_0) = -\partial_{\xi_1} e_0(x_0, \xi_0)\partial_{\xi_k} p_1(x_0, \xi_0)/\partial_{\xi_1} p_1(x_0, \xi_0) = 0$$

for $2 \le k \le n$. Now

$$\sigma_{E^*P^*}(x, \xi) \sim \sum \frac{1}{\alpha!}\partial_\xi^\alpha \sigma_{E^*}\, D_x^\alpha(p_1(x, \xi) + p_0(x)),$$

and since we have a bilinear map

$$S_{\mathrm{cl}}^{m'}/S^{-\infty} \times S_{\mathrm{cl}}^{m''}/S^{-\infty} \ni (a, b) \mapsto a \,\#\, b \in S_{\mathrm{cl}}^{m'+m''}/S^{-\infty}$$

with

$$(a \,\#\, b)(x, \xi) \sim \sum \frac{1}{\alpha!}\partial_\xi^\alpha a(x, \xi)\, D_x^\alpha b(x, \xi),$$

we find that the term of order 0 in the symbol of E^*P^* is

$$\sigma_0(E^*P^*)(x, \xi) = e_{-1}(x, \xi)p_1(x, \xi) + e_0(x, \xi)p_0(x) + \sum_{k=1}^{n} \partial_{\xi_k} e_0(x, \xi)\, D_k p_1(x, \xi). \tag{2-25}$$

Since $\partial_{\xi_k} e_0$ and p_1 vanish at (x_0, ξ_0) we find that $q_0(x_0) = \mu p_0(x_0)$ at the fixed point x_0, which completes the proof. □

Having proved this result, we immediately obtain the following after making the obvious adjustments to [Hörmander 1963, Theorem 6.2.2]. The fact that we require higher regularity on the coefficients of Q then yields higher regularity on the propertionality factor. Since the proof remains the same, it is omitted.

Corollary 2.23. *Let $\Omega \subset \mathbb{R}^n$ be open, and let $P(x, D)$ and $Q(x, D)$ be two first order differential operators with coefficients in $C^\infty(\Omega)$. Let p be the principal symbol of P, and assume that the coefficients of $p(x, D)$ do not vanish simultaneously in Ω. If for a dense set of points x in Ω one can find $\xi \in \mathbb{R}^n$ such that* (2-22) *is fulfilled, and if for each (x, ξ) there is a compactly based cone $K \subset T^*(\Omega) \smallsetminus 0$ containing (x, ξ) such that the range of Q is microlocally contained in the range of P at K, then there exists a function $e \in C^\infty(\Omega)$ such that*

$$Q(x, D)u \equiv P(x, D)(eu). \tag{2-26}$$

In stating Corollary 2.23 we could replace the assumption that the coefficients of $p(x, D)$ do not vanish simultaneously in Ω with the condition that P is of principal type. Indeed, if $dp \neq 0$ then by a canonical transformation we find that condition (1-6) holds. Since $p \neq 0$ implies $\partial_\xi p \neq 0$ by the Euler homogeneity equation we then have $\partial_\xi p \neq 0$ everywhere, that is, the coefficients of $p(x, D)$ do not vanish simultaneously in Ω. The converse is obvious.

As shown in Example 2.25 below, we also recover the result for higher order differential operators mentioned in the introduction as a special case of the following corollary to Theorem 2.19, although we again need to assume higher regularity in order to apply our results.

Proposition 2.24. *Let X be a smooth manifold, and let $P \in \Psi^k_{\mathrm{cl}}(X)$ and $Q \in \Psi^{k'}_{\mathrm{cl}}(X)$ be properly supported such that the range of $Q \circ P$ is microlocally contained in the range of P at a compactly based cone $K \subset T^*(X) \smallsetminus 0$. Let p and q be the principal symbols of P and Q, respectively, and assume that P is of principal type microlocally near K. If $\gamma : I \to T^*(X) \smallsetminus 0$ is a minimal characteristic point or a minimal bicharacteristic interval of p contained in K then it follows that*

$$H_p^m(q) = 0 \quad \textit{for all } (x,\xi) \in \gamma(I) \textit{ and } m \geq 1.$$

Here $H_p^m(q)$ is defined recursively by $H_p(q) = \{p, q\}$ and $H_p^m(q) = \{p, H_p^{m-1}(q)\}$ for $m \geq 2$.

Proof. First note that if the range of $Q \in \Psi^{k'}_{\mathrm{cl}}(X)$ is microlocally contained in the range of $P \in \Psi^k_{\mathrm{cl}}(X)$ at K and both operators are properly supported, then it follows that the range of $Q \circ P$ is microlocally contained in the range of P at K. (The converse is not true in general.) Indeed, let N be the integer given by Definition 2.1, and let $f \in H^{\mathrm{loc}}_{(N+k)}(X)$. Since $P : H^{\mathrm{loc}}_{(N+k)}(X) \to H^{\mathrm{loc}}_{(N)}(X)$ is continuous, we have $g = Pf \in H^{\mathrm{loc}}_{(N)}(X)$. Thus, there exists a $u \in \mathscr{D}'(X)$ such that

$$\varnothing = K \cap WF(Qg - Pu) = K \cap WF(QPf - Pu),$$

so the conditions of Definition 2.1 are satisfied with N replaced with $N + k$.

Let $(x, \xi) \in \gamma(I)$. The range of PQ is easily seen to be microlocally contained in the range of P for any properly supported pseudodifferential operator Q. The assumptions of the proposition therefore imply that the range of the commutator

$$R_1 = P \circ Q - Q \circ P \in \Psi^{k+k'-1}_{\mathrm{cl}}(X) \tag{2-27}$$

is microlocally contained in the range of P at K. Hence, by Theorem 2.19 there exists an operator $E \in \Psi^{k'-1}_{\mathrm{cl}}(X)$ such that, in particular, the principal symbol of $R_1 - PE$ vanishes at (x, ξ). If e is the principal symbol of E, homogeneous of degree $k' - 1$, then the principal symbol of PE satisfies $p(x, \xi)e(x, \xi) = 0$ since $p \circ \gamma = 0$. Since the principal symbol of R_1 is

$$\sigma_{k+k'-1}(R_1) = \frac{1}{i}\{p, q\},$$

the result follows for $m = 1$.

Let R_m be defined recursively by $R_m = [P, R_{m-1}]$ for $m \geq 2$ with R_1 given by (2-27). Arguing by induction, we conclude in view of the first paragraph of the proof that the range of R_m is microlocally contained in the range of P at K for $m = 1, 2 \ldots$ since this holds for R_1. Assuming the proposition holds for some $m \geq 1$, we can repeat the arguments above to show that the principal symbol of R_{m+1} must vanish at (x, ξ). Since the principal symbol of R_{m+1} equals $-i\{p, H_p^m(q)\}$, this completes the proof. □

Example 2.25. Let $\Omega \subset \mathbb{R}^n$ be open, $P(x, D)$ be a differential operator of order m with coefficients in $C^\infty(\Omega)$, and let μ be a function in $C^\infty(\Omega)$ such that the equation

$$P(x, D)u = \mu P(x, D)f$$

has a solution $u \in \mathscr{D}'(\Omega)$ for every $f \in C_0^\infty(\Omega)$. If p is the principal symbol of P then it follows that

$$\sum_{j=1}^{n} \partial_{\xi_j} p(x,\xi) D_{x_j}\mu(x) = 0 \tag{2-28}$$

for all $x \in \Omega$ and $\xi \in \mathbb{R}^n$ such that

$$\{p, \bar{p}\}(x,\xi) \neq 0 \quad \text{and} \quad p(x,\xi) = 0. \tag{2-29}$$

Indeed, if (x,ξ) satisfies (2-29) then we may assume that

$$\{\operatorname{Re} p, \operatorname{Im} p\}(x,\xi) = -\frac{1}{2i}\{p, \bar{p}\}(x,\xi) > 0$$

since otherwise we just regard $(x,-\xi)$ instead, as per the remarks preceding Corollary 2.22. By the same discussion it is also clear that (x,ξ) is a minimal characteristic point of p. Now the conditions above imply that there exists a compactly based cone $K \subset T^*(\Omega) \smallsetminus 0$ containing (x,ξ) such that the range of μP is microlocally contained in the range of P at K. By condition (2-29), P is of principal type near (x,ξ), so Proposition 2.24 implies that $\{p,\mu\} = 0$ at (x,ξ), that is,

$$\sum_{j=1}^{n} \partial_{\xi_j} p(x,\xi) \partial_{x_j}\mu(x) - \partial_{x_j} p(x,\xi) \partial_{\xi_j}\mu(x) = 0.$$

Since μ is independent of ξ we find that (2-28) holds at (x,ξ). By homogeneity it then also holds at $(x,-\xi)$.

3. Proof of Theorem 2.20

Throughout this section we assume that the hypotheses of Theorem 2.20 hold. We shall prove the theorem by using Lemma 2.3 on approximate solutions of the equation $P^*v = 0$ concentrated near $\Gamma' = \{(0, \varepsilon_n)\}$. We take as starting point the construction on [Hörmander 1985b, page 103], but some modifications need to be made in particular to the amplitude function ϕ, so the results there concerning the estimates for the right side of (2-1) cannot be used immediately. To obtain the desired estimates we will instead have to use [Hörmander 1985b, Lemma 26.4.15]. Set

$$v_\tau(x) = \phi(x) e^{i\tau w(x)}, \tag{3-1}$$

where

$$w(x) = x_n + i(x_1^2 + x_2^2 + \cdots + x_{n-1}^2 + (x_n + i x_1^2/2)^2)/2$$

satisfies $P^*w = 0$ and $\phi \in C_0^\infty(\mathbb{R}^n)$. By the Cauchy–Kovalevsky theorem we can solve $D_1\phi - i x_1 D_n \phi = 0$ in a neighborhood of 0 for any analytic initial data $\phi(0,x') = f(x') \in C^\omega(\mathbb{R}^{n-1})$; in particular we are free to specify the Taylor coefficients of $f(x')$ at $x' = 0$. We take ϕ to be such a solution. If need be we can reduce the support of ϕ by multiplying by a smooth cutoff function χ, where χ is equal to 1 in some smaller neighborhood of 0 so that $\chi\phi$ solves the equation there. We assume this to be done and note that if $\operatorname{supp}\phi$ is small enough then

$$\operatorname{Im} w(x) \geq |x|^2/4 \quad \text{for } x \in \operatorname{supp}\phi. \tag{3-2}$$

Since

$$d \operatorname{Re} w(x) = -x_1 x_n dx_1 + (1 - x_1^2/2) dx_n,$$

we may similarly assume that $d \operatorname{Re} w(x) \neq 0$ in the support of ϕ. We then have the following result.

Lemma 3.1. *Suppose $P = D_1 + ix_1 D_n$ and let v_τ be defined by* (3-1). *Then ϕ and w can be chosen so that for any $f \in C^\omega(\mathbb{R}^{n-1})$ and any positive integers k and m we have $\phi(0, x') = f(x')$ in a neighborhood of* $(0, 0)$, $\tau^k \| P^* v_\tau \|_{(m)} \to 0$ *as $\tau \to \infty$, and*

$$\| v_\tau \|_{(-m)} \leq C_m \tau^{-m}. \tag{3-3}$$

If $\tilde{\Gamma}$ is the cone generated by

$$\{(x, w'(x)) : x \in \operatorname{supp}\phi,\ \operatorname{Im} w(x) = 0\},$$

then $\tau^k v_\tau \to 0$ in $\mathscr{D}'_{\tilde{\Gamma}}$ as $\tau \to \infty$; hence $\tau^k A v_\tau \to 0$ in $C^\infty(\mathbb{R}^n)$ if A is a pseudodifferential operator with $WF(A) \cap \tilde{\Gamma} = \varnothing$.

Here $\mathscr{D}'_{\tilde{\Gamma}}(X) = \{u \in \mathscr{D}'(X) : WF(u) \subset \tilde{\Gamma}\}$, equipped with the topology given by all the seminorms on $\mathscr{D}'(X)$ for the weak topology, together with all seminorms of the form

$$P_{\phi,V,N}(u) = \sup_{\xi \in V} |\widehat{\phi u}(\xi)|(1 + |\xi|)^N$$

where $N \geq 0$, $\phi \in C_0^\infty(X)$, and $V \subset \mathbb{R}^n$ is a closed cone with $(\operatorname{supp}\phi \times V) \cap \tilde{\Gamma} = \varnothing$. Note that $u_j \to u$ in $\mathscr{D}'_{\tilde{\Gamma}}(X)$ is equivalent to $u_j \to u$ in $\mathscr{D}'(X)$ and $Au_j \to Au$ in C^∞ for every properly supported pseudodifferential operator A with $\tilde{\Gamma} \cap WF(A) = \varnothing$; see the remark following [Hörmander 1985a, Theorem 18.1.28].

Proof. We observe that $\tau^k P^* v_\tau = \tau^k (P^*\phi) e^{i\tau w} \to 0$ in $C_0^\infty(\mathbb{R}^n)$ for any k as $\tau \to \infty$, if w and ϕ are chosen in the way given above. Hence $\tau^k \| P^* v_\tau \|_{(m)} \to 0$ for any positive integers k and m. In view of (3-2) and the fact that $d \operatorname{Re} w \neq 0$ in the support of ϕ we can apply [Hörmander 1985b, Lemma 26.4.15] to v_τ. This immediately yields (3-3) and also that $\tau^k v_\tau \to 0$ in $\mathscr{D}'_{\tilde{\Gamma}}$ as $\tau \to \infty$, proving the lemma. □

We are now ready to proceed with a tool that will be instrumental in proving Theorem 2.21. The idea is based on techniques found in [Hörmander 1963].

Let R be the operator given by Theorem 2.20. By assumption there exists a compactly based cone $K \subset T^*(\mathbb{R}^n) \setminus 0$ such that the range of R is microlocally contained in the range of P at K. If N is the integer given by Definition 2.1, let $H(x) \in C_0^\infty(\mathbb{R}^n)$ and set

$$h_\tau(x) = \tau^{-N} H(\tau x). \tag{3-4}$$

Since $\hat{h}_\tau(\xi) = \tau^{-N-n}\hat{H}(\xi/\tau)$ it is clear that for $\tau \geq 1$ we have $h_\tau \in H_{(N)}(\mathbb{R}^n)$ and $\|h_\tau\|_{(N)} \leq C\tau^{-n/2}$. In particular, $\|h_\tau\|_{(N)} \leq C$ for $\tau \geq 1$, where the constant depends on H but not on τ. Now denote by I_τ the integral

$$I_\tau = \tau^n \int H(\tau x) R^* v_\tau(x)\, dx = \tau^{N+n}(R^* v_\tau, \overline{h_\tau}), \tag{3-5}$$

where R^* is the adjoint of R. For any κ we then have by the second equality and Lemma 2.3 that

$$|I_\tau| \leq \tau^{N+n}\|h_\tau\|_{(N)}\|R^* v_\tau\|_{(-N)} \leq C_\kappa \tau^{N+n}(\|P^* v_\tau\|_{(\nu)} + \|v_\tau\|_{(-N-\kappa-n)} + \|A v_\tau\|_{(0)})$$

for some positive integer ν and properly supported pseudodifferential operator A with $WF(A) \cap K = \varnothing$. By Lemma 3.1 this implies

$$|I_\tau| \leq C_\kappa \tau^{-\kappa} \tag{3-6}$$

for any positive integer κ if τ is sufficiently large.

Recall that $R(x, D_{x'})$ is a pseudodifferential operator in x' depending on x_1 as a parameter. Its symbol is given by the asymptotic sum

$$\sigma_R(x, \xi') = r_1(x, \xi') + r_0(x, \xi') + \cdots,$$

where $r_{-j}(x, \xi')$ is homogeneous of degree $-j$ in ξ'. The symbol of R^* has the asymptotic expansion

$$\sigma_{R^*} = \sum \partial_\xi^\alpha D_x^\alpha \overline{\sigma_R(x, \xi')}/\alpha!,$$

which shows that R^* is also a pseudodifferential operator in x' depending on x_1 as a parameter. If we sort the terms above with respect to homogeneity we can write

$$\sigma_{R^*} = q_1(x, \xi') + q_0(x, \xi') + \cdots, \tag{3-7}$$

where q_{-j} is homogeneous of order $-j$, $q_1(x, \xi') = \overline{r_1(x, \xi')}$ and

$$q_0(x, \xi') = \overline{r_0(x, \xi')} + \sum_{k=2}^{n} \partial_{\xi_k} D_{x_k} \overline{r_1(x, \xi')}.$$

A moment's reflection shows that if all the terms in (3-7) have vanishing Taylor coefficients at some point (x, ξ'), then the same must hold for σ_R.

Our goal is to show that if $q_{-j\,(\alpha)}^{(\beta)}(0, \xi^0)$ does not vanish for all $j \geq -1$ and all $\alpha, \beta \in \mathbb{N}^n$, then (3-6) cannot hold. For this purpose, we introduce a total well-ordering $>_t$ on the Taylor coefficients by means of an ordering of the indices (j, α, β) as follows.

Definition 3.2. Let $\alpha_i, \beta_i \in \mathbb{N}^n$ and $j_i \geq -1$ for $i = 1, 2$. We say that

$$q_{-j_1\,(\alpha_1)}^{(\beta_1)}(0, \xi^0) >_t q_{-j_2\,(\alpha_2)}^{(\beta_2)}(0, \xi^0) \quad \text{if } j_1 + |\alpha_1| + |\beta_1| > j_2 + |\alpha_2| + |\beta_2|.$$

To "break ties", we say that if $j_1 + |\alpha_1| + |\beta_1| = j_2 + |\alpha_2| + |\beta_2|$, then

$$q_{-j_1\,(\alpha_1)}^{(\beta_1)}(0, \xi^0) >_t q_{-j_2\,(\alpha_2)}^{(\beta_2)}(0, \xi^0) \quad \text{if } |\beta_2| > |\beta_1|.$$

Note the reversed order. If also $|\beta_1| = |\beta_2|$, then we use a monomial ordering on the β index to break ties. Recall that this is any relation $>$ on $\mathbb{N}^n$ such that $>$ is a total well-ordering on $\mathbb{N}^n$ and $\beta_1 > \beta_2$ and $\gamma \in \mathbb{N}^n$ implies $\beta_1 + \gamma > \beta_2 + \gamma$. Having come this far, the actual order turns out not to matter for the proof of Theorem 2.20, but it will have bearing on the proof of Theorem 2.21. Which monomial ordering we use on the β index will not be important, but for completeness let us choose lexicographic order since this will be used at a later stage in the definition. Here we by lexicographic order refer to the usual one, corresponding to the variables being ordered $x_1 > \cdots > x_n$. That is to say, if $\alpha_i \in \mathbb{N}^n$ for $i = 1, 2$, then $\alpha_1 >_{\text{lex}} \alpha_2$ if, in the vector difference $\alpha_1 - \alpha_2 \in \mathbb{Z}^n$, the leftmost nonzero entry is positive. Thus, if $j_1 + |\alpha_1| + |\beta_1| = j_2 + |\alpha_2| + |\beta_2|$ and $\beta_1 = \beta_2$, then we first say that

$$q^{(\beta_1)}_{-j_1\,(\alpha_1)}(0, \xi^0) >_t q^{(\beta_2)}_{-j_2\,(\alpha_2)}(0, \xi^0) \quad \text{if } |\alpha_2| > |\alpha_1| \tag{3-8}$$

and then use lexicographic order on the n-tuples α to break ties at this stage. Using the lexicographic order on both multiindices (separately) we get

$$q_1 <_t q_1^{(\varepsilon_n)} <_t \cdots <_t q_1^{(\varepsilon_1)} <_t q_{1(\varepsilon_n)} <_t \cdots <_t q_{1(\varepsilon_1)} <_t q_0 <_t \cdots.$$

As indicated above we will prove Theorem 2.20 by a contradiction argument, so in the sequel we let κ denote an integer such that

$$j + |\alpha| + |\beta| < \kappa \tag{3-9}$$

if $q^{(\beta)}_{-j\,(\alpha)}(0, \xi^0)$ is the first nonvanishing Taylor coefficient with respect to the ordering $>_t$. Since $j \geq -1$ we will thus have $\kappa \geq 0$.

To simplify notation, we shall in what follows write t instead of x_1 and x instead of x'. Then v_τ takes the form

$$v_\tau(t, x) = \phi(t, x)e^{i\tau w(t,x)},$$

where

$$w(t, x) = x_{n-1} + i(t^2 + x_1^2 + \cdots + x_{n-2}^2 + (x_{n-1} + it^2/2)^2)/2. \tag{3-10}$$

We shall as before use the notation $\xi^0 = (0, \ldots, 0, 1) \in \mathbb{R}^{n-1}$ when in this context. To interpret the integral I_τ we will need a formula for how $R^*(t, x, D)$ acts on the functions v_τ. This is given by the following lemma, where the parameter t has been suppressed to simplify notation.

Lemma 3.3 [Hörmander 1985b, Lemma 26.4.16]. *Let $q(x, \xi) \in S^\mu(\mathbb{R}^{n-1} \times \mathbb{R}^{n-1})$, let $\phi \in C_0^\infty(\mathbb{R}^{n-1})$ and $w \in C^\infty(\mathbb{R}^{n-1})$, and assume that* $\operatorname{Im} w > 0$ *except at a point y where $w'(y) = \eta \in \mathbb{R}^{n-1} \setminus 0$ and* $\operatorname{Im} w''$ *is positive definite. Then*

$$\Big|q(x, D)(\phi e^{i\tau w}) - \sum_{|\alpha|<k} q^{(\alpha)}(x, \tau\eta)(D - \tau\eta)^\alpha(\phi e^{i\tau w})/\alpha!\Big| \leq C_k \tau^{\mu - k/2} \tag{3-11}$$

for $\tau > 1$ and $k = 1, 2, \ldots$.

An inspection of the proof of [Hörmander 1985b, Lemma 26.4.16] shows that the result is still applicable if $\operatorname{Im} w > 0$ everwhere. This is also used without mention in [Hörmander 1985b] when proving the necessity of condition (Ψ). Thus the statement holds if $\operatorname{Im} w > 0$ except *possibly* at a point y where

$w'(y) = \eta \in \mathbb{R}^{n-1} \setminus 0$ and $\operatorname{Im} w''$ is positive definite. We will also use this fact, but we have refrained from altering the statement of the lemma.

If q is homogeneous of degree μ, then the sum in (3-11) consists (apart from the factor $e^{i\tau w}$) of terms that are homogeneous in τ of degree $\mu, \mu - 1, \dots$. The terms of degree μ are those in

$$\phi \sum q^{(\alpha)}(x, \tau\eta)(\tau w'(x) - \tau\eta)^{\alpha}/\alpha!, \tag{3-12}$$

which is the Taylor expansion at $\tau\eta$ of $q(x, \tau w')$. In this way one can give meaning to the expression $q(x, \tau w')$ even though $q(x, \xi)$ may not be defined for complex ξ. The terms of degree $\mu - 1$ where ϕ is differentiated are similarly

$$\sum_{k=1}^{n-1} q^{(k)}(x, \tau w'(x)) D_k \phi,$$

where $q^{(k)}$ should be replaced by the Taylor expansion at $\tau\eta$ representing the value at $\tau w'(x)$, as in (3-12). In the present case we have

$$w'_x(t, x) - \xi^0 = ix - (t^2/2)\xi^0,$$

so the expression $q_{-j}(t, x, w'_x(t, x))$ is given meaning if it is replaced by a finite Taylor expansion

$$\sum_{\beta} q_{-j}^{(\beta)}(t, x, \xi^0)(w'_x(t, x) - \xi^0)^{\beta}/|\beta|!$$

of sufficiently high order.

Using the classicality of R^* we have

$$\sigma_{R^*}(t, x, \xi) - \sum_{j=-1}^{M} q_{-j}(t, x, \xi) \in \Psi_{\mathrm{cl}}^{-M-1}(\mathbb{R}^n),$$

so there is a symbol $a \in S_{\mathrm{cl}}^{-M-1}(\mathbb{R}^n \times \mathbb{R}^{n-1})$ such that

$$a(t, x, D) = R^*(t, x, D) - \sum_{j=-1}^{M} q_{-j}(t, x, D) \mod \Psi^{-\infty}(\mathbb{R}^n).$$

By (3-2) and (3-10) it is clear that w satisfies the conditions of Lemma 3.3, so

$$a(t, x, D)v_\tau = a(t, x, \tau\xi^0)v_\tau + \mathcal{O}(\tau^{-M-3/2}) = \tau^{-M-1} a(t, x, \xi^0)v_\tau + \mathcal{O}(\tau^{-M-3/2}),$$

which implies that $|a(t, x, D)v_\tau| \le C\tau^{-M-1}$. If we for each $-1 \le j \le M$ write

$$\Bigl| q_{-j}(t, x, D)v_\tau - \sum_{|\alpha| < k_j} q_{-j}^{(\alpha)}(t, x, \tau\xi^0)(D_x - \tau\xi^0)^{\alpha} v_\tau/\alpha! \Bigr| \le C_{k_j} \tau^{-j-k_j/2}$$

with $k_j = 2M - 2j + 1$, then

$$R^*(t, x, D)v_\tau = \sum_{j=-1}^{M} \sum_{|\alpha| < k_j} q_{-j}^{(\alpha)}(t, x, \tau\xi^0)(D_x - \tau\xi^0)^{\alpha} v_\tau/\alpha! + \mathcal{O}(\tau^{-M-1/2}).$$

Now recall the discussion above regarding the homogeneity of the terms in (3-11), and choose $M \geq \kappa$, where κ is an integer satisfying (3-9). Then

$$\begin{aligned} R^*(t,x,D)v_\tau &= e^{i\tau w} \sum_{j=-1}^{M} \sum_{|\alpha|\leq 2M-2j} q_{-j}^{(\alpha)}(t,x,\tau w_x'(t,x)) D^\alpha \phi \\ &= e^{i\tau w} \sum_{j=-1}^{M} \sum_{|\alpha|\leq 2M-2j} \tau^{-j-|\alpha|} q_{-j}^{(\alpha)}(t,x,w_x'(t,x)) D^\alpha \phi \\ &= e^{i\tau w} \sum_{J=-1}^{M} \tau^{-J} \lambda_J(t,x) \end{aligned}$$

with an error of order $\mathcal{O}(\tau^{-\kappa-1/2})$, where

$$\lambda_J(t,x) = \sum_{j+|\alpha|=J} q_{-j}^{(\alpha)}(t,x,w_x'(t,x)) D^\alpha \phi \quad \text{for } j \geq -1. \tag{3-13}$$

As before, $q_{-j}^{(\alpha)}(t,x,w_x'(t,x))$ should be replaced by a finite Taylor expansion at ξ^0 of sufficiently high order representing the value at $w_x'(t,x)$. In view of (3-5), this yields

$$I_\tau = \tau^n \int H(\tau t, \tau x) e^{i\tau w(t,x)} \Big(\sum_{J=-1}^{\kappa} \tau^{-J} \lambda_J(t,x) + \mathcal{O}(\tau^{-\kappa-1/2}) \Big)\, dt\, dx.$$

After the change of variables $(\tau t, \tau x) \mapsto (t,x)$ we find that

$$I_\tau = \int H(t,x) e^{i\tau w(t/\tau, x/\tau)} \Big(\sum_{J=-1}^{\kappa} \tau^{-J} \lambda_J(t/\tau, x/\tau) + \mathcal{O}(\tau^{-\kappa-1/2}) \Big)\, dt\, dx. \tag{3-14}$$

To illustrate how we will proceed to prove Theorem 2.20 by contradiction, let us for the moment assume that $q_1(0,0,\xi^0) \neq 0$, where $\xi^0 = (0,\dots,0,1) \in \mathbb{R}^{n-1}$. Since

$$\lambda_{-1}(t/\tau, x/\tau) = \phi(t/\tau, x/\tau) \sum_\beta q_1^{(\beta)}(t/\tau, x/\tau, \xi^0)(w_x'(t/\tau, x/\tau) - \xi^0)^\beta / |\beta|! \tag{3-15}$$

where

$$w_x'(t/\tau, x/\tau) - \xi^0 = ix/\tau - (t^2/(2\tau^2))\xi^0 = \mathcal{O}(\tau^{-1}), \tag{3-16}$$

and (3-10) implies that $\tau w(t/\tau, x/\tau) \to x_{n-1}$ as $\tau \to \infty$, we obtain

$$\lim_{\tau\to\infty} I_\tau/\tau = \int H(t,x) e^{ix_{n-1}} \phi(0,0) q_1(0,0,\xi^0)\, dt\, dx.$$

Since we may choose $\phi \neq 0$ at the origin, the limit above will then not be equal to 0 for a suitable choice of H. However, this contradicts (3-6).

Now assume that $\partial_t^{k_0} q_{-j_0(\alpha_0)}^{(\beta_0)}(0,0,\xi^0)$ is the first nonvanishing Taylor coefficient with respect to the ordering $>_t$, and let

$$m = j_0 + k_0 + |\alpha_0| + |\beta_0| \tag{3-17}$$

so that $m < \kappa$ by (3-9). Note that $\alpha_0, \beta_0 \in \mathbb{N}^{n-1}$ and that the integer k_0 accounts for derivatives in t while there is no corresponding term for derivatives in the Fourier transform of t since the q_{-j} are independent of this variable. Note also that since j_0 is permitted to be -1, we have $0 \le k_0, |\alpha_0|, |\beta_0| \le m+1$.

To use our assumption we will need for each term $q_{-j}^{(\beta+\gamma)}(t/\tau, x/\tau, \xi^0)$ in the Taylor expansion of $q_{-j}^{(\gamma)}(t/\tau, x/\tau, w_x'(t/\tau, x/\tau))$ (as it appears in (3-13)) at ξ^0 to consider Taylor expansions in t and x at the origin. Note that for given j and γ, it suffices to consider finite Taylor expansions of $q_{-j}^{(\gamma)}$ of order $\kappa - j - |\gamma|$ by (3-14) and (3-16). For each j and γ we thus write

$$\begin{aligned} &q_{-j}^{(\gamma)}(t/\tau, x/\tau, w_x'(t/\tau, x/\tau)) \\ &= \sum_{k+|\alpha|+|\beta|\le\kappa-j-|\gamma|} (\partial_t^k q_{-j\,(\alpha)}^{(\beta+\gamma)})(0,0,\xi^0)\tau^{-k-|\alpha|} t^k x^\alpha (w_x'(t/\tau, x/\tau)-\xi^0)^\beta/(k!|\alpha|!|\beta|!) + \mathcal{O}(\tau^{-\kappa-1+j+|\gamma|}), \end{aligned}$$

where $(w_x'(t/\tau, x/\tau) - \xi^0)^\beta$ should be interpreted by means of (3-16). As we shall see, the term $(t^2/(2\tau^2))\xi^0$ will not pose any problem, since it is $\mathcal{O}(\tau^{-2})$. We have

$$\begin{aligned} \lambda_J(t/\tau, x/\tau) = \sum_{j+|\gamma|=J} \sum_{k+|\alpha|+|\beta|\le\kappa-J} &(\partial_t^k q_{-j\,(\alpha)}^{(\beta+\gamma)})(0,0,\xi^0) D^\gamma\phi(t/\tau, x/\tau) \\ &\times \tau^{-k-|\alpha|} t^k x^\alpha (w_x'(t/\tau, x/\tau)-\xi^0)^\beta/(k!|\alpha|!|\beta|!) + \mathcal{O}(\tau^{-\kappa-1+J}), \end{aligned}$$

where $-1 \le j \le J$. If we are only interested in terms of order τ^{-m} in (3-14), we can use the assumption that $\partial_t^k q_{-j\,(\alpha)}^{(\beta+\gamma)}(0,0,\xi^0) = 0$ for all $-1 \le j+k+|\alpha|+|\beta|+|\gamma| < m$ to let the term $(t^2/(2\tau^2))\xi^0$ from (3-16) be absorbed by the error term in the expression above. This yields

$$\begin{aligned} \sum_{J=-1}^{m} \tau^{-J}\lambda_J(t/\tau, x/\tau) = \sum_{j+k+|\alpha|+|\beta|+|\gamma|=m} &(\partial_t^k q_{-j\,(\alpha)}^{(\beta+\gamma)})(0,0,\xi^0) \\ &\times D^\gamma\phi(t/\tau, x/\tau)\tau^{-m} t^k x^\alpha (ix)^\beta/(k!|\alpha|!|\beta|!) + \mathcal{O}(\tau^{-m-1}), \end{aligned}$$

where we use $J = j + |\gamma|$ together with the fact that we get a factor $\tau^{-|\beta|}$ from $(w_x'(t/\tau, x/\tau) - \xi^0)^\beta$ by (3-16). Thus,

$$\begin{aligned} \lim_{\tau\to\infty} \tau^m I_\tau = \int H(t,x) e^{ix_{n-1}} & \\ \times \Big(\sum_{j+k+|\alpha|+|\beta|+|\gamma|=m} & t^k x^\alpha (ix)^\beta (\partial_t^k q_{-j\,(\alpha)}^{(\beta+\gamma)})(0,0,\xi^0) D^\gamma\phi(0,0)/(k!|\alpha|!|\beta|!)\Big)\,dt\,dx. \end{aligned}$$

Now choose ϕ such that $D^{\beta_0}\phi(0,0) = 1$, but $D^\gamma\phi(0,0) = 0$ for all other γ such that $|\gamma| \le |\beta_0|$. This is possible by the discussion following (3-1). By (3-17) and our choice of the ordering $>_t$, we have $\partial_t^k q_{-j\,(\alpha)}^{(\beta+\beta_0)}(0,0,\xi^0) = 0$ for all β such that $|\beta| > 0$ as long as $j+k+|\alpha|+|\beta|+|\beta_0| = m$. Hence, with this choice of ϕ, the last expression takes the form

$$\lim_{\tau\to\infty} \tau^m I_\tau = \int H(t,x) e^{ix_{n-1}} \Big(\sum_{j+k+|\alpha|+|\beta_0|=m} t^k x^\alpha (\partial_t^k q_{-j\,(\alpha)}^{(\beta_0)})(0,0,\xi^0)/(k!|\alpha|!)\Big)\,dt\,dx, \tag{3-18}$$

where as usual j is allowed to be -1 so that $j \in [-1, m - |\beta_0|]$ in (3-18). Now some of the Taylor coefficients in (3-18) may be zero, in particular, the expression may well contain Taylor coefficients that preceed $\partial_t^{k_0} q_{-j_0\,(\alpha_0)}^{(\beta_0)}(0,0,\xi^0)$, and those are by assumption zero. However, we claim that if at least one of the Taylor coefficients above are nonzero, then we may choose H so that the limit is nonzero. Indeed, if that were not the case then the expression within brackets in (3-18) would be a polynomial with infinitely many zeros, and thus it would have to have vanishing coefficients. Since this violates our assumption, we conclude that the limit is nonzero. However, this contradicts (3-6), which proves Theorem 2.20.

4. Proof of Theorem 2.21

In this section we shall give the proof of Theorem 2.21, using ideas taken from [Hörmander 1963] together with the approach used to prove [Hörmander 1985b, Theorem 26.4.7′]. As in the previous section, we aim to use Lemma 2.3 to estimate the operator $R(x, D_{x'})$ on approximate solutions of the equation $P^*v = 0$, concentrated near

$$\Gamma' = \{(x_1, x', 0, \xi') : x_1 \in I'\} \subset T^*(\mathbb{R}^n) \setminus 0. \tag{4-1}$$

The proofs will be similar, but the situation is more complicated now, which will affect the construction of the approximate solutions. We will also have to make some adjustments to the proof of [Hörmander 1985b, Theorem 26.4.7′] to make it work, so a lot of the details will have to be revisited. Our approximate solutions will also differ slightly from the ones used to prove [Hörmander 1985b, Theorem 26.4.7′], so although we will refer directly to results in [ibid.] whenever possible, the formulation of some of these results will be affected. For a more complete description of the approximate solutions, we refer the reader to [Hörmander 1981] or [Hörmander 1985b], where their construction is carried out in greater detail. When proving Theorem 2.21 we may without loss of generality assume that $x' = 0$ and $\xi' = \xi^0$ in (4-1). In accordance with the notation in the proof of Theorem 2.19, we shall therefore throughout this section refer to Γ' simply by Γ, and we will let $I' = [a_0, b_0]$.

To simplify notation we shall in what follows write t instead of x_1 and x instead of x'. If N is the integer given by Definition 2.1, and n is the dimension, the approximate solutions v_τ will be taken of the form

$$v_\tau(t, x) = \tau^{N+n} e^{i\tau w(t,x)} \sum_0^M \phi_j(t, x)\tau^{-j}. \tag{4-2}$$

Here $\phi_0, \phi_1, \dots$ are amplitude functions, and w is a phase function that should satisfy the eikonal equation

$$\partial w/\partial t - if(t, x, \partial w/\partial x) = 0 \tag{4-3}$$

approximately, where f is the imaginary part of the principal symbol of P. We take w of the form

$$w(t, x) = w_0(t) + \langle x - y(t), \eta(t)\rangle + \sum_{2\le|\alpha|\le M} w_\alpha(t)(x - y(t))^\alpha/|\alpha|!, \tag{4-4}$$

where M is a large integer to be determined later and $x = y(t)$ is a smooth real curve. When discussing the functions w_α we shall permit ourselves to use the notation $\alpha = (\alpha_1, \dots, \alpha_s)$ for a sequence of $s = |\alpha|$ indices between 1 and the dimension $n - 1$ of the x variable, and w_α will be symmetric in these indices. If we take $\eta(t)$ to be real-valued and make sure the matrix $(\operatorname{Im} w_{jk})$ is positive definite, then $\operatorname{Im} w$ will have a strict minimum when $x = y(t)$ as a function of the x variables.

On the curve $x = y(t)$ the eikonal equation (4-3) is reduced to

$$w_0'(t) = \langle y'(t), \eta(t)\rangle + if(t, y(t), \eta(t)), \tag{4-5}$$

which is the only equation where w_0 occurs. Hence it can be used to determine w_0 after y and η have been chosen. In particular

$$d \operatorname{Im} w_0(t)/dt = f(t, y(t), \eta(t)). \tag{4-6}$$

In the proof of Theorem 2.20 we could solve the corresponding eikonal equation explicitly. Here this is not possible, so our goal will instead be to make (4-3) valid apart from an error of order $M+1$ in $x - y(t)$. Note that $f(t, x, \xi)$ is not defined for complex ξ, but since

$$\partial w(t, x)/\partial x_j - \eta_j(t) = \sum w_{\alpha,j}(t)(x - y(t))^\alpha/|\alpha|!,$$

(4-3) is given meaning if $f(t, x, \partial w/\partial x)$ is replaced by the finite Taylor expansion

$$\sum_{|\beta|\le M} f^{(\beta)}(t, x, \eta(t))(\partial w(t, x)/\partial x - \eta(t))^\beta/|\beta|!. \tag{4-7}$$

To compute the coefficient of $(x - y(t))^\alpha$ in (4-7) we just have to consider the terms with $|\beta| \le |\alpha|$. Since

$$\begin{aligned}\partial w/\partial t = w_0' - \langle y', \eta\rangle + \langle x - y, \eta'\rangle \\ + \sum_{2\le|\alpha|\le M} w_\alpha'(t)(x - y)^\alpha/|\alpha|! - \sum_k \sum_{1\le|\alpha|\le M-1} w_{\alpha,k}(t)(x - y)^\alpha dy_k/dt/|\alpha|!,\end{aligned}$$

the first order terms in the equation (4-3) give

$$d\eta_j/dt - \sum_k w_{jk}(t)dy_k/dt = i(f_{(j)}(t, y, \eta) + \sum_k f^{(k)}(t, y, \eta)w_{jk}(t)). \tag{4-8}$$

Note that this is a system of $2n$ equations

$$d\eta_j/dt - \sum_k \operatorname{Re} w_{jk}(t)dy_k/dt = -\sum_k \operatorname{Im} w_{jk}(t) f^{(k)}(t, y, \eta), \tag{4-8$'$}$$

$$\sum_k \operatorname{Im} w_{jk}(t)dy_k/dt = -f_{(j)}(t, y, \eta) - \sum_k \operatorname{Re} w_{jk}(t) f^{(k)}(t, y, \eta), \tag{4-8$''$}$$

since y and η are real, and under the assumption that $\operatorname{Im} w_{jk}$ is positive definite these equations can be solved for dy/dt and $d\eta/dt$. We observe that at a point where $f = df = 0$ they just mean that $dy/dt = d\eta/dt = 0$.

When $2 \le |\alpha| \le M$ we obtain a differential equation

$$dw_\alpha/dt - \sum_k w_{\alpha,k} dy_k/dt = F_\alpha(t, y, \eta, \{w_\beta\}) \tag{4-9}$$

from (4-3). Here F_α is a linear combination of the derivatives of f of order $|\alpha|$ or less, multiplied with polynomials in w_β with $2 \le |\beta| \le |\alpha| + 1$. Of course, when $|\alpha| = M$ the sum on the left side of (4-9) should be dropped, and β should satisfy $|\beta| \le |\alpha|$ instead. Altogether (4-8)$'$, (4-8)$''$ and (4-9) form a quasilinear system of differential equations with as many equations as unknowns. Hence we have local solutions with prescribed initial data. According to [Hörmander 1985b, pages 105–106] we can find a $c > 0$ such that the equations (4-8) and (4-9) with initial data

$$w_{jk} = i\delta_{jk}, \qquad w_\alpha = 0, \qquad \text{when } 2 < |\alpha| \le M \text{ and } t = (a_0 + b_0)/2, \tag{4-10}$$

$$y = x, \qquad \eta = \xi, \qquad \text{when } t = (a_0 + b_0)/2 \tag{4-11}$$

have a unique solution in $(a_0 - c, b_0 + c)$ for all x and ξ with $|x| + |\xi - \xi^0| < c$. (Here δ_{jk} is the Kronecker delta.) Moreover,

(i) $(\operatorname{Im} w_{jk} - \delta_{jk}/2)$ is positive definite,

(ii) the map

$$(x, \xi, t) \mapsto (y, \eta, t), \quad \text{where } |x| + |\xi - \xi^0| < c,\ a_0 - c < t < b_0 + c,$$

is a diffeomorphism.

In the range X_c of the map (ii) we let v denote the image of the vector field $\partial/\partial t$ under the map. Thus v is the tangent vector field of the integral curves, and when $f = df = 0$ we have $v = \partial/\partial t$. By assumption $f = 0$ implies $\partial f/\partial t \le 0$ in a neighborhood of Γ (see (2-18)), so if c is small enough this also holds in X_c. An application of [Hörmander 1985b, Lemma 26.4.11] now yields that f must have a change of sign from $-$ to $+$ along an integral curve of v in X_c, for otherwise there would be no such sign change for increasing t and fixed (x, ξ), and that contradicts the hypothesis in Theorem 2.21. By (4-6) this means that $\operatorname{Im} w_0(t)$ will start decreasing and end increasing, so the minimum is attained at an interior point. We can normalize the minimum value to zero and have then for a suitable interval of t that $\operatorname{Im} w_0 > 0$ at the end points and $\operatorname{Im} w_0 = 0$ at some interior point. Since $\operatorname{Re} w_0$ is given by (4-5) we can at this interior point also normalize the value of $\operatorname{Re} w_0$ to zero. This completes the proof of [Hörmander 1985b, Lemma 26.4.14]. However, in order to prove Theorem 2.21 when $a_0 < b_0$ we shall need the following stronger result.

Lemma 4.1. *Assume that the hypotheses of Theorem 2.21 are fulfilled, the variables being denoted* (t, x) *now. Then given* $M \in \mathbb{N}$ *we can find*

(i) *a curve* $t \mapsto (t, y(t), 0, \eta(t)) \in \mathbb{R}^{2n}$, *with* $a' \le t \le b'$ *as close to* Γ *as desired,*

(ii) C^∞ *functions* $w_\alpha(t)$ *for* $2 \le |\alpha| \le M$, *with* $(\operatorname{Im} w_{jk} - \delta_{jk}/2)$ *positive definite when* $a' \le t \le b'$,

(iii) *a function* $w_0(t)$ *with* $\operatorname{Im} w_0(t) \ge 0$ *for* $a' \le t \le b'$, $\operatorname{Im} w_0(a') > 0$, $\operatorname{Im} w_0(b') > 0$ *and* $\operatorname{Re} w_0(c') = \operatorname{Im} w_0(c') = 0$ *for some* $c' \in (a', b')$

such that (4-4) *is a formal solution to* (4-3) *with an error of order* $\mathcal{O}(|x - y(t)|^{M+1})$. *If* $a_0 < b_0$ *then* (iii) *can be improved in the sense that if* $\varrho \ge 0$ *is the number given by Theorem 2.21, then we can for any* $\varepsilon > \varrho$ *find*

(iii)′ *a function* $w_0(t)$ *with* $\operatorname{Im} w_0(t) \ge 0$, $a' \le t \le b'$, $\operatorname{Im} w_0(a') > 0$, $\operatorname{Im} w_0(b') > 0$ *and* $\operatorname{Re} w_0(t) = \operatorname{Im} w_0(t) = 0$ *for all* $t \in [a_0 + \varepsilon, b_0 - \varepsilon]$.

Proof. In view of [Hörmander 1985b, Lemma 26.4.14] we only need to prove (iii)$'$.

Let $\varepsilon > \varrho$, and let $I_\varepsilon = [a_0 + \varepsilon, b_0 - \varepsilon]$. By the hypotheses of Theorem 2.21, there is a neighborhood $\mathcal{U}$ of

$$\Gamma_\varepsilon = \{(t, 0, 0, \xi^0) : t \in I_\varepsilon\}$$

where f vanishes identically. Take $\delta > 0$ sufficiently small so that

$$t \in I_\varepsilon,\ |x| + |\xi - \xi^0| < \delta \quad \text{implies} \quad (t, x, 0, \xi) \in \mathcal{U}.$$

As above we can find $c > 0$ such that the equations (4-8) and (4-9) with initial data (4-10) and (4-11) have a unique solution in $(a_0 - c, b_0 + c)$ for all x and ξ with $|x| + |\xi - \xi^0| < c$. Since the map

$$(x, \xi, t) \mapsto (y, \eta, t), \quad \text{where } |x| + |\xi - \xi^0| < c,\ a_0 - c < t < b_0 + c,$$

is a diffeomorphism, we can choose c small enough so that if (y, η, t) is in the range X_c of this map, then $|y| + |\eta - \xi^0| < \delta$. As we have seen, f must change sign from $-$ to $+$ along an integral curve of v in X_c if c is small enough, where in X_c we denote by v the image of the vector field $\partial/\partial t$ under the map. Let this integral curve be given by

$$\gamma(t) = (t, y(t), 0, \eta(t)) \in \mathbb{R}^{2n} \quad \text{for } a' \le t \le b',$$

for some choice of a' and b' such that $a_0 - c < a'$, $b' < b_0 + c$ and

$$f(a', y(a'), \eta(a')) < 0 < f(b', y(b'), \eta(b')).$$

Recall that at a point where $f = df = 0$ the equations (4-8)$'$ and (4-8)$''$ imply that $dy/dt = d\eta/dt = 0$. Since f vanishes identically on γ for $t \in I_\varepsilon$ and the function w_0 is determined by (4-5), this proves the lemma after a suitable normalization. □

Note that if Γ is a point then by Lemma 4.1 we can obtain a sequence $\{\gamma_j\}$ of curves

$$\gamma_j(t) = (t, y_j(t), 0, \eta_j(t)) \quad \text{for } a'_j \le t \le b'_j$$

approaching Γ, which implies that at $t = c'_j$ we have

$$(c'_j, y_j(c'_j), 0, \eta_j(c'_j)) \to \Gamma \quad \text{as } j \to \infty$$

in $T^*(\mathbb{R}^n) \setminus 0$, where c'_j is the point where $\operatorname{Re} w_{0j} = \operatorname{Im} w_{0j} = 0$. Similarly, if Γ is an interval and $\varrho \ge 0$ is the number given by Theorem 2.21, then for any point ω in the interior of Γ_ϱ we can use Lemma 4.1 to obtain a sequence $\{\gamma_j\}$ of curves approaching Γ and a sequence $\{w_{0j}\}$ of functions such that for each j there exists a point $\omega_j \in \gamma_j$ with $\omega_j = \gamma_j(t_j)$ that can be chosen so that $\operatorname{Re} w_{0j}(t_j) = \operatorname{Im} w_{0j}(t_j) = 0$ and $\omega_j \to \omega$ as $j \to \infty$. This will be crucial in proving Theorem 2.21. Our strategy is to show that all the terms in the asymptotic sum of the symbol of R have vanishing Taylor coefficients at ω_j, or at $(c'_j, y_j(c'_j), 0, \eta_j(c'_j))$ when Γ is a point. Theorem 2.21 will then follow by continuity. In what follows we will suppress the index j to simplify notation.

Let K and Ω be the cones given by Theorem 2.21, and suppose that the function w given by (4-4) is a formal solution to (4-3) with an error of order $\mathbb{O}(|x-y(t)|^{M+1})$ in a neighborhood Y of

$$\{(t,0): a_0 \le t \le b_0\} \subset \mathbb{R}^n$$

with $K \subset T^*(Y)$, such that $\operatorname{Im} w > 0$ in Y except on a compact nonempty subset T of the curve $x = y(t)$, with $(t_0, y(t_0)) \in T$ and $w = 0$ on T. We want to show that all the terms in the asymptotic sum of the symbol of R have vanishing Taylor coefficients at $(t_0, y(t_0), 0, \eta(t_0))$. By part (i) of Lemma 4.1 we can choose w so that

$$\Gamma_0 = \{(t, x, \partial w(t,x)/\partial t, \partial w(t,x)/\partial x) : (t,x) \in T\} \tag{4-12}$$

is contained in Ω. This is done to ensure that if A is a given pseudodifferential operator with wavefront set contained in the complement of K, then $WF(A)$ does not meet the cone generated by Γ_0.

We now turn our attention to the amplitude functions ϕ_j. With the exception of ϕ_0, which will be of great interest to us, we will not be very thorough in describing them. Suffice it to say that these functions can be chosen so that if P^* is the adjoint of P then

$$\|P^* v_\tau\|_{(\nu)} \le C\tau^{N+n+\nu+(1-M)/2}, \tag{4-13}$$

where M is the number given by (4-2). The procedure begins by setting

$$\phi_0(t,x) = \sum_{|\alpha|<M} \phi_{0\alpha}(t)(x-y(t))^\alpha$$

with $y(t)$ as above, and having $\phi_{0\alpha}$ satisfy the linear system of ordinary differential equations

$$D_t \phi_{0\alpha} + \sum_{|\beta|<M} a_{\alpha\beta}\phi_{0\beta} = 0. \tag{4-14}$$

In the same way we then successively choose ϕ_j and obtain (4-13). The precise details can be found in [Hörmander 1981, pages 87–89], or in [Hörmander 1985b, pages 107–110]. Note that we for any positive integer $J < M$ can solve the equations that determine ϕ_0 so that at the point $(t_0, y(t_0)) \in T$ we have $D_x^\alpha \phi_0(t_0, y(t_0)) = 0$ for all $|\alpha| \le J$ except for one index α, $|\alpha| = J$. This will be important later on. Note also that the estimate (4-13) is not affected if the functions ϕ_j are multiplied by a cutoff function in $C_0^\infty(Y)$ that is 1 in a neighborhood of T. Since the ϕ_j will be irrelevant outside of Y for large τ by construction, we can in this way choose them to be supported in Y so that $v_\tau \in C_0^\infty(Y)$.

Having completed the construction of the approximate solutions, we are now ready to start to follow the proof of Theorem 2.20. To get the estimates for the right side of (2-1) when v is an approximate solution, we shall need the following two results. The first, corresponding to Lemma 3.1, is taken from [Hörmander 1985b]. Observe that here it is stated for our approximate solutions which differ from those in [ibid.] by a factor of τ^{N+n}, which explains the difference in appearance. Note also that although we will not use the lower bound for the approximate solutions, that estimate is included so as not to alter the statement.

Lemma 4.2 [Hörmander 1985b, Lemma 26.4.15]. *Let $X \subset \mathbb{R}^n$ be open, and let v_τ be defined by* (4-2), *where $w \in C^\infty(X)$, $\phi_j \in C_0^\infty(X)$, $\operatorname{Im} w \geq 0$ in X and $d \operatorname{Re} w \neq 0$. For any positive integer m we then have*

$$\|v_\tau\|_{(-m)} \leq C\tau^{N+n-m} \quad \text{for } \tau > 1. \tag{4-15}$$

If $\operatorname{Im} w(t_0, x_0) = 0$ and $\phi_0(t_0, x_0) \neq 0$ for some $(t_0, x_0) \in X$ then

$$\|v_\tau\|_{(-m)} \geq c\tau^{N+n/2-m} \quad \text{for } \tau > 1$$

and for some $c > 0$. If $\tilde{\Gamma}$ is the cone generated by

$$\bigl\{(t, x, \partial_t w(t, x), \partial_x w(t, x)) : (t, x) \in \textstyle\bigcup_j \operatorname{supp} \phi_j,\ \operatorname{Im} w(t, x) = 0\bigr\},$$

then $\tau^k v_\tau \to 0$ in $\mathscr{D}'_{\tilde{\Gamma}}$ as $\tau \to \infty$; hence $\tau^k A v_\tau \to 0$ in $C^\infty(\mathbb{R}^n)$ if A is a pseudodifferential operator with $WF(A) \cap \tilde{\Gamma} = \varnothing$ and k is any real number.

Proposition 4.3. *Assume that the hypotheses of Theorem 2.21 are fulfilled, the variables being denoted (t, x) now, and let v_τ be given by* (4-2), *where $w \in C^\infty(Y)$, $\phi_j \in C_0^\infty(Y)$, $\operatorname{Im} w \geq 0$ in Y and $d \operatorname{Re} w \neq 0$. Here Y is a neighborhood of $\{(t, 0) : a_0 \leq t \leq b_0\}$ such that $K \subset T^*(Y)$. Let $H(t, x) \in C_0^\infty(\mathbb{R} \times \mathbb{R}^{n-1})$ and set*

$$h_\tau(t, x) = \tau^{-N} H(\tau(t - t_0), \tau(x - y(t))), \tag{4-16}$$

where N is the positive integer given by Definition 2.1 for the operators R and P in Theorem 2.21. Then $h_\tau \in H_{(N)}(\mathbb{R}^n)$ for all $\tau \geq 1$ and $\|h_\tau\|_{(N)} \leq C$, where the constant depends on H but not on τ. Furthermore, if M is the integer given by the definition of v_τ in (4-2) *so that* (4-13) *holds, and I_τ is the integral*

$$I_\tau = (R^* v_\tau, \overline{h_\tau}), \tag{4-17}$$

where R^* is the adjoint of $R(t, x, D)$, then for any positive integer κ there exists a constant C such that $|I_\tau| \leq C\tau^{-\kappa}$ if $M = M(\kappa)$ is sufficiently large.

Proof. In Section 3, one easily obtains a formula for the Fourier transform of the corresponding function h_τ (see (3-4)), which yields the estimates needed to show that $h_\tau \in H_{(N)}$. Here we shall instead use the equality

$$\iint |h_\tau(t, x)|^2 dt\, dx = \tau^{-2N} \iint |H(\tau(t - t_0), \tau(x - y(t)))|^2 dt\, dx$$

which shows that if $\tau \geq 1$ then $D_t^j D_x^\alpha h_\tau \in L^2(\mathbb{R}^n)$ for all $(j, \alpha) \in \mathbb{N} \times \mathbb{N}^{n-1}$ such that $j + |\alpha| \leq N + [n/2]$. Hence, by using the equivalent norm on $H_{(N)}(\mathbb{R}^n)$ given by

$$\|h_\tau\|_{(N)} = \sum_{j+|\alpha| \leq N} \|D_t^j D_x^\alpha h_\tau\|_{(0)},$$

we find that $\{h_\tau\}_{\tau \geq 1}$ is a bounded one parameter family in $H_{(N)}(\mathbb{R}^n)$, which proves the first assertion of the proposition.

To prove the second part, let κ be an arbitrary positive integer, and let ν be the positive integer given by Lemma 2.3 (applied to the operator R instead of Q) so that (2-1) holds for the choice of seminorm $\|P^* v\|_{(\nu)}$ in the right side. If we choose

$$(1-M)/2 \le -N-n-\nu-\kappa, \tag{4-18}$$

and recall (4-13), then

$$\|P^* v_\tau\|_{(\nu)} \le C\tau^{-\kappa}. \tag{4-19}$$

Since $\operatorname{supp} H$ is compact, we can find a bounded open ball containing $\operatorname{supp} h_\tau$ for all $\tau \ge 1$. Hence $h_\tau \in H_{(N)}(\mathbb{R}^n)$ has compact support and $v_\tau \in C_0^\infty(Y)$, so the result now follows by the estimate (2-4) together with Lemma 4.2. ☐

To shorten the notation we will from now on assume that $t_0 = 0$, so that $w(0, y(0)) = 0$. As in the proof of Theorem 2.20 it suffices to show that all terms in the asymptotic expansion of the symbol of R^*, given by

$$\sigma_{R^*} = q_1(t,x,\xi) + q_0(t,x,\xi) + \cdots,$$

with q_j homogeneous of degree j in ξ, have vanishing Taylor coefficients at $(0, y(0), \eta(0))$. The method will be to argue by contradiction that if not, then Proposition 4.3 does not hold. Therefore, let us assume that $\partial_t^{k_0} q_{-j_0(\alpha_0)}^{(\beta_0)}(0, y(0), \eta(0))$ is the first nonvanishing Taylor coefficient with respect to the ordering $>_t$ given by Definition 3.2, and let

$$m = j_0 + k_0 + |\alpha_0| + |\beta_0|. \tag{4-20}$$

Now let κ be a positive integer such that $m < \kappa$, and sort the terms in I_τ, given by (4-17), with respect to homogeneity degree in τ. We can use Lemma 3.3 and the classicality of the symbol σ_{R^*} to write

$$\begin{aligned} R^*(t,x,D)v_\tau &= \sum_{j=-1}^{M'} q_{-j}(t,x,D)v_\tau + \mathcal{O}(\tau^{N+n-M'-1}) \\ &= \sum_{j=-1}^{M'} \sum_{l=0}^{M} \tau^{N+n-l} q_{-j}(t,x,D)(e^{i\tau w}\phi_l) + \mathcal{O}(\tau^{N+n-M'-1}) \end{aligned}$$

for some large number M'. Note that (4-18) implies a lower bound on M, but as we shall see below, we must also make sure to pick $M > 2M'+1$. For each j we then estimate $q_{-j}(t,x,D)(e^{i\tau w}\phi_l)$ using (3-11) with $k = M-1-2j$, so that

$$q_{-j}(t,x,D)(e^{i\tau w}\phi_l) = \sum_{|\alpha|<M-1-2j} q_{-j}^{(\alpha)}(t,x,\tau\eta)(D-\tau\eta)^\alpha(\phi_l e^{i\tau w})/\alpha!$$

with an error of order $\mathbb{O}(\tau^{(1-M)/2})$. Recalling (4-18) and the discussion following Lemma 3.3 regarding the homogeneity of the terms in (3-11), this yields, for sufficiently large M',

$$\begin{aligned} R^*(t,x,D)v_\tau &= \sum_{j=-1}^{M'}\sum_{l=0}^{M}\tau^{N+n-l}e^{i\tau w}\sum_{|\alpha|<M-1-2j} q_{-j}^{(\alpha)}(t,x,\tau w_x')D^\alpha\phi_l + \mathbb{O}(\tau^{-\kappa-1}) \\ &= \tau^{N+n}e^{i\tau w}\sum_{j=-1}^{M'}\sum_{l=0}^{M}\sum_{|\alpha|<M-1-2j}\tau^{-j-|\alpha|-l}q_{-j}^{(\alpha)}(t,x,w_x')D^\alpha\phi_l + \mathbb{O}(\tau^{-\kappa-1}). \end{aligned} \tag{4-21}$$

Note that $\tau^{-j-|\alpha|-l}q_{-j}^{(\alpha)}(t,x,w_x')D^\alpha\phi_l$ is now homogeneous of order $-j-|\alpha|-l$ in τ, and that as before, $q_{-j}^{(\alpha)}(t,x,w_x')$ should be replaced by a finite Taylor expansion at η of sufficiently high order. For each $-1\le J\le\kappa$, collect all terms of the form $\tau^{-j-|\alpha|-l}q_{-j}^{(\alpha)}(t,x,w_x')D^\alpha\phi_l$ in (4-21) that are homogeneous of order $-J$ in τ, that is, all terms that satisfy $j+|\alpha|+l=J$ for $j\ge -1$, and $|\alpha|,l\ge 0$. If

$$\lambda_J(t,x)=\sum_{j+|\alpha|+l=J} q_{-j}^{(\alpha)}(t,x,w_x'(t,x))D^\alpha\phi_l(t,x)$$

for the permitted values of j and l, then

$$I_\tau=\tau^n\iint H(\tau t,\tau(x-y(t)))\Big(e^{i\tau w(t,x)}\sum_{J=-1}^{\kappa}\tau^{-J}\lambda_J(t,x)+\mathbb{O}(\tau^{-\kappa-1})\Big)\,dt\,dx.$$

After the change of variables $(\tau t,\tau(x-y(t)))\mapsto(t,x)$ we obtain

$$I_\tau=\iint H(t,x)\Big(e^{i\tau w(t/\tau,x/\tau+y(t/\tau))}\sum_{J=-1}^{\kappa}\tau^{-J}\lambda_J(t/\tau,x/\tau+y(t/\tau))+\mathbb{O}(\tau^{-\kappa-1})\Big)\,dt\,dx, \tag{4-22}$$

where

$$\begin{aligned} &\lambda_J(t/\tau,x/\tau+y(t/\tau)) \\ &\quad=\sum_{j+|\alpha|+l=J} D^\alpha\phi_l(t/\tau,x/\tau+y(t/\tau))q_{-j}^{(\alpha)}(t/\tau,x/\tau+y(t/\tau),w_x'(t/\tau,x/\tau+y(t/\tau))). \end{aligned} \tag{4-23}$$

Recall that $w_0(0)=0$, which together with (4-4) implies

$$i\tau w(t/\tau,x/\tau+y(t/\tau))=itw_0'(0)+i\langle x,\eta(t/\tau)\rangle+\mathbb{O}(\tau^{-1}).$$

Hence

$$\lim_{\tau\to\infty} e^{i\tau w(t/\tau,x/\tau+y(t/\tau))}=e^{itw_0'(0)+i\langle x,\eta(0)\rangle}. \tag{4-24}$$

In the sequel we shall also need

$$\partial w/\partial x_j(t/\tau,x/\tau+y(t/\tau))-\eta_j(t/\tau)=\sum_{k=1}^{n-1}w_{j,k}(t/\tau)(x_k/\tau)+\mathbb{O}(\tau^{-2}), \tag{4-25}$$

which follows from the definition of w and the fact that w_α is symmetric in these special indices α. In particular, $w_{j,k}(t)=w_{k,j}(t)$ for all $j,k\in[1,n-1]$.

Recall that we chose the integer κ such that $m < \kappa$. By Proposition 4.3 there is a constant C such that

$$|I_\tau| \le C\tau^{-\kappa}, \tag{4-26}$$

and we shall now show that if $\partial_t^{k_0} q_{-j_0(\alpha_0)}^{(\beta_0)}(0, y(0), \eta(0))$ is the first nonvanishing Taylor coefficient with respect to the ordering $>_t$, where $m = j_0 + k_0 + |\alpha_0| + |\beta_0|$, then (4-26) cannot hold. (Since we are denoting the variables by (t, x) now, the index α in Definition 3.2 will be replaced by the pair $(k, \alpha) \in \mathbb{N} \times \mathbb{N}^{n-1}$.) We will do this by determining the limit of $\tau^m I_\tau$ as $\tau \to \infty$. To see what is needed, consider $\lambda_{-1}(t/\tau, x/\tau + y(t/\tau))$ and recall that this is

$$q_1(t/\tau, x/\tau + y(t/\tau), w_x'(t/\tau, x/\tau + y(t/\tau)))\phi_0(t/\tau, x/\tau + y(t/\tau)),$$

which should be regarded as a Taylor expansion in ξ of q_1 at $\eta(t/\tau)$ of finite order. The same applies to all the other terms of the form $q_{-j}^{(\alpha)}$. For given j and α, we only ever need to consider Taylor expansions of $q_{-j}^{(\alpha)}$ of order $\kappa - j - |\alpha|$ in view of (4-22) and (4-25). To keep things simple, we shall first only consider q_1; it will be clear by symmetry what the corresponding expressions for the other terms should be. Thus,

$$\begin{aligned} &q_1(t/\tau, x/\tau + y(t/\tau), w_x'(t/\tau, x/\tau + y(t/\tau))) \\ &= \sum_{|\beta| \le \kappa+1} q_1^{(\beta)}(t/\tau, x/\tau + y(t/\tau), \eta(t/\tau))(w_x'(t/\tau, x/\tau + y(t/\tau)) - \eta(t/\tau))^\beta/|\beta|! + \mathbb{O}(\tau^{-\kappa-2}), \end{aligned} \tag{4-27}$$

which shows that to use our assumption regarding the Taylor coefficient $\partial_t^{k_0} q_{-j_0(\alpha_0)}^{(\beta_0)}(0, y(0), \eta(0))$, we have to for each β write $q_1^{(\beta)}(t/\tau, x/\tau + y(t/\tau), \eta(t/\tau))$ as a Taylor series at $\eta(0)$, in addition to having to expand each term as a Taylor series in t and x. However, it is immediate from (4-25) that if β is an $(n-1)$-tuple corresponding to a given differential operator D_ξ^β, then there is a sequence $\tilde\beta = (\tilde\beta_1, \dots, \tilde\beta_s)$ of $s = |\beta|$ indices between 1 and the dimension $n-1$ of the x variable, such that

$$g_\tau^\beta(t, x) = (w_x'(t/\tau, x/\tau + y(t/\tau)) - \eta(t/\tau))^\beta, \tag{4-28}$$

as it appears in (4-27), satisfies

$$g_\tau^\beta(t, x) = c_\beta(t/\tau, x/\tau) + \mathbb{O}(\tau^{-|\beta|-1}),$$

where

$$c_\beta(t/\tau, x/\tau) = \prod_{j=1}^{s}\Big(\sum_{k=1}^{n-1} w_{k,\tilde\beta_j}(t/\tau)x_k/\tau\Big) \quad \text{and} \quad c_\beta(0, x/\tau) = \tau^{-|\beta|}c_\beta(0, x).$$

These expressions make sense if we choose the sequence $\tilde\beta$ to be increasing, for then it is uniquely determined by β. If for instance $D_\xi^\beta = -\partial^2/\partial\xi_i\partial\xi_j$, then $\tilde\beta = (i, j)$ if $i \le j$ (see the indices α used in connection with w_α in (4-4)). Thus (4-27) takes the form

$$\begin{aligned} &q_1(t/\tau, x/\tau + y(t/\tau), w_x'(t/\tau, x/\tau + y(t/\tau))) \\ &\qquad = \sum_{|\beta| \le \kappa+1} q_1^{(\beta)}(t/\tau, x/\tau + y(t/\tau), \eta(t/\tau))g_\tau^\beta(t, x)/|\beta|! + \mathbb{O}(\tau^{-\kappa-2}), \end{aligned}$$

and if we expand each term in this expression as a Taylor series at $\eta(0)$ we obtain

$$q_1(t/\tau, x/\tau + y(t/\tau), w'_x(t/\tau, x/\tau + y(t/\tau))) \\ = \sum_{|\beta|\le\kappa+1} \sum_{|\gamma|\le\kappa+1-|\beta|} q_1^{(\beta+\gamma)}(t/\tau, x/\tau + y(t/\tau), \eta(0)) g_\tau^\beta(t,x)(\eta(t/\tau)-\eta(0))^\gamma/(|\beta|!|\gamma|!) \\ + \mathcal{O}(\tau^{-\kappa-2}), \quad (4\text{-}29)$$

where we regard $\eta(t/\tau)-\eta(0)$ as a finite Taylor series $\eta'(0)t/\tau + \eta''(0)t^2/(2\tau^2) + \cdots$ of sufficiently high order to maintain control of the error term in (4-29). If we for each multiindex β let $G_\tau^\beta(t,x)$ be given by

$$G_\tau^\beta(t,x) = \sum_{\gamma_1+\gamma_2=\beta} (\eta(t/\tau)-\eta(0))^{\gamma_1} g_\tau^{\gamma_2}(t,x)/(|\gamma_1|!|\gamma_2|!) \quad \text{for } \gamma_j \in \mathbb{N}^{n-1},$$

then the required order of the Taylor expansion $\eta(t/\tau)-\eta(0)$ will ultimately depend on β, so we can write

$$q_1(t/\tau, x/\tau + y(t/\tau), w'_x(t/\tau, x/\tau + y(t/\tau))) \\ = \sum_{|\beta|\le\kappa+1} q_1^{(\beta)}(t/\tau, x/\tau + y(t/\tau), \eta(0)) G_\tau^\beta(t,x) + \mathcal{O}(\tau^{-\kappa-2}) \quad (4\text{-}30)$$

and we can always bound $G_\tau^\beta(t,x)$ by a constant times $\tau^{-|\beta|}$. As will be evident in a moment, the value of $G_\tau^\beta(t,x)$ for $|\beta|>0$ is not important. For notational purposes, denote by $G_0^\beta(t,x)$ the limit of $\tau^{|\beta|}G_\tau^\beta(t,x)$ as $\tau\to\infty$. Since $G_\tau^\beta(t,x)=1$ when $\beta=0$ it is clear that $G_0^0(t,x)=1$.

For each β we must now write $q_1^{(\beta)}(t/\tau, x/\tau + y(t/\tau), \eta(0))$ as a Taylor expansion in t and x at 0 and $y(0)$, respectively. As before, for given j and α, we will only have to consider Taylor expansions of $q_{-j}^{(\alpha)}$ of order $\kappa - j - |\alpha|$. By (4-23) and (4-30) we have

$$\lambda_{-1}(t/\tau, x/\tau + y(t/\tau)) \\ = \sum_{k+|\alpha|+|\beta|\le\kappa+1} \phi_0(t/\tau, x/\tau + y(t/\tau))\big((t/\tau)^k (x/\tau + y(t/\tau) - y(0))^\alpha G_\tau^\beta(t,x) \\ \times \partial_t^k q_{1(\alpha)}^{(\beta)}(0, y(0), \eta(0))/(k!|\alpha|!) + \mathcal{O}(\tau^{-\kappa-2})\big), \quad (4\text{-}31)$$

where we in $(x/\tau + y(t/\tau) - y(0))^\alpha$ regard $y(t/\tau) - y(0)$ as a finite Taylor series of sufficiently high order to maintain control of the error terms.

In the way that we expressed the term $q_1(t/\tau, x/\tau+y(t/\tau), w'_x(t/\tau, x/\tau+y(t/\tau)))$ by (4-31), we can get similar expressions of appropriate order for the terms $q_{-j}^{(\gamma)}(t/\tau, x/\tau+y(t/\tau), w'_x(t/\tau, x/\tau+y(t/\tau)))$ that appear in (4-23). For each j and γ we have

$$q_{-j}^{(\gamma)}(t/\tau, x/\tau + y(t/\tau), w'_x(t/\tau, x/\tau + y(t/\tau))) \\ = \sum_{k+|\alpha|+|\beta|\le\kappa-j-|\gamma|} (t/\tau)^k (x/\tau + y(t/\tau) - y(0))^\alpha G_\tau^\beta(t,x) \partial_t^k q_{-j(\alpha)}^{(\beta+\gamma)}(0, y(0), \eta(0))/(k!|\alpha|!) \\ + \mathcal{O}(\tau^{-\kappa-1+j+|\gamma|}). \quad (4\text{-}32)$$

This together with (4-23) gives

$$\begin{aligned}\lambda_J(t/\tau, x/\tau + y(t/\tau)) = \sum_{j+l+|\gamma|=J} \sum_{k+|\alpha|+|\beta|\le\kappa-j-|\gamma|} & (t/\tau)^k (x/\tau + y(t/\tau) - y(0))^\alpha G_\tau^\beta(t,x) \\ & \times D_x^\gamma \phi_l(t/\tau, x/\tau + y(t/\tau)) \frac{\partial_t^k q_{-j(\alpha)}^{(\beta+\gamma)}(0, y(0), \eta(0))}{k!\,|\alpha|!} \\ & + \mathbb{O}(\tau^{-\kappa-1+j+|\gamma|}),\end{aligned}$$

where $-1 \le j \le J$ and $l \ge 0$. Using that by assumption the Taylor coefficients $\partial_t^k q_{-j(\alpha)}^{(\beta+\gamma)}(0, y(0), \eta(0))$ vanish for all $-1 \le j+k+|\alpha|+|\beta|+|\gamma| < m$, and $\tau^{-J-k-|\alpha|} = \tau^{|\beta|}\tau^{-j-k-|\alpha|-|\beta|-|\gamma|-l}$ when $J = j+l+|\gamma|$, the equation above yields

$$\begin{aligned}\sum_{J=-1}^{m} \tau^{-J}\lambda_J(t/\tau, x/\tau + y(t/\tau)) = \sum_{j+l+|\gamma|=-1}^{m} \sum_{j+k+|\alpha|+|\beta|+|\gamma|=m} & \tau^{-m-l} t^k (x + y'(0)t)^\alpha \tau^{|\beta|} G_\tau^\beta(t,x) \\ & \times D_x^\gamma \phi_l(t/\tau, x/\tau + y(t/\tau)) \frac{\partial_t^k q_{-j(\alpha)}^{(\beta+\gamma)}(0, y(0), \eta(0))}{k!\,|\alpha|!} \\ & + \mathbb{O}(\tau^{-m-1-l}),\end{aligned}$$

where $\tau^{|\beta|} G_\tau^\beta(t,x) \to G_0^\beta(t,x)$ as $\tau \to \infty$. As we can see, the expression above is $\mathbb{O}(\tau^{-m-1})$ as soon as $l > 0$, so in view of (4-22) and (4-24) we obtain

$$\begin{aligned}\lim_{\tau\to\infty} \tau^m I_\tau = \iint & H(t,x) e^{itw_0'(0)+i\langle x,\eta(0)\rangle} \\ & \times \Big(\sum_{j+k+|\alpha|+|\beta|+|\gamma|=m} t^k (x + y'(0)t)^\alpha G_0^\beta(t,x) D_x^\gamma \phi_0(0, y(0)) \\ & \qquad \times \partial_t^k q_{-j(\alpha)}^{(\beta+\gamma)}(0, y(0), \eta(0))/(k!|\alpha|!) \Big)\, dt\, dx. \qquad (4\text{-}33)\end{aligned}$$

Recall (4-20) and choose ϕ_0 so that $D_x^{\beta_0}\phi_0(0, y(0)) = 1$, but so that $D_x^\gamma \phi_0(0, y(0)) = 0$ for all other γ such that $|\gamma| \le |\beta_0|$ (see (4-14)). By the choice of our ordering $>_t$ we have $\partial_t^k q_{-j(\alpha)}^{(\beta+\beta_0)}(0, y(0), \eta(0)) = 0$ for all β such that $|\beta| > 0$ as long as $j+k+|\alpha|+|\beta|+|\beta_0| = m$. Hence, with this choice of ϕ_0, equation (4-33) takes the form

$$\begin{aligned}\lim_{\tau\to\infty} \tau^m I_\tau = \iint & H(t,x) e^{itw_0'(0)+i\langle x,\eta(0)\rangle} \\ & \times \Big(\sum_{j+k+|\alpha|+|\beta_0|=m} t^k (x + y'(0)t)^\alpha \partial_t^k q_{-j(\alpha)}^{(\beta_0)}(0, y(0), \eta(0))/(k!|\alpha|!) \Big)\, dt\, dx, \qquad (4\text{-}34)\end{aligned}$$

so as promised, the value of $G_0^\beta(t,x)$ for $|\beta| > 0$ does not matter. (Note that $G_0^0(t,x)$ is present in (4-34) as the constant factor 1.) As in the proof of Theorem 2.20, some of the Taylor coefficients in (4-34) may be zero. In particular, the expression may well contain Taylor coefficients that preceed $\partial_t^{k_0} q_{-j_0(\alpha_0)}^{(\beta_0)}(0, y(0), \eta(0))$ in the ordering, and those are by assumption zero. In contrast to the proof of Theorem 2.20 we shall have to exploit this fact, since the coefficients of most of the monomials in (4-34) will be linear combinations of the Taylor coefficients due to the factor $(x + y'(0)t)^\alpha$. However, the ordering $>_t$ was chosen so that there can be no nonzero Taylor coefficient $\partial_t^k q_{-j(\alpha)}^{(\beta_0)}(0, y(0), \eta(0))$ such that $k + |\alpha| > k_0 + |\alpha_0|$, or $k + |\alpha| = k_0 + |\alpha_0|$ and $k < k_0$. This follows immediately from the choice of lexicographic order on the n-tuple $(k,\alpha)\in\mathbb{N}^n$. (Recall that in the definition of the ordering $>_t$, x denoted all the variables in $\mathbb{R}^n$, while here we denote those variables by (t,x).) Hence, the only coefficient of the monomial $t^{k_0}x^{\alpha_0}$ in (4-34) is $\partial_t^{k_0}q_{-j_0\,(\alpha_0)}^{(\beta_0)}(0,y(0),\eta(0))$. We may therefore, as in the proof of Theorem 2.20, choose H so that the limit in (4-34) is nonzero. Since this contradicts (4-26), Theorem 2.21 follows in view of the discussion following Lemma 4.1.

Appendix A.

Here we prove a few results used in the main text, related to how the property that all terms in the asymptotic expansion of the total symbol have vanishing Taylor coefficients is affected by various operations.

Lemma A.1. *Suppose X and Y are two C^∞ manifolds of the same dimension n. Let $K\subset T^*(X)\smallsetminus 0$ and $K'\subset T^*(Y)\smallsetminus 0$ be compactly based cones and let χ be a homogeneous symplectomorphism from a conic neighborhood of K' to one of K such that $\chi(K')=K$. Let $A\in I^{m'}(X\times Y,\Gamma')$ and $B\in I^{m''}(Y\times X,(\Gamma^{-1})')$, where Γ is the graph of χ, and assume that A and B are properly supported and noncharacteristic at the restriction of the graphs of χ and χ^{-1} to K' and to K respectively, while $WF'(A)$ and $WF'(B)$ are contained in small conic neighborhoods. If R is a properly supported classical pseudodifferential operator in Y, then each term in the asymptotic expansion of the total* (*left*) *symbol of R has vanishing Taylor coefficients at a point $(y,\eta)\in K'$ if and only if each term in the asymptotic expansion of the total* (*left*) *symbol of the pseudodifferential operator ARB in X has vanishing Taylor coefficients at $\chi(y,\eta)\in K$.*

Proof. We may assume that we have a homogeneous generating function $\varphi\in C^\infty$ for the symplectomorphism χ; see [Grigis and Sjöstrand 1994, pages 101–103]. Then χ is locally of the form

$$(\partial\varphi(x,\eta)/\partial\eta,\eta)\mapsto(x,\partial\varphi(x,\eta)/\partial x),$$

and A and B are given by

$$Au(x)=\frac{1}{(2\pi)^n}\iint e^{i(\varphi(x,\zeta)-z\cdot\zeta)}a(x,z,\zeta)u(z)\,dz\,d\zeta,$$
$$Bv(y)=\frac{1}{(2\pi)^n}\iint e^{i(y\cdot\theta-\varphi(s,\theta))}b(y,s,\theta)v(s)\,ds\,d\theta.$$

Since R is properly supported we may assume that

$$Ru(z)=\frac{1}{(2\pi)^n}\int e^{iz\cdot\eta}r(z,\eta)\hat{u}(\eta)\,d\eta\quad\text{for } u\in C_0^\infty(Y),\tag{A-1}$$

where $r(z,\eta)=\sigma_R$ is the total symbol of R. Hence

$$ARBu(x)=\frac{1}{(2\pi)^{3n}}\int e^{i(\varphi(x,\zeta)-z\cdot\zeta+(z-y)\cdot\sigma+y\cdot\theta-\varphi(s,\theta))}a(x,z,\zeta)r(z,\sigma)b(y,s,\theta)u(s)\,ds\,d\theta\,dy\,d\sigma\,dz\,d\zeta,\tag{A-2}$$

since B being properly supported implies that $Bu\in C_0^\infty(Y)$ when $u\in C_0^\infty(Y)$. Using integration by parts in z, we see that we can insert a cutoff $\phi((\zeta-\sigma)/|\sigma|)$ in the last integral without changing the operator

$ARB \bmod \Psi^{-\infty}$. If we make the change of variables $\tau = \zeta - \sigma$, then (A-2) takes the form

$$ARBu(x) = \frac{1}{(2\pi)^{3n}} \int \phi(\tau/|\sigma|) e^{i(\varphi(x,\tau+\sigma) - z\cdot(\tau+\sigma) + (z-y)\cdot\sigma + y\cdot\theta - \varphi(s,\theta))} \\ \times a(x,z,\tau+\sigma) r(z,\sigma) b(y,s,\theta) u(s)\, ds\, d\theta\, dy\, d\sigma\, dz\, d\tau + Lu,$$

with $L \in \Psi^{-\infty}$. If $\Omega \subset \mathbb{R}^{2n}$ is open and $\tilde{\varphi} \in C^\infty(\Omega, \mathbb{R})$ is a phase function with a nondegenerate critical point $x_0 \in \Omega$ such that $d\tilde{\varphi} \neq 0$ everywhere else, then [Grigis and Sjöstrand 1994, Proposition 2.3] states, in particular, that for every compact $M \subset \Omega$ and every $u \in C^\infty(\Omega) \cap \mathscr{E}'(M)$ we have

$$\Bigl| \int e^{i\lambda\tilde{\varphi}(x)} u(x) dx - e^{i\lambda\tilde{\varphi}(x_0)} A_0 u(x_0) \lambda^{-n} \Bigr| \le C_M \lambda^{-n-1} \sum_{|\alpha| \le 2n+3} \sup |\partial^\alpha u(x)| \quad \text{for } \lambda \ge 1, \tag{A-3}$$

where

$$A_0 = \frac{(2\pi)^n \cdot e^{i\pi \operatorname{sgn} \tilde{\varphi}''(x_0)/4}}{|\det \tilde{\varphi}''(x_0)|^{1/2}}. \tag{A-4}$$

It is clear that the result extends to the setting $\Omega = T^*(\mathcal{N}) \smallsetminus 0$, where $\mathcal{N}$ is a C^∞ manifold of dimension n. In order to apply the result, we put $\sigma = \lambda\omega$, and make the change of variables $\tau = \lambda\tilde{\tau}$. After dropping the tilde we obtain

$$ARBu(x) = \frac{\lambda^{2n}}{(2\pi)^{3n}} \int \phi(\tau/|\omega|) e^{i\lambda(\varphi(x,\tau+\omega) - z\cdot(\tau+\omega) + y\cdot\theta/\lambda + (z-y)\cdot\omega - \varphi(s,\theta)/\lambda)} \\ \times a(x,z,\lambda(\tau+\omega)) r(z,\lambda\omega) b(y,s,\theta) u(s)\, ds\, d\theta\, dy\, d\omega\, dz\, d\tau + Lu,$$

where we have used the fact that φ is homogeneous of degree 1 in the fiber. For the z, τ-integration we have the nondegenerate critical point given by $\tau = 0$, $z = \varphi'_\zeta(x, \tau+\omega)$. Note that since φ'_ζ is homogeneous of degree 0 in the fiber we have $\varphi'_\zeta(x, \sigma/\lambda) = \varphi'_\zeta(x,\sigma)$, so this critical point corresponds to the critical point for the z, ζ-integration given by $\zeta = \sigma$, $z = \varphi'_\zeta(x,\sigma)$. Hence the expression above together with (A-3) imply that

$$ARBu(x) = C\lambda^{2n} \int e^{i(\varphi(x,\lambda\omega) + y\cdot\theta - y\cdot\lambda\omega - \varphi(s,\theta))} w(x,y,s,\omega,\theta) u(s)\, ds\, d\theta\, dy\, d\omega + Lu,$$

where

$$\begin{aligned} w(x,y,s,\omega,\theta) &= \frac{A_0}{\lambda^n} a(x,z,\lambda(\tau+\omega)) r(z,\lambda\omega) b(y,s,\theta) \phi(\tau/|\omega|) \Big|_{\substack{\tau=0,\\ z=\varphi'_\zeta(x,\omega)}} \\ &= \frac{A_0}{\lambda^n} a(x, \varphi'_\zeta(x,\omega), \lambda\omega) r(\varphi'_\zeta(x,\omega), \lambda\omega) b(y,s,\theta) \end{aligned}$$

with an error of order $\mathcal{O}(\lambda^{-n-1})$. Note that A_0 is now a function of x and ω, since the matrix corresponding to $\tilde{\varphi}''(x_0)$ in (A-4) is given by the block matrix

$$F = \begin{pmatrix} 0 & -\mathrm{Id}_n \\ -\mathrm{Id}_n & \varphi''_{\zeta\zeta}(x,\omega) \end{pmatrix}, \tag{A-5}$$

where Id_n is the identity matrix on $\mathbb{R}^n$. Clearly the determinant of F is either 1 or -1, so F is nonsingular. Furthermore, F depends smoothly on the parameters x and ω since $\varphi \in C^\infty$, so the eigenvalues of F

are continuous in x and ω. Hence it follows that the signature of F is constant, for if not there has to exist an eigenvalue vanishing at some point (x, ω), contradicting the nonsingularity of F. Reverting to the variable $\sigma = \lambda\omega$ we thus obtain

$$ARBu(x) = C \int e^{i(\varphi(x,\sigma)+y\cdot(\theta-\sigma)-\varphi(s,\theta))}\tilde{w}(x, y, s, \sigma, \theta)u(s)\, ds\, d\theta\, dy\, d\sigma + Lu,$$

where

$$\tilde{w}(x, y, s, \sigma, \theta) = a(x, \varphi'_\zeta(x, \sigma), \sigma)r(\varphi'_\zeta(x, \sigma), \sigma)b(y, s, \theta)$$

with an error of order $\mathbb{O}(\lambda^{-1})$. Taking the limit as $\lambda \to \infty$ yields

$$ARBu(x)=C\int e^{i(\varphi(x,\sigma)+y\cdot(\theta-\sigma)-\varphi(s,\theta))}a(x, \varphi'_\zeta(x, \sigma), \sigma)r(\varphi'_\zeta(x, \sigma), \sigma)b(y, s, \theta)u(s)\,ds\,d\theta\,dy\,d\sigma+Lu.$$

We can now repeat the procedure. Indeed, we can insert a cutoff $\phi((\sigma-\theta)/|\theta|)$ without changing the operator mod $\Psi^{-\infty}$, and after making the corresponding changes of variables in order to apply [Grigis and Sjöstrand 1994, Proposition 2.3] we find that for the y, σ-integration we have the nondegenerate critical point given in the original variables by $\sigma = \theta$, $y = \varphi'_\sigma(x, \sigma)$. After taking the limit as $\lambda \to \infty$ we obtain

$$ARBu(x) = C \int e^{i(\varphi(x,\theta)-\varphi(s,\theta))}w_1(x, s, \theta)u(s)\, ds\, d\theta + L_1u,$$

where $L_1 \in \Psi^{-\infty}$ and

$$w_1(x, s, \theta) = a(x, \varphi'_\theta(x, \theta), \theta)r(\varphi'_\theta(x, \theta), \theta)b(\varphi'_\theta(x, \theta), s, \theta). \tag{A-6}$$

As before we let the factor A_0 from (A-4) be included in the constant C. In a conic neighborhood of $\operatorname{supp} w_1$ we can write

$$\varphi(x, \theta) - \varphi(s, \theta) = (x - s)\Xi(x, s, \theta).$$

Then $\Xi(x, x, \theta) = \varphi'_x(x, \theta)$ so $\partial\Xi(x, x, \theta)/\partial\theta = \varphi''_{x\theta}(x, \theta)$ is invertible, since $\varphi''_{x\theta}(x, \theta) \neq 0$ is equivalent to the fact that the graph of χ is (locally) the graph of a smooth map. Hence $\theta \mapsto \Xi(x, s, \theta)$ is C^∞, homogeneous of degree 1 and with an inverse having the same properties. For s close to x, the equation $\Xi(x, s, \theta) = \xi$ then defines $\theta = \Theta(x, s, \xi)$. After a change of variables, the last integral therefore takes the form

$$ARBu(x) = C \int e^{i(x-s)\cdot\xi}\tilde{w}_1(x, s, \xi)u(s)\, ds\, d\xi + L_1u, \tag{A-7}$$

where $\tilde{w}_1(x, s, \xi)$ is just $w_1(x, s, \Theta(x, s, \xi))$ multiplied by a Jacobian. We note in passing that evaluating $\tilde{w}_1$ at a point (x, x, ξ) where ξ is of the form $\xi = \varphi'_x(x, \eta)$ therefore involves evaluating w_1 at the point (x, x, η). The integral (A-7) defines a pseudodifferential operator with total symbol $\rho(x, \xi)$ satisfying

$$\rho(x, \xi) \sim \sum \frac{i^{-|\alpha|}}{\alpha!}(\partial^\alpha_\xi \partial^\alpha_y \tilde{w}_1(x, y, \xi))|_{y=x}. \tag{A-8}$$

If the total symbol $r = \sigma_R$ of R has vanishing Taylor coefficients at a point $(y, \eta) = (\varphi'_\eta(x, \eta), \eta)$, then by examining (A-8) in decreasing order of homogeneity we find that each term of ρ must have vanishing

Taylor coefficients at $(x,\xi)=(x,\varphi'_x(x,\eta))$, since by what we have shown this would involve evaluating $r(z,\sigma)$ and its derivatives at $(\varphi'_\eta(x,\eta),\eta)$.

To prove the converse, choose $A_1\in I^{-m''}(X\times Y,\Gamma')$ and $B_1\in I^{-m'}(Y\times X,(\Gamma^{-1})')$ properly supported such that

$$K'\cap WF(BA_1-I)=\varnothing,\qquad K\cap WF(A_1B-I)=\varnothing,$$
$$K'\cap WF(B_1A-I)=\varnothing,\qquad K\cap WF(AB_1-I)=\varnothing.$$

Then a repetition of the arguments above shows that all the terms in the asymptotic expansion of the total symbol of B_1ARBA_1 has vanishing Taylor coefficients at a point $(y,\eta)=(\varphi'_\eta(x,\eta),\eta)$ if all the terms in the asymptotic expansion of the total symbol of ARB has vanishing Taylor coefficients at $(x,\xi)=(x,\varphi'_x(x,\eta))$. Since R and B_1ARBA_1 have the same total symbol in K' mod $\Psi^{-\infty}$, the same must hold for the total symbol of R. □

Let $\{e_k:k=1,\dots,n\}$ be a basis for $\mathbb{R}^n$, let (U,x) be local coordinates on a smooth manifold M of dimension n, and let

$$\left\{\frac{\partial}{\partial x_k}:k=1,\dots,n\right\}$$

be the induced local frame for the tangent bundle TM. Since the local frame fields commute, we can use standard multiindex notation to express the partial derivatives $\partial_x^\alpha f$ of $f\in C^\infty(U)$.

Lemma A.2. *Let M be a smooth manifold of dimension n, and for $j\ge 1$ let $p,q_j,g_j\in C^\infty(M)$. Let $\{\gamma_j\}_{j=1}^\infty$ be a sequence in M such that $\gamma_j\to\gamma$ as $j\to\infty$, and assume that $p(\gamma)=p(\gamma_j)=0$ for all j, and that $dp(\gamma)\neq 0$. Let (U,x) be local coordinates on M near γ, and suppose that there exists a smooth function $q\in C^\infty(M)$ such that*

$$\partial_x^\alpha q(\gamma)=\lim_{j\to\infty}\partial_x^\alpha q_j(\gamma_j)\quad \textit{for all } \alpha\in\mathbb{N}^n.$$

If q_j-pg_j vanishes of infinite order at γ_j for all j, then there exists a smooth function $g\in C^\infty(M)$ such that $q-pg$ vanishes of infinite order at γ. Furthermore,

$$\partial_x^\alpha g(\gamma)=\lim_{j\to\infty}\partial_x^\alpha g_j(\gamma_j)\quad \textit{for all } \alpha\in\mathbb{N}^n. \tag{A-9}$$

Proof. We have stated the result for a manifold, but since the result is purely local we may assume that $M\subset\mathbb{R}^n$ in the proof. It is also clear that we may assume that there exists an open neighborhood $\mathcal{U}$ of γ such that $\gamma_j\in\mathcal{U}$ for $j\ge 1$, and that $dp\neq 0$ in $\mathcal{U}$. By shrinking $\mathcal{U}$ if necessary, we can then find a unit vector $\nu\in\mathbb{R}^n$ such that $\partial_\nu p(w)=\langle\nu,dp(w)\rangle\neq 0$ for $w\in\mathcal{U}$. (We will identify a tangent vector $\nu\in\mathbb{R}^n$ at γ with $\partial_\nu\in T_\gamma\mathbb{R}^n$ through the usual vector space isomorphism.) Hence $\partial_\nu p(w)$ is invertible in $\mathcal{U}$, and we let $(\partial_\nu p(w))^{-1}\in C^\infty(\mathcal{U})$ denote its inverse. By an orthonormal change of coordinates we may even assume that $\partial_\nu p(w)=\partial_{e_1}p(w)$. In accordance with the notation used in the statement of the lemma, we shall write $\partial_{x_k}p(w)$ for the partial derivatives $\partial_{e_k}p(w)$ and denote by $(\partial_{x_1}p(w))^{-1}$ the inverse of $\partial_\nu p(w)=\partial_{x_1}p(w)$ in $\mathcal{U}$.

Now

$$0=\partial_{x_1}(q_j-pg_j)(\gamma_j)=\partial_{x_1}q_j(\gamma_j)-\partial_{x_1}p(\gamma_j)g_j(\gamma_j) \tag{A-10}$$

for all j since $p(\gamma_j)=0$. Since $\lim_j \partial_{x_1} q_j(\gamma_j)=\partial_{x_1} q(\gamma)$ by assumption, equation (A-10) yields

$$\lim_{j\to\infty} g_j(\gamma_j)=(\partial_{x_1} p(\gamma))^{-1}\partial_{x_1} q(\gamma)=a\in\mathbb{C}. \tag{A-11}$$

We claim that we can in the same way determine

$$\lim_{j\to\infty}(\partial_x^\alpha g_j)(\gamma_j)=a_{(\alpha)}\in\mathbb{C}\quad\text{for any } \alpha\in\mathbb{N}^n.$$

We start by determining $\lim_{j\to\infty}\partial g_j(\gamma_j)/\partial x_k=a_{(k)}$ for $1\le k\le n$. By the hypotheses of the lemma we have

$$\begin{aligned}0&=\partial_{x_k}\partial_{x_l}(q_j-pg_j)(\gamma_j)\\&=\partial_{x_k}\partial_{x_l}q_j(\gamma_j)-\partial_{x_k}\partial_{x_l}p(\gamma_j)g_j(\gamma_j)-\partial_{x_k}p(\gamma_j)\partial_{x_l}g_j(\gamma_j)-\partial_{x_l}p(\gamma_j)\partial_{x_k}g_j(\gamma_j)\end{aligned} \tag{A-12}$$

since $p(\gamma_j)=0$. For $k=l=1$ we obtain from (A-11) and (A-12)

$$\lim_{j\to\infty}\partial_{x_1}g_j(\gamma_j)=(\partial_{x_1}p(\gamma))^{-1}\big(\partial_{x_1}^2 q(\gamma)-\partial_{x_1}^2 p(\gamma)a\big)/2. \tag{A-13}$$

This allows us to solve for $\partial_{x_k}g_j(\gamma_j)$ in (A-12) by choosing $l=1$. If $b\in\mathbb{C}$ denotes the limit in (A-13) and $a\in\mathbb{C}$ is given by (A-11) we thus obtain

$$\lim_{j\to\infty}\partial_{x_k}g_j(\gamma_j)=(\partial_{x_1}p(\gamma))^{-1}\big(\partial_{x_1}\partial_{x_k}q(\gamma)-\partial_{x_1}\partial_{x_k}p(\gamma)a-\partial_{x_k}p(\gamma)b\big)\quad\text{for } 2\le k\le n.$$

Now assume that for some $m\ge 3$ we have in this way determined

$$\lim_{j\to\infty}\partial_{x_{k_1}}\dots\partial_{x_{k_{m-2}}}g_j(\gamma_j),\qquad\text{for } k_i\in[1,n],\text{ with } i\in[1,m-2].$$

To shorten notation, we will use the (nonstandard) multiindex notation introduced on page 465; to every $\alpha\in\mathbb{N}^n$ with $|\alpha|=m$ corresponds precisely one m-tuple $\beta=(k_1,\dots,k_m)$ of nondecreasing numbers $1\le k_1\le\dots\le k_m\le n$ such that ∂_x^β equals ∂_x^α. Throughout the rest of this proof we shall let β represent such an m-tuple, and we let

$$\hat\beta_i=(k_1,\dots,k_{i-1},k_{i+1},\dots,k_m).$$

As before we have

$$0=\partial_x^\beta(q_j-pg_j)(\gamma_j)=\partial_x^\beta q_j(\gamma_j)-\partial_x^\beta p(\gamma_j)g_j(\gamma_j)-\dots-\sum_{i=1}^m\partial_{x_{k_i}}p(\gamma_j)\partial_x^{\hat\beta_i}g_j(\gamma_j) \tag{A-14}$$

by assumption. If we choose $k_i=1$ for all $1\le i\le m$, the last sum is just $m\partial_{x_1}p(\gamma_j)\partial_{x_1}^{m-1}g_j(\gamma_j)$, and since the limit of all other terms on the right side are known by the induction hypothesis, we thus obtain the value of the limit of $\partial_{x_1}^{m-1}g_j(\gamma_j)$ from (A-14) by first multiplying by $m^{-1}(\partial_{x_1}p(\gamma_j))^{-1}$ and then letting $j\to\infty$. Denote this limit by $c\in\mathbb{C}$. If we choose $k_i\ne 1$ for precisely one $i\in[1,m]$, say $k_m=k$, then the last sum in (A-14) satisfies

$$\sum_{i=1}^m\partial_{x_{k_i}}p(\gamma_j)\partial_x^{\hat\beta_i}g_j(\gamma_j)=\partial_{x_k}p(\gamma_j)\partial_{x_1}^{m-1}g_j(\gamma_j)+(m-1)\partial_{x_1}p(\gamma_j)\partial_{x_1}^{m-2}\partial_{x_k}g_j(\gamma_j),$$

so by the same argument as before we can obtain the value of $\lim_{j\to\infty}\partial_{x_1}^{m-2}\partial_{x_k}g_j(\gamma_j)$ for $2\le k\le n$ by multiplying by $(m-1)^{-1}(\partial_{x_1}p(\gamma_j))^{-1}$ and using $\partial_{x_1}^{m-1}g_j(\gamma_j)\to c$ when taking the limit as $j\to\infty$ in (A-14). Continuing this way it is clear that we can successively determine

$$\lim_{j\to\infty}\partial_{x_{k_1}}\dots\partial_{x_{k_{m-1}}}g_j(\gamma_j)\quad\text{for any } 1\le k_1\le\dots\le k_{m-1}\le n,$$

which completely determines $\lim_{j\to\infty}\partial_x^\alpha g_j(\gamma_j)=a_{(\alpha)}$ for $\alpha\in\mathbb{N}^n$ with $|\alpha|=m-1$, proving the claim.

By Borel's theorem there exists a smooth function $g\in C^\infty(M)$ such that

$$\partial_x^\alpha g(\gamma)=a_{(\alpha)}=\lim_{j\to\infty}\partial_x^\alpha g_j(\gamma_j)\quad\text{for all } \alpha\in\mathbb{N}^n.$$

Since $q-pg$ vanishes of infinite order at γ by construction, this completes the proof. □

The lemma will be used to prove the following result for homogeneous smooth functions on the cotangent bundle.

Proposition A.3. *For $j\ge 1$ let $p, q_j, g_j\in C^\infty(T^*(\mathbb{R}^n)\smallsetminus 0)$, where p and q_j are homogeneous of degree m and the g_j are homogeneous of degree* 0. *Let $\{\gamma_j\}_{j=1}^\infty$ be a sequence in $T^*(\mathbb{R}^n)\smallsetminus 0$ such that $\gamma_j\to\gamma$ as $j\to\infty$, and assume that $p(\gamma)=p(\gamma_j)=0$ for all j, and that $dp(\gamma)\neq 0$. If there exists a smooth function $q\in C^\infty(T^*(\mathbb{R})^n\smallsetminus 0)$, homogeneous of degree m, such that*

$$\partial_x^\alpha\partial_\xi^\beta q(\gamma)=\lim_{j\to\infty}\partial_x^\alpha\partial_\xi^\beta q_j(\gamma_j)\quad\textit{for all } (\alpha,\beta)\in\mathbb{N}^n\times\mathbb{N}^n,$$

and if q_j-pg_j vanishes of infinite order at γ_j for all j, then there exists a $g\in C^\infty(T^*(\mathbb{R}^n)\smallsetminus 0)$, homogeneous of degree 0, *such that $q-pg$ vanishes of infinite order at γ. Furthermore,*

$$\partial_x^\alpha\partial_\xi^\beta g(\gamma)=\lim_{j\to\infty}\partial_x^\alpha\partial_\xi^\beta g_j(\gamma_j)\quad\textit{for all } (\alpha,\beta)\in\mathbb{N}^n\times\mathbb{N}^n.\tag{A-15}$$

Proof. Let $\pi: T^*(\mathbb{R}^n)\smallsetminus 0\to S^*(\mathbb{R}^n)$ be the projection. Since $dp(\gamma)\neq 0$ it follows from homogeneity that $dp(\pi(\gamma))\neq 0$. By using the homogeneity of q, q_j and g_j we may even assume that γ and γ_j belong to $S^*(\mathbb{R}^n)$ for $j\ge 1$ to begin with.

Now, the radial vector field $\xi\partial_\xi$ applied k times to $a\in C^\infty(T^*(\mathbb{R}^n)\smallsetminus 0)$ equals $l^k a$ if a is homogeneous of degree l. For any point $w\in S^*(\mathbb{R}^n)$ with $w=(w_x,w_\xi)$ in local coordinates on $T^*(\mathbb{R}^n)$ it is easy to see that

$$T_wS^*(\mathbb{R}^n)=\{(u,v)\in\mathbb{R}^n\times\mathbb{R}^n:\langle w_\xi,v\rangle=0\}.$$

Therefore a basis for $T_wS^*(\mathbb{R}^n)$ together with the radial vector field $(\xi\partial_\xi)_w$ at w constitutes a basis for $T_wT^*(\mathbb{R}^n)$. This implies that if we can find a homogeneous function g such that $q-pg$ vanishes of infinite order in the directions $T_\gamma S^*(\mathbb{R}^n)$, then $q-pg$ vanishes of infinite order at γ, for the derivatives involving the radial direction are determined by lower order derivatives in the directions $T_\gamma S^*(\mathbb{R}^n)$.

By the hypotheses of the proposition together with an application of Lemma A.2, we find that there exists a function $\tilde g\in C^\infty(T^*(\mathbb{R}^n))$, not necessarily homogeneous, such that $q-p\tilde g$ vanishes of infinite order at γ and (A-15) holds for $\tilde g$. The function $g(x,\xi)=\tilde g(x,\xi/|\xi|)$ coincides with $\tilde g$ on $S^*(\mathbb{R}^n)$. In particular, all derivatives of g and $\tilde g$ in the directions $T_\gamma S^*(\mathbb{R}^n)$ are equal at γ. Thus, by the arguments above we conclude that $q - pg$ vanishes of infinite order at γ. Since g and g_j are homogeneous of degree 0, the same arguments also imply that (A-15) holds for g, which completes the proof. $\square$

Acknowledgement

I am grateful to Professor Nils Dencker at Lund University for suggesting the problem that led to the results presented here, and also for many helpful discussions on the subject.

References

[Dencker 2006] N. Dencker, "The resolution of the Nirenberg–Treves conjecture", *Ann. of Math.* (2) **163**:2 (2006), 405–444. MR 2006i:35386 Zbl 1104.35080

[Grigis and Sjöstrand 1994] A. Grigis and J. Sjöstrand, *Microlocal analysis for differential operators: An introduction*, London Mathematical Society Lecture Note Series **196**, Cambridge University Press, 1994. MR 95d:35009 Zbl 0804.35001

[Hörmander 1960a] L. Hörmander, "Differential equations without solutions", *Math. Ann.* **140** (1960), 169–173. MR 26 #5279 Zbl 0093.28903

[Hörmander 1960b] L. Hörmander, "Differential operators of principal type", *Math. Ann.* **140** (1960), 124–146. MR 24 #A434 Zbl 0090.08101

[Hörmander 1963] L. Hörmander, *Linear partial differential operators*, Grundlehren der mathematischen Wissenschaften **116**, Springer, Berlin, 1963. Zbl 0108.09301

[Hörmander 1966] L. Hörmander, "Pseudo-differential operators and non-elliptic boundary problems", *Ann. of Math.* (2) **83** (1966), 129–209. MR 38 #1387 Zbl 0132.07402

[Hörmander 1981] L. Hörmander, "Pseudo-differential operators of principal type", pp. 69–96 in *Singularities in boundary value problems: Proc. NATO Adv. Study Inst., 1980* (Maratea, Italy: September 22–October 3, 1980), NATO Adv. Study Inst., Ser. C: Math. Phys. Sci. **65**, Reidel, Dordrecht-Boston, MA, 1981. Zbl 0459.35096

[Hörmander 1983a] L. Hörmander, *The analysis of linear partial differential operators, I: Distribution theory and Fourier analysis*, Grundlehren der Mathematischen Wissenschaften **256**, Springer, Berlin, 1983. MR 85g:35002a

[Hörmander 1983b] L. Hörmander, *The analysis of linear partial differential operators, II: Differential operators with constant coefficients*, Grundlehren der Mathematischen Wissenschaften **257**, Springer, Berlin, 1983. MR 85g:35002b

[Hörmander 1985a] L. Hörmander, *The analysis of linear partial differential operators, III: Pseudodifferential operators*, Grundlehren der Mathematischen Wissenschaften **274**, Springer, Berlin, 1985. MR 87d:35002a Zbl 0601.35001

[Hörmander 1985b] L. Hörmander, *The analysis of linear partial differential operators, IV: Fourier integral operators*, Grundlehren der Mathematischen Wissenschaften **275**, Springer, Berlin, 1985. MR 87d:35002b

[Lewy 1957] H. Lewy, "An example of a smooth linear partial differential equation without solution", *Ann. of Math.* (2) **66** (1957), 155–158. MR 19,551d Zbl 0078.08104

[Moyer 1978] R. D. Moyer, "Local solvability in two dimensions: Necessary conditions for the principal-type case", mimeographed manuscript, University of Kansas, 1978.

[Nirenberg and Treves 1970a] L. Nirenberg and F. Treves, "On local solvability of linear partial differential equations, I: Necessary conditions", *Communications on Pure and Applied Mathematics* **23**:1 (1970), 1–38. Zbl 0191.39103

[Nirenberg and Treves 1970b] L. Nirenberg and F. Treves, "On local solvability of linear partial differential equations, II: Sufficient conditions", *Communications on Pure and Applied Mathematics* **23**:3 (1970), 459–509.

[Nirenberg and Treves 1971] L. Nirenberg and F. Treves, "Corrections to On local solvability of linear partial differential equations", *Communications on Pure and Applied Mathematics* **24**:2 (1971), 279–288.

Received 14 Jul 2010. Revised 10 Dec 2010. Accepted 5 Feb 2011.

JENS WITTSTEN: jens.wittsten@math.lu.se
Center for Mathematical Sciences, Lund University, Box 118, SE-221 00 Lund, Sweden

ANALYSIS AND PDE
Vol. 5, No. 3, 2012
dx.doi.org/10.2140/apde.2012.5.475

BLOW-UP SOLUTIONS ON A SPHERE FOR THE 3D QUINTIC NLS IN THE ENERGY SPACE

JUSTIN HOLMER AND SVETLANA ROUDENKO

We prove that if $u(t)$ is a log-log blow-up solution, of the type studied by Merle and Raphaël, to the L^2 critical focusing NLS equation $i\partial_t u+\Delta u+|u|^{4/d}u=0$ with initial data $u_0\in H^1(\mathbb{R}^d)$ in the cases $d=1,2$, then $u(t)$ remains bounded in H^1 away from the blow-up point. This is obtained without assuming that the initial data u_0 has any regularity beyond $H^1(\mathbb{R}^d)$. As an application of the $d=1$ result, we construct an open subset of initial data in the radial energy space $H^1_{\text{rad}}(\mathbb{R}^3)$ with corresponding solutions that blow up on a sphere at positive radius for the 3D quintic ($\dot H^1$-critical) focusing NLS equation $i\partial_t u+\Delta u+|u|^4u=0$. This improves the results of Raphaël and Szeftel [2009], where an open subset in $H^3_{\text{rad}}(\mathbb{R}^3)$ is obtained. The method of proof can be summarized as follows: On the whole space, high frequencies above the blow-up scale are controlled by the bilinear Strichartz estimates. On the other hand, outside the blow-up core, low frequencies are controlled by finite speed of propagation.

1. Introduction

Consider the L^2 critical focusing nonlinear Schrödinger equation (NLS)

$$i\partial_t u+\Delta u+|u|^{4/d}u=0, \tag{1-1}$$

where $u=u(x,t)\in\mathbb{C}$ and $x\in\mathbb{R}^d$, in dimensions $d=1$ and $d=2$. It is locally well-posed in $H^1(\mathbb{R}^d)$ and its solutions satisfy conservation of mass $M(u)$, momentum $P(u)$, and energy $E(u)$:

$$M(u)=\|u\|^2_{L^2},\quad P(u)=\operatorname{Im}\int\bar u\,\nabla u\,dx,\quad E(u)=\tfrac12\|\nabla u\|^2_{L^2}-\frac{1}{4/d+2}\|u\|^{4/d+2}_{L^{4/d+2}}; \tag{1-2}$$

see [Tao 2006, Chapter 3] and [Cazenave 2003, Chapter 4] for exposition and references. The Galilean identity (see [Tao 2006, Exercise 2.5]) transforms any solution to one with zero momentum, so there is no loss in considering only solutions $u(t)$ such that $P(u)=0$.

The unique (up to translation) minimal mass H^1 solution of

$$-Q+\Delta Q+|Q|^{4/d}Q=0,\quad \text{with } Q=Q(x), \tag{1-3}$$

is called the *ground state*. It is smooth, radial, real-valued and positive, and exponentially decaying; see [Tao 2006, Appendix B]. In the case $d=1$, we have explicitly

$$Q(x)=3^{1/4}\operatorname{sech}^{1/2}(x). \tag{1-4}$$

MSC2000: 35Q55.
Keywords: blow-up, nonlinear Schrödinger equation.

Weinstein [1982] proved that solutions to (1-1) with $M(u) < M(Q)$ necessarily satisfy $E(u) > 0$ and remain bounded in H^1 globally in time (that is, they do not blow up in finite time).

Building upon the earlier heuristic and numerical result of Landman, Papanicolaou, Sulem and Sulem [Landman et al. 1988] and the first analytical result of Perelman [2001], Merle and Raphaël in a series of papers (see [Merle and Raphaël 2005] and references therein) studied H^1 solutions to (1-1) such that

$$E(u) < 0, \quad P(u) = 0, \quad M(Q) < M(u) < M(Q) + \alpha^* \tag{1-5}$$

for some small absolute constant $\alpha^* > 0$. They showed that any such solution blows up in finite time at the *log-log rate* — more precisely, they proved that there exists a *threshold time* $T_0(u_0) > 0$ and a *blow-up time* $T(u_0) > T_0(u_0)$ such that

$$\|\nabla u(t)\|_{L^2_x} \sim \left(\frac{\log|\log(T-t)|}{T-t}\right)^{1/2} \quad \text{for } T_0 \le t < T, \tag{1-6}$$

where the implicit constant in (1-6) is universal. Also, with scale parameter $\lambda(t) = \|\nabla Q\|_{L^2} / \|\nabla u(t)\|_{L^2}$, there exist parameters of position $x(t) \in \mathbb{R}^d$ and phase $\gamma(t) \in \mathbb{R}$ such that if we define the *blow-up core*

$$u_{\text{core}}(x,t) = \frac{e^{i\gamma(t)}}{\lambda(t)^{d/2}} Q\left(\frac{x - x(t)}{\lambda(t)}\right), \tag{1-7}$$

and *remainder* $\tilde{u} = u - u_{\text{core}}$, then $\|\tilde{u}\|_{L^2} \le \alpha_*$ and

$$\|\nabla \tilde{u}(t)\|_{L^2} \lesssim \left(\frac{1}{|\log(T-t)|^C (T-t)}\right)^{1/2} \tag{1-8}$$

for some $C > 1$. There is, in addition, a well-defined *blow-up point* $x_0 := \lim_{t \nearrow T} x(t)$. We refer to the region of space $\{x \in \mathbb{R}^d \mid |x - x_0| > R\}$, for any fixed $R > 0$, as the *external* region. While the Merle–Raphaël analysis accurately describes the activity of the solution in the blow-up core, the only information it directly yields about the external region is the bound (1-8).

However, it is a consequence of the analysis in [Raphaël 2006] that in the case $d = 1$, H^1 solutions in the class (1-5) have bounded $H^{1/2}$ norm in the external region all the way up to the blow-up time T. In [Holmer and Roudenko 2011], we extended this result to the case $d = 2$. Raphaël and Szeftel [2009] established for $d = 1$ that solutions with regularity H^N for $N \ge 3$ satisfying (1-5) remain bounded in the $H^{(N-1)/2}$ norm in the external region, and Zwiers [2011] extended this result to the case $d = 2$. These results leave open the possibility that there is a loss of roughly half the regularity in passing from the initial data to the solution in the external region at blow-up time. The first main result of this paper is that such a loss *does not occur*. Specifically, we prove that H^1 solutions in the class (1-5) remain bounded in the H^1 norm in the external region all the way up to the blow-up time, resolving an open problem posed in [Raphaël and Szeftel 2009, Comment 1 on page 976].

Theorem 1.1. *Consider dimension $d = 1$ or $d = 2$. Suppose that $u(t)$ is an H^1 solution to* (1-1) *in the Merle–Raphaël class* (1-5) (*no higher regularity is assumed*). *Let $T > 0$ be the blow-up time and $x_0 \in \mathbb{R}^d$*

the blow-up point. Then for any $R > 0$,

$$\|\nabla u(t)\|_{L^\infty_{[0,T]} L^2_{|x-x_0|\geq R}} \leq C, \quad \textit{where } C \textit{ depends on } R,\ T_0(u_0), \textit{ and } \|\nabla u_0\|_{L^2}.^1$$

We remark that H^1, the *energy space*, is a natural space in which to study the equation (1-1) since the conservation laws (1-2) are defined and Lyapunov–Hamiltonian type methods, such as those used by Merle and Raphaël in their blow-up theory, naturally yield coercivity on H^1 quantities.

The retention of regularity in the external region has applications to the construction of new blow-up solutions, with special geometry, for L^2 supercritical NLS equations. Using their partial regularity methods, Raphaël [2006] and Raphaël and Szeftel [2009] constructed spherically symmetric finite-time blow-up solutions to the quintic NLS

$$i\partial_t u + \Delta u + |u|^4 u = 0 \tag{1-9}$$

in dimension $d \geq 2$ that contract toward a sphere $|x| = r_0 \sim 1$ following the one-dimensional quintic blow-up dynamics (1-6)(1-7) in the radial variable near $r = r_0$. Specifically, they showed there exists an open subset of initial data in some radial function class with corresponding solutions adhering to the blow-up dynamics described above. In [Raphaël 2006], for $d = 2$, an open subset of initial data in the radial energy space $H^1_{\mathrm{rad}}(\mathbb{R}^2)$ was obtained. For $d = 3$, in which case (1-9) is $\dot{H}^1$ critical, Raphaël and Szeftel [2009] obtained an open subset of initial data in a comparably "thin" subset $H^3_{\mathrm{rad}}(\mathbb{R}^3)$ of the radial energy space $H^1_{\mathrm{rad}}(\mathbb{R}^3)$.

As an application of the techniques used to prove Theorem 1.1, we prove, for $d = 3$, the existence of an open subset of initial data in the full radial energy space $H^1_{\mathrm{rad}}(\mathbb{R}^3)$. For the statement, take Q to be the solution to (1-3) in the case $d = 1$, explicitly given by (1-4). The following theorem follows the motif of the $d = 3$ case of [Raphaël and Szeftel 2009, Theorem 1] except that $\mathcal{P}$, the initial data, is an open subset of $H^1_{\mathrm{rad}}(\mathbb{R}^3)$ rather than $H^3_{\mathrm{rad}}(\mathbb{R}^3)$.

Theorem 1.2. *There exists an open subset $\mathcal{P} \subset H^1_{\mathrm{rad}}(\mathbb{R}^3)$ such that the following holds true. Let $u_0 \in \mathcal{P}$ and let $u(t)$ denote the corresponding solution to* (1-9) *in the case $d = 3$. Then there exist a blow-up time $0 < T < +\infty$ and parameters of scale $\lambda(t) > 0$, radial position $r(t) > 0$, and phase $\gamma(t) \in \mathbb{R}$ such that if we take*

$$u_{\mathrm{core}}(t, r) := \frac{1}{\lambda(t)^{1/2}} Q\left(\frac{r - r(t)}{\lambda(t)}\right) e^{i\gamma(t)}$$

and the remainder $\tilde{u}(t) := u(t) - u_{\mathrm{core}}(t)$, then the following hold:

(1) *The remainder converges in L^2: $\tilde{u}(t) \to u^*$ in $L^2(\mathbb{R}^3)$ as $t \nearrow T$.*

(2) *The position of the singular sphere converges: $r(t) \to r_0 > 0$ as $t \nearrow T$.*

[1] We did not see in the Merle–Raphaël papers the threshold time $T_0(u_0)$ or the blow-up time $T(u_0)$ estimated quantitatively in terms of properties of the initial data ($\|\nabla u_0\|_{L^2}$, $E(u_0)$, etc.). If such dependence could be quantified, then the constant C in Theorem 1.1 could be quantified.

(3) *The solution contracts toward the sphere at the log-log rate*:

$$\lambda(t)\Big(\frac{\log|\log(T-t)|}{T-t}\Big)^{1/2} \to \frac{\sqrt{2\pi}}{\|Q\|_{L^2}} \quad \textit{as } t \nearrow T.$$

(4) *The solution remains H^1-small away from the singular sphere*: *For each* $R > 0$,

$$\|u(t)\|_{H^1_{|r-r(T)|\geq R}(\mathbb{R}^3)} \leq \epsilon.$$

The 3D quintic NLS equation (1-9) is energy-critical, and the global well-posedness and scattering problem is one of several critical regularity problems that has received a lot of attention in the last decade [Bourgain 1999; Colliander et al. 2008; Kenig and Merle 2006]. The global well-posedness for small data in $\dot{H}^1$ is classical and follows from the Strichartz estimates. Our Theorem 1.2 takes a large, but special "prefabricated" approximate blow-up solution, and installs it near radius $r = 1$ on top of a small global H^1 background. The main difficulty, of course, is showing that the two different components — the blow-up portion on the one hand, and the evolution of the small $\dot{H}^1$ background on the other — have limited interaction and can effectively evolve separately. Thus, it is not surprising that the techniques to prove Theorem 1.1 are relevant to this analysis.

We now outline the method used to prove Theorem 1.1. We start with a given blow-up solution $u(t)$ in the Merle–Raphaël class, and by scaling and shifting this solution, it suffices to assume that the blow-up point is $x_0 = 0$ and the blow-up time is $T = 1$, and moreover, (1-6) holds over times $0 \leq t < 1$. Since (1-1) is L^2 critical, the size of the L^2 norm is highly relevant. By mass conservation, we know that $\|P_N u(t)\|_{L^2_x} \lesssim 1$ for all N and all $0 \leq t < 1$, where P_N denotes the Littlewood–Paley frequency projection. However, (1-6) shows that for $N \gg (1-t)^{-(1+\delta)/2}$, we have $\|P_N u(t)\|_{L^2_x} \lesssim N^{-1}(1-t)^{-(1+\delta)/2}$, which is a better estimate for these large frequencies N. In Section 3, we show that this smallness of high frequencies reinforces itself and ultimately proves that for $N \gg (1-t)^{-(1+\delta)/2}$, the solution is H^1 bounded. This is achieved using dispersive estimates typically employed in local well-posedness arguments — the Strichartz and Bourgain's bilinear Strichartz estimates — after the equation has been restricted to high frequencies. We note that this improvement of regularity at high frequencies is proved *globally in space*.

For the Schrödinger equation, frequencies of size N propagate at speed N, and thus, travel a distance $O(1)$ over a time N^{-1}. Therefore, at time $t < 1$, a component of the solution in the blow-up core at frequency N will effectively only make it out of the blow-up core and into the external region before the blow-up time, provided $N \gtrsim (1-t)^{-1}$. Thus, we expect that the blow-up action, which is taking place at frequency $\sim (1-t)^{-1/2}\log|\log(1-t)| \ll (1-t)^{-1}$, will not be able to exit the blow-up core before blow-up time. This is the philosophy behind the analysis in Section 4. Recall that in Section 3, we have controlled the solution at frequencies above $(1-t)^{-(1+\delta)/2}$. In Section 4, we apply a spatial localization to the external region, and then look to control the remaining low frequencies, i.e., those frequencies below $(1-t)^{-(1+\delta)/2}$. We examine the equation solved by $P_{\leq(1-t)^{-3/4}}\psi u(t)$, where ψ is a spatial restriction to the external region. In estimating the inhomogeneous terms, we can make use of the frequency restriction to exchange α-spatial derivatives for a time factor $(1-t)^{-3\alpha/4}$. This enables us to prove a low-frequency recurrence: The H^s size of the solution in the external region is bounded by the $H^{s-1/8}$ size of the solution in a slightly larger external region. Iteration gives the H^1 boundedness.

The structure of the paper is as follows. Preliminaries on the Strichartz and bilinear Strichartz estimates appear in Section 2. The proof of Theorem 1.1 is carried out in Sections Section 3 and 4. The proof of Theorem 1.2 is carried out in Section 5.

2. Standard estimates

All of the estimates outlined in this section are now classical and well known. Let P_N, $P_{\le N}$, and $P_{\ge N}$ denote the Littlewood–Paley frequency projections.

We say that (q, p) is an *admissible* pair if $2 \le p \le \infty$ and

$$\frac{2}{q} + \frac{d}{p} = \frac{d}{2},$$

excluding the case $d = 2$, $q = 2$, and $p = \infty$.

Lemma 2.1 (Strichartz estimate). *If (q, p) is an admissible pair, then*

$$\|e^{it\Delta}\phi\|_{L_t^q L_x^p} \lesssim \|\phi\|_{L_x^2}.$$

Proof. See [Strichartz 1977] and [Keel and Tao 1998]. □

Lemma 2.2 (Bourgain bilinear Strichartz estimate). *Suppose that $N_1 \ll N_2$. Then*

$$\|P_{N_1} e^{it\Delta}\phi_1 P_{N_2} e^{it\Delta}\phi_2\|_{L_t^2 L_x^2} \lesssim \left(\frac{N_1^{d-1}}{N_2}\right)^{1/2} \|\phi_1\|_{L_x^2} \|\phi_2\|_{L_x^2}, \tag{2-1}$$

$$\|P_{N_1} e^{it\Delta}\phi_1 \overline{P_{N_2} e^{it\Delta}\phi_2}\|_{L_t^2 L_x^2} \lesssim \left(\frac{N_1^{d-1}}{N_2}\right)^{1/2} \|\phi_1\|_{L_x^2} \|\phi_2\|_{L_x^2}. \tag{2-2}$$

Proof. For the 2D estimate (2-1), see [Bourgain 1998, Lemma 111]; the 1D case appears in [Colliander et al. 2001, Lemma 7.1]; another nice proof is given in [Koch and Tataru 2007, Proposition 3.5], the other dimensions are analogous. We review the 1D proof to show that the second estimate (2-2) holds as well.

Denote $u = e^{it\Delta}(P_{N_1}\phi_1)$ and $v = e^{\pm it\Delta}(P_{N_2}\phi_2)$. Then in the 1D case,

$$\widehat{uv}(\xi, \tau) = \int_{\xi_1+\xi_2=\xi} \widehat{P_{N_1}\phi_1}(\xi_1)\widehat{P_{N_2}\phi_2}(\xi_2)\delta(\tau - (\xi_1^2 \pm \xi_2^2))\, d\xi_1 \tag{2-3}$$

$$= \frac{1}{|g'_{\xi_1}(\xi_1, \xi_2)|} \widehat{P_{N_1}\phi_1}\,\widehat{P_{N_2}\phi_2}|_{(\xi_1,\xi_2)}, \tag{2-4}$$

where $g(\xi_1, \xi_2) = \tau - (\xi_1^2 \pm \xi_2^2)$, thus, $|g'_{\xi_1}| = 2|\xi_1 \pm \xi_2|$. To estimate the $L^2_{\xi,\tau}$ norm of uv, we square the expression above and integrate in τ and ξ. Changing variables (τ, ξ) to (ξ_1, ξ_2) with $\tau = \xi_1^2 \pm \xi_2^2$ and $\xi = \xi_1 + \xi_2$, we obtain $d\tau d\xi = J\, d\xi_1 d\xi_2$ with the Jacobian $J = 2|\xi_1 \pm \xi_2|$, which is of size N_2 (note that $\pm$ does not matter here, since $N_2 \gg N_1$). Bringing the square inside, we get

$$\|uv\|_{L_x^2}^2 \lesssim \int_{|\xi_1|\sim N_1, |\xi_2|\sim N_2} |\widehat{\phi_1}(\xi_1)|^2 |\widehat{\phi_2}(\xi_2)|^2 \frac{d\xi_1\, d\xi_2}{|\xi_1 \pm \xi_2|} \lesssim \frac{1}{N_2} \|\phi_1\|_{L_x^2}^2 \|\phi_2\|_{L_x^2}^2.$$ □

Now we introduce the Fourier restriction norms. For $\tilde{u} \in \mathcal{S}(\mathbb{R}^{1+d})$,

$$\|\tilde{u}\|_{X_{s,b}} = \big\| \langle D_t \rangle^b \langle D_x \rangle^s e^{-it\Delta} \tilde{u}(\,\cdot\,, t) \big\|_{L_t^2 L_x^2} = \Big(\int_\xi \int_\tau |\widehat{\tilde{u}}(\xi, \tau)|^2 \langle \xi \rangle^{2s} \langle \tau + |\xi|^2 \rangle^{2b} \, d\xi \, d\tau \Big)^{1/2}.$$

If $I \subset \mathbb{R}$ is an open subinterval and $u \in \mathcal{D}'(I \times \mathbb{R}^d)$, define

$$\|u\|_{X_{s,b}(I)} = \inf_{\tilde{u}} \|\tilde{u}\|_{X_{s,b}},$$

where the infimum is taken over all distributions $\tilde{u} \in \mathcal{S}'(\mathbb{R}^{1+d})$ such that $\tilde{u}|_I = u$.

Lemma 2.3. *If θ is a function such that* $\operatorname{supp} \theta \subset I$, *then for all* $0 < b < 1$,

$$\|\theta u\|_{X_{s,b}} \lesssim (\|\theta\|_{L^\infty} + \|D_t^{\max(1/2,b)} \theta\|_{L^2}) \|u\|_{X_{s,b}(I)}. \tag{2-5}$$

If $0 \le b < \frac{1}{2}$ and χ_I is the (*sharp*) *characteristic function of the time interval I, then*

$$\|\chi_I u\|_{X_{s,b}} \sim \|u\|_{X_{s,b}(I)}. \tag{2-6}$$

Proof. It suffices to take $s = 0$. The inequality (2-5) follows from the fractional Leibniz rule. To address (2-6), we note that Jerison and Kenig [1995] prove that $\|\chi_{(0,+\infty)} f\|_{H_t^b} \lesssim \|f\|_{H_t^b}$ for $-\frac{1}{2} < b < \frac{1}{2}$. Consequently, $\|\chi_I f\|_{H_t^b} \lesssim \|f\|_{H_t^b}$ for any time interval I. Let $\tilde{u}$ be an extension of u (meaning $\tilde{u}|_I = u$) so that $\|\tilde{u}\|_{X_{0,b}} \le 2\|u\|_{X_{0,b}(I)}$. Then

$$\begin{aligned} \|\chi_I u\|_{X_{0,b}} &= \|\langle D_t \rangle^b e^{-it\Delta} \chi_I \tilde{u}\|_{L_t^2 L_x^2} \\ &= \big\| \|\chi_I e^{-it\Delta} \tilde{u}\|_{H_t^b} \big\|_{L_x^2} \lesssim \big\| \|e^{-it\Delta} \tilde{u}\|_{H_t^b} \big\|_{L_x^2} \\ &= \|\tilde{u}\|_{X_{0,b}} \le 2\|u\|_{X_{0,b}(I)}. \end{aligned}$$

On the other hand, the inequality $\|u\|_{X_{0,b}(I)} \lesssim \|\chi_I u\|_{X_{0,b}}$ is trivial, since $\chi_I u$ is an extension of $u|_I$. □

Lemma 2.4. *If $i\partial_t u + \Delta u = f$ on a time interval $I = (a_1, a_2)$ with $|I| = O(1)$, then*

(1) *For $\frac{1}{2} < b \le 1$, taking $I' = (a_1 - \omega, a_2 + \omega)$, $0 < \omega \le 1$, we have*

$$\|u(t) - e^{i(t-a_1)\Delta} u(a_1)\|_{X_{0,b}(I)} \lesssim \omega^{1/2-b} \|f\|_{X_{0,b-1}(I')}. \tag{2-7}$$

(2) *For $0 \le b < \frac{1}{2}$,*

$$\|u(t) - e^{i(t-a_1)\Delta} u(a_1)\|_{X_{0,b}(I)} \lesssim \|f\|_{L_I^1 L_x^2}. \tag{2-8}$$

Moreover, for all b,

$$\|e^{i(t-a_1)\Delta} \phi\|_{X_{0,b}(I)} \lesssim \|\phi\|_{L_x^2}.$$

Proof. Without loss, we take $a_1 = 0$. First we consider (2-7). Since, for $t \in I$,

$$e^{-it\Delta} u(\,\cdot\,, t) = u(0) - i\theta(t) \int_0^t e^{-it'\Delta} \theta(t') f(\,\cdot\,, t') \, dt',$$

where θ is a cutoff function such that $\theta(t)=1$ on I and $\operatorname{supp}\theta\subset I'$, the estimate reduces to the space-independent estimate

$$\Big\|\theta(t)\int_0^t h(t')\,dt'\Big\|_{H_t^b}\lesssim\|h\|_{H_t^{b-1}}\quad\text{for }\tfrac12<b\le 1 \tag{2-9}$$

by (2-5). Now we prove estimate (2-9). Divide $h=P_{\le 1}h+P_{\ge 1}h$ and use that

$$\int_0^t P_{\ge 1}h(t')=\frac12\int(\operatorname{sgn}(t-t')+\operatorname{sgn}(t'))P_{\ge 1}h(t')\,dt'$$

to obtain the decomposition

$$\theta(t)\int_0^t h(t')\,dt'=H_1(t)+H_2(t)+H_3(t),$$

where

$$\begin{aligned}H_1(t)&=\theta(t)\int_0^t P_{\le 1}h(t')\,dt',\\ H_2(t)&=\tfrac12\theta(t)[\operatorname{sgn}*P_{\ge 1}h](t)\,dt',\\ H_3(t)&=\tfrac12\theta(t)\int_{-\infty}^{+\infty}\operatorname{sgn}(t')P_{\ge 1}h(t')\,dt'.\end{aligned}$$

We begin by addressing term H_1. By Sobolev embedding (recall $\frac12<b\le 1$) and the $L^p\to L^p$ boundedness of the Hilbert transform for $1<p<\infty$,

$$\|H_1\|_{H_t^b}\lesssim\|H_1\|_{L_t^2}+\|\partial_t H_1\|_{L_t^{2/(3-2b)}}.$$

Using that $|I|=O(1)$ and $\|P_{\le 1}h\|_{L_t^\infty}\lesssim\|h\|_{H_t^{b-1}}$, we thus conclude

$$\|H_1\|_{H_t^b}\lesssim\big(\|\theta\|_{L_t^2}+\|\theta\|_{L_t^{2/(3-2b)}}+\|\theta'\|_{L_t^{2/3-2b}}\big)\|h\|_{H_t^{b-1}}.$$

Next we address the term H_2. By the fractional Leibniz rule,

$$\|H_2\|_{H_t^b}\lesssim\|\langle D_t\rangle^b\theta\|_{L_t^2}\|\operatorname{sgn}*P_{\ge 1}h\|_{L_t^\infty}+\|\theta\|_{L_t^\infty}\|\langle D_t\rangle^b(\operatorname{sgn}*P_{\ge 1}h)\|_{L_t^2}.$$

However,

$$\|\operatorname{sgn}*P_{\ge 1}h\|_{L_t^\infty}\lesssim\|\langle\tau\rangle^{-1}\hat h(\tau)\|_{L_\tau^1}\lesssim\|h\|_{H_t^{b-1}}.$$

On the other hand,

$$\|\langle D_t\rangle^b\operatorname{sgn}*P_{\ge 1}h\|_{L_t^2}\lesssim\|\langle\tau\rangle^b\langle\tau\rangle^{-1}\hat h(\tau)\|_{L_\tau^2}\lesssim\|h\|_{H_t^{b-1}}.$$

Consequently,

$$\|H_2\|_{H_t^b}\lesssim(\|\langle D_t\rangle^b\theta\|_{L_t^2}+\|\theta\|_{L_t^\infty})\|h\|_{H_t^{b-1}}.$$

For term H_3, we have

$$\|H_3\|_{H_t^b}\lesssim\|\theta\|_{H_t^b}\Big\|\int_{-\infty}^{+\infty}\operatorname{sgn}(t')P_{\ge 1}h(t')\,dt'\Big\|_{L_t^\infty}.$$

However, the second term is handled via Parseval's identity

$$\int_{t'} \operatorname{sgn}(t') P_{\geq 1} h(t')\, dt' = \int_{|\tau|\geq 1} \tau^{-1} \hat{h}(\tau)\, d\tau,$$

from which the appropriate bounds follow again by Cauchy–Schwarz. Collecting our estimates for H_1, H_2, and H_3, we have

$$\left\| \theta(t) \int_0^t h(t')\, dt' \right\|_{H_t^b} \lesssim C_\theta \|h\|_{H_t^{b-1}},$$

where

$$C_\theta = \|\theta\|_{L_t^2} + \|\theta'\|_{L_t^{2/(3-2b)}} + \|\langle D_t\rangle^b \theta\|_{L_t^2} + \|\theta\|_{L_t^{2/(3-2b)}} + \|\theta\|_{L_t^\infty} \lesssim \omega^{1/2-b}.$$

This completes the proof of (2-7). Next, we prove (2-8). We have

$$e^{-it\Delta} u(\,\cdot\,, t) = u(0) - i \int_0^t e^{-it'\Delta} f(\,\cdot\,, t')\, dt',$$

and thus, (2-8) reduces, by (2-6), to

$$\left\| \chi_I \int_0^t g(t')\, dt' \right\|_{H_t^b} \lesssim \|g\|_{L_I^1}, \quad \text{for } 0 \leq b < \tfrac{1}{2}. \tag{2-10}$$

To prove (2-10), note that

$$\chi_I(t) \int_0^t g(t')\, dt' = \chi_I(t)[\chi_I * (g\chi_I)](t).$$

Hence,

$$\|\chi_I \int_0^t g(t')\, dt'\|_{H_t^b} \lesssim \|\langle D\rangle^b \chi_I\|_{L_t^2} \|g\|_{L_I^1}.$$

The Fourier transform of χ_I is smooth and decays like $|\tau|^{-1}$ as $|\tau| \to \infty$, and hence, $\|\langle D\rangle^b \chi_I\|_{L_t^2} < \infty$ for $0 \leq b < \frac{1}{2}$. $\square$

Lemma 2.5 (Strichartz estimate). *If (q, r) is an admissible pair, then we have the embedding*

$$\|u\|_{L_I^q L_x^p} \lesssim \|u\|_{X_{0,1/2+\delta}(I)}.$$

Proof. We reproduce the well-known argument. Replace u by an extension to $t \in \mathbb{R}$ such that $\|u\|_{X_{0,1/2+\delta}} \leq 2\|u\|_{X_{0,1/2+\delta}(I)}$. Write

$$u(x, t) = \int_\xi \int_\tau e^{it\tau} e^{ix\cdot\xi} \hat{u}(\xi, \tau)\, d\tau\, d\xi.$$

Change variables $\tau \mapsto \tau - |\xi|^2$ and apply Fubini to obtain

$$u(x, t) = \int_\tau e^{it\tau} \int_\xi e^{-it|\xi|^2} e^{ix\cdot\xi} \hat{u}(\xi, \tau - |\xi|^2)\, d\xi\, d\tau.$$

Define $f_\tau(x)$ by $\hat{f}_\tau(\xi) = \hat{u}(\xi, \tau - |\xi|^2)$. Then the above reads

$$u(x, t) = \int_\tau e^{it\tau} e^{it\Delta} f_\tau(x)\, d\tau,$$

and hence,

$$|u(x,t)| \le \int_\tau |e^{it\Delta} f_\tau(x)|\, d\tau.$$

Apply the Strichartz norm, the Minkowski integral inequality, appeal to Lemma 2.1, and invoke Plancherel to obtain

$$\|u\|_{L_I^q L_x^p} \lesssim \int_\tau \|\hat{f}_\tau(\xi)\|_{L_\xi^2}\, d\tau.$$

The argument is completed using Cauchy–Schwarz in τ (note that we need $b > \frac{1}{2}$, since $\int_{\mathbb{R}} \langle\tau\rangle^{-2b}\, d\tau$ has to be finite). □

Lemma 2.6 (Bourgain bilinear Strichartz estimate). *Let $N_1 \ll N_2$. Then*

$$\|P_{N_1}u_1 P_{N_2}u_2\|_{L_I^2 L_x^2} \lesssim \left(\frac{N_1^{d-1}}{N_2}\right)^{1/2} \|u_1\|_{X_{0,1/2+\delta}(I)} \|u_2\|_{X_{0,1/2+\delta}(I)},$$
$$\|P_{N_1}u_1 \overline{P_{N_2}u_2}\|_{L_I^2 L_x^2} \lesssim \left(\frac{N_1^{d-1}}{N_2}\right)^{1/2} \|u_1\|_{X_{0,1/2+\delta}(I)} \|u_2\|_{X_{0,1/2+\delta}(I)}.$$

Proof. We reproduce the well-known argument. As in the proof of Lemma 2.5, taking $f_{j,\tau}(x)$ defined by $\hat{f}_{j,\tau}(\xi) = \hat{u}_1(\xi, \tau - |\xi|^2)$, we have

$$u_j(x,t) = \int_\tau e^{it\tau}\, e^{it\Delta} f_{j,\tau}(x)\, d\tau.$$

Plug these into the expression $\|P_{N_1}u_1\ P_{N_2}u_2\|_{L_t^2 L_x^2}$, and then estimate using Lemma 2.2. □

We need to take $b = \frac{1}{2} - \delta$ in some places. In those situations, we use this:

Lemma 2.7 (interpolated Strichartz). *Take $d = 1$ or $d = 2$ and suppose that $0 \le b < \frac{1}{2}$ and $2 \le p \le \infty$ and $2 < q \le \infty$ satisfy*

$$\frac{2}{q} + \frac{d}{p} > \frac{d}{2} + (1-2b), \tag{2-11}$$
$$\frac{2}{q} - \frac{1}{p} \le \frac{1}{2} \qquad \textit{in the case } d = 1 \textit{ only} \tag{2-12}$$

(*see Figure 1*). *Then*

$$\|u\|_{L_I^q L_x^p} \lesssim \|u\|_{X_{0,b}(I)}. \tag{2-13}$$

with implicit constant dependent upon the size of the gap from equality in (2-11).

Proof. Let

$$\alpha := \frac{1}{2}\left(\frac{2}{q} + \frac{d}{p} - \frac{d}{2} - (1-2b)\right) > 0. \tag{2-14}$$

Using $0 \le \theta \le 1$ as an interpolation parameter, we aim to deduce (2-13) by interpolation between

$$\|u\|_{L_t^{\tilde{q}} L_x^{\tilde{p}}} \lesssim \|u\|_{X_{0,b/(2(b-\alpha))}}, \tag{2-15}$$

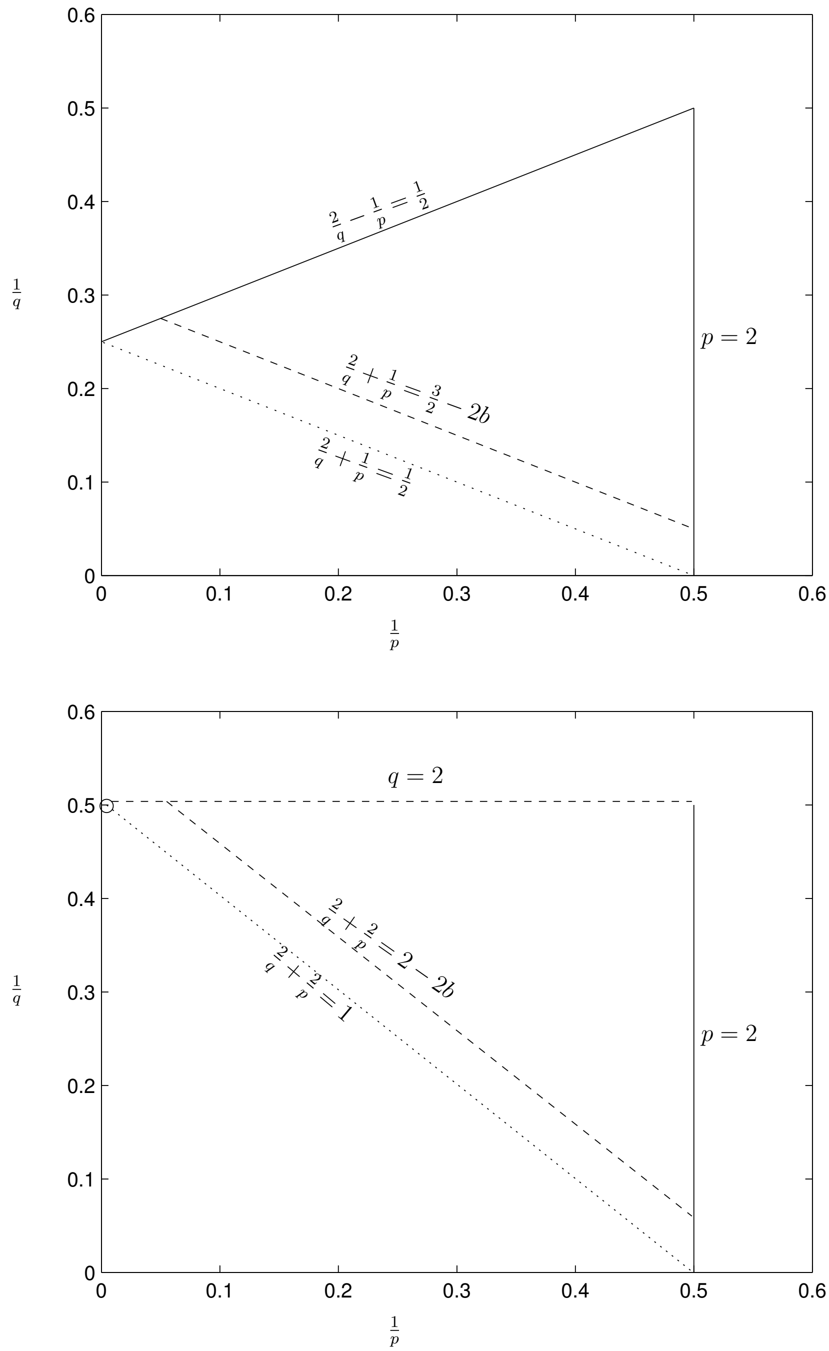

Figure 1. The enclosed triangular region gives the values of $(1/q, 1/p)$ meeting the hypotheses of Lemma 2.7. The top frame is the case $d = 1$ and the bottom frame is the case $d = 2$. The proof of Lemma 2.7 involves interpolating between a point on the line $2/q + d/p = d/2$ and the point $(1/2, 1/2)$.

with weight θ, for some Strichartz admissible pair $(\tilde{q}, \tilde{p})$, and the trivial estimate (equality, in fact)

$$\|u\|_{L_t^2 L_x^2} \lesssim \|u\|_{X_{0,0}}, \tag{2-16}$$

with weight $1-\theta$. The interpolation conditions read

$$\frac{1}{q} = \frac{\theta}{\tilde{q}} + \frac{1-\theta}{2} \quad \text{and} \quad \frac{1}{p} = \frac{\theta}{\tilde{p}} + \frac{1-\theta}{2}. \tag{2-17}$$

Multiplying the first of these relations by 2 and adding d times the second, and using the Strichartz admissibility condition for $(\tilde{q}, \tilde{p})$, we obtain

$$\frac{2}{q} + \frac{d}{p} = \frac{d}{2} + (1-\theta).$$

Combining this relation with (2-14), we get $\theta = 2b - 2\alpha$. We can then solve for $\tilde{q}$ and $\tilde{p}$ using (2-17). □

Lemma 2.8 (interpolated bilinear Strichartz). *Let $d = 1$ or $d = 2$ and $N_1 \ll N_2$. Then*

$$\|P_{N_1} u_1 \; P_{N_2} u_2\|_{L_I^2 L_x^2} \lesssim \frac{N_1^{(d-1)/2}}{N_2^{1/2-\delta'}} \|u_1\|_{X_{0,1/2-\delta}(I)} \|u_2\|_{X_{0,1/2-\delta}(I)}.$$

Proof. First, observe that

$$\|P_{N_1} u_1 \; P_{N_2} u_2\|_{L_I^2 L_x^2} \lesssim \|u_1\|_{L_I^4 L_x^4} \|u_2\|_{L_I^4 L_x^4}. \tag{2-18}$$

In the case $d = 1$, $L_I^4 L_x^4$ interpolates between $L_I^6 L_x^6$ and $L_I^2 L_x^2$, and thus $\|u_j\|_{L_I^4 L_x^4} \lesssim \|u_j\|_{X_{0,3/8+\delta}(I)}$ by Lemma 2.7. We conclude that

$$\|P_{N_1} u_1 \; P_{N_2} u_2\|_{L_I^2 L_x^2} \lesssim \|u_1\|_{X_{0,3/8+\delta}(I)} \|u_2\|_{X_{0,3/8+\delta}(I)}.$$

Interpolating this with the result of Lemma 2.6 completes the proof in the case $d = 1$.

In the case $d = 2$, we still begin with (2-18). Fix $\epsilon > 0$ small. By Sobolev embedding,

$$\|P_{N_j} u_j\|_{L_I^4 L_x^4} \lesssim N_j^{\epsilon} \|P_{N_j} u_j\|_{L_I^4 L_x^{4/(1+2\epsilon)}}.$$

By Lemma 2.7, we have

$$\|P_{N_j} u_j\|_{L_I^4 L_x^{4/(1+2\epsilon)}} \lesssim \|u_j\|_{X_{0,b}}$$

for any $b > \frac{1}{2}(1-\epsilon)$. Plugging into (2-18), we obtain

$$\|P_{N_1} u_1 \; P_{N_2} u_2\|_{L_I^2 L_x^2} \lesssim N_2^{2\epsilon} \|u_1\|_{X_{0,b}} \|u_2\|_{X_{0,b}} \quad \text{for any } b > \tfrac{1}{2}(1-\epsilon).$$

Interpolating this with the result of Lemma 2.6 completes the proof in the case $d = 2$. □

Remark 2.9. After this section we will adopt new notation: Instead of $X_{s,1/2+\delta}$ we will simply write $X_{s,1/2+}$. If an expression has two different Bourgain spaces, it will mean that the delta's will be different. Similarly, if an expression involves δ in the estimate on the right side, it will mean that this δ will be different from the one that would be chosen for spaces such as $X_{s,1/2+}$ or L^{p-}.

The following is a simple consequence of the pseudodifferential calculus; see [Stein 1993, Theorem1 on page 234 and Theorem 2 on page 237]; see also [Evans and Zworski 2003].

Lemma 2.10. *Suppose that ϕ is a smooth function on $\mathbb{R}$ such that $\|\partial_x^\alpha \phi\|_{L^\infty} \le c_\alpha$ for all $\alpha \ge 0$. Then*

$$\|P_{\ge N}(\phi g) - \phi P_{\ge N} g\|_{L^2} \lesssim N^{-1}\|g\|_{L^2} \quad \text{for } N \ge 1.$$

Proof. Let $\chi(\xi)$ be a smooth function that is 1 for $|\xi| \ge 1$ and is 0 for $|\xi| \le \frac{1}{2}$. $P_{\ge N}$ is a pseudodifferential operator with symbol $\chi(N^{-1}\xi)$ and M_ϕ, the operator of multiplication by ϕ, is a pseudodifferential operator with symbol $\phi(x)$. The commutator $[P_N, M_\phi]$ has symbol with top-order asymptotic term $N^{-1}\chi'(N^{-1}\xi)\phi'(x)$. The result then follows from the $L^2 \to L^2$ boundedness of 0-order operators. □

3. Additional high-frequency regularity

In this section, we begin the proof of Theorem 1.1 by showing improved regularity at high frequencies, above the blow-up scale, *with no restriction in space* — this appears as Proposition 3.4 below. In Section 4 below, we will complete the proof of Theorem 1.1 by appealing to a finite-speed of propagation argument for lower frequencies *after we have restricted in space* to outside the blow-up core.

Consider a solution $u(t)$ to (1-1) in the Merle–Raphaël class (1-5); let $T_0 > 0$ be the threshold time, $T > T_0$ the blow-up time and x_0 the blow-up point, as described in the introduction. Our analysis focuses on the time interval $[T_0, T)$ on which the log-log asymptotics (1-6) kick in. Apply a space-time (rescaling) shift, in which $x = x_0$ is sent to $x = 0$ and the time interval $[T_0, T)$ is sent to $[0, 1)$, to obtain a transformed solution that we henceforth still denote by $u(t)$. Now the blow-up time is $T = 1$, the blow-up point is $x = 0$, and (1-6) becomes[2]

$$\|\nabla u(t)\|_{L_x^2} \sim \left(\frac{\log|\log(1-t)|}{1-t}\right)^{1/2}, \tag{3-1}$$

which is now valid for all $0 \le t < 1$. Note that now, however, the time $t = 0$ "initial data", which we henceforth denote u_0, does not correspond to the original initial data u_0 in Theorem 1.1. We remark that the estimate (1-8) on the remainder $\tilde{u}(t)$ becomes

$$\|\nabla \tilde{u}(t)\|_{L_x^2} \lesssim \frac{1}{(1-t)^{1/2}|\log(1-t)|}. \tag{3-2}$$

In our analysis, the norm $L_I^\infty L_x^2$ for an interval $I = [0, T']$, $T' < T$, will be replaced by the norm $X_{0,1/2+}(I)$. While we have, from Lemma 2.5, the bound

$$\|u\|_{L_I^\infty L_x^2} \lesssim \|u\|_{X_{0,1/2+}(I)},$$

the reverse bound does not in general hold. Nevertheless, (3-1) indicates that the solution is blowing up close to the scale rate $(1-t)^{-1/2}$. Thus, the local theory combined with (3-1) implies a bound on $\|u\|_{X_{1,1/2+}(I)}$, where $\log|\log(1-T')|$ is weakened to $(1-T')^{-\delta}$.

[2] The rescaling is the following. If we take $u(x,t)$ in the original frame (for $T_0 \le t < T$), and let

$$u(x,t) = \mu^{d/2} v(\mu(x-x_0), \mu^2(t-T_0))$$

with $\mu = (T-T_0)^{-1/2}$, then $v(y,s)$ is defined in the modified frame (for $0 \le s < 1$). Moreover, we have $\|\nabla v(s)\|_{L_x^2} \sim (\log|\log \mu^{-2}(1-s)|)^{1/2}(1-s)^{-1/2}$, so now the implicit constant of comparability in (3-1) depends on $T - T_0$.

Lemma 3.1. *For $I = [0, T']$ with $T' < T$, for $0 < s \le 1$, we have*

$$\|u\|_{X_{s,\frac{1}{2}+}(I)} \le c_s(1-T')^{-s(1+\delta)/2}, \quad \text{with } c_s \nearrow +\infty \text{ as } s \searrow 0.$$

The fact that c_s diverges as $s \searrow 0$ results from the fact that (1-1) is L^2-critical, and thus, the local theory estimates break down at $s = 0$. At the technical level, some slack is needed in applying the Strichartz and bilinear Strichartz estimates; hence, we need to take $b = 1/2 - \delta$ in place of $b = 1/2 + \delta'$.

Proof. We just carry out the argument for $s = 1$. Let $\lambda(t) = \|\nabla u(t)\|_{L^2}^{-1}$. Let s_k be the increasing sequence of times[3] such that $\lambda(s_k) = 2^{-k}$, so that $\|\nabla u(t)\|_{L^2}$ doubles over $[s_k, s_{k+1}]$. From (3-1), we compute that $s_k = 1 - 2^{-2k} \log k$. Note that $s_{k+1} - s_k \approx 2^{-2k} \log k$. Hence, we can rescale the cutoff solution $u(t)$ on the time interval $[s_k, s_{k+1}]$ to a solution u' on the time interval $[0, \log k]$ so that $\|u'\|_{L^\infty_{[0,\log k]} H^1_x} \sim 1$. We invoke the local theory over $\sim \log k$ time intervals J each of unit size to obtain $\|u'\|_{X_{1,1/2+}(J)} \sim 1$, which are square summed to obtain $\|u'\|_{X_{1,1/2+}(0,\log k)} \sim (\log k)^{1/2}$. Returning to the original frame of reference, we conclude that

$$\|u\|_{X_{1,1/2+}(s_k,s_{k+1})} \lesssim 2^{k(1+\delta)},$$

where a δ-loss is incurred in part from the $(\log k)^{1/2}$ factor but also from the $b = \frac{1}{2} + \delta$ weight in the X norm. Thus,

$$\|u\|_{X_{1,1/2+}(0,s_K)} = \left(\sum_{k=1}^{K-1} 2^{2k(1+\delta)}\right)^{1/2} \sim 2^{K(1+\delta)}. \qquad \square$$

Now suppose that $u(t)$ satisfies (3-1). Let $t_k = 1 - 2^{-k}$ and $I_k = [0, t_k]$. Then from (3-1) and mass conservation, we have

$$\|P_{\ge N} u(t)\|_{L^\infty_{I_k} L^2_x} \lesssim \begin{cases} 2^{k(1+\delta)/2} N^{-1} & \text{for } N \ge 2^{k(1+\delta)/2}, \\ 1 & \text{for } N \le 2^{k(1+\delta)/2}. \end{cases} \tag{3-3}$$

To refine (3-3), we will work with local-theory estimates and thus use the analogous bound on the Bourgain norm $X_{0,1/2+}(I_k)$. From Lemma 3.1 we obtain

$$\|P_{\ge N} u\|_{X_{0,1/2+}(I_k)} \lesssim N^{-s} \|P_{\ge N} u\|_{X_{s,1/2+}(I_k)} \le c_s N^{-s} 2^{ks(1+\delta)/2}. \tag{3-4}$$

We obtain from (3-4) that

$$\|P_{\ge N} u\|_{X_{0,1/2+}(I_k)} \lesssim \begin{cases} 2^{k(1+\delta)/2} N^{-1} & \text{for } N \ge 2^{k(1+\delta)/2}, \\ 2^{k\delta'} & \text{for } N \le 2^{k(1+\delta)/2}. \end{cases} \tag{3-5}$$

The next step is to run local-theory estimates to improve (3-5) at *high* frequencies. Frequencies $N \lesssim 2^k \sim (1 - t_k)^{-1}$ on I_k effectively do not make it out of the blow-up core before blow-up time due to the finite speed of propagation for such frequencies.[4] Hence, these *low* frequencies can be controlled by spatial location, which we address in Section 4. On the other hand, (3-5) shows that the solution at

[3] One of the conclusions of the Merle–Raphaël analysis is the almost monotonicity of the scale parameter $\lambda(t) = \|\nabla u(t)\|_{L^2}^{-1}$: $\lambda(t_2) < 2\lambda(t_1)$ for all $t_2 \ge t_1$.

[4] Recall that for the Schrödinger equation, frequencies of size N propagate at speed N and thus travel a distance $O(1)$ in time N^{-1}.

frequencies $N \gtrsim 2^{k(1+\delta)/2}$ is small. Thus, for these *high* frequencies, dispersive estimates might be able, upon iteration, to show that the solution is even smaller at these high frequencies.

To chose an intermediate dividing point between the high frequencies that are capable of exiting the blow-up core before blow-up time ($N \gtrsim 2^k$) and the frequency scale at which the blow-up is taking place ($N \sim 2^{k/2}(\log k)^{1/2}$), we consider frequencies $\geq 2^{3k/4}$ to be *high* frequencies and frequencies $\leq 2^{3k/4}$ to be *low* frequencies. The goal of this section is Proposition 3.4 below, which shows that the high frequencies are bounded in H^1. In Section 4 below, we will localize in space to the external region and then control the low frequencies.

We first address the dimension $d = 1$ case.

Lemma 3.2 (high frequency recurrence in one dimension). *Take $d = 1$. Let $t_k = 1 - 2^{-k}$ and $I_k = [0, t_k]$. Let $u(t)$ be a solution such that* (3-1) *holds, and define*

$$\alpha(k, N) = \|P_{\geq N} u\|_{X_{0,1/2+}(I_k)}. \tag{3-6}$$

Then there exists an absolute constant $0 < \mu \ll 1$ such that for $N \geq 2^{k(1+\delta)/2}$,

$$\|P_{\geq N}(u - e^{it\partial_x^2}u_0)\|_{X_{0,1/2+}(I_k)} \lesssim 2^{k(1+\delta)/2} N^{-1+\delta} \alpha(k+1, \mu N) + 2^{k\delta}\alpha(k+1, \mu N)^2. \tag{3-7}$$

In particular, by Lemma 2.4,

$$\alpha(k, N) \lesssim \|P_{\geq N} u_0\|_{L^2_x} + 2^{k(1+\delta)/2} N^{-1+\delta} \alpha(k+1, \mu N) + 2^{k\delta}\alpha(k+1, \mu N)^2. \tag{3-8}$$

Proof. By (2-7) of Lemma 2.4 with $\omega = 2^{-k-1}$ and $I = I_k$,

$$\|P_{\geq N}(u - e^{it\partial_x^2}u_0)\|_{X_{0,1/2+}(I_k)} \lesssim 2^{k\delta} \|P_{\geq N}(|u|^4 u)\|_{X_{0,-1/2+}(I_{k+1})}.$$

In the rest of the proof, we estimate the right side of the estimate above, and we will just write I_k instead of I_{k+1} for convenience. By duality,

$$\|P_{\geq N}(|u|^4 u)\|_{X_{0,-1/2+}(I_k)} = \sup_{\|w\|_{X_{0,1/2-}(I_k)}=1} \int_{I_k}\int_{x\in\mathbb{R}} P_{\geq N}(|u|^4 u)\, w \, dx\, dt.$$

Fix w with $\|w\|_{X_{0,1/2-}(I_k)} = 1$ and let

$$J := \int_{I_k}\int_{x\in\mathbb{R}} P_{\geq N}(|u|^4 u)\, w\, dx\, dt.$$

Then J can be decomposed into a finite sum of terms J_α, each of the form (we have dropped complex conjugates, since they are unimportant in the analysis)

$$J_\alpha := \int_0^{t_k}\int_{x\in\mathbb{R}} P_{\geq N}(u_1 u_2 u_3 u_4 u_5)\, w\, dx\, dt$$

such that each term (after a relabeling of the u_j for $1 \leq j \leq 5$) falls into exactly one of the following two categories.[5]

[5]Indeed, decompose each u_j as $u_j = u_{j,\mathrm{lo}} + u_{j,\mathrm{med}} + u_{j,\mathrm{hi}}$, where $u_{j,\mathrm{lo}} = P_{\leq N/160} u_j$, $u_{j,\mathrm{med}} = P_{N/160 \leq \cdot \leq N/20}$, and $u_{j,\mathrm{hi}} = P_{\geq N/20} u_j$. Then in the expansion of $u_1 u_2 u_3 u_4 u_5$, at least one term must be "hi"; without loss take this to be u_5.

Note that w is frequency supported in $|\xi| \gtrsim N$.

Case 1 (exactly one high). Each u_j for $1 \le j \le 4$ is frequency supported in $|\xi| \le \mu N$ and u_5 is frequency supported in $|\xi| \ge 8\mu N$. In this case, we estimate as

$$|J_\alpha| \le \|u_1\|_{L^\infty_{I_k} L^\infty_x} \|u_2\|_{L^\infty_{I_k} L^\infty_x} \|u_3 u_5\|_{L^2_{I_k} L^2_x} \|u_4 w\|_{L^2_{I_k} L^2_x}. \tag{3-9}$$

For $j = 1, 2$, Gagliardo–Nirenberg and (3-1) implies

$$\|u_j\|_{L^\infty_{I_k} L^\infty_x} \lesssim \|u_j\|^{1/2}_{L^\infty_{I_k} L^2_x} \|\partial_x u_j\|^{1/2}_{L^\infty_{I_k} L^2_x} \lesssim 2^{k(1+\delta)/4}. \tag{3-10}$$

The bilinear Strichartz estimate (Lemma 2.6) yields

$$\|u_3 u_5\|_{L^2_{I_k} L^2_x} \lesssim N^{-1/2} \|u_3\|_{X_{0,1/2+}(I_k)} \|u_5\|_{X_{0,1/2+}(I_k)} \lesssim N^{-1/2} 2^{k\delta} \alpha(k, \mu N). \tag{3-11}$$

The interpolated bilinear Strichartz estimate (Lemma 2.8) yields

$$\|u_4 w\|_{L^2_{I_k} L^2_x} \lesssim N^{-1/2+\delta} \|u_4\|_{X_{0,1/2+}(I_k)} \|w\|_{X_{0,1/2-}(I_k)} \lesssim N^{-1/2+\delta} 2^{k\delta}. \tag{3-12}$$

Substituting (3-10), (3-11), and (3-12) into (3-9), we obtain

$$|J_\alpha| \lesssim 2^{k(1+\delta)/2} N^{-1+\delta} \alpha(k, \mu N).$$

Case 2 (at least two high). Both u_4 and u_5 are frequency supported in $|\xi| \ge \mu N$ (no restrictions on u_j for $1 \le j \le 3$). Then we estimate as

$$|J_\alpha| \le \|u_1\|_{L^6_{I_k} L^{6+\delta}_x} \|u_2\|_{L^6_{I_k} L^6_x} \|u_3\|_{L^6_{I_k} L^6_x} \|u_4\|_{L^6_{I_k} L^6_x} \|u_5\|_{L^6_{I_k} L^6_x} \|w\|_{L^6_{I_k} L^{6-\delta'}_x}. \tag{3-13}$$

For $2 \le j \le 3$ we invoke the Strichartz estimate (Lemma 2.5) and (3-5) to obtain

$$\|u_j\|_{L^6_{I_k} L^6_x} \lesssim \|u_j\|_{X_{0,1/2+}(I_k)} \le 2^{k\delta}. \tag{3-14}$$

For $4 \le j \le 5$ we invoke the Strichartz estimate (Lemma 2.5) and (3-6) to obtain

$$\|u_j\|_{L^6_{I_k} L^6_x} \lesssim \|u_j\|_{X_{0,1/2+}} \le \alpha(k, \mu N). \tag{3-15}$$

For $j = 1$, by Sobolev embedding, the Strichartz estimate (Lemma 2.5), and (3-5),

$$\|u_1\|_{L^6_{I_k} L^{6+}_x} \lesssim \|D^\delta_x u_1\|_{L^6_{I_k} L^6_x} \lesssim \|u_1\|_{X_{\delta,1/2+}(I_k)} \lesssim 2^{k\delta}. \tag{3-16}$$

By the interpolated Strichartz estimate (Lemma 2.7), we have

$$\|w\|_{L^6_t L^{6-}_x} \lesssim \|w\|_{X_{0,1/2-}(I_k)} = 1. \tag{3-17}$$

Using (3-14)–(3-17) in (3-13),

$$|J_\alpha| \lesssim 2^{k\delta} \alpha(k, \mu N)^2.$$ □

In the 2D case, we will just go ahead and assume that $N \ge 2^{3k/4}$ to reduce confusion with deltas.

Case 1 corresponds to $u_{1,\mathrm{lo}} u_{2,\mathrm{lo}} u_{3,\mathrm{lo}} u_{4,\mathrm{lo}} u_{5,\mathrm{hi}}$ and Case 2 corresponds to everything else (at least one u_j for $1 \le j \le 4$ must be “med” or “hi”. Hence, we can take $\mu = 1/160$.

Lemma 3.3 (high frequency recurrence, 2D). *Take $d = 2$. Let $t_k = 1 - 2^{-k}$ and $I_k = [0, t_k]$. Let $u(t)$ be a solution such that* (3-1) *holds and define*

$$\alpha(k, N) := \|P_{\geq N} u\|_{X_{0,1/2+}(I_k)}. \tag{3-18}$$

Then there exists an absolute constant $0 < \mu \ll 1$ such that for $N \gtrsim 2^{3k/4}$,

$$\|P_{\geq N}(u - e^{it\Delta}u_0)\|_{X_{0,1/2+}(I_k)} \lesssim 2^{k\delta} N^{-1/6+\delta} \alpha(k+1, \mu N). \tag{3-19}$$

In particular, by Lemma 2.4,

$$\alpha(k, N) \lesssim \|P_{\geq N} u\|_{L^2_x} + 2^{k\delta} N^{-1/6+\delta} \alpha(k+1, \mu N). \tag{3-20}$$

Proof. By Lemma 2.4 (2-7) with $I = I_k$ and $\omega = 2^{-k-1}$,

$$\|P_{\geq N}(u - e^{it\Delta}u_0)\|_{X_{0,1/2+}(I_k)} \lesssim 2^{k\delta} \|P_{\geq N}(|u|^2 u)\|_{X_{0,-1/2+}(I_{k+1})}.$$

In the remainder of the proof, we estimate the right side, and for convenience take I_{k+1} to be I_k. By duality,

$$\|P_{\geq N}(|u|^2 u)\|_{X_{0,-1/2+}(I_k)} = \sup_{\|w\|_{X_{0,1/2-}(I_k)}=1} \int_{I_k} \int_{x \in \mathbb{R}} P_{\geq N}(|u|^2 u)\, w \, dx \, dt.$$

Fix w with $\|w\|_{X_{0,1/2-}(I_k)} = 1$ and let

$$J := \int_{I_k} \int_{x \in \mathbb{R}} P_{\geq N}(|u|^2 u)\, w \, dx \, dt.$$

Then J can be decomposed into a finite sum of terms J_α, each of the form (we have dropped complex conjugates, since they are unimportant in the analysis)

$$J_\alpha := \int_0^{t_k} \int_{x \in \mathbb{R}} P_{\geq N}(u_1 u_2 u_3)\, w \, dx \, dt$$

such that each term (after a relabeling of the u_j for $1 \leq j \leq 3$) falls into exactly one of the following two categories.[6] Note that w is frequency supported in $|\xi| \gtrsim N$.

Case 1′ (exactly one high). Both u_1 and u_2 are frequency supported in $|\xi| \leq N^{5/6}$ and u_3 is frequency supported in $|\xi| \geq N/12$. In this case, we estimate as

$$|J_\alpha| \lesssim \|u_1 w\|_{L^2_{I_k} L^2_x} \|u_2 u_3\|_{L^2_{I_k} L^2_x}.$$

By the interpolated bilinear Strichartz estimate (Lemma 2.8),

$$\|u_1 w\|_{L^2_{I_k} L^2_x} \lesssim (N^{5/6})^{1/2} N^{-1/2+\delta} \|u_1\|_{X_{0,1/2-}(I_k)} \|w\|_{X_{0,1/2-}(I_k)} \lesssim N^{-1/12+\delta} 2^{k\delta},$$

[6] Indeed, decompose $u_j = u_{j,\text{lo}} + u_{j,\text{med}} + u_{j,\text{hi}}$, where $u_{j,\text{lo}} = P_{\leq N^{5/6}} u_j$, $u_{j,\text{med}} = P_{N^{5/6} \leq \cdot \leq N/12}$, and $u_{j,\text{hi}} = P_{\geq N/12} u_j$. Then at least one term must be "hi"; take it to be u_3. Case 1′ corresponds to $u_{1,\text{lo}} u_{2,\text{lo}} u_{3,\text{hi}}$ and Case 2′ corresponds to all other possibilities. Hence, we can take $\mu = 1/12$.

and by Lemma 2.6 directly,

$$\|u_2u_3\|_{L^2_{I_k}L^2_x} \lesssim (N^{5/6})^{1/2}N^{-1/2+\delta}\|u_2\|_{X_{0,1/2+}(I_k)}\|u_3\|_{X_{0,1/2+}(I_k)} \lesssim N^{-1/12+\delta}2^{k\delta}\alpha(k,\mu N).$$

Combining yields

$$|J_\alpha| \lesssim N^{-1/6+\delta}2^{k\delta}\alpha(k,\mu N).$$

Case 2′ (at least two high). Here we suppose that u_2 is frequency supported in $|\xi| \geq N^{5/6}$ and u_3 is frequency supported in $|\xi| \geq \mu N$; we make no assumptions about u_1. Then we estimate as

$$|J_\alpha| \lesssim \|u_1\|_{L^4_{I_k}L^{4+\delta}_x}\|u_2\|_{L^4_{I_k}L^4_x}\|u_3\|_{L^4_{I_k}L^4_x}\|w\|_{L^4_{I_k}L^{4-\delta}_x}.$$

For u_1, we use Sobolev embedding and (3-5) to obtain

$$\|u_1\|_{L^4_{I_k}L^{4+\delta}_x} \lesssim \|D^\delta_x u_1\|_{L^4_{I_k}L^4_x} \lesssim \|u_1\|_{X_{\delta,\frac{1}{2}+}(I_k)} \lesssim 2^{k\delta}.$$

Since $N \gtrsim 2^{3k/4}$, we have $N^{5/6} \gtrsim 2^{5k/8} \gg 2^{k(1+\delta)/2}$, and thus by Lemma 2.5 and (3-5),

$$\begin{aligned}\|u_2\|_{L^4_{I_k}L^4_x} &\lesssim 2^{k(1+\delta)/2}N^{-5/6} \lesssim (2^{k(1+\delta)}N^{-2/3})N^{-1/6} \\ &\lesssim 2^{k\alpha}N^{-1/6}, \quad \text{since } N \gtrsim 2^{3k/4}.\end{aligned}$$

For u_3, we use Lemma 2.5 and (3-18) to obtain

$$\|u_3\|_{L^4_{I_k}L^4_x} \lesssim \alpha(k,\mu N).$$

Combining, we obtain (changing deltas)

$$|J_\alpha| \lesssim 2^{k\delta}N^{-1/6}\alpha(k,\mu N). \qquad \square$$

The main result of this section is the following. It states that high frequencies (those strictly above $2^{3k/4}$) are H^1 bounded on I_k. Moreover, if we subtract the linear flow, we obtain $H^{4/3-\delta}$ boundedness for frequencies above $2^{3k/4}$ in the case $d=1$ and $H^{7/6-\delta}$ boundedness for frequencies above $2^{3k/4}$ in the case $d=2$.[7]

Proposition 3.4. *Let $t_k = 1-2^{-k}$, $I_k = [0,t_k]$, and let $u(t)$ be a solution to* (1-1) *such that* (3-1) *holds. Then we have*

$$\|P_{\geq 2^{3k/4}}u(t)\|_{L^\infty_{I_k}H^1_x} \lesssim \|P_{\geq 2^{3k/4}}u(t)\|_{X_{1,1/2+}(I_k)} \lesssim 1.$$

Moreover, we have the following regularity above H^1 *after the linear flow of the initial data is removed: For any $0 \leq s \leq \frac{4}{3}-\delta$ in the case $d=1$ and for any $0 \leq s \leq \frac{7}{6}-\delta$ in the case $d=2$, we have*

$$\|P_{\geq 2^{3k/4}}(u(t)-e^{it\Delta}u_0)\|_{L^\infty_{I_k}H^s_x} \lesssim \|P_{\geq 2^{3k/4}}(u(t)-e^{it\Delta}u_0)\|_{X_{s,1/2+\delta}(I_k)} \lesssim 1. \tag{3-21}$$

[7] In fact, the threshold $\geq 2^{3k/4}$, to obtain H^1 boundedness (but not (3-21)), can be replaced by $2^{k(1+\delta)/2}$ for any $\delta > 0$; in the $d=1$ case, one can appeal to Lemma 3.2 with a strictly smaller choice of δ in order to obtain a nontrivial gain upon each application of Lemma 3.2. The number of applications of Lemma 3.2 is still finite number but δ-dependent. In the 2D case, Lemma 3.3 would first need to be rewritten. We have stated the proposition with threshold $\geq 2^{3k/4}$ because this is all that is needed in Section 4, and it allows us to avoid confusion with multiple small parameters.

Proof. We carry out the $d = 1$ case in full, which is a consequence of Lemma 3.2. The $d = 2$ case follows from Lemma 3.3 in a similar way.

By (3-5), we start with the knowledge that $\alpha(k, N) \lesssim 2^{k(1+\delta)/2} N^{-1}$ for $N \geq 2^{k(1+\delta)/2}$. Note

$$\|P_{\geq N} u_0\|_{L_x^2} \lesssim N^{-1} \|\nabla u_0\|_{L_x^2} \lesssim N^{-1}.$$

By (3-8) in Lemma 3.2,

$$\alpha(k, N) \lesssim N^{-1} + 2^{k(1+\delta)/2} N^{-1+\delta} \alpha(k+1, \mu N). \tag{3-22}$$

Application of (3-22) m times gives

$$\alpha(k, N) \lesssim N^{-1} \biggl(\sum_{j=0}^{m-1} (2^{k(1+\delta)/2} N^{-1+\delta})^j \biggr) + (2^{k(1+\delta)/2} N^{-1+\delta})^m \alpha(k+m, \mu^m N).$$

Since $N \geq 2^{3k/4}$, we have $2^{k/2} N^{-1} \lesssim N^{-1/3}$. Taking $m = 7$ we obtain $\alpha(k, N) \lesssim N^{-1}$. Substituting this into (3-7) of Lemma 3.2, we obtain

$$\|P_{\geq N}(u(t) - e^{it\partial_x^2} u_0)\|_{X_{0,1/2+}(I_k)} \lesssim 2^{k(1+\delta)/2} N^{-2+\delta} \lesssim N^{-4/3+\delta}. \qquad \square$$

4. Finite speed of propagation

Recall that the main result of the last section was Proposition 3.4, which showed that the solution at frequencies $\geq 2^{3k/4}$ is H^1 bounded on I_k. This was achieved without applying any restriction in space. In this section, we apply a spatial restriction to $|x| \geq R$ (outside the blow-up core), and study the low frequencies $\leq 2^{3k/4}$ on I_k. Since frequencies of size N propagate at speed N, and thus travel a distance $O(1)$ over a time N^{-1}, we expect that frequencies of size $\lesssim 2^k$ involved in the blow-up dynamics will be incapable of exiting the blow-up core $|x| \leq R$ before blow-up time.

Since $I_k = [0, t_k]$ and $t_k = 1 - 2^{-k}$, restricting to frequencies $\leq 2^{3k/4}$ on I_k for each k is effectively equivalent to inserting a time-dependent spatial frequency projection $P_{\leq(1-t)^{-3/4}}$. The main technical Lemma 4.3 below shows that, for $0 < r_1 < r_2 < \infty$, the H^s size of the solution in the external region $|x| \geq r_2$ is bounded by the $H^{s-1/8}$ size of the solution in the slightly larger external region $|x| \geq r_1$. This lemma is proved by studying the equation solved by $P_{\leq(1-t)^{-3/4}} \psi u$, where ψ is a spatial cutoff. In estimating the inhomogeneous terms of this equation, we use that the presence of the $P_{\leq(1-t)^{-3/4}}$ projection enables an exchange of α spatial derivatives for a factor of $(1-t)^{-3\alpha/4}$. This is the manner in which finite speed of propagation is implemented. Lemma 4.3 is the main recurrence device for proving Proposition 4.4, giving the H^1 boundedness of the solution in the external region, completing the proof of Theorem 1.1.

Before getting to Lemma 4.3, we begin by using the method of Raphaël [2006], based on the use of local smoothing and (3-2), to achieve a small gain of regularity.[8]

[8] In the $d = 1$ case, we obtain a gain of $2/5$ derivatives in this first step, but in fact the proof could be rewritten to achieve a gain of $s < 1/2$ derivatives. The reason $s = 1/2$ derivatives cannot be achieved in one step is the failure of the $H^{1/2} \hookrightarrow L^\infty$ embedding needed to estimate the nonlinear term. One could achieve $1/2$ derivatives by running the same argument twice, but

Lemma 4.1 (a little regularity, $d=1$ case). *Suppose $d=1$. Suppose that $u(t)$ solving* (1-1) *with H^1 initial data satisfies* (3-1). *Fix $R>0$. Then*

$$\|\langle D_x\rangle^{2/5}\psi_R u\|_{L^\infty_{[0,1)}L^2_x} \lesssim 1,$$

where $\psi_R(x)=\psi(x/R)$ and $\psi(x)$ is a smooth cutoff with $\psi(x)=1$ for $|x|\ge 1/2$ and $\psi(x)=0$ for $|x|\le 1/4$.

Proof. Let $w=\psi_R u$ and $q=\psi_{R/2}u$. Then w solves the equation

$$i\partial_t w+\partial_x^2 w=-|q|^4 w+2\partial_x(\psi_R' u)-\psi_R'' u = F_1+F_2+F_3.$$

Apply $\langle D_x\rangle^{2/5}$, and estimate with $I=[T_1,1)$ using the (dual) local smoothing estimate for the F_2 term:

$$\begin{aligned}\|\langle D_x\rangle^{2/5}w\|_{L^\infty_I L^2_x} \lesssim \|\langle D_x\rangle^{2/5}w(T_1)\|_{L^2_x}+\|\langle D_x\rangle^{2/5}F_1\|_{L^1_I L^2_x}&\\ +\|\langle D_x\rangle^{2/5}\langle D_x\rangle^{-1/2}F_2\|_{L^2_I L^2_x}+\|\langle D_x\rangle^{2/5}F_3\|_{L^1_I L^2_x}&.\end{aligned}$$

We begin by estimating term F_1. By the fractional Leibniz rule,

$$\begin{aligned}\|D_x^{2/5}F_1\|_{L^1_I L^2_x} &\lesssim \||q|^4\|_{L^1_I L^\infty_x}\|D_x^{2/5}w\|_{L^\infty_I L^2_x}+\|D_x^{2/5}|q|^4\|_{L^1_I L^{5/2}_x}\|w\|_{L^\infty_I L^{10}_x}.\\ &\lesssim \big(\||q|^4\|_{L^1_I L^\infty_x}+\|D_x^{2/5}|q|^4\|_{L^1_I L^{5/2}_x}\big)\|D_x^{2/5}w\|_{L^\infty_I L^2_x}.\end{aligned}$$

By Sobolev/Gagliardo–Nirenberg embedding and (3-2),

$$\||q|^4\|_{L^\infty_x}+\|D_x^{2/5}|q|^4\|_{L^{5/2}_x}\lesssim \|q\|_{L^2_x}^2\|\partial_x q\|_{L^2_x}^2\lesssim (1-t)^{-1}(\log(1-t)^{-1})^{-2}.$$

Applying the L^1_I time norm, we obtain a bound by $(\log(1-T_1)^{-1})^{-1}$. Hence,

$$\|\langle D_x\rangle^{2/5}F_1\|_{L^1_I L^2_x}\lesssim (\log(1-T_1)^{-1})^{-1}\|\langle D_x\rangle^{2/5}w\|_{L^\infty_I L^2_x}.$$

Next, we address term F_2. We have

$$\|\langle D_x\rangle^{2/5}\langle D_x\rangle^{-1/2}F_2\|_{L^2_I L^2_x}\lesssim \|\langle D_x\rangle^{9/10}q\|_{L^2_I L^2_x}\lesssim \|q\|_{L^\infty_I L^2_x}^{1/10}\|\|\langle\partial_x\rangle q\|_{L^2_x}^{9/10}\|_{L^2_I}.$$

From (3-2), we have $\|\partial_x q\|_{L^2_x}\lesssim (T-t)^{-1/2}|\log(1-t)|^{-1}$ and hence

$$\|\langle D_x\rangle^{2/5}\langle D_x\rangle^{-1/2}F_2\|_{L^2_I L^2_x}\lesssim (1-T_1)^{1/10}.$$

Term F_3 is comparatively straightforward. Indeed, we obtain

$$\|\langle D_x\rangle^{2/5}F_3\|_{L^1_I L^2_x}\lesssim \|u\|_{L^\infty_I L^2_x}^{3/5}\|\|\langle\partial_x\rangle\psi_2 u\|_{L^2_x}^{2/5}\|_{L^1_I}\lesssim (1-T_1)^{4/5}.$$

Collecting the estimates above, we obtain

$$\|\langle D_x\rangle^{2/5}w\|_{L^\infty_I L^2_x}\lesssim \|\langle D_x\rangle^{2/5}w(T_1)\|_{L^2_x}+(\log(1-T_1)^{-1})^{-1}\|\langle D_x\rangle^{2/5}w\|_{L^\infty_I L^2_x}+(1-T_1)^{1/10}.$$

---

this is unnecessary since we only need a small gain of $s>0$ to complete the proof of our main new Lemma 4.3/Proposition 4.4 below, which enables us to reach the full $s=1$ gain. One cannot achieve a gain of $s>1/2$ by the method employed in the proof of Lemma 4.1 alone due to the term $\partial_x(\psi_R' u)$.

By taking T_1 sufficiently close to 1 so that $(\log(1-T_1)^{-1})^{-1}$ beats out the (absolute) implicit constants furnished by the estimates, we obtain

$$\|\langle D_x\rangle^{2/5} w\|_{L_I^\infty L_x^2} \lesssim \|\langle D_x\rangle^{2/5} w(T_1)\|_{L_x^2} + (1-T_1)^{1/10}. \qquad \square$$

Lemma 4.2 (a little regularity, $d=2$ case). *Suppose $d=2$. Suppose that $u(t)$ solving* (1-1) *with H^1 initial data satisfies* (3-1). *Fix $R>0$. Then*

$$\|\langle D_x\rangle^{1/2}\psi_R u\|_{L^\infty_{[0,1)}L_x^2} \lesssim 1,$$

where $\psi_R(x)=\psi(x/R)$ and $\psi(x)$ is a smooth cutoff with $\psi(x)=1$ for $|x|\geq \frac{1}{2}$ and $\psi(x)=0$ for $|x|\leq \frac{1}{4}$.

Proof. Let $w=\psi_R u$ and $q=\psi_{R/2}u$, and take $\tilde{\psi}=\nabla_x\psi_R$ and $\tilde{\tilde{\psi}}=\Delta_x\psi_R$. Then w solves the equation

$$i\partial_t w + \Delta w = -|q|^2 w + 2\nabla_x\cdot(\tilde{\psi}\, u) - \tilde{\tilde{\psi}}\, u = F_1+F_2+F_3.$$

Apply $\langle D_x\rangle^{1/2}$, and estimate with $I=[T_1,1)$ using the (dual) local smoothing estimate for the term F_2:

$$\begin{aligned}\|\langle D_x\rangle^{1/2} w\|_{L_I^\infty L_x^2} &+ \|\langle D_x\rangle^{1/2} w\|_{L_I^4 L_x^4}\\ &\lesssim \|\langle D_x\rangle^{1/2} w_0\|_{L_x^2} + \|\langle D_x\rangle^{1/2}F_1\|_{L_I^{4/3}L_x^{4/3}} + \|F_2\|_{L_I^2L_x^2} + \|\langle D_x\rangle^{1/2}F_3\|_{L_I^1L_x^2}.\end{aligned}$$

Before we begin treating term F_1, let us note that by (3-2), $\|\nabla q\|_{L_x^2}\lesssim (1-t)^{-1/2}(\log(1-t)^{-1})^{-1}$ and hence $\|\nabla q\|_{L_I^2L_x^2}\lesssim (\log(1-T_1)^{-1})^{-1/2}$. By the fractional Leibniz rule and Sobolev/Gagliardo–Nirenberg embedding,

$$\|D_x^{1/2}|q|^2\|_{L_x^2} \lesssim \|D_x^{1/2}q\|_{L_x^4}\|q\|_{L_x^4} \lesssim \|q\|_{L_x^2}^{1/2}\|\nabla q\|_{L_x^2}^{3/2}.$$

Hence,

$$\|D_x^{1/2}|q|^2\|_{L_I^{4/3}L_x^2} \lesssim \|q\|_{L_I^\infty L_x^2}^{1/2}\|\nabla q\|_{L_I^2L_x^2}^{3/2} \lesssim (\log(1-T_1)^{-1})^{-3/4}. \tag{4-1}$$

Also, we have

$$\|q\|_{L_x^4} \lesssim \|D_x^{1/2}q\|_{L_x^2} \lesssim \|q\|_{L_x^2}^{1/2}\|\nabla q\|_{L_x^2}^{1/2},$$

and hence

$$\|q\|_{L_I^4L_x^4}^2 \lesssim \|q\|_{L_I^\infty L_x^2}\|\nabla q\|_{L_I^2L_x^2} \lesssim (\log(1-T_1)^{-1})^{-1/2}. \tag{4-2}$$

Now we proceed with the estimates for term F_1. By the fractional Leibniz rule (in x),

$$\|\langle D_x\rangle^{1/2}F_1\|_{L_I^{4/3}L_x^{4/3}} \lesssim \|\langle D_x\rangle^{1/2}|q|^2\|_{L_I^{4/3}L_x^2}\|w\|_{L_I^\infty L_x^4} + \||q|^2\|_{L_I^2L_x^2}\|\langle D_x\rangle^{1/2}w\|_{L_I^4L_x^4}.$$

By (4-1) and (4-2), we obtain

$$\|\langle D_x\rangle^{1/2}F_1\|_{L_I^{4/3}L_x^{4/3}} \lesssim (\log(1-T_1)^{-1})^{-1/2}(\|\langle D_x\rangle^{1/2}w\|_{L_I^\infty L_x^2} + \|\langle D_x\rangle^{1/2}w\|_{L_I^4L_x^4}).$$

Next, we treat the F_2 term. Again since $\|\nabla q\|_{L_x^2}\lesssim (1-t)^{-1/2}(\log(1-t)^{-1})^{-1}$,

$$\|F_2\|_{L_I^2L_x^2} \lesssim (\log(1-T_1)^{-1})^{-1}.$$

The F_3 term is comparatively straightforward.

Collecting the estimates above, we have

$$\begin{aligned}\|\langle D_x\rangle^{1/2}w\|_{L_I^\infty L_x^2}+\|\langle D_x\rangle^{1/2}w\|_{L_I^4L_x^4}&\\ \lesssim \|\langle D_x\rangle^{1/2}w(T_1)\|_{L_x^2}&+(\log(1-T_1)^{-1})^{-1}\\ &+(\log(1-T_1)^{-1})^{-1/2}(\|\langle D_x\rangle^{1/2}w\|_{L_I^\infty L_x^2}+\|\langle D_x\rangle^{1/2}w\|_{L_I^4L_x^4}).\end{aligned}$$

By taking T_1 sufficiently close to 1, we obtain

$$\|\langle D_x\rangle^{1/2}w\|_{L_I^\infty L_x^2}\lesssim \|\langle D_x\rangle^{1/2}w(T_1)\|_{L_x^2}+(\log(1-T_1)^{-1})^{-1}. \qquad \square$$

Lemma 4.3 (low frequency recurrence). *Let $d=1$ or $d=2$, $0<R\le r_1<r_2$ and $\frac18\le s\le 1$. Let $\psi_1(x)$ and $\psi_2(x)$ be smooth radial cutoff functions such that*

$$\psi_1(x)=\begin{cases}0 & \text{on } |x|\le r_1,\\ 1 & \text{on } |x|\ge \frac12(r_1+r_2)\end{cases}\qquad \text{and}\qquad \psi_2(x)=\begin{cases}0 & \text{on } |x|\le \frac12(r_1+r_2),\\ 1 & \text{on } |x|\ge r_2.\end{cases}$$

Then

$$\|D_x^s\psi_2u\|_{L^\infty_{[0,1)}L_x^2}\lesssim 1+\|\langle D_x\rangle^{s-1/8}\psi_1u\|_{L^\infty_{[0,1)}L_x^2}.$$

Proof. Let $\chi(\rho)$ be a smooth function such that $\chi(\rho)=1$ for $|\rho|\le 1$ for $\chi(\rho)=0$ for $|\rho|\ge 2$. Let $P_-=P_{\le (T-t)^{-3/4}}$ be the time-dependent multiplier operator defined by $\widehat{Pf}(\xi)=\chi((T-t)^{3/4}|\xi|)\hat f(\xi)$ (where the Fourier transform is in space only). Note that the Fourier support of P at time $t_k=1-2^{-k}$ is $\lesssim 2^{3k/4}$. We further have that

$$\partial_tP_-f=\tfrac34 i(1-t)^{-1/4}QD_xf+P\partial_tf,$$

where $Q=Q_{(1-t)^{-3/4}}$ is the time-dependent multiplier

$$\widehat{Qf}(\xi)=\chi'((1-t)^{3/4}|\xi|)\hat f(\xi).$$

Note that the Fourier support of Q at time $t_k=1-2^{-k}$ is $\sim 2^{3k/4}$. Note also that if $g=g(x)$ is any function, then

$$\|PD_x^\alpha g\|_{L_x^2}\le (1-t)^{-3\alpha/4}\|g\|_{L_x^2}. \tag{4-3}$$

Let $w=P_-\psi_2u$. Taking $\tilde\psi_2=\nabla_x\psi_2$ and $\tilde{\tilde\psi}_2=\Delta_x\psi_2$, we have

$$\begin{aligned}i\partial_tw+\Delta w&=-i(1-t)^{-1/4}Q\cdot\nabla_x\,w-P_-\psi_2|u|^{4/d}u+2P_-\nabla_x\cdot[\tilde\psi_2u]-P_-\tilde{\tilde\psi}_2u\\ &=F_1+F_2+F_3+F_4.\end{aligned}$$

By the energy method,

$$\|D_x^sw\|^2_{L^\infty_{[0,1)}L_x^2}\lesssim \|D_x^sw(0)\|^2_{L_x^2}+\int_0^1|\langle D_x^sF_1(s),D_x^sw(s)\rangle_{L_x^2}|\,ds+10\sum_{j=2}^4\|D_x^sF_j\|^2_{L^1_{[0,1)}L_x^2}.$$

For term F_1, we argue as follows. Let $\tilde Q$ be a projection onto frequencies of size $(1-t)^{-3/4}$. Then

$$\int_0^1|\langle D_x^sF_1(s),D_x^sw(s)\rangle_{L_x^2}|\,ds\lesssim\int_0^1(1-s)^{-1/4}\|D_x^{1/2+s}\tilde Q\psi_2u(s)\|^2_{L_x^2}\,ds.$$

Applying (4-3) with $\alpha = \frac{1}{2}$, we can control the above by

$$\int_0^1 (1-s)^{-1} \| D_x^s \tilde{Q} \psi_2 u(s) \|_{L_x^2}^2 \, ds.$$

Dividing the time interval $[0,1) = \bigcup_{k=1}^{\infty} [t_k, t_{k+1})$, we bound the above by

$$\sum_{k=1}^{+\infty} 2^k \int_{t_k}^{t_{k+1}} \| D_x^s P_{2^{3k/4}} \psi_2 u(s) \|_{L_x^2}^2 \, ds \lesssim \sum_{k=1}^{+\infty} \| D_x^s P_{2^{3k/4}} \psi_2 u(s) \|_{L^\infty_{[t_k, t_{k+1})} L_x^2}^2,$$

where $P_{2^{3k/4}}$ is the projection onto frequencies of size $\sim 2^{3k/4}$ (and not $\lesssim 2^{3k/4}$). However, writing $u(t) = e^{it\Delta} u_0 + (u(t) - e^{it\Delta} u_0)$, the above is controlled by (taking $s = 1$, the worst case)

$$\sum_{k=1}^{\infty} \| \nabla_x P_{2^{3k/4}} u_0 \|_{L_x^2}^2 + \sum_{k=1}^{+\infty} \| \nabla_x P_{2^{3k/4}} (u(t) - e^{it\Delta} u_0) \|_{L_x^2}^2.$$

By (3-21) of Proposition 3.4,

$$\| \nabla_x u_0 \|_{L_x^2}^2 + \sum_{k=1}^{+\infty} 2^{-k/8} \lesssim 1.$$

In conclusion, for term F_1 we obtain

$$\int_0^1 | \langle D_x^s F_1(s), D_x^s w(s) \rangle_{L_x^2} | \, ds \lesssim 1.$$

We next address term F_2. Insert $\psi_2 \psi_1^{4/d+1} = \psi_2$, then apply (4-3) with $\alpha = s$ to obtain (in the worst case $s = 1$),

$$\| D_x^s F_2 \|_{L^1_{[0,1)} L_x^2} \lesssim \| (1-t)^{-3/4} \psi_2 |u|^{4/d} u \|_{L^1_{[0,1)} L_x^2} \lesssim \| (1-t)^{-3/4} \| \psi_1 u \|_{L_x^{2(4/d+1)}}^{4/d+1} \|_{L^1_{[0,1)}}.$$

We consider the cases $d = 1$ and $d = 2$ separately. When $d = 1$,

$$\| \psi_1 u \|_{L_x^{10}} \lesssim \| D_x^{2/5} \psi_1 u \|_{L_x^2} \lesssim 1,$$

by Lemma 4.1. Consequently,

$$\| D_x^s F_2 \|_{L^1_{[0,1)} L_x^2} \lesssim \| (1-t)^{-3/4} \|_{L^1_{[0,1)}} \lesssim 1.$$

On the other hand, when $d = 2$, we have

$$\| \psi_1 u \|_{L_x^6} \lesssim \| D_x^{2/3} \psi_1 u \|_{L_x^2} \lesssim \| D_x^{1/2} \psi_1 u \|_{L_x^2}^{2/3} \| \nabla_x \psi_1 u \|_{L_x^2}^{1/3} \lesssim (1-t)^{-1/6}$$

by Lemma 4.2 and (3-2). Consequently,

$$\| D_x^s F_2 \|_{L^1_{[0,1)} L_x^2} \lesssim \| (1-t)^{-3/4} (1-t)^{-1/6} \|_{L^1_{[0,1)}} \lesssim 1.$$

Next, we address term F_3. By (4-3) with $\alpha = 9/8$,

$$\| D_x^s F_3 \|_{L^1_{[0,1)} L_x^2} \lesssim \| (1-t)^{-27/32} \|_{L^1_{[0,1)}} \| D_x^{s-1/8} (\tilde{\psi}_2 u) \|_{L^\infty_{[0,1)} L_x^2}.$$

Since $\|(1-t)^{-27/32}\|_{L^1_{[0,1)}} \sim 1$ and the support of $\widetilde{\psi}_2$ is contained in the set where $\psi_1 = 1$, we have

$$\|D_x^s F_3\|_{L^1_{[0,1)} L^2_x} \lesssim \|\langle D_x\rangle^{s-1/8} \psi_1 u\|_{L^\infty_{[0,1)} L^2_x}.$$

Finally, we consider F_4. We have

$$\|D_x^s F_4\|_{L^1_{[0,1)} L^2_x} \lesssim \|\langle \nabla_x\rangle P_- \psi_1 u\|_{L^1_{[0,1)} L^2_x} \lesssim \|(1-t)^{-3/4}\|_{L^1_{[T_1,1)}} \|u\|_{L^\infty_{[0,1)} L^2_x} \lesssim 1$$

by (4-3) with $\alpha = 1$. □

Proposition 4.4. *Suppose that $u(t)$ solving* (1-1) *with H^1 initial data satisfies* (3-1). *Fix $R > 0$. Then*

$$\|u\|_{L^\infty_{[0,1)} H^1_{|x|\geq R}} \lesssim 1.$$

Proof. Iterate Lemma 4.3 eight times on successively larger external regions. □

Proposition 4.4 completes the proof of Theorem 1.1.

5. Application to 3D standing sphere blow-up

We now outline the proof of Theorem 1.2 utilizing the techniques of Section 3 and 4. Theorem 1.2 pertains to radial solutions of (1-9). We define the initial data set $\mathcal{P}$ as in[9] Raphaël and Szeftel [2009, Definition 1, page 980–1], except that condition (v) is replaced by $\|\tilde{u}_0\|_{H^1(|r-1|\geq 1/10)} \leq \epsilon^5$. The goal then becomes to complete the proof of the bootstrap Proposition 1 on page 982, where the "improved regularity estimates" (35)–(37) are effectively replaced with

$$\|u(t)\|_{L^\infty_{[0,t_1]} H^1_{|x|\leq 1/2}} \leq \epsilon.$$

Let us formulate a more precise statement:

Proposition 5.1 (partial bootstrap argument). *Let Q be the* 1*D ground state given by* (1-4), *and let $\epsilon > 0$, $T > 0$ be fixed with $T \leq \epsilon^{200}$. Suppose that $u(t)$ is a radial* 3*D solution to*

$$i\partial_t u + \Delta u + |u|^4 u = 0$$

on an interval $[0, T'] \subset [0, T)$ such that the following "bootstrap inputs" hold:

(1) *There exist parameters $\lambda(t) > 0$, $\gamma(t) \in \mathbb{R}$, and $|r(t) - 1| \leq 1/10$, such that if we define*

$$\tilde{u}(r,t) = u(r,t) - \frac{1}{\lambda(t)^{1/2}} Q\Big(\frac{r - r(t)}{\lambda(t)}\Big), \tag{5-1}$$

then, for $0 \leq t \leq T'$,

$$\|\nabla u(t)\|_{L^2_x} = \lambda(t)^{-1} \sim \Big(\frac{\log|\log(T-t)|}{T-t}\Big)^{1/2}, \tag{5-2}$$

and

$$\|\nabla \tilde{u}(t)\|_{L^2_x} \lesssim \frac{1}{|\log(T-t)|^{1+}(T-t)^{1/2}}. \tag{5-3}$$

[9]We are considering the case dimension $d = 3$ (in their notation $N = 3$).

(2) *Interior Strichartz control*: $\|\langle\nabla\rangle u(t)\|_{L^5_{[0,T']}L^{30/11}_{|x|\le 1/2}} \le \epsilon$.

(3) *Initial data remainder control*: $\|\langle\nabla\rangle \tilde{u}_0\|_{L^2_x} \le \epsilon^5$.

Then we have the following "bootstrap output":

$$\|\langle\nabla\rangle u(t)\|_{L^\infty_{[0,T']}L^2_{|x|\le 1/2}} + \|\langle\nabla\rangle u(t)\|_{L^5_{[0,T']}L^{30/11}_{|x|\le 1/2}} \lesssim \epsilon^5. \tag{5-4}$$

The goal of this section is to prove Proposition 5.1, which shows that the bootstrap input (2) is reinforced. Proposition 5.1 is, however, an incomplete bootstrap and by itself does not establish Theorem 1.2. The analysis which uses (5-4) to reinforce the bootstrap assumption (1) is rather elaborate but will be omitted here as it follows the arguments in [Raphaël 2006] and [Raphaël and Szeftel 2009]. Moreover, these papers demonstrate how the assertions in Theorem 1.2 follow.

The proof of Proposition 5.1 follows the methods developed in Section 3–4 used to prove Theorem 1.1. We do not, however, rescale the solution so that $T = 1$ as was done in Section 3.

Remark 5.2. Let us list some notational conventions for the rest of the section. We take $t_k = T - 2^{-k}$ and denote $I_k = [0, t_k]$. Let $v(r,t) = ru(r,t)$, and consider v as a 1D function in r extended to $r < 0$ as an odd function. Note that v solves

$$i\partial_t v + \partial_r^2 v = -r^{-4}|v|^4 v.$$

The frequency projection P_N will always refer to the 1D frequency projection in the r-variable. The Bourgain norm $\|v\|_{X_{s,b}}$ refers to the 1D norm in the r-variable.

Let $\lambda_0 = \lambda(0)$ and take $k_0 \in \mathbb{N}$ such that $2^{-k_0/2}(\log k_0)^{-1/2} \sim \lambda_0$. We then have $T \sim 2^{-k_0}$. The assumption $T \le \epsilon^{40}$ equates to $2^{-k_0/8} \le \epsilon^5$. Note that $\lambda(t_k) = 2^{-k/2}(\log k)^{-1/2}$.

Lemma 5.3 (smallness of initial data). *Under the assumption (3) in Proposition 5.1 on the initial data, and with $v_0 = ru_0$, we have*

$$\|P_{\ge 2^{3k_0/4}}\partial_r v_0\|_{L^2_r} + \|\partial_r v_0\|_{L^2_{r\le 1/2}} \lesssim \epsilon^5.$$

Proof. Let $\tilde{v}_0 = r\tilde{u}_0$. Since $\partial_r \tilde{v}_0 = \tilde{u}_0 + r\partial_r\tilde{u}_0$, we have by Hardy's inequality

$$\|\partial_r\tilde{v}_0\|_{L^2_r} \lesssim \||x|^{-1}\tilde{u}_0\|_{L^2_x} + \|\nabla\tilde{u}_0\|_{L^2_x} \lesssim \|\nabla\tilde{u}_0\|_{L^2_x} \lesssim \epsilon^5.$$

Recalling the definition of $\tilde{u}_0 = \tilde{u}(0)$ in (5-1) (with $t = 0$), we have

$$v_0 = \frac{r}{\lambda_0^{1/2}} Q\left(\frac{r - r_0}{\lambda_0}\right) + \tilde{v}_0.$$

The result then follows from the exponential localization and smoothness of Q. □

Lemma 5.4 (radial Strichartz). *Suppose that $u(t)$ is a 3D radial solution to*

$$i\partial_t u + \Delta u = f.$$

Let $v(r,t) = ru(r,t)$ and $g(r,t) = rf(r,t)$ and consider v as a 1D function in r (extended to be odd), so that

$$i\partial_t v + \partial_r^2 v = g.$$

Then for (q,r) *and* $(\tilde{q},\tilde{r})$ *satisfying the* 3D *admissibility condition*,

$$\|r^{2/p-1}v\|_{L_t^q L_r^p} \lesssim \|v_0\|_{L_r^2} + \|r^{2/p'-1}g\|_{L_t^{\tilde{q}'}L_r^{\tilde{p}'}}.$$

Proof. The left side is equivalent to $\|\nabla u\|_{L_t^q L_x^p}$ and the right side is equivalent to $\|u_0\|_{L_x^2} + \|f\|_{L_t^{\tilde{q}'}L_x^{\tilde{p}}}$, so it is just a restatement of the 3D Strichartz estimates. □

Lemma 5.5 (3D to 1D conversion). *Suppose that* $u(x)$ *is a* 3D *radial function*, *and write* $u(r) = u(x)$. *Let* $v(r) = ru(r)$. *Then for* $1 < p < 3$, *we have*

$$\|r^{2/p-1}\partial_r v\|_{L_r^p} \lesssim \|\nabla_x u\|_{L_x^p}. \tag{5-5}$$

Also for $\frac{3}{2} < p < +\infty$, *we have*

$$\|\nabla_x u\|_{L_x^p} \lesssim \|r^{2/p-1}\partial_r v\|_{L_r^p}. \tag{5-6}$$

Consequently, *for* 3D *admissible pairs* (q,p) *such that* $2 \le p < 3$, *we have*

$$\|\nabla u\|_{L_t^q L_x^p} \sim \|r^{2/p-1}\partial_r v\|_{L_t^q L_r^p}. \tag{5-7}$$

We remark that $q = 5$ and $p = \frac{30}{11}$ falls within the range of validity for (5-7).

Proof. The proof of (5-5) and (5-6) is a standard application of the Hardy inequality.

First, we prove (5-5). Using $v = ru$,

$$r^{2/p-1}\partial_r v = r^{2/p}\partial_r u + r^{2/p-1}u,$$

and thus,

$$\|r^{2/p-1}\partial_r v\|_{L_r^p} \le \|r^{2/p}\partial_r u\|_{L_r^p} + \|r^{2/p-1}u\|_{L_r^p}.$$

We have, for $r > 0$,

$$u(r) = -(u(+\infty) - u(r)) = \int_{s=1}^{+\infty} \frac{d}{ds}(u(sr))\,ds = \int_{s=1}^{+\infty} u'(sr)r\,ds.$$

By the Minkowski integral inequality,

$$\|r^{2/p-1}u\|_{L_r^p} \le \int_{s=1}^{+\infty} \|u'(sr)r^{2/p}\|_{L_{r>0}^p}\,ds.$$

Changing variable $r \mapsto s^{-1}r$, we obtain that the right-hand side is bounded by

$$\left(\int_{s=1}^{+\infty} s^{-3/p}\,ds\right)\|r^{2/p}u'\|_{L_{r>0}^p}$$

and the s integral is finite provided $p < 3$.

Next, we prove (5-6). We have

$$r^{2/p}\partial_r u = r^{2/p}\partial_r(r^{-1}v) = -r^{2/p-2}v + r^{2/p-1}\partial_r v,$$

and hence,

$$\|r^{2/p}\partial_r u\|_{L_r^p} \le \|r^{2/p-2}v\|_{L_r^p} + \|r^{2/p-1}\partial_r v\|_{L_r^p}.$$

We have

$$v(r) = v(r) - v(0) = \int_{s=0}^{1} \frac{d}{ds}(v(sr))\,ds = \int_{s=0}^{1} v'(sr)r\,ds.$$

By the Minkowski integral inequality,

$$\|r^{2/p-2}v\|_{L^p_r} \le \int_{s=0}^{1} \|v'(sr)r^{2/p-1}\|_{L^p_r}\,ds.$$

Changing variable $r \mapsto s^{-1}r$ in the right side, we obtain

$$\|r^{2/p-2}v\|_{L^p_r} \le \Big(\int_{s=0}^{1} s^{-3/p+1}ds\Big)\|v'(r)r^{2/p-1}\|_{L^p_r}$$

and the s integral is finite provided $p > \frac{3}{2}$. □

The replacement for Lemma 3.1 is Lemma 5.6 below. The difference is that in Lemma 5.6, we only use $b < \frac{1}{2}$ when working at $\dot{H}^1$ regularity.

Lemma 5.6. *Suppose that the assumptions of Proposition 5.1 and Remark 5.2 hold. Then for* $\frac{1}{2} - \delta \le b < \frac{1}{2}$,

$$\|\partial_r v\|_{X_{0,b}(I_k)} \lesssim 2^{kb}(\log k)^{b+1/2} = (T-t)^{-b}(\log|\log(T-t)|)^{b+1/2}. \tag{5-8}$$

Also, for $\frac{1}{2} - \delta < b < \frac{1}{2} + \delta$,

$$\|v\|_{X_{0,b}(I_k)} \lesssim_\delta 2^{k\delta} = (T-t)^{-\delta}. \tag{5-9}$$

Proof. We will only carry out the proof of (5-8), which stems from (5-2).[10] The proof of (5-9) is similar, and stems from the bound on $\|u(t)\|_{H^\delta}$ obtained from interpolation between (5-2) and mass conservation.

In the proof below, T has no relation to the T representing blow-up time in the rest of the article.

Let $\lambda = \lambda(t_k) = 2^{-k/2}(\log k)^{-1/2}$. Let $r = \lambda R$, $x = \lambda X$, and $t = \lambda^2 T + t_k$. Define the functions

$$V(R,T) = \lambda^{1/2}v(\lambda R, \lambda^2 T + t_k) = \lambda^{1/2}v(r,t),$$
$$U(X,T) = \lambda^{1/2}u(\lambda X, \lambda^2 T + t_k) = \lambda^{1/2}u(x,t).$$

Note that the identity $v(r) = ru(r)$ corresponds to $V(R) = \lambda R U(R)$.

We study $V(R,T)$ on $T \in [0, \log k]$, which corresponds to $t \in [t_k, t_{k+1}]$. We have $\|V\|_{L^2_R} = \|v\|_{L^2_r} \sim O(1)$ (by mass conservation) and $\|\partial_R V\|_{L^2_R} = \lambda\|\partial_r v\|_{L^2_r}$. Hence, $\|\partial_R V\|_{L^\infty_{[0,\log k]}L^2_R} = O(1)$. The equation satisfied by V is

$$i\partial_T V + \partial_R^2 V = -\lambda^{-4}R^{-4}|V|^4 V.$$

Let $J = [a,b]$ be a unit-sized time interval in $[0, \log k]$. Then by Lemma 2.4,

$$\|\partial_R V\|_{X_{0,b}(J)} \lesssim \|\partial_R V(a)\|_{L^2} + \|\partial_R(\lambda^{-4}R^{-4}|V|^4 V)\|_{L^1_J L^2_R}.$$

[10] The need to take $b < 1/2$ comes from Lemma 2.4, (2-7) versus (2-8); when working at $\dot{H}^1$ regularity near the origin, we cannot suffer any loss of derivatives. The fact that $\|\partial_r v\|_{X_{0,b}(I_k)}$ for $b < 1/2$ is only a $\dot{H}^1$ subcritical quantity is of no harm as the only application of (5-8) in the subsequent arguments is to control the solution for $r \ge 1/2$, where the equation is effectively L^2 critical.

Let $\chi_1(r)=1$ for $r\le\frac14$ and $\operatorname{supp}\chi_1\subset B(0,\frac38)$. Let $\chi_2=1-\chi_1$. Let $g_1=\partial_R(\lambda^{-4}R^{-4}\chi_1(\lambda R)|V|^4V)$ and $g_2=\partial_R(\lambda^{-4}R^{-4}\chi_2(\lambda R)|V|^4V)$, so that the above becomes

$$\|\partial_R V\|_{X_{0,b}(J)}\lesssim\|\partial_R V(a)\|_{L^2}+\|g_1\|_{L^1_JL^2_R}+\|g_2\|_{L^1_JL^2_R}. \tag{5-10}$$

We begin with estimating $\|g_2\|_{L^1_JL^2_R}$. We have

$$\|g_2\|_{L^1_JL^2_R}\lesssim\|V^5\|_{L^1_JL^2_R}+\|V^4(\partial_R V)\|_{L^1_JL^2_R}. \tag{5-11}$$

We now treat the first term in (5-11). Of course, $\|V^5\|_{L^1_JL^2_R}=\|V\|^5_{L^5_JL^{10}_R}$. By Sobolev embedding $\|V\|_{L^{10}_R}\lesssim\|D_R^{2/5}V\|_{L^2_R}$ and by Hölder,

$$\begin{aligned}\|V\|_{L^5_JL^{10}_R}\lesssim|J|^{1/10}\|D_R^{2/5}V\|_{L^{10}_JL^2_R}&\lesssim|J|^{1/10}(\|V\|_{L^{10}_JL^2_R}+\|\partial_R V\|_{L^{10}_JL^2_R})\\&\le|J|^{1/10}(|J|^{1/10}\,\|V\|_{L^\infty_JL^2_R}+\|\partial_R V\|_{L^{10}_JL^2_R}).\end{aligned}$$

Using that $\|V\|_{L^\infty_JL^2_R}\sim1$, that $|J|\sim1$ and Lemma 2.7, provided $\frac25<b<\frac12$, we have

$$\|V\|_{L^5_JL^{10}_R}\lesssim|J|^{1/10}(1+\|\partial_R V\|_{X_{0,b}}). \tag{5-12}$$

We now treat the second term in (5-11), similarly estimating the term $\|V\|_{L^{10}_R}$. We have

$$\begin{aligned}\|V^4\partial_R V\|_{L^1_JL^2_R}&\lesssim|J|^{7/20}\|V\|^4_{L^{10}_JL^{10}_R}\|\partial_R V\|_{L^4_JL^{10}_R}\\&\lesssim|J|^{7/20}(1+\|\partial_R V\|_{L^{10}_JL^2_R})^4\|\partial_R V\|_{L^4_JL^{10}_R}.\end{aligned}$$

Appealing to Lemma 2.7, provided $\frac{9}{20}<b<\frac12$, we obtain

$$\|V^4\partial_R V\|_{L^1_JL^2_R}\lesssim|J|^{7/20}(1+\|\partial_R V\|_{X_{0,b}})^5. \tag{5-13}$$

Combining (5-12) and (5-13), we have

$$\|g_2\|_{L^1_JL^2_R}\lesssim|J|^{7/20}(1+\|\partial_R V\|_{X_{0,b}})^5. \tag{5-14}$$

Next we estimate $\|g_1\|_{L^1_JL^2_R}$. By rescaling,

$$\|g_1\|_{L^1_JL^2_R}=\lambda\|\partial_r(\chi_1r^{-4}|v|^4v)\|_{L^1_{[t_k,t_{k+1}]}L^2_r}.$$

Let $w=\tilde\chi_1u$, where $\tilde\chi_1=1$ on $\operatorname{supp}\chi_1$ but $\operatorname{supp}\tilde\chi_1\subset B(0,\frac12)$. Replacing $u=r^{-1}v$, we obtain $\partial_r(r\chi_1u^5)=\partial_r(r\chi_1w^5)$, and hence,

$$\|g_1\|_{L^2_R}\lesssim\lambda(\|w\|^5_{L^{10}_r}+\|rw^4\partial_r w\|_{L^2_r})\lesssim\lambda(\||x|^{-1/5}w\|^5_{L^{10}_x}+\|w^4\nabla w\|_{L^2_x}). \tag{5-15}$$

By Hardy's inequality and 3D Sobolev embedding,

$$\||x|^{-1/5}w\|_{L^{10}_x}\lesssim\|D_x^{1/5}w\|_{L^{10}_x}\lesssim\|\nabla w\|_{L^{30/11}_x}.$$

By Hölder's inequality and 3D Sobolev embedding,

$$\|w^4\nabla w\|_{L^2_x}\le\|w\|^4_{L^{30}_x}\|\nabla w\|_{L^{30/11}_x}\lesssim\|\nabla w\|^5_{L^{30/11}_x}.$$

Returning to (5-15) and invoking (2) of Proposition 5.1,

$$\|g_1\|_{L^1_{I_k}L^2_r} \lesssim \lambda \|\nabla w\|^5_{L^5_{I_k}L^{30/11}_x} \lesssim \lambda\epsilon^5. \tag{5-16}$$

By putting (5-14) and (5-16) into (5-10), we obtain

$$\|\partial_R V\|_{X_{0,b}(J)} \lesssim \|\partial_R V(a)\|_{L^2} + |J|^{7/20}(1+\|\partial_R V\|_{X_{0,b}(J)})^5 + \lambda\epsilon^5.$$

From this, we conclude that we can take $|J|$ sufficiently small (but still "unit-sized"[11]) so that it follows that

$$\|\partial_R V\|_{X_{0,b}(J)} \le O(1).$$

Square summing over unit-sized intervals J filling $[0, \log k]$,

$$\|\partial_R V\|_{X_{0,b}([0,\log k])} \lesssim (\log k)^{1/2}.$$

This estimate scales back to

$$\|\partial_r v\|_{X_{0,b}([t_k,t_{k+1}])} \lesssim (\log k)^{1/2}\lambda(t_k)^{-2b} = 2^{kb}(\log k)^{b+1/2}.$$

Now square sum over k from $k=0$ to $k=K$ to obtain a bound of $2^{Kb}(\log K)^{b+1/2}$ over the time interval I_K, which is the claimed estimate (5-8). □

The analogue of Lemma 3.2 will be Lemma 5.7 below. We note that as a consequence of Lemma 5.6, the hypothesis of Lemma 5.7 below is satisfied with $\alpha(k,N) = 2^{-k/2}N^{-1}$.

Lemma 5.7 (high-frequency recurrence). *Let the assumptions of Proposition 5.1 and Remark 5.2 hold, and let*[12]

$$\beta(k,N) := \|P_{\ge N}\partial_r v\|_{X_{0,1/2-}(I_k)}.$$

Then there exists an absolute constant $0<\mu\ll 1$ *such that for* $N \ge 2^{k(1+\delta)/2}$, *we have*

$$\beta(k,N) + \|r^{2/p-1}P_{\ge N}\partial_r v\|_{L^q_{I_k}L^p_r} \lesssim \|P_{\ge N}\partial_r v_0\|_{L^2_r} + 2^{k(1+\delta)/2}N^{-1+\delta}\beta(k,\mu N) + N^{-1+\delta}2^{k\delta}\beta(k,\mu N)^2 + 2^{-k\delta} + \epsilon^5 \tag{5-17}$$

for all 3D *admissible* (q,p).

Proof. Note that v solves

$$i\partial_t v + \partial_r^2 v = -r|u|^4 u = -r^{-4}|v|^4 v.$$

Let $\chi_1(r)$ be a smooth function such that $\chi_1(r)=1$ for $|r|\le \frac14$ and χ_1 is supported in $|r|\le\frac38$. Let $\chi_2 = 1-\chi_1$. Apply $P_{\ge N}\partial_r$ to obtain

$$(i\partial_t + \partial_r^2)P_{\ge N}\partial_r v = g_1 + g_2,$$

[11] Meaning: with size independent of any small parameters like ϵ or λ

[12] Note the inclusion of one derivative in the definition of β, in contrast to the choice of definition for α in Proposition 3.4.

where

$$g_j(r) = -P_{\geq N}\partial_r(\chi_j\, r^{-4}\,|v|^4 v) \quad \text{for } j = 1, 2.$$

Then by Lemma 2.4[13] and Lemma 5.4,

$$\|P_{\geq N}\partial_r v\|_{X_{0,1/2-}(I_k)} + \|r^{2/p-1}P_{\geq N}\partial_r v\|_{L^q_{I_k}L^p_r} \lesssim \|P_{\geq N}\partial_r v_0\|_{L^2_r} + \|g_1\|_{L^1_{I_k}L^2_r} + \|g_2\|_{L^1_{I_k}L^2_r}.$$

The term $\|g_2\|_{L^1_t L^2_r}$ is controlled in a manner similar to the analysis in the proof of Lemma 3.2. For this term, $\chi_2\, r^{-4}$ and $\partial_r(\chi_2\, r^{-4})$ are smooth bounded functions, with all derivatives bounded. By Lemma 2.10,

$$\|g_2\|_{L^2_r} \lesssim \|P_{\geq N}\langle\partial_r\rangle v^5\|_{L^2_r} + N^{-1}\|\langle\partial_r\rangle v^5\|_{L^2_r}. \tag{5-18}$$

By an analysis similar to the proof of Lemma 3.2, utilizing the bounds in Lemma 5.6, we obtain

$$\|P_{\geq N}\langle\partial_r\rangle v^5\|_{L^1_{I_k}L^2_r} \lesssim 2^{k(1+\delta)/2}N^{-1+\delta}\beta(k,\mu N) + N^{-1+\delta}2^{k\delta}\beta(k,\mu N)^2. \tag{5-19}$$

Also by the Strichartz estimates, as in the proof of Lemma 5.6 above,

$$\|\langle\partial_r\rangle v^5\|_{L^1_{I_k}L^2_r} \lesssim \|D^\delta v\|^4_{X_{0,b}}\|\partial_R v\|_{X_{0,b}} \lesssim 2^{k(1+\delta)/2}. \tag{5-20}$$

Inserting (5-19) and (5-20) into (5-18), we obtain

$$\|g_2\|_{L^1_{I_k}L^2_r} \lesssim 2^{k(1+\delta)/2}N^{-1+\delta}\beta(k,\mu N) + N^{-1+\delta}2^{k\delta}\beta(k,\mu N)^2 + N^{-1}2^{k(1+\delta)/2}. \tag{5-21}$$

The last term, $N^{-1}2^{k(1+\delta)/2}$, gives the contribution $2^{-k\delta}$ in (5-17) due to the restriction $N \geq 2^{k(1+\delta)/2}$ (different deltas).

Next we address $\|g_1\|_{L^1_{I_k}L^2_r}$. We estimate away $P_{\geq N}$ by

$$\|g_1\|_{L^1_{I_k}L^2_r} \lesssim \|\tilde{g}_1\|_{L^1_{I_k}L^2_r}, \tag{5-22}$$

where (ignoring complex conjugates)

$$\tilde{g}_1 = \partial_r(r^{-4}\chi_1 v^5).$$

Let $w = \tilde{\chi}_1 u$, where $\tilde{\chi}_1 = 1$ on $\operatorname{supp}\chi_1$ but $\operatorname{supp}\tilde{\chi}_1 \subset B(0,\tfrac{1}{2})$. Replacing $u = r^{-1}v$, we obtain $\tilde{g}_1 = \partial_r(r\chi_1 u^5) = \partial_r(r\chi_1 w^5)$, and hence,

$$\|\tilde{g}_1\|_{L^2_r} \lesssim \|w\|^5_{L^{10}_r} + \|rw^4\partial_r w\|_{L^2_r} \lesssim \||x|^{-1/5}w\|^5_{L^{10}_x} + \|w^4\nabla w\|_{L^2_x}.$$

By Hardy's inequality and 3D Sobolev embedding,

$$\||x|^{-1/5}w\|_{L^{10}_x} \lesssim \|D_x^{1/5}w\|_{L^{10}_x} \lesssim \|\nabla w\|_{L^{30/11}_x}.$$

By Hölder's inequality and 3D Sobolev embedding,

$$\|w^4\nabla w\|_{L^2_x} \leq \|w\|^4_{L^{30}_x}\|\nabla w\|_{L^{30/11}_x} \lesssim \|\nabla w\|^5_{L^{30/11}_x}.$$

[13] We were able to obtain the $L^1_{I_k}L^2_r$ right side (without δ loss), because we took $b < 1/2$ in the Bourgain norm.

Hence, $\|\tilde{g}_1\|_{L_r^2} \lesssim \|\nabla w\|^5_{L_x^{30/11}}$. Returning to (5-22) and invoking (2) of Proposition 5.1,

$$\|g_1\|_{L^1_{I_k} L^2_r} \lesssim \|\nabla w\|^5_{L^5_{I_k} L_x^{30/11}} \lesssim \epsilon^5. \qquad \square$$

The analogue of Proposition 3.4 is this:

Proposition 5.8 (high-frequency control). *Let the assumptions of Proposition 5.1 and Remark 5.2 hold. Then for any* 3D *Strichartz admissible pair* (q, p), *we have*

$$\|P_{\geq 2^{3k/4}} \partial_r v\|_{X_{0,1/2-}(I_k)} + \|r^{2/p-1} P_{\geq 2^{3k/4}} \partial_r v\|_{L^q_{I_k} L^p_r} \lesssim \epsilon^5.$$

Proof. Several applications of Lemma 5.7, just as Proposition 3.4 is deduced from Lemma 3.2. $\square$

Due to the $\dot{H}^1$ criticality of the problem, we do not have improved regularity of $v(t) - e^{it\partial_r^2} v_0$ as was the case in Proposition 3.4. As a substitute, we can use the methods of Lemma 5.7 to obtain the following lemma:

Lemma 5.9 (additional high-frequency control). *Suppose that the assumptions of Proposition 5.1 and Remark 5.2 hold. Then*

$$\left(\sum_{k=k_0}^{+\infty} \|P_{2^{3k/4}} \partial_r v\|^2_{L^\infty_{[t_{k-1}, t_k]} L^2_r} \right)^{1/2} \lesssim \epsilon^5. \tag{5-23}$$

Proof. It suffices to prove the estimate with the sum terminating at $k = K$, provided we obtain a bound independent of K. For each k in $k_0 \leq k \leq K$, write the integral equation on I_k. For $t \in [t_{k-1}, t_k]$

$$v(t) = e^{it\partial_r^2} v_0 - i \int_0^t e^{i(t-t')\partial_r^2} (r^{-4} |v|^4 v(t'))\, dt'.$$

Apply $P_{2^{3k/4}} \partial_r$ to obtain

$$P_{2^{3k/4}} \partial_r v(t) = P_{2^{3k/4}} e^{it\partial_r^2} \partial_r v_0 - i \int_0^t e^{i(t-t')\partial_r^2} P_{2^{3k/4}} \partial_r (r^{-4} |v|^4 v(t'))\, dt'.$$

Estimate

$$\|P_{2^{3k/4}} \partial_r v\|_{L^\infty_{[t_{k-1}, t_k]} L^2_r} \leq \|P_{2^{3k/4}} \partial_r v_0\|_{L^2_r} + \|P_{2^{3k/4}} \partial_r (r^{-4} |v|^4 v)\|_{L^1_{I_k} L^2_r}.$$

By the inequality $(a+b)^2 \leq 2a^2 + 2b^2$, this implies

$$\|P_{2^{3k/4}} \partial_r v\|^2_{L^\infty_{[t_{k-1}, t_k]} L^2_r} \lesssim \|P_{2^{3k/4}} \partial_r v_0\|^2_{L^2_r} + \|P_{2^{3k/4}} \partial_r (r^{-4} |v|^4 v)\|^2_{L^1_{I_k} L^2_r}.$$

Let $\chi_1(r)$ be a smooth function such that $\chi_1(r) = 1$ for $|r| \leq \frac{1}{4}$ and χ_1 is supported in $|r| \leq \frac{3}{8}$. Let $\chi_2 = 1 - \chi_1$. Let $g_j = P_{2^{3k/4}} \partial_r (\chi_j r^{-4} |v|^4 v)$ for $j = 1, 2$.

Recall that in the proof of Lemma 5.7, we showed that

$$\|P_{\geq N} \partial_r \chi_2 r^{-4} |v|^4 v\|_{L^1_{I_k} L^2_r} \lesssim 2^{k(1+\delta)/2} N^{-1+\delta} \beta(k, \mu N) + N^{-1+\delta} 2^{k\delta} \beta(k, \mu N)^2 + N^{-1} 2^{k(1+\delta)/2},$$

and Proposition 5.8 showed that $\beta(k, 2^{3k/4}) \lesssim 1$. Combining gives $\|g_2\|_{L^1_{I_k} L^2_r} \lesssim 2^{-k/8}$, and hence,

$$\left(\sum_{k=k_0}^{K} \|g_2\|^2_{L^1_{I_k} L^2_r} \right)^{1/2} \lesssim 2^{-k_0/8} \le \epsilon^5.$$

Now we address g_1. Let $w = \tilde{\chi}_1 u$. For each k, lengthen I_k to $I := I_K$ to obtain

$$\sum_{k=k_0}^{K} \|g_1\|^2_{L^1_{I_k} L^2_r} \lesssim \|P_{2^{3k/4}} \partial_r (r^{-4} \chi_1 |w|^4 w)\|^2_{\ell^2_k L^1_I L^2_r}.$$

By the Minkowski inequality, for any space-time function F, we have

$$\|P_{2^{3k/4}} F\|_{\ell^2_k L^1_I L^2_r} \le \|P_{2^{3k/4}} F\|_{L^1_I \ell^2_k L^2_r} \lesssim \|F\|_{L^1_I L^2_r}.$$

Hence,

$$\sum_{k=k_0}^{K} \|g_1\|^2_{L^1_{I_k} L^2_r} \lesssim \|\partial_r (\chi_1 r^{-4} |w|^4 w)\|^2_{L^1_I L^2_r}.$$

At this point we proceed as in Lemma 5.7 to obtain a bound by ϵ^5. □

Now we begin to insert spatial cutoffs away from the blow-up core and obtain the missing low frequency bounds. The first step is to obtain a little regularity above L^2, since it is needed in the proof of Lemma 5.11.

Lemma 5.10 (small regularity gain). *Suppose that the assumptions of Proposition 5.1 and Remark 5.2 hold. Let $\psi_{3/4}(r)$ be a smooth function such that $\psi_{3/4}(r) = 1$ for $|r| \le \frac{3}{4}$ and $\psi_{3/4}(r) = 0$ for $|r| \ge \frac{7}{8}$. Then*

$$\|\langle D_r \rangle^{3/7} \psi_{3/4} v\|_{L^\infty_{[0,T)} L^2_r} \lesssim \epsilon^5.$$

Proof. Taking $\psi = \psi_{3/4}$, let $w = \psi v$. Then

$$\begin{aligned} i\partial_t w + \partial_r^2 w &= \psi (i\partial_t + \partial_r^2) v + 2\partial_r (\psi' v) - \psi'' v \\ &= -r^{-4} \psi |v|^4 v + 2\partial_r(\psi' v) - \psi'' v = F_1 + F_2 + F_3. \end{aligned}$$

Local smoothing and energy estimates provide the estimate

$$\begin{aligned} &\|D_r^{3/7} w\|_{L^\infty_{[0,T)} L^2_r} \\ &\qquad \lesssim \|D_r^{3/7} w_0\|_{L^2_r} + \|D_r^{3/7} F_1\|_{L^1_{[0,T)} L^2_r} + \|D_r^{-1/2} D_r^{3/7} F_2\|_{L^2_{[0,T)} L^2_r} + \|D_r^{3/7} F_3\|_{L^1_{[0,T)} L^2_r}. \end{aligned} \tag{5-24}$$

We begin with the F_1 estimate. Let $\tilde{\psi}$ be a smooth function such that

$$\tilde{\psi}(r) = \begin{cases} 0 & \text{if } r \le \frac{1}{4}, \\ 1 & \text{if } \frac{1}{2} \le r \le \frac{7}{8}, \\ 0 & \text{if } r \ge \frac{7}{8}. \end{cases}$$

Let $q = r^{-1} \tilde{\psi} v$. By writing $1 = (1 - \tilde{\psi}^4) + \tilde{\psi}^4$, we obtain

$$F_1 = -(1 - \tilde{\psi}^4) \psi r^{-4} |v|^4 v - |q|^4 w.$$

Note that $(1-\tilde{\psi}^4)\psi$ is supported in $|r| \le \frac{1}{2}$ and $\tilde{\psi}^4\psi$ is supported in $\frac{1}{4} \le |r| \le \frac{15}{16}$.

For the term $(1-\tilde{\psi}^4)\psi r^{-4}|v|^4 v$, we appeal to the bootstrap hypothesis (2) in the same way we did in the proof of Lemma 5.7 to obtain a bound by ϵ^5. As for the term $|q|^4 w$, by the fractional Leibniz rule,

$$\|D_r^{3/7}(|q|^4 w)\|_{L^1_{[0,T)}L^2_r} \lesssim \|D_r^{3/7}|q|^4\|_{L^1_{[0,T)}L^{7/3}_r}\|w\|_{L^\infty_{[0,T)}L^{14}_r} + \||q|^4\|_{L^1_{[0,T)}L^\infty_r}\|D_r^{3/7}w\|_{L^\infty_{[0,T)}L^2_r}.$$

By Sobolev embedding and Gagliardo–Nirenberg,

$$\|D_r^{3/7}|q|^4\|_{L^{7/3}_r} + \||q|^4\|_{L^\infty_r} \lesssim \|q\|^2_{L^2_r}\|\partial_r q\|^2_{L^2_r} \quad \text{and} \quad \|w\|_{L^{14}_r} \lesssim \|D_r^{3/7}w\|_{L^2_r}.$$

Hence,

$$\|D_r^{3/7}(|q|^4 w)\|_{L^1_{[0,T)}L^2_r} \lesssim \|q\|^2_{L^\infty_{[0,T)}L^2_r}\|\partial_r q\|^2_{L^2_{[0,T)}L^2_r}\|D_r^{3/7}w\|_{L^\infty_{[0,T)}L^2_r}.$$

By (5-3), $\|\partial_r q\|_{L^2_{[0,T)}L^2_r} \lesssim (|\log T|)^{-1} \lesssim (\log \epsilon^{-1})^{-1}$. Consequently, we obtain

$$\|D_r^{3/7}F_1\|_{L^1_{[0,T)}L^2_r} \lesssim \epsilon^5 + (\log\epsilon^{-1})^{-1}\|D_r^{3/7}w\|_{L^\infty_{[0,T)}L^2_r}.$$

As for F_2, we start by bounding

$$\|D_r^{-1/2}D_r^{3/7}F_2\|_{L^2_{[0,T)}L^2_r} \lesssim \|D_r^{13/14}(\psi' v)\|_{L^2_{[0,T)}L^2_r}.$$

On the support of ψ', we have $v = rq$. Noting that on the support of ψ' we have $r \sim 1$ and using the interpolation, we get

$$\|D_r^{13/14}(\psi' rq)\|_{L^2_r} \lesssim \|q\|_{L^2_r} + \|q\|^{1/14}_{L^2_r}\|\partial_r q\|^{13/14}_{L^2_r}.$$

By (5-3),

$$\|\|\partial_r q\|^{13/14}_{L^2_r}\|_{L^2_{[0,T)}} \lesssim T^{1/28} \lesssim \epsilon^5.$$

Consequently,

$$\|D_r^{-1/2}D_r^{3/7}F_2\|_{L^2_{[0,T)}L^2_r} \lesssim T^{1/2} + T^{1/28} \lesssim \epsilon^5.$$

Finally, for the term F_3, we estimate

$$\|D_r^{3/7}F_3\|_{L^1_{[0,T)}L^2_r} \lesssim \|q\|_{L^1_{[0,T)}L^2_r} + \|\partial_r q\|_{L^1_{[0,T)}L^2_r} \lesssim T + T^{1/2} \lesssim \epsilon^5.$$

Collecting the above estimates and inserting into (5-24), we obtain

$$\|D_r^{3/7}w\|_{L^2_{[0,T)}L^2_r} \lesssim \|D_r^{3/7}w_0\|_{L^2_r} + (\log\epsilon^{-1})^{-1}\|D_r^{3/7}w\|_{L^\infty_{[0,T)}L^2_r} + \epsilon^5,$$

and the result follows (by bootstrap assumption (3), $\|D_r^{3/7}w_0\|_{L^2_r} \lesssim \epsilon^5$). □

We will need to apply the following lemma eight times in the proof of Proposition 5.12 below. As in Section 4, the use of the frequency projection $P_{\lesssim (T-t)^{-3/4}}$ and the process of exchanging derivatives for time factors via (5-25) is essentially an appeal to the finite speed of propagation for low frequencies.

Lemma 5.11 (low frequency recurrence). *Let the assumptions of Proposition 5.1 and Remark 5.2 hold. Let $\frac{5}{8} < r_1 < r_2 < \frac{3}{4}$ and $\frac{1}{8} \le s \le 1$. Let $\psi_1(r)$ and $\psi_2(r)$ be smooth cutoff functions such that*

$$\psi_1(r) = \begin{cases} 1 & \text{on } |r| \le r_1, \\ 0 & \text{on } |r| \ge \frac{1}{2}(r_1+r_2) \end{cases} \quad \text{and} \quad \psi_2(r) = \begin{cases} 1 & \text{on } |r| \le \frac{1}{2}(r_1+r_2), \\ 0 & \text{on } |r| \ge r_2. \end{cases}$$

Then

$$\|D_r^s(\psi_1 v)\|_{L^\infty_{[0,T)}L^2_r} \lesssim \|D_r^{s-1/8}(\psi_2 v)\|_{L^\infty_{[0,T)}L^2_r} + \epsilon^5.$$

Proof. Let $\chi(\xi) = 1$ for $|\xi| \le 1$ and $\chi(\xi) = 0$ for $|\xi| \ge 2$ be a smooth function. Let $P = P_{\le (T-t)^{-3/4}}$ be the time-dependent multiplier operator defined by $\widehat{Pf}(\xi) = \chi((T-t)^{3/4}\xi)\hat{f}(\xi)$ (where Fourier transform is in space only). Note that the Fourier support of P at time $T-t = 2^{-k}$ is $\lesssim 2^{3k/4}$. We further have that

$$\partial_t Pf = \tfrac{3}{4} i (T-t)^{-1/4} Q \partial_r f + P \partial_t f,$$

where $Q = Q_{(T-t)^{-3/4}}$ is the time-dependent multiplier

$$\widehat{Qh}(\xi) = \chi'((T-t)^{3/4}\xi)\,\widehat{h}(\xi).$$

Note that the Fourier support of Q at time $t = T - 2^{-k}$ is $\sim 2^{3k/4}$. Note also that if $g = g(r)$ is any function, then

$$\|P D_r^\alpha g\|_{L^2_r} \le (T-t)^{-3\alpha/4}\|g\|_{L^2_r}. \tag{5-25}$$

Let $\tilde{\psi}$ be a smooth function such that

$$\tilde{\psi}(r) = \begin{cases} 0 & \text{if } |r| \le \frac{1}{4}, \\ 1 & \text{if } \frac{1}{2} \le |r| \le \frac{1}{2}(r_1+r_2), \\ 0 & \text{if } |r| \ge r_2. \end{cases}$$

Let $w = P_{\le (T-t)^{-3/4}} D_r^s(\psi_1 v)$. By Proposition 5.8, it suffices to show that

$$\|w\|_{L^\infty_{[0,T)}L^2_r} \lesssim \|D_r^{s-1/8}(\psi_2 v)\|_{L^\infty_{[0,T)}L^2_r} + \epsilon^5.$$

Note that w solves

$$\begin{aligned} i\partial_t w + \partial_r^2 w &= -\tfrac{3}{4}(T-t)^{-1/4} Q \partial_r D_r^s(\psi_1 v) - P D_r^s(\psi_1 r^{-4}|v|^4 v) + 2P\partial_r D_r^s(\psi_1' v) - P D_r^s(\psi_1'' v) \\ &= F_1 + F_2 + F_3 + F_4. \end{aligned}$$

By the energy method, we obtain

$$\|w\|^2_{L^\infty_t L^2_r} \le \|w_0\|^2_{L^2_r} + \int_0^T |\langle F_1, w\rangle_{L^2_r}| + 10\sum_{j=2}^4 \|F_j\|^2_{L^1_{[0,T)}L^2_r}.$$

We estimate F_1 using Lemma 5.9 as follows.[14] Let $\tilde{Q}$ be a projection onto frequencies of size $\sim (T-t)^{-3/4}$ (importantly, *not* $\lesssim (T-t)^{-3/4}$). Then

$$\int_0^T |\langle F_1, w\rangle_{L_r^2}| \lesssim \int_0^T (T-t)^{-1/4} \|\tilde{Q} D_r^{1/2+s}(\psi_1 v)\|_{L_r^2}^2.$$

It suffices to take $s=1$, the worst case. The presence of $\tilde{Q}$ allows for the exchange $D_r^{1/2} \sim (T-t)^{-3/8}$, which gives

$$\int_0^T |\langle F_1, w\rangle_{L_r^2}| \lesssim \int_0^T (T-t)^{-1} \|\tilde{Q}\partial_r(\psi_1 v)\|_{L_r^2}^2.$$

By decomposing $[0,T) = \bigcup_{k=k_0}^{\infty}[t_k, t_{k+1}]$, and using that $(T-t)^{-1} = 2^k$ on $[t_k, t_{k+1}]$, we have

$$\int_0^T (T-t)^{-1}\|\tilde{Q}\partial_r(\psi_1 v)\|_{L_r^2}^2 = \sum_{k=k_0}^{\infty} \int_{[t_k,t_{k+1}]} 2^k \|P_{2^{3k/4}}\partial_r(\psi_1 v)\|_{L_r^2}^2.$$

Since $|[t_k, t_{k+1}]| = 2^{-k}$, the above is controlled by $\sum_{k=k_0}^{\infty} \|P_{2^{3k/4}}\partial_r(\psi_1 v)\|_{L^\infty_{[t_k,t_{k+1}]}L_r^2}^2$, the square root of which is bounded by ϵ^5 (by Lemma 5.9).

For the nonlinear term F_2, by writing $1 = 1 - \tilde{\Psi}^4 + \tilde{\Psi}^4$, we have

$$F_2 = -PD_r^s(r^{-4}(1-\tilde{\Psi}^4)\psi_1|v|^4 v) - PD_r^s(r^{-4}\tilde{\Psi}^4\psi_1|v|^4 v) = F_{21} + F_{22}.$$

The support of $(1-\tilde{\Psi}^4)\psi_1$ is contained in $|r| \le \frac{1}{2}$, and we can use the bootstrap hypothesis (2) to obtain

$$\|F_{21}\|_{L^1_{[0,T)}L_r^2} \lesssim \epsilon^5,$$

as was done in the proof of Lemma 5.7 (for any $s \le 1$). For F_{22}, taking $\tilde{v} = \psi_2 v$ and noting that $\psi_1\psi_2 = \psi_1$, we have $F_{22} = PD_r^s(r^{-4}\tilde{\Psi}^4\psi_1|\tilde{v}|^4\tilde{v})$. By (5-25) with $\alpha = \frac{1}{8}$,

$$\|F_{22}\|_{L^1_{[0,T)}L_r^2} \le \big\|(T-t)^{-3/32}\|D_r^{s-1/8}(r^{-4}\tilde{\Psi}^4\psi_1|\tilde{v}|^4\tilde{v})\|_{L_r^2}\big\|_{L^1_{[0,T)}}.$$

Since $\tilde{\Psi}$ is supported in $\frac{1}{4} \le |r| \le r_2$, the function $\tilde{\Psi}^4\psi_1 r^{-4}$ is smooth and compactly supported. By the fractional Leibniz rule,

$$\|D_r^{s-1/8}(r^{-4}\tilde{\Psi}^4\psi_1|\tilde{v}|^4\tilde{v})\|_{L_r^2} \lesssim \|\tilde{v}\|_{L_r^\infty}^4 \|\langle D_r\rangle^{s-1/8}\tilde{v}\|_{L_r^2} \lesssim \|D_r^{3/7}\tilde{v}\|_{L_r^2}^{7/2}\|\partial_r\tilde{v}\|_{L_r^2}^{1/2}\|\langle D_r\rangle^{s-\frac{1}{8}}\tilde{v}\|_{L_r^2}.$$

Using the bound $\|\partial_r\tilde{v}\|_{L_r^2} \le (T-t)^{-1/2}$ from (5-3) and the bound on $\|D_r^{3/7}\tilde{v}\|_{L^\infty_{[0,T)}L_r^2}$ from Lemma 5.10, we obtain

$$\|F_{22}\|_{L^1_{[0,T)}L_r^2} \lesssim \|(T-t)^{-3/32}(T-t)^{-1/4}\|_{L^1_{[0,T)}}\|\langle D_r\rangle^{s-1/8}\tilde{v}\|_{L^\infty_{[0,T)}L_r^2} \lesssim \epsilon^5\|\langle D_r\rangle^{s-1/8}\tilde{v}\|_{L^\infty_{[0,T)}L_r^2}.$$

To bound F_3, we use (5-25) with $\alpha = \frac{9}{8}$ to obtain

$$\|F_3\|_{L^1_{[0,T)}L_r^2} \lesssim \|(T-t)^{-27/32}\|_{L^1_{[0,T)}}\|D_r^{s-1/8}\tilde{v}\|_{L^\infty_{[0,T)}L_r^2}.$$

[14] It seems that the energy method is needed here, since it furnishes $\int_0^T |\langle F_1, w\rangle_{L_r^2}|$; we cannot see a way to estimate $\|F_1\|_{L^1_{[0,T)}L_r^2}$. Indeed, by pursuing the method here, one ends up with a bound $\|F_1\|_{L^1_{[0,T)}L_r^2} \lesssim \sum_{k=k_0}^{\infty}\|P_{2^{3k/4}}\psi_1 v\|_{L_r^2}$, which is not controlled by Lemma 5.9, since it is not a *square* sum.

The F_4 term is more straightforward than F_3, since there is one fewer derivative. □

The H^1 control will complete part of the bootstrap estimate (5-4) in Proposition 5.1:

Proposition 5.12 (H^1 control). *Suppose that the assumptions of Proposition 5.1 and Remark 5.2 hold. Then*

$$\|\partial_r v\|_{L^\infty_{[0,T)} L^2_{|r|\le 5/8}} \lesssim \epsilon^5.$$

Proof. Let $r_k = \frac{5}{8} + \frac{1}{64}(k-1)$. Apply Lemma 5.11 on $[r_k, r_{k+1}]$ for $k = 1, \dots, 8$ to obtain collectively by Lemma 5.10 that

$$\|\partial_r v\|_{L^\infty_{[0,T)} L^2_{|r|\le 5/8}} \lesssim \epsilon^5 + \|v\|_{L^2_{|r|\le 3/4}} \le \epsilon^5. \qquad \square$$

Proposition 5.13 (local smoothing control). *Let the assumptions of Proposition 5.1 and Remark 5.2 hold. Let $\psi_{9/16}$ be a smooth function such that $\psi_{9/16}(r) = 1$ for $|r| \le \frac{9}{16}$ and $\psi_{9/16}(r) = 0$ for $|r| \ge \frac{5}{8}$. Then*

$$\|D_r^{3/2}(\psi_{9/16} v)\|_{L^2_{[0,T)} L^2_r} \lesssim \epsilon^5.$$

Proof. Let $\chi(\xi) = 1$ for $|\xi| \le 1$ and $\chi(\xi) = 0$ for $|\xi| \ge 2$ be a smooth function. Let $\chi_- = \chi$ and $\chi_+ = 1 - \chi$. Let P_- be the Fourier multiplier with symbol $\chi_-((T-t)^{3/4}\xi)$ and P_+ be the Fourier multiplier with symbol $\chi_+((T-t)^{3/4}\xi)$. Then $I = P_- + P_+$ for each t, and P_- projects onto frequencies $\lesssim (T-t)^{-3/4}$, while P_+ projects onto frequencies $\gtrsim (T-t)^{-3/4}$. Letting Q be the Fourier multiplier with symbol $\frac{3}{4}\chi'((T-t)^{3/4}\xi)$, we have $\partial_t P_\pm f = \pm i(T-t)^{-1/4} Q \partial_r f + P \partial_t f$. Note that Q has Fourier support in $|\xi| \sim (T-t)^{-3/4}$.

First, we can discard low frequencies. From Proposition 5.12 and (5-25) with $\alpha = \frac{1}{2}$,

$$\|D_r^{3/2} P_- \psi_{9/16} v\|_{L^2_{[0,T)} L^2_r} \lesssim \|(T-t)^{-3/8} \partial_r \psi_{9/16} v\|_{L^2_{[0,T)} L^2_r} \lesssim T^{1/8} \|\partial_r \psi_{9/16} v\|_{L^\infty_{[0,T)} L^2_r} \lesssim \epsilon^5.$$

For the high-frequency portion, $D_r^{3/2} P_+ \psi_{9/16} v$, we first need to dispose of the spatial cutoff. We have

$$D_r^{3/2} P_+ \psi_{9/16} = \psi_{9/16} D_r^{3/2} P_+ + [D_r^{3/2} P_+, \psi_{9/16}].$$

The leading order term in the symbol of the commutator $[D_r^{3/2} P_+, \psi_{9/16}]$, by the pseudodifferential calculus, is $\xi^{1/2}\chi_+(\xi(T-t)^{3/4})\psi'(r) + \xi^{3/2}(T-t)^{3/4}\chi_+'(\xi(T-t)^{3/4})\psi'(r)$. Hence, we obtain the bound

$$\|[D_r^{3/2} P_+, \psi_{9/16}] \langle D_r \rangle^{-1/2}\|_{L^2_r \to L^2_r} \lesssim 1,$$

independently of t. Thus, $\|[D_r^{3/2} P_+, \psi_{9/16}] v\|_{L^2_{[0,T)} L^2_r}$ is easily bounded by Proposition 5.12.

It remains to show that $\|\psi_{9/16} D_r^{3/2} P_+ v\|_{L^2_{[0,T)} L^2_r} \lesssim \epsilon^5$, the estimate for the high-frequency portion with no spatial cutoff to the right of the frequency cut-off. To obtain local smoothing via the energy method, we need to introduce the pseudodifferential operator A of order 0 with symbol $\exp(-(\operatorname{sgn}\xi)(\tan^{-1} r))$, where $\operatorname{sgn}\xi$ is a smoothed signum function. Note that by the sharp Gärding inequality, A is positive. The key property of A is

$$\partial_r^2 A f = A \partial_r^2 f - 2i(1+r^2)^{-1} D_r A f + B f,$$

where B is an order 0 pseudodifferential operator. The first-order term $i(1+r^2)^{-1}D_rAf$ will generate the local smoothing estimate.

Let $w=AP_+v$. By the sharp Gärding inequality,

$$\|\psi_{9/16}D_r^{3/2}P_+v\|_{L^2_{[0,T)}L^2_r}\lesssim\|(1+r^2)^{-1/2}D_r^{3/2}w\|_{L^2_{[0,T)}L^2_r}$$

and it suffices to prove that $\|(1+r^2)^{-1/2}D_r^{3/2}w\|_{L^2_{[0,T)}L^2_r}\lesssim\epsilon^5$. The equation satisfied by w is

$$i\partial_t w+\partial_r^2 w+2i(1+r^2)^{-1}D_rw=(T-t)^{-1/4}AQ\partial_rv-AP_+r^{-4}|v|^4v+Bv=F_1+F_2+F_3,$$

where B is a order 0 operator (satisfying bounds independent of t). By applying ∂_r and pairing this equation with $\partial_r w$ (energy method), we obtain, upon time integration,

$$\|\partial_r w\|^2_{L^\infty_{[0,T)}L^2_r}+\|(1+r^2)^{-1/2}D_r^{3/2}w\|^2_{L^2_{[0,T)}L^2_r}\lesssim\int_0^T|\langle\partial_rF_1,w\rangle|+10\|\partial_rF_2\|^2_{L^1_{[0,T)}L^2_r}+10\|\partial_rF_3\|^2_{L^1_{[0,T)}L^2_r}.$$

The F_3 term is easily controlled using Proposition 5.12.

The F_1 term is controlled as in the proof of Lemma 5.11 (a similar first term). For the F_2 term, let ψ be a smooth function such that $\psi(r)=1$ for $|r|\le\frac{1}{4}$ and $\psi(r)=0$ for $|r|\le\frac{1}{2}$. Writing $1=\psi^5+(1-\psi^5)$, we have

$$F_2=AP_+\psi^5r^{-4}|v|^4v+AP_+(1-\psi^5)r^{-4}|v|^4v=F_{21}+F_{22}.$$

We estimate $\|\partial_rF_{21}\|_{L^1_{[0,T)}L^2_r}$ as we did in the proof of Lemma 5.7. For the term F_{22}, take $\psi_+=(1-\psi^5)r^{-4}$, and note that ψ_+ is smooth and well localized. In the proof of Lemma 5.7 (see (5-18) and (5-21)), we showed that

$$\|P_{\ge N}\partial_r\psi_+|v|^4v\|_{L^1_{I_k}L^2_r}\lesssim 2^{k(1+\delta)/2}N^\delta\beta(k,\mu N)+N^{-1+\delta}2^{k\delta}\beta(k,\mu N)^2+N^{-1}2^{k(1+\delta)/2}.$$

Furthermore, Proposition 5.8 showed that $\beta(k,2^{3k/4})\lesssim 1$. Combining with the above gives

$$\|P_{\ge 2^{3k/4}}\partial_r\psi_+|v|^4v\|_{L^1_{I_k}L^2_r}\lesssim 2^{-k/8}.$$

Thus,

$$\|\partial_rF_{22}\|_{L^1_{[0,T)}L^2_r}\lesssim\sum_{k=k_0}^\infty\|P_{\ge 2^{3k/4}}\partial_r\psi_+|v|^4v\|_{L^1_{I_k}L^2_r}\lesssim\sum_{k=k_0}^\infty\|P_{\ge 2^{3k/4}}\partial_r\psi_+|v|^4v\|_{L^1_{I_k}L^2_r}\lesssim 2^{-k_0/8}\lesssim\epsilon^5.\qquad\square$$

Proposition 5.14 (Strichartz control). *Suppose that the assumptions of Proposition 5.1 and Remark 5.2 hold. Then*

$$\|r^{2/p-1}\partial_rv\|_{L^q_{[0,T)}L^p_{|r|\le 1/2}}\lesssim\epsilon^5.$$

Proof. Let ψ be a smooth function such that $\psi(r)=1$ for $|r|\le\frac{1}{2}$ and $\psi(r)=0$ for $|r|\ge\frac{9}{16}$. Let $w=\psi v$. Then w solves

$$i\partial_tw+\partial_r^2w=-\psi r^{-4}|v|^4v+2\partial_r(\psi'v)-\psi''v=F_1+F_2+F_3.$$

By the Strichartz estimate and dual local smoothing estimate, we obtain

$$\|r^{2/p-1}\partial_r w\|_{L^q_{[0,T)}L^p_r} \lesssim \|\partial_r w_0\|_{L^2_r} + \|\partial_r F_1\|_{L^1_{[0,T)}L^2_r} + \|D_r^{-1/2}\partial_r F_2\|_{L^2_{[0,T)}L^2_r} + \|\partial_r F_3\|_{L^1_{[0,T)}L^2_r}.$$

Let $\tilde{\psi}$ be a smooth function such that $\tilde{\psi}(r) = 1$ for $|r| \le \frac{1}{4}$ and $\tilde{\psi}(r) = 0$ for $|r| \ge \frac{1}{2}$. By writing $1 = \tilde{\psi}^5 + (1 - \tilde{\psi}^5)$, we have

$$F_1 = -\psi\tilde{\psi}^5 r^{-4}|v|^4 v - \psi(1-\tilde{\psi}^5)r^{-4}|v|^4 v = F_{11} + F_{12}.$$

Since the support of $\psi\tilde{\psi}^5$ is contained in $|r| \le \frac{1}{2}$, we can estimate the term $\|\partial_r F_{11}\|_{L^1_{[0,T)}L^2_r}$ by ϵ^5 using bootstrap assumption (2) as in the proof of Lemma 5.7. Since $(1-\tilde{\psi}^5)\psi r^{-4}$ is a bounded and smooth function,

$$\|\partial_r F_{12}\|_{L^1_{[0,T)}L^2_r} \lesssim \|\langle\partial_r\rangle v^5\|_{L^1_{[0,T)}L^2_{|r|\le 5/8}} \lesssim T\|\langle\partial_r\rangle v\|^5_{L^\infty_{[0,T)}L^2_{|r|\le 5/8}} \lesssim \epsilon^5.$$

Also, by Proposition 5.13,

$$\|D_r^{1/2}F_2\|_{L^2_{[0,T)}L^2_r} \lesssim \|\langle D_r\rangle^{3/2}\psi_{9/16}v\|_{L^2_{[0,T)}L^2_r} \lesssim \epsilon^5.$$

Finally,

$$\|\partial_r F_3\|_{L^1_{[0,T)}L^2_r} \lesssim T\|\langle\partial_r\rangle v\|_{L^\infty_{[0,T)}L^2_{|r|\le 5/8}} \lesssim \epsilon^5$$

by Proposition 5.12. Collecting the estimates above, we obtain the claimed bound. □

This completes the proof of Proposition 5.1 (via Lemma 5.5).

Acknowledgments

Holmer is partially supported by a Sloan fellowship and NSF grant DMS-0901582. Roudenko is partially supported by NSF grant DMS-0808081. Holmer thanks Mike Christ and Daniel Tataru for patient mentorship, in work related to the paper [Koch and Tataru 2007], on the use of the bilinear Strichartz estimates.

References

[Bourgain 1998] J. Bourgain, "Refinements of Strichartz' inequality and applications to 2D-NLS with critical nonlinearity", *Internat. Math. Res. Notices* **1998**:5 (1998), 253–283. MR 99f:35184 Zbl 0917.35126

[Bourgain 1999] J. Bourgain, "Global wellposedness of defocusing critical nonlinear Schrödinger equation in the radial case", *J. Amer. Math. Soc.* **12**:1 (1999), 145–171. MR 99e:35208 Zbl 0958.35126

[Cazenave 2003] T. Cazenave, *Semilinear Schrödinger equations*, Courant Lecture Notes in Mathematics **10**, Courant Institute of Mathematical Sciences, New York, 2003. MR 2004j:35266 Zbl 1055.35003

[Colliander et al. 2001] J. Colliander, M. Keel, G. Staffilani, H. Takaoka, and T. Tao, "Global well-posedness for Schrödinger equations with derivative", *SIAM J. Math. Anal.* **33**:3 (2001), 649–669. MR 2002j:35278 Zbl 1002.35113

[Colliander et al. 2008] J. Colliander, M. Keel, G. Staffilani, H. Takaoka, and T. Tao, "Global well-posedness and scattering for the energy-critical nonlinear Schrödinger equation in $\mathbb{R}^3$", *Ann. of Math.* (2) **167**:3 (2008), 767–865. MR 2009f:35315 Zbl 1178.35345

[Evans and Zworski 2003] L. C. Evans and M. Zworski, "Lectures on semiclassical analysis", University of California, Berkeley, CA, 2003, Available at http://math.berkeley.edu/∼evans/semiclassical.pdf.

[Holmer and Roudenko 2011] J. Holmer and S. Roudenko, "A class of solutions to the 3D cubic nonlinear Schrödinger equation that blows up on a circle", *Appl. Math. Res. Express* **2011**:1 (2011), 23–94. MR 2012b:35323 Zbl 1220.35159 arXiv 1002.2407

[Jerison and Kenig 1995] D. Jerison and C. E. Kenig, "The inhomogeneous Dirichlet problem in Lipschitz domains", *J. Funct. Anal.* **130**:1 (1995), 161–219. MR 96b:35042 Zbl 0832.35034

[Keel and Tao 1998] M. Keel and T. Tao, "Endpoint Strichartz estimates", *Amer. J. Math.* **120**:5 (1998), 955–980. MR 2000d:35018 Zbl 0922.35028

[Kenig and Merle 2006] C. E. Kenig and F. Merle, "Global well-posedness, scattering and blow-up for the energy-critical, focusing, non-linear Schrödinger equation in the radial case", *Invent. Math.* **166**:3 (2006), 645–675. MR 2007g:35232 Zbl 1115.35125

[Koch and Tataru 2007] H. Koch and D. Tataru, "A priori bounds for the 1D cubic NLS in negative Sobolev spaces", *Internat. Math. Res. Notices* **2007**:16 (2007), Art. ID rnm053. MR 2010d:35307 Zbl 1169.35055

[Landman et al. 1988] M. J. Landman, G. C. Papanicolaou, C. Sulem, and P.-L. Sulem, "Rate of blowup for solutions of the nonlinear Schrödinger equation at critical dimension", *Phys. Rev. A* (3) **38**:8 (1988), 3837–3843. MR 89k:35218

[Merle and Raphaël 2005] F. Merle and P. Raphaël, "Profiles and quantization of the blow up mass for critical nonlinear Schrödinger equation", *Comm. Math. Phys.* **253**:3 (2005), 675–704. MR 2006m:35346 Zbl 1062.35137

[Perelman 2001] G. Perelman, "On the formation of singularities in solutions of the critical nonlinear Schrödinger equation", *Ann. Henri Poincaré* **2**:4 (2001), 605–673. MR 2002m:35205 Zbl 1007.35087

[Raphaël 2006] P. Raphaël, "Existence and stability of a solution blowing up on a sphere for an L^2-supercritical nonlinear Schrödinger equation", *Duke Math. J.* **134**:2 (2006), 199–258. MR 2007k:35472 Zbl 1117.35077

[Raphaël and Szeftel 2009] P. Raphaël and J. Szeftel, "Standing ring blow up solutions to the N-dimensional quintic nonlinear Schrödinger equation", *Comm. Math. Phys.* **290**:3 (2009), 973–996. MR 2010k:35469 Zbl 1184.35295

[Stein 1993] E. M. Stein, *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, Princeton Mathematical Series **43**, Princeton University Press, Princeton, NJ, 1993. MR 95c:42002 Zbl 0821.42001

[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 58 #23577 Zbl 0372.35001

[Tao 2006] T. Tao, *Nonlinear dispersive equations: local and global analysis*, CBMS Regional Conference Series in Mathematics **106**, American Mathematical Society, Providence, RI, 2006. MR 2008i:35211 Zbl 1106.35001

[Weinstein 1982] M. I. Weinstein, "Nonlinear Schrödinger equations and sharp interpolation estimates", *Comm. Math. Phys.* **87**:4 (1982), 567–576. MR 84d:35140 Zbl 0527.35023

[Zwiers 2011] I. Zwiers, "Standing ring blowup solutions for cubic nonlinear Schrödinger equations", *Anal. PDE* **4**:5 (2011), 677–727. MR 2901563

Received 23 Jul 2010. Revised 10 Jan 2011. Accepted 21 Feb 2011.

JUSTIN HOLMER: holmer@math.brown.edu
Mathematics, Brown University, Box 1917, 151 Thayer St, Providence, RI 02912, United States

SVETLANA ROUDENKO: roudenko@gwu.edu
Mathematics, George Washington University, 2115 G Street NW, George Washington University, Washington, DC 20052, United States

ANALYSIS AND PDE
Vol. 5, No. 3, 2012
dx.doi.org/10.2140/apde.2012.5.513

SHARP GEOMETRIC UPPER BOUNDS ON RESONANCES FOR SURFACES WITH HYPERBOLIC ENDS

DAVID BORTHWICK

We establish a sharp geometric constant for the upper bound on the resonance counting function for surfaces with hyperbolic ends. An arbitrary metric is allowed within some compact core, and the ends may be of hyperbolic planar, funnel, or cusp type. The constant in the upper bound depends only on the volume of the core and the length parameters associated to the funnel or hyperbolic planar ends. Our estimate is sharp in that it reproduces the exact asymptotic constant in the case of finite-area surfaces with hyperbolic cusp ends, and also in the case of funnel ends with Dirichlet boundary conditions.

1. Introduction

For a compact Riemannian surface, the Weyl law shows that the asymptotic distribution of eigenvalues is determined by global geometric quantities. In the compact hyperbolic case, Weyl asymptotics follow easily from the Selberg trace formula; see, e.g, [McKean 1972], This approach extends also to noncompact hyperbolic surfaces of finite area [Venkov 1990]. Some reinterpretation of the spectral counting is needed for the noncompact case, however. One can either supplement the counting function for the discrete spectrum by a term related to the scattering phase, or else use the counting function for resonances instead of eigenvalues. Weyl asymptotics, in this extended sense, were established for general finite-area surfaces with hyperbolic cusp ends by Müller [1992] and Parnovski [1995].

For infinite-area surfaces with hyperbolic ends, the discrete spectrum is finite and possibly empty, and therefore plays no role in the spectral asymptotics. One could look for analogies to the finite-area results in the asymptotics of either the scattering phase or the resonance counting function. For the scattering phase of a surface with hyperbolic ends, Weyl asymptotics were proven by Guillopé and Zworski [1997]. One does not necessarily expect a corresponding result to hold for the resonance counting function — see e.g., [Guillopé and Zworski 1997, Remark 1.6] — but neither can we rule out the possibility at this point. Understanding the role that global geometric properties play in the distribution of resonances remains a compelling problem.

In the context of infinite-area hyperbolic surfaces, only the order of growth of the resonance counting function is currently well understood. Guillopé and Zworski [1995; 1997] showed the resonance counting function for infinite-area surfaces with hyperbolic ends satisfies $N_g(t) \asymp t^2$ (with the caveat that the lower bound is proportional to the 0-volume, which might be zero in exceptional cases). These results have been

Supported in part by NSF grant DMS-0901937.
MSC2000: primary 35P25, 58J50; secondary 47A40.
Keywords: resonances, hyperbolic surfaces, scattering theory.

extended to higher-dimensional manifolds with hyperbolic ends in [Borthwick 2008]. Unfortunately, the methods used in these proofs yield only an ineffective constant for the upper bound, with no clear geometric content. Moreover, the derivation of the lower bound depends explicitly on the upper bound, so the geometric dependence of the lower bound is likewise undetermined.

In this paper we present a geometric constant for the upper bound on the resonance counting function for infinite-area surfaces with hyperbolic ends. This constant is sharp in the sense that it agrees with the exact asymptotics in the cases of finite area surfaces or truncated funnels. Our approach is inspired by Stefanov's recent paper [2006] on compactly supported perturbations of the Laplacian on $\mathbb{R}^n$ for n odd, and similar techniques were applied to compactly supported perturbations of $\mathbb{H}^{n+1}$ in [Borthwick 2010].

We can state the cleanest result for a hyperbolic surface $(X, g) \cong \mathbb{H}^2/\Gamma$. Let $\mathcal{R}_g$ denote the associated resonance set (poles of the meromorphic continuation of $(\Delta_g - s(1-s))^{-1}$), with counting function

$$N_g(t) := \#\big\{\zeta \in \mathcal{R}_g : |\zeta - \tfrac{1}{2}| \le t\big\}.$$

The sharp version of our bound involves a regularization of the counting function,

$$\widetilde{N}_g(a) := \int_0^a \frac{2(N_g(t) - N_g(0))}{t^2}\, dt. \tag{1-1}$$

This type of regularization is standard in the theory of zeros of entire functions, and there is a natural connection to the asymptotics of $N_g(t)$,

$$\widetilde{N}_g(a) \sim Ba^2 \quad \Longleftrightarrow \quad N_g(t) \sim Bt^2;$$

see [Stefanov 2006, Lemma 1]. If we have only the upper bound on $\widetilde{N}_g$, then we lose some sharpness in the estimate of N_g:

$$\widetilde{N}_g(a) \le Ba^2 \quad \Longrightarrow \quad N_g(t) \le eBt^2.$$

Theorem 1.1. *Suppose (X, g) is a smooth geometrically finite hyperbolic surface with $\chi(X) < 0$. Let $\ell_1, \dots, \ell_{n_f}$ denote the diameters of the geodesic boundaries of the funnels of X. The regularized counting function for the resonances of Δ_g satisfies*

$$\frac{\widetilde{N}_g(a)}{a^2} \le |\chi(X)| + \sum_{j=1}^{n_f} \frac{\ell_j}{4} + o(1). \tag{1-2}$$

We can see that this result is sharp in two extreme cases. For a finite-area hyperbolic surface (that is, with $n_f = 0$), our upper bound agrees with the known asymptotic $N_g(t)/t^2 \sim |\chi(X)|$. On the other hand, for an isolated hyperbolic funnel F_ℓ of boundary length ℓ, under Dirichlet boundary conditions, the resonances form a half lattice. It is easy to see that $N_{F_\ell}(t)/t^2 \sim \ell/4$, so the funnel portion of (1-2) is also sharp.

The restriction to $\chi(X) < 0$ in Theorem 1.1 leaves out just a few cases. The complete (smooth) hyperbolic surfaces for which $\chi(X) \ge 0$ are the hyperbolic plane $\mathbb{H}^2$, the hyperbolic cylinder $C_\ell := \mathbb{H}^2/\langle z \mapsto e^\ell z\rangle$, and the parabolic cylinder $C_\infty := \mathbb{H}^2/\langle z \mapsto z+1\rangle$. Resonance sets can be computed explicitly in these cases (see [Borthwick 2007, Sections 4–5]), and exact asymptotics for the counting function are easily obtained:

$$N_{\mathbb{H}^2}(t) \sim t^2, \quad N_{C_\ell}(t) \sim \tfrac{1}{2}\ell t^2, \quad N_{C_\infty}(t) = 1.$$

If we interpret C_ℓ as the union of 2 funnel ends, then (1-2) would also give a sharp estimate for this case.

Using Theorem 1.1 in conjunction with the argument of Guillopé and Zworski [1997] for the lower bound, we can deduce the following:

Corollary 1.2. *For $k \in \mathbb{N}$ there exists a constant c_k such that for any geometrically finite hyperbolic surface (X, g) with $\chi(X) < 0$,*

$$\frac{N_g(t)}{t^2} \geq c_k |\chi(X)| \biggl(1 + \frac{1}{|\chi(X)|} \sum_{j=1}^{n_\mathrm{f}} \frac{\ell_j}{4}\biggr)^{-2/k} \quad \textit{for } t \geq 1.$$

The constant c_k obtained in this way (see Section 4 for the derivation) is rather ineffective; the point here is just that there exists a lower bound that depends only on $\chi(X)$ and $\{\ell_j\}$.

We will obtain Theorem 1.1 as a consequence of a somewhat more general estimate. Consider a smooth Riemannian surface (X, g), possibly with boundary, that has finitely many ends that are assumed to be of hyperbolic planar, funnel, or cusp type. That is, X admits the decomposition

$$X = K \sqcup Y_1 \sqcup \cdots \sqcup Y_{n_\mathrm{f}} \sqcup C_{n_\mathrm{f}+1} \sqcup \cdots \sqcup C_{n_\mathrm{f}+n_\mathrm{c}} \tag{1-3}$$

illustrated in Figure 1, where the core K is a smooth compact manifold with boundary. The metric in K is arbitrary. The Y_j are infinite-area ends: either hyperbolic planar,

$$Y_j \cong [b_j, \infty) \times S^1, \quad g|_{Y_j} = dr^2 + \sinh^2 r \, d\theta^2, \quad \text{where } b_j > 0, \tag{1-4}$$

or hyperbolic funnels,

$$Y_j \cong [b_j, \infty) \times S^1, \quad g|_{Y_j} = dr^2 + \ell_j^2 \cosh^2 r \frac{d\theta^2}{(2\pi)^2}, \quad \text{where } b_j \geq 0 \text{ and } \ell_j > 0. \tag{1-5}$$

The C_j are hyperbolic cusps,

$$C_j \cong [b_j, \infty) \times S^1, \quad g|_{C_j} = dr^2 + e^{-2r} \frac{d\theta^2}{(2\pi)^2}, \quad \text{where } b_j \geq 0. \tag{1-6}$$

The finite-area portion of X consisting of the core plus the cusps is denoted by

$$X_\mathrm{c} := K \sqcup C_{n_\mathrm{f}+1} \sqcup \cdots \sqcup C_{n_\mathrm{f}+n_\mathrm{c}}. \tag{1-7}$$

Any geometrically finite hyperbolic surface, with the exception of the parabolic cylinder C_∞, admits a decomposition of the form (1-3). In such surfaces, aside from $\mathbb{H}^2$ itself, only funnel or cusp ends can occur.

We let Δ_g denote the positive Laplacian on (X, g). In general we may consider the operator

$$P := \Delta_g + V,$$

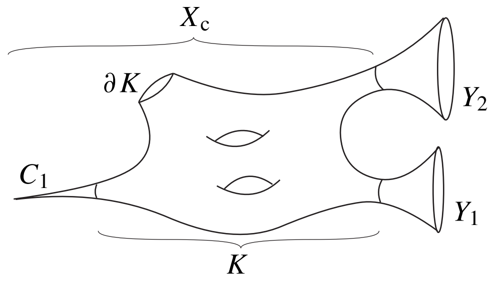

Figure 1. Surface X with boundary and hyperbolic ends.

where $V \in C_0^\infty(X)$ with $\operatorname{supp}(V) \subset K$. We denote by $\mathcal{R}_P$ the resonance set associated to P. These resonances are the poles of the analytically continued resolvent

$$R_P(s) := (P - s(1-s))^{-1},$$

counted according to multiplicity. The associated resonance counting function is

$$N_P(t) := \#\{\zeta \in \mathcal{R}_P : |\zeta - \tfrac{1}{2}| \le t\}.$$

Our context is essentially that of Guillopé and Zworski [1995; 1997], and so we already know that $N_P(t) \asymp t^2$ (see Section 2 for details). It is thus natural to define the regularized counting function $\widetilde{N}_P(a)$ just as in (1-1).

Before stating the upper bound, we introduce the asymptotic constants associated to the resonance count for isolated hyperbolic planar or funnel ends.

Theorem 1.3. *For a hyperbolic planar or funnel end $Y \cong [b, \infty) \times S^1$, with metric as in* (1-4) *or* (1-5), *the resonance counting function for the Laplacian with Dirichlet boundary conditions at $r = b$ satisfies an asymptotic as $t \to \infty$,*

$$N_Y(t) \sim A(Y) t^2.$$

We will write these constants $A(Y)$ explicitly in a moment. First let us state our main result.

Theorem 1.4. *For (X, g) a surface with hyperbolic ends as in* (1-3) *and $V \in C_0^\infty(X)$, the regularized counting function for $P = \Delta_g + V$ satisfies*

$$\frac{\widetilde{N}_P(a)}{a^2} \le \frac{1}{2\pi} \operatorname{vol}(X_c, g) + \sum_{j=1}^{n_f} A(Y_j) + o(1), \tag{1-8}$$

where X_c is the subset (1-7).

If (X, g) is a finite-area surface with hyperbolic cusp ends (and arbitrary metric in the interior), Parnovski [1995] proved that

$$N_g(t) \sim \frac{1}{2\pi} \operatorname{vol}(X, g) t^2.$$

This shows that Theorem 1.4 is sharp in the case $n_{\mathrm{f}} = 0$. It also suggests an intriguing interpretation of the constants appearing in (1-8). Suppose we split X into a disjoint union $X_{\mathrm{c}} \cup Y_1 \cup \cdots \cup Y_{n_{\mathrm{f}}}$ at the boundary of X_{c} and impose Dirichlet boundary conditions at the newly created boundaries. The constant on the right side of (1-8) is the sum of the asymptotic constants for the resonance counting function of the resulting components.

To obtain Theorem 1.1 from Theorem 1.4, we take the Y_j to be standard funnels with boundaries at $b_j = 0$, in which case $A(Y_j) = \ell_j/4$. And under the assumptions that X_{c} has geodesic boundary and hyperbolic interior, the Gauss–Bonnet theorem gives $\mathrm{vol}(X_{\mathrm{c}}, g) = -2\pi\chi(X)$.

As in Corollary 1.2, combining Theorem 1.4 with the Guillopé–Zworski argument gives a lower bound on $N_P(t)$ with a constant that depends only on 0-vol(X, g) and the end parameters ℓ_j and b_j for $j = 1, \dots, n_{\mathrm{f}}$, assuming that 0-vol$(X, g) \neq 0$.

The asymptotic constants $A(Y)$ appearing in Theorem 1.3 have a somewhat complicated form. Consider first a model funnel end F_{ℓ,r_0} defined by

$$F_{\ell,r_0} \cong [r_0, \infty) \times S^1 \quad \text{and} \quad ds^2 = dr^2 + \ell^2 \cosh^2 r \, \frac{d\theta^2}{(2\pi)^2}. \tag{1-9}$$

The case $r_0 = 0$, a standard funnel with geodesic boundary, is simply denoted by F_ℓ. The resonance set for the Laplacian on F_{ℓ,r_0} with Dirichlet boundary conditions at $r = r_0$ is denoted $\mathcal{R}_{F_{\ell,r_0}}$.

In Section 7 we will show that for $r_0 \geq 0$,

$$A(F_{\ell,r_0}) = -\frac{\ell}{2\pi} \sinh r_0 + \frac{4}{\pi} \int_0^{\pi/2} \int_0^{\infty} \frac{[I(xe^{i\theta}, \ell, r_0)]_+}{x^3} \, dx \, d\theta, \tag{1-10}$$

where $[\,\cdot\,]_+$ denotes the positive part and, with $\omega := 2\pi/\ell$,

$$I(\alpha, \ell, r) := \mathrm{Re}\biggl(2\alpha \log\biggl(\frac{\alpha \sinh r + \sqrt{\omega^2 + \alpha^2 \cosh^2 r}}{\sqrt{\omega^2 + \alpha^2}}\biggr)\biggr) + \omega \arg\biggl(\frac{\sqrt{\omega^2 + \alpha^2 \cosh^2 r} - i\omega \sinh r}{\sqrt{\omega^2 + \alpha^2 \cosh^2 r} + i\omega \sinh r}\biggr) + \pi(\mathrm{Im}\,\alpha - \omega). \tag{1-11}$$

(We will use the principal branch of log in all such formulas.) The integral in (1-10) is explicitly computable in the case $r_0 = 0$, since $I(xe^{i\theta}, \ell, 0) = \pi(x \sin\theta - \omega)$. In this case we recover the asymptotic constant for the standard funnel, $A(F_\ell) = \ell/4$.

It is interesting to compare the resonance sets of truncated funnels F_{ℓ,r_0} with $r_0 > 0$ to extended funnels with $r_0 < 0$. The two cases are quite different in terms of the classical dynamics; an extended funnel contains a trapped geodesic, while truncated funnels are nontrapping. Because of this change in dynamics, we expect the distribution of resonances near the critical line to change dramatically as r_0 switches from positive to negative. Figure 2 illustrates these differences. In the nontrapping case at left, the distance from the resonances to the critical line increases logarithmically as $\mathrm{Im}\, s \to \infty$. For the trapping case at right, the distance decreases exponentially. These behaviors are consistent with results on resonance-free regions for asymptotically hyperbolic manifolds by Guillarmou [2005].

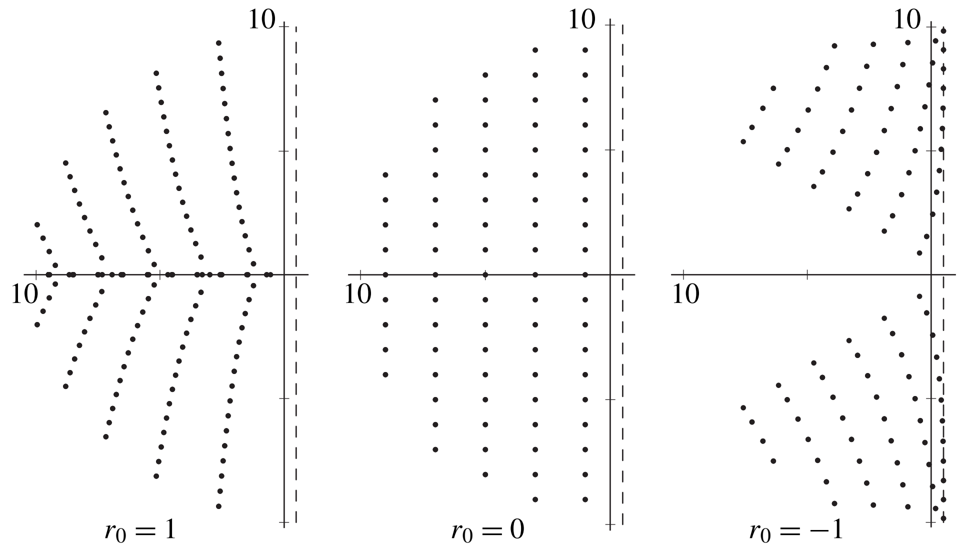

Figure 2. Resonance sets of the funnel F_{ℓ,r_0} with different boundary locations r_0, shown for $\ell = 2\pi$.

Of course, the asymptotics of the global counting function $N_P(t)$ are not expected to be sensitive to the dynamics. Indeed, we will show in Section 8 that the formula (1-10) for the asymptotic constant of $N_{F_{\ell,r_0}}(t)$ remains valid for $r_0 < 0$. This exact asymptotic can be compared to the upper bound obtained for the extended funnel from Theorem 1.4, which is

$$\frac{\widetilde{N}_{F_{\ell,r_0}}(a)}{a^2} \le -\frac{\ell}{2\pi}\sinh r_0 + \frac{\ell}{4} \quad \text{for } r_0 \le 0. \tag{1-12}$$

Figure 3 illustrates the difference between the upper bound (1-12) and the sharp asymptotic in this situation. Given this discrepancy, one might think that the bound in Theorem 1.4 could be improved by moving the boundary of K further into the interior of the surface (that is, by allowing $b_j < 0$ in the definition (1-5)). Unfortunately, for reasons that we will explain in Section 4, it does not seem possible to obtain any improvement this way.

In the hyperbolic planar case, the model problem for Y_j is scattering by a spherical obstacle in $\mathbb{H}^2$, that is, on the exterior Dirichlet domain $\Omega_{r_0} := \{r \ge r_0\} \subset \mathbb{H}^2$. The resonance asymptotics for this spherical obstacles in $\mathbb{H}^{n+1}$ were worked out in Borthwick [2010, Theorem 1.2]. In two dimensions the result is

$$A(\Omega_{r_0}) = 2 - \cosh r_0 + \frac{4}{\pi}\int_0^{\pi/2}\int_0^{\infty} \frac{[H(xe^{i\theta}, r_0)]_+}{x^3}\, dx\, d\theta, \tag{1-13}$$

where

$$H(\alpha, r) := \operatorname{Re}\left(2\alpha \log\left(\frac{\alpha\cosh r + \sqrt{1+\alpha^2\sinh^2 r}}{\sqrt{\alpha^2-1}}\right)\right) + \log\left|\frac{\cosh r - \sqrt{1+\alpha^2\sinh^2 r}}{\cosh r + \sqrt{1+\alpha^2\sinh^2 r}}\right|. \tag{1-14}$$

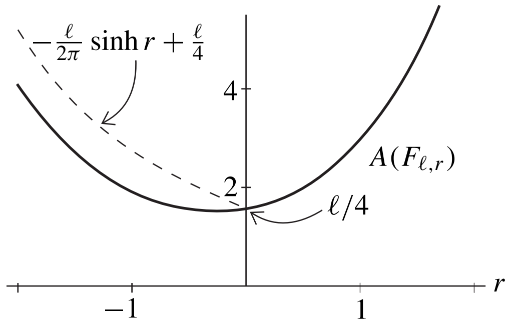

Figure 3. The exact asymptotic constant for $F_{\ell,r}$ as a function of boundary location r, shown for $\ell = 2\pi$. The dotted line shows the bound from Theorem 1.4.

The paper is organized as follows. The basic material on the resolvent and resonances of the operator P is reviewed in Section 2. In Section 3 we present the factorization formula for the relative scattering determinant and show that this leads to Weyl asymptotics for the scattering phase and a counting formula for resonances based on contour integration. The growth estimates on the scattering determinant and the resulting proof of Theorem 1.4 are given in Section 4, assuming certain estimates to be developed in later sections. The derivation of Corollary 1.2 is also given in Section 4. In Section 5, we develop the asymptotic analysis of Dirichlet eigenmodes on hyperbolic funnels. These asymptotics are applied in Section 6 to prove the Poisson operator estimates needed for Section 4. Finally, in Section 7 and Section 8 we establish the exact asymptotic constant (1-10) for the truncated and extended funnel cases, respectively, and prove the funnel part of Theorem 1.3 in particular.

2. Resonances

The context introduced in Section 1 differs from that of Guillopé and Zworski [1995; 1997] in two relatively minor ways: Hyperbolic planar ends are allowed in addition to funnels, and a compactly supported potential V is possibly added to Δ_g. The latter addition really is trivial, but the inclusion of hyperbolic planar ends requires a few extra estimates on model terms. In this section we will briefly review the theory [Guillopé and Zworski 1995; 1997], in order to explain those additional estimates.

To define resonances we need analytic continuation of the resolvent, $R_P(s) := (P - s(1-s))^{-1}$, from its original domain $\operatorname{Re} s > \frac{1}{2}$. Each end Y_j is isometric to a portion of either $\mathbb{H}$ or the model funnel F_{ℓ_j}, and we can use this identification to pullback model resolvents $R^0_{Y_j}(s)$. After appropriate cutoffs are applied, we can treat these model terms as operators on X, whose kernels have support only in the corresponding ends Y_j. Similarly, we define $R^0_{C_j}(s)$ by pullback from the model cusp. Suppose that $\chi_k^j \in C^\infty(X)$ are cutoff functions for $j = 1, \ldots, n_{\mathrm{f}} + n_{\mathrm{c}}$ and $k = 0, 1, 2$, such that

$$\chi_k^j = \begin{cases} 0 & \text{for } r \geq k+1 \text{ in end } j, \\ 1 & \text{for } r \leq k \text{ in end } j, \\ 1 & \text{outside of end } j. \end{cases}$$

We also set $\chi_k := \prod_j \chi_k^j$.

For some s_0 with $\operatorname{Re} s_0$ sufficiently large, so that $R_P(s_0)$ is defined, we set

$$M(s) := \chi_2 R_P(s_0)\chi_1 + \sum_{j=1}^{n_{\mathrm{f}}}(1-\chi_0^j)R^0_{Y_j}(s)(1-\chi_1^j) + \sum_{j=n_{\mathrm{f}}+1}^{n_{\mathrm{f}}+n_{\mathrm{c}}}(1-\chi_0^j)R^0_{C_j}(s)(1-\chi_1^j).$$

This parametrix satisfies

$$(P - s(1-s))M(s) = I - L(s),$$

where

$$L(s) := -[\Delta_g, \chi_2]R_P(s_0)\chi_1 + (s(1-s) - s_0(1-s_0))\chi_2 R_P(s_0)\chi_1 + \sum_{j=1}^{n_{\mathrm{f}}}[\Delta_g, \chi_0^j]R^0_{Y_j}(s)(1-\chi_1^j) + \sum_{j=n_{\mathrm{f}}+1}^{n_{\mathrm{f}}+n_{\mathrm{c}}}[\Delta_g, \chi_0^j]R^0_{C_j}(s)(1-\chi_1^j).$$

There are two differences here from the construction of [Guillopé and Zworski 1995]. First of all, some of our model terms $R^0_{Y_j}(s)$ will be copies of $R_{\mathbb{H}}(s)$ instead of the funnel resolvent. Second, we follow the treatment in [Borthwick 2007] in using the model resolvent for a full cusp, rather than modifying the original Hilbert space.

Let $\rho \in C^\infty(X)$ be proportional to e^{-r} in the ends Y_j and C_j, with respect to the coordinate systems given in (1-4)–(1-6). The operator $L(s)$ is compact on $\rho^N L^2(X, dg)$ for $\operatorname{Re} s > \frac{1}{2} - N$ and defines a meromorphic family with poles of finite rank. (The structure of the kernel of $R^0_{Y_j}(s)$ at infinity is the same whether Y_j is a funnel or hyperbolic planar, so this part of the argument is unaffected by the addition of hyperbolic planar ends.)

By choosing s and s_0 appropriately we can insure that $I - L(s)$ is invertible at some s, and then the analytic Fredholm yields

$$R_P(s) = M(s)(I - L(s))^{-1}. \tag{2-1}$$

This proves the following result, a slight generalization of [Guillopé and Zworski 1995, Theorem 1]:

Theorem 2.1 (Guillopé and Zworski). *The formula* (2-1) *defines a meromorphic extension of $R_P(s)$ to a bounded operator on $\rho^N L^2(X, dg)$ for* $\operatorname{Re} s > \frac{1}{2} - N$, *with poles of finite rank.*

Meromorphic continuation allows us to define $\mathcal{R}_P$ as the set of poles of $R_P(s)$, listed according to multiplicities given by

$$m_P(\zeta) := \operatorname{rank} \operatorname{Res}_\zeta R_P(s).$$

The same parametrix construction also leads to an estimate of the order of growth of the resonance counting function. The following is a slight generalization of [Guillopé and Zworski 1995, Theorem 2]:

Theorem 2.2 (Guillopé and Zworski). *The resonance counting function satisfies a bound*

$$N_P(t) = O(t^2).$$

Our version requires just a few additional estimates. To obtain this bound on the counting function, Guillopé and Zworski [1995] introduced a Fredholm determinant

$$D(s) := \det(I - L_3(s)^3), \quad \text{where } L_3(s) := L(s)\chi_3.$$

Using the relation

$$R_P(s)\chi_3 = M(s)\chi_3(I + L_3(s) + L_3(s)^2)(I - L_3(s)^3)^{-1},$$

and a result of Vodev [1994, Appendix], they showed that $\mathcal{R}_P$ is included in the union of the set of poles of $D(s)$ with 3 copies of the union of the sets of poles of $M(s)$ and $L_3(s)$.

The only change that the inclusion of hyperbolic planar ends requires in this argument is that for each hyperbolic planar end we include a copy of $\mathcal{R}_{\mathbb{H}}$ among the possible poles of $M(s)$ and $L_3(s)$. Since $N_{\mathbb{H}}(t) = O(t^2)$, just as for funnels, the problem reduces as in [Guillopé and Zworski 1995] to an estimate of the growth of $D(s)$. Through Weyl's inequality, the estimate of $D(s)$ is broken up into estimates on the singular values of various model terms. We must check that the relevant estimates are satisfied by the hyperbolic planar model terms.

There are three estimates to consider. The first concerns the resolvent $R_{\mathbb{H}}(s)$. If Q_1 and Q_2 are compactly supported differential operators of orders q_1 and q_2, with disjoint supports, then for $\varepsilon > 0$,

$$\|Q_1 R_{\mathbb{H}}(s) Q_2\| \le C(q_j, \varepsilon)\langle s\rangle^{q_1+q_2} \quad \text{for } \operatorname{Re} s > \varepsilon, \tag{2-2}$$

and

$$\|Q_1 R_{\mathbb{H}}(s) Q_2\| \le C(q_j, \varepsilon)\langle s\rangle^{q_1+q_2-1} \quad \text{for } \operatorname{Re} s > \tfrac{1}{2} + \varepsilon. \tag{2-3}$$

To prove either of these, one can simply use the explicit formula

$$R_{\mathbb{H}}(s; z, z') = \frac{1}{4\pi}\int_0^1 \frac{(t(1-t))^{s-1}}{(t + \sinh^2 d(z, z'))^s}\, dt,$$

and repeat the argument from [Guillopé and Zworski 1995, Lemma 3.2].

The next estimate is for the Poisson kernel $E_{\mathbb{H}}(s)$. In the Poincaré ball model $\mathbb{B}$, this kernel is given by

$$E_{\mathbb{B}}(s; z, \theta) = \frac{1}{4\pi}\frac{\Gamma(s)^2}{\Gamma(2s)}\frac{(1-|z|^2)^s}{|e^{i\theta} - z|^{2s}} \quad \text{for } z \in \mathbb{B},\ \theta \in \mathbb{R}/(2\pi\mathbb{Z}).$$

Given a compact set $K \subset \mathbb{B}$ and $\varepsilon > 0$, we have

$$\left|\partial_\theta^k E_{\mathbb{B}}(s; z, \theta)\right| \le C(K, \varepsilon)^k k!\, e^{c\langle s\rangle} \quad \text{for } z \in K,\ k \in \mathbb{N}. \tag{2-4}$$

This is not difficult to prove directly by induction, or one can use an analyticity argument as in [Guillopé and Zworski 1995, Lemma 3.1].

Finally, we must estimate the scattering matrix $S_{\mathbb{H}}(s)$. We can write this explicitly in terms of Fourier modes,

$$S_{\mathbb{H}}(s) = \sum_{k\in\mathbb{Z}} [S_{\mathbb{H}}(s)]_k e^{ik(\theta-\theta')}, \quad \text{where } [S_{\mathbb{H}}(s)]_k = 2^{1-2s}\frac{\Gamma(\frac{1}{2} - s)}{\Gamma(s - \frac{1}{2})}\frac{\Gamma(s + |k|)}{\Gamma(1 - s + |k|)}$$

Using Stirling's formula, it is easy to use this expression for the eigenvalues to estimate the singular values of $S_{\mathbb{H}}(s)$. Assuming that $\operatorname{Re} s < \frac{1}{2} - \varepsilon$ and $\operatorname{dist}(s, -\mathbb{N}_0) > \eta$, we have

$$\mu_j(S_{\mathbb{H}}(s)) \le \exp(C(\eta)\langle s\rangle + \operatorname{Re}(1-2s)\log(\langle s\rangle/j)). \tag{2-5}$$

This is the analog of [Guillopé and Zworski 1997, Lemma 4.2].

With these model estimates in place, one can simply apply Guillopé and Zworski's original argument (treating the cusp contributions as in [Borthwick 2007, Section 9.4]) to prove that

$$|g(s)D(s)| \le e^{C\langle s\rangle^2},$$

where $g(s)$ is a entire function of order 2 and finite type, with zeros derived from $\mathcal{R}_{\mathbb{H}}$ and the model resolvent sets for the funnels and cusps. This yields the proof of Theorem 2.2.

3. Relative scattering determinant

To define scattering matrices, we will fix a function $\rho \in C^\infty(X)$ that serves as a boundary defining function for a suitable compactification of X. We start with smooth positive functions ρ_{f}, ρ_{c} satisfying

$$\rho_{\mathrm{f}} = \begin{cases} 2e^{-r} & \text{in each } Y_j, \\ 1 & \text{in each } C_j \end{cases} \quad \text{and} \quad \rho_{\mathrm{c}} = \begin{cases} 1 & \text{in each } Y_j, \\ e^{-r} & \text{in each } C_j. \end{cases}$$

Then we set $\rho = \rho_{\mathrm{f}}\rho_{\mathrm{c}}$ for the global boundary defining function.

The ends Y_j are conformally compact, and we distinguish between the internal boundary ∂Y_j, and the boundary at infinity $\partial_\infty Y_j$ induced by the conformal compactification. The funnel ends Y_j come equipped with a length parameter ℓ_j, the length of the closed geodesic bounding the finite end. If we assign length $\ell_j = 2\pi$ to a hyperbolic planar end, for consistency, then the metric induced by $\rho^2 g$ on the boundary of Y_j at infinity gives an isometry

$$\partial_\infty Y_j \cong \mathbb{R}/\ell_j\mathbb{Z}.$$

The cusp ends can be compactified naturally by lifting to $\mathbb{H}$ and invoking the Riemann-sphere topology, as described in [Borthwick 2007, Section 6.1]. The resulting boundary $\partial_\infty C_j$ consists of a single point.

Despite the discrepancy in dimensions, it will be convenient to group all of the infinite boundaries together as

$$\partial_\infty X := \partial_\infty Y_1 \cup \cdots \cup \partial_\infty Y_{n_{\mathrm{f}}} \cup \partial_\infty C_{n_{\mathrm{f}}+1} \cup \cdots \cup \partial_\infty C_{n_{\mathrm{f}}+n_{\mathrm{c}}}.$$

Then we have

$$C^\infty(\partial_\infty X) := C^\infty(\mathbb{R}/\ell_1\mathbb{Z}) \oplus \cdots \oplus C^\infty(\mathbb{R}/\ell_{n_{\mathrm{f}}}\mathbb{Z}) \oplus \mathbb{C}^{n_{\mathrm{c}}},$$

and similarly for $L^2(\partial_\infty X)$.

In Section 2, $R^0_{Y_j}(s)$ denoted the pullback of the model resolvent in the parametrix construction. Carrying on with this notation, we also define the model Poisson operators

$$E^0_{Y_j}(s) : C^\infty(\partial_\infty Y_j) \to L^2(Y_j),$$

and scattering matrices

$$S^0_{Y_j}(s) : C^\infty(\partial_\infty Y_j) \to C^\infty(\partial_\infty Y_j).$$

Similarly, for the cusp ends we have the Poisson kernels

$$E^0_{C_j}(s) : \mathbb{C} \to L^2(C_j).$$

There is no analog of the model scattering matrix for a cusp; see [Borthwick 2007, Section 7.5] for an explanation of this.

The scattering matrix $S_P(s)$ is defined as a map on $C^\infty(\partial_\infty X)$, which we can write as

$$S_P(s) = \begin{pmatrix} S^{\mathrm{ff}}(s) & S^{\mathrm{fc}}(s) \\ S^{\mathrm{cf}}(s) & S^{\mathrm{cc}}(s) \end{pmatrix}, \tag{3-1}$$

where the blocks are split between the "funnel-type" ends Y_j and the cusps C_j. The block $S^{\mathrm{ff}}(s)$ is a matrix of pseudodifferential operators; all other blocks have finite rank. To define a scattering determinant, we normalize using the background operator

$$S_0(s) = \begin{pmatrix} S^0_Y(s) & 0 \\ 0 & I \end{pmatrix}, \quad \text{where } S^0_Y(s) = S^0_{Y_1}(s) \oplus \cdots \oplus S^0_{Y_{n_{\mathrm{f}}}}(s).$$

The relative scattering determinant is then defined by

$$\tau(s) = \det S_P(s) S_0(s)^{-1}. \tag{3-2}$$

The poles of the background scattering matrix $S_0(s)$ define a background resonance set

$$\mathcal{R}_0 = \bigcup_{j=1}^{n_{\mathrm{f}}} \begin{cases} \mathcal{R}_{F_{\ell_j}} & \text{for a funnel end,} \\ \mathcal{R}_{\mathbb{H}} & \text{for a hyperbolic planar end.} \end{cases} \tag{3-3}$$

For $* = 0$ or P let $H_*(s)$ denote the Hadamard product over $\mathcal{R}_*$,

$$H_*(s) := \prod_{\zeta \in \mathcal{R}_*} (1 - s/\zeta) e^{s/\zeta + s^2/(2\zeta^2)}.$$

Theorem 2.2 implies that the product for $H_P(s)$ converges, and for $H_0(s)$ this is clear from the definition of $\mathcal{R}_0$.

Proposition 3.1. *For $P = \Delta_g + V$, the relative scattering determinant admits a factorization*

$$\tau(s) = e^{q(s)} \frac{H_P(1-s)}{H_P(s)} \frac{H_0(s)}{H_0(1-s)},$$

where $q(s)$ is a polynomial of degree at most 2.

Proof. If the ends Y_j are all hyperbolic funnels, then Guillopé and Zworski [1997, Proposition 3.7] proved the factorization formula of with $q(s)$ a polynomial of degree at most 4. The first part of the proof, the characterization of the divisor of $\tau(s)$ obtained in [Guillopé and Zworski 1997, Proposition 2.14], remains valid if hyperbolic planar ends are included.

To extend the more difficult part of the argument, which is the estimate that shows $q(s)$ is polynomial, we require only the extra estimates on model terms given in (2-2), (2-3), (2-4), and (2-5). With these estimates one can easily extend the proof of [Guillopé and Zworski 1997, Proposition 3.7]. We refer the reader also to [Borthwick 2007, Section 10.5], for an expository treatment of these details.

To see that the maximal order of $q(s)$ is 2, we could prove an estimate analogous to [Borthwick 2008, Lemma 5.2]. However, we will be proving a sharper version of this estimate later in this paper. From (4-12) in the proof of Theorem 4.1, it will follow that for some sequence $a_i \to \infty$,

$$\log|\tau(s)| \le O(a_i^2) \quad \text{for } |s-\tfrac{1}{2}| = a_i,\ |\arg(s-\tfrac{1}{2})| \le \tfrac{1}{2}\pi - \delta.$$

Because the Hadamard products $H_*(s)$ have order 2, this implies a bound $|q(s)| = O(|s|^{2+\varepsilon})$ in the sector $|\arg(s-\frac{1}{2})| \le \frac{1}{2}\pi - \delta$. Hence $q(s)$ has degree at most 2, since it is already known to be polynomial. (The derivations leading to (4-12) require only that $q(s)$ is polynomial, so this argument is not circular.) □

To apply the factorization of $\tau(s)$ to resonance counting we introduce the relative scattering phase of P, defined as

$$\sigma(\xi) := \frac{i}{2\pi} \log \tau(\tfrac{1}{2} + i\xi), \tag{3-4}$$

with branches of the log chosen so that $\sigma(\xi)$ is continuous and $\sigma(0) = 0$. By the properties of the relative scattering matrix, $\sigma(\xi)$ is real and $\sigma(-\xi) = -\sigma(\xi)$.

To state the relative counting formula, we let N_0 denote the counting function associated to $\mathcal{R}_0$,

$$N_0(t) := \#\big\{\zeta \in \mathcal{R}_0 : |\zeta - \tfrac{1}{2}| \le t\big\},$$

and $\widetilde{N}_0(a)$ the corresponding regularized counting function.

Corollary 3.2. *As $a \to \infty$,*

$$\widetilde{N}_P(a) - \widetilde{N}_0(a) = 4\int_0^a \frac{\sigma(t)}{t}\,dt + \frac{2}{\pi}\int_0^{\pi/2} \log|\tau(\tfrac{1}{2} + ae^{i\theta})|\,d\theta + O(\log a). \tag{3-5}$$

The proof is by contour integration of $\tau'/\tau(s)$ around a half-circle centered at $s = \frac{1}{2}$. See [Borthwick 2010, Proposition 3.2] for the details of the derivation of (3-5) from Proposition 3.1. This is the analog of a formula developed by Froese [1998] for Schrödinger operators in the Euclidean setting.

The other consequence we need from Proposition 3.1 is essentially also already proven. To analyze the first term on the right side of (3-5), we will invoke the Weyl-type asymptotics satisfied by the scattering phase:

Theorem 3.3 (Guillopé and Zworski). *As $\xi \to +\infty$,*

$$\sigma(\xi) = \Big(\frac{1}{4\pi}\,0\text{-vol}(X,g) - \frac{n_{\mathrm{hp}}}{2}\Big)\xi^2 - \frac{n_{\mathrm{c}}}{\pi}\xi\log\xi + O(\xi),$$

where n_{hp} denotes the number of the Y_j that are hyperbolic planar.

For surfaces with hyperbolic funnel or cusp ends, this result was established by Guillopé and Zworski [1997, Theorem 1.5]. As in the other cases discussed above, the modifications needed to adapt the proof to our slightly more general setting are fairly simple. The first point is that the addition of a compactly supported potential V does not change the argument at all, since it does not affect the leading term in the wave trace asymptotics as derived in [Guillopé and Zworski 1997, Lemma 6.2]. The second issue is that we allow hyperbolic planar ends in addition to funnels. However, for $|t| < \ell$ the restriction to the diagonal of the wave kernel on a model funnel F_ℓ is identical to that of $\mathbb{H}^2$. This is the content of [Guillopé and Zworski 1997, Equation (6.1)]. So hyperbolic planar ends may also be included without modifying the argument. Such ends do affect the final calculation, however, because 0-vol$(\mathbb{H}^2) = -2\pi$ whereas the model funnels had 0-vol$(F_\ell) = 0$. This difference accounts for the n_{hp} term.

4. Scattering determinant asymptotics

To state the asymptotic estimate for the scattering determinant contribution to the resonance counting formula (3-5), we introduce the following constants. If Y_j is a funnel with parameters ℓ_j, b_j, then we set

$$B(Y_j) := \frac{4}{\pi} \int_0^{\pi/2} \int_0^\infty \frac{[I(xe^{i\theta}, \ell_j, b_j)]_+}{x^3}\, dx\, d\theta - \frac{\ell_j}{4},$$

where $I(\alpha, \ell, r)$ was defined in (1-11). If Y_j is a hyperbolic planar end with parameter b_j, then

$$B(Y_j) := \frac{4}{\pi} \int_0^{\pi/2} \int_0^\infty \frac{[H(xe^{i\theta}, b_j)]_+}{x^3}\, dx\, d\theta,$$

where $H(\alpha, \ell, r)$ was defined in (1-14). The cusps do not contribute to the asymptotics of $\tau(s)$ to leading order, so we make no analogous definition for C_j.

Theorem 4.1. *For* (X, g) *a surface with hyperbolic ends as in* (1-3), *there exists an unbounded set* $\Lambda \subset [1, \infty)$ *such that*

$$\frac{2}{\pi} \int_0^{\pi/2} \log\left|\tau\left(\tfrac{1}{2} + ae^{i\theta}\right)\right| d\theta \le \sum_{j=1}^{n_{\mathrm{f}}} B(Y_j) a^2 + o(a^2) \quad \textit{for all } a \in \Lambda.$$

Before undertaking the proof of Theorem 4.1, we will show how this theorem leads to the proof of the main result stated in Section 1:

Proof of Theorem 1.4. Starting from the counting formula from Corollary 3.2, we apply Theorem 3.3 to the scattering phase term and Theorem 4.1 to the scattering determinant contribution. This yields

$$\widetilde{N}_P(a) \le \widetilde{N}_0(a) + \frac{1}{2\pi}\, \text{0-vol}(X, g) a^2 + \sum_{j=1}^{n_{\mathrm{f}}} B(Y_j) a^2 + o(a^2), \tag{4-1}$$

as $a \to \infty$. From the explicit definition (3-3) of $\mathcal{R}_0$, we see that

$$\frac{N_0(t)}{t^2} \sim \sum_{j=1}^{n_{\mathrm{f}}} \begin{cases} 1 & \text{for a hyperbolic planar end,} \\ \ell_j/4 & \text{for a funnel end,} \end{cases}$$

and so $\widetilde{N}_0(a)$ satisfies the same asymptotic. Also, we have

$$\text{0-vol}(X, g) = \text{vol}(X_{\text{c}}, g) + \sum_{j=1}^{n_{\text{f}}} \text{0-vol}(Y_j, g).$$

The 0-volumes of the Y_j are easily computed. For a hyperbolic planar end,

$$\text{0-vol}(Y_j, g) = 2\pi \underset{\varepsilon\to 0}{\text{FP}} \int_{b_j}^{\log(2/\varepsilon)} \sinh r\, dr = -2\pi \cosh b_j,$$

and for a funnel end,

$$\text{0-vol}(Y_j, g) = \ell_j \underset{\varepsilon\to 0}{\text{FP}} \int_{b_j}^{\log(2/\varepsilon)} \cosh r\, dr = -\ell_j \sinh b_j.$$

By the formulas (1-13) and (1-10) for $A(Y_j)$, we see that (4-1) is equivalent to the claimed estimate. □

The derivation of Theorem 1.1 from Theorem 1.4 was already explained in Section 1. To prove Corollary 1.2 we simply recall a few details of the proof of the lower bound in Guillopé and Zworski [1997, Theorem 1.3]. For a test function $\phi \in C_0^\infty(\mathbb{R}_+)$ with $\phi \geq 0$ and $\phi(1) > 0$, we have estimates

$$|\hat{\phi}(\xi)| \leq C_k(1+|\xi|)^{-k-2} \quad \text{for } k \in \mathbb{N} \text{ and } \operatorname{Im}\xi \leq 0.$$

Pairing the distributional Poisson formula [Guillopé and Zworski 1997, Theorem 5.7] with $\lambda\phi(\lambda\,\cdot\,)$ yields

$$|\text{0-vol}(X, g)|\,\lambda^2 \leq C_k \int_0^\infty (1+r)^{-k-3} N_P(\lambda r)\, dr.$$

If we have $N_P(t) \leq At^2$ for $t \geq 1$, then splitting the integral at a gives

$$|\text{0-vol}(X, g)|\,\lambda^2 \leq C_k\big(N(\lambda a) + A\lambda^2 a^{-k}\big).$$

Setting $t = \lambda a$, we have

$$N(t) \geq \big(c_k |\text{0-vol}(X, g)|\, a^{-2} - Aa^{-2-k}\big)\, t^2,$$

and optimizing with respect to a then yields

$$N(t) \geq c_k |\text{0-vol}(X, g)|^{1+k/2} A^{-k/2}.$$

Corollary 1.2 is then proven by substituting the constant obtained in Theorem 1.1 for A.

The rest of this section is devoted to the proof of Theorem 4.1. To produce a formula convenient for estimation, we introduce cutoff functions as follows. Fix some $\eta \in (0, 1)$. For $j = 1, \ldots, n_{\text{f}} + n_{\text{c}}$ and $k = 1, 2$, we define $\chi_k^j \in C^\infty(X)$ so that $\chi_k^j = 1$ outside the j-th end (Y_j or C_j), and inside the j-th end we have

$$\chi_k^j = \begin{cases} 0 & \text{for } r \geq b_j + (k+1)\eta, \\ 1 & \text{for } r \leq b_j + k\eta. \end{cases} \tag{4-2}$$

Note that $\chi_2^j = 1$ on the support of χ_1^j, as illustrated in Figure 4.

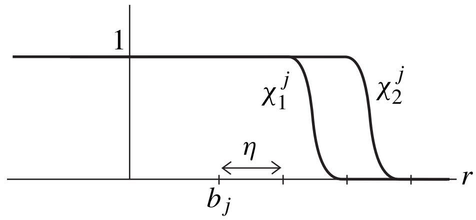

Figure 4. The cutoff functions χ_k^j in the j-th end.

Proposition 4.2. *With cutoffs defined as in* (4-2), *we have*

$$S_X(s)S_0(s)^{-1} = I + Q(s),$$

where the components of $Q(s)$, in terms of the block decomposition introduced in (3-1), *are*

$$\begin{aligned}
Q_{ij}^{\mathrm{ff}}(s) &= (2s-1)E_{Y_i}^0(s)^t[\Delta_{Y_i}, \chi_2^i]R_P(s)[\Delta_{Y_j}, \chi_1^j]E_{Y_j}^0(1-s),\\
Q_{ij}^{\mathrm{cf}}(s) &= (2s-1)E_{C_i}^0(s)^t[\Delta_{C_i}, \chi_2^i]R_P(s)[\Delta_{Y_j}, \chi_1^j]E_{Y_j}^0(1-s),\\
Q_{ij}^{\mathrm{fc}}(s) &= -(2s-1)E_{Y_i}^0(s)^t[\Delta_{Y_i}, \chi_2^i]R_P(s)[\Delta_{C_j}, \chi_1^j]E_{C_j}^0(s),\\
Q_{ij}^{\mathrm{cf}}(s) &= -(2s-1)E_{C_i}^0(s)^t[\Delta_{C_i}, \chi_2^i]R_P(s)[\Delta_{C_j}, \chi_1^j]E_{C_j}^0(s),
\end{aligned}$$

Proof. One can characterize the scattering matrix $S_X(s)$ through the boundary behavior of solutions of $(\Delta_g - s(1-s))u = 0$. For $\psi \in C^\infty(\partial_\infty X)$ and $\operatorname{Re} s \geq \frac{1}{2}$, with $s \neq \mathbb{N}/2$, there is a unique generalized eigenfunction $u \in C^\infty(X)$ with the asymptotic behavior

$$u \sim \rho_{\mathrm{f}}^{1-s}\rho_{\mathrm{c}}^{-s}\psi + \rho_{\mathrm{f}}^{s}\rho_{\mathrm{c}}^{s-1}S_X(s)\psi. \tag{4-3}$$

For hyperbolic surfaces with cusps, a proof is given in Borthwick [2007, Proposition 7.13]. The essential analysis takes place in the ends, so including smooth metric or potential perturbations within K requires only trivial modifications to the proof. Likewise, hyperbolic planar ends may be included without much change to the argument.

Suppose $f_j \in C^\infty(\partial_\infty Y_j)$. Then we can use the model Poisson kernel $E_{Y_j}^0(s)$ to create a partial solution $(1-\chi_1^j)E_{Y_j}^0(s)f_j$ supported in Y_j. As $\rho \to 0$ in Y_j, this function has the asymptotic behavior

$$(1-\chi_1^j)E_{Y_j}^0(s)f_j \sim \frac{1}{2s-1}(\rho_{\mathrm{f}}^{1-s}f_j + \rho_{\mathrm{f}}^{s}S_{Y_j}^0(s)f_j). \tag{4-4}$$

To create a full solution, we will take the ansatz

$$u = (1-\chi_1^j)E_{Y_j}^0(s)f_j + u'$$

and then solve $(\Delta_g - s(1-s))u = 0$ for u' by applying the resolvent. The result is

$$u' = R_P(s)[\Delta_{Y_j}, \chi_1^j]E_{Y_j}^0(s)f_j.$$

In the end Y_i, we can use the fact that $(1-\chi_2^i)[\Delta_{Y_j},\chi_1^j]=0$ to deduce

$$(\Delta_{Y_i}-s(1-s))(1-\chi_2^i)u' = -[\Delta_{Y_i},\chi_2^i]u',$$

and hence that

$$(1-\chi_2^i)u' = -R^0_{Y_i}(s)[\Delta_{Y_i},\chi_2^i]R_P(s)[\Delta_{Y_j},\chi_1^j]E^0_{Y_j}(s)f_j.$$

This gives the asymptotic behavior in Y_i:

$$u' \sim -\rho_{\mathrm{f}}^s E^0_{Y_i}(s)^t[\Delta_{Y_i},\chi_2^i]R_P(s)[\Delta_{Y_j},\chi_1^j]E^0_{Y_j}(s)f_j. \tag{4-5}$$

By comparing the asymptotics (4-4) and (4-5) to the general form (4-3), we see that

$$S^{\mathrm{ff}}_{ij}(s) = \delta_{ij}S^0_{Y_j}(s) - (2s-1)E^0_{Y_i}(s)^t[\Delta_{Y_i},\chi_2^i]R_P(s)[\Delta_{Y_j},\chi_1^j]E^0_{Y_j}(s)$$

We then obtain $Q^{\mathrm{ff}}_{ij}(s)$ by noting that

$$E^0_{Y_j}(s)S^0_{Y_j}(s)^{-1} = -E^0_{Y_j}(1-s).$$

To find $Q^{\mathrm{cf}}_{ij}(s)$ we use the same setup starting from $f_j\in C^\infty(\partial_\infty Y_j)$, but then analyze u' by restricting to the cusp end C_i. This yields

$$(1-\chi_2^i)u' = -R^0_{C_i}(s)[\Delta_{C_i},\chi_2^i]R_P(s)[\Delta_{Y_j},\chi_1^j]E^0_{Y_j}(s)f_j.$$

The asymptotic behavior in C_i is given by

$$(1-\chi_2^i)u' \sim -\rho^{s-1}E^0_{C_i}(s)^t[\Delta_{C_i},\chi_2^i]R_P(s)[\Delta_{Y_j},\chi_1^j]E^0_{Y_j}(s)f_j,$$

so that

$$S^{\mathrm{cf}}_{ij}(s) = -(2s-1)E^0_{C_i}(s)^t[\Delta_{C_i},\chi_2^i]R_P(s)[\Delta_{Y_j},\chi_1^j]E^0_{Y_j}(s).$$

Next take $a_j\in C^\infty(\partial_\infty C_j)=\mathbb{C}$. Since $E^0_{C_j}(s;r)=\rho_{\mathrm{c}}^{-s}/(2s-1)$, our ansatz for a generalized eigenfunction satisfying (4-3) starts from

$$(1-\chi_1^j)E^0_{C_j}(s)a_j \sim \frac{1}{2s-1}\rho_{\mathrm{c}}^{-s}a_j.$$

The corresponding generalized eigenfunction is

$$u = (1-\chi_1^j)E^0_{C_j}(s)a_j + u', \quad \text{where } u' = R_P(s)[\Delta_{C_\infty},\chi_1^j]E^0_{C_j}(s)a_j.$$

arguing as above, we find that

$$u' \sim -\rho_{\mathrm{f}}^s E^0_{Y_i}(s)^t[\Delta_{Y_i},\chi_2^i]R_P(s)[\Delta_{C_\infty},\chi_1^j]E^0_{C_j}(s)a_j$$

in the funnel Y_i, and

$$u' \sim -\rho_{\mathrm{c}}^{1-s}E^0_{C_i}(s)^t[\Delta_{C_i},\chi_2^i]R_P(s)[\Delta_{C_j},\chi_1^j]E^0_{C_j}(s)a_j$$

in the cusp C_i. We can then read off the matrix elements, $S^{\mathrm{fc}}_{ij}(s)$ and $S^{\mathrm{cc}}_{ij}(s)$, as above. □

In conjunction with the cutoffs defined in (4-2), we introduce projections $\mathbb{1}_k^j$ on $L^2(X, dg)$, where

$$\mathbb{1}_k^j f = \begin{cases} f & \text{for } r \in [b_j + k\eta, b_j + (k+1)\eta] \text{ in end } j, \\ 0 & \text{otherwise.} \end{cases} \tag{4-6}$$

As with the cutoffs, these projections depend on b_j and also on the choice of $\eta > 0$. We then introduce operators on $L^2(X, dg)$ given by

$$G_j(s) := (2s-1)\mathbb{1}_1^j E^0_{Y_j}(1-s) E^0_{Y_j}(s)^t \mathbb{1}_2^j \quad \text{for } j = 1, \dots, n_{\mathrm{f}}, \tag{4-7}$$

$$G_j(s) := -(2s-1)\mathbb{1}_1^j E^0_{C_j}(s) E^0_{C_j}(s)^t \mathbb{1}_2^j \quad \text{for } j = n_{\mathrm{f}}+1, \dots, n_{\mathrm{f}}+n_{\mathrm{c}}. \tag{4-8}$$

Proposition 4.3. *The relative scattering phase is bounded by*

$$\log|\tau(s)| \le \sum_{j=1}^{n_{\mathrm{f}}+n_{\mathrm{c}}} \log\det(I + C(\eta, \varepsilon)|G_j(s)|)$$

for $\operatorname{Re} s \ge \frac{1}{2}$ *with* $\operatorname{dist}(s(1-s), \sigma(P)) \ge \varepsilon$.

Proof. In the formula for the relative scattering matrix given in Proposition 4.2, we can write $Q(s)$ as the composition of three operators,

$$Q(s) : L^2(\partial_\infty X) \xrightarrow{Q_3} L^2(X, dg) \xrightarrow{Q_2} L^2(X, dg) \xrightarrow{Q_1} L^2(\partial_\infty X),$$

where

$$Q_1 := \sum_{j=1}^{n_{\mathrm{f}}} E^0_{Y_j}(s)^t \mathbb{1}_2^j + \sum_{j=n_{\mathrm{f}}+1}^{n_{\mathrm{f}}+n_{\mathrm{c}}} E^0_{C_j}(s)^t \mathbb{1}_2^j, \quad Q_2 := \sum_{i,j=1}^{n_{\mathrm{f}}+n_{\mathrm{c}}} [\Delta_g, \chi_2^i] R_P(s) [\Delta_g, \chi_1^j],$$

$$Q_3\big|_{L^2(\partial_\infty Y_j)} := \mathbb{1}_1^j E^0_{Y_j}(1-s), \quad Q_3\big|_{L^2(\partial_\infty C_j)} := \mathbb{1}_1^j E^0_{C_j}(s).$$

By the cyclicity of the trace,

$$\tau(s) = \det(I + Q(s)) = \det(I + Q_2 \circ Q_3 \circ Q_1).$$

Under the assumptions $\operatorname{Re} s \ge \frac{1}{2}$ with $\operatorname{dist}(s(1-s), \sigma(P)) \ge \varepsilon$, we can apply the spectral theorem and standard elliptic estimates to prove that $\|Q_2\| \le C(\eta, \varepsilon)$. By the Weyl estimate this then gives

$$|\tau(s)| \le \prod_{j=1}^{\infty} (1 + C(\eta, \varepsilon)\mu_j(Q_3 \circ Q_1)) = \det(1 + C(\eta, \varepsilon)|Q_3 \circ Q_1|)$$

The result follows immediately from

$$Q_3 \circ Q_1 = G_1 \oplus \cdots \oplus G_{n_{\mathrm{f}}+n_{\mathrm{c}}},$$

where the $G_j(s)$ are given by (4-7) and (4-8). □

The right side of the estimate from Proposition 4.3 is always positive. It is therefore impossible to obtain a sharp estimate by this approach in cases where the leading asymptotic behavior of $\log|\tau(s)|$ is negative. The extended funnel, whose resonance asymptotics are studied in Section 8, gives an example of this situation.

Proof of Theorem 4.1. Let $\mathcal{R}_0$ be the background resonance set as defined in (3-3). To avoid poles, we will restrict our attention to radii in the set

$$\Lambda := \big\{a \geq 1 : \operatorname{dist}\big(\{|s-\tfrac{1}{2}|=a\}, \mathcal{R}_0 \cup \mathcal{R}_P\big) \geq a^{-3}\big\}.$$

Since $N_0(t)$ and $N_P(t)$ are $O(t^2)$, the density of Λ in $[1,r)$ approaches 1 as $r\to\infty$.

If we assume that $0\leq\theta\leq\pi/2-\varepsilon a^{-2}$, then $s=\frac{1}{2}+ae^{i\theta}$ will satisfy the hypothesis that

$$\operatorname{dist}(s(1-s),\sigma(P)) \geq \varepsilon$$

for Proposition 4.3. We also assume $a\in\Lambda$ throughout this argument. If Y_j is a funnel end, then Proposition 6.3 gives

$$\log\det(I+C(\eta,\varepsilon)|G_j(\tfrac{1}{2}+ae^{i\theta})|) \leq \kappa_j(\theta, b_j+4\eta)a^2 + C(\eta,\varepsilon,b_j)a\log a, \tag{4-9}$$

where

$$\kappa_j(\theta,r) := 2\int_0^\infty \frac{[I(xe^{i\theta},\ell_j,r)]_+}{x^3}\,dx - \tfrac{1}{2}\ell_j\sin^2\theta,$$

If Y_j is hyperbolic planar, the corresponding estimate follows from [Borthwick 2010, Proposition 5.4], with

$$\kappa_j(\theta,r) := 2\int_0^\infty \frac{[H(xe^{i\theta},r)]_+}{x^3}\,dx,$$

(A slight modification of the original proof is required, replacing the assumption $a\in\mathbb{N}$ with an estimate based on $\operatorname{dist}(\frac{1}{2}-ae^{i\theta},-\mathbb{N})$.)

For a cusp end C_j, it is easy to estimate directly since

$$E^0_{C_j}(s) = \frac{e^{sr}}{2s-1},$$

which gives

$$G_j(s;r,\theta,r',\theta') = -\frac{1}{2s-1}\mathbb{1}_{j,1}(r)e^{s(r+r')}\mathbb{1}_{j,2}(r').$$

This operator has rank one, so that

$$\det(I+c|G_j(s)|) = 1+c\mu_1(G_j(s)),$$

where the sole singular value is given by

$$\mu_1(G_j(s)) = \frac{1}{|2s-1|}\Big(\int_{b_j+\eta}^{b_j+2\eta} e^{2r\operatorname{Re}s}e^{-r}\,dr\Big)^{1/2}\Big(\int_{b_j+2\eta}^{b_j+3\eta} e^{2r\operatorname{Re}s}e^{-r}\,dr\Big)^{1/2}.$$

Hence we have

$$\det(I+c|G_j(\tfrac{1}{2}+ae^{i\theta})|) \leq 1+\frac{c}{2a}e^{2a(b_j+3\eta)}.$$

For a sufficiently large,

$$\log\det(I + C(\eta,\varepsilon)|G_j(\tfrac{1}{2} + ae^{i\theta})|) \le C(\eta,\varepsilon,b_j)a \quad \text{for all } |\theta| \le \pi/2. \tag{4-10}$$

From (4-9) and (4-10) we conclude that

$$\frac{\log|\tau(\frac{1}{2}+ae^{i\theta})|}{a^2} \le \sum_{j=1}^{n_{\mathrm{f}}} \kappa_j(\theta, b_j + 4\eta) + C(\eta,\varepsilon,b_j)a^{-1}\log a \tag{4-11}$$

for $a \in \Lambda$ and $0 \le \theta \le \pi/2 - \varepsilon a^{-2}$. Since the $\kappa_j(\theta, r)$ are uniformly continuous on $[0,\pi/2]\times[b_j, b_j+1]$, we can take $\eta \to 0$ in (4-11), to obtain

$$\frac{\log|\tau(\frac{1}{2}+ae^{i\theta})|}{a^2} \le \sum_{j=1}^{n_{\mathrm{f}}} \kappa_j(\theta, b_j) + o(a^2), \tag{4-12}$$

uniformly for $0 \le \theta \le \pi/2 - \varepsilon a^{-2}$.

By integrating the estimate (4-12) over θ, we obtain

$$\frac{2}{\pi}\int_0^{\pi/2-\varepsilon a^{-2}} \log|\tau(\tfrac{1}{2}+ae^{i\theta})|\,d\theta \le \sum_{j=1}^{n_{\mathrm{f}}} B(Y_j)a^2 + o(a^2).$$

It remains to fill in the small gap where $|\theta|$ is close to $\pi/2$. The factorization given by Proposition 3.1, together with the minimum modulus theorem [Boas 1954, Theorem 3.7.4], implies that for any $\eta > 0$,

$$|\tau(\tfrac{1}{2}+ae^{i\theta})| \le C_\eta \exp(a^{2+\eta}), \tag{4-13}$$

provided $a \in \Lambda$. (This was the reason that $\mathcal{R}_P$ was included in the definition of Λ.) Thus,

$$\frac{2}{\pi}\int_{\pi/2-\varepsilon a^{-2}}^{\pi/2} \log|\tau(\tfrac{1}{2}+ae^{i\theta})|\,d\theta = O(a^\eta \varepsilon),$$

and so this term can be absorbed into the $o(a^2)$ error. □

To conclude this section, we'll derive some uniform upper and lower bounds on the growth of $\tau(s)$ for $s \in \mathbb{C}$, refining the estimates that one could obtain directly from Proposition 3.1. These will prove useful in Section 7 and Section 8, in particular.

Lemma 4.4. *Let $\mathfrak{Q}$ denote the joint set of zeros and poles of $\tau(\frac{1}{2}+z)$ and $\tau(\frac{1}{2}-iz)$. Assuming $|z| \ge 1$ and $\operatorname{dist}(z,\mathfrak{Q}) > |z|^{-\beta}$ with $\beta > 2$, we have*

$$-c(\beta)|z|^2 \le \log|\tau(\tfrac{1}{2}+z)| \le C(\beta)|z|^2.$$

Proof. Since $\tau(\frac{1}{2}-z) = 1/\tau(\frac{1}{2}+z)$ and $\tau(\frac{1}{2}+\bar{z}) = \overline{\tau(\frac{1}{2}+z)}$, it suffices to prove the bounds for z in the first quadrant.

For $\operatorname{Re} z \ge \delta$ with $\delta > 0$, the upper bound is given in (4-11). As long as $\delta < 1$, the function $\tau(s)$ is analytic in the strip $\operatorname{Re} z \in [0,\delta]$. And since $\log|\tau(\frac{1}{2}+z)| = 1$ for $\operatorname{Re} z = 0$, the bound $\log|\tau(\frac{1}{2}+z)| = O(|z|^2)$ extends to the strip $\operatorname{Re} z \in [0,\delta]$ by (4-13) and the Phragmén–Lindelöf theorem.

To prove the lower bound, consider the Hadamard products appearing in the factorization of $\tau(s)$ given in Proposition 3.1. These products are of order 2 but not finite type, so applying the minimum modulus theorem directly would give $-\log|\tau(\frac{1}{2}+z)| = O(|z|^{2+\eta})$, away from the zeros. However, Lindelöf's theorem (see e.g., [Boas 1954, Theorem 2.10.1]) shows that products of the form $H_*(\frac{1}{2}+z)H_*(\frac{1}{2}\pm iz)$ are of finite type. In other words,

$$\log|H_*(\tfrac{1}{2}+z)H_*(\tfrac{1}{2}\pm iz)| \le C|z|^2,$$

as $|z|\to\infty$. Using these estimates, and their implications via the minimum modulus theorem, we can prove a lower bound

$$\log|\tau(\tfrac{1}{2}+z)| \ge -c(\beta)|z|^2 - \log|\tau(\tfrac{1}{2}\pm iz)|, \tag{4-14}$$

provided $\frac{1}{2}+z$ and $\frac{1}{2}\pm iz$ stay at least a distance $|z|^{-\beta}$ away from the sets $1-\mathcal{R}_{F_{\ell,r_0}}$ and $\mathcal{R}_{F_\ell}$, with $\beta>2$.

Assuming $\arg z\in[0,\pi/2]$, we already know $\log|\tau(\frac{1}{2}-iz)| \le C(\beta)|z|^2$ from above, provided $\frac{1}{2}-iz$ stays at least a distance $|z|^{-\beta}$ away from the sets $\mathcal{R}_{F_{\ell,r_0}}$ and $1-\mathcal{R}_{F_\ell}$. The lower bound in the first quadrant then follows from (4-14). □

5. Funnel eigenmodes

Let F_ℓ be a hyperbolic funnel of diameter ℓ. In geodesic coordinates $(r,\theta)\in\mathbb{R}_+\times S^1$, defined with respect to the closed geodesic neck, the metric is

$$g_0 = dr^2 + \cosh^2 r\,\frac{d\theta^2}{\omega^2}, \quad \text{where } \omega := \frac{2\pi}{\ell}. \tag{5-1}$$

The Laplacian is given by

$$\Delta_{F_\ell} = -\partial_r^2 - \tanh r\,\partial_r - \frac{\omega^2}{\cosh^2 r}\partial_\theta^2. \tag{5-2}$$

In this section we will consider asymptotic properties of the Fourier modes of generalized eigenfunctions of Δ_{F_ℓ}.

The restriction of eigenvalue equation $(\Delta_{F_\ell} - s(1-s))u=0$ to the k-th Fourier mode, $u=w(r)e^{ik\theta}$, yields the equation

$$-\partial_r^2 w - \tanh r\,\partial_r w + \left(\frac{k^2\omega^2}{\cosh^2 r} - s(1-s)\right)w = 0. \tag{5-3}$$

This is essentially a hypergeometric equation. With respect to the symmetry $r\mapsto -r$, we have an even solution,

$$w_k^+(s;r) := (\cosh r)^{i\omega k}\mathbf{F}(\tfrac{1}{2}(s+i\omega k), \tfrac{1}{2}(1-s+i\omega k); \tfrac{1}{2}; -\sinh^2 r), \tag{5-4}$$

and an odd solution,

$$w_k^-(s;r) := \sinh r(\cosh r)^{i\omega k}\mathbf{F}(\tfrac{1}{2}(1+s+i\omega k), \tfrac{1}{2}(2-s+i\omega k); \tfrac{3}{2}; -\sinh^2 r). \tag{5-5}$$

(We follow Olver's convention in using $\mathbf{F}(a,b;c;z) := F(a,b;c;z)/\Gamma(c)$, where $F(a,b;c;z)$ is the standard Gauss hypergeometric function.)

By symmetry, we can and will assume that $k \geq 0$. If we substitute $w = (\cosh r)^{-1/2} U$ and introduce the parameter α defined by $s = \frac{1}{2} + k\alpha$, the coefficient equation (5-3) becomes

$$\partial_r^2 U = (k^2 f + g) U, \tag{5-6}$$

where

$$f := \frac{\omega^2 + \alpha^2 \cosh^2 r}{\cosh^2 r} \quad \text{and} \quad g := \frac{1}{4\cosh^2 r}.$$

This equation has turning points when $\alpha = \pm i\omega / \cosh r$. We will restrict our attention to $\arg \alpha \in [0, \frac{1}{2}\pi]$, so that we only consider the upper turning point. The Liouville transformation involves a new variable ζ defined by integrating

$$\sqrt{\zeta}\, d\zeta := \sqrt{f}\, dr, \tag{5-7}$$

on a contour that starts from the upper turning point. Integrating (5-7) yields

$$(2/3)\zeta^{3/2} = \phi, \tag{5-8}$$

where $\phi(\alpha, r)$, the integral of $\sqrt{f}\, dr$ from the turning point, is given explicitly by

$$\phi(\alpha, r) := \alpha \log\left(\frac{\alpha \sinh r + \sqrt{\omega^2 + \alpha^2 \cosh^2 r}}{\sqrt{\omega^2 + \alpha^2}}\right) + \frac{i\omega}{2} \log\left(\frac{\sqrt{\omega^2 + \alpha^2 \cosh^2 r} - i\omega \sinh r}{\sqrt{\omega^2 + \alpha^2 \cosh^2 r} + i\omega \sinh r}\right) + \phi_0(\alpha) \tag{5-9}$$

for $\alpha \neq i\omega$, where

$$\phi_0(\alpha) = \phi(\alpha; 0) = -\tfrac{1}{2}\pi(i\alpha + \omega). \tag{5-10}$$

By continuity, the definition of ϕ extends to $\alpha = i\omega$, with

$$\phi(i\omega, r) = i\omega \log \cosh r.$$

To complete the Liouville transformation, we set $W = (f/\zeta)^{1/4} U$, so that (5-6) becomes an approximate Airy equation,

$$\partial_\zeta^2 W = (k^2 \zeta + \psi) W, \tag{5-11}$$

with the extra term given by

$$\psi = \frac{\zeta}{4f^2} \partial_r^2 f - \frac{5\zeta}{16 f^3} (\partial_r f)^2 + \frac{\zeta g}{f} + \frac{5}{16\zeta^2}. \tag{5-12}$$

The solutions of (5-11) are of the form

$$W_\sigma := \mathrm{Ai}(k^{2/3} e^{2\pi i \sigma/3} \zeta) + h_\sigma(k, \alpha, r), \tag{5-13}$$

where $\sigma = 0$ or ± 1, and the error term satisfies the differential equation

$$\partial_\zeta^2 h_\sigma - k^2 \zeta h_\sigma = (h_\sigma + \mathrm{Ai}(k^{2/3} e^{2\pi i \sigma/3} \zeta)) \psi. \tag{5-14}$$

Using methods from Olver [1974] we can control this error term.

Lemma 5.1. *The error equation* (5-14) *admits solutions that satisfy* $\lim_{r\to\infty} h_\sigma(r) = 0$ *and*

$$|h_\sigma| \le Ck^{-1}|\alpha|^{-2/3}(1+|k\phi|^{1/6})^{-1}e^{(-1)^{\sigma+1}k\,\mathrm{Re}\,\phi},$$

with C independent of r, k and α.

We will defer the rather technical proof of Lemma 5.1 to the end of this section, in order to concentrate on the implications of (5-13). The asymptotics of the Airy function are well known; see for example [Olver 1974, Section 11.8]. Uniformly for $|\arg z| < \pi - \varepsilon$, we have

$$\mathrm{Ai}(z) = \frac{1}{2\pi^{1/2}} z^{-1/4}\exp(-\tfrac{2}{3}z^{3/2})(1+O(|z|^{-3/2})). \tag{5-15}$$

And uniformly for $|\arg z| \ge \frac{1}{3}\pi + \varepsilon$,

$$\mathrm{Ai}(z) = \frac{1}{\pi^{1/2}}(-z)^{-1/4}\cos(\tfrac{2}{3}(-z)^{3/2} - \tfrac{1}{4}\pi)(1+O(|z|^{-3/2})). \tag{5-16}$$

These asymptotics make it convenient to introduce a pair of solutions of the eigenvalue equation (5-3) defined by

$$w_\sigma = 2\pi^{1/2}e^{i\pi\sigma/6}k^{1/6}\zeta^{1/4}(\omega^2+\alpha^2\cosh^2 r)^{-1/4}W_\sigma, \tag{5-17}$$

where W_σ is the ansatz (5-13) for $\sigma = 0$ or 1.

Proposition 5.2. *Consider the solutions of the equation*

$$(\Delta_{F_\ell} - \tfrac{1}{4} - k^2\alpha^2)e^{ik\theta}w_\sigma(r) = 0$$

given by (5-17) *with $\sigma = 0$ or* 1. *Assuming $k \ge 1$ and* $\arg\alpha \in [0, \frac{1}{2}\pi - \varepsilon]$, *we have asymptotics*

$$w_\sigma = (\omega^2+\alpha^2\cosh^2 r)^{-1/4}\exp\big((-1)^{\sigma+1}k\phi\big)\big(1+O(|k\alpha|^{-1})\big), \tag{5-18}$$

with constants that depend only on ε. In addition, for $\arg\alpha \in [0, \pi/2]$ *and $|k\alpha|$ sufficiently large, we have the upper bounds*

$$|w_\sigma| \le Ck^{1/6}\exp((-1)^{\sigma+1}k\,\mathrm{Re}\,\phi), \tag{5-19}$$

and the lower bound

$$|w_0| \ge ce^{-k\,\mathrm{Re}\,\phi}. \tag{5-20}$$

Proof. The assumption that $\arg\alpha \in [0, \pi/2-\varepsilon]$ implies that $\arg\zeta \in [-2\pi/3, \pi/3-\varepsilon]$, so that (5-15) applies to both w_0 and w_1 in this case. It also implies that $|\phi| \ge c(\varepsilon)(|\alpha|+1)$, so that the error term $O(|w|^{-3/2})$ from (5-15) becomes $O(|k\alpha|^{-1})$ when applied to $|w| = k^{2/3}|\zeta|$. In combination with Lemma 5.1, this proves (5-18), and also (5-19) and (5-20) in the case where $\arg\alpha$ is bounded away from $\pi/2$.

If $\arg\alpha \in [\pi/2-\varepsilon, \pi/2]$, then (5-15) and (5-16), together with Lemma 5.1, give the estimates

$$|k^{1/6}\zeta^{1/4}W_\sigma| \le C\exp\big((-1)^{\sigma+1}k\,\mathrm{Re}\,\phi\big), \tag{5-21}$$

and

$$|k^{1/6}\zeta^{1/4}W_0| \geq ce^{-k\,\mathrm{Re}\,\phi}, \tag{5-22}$$

If $|\omega^2+\alpha^2\cosh^2 r| \geq 1$, which bounds ϕ away from 0, then this gives (5-19) immediately. This leaves the case $|\omega^2+\alpha^2\cosh^2 r| \leq 1$, which puts ϕ close to zero. In this case, $\zeta \asymp (\omega^2+\alpha^2\cosh^2 r)$, so that $w_\sigma \asymp k^{1/6}W_\sigma$. Then if $|k\phi| \geq 1$ we can derive the estimates from (5-21) and (5-22), while for $|k\phi| \leq 1$ we simply note that W_σ is bounded and nonzero near the origin. ☐

Another detail we will need later is the asymptotic behavior of w_σ as $r \to \infty$.

Lemma 5.3. *For* $\mathrm{Re}\,\alpha \geq 0$, *as* $r \to \infty$,

$$w_0 \sim \alpha^{-1/2}e^{-k(\phi_0(\alpha)+\gamma(\alpha))}\rho^{1/2+k\alpha},$$

and

$$w_1 \sim \alpha^{-1/2}e^{k(\phi_0(\alpha)+\gamma(\alpha))}(\rho^{1/2-k\alpha}+i\rho^{1/2+k\alpha})$$

where $\rho := 2e^{-r}$, *and*

$$\gamma(\alpha) := \alpha\log\frac{2\alpha}{\sqrt{\omega^2+\alpha^2}}+\frac{i\omega}{2}\log\frac{\alpha-i\omega}{\alpha+i\omega}. \tag{5-23}$$

Proof. The results follow immediately from (5-15) and (5-16), in combination with the asymptotic

$$\phi(\alpha;r) = \alpha r+\phi_0(\alpha)+\alpha\log\frac{\alpha}{\sqrt{\omega^2+\alpha^2}}+\frac{i\omega}{2}\log\frac{\alpha-i\omega}{\alpha+i\omega}+O(r^{-1}), \tag{5-24}$$

as $r \to \infty$. ☐

We conclude the section with the proof of the error estimate that is the basis of Proposition 5.2 and Lemma 5.3.

Proof of Lemma 5.1. The cases of different σ are all very similar, so we consider only $\sigma = 0$. In this case combining the boundary condition with variation of parameters allows us to transform (5-14) into an integral equation:

$$h_0(k,\alpha,r) = \frac{2\pi e^{-i\pi/6}}{k^{2/3}}\int_r^\infty K_0(r,r')\psi(r')\big(h_0(k,\alpha,r')+\mathrm{Ai}(k^{2/3}\zeta(r'))\big)\frac{f(r')^{1/2}}{\zeta(r')^{1/2}}\,dr',$$

where

$$K_0(r,r') := \mathrm{Ai}(k^{2/3}\zeta(r'))\mathrm{Ai}(k^{2/3}e^{-2\pi i/3}\zeta(r)) - \mathrm{Ai}(k^{2/3}e^{-2\pi i/3}\zeta(r'))\mathrm{Ai}(k^{2/3}\zeta(r)).$$

Then, using the method of successive approximations as in [Olver 1974, Theorem 6.10.2], together with the bounds on the Airy function and its derivatives developed in [Olver 1974, Section 11.8], we obtain the bound

$$|h_0| \leq Ce^{-k\,\mathrm{Re}\,\phi}(1+k^{1/6}|\zeta|^{1/4})^{-1}(e^{ck^{-1}\Psi(r)}-1), \tag{5-25}$$

where

$$\Psi(r) := \int_r^\infty |\psi f^{1/2}\zeta^{-1/2}|\,dr'. \tag{5-26}$$

From (5-12), we compute

$$\psi f^{1/2}\zeta^{-1/2} = \left(\frac{\alpha^4\cosh^2 r + 4\alpha^2\omega^2\sinh^2 r - \omega^4}{4(\omega^2+\alpha^2\cosh^2 r)^{5/2}}\right)\zeta^{1/2}\cosh r + \frac{5}{16}\frac{(\omega^2+\alpha^2\cosh^2 r)^{1/2}}{\zeta^{5/2}\cosh r}. \tag{5-27}$$

The estimate must be broken into various regions. Fix some $c > 0$.

Case 1. Assume $|\alpha| \geq 1$ and $|\omega^2 + \alpha^2\cosh^2(r)| \geq c$. Under these conditions, we can estimate

$$|\phi| \asymp |\alpha|(r+1).$$

Then from (5-27), we find

$$|\psi f^{1/2}\zeta^{-1/2}| \leq C_1|\alpha|^{-2/3}e^{-2r}(r+1)^{1/3} + C_2|\alpha|^{-2/3}(r+1)^{-5/3}.$$

We easily conclude that for $|\alpha| \geq 1$,

$$\int_{|\omega^2+\alpha^2\cosh^2(r)|\geq c} |\psi f^{1/2}\zeta^{-1/2}|\,dr = O(|\alpha|^{-2/3}). \tag{5-28}$$

Case 2. Assume $|\alpha| \leq 1$ and $|\omega^2+\alpha^2\cosh^2(r)| \geq c$. The behavior of ϕ is now slightly more complicated, depending on the size of r relative to $|\alpha|$,

$$|\phi| \asymp \begin{cases} |\alpha| + e^{-r} & \text{for } |\alpha|\sinh r \leq 1, \\ |\alpha|(r+\log|\alpha|) & \text{for } |\alpha|\sinh r \geq 1. \end{cases}$$

In this case, we estimate (5-27) by

$$|\psi f^{1/2}\zeta^{-1/2}| \leq \begin{cases} C_1(|\alpha|+e^{-r})^{1/3}e^r + C_2e^{-r}(|\alpha|+e^{-r})^{-5/3} & \text{for } |\alpha|\sinh r \leq 1, \\ C_1(1+|\alpha|e^r)^{-3}|\alpha|^{1/3}(r+\log|\alpha|)^{1/3}e^r + C_2|\alpha|^{-5/2}r^{-5/3}e^{-r} & \text{for } |\alpha|\sinh r \geq 1. \end{cases}$$

It is then straightforward to bound, for $|\alpha| \leq 1$,

$$\int_{|\omega^2+\alpha^2\cosh^2(r)|\geq c} |\psi f^{1/2}\zeta^{-1/2}|\,dr = O(|\alpha|^{-2/3}). \tag{5-29}$$

Case 3. Assume $|\omega^2 + \alpha^2\cosh^2(r)| \leq c$. In this case we are near the turning point, where ϕ and ζ are small. Since $|\omega^2 + \alpha^2\cosh^2(r)| \leq c$ implies $|\alpha|^2 \leq \omega^2 + c$, we are only concerned with small $|\alpha|$ here. We proceed as in [Borthwick 2010, Appendix]. In the coordinate $z = \sinh r$, the turning point occurs at

$$z_0 = \sqrt{-1 - \frac{\omega^2}{\alpha^2}}.$$

Set

$$p(z) := \left(\frac{f}{z - z_0}\right)^{1/2} = \frac{\alpha\sqrt{z+z_0}}{\sqrt{z^2+1}}. \tag{5-30}$$

Because $|\omega^2 + \alpha^2\cosh^2(r)| = |\alpha^2(z^2 - z_0^2)|$, the assumption $|\omega^2 + \alpha^2\cosh^2(r)| \leq c$ implies

$$z \asymp z_0 \asymp |\alpha|^{-1}, \tag{5-31}$$

with constants that depend only on c. This makes it easy to estimate

$$|\partial_z^k p(z)| \asymp |\alpha|^{3/2+k}, \tag{5-32}$$

with constants that depend only on c and k. If we define

$$q(z) := \frac{\phi}{(z-z_0)^{3/2}},$$

then by writing

$$q(z) = \int_0^1 t^{1/2} \frac{p(z_0 + t(z-z_0))}{\sqrt{((1-t)z_0 + tz)^2 + 1}}\, dt,$$

we can deduce from (5-32) that

$$|\partial_z^k q(z)| \asymp |\alpha|^{5/2+k}. \tag{5-33}$$

To apply these estimates, we note that $f/\zeta = p^2(\frac{3}{2}q)^{-2/3}$. We can use this identification to apply the bounds (5-32) and (5-33) to the formula (5-31) for ψ, obtaining

$$|\psi f^{1/2} \zeta^{-1/2}| \asymp |\alpha|^{-2/3} \quad \text{for } |\omega^2 + \alpha^2 \cosh^2(r)| \le c.$$

The bound

$$\int_{|\omega^2+\alpha^2\cosh^2(r)|\le c} |\psi f^{1/2} \zeta^{-1/2}| dr = O(|\alpha|^{-2/3}), \tag{5-34}$$

follows immediately, since the range of integration for r is $O(1)$.

Combining the bounds (5-28), (5-29), and (5-34) gives

$$\Phi(0) = O(|\alpha|^{-2/3}),$$

and the claimed estimate follows from (5-25). □

6. Funnel determinant estimates

For the model funnel F_ℓ, fix $r_0 \ge 0$ and for some $\eta > 0$ set

$$r_k = r_0 + k\eta.$$

Let $\mathbb{1}_k$ denote the multiplication operator for the characteristic function of the interval $r \in [r_k, r_{k+1}]$ in $L^2(F_\ell)$. The operator $G_j(s)$ defined in (4-7) can be represented in the model funnel case by

$$G(s) := (2s-1)\mathbb{1}_1 E_{F_\ell}(1-s) E_{F_\ell}(s)^t \mathbb{1}_2 \tag{6-1}$$

Our goal in this section is to prove the sharp bound on $\log\det(1 + c|G(s)|)$ used in the proof of Theorem 4.1.

To proceed we must analyze the Fourier decomposition of $E_{F_\ell}(s)$. Because of the circular symmetry, the Poisson kernel on F_ℓ admits a diagonal expansion into Fourier modes:

$$E_{F_\ell}(s; r, \theta, \theta') = \frac{1}{\ell} \sum_{k\in\mathbb{Z}} a_k(s; r) e^{ik(\theta-\theta')} \tag{6-2}$$

The coefficients $a_k(s;r)$ satisfy (5-3) with the boundary condition $a_k(s;0)=0$, so we must have

$$a_k(s;r) = c_k(s) w_k^-(s;r), \tag{6-3}$$

where w_k^- is the odd solution (5-5). To compute the normalization constant $c_k(s)$, we use the fact that

$$(2s-1)a_k(s;r) \sim \rho^{1-s} + [S_{F_\ell}(s)]_k \rho^s \quad \text{as } \rho \to 0, \tag{6-4}$$

where $[S_{F_\ell}(s)]_k$ is the k-th matrix element of the scattering matrix $S_\ell(s)$. Applying the appropriate Kummer identity [Olver 1974, Equation (5.10.16)] to the hypergeometric function in (5-5) gives

$$a_k(s;r) \sim c_k(s)\big(\Gamma(\tfrac{1}{2}-s)\beta_k(2-s)\rho^s + \Gamma(s-\tfrac{1}{2})\beta_k(1+s)\rho^{1-s}\big),$$

where

$$\beta_k(s) := \frac{1}{\Gamma(\frac{1}{2}(s+ik\omega)))\Gamma(\frac{1}{2}(s-ik\omega)))}. \tag{6-5}$$

By comparing this asymptotic to (6-4), we can read off the coefficient

$$c_k(s) = \frac{2s-1}{\Gamma(s-\frac{1}{2})\beta_k(1+s)},$$

as well as the scattering matrix element

$$[S_{F_\ell}(s)]_k = \frac{\Gamma(\frac{1}{2}-s)\beta_k(2-s)}{\Gamma(s-\frac{1}{2})\beta_k(1+s)}. \tag{6-6}$$

For future reference we note also that

$$a_k(1-s;r) = -\frac{a_k(s;r)}{[S_{F_\ell}(s)]_k}. \tag{6-7}$$

and

$$a_k(s;r) = a_{-k}(s;r) \tag{6-8}$$

We can express the singular values of $G(s)$ in terms of the coefficients $a_k(s;r)$. Up to reordering, these singular values are given by

$$\lambda_k(s) := |2s-1|\Big(\int_{r_1}^{r_2} |a_k(1-s;r)|^2 \cosh r\, dr\Big)^{1/2}\Big(\int_{r_2}^{r_3} |a_k(s;r)|^2 \cosh r\, dr\Big)^{1/2} \quad \text{for } k \in \mathbb{Z}. \tag{6-9}$$

To prove this, we note that $\lambda_k(s)^2$ is the eigenvalue of $G^*G(s)$ corresponding to the eigenfunction $\chi_{[r_2,r_3]}(r)\overline{a_k(s;r)}e^{-ik\theta}$. Also, it is easy to see from (6-1) and (6-2) that these are the only nonzero eigenvalues.

Using (6-7) to replace $a_k(1-s)$ by $a_k(s)$, and assuming $\eta \le 1$, we can estimate

$$\lambda_k(\tfrac{1}{2}+k\alpha) \le \big|2k\alpha a_k(\tfrac{1}{2}+k\alpha; r_3)^2 [S_{F_\ell}(\tfrac{1}{2}-k\alpha)]_k \cosh r_3\big|. \tag{6-10}$$

We will first estimate the various components. Recall that the matrix elements of $S_{F_\ell}(s)$ were expressed in terms of the function β_k defined in (6-5).

Lemma 6.1. *For* $k>0$ *and* $\arg\alpha\in[0,\frac{1}{2}\pi]$, *if we assume* $\operatorname{dist}(k\alpha,\mathbb{N}_0)\geq\delta$ *then we have*

$$\log\bigl|[S_{F_\ell}(\tfrac{1}{2}-k\alpha)]_k\bigr|\geq 2k\operatorname{Re}\gamma+2k[\operatorname{Re}\phi_0]_- - C(\delta),$$

where $\gamma(\alpha)$ *was defined in* (5-23). *If instead we assume that* $\operatorname{dist}(\frac{1}{2}-k\alpha,\mathcal{R}_{F_\ell})\leq|k\alpha|^{-\beta}$, *then*

$$\log\bigl|[S_{F_\ell}(\tfrac{1}{2}-k\alpha)]_k\bigr|\leq 2k\operatorname{Re}\gamma+2k[\operatorname{Re}\phi_0]_- + C(\beta)\log|k\alpha|.$$

Proof. Consider the matrix element (6-6). For $\operatorname{Re}\alpha\geq 0$, we can apply Stirling's formula directly to obtain

$$\log\Gamma(k\alpha)\beta_k(\tfrac{3}{2}+k\alpha)=k\gamma(\alpha)-\tfrac{1}{2}\log\pi k^2\alpha\sqrt{\omega^2+\alpha^2}+O(|k\alpha|^{-1}),$$

To estimate the other term, we must avoid zeros and poles. For $\operatorname{Re}z\leq 0$, applying Stirling via the reflection formula gives

$$\log|\Gamma(z)|\leq\operatorname{Re}\bigl((z-\tfrac{1}{2})\log(-z)-z\bigr)-\pi|\operatorname{Im}z|+\log(1+\operatorname{dist}(z,-\mathbb{N}_0)^{-1})+O(1),$$

and

$$\log|\Gamma(z)|\geq\operatorname{Re}\bigl((z-\tfrac{1}{2})\log(-z)-z\bigr)-\pi|\operatorname{Im}z|+O(1).$$

If we assume that $\operatorname{dist}(k\alpha,\mathbb{N}_0)\geq\delta$, then we obtain the upper bound

$$\log\bigl|\Gamma(-k\alpha)\beta_k(\tfrac{3}{2}-k\alpha)\bigr|\leq -k\operatorname{Re}\gamma(\alpha)-2k[\operatorname{Re}\phi_0]_- -\tfrac{1}{2}\log k^2\alpha\sqrt{\omega^2+\alpha^2}+C(\delta).$$

For a lower bound, we need to assume that $\operatorname{dist}(k\alpha,\mathcal{R}_{F_\ell})\geq|k\alpha|^{-\beta}$, and then we find that

$$\log\bigl|\Gamma(-k\alpha)\beta_k(\tfrac{3}{2}-k\alpha)\bigr|\geq -k\operatorname{Re}\gamma(\alpha)-2k[\operatorname{Re}\phi_0]_- -\tfrac{1}{2}\log k^2\alpha\sqrt{\omega^2+\alpha^2}-C(\beta)\log|k\alpha|. \qquad\square$$

Lemma 6.2. *Assuming that* $\operatorname{Re}\alpha\geq 0$, $k>0$, *and* $\operatorname{dist}(\frac{1}{2}-k\alpha,\mathcal{R}_{F_\ell})\leq|k\alpha|^{-\beta}$, *we have*

$$\log\lambda_k(\tfrac{1}{2}+k\alpha)\leq 2k\operatorname{Re}\phi(\alpha;r_3)-2k[\operatorname{Re}\phi_0(\alpha)]_+ + O(\log|k\alpha|).$$

Proof. By conjugation we can assume $\arg\alpha\in[0,\frac{1}{2}\pi]$. Then $a_k(\frac{1}{2}+k\alpha;r)$ can be expressed in terms of the solutions w_σ from Proposition 5.2. To satisfy the Dirichlet boundary condition, it must be a constant multiple of $w_0(0)w_1(r)-w_1(0)w_0(r)$. Lemma 5.3 gives the asymptotic behavior of this expression as $r\to\infty$, allowing us to deduce the constant. After comparing to (6-4), we find that

$$a_k(\tfrac{1}{2}+k\alpha;r)=\frac{1}{2kw_0(0)}\alpha^{-1/2}e^{-k(\phi_0(\alpha)+\gamma(\alpha))}\bigl(w_0(0)w_1(r)-w_1(0)w_0(r)\bigr) \tag{6-11}$$

The estimate

$$\bigl|a_k(\tfrac{1}{2}+k\alpha;r)\bigr|\leq Ck^{1/6}e^{k\operatorname{Re}(\phi(\alpha,r)-\phi_0(\alpha)-\gamma(\alpha))}, \tag{6-12}$$

for $|k\alpha|$ sufficiently large, then follows immediately from (5-19) and (5-20). The result now follows from applying Lemma 6.1 and (6-12) in (6-10). $\square$

Proposition 6.3. *Assuming that* $\eta\leq 1$, $0\leq\theta\leq\pi/2$, *and* $\operatorname{dist}(\frac{1}{2}-ae^{i\theta},\mathcal{R}_{F_\ell})\geq a^{-\beta}$ *for some fixed* $\beta>1$, *we have*

$$\log\det\bigl(I+c|G(\tfrac{1}{2}+ae^{i\theta})|\bigr)\leq\kappa(\theta,r_4)a^2+C(c,r_0,\beta)a\log a,$$

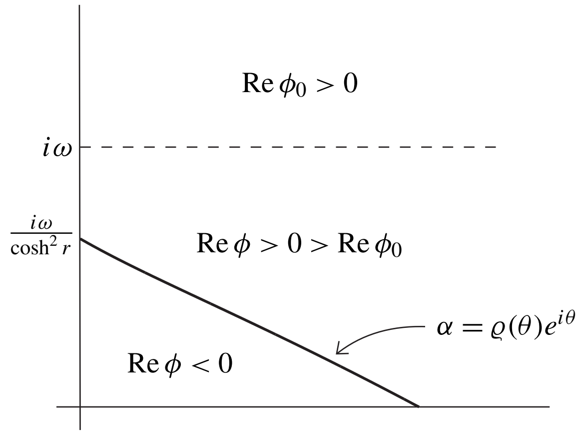

Figure 5. Positive and negative regions for $\operatorname{Re}\phi(\alpha; r)$ and $\operatorname{Re}\phi_0(\alpha)$, shown for $r = 1$.

where

$$\kappa(\theta, r) = 2\int_0^\infty \frac{[I(xe^{i\theta}, \ell, r)]_+}{x^3}\,dx - \tfrac{1}{2}\ell\sin^2\theta, \tag{6-13}$$

with $I(xe^{i\theta}, \ell, r) := 2\operatorname{Re}\phi(xe^{i\theta}; r)$, *which agrees with the definition* (1-11).

Proof. We start from the expression for the determinant in terms of the singular values,

$$\det(I + c|G(\tfrac{1}{2} + ae^{i\theta})|) = \prod_{k\in\mathbb{Z}}(1 + c\lambda_k(\tfrac{1}{2} + ae^{i\theta})).$$

By the conjugation symmetry, we can assume $\theta \in [0, \frac{1}{2}\pi]$. Let $\varrho(\theta)$ be the implicit solution of the equation $\operatorname{Re}\phi(\varrho(\theta)e^{i\theta}, r_3) = 0$, as illustrated in Figure 5.

Note that $\operatorname{Re}\phi_0(xe^{i\theta}) = 0$ in a neighborhood of $x = \varrho(\theta)$. For some $\delta > 0$, we subdivide the sum in

$$\log\det\bigl(I + c|G(\tfrac{1}{2} + ae^{i\theta})|\bigr) = 2\sum_{k=1}^{\infty}\log(1 + c\lambda_k(\tfrac{1}{2} + a_i e^{i\theta})) + O(a\log a)$$

at values where $a_i/k = \varrho(\theta)$ and $(1-\delta)\varrho(\theta)$. The dominant part of the sum is

$$\Sigma_+ := \sum_{1\le k\le a/\varrho(\theta)} \log(1 + c\lambda_k(\tfrac{1}{2} + ae^{i\theta})).$$

Assuming that $a \in \{a_i\}$, Lemma 6.2 gives the bound

$$\Sigma_+ \le \sum_{1\le k\le a/\varrho(\theta)} 2k\bigl(\operatorname{Re}\phi(ae^{i\theta}/k; r_3) - [\operatorname{Re}\phi_0(ae^{i\theta}/k)]_+\bigr) + C(c, r_0, \beta)a\log a.$$

Because the summand is a decreasing function of k, we may estimate the sum by the integral

$$\Sigma_+ \le \int_0^{a/\varrho(\theta)} 2k\bigl(\operatorname{Re}\phi(ae^{i\theta}/k; r_3) - [\operatorname{Re}\phi_0(ae^{i\theta}/k)]_+\bigr) + C(c, r_0, \beta)a\log a$$

Substituting $x = a/k$ gives

$$\int_0^{a/\varrho(\theta)} 2k \operatorname{Re} \phi(ae^{i\theta}/k; r_3)\, dk = 2a^2 \int_{\varrho(\theta)}^{\infty} \frac{\operatorname{Re} \phi(xe^{i\theta}; r_3)}{x^3}\, dx.$$

We can also compute that

$$\int_0^{a/\varrho(\theta)} 2k[\operatorname{Re} \phi_0(ae^{i\theta}/k)]_+\, dk = \pi a^2 \int_{\omega/\sin(\theta)}^{\infty} \frac{x \sin\theta - \omega}{x^3}\, dx = \frac{\pi a^2}{2\omega} \sin\theta.$$

Comparing to (6-13), we conclude that

$$\Sigma_+ \le \kappa(\theta, r_3) a^2 + C(c, r_0, \beta) a \log a.$$

The middle term is given by

$$\Sigma_0 := \sum_{a/\varrho(\theta) \le k \le a/(1-\delta)\varrho(\theta)} \log(1 + c\lambda_k(\tfrac{1}{2} + ae^{i\theta})),$$

Since $I(\alpha, \ell, r_3) = O(\delta)$ for k in this range, the same integral estimate used for Σ_+ gives

$$|\Sigma_0| \le C(c, r_0, \beta)\delta a^2 + C(c, r_0, \beta) a \log a.$$

Finally, we set

$$\Sigma_- := \sum_{k \ge a/(1-\delta)\varrho(\theta)} \log(1 + c\lambda_k(\tfrac{1}{2} + ae^{i\theta})).$$

For k in this range, $I(\alpha, \ell, r_3) \le -C\delta$ and we can estimate

$$|\Sigma_-| \le C(c, r_0, \beta, \delta) e^{-ca} \quad \text{for some } c > 0.$$

Adding together the estimates for Σ_+, Σ_0, and Σ_- gives

$$\log\det\bigl(I + C|G(\tfrac{1}{2} + ae^{i\theta})|\bigr) \le \kappa(\theta, r_3)a^2 + C(c, r_0, \beta)(\delta a^2 + a\log a) + C(c, r_0, \beta, \delta)e^{-ca}$$

We can absorb the δa^2 term into the first term by replacing r_3 by r_4, assuming that $\eta = O(\delta)$, since $\kappa(\theta, \cdot)$ is strictly increasing. This yields the claimed estimate. ☐

7. Resonance asymptotics for truncated funnels

Inside the model funnel F_ℓ, with metric given by (5-1), we let F_{ℓ,r_0} denote the truncated region $\{r \ge r_0\}$, with the Laplacian defined by imposing Dirichlet boundary conditions at $r = r_0$. To compute the associated scattering matrix elements exactly, we consider the solutions of the Fourier mode equation (5-3) given by (5-4) and (5-5). To impose the boundary condition at $r = r_0$, we set

$$u_k(s; r) := w_k^+(s; r_0) w_k^-(s; r) - w_k^-(s; r_0) w_k^+(s; r). \tag{7-1}$$

The scattering matrix element may be obtained from the asymptotics of $u_k(s;r)$ as $r\to\infty$ be noting that for any generalized eigenmode we have

$$u_k(s;r) \sim c_{k,s}(\rho^{1-s} + [S_{F_{\ell,r_0}}(s)]_k \rho^s) \tag{7-2}$$

as $r\to\infty$, where $\rho := 2e^{-r}$ as before. The solutions $w_k^\pm$ have leading asymptotics,

$$\begin{aligned} w_k^+(s;r) &\sim \Gamma(s-\tfrac12)\beta_k(s)\rho^{1-s} + \Gamma(\tfrac12 - s)\beta_k(1-s)\rho^s,\\ w_k^-(s;r) &\sim \Gamma(s-\tfrac12)\beta_k(1+s)\rho^{1-s} + \Gamma(\tfrac12 - s)\beta_k(2-s)\rho^s \end{aligned} \tag{7-3}$$

as $r\to\infty$, where $\beta_k(s)$ was defined in (6-5).

If we set

$$f_k(s;r) := \Gamma(s-\tfrac12)\big(\beta_k(1+s)w_k^+(s;r) - \beta_k(s)w_k^-(s;r)\big), \tag{7-4}$$

Then from (7-2) we can read off that

$$[S_{F_{\ell,r_0}}(s)]_k = \frac{f_k(1-s;r_0)}{f_k(s;r_0)}. \tag{7-5}$$

The k-th Fourier mode thus contributes scattering poles at the values of s for which

$$\beta_k(1+s)w_k^+(s;r_0) - \beta_k(s)w_k^-(s;r_0) = 0.$$

This function can be written in terms of a single normalized hypergeometric function, via the standard identities, yielding

$$\mathcal{R}_{F_{\ell,r_0}} = \bigcup_{k\in\mathbb{Z}} \big\{ s : \mathbf{F}(\tfrac12(1+s+i\omega k), \tfrac12(s+i\omega k); \tfrac12 + s; -\sinh^{-2} r_0) = 0 \big\}.$$

A sample resonance counting function is shown in Figure 6.

Theorem 7.1. *For the truncated funnel with Dirichlet boundary conditions*,

$$N_{F_{\ell,r_0}}(t) \sim A(F_{\ell.r_0})t^2,$$

where $A(F_{\ell.r_0})$ is given by (1-10).

In conjunction with [Borthwick 2010, Theorem 1.2] for the hyperbolic planar case, this will complete the proof of Theorem 1.3. Before giving the proof, we need some estimates of scattering matrix elements.

Lemma 7.2. *Assuming that* $\arg\alpha \in [0, \pi/2 - \varepsilon]$ *with* $\operatorname{dist}(k\alpha, \mathbb{N}_0) \ge \eta$, *we can have*

$$\log\left| \frac{[S_{F_{\ell,r_0}}(\frac12 + k\alpha)]_k}{[S_{F_\ell}(\frac12 + k\alpha)]_k} - 1 \right| \ge 2k(\operatorname{Re}\phi(\alpha;r_0) - [\operatorname{Re}\phi_0(\alpha)]_+) - C(\eta)$$

for $|k\alpha|$ sufficiently large.

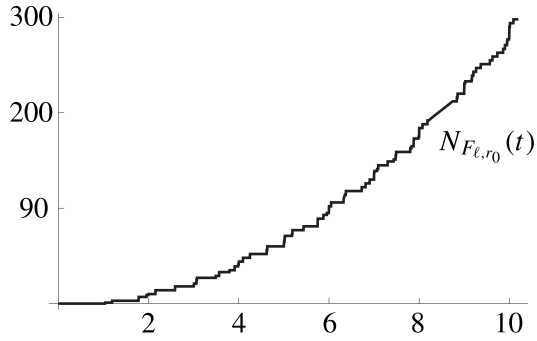

Figure 6. The resonance counting function for F_{ℓ,r_0}, shown for $\ell = 2\pi$ and $r_0 = 1$.

Proof. To estimate $[S_{F_{\ell,r_0}}(s)]_k$, as given in (7-5), we must connect f_k to the solutions w_σ introduced in (5-17). Since $f_k(\frac{1}{2}+k\alpha; r)$ is recessive as $r \to \infty$, this solution must be proportional to w_0. From (7-3), we can use the reflection formula for the gamma function to see that

$$f_k(\tfrac{1}{2}+k\alpha; r) \sim \frac{\rho^s}{\pi k\alpha} \quad \text{as } r \to \infty.$$

By comparing this to the asymptotic from Lemma 5.3, we find that

$$f_k(\tfrac{1}{2}+k\alpha; r) = A_0^+ w_0(r). \tag{7-6}$$

where

$$A_0^+ := \frac{1}{\pi k\sqrt{\alpha}} e^{k(\phi_0+\gamma)}.$$

We may also express $f_k(\frac{1}{2}-k\alpha; r)$ in terms of the w_σ,

$$f_k(\tfrac{1}{2}-k\alpha; r) = A_0^- w_0(r) + A_1^- w_1(r), \tag{7-7}$$

for some coefficients A_0^- and A_1^- that are independent of r but do depend on k and α. By (7-3),

$$f_k(\tfrac{1}{2}-k\alpha; r) \sim -\frac{\rho^{1-s}}{\pi k\alpha},$$

and so by Lemma 5.3 we have

$$A_1^- = -\pi^{-1}k^{-1}\alpha^{-1/2}e^{-k(\phi_0+\gamma)}. \tag{7-8}$$

The other coefficient can then be computed by comparing values at $r = 0$,

$$A_0^- = \frac{1}{w_0(0)}\big(f_k(\tfrac{1}{2}-k\alpha; 0) - A_1^- w_1(0)\big). \tag{7-9}$$

Using (7-6) to relate $w_0(0)$ to $f_k(\frac{1}{2}+k\alpha; 0)$, we can then deduce that

$$[S_{F_{\ell,r}}(\tfrac{1}{2}+k\alpha)]_k = [S_{F_\ell}(\tfrac{1}{2}+k\alpha)]_k - e^{-2k(\phi_0+\gamma)}\left(\frac{w_1(r)}{w_0(r)} - \frac{w_1(0)}{w_0(0)}\right). \tag{7-10}$$

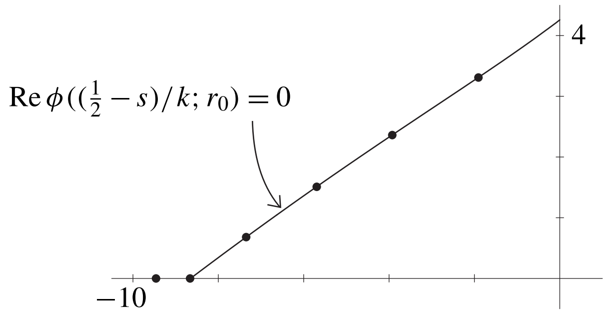

Figure 7. Using the equation $\operatorname{Re}\phi=0$ to locate the resonances of F_{ℓ,r_0} occurring in the $k=7$ Fourier mode, shown for $\ell=2\pi$ and $r_0=1$.

Hence

$$\frac{[S_{F_{\ell,r}}(\frac{1}{2}+k\alpha)]_k}{[S_{F_\ell}(\frac{1}{2}+k\alpha)]_k}-1=-e^{-2k(\phi_0+\gamma)}\left(\frac{w_1(r)}{w_0(r)}-\frac{w_1(0)}{w_0(0)}\right)[S_{F_\ell}(\tfrac{1}{2}-k\alpha)]_k \tag{7-11}$$

For $\arg\alpha\in[0,\pi/2-\varepsilon]$, we deduce from (5-18) (using also the fact that $\operatorname{Re}(\phi-\phi_0)>c(\varepsilon,r)$) that

$$\left(\frac{w_1(r)}{w_0(r)}-\frac{w_1(0)}{w_0(0)}\right)=e^{2k\phi}(1+O(|k\alpha|^{-1})). \tag{7-12}$$

The result then follows from (7-11) and the lower bound on $[S_{F_\ell}(\frac{1}{2}-k\alpha)]_k$ provided by Lemma 6.1. $\square$

The estimates in Lemma 7.2 give approximate locations for the resonances in $\mathcal{R}_{F_{\ell,r_0}}$ arising from the k-th Fourier mode. The zeros of (7-10) correspond to resonances at $s=\frac{1}{2}-k\alpha$. This requires a cancellation between the two terms on the right side of (7-10). If $\operatorname{Re}\phi>0$, then the second term is larger by approximately $e^{2k\phi}$ and cancellation only occurs near the poles of $[S_{F_\ell}(s)]_k$; this explains the poles of $[S_{F_{\ell,r_0}}(s)]_k$ on the negative real axis. For $\operatorname{Re}\phi=0$, the two terms in (7-10) have the same magnitude; the resonances off the real axis in $\mathcal{R}_{F_{\ell,r_0}}$ thus occur near the line $\operatorname{Re}\phi((\frac{1}{2}-s)/k;r_0)=0$ (and its conjugate). Figure 7 illustrates this phenomenon. For $\operatorname{Re}\phi<0$, the first term in (7-10) is always larger than the second and no zeros occur.

Since $[S_{F_{\ell,r}}(\frac{1}{2}+k\alpha)]_k$ may indeed have zeros near the line $\operatorname{Re}\phi=0$, proving a lower bound is more delicate in this region. By focusing on a relatively narrow strip, we can settle for a cruder estimate on the matrix elements in the vicinity of the zeros.

Lemma 7.3. *For* $k\geq 0$ *and* $\operatorname{Re}s\geq\frac{1}{2}$ *and assuming* $\operatorname{dist}(1-s,\mathcal{R}_{F_\ell})\geq|s|^{-\beta}$ *with* $\beta>2$,

$$\log\left|\frac{[S_{F_{\ell,r_0}}(s)]_k}{[S_{F_\ell}(s)]_k}\right|\leq C(r_0,\beta)(k+|s|)\log|s|.$$

If $\operatorname{dist}(1-s,\mathcal{R}_{F_{\ell,r_0}})\geq|s|^{-\beta}$ *with* $\beta>2$, *then we have*

$$\log\left|\frac{[S_{F_{\ell,r_0}}(s)]_k}{[S_{F_\ell}(s)]_k}\right|\geq -c(r_0,\beta)(k+|s|)\log|s|.$$

Proof. From (7-4), we note that $f_k(s;r_0)/\Gamma(s-\frac{1}{2})$ is an entire function of s. By Stirling's formula and the estimate (5-19), we can estimate its growth for large $|s|$ and $k \neq 0$ by

$$\log\left|\frac{f_k(s;r_0)}{\Gamma(s-\frac{1}{2})}\right| \leq C(r_0)(k+|s|)\log|s|),$$

where C is independent of k. The same estimate holds for $k=0$, by the classical asymptotics of the hypergeometric function due to Watson [Erdélyi et al. 1953, Section 2.3.2]. Assuming that $\operatorname{dist}(s, \mathcal{R}_{F_{\ell,r_0}}) \geq |s|^{-\beta}$, where $\beta > 2$, the minimum modulus theorem gives

$$\log\left|\frac{f_k(s;r_0)}{\Gamma(s-\frac{1}{2})}\right| \geq -c(r_0,\beta)(k+|s|)\log|s| \quad \text{for large } |s|.$$

The results follow from applying these estimates to

$$\frac{[S_{F_{\ell,r_0}}(s)]_k}{[S_{F_\ell}(s)]_k} = \frac{f_k(1-s;r_0)}{f_k(s;r_0)}\,\frac{f_k(s;0)}{f_k(1-s;0)}. \qquad \square$$

Proof of Theorem 7.1. We note that

$$N_{F_\ell}(t) \sim \tfrac{1}{4}\ell t^2 \quad \text{and} \quad 0\text{-vol}(F_{\ell,r_0}) = -\ell \sinh r_0.$$

By Corollary 3.2 and Theorem 4.1, the claimed asymptotic will be proved if we can show that there exists an unbounded set $\Lambda \subset [1,\infty)$ such that

$$\frac{2}{\pi}\int_0^{\pi/2} \log|\tau(\tfrac{1}{2}+ae^{i\theta})|\,d\theta \geq \frac{4a^2}{\pi}\int_0^{\pi/2}\int_0^\infty \frac{[I(xe^{i\theta},\ell,r_0)]_+}{x^3}\,dx - \tfrac{1}{4}\ell a^2 - o(a^2) \tag{7-13}$$

for all $a \in \Lambda$. We take

$$\Lambda := \bigl\{a \geq 1 : \operatorname{dist}\bigl(\{|s-\tfrac{1}{2}| = a\},\ \mathcal{R}_{F_\ell} \cup \mathcal{R}_{F_{\ell,r_0}} \cup \mathbb{N}_0\bigr) \geq a^{-3}\bigr\}. \tag{7-14}$$

Using the symmetry of coefficients under $k \to -k$, and estimating the $k=0$ term by Lemma 7.3, we have

$$\log\bigl|\tau(\tfrac{1}{2}+ae^{i\theta})\bigr| = 2\sum_{k=1}^\infty \log\left|\frac{[S_{F_{\ell,r_0}}(\frac{1}{2}+ae^{i\theta})]_k}{[S_{F_\ell}(\frac{1}{2}+ae^{i\theta})]_k}\right| + O(a\log a). \tag{7-15}$$

Define $\varrho(\theta)$ by $\operatorname{Re}\phi(\varrho(\theta)e^{i\theta}, r_0) = 0$, as in the proof of Proposition 6.3, and assume for now that $\theta \leq \frac{1}{2}\pi - \varepsilon$. For $\delta > 0$, we will split the sum (7-15) at $a/k = \varrho(\theta)(1 \pm a^{-1/2})$. Let Σ_+ denote the portion of the sum with $a/k \geq \varrho(\theta)(1+a^{-1/2})$. Under this condition, we want to derive a lower bound from Lemma 7.2 using the inequality

$$\log|1+\lambda| \geq \log|\lambda| - \log 2 \quad \text{for } |\lambda| \geq 2.$$

For a sufficiently large, we will have $\operatorname{Re}\phi(xe^{i\theta}, r_0) \geq ca^{-1/2}$ for $x \geq \varrho(\theta)(1+a^{-1/2})$. Thus, for $k \geq c\sqrt{a}$ we can deduce from Lemma 7.2 that

$$\log\left|\frac{[S_{F_{\ell,r_0}}(\frac{1}{2}+ae^{i\theta})]_k}{[S_{F_\ell}(\frac{1}{2}+ae^{i\theta})]_k}\right| \geq 2k\bigl(\operatorname{Re}\phi(ae^{i\theta}/k; r_3) - [\operatorname{Re}\phi_0(ae^{i\theta}/k)]_+\bigr) + O(1).$$

Arguing as in the proof of Proposition 6.3, we can then obtain

$$\sum_{c\sqrt{a}\le k\le a/(\varrho(\theta)(1+a^{-1/2}))} \log\left|\frac{[S_{F_{\ell,r_0}}(\frac{1}{2}+ae^{i\theta})]_k}{[S_{F_\ell}(\frac{1}{2}+ae^{i\theta})]_k}\right| \ge 2a^2\int_{\varrho(\theta)(1+a^{-1/2})}^{C\sqrt{a}} \frac{\operatorname{Re}\phi(xe^{i\theta},r_0)-[\operatorname{Re}\phi_0(xe^{i\theta})]_+}{x^3}\,dx - O(a\log a).$$

For $k\le c\sqrt{a}$, Lemma 7.3 gives the estimate

$$\sum_{1\le k\le c\sqrt{a}} \log\left|\frac{[S_{F_{\ell,r_0}}(\frac{1}{2}+ae^{i\theta})]_k}{[S_{F_\ell}(\frac{1}{2}+ae^{i\theta})]_k}\right| \ge -O(a^{3/2}\log a).$$

On the other hand, since $|\operatorname{Re}\phi(\alpha,r)| = O(|\alpha|)$ for large $|\alpha|$, we also have

$$2a^2\int_{C\sqrt{a}}^{\infty} \frac{\operatorname{Re}\phi(xe^{i\theta},r_0)-[\operatorname{Re}\phi_0(xe^{i\theta})]_+}{x^3}\,dx = O(a^{3/2}).$$

We can also estimate

$$2a^2\int_{\varrho(\theta)}^{\varrho(\theta)(1+a^{-1/2})} \frac{\operatorname{Re}\phi(xe^{i\theta},r_0)-[\operatorname{Re}\phi_0(xe^{i\theta})]_+}{x^3}\,dx = O(a^{3/2})$$

since $\operatorname{Re}\phi(\alpha,\ell,r_0)$ is $O(\delta)$ in the range of integration. In combination, these estimates give

$$\Sigma_+ \ge 2a^2\int_{\varrho(\theta)}^{\infty} \frac{\operatorname{Re}\phi(xe^{i\theta},r_0)}{x^3}\,dx - \frac{\pi a^2}{2\omega}\sin^2\theta - O(a^{3/2}\log a) \quad \text{for } a\in\Lambda. \tag{7-16}$$

Let Σ_0 denote the portion of the sum in (7-15) for which $\varrho(\theta)(1-a^{-1/2}) < a/k < \varrho(\theta)(1+a^{-1/2})$. Since there are $O(a^{1/2})$ values of k in this range, Lemma 7.3 gives the estimate

$$\Sigma_0 \ge -O(a^{3/2}\log a). \tag{7-17}$$

Finally, we have Σ_-, defined as the portion of (7-15) with $a/k\le\varrho(\theta)(1-a^{-1/2})$. Now we wish to apply Lemma 7.2 using

$$\log|1+\lambda| \ge -|\lambda|\log 4 \quad \text{for } |\lambda|\le\tfrac{1}{2}.$$

Note that $I(xe^{i\theta},\ell,r_0)\le -ca^{-1/2}$ for $x\le\varrho(\theta)(1-a^{-1/2})$ and a sufficiently large, and that $k\ge ca$ in the range of Σ_-. Thus for large a Lemma 7.2 yields

$$\log\left|\frac{[S_{F_{\ell,r_0}}(\frac{1}{2}+ae^{i\theta})]_k}{[S_{F_\ell}(\frac{1}{2}+ae^{i\theta})]_k}\right| \ge -O(e^{-cka^{-1/2}}),$$

within the scope of Σ_-. We conclude that

$$\Sigma_- \ge -O(e^{-ca^{1/2}}). \tag{7-18}$$

Applying the estimates (7-16), (7-17), and (7-18) to the sum (7-15) now proves the lower bound

$$\frac{2}{\pi}\int_0^{\pi/2-\varepsilon}\log|\tau(\tfrac{1}{2}+ae^{i\theta})|\,d\theta \geq \frac{4a^2}{\pi}\int_0^{\pi/2-\varepsilon}\int_0^{\infty}\frac{[2\operatorname{Re}\phi(xe^{i\theta},r_0)]_+}{x^3}\,dx - \frac{2a^2}{\omega}\int_0^{\pi/2-\varepsilon}\sin^2\theta\,d\theta - o(a^2),$$

For the missing sectors, we appeal to Lemma 4.4 to see that

$$\frac{2}{\pi}\int_{\pi/2-\varepsilon}^{\pi/2}\log|\tau(\tfrac{1}{2}+ae^{i\theta})|\,d\theta \geq -c\varepsilon a^2.$$

We can thus take $\varepsilon \to 0$ to complete the proof of (7-13). □

Remark. In the proof of (1-13) given in [Borthwick 2010, Theorem 1.2], the Σ_- term was estimated incorrectly. This term is not necessarily positive, so the upper bound $O(e^{-ca})$ does not imply a corresponding lower bound. Instead, one needs to argue as in the derivation of (7-18) above. The estimates needed for the correct argument were given in [Borthwick 2010, Equations (6.8)–(6.10)].

8. Resonance asymptotics for extended funnels

Using the same notation as in Section 7, we now consider $F_{\ell,-r_0}$, defined as the subset $r \geq -r_0$ in a hyperbolic cylinder of diameter ℓ, where $r_0 \geq 0$. The metric and Laplacian are still given by (5-1) and (5-2), so that the scattering matrix elements are easily computed in terms of hypergeometric functions as before.

With reference to the even/odd solutions $w_k^{\pm}$ defined in (5-4) and (5-5), a solution $u_k(s;r)$ to the k-th eigenmode equation (5-3) satisfying $u_k(s;-r_0)=0$ can be written

$$u_k(s;r) = w_k^+(s;r_0)w_k^-(s;r) + w_k^-(s;r_0)w^+(s;r),$$

where $w_k^{\pm}(s;r)$ are the even/odd hypergeometric solutions defined in (5-4) and (5-5). Using the asymptotic expansions (7-3) as $r\to\infty$, we can read off the scattering matrix elements

$$[S_{F_{\ell,-r_0}}(s)]_k = \frac{\Gamma(\frac{1}{2}-s)}{\Gamma(s-\frac{1}{2})}\frac{\beta_k(2-s)w_k^+(s;r_0)+\beta_k(1-s)w_k^-(s;r_0)}{\beta_k(1+s)w_k^+(s;r_0)+\beta_k(s)w_k^-(s;r_0)}, \tag{8-1}$$

where $\beta_k(s)$ was defined in (6-5).

This shows in particular that

$$\mathcal{R}_{F_{\ell,-r_0}} = \bigcup_{k\in\mathbb{Z}}\bigl\{s : \beta_k(1+s)w_k^+(s;r_0)+\beta_k(s)w_k^-(s;r_0)=0\bigr\}.$$

Theorem 8.1. *For the extended funnel with Dirichlet boundary conditions imposed at $r=-r_0$, for $r_0\geq 0$, we have*

$$N_{F_{\ell,-r_0}}(t) \sim A(F_{\ell,-r_0})t^2,$$

where

$$A(F_{\ell,-r_0}) = \frac{\ell}{2\pi}\sinh r_0 + \frac{4}{\pi}\int_0^{\pi/2}\int_0^{\infty}\frac{[I(xe^{i\theta},\ell,-r_0)]_+}{x^3}\,dx\,d\theta, \tag{8-2}$$

and $I(\alpha,\ell,r)$ was defined in (1-11).

Proof. Since $N_{F_\ell}(t) \sim \frac{1}{4}\ell t^2$ and 0-vol$(F_{\ell,-r_0}) = \ell \sinh r_0$, Theorem 8.1 will follow from Corollary 3.2 and Theorem 3.3, once we establish

$$\frac{2}{\pi}\int_0^{\pi/2} \log|\tau(\tfrac{1}{2}+ae^{i\theta})|\,d\theta = \frac{4a^2}{\pi}\int_0^{\pi/2}\int_0^\infty \frac{[I(xe^{i\theta},\ell,-r_0)]_+}{x^3}\,dx\,d\theta - \tfrac{1}{4}\ell a^2 - o(a^2), \tag{8-3}$$

where Λ is defined again by (7-14).

As in the proof of Theorem 7.1, we start with the Fourier decomposition of the scattering matrices and use Lemma 7.3 to estimate the $k=0$ term, leaving

$$\log|\tau(\tfrac{1}{2}+ae^{i\theta})| = 2\sum_{k=1}^\infty \log\left|\frac{[S_{F_{\ell,-r_0}}(\frac{1}{2}+ae^{i\theta})]_k}{[S_{F_\ell}(\frac{1}{2}+ae^{i\theta})]_k}\right| + O(a\log a). \tag{8-4}$$

If we define

$$g_k(s;r) := \Gamma(s-\tfrac{1}{2})\big(\beta_k(1+s)w_k^+(s;r) + \beta_k(s)w_k^-(s;r)\big),$$

then by (8-1),

$$[S_{F_{\ell,-r_0}}(\tfrac{1}{2}+ae^{i\theta})]_k = g_k(\tfrac{1}{2}-ae^{i\theta})/g_k(\tfrac{1}{2}+ae^{i\theta}).$$

Assuming $k>0$, we set $k\alpha = ae^{i\theta}$. Since $g_k(s;\ \cdot)$ solves (5-3), for $\operatorname{Re}\alpha \ge 0$, we can write

$$g_k(\tfrac{1}{2}\pm k\alpha; r) = B_0^\pm w_0(r) + B_1^\pm w_1(r),$$

where w_σ are the solutions given in (5-17).

As $r\to\infty$, the coefficient of ρ^{1-s} in the expansion of $g_k(\frac{1}{2}+k\alpha;r)$ is

$$2\Gamma(k\alpha)^2\beta_k(\tfrac{1}{2}+k\alpha)\beta_k(\tfrac{3}{2}+k\alpha) = \frac{1}{\pi k\alpha}\left(1-\frac{\cosh\pi k\omega}{\sin\pi k\alpha}\right)[S_{F_\ell}(\tfrac{1}{2}-k\alpha)]_k. \tag{8-5}$$

The coefficient of ρ^{1-s} in $g_k(\frac{1}{2}-k\alpha;r)$ is

$$\Gamma(k\alpha)\Gamma(-k\alpha)\big(\beta_k(\tfrac{1}{2}+k\alpha)\beta_k(\tfrac{3}{2}-k\alpha) + \beta_k(\tfrac{1}{2}-k\alpha)\beta_k(\tfrac{3}{2}+k\alpha)\big) = -\frac{1}{\pi k\alpha}\frac{\cosh\pi k\omega}{\sin\pi k\alpha}. \tag{8-6}$$

Comparing these to the asymptotics for w_σ, as given in Lemma 5.3, we see that

$$B_1^+ = \frac{e^{-k(\phi_0+\gamma)}}{\pi k\sqrt{\alpha}}\left(1-\frac{\cosh\pi k\omega}{\sin\pi k\alpha}\right)[S_{F_\ell}(\tfrac{1}{2}-k\alpha)]_k, \tag{8-7}$$

and

$$B_1^- = -\frac{e^{-k(\phi_0+\gamma)}}{\pi k\sqrt{\alpha}}\frac{\cosh\pi k\omega}{\sin\pi k\alpha} \tag{8-8}$$

We then find the B_0 coefficients by evaluating at $r=0$,

$$B_0^\pm = \frac{1}{w_0(0)}\big(g_k(\tfrac{1}{2}\pm k\alpha;0) - B_1^\pm w_1(0)\big). \tag{8-9}$$

Since f_k and g_k agree at $r=0$, (7-6) shows that

$$w_0(0) = A_0^+ g_k(\tfrac{1}{2}+k\alpha;0), \quad \text{where } A_0^+ := \frac{1}{\pi k\sqrt{\alpha}}e^{k(\phi_0+\gamma)}.$$

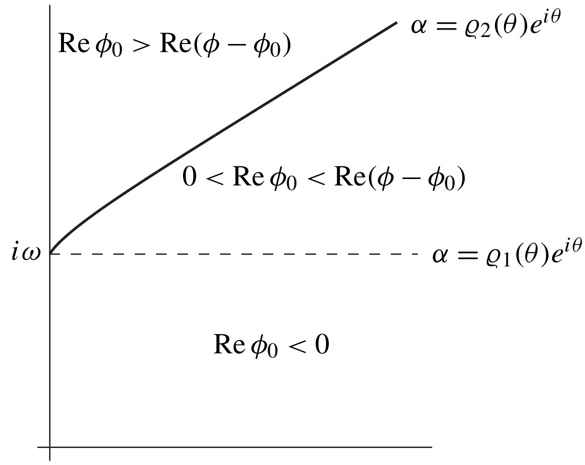

Figure 8. Positive and negative regions for $\operatorname{Re}(\phi(\alpha; r) - \phi_0(\alpha))$, shown for $r = 1$.

Combining these formulas gives

$$g_k(\tfrac{1}{2} + k\alpha; r) = A_0^+ w_0(r) + B_1^+\Big(w_1(r) - \frac{w_1(0)}{w_0(0)} w_0(r)\Big), \tag{8-10}$$

and

$$g_k(\tfrac{1}{2} - k\alpha; r) = [S_{F_\ell}(\tfrac{1}{2} + k\alpha)]_k A_0^+ w_0(r) + B_1^-\Big(w_1(r) - \frac{w_1(0)}{w_0(0)} w_0(r)\Big). \tag{8-11}$$

The asymptotic analysis of (8-10) is straightforward. The $B_1^+ w_1(r)$ term always dominates for $|k\alpha|$ large and $\arg\alpha \in [0, \pi/2 - \varepsilon]$, by Proposition 5.2. By applying Stirling's formula to (8-5) we find that

$$g_k(\tfrac{1}{2} + k\alpha; r) = \frac{1}{\pi k\sqrt{\alpha}} (\omega^2 + \alpha^2 \cosh^2 r)^{-1/4} e^{k(\phi - \phi_0 + \gamma)} (1 + O(|k\alpha|^{-1}). \tag{8-12}$$

The analysis of (8-11) more complicated. This term has both zeros and poles, and different terms can dominate for α in different regions. For $\alpha = xe^{i\theta}$, the borders between these regions will be denoted $x = \varrho_j(\theta)$ for $j = 1, 2$, where

$$\operatorname{Re}\phi_0(\varrho_1(\theta)e^{i\theta}) = 0 \quad \text{and} \quad \operatorname{Re}\big(\phi(\varrho_2(\theta)e^{i\theta}; r) - 2\phi_0(\varrho_2(\theta)e^{i\theta}; r)\big) = 0.$$

For the first curve we can be explicit, with $\varrho_1(\theta) = \omega \csc\theta$.

Consider first the portion of the sum (8-4) with $a/k \geq \varrho_2(\theta)$. In this region, $\operatorname{Re}\phi_0 > \operatorname{Re}(\phi - \phi_0)$ and the first term in (8-11) dominates the asymptotics. In this case, provided $|k\alpha| \in \Lambda$,

$$\log|g_k(\tfrac{1}{2} - k\alpha; r)| = k \operatorname{Re}(-\phi + \phi_0 - \gamma) + O(\log|k\alpha|).$$

For $k \leq a/\varrho_2(\theta)$, we thus have

$$\log\left|\frac{[S_{F_{\ell,-r_0}}(\frac{1}{2} + ae^{i\theta})]_k}{[S_{F_\ell}(\frac{1}{2} + ae^{i\theta})]_k}\right| = -2k \operatorname{Re}\Big(\phi\Big(\frac{ae^{i\theta}}{k}; r_0\Big) - \phi_0\Big(\frac{ae^{i\theta}}{k}\Big)\Big) + O(\log a).$$

This gives the estimate

$$\sum_{1\le k\le a/\varrho_2(\theta)} \log\left|\frac{[S_{F_{\ell,-r_0}}(\frac{1}{2}+ae^{i\theta})]_k}{[S_{F_\ell}(\frac{1}{2}+ae^{i\theta})]_k}\right| = a^2\int_{\varrho_2(\theta)}^{\infty} \frac{2\operatorname{Re}[\phi_0(xe^{i\theta})-\phi(xe^{i\theta};r_0)]}{x^3}\,dx + O(a\log a). \quad (8\text{-}13)$$

The region $\varrho_1(\theta) < a/k < \varrho_2(\theta)$ corresponds to $0 < \operatorname{Re}\phi_0 < \operatorname{Re}(\phi-\phi_0)$. In this case, the $B_1^- w_1(r)$ term dominates the asymptotics of (8-11), and we have

$$\log|g_k(\tfrac{1}{2}-k\alpha;r)| = k\operatorname{Re}(\phi-3\phi_0-\gamma) + O(\log|k\alpha|).$$

Using this along with (8-12) gives

$$\log\left|\frac{[S_{F_{\ell,-r_0}}(\frac{1}{2}+ae^{i\theta})]_k}{[S_{F_\ell}(\frac{1}{2}+ae^{i\theta})]_k}\right| = -2k\operatorname{Re}\phi_0(ae^{i\theta}/k) + O(\log a) \quad \text{for } k\le a/\varrho_2(\theta).$$

We conclude that

$$\sum_{a/\varrho_2(\theta)\le k\le a/\varrho_1(\theta)} \log\left|\frac{[S_{F_{\ell,-r_0}}(\frac{1}{2}+ae^{i\theta})]_k}{[S_{F_\ell}(\frac{1}{2}+ae^{i\theta})]_k}\right| = -a^2\int_{\varrho_2(\theta)}^{\infty} \frac{2\operatorname{Re}\phi_0(xe^{i\theta})}{x^3}\,dx + O(a\log a). \quad (8\text{-}14)$$

The terms with $\operatorname{Re}\phi_0 \le 0$ make only lower order contributions. First of all, we can prove a general estimate,

$$\log\left|\frac{[S_{F_{\ell,-r_0}}(s)]_k}{[S_{F_\ell}(s)]_k}\right| = O((k+|s|)\log|s|),$$

just as in Lemma 7.3, to show that

$$\sum_{\varrho_1(\theta)(1-a^{-1/2})\le a/k\le\varrho_1(\theta)} \log\left|\frac{[S_{F_{\ell,-r_0}}(\frac{1}{2}+ae^{i\theta})]_k}{[S_{F_\ell}(\frac{1}{2}+ae^{i\theta})]_k}\right| = O(a^{3/2}\log a). \quad (8\text{-}15)$$

For the remaining terms, we use (8-10) and (8-11) to write

$$\frac{[S_{F_{\ell,-r}}(\frac{1}{2}+k\alpha)]_k}{[S_{F_\ell}(\frac{1}{2}+k\alpha)]_k} = 1 + \frac{e^{-k(\phi_0+\gamma)}}{\pi k\sqrt{\alpha}}\,\frac{[S_{F_\ell}(\frac{1}{2}-k\alpha)]_k}{g_k(\frac{1}{2}+k\alpha;r)}\Big(w_1(r) - \frac{w_1(0)}{w_0(0)}w_0(r)\Big).$$

This gives the estimate

$$\log\left|\frac{[S_{F_{\ell,-r}}(\frac{1}{2}+k\alpha)]_k}{[S_{F_\ell}(\frac{1}{2}+k\alpha)]_k} - 1\right| \le 2k\operatorname{Re}\phi_0(\alpha) + O(\log|k\alpha|).$$

For a sufficiently large, this gives

$$\sum_{a/k\le\varrho_1(\theta)(1-a^{-1/2})} \log\left|\frac{[S_{F_{\ell,-r_0}}(\frac{1}{2}+ae^{i\theta})]_k}{[S_{F_\ell}(\frac{1}{2}+ae^{i\theta})]_k}\right| = O(e^{-c\sqrt{a}}). \quad (8\text{-}16)$$

The estimates (8-14)–(8-16) cover all terms in the sum (8-4), and together yield

$$\log|\tau(\tfrac{1}{2}+ae^{i\theta})| = 2a^2\int_{\varrho_2(\theta)}^{\infty} \frac{2\operatorname{Re}(2\phi_0(xe^{i\theta})-\phi(xe^{i\theta};r_0))}{x^3}\,dx - \frac{\pi a^2}{\omega}\sin^2\theta + O(a\log a)$$

for $a \in \Lambda$ and $0 \le \theta \le \pi/2 - \varepsilon$.

We now integrate over $\theta \in [0, \frac{1}{2}\pi - \varepsilon]$ and use Lemma 4.4 to control the limit $\varepsilon \to 0$, as in the proof of Theorem 7.1. This yields

$$\frac{2}{\pi}\int_0^{\pi/2} \log|\tau(\tfrac{1}{2} + ae^{i\theta})|\, d\theta = \frac{4a^2}{\pi}\int_0^{\pi/2}\int_{\varrho_2(\theta)}^{\infty} \frac{2\operatorname{Re}(2\phi_0(xe^{i\theta}) - \phi(xe^{i\theta}; r_0))}{x^3}\, dx\, d\theta - \tfrac{1}{4}\ell a^2 - o(a^2).$$

To complete the proof of (8-3), recall the definition of $\phi(\alpha; r)$ as the integral of $\sqrt{f}\, dr$ in (5-7). Since the function f occurring there is an even function of r, the function $\phi - \phi_0$ will be odd in r. (This is not readily apparent from the definition (5-9).) This parity implies that

$$I(\alpha, \ell, -r_0) = 2\operatorname{Re}(2\phi_0(\alpha) - \phi(\alpha; r_0)). \qquad \square$$

Acknowledgment

I would like to thank to Plamen Stefanov for suggesting the extension of his results to the hyperbolic setting. I am also grateful for support from the Banff International Research Station, where some of the work for this project was done.

References

[Boas 1954] R. P. Boas, Jr., *Entire functions*, Academic Press, New York, 1954. MR 16,914f Zbl 0058.30201

[Borthwick 2007] D. Borthwick, *Spectral theory of infinite-area hyperbolic surfaces*, Progress in Mathematics **256**, Birkhäuser, Boston, 2007. MR 2008h:58056 Zbl 1130.58001

[Borthwick 2008] D. Borthwick, "Upper and lower bounds on resonances for manifolds hyperbolic near infinity", *Comm. Partial Differential Equations* **33**:7-9 (2008), 1507–1539. MR 2009i:58039 Zbl 1168.58012

[Borthwick 2010] D. Borthwick, "Sharp upper bounds on resonances for perturbations of hyperbolic space", *Asymptot. Anal.* **69**:1-2 (2010), 45–85. MR 2012a:58052 Zbl 1230.58019 arXiv 0910.2439

[Erdélyi et al. 1953] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher transcendental functions, I*, McGraw-Hill, New York, 1953. MR 15,419i Zbl 0051.30303

[Froese 1998] R. Froese, "Upper bounds for the resonance counting function of Schrödinger operators in odd dimensions", *Canad. J. Math.* **50**:3 (1998), 538–546. MR 99f:35150 Zbl 0918.47005

[Guillarmou 2005] C. Guillarmou, "Absence of resonance near the critical line on asymptotically hyperbolic spaces", *Asymptot. Anal.* **42**:1-2 (2005), 105–121. MR 2006e:35235 Zbl 1083.58019

[Guillopé and Zworski 1995] L. Guillopé and M. Zworski, "Upper bounds on the number of resonances for non-compact Riemann surfaces", *J. Funct. Anal.* **129**:2 (1995), 364–389. MR 96b:58116 Zbl 0841.58063

[Guillopé and Zworski 1997] L. Guillopé and M. Zworski, "Scattering asymptotics for Riemann surfaces", *Ann. of Math.* (2) **145**:3 (1997), 597–660. MR 98g:58181 Zbl 0898.58054

[McKean 1972] H. P. McKean, "Selberg's trace formula as applied to a compact Riemann surface", *Comm. Pure Appl. Math.* **25** (1972), 225–246. MR 57 #12843a

[Müller 1992] W. Müller, "Spectral geometry and scattering theory for certain complete surfaces of finite volume", *Invent. Math.* **109**:2 (1992), 265–305. MR 93g:58151 Zbl 0772.58063

[Olver 1974] F. W. J. Olver, *Asymptotics and special functions*, Academic Press, New York, 1974. MR 55 #8655 Zbl 0303.41035

[Parnovski 1995] L. B. Parnovski, "Spectral asymptotics of Laplace operators on surfaces with cusps", *Math. Ann.* **303**:2 (1995), 281–296. MR 97a:11080 Zbl 0849.35093

[Stefanov 2006] P. Stefanov, “Sharp upper bounds on the number of the scattering poles”, *J. Funct. Anal.* **231**:1 (2006), 111–142. MR 2006i:35267 Zbl 1099.35074

[Venkov 1990] A. B. Venkov, *Spectral theory of automorphic functions and its applications*, Mathematics and its Applications (Soviet Series) **51**, Kluwer Academic, Dordrecht, 1990. MR 93a:11046 Zbl 0719.11030

[Vodev 1994] G. Vodev, “Sharp bounds on the number of scattering poles in even-dimensional spaces”, *Duke Math. J.* **74**:1 (1994), 1–17. MR 95e:35153 Zbl 0813.35075

Received 31 Jul 2010. Accepted 26 Feb 2011.

DAVID BORTHWICK: davidb@mathcs.emory.edu
Department of Mathematics and Computer Science, Emory University, 400 Dowman Drive, Atlanta, GA 30322, United States

ANALYSIS AND PDE
Vol. 5, No. 3, 2012
dx.doi.org/10.2140/apde.2012.5.553

A VECTOR FIELD METHOD APPROACH TO IMPROVED DECAY FOR SOLUTIONS TO THE WAVE EQUATION ON A SLOWLY ROTATING KERR BLACK HOLE

JONATHAN LUK

We prove that sufficiently regular solutions to the wave equation $\Box_{g_K}\Phi = 0$ on the exterior of a sufficiently slowly rotating Kerr black hole obey the estimates $|\Phi| \le C(t^*)^{-3/2+\eta}$ on a compact region of r. This is proved with the help of a new vector field commutator that is analogous to the scaling vector field on Minkowski and Schwarzschild spacetime. This result improves the known robust decay rates that are proved using the vector field method in the region of finite r and along the event horizon.

1. Introduction

A major open problem in general relativity is that of the nonlinear stability of Kerr spacetimes. These spacetimes are stationary axisymmetric asymptotically flat black hole solutions to the vacuum Einstein equations

$$R_{\mu\nu} = 0$$

MSC2010: 83C99, 35R01.
Keywords: Kerr spacetime, wave equation, decay estimates.

in $3+1$ dimensions. They are parametrized by (M, a) representing respectively the mass and the specific angular momentum of a black hole; see Section 2. It is conjectured that Kerr spacetimes are asymptotically stable. In the framework of the initial value problem, the stability of Kerr would mean that for any solution to the vacuum Einstein equations with initial data close to the initial data of a Kerr spacetime, its maximal Cauchy development has an exterior region that approaches a nearby, but possibly different, Kerr spacetime.

To study the stability of Kerr spacetimes, it is important to first understand the corresponding linear problem. One way to approach this is to study the linear scalar wave equation $\Box_{g_K}\Phi = 0$, where g_K is the metric on a fixed Kerr background and $\Box_{g_K}$ is the Laplace–Beltrami operator. This can be compared with the proofs of the nonlinear stability of the Minkowski spacetime in which a robust understanding of the quantitative decay properties of solutions to the linear wave equation plays a fundamental role [Christodoulou and Klainerman 1993; Lindblad and Rodnianski 2005].

The Kerr family of spacetimes contains a one-parameter subfamily known as the Schwarzschild spacetimes for which $a = 0$. It is natural when studying the wave equation on Kerr spacetimes to begin by focusing on the wave equation on Schwarzschild spacetimes. Pointwise boundedness and decay of the solutions to the wave equation on Schwarzschild spacetimes has been proved in [Wald 1979; Kay and Wald 1987; Machedon and Stalker 2002; Blue and Sterbenz 2006; Dafermos and Rodnianski 2009; Kronthaler 2007; Blue and Soffer 2006; Donninger et al. 2011; Tataru 2009]. In particular, Dafermos and Rodnianski used the vector field method to show that on the exterior region of the Schwarzschild spacetimes, including along the event horizon, solutions to the linear wave equation satisfy $|\Phi| \le C(t^*)^{-1}$, where t^* is a regular coordinate (up to the event horizon) that approaches infinity towards null infinity. In an earlier work [Luk 2010], we improved this decay rate. More precisely, we showed that sufficiently regular solutions to the wave equation $\Box_g\Phi = 0$ on the Schwarzschild black hole obey the estimates $|\Phi| \le C_\eta(t^*)^{-3/2+\eta}$ for any $\eta > 0$ on a compact region of r, including along the event horizon and inside the black hole.

This paper generalizes the result above to Kerr spacetimes where $a \ll M$. For Kerr spacetimes satisfying this condition, Dafermos and Rodnianski [2011], and subsequently Andersson and Blue [2009], have proved a decay rate in the exterior region of the Kerr spacetime, including along the event horizon, of $|\Phi| \le C(t^*)^{-1+\eta}$, where t^* is a regular coordinate to be defined later, and with t^* we will define a foliation of the exterior region of Kerr spacetime by the spacelike hypersurfaces Σ_{t^*}. Extending the methods in [Luk 2010], we are able to improve this decay rate using the vector field method.

Theorem 1. *Suppose $\Box_{g_K}\Phi = 0$. Then for all $\eta > 0$ and all $M > 0$ there exists a_0 such that the following estimates hold on Kerr spacetimes with (M, a) for which $a \le a_0$.*

(1) *Improved decay of nondegenerate energy*:

$$\sum_{j=0}^{M} \int_{\Sigma_{t^*}\cap\{r\le R\}} (D^j\Phi)^2 \le C_R E_M (t^*)^{-3+\eta}.$$

(2) *Improved pointwise decay*:

$$\sum_{j=0}^{M}|D^j\Phi| \leq C_R E'_M(t^*)^{-3/2+\eta} \quad \textit{for } r \leq R.$$

Here, D denotes derivatives in a regular coordinate system (See Section 2). E_M and E'_M depend only on M and some weighted Sobolev norm of the initial data.

A more precise version of this theorem will be given in Section 6. Our proof relies on an analogue of the scaling vector field on Minkowski spacetime. Recall that in Minkowski spacetime the vector field $S = t\partial_t + r\partial_r$ is conformally Killing and satisfies $[\Box_m, S] = 2\Box_m$. Hence any estimates that hold for Φ a solution to $\Box_m\Phi = 0$ would also hold for $S\Phi$. However, S has a weight that is increasing with t. Hence one can hope to prove a better estimate for Φ using the estimates for $S\Phi$. (See, for example, [Klainerman and Sideris 1996]).

In [Luk 2010], we introduced an analogue of the scaling vector field on Schwarzschild spacetimes. We defined, in the Regge–Wheeler tortoise coordinate (see Section 2), the vector field $S = t\partial_t + r^*\partial_{r^*}$. In [Luk 2010], we studied the commutator $[\Box_{g_S}, S]$ and showed that all the error terms can be controlled. Thus, up to a loss of t^η (for η arbitrarily small), $S\Phi$ obeys all the estimates of Φ that were proved in [Dafermos and Rodnianski 2009]. In particular, we showed that $S\Phi$, like Φ itself, obeys a local integrated decay estimate

$$\int_{t'}^{t}\int_{r_1}^{r_2}(D^k\Phi)^2\,dr\,dt \leq CE_k(t')^{-2} \qquad \text{for } t' \leq t \leq (1.1)t',$$
$$\int_{t'}^{t}\int_{r_1}^{r_2}(SD^k\Phi)^2\,dr\,dt \leq CE_k(t')^{-2+\eta} \quad \text{for } t' \leq t \leq (1.1)t'.$$

From this we proved an improved decay of the L^2 norm of $D^k\Phi$. We will explain the main idea in the case $k = 0$. Firstly, the local integrated decay for Φ would already imply on a sequence of t_i slices, with $t_i \leq t_{i+1} \leq (1.1)^2 t_i$, that Φ obeys a better decay rate, namely $\Phi(t_i) \leq Ct_i^{-3/2}$. We then introduced a new method to use the estimates for $S\Phi$, which can be explained heuristically as follows. Given any time t, we find the largest $t_i \leq t$ that has a better decay rate. Then we integrated from t_i to t using the vector field S. At this point S has a weight that grows like t. Hence we have, at least schematically,

$$\int_{r_1}^{r_2}\Phi(t)^2 dr \leq C\Big(\int_{r_1}^{r_2}\Phi(t_i)^2 dr + t^{-1}\Big|\int_{t_i}^{t}\int_{r_1}^{r_2}S(\Phi^2)\,dr\,dt\Big|\Big).$$

We then notice that the last term can be estimated by the local integrated decay estimates

$$\Big|\int_{t_i}^{t}\int_{r_1}^{r_2}S(\Phi^2)\,dr\,dt\Big| \leq C\Big(\int_{t_i}^{t}\int_{r_1}^{r_2}\Phi^2\,dr\,dt + \int_{t_i}^{t}\int_{r_1}^{r_2}(S\Phi)^2\,dr\,dt\Big) \leq Ct^{-2+\eta}.$$

Putting these together, we would get

$$\int_{r_1}^{r_2}\Phi(t)^2 dr \leq Ct^{-3+\eta}.$$

Using this method, we also showed the improved decay for the L^2 norm of higher derivatives. Pointwise decay estimate thus followed from standard Sobolev embedding.

In this paper we would like to carry out a similar argument. We introduce a scaling vector field (which we again call S) which is the same as in [Luk 2010] at the asymptotically flat end, but is smooth up to and across the event horizon. We will prove a local integrated decay estimate for $S\Phi$ and use the argument in [ibid.] as outlined above to prove an improved decay rate. The most difficult part of the argument is to control the error terms coming from the commutation of $\Box_{g_K}$ and (the modified) S, that is, the term $[\Box_{g_K}, S]\Phi$. To control this, we need to use estimates for derivatives of Φ, which in turn are provided by the energy estimates for the homogeneous equation $\Box_{g_K}\Phi = 0$ proved in [Dafermos and Rodnianski 2011; 2008]. This term schematically looks like

$$[\Box_{g_K}, S]\Phi = O(1)\Box_{g_K}\Phi + O(r^{-2+\delta})(D^2\Phi + D\Phi + rD\nabla\!\!\!\!/\,\Phi), \tag{1}$$

where $\nabla\!\!\!\!/$ is an angular derivative on the 2-sphere. The term $O(1)\Box_{g_K}\Phi$ vanishes since we are considering $\Box_{g_K}\Phi = 0$. The other terms have the two desirable features. First, although S has a weight in t^*, the commutator is independent of t^*, which is a result of ∂_{t^*} being a Killing vector field. Second, these terms decay as $r \to \infty$, which is a result of the asymptotic flatness of Kerr spacetimes. The last term would appear to have less decay in r, which is also the case in Schwarzschild spacetimes. In that case, we controlled this term in [Luk 2010] by commuting the equation with Ω, the generators of the spherical symmetry of Schwarzschild spacetimes. The quantity $\Omega\Phi$ would then give us control over an extra power of r. One difficulty that arises in the case of Kerr spacetimes is that they are not spherically symmetric. Nevertheless, following [Dafermos and Rodnianski 2008], we can construct an analog of Ω, call it $\tilde{\Omega}$, that is an asymptotic symmetry, that is, the commutator $[\Box_{g_K}, \tilde{\Omega}]$ would decay in r. The nondegenerate energy of $\tilde{\Omega}\Phi$ would then control the last term in the above expression. Moreover, it is sufficient to define $\tilde{\Omega}$ only when r is very large since otherwise the factor in r can be absorbed by constants. However, in a finite region of r, the commutator $[\Box_{g_K}, S]$ would in general be large.

To understand which quantities of $S\Phi$ have to be controlled, we rederive the energy estimates in [Dafermos and Rodnianski 2008] in the slightly more general case of the inhomogeneous equation $\Box_{g_K}\Phi = G$. This would also immediately imply that for the linear inhomogeneous equation $\Box_{g_K}\Phi = G$ with sufficiently regular and sufficiently decaying (both in space and time) G, the solution Φ, assuming that the initial data is sufficiently regular, would decay with a rate of $(t^*)^{-1+\eta}$, precisely as that in [ibid.]. We will then apply this to the equations for $\tilde{\Omega}\Phi$ and $S\Phi$. To derive these energy estimates, we will use the (non-Killing) vector field multipliers N and Z. Here N is a modification of ∂_{t^*} so that it is timelike everywhere, including near the event horizon. The use of N tackles the issue of superradiance, a difficulty that arises from the spacelike nature of ∂_{t^*} near the event horizon. Z is an analogue of the conformal vector field $u^2\partial_u + v^2\partial_v$ in Minkowski spacetime and is used to prove decay.

Since we will use vector field multipliers that have weights in t^* and r, to prove the energy estimates at $t^* = \tau$ for the inhomogeneous equation we would have to control the term (as well as other similar or more easily controlled terms)

$$\iint_{\mathcal{R}(\tau_0,\tau)} (t^*)^2 r^{1+\delta} G^2,$$

where the integration over space and the t^* interval $[\tau_0, \tau]$. To prove the energy estimates for $S\Phi$, we need to show that for G as in (1), this is bounded by $C(\tau)^\eta$. We split this into two parts: $r \le \frac{1}{2}t^*$ and $r \ge \frac{1}{2}t^*$. For $r \le \frac{1}{2}t^*$, we can replace $r^{1+\delta}$ by $r^{-3+2\delta}$ since G decays in r. Then, we use the fact that

$$\sum_{k=1}^{N} \iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)\cap\{r\le\frac{1}{2}t^*\}} r^{-1+\delta}(D^k\Phi)^2 \le C\tau^{-2+\eta}.$$

Hence if we sum up the whole integral by integrating in $[\tau_0, (1.1)\tau_0]$, $[(1.1)\tau_0, (1.1)^2\tau_0]$ etc., we will get a bound of

$$\sum_{i=0}^{\lfloor\log\tau\rfloor+1} (1.1)^i\tau_0 \sim_\eta \tau^\eta.$$

For $r \ge \frac{1}{2}t^*$, we do not have a decay estimate for the integrated in time estimate. However, we would still have an almost boundedness estimate:

$$\sum_{k=1}^{N} \iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)\cap\{r\ge\frac{1}{2}t^*\}} r^{-1+\delta}(D^k\Phi)^2 \le C\tau^{\eta}.$$

Notice, moreover, that $G^2 \sim r^{-3+\delta}(D^k\Phi)^2$ and this region we have $r^{-3+\delta} \le (t^*)^{-2}r^{-1+\delta}$. Hence we again have

$$\sum_{k=1}^{N} \iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)\cap\{r\ge\frac{1}{2}t^*\}} G^2 \le C\tau^{-2+\eta},$$

and the required estimate followed in the same manner as in the case $r \le \frac{1}{2}t^*$.

With the modified S, which is smooth up to the event horizon (contrary to [Luk 2010]), we can prove the improved decay estimates for the L^2 norm of Φ and $D\Phi$ once these error terms are controlled. Using the commutation with the Killing vector ∂_{t^*}, we would also have control for L^2 norm of $D\partial_{t^*}^k\Phi$. Away from the event horizon, this is sufficient to control all other derivatives by elliptic estimates. However, since near the event horizon, ∂_{t^*} is not Killing, we would not have control over other derivatives. Here, we follow [Dafermos and Rodnianski 2008; 2011] and commute the equation with a version of the red-shift vector field, $\hat{Y}$. Once we control $D\hat{Y}^k\partial_{t^*}^j\Phi$ we can use the wave equation to control (any derivatives of) $\Delta\!\!\!\!/\,\Phi$, where $\Delta\!\!\!\!/$ is the Laplace–Beltrami operator on the sphere, which is elliptic. We can thus control derivatives of Φ in any directions. We will show, moreover, that the commutator $[\Box_{g_K}, \hat{Y}]$ has the property that the inhomogeneous terms can be controlled once we have controlled the L^2 norm of $D\partial_{t^*}^k\Phi$. This implies that $\hat{Y}\Phi$ would decay in the same rate as $\partial_{t^*}\Phi$ for which we have already derived an improved decay rate.

We now turn to some history of this problem. We mention some results on Kerr spacetimes with $a > 0$ here and refer the readers to [Dafermos and Rodnianski 2008; Luk 2010] for references on the corresponding problem on Schwarzschild spacetimes. There has been a large literature on the mode stability and nonquantitative decay of azimuthal modes. See for example [Press and Teukolsky 1973; Hartle and Wilkins 1974; Whiting 1989; Finster et al. 2008; 2006] and references in [Dafermos and Rodnianski 2008]. The first global result for the Cauchy problem was obtained by Dafermos and Rodnianski [2011], who proved that for a class of small, axisymmetric, stationary perturbations of Schwarzschild spacetime, which include Kerr spacetimes that rotate sufficiently slowly, solutions to the wave equation are uniformly bounded. Similar results were obtained later using an integrated decay estimate on slowly rotating Kerr spacetimes by Tataru and Tohaneanu [2011]. Using the integrated decay estimate, Tohaneanu [2012] also proved Strichartz estimates.

Decay for general solutions to the wave equation on sufficiently slowly rotating Kerr spacetimes was first proved by Dafermos and Rodnianski [2008] with a quantitative rate of $|\Phi| \leq C(t^*)^{-1+Ca}$. A similar result was later obtained by Andersson and Blue [2009] using a physical space construction to obtain an integrated decay estimate. In all of [Tataru and Tohaneanu 2011; Dafermos and Rodnianski 2008; Andersson and Blue 2009], the integrated decay estimate is proved and plays an important role. All proofs of such estimates rely heavily on the separability of the wave equation, or equivalently, the existence of a Killing tensor on Kerr spacetime. In a recent work, Dafermos and Rodnianski [2010] show that assuming the integrated decay estimate (nondegenerate up to the event horizon if it exists) and boundedness for the wave equation on an asymptotically flat spacetime, the decay rate $|\Phi| \leq C(t^*)^{-1}$ holds. This in particular improves the rates in [Dafermos and Rodnianski 2008; Andersson and Blue 2009]. In a similar framework, but assuming in addition exact stationarity, Tataru [2009] proved a local decay rate of $(t^*)^{-3}$ using Fourier-analytic methods. This applies in particular to sufficiently slowly rotating Kerr spacetimes. Dafermos and Rodnianski have recently announced a proof for the decay of solutions to the wave equation on the full range of subextremal Kerr spacetimes $a < M$.

In view of the nonlinear problem, it is important to understand decay in a robust manner. In particular, past experience shows that refined decay estimates might not be stable in nonlinear problems. The vector field method is known to be robust and culminated in the proof of the stability of the Minkowski spacetime [Christodoulou and Klainerman 1993; Lindblad and Rodnianski 2005]. We prove our decay result using the vector field method with the expectation that the method will be useful for studying nonlinear problems. As a model problem, we will study the semilinear equation with a null condition on a fixed slowly rotating Kerr background. In a forthcoming paper that we will show the global existence of solutions with small initial data for this class of equations. We will also study the asymptotic behavior of these solutions. The null condition, which is a special structure of the nonlinearity, has served as an important model for the proofs of the nonlinear stability of Minkowski spacetime and we hope that it will find relevance to the problem of the nonlinear stability of Kerr spacetime.

We end the introduction with an overview of the paper. In Section 2, we will introduce the Kerr geometry, including a few different coordinate systems that we will find useful in the rest of the paper. In Section 4, we introduce the (non-Killing) vector field commutators that will be used. These include the scaling vector field S, which is the main tool for obtaining improved decay rates. In Section 5, we introduce the formalism for vector field multipliers. We then have all the notation necessary to state the precise form of our main theorem in Section 6. In Sections 7, 8 and 9, we prove the main energy estimates using the vector field multipliers N, X and Z. We write down the energy estimates in the most general form, allowing for the possibility of controlling the inhomogeneous terms in different energy norms. Such generality is unnecessary for the result in this paper, but will be useful in studying the null condition. Starting from Section 10, we return to the homogeneous equation. In Section 10, we write down the energy estimates proved in [Dafermos and Rodnianski 2008]. We then derive the energy estimates after commuting with $\hat{Y}$, $\widetilde{\Omega}$ and S in Sections 11, 12 and 13 respectively. Finally, using the estimates for $S\Phi$, we prove the main theorem in Section 14.

2. Geometry of Kerr spacetime

2.1. ***Kerr coordinates.*** The Kerr metric in the Boyer–Lindquist coordinates takes the form

$$g_K = -\left(1 - \frac{2M}{r\left(1 + \frac{a^2\cos^2\theta}{r^2}\right)}\right) dt^2 + \frac{1 + \frac{a^2\cos^2\theta}{r^2}}{1 - \frac{2M}{r} + \frac{a^2}{r^2}}\, dr^2 + r^2\left(1 + \frac{a^2\cos^2\theta}{r^2}\right) d\theta^2$$
$$+ r^2\left(1 + \frac{a^2}{r^2} + \frac{2M}{r}\frac{a^2\sin^2\theta}{r^2\left(1 + \frac{a^2\cos^2\theta}{r^2}\right)}\right)\sin^2\theta\, d\phi^2 - 4M\frac{a\sin^2\theta}{r\left(1 + \frac{a^2\cos^2\theta}{r^2}\right)}\, dt\, d\phi. \quad (2)$$

In this paper, we will consider Kerr spacetimes with a small. It can then be thought of as a small perturbation of Schwarzschild spacetimes because by setting $a = 0$, we recover the Schwarzschild metric:

$$g_S = -\left(1 - \frac{2M}{r_S}\right) dt_S^2 + \left(1 - \frac{2M}{r_S}\right)^{-1} dr_S^2 + r_S^2\, d\theta^2 + r_S^2 \sin^2\theta\, d\phi^2.$$

The Cauchy development of Kerr spacetimes can be described schematically by taking a two-dimensional slice as in Figure 1.

Notice that (2) represents the metric on the exterior region (the right side in the diagram). In the coordinate system above, this is the region $\{r \geq r_+\}$, where r_+ is the larger root of $\Delta = r^2 - 2Mr + a^2$. This is the region that we will study. We foliate the exterior region of the Kerr spacetime by hypersurfaces Σ_τ as depicted in the diagram. A precise definition of the hypersurface Σ_τ will be given in Section 3.3. The coordinates in (2) are not regular at the event horizon $\mathcal{H}^+ = \{r = r_+\}$. It will be helpful in the sequel to use different coordinate systems on Kerr spacetimes. From now on we will call the coordinate system on which the metric (2) is defined the Kerr (t, r, θ, ϕ) coordinates. We define a new coordinate system, the Kerr (t, r^*, θ, ϕ) coordinates, by

$$\frac{dr^*}{dr} = \frac{r^2 + a^2}{\Delta},$$

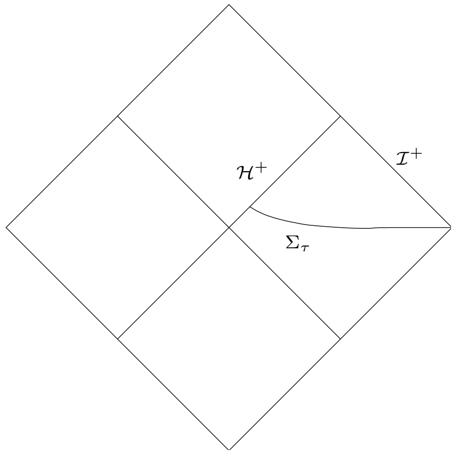

Figure 1. Kerr spacetime.

where $\Delta = r^2 - 2Mr + a^2$ is zero at the event horizon. In this coordinate system, the metric looks like

$$g_K = -\left(1 - \frac{2M}{r\left(1+\frac{a^2\cos^2\theta}{r^2}\right)}\right)dt^2 + \Delta(r^2+a^2)^2(r^2+a^2\cos^2\theta)dr^{*2} + r^2\left(1+\frac{a^2\cos^2\theta}{r^2}\right)d\theta^2$$
$$+ r^2\left(1+\frac{a^2}{r^2}+\frac{2M}{r}\frac{a^2\sin^2\theta}{r^2\left(1+\frac{a^2\cos^2\theta}{r^2}\right)}\right)\sin^2\theta\, d\phi^2 - 4M\frac{a\sin^2\theta}{r\left(1+\frac{a^2\cos^2\theta}{r^2}\right)}\,dt\,d\phi.$$

Since the definition of r^* depends only on r, it is unambiguous to talk about the vector ∂_t.

2.2. ***Schwarzschild coordinates.*** In order to compare calculations on Kerr spacetimes to calculations on Schwarzschild spacetimes, it is helpful to exhibit a diffeomorphism between the two. We do so by defining an explicit map between the coordinate functions (t, r, θ, ϕ) on a Kerr spacetime and the coordinate functions $(t_S, r_S, \theta_S, \phi_S)$ on a Schwarzschild spacetime with the same mass. These will be defined differently near and away from the event horizon. Take

$$\chi(r) = \begin{cases} 1 & \text{if } r \le r_Y^- - \frac{1}{2}(r_Y^- - r_+), \\ 0 & \text{if } r \ge r_Y^- - \frac{1}{4}(r_Y^- - r_+), \end{cases}$$

where r_+, as above, is the larger root of $\Delta = r^2 - 2Mr + a^2$ and $r_Y^- > r_+$ is a constant to be determined later. With this $\chi(r)$, we can then define

$$\begin{aligned} r_S^2 - 2Mr_S &= r^2 - 2Mr + a^2, \\ t_S + \chi(r_S)2M\log(r_S - 2M) &= t + \chi(r)h(r), \qquad \text{where } \frac{dh(r)}{dr} = \frac{2Mr}{r^2 - 2Mr + a^2}, \\ \theta_S &= \theta, \\ \phi_S &= \phi + \chi(r)P(r), \qquad \text{where } \frac{dP(r)}{dr} = \frac{a}{r^2 - 2Mr + a^2}. \end{aligned}$$

Then, by identifying $(t_S, r_S, \theta_S, \phi_S)$ with the corresponding coordinate functions on Schwarzschild spacetimes, we have a diffeomorphism between Kerr spacetimes and Schwarzschild spacetimes. This coordinate system will be used and will be called the Schwarzschild $(t_S, r_S, \theta_S, \phi_S)$ coordinates on Kerr spacetimes. Once we have this diffeomorphism, we can put any system of Schwarzschild coordinates on Kerr spacetimes. These include the Schwarzschild $(t_S^*, r_S, \theta_S, \phi_S)$ coordinates, where $t_S^* = t_S + \chi(r_S) 2M \log(r_S - 2M)$ and r_S, θ_S, ϕ_S are defined as above. We also define

$$t^* = t_S^* = t_S + \chi(r_S) 2M \log(r_S - 2M)$$

and use the Kerr (t^*, r, θ, ϕ^*) coordinates. Notice that $\partial_{t^*} = \partial_{t_S^*}$.

It is common to denote on Schwarzschild spacetimes $\mu = 2M/r_S$. We will take the same notation on Kerr spacetimes, with the understanding that it is always defined with respect to the Schwarzschild r_S coordinates. In particular $(1-\mu)$ approaches 0 as $r \to r_+$ (the event horizon).

Another system of Schwarzschild coordinates can be defined by considering two coordinate charts on the standard unit 2-sphere and introducing a system of coordinates (x_S^A, x_S^B) on each of them. We then define the Schwarzschild $(t_S^*, r_S, x_S^A, x_S^B)$ coordinates in the obvious manner. Using this coordinate system and the diffeomorphism as above, we have, for small a,

$$|(g_K)_{\alpha\beta} - (g_S)_{\alpha\beta}| \le \epsilon r^{-2}. \tag{3}$$

This smallness assumption will be used throughout this paper.

2.3. ***Null frame near event horizon.*** Some extra cancellations for the estimates near the event horizon are best captured using a null frame. Hence we define a null frame $\{\hat{V}, \hat{Y}, E_1, E_2\}$ in the region $r \le r_Y^-$, where r_Y^- is to be determined later. On the event horizon,

$$V = \partial_{t^*} + \frac{a}{2Mr_+}\partial_{\phi^*}$$

is the Killing null generator. A direct computation shows that it satisfies

$$\nabla_V V = \kappa V,$$

where κ is a strictly positive number on the event horizon. We want to extend V to a null frame. On the event horizon, define $\hat{Y}$ first on a 2-sphere given by a fixed t^* to be null, orthogonal to the 2-sphere and require that $g_K(V, \hat{Y}) = -2$. Define also locally an orthonormal frame $\{E_1, E_2\}$ tangent to the fixed 2-sphere. In the sequel, we will only need to work with a local null frame. We then extend this null frame off the fixed 2-sphere on the event horizon (with $\hat{V}|_{\mathcal{H}} = V$) by requiring

$$\nabla_{\hat{Y}} \hat{Y} = \nabla_{\hat{Y}} \hat{V} = \nabla_{\hat{Y}} E_A = 0, \tag{4}$$

where $A \in \{1, 2\}$. Then extend this null frame using the isomorphisms generated by V. The equations above hold everywhere. If we choose r_Y^- close enough to r_+, we still have, by Taylor's theorem,

$$\nabla_{\hat{V}} \hat{V} = \kappa \hat{V} + b^Y \hat{Y} + b^1 E_1 + b^2 E_2, \tag{5}$$

where κ is a strictly positive function in $r_+ \le r \le r_Y^-$ bounded away from 0, and $|b^\alpha| \le C\,(1-\mu)$.

In Schwarzschild spacetime, consider the frame on $\mathcal{S}^2$ given by $\{(r_S^2 \sin\theta_S)^{-1}\partial_{\phi_S}, r_S^{-1}\partial_{\theta_S}\}$. Then we get

$$\hat{V} = (1+\mu)\partial_{t_S^*} + (1-\mu)\partial_{r_S}, \quad \hat{Y} = \partial_{t_S^*} - \partial_{r_S}, \quad E_1 = r_S^{-1}\partial_{\theta_S}, \quad E_2 = (r_S \sin\theta_S)^{-1}\partial_{\phi_S}$$

Since we consider Kerr spacetimes on which the metric is close to that on a Schwarzschild spacetime, the null frame can be expressed in (t^*, r, θ, ϕ^*) coordinates as

$$\begin{aligned} \hat{V} &= (1+\mu)\partial_{t^*} + (1-\mu)\partial_r + O_1(\epsilon)\partial, & E_1 &= r^{-1}\partial_\theta + O_1(\epsilon)\partial, \\ \hat{Y} &= \partial_{t^*} - \partial_r + O_1(\epsilon)\partial, & E_2 &= (r\sin\theta)^{-1}\partial_{\phi^*} + O_1(\epsilon)\partial. \end{aligned}$$

Alternatively, if we write E_α, where $\alpha = 1, 2, 3, 4$, for the null frame, we have

$$\begin{aligned} (1+\mu)\partial_{t^*} + (1-\mu)\partial_r &= \hat{V} + O_1(\epsilon)E_\alpha, & \partial_\theta &= rE_1 + O_1(\epsilon)E_\alpha, \\ \partial_{t^*} - \partial_r &= \hat{Y} + O_1(\epsilon)E_\alpha, & \partial_{\phi^*} &= r\sin\theta E_2 + O_1(\epsilon)E_\alpha. \end{aligned}$$

We also define the vector fields $\hat{V}$, $\hat{Y}$, E_1, E_2 outside $\{r \le r_Y^-\}$ by requiring them to be compactly supported in $\{r \le r_Y^+\}$ (for some r_Y^+ to be determined) and invariant under the one-parameter families of isometries generated by ∂_{t^*} and ∂_{ϕ^*}. Notice that in particular there is no requirement that the vector fields form a null frame in the region $\{r_Y^- \le r \le r_Y^+\}$.

3. Notation

3.1. ***Constants.*** Throughout this paper, we will use C to denote a large constant and c to denote a small constant. They can be different from line to line. We will also use A to denote bootstrap constants and we think of A to be large, that is, $A \gg C$. We also use the notation $O_i(1)$ and $O_i(\epsilon)$ to denote terms that are bounded up to a constant by 1 and ϵ, with bounds that improve by r^{-1} for each derivative up to the i-th derivative. We will also use the notation $f \sim g$ to denote $cf \le g \le Cf$.

There are some constants that we will choose in the proof. The following are values of r in the Kerr coordinates:

$$r_+ < r_Y^- < r_Y^+ < \tfrac{11}{4}M < R_\Omega.$$

We will fix r_Y^+ and r_Y^- in Remarks and, respectively.

There are also smallness parameters that can be thought of as obeying

$$0 < \delta < \epsilon \ll \eta \ll e.$$

We use ϵ to denote the smallness of the specific angular momentum a of the spacetimes that we are working on. We use $\eta \sim C\epsilon$ to denote the loss in the decay rate of the solutions to the wave equation as compared to that on Schwarzschild spacetimes. We use e to construct the nondegenerate energy, and use δ and δ' as small parameters whenever they are needed. The parameters δ and δ' need not be fixed from line to line.

3.2. ***Values of t^*.*** We will adopt the following as much as possible: We denote by t^* a general value of t^*. In particular, it will be used for integration variables. We denote by τ the t^* value for which we want an estimate and by τ_0 the t^* value where the initial data is posed. We will always assume $\tau_0 \geq 1$ and the reader can think of $\tau_0 = 1$. When integrating, we will often denote the endpoints by τ' and τ. Finally, at a few places we will need to choose a particular value of t^* in an interval. This is usually done to achieve the maximum or minimum of the energy quantities. We often denote such choices as $\tilde{\tau}$.

3.3. ***Integration.***

Definition 1. Define the following sets:

- $\Sigma_\tau = \{t^* = \tau\}$.
- $\mathcal{R}(\tau', \tau) = \{\tau' \leq t^* \leq \tau\}$.
- $\mathcal{H}(\tau', \tau) = \{r = r_+, \tau' \leq t^* \leq \tau\}$.

When integrating on these sets, we will normally integrate with respect to the volume form, which we suppress. On Σ_τ the volume form is $\sqrt{\det g_K|_{\Sigma_\tau}}$. On $\mathcal{R}(\tau', \tau)$, the volume form is $\sqrt{\det g_K}$. However, on the event horizon $\mathcal{H}(\tau', \tau)$, the surface is null and the metric is degenerate. Nevertheless, on $\mathcal{H}(\tau', \tau)$, the integrand will always be of the form $J_\mu n^\mu_{\Sigma_{\mathcal{H}^+}}$, where $n^\mu_{\Sigma_{\mathcal{H}^+}}$ is the normal to $\mathcal{H}(\tau', \tau)$. We will hence take the volume form corresponding to the (arbitrarily) chosen normal. Occasionally, we will also integrate over the topological 2-spheres given by fixing t and r. We will denote the volume form by $dA = \sqrt{\det g_K|_{\mathbb{S}^2}}$.

For some computations, however, it is more convenient to write down the volume form explicitly in coordinates. In our notation, the following two expressions denote the same integral:

$$\int_{\Sigma_\tau} f = \int_{\Sigma_\tau} f \sqrt{\det g_K|_{\Sigma_\tau}}\, dr d\theta\, d\phi.$$

When we write the integrals, we will often use $\iint$ to denote an integral over a spacetime region and use $\int$ denote an integral over a spacelike or null hypersurface.

The volume form on Σ_{t^*} can be compared with that on $\mathcal{R}(\tau', \tau)$. In particular, we have

$$\iint_{\mathcal{R}(\tau',\tau)} f \sim \int_{\tau'}^{\tau} \Big(\int_{\Sigma_{t^*}} f \Big)\, dt^*.$$

4. Vector field commutators

In this section, we discuss the vector field commutators that we will use in this article. One obvious such vector field is the Killing vector field ∂_{t^*}, which satisfies

$$[\Box_{g_K}, \partial_{t^*}] = 0.$$

In addition to ∂_{t^*}, we will use three *non-Killing* vector fields S, $\hat{Y}$ and Ω_i to control higher derivatives of the solution Φ. We introduce S, a new vector field, to obtain the improved decay rate for the solution Φ. We will follow [Dafermos and Rodnianski 2008; 2011], defining the commutator $\hat{Y}$ to estimate Φ near the event horizon. We will use the vector fields Ω_i, which are analogues of the angular momentum vector fields in Schwarzschild spacetime, to control the error terms coming from the commutator $[\Box_{g_K}, S]$.

4.1. ***Vector field commutators under metric perturbations.*** Some computations are easier in Schwarzschild spacetime than in Kerr spacetime. In the sequel, we will often consider a fixed vector field on the differentiable structure of the Schwarzschild exterior. We now show that for such vector fields, the commutators with $\Box_{g_S}$ and $\Box_{g_K}$ are close to each other as long as a is chosen to be sufficiently small:

Proposition 2. *Consider either the Schwarzschild $(t_S^*, r_S, x_S^A, x_S^B)$ coordinates or $(t_S, r_S \geq r_Y^-, x_S^A, x_S^B)$ coordinates. Suppose V is a vector field defined on either of these coordinates. Then*

$$\big|[\Box_{g_K} - \Box_{g_S}, V]\Phi\big| \leq C\epsilon r^{-2}\Big(\sum_{m=1}^{2}\sum_{k=1}^{2}\max_{\alpha}|\partial^m V^\alpha||\partial^k\Phi|\Big),$$

where ∂ is the coordinate derivative for the coordinate system on which V is defined.

Proof. We rewrite

$$\Box_{g_S} = g_S^{\alpha\beta}\partial_\alpha\partial_\beta + \eta_S^\alpha\partial_\alpha \quad \text{and} \quad \Box_{g_K} = g_K^{\alpha\beta}\partial_\alpha\partial_\beta + \eta_K^\alpha\partial_\alpha.$$

Using $|(g_K)_{\alpha\beta} - (g_S)_{\alpha\beta}| \leq \epsilon r^{-2}$ and $|\partial_\gamma((g_K)_{\alpha\beta} - (g_S)_{\alpha\beta})| \leq \epsilon r^{-2}$, we have $|\sqrt{-\det g_K} - \sqrt{-\det g_S}| \leq \epsilon r^{-2}$ and $|\partial_\alpha(\sqrt{-\det g_K} - \sqrt{-\det g_S})| \leq \epsilon r^{-2}$. Therefore,

$$\sup_{\alpha,\beta}|g_S^{\alpha\beta} - g_K^{\alpha\beta}| + \sup_{\alpha}|\eta_S^\alpha - \eta_K^\alpha| \leq C\epsilon r^{-2}.$$

Therefore,

$$\begin{aligned}\big|[\Box_{g_K} - \Box_{g_S}, V]\Phi\big| &\leq \big|(g_K^{\alpha\beta} - g_S^{\alpha\beta})(\partial_\alpha\partial_\beta V^\gamma)\partial_\gamma\Phi\big| + 2\big|(g_K^{\alpha\beta} - g_S^{\alpha\beta})\partial_\alpha V^\gamma\partial_\beta\partial_\gamma\Phi\big| \\ &\qquad + \big|(\eta_S^\alpha - \eta_K^\alpha)(\partial_\alpha V^\gamma)\partial_\gamma\Phi\big| \\ &\leq C\epsilon r^{-2}\Big(\sum_{m=1}^{2}\sum_{k=1}^{2}\sup_{\alpha}|\partial^m V^\alpha||\partial^k\Phi|\Big). \end{aligned}$$

□

4.2. ***Commutator S.*** We construct a commuting vector field S on Schwarzschild that is different from [Luk 2010] and is stable under perturbation.

Define $S = t_S^*\partial_{t_S^*} + h(r_S)\partial_{r_S}$, where

$$h(r) = \begin{cases} r - 2M & \text{if } r \sim 2M, \\ (r + 2M\log(r-2M) - 3M - 2M\log M)(1-\mu) & \text{if } r \geq R, \end{cases}$$

for some large R, and it is interpolated so that it is smooth and nonnegative. For $r \geq R$, since $t^* = t$, this agrees with the definition in [Luk 2010]. Therefore we have

$$[\Box_{g_S}, S] = \Big(2 + \frac{r^*\mu}{r}\Big)\Box_{g_S} + \frac{2}{r}\Big(\frac{r^*}{r} - 1 - \frac{2r^*\mu}{r}\Big)\partial_{r^*} + 2\Big(\Big(\frac{r^*}{r} - 1\Big) - \frac{3r^*\mu}{2r}\Big)\not{\Delta}, \tag{6}$$

where $\not{\Delta}$ is the Laplace–Beltrami operator on the standard sphere. In the coordinates (t^*, r, θ, ϕ),

$$\Box_{g_S} = -\alpha_1(r)\,\partial_t^2 + \alpha_2(r)\partial_r^2 + \alpha_3(r)\partial_r\partial_t + \alpha_4(r)\partial_t + \alpha_5(r)\partial_r + \not{\Delta}.$$

The crucial observation is that all α_i are smooth and bounded and depend only on r. Noting that α_i does not depend on t, we have

$$[\Box_{g_S}, S] = \beta_1(r)\,\partial_t^2 + \beta_2(r)\partial_r^2 + \beta_3(r)\partial_r\partial_t + \beta_4(r)\partial_t + \beta_5(r)\partial_r + \beta_6(r)\not\Delta.$$

Again, it is important to note that all β_i are smooth, bounded and depend only on r. The form of β_i for $r \geq R$ is given by (6).

We consider the same vector field S on Kerr. Using Proposition 2, and noting that $\partial^m S^\alpha$ is bounded for $m \geq 1$, we have for $r > R$

$$\left|[\Box_{g_K}, S]\Phi - \left(2+\frac{r^*\mu}{r}\right)\Box_{g_K}\Phi - \frac{2}{r}\left(\frac{r^*}{r} - 1 - \frac{2r^*\mu}{r}\right)\partial_{r^*}\Phi - 2\left(\left(\frac{r^*}{r}-1\right) - \frac{3r^*\mu}{2r}\right)\not\Delta\Phi\right| \leq C\epsilon r^{-2}\left(\sum_{k=1}^{2}|\partial^k\Phi|\right),$$

and for $r \leq R$,

$$\left|[\Box_{g_K}, S]\Phi\right| \leq C\sum_{k=1}^{2}|D^k\Phi|.$$

4.3. ***Commutator*** $\widetilde{\Omega}_i$**.** Let Ω_i be a basis of vector fields of rotations in Schwarzschild spacetimes. An explicit realization can be

$$\Omega = \left\{\partial_\phi, \sin\phi\partial_\theta + \frac{\cos\phi\cos\theta}{\sin\theta}\partial_\phi, \cos\phi\partial_\theta - \frac{\sin\phi\cos\theta}{\sin\theta}\partial_\phi\right\}.$$

Define $\widetilde{\Omega}_i = \chi(r)\Omega_i$ to be cutoff so that it is supported in $\{r > R_\Omega\}$ and equals Ω_i for $r > R_\Omega + 1$ for some large R_Ω. On Schwarzschild spacetimes, Ω_i is Killing and therefore $\widetilde{\Omega}_i$ is Killing for $r > R_\Omega + 1$. Therefore,

$$[\Box_{g_S}, \widetilde{\Omega}_i] = \tilde{\chi}(r)(\partial^2 + \partial),$$

where $\tilde{\chi}$ is some function that depends only on r and is supported in $\{R_\Omega < r < R_\Omega + 1\}$.

Using Proposition 2, we have

$$|[\Box_{g_K}, \widetilde{\Omega}_i]\Phi| \leq Cr^{-2}(|\partial^2\Phi| + |\partial\Phi|).$$

Moreover, since $\widetilde{\Omega}_i$ vanishes for $r < R_\Omega$, we have trivially

$$[\Box_{g_K}, \widetilde{\Omega}_i]\Phi = 0 \quad \text{for } r < R_\Omega.$$

From now on, we write $\widetilde{\Omega}$ to denote any one of the Ω_i, while taking the norm to be $|\widetilde{\Omega}\Phi| = \sum_i |\Omega_i\Phi|$. This commutator is useful for gaining powers of r near spatial infinity. In particular we have

$$|\not\nabla\Phi| \leq Cr^{-1}|\widetilde{\Omega}\Phi|.$$

This extra power of r is essential for controlling the error terms arising from the commutation of $\Box_{g_K}$ with S.

4.4. ***Commutator $\hat{Y}$.*** Let $\hat{Y}$, as in Section 2.3, be a vector field that is null near the event horizon, is normalized with respect to another null vector $\hat{V}$ and is cut off to be compactly supported in $\{r \leq r_Y^+\}$.

Proposition 3. *On Kerr spacetimes such that ϵ is small, we have*

$$\left|[\Box_{g_K}, \hat{Y}]\Phi - \kappa \hat{Y}^2\Phi\right| \leq C\big(|D\partial_{t^*}\Phi| + \epsilon|D^2\Phi| + |D\Phi|\big) \quad \textit{for } r \leq r_Y^-,$$

where $\kappa > c > 0$ is as in (5).

Proof. The principal term for the commutator $[\Box_{g_K}, \hat{Y}]\Phi$ is $2^{(\hat{Y})}\pi^{\mu\nu} D_\mu D_\nu \Phi$, where ${}^{(\hat{Y})}\pi_{\mu\nu}$ is the deformation tensor defined by ${}^{(\hat{Y})}\pi_{\mu\nu} = \frac{1}{2}(D_\mu \hat{Y}_\nu + D_\nu \hat{Y}_\mu)$. We look at three terms that are useful in deriving the estimates.

$$\begin{aligned}
{}^{(\hat{Y})}\pi_{\hat{V}\hat{V}} &= g(D_{\hat{V}}\hat{Y}, \hat{V}) = -g(\hat{Y}, D_{\hat{V}}\hat{V}) = 2\kappa,\\
\left|{}^{(\hat{Y})}\pi_{\hat{V}E_A}\right| &= \left|\tfrac{1}{2}\big(g(D_{\hat{V}}\hat{Y}, E_A) + g(D_{E_A}\hat{Y}, \hat{V})\big)\right| \leq C\epsilon,\\
\left|{}^{(\hat{Y})}\pi_{E_A E_B}\right| &= \left|\tfrac{1}{2}\big(g(D_{E_B}\hat{Y}, E_A) + g(D_{E_A}\hat{Y}, E_B)\big)\right| \leq C\epsilon,
\end{aligned}$$

where the smallness in the second and third line come from the assumption that we are close to Schwarzschild. Notice also that for $r \leq r_Y^-$, $\hat{V}$ is C^0 close to ∂_{t^*}. Therefore, in the commutator, the main term is

$$\kappa \hat{Y}^2 \Phi.$$

All the other second order terms either have a ∂_{t^*} derivative or small. $\square$

5. The basic identities for currents

5.1. ***Vector field multipliers.*** We consider the conservation laws for Φ satisfying $\Box_g \Phi = 0$. Define the energy-momentum tensor

$$T_{\mu\nu} = \partial_\mu \Phi \partial_\nu \Phi - \tfrac{1}{2} g_{\mu\nu} \partial^\alpha \Phi \partial_\alpha \Phi.$$

We note that $T_{\mu\nu}$ is symmetric and the wave equation implies that $D^\mu T_{\mu\nu} = 0$. Given a vector field V^μ, we define the associated currents

$$J_\mu^V(\Phi) = V^\nu T_{\mu\nu}(\Phi) \quad \text{and} \quad K^V(\Phi) = {}^{(V)}\pi_{\mu\nu} T^{\mu\nu}(\Phi),$$

where ${}^{(V)}\pi_{\mu\nu}$ is the deformation tensor defined by

$${}^{(V)}\pi_{\mu\nu} = \tfrac{1}{2}(D_\mu V_\nu + D_\nu V_\mu).$$

In particular, $K^V(\Phi) = {}^{(V)}\pi_{\mu\nu} = 0$ if V is Killing. Since the energy-momentum tensor is divergence free,

$$D^\mu J_\mu^V(\Phi) = K^V(\Phi).$$

We also define the modified currents

$$\begin{aligned}
J_\mu^{V,w}(\Phi) &= J_\mu^V(\Phi) + \tfrac{1}{8}(w\partial_\mu \Phi^2 - \partial_\mu w \Phi^2),\\
K^{V,w}(\Phi) &= K^V(\Phi) + \tfrac{1}{4} w \partial^\nu \Phi \partial_\nu \Phi - \tfrac{1}{8}\Box_g w \Phi^2.
\end{aligned}$$

Then

$$D^\mu J^{V,w}_\mu(\Phi) = K^{V,w}(\Phi)\,.$$

We integrate by parts with this in the region bounded by Σ_τ, $\Sigma_{\tau'}$ and $\mathscr{H}^+(\tau',\tau)$. We denote this region as $\mathscr{R}(\tau',\tau)$. We denote the future-directed normal to Σ_τ by $n^\mu_{\Sigma_\tau}$.

Proposition 4. *We have*

$$\int_{\Sigma_\tau} J^V_\mu(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathscr{H}(\tau',\tau)} J^V_\mu(\Phi) n^\mu_{\mathscr{H}^+(\tau',\tau)} + \iint_{\mathscr{R}(\tau',\tau)} K^V(\Phi) = \int_{\Sigma_{\tau'}} J^V_\mu(\Phi) n^\mu_{\Sigma_{\tau'}},$$

$$\int_{\Sigma_\tau} J^{V,w^V}_\mu(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathscr{H}^+(\tau',\tau)} J^{V,w^V}_\mu(\Phi) n^\mu_{\mathscr{H}^+(\tau',\tau)} + \iint_{\mathscr{R}(\tau',\tau)} K^{V,w^V}(\Phi) = \int_{\Sigma_{\tau'}} J^{V,w^V}_\mu(\Phi) n^\mu_{\Sigma_{\tau'}}.$$

One can similarly define the quantities above for the inhomogeneous wave equation $\Box_g \Phi = F$. In this case, the energy-momentum is no longer divergence free. Instead, we have

$$D^\mu T_{\mu\nu} = F\partial_\nu\Phi.$$

In this case,

$$D^\mu J^V_\mu(\Phi) = K^V(\Phi) + FV^\nu\partial_\nu\Phi.$$

For the modified current,

$$D^\mu J^{V,w}_\mu(\Phi) = K^{V,w}(\Phi) + \tfrac{1}{4}Fw\Phi + FV^\nu\partial_\nu\Phi.$$

Proposition 5. *We have*

$$\int_{\Sigma_\tau} J^V_\mu(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathscr{H}(\tau',\tau)} J^V_\mu(\Phi) n^\mu_{\mathscr{H}^+(\tau',\tau)} + \iint_{\mathscr{R}(\tau',\tau)} K^V(\Phi) = \int_{\Sigma_{\tau'}} J^V_\mu(\Phi) n^\mu_{\Sigma_{\tau'}} - \iint_{\mathscr{R}(\tau',\tau)} FV^\nu\partial_\nu\Phi,$$

$$\begin{aligned}\int_{\Sigma_\tau} J^{V,w^V}_\mu(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathscr{H}(\tau',\tau)} J^{V,w^V}_\mu(\Phi) n^\mu_{\mathscr{H}^+(\tau',\tau)} + \iint_{\mathscr{R}(\tau',\tau)} K^{V,w^V}(\Phi)&\\ = \int_{\Sigma_{\tau'}} J^{V,w^V}_\mu(\Phi) n^\mu_{\Sigma_{\tau'}} + \iint_{\mathscr{R}(\tau',\tau)} \big(-\tfrac{1}{4}Fw\Phi - FV^\nu\partial_\nu\Phi\big).&\end{aligned}$$

5.2. ***Vector field multipliers under metric perturbations.*** If we consider Kerr spacetimes such that ϵ is small, vector fields multipliers are stable if defined in the Schwarzschild coordinates (t^*, r, x^A, x^B) or $(t, r \geq r_Y^-, x^A, x^B)$. We can consider a fixed vector field defined on the differentiable structure of a Schwarzschild exterior and compare the currents obtained using the Schwarzschild metric and the Kerr metric.

Proposition 6. *Consider either the Schwarzschild $(t_S^*, r_S, x_S^A, x_S^B)$ coordinates or $(t_S, r_S \geq r_Y^-, x_S^A, x_S^B)$ coordinates. Suppose V is a vector field defined on either of these coordinates. Then*

$$\big|(J_S^{V,w^V})_\mu(\Phi) n^\mu_{\Sigma_\tau} - (J_K^{V,w^V})_\mu(\Phi) n^\mu_{\Sigma_\tau}\big| \leq C\epsilon r^{-2}\max_\alpha |V^\alpha|(\partial\Phi)^2$$

and

$$\left|K_S^{V,w^V}(\Phi)-K_K^{V,w^V}(\Phi)\right|\le C\epsilon r^{-2}\Big(\Big(\sum_{k=0,1}\max_{\alpha}|\partial^k V^\alpha|+|w|\Big)(\partial\Phi)^2+\sum_{m=1,2}|\partial^m w|\Phi^2\Big).$$

6. Statement of the main theorem

With the currents defined, we can state our main theorem.

Main Theorem. *Suppose $\Box_{g_K}\Phi=0$. Then for all $\eta>0$, $R>r_+$ and all $M>0$ there exists a_0 such that the following estimates hold in the region $\{r_+\le r\le R\}$ on Kerr spacetimes with (M,a) for which $a\le a_0$.*

(1) *Improved decay of nondegenerate energy*:

$$\sum_{j=0}^{\ell}\int_{\Sigma_\tau\cap\{r\le R\}}(D^j\Phi)^2$$
$$\le C_R\tau^{-3+\eta}\Big(\sum_{m=0}^{\ell+2}\int_{\Sigma_{\tau_0}}J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m S\Phi)n^\mu_{\Sigma_{\tau_0}}+\sum_{m+k+j\le\ell+5}\int_{\Sigma_{\tau_0}}J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m\hat{Y}^k\tilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}\Big).$$

(2) *Improved pointwise decay*:

$$\sum_{j=0}^{\ell}|D^j\Phi|$$
$$\le C_R\tau^{-3/2+\eta}\Big(\sum_{m=0}^{\ell+4}\int_{\Sigma_{\tau_0}}J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m S\Phi)n^\mu_{\Sigma_{\tau_0}}+\sum_{m+k+j\le\ell+7}\int_{\Sigma_{\tau_0}}J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m\hat{Y}^k\tilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}\Big)^{1/2}.$$

Here the vector field N will be defined in Section 7, and the vector field Z with the modifying function w^Z will be defined in Section 9.

Remark. We will show that although $J_\mu^{Z,w^Z}(\Phi)\,n^\mu_{\Sigma_{t^*}}$ is not always nonnegative, $J_\mu^{Z+CN,w^Z}(\Phi)\,n^\mu_{\Sigma_{t^*}}$ is nonnegative for sufficiently large C. Hence all the energy quantities in the theorem are nonnegative.

Remark. Since we have the improved decay of the nondegenerate energy, the theorem above can be extended beyond the event horizon. More precisely, for any $r_b\in(r_-,r_+)$, where r_- is the smaller root of $\Delta=r^2-2Mr-a^2$, the theorem holds up to $r\ge r_b$ for D understood as a regular derivative inside the black hole, and with the constant depending also on r_b. The proof is similar to that in [Luk 2010].

7. Vector field multiplier N_e and mild growth of nondegenerate energy

Kerr spacetime has a Killing vector field ∂_t. The conservation law gives

$$\int_{\Sigma_\tau}J_\mu^T(\Phi)n^\mu_{\Sigma_\tau}+\int_{\mathcal{H}(\tau_0,\tau)}J_\mu^T(\Phi)n^\mu_{\Sigma_\tau}=\int_{\Sigma_{\tau_0}}J_\mu^T(\Phi)n^\mu_{\Sigma_{\tau_0}}+\iint_{\mathcal{R}(\tau_0,\tau)}\partial_t\Phi G.$$

We add to the Killing vector field ∂_t a red-shift vector field. Here, we use the "nonregular" red-shift vector field as in [Dafermos and Rodnianski 2008]. Under this construction, N_e is C^0 but not C^1 at the event horizon $\mathcal{H}^+$. Compared to the smooth construction in [ibid.], this construction will provide extra control for some derivatives near $\mathcal{H}^+$.

Define

$$Y = y_1(r)\hat{Y} + y_2(r)\hat{V},$$

where

$$y_1(r) = 1 - \frac{1}{(\log(r-r_+))^3} \quad \text{and} \quad y_2(r) = -\frac{1}{(\log(r-r_+))^3}.$$

By this definition Y is compactly supported in $\{r \le r_Y^+\}$ and is invariant under the isomorphisms generated by ∂_{t^*} and ∂_{ϕ^*}.

Proposition 7. *Let $N_e = \partial_{t^*} + eY$. For any e, there is a corresponding choice of $\epsilon \ll e$ and r_Y^- such that for every integer p, there exists $c_p > 0$ such that*

$$\begin{aligned}
&J_\mu^{N_e}(\Phi) n^\mu_{\mathcal{H}^+} \sim (D_{\hat{V}}\Phi)^2 + e \sum_{E_A \in \{E_1, E_2\}} (D_{E_A}\Phi)^2 && \textit{on the event horizon},\\
&J_\mu^{N_e}(\Phi) n^\mu_{\Sigma_\tau} \sim \sum_{E_\alpha \in \{E_1, E_2, \hat{V}\}} (D_{E_\alpha}\Phi)^2 + e(D_{E_{\hat{Y}}}\Phi)^2 && \textit{for } r \le r_Y^-,\\
&J_\mu^{N_e}(\Phi) n^\mu_{\Sigma_\tau} \sim \sum (\partial\Phi)^2 && \textit{for } r \ge r_Y^- \textit{ in the } (t^*, r, x^A, x^B) \textit{ coordinates},\\
&K^{N_e}(\Phi) \ge c_p e\Big(|\log(r-r_+)|^p\Big((D_{\hat{V}}\Phi)^2 + \sum_A (D_{E_A}\Phi)^2\Big) + (D_{\hat{Y}}\Phi)^2\Big) && \textit{for } r \le r_Y^-,\\
&K^{N_e}(\Phi) \le Ce J_\mu^{N_e}(\Phi) n^\mu_{\Sigma_\tau} && \textit{for } r_Y^- \le r \le r_Y^+.
\end{aligned}$$

Proof. It is obvious that Y is timelike and future-oriented for $r \le r_Y^-$. Since ∂_{t^*} is casual in the exterior region of Schwarzschild spacetime and is null only on the event horizon, for every small $e > 0$, there exists sufficiently small $\epsilon > 0$ such that N_e is timelike and future-directed on Kerr spacetimes up to the event horizon. The first two estimates hold since in Kerr spacetime, ∂_{t^*} is ϵ-close to $\hat{V}$ on the event horizon. The third estimate holds because outside a small (depending on ϵ) neighborhood of the event horizon, ∂_{t^*} is timelike.

To show that $K^{N_e}(\Phi)$ has the required positivity near the event horizon, we compute the deformation tensor. First, notice that

$$D_{\hat{Y}} y_1 = D_{\hat{Y}} y_2 = \frac{3 D_{\hat{Y}} r}{(r-r_+)(\log(r-r_+))^4}.$$

Using this we have

$$\begin{aligned}
{}^{(Y)}\pi_{\hat{V}\hat{V}} &= g_K(D_{\hat{V}}(y_1\hat{Y} + y_2\hat{V}), \hat{V}) = -g_K(y_1\hat{Y}+y_2\hat{V}, D_{\hat{V}}\hat{V}) = 2y_1\kappa + b^Y y_2,\\
{}^{(Y)}\pi_{\hat{Y}\hat{Y}} &= g_K(D_{\hat{Y}}(y_1\hat{Y}+y_2\hat{V}), \hat{Y}) = -\frac{6 D_{\hat{Y}} r}{(r-r_+)(\log(r-r_+))^4},\\
{}^{(Y)}\pi_{\hat{V}\hat{Y}} &= \tfrac{1}{2} g_K(D_{\hat{V}}(y_1\hat{Y}+y_2\hat{V}), \hat{Y}) + \tfrac{1}{2} g_K(D_{\hat{Y}}(y_1\hat{Y}+y_2\hat{V}), \hat{V}) = -\frac{3 D_{\hat{Y}} r}{(r-r_+)(\log(r-r_+))^4} + y_1\kappa + y_2 b^Y,
\end{aligned}$$

Moreover, we have

$$^{(Y)}\pi_{\hat{V}E_A},\ ^{(Y)}\pi_{\hat{Y}E_A},\ ^{(Y)}\pi_{E_AE_B} = O(1).$$

Notice that

$$T_{\hat{Y}\hat{Y}} \sim (D_{\hat{Y}}\Phi)^2, \quad T_{\hat{V}\hat{V}} \sim (D_{\hat{V}}\Phi)^2, \quad T_{\hat{Y}\hat{V}} \sim |\nabla\!\!\!/\Phi|^2,$$

and that $^{(Y)}\pi_{\hat{V}E_A}$, $^{(Y)}\pi_{\hat{Y}E_A}$ and $^{(Y)}\pi_{E_AE_B}$ have no terms of the form $(D_{\hat{Y}}\Phi)^2$. Hence we can choose r_Y^- sufficiently close to r_+ so that for $r_+ \le r \le r_Y^-$,

$$K^Y(\Phi) \ge c\kappa(D_{\hat{Y}}\Phi)^2 + \frac{c}{(r-r_+)|\log(r-r_+)|^4}\Big((D_{\hat{V}}\Phi)^2 + \sum_A (D_{E_A}\Phi)^2\Big).$$

Since ∂_{t^*} is Killing, and $K^{N_e}(\Phi) = eK^Y(\Phi)$, we have

$$K^{N_e}(\Phi) \ge ce\bigg(\kappa(D_{\hat{Y}}\Phi)^2 + \frac{1}{(r-r_+)|\log(r-r_+)|^4}((D_{\hat{V}}\Phi)^2 + \sum_A (D_{E_A}\Phi)^2)\bigg) \quad \text{for } r \le r_Y^-,$$

Finally, since $J_\mu^{\partial_{t^*}} n^\mu_{\Sigma_{t^*}}$ controls all derivatives in the region $r_Y^- \le r \le r_Y^+$, we have

$$K^{N_e}(\Phi) \le CeJ_\mu^{N_e}(\Phi)\, n^\mu_{\Sigma_\tau} \quad \text{for } r_Y^- \le r \le r_Y^+. \qquad \square$$

Definition 8. We call the positive quantity $\int_{\Sigma_\tau} J_\mu^{N_e}(\Phi)\, n^\mu_{\Sigma_\tau}$ the nondegenerate energy.

The following identity determines how the nondegenerate energy changes with τ.

Proposition 9. *Let Φ satisfy $\Box_{g_K}\Phi = G$. Then*

$$\int_{\Sigma_\tau} J_\mu^{N_e}(\Phi)n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau_0,\tau)} J_\mu^{N_e}(\Phi)n^\mu_{\Sigma_\tau} + \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\le r_Y^-\}} K_\mu^{N_e}(\Phi)$$
$$= \int_{\Sigma_{\tau_0}} J_\mu^{N_e}(\Phi)n^\mu_{\Sigma_{\tau_0}} + e\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r_Y^-\le r\le r_Y^+\}} K^Y(\Phi) + \int_{\mathcal{R}(\tau_0,\tau)} (\partial_{t^*}\Phi + eY\Phi)G.$$

The estimates given by the vector field N are sufficient to show that, modulo inhomogeneous terms, the quantity $\int_{\Sigma_\tau} J_\mu^{N_e}(\Phi)\, n^\mu_{\Sigma_\tau}$ cannot grow too much in a short time interval:

Proposition 10. *Let Φ satisfy $\Box_{g_K}\Phi = G$. For e sufficiently small, $\epsilon \ll e$ and $0 \le \tau - \tau' \le 1$, we have*

$$\int_{\Sigma_\tau} J_\mu^{N_e}(\Phi)n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J_\mu^{N_e}(\Phi)n^\mu_{\mathcal{H}^+} \le 4\int_{\Sigma_{\tau'}} J_\mu^{N_e}(\Phi)n^\mu_{\Sigma_{\tau'}} + C\iint_{\mathcal{R}(\tau',\tau)} G^2.$$

Proof. We first note that

$$\iint_{\mathcal{R}(\tau',\tau)\cap\{r_Y^-\le r\le r_Y^+\}} K^Y(\Phi) \le C\int_{\tau'}^{\tau}\int_{\Sigma_{\bar\tau}} J_\mu^{N_e}(\Phi)\, n^\mu_{\Sigma_{\bar\tau}} d\bar\tau,$$

with C independent of e and ϵ whenever $\epsilon \ll e < 1$. Then, by Proposition 9,

$$\begin{aligned}
&\int_{\Sigma_\tau} J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J^{N_e}_\mu(\Phi) n^\mu_{\mathcal{H}^+} + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r^-_Y\}} K^{N_e}_\mu(\Phi) \\
&= \int_{\Sigma_{\tau'}} J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_{\tau'}} + e \iint_{\mathcal{R}(\tau',\tau)\cap\{r^-_Y\le r\le r^+_Y\}} K^Y(\Phi) + \int_{\mathcal{R}(\tau',\tau)} (\partial_{t^*}\Phi + eY\Phi) G \\
&\le \int_{\Sigma_{\tau'}} J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_{\tau'}} + Ce \int_{\tau'}^{\tau} \int_{\Sigma_{\bar\tau}} J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_{\bar\tau}} d\bar\tau + \delta' \iint_{\mathcal{R}(\tau',\tau)} ((\partial_{t^*}\Phi + eY\Phi))^2 + (\delta')^{-1} \iint_{\mathcal{R}(\tau',\tau)} G^2 \\
&\le \int_{\Sigma_{\tau'}} J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_{\tau'}} + (C\delta' + 2Ce) \int_{\tau'}^{\tau} \int_{\Sigma_{\bar\tau}} J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_{\bar\tau}} d\bar\tau + (\delta')^{-1} \iint_{\mathcal{R}(\tau',\tau)} G^2.
\end{aligned}$$

By Gronwall's inequality and absorbing $(\delta')^{-1}$ into the constant C, we have

$$\int_{\Sigma_\tau} J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_\tau} \le 2 \int_{\Sigma_{\tau'}} J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_{\tau'}} + C \iint_{\mathcal{R}(\tau',\tau)} G^2.$$

Now the estimate for the term horizon follows from Proposition 9. □

8. Integrated decay estimates and boundedness of nondegenerate energy

In this section we would like to show an integrated decay estimate. We first follow [Luk 2010] to construct a vector field and prove an integrated decay estimate for the terms near spatial infinity. That construction is in turn inspired by [Sterbenz 2005]. In [Luk 2010], the decay rate in r of this integrated decay estimate is crucial for controlling the error terms arising from the vector field commutator S. In the sequel, such an estimate will also facilitate many computations as we prove the full integrated decay estimate.

In view of the red shift, all derivatives of Φ can be controlled near the event horizon. However, we would also like to prove an integrated decay estimate that controls Φ itself near the event horizon. This is in contrast to the integrated decay estimate in [Dafermos and Rodnianski 2008], which degenerates near the event horizon. This extra control is useful as we are considering the inhomogeneous problem.

The proof of the integrated decay estimate for a finite region of r away from the horizon follows that in [Dafermos and Rodnianski 2008]. The one difference here is that we do not assume the boundedness of $\int_{\Sigma_\tau} J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_\tau}$ (even after ignoring inhomogeneous terms). We would instead like to prove the boundedness of $\int_{\Sigma_\tau} J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_\tau}$ using the integrated decay estimates. We will, however, use Proposition 10.

The reader should think of this integrated decay estimates as analogous to the estimates associated to the vector field X in [Dafermos and Rodnianski 2009; 2011; Luk 2010]. However, it is impossible to obtain such estimates using a vector field in Kerr spacetimes and we therefore resort to a phase space analysis; see [Alinhac 2009].

To perform the phase space analysis, we will take the Fourier transform in the variable t^*, take the Fourier series in the variable ϕ^* and express the dependence on the θ variable in oblate spheroidal harmonics. Carter [1968] discovered that with this decomposition, the wave equation can be separated.

However, to take the Fourier transform in the variable t^*, we need Φ to be at least in L^2. To this end, we perform a cutoff in the variable t^*.

8.1. ***Estimates near spatial infinity.*** In this subsection, we follow [Luk 2010] to construct a vector field $\tilde{X} = \tilde{f}(r^*)\partial_{r^*}$ such that the spacetime integral that can be controlled with a good weight in r.

Proposition 11 [Luk 2010, Proposition 8]. *In Schwarzschild spacetimes, using (t, r^*, x^A, x^B) coordinates, there exists $\tilde{X}_S = \tilde{f}(r^*)\partial_{r^*}$ and $w_S^{\tilde{X}}$ supported in $r \geq \frac{13}{4}M$, such that*

$$K^{\tilde{X},w^{\tilde{X}}}(\Phi) \geq c(r^{-1-\delta}(\partial_{r^*}\Phi)^2 + r^{-1}|\nabla\!\!\!/\Phi|^2 + r^{-3-\delta}\Phi^2) \quad \textit{for } r^* \geq \max\{100, 100M\}$$

and

$$\left|\int_{\Sigma_\tau} J_\mu^{\tilde{X},w^{\tilde{X}}}(\Phi)n^\mu_{\Sigma_\tau}\right| \leq C\int_{\Sigma_\tau} J_\mu^{N_e}(\Phi)n^\mu_{\Sigma_\tau}.$$

This implies via stability (since the vector field is supported away from the event horizon) the following:

Proposition 12. *In Kerr spacetimes, using (t^*, r, x^A, x^B) coordinates, there exists $\tilde{X}$ and $w^{\tilde{X}}$ supported in $r \geq \frac{25}{8}M$ such that for some large R,*

$$K^{\tilde{X},w^{\tilde{X}}}(\Phi) \geq c_{\tilde{X}}\big(r^{-1-\delta}(\partial_{r^*}\Phi)^2 + r^{-1}|\nabla\!\!\!/\Phi|^2 + r^{-3-\delta}\Phi^2\big) - C_{\tilde{X}}\epsilon r^{-2}(\partial_{t^*}\Phi)^2 \quad \textit{for } r^* \geq R$$

and

$$\left|\int_{\Sigma_\tau} J_\mu^{\tilde{X},w^{\tilde{X}}}(\Phi)n^\mu_{\Sigma_\tau}\right| \leq C\int_{\Sigma_\tau} J_\mu^{N_e}(\Phi)n^\mu_{\Sigma_\tau}.$$

Now it is easy to construct the following vector field on Schwarzschild spacetimes:

Proposition 13. *In Schwarzschild spacetimes, in (t, r^*, x^A, x^B) coordinates, there exists $\check{X}_S = \check{f}(r^*)\,\partial_{r^*}$ supported in $r \geq \frac{13}{4}M$ such that*

$$K^{\check{X}}(\Phi) \geq cr^{-1-\delta}(\partial_{t^*}\Phi)^2 - C(r^{-1-\delta}(\partial_{r^*}\Phi)^2 + r^{-1}|\nabla\!\!\!/\Phi|^2 + r^{-3-\delta}\Phi^2) \quad \textit{for } r^* \geq \max\{100, 100M\}$$

and

$$\left|\int_{\Sigma_\tau} J_\mu^{\check{X}}(\Phi)n^\mu_{\Sigma_\tau}\right| \leq C\int_{\Sigma_\tau} J_\mu^{N_e}(\Phi)n^\mu_{\Sigma_\tau}.$$

Proof. Let $\check{f}$ be supported appropriately and let $\check{f}(r^*) = 1/(1+r^*)^\delta$ whenever r^* is large. □

As before, a stability argument gives this:

Proposition 14. *In Kerr spacetimes, in (t^*, r, x^A, x^B) coordinates, there exists $\check{X}$ supported in $r \geq \frac{25}{8}M$ such that for some large R,*

$$K^{\check{X}}(\Phi) \geq cr^{-1-\delta}(\partial_{t^*}\Phi)^2 - C_{\check{X}}(r^{-1-\delta}(\partial_{r^*}\Phi)^2 + r^{-1}|\nabla\!\!\!/\Phi|^2 + r^{-3-\delta}\Phi^2) \quad \textit{for } r^* \geq R$$

and

$$\left|\int_{\Sigma_\tau} J_\mu^{\check{X}}(\Phi)n^\mu_{\Sigma_\tau}\right| \leq C\int_{\Sigma_\tau} J_\mu^{N_e}(\Phi)n^\mu_{\Sigma_\tau}.$$

Now using the vector field $\tilde{X} + \frac{1}{2}(c_{\tilde{X}}/C_{\check{X}})\check{X}$ and modifying function $w^{\tilde{X}}$, we get the following estimate:

Proposition 15. *For ϵ sufficiently small,*

$$\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\geq R\}} \big(r^{-1-\delta} J^{N_e}_{\mu}(\Phi) n^{\mu}_{\Sigma_{t^*}} + r^{-3-\delta}\Phi^2\big)$$
$$\leq C\bigg(\int_{\Sigma_\tau} J^{N_e}_{\mu}(\Phi) n^{\mu}_{\Sigma_\tau} + \int_{\Sigma_{\tau_0}} J^{N}_{\mu}(\Phi) n^{\mu}_{\Sigma_{\tau_0}} + \iint_{\mathcal{R}(\tau_0,\tau)\cap\{\frac{25}{8}M\leq r\leq R\}} (J^{N_e}_{\mu}(\Phi)\, n^{\mu}_{\Sigma_{t^*}} + \Phi^2) + \iint_{\mathcal{R}(\tau_0,\tau)} (|\partial_r\Phi| + r^{-1}|\Phi|)|G|\bigg).$$

8.2. ***Estimates near the event horizon.*** The integrated decay estimates shown in [Dafermos and Rodnianski 2008] are degenerate around the event horizon. Here we will prove the corresponding estimates near the event horizon. In view of the availability of the red-shift estimate K^{N_e}, we will focus on the zeroth order term Φ^2. It turns out that we can use a construction in [Luk 2010].

Proposition 16. *In Schwarzschild spacetimes, in (t, r^*, x^A, x^B) coordinates, there exists $X_h = f_h(r^*)\partial_{r^*}$ and w^{X_h} supported in $r \leq \frac{23}{8}M$ such that*

$$K^{X_h, w^{X_h}}(\Phi) \geq c((\partial_{r^*}\Phi)^2 + |\nabla\!\!\!/\Phi|^2 + \Phi^2) \quad \textit{for } r \leq r_Y^-$$

and

$$\left|\int_{\Sigma_\tau} J^{X_h, w^{X_h}}_{\mu}(\Phi)\, n^{\mu}_{\Sigma_\tau}\right| \leq C \int_{\Sigma_\tau} J^{N_e}_{\mu}(\Phi)\, n^{\mu}_{\Sigma_\tau} \quad \textit{and} \quad |J^{X_h, w^{X_h}}_{\mu}(\Phi)\, n^{\mu}_{\mathcal{H}^+}| \leq C J^{N_e}_{\mu}(\Phi)\, n^{\mu}_{\mathcal{H}^+}.$$

Proof. Let

$$X_h = f_h(r_S^*)\partial_{r_S^*} = -\chi(r)\frac{M^3}{(1+4\mu^{-2})}\partial_{r_S^*} = -\chi(r)\frac{\mu^3 r^3}{8(1+4\mu^{-2})}\partial_{r_S^*},$$

where $\chi(r)$ is a cutoff function that is compactly supported in $r \leq \frac{23}{8}M$ and is identically 1 for $r \leq r_Y^-$. Also, let

$$w^{X_h} = 2f_h'(r^*) + \frac{4(1-\mu)}{r} f_h(r^*).$$

From now on, we will focus on the behavior when $r \leq r_Y^-$ and treat the terms in $\{r_Y^- \leq r \leq \frac{23}{8}M\}$ as errors. Recall that on Schwarzschild spacetime,

$$K^{X_h, w^{X_h}}(\Phi) = \frac{f'(r^*)}{1-\mu}(\partial_{r^*}\Phi)^2 + \frac{(2-3\mu)f(r^*)}{2r}|\nabla\!\!\!/\Phi|^2$$
$$- \frac{1}{4}\bigg(\frac{1}{1-\mu} f'''(r^*) + \frac{4}{r} f''(r^*) + \frac{\mu}{r^2} f'(r^*) - \frac{2\mu}{r^3}(3-4\mu) f(r^*)\bigg)\Phi^2.$$

We now look at the sign of this expression for $r \leq r_Y^-$. It is easy to see that the coefficient for $(\partial_{r^*}\Phi)^2$ is positive:

$$f'(r^*) = (1-\mu)\partial_r f_0(r) = \frac{\mu r^2(1-\mu)}{(1+4\mu^{-2})^2} \geq \frac{c(1-\mu)}{r^3},$$

The coefficient of $|\nabla\!\!\!/ \Phi|^2$ is also clearly positive. A computation shows that

$$\frac{1}{1-\mu}f''' + \frac{4}{r}f'' + \frac{\mu}{r^2}f' - \frac{2\mu}{r^3}(3-4\mu)f = -\frac{\mu^6(192+\mu(128+\mu(-784+\mu(464+\mu(-28+\mu(52+\mu(-3+4\mu)))))))}{4(4+\mu^2)^4}$$

We want to show that $P(\mu) = 192+\mu(128+\mu(-784+\mu(464+\mu(-28+\mu(52+\mu(-3+4\mu)))))) \geq 1/7$ for $16/23 \leq \mu \leq 1$.

First, $192+128\mu-784\mu^2+464\mu^3 = 16(-12-20\mu+29\mu^2)(\mu-1) \geq 0$.

Now $52-3\mu+4\mu^2$ reaches its minimum at $3/8$. Hence, $52-3\mu+4\mu^2 \geq 823/16$.

Finally $-28+\mu(52-3\mu+4\mu^2) \geq -28+11/20\cdot 823/16 \geq 93/320$.

Therefore, $P(\mu) \geq 1023/6400 \geq 1/7$ for $16/23 \leq \mu \leq 1$. Therefore, for $r \leq r_Y^-$,

$$K^{X_h,w^{X_h}}(\Phi) \geq c((\partial_{r^*}\Phi)^2 + |\nabla\!\!\!/ \Phi|^2 + \Phi^2).$$

The second and third statements,

$$\left|\int_{\Sigma_\tau} J_\mu^{X_h,w^{X_h}}(\Phi)n^\mu_{\Sigma_\tau}\right| \leq C\int_{\Sigma_\tau} J_\mu^{N_e}(\Phi)n^\mu_{\Sigma_\tau} \quad \text{and} \quad |J_\mu^{X_h,w^{X_h}}(\Phi)n^\mu_{\mathcal{H}^+}| \leq C J_\mu^{N_e}(\Phi)n^\mu_{\mathcal{H}^+},$$

follow from the boundedness of f_h and w^{X_h} and that on the Schwarzschild horizon $\partial_t = \partial_{r^*}$. Hence in both estimates, the constants are independent of e for e small. □

Because X_h and w^{X_h} are actually smooth up to the event horizon, we have this via a stability argument:

Proposition 17. *In Kerr spacetimes, using (t_S, r_S, x_S^1, x_S^2) coordinates, there exists X_h and w^{X_h} supported in $r \leq \frac{23}{8}M$ such that*

$$K^{X_h,w^{X_h}}(\Phi) \geq c\Phi^2 - C\epsilon(\partial_{t^*}\Phi)^2 - C\epsilon(\partial_r\Phi)^2 \quad \textit{for } r \leq r_Y^-$$

and

$$\left|\int_{\Sigma_\tau} J_\mu^{X_h,w^{X_h}}(\Phi)n^\mu_{\Sigma_\tau}\right| \leq C\int_{\Sigma_\tau} J_\mu^{N}(\Phi)n^\mu_{\Sigma_\tau} \quad \textit{and} \quad \left|J_\mu^{X_h,w^{X_h}}(\Phi)n^\mu_{\mathcal{H}^+}\right| \leq C J_\mu^{N_e}(\Phi)n^\mu_{\mathcal{H}^+}.$$

Together with the red shift, we then have an integrated decay estimate near the event horizon:

Proposition 18.

$$\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq r_Y^-\}} (\Phi^2 + K^{N_e}(\Phi)) \leq C\Bigg(\int_{\Sigma_\tau} J_\mu^{N_e}(\Phi)n^\mu_{\Sigma_\tau} + \int_{\Sigma_{\tau_0}} J_\mu^{N_e}(\Phi)n^\mu_{\Sigma_{\tau_0}} + \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r_Y^-\leq r\leq \frac{23}{8}M\}} (\Phi^2 + J_\mu^{N_e}(\Phi)n^\mu_{\Sigma_{t^*}}) \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq\frac{23}{8}M\}} (|\partial_{r^*}\Phi| + r^{-1}|\Phi|)|G| + \left|\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq\frac{23}{8}M\}} (\partial_{t^*}\Phi + eY\Phi)G\right|\Bigg).$$

8.3. ***Cutoff, decomposition and separation.*** Following [Dafermos and Rodnianski 2008], we define the cutoff $\Phi^{\tau}_{\tau'} = \xi\Phi$, where $\xi = \chi(t^* - 1 - \tau)\chi(-t^* - 1 + \tau')$, for some smooth cutoff function $\chi(x)$ that is identically 1 for $x \le -1$ and has support on $\{x \le 0\}$. Then

$$\Box_g \Phi^{\tau}_{\tau'} = \xi G + 2D^{\alpha}\Phi D_{\alpha}\xi + \Phi\Box_g\xi =: F.$$

We then decompose in frequency. We decompose the Fourier transform in t of Φ into Fourier series in ϕ and oblate spheroidal harmonics:

$$\hat{\Phi}^{\tau}_{\tau'} = \sum_{m,\ell} R^{\omega}_{m\ell}(r) S_{m\ell}(a\omega, \cos\theta) e^{im\phi}.$$

We also decompose the inhomogeneous term F (which comes both from the original inhomogeneous term G and the cutoff):

$$\hat{F} = \sum_{m,\ell} F^{\omega}_{m\ell}(r) S_{m\ell}(a\omega, \cos\theta) e^{im\phi}.$$

Letting ζ be a sharp cutoff with such that $\zeta = 1$ for $|x| \le 1$ and $\zeta = 0$ for $|x| > 1$, we define

$$\begin{aligned}
\Phi_{\flat} &= \int_{-\infty}^{\infty} \zeta(\omega/\omega_1) \sum_{m,l:\lambda_{ml}(\omega)\le\lambda_1} R^{\omega}_{ml}(r) S_{ml}(a\omega, \cos\theta) e^{im\phi} e^{i\omega t}\, d\omega,\\
\Phi_{\reflectbox{$\flat$}} &= \int_{-\infty}^{\infty} \zeta(\omega/\omega_1) \sum_{m,l:\lambda_{ml}(\omega)>\lambda_1} R^{\omega}_{ml}(r) S_{ml}(a\omega, \cos\theta) e^{im\phi} e^{i\omega t}\, d\omega,\\
\Phi_{\natural} &= \int_{-\infty}^{\infty} (1 - \zeta(\omega/\omega_1)) \sum_{m,l:\lambda_{ml}(\omega)\ge\lambda_2\omega^2} R^{\omega}_{ml}(r) S_{ml}(a\omega, \cos\theta) e^{im\phi} e^{i\omega t}\, d\omega,\\
\Phi_{\sharp} &= \int_{-\infty}^{\infty} (1 - \zeta(\omega/\omega_1)) \sum_{m,l:\lambda_{ml}(\omega)<\lambda_2\omega^2} R^{\omega}_{ml}(r) S_{ml}(a\omega, \cos\theta) e^{im\phi} e^{i\omega t}\, d\omega.
\end{aligned}$$

In this decomposition, we think of ω_1 as large and λ_2 as small.

8.4. ***The trapped frequencies.*** Trapping occurs for $\Phi_{\natural}$. An integrated decay estimate is proved in detail in [Dafermos and Rodnianski 2008, Section 5.3.3]. The first term on right side in the following proposition is different from that in [ibid.], but the inequality still holds as a result of the proof of the corresponding inequality there.

Proposition 19.

$$\begin{aligned}
&\iint_{\mathcal{R}(-\infty,\infty)} (\chi\Phi_{\natural}^2 + \chi(\partial_r\Phi_{\natural})^2 + \chi\mathbb{1}_{\{|r-3M|\ge\frac{1}{8}M\}} J^N_{\mu}(\Phi_{\natural}) n^{\mu}_{\Sigma_{t^*}})\\
&\le C\int_{\mathcal{H}(-\infty,\infty)} (\partial_{t^*}\Phi^{\tau}_{\tau'})^2 + C\epsilon\int_{\mathcal{H}(-\infty,\infty)} (\partial_{\phi^*}\Phi^{\tau}_{\tau'})^2 + \int_{-\infty}^{\infty} dt^* \int_{r\ge R} (2f(r^2+a^2)^{1/2} F_{\natural}\partial_{r^*}((r^2+a^2)^{1/2}\Phi^{\tau}_{\tau'})\\
&\quad + f'(r^2+a^2)F_{\natural}\Phi^{\tau}_{\tau'})\frac{\Delta}{r^2+a^2}\sin\theta\, d\phi\, d\theta\, dr^* + \delta' \iint_{\mathcal{R}\cap\{r\le R\}} (\Phi^{\tau}_{\tau'})^2 + (\partial_{r^*}\Phi^{\tau}_{\tau'})^2 + C(\delta')^{-1}\iint_{\mathcal{R}\cap\{r\le R\}} F^2,
\end{aligned}$$

where χ is a weight that degenerates at infinity and near the event horizon and f is increasing and $f = \tan^{-1}(r^* - \alpha - \sqrt{\alpha})/\alpha - \tan^{-1}(-1-\alpha)^{-1/2}$ for $r > R$ for some fixed α.

8.5. ***The untrapped frequencies.*** For each of the pieces that are untrapped, that is, $\Phi_\bullet$ for $\bullet = \flat$, $\natural$ or $\sharp$, Dafermos and Rodnianski [2008] constructed a vector field $X_\bullet$ such that

$$\iint_{\mathcal{R}(-\infty,\infty)} \chi (J^{N_e}_\mu(\Phi_\bullet) n^\mu_{\Sigma_{t^*}} + \Phi_\bullet^2) \leq C \iint_{\mathcal{R}(-\infty,\infty)} K^{X_\bullet}(\Phi_\bullet),$$

where χ is a weight function that both degenerates at infinity and vanishes around the event horizon. Using this vector field and the conservation identity, they showed the following:

Proposition 20.

$$\begin{aligned}
&\iint_{\mathcal{R}(-\infty,\infty)} \chi\big((J^{N_e}_\mu(\Phi_\flat) + J^{N_e}_\mu(\Phi_\natural) + J^N_\mu(\Phi_\sharp)) n^\mu_{\Sigma_\tau} + (\Phi_\flat^2 + \Phi_\natural^2 + \Phi_\sharp^2)\big) \\
&\quad \leq C \int_{\mathcal{H}(-\infty,\infty)} J^{N_e}_\mu(\Phi) n^\mu_{\mathcal{H}^+} + C(\delta')^{-1} \iint_{\mathcal{R}(-\infty,\infty)\cap\{r\leq R\}} F^2 \\
&\qquad\qquad + C\delta' \iint_{\mathcal{R}(-\infty,\infty)\cap\{r\leq R\}} (\Phi^\tau_{\tau'})^2 + (\partial_{r^*}\Phi^\tau_{\tau'})^2 + \mathbb{1}_{\{r\leq \frac{23}{8}M\}} J^N_\mu(\Phi^\tau_{\tau'}) n^\mu_{\Sigma_{t^*}} \\
&\qquad\qquad + \int_{-\infty}^{\infty} dt^* \int_{\{r\geq R\}} (2f(r^2+a^2)^{1/2}(F_\flat + F_\natural + F_\sharp)\partial_{r^*}((r^2+a^2)^{1/2}\Phi^\tau_{\tau'}) \\
&\qquad\qquad\qquad\qquad + f'(r^2+a^2)(F_\flat + F_\natural + F_\sharp)\Phi^\tau_{\tau'}) \frac{\Delta}{r^2+a^2} \sin\theta \, d\phi \, d\theta \, dr^*,
\end{aligned}$$

where χ and f are exactly as in Proposition 19.

Proof. This inequality is essentially borrowed from Dafermos and Rodnianski [2008, Section 5.3.4]. The only difference is the first term on it right side. They used the estimate

$$\int_{\mathcal{H}(-\infty,\infty)} J^{X_\bullet}_\mu(\Phi_\bullet) n^\mu_{\mathcal{H}^+} \leq C \int_{\Sigma_{\tau'}} J^N_\mu(\Phi^\tau_{\tau'}) n^\mu_{\Sigma_{\tau'}}.$$

Here, we have not proved boundedness of the solution and hence we are content with the estimate

$$\int_{\mathcal{H}(-\infty,\infty)} J^{X_\bullet}_\mu(\Phi_\bullet) n^\mu_{\mathcal{H}^+} \leq C \int_{\mathcal{H}(-\infty,\infty)} J^{N_e}_\mu(\Phi^\tau_{\tau'}) n^\mu_{\mathcal{H}^+}.$$

This estimate holds for C independent of e because $X_\bullet$ is constructed as $f\partial_{r^*}$ and, on the event horizon, $\partial_{r^*} = O(1)\hat{V} + O(\epsilon)E_A$. □

8.6. ***The integrated decay estimates.*** To add up the estimates in the previous sections, we need a Hardy-type inequality:

Proposition 21. *For $R' < R$,*

$$\int_{\Sigma_\tau\cap\{r\geq R\}} r^{\alpha-2}\Phi^2 \leq C \int_{\Sigma_\tau\cap\{r\geq R'\}} r^\alpha J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_\tau}.$$

Proof. Let $k(r)$ be defined by solving $k'(r,\theta,\phi) = r^{\alpha-2}\,\mathrm{vol}$, where $\mathrm{vol} = \mathrm{vol}(r,\theta,\phi)$ is the volume density on Σ_τ with r, θ, ϕ coordinates, with boundary condition $k(R',\theta,\phi) = 0$. Now

$$\begin{aligned}\int_{\Sigma_\tau} r^{\alpha-2}\Phi^2 &= \iiint_{r_+}^{\infty} k'(r)\Phi^2 dr\, d\theta\, d\phi\\ &= -2\iiint k(r)\Phi\partial_r\Phi dr\, d\theta\, d\phi\\ &\le 2\Big(\iiint \frac{k(r)^2}{k'(r)}(\partial_r\Phi)^2 dr\, d\theta\, d\phi\Big)^{1/2}\Big(\iiint k'(r)\Phi^2 dr\, d\theta\, d\phi\Big)^{1/2}.\end{aligned}$$

Since $\mathrm{vol}\sim r^2$, $k(r)\sim r^{\alpha+1}$ and $k'(r)\sim r^\alpha$, we have $(1+k(r)^2)/(1+k'(r))\sim r^\alpha\,\mathrm{vol}$. □

We now add up the estimates for $\Phi_\flat$, $\Phi_\natural$, $\Phi_\natural$ and $\Phi_\sharp$.

Proposition 22.

$$\begin{aligned}&\iint_{\mathcal{R}(\tau',\tau)} (r^{-1-\delta}\mathbb{1}_{\{|r-3M|\ge\frac{1}{8}M\}} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_\tau} + r^{-1-\delta}(\partial_r\Phi)^2 + r^{-3-\delta}\Phi^2)\\ &\quad\le C\Big(\int_{\Sigma_\tau} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_\tau} + \int_{\Sigma_{\tau'}} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_{\tau'}} + \int_{\mathcal{H}(\tau',\tau)} J^{N_e}_\mu(\Phi)n^\mu_{\mathcal{H}^+}\\ &\qquad + \iint_{\mathcal{R}(\tau'-1,\tau+1)} (|\partial_{r^*}\Phi| + r^{-1}|\Phi|)|G| + \Big|\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\le\frac{23}{8}M\}} (\partial_{t^*}\Phi + eY\Phi)G\Big|\\ &\qquad + \iint_{\mathcal{R}(\tau'-1,\tau+1)} G^2\Big).\end{aligned}$$

Proof. Since the function f appears identically in Propositions 19 and 20, we can add up the estimates to obtain

$$\begin{aligned}&\iint_{\mathcal{R}(-\infty,\infty)} \big(\chi(J^{N_e}_\mu(\Phi_\flat) + J^{N_e}_\mu(\Phi_\natural) + \mathbb{1}_{\{|r-3M|\ge\frac{1}{8}M\}} J^{N_e}_\mu(\Phi_\natural) + J^{N_e}_\mu(\Phi_\sharp))n^\mu_{\Sigma_\tau} + \chi(\Phi_\flat^2 + \Phi_\natural^2 + \Phi_\natural^2 + \Phi_\sharp^2)\big)\\ &\quad\le C\int_{\mathcal{H}(\tau'-1,\tau+1)} J^{N_e}_\mu(\Phi)n^\mu_{\mathcal{H}^+} + C(\delta')^{-1}\iint_{\mathcal{R}(-\infty,\infty)\cap\{r\le R\}} F^2\\ &\qquad + C\delta'\iint_{\mathcal{R}(-\infty,\infty)\cap\{r\le R\}} (\Phi^\tau_{\tau'})^2 + (\partial_{r^*}\Phi^\tau_{\tau'})^2 + \mathbb{1}_{\{r\le\frac{23}{8}M\}} J^{N_e}_\mu(\Phi^\tau_{\tau'})n^\mu_{\Sigma_{t^*}}\\ &\qquad + \int_{-\infty}^{\infty} dt^* \int_{\{r\ge R\}} (2f(r^2+a^2)^{1/2}(F_\flat + F_\natural + F_\natural + F_\sharp)\partial_{r^*}((r^2+a^2)^{1/2}\Phi^\tau_{\tau'})\\ &\qquad\qquad + f'(r^2+a^2)(F_\flat + F_\natural + F_\natural + F_\sharp)\Phi^\tau_{\tau'})\frac{\Delta}{r^2+a^2}\sin\theta\, d\phi\, d\theta\, dr^*.\end{aligned}$$

By the definition of the cutoff, we have the pointwise equalities

$$F = F_\flat + F_\natural + F_\natural + F_\sharp.$$

Therefore, we have

$$\begin{aligned}
&\iint_{\mathcal{R}(-\infty,\infty)} \chi\, \mathbb{1}_{\{|r-3M|\geq \frac{1}{8}M\}} J^{N_e}_\mu(\Phi^\tau_{\tau'}) n^\mu_{\Sigma_{t^*}} + \chi(\Phi^\tau_{\tau'})^2 \\
&\leq C\int_{\mathcal{H}(\tau'-1,\tau+1)} J^{N_e}_\mu(\Phi) n^\mu_{\mathcal{H}^+} + C(\delta')^{-1}\iint_{\mathcal{R}(-\infty,\infty)\cap\{r\leq R\}} F^2 \\
&\quad + C\delta' \iint_{\mathcal{R}(-\infty,\infty)\cap\{r\leq R\}} (\Phi^\tau_{\tau'})^2 + (\partial_{r^*}\Phi^\tau_{\tau'})^2 + \mathbb{1}_{\{r\leq \frac{23}{8}M\}} J^{N_e}_\mu(\Phi^\tau_{\tau'}) n^\mu_{\Sigma_{t^*}} \\
&\quad + \int_{-\infty}^{\infty} dt^* \int_{\{r\geq R\}} \big(2f(r^2+a^2)^{1/2} F \partial_{r^*}((r^2+a^2)^{1/2}\Phi^\tau_{\tau'}) + f'(r^2+a^2) F\Phi^\tau_{\tau'}\big)\frac{\Delta}{r^2+a^2}\sin\theta\, d\phi\, d\theta\, dr^*.
\end{aligned}$$

First, by Proposition 10, we have

$$\int_{\mathcal{H}(\tau'-1,\tau+1)} J^{N_e}_\mu(\Phi) n^\mu_{\mathcal{H}^+} \leq C\int_{\Sigma_{\tau'}} J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_{\tau'}} + C\int_{\Sigma_\tau} J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J^{N_e}_\mu(\Phi) n^\mu_{\mathcal{H}^+}.$$

Recall that

$$F = \xi G + 2D^\alpha\Phi D_\alpha\xi + \Phi\Box_{g_K}\xi.$$

By the definition of ξ, the last two terms are supported in the t^* range $(\tau'-1,\tau')\cup(\tau,\tau+1)$. Moreover, since ξ depends only on t^*, the only terms involving $D\Phi$ are $\partial_{t^*}\Phi$ and $O(\epsilon)\partial_{\phi^*}\Phi$. Using this, we immediately have the following with C independent of e as long as $\epsilon \ll e$:

$$\begin{aligned}
&C(\delta')^{-1}\iint_{\mathcal{R}(-\infty,\infty)\cap\{r\leq R\}} F^2 \\
&\qquad\leq C(\delta')^{-1}\bigg(\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\leq R\}} G^2 + \iint_{\mathcal{R}(\tau'-1,\tau')\cup\mathcal{R}(\tau,\tau+1)} (r^{-2}\Phi^2 + J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_{t^*}})\bigg).
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\int_{-\infty}^{\infty} dt^* \int_{\{r\geq R\}} f'(r^2+a^2) F\Phi^\tau_{\tau'}\frac{\Delta}{r^2+a^2}\sin\theta\, d\phi\, d\theta\, dr^* \\
&\qquad\leq C\bigg(\iint_{\mathcal{R}(\tau'-1,\tau+1)} r^{-1}|\Phi||G| + \iint_{\mathcal{R}(\tau'-1,\tau')\cup\mathcal{R}(\tau,\tau+1)} (r^{-2}\Phi^2 + J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_{t^*}})\bigg).
\end{aligned}$$

The other term with F is more delicate to estimate. One of the terms in the expansion does not have sufficient decay in r:

$$\begin{aligned}
&\int_{-\infty}^{\infty} dt^* \int_{\{r\geq R\}} 2f(r^2+a^2)^{1/2} F \partial_{r^*}((r^2+a^2)^{1/2}\Phi^\tau_{\tau'})\frac{\Delta}{r^2+a^2}\sin\theta\, d\phi\, d\theta\, dr^* \\
&\quad\leq C\bigg(\iint_{\mathcal{R}(\tau'-1,\tau+1)} r^{-1}|\Phi||G| + \iint_{\mathcal{R}(\tau'-1,\tau')\cup\mathcal{R}(\tau,\tau+1)} (r^{-2}\Phi^2 + J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_{t^*}})\bigg) \\
&\qquad\qquad + \int_{-\infty}^{\infty} dt^* \int_{\{r\geq R\}} 2f(r^2+a^2)^{1/2}\Phi\Box_{g_K}\xi\, \partial_{r^*}((r^2+a^2)^{1/2}\Phi)\xi\frac{\Delta}{r^2+a^2}\sin\theta\, d\phi\, d\theta\, dr^*
\end{aligned}$$

Nevertheless, noting that ξ is independent of r^*, an integration by parts in r^* would give

$$\begin{aligned}\int_{-\infty}^{\infty} dt^* \int_{\{r\geq R\}} & 2f(r^2+a^2)^{1/2}\Phi\Box_{g_K}\xi\,\partial_{r^*}((r^2+a^2)^{1/2}\Phi)\xi\,\frac{\Delta}{r^2+a^2}\sin\theta\, d\phi\, d\theta\, dr^* \\ &= -\int_{-\infty}^{\infty} dt^* \int_{\{r\geq R\}} (r^2+a^2)\Phi^2\xi\,\Box_{g_K}\xi\,\partial_{r^*}\Big(f\frac{\Delta}{r^2+a^2}\Big)\sin\theta\, d\phi\, d\theta\, dr^* + \text{boundary terms} \\ &\leq C\iint_{\mathcal{R}(\tau'-1,\tau')\cup\mathcal{R}(\tau,\tau+1)} r^{-2}\Phi^2,\end{aligned}$$

where the boundary terms can be controlled (after possibly changing R) by pigeonholing in $r\in[R,\,R+1]$. By the mild growth estimate of Proposition 10, the estimate near the event horizon from Proposition 18 and the Hardy inequality of Proposition 21,

$$\begin{aligned}\iint_{\mathcal{R}(\tau'-1,\tau')\cup\mathcal{R}(\tau,\tau+1)} & (r^{-2}\Phi^2 + J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_{t^*}}) \\ &\leq C\bigg(\int_{\Sigma_{\tau'}} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_{\tau'}} + \int_{\Sigma_\tau} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_{\tau'}} + \iint_{\mathcal{R}(\tau'-1,\tau')\cup\mathcal{R}(\tau,\tau+1)} G^2\bigg).\end{aligned}$$

Therefore, using all the estimates above and noticing the support of ξ, we have

$$\begin{aligned}&\iint_{\mathcal{R}(\tau',\tau)} \chi(\mathbb{1}_{\{|r-3M|\geq\frac{1}{8}M\}}J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_{t^*}}+\Phi^2) \\ &\leq C(\delta')^{-1}\int_{\Sigma_{\tau'}} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_{\tau'}} + C(\delta')^{-1}\int_{\Sigma_\tau} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_\tau} + C\int_{\mathcal{H}(\tau',\tau)} J^{N_e}_\mu(\Phi)n^\mu_{\mathcal{H}^+} \\ &\qquad + C\iint_{\mathcal{R}(\tau'-1,\tau+1)} r^{-1}|\Phi||G| + C(\delta')^{-1}\iint_{\mathcal{R}(\tau'-1,\tau+1)} G^2 \\ &\qquad + C\delta'\iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq R\}} \Phi^2+(\partial_{r^*}\Phi)^2+\mathbb{1}_{\{r\leq\frac{23}{8}M\}}J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_{t^*}}.\end{aligned}$$

We add to this the estimates near spatial infinity and the event horizon, that is, Propositions 15 and 18, to get

$$\begin{aligned}&\iint_{\mathcal{R}(\tau',\tau)} r^{-1-\delta}\mathbb{1}_{\{|r-3M|\geq\frac{1}{8}M\}}J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_{t^*}} + r^{-1-\delta}(\partial_r\Phi)^2 + r^{-3-\delta}\Phi^2 \\ &\leq C(\delta')^{-1}\int_{\Sigma_{\tau'}} J^{N}_\mu(\Phi)n^\mu_{\Sigma_{\tau'}} + C(\delta')^{-1}\int_{\Sigma_\tau} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_\tau} + C\int_{\mathcal{H}(\tau',\tau)} J^{N_e}_\mu(\Phi)n^\mu_{\mathcal{H}^+} \\ &\quad + C\iint_{\mathcal{R}(\tau'-1,\tau+1)} (|\partial_{r^*}\Phi|+r^{-1}|\Phi|)|G| + C\Big|\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\leq\frac{23}{8}M\}} (\partial_{t^*}\Phi+eY\Phi)G\Big| \\ &\quad + C(\delta')^{-1}\iint_{\mathcal{R}(\tau'-1,\tau+1)} G^2 + C\delta'\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\leq R\}} \Phi^2+(\partial_{r^*}\Phi)^2+\mathbb{1}_{\{r\leq\frac{23}{8}M\}}J^{N}_\mu(\Phi)n^\mu_{\Sigma_{t^*}}.\end{aligned}$$

By choosing δ' sufficiently small and absorbing $(\delta')^{-1}$ into the constant C, we can absorb the last term:

$$\begin{aligned}
&\iint_{\mathcal{R}(\tau',\tau)} r^{-1-\delta}\mathbb{1}_{\{|r-3M|\geq\frac{1}{8}M\}} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_{t^*}} + r^{-1-\delta}(\partial_r\Phi)^2 + r^{-3-\delta}\Phi^2 \\
&\quad\leq C\int_{\Sigma_{\tau'}} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_{\tau'}} + C\int_{\Sigma_\tau} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_\tau} + C\int_{\mathcal{H}(\tau',\tau)} J^{N_e}_\mu(\Phi)n^\mu_{\mathcal{H}^+} \\
&\qquad + C\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\leq\frac{23}{8}M\}} (|\partial_{r^*}\Phi| + r^{-1}|\Phi|)|G| + C\Big|\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\leq\frac{23}{8}M\}} (\partial_{t^*}\Phi + eY\Phi)G\Big| \\
&\qquad\qquad + C\iint_{\mathcal{R}(\tau'-1,\tau+1)} G^2.
\end{aligned}$$

using Proposition 10 and 21 at the last step. □

Definition 23. From now on, we write

$$\begin{aligned}
K^{X_0}(\Phi) &= r^{-1-\delta}\mathbb{1}_{\{|r-3M|\geq\frac{1}{8}M\}} J^N_\mu(\Phi)n^\mu_{\Sigma_\tau} + r^{-1-\delta}(\partial_r\Phi)^2 + r^{-3-\delta}\Phi^2, \\
K^{X_1}(\Phi) &= r^{-1-\delta} J^N_\mu(\Phi)n^\mu_{\Sigma_\tau} + r^{-3-\delta}\Phi^2.
\end{aligned}$$

This is a slight abuse of notation because these "currents" do not arise directly from a vector field.

8.7. *Boundedness of the nondegenerate energy.*

Proposition 24. *Let Φ satisfy $\Box_{g_K}\Phi = G$. For e sufficiently small and $\epsilon \ll e$, we have*

$$\begin{aligned}
&\int_{\Sigma_\tau} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J^{N_e}_\mu(\Phi)n^\mu_{\mathcal{H}^+} + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq r_Y^-\}} K^{N_e}_\mu(\Phi) \\
&\quad\leq C\bigg(\int_{\Sigma_{\tau'}} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_{\tau'}} + \Big|\iint_{\mathcal{R}(\tau'-1,\tau+1)} \partial_{t^*}\Phi G\Big| + \Big|\iint_{\mathcal{R}(\tau'-1,\tau+1)} eY\Phi G\Big| \\
&\qquad\qquad + \iint_{\mathcal{R}(\tau'-1,\tau+1)} (|\partial_r\Phi| + r^{-1}|\Phi|)|G| + \iint_{\mathcal{R}(\tau'-1,\tau+1)} G^2\bigg).
\end{aligned}$$

Proof. We recall that

$$\iint_{\mathcal{R}(\tau',\tau)\cap\{r_Y^-\leq r\leq r_Y^+\}} K^Y(\Phi) \leq C\int_{\tau'}^{\tau}\int_{\Sigma_{\bar\tau}} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_{t^*}}dt^*,$$

with C independent of e and ϵ whenever $\epsilon \ll e < 1$. At this point, we choose $r_Y^+ < \frac{11}{4}M < \frac{23}{8}M$. Hence this term can be controlled by the integrated decay estimates. Then, by Proposition 9,

$$\begin{aligned}
&\int_{\Sigma_\tau} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J^{N_e}_\mu(\Phi)n^\mu_{\mathcal{H}^+} + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq r_Y^-\}} K^{N_e}_\mu(\Phi) \\
&\quad= \int_{\Sigma_{\tau'}} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_{\tau'}} + e\iint_{\mathcal{R}(\tau',\tau)\cap\{r_Y^-\leq r\leq r_Y^+\}} K^Y(\Phi) + \iint_{\mathcal{R}(\tau',\tau)} (\partial_{t^*}\Phi + eY\Phi)G \\
&\quad\leq \int_{\Sigma_{\tau'}} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_{\tau'}} + Ce\int_{\tau'}^{\tau}\int_{\Sigma_{\bar\tau}\cap\{r_Y^-\leq r\leq r_Y^+\}} J^{N_e}_\mu(\Phi)n^\mu_{\Sigma_{\bar\tau}}d\bar\tau + \Big|\iint_{\mathcal{R}(\tau',\tau)} (\partial_{t^*}\Phi + eY\Phi)G\Big|
\end{aligned}$$

$$\leq \int_{\Sigma_{\tau'}} J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_{\tau'}} + Ce\Big(\int_{\Sigma_\tau} J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_\tau} + \int_{\Sigma_{\tau'}} J^{N_e}_\mu(\Phi) n^\mu_{\Sigma_{\tau'}} + \int_{\mathcal{H}(\tau',\tau)} J^{N_e}_\mu(\Phi) n^\mu_{\mathcal{H}^+} + \iint_{\mathcal{R}(\tau'-1,\tau+1)} (|\partial_r\Phi| + r^{-1}|\Phi|)|G| + \iint_{\mathcal{R}(\tau'-1,\tau+1)} G^2\Big) + \Big|\iint_{\mathcal{R}(\tau',\tau)} (\partial_{t^*}\Phi + eY\Phi)G\Big|.$$

Hence, the proposition holds if e is chosen to be sufficiently small. □

Remark. From this point on, we will consider r_Y^+ and e to be fixed. After e is fixed, the vector field N_e will be written simply as N.

We now estimate the inhomogeneous terms in Proposition 24:

Proposition 25.

$$\int_{\Sigma_\tau} J^N_\mu(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J^N_\mu(\Phi) n^\mu_{\mathcal{H}^+} + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq r_Y^-\}} K^N(\Phi) + \iint_{\mathcal{R}(\tau',\tau)} K^{X_0}(\Phi) \leq C\Big(\int_{\Sigma_{\tau'}} J^N_\mu(\Phi) n^\mu_{\Sigma_{\tau'}} + \Big(\int_{\tau'-1}^{\tau+1}\Big(\int_{\Sigma_{t^*}} G^2\Big)^{1/2} dt^*\Big)^2 + \iint_{\mathcal{R}(\tau',\tau)} G^2\Big).$$

Proof. Adding the estimates in Propositions 22 and δ times the estimates in Proposition 24, we have

$$\begin{aligned}
&\int_{\Sigma_\tau} J^N_\mu(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J^N_\mu(\Phi) n^\mu_{\mathcal{H}^+} + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq r_Y^-\}} K^N(\Phi) + \delta\iint_{\mathcal{R}(\tau',\tau)} K^{X_0}(\Phi) \\
&\quad\leq C\Big(\int_{\Sigma_{\tau'}} J^N_\mu(\Phi) n^\mu_{\Sigma_{\tau'}} + \iint_{\mathcal{R}(\tau'-1,\tau+1)} (|\partial\Phi| + r^{-1}|\Phi|)|G| + \iint_{\mathcal{R}(\tau'-1,\tau+1)} G^2\Big) \\
&\qquad\qquad + C\delta\Big(\int_{\Sigma_\tau} J^N_\mu(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J^N_\mu(\Phi) n^\mu_{\mathcal{H}^+}\Big) \\
&\quad\leq C\Big(\int_{\Sigma_{\tau'}} J^N_\mu(\Phi) n^\mu_{\Sigma_{\tau'}} + \sup_{t^*\in[\tau'-1,\tau+1]}\Big(\int_{\Sigma_{t^*}} J^N_\mu(\Phi) n^\mu_{\Sigma_{t^*}}\Big)^{1/2} \int_{\tau'-1}^{\tau+1}\Big(\int_{\Sigma_{t^*}} G^2\Big)^{1/2} dt^* \\
&\qquad\qquad + \iint_{\mathcal{R}(\tau'-1,\tau+1)} G^2\Big) + C\delta\Big(\int_{\Sigma_\tau} J^N_\mu(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J^N_\mu(\Phi) n^\mu_{\mathcal{H}^+}\Big),
\end{aligned}$$

where at the last step we have used Proposition 21. Choosing $C\delta \leq \frac{1}{2}$, we can absorb the last term to the left side to get

$$\begin{aligned}
&\int_{\Sigma_\tau} J^N_\mu(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J^N_\mu(\Phi) n^\mu_{\mathcal{H}^+} + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq r_Y^-\}} K^N(\Phi) + \delta\iint_{\mathcal{R}(\tau',\tau)} K^{X_0}(\Phi) \\
&\quad\leq C\Big(\int_{\Sigma_{\tau'}} J^N_\mu(\Phi) n^\mu_{\Sigma_{\tau'}} + \sup_{t^*\in[\tau'-1,\tau+1]}\Big(\int_{\Sigma_{t^*}} J^N_\mu(\Phi) n^\mu_{\Sigma_{t^*}}\Big)^{1/2} \int_{\tau'-1}^{\tau+1}\Big(\int_{\Sigma_{t^*}} G^2\Big)^{1/2} dt^* \\
&\qquad\qquad + \iint_{\mathcal{R}(\tau'-1,\tau+1)} G^2\Big). \qquad (7)
\end{aligned}$$

By considering the estimate above on $[\tau', \tilde{\tau}]$, where $\tilde{\tau}$ is when the supremum on the right-hand side is achieved, and using Proposition 10, we get

$$\begin{aligned}\int_{\Sigma_{\tilde{\tau}}} J_\mu^N(\Phi) n^\mu_{\Sigma_{\tilde{\tau}}} + \int_{\mathcal{H}(\tau',\tilde{\tau})} J_\mu^N(\Phi) n^\mu_{\mathcal{H}^+} + \iint_{\mathcal{R}(\tau',\tilde{\tau})\cap\{r\le r_Y^-\}} K^N(\Phi) + \delta \iint_{\mathcal{R}(\tau',\tilde{\tau})} K^{X_0}(\Phi)\\ \le C\Big(\int_{\Sigma_{\tau'}} J_\mu^N(\Phi) n^\mu_{\Sigma_{\tau'}} + \Big(\int_{\tau'-1}^{\tau+1} (\int_{\Sigma_{t^*}} G^2)^{1/2} dt^*\Big)^2\Big).\end{aligned}$$

We plug this into (7) and apply Cauchy–Schwarz to prove the proposition. □

We can also estimate the inhomogeneous terms not in L^1L^2 but in L^2L^2, provided that we allow some extra factors of r and some loss of derivatives in G. This is especially useful for estimating the commutator terms from S, which do not have sufficient decay in t^* in the interior to be estimated in L^1L^2. More precisely:

Proposition 26. $\displaystyle\int_{\Sigma_\tau} J_\mu^N(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J_\mu^N(\Phi) n^\mu_{\mathcal{H}^+} + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} K^N(\Phi) + \iint_{\mathcal{R}(\tau',\tau)} K^{X_0}(\Phi)$

$$\le C\Big(\int_{\Sigma_{\tau'}} J_\mu^N(\Phi) n^\mu_{\Sigma_{\tau'}} + \sum_{m=0}^{1} \iint_{\mathcal{R}(\tau'-1,\tau+1)} r^{1+\delta} (\partial_{t^*}^m G)^2 + \sup_{t^*\in[\tau'-1,\tau+1]} \int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}} G^2\Big).$$

Proof. By Propositions 22 and 24,

$$\begin{aligned}&\int_{\Sigma_\tau} J_\mu^N(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J_\mu^N(\Phi) n^\mu_{\mathcal{H}^+} + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} K^N(\Phi) + \delta \iint_{\mathcal{R}(\tau',\tau)} K^{X_0}(\Phi)\\ &\le C\Big(\int_{\Sigma_{\tau'}} J_\mu^{N_e}(\Phi) n^\mu_{\Sigma_{\tau'}} + \Big|\iint_{\mathcal{R}(\tau'-1,\tau+1)} \partial_{t^*}\Phi G\Big|\\ &\qquad + \Big|\iint_{\mathcal{R}(\tau'-1,\tau+1)} eY\Phi G\Big| + \iint_{\mathcal{R}(\tau'-1,\tau+1)} (|\partial_r\Phi| + r^{-1}|\Phi|)|G| + \iint_{\mathcal{R}(\tau'-1,\tau+1)} G^2\Big)\\ &\qquad + C\delta'\Big(\int_{\Sigma_\tau} J_\mu^N(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J_\mu^N(\Phi) n^\mu_{\mathcal{H}^+}\Big).\end{aligned}$$

Choosing $C\delta' \le \frac{1}{2}$, we can absorb the last term into the left hand side to get

$$\begin{aligned}&\int_{\Sigma_\tau} J_\mu^N(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J_\mu^N(\Phi) n^\mu_{\mathcal{H}^+} + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} K^N(\Phi) + \iint_{\mathcal{R}(\tau',\tau)} K^{X_0}(\Phi)\\ &\le C\Big(\int_{\Sigma_{\tau'}} J_\mu^N(\Phi) n^\mu_{\Sigma_{\tau'}} + \Big|\iint_{\mathcal{R}(\tau'-1,\tau+1)} \partial_{t^*}\Phi G\Big| + \Big|\iint_{\mathcal{R}(\tau'-1,\tau+1)} eY\Phi G\Big|\\ &\qquad + \iint_{\mathcal{R}(\tau'-1,\tau+1)} (|\partial_r\Phi| + r^{-1}|\Phi|)|G| + \iint_{\mathcal{R}(\tau'-1,\tau+1)} G^2\Big).\end{aligned}$$

For the bulk error term, we focus at the region $\{|r-3M|\leq\frac{1}{8}M\}$ and integrate by parts.

$$\begin{aligned}
&\Big|\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{|r-3M|\leq\frac{1}{8}M\}}\partial_{t^*}\Phi G\Big|\\
&\quad\leq\delta'\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{|r-3M|\leq\frac{1}{8}M\}}\Phi^2+C(\delta')^{-1}\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{|r-3M|\leq\frac{1}{8}M\}}(\partial_{t^*}G)^2\\
&\qquad\qquad+\Big|\int_{\Sigma_{\tau+1}\cap\{|r-3M|\leq\frac{1}{8}M\}}\Phi G\Big|+\Big|\int_{\Sigma_{\tau'-1}\cap\{|r-3M|\leq\frac{1}{8}M\}}\Phi G\Big|\\
&\quad\leq\delta'\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{|r-3M|\leq\frac{1}{8}M\}}r^{-3-\delta}\Phi^2+C(\delta')^{-1}\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{|r-3M|\leq\frac{1}{8}M\}}(\partial_{t^*}G)^2\\
&\qquad\qquad+\sup_{t^*\in[\tau'-1,\tau+1]}\Big(\delta\int_{\Sigma_{t^*}\cap\{|r-3M|\leq\frac{1}{8}M\}}r^{-2}\Phi^2+C(\delta')^{-1}\int_{\Sigma_{t^*}\cap\{|r-3M|\leq\frac{1}{8}M\}}G^2\Big)\\
&\quad\leq\delta'\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{|r-3M|\leq\frac{1}{8}M\}}r^{-3-\delta}\Phi^2+C(\delta')^{-1}\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{|r-3M|\leq\frac{1}{8}M\}}(\partial_{t^*}G)^2\\
&\qquad\qquad+\sup_{t^*\in[\tau'-1,\tau+1]}\Big(\delta'\int_{\Sigma_{t^*}}J^N_\mu(\Phi)n^\mu_{\Sigma_{t^*}}+C(\delta')^{-1}\int_{\Sigma_{t^*}\cap\{|r-3M|\leq\frac{1}{8}M\}}G^2\Big),
\end{aligned}$$

where at the last step we used Proposition 21. Therefore,

$$\begin{aligned}
&\int_{\Sigma_\tau}J^N_\mu(\Phi)n^\mu_{\Sigma_\tau}+\int_{\mathcal{H}(\tau',\tau)}J^N_\mu(\Phi)n^\mu_{\mathcal{H}^+}+\iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq r_Y^-\}}K^N(\Phi)+\iint_{\mathcal{R}(\tau',\tau)}K^{X_0}(\Phi)\\
&\quad\leq C\Big(\int_{\Sigma_{\tau'}}J^N_\mu(\Phi)n^\mu_{\Sigma_{\tau'}}+\Big|\iint_{\mathcal{R}(\tau'-1,\tau+1)}\partial_{t^*}\Phi G\Big|+\Big|\iint_{\mathcal{R}(\tau'-1,\tau+1)}eY\Phi G\Big|\\
&\qquad\qquad+\iint_{\mathcal{R}(\tau'-1,\tau+1)}(|\partial_r\Phi|+r^{-1}|\Phi|)|G|+\iint_{\mathcal{R}(\tau'-1,\tau+1)}G^2\Big)\\
&\quad\leq C\Big(\int_{\Sigma_{\tau'}}J^N_\mu(\Phi)n^\mu_{\Sigma_{\tau'}}+\Big|\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{|r-3M|\leq\frac{1}{8}M\}}\partial_{t^*}\Phi G\Big|\Big)+C(\delta')^{-1}\iint_{\mathcal{R}(\tau'-1,\tau+1)}r^{1+\delta}G^2\\
&\qquad\qquad+\delta'\iint_{\mathcal{R}(\tau'-1,\tau+1)}(r^{-3-\delta}\Phi^2+r^{-1-\delta}(\partial_r\Phi)^2+\mathbb{1}_{\{r\leq r_Y^+\}}J^N_\mu(\Phi)n^\mu_{\Sigma_{t^*}})\\
&\quad\leq C\int_{\Sigma_{\tau'}}J^N_\mu(\Phi)n^\mu_{\Sigma_{\tau'}}+C(\delta')^{-1}\sum_{m=0}^{1}\int_{\mathcal{R}(\tau'-1,\tau+1)}r^{1+\delta}(\partial^m_{t^*}G)^2+\delta'\iint_{\mathcal{R}(\tau'-1,\tau+1)}K^{X_0}(\Phi)\\
&\qquad\qquad+\sup_{t^*\in[\tau'-1,\tau+1]}\Big(\delta'\int_{\Sigma_{t^*}}J^N_\mu(\Phi)n^\mu_{\Sigma_{t^*}}+C(\delta')^{-1}\int_{\Sigma_{t^*}\cap\{|r-3M|\leq\frac{1}{8}M\}}G^2\Big).\qquad(8)
\end{aligned}$$

where at the last step we have used Propositions 10 and 21. Suppose $\sup_{t^*\in[\tau'-1,\tau+1]}\delta'\int_{\Sigma_{t^*}}J^N_\mu(\Phi)n^\mu_{\Sigma_{t^*}}$ is achieved by $t^*=\tilde{\tau}$. Applying (8) on $[\tau',\tilde{\tau}]$, we get

$$\begin{aligned}\int_{\Sigma_{\tilde{\tau}}}J^N_\mu(\Phi)n^\mu_{\Sigma_{\tilde{\tau}}}\leq C\int_{\Sigma_{\tau'}}J^N_\mu(\Phi)n^\mu_{\Sigma_{\tau'}}+C(\delta')^{-1}\sum_{m=0}^{1}\iint_{\mathcal{R}(\tau'-1,\tau+1)}r^{1+\delta}(\partial_{t^*}^m G)^2+\delta'\iint_{\mathcal{R}(\tau'-1,\tau+1)}K^{X_0}(\Phi)\\+\delta'\int_{\Sigma_{\tilde{\tau}}}J^N_\mu(\Phi)n^\mu_{\Sigma_{\tilde{\tau}}}+C\sup_{t^*\in[\tau'-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\leq\frac{1}{8}M\}}G^2,\end{aligned}$$

which, upon choosing $\delta'\leq\frac{1}{2}$ and subtracting the small term on both sides, gives

$$\begin{aligned}\int_{\Sigma_{\tilde{\tau}}}J^N_\mu(\Phi)n^\mu_{\Sigma_{\tilde{\tau}}}\leq C\int_{\Sigma_{\tau'}}J^N_\mu(\Phi)n^\mu_{\Sigma_{\tau'}}+C(\delta')^{-1}\sum_{m=0}^{1}\iint_{\mathcal{R}(\tau'-1,\tau+1)}r^{1+\delta}(\partial_{t^*}^m G)^2+\delta'\iint_{\mathcal{R}(\tau'-1,\tau+1)}K^{X_0}(\Phi)\\+C\sup_{t^*\in[\tau'-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\leq\frac{1}{8}M\}}G^2,\end{aligned}$$

Therefore, plugging this back into (8), we have

$$\begin{aligned}&\int_{\Sigma_\tau}J^N_\mu(\Phi)n^\mu_{\Sigma_\tau}+\int_{\mathcal{H}(\tau',\tau)}J^N_\mu(\Phi)n^\mu_{\mathcal{H}^+}+\iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq r_Y^-\}}K^N(\Phi)+\iint_{\mathcal{R}(\tau',\tau)}K^{X_0}(\Phi)\\&\quad\leq C\int_{\Sigma_{\tau'}}J^N_\mu(\Phi)n^\mu_{\Sigma_{\tau'}}+\delta'\iint_{\mathcal{R}(\tau',\tau)}K^{X_0}(\Phi)+\delta'\iint_{\mathcal{R}(\tau'-1,\tau')\cup\mathcal{R}(\tau,\tau+1)}K^{X_0}(\Phi)\\&\qquad\qquad+C\sum_{m=0}^{1}\iint_{\mathcal{R}(\tau'-1,\tau+1)}r^{1+\delta}(\partial_{t^*}^m G)^2+C\sup_{t^*\in[\tau'-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\leq\frac{1}{8}M\}}G^2\\&\quad\leq C\int_{\Sigma_{\tau'}}J^N_\mu(\Phi)n^\mu_{\Sigma_{\tau'}}+\delta'\iint_{\mathcal{R}(\tau',\tau)}K^{X_0}(\Phi)+C\delta'\int_{\Sigma_\tau}J^N_\mu(\Phi)n^\mu_{\Sigma_\tau}\\&\qquad\qquad+C\sum_{m=0}^{1}\iint_{\mathcal{R}(\tau'-1,\tau+1)}r^{1+\delta}(\partial_{t^*}^m G)^2+C\sup_{t^*\in[\tau'-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\leq\frac{1}{8}M\}}G^2,\end{aligned}$$

where at the last step we have used Proposition 10. Finally, by choosing $C\delta\leq\frac{1}{2}$, we can absorb the small terms into the left-hand side and achieve the conclusion of the proposition. □

In the proof of Proposition 26, there is a loss in derivative for G because we have to integrate by parts in the region $\{|r-3M|\leq\frac{1}{8}M\}$. Therefore, if G is supported away from this region, we can repeat the proof without this loss. In other words:

Proposition 27. *Suppose G is supported away from $\{|r-3M|\leq\frac{1}{8}M\}$. Then*

$$\begin{aligned}&\int_{\Sigma_\tau}J^N_\mu(\Phi)n^\mu_{\Sigma_\tau}+\int_{\mathcal{H}(\tau',\tau)}J^N_\mu(\Phi)n^\mu_{\mathcal{H}^+}+\iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq r_Y^-\}}K^N(\Phi)+\iint_{\mathcal{R}(\tau',\tau)}K^{X_0}(\Phi)\\&\qquad\leq C\Big(\int_{\Sigma_{\tau'}}J^N_\mu(\Phi)n^\mu_{\Sigma_{\tau'}}+\sum_{m=0}^{1}\iint_{\mathcal{R}(\tau'-1,\tau+1)}r^{1+\delta}G^2+\sup_{t^*\in[\tau'-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\leq\frac{1}{8}M\}}G^2\Big).\end{aligned}$$

This will be useful in Section 13.

In applications, it is useful to have both ways of estimating G.

Proposition 28. *Let $G = G_1 + G_2$ be any way to decompose the function G. Then*

$$\begin{aligned}\int_{\Sigma_\tau} J^N_\mu(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J^N_\mu(\Phi) n^\mu_{\mathcal{H}^+} + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} K^N(\Phi) + \iint_{\mathcal{R}(\tau',\tau)} K^{X_0}(\Phi) \\ \le C\bigg(\int_{\Sigma_{\tau'}} J^N_\mu(\Phi) n^\mu_{\Sigma_{\tau'}} + \bigg(\int_{\tau'-1}^{\tau+1}\bigg(\int_{\Sigma_{t^*}} G_1^2\bigg)^{1/2} dt^*\bigg)^2 + \iint_{\mathcal{R}(\tau'-1,\tau+1)} G_1^2 \\ + \sum_{m=0}^{1}\iint_{\mathcal{R}(\tau'-1,\tau+1)} r^{1+\delta}(\partial^m_{t^*}G_2)^2 + \sup_{t^*\in[\tau'-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}} G_2^2\bigg).\end{aligned}$$

In the estimates above, only the function Φ and its ∂_r derivative can be estimated without a loss around the trapped set. To estimate the other derivatives, we need to commute with the Killing vector field ∂_{t^*}.

Proposition 29. *We have*

$$\begin{aligned}\iint_{\mathcal{R}(\tau',\tau)} K^{X_1}(\Phi) \\ \le C\bigg(\sum_{m=0}^{1}\int_{\Sigma_{\tau'}} J^N_\mu(\partial^m_{t^*}\Phi) n^\mu_{\Sigma_{\tau'}} + \sum_{m=0}^{1}\bigg(\int_{\tau'-1}^{\tau+1}\bigg(\int_{\Sigma_{t^*}} (\partial^m_{t^*}G_1)^2\bigg)^{1/2} dt^*\bigg)^2 + \sum_{m=0}^{1}\iint_{\mathcal{R}(\tau'-1,\tau+1)} (\partial^m_{t^*}G_1)^2 \\ + \sum_{m=0}^{2}\iint_{\mathcal{R}(\tau'-1,\tau+1)} r^{1+\delta}(\partial^m_{t^*}G_2)^2 + \sup_{t^*\in[\tau'-1,\tau+1]}\sum_{m=0}^{1}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}} (\partial^m_{t^*}G_2)^2\bigg).\end{aligned}$$

Proof. Using Proposition 28 and the fact that ∂_{t^*} is Killing, we immediately have the following estimate for $\partial_{t^*}\Phi$:

$$\begin{aligned}\iint_{\mathcal{R}(\tau',\tau)} r^{-3-\delta}(\partial_{t^*}\Phi)^2 \\ \le C\bigg(\int_{\Sigma_{\tau'}} J^N_\mu(\partial_{t^*}\Phi) n^\mu_{\Sigma_{\tau'}} + \bigg(\int_{\tau'-1}^{\tau+1}\bigg(\int_{\Sigma_{t^*}} (\partial_{t^*}G_1)^2\bigg)^{1/2} dt^*\bigg)^2 + \iint_{\mathcal{R}(\tau'-1,\tau+1)} (\partial_{t^*}G_1)^2 \\ + \sum_{m=1}^{2}\iint_{\mathcal{R}(\tau'-1,\tau+1)} r^{1+\delta}(\partial^m_{t^*}G_2)^2 + \sup_{t^*\in[\tau'-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}} (\partial_{t^*}G_2)^2\bigg).\end{aligned}$$

This would allow us to estimate all derivatives of Φ except for the fact that the estimates for the angular derivatives of Φ degenerate around $r = 3M$:

$$\begin{aligned}&\iint_{\mathcal{R}(\tau',\tau)} (r^{-1-\delta}\mathbb{1}_{\{|r-3M|\geq\frac18 M\}}|\nabla\!\!\!/\Phi|^2 + r^{-1-\delta}(\partial_r\Phi)^2 + r^{-1-\delta}(\partial_{t^*}\Phi)^2 + r^{-3-\delta}\Phi^2)\\
&\quad\leq C\Bigg(\sum_{m=0}^{1}\int_{\Sigma_{\tau'}} J^N_\mu(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau'}} + \sum_{m=0}^{1}\bigg(\int_{\tau'-1}^{\tau+1}\Big(\int_{\Sigma_{t^*}}(\partial_{t^*}^m G_1)^2\Big)^{1/2}dt^*\bigg)^2 + \sum_{m=0}^{1}\iint_{\mathcal{R}(\tau'-1,\tau+1)}(\partial_{t^*}^m G_1)^2\\
&\qquad\qquad + \sum_{m=0}^{2}\iint_{\mathcal{R}(\tau'-1,\tau+1)} r^{1+\delta}(\partial_{t^*}^m G_2)^2 + \sup_{t^*\in[\tau'-1,\tau+1]}\sum_{m=0}^{1}\int_{\Sigma_{t^*}\cap\{|r-3M|\leq\frac18 M\}}(\partial_{t^*}^m G_2)^2\Bigg).\end{aligned}$$

We now use this known estimate and construct another vector field to control the angular derivatives in the region $r\sim 3M$. The argument is simple because the estimate is only local. Take $f_{an}(r)$ to be compactly support in $3M-\frac14 M\leq r\leq 3M+\frac14 M$ and identically equal to -1 in $3M-\frac18 M\leq r\leq 3M+\frac18 M$. If we consider $X_{an}=f_{an}(r)\partial_{r^*}$ in Schwarzschild spacetime, we get that the coefficient in front of the terms with angular derivatives is $\mu/2r$, which is bounded below in $3M-\frac18 M\leq r\leq 3M+\frac18 M$. In other words, one gets an estimate of the form

$$\begin{aligned}&\iint_{\mathcal{R}(\tau',\tau)} r^{-1-\delta}\mathbb{1}_{\{|r-3M|\leq\frac18 M\}}|\nabla\!\!\!/\Phi|^2\\
&\quad\leq C\bigg(\iint_{\mathcal{R}(\tau',\tau)}(r^{-1-\delta}\mathbb{1}_{\{|r-3M|\geq\frac18 M\}}|\nabla\!\!\!/\Phi|^2 + r^{-1-\delta}(\partial_r\Phi)^2 + r^{-1-\delta}(\partial_{t^*}\Phi)^2 + r^{-3-\delta}\Phi^2)\\
&\qquad + \int_{\Sigma_\tau} J^N_\mu(\Phi)n^\mu_{\Sigma_\tau} + \int_{\Sigma_{\tau'}} J^N_\mu(\Phi)n^\mu_{\Sigma_{\tau'}} + \iint_{\mathcal{R}(\tau',\tau)}(|\partial_r\Phi|+|r^{-1}\Phi|)|G| + \iint_{\mathcal{R}(\tau',\tau)} G^2\bigg). \quad (9)\end{aligned}$$

Using a stability argument, (9) would hold also on Kerr spacetimes. One easily checks that the terms with G on the right-hand side can be estimated in the same manner as before. Hence, the proposition can be proved by applying Proposition 28. □

9. Vector field multiplier Z and decay of nondegenerate energy

We follow the definition of Z in [Dafermos and Rodnianski 2008]. Let $Z=u^2\underline{L}+v^2L$, where u and v are the Schwarzschild coordinates $u=\frac12(t-r_S^*)$ and $v=\frac12(t+r_S^*)$, and $\underline{L}=\partial_u$ and $L=2V-\underline{L}$, where $V=\partial_{t^*}+\chi(r)a/(2Mr_+)\partial_{\phi^*}$ with χ being a cutoff function that is identically 1 for $r\leq r_Y^- - \frac12(r_Y^- - r_+)$ and is compactly supported in $\{r\leq r_Y^- - \frac14(r_Y^- - r_+)\}$. With this definition, V is Killing except in the set $\{r_Y^- - \frac12(r_Y^- - r_+)\leq r\leq r_Y^- - \frac14(r_Y^- - r_+)\}$. Let $w^Z=4tr_S^*(1-\mu)/r$. Notice that while $u\to\infty$ as one approaches the event horizon, Z is continuous up to the event horizon due to the following (however, Z is not C^1 and hence its deformation tensor is not continuous up to the event horizon):

Proposition 30. *In the Kerr (t^*, r, θ, ϕ^*) coordinates,*

$$\underline{L}=(1-\mu)\partial_{t^*}-(1-\mu)\Big(\frac{2r_s-2M}{2r-2M}\Big)\partial_r.$$

In the null frame near the event horizon in Section 2.3, we can write

$$\underline{L}=\underline{L}^{\hat V}\hat V+\underline{L}^{\hat Y}\hat Y+\underline{L}^A E_A,\quad \textit{where } |\underline{L}^\alpha|\leq C(1-\mu).$$

Heuristically, we want to show that in the region $\{r \geq r_Y^-\}$,

$$\int_{\Sigma_\tau \cap \{r \geq r_Y^-\}} J_\mu^{Z,w^Z}(\Phi) n_{\Sigma_\tau}^\mu \geq 0.$$

Moreover, we would like to have

$$\int_{\Sigma_\tau \cap \{r \geq r_Y^-\}} J_\mu^{Z,w^Z}(\Phi) n_{\Sigma_\tau}^\mu \geq \int_{\Sigma_\tau \cap \{r \geq r_Y^-\}} u^2(\underline{L}\Phi)^2 + v^2(L\Phi)^2 + (u^2+v^2)|\nabla\!\!\!/ \Phi|^2 + \left(\frac{u^2+v^2}{r^2}\right)\Phi^2$$

These are true modulo some error terms that can be controlled:

Proposition 31. *We have*

$$\int_{\Sigma_\tau \cap \{r \geq r_Y^-\}} u^2(\underline{L}\Phi)^2 + v^2(L\Phi)^2 + (u^2+v^2)|\nabla\!\!\!/ \Phi|^2 + \left(\frac{u^2+v^2}{r^2}\right)\Phi^2$$
$$\leq C \int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\Phi) n_{\Sigma_\tau}^\mu + C \int_{\Sigma_\tau} J_\mu^N(\Phi) n_{\Sigma_\tau}^\mu + C^2\tau^2 \int_{\Sigma_\tau \cap \{r \leq r_Y^-\}} J_\mu^N(\Phi) n_{\Sigma_\tau}^\mu.$$

Proof. The proof is analogous to that in Minkowski spacetime [Morawetz 1975] and Schwarzschild spacetime [Dafermos and Rodnianski 2009]. Recall from the latter that on Schwarzschild spacetime, on a t slice,

$$(J_S^{Z,w^Z})_\mu(\Phi) n_{\Sigma_t}^\mu$$
$$= \frac{1}{\sqrt{1-\mu}}\Big(v^2(L\Phi)^2 + u^2(\underline{L}\Phi)^2 + (u^2+v^2)|\nabla\!\!\!/ \Phi|^2 + \frac{2tr^*(1-\mu)}{r}\Phi\partial_t\Phi - \frac{r^*(1-\mu)}{r}\Phi^2\Big).$$

Now, since t and r^* are stable under perturbation on $\{r \geq r_Y^- - (r_Y^- - r_+)/4\}$, we have, on this set,

$$(J_K^{Z,w^Z})_\mu(\Phi) n_{\Sigma_\tau}^\mu \geq \frac{1}{\sqrt{1-\mu}}\Big(v^2(L\Phi)^2 + u^2(\underline{L}\Phi)^2 + (u^2+v^2)|\nabla\!\!\!/ \Phi|^2 + \frac{2tr^*(1-\mu)}{r}\Phi\partial_{t^*}\Phi - \frac{r^*(1-\mu)}{r}\Phi^2\Big)$$
$$- C\epsilon r^{-2}((u^2+v^2)(\nabla\Phi)^2 + t^*\Phi^2).$$

We now cut off Φ. Define $\hat{\Phi}$ so that it is supported in $\{r \geq r_Y^- - (r_Y^- - r_+)/4\}$ and equals Φ in $\{r \geq r_Y^-\}$. All the error terms arising from the cutoff will be controlled using the red-shift vector field:

$$\int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\hat{\Phi}) n_{\Sigma_\tau}^\mu \geq \int_{\Sigma_\tau \cap \{r \geq r_Y^- - (r_Y^- - r_+)/4\}} \frac{1}{\sqrt{1-\mu}}\Big(v^2(L\hat{\Phi})^2 + u^2(\underline{L}\hat{\Phi})^2 + (u^2+v^2)|\nabla\!\!\!/ \hat{\Phi}|^2\Big)$$
$$+ \frac{2tr_S^*(1-\mu)^{1/2}}{r}\hat{\Phi}\partial_{t^*}\hat{\Phi} - \frac{r_S^*(1-\mu)^{1/2}}{r}\hat{\Phi}^2 - C\epsilon r^{-2}((u^2+v^2)(\nabla\hat{\Phi})^2 + t^*\hat{\Phi}^2).$$

The term

$$\int_{\Sigma_\tau \cap \{r \geq r_Y^- - (r_Y^- - r_+)/4\}} \frac{2tr_S^*(1-\mu)^{1/2}}{r}\hat{\Phi}\partial_{t^*}\hat{\Phi}$$

is to be handled by two different integrations by parts. Recall [Dafermos and Rodnianski 2009] that on Schwarzschild spacetimes we have

$$t\partial_t\hat{\Phi} = vL\hat{\Phi} + u\underline{L}\hat{\Phi} - r_S^*\partial_{r_S^*}\hat{\Phi} \quad \text{and} \quad t\partial_t\hat{\Phi} = \frac{t}{r_S^*}(vL\hat{\Phi} - u\underline{L}\hat{\Phi}) - \frac{t^2}{r_S^*}\partial_{r_S^*}\hat{\Phi}.$$

Therefore, upon integrating by parts, we have on Schwarzschild spacetimes that

$$\begin{aligned}
&\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- -r_+)/4\}} \frac{t r_S^*(1-\mu)^{1/2}}{r}\hat{\Phi}\partial_{t^*}\hat{\Phi} \\
&\quad= \int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- -r_+)/4\}} \Big((1-\mu)r^2\frac{r_S^*}{r}(vL\hat{\Phi}+u\underline{L}\hat{\Phi})\hat{\Phi} + \tfrac{1}{2}\partial_{r_S^*}((1-\mu)r(r^*)^2)\hat{\Phi}^2\Big)\, d\theta\, d\phi\, dr^* \\
&\quad= \int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- -r_+)/4\}} (1-\mu)r^2\Big(\frac{r_S^*}{r}(vL\hat{\Phi}+u\underline{L}\hat{\Phi})\hat{\Phi} + \Big(\frac{1}{2}\frac{(r_S^*)^2}{r^2}+\frac{r_S^*}{r}\Big)\hat{\Phi}^2\Big)\, d\theta\, d\phi\, dr^*.
\end{aligned}$$

Notice that in the equation above, we suppressed the volume form in our notation in the first line, while when we write in coordinates as in the second and the third line, we write out the volume form explicitly. Alternatively, we have

$$\begin{aligned}
&\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- -r_+)/4\}} \frac{t r_S^*(1-\mu)^{1/2}}{r}\hat{\Phi}\partial_{t^*}\hat{\Phi} \\
&\quad= \int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- -r_+)/4\}} \Big((1-\mu)r^2\frac{t^*}{r}(vL\hat{\Phi}-u\underline{L}\hat{\Phi})\hat{\Phi} + \tfrac{1}{2}\partial_{r^*}((1-\mu)r(t^*)^2)\hat{\Phi}^2\Big)\, d\theta\, d\phi\, dr^* \\
&\quad= \int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- -r_+)/4\}} (1-\mu)r^2\Big(\frac{t^*}{r}(vL\hat{\Phi}-u\underline{L}\hat{\Phi})\hat{\Phi} + \frac{1}{2}\frac{(t^*)^2}{r^2}\hat{\Phi}^2\Big)\, d\theta\, d\phi\, dr^*.
\end{aligned}$$

We would like to imitate this integration by parts on Kerr spacetimes. On the domain of integration, we have

$$t\partial_t\hat{\Phi} = vL\hat{\Phi} + u\underline{L}\hat{\Phi} - r_S^*\partial_{r_S^*}\hat{\Phi}, \tag{10}$$

$$t\partial_t\hat{\Phi} = \frac{t}{r_S^*}(vL\hat{\Phi} - u\underline{L}\hat{\Phi}) - \frac{t^2}{r_S^*}\partial_{r_S^*}\hat{\Phi}. \tag{11}$$

The volume form on a constant t^* slice on a Kerr spacetime is close to that on a Schwarzschild spacetime, including in the region being considered. In other words, for $r \geq r_Y^- - (r_Y^- - r_+)/4$,

$$d\,\mathrm{Vol}_{\Sigma_\tau} = (r^2(1-\mu)^{-1/2} + O_1(\epsilon))\, dr\, dx^A d\, x^B.$$

Moreover, for $r \geq r_Y^- - (r_Y^- - r_+)/4$,

$$\partial_{r_S^*} = ((1-\mu) + O_1(\epsilon r^{-2}))\partial_r.$$

Therefore, using (10), we have

$$\begin{aligned}
&\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- - r_+)/4\}} \frac{t r_S^*(1-\mu)^{1/2}}{r}\hat{\Phi}\partial_{t^*}\hat{\Phi}\\
&\quad=\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- - r_+)/4\}} ((r r_S^*+O(\epsilon))(vL\hat{\Phi}+u\underline{L}\hat{\Phi})\hat{\Phi}+\big(\tfrac{1}{2}\partial_r((1-\mu)r(r_S^*)^2)+O(\epsilon)\big)\hat{\Phi}^2)\,dr\,dx^A dx^B\\
&\quad=\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- - r_+)/4\}} (1-\mu)^{1/2}\Big(\Big(\frac{r_S^*}{r}+O(\epsilon r^{-2})\Big)(vL\hat{\Phi}+u\underline{L}\hat{\Phi})\hat{\Phi}+\Big(\frac{1}{2}\frac{(r_S^*)^2}{r^2}+\frac{r_S^*}{r}+O(\epsilon r^{-2})\Big)\hat{\Phi}^2\Big).
\end{aligned}$$

Alternatively, we can integrate by parts after using (11):

$$\begin{aligned}
&\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- - r_+)/4\}} \frac{t r_S^*(1-\mu)^{1/2}}{r}\hat{\Phi}\partial_{t^*}\hat{\Phi}\\
&\quad=\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- - r_+)/4\}} ((r t^*+O(\epsilon))(vL\hat{\Phi}-u\underline{L}\hat{\Phi})\hat{\Phi}+\big(\tfrac{1}{2}\partial_r((1-\mu)r(t^*)^2)+O(\epsilon)\big)\hat{\Phi}^2)\\
&\quad=\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- - r_+)/4\}} (1-\mu)^{1/2}\Big(\Big(\frac{t^*}{r}+O(\epsilon r^{-2})\Big)(vL\hat{\Phi}-u\underline{L}\hat{\Phi})\hat{\Phi}+\Big(\frac{1}{2}\frac{(t^*)^2}{r^2}+O(\epsilon r^{-2})\Big)\hat{\Phi}^2\Big).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\hat{\Phi})n_{\Sigma_\tau}^\mu\\
&\quad\geq\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- - r_+)/4\}} \frac{1}{\sqrt{1-\mu}}(v^2(L\hat{\Phi})^2+u^2(\underline{L}\hat{\Phi})^2+(u^2+v^2)|\nabla\!\!\!\!/\,\hat{\Phi}|^2)\\
&\qquad\qquad+\frac{2t r_S^*(1-\mu)^{1/2}}{r}\hat{\Phi}\partial_{t^*}\hat{\Phi}-\frac{r_S^*(1-\mu)^{1/2}}{r}\hat{\Phi}^2-C\epsilon r^{-2}((u^2+v^2)(D\hat{\Phi})^2+t^*\hat{\Phi}^2)\\
&\quad\geq\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- - r_+)/4\}} \frac{1}{\sqrt{1-\mu}}(v^2(L\hat{\Phi})^2+u^2(\underline{L}\hat{\Phi})^2+(u^2+v^2)|\nabla\!\!\!\!/\,\hat{\Phi}|^2)\\
&\qquad\qquad+\frac{r_S^*(1-\mu)^{1/2}}{r}(vL\hat{\Phi}+u\underline{L}\hat{\Phi})\hat{\Phi}+\frac{1}{2}\frac{(r_S^*)^2(1-\mu)^{1/2}}{r^2}\hat{\Phi}^2\\
&\qquad\qquad+\frac{t^*(1-\mu)^{1/2}}{r}(vL\hat{\Phi}-u\underline{L}\hat{\Phi})\hat{\Phi}+\frac{1}{2}\frac{(t^*)^2}{r^2}\hat{\Phi}^2-C\epsilon r^{-2}((u^2+v^2)(D\hat{\Phi})^2+t^*\hat{\Phi}^2)\\
&\quad\geq c\Big(\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- - r_+)/4\}} \mu((vL\hat{\Phi}+u\underline{L}\hat{\Phi})^2+(vL\hat{\Phi}-u\underline{L}\hat{\Phi})^2)\\
&\qquad\qquad+(1-\mu)((vL\hat{\Phi}+u\underline{L}\hat{\Phi}+\frac{r_S^*}{r}\hat{\Phi})^2+(vL\hat{\Phi}-u\underline{L}\hat{\Phi}+\frac{t^*}{r}\hat{\Phi})^2+2(u^2+v^2)|\nabla\!\!\!\!/\,\hat{\Phi}|^2)\\
&\qquad\qquad-C\epsilon r^{-2}((u^2+v^2)(D\hat{\Phi})^2+t^*\hat{\Phi}^2)\Big),
\end{aligned}\tag{12}$$

where the last line is obtained by first completing the square and using $c\leq 1-\mu\leq C$ in this region of r. Let us for now ignore the error term and look at the other terms (which are manifestly positive). By exactly the same argument as in [Dafermos and Rodnianski 2009], these positive terms provide good estimates:

$$\begin{aligned}\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- -r_+)/4\}} &\mu((vL\hat{\Phi}+u\underline{L}\hat{\Phi})^2+(vL\hat{\Phi}-u\underline{L}\hat{\Phi})^2)\\ &+(1-\mu)\Big(\Big(vL\hat{\Phi}+u\underline{L}\hat{\Phi}+\frac{r_S^*}{r}\hat{\Phi}\Big)^2+(vL\hat{\Phi}-u\underline{L}\hat{\Phi}+\frac{t^*}{r}\hat{\Phi})^2+2(u^2+v^2)|\nabla\!\!\!\!/\,\hat{\Phi}|^2\Big)\\ \geq c\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- -r_+)/4\}}&\Big(v^2(L\hat{\Phi})^2+u^2(\underline{L}\hat{\Phi})^2+(u^2+v^2)|\nabla\!\!\!\!/\,\hat{\Phi}|^2+\frac{u^2+v^2}{r^2}\hat{\Phi}^2\Big)\\ &+C\epsilon\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- -r_+)/4\}} r^{-2}((u^2+v^2)(D\hat{\Phi})^2+t^*\hat{\Phi}^2).\end{aligned}$$

See [Dafermos and Rodnianski 2009] for the proof. This together with $J_\mu^N(\hat{\Phi})n^\mu_{\Sigma_\tau}$ bounds the error term in (12):

$$\begin{aligned}\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- -r_+)/4\}} &r^{-2}((u^2+v^2)(D\hat{\Phi})^2+\tau\Phi^2)\\ \leq C\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- -r_+)/4\}}&\Big(v^2(L\hat{\Phi})^2+u^2(\underline{L}\hat{\Phi})^2+(u^2+v^2)|\nabla\!\!\!\!/\,\hat{\Phi}|^2+\frac{u^2+v^2}{r^2}\hat{\Phi}^2\Big)\\ &+C\int_{\Sigma_\tau\cap\{r\geq\tau/4\}}(\underline{L}\hat{\Phi})^2\\ \leq C\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- -r_+)/4\}}&\Big(v^2(L\hat{\Phi})^2+u^2(\underline{L}\hat{\Phi})^2+(u^2+v^2)|\nabla\!\!\!\!/\,\hat{\Phi}|^2+\frac{u^2+v^2}{r^2}\hat{\Phi}^2\Big)\\ &+C\int_{\Sigma_\tau}J_\mu^N(\hat{\Phi})n^\mu_{\Sigma_\tau}\\ \leq C\Big(\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- -r_+)/4\}}&\mu((vL\hat{\Phi}+u\underline{L}\hat{\Phi})^2+(vL\hat{\Phi}-u\underline{L}\hat{\Phi})^2)\\ &+(1-\mu)\Big(\Big(vL\hat{\Phi}+u\underline{L}\hat{\Phi}+\frac{r_S^*}{r}\hat{\Phi}\Big)^2+\Big(vL\hat{\Phi}-u\underline{L}\hat{\Phi}+\frac{t^*}{r}\hat{\Phi}\Big)^2+2(u^2+v^2)|\nabla\!\!\!\!/\,\hat{\Phi}|^2\Big)\Big)\\ &+C\int_{\Sigma_\tau}J_\mu^N(\hat{\Phi})n^\mu_{\Sigma_\tau}.\end{aligned}$$

Therefore, if ϵ is chosen to be small enough, then (12) implies that

$$\begin{aligned}\int_{\Sigma_\tau}J_\mu^{Z,w^Z}(\hat{\Phi})n^\mu_{\Sigma_\tau}&+\int_{\Sigma_\tau}J_\mu^N(\hat{\Phi})n^\mu_{\Sigma_\tau}\\ &\geq c\int_{\Sigma_\tau\cap\{r\geq r_Y^- -(r_Y^- -r_+)/4\}}\Big(v^2(L\hat{\Phi})^2+u^2(\underline{L}\hat{\Phi})^2+(u^2+v^2)|\nabla\!\!\!\!/\,\hat{\Phi}|^2+\frac{u^2+v^2}{r^2}\hat{\Phi}^2\Big).\end{aligned}\tag{13}$$

We note that c here is independent of the choice of r_Y^-. With this bound we would like to estimate $\int_{\mathbb{S}^2}\Phi(\tau,r)^2$. Using (13), there exists a $\tilde{r}\in[r_Y^-,r_Y^-+1]$ such that

$$\int_{\mathbb{S}^2}\Phi(\tau,\tilde{r})^2=\int_{\mathbb{S}^2}\hat{\Phi}(\tau,\tilde{r})^2\leq C\tau^{-2}\Big(\int_{\Sigma_\tau}J_\mu^{Z,w^Z}(\hat{\Phi})n^\mu_{\Sigma_\tau}+\int_{\Sigma_\tau}J_\mu^N(\hat{\Phi})n^\mu_{\Sigma_\tau}\Big).$$

Then for every $r \in [r_+, r_Y^-]$, since $\Phi(\tau, \tilde{r}) - \Phi(\tau, r) = \int_r^{\tilde{r}} \partial_r \Phi dr$, we have

$$\begin{aligned}\int_{\mathbb{S}^2} \Phi(\tau, r)^2 &\le \int_{\mathbb{S}^2} \Phi(\tau, \tilde{r})^2 + (\tilde{r} - r) \int_{\Sigma_\tau \cap [r, \tilde{r}]} J_\mu^N(\Phi) n_{\Sigma_\tau}^\mu \\ &\le C \int_{\Sigma_\tau \cap \{r \le r_Y^-\}} J_\mu^N(\Phi) n_{\Sigma_\tau}^\mu + C\tau^{-2} \Big(\int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\hat{\Phi}) n_{\Sigma_\tau}^\mu + \int_{\Sigma_\tau} J_\mu^N(\hat{\Phi}) n_{\Sigma_\tau}^\mu \Big)\end{aligned} \tag{14}$$

Now we need to obtain estimates for Φ from that for $\hat{\Phi}$. It is obvious that

$$\begin{aligned}&\int_{\Sigma_\tau \cap \{r \ge r_Y^-\}} \Big(v^2 (L\Phi)^2 + u^2 (\underline{L}\Phi)^2 + (u^2 + v^2) |\nabla\!\!\!/ \Phi|^2 + \frac{u^2 + v^2}{r^2} \Phi^2 \Big) \\ &\quad \le \int_{\Sigma_\tau \cap \{r \ge r_Y^-\}} \Big(v^2 (L\hat{\Phi})^2 + u^2 (\underline{L}\hat{\Phi})^2 + (u^2 + v^2) |\nabla\!\!\!/ \hat{\Phi}|^2 + \frac{u^2 + v^2}{r^2} \hat{\Phi}^2 \Big) \\ &\quad \le \int_{\Sigma_\tau \cap \{r \ge r_Y^- - (r_Y^- - r_+)/4\}} \Big(v^2 (L\hat{\Phi})^2 + u^2 (\underline{L}\hat{\Phi})^2 + (u^2 + v^2) |\nabla\!\!\!/ \hat{\Phi}|^2 + \frac{u^2 + v^2}{r^2} \hat{\Phi}^2 \Big),\end{aligned}$$

and

$$\begin{aligned}&\int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\hat{\Phi}) n_{\Sigma_\tau}^\mu + \int_{\Sigma_\tau} J_\mu^N(\hat{\Phi}) n_{\Sigma_\tau}^\mu \\ &\quad \le \int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\Phi) n_{\Sigma_\tau}^\mu + \int_{\Sigma_\tau} J_\mu^N(\Phi) n_{\Sigma_\tau}^\mu + C\tau^2 \int_{\Sigma_\tau \cap \{r \le r_Y^-\}} J_\mu^N(\Phi) n_{\Sigma_\tau}^\mu + C\tau^2 \int_{\Sigma_\tau \cap \{r \le r_Y^-\}} \Phi^2\end{aligned}$$

(where we have used Proposition 30 to show that the u^2 factor comes with a factor of $1 - \mu$)

$$\begin{aligned}&\quad \le \int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\Phi) n_{\Sigma_\tau}^\mu + \int_{\Sigma_\tau} J_\mu^N(\Phi) n_{\Sigma_\tau}^\mu + C\tau^2 \int_{\Sigma_\tau \cap \{r \le r_Y^-\}} J_\mu^N(\Phi) n_{\Sigma_\tau}^\mu \\ &\qquad\qquad + C(r_Y^- - r_+) \Big(\int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\hat{\Phi}) n_{\Sigma_\tau}^\mu + \int_{\Sigma_\tau} J_\mu^N(\hat{\Phi}) n_{\Sigma_\tau}^\mu \Big) \\ &\quad \le \int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\Phi) n_{\Sigma_\tau}^\mu + \int_{\Sigma_\tau} J_\mu^N(\Phi) n_{\Sigma_\tau}^\mu + C\tau^2 \int_{\Sigma_\tau \cap \{r \le r_Y^-\}} J_\mu^N(\Phi) n_{\Sigma_\tau}^\mu \\ &\qquad\qquad + \frac{1}{2} \Big(\int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\hat{\Phi}) n_{\Sigma_\tau}^\mu + \int_{\Sigma_\tau} J_\mu^N(\hat{\Phi}) n_{\Sigma_\tau}^\mu \Big)\end{aligned} \tag{15}$$

for r_Y^- chosen to be sufficiently close to r_+. Then

$$\begin{aligned}&\int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\hat{\Phi}) n_{\Sigma_\tau}^\mu + \int_{\Sigma_\tau} J_\mu^N(\hat{\Phi}) n_{\Sigma_\tau}^\mu \\ &\qquad \le \int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\Phi) n_{\Sigma_\tau}^\mu + \int_{\Sigma_\tau} J_\mu^N(\Phi) n_{\Sigma_\tau}^\mu + C\tau^2 \int_{\Sigma_\tau \cap \{r \le r_Y^-\}} J_\mu^N(\Phi) n_{\Sigma_\tau}^\mu. \quad \square\end{aligned}$$

Remark. From this point onward, we consider r_Y^- to be fixed. We note again that r_Y^- is chosen so that (5) and (15) hold.

Remark. The proof of the proposition above in particular shows that

$$\int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\Phi)n^\mu_{\Sigma_\tau} + C\tau^2 \int_{\Sigma_\tau\cap\{r\le r_Y^-\}} J_\mu^N(\Phi)n^\mu_{\Sigma_\tau} \ge 0.$$

To use this proposition, it is helpful to have a localized version of Φ. This follows [Dafermos and Rodnianski 2008; 2009]. The idea is to use the finite speed of propagation and cutoff Φ outside the domain of dependence. Focus now on the time interval $[\tau',\tau]$. Take $\tilde{G}$ to be any smooth function agreeing with G on the domain of dependence of the region $(t^*=\tau, r\le\tau/2)$. Let $\tilde{\Phi}(\tau')=\chi\Phi(\tau')$ and $\partial_{t^*}\tilde{\Phi}(\tau')=\chi\partial_{t^*}\Phi(\tau')$, where χ is a cutoff function identically equal to 1 for $r\le\frac{7}{10}\tau'$ and compactly supported in $r\le\frac{9}{10}\tau'$. The region for which χ is one is inside the domain of dependence of the region $(t^*=\tau, r\le\tau/2)$ if $\tau'\le\tau\le(1.1)\tau'$. We solve for $\Box_{g_K}\tilde{\Phi}=\tilde{G}$.

With this definition of $\tilde{\Phi}$, we have two ways to estimate the nondegenerate energy of $\tilde{\Phi}$:

Proposition 32. *We have*

$$\int_{\Sigma_{\tau'}} J_\mu^N(\tilde{\Phi})n^\mu_{\Sigma_{\tau'}} \le C\int_{\Sigma_{\tau'}} J_\mu^N(\Phi)n^\mu_{\Sigma_{\tau'}},$$

$$\int_{\Sigma_{\tau'}} J_\mu^N(\tilde{\Phi})n^\mu_{\Sigma_{\tau'}} \le C^2\int_{\Sigma_{\tau'}\cap\{r\le r_Y^-\}} J_\mu^N(\Phi)n^\mu_{\Sigma_{\tau'}} + C(\tau')^{-2}\left(\int_{\Sigma_{\tau'}} J_\mu^N(\Phi)n^\mu_{\Sigma_{\tau'}} + \int_{\Sigma_{\tau'}} J_\mu^{Z,w^Z}(\Phi)n^\mu_{\Sigma_{\tau'}}\right).$$

Proof. The first part is an easy application of Proposition 21:

$$\begin{aligned}\int_{\Sigma_{\tau'}} J_\mu^N(\tilde{\Phi})n^\mu_{\Sigma_{\tau'}} &\le C\int_{\Sigma_{\tau'}\cap\{R\le r\le\frac{9}{10}\tau'\}} ((D\Phi)^2+(\tau')^{-2}\Phi^2)\\ &\le C\int_{\Sigma_{\tau'}\cap\{R\le r\le\frac{9}{10}\tau'\}} ((D\Phi)^2+r^{-2}\Phi^2) \le C\int_{\Sigma_{\tau'}} J_\mu^N(\Phi)n^\mu_{\Sigma_{\tau'}}.\end{aligned}$$

Following (14), we have

$$\int_{\Sigma_{\tau'}\cap\{r\le r_Y^-\}} \Phi^2 \le C\left(\int_{\Sigma_{\tau'}\cap\{r\le r_Y^-\}} J_\mu^N(\Phi)n^\mu_{\Sigma_{\tau'}} + \int_{\Sigma_{\tau'}\cap\{r_Y^-\le r\le r_Y^+\}} \Phi^2\right).$$

Using this and Proposition 31, we have

$$\begin{aligned}&\int_{\Sigma_{\tau'}} J_\mu^N(\tilde{\Phi})n^\mu_{\Sigma_{\tau'}}\\ &\quad\le C\int_{\Sigma_{\tau'}\cap\{r\le\frac{9}{10}\tau'\}} ((D\Phi)^2+(\tau')^{-2}\Phi^2)\\ &\quad\le C\int_{\Sigma_{\tau'}\cap\{r\le r_Y^-\}} ((D\Phi)^2+\Phi^2)\\ &\qquad + C(\tau')^{-2}\int_{\Sigma_{\tau'}\cap\{r_Y^-\le r\le\frac{9}{10}\tau'\}} \left(u^2(\underline{L}\Phi)^2+v^2(L\Phi)^2+(u^2+v^2)|\nabla\!\!\!/\Phi|^2+\left(\frac{u^2+v^2}{r^2}\right)\Phi^2\right)\\ &\quad\le C^2\int_{\Sigma_{\tau'}\cap\{r\le r_Y^-\}} J_\mu^N(\Phi)n^\mu_{\Sigma_{\tau'}} + C(\tau')^{-2}\left(\int_{\Sigma_{\tau'}} J_\mu^N(\Phi)n^\mu_{\Sigma_{\tau'}} + \int_{\Sigma_{\tau'}} J_\mu^{Z,w^Z}(\Phi)n^\mu_{\Sigma_{\tau'}}\right).\end{aligned}$$

□

The cutoff procedure above will also allow us to localize the estimates for the bulk term:

Proposition 33. *Let $G = G_1 + G_2$ be any way to decompose the function G. Then for $\tau' \le \tau \le (1.1)\tau'$, we have*

(1) *the localized boundedness estimate*

$$\begin{aligned}
&\int_{\Sigma_\tau \cap \{r \le \tau/2\}} J^N_\mu(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J^N_\mu(\Phi) n^\mu_{\mathcal{H}^+} + \iint_{\mathcal{R}(\tau',\tau)\cap\{r \le r_Y^-\}} K^N(\Phi) + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le \frac{1}{2}t^*\}} K^{X_0}(\Phi) \\
&\quad \le C\bigg(\int_{\Sigma_{\tau'}} J^N_\mu(\Phi) n^\mu_{\Sigma_{\tau'}} + \bigg(\int_{\tau'-1}^{\tau+1}\bigg(\int_{\Sigma_{t^*}\cap\{r\le\frac{9}{10}t^*\}} G_1^2\bigg)^{1/2} dt^*\bigg)^2 \\
&\qquad + \iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\le \frac{9}{10}t^*\}} G_1^2 + \sum_{m=0}^{1} \iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}} r^{1+\delta}(\partial^m_{t^*} G_2)^2 \\
&\qquad + \sup_{t^*\in[\tau'-1,\tau+1]} \int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}\cap\{r\le\frac{9}{10}t^*\}} {G_2}^2\bigg);
\end{aligned}$$

(2) *the localized decay estimate*

$$\begin{aligned}
&\int_{\Sigma_\tau \cap \{r \le \frac{1}{2}\tau\}} J^N_\mu(\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J^N_\mu(\Phi) n^\mu_{\mathcal{H}^+} + \iint_{\mathcal{R}(\tau',\tau)\cap\{r \le r_Y^-\}} K^N(\Phi) + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le \frac{1}{2}t^*\}} K^{X_0}(\Phi) \\
&\quad \le C\bigg(\tau^{-2}\int_{\Sigma_{\tau'}} J^{Z+N,w^Z}_\mu(\Phi) n^\mu_{\Sigma_{\tau'}} + C\int_{\Sigma_{\tau'}\cap\{r\le r_Y^-\}} J^N_\mu(\Phi) n^\mu_{\Sigma_{\tau'}}\bigg) \\
&\qquad + C\bigg(\bigg(\int_{\tau'-1}^{\tau+1}\bigg(\int_{\Sigma_{t^*}\cap\{r\le\frac{9}{10}t^*\}} G_1^2\bigg)^{1/2} dt^*\bigg)^2 + \iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\le \frac{9}{10}t^*\}} G_1^2 \\
&\qquad + \sum_{m=0}^{1} \iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}} r^{1+\delta}(\partial^m_{t^*} G_2)^2 + \sup_{t^*\in[\tau'-1,\tau+1]} \int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}\cap\{r\le\frac{9}{10}t^*\}} G_2^2\bigg).
\end{aligned}$$

Proof. Applying Proposition 28 to the equation $\Box_{g_K}\tilde{\Phi} = \tilde{G}$, we have

$$\begin{aligned}
&\int_{\Sigma_\tau \cap \{r \le \frac{1}{2}\tau\}} J^N_\mu(\tilde{\Phi}) n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J^N_\mu(\tilde{\Phi}) n^\mu_{\mathcal{H}^+} + \iint_{\mathcal{R}(\tau',\tau)\cap\{r \le r_Y^-\}} K^N(\tilde{\Phi}) + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le \frac{1}{2}t^*\}} K^{X_0}(\tilde{\Phi}) \\
&\quad \le C\bigg(\int_{\Sigma_{\tau'}} J^N_\mu(\tilde{\Phi}) n^\mu_{\Sigma_{\tau'}} + \bigg(\int_{\tau'-1}^{\tau+1}\bigg(\int_{\Sigma_{t^*}} \tilde{G}_1^2\bigg)^{1/2} dt^*\bigg)^2 + \iint_{\mathcal{R}(\tau'-1,\tau+1)} \tilde{G}_1^2 \\
&\qquad + \sum_{m=0}^{1} \iint_{\mathcal{R}(\tau'-1,\tau+1)} r^{1+\delta}(\partial^m_{t^*} \tilde{G}_2)^2 + \sup_{t^*\in[\tau'-1,\tau+1]} \int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}} \tilde{G}_2^2\bigg).
\end{aligned}$$

Since by the finite speed of propagation, $\tilde{\Phi}=\Phi$ in $\{r\le \frac{1}{2}t^*\}$, we have

$$\begin{aligned}\int_{\Sigma_\tau\cap\{r\le\frac12\tau\}} J^N_\mu(\Phi)n^\mu_{\Sigma_\tau}+\int_{\mathcal{H}(\tau',\tau)} J^N_\mu(\Phi)n^\mu_{\mathcal{H}^+}+\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} K^N(\Phi)+\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le\frac12 t^*\}} K^{X_0}(\Phi)\\ \le C\bigg(\int_{\Sigma_{\tau'}} J^N_\mu(\tilde{\Phi})n^\mu_{\Sigma_{\tau'}}+\bigg(\int_{\tau'-1}^{\tau+1}\bigg(\int_{\Sigma_{t^*}}\tilde{G}_1^2\bigg)^{1/2}dt^*\bigg)^2+\iint_{\mathcal{R}(\tau'-1,\tau+1)}\tilde{G}_1^2\\ +\sum_{m=0}^{1}\iint_{\mathcal{R}(\tau'-1,\tau+1)} r^{1+\delta}(\partial_{t^*}^m\tilde{G}_2)^2+\sup_{t^*\in[\tau'-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac18 M\}}\tilde{G}_2^2\bigg).\end{aligned}$$

Now, we choose a particular $\tilde{G}$. Define $\tilde{G}$ to be G for $r\le\frac{7}{10}t^*$, and 0 for $r\ge\frac{9}{10}t^*$. It can be easily shown that one has the bounds

$$|\partial_{t^*}^m\tilde{G}|\le C\sum_{k=0}^{m}\bigg|\Big(\frac{r^*}{(t^*)^2}\Big)^k\partial_{t^*}^{m-k}G\bigg|\le C\sum_{k=0}^{m}|(t^*)^{-k}\partial_{t^*}^{m-k}G|\quad\text{for }\tfrac{7}{10}t^*\le r\le\tfrac{9}{10}t^*.$$

Therefore, we have

$$\begin{aligned}\int_{\Sigma_\tau\cap\{r\le\frac12\tau\}} J^N_\mu(\Phi)n^\mu_{\Sigma_\tau}+\int_{\mathcal{H}(\tau',\tau)} J^N_\mu(\Phi)n^\mu_{\mathcal{H}^+}+\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} K^N(\Phi)+\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le\frac12 t^*\}} K^{X_0}(\Phi)\\ \le C\bigg(\int_{\Sigma_{\tau'}} J^N_\mu(\tilde{\Phi})n^\mu_{\Sigma_{\tau'}}+\Big(\int_{\tau'-1}^{\tau+1}\Big(\int_{\Sigma_{t^*}\cap\{r\le\frac{9}{10}t^*\}}G_1^2\Big)^{1/2}dt^*\Big)^2+\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}}G_1^2\\ +\sum_{m=0}^{1}\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}} r^{1+\delta}(\partial_{t^*}^m G_2)^2+\sup_{t^*\in[\tau'-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac18 M\}}G_2^2\bigg).\end{aligned}$$

We can now conclude the proposition using Proposition 32. □

We can remove the degeneracy around $r\sim 3M$ using an extra derivative.

Proposition 34. *Let $G=G_1+G_2$ be any way to decompose the function G. Then for $\tau'\le\tau\le(1.1)\tau'$, we have*

(1) *the localized boundedness estimate*

$$\begin{aligned}&\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le\frac12 t^*\}} K^{X_1}(\Phi)\\ &\le C\bigg(\sum_{m=0}^{1}\int_{\Sigma_{\tau'}} J^N_\mu(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau'}}+\sum_{m=0}^{1}\Big(\int_{\tau'-1}^{\tau+1}\Big(\int_{\Sigma_{t^*}\cap\{r\le\frac{9}{10}t^*\}}(\partial_{t^*}^m G_1)^2\Big)^{1/2}dt^*\Big)^2\\ &\quad+\sum_{m=0}^{1}\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}}(\partial_{t^*}^m G_1)^2+\sum_{m=0}^{2}\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}} r^{1+\delta}(\partial_{t^*}^m G_2)^2\\ &\quad+\sup_{t^*\in[\tau'-1,\tau+1]}\sum_{m=0}^{1}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac18 M\}\cap\{r\le\frac{9}{10}t^*\}}(\partial_{t^*}^m G_2)^2\bigg);\end{aligned}$$

(2) *the localized decay estimate*

$$\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le\frac{1}{2}t^*\}} K^{X_1}(\Phi)$$
$$\le C\Big(\tau^{-2}\sum_{m=0}^{1}\int_{\Sigma_{\tau'}} J_\mu^{Z+N,w^Z}(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau'}} + C\sum_{m=0}^{1}\int_{\Sigma_{\tau'}\cap\{r\le r_Y^-\}} J_\mu^N(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau'}}\Big)$$
$$+C\Big(\sum_{m=0}^{1}\Big(\int_{\tau'-1}^{\tau+1}\Big(\int_{\Sigma_{t^*}\cap\{r\le\frac{9}{10}t^*\}}(\partial_{t^*}^m G_1)^2\Big)^{1/2}dt^*\Big)^2+\sum_{m=0}^{1}\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}}(\partial_{t^*}^m G_1)^2$$
$$+\sum_{m=0}^{2}\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}} r^{1+\delta}(\partial_{t^*}^m G_2)^2+\sup_{t^*\in[\tau'-1,\tau+1]}\sum_{m=0}^{1}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}\cap\{r\le\frac{9}{10}t^*\}}(\partial_{t^*}^m G_2)^2\Big).$$

Proof. We repeat the argument in Proposition 33, using Proposition 29 instead of 28. □

When using the conservation law for Z, we can ignore the part of the bulk term that has a good sign.

Definition 35. Let $K_+^{Z,w^Z}(\Phi)=\max\{K^{Z,w^Z}(\Phi),0\}$.

Using the conservation law for the modified vector field, we have a one-sided bound:

Proposition 36.

$$\int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\Phi)n^\mu_{\Sigma_\tau}+\int_{\mathcal{H}(\tau',\tau)} J_\mu^{Z,w^Z}(\Phi)n^\mu_{\mathcal{H}^+}$$
$$\le C(\tau')^2\int_{\Sigma_{\tau'}\cap\{r\le r_Y^-\}} J_\mu^N(\Phi)n^\mu_{\Sigma_{\tau'}}+\int_{\Sigma_{\tau'}} J_\mu^{Z,w^Z}(\Phi)n^\mu_{\Sigma_{\tau'}}$$
$$+\iint_{\mathcal{R}(\tau',\tau)} K_+^{Z,w^Z}(\Phi)+\Big|\iint_{\mathcal{R}(\tau',\tau)}(u^2L\Phi+v^2\underline{L}\Phi-\tfrac{1}{4}w\Phi)G\Big|.$$

Remark. In the proposition, the left-hand side is not claimed to be positive. Note, however, that the right-hand side is positive by the remark on page 591.

Remark. We note also that

$$\int_{\mathcal{H}(\tau',\tau)} J_\mu^{Z,w^Z}(\Phi)n^\mu_{\mathcal{H}^+}\ge 0$$

because Z and $n^\mu_{\mathcal{H}^+}$ are both null and future directed and $w^Z=0$ on the event horizon.

To show that $\int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\Phi)n^\mu_{\Sigma_\tau}$ is almost bounded, we will have to show that

$$\int_{\Sigma_{\tau'}\cap\{r\le r_Y^-\}} J_\mu^N(\Phi)n^\mu_{\Sigma_{\tau'}}$$

in fact decays:

Proposition 37.

$$
\begin{aligned}
&\int_{\Sigma_\tau\cap\{r\le\frac{1}{2}t^*\}} J^N_\mu(\Phi)n^\mu_{\Sigma_\tau}\\
&\quad\le C^2\tau^{-2}\int_{\Sigma_{(1.1)^{-2}\tau}\cap\{r\le r^-_Y\}} J^N_\mu(\Phi)n^\mu_{\Sigma_{(1.1)^{-2}\tau}}+C\tau^{-2}\int_{\Sigma_{(1.1)^{-2}\tau}} J^{Z,w^Z}_\mu(\Phi)n^\mu_{\Sigma_{(1.1)^{-1}\tau}}\\
&\qquad+C\tau^{-2}\iint_{\mathcal{R}((1.1)^{-2}\tau,\tau)} K^{Z,w^Z}_+(\Phi)+C\tau^{-2}\Big|\iint_{\mathcal{R}((1.1)^{-2}\tau,\tau)}(u^2L\Phi+v^2\underline{L}\Phi-\tfrac{1}{4}w\Phi)G\Big|\\
&\qquad+C\bigg(\Big(\int_{(1.1)^{-2}\tau-1}^{\tau+1}\Big(\int_{\Sigma_{t^*}\cap\{r\le\frac{9}{10}t^*\}}G_1^2\Big)^{1/2}dt^*\Big)^2+\iint_{\mathcal{R}((1.1)^{-2}\tau-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}}G_1^2\bigg)\\
&\quad+C\bigg(\sum_{m=0}^{1}\iint_{\mathcal{R}((1.1)^{-2}\tau-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}} r^{1+\delta}(\partial^m_{t^*}G_2)^2+\sup_{t^*\in[(1.1)^{-2}\tau-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}\cap\{r\le\frac{9}{10}t^*\}}G_2^2\bigg).
\end{aligned}
$$

Proof. By Proposition 33(2) applied to the t^* interval $[(1.1)^{-1}\tau,\tau]$, we have

$$
\begin{aligned}
\iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)\cap\{r\le r^-_Y\}} K^N(\Phi)\le C\tau^{-2}\int_{\Sigma_{(1.1)^{-1}\tau}} J^{Z+N,w^Z}_\mu(\Phi)n^\mu_{\Sigma_{(1.1)^{-1}\tau}}+C^2\int_{\Sigma_{(1.1)^{-1}\tau}\cap\{r\le r^-_Y\}} J^N_\mu(\Phi)n^\mu_{\Sigma_{(1.1)^{-1}\tau}}\\
+C\bigg(\sum_{m=0}^{1}\iint_{\mathcal{R}((1.1)^{-1}\tau-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}} r^{1+\delta}(\partial^m_{t^*}G)^2+\sup_{t^*\in[(1.1)^{-1}\tau-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}}G^2\bigg).
\end{aligned}
$$

By taking the infimum there exists $\tilde\tau\in[(1.1)^{-1}\tau,\tau]$ such that

$$
\int_{\Sigma_{\tilde\tau}\cap\{r\le r^-_Y\}} J^N_\mu(\Phi)n^\mu_{\Sigma_{\tilde\tau}}\le C\tau^{-1}\iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)\cap\{r\le r^-_Y\}} K^N(\Phi).
$$

Hence,

$$
\begin{aligned}
&\int_{\Sigma_{\tilde\tau}\cap\{r\le r^-_Y\}} J^N_\mu(\Phi)n^\mu_{\Sigma_{\tilde\tau}}\\
&\quad\le C\tau^{-2}\int_{\Sigma_{(1.1)^{-1}\tau}} J^{Z+N,w^Z}_\mu(\Phi)n^\mu_{\Sigma_{(1.1)^{-1}\tau}}+C^2\tau^{-1}\int_{\Sigma_{(1.1)^{-1}\tau}\cap\{r\le r^-_Y\}} J^N_\mu(\Phi)n^\mu_{\Sigma_{(1.1)^{-1}\tau}}\\
&\qquad+C\tau^{-1}\bigg(\sum_{m=0}^{1}\iint_{\mathcal{R}((1.1)^{-1}\tau-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}} r^{1+\delta}(\partial^m_{t^*}G)^2+\sup_{t^*\in[(1.1)^{-1}\tau-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}}G^2\bigg).
\end{aligned}
$$

We apply Proposition 33(2) to the t^* interval $[\tilde\tau,\tau]$ and use Proposition 28 and 36, getting

$$
\begin{aligned}
&\int_{\Sigma_\tau\cap\{r\le\frac{1}{2}t^*\}} J^N_\mu(\Phi)n^\mu_{\Sigma_\tau}\\
&\le C\Big(\tau^{-2}\int_{\Sigma_{\tilde\tau}} J^{Z+N,w^Z}_\mu(\Phi)n^\mu_{\Sigma_{\tilde\tau}}+C\int_{\Sigma_{\tilde\tau}\cap\{r\le r_Y^-\}} J^N_\mu(\Phi)n^\mu_{\Sigma_{\tilde\tau}}\Big)\\
&\quad+C\Big(\sum_{m=0}^{1}\iint_{\mathcal{R}((1.1)^{-1}\tau-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}} r^{1+\delta}(\partial^m_{t^*}G)^2+\sup_{t^*\in[(1.1)^{-1}\tau-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}\cap\{r\le\frac{9}{10}t^*\}} G^2\Big)\\
&\le C\tau^{-2}\Big(\int_{\Sigma_{\tilde\tau}} J^{Z+N,w^Z}_\mu(\Phi)n^\mu_{\Sigma_{\tilde\tau}}+\int_{\Sigma_{(1.1)^{-1}\tau}} J^{Z+N,w^Z}_\mu(\Phi)n^\mu_{\Sigma_{(1.1)^{-1}\tau}}\Big)\\
&\quad+C^2\tau^{-1}\int_{\Sigma_{(1.1)^{-1}\tau}\cap\{r\le r_Y^-\}} J^N_\mu(\Phi)n^\mu_{\Sigma_{(1.1)^{-1}\tau}}\\
&\quad+C\Big(\sum_{m=0}^{1}\iint_{\mathcal{R}((1.1)^{-1}\tau-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}} r^{1+\delta}(\partial^m_{t^*}G)^2+\sup_{t^*\in[(1.1)^{-1}\tau-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}\cap\{r\le\frac{9}{10}t^*\}} G^2\Big)\\
&\le C^2\tau^{-1}\int_{\Sigma_{(1.1)^{-1}\tau}\cap\{r\le r_Y^-\}} J^N_\mu(\Phi)n^\mu_{\Sigma_{(1.1)^{-1}\tau}}+C\tau^{-2}\int_{\Sigma_{(1.1)^{-1}\tau}} J^{Z+N,w^Z}_\mu(\Phi)n^\mu_{\Sigma_{(1.1)^{-1}\tau}}\\
&\quad+C\tau^{-2}\iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)} K^{Z,w^Z}_+(\Phi)+C\tau^{-2}\Big|\iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)}(u^2\underline{L}\Phi+v^2L\Phi-\tfrac{1}{4}w\Phi)G\Big|\\
&\quad+C\Big(\sum_{m=0}^{1}\iint_{\mathcal{R}((1.1)^{-1}\tau-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}} r^{1+\delta}(\partial^m_{t^*}G)^2+\sup_{t^*\in[(1.1)^{-1}\tau-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}\cap\{r\le\frac{9}{10}t^*\}} G^2\Big).
\end{aligned}
$$

Replacing $[(1.1)^{-1}\tau,\tau]$ with $[(1.1)^{-2}\tau,(1.1)^{-1}\tau]$, we get also

$$
\begin{aligned}
&\int_{\Sigma_{(1.1)^{-1}\tau}\cap\{r\le\frac{1}{2}t^*\}} J^N_\mu(\Phi)n^\mu_{\Sigma_\tau}\\
&\quad\le C^2\tau^{-1}\int_{\Sigma_{(1.1)^{-2}\tau}\cap\{r\le r_Y^-\}} J^N_\mu(\Phi)n^\mu_{\Sigma_{(1.1)^{-2}\tau}}+C\tau^{-2}\int_{\Sigma_{(1.1)^{-2}\tau}} J^{Z+N,w^Z}_\mu(\Phi)n^\mu_{\Sigma_{(1.1)^{-2}\tau}}\\
&\qquad+C\tau^{-2}\iint_{\mathcal{R}((1.1)^{-2}\tau,(1.1)^{-1}\tau)} K^{Z,w^Z}_+C(\Phi)+C\tau^{-2}\Big|\iint_{\mathcal{R}((1.1)^{-2}\tau,(1.1)^{-1}\tau)}(u^2L\Phi+v^2\underline{L}\Phi-\tfrac{1}{4}w\Phi)G\Big|\\
&\qquad+C\Big(\sum_{m=0}^{1}\iint_{\mathcal{R}((1.1)^{-2}\tau-1,(1.1)^{-1}\tau+1)\cap\{r\le\frac{9}{10}t^*\}} r^{1+\delta}(\partial^m_{t^*}G)^2\\
&\qquad\qquad+\sup_{t^*\in[(1.1)^{-2}\tau-1,(1.1)^{-1}\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}\cap\{r\le\frac{9}{10}t^*\}} G^2\Big)
\end{aligned}
$$

Therefore, plugging this result into the previous, we get

$$
\begin{aligned}
&\int_{\Sigma_\tau\cap\{r\le\frac{1}{2}t^*\}} J^N_\mu(\Phi)n^\mu_{\Sigma_\tau}\\
&\quad\le C^2\tau^{-2}\int_{\Sigma_{(1.1)^{-2}\tau}\cap\{r\le r_Y^-\}} J^N_\mu(\Phi)n^\mu_{\Sigma_{(1.1)^{-2}\tau}}+C\tau^{-2}\int_{\Sigma_{(1.1)^{-2}\tau}} J^{Z,w^Z}_\mu(\Phi)n^\mu_{\Sigma_{(1.1)^{-1}\tau}}\\
&\qquad+C\tau^{-2}\iint_{\mathcal{R}((1.1)^{-2}\tau,\tau)} K^{Z,w^Z}_+(\Phi)+C\tau^{-2}\Big|\iint_{\mathcal{R}((1.1)^{-2}\tau,\tau)}(u^2L\Phi+v^2\underline{L}\Phi-\tfrac{1}{4}w\Phi)G\Big|\\
&\qquad+C\Big(\sum_{m=0}^{1}\iint_{\mathcal{R}((1.1)^{-2}\tau-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}} r^{1+\delta}(\partial^m_{t^*}G)^2\\
&\qquad\qquad+\sup_{t^*\in[(1.1)^{-2}\tau-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}\cap\{r\le\frac{9}{10}t^*\}} G^2\Big). \qquad\square
\end{aligned}
$$

Proposition 37 immediately gives control over the nondegenerate energy and conformal energy using Propositions 31 and 36, respectively:

Corollary 38. *For any* $\gamma<1$,

$$
\begin{aligned}
&\int_{\Sigma_\tau} J^{Z,w^Z}_\mu(\Phi)n^\mu_{\Sigma_\tau}+C\tau^2\int_{\Sigma_\tau\cap\{r\le\gamma\tau\}} J^N_\mu(\Phi)n^\mu_{\Sigma_\tau}\\
&\quad\le C\Big(\int_{\Sigma_{\tau_0}} J^{Z,w^Z}_\mu(\Phi)n^\mu_{\Sigma_{\tau_0}}+C\int_{\Sigma_{\tau_0}} J^N_\mu(\Phi)n^\mu_{\Sigma_{\tau_0}}\\
&\qquad\qquad+\iint_{\mathcal{R}(\tau_0,\tau)} K^{Z,w^Z}_+(\Phi)+\Big|\iint_{\mathcal{R}(\tau_0,\tau)}(u^2\underline{L}\Phi+v^2L\Phi-\tfrac{1}{4}w\Phi)G\Big|\Big)\\
&\quad+C\Big(\sum_{m=0}^{1}\iint_{\mathcal{R}(\tau_0-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}}(t^*)^2r^{1+\delta}(\partial^m_{t^*}G)^2+\sup_{t^*\in[\tau_0-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}\cap\{r\le\frac{9}{10}t^*\}}(t^*)^2G^2\Big).
\end{aligned}
$$

Proof. By Proposition 31,

$$
\tau^2\int_{\Sigma_\tau\cap\{r\le\gamma\tau\}} J^N_\mu(\Phi)n^\mu_{\Sigma_\tau}\le C\int_{\Sigma_\tau} J^{Z+N,w^Z}_\mu(\Phi)n^\mu_{\Sigma_\tau}+C^2\tau^2\int_{\Sigma_\tau\cap\{r\le r_Y^-\}} J^N_\mu(\Phi)n^\mu_{\Sigma_\tau}.
$$

Therefore, by Propositions 36 and 37,

$$
\begin{aligned}
&\int_{\Sigma_\tau} J^{Z,w^Z}_\mu(\Phi)n^\mu_{\Sigma_\tau}+C\tau^2\int_{\Sigma_\tau\cap\{r\le\gamma\tau\}} J^N_\mu(\Phi)n^\mu_{\Sigma_\tau}\\
&\quad\le C\Big(\int_{\Sigma_{(1.1)^{-2}\tau}} J^N_\mu(\Phi)n^\mu_{\Sigma_{(1.1)^{-2}\tau}}+\int_{\Sigma_{(1.1)^{-2}\tau}} J^{Z,w^Z}_\mu(\Phi)n^\mu_{\Sigma_{(1.1)^{-2}\tau}}+\iint_{\mathcal{R}((1.1)^{-2}\tau,\tau)} K^{Z,w^Z}_+(\Phi)\\
&\qquad\qquad+\Big|\iint_{\mathcal{R}((1.1)^{-2}\tau,\tau)}(u^2\underline{L}\Phi+v^2L\Phi-\tfrac{1}{4}w\Phi)G\Big|\Big)\\
&\qquad+C\tau^2\Big(\sum_{m=0}^{1}\iint_{\mathcal{R}((1.1)^{-2}\tau-1,\tau+1)} r^{1+\delta}(\partial^m_{t^*}G)^2+\sup_{t^*\in[(1.1)^{-2}\tau-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}\cap\{r\le\frac{9}{10}t^*\}} G^2\Big).
\end{aligned}
$$

We then use the same estimate for $[(1.1)^{-4}\tau,(1.1)^{-2}\tau]$, $[(1.1)^{-6}\tau,(1.1)^{-4}\tau]$, $\dots$. $\square$

The term $\iint_{\mathcal{R}(\tau_0,\tau)} K_+^{Z,w^Z}(\Phi)$ can be controlled. Here is where the control of the logarithmic divergences from the red-shift vector field is crucially used.

Proposition 39. *We have*

$$\iint_{\mathcal{R}(\tau',\tau)} K_+^{Z,w^Z}(\Phi) \leq C \iint_{\mathcal{R}(\tau',\tau)\cap\{r\geq r_Y^-\}} t^*(r^{-2} J_\mu^N(\Phi) n^\mu_{\Sigma_{\bar\tau}} + r^{-4}\Phi) + \epsilon \iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq r_Y^-\}} (t^*)^2 K^N(\Phi).$$

Proof. See [Dafermos and Rodnianski 2008]. □

The bulk term arising from the inhomogeneous term G can also be controlled.

Proposition 40.

$$\begin{aligned}
&\Big|\iint_{\mathcal{R}(\tau_0,\tau)} (u^2\underline{L}\Phi + v^2 L\Phi - \tfrac14 w\Phi)G\Big| \\
&\quad\leq \delta' \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq\frac12 t^*\}} (t^*)^2 K^{X_0}(\Phi) + \delta' \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq r_Y^-\}} (t^*)^2 K^N(\Phi) \\
&\qquad + \delta' \sup_{t^*\in[\tau_0,\tau]} \Big(\int_{\Sigma_{t^*}\cap\{r\geq\frac12 t^*\}} J_\mu^{Z+N,w^Z}(\Phi) n^\mu_{\Sigma_{t^*}} + (t^*)^2 \int_{\Sigma_{t^*}\cap\{r\leq\frac{23}{8}M\}} J_\mu^N(\Phi) n^\mu_{\Sigma_{t^*}} \Big) \\
&\qquad + C(\delta')^{-1} \sum_{m=0}^{1} \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq\frac12 t^*\}} (t^*)^2 r^{1+\delta} (\partial_{t^*}^m G)^2 + C(\delta')^{-1} \Big(\int_{\tau_0}^{\tau} \Big(\int_{\Sigma_{t^*}\cap\{r\geq\frac12 t^*\}} r^2 G^2 \Big)^{1/2} dt^* \Big)^2 \\
&\qquad + C(\delta')^{-1} \sup_{t^*\in[\tau_0,\tau]} \int_{\Sigma_{t^*}\cap\{r_Y^-\leq r\leq\frac{25}{8}M\}} (t^*)^2 G^2.
\end{aligned}$$

Proof. Two regions require particular care. The first is the region $\{r \leq r_Y^-\}$, since the coefficients of the vector field Z are not bounded as $r \to r_+$. The other is the region $\{|r - 3M| \leq \frac18 M\}$. This is where trapping occurs and where the integrated decay estimate degenerates or loses derivatives. We first look at the region $\{r \leq r_+\}$ using the null frame:

$$\begin{aligned}
&\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq r_Y^-\}} (u^2\underline{L}\Phi + v^2 L\Phi - \tfrac14 w\Phi)G \\
&\quad\leq C \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq r_Y^-\}} ((t^*)^2 + (r_S^*)^2)(|\nabla_{\hat V}\Phi G| + (1-\mu)|\nabla_{\hat Y}\Phi G| + (1-\mu)\sum_A |\nabla_{E_A}\Phi G|) \\
&\qquad\qquad \text{(using Proposition 30)} \\
&\quad\leq C \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq r_Y^-\}} (t^*)^2 (|\log|r-r_+||^2 |\nabla_{\hat V}\Phi G| + |\nabla_{\hat Y}\Phi G| + \sum_A |\nabla_{E_A}\Phi G|) \\
&\quad\leq \delta' \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq r_Y^-\}} (t^*)^2 (|\log|r-r_+||^4 (\nabla_{\hat V}\Phi)^2 + (\nabla_{\hat Y}\Phi)^2 + \sum_A (\nabla_{E_A}\Phi)^2) \\
&\qquad + C(\delta')^{-1} \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq r_Y^-\}} (t^*)^2 G^2 \\
&\quad\leq \delta' \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq r_Y^-\}} (t^*)^2 K^N(\Phi) + C(\delta')^{-1} \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq r_Y^-\}} (t^*)^2 G^2.
\end{aligned}$$

For the region $\{r_Y^- \le r \le \frac{25}{8}M\}$, where trapping occurs, we integrate by parts in t^* so that the bulk term does not have $\partial_{t^*}\Phi$, which cannot be controlled by the integrated decay estimate.

$$\begin{aligned}
&\Big|\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r_Y^-\le r\le\frac{25}{8}M\}} (u^2\underline{L}\Phi + v^2 L\Phi - \tfrac{1}{4}w\Phi)G\Big| \\
&\quad\le C\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r_Y^-\le r\le\frac{25}{8}M\}} (t^*)^2|\partial_r\Phi G| + (t^*)^2|\Phi\partial_{t^*}G| + t^*|\Phi G| + \int_{\Sigma_\tau}\tau^2|\Phi G| + \int_{\Sigma_{\tau_0}}\tau_0^2|\Phi G| \\
&\quad\le C\Big(\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\le\frac{25}{8}M\}} (t^*)^2(\Phi^2 + (\partial_r\Phi)^2)\Big)^{1/2}\Big(\sum_{m=0}^{1}\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r_Y^-\le r\le\frac{25}{8}M\}} (t^*)^2(\partial_{t^*}^m G)^2\Big)^{1/2} \\
&\qquad\qquad + \delta'\int_{\Sigma_\tau\cap\{r_Y^-\le r\le r_Y^-\le\frac{25}{8}M\}} \tau^2 J_\mu^N(\Phi)n^\mu_{\Sigma_\tau} + \delta'\int_{\Sigma_{\tau_0}\cap\{r_Y^-\le r\le\frac{25}{8}M\}} \tau_0^2 J_\mu^N(\Phi)n^\mu_{\Sigma_{\tau_0}} \\
&\qquad\qquad + C(\delta')^{-1}\sup_{t^*\in[\tau_0,\tau]}\int_{\Sigma_{t^*}\cap\{r_Y^-\le r\le\frac{25}{8}M\}} (t^*)^2G^2,
\end{aligned}$$

using Proposition 21. We then move to the region $\{\frac{25}{8}M \le r \le \frac{1}{2}t^*\}$:

$$\begin{aligned}
&\Big|\iint_{\mathcal{R}(\tau_0,\tau)\cap\{\frac{25}{8}M\le r\le\frac{1}{2}t^*\}} (u^2\underline{L}\Phi + v^2 L\Phi - \tfrac{1}{4}w\Phi)G\Big| \\
&\quad\le C\iint_{\mathcal{R}(\tau_0,\tau)\cap\{\frac{25}{8}M\le r\le\frac{1}{2}t^*\}} ((t^*)^2|\partial\Phi| + t^*|\Phi|)|G| \\
&\quad\le C\Big(\iint_{\mathcal{R}(\tau_0,\tau)\cap\{\frac{25}{8}M\le r\le\frac{1}{2}t^*\}} (t^*)^2(r^{-3-\delta}\Phi^2 + r^{-1-\delta}J_\mu^N(\Phi)n^\mu_{\Sigma_{t^*}})\Big)^{1/2} \\
&\qquad\qquad\times\Big(\iint_{\mathcal{R}(\tau_0,\tau)\cap\{\frac{25}{8}M\le r\le\frac{1}{2}t^*\}} (r^{3+\delta} + (t^*)^2r^{1+\delta})G^2\Big)^{1/2} \\
&\quad\le C\Big(\iint_{\mathcal{R}(\tau_0,\tau)\cap\{\frac{25}{8}M\le r\le\frac{1}{2}t^*\}} (t^*)^2(r^{-3-\delta}\Phi^2 + r^{-1-\delta}J_\mu^N(\Phi)n^\mu_{\Sigma_{t^*}})\Big)^{1/2} \\
&\qquad\qquad\times\Big(\iint_{\mathcal{R}(\tau_0,\tau)\cap\{\frac{25}{8}M\le r\le\frac{1}{2}t^*\}} (t^*)^2r^{1+\delta}G^2\Big)^{1/2}
\end{aligned}$$

Finally, we estimate in the region $\{r \ge \frac{1}{2}t^*\}$:

$$\begin{aligned}
&\Big|\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\ge\frac{1}{2}t^*\}} (u^2\underline{L}\Phi + v^2 L\Phi - \tfrac{1}{4}w\Phi)G\Big| \\
&\qquad\le C\sup_{t^*\in[\tau_0,\tau]}\Big(\int_{\Sigma_{t^*}\cap\{r\ge\frac{1}{2}t^*\}} J_\mu^{Z,w^Z}(\Phi)n^\mu_{\Sigma_{t^*}} + (t^*)^2\int_{\Sigma_{t^*}\cap\{r\le r_Y^-\}} J_\mu^N(\Phi)n^\mu_{\Sigma_{t^*}}\Big)^{1/2} \\
&\qquad\qquad\times\int_{\tau_0}^{\tau}\Big(\int_{\Sigma_{t^*}\cap\{r\ge\frac{1}{2}t^*\}} r^2G^2\Big)^{1/2} dt^*,
\end{aligned}$$

where we have used Proposition 31. The proposition follows from Cauchy–Schwarz. □

We have therefore proved the following decay result associated to the vector field Z.

Proposition 41. *For sufficiently small positive δ and δ' and $0 \le \gamma < 1$, there exist $c = c(\delta, \gamma)$ and $C = C(\delta, \gamma)$ such that the following estimate holds for any solution to $\Box_{g_K}\Phi = G$:*

$$\begin{aligned}
c\int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\Phi)n^\mu_{\Sigma_\tau} + \tau^2\int_{\Sigma_\tau\cap\{r\le\gamma\tau\}} J_\mu^N(\Phi)n^\mu_{\Sigma_\tau}
&\le C\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\Phi)n^\mu_{\Sigma_{\tau_0}} + C\iint_{\mathcal{R}(\tau_0,\tau)} t^* r^{-1+\delta}K^{X_1}(\Phi)\\
&\quad + C\delta'\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\le\frac12 t^*\}}(t^*)^2K^{X_0}(\Phi) + C(\delta'+\epsilon)\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\le r_Y^-\}}(t^*)^2K^N(\Phi)\\
&\quad + C(\delta')^{-1}\left(\int_{\tau_0}^{\tau}\left(\int_{\Sigma_{t^*}\cap\{r\ge\frac12 t^*\}} r^2G^2\right)^{1/2}dt^*\right)^2 + C(\delta')^{-1}\sum_{m=0}^{1}\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\le\frac{9}{10}t^*\}}(t^*)^2r^{1+\delta}(\partial_{t^*}^m G)^2\\
&\quad + C(\delta')^{-1}\sup_{t^*\in[\tau_0,\tau]}\int_{\Sigma_{t^*}\cap\{r_Y^-\le r\le\frac{25}{8}M\}}(t^*)^2G^2.
\end{aligned}$$

10. Estimates for solutions to $\Box_{g_K}\Phi = 0$

From this point on, we consider $\Box_{g_K}\Phi = 0$. In this section, we write down the energy estimates derived by Dafermos and Rodnianski [2008]. These will be used in later sections.

Proposition 42.

$$\tau^2\int_{\Sigma_\tau\cap\{r\le\frac12\tau\}} J_\mu^N(\Phi)n^\mu_{\Sigma_\tau} + c\int_{\Sigma_\tau} J_\mu^{Z+N,w^Z}(\Phi)n^\mu_{\Sigma_\tau} \le C\tau^\eta\sum_{m=0}^{2}\left(\int_{\Sigma_{\tau_0}} J_\mu^{Z,w^Z}(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}} + \int_{\Sigma_{\tau_0}} J_\mu^N(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}}\right).$$

Proof. We introduce the bootstrap assumptions:

$$\tau^2\int_{\Sigma_\tau\cap\{r\le\frac12\tau\}} J_\mu^N(\Phi)n^\mu_{\Sigma_\tau} + c\int_{\Sigma_\tau} J_\mu^{Z+N,w^Z}(\Phi)n^\mu_{\Sigma_\tau} \le A^2\tau^\eta\sum_{m=0}^{2}\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}}, \tag{16}$$

$$\tau^2\int_{\Sigma_\tau\cap\{r\le\frac12\tau\}} J_\mu^N(\partial_{t^*}\Phi)n^\mu_{\Sigma_\tau} + c\int_{\Sigma_\tau} J_\mu^{Z+N,w^Z}(\Phi)n^\mu_{\Sigma_\tau} \le A\tau^{1+\eta}\sum_{m=0}^{2}\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}}. \tag{17}$$

Here we think of η as a small positive number. We divide the interval $[\tau_0, \tau]$ dyadically into $\tau_0 \le \tau_1 \le \cdots \le \tau_{n-1} \le \tau_n = \tau$ with $\tau_{i+1} \le (1.1)\tau_i$ and n the smallest integer for doing such division. We then have $n \sim \log|\tau - \tau'|$. We can now apply Proposition 33 on the intervals $[\tau_{i-1}, \tau_i]$ and use the bootstrap assumption (16):

$$\iint_{\mathcal{R}(\tau_{i-1},\tau_i)\cap\{r\le\frac12 t^*\}} K^{X_0}(\Phi)+\iint_{\mathcal{R}(\tau_{i-1},\tau_i)\cap\{r\le r_Y^-\}} K^N(\Phi)$$
$$\le C\left(\tau_i^{-2}\int_{\Sigma_{\tau_{i-1}}} J^{Z,w^Z}_\mu(\Phi)n^\mu_{\Sigma_{\tau_{i-1}}}+C\int_{\Sigma_{\tau_{i-1}}\cap\{r\le r_Y^-\}} J^N_\mu(\Phi)n^\mu_{\Sigma_{\tau_{i-1}}}\right)$$
$$\le CA^2\tau_i^{-2+\eta}\sum_{m=0}^{2}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}}.$$

Similarly, we can apply Proposition 33 on the intervals $[\tau_{i-1},\tau_i]$ for $\partial_{t^*}\Phi$ and use the bootstrap assumption (17):

$$\iint_{\mathcal{R}(\tau_{i-1},\tau_i)\cap\{r\le\frac12 t^*\}} K^{X_0}(\partial_{t^*}\Phi)+\iint_{\mathcal{R}(\tau_{i-1},\tau_i)\cap\{r\le r_Y^-\}} K^N(\partial_{t^*}\Phi)$$
$$\le C\left(\tau_i^{-2}\int_{\Sigma_{\tau_{i-1}}} J^{Z,w^Z}_\mu(\partial_{t^*}\Phi)n^\mu_{\Sigma_{\tau_{i-1}}}+C\int_{\Sigma_{\tau_{i-1}}\cap\{r\le r_Y^-\}} J^N_\mu(\partial_{t^*}\Phi)n^\mu_{\Sigma_{\tau_{i-1}}}\right)$$
$$\le CA\tau_i^{-1+\eta}\sum_{m=0}^{2}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}}.$$

By Proposition 29, we have

$$\iint_{\mathcal{R}(\tau_{i-1},\tau_i)} r^{-1+\delta}K^{X_1}(\partial_{t^*}\Phi)\le C\sum_{m=0}^{1}\int_{\Sigma_{\tau_0}} J^N_\mu(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}}.$$

By Propositions 29 and 34, we have

$$\iint_{\mathcal{R}(\tau_{i-1},\tau_i)} r^{-1+\delta}K^{X_1}(\Phi)\le C\iint_{\mathcal{R}(\tau_{i-1},\tau_i)\cap\{r\le\frac12 t^*\}} K^{X_1}(\Phi)+C\tau_i^{-1+\delta}\iint_{\mathcal{R}(\tau_{i-1},\tau_i)\cap\{r\ge\frac12 t^*\}} K^{X_1}(\Phi)$$
$$\le CA\tau_i^{-1+\eta}\sum_{m=0}^{2}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}}.$$

Applying Proposition 41, we get

$$c\int_{\Sigma_\tau} J^{Z+N,w^Z}_\mu(\Phi)n^\mu_{\Sigma_\tau}+\tau^2\int_{\Sigma_\tau\cap\{r\le\gamma\tau\}} J^N_\mu(\Phi)n^\mu_{\Sigma_\tau}$$
$$\le C\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\Phi)n^\mu_{\Sigma_{\tau_0}}+C\iint_{\mathcal{R}(\tau_0,\tau)} t^*r^{-1+\delta}K^{X_1}(\Phi)$$
$$+C\delta'\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\le\frac12 t^*\}}(t^*)^2K^{X_0}(\Phi)+C(\delta'+\epsilon)\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\le r_Y^-\}}(t^*)^2K^N(\Phi)$$
$$\le\left(C+\left(C+CA+CA^2(2\delta'+\epsilon)\right)\sum_{i=0}^{n-1}\tau_i^\eta\right)\sum_{m=0}^{2}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}}$$

$$\leq (C+\eta^{-1}(C+CA+CA^2(2\delta'+\epsilon))\tau^{\eta})\sum_{m=0}^{2}\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}}.$$

Now take A large, $\epsilon=\eta/(4C)$ and $\delta'=\epsilon/2$, we improve (16). Applying Proposition 41 again, this time to $\partial_{t^*}\Phi$, we have

$$\begin{aligned}
c\int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\partial_{t^*}\Phi)n^\mu_{\Sigma_\tau}+\tau^2\int_{\Sigma_\tau\cap\{r\leq\gamma\tau\}} J_\mu^{N}(\partial_{t^*}\Phi)n^\mu_{\Sigma_\tau}&\\
\leq C\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}\Phi)n^\mu_{\Sigma_{\tau_0}}+C\iint_{\mathcal{R}(\tau_0,\tau)} t^*r^{-1+\delta}K^{X_1}(\partial_{t^*}\Phi)&\\
+C\delta'\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq\frac12 t^*\}}(t^*)^2K^{X_0}(\partial_{t^*}\Phi)+C(\delta'+\epsilon)\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq r_Y^-\}}(t^*)^2K^{N}(\partial_{t^*}\Phi)&\\
\leq\left(C+C\sum_{i=0}^{n-1}\tau_i+CA(2\delta'+\epsilon)\sum_{i=0}^{n-1}\tau_i^{1+\eta}\right)\sum_{m=0}^{2}\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}}&\\
\leq (C+C\tau+CA(2\delta'+\epsilon)\tau^{1+\eta})\sum_{m=0}^{2}\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}}.&
\end{aligned}$$

Now taking A large, $\delta'=\epsilon$ and ϵ sufficiently small, we also improve (17). □

In particular, the theorem of Dafermos and Rodnianski [2008] is retrieved.

Corollary 43 (Dafermos and Rodnianski). *Suppose $\Box_{g_K}\Phi=0$. Then for all $\eta>0$ and all $M>0$ there exists a_0 such that the following estimates hold on Kerr spacetimes with (M,a) for which $a\leq a_0$:*

(1) *The boundedness of nondegenerate energy*:

$$\int_{\Sigma_\tau} J_\mu^N(\Phi)n^\mu_{\Sigma_\tau}+\int_{\mathcal{H}(\tau_0,\tau)} J_\mu^N(\Phi)n^\mu_{\mathcal{H}^+}+\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq r_Y^-\}}K^N(\Phi)+\iint_{\mathcal{R}(\tau',\tau)}K^{X_0}(\Phi)\leq C\int_{\Sigma_{\tau_0}} J_\mu^N(\Phi)n^\mu_{\Sigma_{\tau_0}};$$

(2) *The decay of nondegenerate energy*:

$$\tau^2\int_{\Sigma_\tau\cap\{r\leq\gamma\tau\}} J_\mu^N(\Phi)n^\mu_{\Sigma_\tau}+c\int_{\Sigma_\tau} J_\mu^{Z+N,w^Z}(\Phi)n^\mu_{\Sigma_\tau}\leq C\tau^{1+\eta}\sum_{m=0}^{1}\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}},$$

$$\tau^2\int_{\Sigma_\tau\cap\{r\leq\gamma\tau\}} J_\mu^N(\Phi)n^\mu_{\Sigma_\tau}+c\int_{\Sigma_\tau} J_\mu^{Z+N,w^Z}(\Phi)n^\mu_{\Sigma_\tau}\leq C\tau^{\eta}\sum_{m=0}^{2}\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}}.$$

(3) *The decay of local integrated energy*: *For $\tau'\leq\tau\leq(1.1)\tau'$,*

$$\iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq\frac12 t^*\}}K^{X_0}(\Phi)+\iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq r_Y^-\}}K^N(\Phi)\leq C\tau^{-2+\eta}\sum_{m=0}^{2}\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}},$$

$$\iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq\frac12 t^*\}}K^{X_1}(\Phi)\leq C\tau^{-2+\eta}\sum_{m=0}^{3}\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}}.$$

Proof. Part 1 follows directly from Proposition 25. Part 2 contains two statements. The second is a restatement of Proposition 42. The first is evident from the proof of Proposition 42. Part 3 again has two statements. For the first, we revisit the proof of Proposition 42. The bootstrap assumptions are true; hence it holds. For the second statement, we note by comparing Propositions 33 and 34 that K^{X_1} can be estimated in the same way as K^{X_0} except for an extra derivative. The second statement in 3 can then be proved by rerunning the argument in Proposition 42 with an extra derivative. □

11. Estimates for $\hat{Y}\Phi$ and elliptic estimates

Away from the event horizon, we can control all higher order derivatives simply by commuting with ∂_{t^*} and using standard elliptic estimates. We write down a general version of the estimates in which we have inhomogeneous terms.

Proposition 44. *Suppose $\Box_{g_K}\Phi = G$. For $m \geq 1$ and for any α, we have*

(1) *the boundedness of weighted energy*,

$$\int_{\Sigma_\tau\cap\{r\geq r_Y^-\}} r^\alpha (D^m\Phi)^2 \leq C_{\alpha,m}\Big(\sum_{j=0}^{m-1}\int_{\Sigma_\tau} r^\alpha J_\mu^N(\partial_{t^*}^j\Phi)n^\mu_{\Sigma_\tau} + \sum_{j=0}^{m-2}\int_{\Sigma_\tau} r^\alpha (D^jG)^2\Big),$$

(2) *and the boundedness of local energy*, *that is*, *for any* $0<\gamma<\gamma'$,

$$\int_{\Sigma_\tau\cap\{r_Y^-\leq r\leq \gamma t^*\}} r^\alpha (D^m\Phi)^2 \leq C_{\alpha,m,\gamma,\gamma'}\Big(\sum_{j=0}^{m-1}\int_{\Sigma_\tau\cap\{r\leq\gamma' t^*\}} r^\alpha J_\mu^N(\partial_{t^*}^j\Phi)n^\mu_{\Sigma_\tau} + \tau^{\alpha-\beta-2}\int_{\Sigma_\tau} r^\beta J_\mu^N(\Phi)n^\mu_{\Sigma_\tau} + \sum_{j=0}^{m-2}\int_{\Sigma_\tau\cap\{r\leq\gamma' t^*\}} r^\alpha (D^jG)^2\Big).$$

Proof. This is obvious for $m=1$ (even without the restriction $r\geq r_Y^-$). We will proceed by induction. Take $\delta \ll (r_Y^- - r_+)/4$. Assume

$$\sum_{j=1}^{m-1}\int_{\Sigma_\tau\cap\{r\geq r_Y^- -2\delta\}} r^\alpha (D^j\Phi)^2 \leq C\Big(\sum_{j=0}^{m-2}\int_{\Sigma_\tau} r^\alpha J_\mu^N(\partial_{t^*}^j\Phi)n^\mu_{\Sigma_\tau} + \sum_{j=0}^{m-3}\int_{\Sigma_\tau} r^\alpha (D^jG)^2\Big).$$

We want to show

$$\int_{\Sigma_\tau\cap\{r\geq r_Y^- -\delta\}} r^\alpha (D^m\Phi)^2 \leq C\Big(\sum_{j=0}^{m-1}\int_{\Sigma_\tau} r^\alpha J_\mu^N(\partial_{t^*}^j\Phi)n^\mu_{\Sigma_\tau} + \sum_{j=0}^{m-2}\int_{\Sigma_\tau} r^\alpha (D^jG)^2\Big),$$

which would then imply the conclusion. Denote by Δ_{g_K} the Laplace–Beltrami operator for the metric g_K restricted on the spacelike hypersurface on which t^* is constant. Since ∂_{t^*} is Killing, the operator is defined independent of t^*. Then we have

$$\big|[\Delta_{g_K}, D^k]\Phi\big| \leq C\sum_{j=1}^{k+1}|D^j\Phi|.$$

Denote by ∇ the spatial derivatives with respect to the spatial coordinate variables in the Schwarzschild $(t^*_S, r_S, x^1_S, x^2_S)$ coordinate system. On the set $\{r \geq r^-_Y - (r^-_Y - r_+)/4\}$, Δ_{g_K} is elliptic and therefore controls all spatial derivatives:

$$\begin{aligned}
&\int_{\Sigma_\tau \cap \{r \geq r^-_Y - \delta\}} r^\alpha (D^m \Phi)^2 \\
&\quad \leq C \int_{\Sigma_\tau \cap \{r \geq r^-_Y - 2\delta\}} r^\alpha ((\Delta_{g_K} D^{m-2}\Phi)^2 + (D^{m-1}\Phi)^2 + (\partial^{m-1}_{t^*}\nabla\Phi)^2 + (\partial^m_{t^*}\Phi)^2) \\
&\quad \leq C \int_{\Sigma_\tau \cap \{r \geq r^-_Y - 2\delta\}} r^\alpha \Big((D^{m-2}\Delta_{g_K}\Phi)^2 + \sum_{j=1}^{m-1} (D^j\Phi)^2 + r^{-2}\Phi^2 + (\partial^{m-1}_{t^*}\nabla\Phi)^2 + (\partial^m_{t^*}\Phi)^2 \Big)
\end{aligned}$$

The last two terms are obviously bounded by $C\int_{\Sigma_\tau} J^N_\mu(\partial^{m-1}_{t^*}\Phi) n^\mu_{\Sigma_\tau}$. The second term can be bounded using the induction hypothesis. The third term can be bounded using the Hardy inequality in Proposition 21. Finally, to estimate the first term we use the equation $\Box_{g_K}\Phi = G$. Then, by the form of the Kerr metric, $\Delta_{g_K}\Phi = G - g^{t^*t^*}\partial^2_{t^*}\Phi - 2g^{t^*\phi^*}\partial_{t^*}\partial_{\phi^*}\Phi$. Therefore,

$$\begin{aligned}
&\int_{\Sigma_\tau \cap \{r \geq r^-_Y - 2\delta\}} r^\alpha (D^{m-2}\Delta_{g_K}\Phi)^2 \\
&\quad \leq C \int_{\Sigma_\tau \cap \{r \geq r^-_Y - 2\delta\}} r^\alpha ((D^{m-1}\partial_{t^*}\Phi)^2 + (D^{m-2}G)^2) \\
&\quad \leq C \Big(\sum_{j=0}^{m-1} \int_{\Sigma_\tau} r^\alpha J^N_\mu(\partial^j_{t^*}\Phi) n^\mu_{\Sigma_\tau} + \sum_{j=0}^{m-2} \int_{\Sigma_\tau} r^\alpha (D^j G)^2 \Big),
\end{aligned}$$

where at the last step we have used the induction hypothesis for $\partial_{t^*}\Phi$. We have thus proved the boundedness of weighted energy. To prove the second part of the proposition, consider the function $\chi(r/\tau)\Phi(\tau)$ for a fixed time $t^* = \tau$, where $\chi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is supported in $\{x \leq \gamma'\}$ and is identically 1 in $\{x \leq \gamma\}$. Now

$$\Box_{g_K}\Phi = \chi G + \tau^{-1}\tilde{\chi}\partial_r\Phi + \tau^{-2}\tilde{\tilde{\chi}}\Phi,$$

where $\tilde{\chi}$ and $\tilde{\tilde{\chi}}$ are supported in $\{\gamma \leq t^*/r \leq \gamma'\}$. Thus, by the estimate just proved,

$$\begin{aligned}
&\int_{\Sigma_\tau \cap \{r^-_Y \leq r \leq \gamma t^*\}} r^\alpha (D^m\Phi)^2 \\
&\quad \leq C_\alpha \Big(\sum_{j=0}^{m-1} \int_{\Sigma_\tau \cap \{r \leq \gamma' t^*\}} r^\alpha J^N_\mu(\partial^j_{t^*}\Phi) n^\mu_{\Sigma_\tau} + \int_{\Sigma_\tau \cap \{\gamma t \leq r \leq \gamma' t^*\}} r^\alpha \tau^{-4}\Phi^2 + \sum_{j=0}^{m-2} \int_{\Sigma_\tau \cap \{r \leq \gamma' t^*\}} r^\alpha (D^j G)^2 \Big) \\
&\quad \leq C_\alpha \Big(\sum_{j=0}^{m-1} \int_{\Sigma_\tau \cap \{r \leq \gamma' t^*\}} r^\alpha J^N_\mu(\partial^j_{t^*}\Phi) n^\mu_{\Sigma_\tau} + \tau^{\alpha-\beta-2} \int_{\Sigma_\tau} r^\beta J^N_\mu(\Phi) n^\mu_{\Sigma_\tau} + \sum_{j=0}^{m-2} \int_{\Sigma_\tau \cap \{r \leq \gamma' t^*\}} r^\alpha (D^j G)^2 \Big),
\end{aligned}$$

by the Hardy inequality in Proposition 21. $\Box$

Remark. The boundedness of local energy should be seen as a decay result because for example for the homogeneous equation, the right-hand side of the inequality decays.

Near the event horizon, higher order derivatives can be controlled by commuting with the red-shift vector field as in [Dafermos and Rodnianski 2008; 2011]. The computation here will be completely local, that is, only in the region $\{r \le r_Y^-\}$.

We have the following estimate for higher order derivatives:

Proposition 45. *Suppose $\Box_{g_K}\Phi = G$. For every $m \ge 1$,*

$$\int_{\Sigma_\tau\cap\{r\le r_Y^-\}} (D^m\Phi)^2 \le C\bigg(\sum_{j+k\le m-1}\int_{\Sigma_\tau\cap\{r\le r_Y^-\}} J^N_\mu(\partial_{t^*}^j\hat{Y}^k\Phi)n^\mu_{\Sigma_\tau} + \sum_{j=0}^{m-2}\int_{\Sigma_\tau\cap\{r\le r_Y^-\}}(D^jG)^2\bigg).$$

Proof. This is obvious for $m = 1$. We will proceed by induction. Suppose, for some $m \ge 2$ that

$$\sum_{j=0}^{m-1}\int_{\Sigma_\tau\cap\{r\le r_Y^-\}} (D^j\Phi)^2 \le C\bigg(\sum_{j+k\le m-2}\int_{\Sigma_\tau\cap\{r\le r_Y^-\}} J^N_\mu(\partial_{t^*}^j\hat{Y}^k\Phi)n^\mu_{\Sigma_\tau} + \sum_{j=0}^{m-3}\int_{\Sigma_\tau\cap\{r\le r_Y^-\}}(D^jG)^2\bigg). \tag{18}$$

Since $\Box_{g_K}(\partial_{t^*}\Phi) = \partial_{t^*}G$, this immediately implies

$$\int_{\Sigma_\tau\cap\{r\le r_Y^-\}} (\partial_{t^*}D^{m-1}\Phi)^2 \le C\bigg(\sum_{j+k\le m-1}\int_{\Sigma_\tau\cap\{r\le r_Y^-\}} J^N_\mu(\partial_{t^*}^j\hat{Y}^k\Phi)n^\mu_{\Sigma_\tau} + \sum_{j=0}^{m-2}\int_{\Sigma_\tau\cap\{r\le r_Y^-\}}(D^jG)^2\bigg). \tag{19}$$

Since $\Box_{g_K}\Phi = G$, we have $\Box_{g_K}(\hat{Y}\Phi) = \hat{Y}G + O(1)(D^2\Phi + D\Phi)$. Then using the induction hypothesis (18) (both on $\hat{Y}\Phi$ and Φ), we have

$$\begin{aligned}
&\sum_{j=0}^{m-1}\int_{\Sigma_\tau\cap\{r\le r_Y^-\}} (D^j\hat{Y}\Phi)^2 \\
&\quad\le C\bigg(\sum_{j+k\le m-2}\int_{\Sigma_\tau\cap\{r\le r_Y^-\}} J^N_\mu(\partial_{t^*}^j\hat{Y}^{k+1}\Phi)n^\mu_{\Sigma_\tau} + \sum_{j=0}^{m-3}\int_{\Sigma_\tau}(D^j\hat{Y}G)^2 + \sum_{j=0}^{m-1}\int_{\Sigma_\tau\cap\{r\le r_Y^-\}}(D^j\Phi)^2\bigg) \\
&\quad\le C\bigg(\sum_{j+k\le m-1}\int_{\Sigma_\tau\cap\{r\le r_Y^-\}} J^N_\mu(\partial_{t^*}^j\hat{Y}^k\Phi)n^\mu_{\Sigma_\tau} + \sum_{j=0}^{m-2}\int_{\Sigma_\tau\cap\{r\le r_Y^-\}}(D^jG)^2\bigg).
\end{aligned}\tag{20}$$

Using the null frame $\{\hat{V}, \hat{Y}, E_1, E_2\}$,

$$\Box_{g_K}(D^{m-2}\Phi) = -4\nabla_{\hat{Y}}\nabla_{\hat{V}}D^{m-2}\Phi + \not{\Delta}D^{m-2}\Phi + P_1D^{m-2}\Phi,$$

where P_1 denotes a first order differential operator. Notice that we also have

$$|\Box_{g_K}(D^{m-2}\Phi)| = \big|[\Box_{g_K}, D^{m-2}]\Phi + D^{m-2}G\big| \le C\bigg(\sum_{j=0}^{m-1}|D^j\Phi| + |D^{m-2}G|\bigg).$$

Now using a standard L^2 elliptic estimate on the sphere, we have

$$\int_{\mathbb{S}^2} |\nabla\!\!\!/^2 D^{m-2}\Phi|^2 dA \le C\int_{\mathbb{S}^2}\Big((D^{m-2}G)^2+\sum_{j=0}^{m-1}(D^j\Phi)^2+(D^{m-1}\nabla_{\hat{Y}}\Phi)^2\Big)\,dA,$$

where we notice that the constant can be chosen uniformly because the metric on the sphere is everywhere close to that of the standard metric. Therefore, after integrate over $\{r_+\le r\le r_Y^-\}$ and applying (18) and (20), we have

$$\begin{aligned}\int_{\Sigma_\tau\cap\{r\le r_Y^-\}} |\nabla\!\!\!/^2 D^{m-2}\Phi|^2 &\le C\int_{\Sigma_\tau\cap\{r\le r_Y^-\}}\Big((D^{m-2}G)^2+\sum_{j=0}^{m-1}(D^j\Phi)^2+(D^{m-1}\nabla_{\hat{Y}}\Phi)^2\Big)\\ &\le C\Big(\sum_{j+k\le m-1}\int_{\Sigma_\tau\cap\{r\le r_Y^-\}} J^N_\mu(\partial_{t^*}^j\hat{Y}^k\Phi)n^\mu_{\Sigma_\tau}+\sum_{j=0}^{m-2}\int_{\Sigma_\tau\cap\{r\le r_Y^-\}}(D^jG)^2\Big).\end{aligned}\tag{21}$$

Combining (19), (20) and (21), we have

$$\int_{\Sigma_\tau\cap\{r\le r_Y^-\}}(D^m\Phi)^2\le C\Big(\sum_{j+k\le m}\int_{\Sigma_\tau\cap\{r\le r_Y^-\}} J^N_\mu(\partial_{t^*}^j\hat{Y}^k\Phi)n^\mu_{\Sigma_\tau}+\sum_{j=0}^{m-2}\int_{\Sigma_\tau\cap\{r\le r_Y^-\}}(D^jG)^2\Big).\qquad\square$$

We show that the currents associated to $\hat{Y}^k\Phi$ can actually be controlled. Again, in view of the nonlinear problem, we work in the setting of an inhomogeneous equation.

Proposition 46. *Suppose $\Box_{g_K}\Phi=G$. For every $k\ge 0$,*

$$\begin{aligned}&\int_{\Sigma_\tau\cap\{r\le r_Y^+\}} J^N_\mu(\hat{Y}^k\Phi)n^\mu_{\Sigma_\tau}+\int_{\mathcal{H}(\tau',\tau)} J^N_\mu(\hat{Y}^k\Phi)n^\mu_{\Sigma_\tau}+\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} K^N(\hat{Y}^k\Phi)\\ &\quad\le C\Big(\sum_{j+m\le k}\int_{\Sigma_{\tau'}\cap\{r\le r_Y^+\}} J^N_\mu(\partial_{t^*}^j\hat{Y}^m\Phi)n^\mu_{\Sigma_{\tau'}}+\sum_{j=0}^{k}\int_{\Sigma_\tau\cap\{r\le r_Y^+\}} J^N_\mu(\partial_{t^*}^j\Phi)n^\mu_{\Sigma_\tau}\\ &\qquad+\sum_{j=0}^{k}\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le\frac{23}{8}M\}}(\Phi^2+J^N_\mu(\partial_{t^*}^j\Phi)n^\mu_{\Sigma_{t^*}})+\sum_{j=0}^{k}\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le\frac{23}{8}M\}}(D^jG)^2\Big).\end{aligned}$$

Proof. We prove the proposition by induction on k. The $k=0$ case is trivial because the right-hand side simply contains more terms than the left hand side. We suppose the proposition is true for $k\le k_0-1$ for some $k_0\ge 1$. Commuting $\Box_{g_K}$ with $\hat{Y}$ for k_0 times, we get

$$\Box_{g_K}\hat{Y}^{k_0}\Phi=\kappa^{k_0}\hat{Y}^{k_0+1}\Phi+O(1)\hat{Y}^{k_0}\partial_{t^*}\Phi+O(\epsilon)D^{k_0+1}\Phi+O(1)\sum_{j=1}^{k_0}D^j\Phi+\hat{Y}^{k_0}G.$$

We now use the energy identity for the vector field N, that is, Proposition 9 for $\hat{Y}^k\Phi$. Notice that $\hat{Y}$ is supported in $\{r \le r_Y^+\}$ and therefore each term is supported in the same set. Then

$$\begin{aligned}
&\int_{\Sigma_\tau\cap\{r\le r_Y^+\}} J^N_\mu(\hat{Y}^{k_0}\Phi)n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau_0,\tau)} J^N_\mu(\hat{Y}^{k_0}\Phi)n^\mu_{\Sigma_\tau} + \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\le r_Y^-\}} K^N(\hat{Y}^{k_0}\Phi)\\
&\quad = \int_{\Sigma_{\tau_0}} J^N_\mu(\hat{Y}^{k_0}\Phi)n^\mu_{\Sigma_{\tau_0}} + e\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r_Y^-\le r\le r_Y^+\}} K^Y(\hat{Y}^{k_0}\Phi)\\
&\qquad + \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\le r_Y^+\}} (\partial_{t^*}\hat{Y}^{k_0}\Phi + e\hat{Y}^{k_0+1}\Phi)\Big(-\kappa^{k_0}\hat{Y}^{k_0+1}\Phi + O(1)\hat{Y}^{k_0}\partial_{t^*}\Phi\\
&\qquad\qquad + O(\epsilon)D^{k_0+1}\Phi + O(1)\sum_{j=1}^{k_0} D^j\Phi + \hat{Y}^k G\Big).
\end{aligned}$$

The crucial observation in [Dafermos and Rodnianski 2011] is that one of the inhomogeneous terms has a good sign and thus gives

$$\begin{aligned}
&\int_{\Sigma_\tau} J^N_\mu(\hat{Y}^{k_0}\Phi)n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau',\tau)} J^N_\mu(\hat{Y}^{k_0}\Phi)n^\mu_{\Sigma_\tau} + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} K^N(\hat{Y}^{k_0}\Phi) + \iint_{\mathcal{R}(\tau',\tau)} (\hat{Y}^{k_0+1}\Phi)^2\\
&\quad \le C\Big(\int_{\Sigma_{\tau'}\cap\{r\le r_Y^+\}} J^N_\mu(\hat{Y}^{k_0}\Phi)n^\mu_{\Sigma_{\tau_0}} + \iint_{\mathcal{R}(\tau',\tau)\cap\{r_Y^-\le r\le r_Y^+\}} K^N(\hat{Y}^{k_0}\Phi) + \epsilon\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^+\}} (D^{k_0+1}\Phi)^2\\
&\quad + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^+\}} J^N_\mu(\partial_{t^*}\hat{Y}^{k_0-1}\Phi)n^\mu_{\Sigma_{t^*}} + \sum_{j=1}^{k_0}\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^+\}} (D^j\Phi)^2 + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^+\}} (\hat{Y}^{k_0}G)^2\Big)\\
&\quad \le C\Big(\int_{\Sigma_{\tau'}\cap\{r\le r_Y^+\}} J^N_\mu(\hat{Y}^{k_0}\Phi)n^\mu_{\Sigma_{\tau_0}} + \sum_{j=1}^{k_0+1}\iint_{\mathcal{R}(\tau',\tau)\cap\{r_Y^-\le r\le r_Y^+\}} (D^j\Phi)^2 + \epsilon\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} (D^{k_0+1}\Phi)^2\\
&\quad + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} J^N_\mu(\partial_{t^*}\hat{Y}^{k_0-1}\Phi)n^\mu_{\Sigma_{t^*}} + \sum_{j=1}^{k_0}\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} (D^j\Phi)^2 + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^+\}} (\hat{Y}^{k_0}G)^2\Big).
\end{aligned}$$

Using Proposition 44 with an appropriate cutoff, we have

$$\begin{aligned}
&\sum_{j=1}^{k_0+1}\iint_{\mathcal{R}(\tau',\tau)\cap\{r_Y^-\le r\le r_Y^+\}} (D^j\Phi)^2\\
&\qquad \le C\Big(\sum_{j=0}^{k_0}\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le\frac{23}{8}M\}} (\Phi^2 + J^N_\mu(\partial^j_{t^*}\Phi)n^\mu_{\Sigma_{t^*}}) + \sum_{j=0}^{k_0-1}\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le\frac{23}{8}M\}} (D^jG)^2\Big).
\end{aligned}$$

Using Proposition 45, we get

$$\begin{aligned}
&\sum_{j=1}^{k_0} \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} (D^j\Phi)^2 \\
&\quad \le C\Big(\sum_{j+m\le k_0-1} \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} J^N_\mu(\partial_{t^*}^j \hat{Y}^m\Phi) n^\mu_{\Sigma_{t^*}} + \sum_{j=0}^{k_0-2} \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} (D^j G)^2\Big) \\
&\quad \le C\Big(\sum_{j+m\le k_0-1} \int_{\Sigma_{\tau'}\cap\{r\le r_Y^+\}} J^N_\mu(\partial_{t^*}^j \hat{Y}^m\Phi) n^\mu_{\Sigma_{\tau'}} + \sum_{j=0}^{k-1} \int_{\Sigma_{\tau}\cap\{r\le r_Y^+\}} J^N_\mu(\partial_{t^*}^j\Phi) n^\mu_{\Sigma_{\tau}} \\
&\qquad\qquad + \sum_{j+m\le k_0-1} \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le \frac{23}{8}M\}} J^N_\mu(\partial_{t^*}^j \hat{Y}^m\Phi) n^\mu_{\Sigma_{t^*}} + \sum_{j=0}^{k_0-1} \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le \frac{23}{8}M\}} (D^j G)^2\Big),
\end{aligned}$$

using the induction hypothesis (on $\partial_{t^*}^m\Phi$ instead of Φ) at the last step. Similarly, using Proposition 45,

$$\begin{aligned}
&\epsilon \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} (D^{k_0+1}\Phi)^2 \\
&\quad \le C\epsilon\Big(\sum_{j+m\le k_0} \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} J^N_\mu(\partial_{t^*}^j \hat{Y}^m\Phi) n^\mu_{\Sigma_{t^*}} + \sum_{j=0}^{k-1} \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} (D^j G)^2\Big) \\
&\quad \le C\epsilon\Big(\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^+\}} J^N_\mu(\hat{Y}^{k_0}\Phi) n^\mu_{\Sigma_{t^*}} + \sum_{j=0}^{k_0} \int_{\Sigma_{\tau'}\cap\{r\le r_Y^+\}} J^N_\mu(\partial_{t^*}^j\Phi) n^\mu_{\Sigma_{\tau'}} \\
&\qquad\qquad + \sum_{j=0}^{k_0} \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le \frac{23}{8}M\}} J^N_\mu(\partial_{t^*}^j\Phi) n^\mu_{\Sigma_{t^*}} + \sum_{j=0}^{k_0} \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le \frac{23}{8}M\}} (D^j G)^2\Big),
\end{aligned}$$

where again the induction hypotheses is used at the last step. All these together give

$$\begin{aligned}
&\int_{\Sigma_\tau} J^N_\mu(\hat{Y}^{k_0}\Phi) n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau_0,\tau)} J^N_\mu(\hat{Y}^{k_0}\Phi) n^\mu_{\Sigma_\tau} + \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\le r_Y^-\}} K^N(\hat{Y}^{k_0}\Phi) \\
&\quad \le C\Big(\sum_{j+m\le k_0} \int_{\Sigma_{\tau'}\cap\{r\le r_Y^+\}} J^N_\mu(\partial_{t^*}^j \hat{Y}^m\Phi) n^\mu_{\Sigma_{\tau'}} + \sum_{j=0}^{k_0} \int_{\Sigma_{\tau'}\cap\{r\le r_Y^+\}} J^N_\mu(\partial_{t^*}^j\Phi) n^\mu_{\Sigma_{\tau'}} \\
&\qquad + \sum_{j=0}^{k_0} \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le \frac{23}{8}M\}} (\Phi^2 + J^N_\mu(\partial_{t^*}^j\Phi) n^\mu_{\Sigma_{t^*}}) + \sum_{j=0}^{k_0} \int_{\mathcal{R}(\tau',\tau)\cap\{r\le \frac{23}{8}M\}} (D^j G)^2 \\
&\qquad + \epsilon \iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} J^N_\mu(\hat{Y}^{k_0}\Phi) n^\mu_{\Sigma_{t^*}}\Big).
\end{aligned}$$

The proposition can be proved by noticing that

$$\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} J^N_\mu(\hat{Y}^{k_0}\Phi) n^\mu_{\Sigma_{t^*}} \le C \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\le r_Y^-\}} K^N(\hat{Y}^{k_0}\Phi).$$

and absorbing the small term to the left-hand side. □

We now specialize to the case $\Box_{g_K}\Phi = 0$. The proposition above implies that the behavior of $\hat{Y}^k\Phi$ is determined by the behavior of $\partial_{t^*}^m\Phi$ in the region $\{r \le \frac{23}{8}M\}$.

Proposition 47. *Fix $k \ge 0$. Suppose $\Box_{g_K}\Phi = 0$ and suppose for some constants α, $B > 0$ (independents of τ) that*

$$\sum_{j=0}^{k}\int_{\Sigma_\tau\cap\{r\le\frac{23}{8}M\}}(\Phi^2 + J^N_\mu(\partial_{t^*}^j\Phi)n^\mu_{\Sigma_\tau}) \le CB\tau^{-\alpha}.$$

Then

$$\sum_{j+m\le k}\int_{\Sigma_\tau} J^N_\mu(\partial_{t^*}^j\hat{Y}^m\Phi)n^\mu_{\Sigma_\tau} \le C\tau^{-\alpha}\bigg(\sum_{j+m\le k}\int_{\Sigma_{\tau_0}} J^N_\mu(\partial_{t^*}^j\hat{Y}^m\Phi)n^\mu_{\Sigma_{\tau_0}} + B\bigg),$$

and

$$\sum_{j+m\le k}\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} K^N(\partial_{t^*}^j\hat{Y}^m\Phi) \le C(\tau')^{-\alpha}\bigg(\sum_{j+m\le k}\int_{\Sigma_{\tau_0}} J^N_\mu(\partial_{t^*}^j\hat{Y}^m\Phi)n^\mu_{\Sigma_{\tau_0}} + B\bigg).$$

Remark. In the applications, we will apply this proposition with B being some energy quantity of the initial condition.

Proof. We will proof this with a bootstrap argument. Suppose for all τ that

$$\sum_{j+m\le k}\int_{\Sigma_\tau\cap\{r\le r_Y^+\}} J^N_\mu(\partial_{t^*}^j\hat{Y}^m\Phi)n^\mu_{\Sigma_\tau} \le A\tau^{-\alpha}\bigg(\sum_{j+m\le k}\int_{\Sigma_{\tau_0}} J^N_\mu(\partial_{t^*}^j\hat{Y}^m\Phi)n^\mu_{\Sigma_{\tau_0}} + B\bigg). \tag{22}$$

This obviously holds initially for any $A \ge 1$ (and in particular independent of Φ). By taking $\tau' = \tau - K$, for some (large and to be chosen) constant K and $\tau \ge 2K$, Proposition 46 implies

$$\begin{aligned}
&\sum_{j+m\le k}\bigg(\int_{\Sigma_\tau\cap\{r\le r_Y^+\}} J^N_\mu(\partial_{t^*}^j\hat{Y}^m\Phi)n^\mu_{\Sigma_\tau} + \iint_{\mathcal{R}(\tau-K,\tau)\cap\{r\le r_Y^-\}} J^N_\mu(\partial_{t^*}^j\hat{Y}^m\Phi)n^\mu_{\Sigma_{t^*}}\bigg)\\
&\quad\le C\bigg(\sum_{j+m\le k}\int_{\Sigma_{\tau-K}\cap\{r\le r_Y^+\}} J^N_\mu(\partial_{t^*}^j\hat{Y}^m\Phi)n^\mu_{\Sigma_{\tau-K}} + \sum_{j=0}^{k}\int_{\Sigma_\tau\cap\{r\le r_Y^+\}} J^N_\mu(\partial_{t^*}^j\Phi)n^\mu_{\Sigma_\tau}\\
&\qquad\qquad + \sum_{j=0}^{k}\iint_{\mathcal{R}(\tau-K,\tau)\cap\{r\le\frac{23}{8}M\}}(\Phi^2 + J^N_\mu(\partial_{t^*}^j\Phi)n^\mu_{\Sigma_{t^*}})\bigg)\\
&\quad\le C\bigg(\sum_{j+m\le k}\int_{\Sigma_{\tau-K}\cap\{r\le r_Y^+\}} J^N_\mu(\partial_{t^*}^j\hat{Y}^m\Phi)n^\mu_{\Sigma_{\tau-K}} + KB\tau^{-\alpha}\bigg),
\end{aligned}$$

using the assumption of the proposition and using Proposition 44). Using the bootstrap assumption, we further see that this expression satisfies the bound

$$\begin{aligned}&\leq C\left(A(\tau-K)^{-\alpha}\left(\sum_{j+m\leq k}\int_{\Sigma_{\tau_0}} J^N_\mu(\partial^j_{t^*}\hat{Y}^m\Phi)n^\mu_{\Sigma_{\tau_0}}+B\right)+KB\tau^{-\alpha}\right)\\ &\leq C\tau^{-\alpha}\left(\sum_{j+m\leq k}A\int_{\Sigma_{\tau_0}} J^N_\mu(\partial^j_{t^*}\hat{Y}^m\Phi)n^\mu_{\Sigma_{\tau_0}}+AB+KB\right).\end{aligned}$$

Notice that C is independent of K. By selecting a t^* slice, we have that for some $\tilde{\tau}$,

$$\int_{\Sigma_{\tilde{\tau}}\cap\{r\leq r^-_Y\}} J^N_\mu(\hat{Y}^k\Phi)n^\mu_{\Sigma_{\tilde{\tau}}}\leq CK^{-1}\tau^{-\alpha}\left(\sum_{j+m\leq k}A\int_{\Sigma_{\tau_0}} J^N_\mu(\partial^j_{t^*}\hat{Y}^m\Phi)n^\mu_{\Sigma_{\tau_0}}+AB+KB\right).$$

Now apply Proposition 46 on $[\tilde{\tau},\tau]$ to get

$$\begin{aligned}&\sum_{j+m\leq k}\int_{\Sigma_\tau\cap\{r\leq r^+_Y\}} J^N_\mu(\partial^j_{t^*}\hat{Y}^m\Phi)n^\mu_{\Sigma_\tau}\\ &\quad\leq C\left(\sum_{j+m\leq k}\int_{\Sigma_{\tilde{\tau}}\cap\{r\leq r^+_Y\}} J^N_\mu(\partial^j_{t^*}\hat{Y}^m\Phi)n^\mu_{\Sigma_{\tilde{\tau}}}+\sum_{j=0}^{k}\int_{\Sigma_\tau\cap\{r\leq r^+_Y\}} J^N_\mu(\partial^j_{t^*}\Phi)n^\mu_{\Sigma_\tau}\right.\\ &\qquad\qquad\left.+\sum_{j=0}^{k}\iint_{\mathcal{R}(\tilde{\tau},\tau)\cap\{r\leq\frac{23}{8}M\}}(\Phi^2+J^N_\mu(\partial^j_{t^*}\Phi)n^\mu_{\Sigma_{t^*}})\right)\\ &\quad\leq CK^{-1}\tau^{-\alpha}\left(\sum_{j+m\leq k}A\int_{\Sigma_{\tau_0}} J^N_\mu(\partial^j_{t^*}\hat{Y}^m\Phi)n^\mu_{\Sigma_{\tau_0}}+AB+BK\right)+CB(K+1)\tau^{-\alpha}\\ &\quad\leq CAK^{-1}\tau^{-\alpha}\sum_{j+m\leq k}\int_{\Sigma_{\tau_0}} J^N_\mu(\partial^j_{t^*}\hat{Y}^m\Phi)n^\mu_{\Sigma_{\tau_0}}+(CAK^{-1}+CK+C)B\tau^{-\alpha}.\end{aligned}$$

This will improve (22) if we choose $K=4C$ and A sufficiently large. Hence we have proved

$$\sum_{j+m\leq k}\int_{\Sigma_\tau\cap\{r\leq r^+_Y\}} J^N_\mu(\partial^j_{t^*}\hat{Y}^m\Phi)n^\mu_{\Sigma_\tau}\leq C\tau^{-\alpha}\left(\sum_{j+m\leq k}\int_{\Sigma_{\tau_0}} J^N_\mu(\partial^j_{t^*}\hat{Y}^m\Phi)n^\mu_{\Sigma_{\tau_0}}+B\right).$$

To prove the second statement, we simply use the first statement and Proposition 46. □

We can use Corollary 43 to show the decay of $\hat{Y}^k\Phi$.

Corollary 48. *Suppose $\Box_{g_K}\Phi=0$. Then for $\tau'\leq\tau\leq(1.1)\tau'$,*

$$\begin{aligned}&\sum_{j+m\leq k}\int_{\Sigma_\tau\cap\{r\leq r^+_Y\}} J^N_\mu(\partial^j_{t^*}\hat{Y}^m\Phi)n^\mu_{\Sigma_\tau}\\ &\qquad\leq C\tau^{-2+\eta}\left(\sum_{j+m\leq k}\int_{\Sigma_{\tau_0}} J^N_\mu(\partial^j_{t^*}\hat{Y}^m\Phi)n^\mu_{\Sigma_{\tau_0}}+\sum_{j=0}^{k+2}\int_{\Sigma_{\tau_0}} J^{Z+CN}_\mu(\partial^j_{t^*}\Phi)n^\mu_{\Sigma_{\tau_0}}\right),\end{aligned}$$

and

$$\sum_{j+m\le k}\iint_{\mathcal{R}(\tau',\tau)\cap\{r\le r_Y^-\}} K^N(\partial_{t^*}^j\hat{Y}^m\Phi)$$
$$\le C\tau^{-2+\eta}\Big(\sum_{j+m\le k}\int_{\Sigma_{\tau_0}} J^N_\mu(\partial_{t^*}^j\hat{Y}^m\Phi)n^\mu_{\Sigma_{\tau_0}}+\sum_{j=0}^{k+2}\int_{\Sigma_{\tau_0}} J^{Z+CN}_\mu(\partial_{t^*}^j\Phi)n^\mu_{\Sigma_{\tau_0}}\Big).$$

12. Estimates for $\widetilde{\Omega}\Phi$

In this section, we would like to prove estimates for $\widetilde{\Omega}^\ell\Phi$. The estimates for $\widetilde{\Omega}\Phi$ are useful to provide an extra factor of r in the energy estimates.

Proposition 49. *We have*

$$\int_{\Sigma_\tau\cap\{r\ge r_Y^-\}} r^2|\nabla\!\!\!/^2\Phi|^2\le C\int_{\Sigma_\tau} J^N_\mu(\Phi,\partial_{t^*}\Phi,\widetilde{\Omega}\Phi)n^\mu_{\Sigma_\tau},$$
$$\iint_{\mathcal{R}(\tau',\tau)\cap\{r\ge r_Y^-\}} r^{1-\delta}|\nabla\!\!\!/^2\Phi|^2\le C\iint_{\mathcal{R}(\tau',\tau)} r^{-1-\delta}J^N_\mu(\Phi,\partial_{t^*}\Phi,\widetilde{\Omega}\Phi)n^\mu_{\Sigma_{t^*}},$$
$$\iint_{\mathcal{R}(\tau',\tau)\cap\{r\ge r_Y^-\}} r^{1-\delta}|\nabla\!\!\!/^2\Phi|^2\le C\iint_{\mathcal{R}(\tau',\tau)} K^{X_1}(\Phi,\partial_{t^*}\Phi)+K^{X_0}(\widetilde{\Omega}\Phi).$$

To prove such estimates, we commute $\Box_{g_K}$ with $\widetilde{\Omega}$. Recall from Section 4.3 that

$$|[\Box_{g_K},\widetilde{\Omega}]\Phi|\le Cr^{-2}(|D^2\Phi|+|D\Phi|)$$

everywhere, and

$$[\Box_{g_K},\widetilde{\Omega}]\Phi=0$$

for $r<R_\Omega$. Now suppose $\Box_{g_K}\Phi=0$. We have

$$\Box_{g_K}(\widetilde{\Omega}^\ell\Phi)=\sum_{j=0}^{\ell-1}\widetilde{\Omega}^j[\Box_{g_K},\widetilde{\Omega}]\widetilde{\Omega}^{\ell-j-1}\Phi=:G_{\Omega,\ell}.$$

Since $[D,\widetilde{\Omega}]=D$, we have

$$|G_{\Omega,\ell}|\le Cr^{-2}\Big(\sum_{j=0}^{\ell-1}(|D^2\widetilde{\Omega}^j\Phi|+|D\widetilde{\Omega}^j\Phi|)+\sum_{j=0}^{\ell+1}|D^j\Phi|\Big),$$

and $G_{\Omega,\ell}$ is supported in $\{r\ge R_\Omega\}$.

Definition 50.
$$E_{\Omega,\ell}=\iint_{\mathcal{R}(\tau'-1,\tau+1)} r^{1+\delta}G^2_{\Omega,\ell}+\sup_{t^*\in[\tau'-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac18 M\}} {G_{\Omega,\ell}}^2.$$

This is the error term for the energy estimates for $\widetilde{\Omega}^\ell\Phi$. We show that this can be controlled.

Proposition 51. *We have*

$$E_{\Omega,\ell} \leq C \sum_{m=0}^{1} \sum_{j=0}^{\ell-1} \iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\geq R_\Omega\}} K^{X_0}(\partial_{t^*}^m \widetilde{\Omega}^j \Phi) + C \sum_{m=0}^{\ell} \iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\geq R_\Omega\}} K^{X_0}(\partial_{t^*}^m \Phi).$$

Proof.

$$\begin{aligned}
&\iint_{\mathcal{R}(\tau'-1,\tau+1)} r^{1+\delta} G_{\Omega,\ell}^2 \\
&\leq C \sum_{j=0}^{\ell-1} \iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\geq R_\Omega\}} r^{-3+\delta}((D^2\widetilde{\Omega}^j\Phi)^2 + (D\widetilde{\Omega}^j\Phi)^2) + C \sum_{j=0}^{\ell+1} \iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\geq R_\Omega\}} r^{-3+\delta}(D^j\Phi)^2 \\
&\leq C \sum_{m=0}^{1} \sum_{j=0}^{\ell-1} \iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\geq R_\Omega\}} r^{-3+\delta} J_\mu^N(\partial_{t^*}^m \widetilde{\Omega}^j \Phi) n_{\Sigma t^*}^\mu + C \sum_{m=0}^{\ell} \iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\geq R_\Omega\}} r^{-3+\delta} J_\mu^N(\partial_{t^*}^m \Phi) n_{\Sigma t^*}^\mu \\
&\leq C \sum_{m=0}^{1} \sum_{j=0}^{\ell-1} \iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\geq R_\Omega\}} K^{X_0}(\partial_{t^*}^m \widetilde{\Omega}^j \Phi) + C \sum_{m=0}^{\ell} \iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\geq R_\Omega\}} K^{X_0}(\partial_{t^*}^m \Phi).
\end{aligned}$$

By choosing R_Ω sufficiently large, the second term of $E_{\Omega,\ell}$ vanishes. □

We can show that the nondegenerate energy of $\widetilde{\Omega}^\ell \Phi$ is almost bounded.

Proposition 52. *Suppose* $\Box_{g_K} \Phi = 0$. *Then*

$$\int_{\Sigma_\tau} J_\mu^N(\widetilde{\Omega}^\ell \Phi) n_{\Sigma_\tau}^\mu + \int_{\mathcal{H}(\tau_0,\tau)} J_\mu^N(\widetilde{\Omega}^\ell \Phi) n_{\mathcal{H}^+}^\mu + \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq r_Y^-\}} K^N(\widetilde{\Omega}^\ell \Phi) + \iint_{\mathcal{R}(\tau_0,\tau)} K^{X_0}(\widetilde{\Omega}^\ell \Phi) \leq C \sum_{i+j\leq\ell} \int_{\Sigma_{\tau_0}} J_\mu^N(\partial_{t^*}^i \widetilde{\Omega}^j \Phi) n_{\Sigma_{\tau_0}}^\mu.$$

Proof. We prove this by induction on ℓ. The $\ell = 0$ case is true by setting $G = 0$ in Proposition 28. We assume that the proposition is true for $\ell \leq \ell_0 - 1$. This in particular implies, after a commutations with the Killing vector field ∂_{t^*}, that

$$\sum_{j=0}^{\ell_0-1} \iint_{\mathcal{R}(\tau_0,\tau)} K^{X_0}(\partial_{t^*}^m \widetilde{\Omega}^j \Phi) \leq C \sum_{i+j\leq m+\ell_0-1} \int_{\Sigma_{\tau_0}} J_\mu^N(\partial_{t^*}^i \widetilde{\Omega}^j \Phi) n_{\Sigma_{\tau_0}}^\mu.$$

By Propositions 27 and 51,

$$
\begin{aligned}
&\int_{\Sigma_\tau} J^N_\mu(\widetilde{\Omega}^{\ell_0}\Phi)n^\mu_{\Sigma_\tau} + \int_{\mathcal{H}(\tau_0,\tau)} J^N_\mu(\widetilde{\Omega}^{\ell_0}\Phi)n^\mu_{\mathcal{H}^+} + \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\le r_Y^-\}} K^N(\widetilde{\Omega}^{\ell_0}\Phi) + \iint_{\mathcal{R}(\tau_0,\tau)} K^{X_0}(\widetilde{\Omega}^{\ell_0}\Phi)\\
&\qquad\le C\Big(\int_{\Sigma_{\tau_0}} J^N_\mu(\widetilde{\Omega}^{\ell_0}\Phi)n^\mu_{\Sigma_{\tau_0}} + \iint_{\mathcal{R}(\tau'-1,\tau+1)} r^{1+\delta}G^2_{\Omega,\ell_0} + \sup_{t^*\in[\tau'-1,\tau+1]}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac{1}{8}M\}} G^2_{\Omega,\ell_0}\Big)\\
&\qquad\le C\Big(\int_{\Sigma_{\tau_0}} J^N_\mu(\widetilde{\Omega}^{\ell_0}\Phi)n^\mu_{\Sigma_{\tau_0}} + C\sum_{m=0}^{1}\sum_{j=0}^{\ell_0-1}\iint_{\mathcal{R}(\tau_0-1,\tau+1)\cap\{r\ge R_\Omega\}} K^{X_0}(\partial_{t^*}^m\widetilde{\Omega}^j\Phi)\\
&\qquad\qquad\qquad + C\sum_{m=0}^{\ell_0}\iint_{\mathcal{R}(\tau_0-1,\tau+1)\cap\{r\ge R_\Omega\}} K^{X_0}(\partial_{t^*}^m\Phi)\Big)\\
&\qquad\le C\sum_{i+j\le\ell_0}\int_{\Sigma_{\tau_0}} J^N_\mu(\partial_{t^*}^i\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}.
\end{aligned}
$$

□

Remark. Only the $\ell = 1$ case will be used.

13. Estimates for $S\Phi$

We will now use the energy estimates that we have derived to control $S\Phi$. In particular, we would like to prove a local integrated decay estimate for $S\Phi$. This will be used in the next section where we prove our main theorem. Recall from Section 4.2 that for r large

$$
\begin{aligned}
&\Big|[\Box_{g_K}, S]\Phi - \Big(2+\frac{r^*\mu}{r}\Big)\Box_{g_K}\Phi - \frac{2}{r}\Big(\frac{r^*}{r}-1-\frac{2r^*\mu}{r}\Big)\partial_{r^*}\Phi - 2\Big(\Big(\frac{r^*}{r}-1\Big)-\frac{3r^*\mu}{2r}\Big)\Delta\!\!\!\!/\,\Phi\Big|\\
&\qquad\qquad\qquad\qquad\le C\epsilon r^{-2}\Big(\sum_{k=1}^{2}|\partial^k\Phi|\Big),
\end{aligned}
$$

and that for $r\le R$, we have

$$|[\Box_{g_K}, S]\Phi| \le C\Big(\sum_{k=1}^{2}|D^k\Phi|\Big).$$

From now on we will prove estimates for $S\Phi$ by considering the wave equation that it satisfies. We will assume, as before, $\Box_{g_K}\Phi = 0$ and let G denote the commutator term, that is, $\Box_{g_K}(S\Phi) = G$. If we look at our estimates in the previous sections, we will need to control G in three different norms. We now consider them separately.

Proposition 53. *Let $\tau'\le\tau\le(1.1)\tau'$. Then*

$$
\begin{aligned}
&\sum_{m=0}^{\ell}\iint_{\mathcal{R}(\tau'-1,\tau+1)} r^{1+\delta}(\partial_{t^*}^m G)^2\\
&\qquad\le C\tau^{-1+\eta}\sum_{m+k+j\le\ell+3}\Big(\int_{\Sigma_{\tau_0}} J^{Z,w^Z}_\mu(\partial_{t^*}^m\hat{Y}^k\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}} + C\int_{\Sigma_{\tau_0}} J^N_\mu(\partial_{t^*}^m\hat{Y}^k\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}\Big).
\end{aligned}
$$

and

$$\sum_{m=0}^{\ell}\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}} r^{1+\delta}(\partial_{t^*}^m G)^2 \le C\tau^{-2+\eta}\sum_{m+k+j\le\ell+4}\Big(\int_{\Sigma_{\tau_0}} J_\mu^{Z,w^Z}(\partial_{t^*}^m\hat{Y}^k\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}} + C\int_{\Sigma_{\tau_0}} J_\mu^N(\partial_{t^*}^m\hat{Y}^k\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}\Big).$$

In other words, we can get more decay if we localize and allow an extra derivative.

Proof.

$$\begin{aligned}
&\sum_{m=0}^{\ell}\iint_{\mathcal{R}(\tau'-1,\tau+1)} r^{1+\delta}(\partial_{t^*}^m G)^2\\
&\le C\sum_{m=0}^{\ell}\iint_{\mathcal{R}(\tau'-1,\tau+1)} r^{-3+\delta}((\partial_{t^*}^m D^2\Phi)^2+(\partial_{t^*}^m D\Phi)^2+(r\partial_{t^*}^m \not\Delta\Phi)^2)\\
&\qquad\qquad\text{(noting that the } \delta \text{ in the two lines are different)}\\
&\le C\sum_{m+k\le\ell+1}\Big(\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\le\frac12 t^*\}} r^{-1+\delta}J_\mu^N(\partial_{t^*}^m\hat{Y}^k\Phi)n^\mu_{\Sigma^*_t}\\
&\qquad\qquad + \iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\ge\frac12 t^*\}} r^{-3+\delta}J_\mu^N(\partial_{t^*}^m\widetilde{\Omega}^k\Phi)n^\mu_{\Sigma^*_t}\Big)\\
&\qquad\qquad\text{(by Proposition 44, 45 and 49)}\\
&\le C\sum_{m+k\le\ell+3}\tau^{-1+\eta}\Big(\int_{\Sigma_{\tau_0}} J_\mu^{Z,w^Z}(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}} + C\int_{\Sigma_{\tau_0}} J_\mu^N(\partial_{t^*}^m\hat{Y}^k\Phi)n^\mu_{\Sigma_{\tau_0}}\Big)\\
&\qquad + C\sum_{m+k\le\ell+3}\int_{\tau'-1}^{\tau+1}(t^*)^{-3+\delta}\Big(\int_{\Sigma_{\tau_0}} J_\mu^{Z,w^Z}(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{\tau_0}} + C\int_{\Sigma_{\tau_0}} J_\mu^N(\partial_{t^*}^m\hat{Y}^k\Phi)n^\mu_{\Sigma_{\tau_0}}\Big)\,dt^*\\
&\qquad + C\sum_{m\le\ell+1}\int_{\tau'-1}^{\tau+1}(t^*)^{-3+\delta}\Big(\int_{\Sigma_{\tau_0}} J_\mu^N(\partial_{t^*}^m\widetilde{\Omega}\Phi)n^\mu_{\Sigma_{\tau_0}}\Big)\,dt^*\\
&\qquad\qquad\text{(using Corollaries 43, 48 and Proposition 52)}\\
&\le C\tau^{-1+\eta}\sum_{m+k+j\le\ell+3}\Big(\int_{\Sigma_{\tau_0}} J_\mu^{Z,w^Z}(\partial_{t^*}^m\hat{Y}^k\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}} + C\int_{\Sigma_{\tau_0}} J_\mu^N(\partial_{t^*}^m\hat{Y}^k\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}\Big).
\end{aligned}$$

We then move on to the localized version:

$$\begin{aligned}
&\sum_{m=0}^{\ell}\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}} r^{1+\delta}(\partial_{t^*}^m G)^2\\
&\qquad\le C\sum_{m=0}^{\ell}\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\le\frac{9}{10}t^*\}} r^{-3+\delta}((\partial_{t^*}^m D^2\Phi)^2+(\partial_{t^*}^m D\Phi)^2+(r\partial_{t^*}^m \not\Delta\Phi)^2)
\end{aligned}$$

$$
\begin{aligned}
&\le C \sum_{m+k\le \ell+1}\Bigg(\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{r\le \frac12 t^*\}} r^{-1+\delta} J^N_\mu(\partial_{t^*}^m \hat{Y}\Phi) n^\mu_{\Sigma^*_t} + \iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{\frac12 t^*\le r\le \frac{19t^*}{20}\}} r^{-3+\delta} J^N_\mu(\partial_{t^*}^m \hat{Y}\Phi) n^\mu_{\Sigma^*_t}\Bigg)\\
&\qquad + C\sum_{m=0}^{1}\Bigg(\iint_{\mathcal{R}(\tau'-1,\tau+1)\cap\{\frac12 t^*\le r\le \frac{19t^*}{20}\}} r^{-3+\delta} J^N_\mu(\partial_{t^*}^m \widetilde{\Omega}\Phi) n^\mu_{\Sigma^*_t}\Bigg) \quad \text{(by Proposition 44, 45 and 49)}\\
&\le C\sum_{m+k\le \ell+4} \tau^{-2+\eta}\Bigg(\int_{\Sigma_{\tau_0}} J^{Z,w^Z}_\mu(\partial_{t^*}^m\Phi) n^\mu_{\Sigma_{\tau_0}} + C\int_{\Sigma_{\tau_0}} J^N_\mu(\partial_{t^*}^m \hat{Y}\Phi) n^\mu_{\Sigma_{\tau_0}}\Bigg)\\
&\qquad + C\sum_{m+k\le \ell+3}\int_{\tau'-1}^{\tau+1} (t^*)^{-3+\delta}\Bigg(\int_{\Sigma_{\tau_0}} J^{Z,w^Z}_\mu(\partial_{t^*}^m\Phi) n^\mu_{\Sigma_{\tau_0}} + C\int_{\Sigma_{\tau_0}} J^N_\mu(\partial_{t^*}^m \hat{Y}\Phi) n^\mu_{\Sigma_{\tau_0}}\Bigg)dt^*\\
&\qquad + C\sum_{m=0}^{\ell}\int_{\tau'-1}^{\tau+1} (t^*)^{-3+\delta}\Bigg(\int_{\Sigma_{\tau_0}} J^N_\mu(\partial_{t^*}^m \widetilde{\Omega}\Phi) n^\mu_{\Sigma_{\tau_0}}\Bigg)dt^* \quad \text{(using Corollaries 43, 48 and Proposition 52)}\\
&\le C\tau^{-2+\eta}\sum_{m+k+j\le \ell+4}\Bigg(\int_{\Sigma_{\tau_0}} J^{Z,w^Z}_\mu(\partial_{t^*}^m \hat{Y}^k\widetilde{\Omega}^j\Phi) n^\mu_{\Sigma_{\tau_0}} + C\int_{\Sigma_{\tau_0}} J^N_\mu(\partial_{t^*}^m \hat{Y}^k\widetilde{\Omega}^j\Phi) n^\mu_{\Sigma_{\tau_0}}\Bigg). \qquad \square
\end{aligned}
$$

To estimate the inhomogeneous term in the region $r \le \frac12 t^*$, we will also need to estimate a term not integrated over t^*, which arises from the integration by parts.

Proposition 54. *For $\tau' \le \tau \le (1.1)\tau'$,*

$$
\sup_{t^*\in[\tau'-1,\tau+1]}\sum_{m=0}^{\ell}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac18 M\}} (\partial_{t^*}^m G)^2 \le C\tau^{-2+\eta}\sum_{m+j\le \ell+3}\Bigg(\int_{\Sigma_{\tau_0}} J^{Z,w^Z}_\mu(\partial_{t^*}^m\widetilde{\Omega}^j\Phi) n^\mu_{\Sigma_{\tau_0}} + C\int_{\Sigma_{\tau_0}} J^N_\mu(\partial_{t^*}^m\widetilde{\Omega}^j\Phi) n^\mu_{\Sigma_{\tau_0}}\Bigg).
$$

Proof.

$$
\begin{aligned}
&\sup_{t^*\in[\tau'-1,\tau+1]}\sum_{m=0}^{\ell}\int_{\Sigma_{t^*}\cap\{|r-3M|\le\frac18 M\}} (\partial_{t^*}^m G)^2\\
&\quad\le C\sup_{t^*\in[\tau'-1,\tau+1]}\sum_{m=0}^{\ell}\int_{\Sigma_{t^*}\cap\{r_Y^-\le r\le\frac{25}{8}M\}} \big((D^2\partial_{t^*}^m\Phi)^2 + (D\partial_{t^*}^m\Phi)^2 + (r\not{\Delta}\partial_{t^*}^m\Phi)^2\big)\\
&\quad\le C\sup_{t^*\in[\tau'-1,\tau+1]}\Bigg(\sum_{m=0}^{\ell+1}\int_{\Sigma_{t^*}\cap\{r_Y^-\le r\le\frac{25}{8}M\}} J^N_\mu(\partial_{t^*}^m\Phi) n^\mu_{\Sigma^*_t} + \sum_{m=0}^{\ell}\int_{\Sigma_{t^*}\cap\{r_Y^-\le r\le\frac{25}{8}M\}} J^N_\mu(\widetilde{\Omega}\partial_{t^*}^m\Phi) n^\mu_{\Sigma^*_t}\Bigg)\\
&\qquad\qquad \text{(by Proposition 44 and 49)}\\
&\quad\le C\tau^{-2+\eta}\sum_{m+j\le \ell+3}\Bigg(\int_{\Sigma_{\tau_0}} J^{Z,w^Z}_\mu(\partial_{t^*}^m\widetilde{\Omega}^j\Phi) n^\mu_{\Sigma_{\tau_0}} + C\int_{\Sigma_{\tau_0}} J^N_\mu(\partial_{t^*}^m\widetilde{\Omega}^j\Phi) n^\mu_{\Sigma_{\tau_0}}\Bigg),\\
&\qquad\qquad \text{(using Corollary 43 and Proposition 52).} \qquad \square
\end{aligned}
$$

Finally, we estimate the third norm:

Proposition 55. $$\sum_{m=0}^{\ell}\left(\int_{\tau_0}^{\tau}\left(\int_{\Sigma_{t^*}\cap\{r\geq\frac{1}{2}t^*\}} r^2(\partial_{t^*}^m G)^2\right)^{1/2} dt^*\right)^2 \leq C\tau^{\eta}\sum_{m+j\leq\ell+3}\int_{\Sigma_{\tau_0}} J_\mu^N(\partial_{t^*}^m\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}.$$

Proof.

$$\begin{aligned}
&\sum_{m=0}^{\ell}\left(\int_{\tau_0}^{\tau}\left(\int_{\Sigma_{t^*}\cap\{r\geq\frac{1}{2}t^*\}} r^2(\partial_{t^*}^m G)^2\right)^{1/2} dt^*\right)^2\\
&\quad\leq C\sum_{m=0}^{\ell}\left(\int_{\tau_0}^{\tau}(t^*)^{-1+\delta}\left(\int_{\Sigma_{t^*}\cap\{r\geq\frac{1}{2}t^*\}}((D^2\Phi)^2+(D\partial_{t^*}^m\Phi)^2+(r\not\Delta\partial_{t^*}^m\Phi)^2)\right)^{1/2} dt^*\right)^2\\
&\quad\leq C\left(\int_{\tau_0}^{\tau}(t^*)^{-1+\delta}\left(\sum_{m=0}^{\ell+1}\int_{\Sigma_{t^*}} J_\mu^N(\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{t^*}}+\sum_{m=0}^{\ell}\int_{\Sigma_{t^*}} J_\mu^N(\widetilde{\Omega}\partial_{t^*}^m\Phi)n^\mu_{\Sigma_{t^*}}\right)^{1/2} dt^*\right)^2\\
&\quad\leq C\tau^{\eta}\sum_{m+j\leq\ell+3}\int_{\Sigma_{\tau_0}} J_\mu^N(\partial_{t^*}^m\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}. \qquad\square
\end{aligned}$$

Now that we have control of the inhomogeneous terms in the equation $\Box_{g_K}\Phi = G$, we can prove the decay of $S\Phi$. To this end, we will introduce the bootstrap assumptions:

$$\begin{aligned}
&c\int_{\Sigma_\tau} J_\mu^{Z,w^Z}(\partial_{t^*}S\Phi)n^\mu_{\Sigma_\tau}+\tau^2\int_{\Sigma_\tau\cap\{r\leq\gamma\tau\}} J_\mu^N(\partial_{t^*}S\Phi)n^\mu_{\Sigma_\tau}\\
&\qquad\leq A\tau\sum_{m=1}^{2}\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m S\Phi)n^\mu_{\Sigma_{\tau_0}}+A\tau^{1+\eta}\sum_{m+k+j\leq 5}\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m\hat{Y}^k\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}, \qquad (23)\\
&c\int_{\Sigma_\tau} J_\mu^{Z,w^Z}(S\Phi)n^\mu_{\Sigma_\tau}+\tau^2\int_{\Sigma_\tau\cap\{r\leq\gamma\tau\}} J_\mu^N(S\Phi)n^\mu_{\Sigma_\tau}\\
&\qquad\leq A^2\tau^{\eta}\left(\sum_{m=0}^{2}\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m S\Phi)n^\mu_{\Sigma_{\tau_0}}+\sum_{m+k+j\leq 5}\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m\hat{Y}^k\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}\right). \qquad (24)
\end{aligned}$$

We think of A as some large constant to be chosen. We will improve the constants A and A^2 in the assumptions above. Under these two assumptions, we will get the following three estimates for the bulk terms:

Proposition 56.

$$\iint_{\mathcal{R}(\tau_0,\tau)} K^{X_1}(\partial_{t^*}S\Phi)\leq C\left(\sum_{m=1}^{2}\int_{\Sigma_{\tau_0}} J_\mu^N(\partial_{t^*}^m S\Phi)n^\mu_{\Sigma_{\tau_0}}+\sum_{m+k+j\leq 5}\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m\hat{Y}\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}\right).$$

Proof. The follows by Proposition 29 applied to the equation $\Box_{g_K}(\partial_{t^*}S\Phi)=\partial_{t^*}G$, taking $\tau'=\tau_0$, $G_1=0$, and $G_2=\partial_{t^*}G$. Then use Propositions 53 and 54 to estimate the terms with G. $\square$

Proposition 57. *For $\tau' \leq \tau \leq (1.1)\tau'$, we have*

$$\iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq r_Y^-\}} K^N(\partial_{t^*}S\Phi) + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq\frac{1}{2}t^*\}} K^{X_0}(\partial_{t^*}S\Phi) + \iint_{\mathcal{R}(\tau_0,\tau)} r^{-1+\delta}K^{X_1}(S\Phi)$$
$$\leq CA\bigg(\tau^{-2}\int_{\Sigma_{\tau'}} J_\mu^{Z,w^Z}(S\Phi)n^\mu_{\Sigma_{\tau'}} + C\int_{\Sigma_{\tau'}\cap\{r\leq r_Y^-\}} J_\mu^N(S\Phi)n^\mu_{\Sigma_{\tau'}}\bigg)$$
$$+ CA\tau^{-1+\eta}\sum_{m+k+j\leq 5}\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m\hat{Y}\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}.$$

Proof. By Propositions 33 and 34, taking $G_1 = 0$ and $G_2 = G$, and using Propositions 53 and 54 to estimate the terms with G, we have

$$\iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq r_Y^-\}} K^N(\partial_{t^*}S\Phi) + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq\frac{1}{2}t^*\}} K^{X_0}(\partial_{t^*}S\Phi) + \iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq\frac{1}{2}t^*\}} K^{X_1}(S\Phi)$$
$$\leq CA\bigg(\tau^{-2}\int_{\Sigma_{\tau'}} J_\mu^{Z,w^Z}(S\Phi)n^\mu_{\Sigma_{\tau'}} + C\int_{\Sigma_{\tau'}\cap\{r\leq r_Y^-\}} J_\mu^N(S\Phi)n^\mu_{\Sigma_{\tau'}}\bigg)$$
$$+ CA\tau^{-1+\eta}\sum_{m+k+j\leq 5}\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m\hat{Y}\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}.$$

It remains to estimate $r^{-1+\delta}K^{X_1}$ in the region $r \geq \frac{1}{2}t^*$. Here, we will use crucially the decay in r. Clearly,

$$\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\geq\frac{1}{2}t^*\}} r^{-1+\delta}K^{X_1}(S\Phi) \leq C\tau^{-1+\delta}\iint_{\mathcal{R}(\tau_0,\tau)} K^{X_1}(S\Phi).$$

Then we can estimate the right-hand side by Proposition 29, taking $\tau' = \tau_0$, $G_1 = 0$ and $G_2 = \partial_{t^*}G$. Then use Propositions 53 and 54 to estimate the terms with G. □

Proposition 58. *For $\tau' \leq \tau \leq (1.1)\tau'$,*

$$\iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq r_Y^-\}} K^N(S\Phi) + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq\frac{1}{2}t^*\}} K^{X_0}(S\Phi)$$
$$\leq CA^2\bigg(\tau^{-2}\int_{\Sigma_{\tau'}} J_\mu^{Z,w^Z}(S\Phi)n^\mu_{\Sigma_{\tau'}} + C\int_{\Sigma_{\tau'}\cap\{r\leq r_Y^-\}} J_\mu^N(S\Phi)n^\mu_{\Sigma_{\tau'}}\bigg)$$
$$+ CA^2\tau^{-2+\eta}\sum_{m+k+j\leq 4}\int_{\Sigma_{\tau_0}} J_\mu^{Z+CN,w^Z}(\partial_{t^*}^m\hat{Y}\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}.$$

Proof. This follows from using Proposition 33, taking $G_1 = 0$ and $G_2 = G$, and using Propositions 53 and 54 to estimate the terms with G. □

We are now ready to retrieve the bootstrap assumptions. First, we retrieve the assumption (23):

Proposition 59.

$$
\begin{aligned}
&c\int_{\Sigma_\tau} J^{Z,w^Z}_\mu(\partial_{t^*}S\Phi)n^\mu_{\Sigma_\tau}+\tau^2\int_{\Sigma_\tau\cap\{r\le\gamma\tau\}} J^N_\mu(\partial_{t^*}S\Phi)n^\mu_{\Sigma_\tau}\\
&\qquad\le \tfrac12 A\tau\sum_{m=1}^{2}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial^m_{t^*}S\Phi)n^\mu_{\Sigma_{\tau_0}}+\tfrac12 A\tau^{1+\eta}\sum_{m+k+j\le 5}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial^m_{t^*}\hat{Y}^k\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}.
\end{aligned}
$$

Proof. By Proposition 41,

$$
\begin{aligned}
&c\int_{\Sigma_\tau} J^{Z,w^Z}_\mu(\partial_{t^*}S\Phi)n^\mu_{\Sigma_\tau}+\tau^2\int_{\Sigma_\tau\cap\{r\le\gamma\tau\}} J^N_\mu(\partial_{t^*}S\Phi)n^\mu_{\Sigma_\tau}\\
&\quad\le C\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial_{t^*}S\Phi)n^\mu_{\Sigma_{\tau_0}}+C\iint_{\mathcal{R}(\tau_0,\tau)} t^*r^{-1+\delta}K^{X_1}(\partial_{t^*}S\Phi)\\
&\qquad+C\delta'\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\le\frac12 t^*\}}(t^*)^2K^{X_0}(\partial_{t^*}S\Phi)+C(\delta'+\epsilon)\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\le r_Y^-\}}(t^*)^2K^N(\partial_{t^*}S\Phi)\\
&\qquad+C(\delta')^{-1}\left(\int_{\tau_0}^{\tau}\left(\int_{\Sigma_{t^*}\cap\{r\ge\frac12 t^*\}} r^2(\partial_{t^*}G)^2\right)^{1/2}dt^*\right)^2+C(\delta')^{-1}\sum_{m=1}^{2}\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\le\frac{9}{10}t^*\}}(t^*)^2r^{1+\delta}(\partial^m_{t^*}G)^2\\
&\qquad\qquad+C(\delta')^{-1}\sup_{t^*\in[\tau_0,\tau]}\int_{\Sigma_{t^*}\cap\{r_Y^-\le r\le\frac{25}{8}M\}}(t^*)^2(\partial_{t^*}G)^2.
\end{aligned}
$$

It suffices to check that by Propositions 53, 54 and 55, all terms are acceptable. □

We can now retrieve the bootstrap assumption (24).

Proposition 60.

$$
\begin{aligned}
&c\int_{\Sigma_\tau} J^{Z,w^Z}_\mu(S\Phi)n^\mu_{\Sigma_\tau}+\tau^2\int_{\Sigma_\tau\cap\{r\le\gamma\tau\}} J^N_\mu(S\Phi)n^\mu_{\Sigma_\tau}\\
&\qquad\le A^2\tau^\eta\left(\sum_{m=0}^{2}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial^m_{t^*}S\Phi)n^\mu_{\Sigma_{\tau_0}}+\sum_{m+k+j\le5}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial^m_{t^*}\hat{Y}^k\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}\right).
\end{aligned}
$$

Proof. By Proposition 41,

$$
\begin{aligned}
c\int_{\Sigma_\tau} J^{Z,w^Z}_\mu(S\Phi)n^\mu_{\Sigma_\tau} + \tau^2\int_{\Sigma_\tau\cap\{r\leq\gamma\tau\}} J^N_\mu(S\Phi)n^\mu_{\Sigma_\tau}
&\leq C\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(S\Phi)n^\mu_{\Sigma_{\tau_0}} + C\iint_{\mathcal{R}(\tau_0,\tau)} t^* r^{-1+\delta}K^{X_1}(S\Phi)\\
&\quad + C\delta'\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq\frac12 t^*\}}(t^*)^2K^{X_0}(S\Phi) + C(\delta'+\epsilon)\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq r_Y^-\}}(t^*)^2K^N(S\Phi)\\
&\quad + C(\delta')^{-1}\left(\int_{\tau_0}^{\tau}\left(\int_{\Sigma_{t^*}\cap\{r\geq\frac12 t^*\}} r^2G^2\right)^{1/2}dt^*\right)^2 + C(\delta')^{-1}\sum_{m=0}^{1}\iint_{\mathcal{R}(\tau_0,\tau)\cap\{r\leq\frac{9}{10}t^*\}}(t^*)^2 r^{1+\delta}(\partial_{t^*}^m G)^2\\
&\quad + C(\delta')^{-1}\sup_{t^*\in[\tau_0,\tau]}\int_{\Sigma_{t^*}\cap\{r_Y^-\leq r\leq\frac{25}{8}M\}}(t^*)^2G^2.
\end{aligned}
$$

It suffices to check that by Propositions 53, 54 and 55, all terms are acceptable. ☐

We have thus shown the following:

Proposition 61. *For all $\eta>0$, there exists $\epsilon>0$ small enough such that for Kerr spacetimes satisfying* (3), *the following estimates hold*:

$$
\begin{aligned}
c\int_{\Sigma_\tau} J^{Z,w^Z}_\mu(S\Phi)n^\mu_{\Sigma_\tau} + \tau^2\int_{\Sigma_\tau\cap\{r\leq\gamma\tau\}} J^N_\mu(S\Phi)n^\mu_{\Sigma_\tau}
&\leq C\tau^\eta\sum_{m=0}^{2}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m S\Phi)n^\mu_{\Sigma_{\tau_0}} + C\tau^\eta\sum_{m+k+j\leq 5}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m\hat{Y}\tilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}.
\end{aligned}
$$

Moreover, for $\tau'\leq\tau\leq(1.1)\tau'$,

$$
\begin{aligned}
\iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq r_Y^-\}} K^N(S\Phi) + \iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq\frac12 t^*\}} K^{X_0}(S\Phi)
&\leq C\tau^{-2+\eta}\sum_{m=0}^{2}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m S\Phi)n^\mu_{\Sigma_{\tau_0}} + C\tau^{-2+\eta}\sum_{m+k+j\leq 5}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m\hat{Y}\tilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}.
\end{aligned}
$$

and

$$
\begin{aligned}
\iint_{\mathcal{R}(\tau',\tau)\cap\{r\leq\frac12 t^*\}} K^{X_1}(S\Phi)
&\leq C\tau^{-2+\eta}\sum_{m=0}^{3}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m S\Phi)n^\mu_{\Sigma_{\tau_0}} + C\tau^{-2+\eta}\sum_{m+k+j\leq 6}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m\hat{Y}\tilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}.
\end{aligned}
$$

Proof. The first statement is proved by the bootstrap above. Since the bootstrap assumptions are true, the conclusion in Proposition 58 is also true; hence the second statement is true. The third statement makes use of the fact that K^{X_1} can be estimated in the same way as K^{X_0} with an extra derivative. ☐

14. Improved decay for the linear homogeneous wave equation

To use the estimates for $S\Phi$, we need to integrate along integral curves of S. We first find the integral curves by solving the ordinary differential equation

$$\frac{dr_S}{dt_S^*} = \frac{h(r_S)}{t_S^*}$$

where $h(r_S)$ is as in the definition of S. Hence the integral curves are given by

$$\rho := \frac{\exp\left(\int_{(r_S)_0}^{r_S} \frac{dr_S'}{h(r_S')}\right)}{t_S^*} = \text{constant},$$

where $r_0 > 2M$ can be chosen arbitrarily. Let $\sigma = t^*$, and consider (σ, ρ, x^A, x^B) as a new system of coordinates. Notice that

$$\partial_\sigma = \frac{h(r_S)}{t^*}\partial_{r_S} + \partial_{t_S^*} = \frac{1}{t^*}S.$$

Now for each fixed ρ, we have

$$\Phi^2(\tau) \le \Phi^2(\tau') + \left|\int_{\tau'}^{\tau} \frac{1}{\sigma} S(\Phi^2)\, d\sigma\right|.$$

Integrating along a finite region of ρ, we get

$$\int_{\rho_1}^{\rho_2} \Phi^2(\tau)\, d\rho \le \int_{\rho_1}^{\rho_2} \Phi^2(\tau')\, d\rho + \int_{\rho_1}^{\rho_2}\int_{\tau'}^{\tau} \left|\frac{2}{\sigma}\Phi S\Phi\right| d\sigma\, d\rho.$$

We would like to change coordinates back to $(t_S^*, r_S, x_S^A, x_S^B)$. Since $h(r_S)$ is everywhere positive, (ρ, τ) would correspond to a point with a larger value of r than (ρ, τ'). Therefore,

$$\begin{aligned}\int_{r_+}^{r_2} \Phi^2(\tau) \frac{\exp\left(\int_{(r_S)_0}^{r_S} \frac{dr_S'}{h(r_S')}\right)}{\tau h(r_S)} dr \\ \le \int_{r_+}^{r_2} \Phi^2(\tau') \frac{\exp\left(\int_{(r_S)_0}^{r_S} \frac{dr_S'}{h(r_S')}\right)}{\tau' h(r_S)} dr + \int_{\tau'}^{\tau}\int_{r_+}^{r_2} \left|\frac{2}{\sigma}\Phi S\Phi\right| \frac{\exp\left(\int_{(r_S)_0}^{r_S} \frac{dr_S'}{h(r_S')}\right)}{t_S^* h(r_S)} dr\, dt^*.\end{aligned}$$

We have to compare $\exp\left(\int_{(r_S)_0}^{r_S} \frac{dr_S'}{h(r_S')}\right)/h(r_S)$ with the volume form. Very close to the horizon, we have $h(r_S) = r_S - 2M$. Hence

$$\frac{\exp\left(\int_{(r_S)_0}^{r_S} \frac{dr_S'}{h(r_S')}\right)}{h(r_S)} = e^{\int_{(r_S)_0}^{r_S} \frac{dr_S'}{h(r_S')}}\left(\frac{1}{r_S - 2M}\right) \sim 1.$$

The corresponding expression on the compact set $[r_Y^-, R]$ is obviously bounded. Hence we have

$$\int_{\Sigma_\tau \cap \{r<r_2\}} \frac{\Phi^2(\tau)}{\tau} \le C\left(\int_{\Sigma_{\tau'} \cap \{r<r_2\}} \frac{\Phi^2(\tau')}{\tau'} + \iint_{\mathcal{R}(\tau',\tau)\cap\{r<r_2\}} \left|\frac{2}{(t^*)^2}\Phi S\Phi\right|\right). \tag{25}$$

This easily implies the following improved decay for the nondegenerate energy:

Proposition 62. $$\int_{\Sigma_\tau\cap\{r<R\}}\Phi^2\le C_R\tau^{-1}\Big(\iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)\cap\{r<R\}}\Phi^2+\iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)\cap\{r<R\}}(S\Phi)^2\Big)R.$$

Proof. By choosing an appropriate $\tilde{\tau}\in[(1.1)^{-1}\tau,\tau]$, we have

$$\int_{\Sigma_{\tilde{\tau}}\cap\{r<R\}}\Phi^2\le C\tau^{-1}\iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)\cap\{r<R\}}\Phi^2.$$

Now, apply (25) with $\tau'=\tilde{\tau}$, we have

$$\begin{aligned}\int_{\Sigma_\tau\cap\{r<R\}}\Phi^2&\le C\tau\Big(\int_{\Sigma_{\tilde{\tau}}\cap\{r<R\}}\frac{\Phi^2}{\tilde{\tau}}+\iint_{\mathcal{R}(\tilde{\tau},\tau)\cap\{r<R\}}\Big|\frac{2}{(t^*)^2}\Phi S\Phi\Big|\Big)\\&\le C\tau^{-1}\Big(\iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)\cap\{r<R\}}\Phi^2+\iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)\cap\{r<R\}}(S\Phi)^2\Big),\end{aligned}$$

using Cauchy–Schwarz for the second term. □

We can now conclude with the improved decay for solutions to the homogeneous wave equation.

Proof of Main Theorem. By Corollaries 43, 62 and 61, we have

$$\begin{aligned}\int_{\Sigma_\tau\cap\{r<R\}}\Phi^2&\le C_R\tau^{-1}\Big(\iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)\cap\{r<R\}}\Phi^2+\iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)\cap\{r<R\}}(S\Phi)^2\Big)\\&\le C_R\tau^{-1}\iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)\cap\{r<R\}}(K^{X_0}(\Phi)+K^{X_0}(S\Phi))\\&\le C_R\tau^{-3+\eta}\Big(\sum_{m=0}^{2}\int_{\Sigma_{\tau_0}}J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m S\Phi)n^\mu_{\Sigma_{\tau_0}}+\sum_{m+k+j\le 5}\int_{\Sigma_{\tau_0}}J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m\hat{Y}\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}\Big)\end{aligned}$$

Similarly we can use Proposition 62 for the derivatives of Φ. By Corollary 43, 62 and 61, we have

$$\begin{aligned}&\int_{\Sigma_\tau\cap\{r<R\}}(D\Phi)^2\\&\quad\le C_R\tau^{-1}\Big(\iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)\cap\{r<R\}}(D\Phi)^2+\iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)\cap\{r<R\}}(SD\Phi)^2\Big)\\&\quad\le C_R\tau^{-1}\iint_{\mathcal{R}((1.1)^{-1}\tau,\tau)\cap\{r<R\}}(K^{X_1}(\Phi)+K^{X_1}(S\Phi))\\&\qquad\qquad\qquad\qquad\text{(since we have the commutation } [D,S]=D)\\&\quad\le C_R\tau^{-3+\eta}\Big(\sum_{m=0}^{3}\int_{\Sigma_{\tau_0}}J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m S\Phi)n^\mu_{\Sigma_{\tau_0}}+\sum_{m+k+j\le 6}\int_{\Sigma_{\tau_0}}J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m\hat{Y}\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}\Big).\end{aligned}$$

By commuting with ∂_{t^*}, we get

$$\begin{aligned}&\int_{\Sigma_\tau\cap\{r<R\}}(D\partial_{t^*}^\ell\Phi)^2\\&\quad\le C_R\tau^{-3+\eta}\Big(\sum_{m=0}^{\ell+3}\int_{\Sigma_{\tau_0}}J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m S\Phi)n^\mu_{\Sigma_{\tau_0}}+\sum_{m+k+j\le \ell+6}\int_{\Sigma_{\tau_0}}J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m\hat{Y}\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}\Big).\end{aligned}$$

Without loss of generality, we can take $R > \frac{23}{8}M$. Then, by Proposition 47,

$$\sum_{j+m\leq \ell}\int_{\Sigma_\tau\cap\{r\leq r_Y^+\}} J^N_\mu(\partial_{t^*}^j \hat{Y}^m\Phi)n^\mu_{\Sigma_\tau}$$
$$\leq C_R\tau^{-3+\eta}\biggl(\sum_{m=0}^{\ell+3}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m S\Phi)n^\mu_{\Sigma_{\tau_0}} + \sum_{m+k+j\leq \ell+6}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m \hat{Y}\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}\biggr).$$

Hence, by Proposition 44 and 45,

$$\sum_{j=0}^{\ell}\int_{\Sigma_\tau\cap\{r\leq R\}} (D^j\Phi)^2$$
$$\leq C_R\tau^{-3+\eta}\biggl(\sum_{m=0}^{\ell+2}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m S\Phi)n^\mu_{\Sigma_{\tau_0}} + \sum_{m+k+j\leq \ell+5}\int_{\Sigma_{\tau_0}} J^{Z+CN,w^Z}_\mu(\partial_{t^*}^m \hat{Y}\widetilde{\Omega}^j\Phi)n^\mu_{\Sigma_{\tau_0}}\biggr).$$

The pointwise decay statement follows from standard Sobolev embedding. □

15. Discussion

Out main paper holds in the set $\{r_+ \leq r \leq R\}$ for any fixed R. It is however interesting also to derive the same estimates, for example, in the set $\{r_+ \leq r \leq \frac{1}{2}t^*\}$. This can be achieved by proving the full decay result when we commuted the equation with $\widetilde{\Omega}^\ell$. Using this we can prove (with more loss in derivatives) that

$$|\Phi| \leq CE(t^*)^{-3/2+\eta}r^\eta \quad \text{and} \quad |D\Phi| \leq CE(t^*)^{-3/2+\eta}r^{-\frac{1}{2}+\eta},$$

for $r \leq \frac{1}{2}t^*$. This will be useful in studying nonlinear problems. This decay rate will be proved as a corollary in our forthcoming paper on the null condition.

Acknowledgments

I thank my advisor Igor Rodnianski for continual support and encouragement and for many enlightening discussions. I thank Gustav Holzegel for very helpful comments on the manuscript. I thank also an anonymous referee for many suggestions to improve the manuscript.

References

[Alinhac 2009] S. Alinhac, "Energy multipliers for perturbations of the Schwarzschild metric", *Comm. Math. Phys.* **288**:1 (2009), 199–224. MR 2010b:58038 Zbl 1196.53053

[Andersson and Blue 2009] L. Andersson and P. Blue, "Hidden symmetries and decay for the wave equation on the Kerr spacetime", preprint, 2009. arXiv 0908.2265

[Blue and Soffer 2006] P. Blue and A. Soffer, "Improved decay rates with small regularity loss for the wave equation about a Schwarzschild black hole", preprint, 2006. arXiv math.AP/0612168

[Blue and Sterbenz 2006] P. Blue and J. Sterbenz, "Uniform decay of local energy and the semi-linear wave equation on Schwarzschild space", *Comm. Math. Phys.* **268**:2 (2006), 481–504. MR 2007i:58037 Zbl 1123.58018

[Carter 1968] B. Carter, "Global structure of the Kerr family of gravitational fields", *Phys. Rev., II. Ser.* **174**:5 (1968), 1559–1571. Zbl 0167.56301

[Christodoulou and Klainerman 1993] D. Christodoulou and S. Klainerman, *The global nonlinear stability of the Minkowski space*, Princeton Mathematical Series **41**, Princeton University Press, Princeton, NJ, 1993. MR 95k:83006 Zbl 0827.53055

[Dafermos and Rodnianski 2008] M. Dafermos and I. Rodnianski, "Lectures on black holes and linear waves", preprint, 2008. arXiv 0811.0354

[Dafermos and Rodnianski 2009] M. Dafermos and I. Rodnianski, "The red-shift effect and radiation decay on black hole spacetimes", *Comm. Pure Appl. Math.* **62**:7 (2009), 859–919. MR 2011b:83059 Zbl 1169.83008

[Dafermos and Rodnianski 2010] M. Dafermos and I. Rodnianski, "A new physical-space approach to decay for the wave equation with applications to black hole spacetimes", pp. 421–432 in *XVIth International Congress on Mathematical Physics* (Prague, 2009), edited by P. Exner, World Scientific, Hackensack, NJ, 2010. MR 2012e:58051 Zbl 1211.83019

[Dafermos and Rodnianski 2011] M. Dafermos and I. Rodnianski, "A proof of the uniform boundedness of solutions to the wave equation on slowly rotating Kerr backgrounds", *Invent. Math.* **185**:3 (2011), 467–559. MR 2827094 Zbl 1226.83029

[Donninger et al. 2011] R. Donninger, W. Schlag, and A. Soffer, "A proof of Price's law on Schwarzschild black hole manifolds for all angular momenta", *Adv. Math.* **226**:1 (2011), 484–540. MR 2012d:58043 Zbl 1205.83041 arXiv 0908.4292

[Finster et al. 2006] F. Finster, N. Kamran, J. Smoller, and S.-T. Yau, "Decay of solutions of the wave equation in the Kerr geometry", *Comm. Math. Phys.* **264**:2 (2006), 465–503. MR 2007b:83019 Zbl 1194.83015

[Finster et al. 2008] F. Finster, N. Kamran, J. Smoller, and S.-T. Yau, "Erratum to [Finster et al. 2006]", *Comm. Math. Phys.* **280**:2 (2008), 563–573. MR 2009c:83008 Zbl 1194.83014

[Hartle and Wilkins 1974] J. B. Hartle and D. C. Wilkins, "Analytic properties of the Teukolsky equation", *Comm. Math. Phys.* **38** (1974), 47–63. MR 50 #3848

[Kay and Wald 1987] B. S. Kay and R. M. Wald, "Linear stability of Schwarzschild under perturbations which are nonvanishing on the bifurcation 2-sphere", *Classical Quantum Gravity* **4**:4 (1987), 893–898. MR 88m:83043 Zbl 0647.53065

[Klainerman and Sideris 1996] S. Klainerman and T. C. Sideris, "On almost global existence for nonrelativistic wave equations in 3D", *Comm. Pure Appl. Math.* **49**:3 (1996), 307–321. MR 96m:35231 Zbl 0867.35064

[Kronthaler 2007] J. Kronthaler, "Decay rates for spherical scalar waves in the Schwarzschild geometry", preprint, 2007. arXiv 0709.3703

[Lindblad and Rodnianski 2005] H. Lindblad and I. Rodnianski, "Global existence for the Einstein vacuum equations in wave coordinates", *Comm. Math. Phys.* **256**:1 (2005), 43–110. MR 2006b:83020 Zbl 1081.83003

[Luk 2010] J. Luk, "Improved decay for solutions to the linear wave equation on a Schwarzschild black hole", *Ann. Henri Poincaré* **11**:5 (2010), 805–880. MR 2736525 Zbl 1208.83068

[Machedon and Stalker 2002] M. Machedon and J. Stalker, "Decay of solutions to the wave equation on a spherically symmetric background", preprint, 2002.

[Morawetz 1975] C. S. Morawetz, *Notes on time decay and scattering for some hyperbolic problems*, Regional Conference Series in Applied Mathematics **19**, Society for Industrial and Applied Mathematics, Philadelphia, 1975. MR 58 #11968 Zbl 0303.35002

[Press and Teukolsky 1973] W. H. Press and S. A. Teukolsky, "Pertubations of a rotating black hole, II: Dynamical stability of the Kerr metric", *Astrophys. J.* **185** (1973), 649–673.

[Sterbenz 2005] J. Sterbenz, "Angular regularity and Strichartz estimates for the wave equation", *Int. Math. Res. Not.* **2005**:4 (2005), 187–231. MR 2006i:35212 Zbl 1072.35048

[Tataru 2009] D. Tataru, "Local decay of waves on asymptotically flat stationary space-times", preprint, 2009. arXiv 0910.5290

[Tataru and Tohaneanu 2011] D. Tataru and M. Tohaneanu, "A local energy estimate on Kerr black hole backgrounds", *Int. Math. Res. Not.* **2011**:2 (2011), 248–292. MR 2012a:58050 Zbl 1209.83028 arXiv 0810.5766

[Tohaneanu 2012] M. Tohaneanu, "Strichartz estimates on Kerr black hole backgrounds", *Trans. Amer. Math. Soc.* **364**:2 (2012), 689–702. MR 2846348 Zbl 1234.35275 arXiv 0910.1545

[Wald 1979] R. M. Wald, "Note on the stability of the Schwarzschild metric", *J. Math. Phys.* **20**:6 (1979), 1056–1058. MR 80c:83012a

[Whiting 1989] B. F. Whiting, "Mode stability of the Kerr black hole", *J. Math. Phys.* **30**:6 (1989), 1301–1305. MR 90m:83038 Zbl 0689.53041

Received 5 Oct 2010. Revised 25 Mar 2011. Accepted 21 Apr 2011.

JONATHAN LUK: jluk@math.princeton.edu
Department of Mathematics, Princeteon University, Fine Hall, Washington Road, Princeton, NJ 08544, United States

ANALYSIS AND PDE
Vol. 5, No. 3, 2012
dx.doi.org/10.2140/apde.2012.5.627

ON THE BOGOLYUBOV–RUZSA LEMMA

TOM SANDERS

Our main result is that if A is a finite subset of an abelian group with $|A+A| \leqslant K|A|$, then $2A-2A$ contains an $O(\log^{O(1)} 2K)$-dimensional coset progression M of size at least $\exp(-O(\log^{O(1)} 2K))|A|$.

1. Introduction

Croot and Sisask [2010] introduced a fundamental new method to additive combinatorics and, although they have already given a number of applications, our present purpose is to give another. Specifically, we shall prove the following.

Theorem 1.1 (Bogolyubov–Ruzsa lemma for abelian groups). *Suppose that G is an* (*discrete*) *abelian group and $A, S \subset G$ are finite nonempty sets such that $|A+S| \leqslant K \min\{|A|, |S|\}$. Then $(A-A)+(S-S)$ contains a proper symmetric $d(K)$-dimensional coset progression M of size* $\exp(-h(K))|A+S|$. *Moreover*, *we may take $d(K) = O(\log^6 2K)$ and $h(K) = O(\log^6 2K \log 2 \log 2K)$.*

We should take a moment to justify the name, which is slightly nonstandard. Bogolyubov's lemma (the idea for which originates in [Bogolyubov 1939]) is usually stated for sets of large density in the ambient group, rather than small doubling, and asserts that the fourfold sumset of a thick set contains a large Bohr set.

Ruzsa [1994], on his way to proving Freĭman's theorem, showed that a set with small doubling could be sensibly embedded into a group where it is thick. He then applied Bogolyubov's lemma and proceeded to show that a Bohr set contains a large generalised arithmetic progression which could then be pulled back. In doing all this he implicitly proved the first version of Theorem 1.1 in $\mathbb{Z}$ — although, with different bounds — and this motivates the name.

This result has many variants (although the form given above seems to be a fairly useful one) and in light of this the history is not completely transparent. Certainly most proofs of Freĭman's theorem broadly following the model of [Ruzsa 1994] will implicitly prove a result of this shape. With this in mind the extension from $\mathbb{Z}$ to arbitrary abelian groups is due to Green and Ruzsa [2007], and the first good bounds to Schoen [2011] for certain classes of groups.

There are many applications of results of this type, particularly since their popularisation by Gowers [1998], and we shall deal with a number of these in Section 11 at the end of the paper. To help explain the main ideas we include a discursive sketch of the paper after the next section, which simply sets some notation.

MSC2010: 11L07.
Keywords: Freiman, Fourier analysis, sumsets, generalised arithmetic progressions, coset progressions, small doubling.

2. Notation

The main tool used in the paper is Fourier analysis on groups for which the classic reference is [Rudin 1990]. We deal almost exclusively with finite groups in the paper, but to be complete we shall need slightly more generality.

Suppose that G is a locally compact topological group. We write $C(G)$ for the space of continuous complex-valued functions on G. More generally if $R \subset \mathbb{C}$ we write $C(G, R)$ for the continuous R-valued functions on G.

The group structure on G induces an action of G on $C(G)$ called translation. In particular if $x \in G$ and $f \in C(G)$ then we write

$$\rho_x(f)(y) := f(yx) \quad \text{for all } y \in G. \tag{2-1}$$

We also write $M(G)$ for the space of regular Borel measures on G and can extend ρ to these in the natural way: for $x \in G$ and $\mu \in M(G)$, $\rho_x(\mu)$ is the measure induced by

$$C(G) \to C(G); \quad f \mapsto \int f(x)\, d\mu(yx).$$

The group structure on G is reflected in $M(G)$ in a fairly natural way and we define the convolution of two measures $\mu, \nu \in M(G)$ to be the measure $\mu * \nu$ induced by

$$C(G) \to C(G); \quad f \mapsto \int f(xy)\, d\mu(x)\, d\nu(y).$$

There is a family of privileged measures on G called Haar measures. These are the translation-invariant measures on G: $\mu \in M(G)$ is a Haar measure on G if $\rho_x(\mu) = \mu$ for all $x \in G$.

Given a Haar measure μ on G we can extend ρ in the obvious way from (2-1) to define the right regular representation $\rho : G \to \mathrm{Aut}(L^2(\mu))$. More than this we can define the convolution of two functions $f, g \in L^1(\mu)$ by

$$f * g(x) := \int f(y)g(y^{-1}x)\, d\mu(y) \quad \text{for all } x \in G.$$

There are two particularly useful instances of Haar measure depending on the topology on G: if G is compact we write μ_G for the Haar probability measure on G, while if G is discrete we write δ_G for the Haar counting measure on G, which assigns mass 1 to each element of G.

Of course, if G is finite it is both discrete and compact so one has both probability measure and counting measure to choose from. The measures are multiples of each other as μ_G is just the measure assigning mass $|G|^{-1}$ to each element of G. More generally given a finite set X we write μ_X for the measure assigning mass $|X|^{-1}$ to each $x \in X$.

When it is relevant we shall indicate whether we are taking a finite group G to be compact or discrete by declaring the group either compact, so that μ_G is to be used, or discrete so that δ_G is to be used. The reader should be aware that this has the effect of changing the normalisations in convolutions.

The above all works for general finite groups G, but when G is also abelian convolution operators can be written in a particularly simple form with respect to the Fourier basis which we now recall.

We write $\widehat{G}$ for the dual group, that is the finite abelian group of homomorphisms $\gamma : G \to S^1$, where $S^1 := \{z \in \mathbb{C} : |z| = 1\}$. Given $\mu \in M(G)$ we define $\widehat{\mu} \in \ell^\infty(\widehat{G})$ by

$$\widehat{\mu}(\gamma) := \int \overline{\gamma}\, d\mu \quad \text{ for all } \gamma \in \widehat{G},$$

and extend this to $f \in L^1(\mu_G)$ by $\widehat{f} := \widehat{f d\mu_G}$. It is easy to check that $\widehat{\mu * \nu} = \widehat{\mu} \cdot \widehat{\nu}$ for all $\mu, \nu \in M(G)$ and $\widehat{f * g} = \widehat{f} \cdot \widehat{g}$ for all $f, g \in L^1(\mu_G)$.

3. A sketch of the argument

Assuming the hypotheses of Theorem 1.1 our objective will be to show that there is a large, low-dimensional coset progression M correlated with $A + S$, meaning such that

$$\|1_{A+S} * \mu_M\|_{\ell^\infty(G)} > 1 - o(1).$$

This is essentially the statement of Theorem 10.1 later, and Theorem 1.1 can be derived from it by a simple pigeonholing argument.

A simplified argument: the case of good modelling. We shall assume that we have good modelling in the sense of [Green and Ruzsa 2007], meaning that we shall assume that the sets A and S have density $K^{-O(1)}$ in the ambient group. This can actually be arranged in the two cases of greatest interest: $\mathbb{F}_2^n$ and $\mathbb{Z}$ and facilitates considerable simplifications.

A very useful observation in [López and Ross 1975] is that because the support of $\mu_A * \mu_S$ is contained in $A + S$ we have the identity

$$\langle 1_{A+S} * \mu_{-S}, \mu_A \rangle = 1.$$

Now, suppose we had a coset progression M over which $1_{A+S} * \mu_{-S}$ was in some sense invariant, meaning

$$\|1_{A+S} * \mu_{-S} * \mu_M - 1_{A+S} * \mu_{-S}\|_{\ell^p(G)} \leqslant \epsilon \|1_{A+S}\|_{\ell^p(G)}. \tag{3-1}$$

Then Hölder's inequality and the López–Ross identity tell us that

$$|\langle 1_{A+S} * \mu_{-S} * \mu_M, \mu_A \rangle - 1| \leqslant \epsilon \|1_{A+S}\|_{\ell^p(G)} \|\mu_A\|_{\ell^{p/(p-1)}(G)} \leqslant \epsilon K^{1/p},$$

and it follows by averaging that $A + S$ is correlated with M provided that $\epsilon \sim K^{-1/p}$.

The traditional Fourier analytic approach to finding an M such that (3-1) holds is not particularly efficient, but recently Croot and Sisask showed that there is, at least, a set Z such that we have (3-1) with Z in place of M and

$$\mu_G(Z) \geqslant \exp(-O(\epsilon^{-2} p \log K))\mu_G(A).$$

Moreover, they noted by the triangle inequality that one can endow Z with the structure of a k-fold sumset, so that we have (3-1) with kX in place of M and

$$\mu_G(X) \geqslant \exp(-O(k^2\epsilon^{-2} p \log K))\mu_G(A) = \exp(-O(k^2 \log^2 K))\mu_G(A), \tag{3-2}$$

where the third term is by optimising the choice of $p \sim \log K$ given that $\epsilon \sim K^{-1/p}$.

What we actually end up with after all this is a set X with density as described in (3-2) such that

$$\langle 1_{A+S} * \mu_{-S} * \mu_X^{(k)}, \mu_A\rangle > 1 - o(1). \tag{3-3}$$

Now, by the usual sorts of applications of Plancherel's theorem and Cauchy–Schwarz we find that most of the Fourier mass of the inner product is concentrated on those characters in $\mathrm{Spec}_{1/2}(1_X)$ provided $2^k \sim K$, and so we choose $k \sim \log K$.

With most of the Fourier mass supported on $\mathrm{Spec}_{1/2}(1_X)$, it follows that the integrand in (3-3) correlates with any set which approximately annihilates $\mathrm{Spec}_{1/2}(1_X)$. It remains to show that the approximate annihilator of $\mathrm{Spec}_{1/2}(1_X)$ — that is the Bohr set B with $\mathrm{Spec}_{1/2}(1_X)$ as its frequency set — contains a large coset progression.

We can now apply Chang's theorem to get that B is low-dimensional and then the usual geometry of numbers argument tells us that this Bohr set contains a large coset progression, and the result is proved.

Extending the argument: the case of bad modelling. We now drop the assumption of good modelling, and the argument proceeds in essentially the same way up until the application of Chang's theorem above.

In this case Chang's theorem does not provide good bounds. Instead what we do is note that the set X satisfies a relative polynomial growth condition

$$|nX| \leqslant n^{O(\log^4 K)}|X| \quad \text{for all } n \geqslant 1.$$

This lets us produce a Bohr set containing X which behaves enough like a group for a relative version of Chang's theorem to hold, whilst at the same time X is much denser in the Bohr set than it would be in the modelling group.

Since we are not using modelling what we have just done does not actually give us a Bohr set of low dimension, but rather a Bohr set of size comparable to X which has a lower order of polynomial growth on a certain range. It turns out that the usual argument that shows a low-dimensional Bohr set contains a large coset progression can be adapted relatively easily to this more general setting and this gives us our final ingredient.

These arguments are spread over the paper as follows. The simplified argument up to (3-3) is essentially contained in Section 4. Then, in Section 5, we record the basic properties of Bohr sets we need before Section 6, which has the relative version of Chang's theorem, and Section 7, which puts the material together to take a set satisfying a relative polynomial growth condition and produce a large Bohr superset.

After the material on Bohr sets we have Section 8 which records some standard covering lemmas and then Section 9 where we show how to find a large coset progression in a Bohr set with relative polynomial growth. Finally the argument is all put together in Section 10.

4. Freĭman-type theorems in arbitrary groups

In this section we are interested in Freĭman-type theorems in arbitrary, possibly nonabelian, groups. There has been considerable work towards such results, although often with restrictions on the type of nonabelian groups considered, or rather weak bounds. We direct the reader to [Green 2009] for a survey, but our interest is narrower, lying with a crucial result of Tao [2010, Proposition C.3] which inspires the following.

Proposition 4.1. *Suppose that G is a (discrete) group, $A, S \subset G$ are finite nonempty sets such that $|AS| \leqslant K \min\{|A|, |S|\}$, and $k \in \mathbb{N}$ is a parameter. Then $A^{-1}ASS^{-1}$ contains X^k where X is a symmetric neighbourhood of the identity with size $\delta(k, K)|AS|$. Moreover, we may take $\delta(k, K) \geqslant \exp(-O(k^2 \log^2 2K))$.*

Note that this result is a very weak version of Theorem 1.1 but for any group, not just abelian groups, and despite its weaknesses, its generality makes it useful in some situations.

Proposition 4.1 was essentially proved in [Croot and Sisask 2010, Theorem 1.6] with weaker K-dependence in the bound, using the $p = 2$ version of their Lemma 4.3 below. It turns out that we shall be able to show the above bound by coupling the large p case of their result with the López–Ross identity.

The key proposition of this section, then, is the following.

Proposition 4.2. *Suppose that G is (discrete) a group, $A, S, T \subset G$ are finite nonempty sets such that $|AS| \leqslant K|A|$ and $|TS| \leqslant L|S|$, and $k \in \mathbb{N}$ and $\epsilon \in (0, 1]$ are a pair of parameters. Then there is a symmetric neighbourhood of the identity $X \subset G$ with*

$$|X| \geqslant \exp(-O(\epsilon^{-2} k^2 \log 2K \log 2L))|T|$$

such that

$$|\mu_{A^{-1}} * 1_{AS} * \mu_{S^{-1}}(x) - 1| \leqslant \epsilon \text{ for all } x \in X^k.$$

The main ingredient in the proof of this is the following result, which is essentially due to Croot and Sisask [2010, Proposition 3.3]. To prove it they introduced the idea of sampling from physical space rather than Fourier space — sampling in Fourier space can be seen as the main idea in Chang's theorem. Not only does this work in settings where the Fourier transform is less well behaved, but it also runs much more efficiently, which leads to the superior bounds.

We include the proof since it is the pivotal ingredient of this paper, and we frame it in such a way as to emphasise the parallels with Chang's theorem.

Lemma 4.3 (Croot–Sisask). *Suppose that G is a (discrete) group, $f \in \ell^p(G)$ for $p \geqslant 2$ and $S, T \subset G$ are nonempty with $|ST| \leqslant K|S|$. Then there is a $t \in T$ and a set $X \subset Tt^{-1}$ with $|X| \geqslant (2K)^{-O(\epsilon^{-2}p)}|T|$ such that*

$$\|\rho_x(f * \mu_S) - f * \mu_S\|_{\ell^p(G)} \leqslant \epsilon \|f\|_{\ell^p(G)} \quad \text{for all } x \in X.$$

Proof. Let $z_1, \dots, z_k$ be independent uniformly distributed S-valued random variables, and for each $y \in G$ define $Z_i(y) := \rho_{z_i^{-1}}(f)(y) - f * \mu_S(y)$. For fixed y, the variables $Z_i(y)$ are independent and have mean zero, so it follows by the Marcinkiewicz–Zygmund inequality and Hölder's inequality that

$$\left\|\sum_{i=1}^{k} Z_i(y)\right\|_{L^p(\mu_S^k)}^p \leqslant O(p)^{p/2}\int\left(\sum_{i=1}^{k}|Z_i(y)|^2\right)^{p/2} d\mu_S^k \leqslant O(p)^{p/2}k^{p/2-1}\sum_{i=1}^{k}\int |Z_i(y)|^p\, d\mu_S^k.$$

Summing over y and interchanging the order of summation we get

$$\sum_{y\in G}\left\|\sum_{i=1}^{k} Z_i(y)\right\|_{L^p(\mu_S^k)}^p \leqslant O(p)^{p/2}k^{p/2-1}\int\sum_{i=1}^{k}\sum_{y\in G}|Z_i(y)|^p\, d\mu_S^k. \tag{4-1}$$

On the other hand,

$$\left(\sum_{y\in G}|Z_i(y)|^p\right)^{1/p} = \|Z_i\|_{\ell^p(G)} \leqslant \|\rho_{z_i^{-1}}(f)\|_{\ell^p(G)} + \|f*\mu_S\|_{\ell^p(G)} \leqslant 2\|f\|_{\ell^p(G)}$$

by the triangle inequality. Dividing (4-1) by k^p and inserting the above and the expression for the Z_is we get

$$\int\sum_{y\in G}\left|\frac{1}{k}\sum_{i=1}^{k}\rho_{z_i^{-1}}(f)(y) - f*\mu_S(y)\right|^p d\mu_S^k(z) = O(pk^{-1}\|f\|_{\ell^p(G)}^2)^{p/2}.$$

Pick $k = O(\epsilon^{-2}p)$ such that the right-hand side is at most $(\epsilon\|f\|_{\ell^p(G)}/4)^p$ and write L for the set of $x \in S\times\cdots\times S$ (where the Cartesian product is k-fold) for which the integrand above is at most $(\epsilon\|f\|_{\ell^p(G)}/2)^p$; by averaging $\mu_S^k(L^c) \leqslant 2^{-p}$ and so $\mu_S^k(L) \geqslant 1-2^{-p} \geqslant \frac{1}{2}$.

Now, $\Delta := \{(t,\dots,t) : t \in T\}$ has $L\Delta \subset ST\times\cdots\times ST$, whence $|L\Delta| \leqslant 2K^k|L|$ and so

$$\langle 1_\Delta * 1_{\Delta^{-1}}, 1_{L^{-1}} * 1_L\rangle_{\ell^2(G\times\cdots\times G)} = \|1_L * 1_\Delta\|_{\ell^2(G\times\cdots\times G)}^2 \geqslant |\Delta|^2|L|/2K^k,$$

by the Cauchy–Schwarz inequality since the adjoint of $g \mapsto 1_L * g$ is $g \mapsto 1_{L^{-1}} * g$ and similarly for $g \mapsto g * 1_\Delta$.

By averaging it follows that at least $|\Delta|^2/2K^k$ pairs $(z,y) \in \Delta\times\Delta$ have $1_{L^{-1}} * 1_L(zy^{-1}) > 0$, and hence there is some $t \in T$ such that there is a set $X \subset Tt^{-1}$ of size at least $|T|/2K^k$ elements with $1_{L^{-1}} * 1_L(x,\dots,x) > 0$ for all $x \in X$.

Thus for each $x \in X$ there is some $z(x) \in L$ and $y(x) \in L$ such that $y(x)_i = z(x)_i x$. But then by the triangle inequality we get

$$\begin{aligned}&\|\rho_{x^{-1}}(f*\mu_S) - f*\mu_S\|_{\ell^p(G)}\\ &\quad\leqslant \left\|\rho_{x^{-1}}\left(\frac{1}{k}\sum_{i=1}^{k}\rho_{z(x)_i^{-1}}(f)\right) - f*\mu_S\right\|_{\ell^p(G)} + \left\|\rho_{x^{-1}}\left(\frac{1}{k}\sum_{i=1}^{k}\rho_{z(x)_i^{-1}}(f) - f*\mu_S\right)\right\|_{\ell^p(G)}.\end{aligned}$$

However, since ρ_x is isometric on $\ell^p(G)$ we see that

$$\left\|\rho_x(f*\mu_S)-f*\mu_S\right\|_{\ell^p(G)} \leqslant \left\|\frac{1}{k}\sum_{i=1}^{k}\rho_{y(x)_i^{-1}}(f)-f*\mu_S\right\|_{\ell^p(G)} + \left\|\frac{1}{k}\sum_{i=1}^{k}\rho_{z(x)_i^{-1}}(f)-f*\mu_S\right\|_{\ell^p(G)},$$

and we are done since $z(x), y(x) \in L$. $\square$

The important thing to note about the Croot–Sisask lemma is that the p-dependence of the size of the set X is very good. The natural Fourier analytic analogue (essentially given in [Bourgain 1990], and clearly exposited in [Sisask 2009]) gives an exponentially worse bound. To make use of this strength we use the aforementioned López–Ross identity.

Proof of Proposition 4.2. We apply Lemma 4.3 to the function $f := 1_{AS}$ and with the set S^{-1} (so that $|S^{-1}T^{-1}| \leqslant L|S^{-1}|$) to get a set X with $|X| \geqslant (2L)^{O(\epsilon^{-2}k^2p)}|T|$ such that

$$\|\rho_x(1_{AS} * \mu_{S^{-1}}) - 1_{AS} * \mu_{S^{-1}}\|_{\ell^p(G)} \leqslant \frac{\epsilon\|1_{AS}\|_{\ell^p(G)}}{ek} \quad \text{for all } x \in X.$$

Since ρ is isometric on $\ell^p(G)$ and ρ_{1_G} is the identity we may certainly assume that X is a symmetric neighbourhood of the identity. Furthermore, by the triangle inequality we have

$$\|\rho_x(1_{AS} * \mu_{S^{-1}}) - 1_{AS} * \mu_{S^{-1}}\|_{\ell^p(G)} \leqslant \epsilon e^{-1}\|1_{AS}\|_{\ell^p(G)} \quad \text{for all } x \in X^k.$$

Now for any (real) function g we have

$$\mu_{A^{-1}} * g(x) - \mu_{A^{-1}} * g(1_G) = \mu_{A^{-1}} * (\rho_x(g) - g)(1_G) = \langle \mu_A, \rho_x(g) - g\rangle.$$

Thus by Hölder's inequality we have

$$|\mu_{A^{-1}} * g(x) - \mu_{A^{-1}} * g(1_G)| \leqslant \|\mu_A\|_{\ell^{p'}(G)}\|\rho_x(g) - g\|_{\ell^p(G)}.$$

Putting $g = 1_{AS} * \mu_{S^{-1}}$ we conclude that

$$\begin{aligned}|\mu_{A^{-1}} * 1_{AS} * \mu_{S^{-1}}(x) - \mu_{A^{-1}} * 1_{AS} * \mu_{S^{-1}}(1_G)| &\leqslant \frac{\epsilon\|\mu_A\|_{\ell^{p'}(G)}\|1_{AS}\|_{\ell^p(G)}}{e} \\ &\leqslant \frac{\epsilon|A|^{1/p'}|AS|^{1/p}}{e|A|} \leqslant \frac{\epsilon K^{1/p}}{e}\end{aligned}$$

for all $x \in X^k$. Putting $p := 2 + \log K$ we get the conclusion. $\square$

Proof of Proposition 4.1. We simply take $T = A$, $L = K$ and $\epsilon = \frac{1}{2}$ in Proposition 4.2. $\square$

5. Basic properties of Bohr sets

Following [Bourgain 2008] we use a slight generalisation of the traditional notion of Bohr set, letting the width parameter vary according to the character. The advantage of this definition is that the meet of two Bohr sets in the lattice of Bohr sets is then just their intersection.

Throughout the section we let G be a finite (compact) abelian group. A set B is called a *Bohr set* if there is a *frequency set* Γ of characters on G, and a *width function* $\delta \in (0, 2]^\Gamma$ such that

$$B = \{x \in G : |1 - \gamma(x)| \leqslant \delta_\gamma \text{ for all } \gamma \in \Gamma\}.$$

Technically the same Bohr set can be defined by different frequency sets and width functions; we make the standard abuse that when we introduce a Bohr set we are implicitly fixing a frequency set and width function.

There is a natural way of dilating Bohr sets which will be of particular use to us. For a Bohr set B and $\rho \in \mathbb{R}^+$ we denote by B_ρ the Bohr set with frequency set Γ and width function[1] $\rho\delta$ so that, in particular, $B = B_1$ and more generally $(B_\rho)_{\rho'} = B_{\rho\rho'}$.

Given two Bohr sets B and B' we define their *intersection* to be the Bohr set with frequency set $\Gamma \cup \Gamma'$ and width function $\delta \wedge \delta'$. A simple averaging argument (see [Tao and Vu 2006, Lemma 4.20] but also the end of Lemma 4.3) can be used to see that the intersection of several Bohr sets is large.

Lemma 5.1 (intersections of Bohr sets). *Suppose that $(B^{(i)})_{i=1}^k$ is a sequence of Bohr sets. Then*

$$\mu_G(\textstyle\bigwedge_{i=1}^k B^{(i)}) \geqslant \prod_{i=1}^k \mu_G(B^{(i)}_{1/2}).$$

Proof. Let $\Delta := \{(x, \dots, x) \in G^k : x \in G\}$ and $S := B^{(1)}_{1/2} \times \cdots \times B^{(k)}_{1/2}$. Then

$$\int 1_\Delta * 1_{-\Delta} 1_S * 1_{-S} \, d\mu_{G^k} = \int (1_\Delta * 1_S)^2 d\mu_{G^k} \geqslant \mu_{G^k}(\Delta)^2 \mu_{G^k}(S)^2 \tag{5-1}$$

by Cauchy–Schwarz. The integrand on the left-hand side is at most $\mu_{G^k}(\Delta)\mu_{G^k}(S)$ and it is supported on the set of $x \in \Delta - \Delta = \Delta$ such that $1_S * 1_{-S}(x) > 0$. But if $1_S * 1_{-S}(y, \dots, y) > 0$ then

$$y \in \textstyle\bigcap_{i=1}^k (B^{(i)}_{1/2} - B^{(i)}_{1/2}) \subset \bigcap_{i=1}^k B^{(i)}_1 = (\bigwedge_{i=1}^k B^{(i)})_1.$$

Hence

$$\mu_{G^k}(\operatorname{supp} 1_\Delta * 1_{-\Delta} 1_S * 1_{-S}) \leqslant \mu_G((\textstyle\bigwedge_{i=1}^k B^{(i)})_1)\mu_{G^k}(\Delta),$$

and inserting this in (5-1) we get

$$\mu_G((\textstyle\bigwedge_{i=1}^k B^{(i)})_1)\mu_{G^k}(\Delta)^2\mu_{G^k}(S) \geqslant \mu_{G^k}(\Delta)^2\mu_{G^k}(S)^2.$$

The result follows after some cancelation and noting that $\mu_{G^k}(S)$ is just the right-hand side of the inequality in the statement of the lemma. □

Note that if B is a Bohr set whose frequency set has one element, and whose width function is the constant function 2 then there is an easy lower bound for $\mu_G(B_\eta)$ as the length of a certain arc on a circle:

$$\mu_G(B_\eta) \geqslant \frac{1}{\pi} \arccos(1 - 2\eta^2) \geqslant \frac{1}{\pi} \min\{\eta, 2\}. \tag{5-2}$$

From this we immediately recover the usual lower bound on the size of a Bohr set with a larger frequency set from this and the preceding lemma.[2]

[1] Technically width function $\gamma \mapsto \min\{\rho\delta_\gamma, 2\}$.

[2] To recover the bound in [Tao and Vu 2006, Lemma 4.20] some adjustments need to be made as our definition of a Bohr set is in terms of $\gamma(x)$ being close to 1 rather than $\arg \gamma(x)$ being close to 0.

Bourgain [1999] developed the idea of Bohr sets as approximate substitutes for groups, and since then his techniques have become an essential tool in additive combinatorics. To begin with we define the *entropy* of a Bohr set B to be

$$h(B) := \log \frac{\mu_G(B_2)}{\mu_G(B_{1/2})}.$$

A trivial covering argument shows that B_2 can be covered by $\exp(h(B))$ translates of B, and if B is actually a subgroup then $h(B) = 0$. It is often desirable to have a uniform bound on $h(B_\delta)$ for all $\delta \in (0, 2]$, and such a bound is called the dimension of B in other work. Here, however, it is crucial that we do not insist on this.

We shall be particularly interested in Bohr sets which grow in a reasonably regular way because they will function well as approximate groups. In light of the definition of entropy (which encodes growth over a fixed range) we say that a Bohr set B is *C-regular* if

$$\frac{1}{1 + Ch(B)|\eta|} \leqslant \frac{\mu_G(B_{1+\eta})}{\mu_G(B)} \leqslant 1 + Ch(B)|\eta|$$

for all η with $|\eta| \leqslant 1/Ch(B)$. Crucially such Bohr sets are commonplace.

Lemma 5.2. *There is an absolute constant $C_{\mathfrak{R}}$ such that if B is a Bohr set then there is some $\lambda \in [1, 2]$ such that B_λ is $C_{\mathfrak{R}}$-regular.*

The proof is by a covering argument and follows [Tao and Vu 2006, Lemma 4.24], for example. From now on we say that a Bohr set B is *regular* if it is $C_{\mathfrak{R}}$-regular.

Finally, we write β_ρ for the probability measure induced on B_ρ by μ_G, and β for β_1. These measures function as approximate analogues for Haar measure, and the following useful lemma of Green and Konyagin [2009] shows how they can used to describe a sensible version of the annihilator of a Bohr set.

Lemma 5.3. *Suppose that B is a regular Bohr set. Then*

$$\{\gamma : |\widehat{\beta}(\gamma)| \geqslant \kappa\} \subset \{\gamma : |1 - \gamma(x)| = O(h(B)\kappa^{-1}\rho) \text{ for all } x \in B_\rho\}.$$

Proof. First, suppose that $|\widehat{\beta}(\gamma)| \geqslant \kappa$ and $y \in B_\rho$. Then

$$|1 - \gamma(y)|\kappa \leqslant \left|\int \gamma(x)\, d\beta(x) - \int \gamma(x + y)\, d\beta(x)\right| \leqslant \frac{\mu_G(B_{1+\rho} \setminus B_{1-\rho})}{\mu_G(B_1)} = O(h(B)\rho)$$

provided $\rho \leqslant 1/C_{\mathfrak{R}}h(B)$. The result is proved. □

6. The large spectrum and Chang's theorem

Given a probability measure μ, a function $f \in L^1(\mu)$ and a parameter $\epsilon \in (0, 1]$ we define the *ϵ-spectrum of f w.r.t. μ* to be the set

$$\mathrm{Spec}_\epsilon(f, \mu) := \{\gamma \in \widehat{G} : |(f d\mu)^\wedge(\gamma)| \geqslant \epsilon \|f\|_{L^1(\mu)}\}.$$

This definition extends the usual one from the case $\mu = \mu_G$. We shall need a local version of a result of Chang [2002] for estimating the "complexity" or "entropy" of the large spectrum.

Given a set of characters Λ and a function $\omega : \Lambda \to D := \{z \in \mathbb{C} : |z| \leqslant 1\}$ we define

$$p_{\omega,\Lambda} := \prod_{\lambda\in\Lambda} (1 + \operatorname{Re}\omega(\lambda)\lambda),$$

and call such a function a *Riesz product for* Λ. It is easy to see that all Riesz products are real nonnegative functions. They are at their most useful when they also have mass close to 1: the set Λ is said to be *K-dissociated w.r.t. μ* if

$$\int p_{\omega,\Lambda}\, d\mu \leqslant \exp(K) \quad \text{ for all } \omega : \Lambda \to D.$$

In particular, being 0-dissociated w.r.t. μ_G is the usual definition of being dissociated. This relativised version of dissociativity has a useful monotonicity property.

Lemma 6.1 (monotonicity of dissociativity). *Suppose that μ' is another probability measure, Λ is K-dissociated w.r.t. μ, $\Lambda' \subset \Lambda$ and $K' \geqslant K$. Then Λ' is K'-dissociated w.r.t. $\mu' * \mu$.*

Conceptually the next definition is inspired by the discussion of quadratic rank Gowers and Wolf give in [Gowers and Wolf 2011]. The *(K, μ)-relative entropy* of a set Γ is the size of the largest subset $\Lambda \subset \Gamma$ such that Λ is K-dissociated w.r.t. μ.

Lemma 6.2 (Chang bound [Sanders 2012, Lemma 4.6]). *Suppose that $0 \not\equiv f \in L^2(\mu)$ and write $L_f := \|f\|_{L^2(\mu)}\|f\|_{L^1(\mu)}^{-1}$. Then the set $\operatorname{Spec}_\epsilon(f, \mu)$ has $(1, \mu)$-relative entropy $O(\epsilon^{-2} \log 2L_f)$.*

The proof of this goes by a Chernoff-type estimate, the argument for which follows [Green and Ruzsa 2007, Proposition 3.4], and then the usual argument from [Chang 2002].

Although Chang's theorem cannot be significantly improved (see [Green 2003; 2004] for a discussion), there are some small refinements and discussions of their limitations in [Shkredov 2006; 2007; 2008].

Low entropy sets of characters are majorised by large Bohr sets, a fact encoded in the following lemma. The proof is a minor variant of [Sanders 2012, Lemma 6.3].

Lemma 6.3 (annihilating dissociated sets). *Suppose that B is a regular Bohr set and Δ is a set of characters with (η, β)-relative entropy k. Then there is a set Λ of size at most k and some*

$$\rho = \Omega(\eta/(1 + h(B))(k + \log 2\eta^{-1}))$$

such that for all $\gamma \in \Delta$ we have

$$|1 - \gamma(x)| = O(k\nu + \rho'\rho^{-1}h(B_\rho)) \quad \text{for all } x \in B_{\rho'} \wedge B'_\nu,\ \rho', \nu \in \mathbb{R}^+$$

where B' is the Bohr set with constant width function 2 and frequency set Λ.

Proof. Let $L := \lceil \log_2 3^k 2(k + 1)\eta^{-1}\rceil$, the reason for which choice will become apparent, and define

$$\beta^+ := \beta_{1+L\rho} * \beta_\rho * \cdots * \beta_\rho,$$

where β_ρ occurs L times in the expression. By regularity (of B) we can pick $\rho \in (\Omega(\eta/(1+h(B))L), 1]$

such that B_ρ is regular and we have the pointwise inequality

$$\beta \leqslant \frac{\mu_G(B_{1+L\rho})}{\mu_G(B)}\beta^+ \leqslant (1+\eta/3)\beta^+.$$

It follows that if Λ is $\eta/2$-dissociated w.r.t. β^+ then Λ is η-dissociated w.r.t. β, and hence Λ has size at most k. From now on all dissociativity will be w.r.t. β^+.

We put $\eta_i := i\eta/2(k+1)$ and begin by defining a sequence of sets $\Lambda_0, \Lambda_1, \dots$ iteratively such that Λ_i is η_i-dissociated. We let $\Lambda_0 := \varnothing$ which is easily seen to be 0-dissociated. Now, suppose that we have defined Λ_i as required. If there is some $\gamma \in \Delta \setminus \Lambda_i$ such that $\Lambda_i \cup \{\gamma\}$ is η_{i+1}-dissociated then let $\Lambda_{i+1} := \Lambda_i \cup \{\gamma\}$. Otherwise, terminate the iteration.

Note that for all $i \leqslant k+1$, if the set Λ_i is defined then it is certainly $\eta/2$-dissociated and so $|\Lambda_i| \leqslant k$. However, if the iteration had continued for $k+1$ steps then $|\Lambda_{k+1}| > k$. This contradiction means that there is some $i \leqslant k$ such that $\Lambda := \Lambda_i$ is η_i-dissociated and $\Lambda_i \cup \{\gamma\}$ is not η_{i+1}-dissociated for any $\gamma \in \Delta \setminus \Lambda_i$.

It follows that we have a set Λ of at most k characters such that for all $\gamma \in \Delta \setminus \Lambda$ there is a function $\omega : \Lambda \to D$ and $\nu \in D$ such that

$$\int p_{\omega,\Lambda}(1 + \operatorname{Re} \nu\gamma)\, d\beta^+ > \exp(\eta_{i+1}).$$

Now, suppose that $\gamma \in \Delta$. If $\gamma \in \Lambda$ then the conclusion is immediate, so we may assume that $\gamma \in \Delta \setminus \Lambda$. Then, since Λ is η_i-dissociated, we see that

$$\left| \int p_{\omega,\Lambda}\overline{\gamma}\, d\beta^+ \right| > \exp(\eta_{i+1}) - \exp(\eta_i) \geqslant \frac{\eta}{2(k+1)}.$$

Applying Plancherel's theorem we get

$$\frac{\eta}{2(k+1)} \leqslant \left| \sum_{\lambda \in \operatorname{Span}(\Lambda)} \widehat{p_{\omega,\Lambda}}(\lambda)\widehat{\beta}^+(\gamma - \lambda) \right| \leqslant 3^k \sup_{\lambda \in \operatorname{Span}(\Lambda)} |\widehat{\beta}_\rho(\gamma-\lambda)|^L.$$

Given the choice of L there is some $\lambda \in \operatorname{Span}(\Lambda)$ such that $|\widehat{\beta}_\rho(\gamma - \lambda)| \geqslant \frac{1}{2}$. By Lemma 5.3 we see that

$$\gamma - \lambda \in \{\gamma' : |1 - \gamma'(x)| = O(\rho'' h(B_\rho)) \text{ for all } x \in (B_\rho)_{\rho''}\}.$$

On the other hand, by the triangle inequality if $\lambda \in \operatorname{Span}(\Lambda)$ then

$$\lambda \in \{\gamma' : |1 - \gamma'(x)| \leqslant k\nu \text{ for all } x \in B'_\nu\},$$

and the result follows from a final application of the triangle inequality. □

7. Containment in a Bohr set

The object of this section is to show the following result.

Proposition 7.1. *Suppose that G is a finite (compact) abelian group, $d \geqslant 1$ and X is a finite subset of G with $\mu_G(nX) \leqslant n^d \mu_G(X)$ for all $n \geqslant 1$ and $\kappa \in (0, 1]$ is a parameter. Then there is a regular Bohr set B such that*

$$X - X \subset B_\kappa \text{ and } \mu_G(B_2) \leqslant \exp(O(d \log 2d\kappa^{-1}))\mu_G(X).$$

What is important here is that given a set of relative polynomial growth we have produced a Bohr set which contains the original set, and which has controlled growth over a fixed range of dilations. Extending this range down to zero can be done but involves considerable additional work as well as being unnecessary for our arguments.

The next lemma is the key ingredient that provides us with an appropriate Bohr set. The idea originates with [Green and Ruzsa 2007, Lemma 2.3], but the lemma we record is more obviously related to [Tao and Vu 2006, Proposition 4.39].

Lemma 7.2. *Suppose that G is a finite (compact) abelian group, $A, S \subset G$ have*

$$\mu_G(A + S) \leqslant K\mu_G(A) \text{ and } |\widehat{1_{A+S}}(\gamma)| \geqslant (1 - \epsilon)\mu_G(A + S).$$

Then $|1 - \gamma(s)| \leqslant \sqrt{2^3 K\epsilon}$ for all $s \in S - S$.

Proof. By hypothesis there is a phase $\omega \in S^1$ such that

$$\int 1_{A+S}\omega\gamma d\mu_G = |\widehat{1_{A+S}}(\gamma)| \geqslant (1 - \epsilon)\mu_G(A + S).$$

It follows that

$$\int 1_{A+S}|1 - \omega\gamma|^2 d\mu_G = 2\int 1_{A+S}(1 - \omega\gamma)\, d\mu_G \leqslant 2\epsilon\mu_G(A + S),$$

and so if $y_0, y_1 \in S$ then

$$\int 1_A|1 - \omega\gamma(y_i)\gamma|^2 d\mu_G \leqslant \int 1_{A+S}|1 - \omega\gamma|^2 d\mu_G \leqslant 2\epsilon\mu_G(A + S).$$

However, the Cauchy–Schwarz inequality tells us that

$$|1 - \gamma(y_0 - y_1)|^2 \leqslant 2(|1 - \omega\gamma(y_0)\gamma(x)|^2 + |1 - \omega\gamma(y_1)\gamma(x)|^2)$$

for all $x \in G$, whence

$$\int 1_A|1 - \gamma(y_0 - y_1)|^2 d\mu_G \leqslant 2^3\epsilon\mu_G(A + S),$$

and the result follows. □

To prove the proposition we use an idea from [Schoen 2003], first introduced to Freĭman-type problems in [Green and Ruzsa 2007]. The essence is that if we have sub-exponential growth of a set then we can apply the Cauchy–Schwarz inequality and Parseval's theorem in a standard way to get a Fourier coefficient of very close to maximal value.

Proof of Proposition 7.1. By the pigeonhole principle there is some $l = O(d \log 2d)$ such that $\mu_G(lX) \leqslant 2\mu_G((l-1)X)$. We let B' be the Bohr set with width function the constant function $\frac{1}{2}$ and frequency set $\Gamma := \mathrm{Spec}_{1-\epsilon}(1_{lX})$ where we pick $\epsilon := 2^{-10}\kappa^2$.

It follows by Lemma 7.2 applied to $A = (l-1)X$ and $S = X$ that

$$|1-\gamma(x)| \leqslant \sqrt{2^3.2.\epsilon} = \kappa/8 \quad \text{for all } x \in X - X \text{ and } \gamma \in \mathrm{Spec}_{1-\epsilon}(1_{lX}),$$

and hence that $X - X \subset B'_{\kappa/4}$.

It remains to show that the Bohr set is not too large. Begin by noting that

$$\int (1_{lX}^{(k)})^2 d\mu_G \geqslant \frac{1}{\mu_G(k(lX))}\left(\int 1_{lX}^{(k)}\, d\mu_G\right)^2 \geqslant \frac{\mu_G(lX)^{2k-1}}{(kl)^d}, \tag{7-1}$$

where $1_{lX}^{(k)}$ denotes the k-fold convolution of 1_{lX} with itself, and the inequality is Cauchy–Schwarz and then the hypothesis. On the other hand, by Parseval's theorem

$$\begin{aligned}\sum_{\gamma \notin \mathrm{Spec}_{1-\epsilon}(1_{lX})} |\widehat{1_{lX}}(\gamma)|^{2k} &\leqslant ((1-\epsilon)\mu_G(lX))^{2k-2} \sum_{\gamma \in \widehat{G}} |\widehat{1_{lX}}(\gamma)|^2 \\ &\leqslant \exp(-\Omega(k\kappa))\mu_G(lX)^{2k-1} \leqslant \frac{\mu_G(lX)^{2k-1}}{2(kl)^d}\end{aligned}$$

for some $k = O(d\kappa^{-1} \log 2d\kappa^{-1})$. In particular, from (7-1) we have

$$\sum_{\gamma \notin \mathrm{Spec}_{1-\epsilon}(1_{lX})} |\widehat{1_{lX}}(\gamma)|^{2k} \leqslant \frac{1}{2}\int (1_{lX}^{(k)})^2 d\mu_G.$$

It then follows from Parseval's theorem and the triangle inequality that

$$\begin{aligned}\sum_{\gamma \in \mathrm{Spec}_{1-\epsilon}(1_{lX})} |\widehat{1_{lX}}(\gamma)|^{2k} &= \sum_{\gamma \in \widehat{G}} |\widehat{1_{lX}}(\gamma)|^{2k} - \sum_{\gamma \notin \mathrm{Spec}_{1-\epsilon}(1_{lX})} |\widehat{1_{lX}}(\gamma)|^{2k} \\ &\geqslant \int (1_{lX}^{(k)})^2 d\mu_G - \frac{1}{2}\int (1_{lX}^{(k)})^2 d\mu_G = \frac{1}{2}\int (1_{lX}^{(k)})^2 d\mu_G.\end{aligned}$$

On the other hand by the triangle inequality $|\widehat{\beta'}(\gamma)| \geqslant \frac{1}{2}$ if $\gamma \in \Gamma$ since $\delta \leqslant \frac{1}{2}$, whence

$$\sum_{\gamma \in \widehat{G}} |\widehat{1_{lX}}(\gamma)|^{2k} |\widehat{\beta'}(\gamma)|^2 \geqslant \frac{1}{4} \sum_{\gamma \in \mathrm{Spec}_{1-\epsilon}(1_{lX})} |\widehat{1_{lX}}(\gamma)|^{2k} \geqslant \frac{\mu_G(lX)^{2k-1}}{8(kl)^d}.$$

But, by Parseval's theorem and Hölder's inequality we have

$$\begin{aligned}\sum |\widehat{1_{lX}}(\gamma)|^{2k} |\widehat{\beta'}(\gamma)|^2 &= \int (1_{lX}^{(k)} * \beta')^2 d\mu_G \\ &\leqslant \|1_{lX}^{(k)} * 1_{-lX}^{(k)}\|_{L^1(G)} \|\beta' * \beta'\|_{L^\infty(G)} = \frac{\mu_G(lX)^{2k}}{\mu_G(B')},\end{aligned}$$

and so

$$\mu_G(B') \leqslant (kl)^d \mu_G(lX) \leqslant \exp(O(d \log 2d\kappa^{-1}))\mu_G(X).$$

Finally we apply Lemma 5.2 to get a regular Bohr set B with $B_2 \subset B'_1$ and $B_\kappa \supset B'_{\kappa/4}$ so the result is proved. □

8. Covering and growth in abelian groups

Covering lemmas are a major tool in additive combinatorics and have been since their development in [Ruzsa 1999]. This was further extended in [Green and Ruzsa 2006], and such lemmas play a pivotal role in the nonabelian theory as was highlighted by Tao [2008a], where we do not have many other techniques.

While the most basic form of covering lemmas do work in the nonabelian setting, there is a refined argument due to Chang [2002] that does not port over so easily.

Lemma 8.1 (Chang's covering lemma [Tao and Vu 2006, Lemma 5.31]). *Suppose that G is an (discrete) abelian group and $A, S \subset G$ are finite sets with $|nA| \leqslant K^n|A|$ for all $n \geqslant 1$ and $|A + S| \leqslant L|S|$. Then there is a set T with $|T| = O(K \log 2KL)$ such that*[3]

$$A \subset \mathrm{Span}(T) + S - S.$$

We shall also need the following slight variant which provides a way in abelian groups to pass from relative polynomial growth on one scale to all scales.

Lemma 8.2 (variant of Chang's covering lemma). *Suppose that G is an (discrete) abelian group and $A, S \subset G$ are finite sets with $|kA + S| < 2^k|S|$. Then there is a set $T \subset A$ with $|T| < k$ such that $A \subset \mathrm{Span}(T) + S - S$.*

Proof. Let T be a maximal S-*dissociated* subset of A, that is a maximal subset of A such that

$$(\sigma.T + S) \cap (\sigma'.T + S) = \varnothing \quad \text{for all } \sigma \neq \sigma' \in \{0, 1\}^T.$$

Now suppose that $x' \in A \setminus T$ and write $T' := T \cup \{x'\}$. By the maximality of T there are elements σ, σ' in $\{0, 1\}^{T'}$ such that $(\sigma.T' + S) \cap (\sigma'.T' + S) \neq \varnothing$. Now if $\sigma_{x'} = \sigma'_{x'}$ then $(\sigma|_T.T + S) \cap (\sigma'|_T.T + A) \neq \varnothing$, contradicting the fact that T is S-dissociated. Hence, without loss of generality, $\sigma_{x'} = 1$ and $\sigma'_{x'} = 0$, whence

$$x' \in \sigma'|_T.T - \sigma|_T.T + S - S \subset \mathrm{Span}(T) + S - S.$$

We are done unless $|T| \geqslant k$; assume it is and let $T' \subset T$ be a set of size k. Denote $\{\sigma.T' : \sigma \in \{0, 1\}^{T'}\}$ by P and note that $P \subset kA$, whence

$$2^k|S| = |P + S| \leqslant |kA + S| < 2^k|S|.$$

This contradiction completes the proof. □

[3] Recall that $\mathrm{Span}(T) := \{\sum_{t \in T} \sigma_t.t : \sigma \in \{-1, 0, 1\}^t\}$.

Although this is a result in abelian groups, it has many parallels with Milnor's proof [1968] establishing the dichotomy between polynomial growth and exponential growth in solvable groups.

The above lemma is particularly useful for controlling the order of relative polynomial growth through the next result, an idea introduced in [Green and Ruzsa 2006].

Lemma 8.3. *Suppose that G is an (discrete) abelian group, $X \subset G$ and $2X - X \subset \operatorname{Span}(T) + X - X$ for some set T of size k. Then*

$$|(n+1)X - X| \leqslant (2n+1)^k |X - X| \quad \textit{for all } n \geqslant 1.$$

Proof. By induction it is immediate that

$$(n+1)X - X \subset n\operatorname{Span}(T) + X - X,$$

and it is easy to see that $|n\operatorname{Span}(T)| \leqslant (2n+1)^k$ from which the result follows. □

9. Lattices and coset progressions

The geometry of numbers seems to play a pivotal role in proofs of Freĭman-type theorems, and we direct the reader to [Tao and Vu 2006, Chapter 3.5] or [Green 2002b] for a much more comprehensive discussion.

Recall that Λ is a *lattice* in $\mathbb{R}^k$ if there are linearly independent vectors $v_1, \ldots, v_k$ such that $\Lambda = v_1\mathbb{Z} + \cdots + v_k\mathbb{Z}$; we call $v_1, \ldots, v_k$ a basis for Λ. Furthermore, a set K in $\mathbb{R}^k$ is called a *convex body* if it is convex, open, nonempty and bounded.

We require the following application of John's theorem and Minkowski's second theorem, which provides us with a way of producing a generalised arithmetic progression from some sort of "convex progression".[4]

Lemma 9.1 [Tao and Vu 2006, Lemma 3.33]. *Suppose that K is a symmetric convex body and Λ is a lattice, both in $\mathbb{R}^d$. Then there is a proper d-dimensional progression P in $K \cap \Lambda$ such that $|P| \geqslant \exp(-O(d\log 2d))|K \cap \Lambda|$.*

The $\exp(-O(d\log d))$ factor should not come as a surprise: consider packing a d-dimensional cube (playing the role of the generalised progression) inside a d-dimensional sphere.

The question remains of how to find a "convex progression", and to do this Ruzsa [1994] introduced an important embedding. Suppose that G is a (discrete) finite abelian group and $\Gamma \subset \widehat{G}$. Then we define a map

$$\begin{aligned} R_\Gamma &: G \to C(\Gamma, \mathbb{R}) \\ &x \mapsto R_\Gamma(x) : \Gamma \to \mathbb{R}; \gamma \mapsto \frac{1}{2\pi}\arg(\gamma(x)), \end{aligned}$$

[4] A more formal notion of convex progression is introduced by Green [2002b], where a detailed discussion and literature survey may be found.

where the argument is taken to lie in $(-\pi, \pi]$. Note that R_Γ preserves inverses, meaning that $R_\Gamma(-x) = -R_\Gamma(x)$, and furthermore if[5]

$$\|R_\Gamma(x_1)\|_{C(\Gamma,\mathbb{R})} + \cdots + \|R_\Gamma(x_d)\|_{C(\Gamma,\mathbb{R})} < \tfrac{1}{2},$$

then

$$R_\Gamma(x_1 + \cdots + x_d) = R_\Gamma(x_1) + \cdots + R_\Gamma(x_d).$$

This essentially encodes the idea that R_Γ behaves like a Freĭman morphism.[6] We shall use this embedding to establish the following proposition.

Proposition 9.2. *Suppose that G is a finite abelian group, $d \in \mathbb{N}$ and B is a Bohr set such that*

$$\mu_G(B_{(3d+1)\delta}) < 2^d \mu_G(B_\delta) \textit{ for some } \delta < \tfrac{1}{4}(3d+1).$$

Then B_δ contains a proper coset progression M of dimension at most d satisfying the estimate $\beta_\delta(M) = \exp(-O(d \log 2d))$.

Proof. We write Γ for the frequency set of B and note that we may assume that $L := \bigcap\{\ker\gamma : \gamma \in \Gamma\}$ is trivial. Indeed, if it is nontrivial we may quotient out by it without impacting the hypotheses of the proposition; we call the quotiented Bohr set B' and note that $B_\delta = B'_\delta + L$ from which the result follows.

To start with note that if $x \in B_\eta$ then

$$\|R_\Gamma(x)\|_{C(\Gamma,\mathbb{R})} \leqslant \frac{1}{2\pi} \arccos(1 - \eta^2/2) \leqslant 2\eta,$$

and so since $2(3d+1)\delta < \frac{1}{2}$ we have that if $x_1, \ldots, x_{3d+1} \in B_\delta$ then

$$R_\Gamma(x_1 + \cdots + x_{3d+1}) = R_\Gamma(x_1) + \cdots + R_\Gamma(x_{3d+1}). \tag{9-1}$$

By hypothesis we then have

$$\begin{aligned}|(3d+1)R_\Gamma(B_\delta)| = |R_\Gamma((3d+1)B_\delta)| &\leqslant |(3d+1)B_\delta| \\ &\leqslant |B_{(3d+1)\delta}| < 2^d|B_\delta| = 2^d|R_\Gamma(B_\delta)|.\end{aligned}$$

Apply the variant of Chang's covering lemma in Lemma 8.2 to the set $R_\Gamma(B_\delta)$ (which is symmetric since R_Γ preserves inverses and B_δ is symmetric) to get a set $X \subset R_\Gamma(B_\delta)$ with $|X| \leqslant d$ such that

$$3R_\Gamma(B_\delta) \subset \operatorname{Span}(X) + 2R_\Gamma(B_\delta).$$

Writing V for the real subspace of $C(\Gamma, \mathbb{R})$ generated by X we see that $\dim V \leqslant d$ and (by induction) that

$$nR_\Gamma(B_\delta) \subset V + 2R_\Gamma(B_\delta)$$

for all n. Now, suppose that $v \in 2R_\Gamma(B_\delta)$. It follows that

$$n.v \in 2nR_\Gamma(B_\delta) \subset V + 2R_\Gamma(B_\delta).$$

[5] Recall that if X is a normed space then $\|\cdot\|_X$ denotes the norm on that space, so that $\|f\|_{C(\Gamma,\mathbb{R})} = \|f\|_{L^\infty(\Gamma)}$.

[6] We direct the unfamiliar reader to [Tao and Vu 2006, Chapter 5.3].

for all naturals n. Since $2R_\Gamma(B_\delta)$ is finite we see that there are two distinct naturals n and n' and some element $w \in 2R_\Gamma(B_\delta)$ such that $n.v, n'.v \in V + w$. It follows that $(n-n').v \in V$ whence $v \in V$ since V is a vector space and $n \neq n'$. We conclude that $R_\Gamma(B_\delta) \subset V$.

Let E be the group generated by B_δ which is finite, and note that $H := R_\Gamma(E) + C(\Gamma, \mathbb{Z})$ is a closed discrete subgroup of $C(\Gamma, \mathbb{R})$, where $C(\Gamma, \mathbb{Z})$ is the group of $\mathbb{Z}$-valued functions on Γ. Since H is a closed discrete subgroup of $C(\Gamma, \mathbb{R})$ contained in V, it is also a closed discrete subgroup of V. Since V is certainly generated by $R_\Gamma(B_\delta)$ and $H \supset R_\Gamma(B_\delta)$ we see that $\Lambda := H \cap V$ has finite covolume and so is a lattice in V.

Let ρ be the unique solution to $|1-\exp(2\pi i\rho)| = \eta$ in the range $\left[0, \frac{1}{2}\right]$, and write Q_ρ for the ρ-cube in $C(\Gamma, \mathbb{R})$, which is a symmetric convex body in $C(\Gamma, \mathbb{R})$, and so $K := V \cap Q_\rho$ is a symmetric convex body in V. Now, by Lemma 9.1 the set $K \cap \Lambda$ contains a proper d-dimensional progression P of size $\exp(-O(d \log 2d))|K \cap \Lambda|$.

To see this note that by (9-1), $R_\Gamma|_{B_\delta}$ is a Freĭman 2-homomorphism. Now, if $x_1, x_2, x_3, x_4 \in B_\delta$ satisfy

$$R_\Gamma(x_1) + R_\Gamma(x_2) = R_\Gamma(x_3) + R_\Gamma(x_4)$$

then

$$R_\Gamma(x_1 + x_2 - x_3 - x_4) = R_\Gamma(x_1) + R_\Gamma(x_2) + R_\Gamma(-x_3) + R_\Gamma(-x_4) = 0.$$

However, $R_\Gamma(x) = 0$ if and only if $\gamma(x) = 1$ for all $\gamma \in \Gamma$, which is to say if and only if $x \in L$. Since L is trivial we conclude that $x_1 + x_2 = x_3 + x_4$ and hence that R_Γ is injective on B_δ, and $R_\Gamma^{-1} : R_\Gamma(B_\delta) \to B_\delta$ is a Freĭman 2-homomorphism.

On the other hand, by (9-1) $R_\Gamma : B_\delta \to R_\Gamma(B_\delta)$ is a Freĭman 2-homomorphism, and therefore also a Freĭman 2-isomorphism; hence its inverse $R_\Gamma^{-1} : R_\Gamma(B_\delta) \to B_\delta$ is one as well.

Since $B_\delta = R_\Gamma^{-1}(K \cap \Lambda)$, we are done by, for example, [Tao and Vu 2006, Proposition 5.24], which simply says that the image of a proper coset progression under a Freĭman isomorphism of order at least 2 is a proper coset progression of the same size and dimension; in particular $R_\Gamma^{-1}(P)$ is a proper coset progression of size $\exp(-O(d \log 2d))|B_\delta|$ and dimension at most d. □

10. Proof of the main theorem

The result driving Theorem 1.1 is the following which brings together all the ingredients of the paper.

Theorem 10.1. *Suppose that G is a finite abelian group, $A, S \subset G$ have $|A + S| \leqslant K \min\{|A|, |S|\}$, and $\epsilon \in (0, 1]$ is a parameter. Then there is a proper coset progression M with*

$$\dim M = O(\epsilon^{-2} \log^6 2\epsilon^{-1} K) \text{ and } |M| \geqslant \left(\frac{\epsilon}{2 \log K}\right)^{O(\epsilon^{-2} \log^6 2\epsilon^{-1} K)} |A + S|,$$

such that for any probability measure μ supported on M we have

$$\|1_{A+S} * \mu\|_{\ell^\infty(G)} \geqslant 1 - \epsilon \text{ and } \|1_A * \mu\|_{\ell^\infty(G)} \geqslant (1-\epsilon)\frac{|A|}{|A+S|}.$$

Proof. We start by thinking of G as discrete and using counting measure. By Plünnecke's inequality [Tao and Vu 2006, Corollary 6.28] there is a nonempty set $S' \subset S$ such that

$$|A+A+S'| \leqslant \left(\frac{K\min\{|A|,|S|\}}{|S|}\right)^2 |S'| \leqslant K^2\frac{|A||S'|}{|S|} \leqslant K^2|A|.$$

Note, in particular, that since $|A+A+S'| \geqslant |A|$ we have $|S'| \geqslant |S|/K^2$ from the second inequality. Applying the inequality again we get a nonempty set $A' \subset A$ such that

$$|A'+(A+S')+(A+S')| \leqslant K^4|A'|,$$

and it follows that

$$|(A+S')+(A+S')| \leqslant K^4|A+S'|. \tag{10-1}$$

Now we apply Proposition 4.2 with $T = A$ to get a symmetric neighbourhood of the identity X such that

$$|X| \geqslant \exp(-O(\epsilon^{-2}k^2\log^2 2K))|A+S|$$

since $|A| \geqslant |A+S|/K$, and

$$|\mu_{-A} * 1_{A+S'} * \mu_{-S'}(x) - 1| \leqslant \epsilon/4 \quad \text{for all } x \in kX. \tag{10-2}$$

In the first instance it follows that $kX \subset (A+S')-(A+S')$. On the other hand, by the Plünnecke–Ruzsa estimates [Tao and Vu 2006, Corollary 6.29] applied to (10-1) we have

$$|4l((A+S')-(A+S'))| \leqslant K^{32l}|A+S'| = \exp(O(l\log K + \epsilon^{-2}k^2\log^2 K))|X|,$$

and hence

$$|4lkX| \leqslant \exp(O(l\log 2K + \epsilon^{-2}k^2\log^2 2K))|X|.$$

We put $l = \lceil \epsilon^{-2}k^2\log 2K\rceil$, so that

$$|(3kl+1)X| \leqslant |4klX| \leqslant 2^{kl.O(k^{-1}\log 2K)}|X|.$$

Hence we can pick k such that

$$1+\log\epsilon^{-1}K \leqslant k = O(\log 2\epsilon^{-1}K) \text{ and } |(3kl+1)X| < 2^{kl}|X|.$$

By the variant of Chang's covering lemma in Lemma 8.2 there is some set T of size at most $kl = O(\epsilon^{-2}\log^4 2\epsilon^{-1}K)$ such that $3X \subset \mathrm{Span}(T)+2X$, and hence (by Lemma 8.3)

$$|(n+2)X)| \leqslant n^{O(\epsilon^{-2}\log^4 2\epsilon^{-1}K)}|2X| \text{ for all } n \geqslant 1.$$

On the other hand $|2X| \leqslant 2^{kl}|X|$, and so (rescaling the measure to think of G as compact) we have

$$\mu_G(nX) \leqslant n^{O(\epsilon^{-2}\log^4 2\epsilon^{-1}K)}\mu_G(X) \text{ for all } n \geqslant 1.$$

Now, by Proposition 7.1 applied to the set X there is a $d = O(kl\log 2kl\kappa^{-1})$ (which we may also assume is at least 1) and a regular Bohr set B such that

$$X - X \subset B_{\kappa/2} \text{ and } \mu_G(B_2) \leqslant \exp(d)\mu_G(X).$$

Let c be the absolute constant in the following technical lemma and note that since X is a neighbourhood of the identity, $X \subset B$ and $\beta(X) \geqslant \exp(-d)$.

We apply Chang's theorem relative to B to get that $\mathrm{Spec}_c(1_X, \beta) = \mathrm{Spec}_c(\mu_X)$ has $(1, \beta)$-relative entropy

$$r = O(c^{-2} \log 2\|1_X\|_{L^2(\beta)} \|1_X\|_{L^1(\beta)}^{-1}) = O(d).$$

It follows from Lemma 6.3 that there is a set of characters Λ of size r and a $\rho = \Omega(1/(1+h(B))r)$ such that for all $\gamma \in \mathrm{Spec}_c(\mu_X)$ we have

$$|1-\gamma(x)| = O(\nu r + \rho' r h(B) h(B_\rho)) \quad \text{for all } x \in B_{\rho'} \wedge B'_\nu,$$

where B' is the Bohr set with width function the constant function 2 and frequency set Λ. Provided $\rho \geqslant \kappa$ we see that

$$\mu_G(X) \leqslant \mu_G(B_{\rho/2}) \leqslant \mu_G(B_{1/2}) \text{ and } \mu_G(B_{2\rho}) \leqslant \mu_G(B_2) \leqslant \exp(d)\mu_G(X),$$

and so it follows that $h(B), h(B_\rho) \leqslant d$. It follows that $\rho = \Omega(1/d^2)$ and

$$|1-\gamma(x)| = O(\nu d + \rho' d^3) \quad \text{for all } x \in B_{\rho'} \wedge B'_\nu \text{ and } \gamma \in \mathrm{Spec}_c(\mu_X).$$

Pick $\rho' = \Omega(\epsilon/d^3 K^2)$ and $\nu = \Omega(\epsilon/K^2 d)$ such that $B'' := B_{\rho'} \wedge B'_\nu$ has

$$|1-\gamma(x)| \leqslant \epsilon/4K^2 \quad \text{for all } x \in B'' \text{ and } \gamma \in \mathrm{Spec}_c(\mu_X).$$

In particular

$$\rho', \nu = \Omega(1/K^2\, d^{O(1)}).$$

For each $\lambda \in \Lambda$ write $B^{(\lambda)}$ for the Bohr set with frequency set $\{\lambda\}$ and width function the constant function 2, thus $B'_\nu = \bigwedge_{\lambda \in \Lambda} B^{(\lambda)}_\nu$. By Lemma 5.1 we see that

$$\mu_G(B''_\eta) \geqslant \mu_G(B_{\eta\rho'/2}) \prod_{\lambda \in \Lambda} \mu_G(B^{(\lambda)}_{\eta\nu/2}).$$

On the other hand, since $B^{(\lambda)}$ has a frequency set of size 1 we see (from (5-2)) that

$$\mu_G(B^{(\lambda)}_{\eta'}) \geqslant \frac{1}{\pi} \min\{\eta', 2\}.$$

Now, if $\eta\rho'/2 \geqslant \kappa$ we have

$$\mu_G(B''_\eta) \geqslant (\eta\nu/2\pi)^r \mu_G(X),$$

and on the other we have $\mu_G(B) \leqslant \exp(d)\mu_G(X)$. Let $t \geqslant 1$ be a natural such that

$$(16\pi(3t+1)\nu^{-1})^r \exp(d) < 2^t \text{ and } t = O(d \log 2dK).$$

Then if $\eta \in \left[\frac{1}{8}(3t+1), \frac{1}{4}(3t+1)\right)$ we have

$$\mu_G(B''_{(3t+1)\eta}) < 2^t \mu_G(B''_\eta).$$

We now apply Proposition 9.2 to get that B''_η contains a proper coset progression M of dimension at most t and size $(2t)^{-O(t)}\mu_G(X)$. The result is proved on an application of the next lemma provided such a choice of η is possible. This can be done if κ can be chosen such that

$$\frac{\rho'}{8(3t+1)} > \kappa,$$

which can be done with $\kappa = \Omega(\epsilon^{O(1)}K^{-O(1)})$, and working this back gives that $t = O(\epsilon^{-2}\log^6 2\epsilon^{-1}K)$ and the result. □

The next lemma is here simply to avoid interrupting the flow of the previous argument, and the hypotheses are set up purely for that setting. The proof is simply a series of standard Fourier manipulations.

Lemma 10.2. *There is an absolute constant $c > 0$ such that if G is a finite abelian group, $A, S, X \subset G$ have $|A+S| \leqslant K\min\{|A|,|S|\}$, $S' \subset S$ has $|S'| \geqslant |S|/K^2$, $k \geqslant \log\epsilon^{-1}K$ is a natural number such that*

$$|\mu_{-A} * 1_{A+S'} * \mu_{-S'}(x) - 1| \leqslant \epsilon/4 \quad \textit{for all } x \in kX,$$

and M is a set such that

$$|1-\gamma(x)| \leqslant \epsilon/4K^2 \quad \textit{for all } x \in M \textit{ and } \gamma \in \operatorname{Spec}_c(\mu_X), \tag{10-3}$$

then for any probability measure μ supported on M we have

$$\|1_{A+S} * \mu\|_{\ell^\infty(G)} \geqslant 1-\epsilon \quad \textit{and} \quad \|1_A * \mu\|_{\ell^\infty(G)} \geqslant (1-\epsilon)\frac{|A|}{|A+S|}.$$

Proof. Integrating the first hypothesis we get

$$|\langle \mu_{-A} * 1_{A+S'} * \mu_{-S'}, \mu_X^{(k)}\rangle - 1| \leqslant \epsilon/4,$$

where $\mu_X^{(k)}$ denotes the k-fold convolution of μ_X with itself. By Fourier inversion we have

$$\left|\sum_{\gamma\in\widehat{G}} \widehat{1_{A+S'}}(\gamma)\overline{\widehat{\mu_A}(\gamma)\widehat{\mu_{S'}}(\gamma)\widehat{\mu_X}(\gamma)^k} - 1\right| \leqslant \epsilon/4. \tag{10-4}$$

The triangle inequality, Cauchy–Schwarz and Parseval's theorem in the usual way tell us that

$$\sum_{\gamma\in\widehat{G}} \left|\widehat{1_{A+S'}}(\gamma)\widehat{\mu_A}(\gamma)\widehat{\mu_{S'}}(\gamma)\right| \leqslant \mu_G(A+S')\|\widehat{\mu_A}\|_{\ell^2(\widehat{G})}\|\widehat{\mu_{S'}}\|_{\ell^2(\widehat{G})} = \frac{\mu_G(A+S')}{\sqrt{\mu_G(A)\mu_G(S')}} \leqslant K^2. \tag{10-5}$$

Then, by the triangle inequality, for any probability measure μ supported on M we have

$$|\widehat{\mu}(\gamma) - 1| \leqslant \epsilon/4K^2 \quad \text{for all } \gamma \in \operatorname{Spec}_c(\mu_X). \tag{10-6}$$

We conclude that

$$E := \left|\langle 1_{A+S'} * \mu, \mu_A * \mu_{S'} * \mu_X^{(k)} * \mu\rangle - 1\right| = \left|\sum_{\gamma\in\widehat{G}} \widehat{1_{A+S'}}(\gamma)\widehat{\mu}(\gamma)\overline{\widehat{\mu_A}(\gamma)\widehat{\mu_{S'}}(\gamma)\widehat{\mu_X}(\gamma)^k\widehat{\mu}(\gamma)} - 1\right|$$

is at most $S_1 + S_2 + S_3$, where

$$S_1 := \Bigg| \sum_{\gamma \notin \mathrm{Spec}_c(\mu_X)} \widehat{1_{A+S'}}(\gamma)\overline{\widehat{\mu_A}(\gamma)\widehat{\mu_{S'}}(\gamma)\widehat{\mu_X}(\gamma)^k}(|\widehat{\mu}(\gamma)|^2 - 1)\Bigg|,$$

$$S_2 := \Bigg| \sum_{\gamma \in \mathrm{Spec}_c(\mu_X)} \widehat{1_{A+S'}}(\gamma)\overline{\widehat{\mu_A}(\gamma)\widehat{\mu_{S'}}(\gamma)\widehat{\mu_X}(\gamma)^k}(|\widehat{\mu}(\gamma)|^2 - 1)\Bigg|,$$

$$S_3 := \Bigg| \sum_{\gamma \in \widehat{G}} \widehat{1_{A+S'}}(\gamma)\overline{\widehat{\mu_A}(\gamma)\widehat{\mu_{S'}}(\gamma)\widehat{\mu_X}(\gamma)^k} - 1\Bigg|.$$

By the triangle inequality and (10-5) we see that

$$S_1 \leqslant \sup_{\gamma \notin \mathrm{Spec}_c(\mu_X)} |\widehat{\mu_X}(\gamma)|^k \cdot \sum_{\gamma \in \widehat{G}} |\widehat{1_{A+S'}}(\gamma)\widehat{\mu_A}(\gamma)\widehat{\mu_{S'}}(\gamma)| \leqslant c^k K^2 \leqslant \epsilon/4$$

for a suitable choice of $c = \Omega(1)$, since $k \geqslant \log \epsilon^{-1} K$; by (10-5) and (10-6) we see that

$$S_2 \leqslant 2 \sup_{\gamma \in \mathrm{Spec}_c(\mu_X)} |\widehat{\mu}(\gamma) - 1| \cdot \sum_{\gamma \in \widehat{G}} \left|\widehat{1_{A+S'}}(\gamma)\widehat{\mu_A}(\gamma)\widehat{\mu_{S'}}(\gamma)\right| \leqslant 2(\epsilon/4K^2)K^2 \leqslant \epsilon/2;$$

and finally by (10-4) we see that $S_3 \leqslant \epsilon/4$, so that $E \leqslant \epsilon$. It follows from this that

$$\langle 1_{A+S'} * \mu, \mu_A * \mu_{S'} * \mu_X^{(k)} * \mu \rangle \geqslant 1 - \epsilon,$$

and hence by averaging that

$$\|1_{A+S'} * \mu\|_{L^\infty(G)} \geqslant 1 - \epsilon \quad \text{and} \quad \|1_A * \mu\|_{L^\infty(G)} \geqslant (1-\epsilon)\frac{\mu_G(A)}{\mu_G(A+S')}.$$

The lemma is proved. □

It is worth making a couple of remarks before continuing. First, Theorem 10.1 can be extended to infinite abelian groups by embedding the sets there in a finite group via a sufficiently large Freĭman isomorphism. This is the finite modelling argument of [Green and Ruzsa 2007, Lemma 2.1], but we shall not pursue it here.

The expected ϵ-dependence in Theorem 10.1 may be less clear than the K-dependence. The argument we have given works equally well for the so-called popular difference set in place of 1_{A+S}, that is the set

$$D(A, S) := \{x \in G : 1_A * 1_S(x) \geqslant c\epsilon/K\}$$

for sufficiently small c. On the other hand Wolf [2010], developing the niveau set construction of Ruzsa [1987; 1991], showed that even finding a large sumset in such popular difference sets is hard, and it seems likely that her arguments can be adapted to cover the case of $D(A, S)$ containing a proportion $1 - \epsilon$ of a sumset.

Understanding this, even in the model setting of $G=\mathbb{F}_2^n$, would be of great interest since a better ϵ-dependence would probably yield better analysis of inner products of the form $\langle 1_A * 1_S, 1_T\rangle$ which are of importance in, for example, Roth's theorem [Roth 1953; 1952].

We are now in a position to prove Theorem 1.1 by an easy pigeonhole argument.

Proof of Theorem 1.1. Freĭman 2-embed the sets A and S into a finite group (via, for example, the method of [Green and Ruzsa 2007, Lemma 2.1]); if we can prove the result there then it immediately pulls back.

Apply Theorem 10.1 with $\epsilon=\frac{1}{2}(1+\sqrt{2})$ to get a proper d-dimensional coset progression M. Note that we may assume the progression is symmetric by translating it and possibly shrinking it by a factor of $\exp(d)$; this has no impact on the bounds. Thus we put

$$M=H+\{x_1.l_1+\cdots+x_d.l_d : |l_i|\leqslant L_i \quad \text{for all } 1\leqslant i\leqslant d\}$$

where $L_1,\dots,L_d\in\mathbb{N}$, $H\leqslant G$ and $x_1,\dots,x_d\in G$. Write

$$M_\eta := H+\{x_1.l_1+\cdots+x_d.l_d : |l_i|\leqslant \eta L_i \quad \text{for all } 1\leqslant i\leqslant d\},$$

and note that $|M_1|\leqslant\exp(O(d))|M_{1/2}|$. On the other hand if $j\eta\leqslant\frac{1}{2}$ we have

$$M_{1/2}\subset M_{1/2+\eta}\subset\cdots\subset M_{1/2+j\eta}=M_1,$$

so it follows that there is some $\eta=\Omega(1/d)$ and $i\leqslant j=O(d)$ such that $|M_{1/2+i\eta}|\leqslant 2^{1/2}|M_{1/2+(i-1)\eta}|$. Since $\eta=\Omega(1/d)$ we easily have that $|M_\eta|=\exp(-O(d\log d))|M_1|$. On the other hand if we apply the conclusion of Theorem 10.1 with

$$\mu=\frac{1_{M_{1/2+i\eta}}+1_{M_{1/2+(i-1)\eta}}}{|M_{1/2+i\eta}|+|M_{1/2+(i-1)\eta}|}$$

we get an element x such that

$$|(x+A+S)\cap M_{1/2+i\eta}|+|(x+A+S)\cap M_{1/2+(i-1)\eta}|$$

is at least

$$(1-\epsilon)(|M_{1/2+i\eta}|+|M_{1/2+(i-1)\eta}|).$$

But then if $z\in M_\eta$ we get

$$\begin{aligned}
1_{A+S}*1_{-(A+S)}(z) &= 1_{x+A+S}*1_{-(x+A+S)}(z)\\
&\geqslant 1_{(x+A+S)\cap M_{1/2+i\eta}}*1_{-(x+A+S)\cap M_{1/2+(i-1)\eta}}(z)\\
&\geqslant \big|(x+A+S)\cap M_{1/2+i\eta}\big|+\big|z+((x+A+S)\cap M_{1/2+(i-1)\eta})\big|\\
&\quad -\big|((x+A+S)\cap M_{1/2+i\eta})\cup(z+((x+A+S)\cap M_{1/2+(i-1)\eta}))\big|\\
&\geqslant \big|(x+A+S)\cap M_{1/2+i\eta}\big|+\big|(x+A+S)\cap M_{1/2+(i-1)\eta}\big|-|M_{1/2+i\eta}|\\
&\geqslant (1-(1+\sqrt{2})\epsilon)|M_{1/2+(i-1)\eta}|>0,
\end{aligned}$$

and it follows that $(A-A)+(S-S)$ contains M_η. Tracking through the bounds we get the result. □

11. Concluding remarks and applications

To begin with we should remark that in the case when G has bounded exponent or is torsion-free, we can get slightly better bounds and the argument is much simpler because of the presence of a good modelling lemmas. In the first case we get the following result, a proof of which (in the case $G = \mathbb{F}_2^n$) is contained in the Appendix as it is so short.

Theorem 11.1 (Bogolyubov–Ruzsa lemma for bounded exponent abelian groups). *Suppose G is an abelian group of exponent r and $A, S \subset G$ are finite nonempty sets such that $|A+S| \leqslant K \min\{|A|, |S|\}$. Then $(A-A)+(S-S)$ contains a subspace V of size* $\exp(-O_r(\log^4 2K))|A+S|$.

In the second, the material of Sections 5–9 can be replaced by similar but more standard arguments because of the following modelling lemma.

Lemma 11.2 (modelling for torsion-free abelian groups [Ruzsa 2009, Theorem 3.5]). *Suppose that G is a torsion-free abelian group, $A \subset G$ is a finite nonempty set and $k \geqslant 2$ is a natural. Then for every $q \geqslant |kA-kA|$ there is a set $A' \subset A$ with $|A'| \geqslant |A|/k$ such that A' is Freĭman k-isomorphic to a subset of $\mathbb{Z}/q\mathbb{Z}$.*

Theorem 11.3 (Bogolyubov–Ruzsa lemma for torsion-free abelian groups). *Suppose that G is a torsion-free abelian group and $A, S \subset G$ are finite nonempty sets such that $|A+S| \leqslant K \min\{|A|, |S|\}$. Then $(A-A)+(S-S)$ contains a proper symmetric $d(K)$-dimensional coset progression M of size*

$$\exp(-h(K))|A+S|.$$

Moreover, we may take $d(K) = O(\log^4 2K)$ and $h(K) = O(\log^4 2K \log 2 \log 2K)$.

Returning to Theorem 1.1 it is easy to see that we must have $d(K), h(K) = \Omega(\log K)$ by considering a union of $\sqrt{K}$ coset progressions of dimension $\log_2 \sqrt{K}$, and even achieving this bound may be hard without refining the definition of a coset progression. (See the comments of Green in [Tao 2008b] for a discussion of this.)

The paper [Schoen 2011] was a major breakthrough in proving the first good bounds for (a slight variant of) Theorem 1.1; it was essentially shown that one could take

$$d(K), h(K) = O(\exp(O(\sqrt{\log K})))$$

for torsion-free or bounded-exponent abelian groups.

Indeed, it should be clear that while we do not use [Schoen 2011] directly in the proof of Theorem 1.1, it has had a considerable influence on the present work and the applications which now follow are from the end of that paper as well.

Freĭman's theorem. As an immediate corollary of Theorem 1.1 and Chang's covering lemma we have the following.

Theorem 11.4 (Freĭman's theorem for abelian groups). *Suppose that G is an (discrete) abelian group and $A \subset G$ is finite with $|A \pm A| \leqslant K|A|$. Then A is contained in a $d(K)$-dimensional coset progression M of size at most* $\exp(h(K))|A|$. *Moreover, we may take $d(K), h(K) = O(K \log^{O(1)} 2K)$.*

By considering a union of K dissociated translates of a coset progression it is easy to see that we must have $d(K), h(K) = \Omega(K)$, so the result is close to best possible.

Green and Ruzsa [2007] provided the first bounds of $d(K), h(K) = O(K^{4+o(1)})$, and the peppering of their work throughout this paper should indicate the importance of their ideas.

Schoen [2011] improved the bounds to $O(K^{3+o(1)})$ and to $O(K^{1+o(1)})$ for certain classes of groups, and in [Cwalina and Schoen 2010] the structure is further elucidated with particular emphasis on getting good control on the dimension.

The U^3-inverse theorem. Theorem 1.1 can be inserted into the various U^3-inverse theorems of Tao and Green [2008] for finite abelian groups of odd order, and Samorodnitsky [2007] (see also [Wolf 2009]) for $\mathbb{F}_2^n$ to improve the bounds there. In particular one gets the following.

Theorem 11.5 ($U^3(\mathbb{F}_2^n)$-inverse theorem). *Suppose that $f \in L^\infty(\mathbb{F}_2^n)$ has $\|f\|_{U^3(\mathbb{F}_2^n)} \geqslant \delta\|f\|_{L^\infty(\mathbb{F}_2^n)}$. Then there is a quadratic polynomial $q : \mathbb{F}_2^n \to \mathbb{F}_2$ such that*

$$|\langle f, (-1)^q\rangle_{L^2(\mathbb{F}_2^n)}| \geqslant \exp(-O(\log^{O(1)} 2\delta^{-1}))\|f\|_{L^\infty(\mathbb{F}_2^n)}.$$

In fact the connection between good bounds in results of this type and good bounds in Freĭman-type theorems is quite clearly developed by Green and Tao [2010] and Lovett [2010].

Long arithmetic progressions in sumsets. The question of finding long arithmetic progressions in sets of integers is one of central interest in additive combinatorics. The basic question has the following form: suppose that $A_1, \dots, A_k \subset \{1, \dots, N\}$ all have density at least α. How long an arithmetic progression can we guarantee that $A_1 + \dots + A_k$ contain?

For one set this is addressed by the notoriously difficult Szemerédi's theorem [1969; 1975], where the best quantitative work is that of Gowers [1998; 2001]; for two sets the longest progression is much longer with the state of the art due to Green [2002a]; for three sets or more the results get even stronger with the work of Freĭman, Halberstam and Ruzsa [Freĭman et al. 1992]; and finally for eight sets or more, longer again by the recent work of Schoen [2011].

Theorem 1.1 yields an immediate improvement for the case of four sets or more.

Theorem 11.6. *Suppose that $A_1, \dots, A_4 \subset \{1, \dots, N\}$ all have density at least α. Then $A_1 + \dots + A_4$ contains an arithmetic progression of length $N^{O(\log^{-O(1)} 2\alpha^{-1})}$.*

Proof. Since $|A_i + A_j| \leqslant 2\alpha^{-1}|A_i|$ for all i, j we have, by averaging, that there is a symmetric set A of density $\alpha^{O(1)}$ such that $A_1, \dots, A_4$ each contains a translate of A. In particular, the longest progression in $A - A + A - A$ is contained in a translate of $A_1 + A_2 + A_3 + A_4$.

Now, by Theorem 1.1 the set $A - A + A - A$ contains an $O(\log^{O(1)} \alpha^{-1})$-dimensional coset progression M of size $\exp(-O(\log^{O(1)} \alpha^{-1}))N$. Since $\mathbb{Z}$ is torsion-free the progression is just a generalised progression which certainly contains a 1-dimensional progression of length $|M|^{1/\dim M}$. The result is proved. □

It is not clear that this result gives the best possible conclusion for k sets as k tends to infinity, but if one were interested in this no doubt some improvement could be squeezed out by delving into the main proof.

$\Lambda(4)$-estimate for the squares. Inserting Theorem 1.1 into a result from [Chang 2004] (itself developed from an argument of Bourgain in [Johnson and Lindenstrauss 2001]) yields the following $\Lambda(4)$-estimate for the squares.

Theorem 11.7. *Suppose that $n_1, \ldots, n_k$ are naturals. Then*

$$\int \left| \sum_{i=1}^{k} \exp(2\pi i n_i^2 \theta) \right|^4 d\theta = O(k^3 \exp(-\Omega(\log^{\Omega(1)} 2k))).$$

This is essentially equivalent to inserting Theorem 1.1 into the proof of [Schoen 2011, Theorem 8] and Gowers' [1998] version of the Balog–Szemerédi lemma [1994]. In any case a conjecture of Rudin [1960] suggests that the bound $O(k^{2+o(1)})$ is likely to be true, and the above is not even a power-type improvement on the trivial upper bound of k^3.

The Konyagin–Łaba theorem. Theorem 1.1 inserted into the argument at the end of [Schoen 2011] yields the following quantitative improvement to a result from [Konyagin and Łaba 2006].

Theorem 11.8 (Konyagin–Łaba theorem). *Suppose that A is a set of reals and $\alpha \in \mathbb{R}$ is transcendental. Then*

$$|A + \alpha.A| = \exp(\Omega(\log^{\Omega(1)} 2|A|))|A|.$$

What is particularly interesting here is that there is a simple construction which shows that there are arbitrarily large sets A with $|A + \alpha.A| = \exp(O(\sqrt{\log |A|}))|A|$.

Appendix: Proof of Theorem 11.1

Our objective in this appendix is to prove the following result.

Theorem A.1. *Suppose that $G := \mathbb{F}_2^n$, and $A \subset G$ has density $\alpha > 0$. Then there is a subspace $V \leqslant G$ with* $\operatorname{cod} V = O(\log^4 2\alpha^{-1})$ *such that $V \subset 4A$.*

We have distilled this argument out because it is short and just uses the two ingredients of the Croot–Sisask lemma and Chang's theorem. For the reader interested in a little more motivation the sketch after the introduction may be of more interest.

In the rather special setting of $\mathbb{F}_2^n$ it is known from [Green and Ruzsa 2007, Proposition 6.1] that if $|A + A| \leqslant K|A|$ then A is Freĭman 8-isomorphic to a set A' of density $K^{-O(1)}$ in some $\mathbb{F}_2^m$, from which we get the following corollary of Theorem A.1.

Corollary A.2. *Suppose that $G := \mathbb{F}_2^n$, and $A \subset G$ has $|A + A| \leqslant K|A|$. Then there is a subspace $V \leqslant G$ with $|V| \geqslant \exp(-O(\log^4 2K))|A|$ such that $V \subset 4A$.*

In this setting the result of Croot and Sisask is the following.

Lemma A.3 (Croot–Sisask). *Suppose that $G := \mathbb{F}_2^n$, $f \in L^p(G)$ and $A \subset G$ has density $\alpha > 0$. Then there is an $a \in A$ and a set T with $\mu_G(T) \geqslant (\alpha/2)^{O(\epsilon^{-2} p)}$ such that*

$$\|\rho_t(f * \mu_A) - f * \mu_A\|_{L^p(G)} \leqslant \epsilon \|f\|_{L^p(G)} \quad \textit{for all } t \in T.$$

Additionally we have:

Lemma A.4 (Chang's theorem). *Suppose that $G := \mathbb{F}_2^n$ and $A \subset G$ has density $\alpha > 0$. Then*

$$\operatorname{cod} \operatorname{Spec}_\epsilon(\mu_A)^\perp = O(\epsilon^{-2} \log 2\alpha^{-1}).$$

Proof of Theorem A.1. We begin by noting that

$$\langle 1_{2A} * 1_A, 1_A \rangle = \langle 1_{2A}, 1_A * 1_A \rangle = \alpha^2. \tag{A-1}$$

By the Croot–Sisask lemma applied with $f := 1_{2A}$ we get a set $T \subset G$ with $\mu_G(T) \geqslant (\alpha/2)^{O(k^2 p)}$ such that

$$\|\rho_t(1_{2A} * 1_A) - 1_{2A} * 1_A\|_{L^p(G)} \leqslant \alpha/4ke \quad \text{for all } t \in T.$$

By the triangle inequality this gives

$$\|\rho_t(1_{2A} * 1_A) - 1_{2A} * 1_A\|_{L^p(G)} \leqslant \alpha/4e \quad \text{for all } t \in kT,$$

and so on integrating (and applying the triangle inequality again) we have

$$\|1_{2A} * 1_A * \mu_T^{(k)} - 1_{2A} * 1_A\|_{L^p(G)} \leqslant \alpha/4e.$$

By Hölder's inequality we get

$$|\langle 1_{2A} * 1_A * \mu_T^{(k)}, 1_A \rangle - \langle 1_{2A} * 1_A, 1_A \rangle| \leqslant \alpha\alpha^{1+1/(p-1)}/4e.$$

Choosing $p = 1 + \log \alpha^{-1}$ and inserting (A-1) we have

$$|\langle 1_{2A} * 1_A * \mu_T^{(k)}, 1_A \rangle - \alpha^2| \leqslant \alpha^2/4,$$

and so by the triangle inequality

$$\langle 1_{2A} * 1_A * \mu_T^{(k)}, 1_A \rangle_{L^p(G)} \geqslant 3\alpha^2/4.$$

Now, put $V := \operatorname{Spec}_{1/2}(\mu_T)^\perp$ and $g := 1_{2A} * 1_A * \mu_T^{(k)}$, so that

$$\big|\langle g, 1_A \rangle - \langle g * \mu_V, 1_A \rangle\big| = \left| \sum_{\gamma \notin V^\perp} \widehat{1_{2A}}(\gamma) |\widehat{1_A}(\gamma)|^2 \hat{\mu}_T(\gamma)^k \right| \leqslant \alpha 2^{-k} \leqslant \alpha^2/8,$$

by Parseval's theorem, the definition of V and by taking $k = O(\log 2\alpha^{-1})$ a sufficiently large natural. It follows by the triangle inequality that

$$\langle 1_{2A} * 1_A * \mu_T^{(k)} * \mu_V, 1_A \rangle > \alpha^2/2,$$

and so, by averaging, that $\|1_{2A} * \mu_V\|_{L^\infty(G)} > \frac{1}{2}$. We conclude that $4A$ contains V by the pigeon-hole principle and the result is proved on applying Chang's theorem to see that

$$\operatorname{cod} V = O(\log 2\mu_G(T)^{-1}) = O(\log^4 2\alpha^{-1}). \qquad \square$$

Acknowledgement

The author should like to thank Julia Wolf for useful discussions surrounding the $U^3(\mathbb{F}_2^n)$-inverse theorem, and an anonymous referee for a thorough reading of the paper and numerous useful suggestions.

References

[Balog and Szemerédi 1994] A. Balog and E. Szemerédi, "A statistical theorem of set addition", *Combinatorica* **14**:3 (1994), 263–268. MR 95m:11019 Zbl 0812.11017

[Bogolyubov 1939] N. N. Bogolyubov, "On some arithmetic properties of quasi-periods", *Zap. kaf. matem. fiz. Akad. Nauk Ukr. = Ann. Chaire Phys. Math. Acad. Sci. Ukraine* (*Kiev*) **4** (1939), 185–194 (in Ukrainian), 195–205 (in French). MR 8,512b JFM 65.0268.01

[Bourgain 1990] J. Bourgain, "On arithmetic progressions in sums of sets of integers", pp. 105–109 in *A tribute to Paul Erdős*, edited by A. Baker et al., Cambridge University Press, 1990. MR 92e:11011 Zbl 0715.11006

[Bourgain 1999] J. Bourgain, "On triples in arithmetic progression", *Geom. Funct. Anal.* **9**:5 (1999), 968–984. MR 2011h:11132 Zbl 0959.11004

[Bourgain 2008] J. Bourgain, "Roth's theorem on progressions revisited", *J. Anal. Math.* **104** (2008), 155–192. MR 2009g:11011 Zbl 1155.11011

[Chang 2002] M.-C. Chang, "A polynomial bound in Freiman's theorem", *Duke Math. J.* **113**:3 (2002), 399–419. MR 2033d:11151 Zbl 1035.11048

[Chang 2004] M.-C. Chang, "On problems of Erdős and Rudin", *J. Funct. Anal.* **207**:2 (2004), 444–460. MR 2004j:11022 Zbl 1046.11012

[Croot and Sisask 2010] E. Croot and O. Sisask, "A probabilistic technique for finding almost-periods of convolutions", *Geom. Funct. Anal.* **20**:6 (2010), 1367–1396. MR 2012d:11019 Zbl 1234.11013

[Cwalina and Schoen 2010] K. Cwalina and T. Schoen, "A linear bound on the dimension in Green–Ruzsa's theorem", preprint, 2010, available at http://ssdnm.mimuw.edu.pl/pliki/prace-studentow/karol-cwalina-1.pdf.

[Freĭman et al. 1992] G. A. Freiman, H. Halberstam, and I. Z. Ruzsa, "Integer sum sets containing long arithmetic progressions", *J. London Math. Soc.* (2) **46**:2 (1992), 193–201. MR 93j:11008 Zbl 0768.11005

[Gowers 1998] W. T. Gowers, "A new proof of Szemerédi's theorem for arithmetic progressions of length four", *Geom. Funct. Anal.* **8**:3 (1998), 529–551. MR 2000d:11019 Zbl 0907.11005

[Gowers 2001] W. T. Gowers, "A new proof of Szemerédi's theorem", *Geom. Funct. Anal.* **11**:3 (2001), 465–588. MR 2002k:11014 Zbl 1028.11005

[Gowers and Wolf 2011] W. T. Gowers and J. Wolf, "Linear forms and quadratic uniformity for functions on $\mathbb{Z}_N$", *J. Anal. Math.* **115** (2011), 121–186. MR 2855036 arXiv 1002.2210

[Green 2002a] B. J. Green, "Arithmetic progressions in sumsets", *Geom. Funct. Anal.* **12**:3 (2002), 584–597. MR 2003i:11148 Zbl 1020.11009

[Green 2002b] B. J. Green, "Structure theory of set addition: Edinburgh–MIT lecture notes on Freĭman's theorem", preprint, 2002, available at http://www.dpmms.cam.ac.uk/~bjg23/papers/icmsnotes.pdf.

[Green 2003] B. J. Green, "Some constructions in the inverse spectral theory of cyclic groups", *Combin. Probab. Comput.* **12**:2 (2003), 127–138. MR 2004b:11144 Zbl 1097.11051

[Green 2004] B. J. Green, "Spectral structure of sets of integers", pp. 83–96 in *Fourier analysis and convexity*, edited by L. Brandolini et al., Birkhäuser, Boston, 2004. MR 2005i:11015 Zbl 1093.11004

[Green 2009] B. J. Green, "Approximate groups and their applications: work of Bourgain, Gamburd, Helfgott and Sarnak", preprint, 2009. arXiv 0911.3354

[Green and Konyagin 2009] B. J. Green and S. Konyagin, "On the Littlewood problem modulo a prime", *Canad. J. Math.* **61**:1 (2009), 141–164. MR 2010a:42001 Zbl 1232.11013

[Green and Ruzsa 2006] B. J. Green and I. Z. Ruzsa, "Sets with small sumset and rectification", *Bull. London Math. Soc.* **38**:1 (2006), 43–52. MR 2006i:11027 Zbl 1155.11307

[Green and Ruzsa 2007] B. J. Green and I. Z. Ruzsa, "Freiman's theorem in an arbitrary abelian group", *J. London Math. Soc.* (2) **75**:1 (2007), 163–175. MR 2007m:20087 Zbl 1133.11058

[Green and Tao 2008] B. J. Green and T. Tao, "An inverse theorem for the Gowers $U^3(G)$ norm", *Proc. Edinb. Math. Soc.* (2) **51**:1 (2008), 73–153. MR 2009g:11012 Zbl 1202.11013

[Green and Tao 2010] B. J. Green and T. Tao, "An equivalence between inverse sumset theorems and inverse conjectures for the U^3 norm", *Math. Proc. Cambridge Philos. Soc.* **149**:1 (2010), 1–19. MR 2011g:11019 Zbl 1229.11132

[Johnson and Lindenstrauss 2001] W. B. Johnson and J. Lindenstrauss (editors), *Handbook of the geometry of Banach spaces, I*, North-Holland, Amsterdam, 2001. MR 2003a:46001 Zbl 0970.46001

[Konyagin and Łaba 2006] S. Konyagin and I. Łaba, "Distance sets of well-distributed planar sets for polygonal norms", *Israel J. Math.* **152** (2006), 157–179. MR 2006m:11032 Zbl 1127.52021

[López and Ross 1975] J. M. López and K. A. Ross, *Sidon sets*, Lecture Notes in Pure and Applied Mathematics **13**, Marcel Dekker, New York, 1975. MR 55 #13173 Zbl 0351.43008

[Lovett 2010] S. Lovett, "Equivalence of polynomial conjectures in additive combinatorics", preprint, 2010. To appear in *Combinatorica*. arXiv 1001.3356

[Milnor 1968] J. Milnor, "Growth of finitely generated solvable groups", *J. Differential Geometry* **2** (1968), 447–449. MR 39 #6212 Zbl 0176.29803

[Roth 1952] K. F. Roth, "Sur quelques ensembles d'entiers", *C. R. Acad. Sci. Paris* **234** (1952), 388–390. MR 13,724d Zbl 0046.04302

[Roth 1953] K. F. Roth, "On certain sets of integers", *J. London Math. Soc.* **28** (1953), 104–109. MR 14,536g Zbl 0050.04002

[Rudin 1960] W. Rudin, "Trigonometric series with gaps", *J. Math. Mech.* **9** (1960), 203–227. MR 22 #6972 Zbl 0091.05802

[Rudin 1990] W. Rudin, *Fourier analysis on groups*, Wiley, New York, 1990. MR 91b:43002 Zbl 0698.43001

[Ruzsa 1987] I. Z. Ruzsa, "Essential components", *Proc. London Math. Soc.* (3) **54**:1 (1987), 38–56. Zbl 0609.10042

[Ruzsa 1991] I. Z. Ruzsa, "Arithmetic progressions in sumsets", *Acta Arith.* **60**:2 (1991), 191–202. MR 92k:11009 Zbl 0728.11009

[Ruzsa 1994] I. Z. Ruzsa, "Generalized arithmetical progressions and sumsets", *Acta Math. Hungar.* **65**:4 (1994), 379–388. MR 95k:11011 Zbl 0816.11008

[Ruzsa 1999] I. Z. Ruzsa, "An analog of Freiman's theorem in groups", pp. 323–326 in *Structure theory of set addition*, edited by J.-M. Deshouillers et al., Astérisque **258**, Société de Mathématique de France, Paris, 1999. MR 2000h:11111 Zbl 0946.11007

[Ruzsa 2009] I. Z. Ruzsa, "Sumsets and structure", pp. 87–210 in *Combinatorial number theory and additive group theory*, edited by A. Geroldinger and I. Z. Ruzsa, Birkhäuser, Basel, 2009. MR 2010m:11013 Zbl 1221.11026

[Samorodnitsky 2007] A. Samorodnitsky, "Low-degree tests at large distances", pp. 506–515 in *STOC'07: Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, edited by D. Johnson, ACM, New York, 2007. MR 2009f:68077 Zbl 1232.68175

[Sanders 2012] T. Sanders, "On certain other sets of integers", *J. Anal. Math.* **116** (2012), 53–82. MR 2892617

[Schoen 2003] T. Schoen, "Multiple set addition in $\mathbb{Z}_p$", *Integers* **3** (2003), Paper #A17. MR 2004j:11012 Zbl 1089.11009

[Schoen 2011] T. Schoen, "Near optimal bounds in Freiman's theorem", *Duke Math. J.* **158**:1 (2011), 1–12. MR 2012f:11018 Zbl 05904309

[Shkredov 2006] I. D. Shkredov, "On sets of large exponential sums", *Dokl. Akad. Nauk* **411**:4 (2006), 455–459. In Russian; translated in *Dokl. Math.* **74**:3 (2006), 860–864. MR 2009j:11130 Zbl 1211.11093 arXiv math/0605689

[Shkredov 2007] I. D. Shkredov, "Some examples of sets of large trigonometric sums", *Mat. Sb.* **198**:12 (2007), 105–140. In Russian; translated in *Sb. Math.* **198**:12 (2007), 1805–1838. MR 2009e:11150 Zbl 1176.11038

[Shkredov 2008] I. D. Shkredov, "On sets of large trigonometric sums", *Izv. Ross. Akad. Nauk Ser. Mat.* **72**:1 (2008), 161–182. In Russian; translated in *Izv. Math.* **72**:1 (2008), 149–168. MR 2009e:11151 Zbl 1148.11040

[Sisask 2009] O. Sisask, “Bourgain’s proof of the existence of long arithmetic progressions in $A+B$”, preprint, 2009, available at http://www.maths.qmul.ac.uk/~olof/files/dl/bourgain-APs.pdf.

[Szemerédi 1969] E. Szemerédi, “On sets of integers containing no four elements in arithmetic progression”, *Acta Math. Acad. Sci. Hungar.* **20** (1969), 89–104. MR 39 #6861 Zbl 0175.04301

[Szemerédi 1975] E. Szemerédi, “On sets of integers containing no k elements in arithmetic progression”, *Acta Arith.* **27** (1975), 199–245. MR 51 #5547 Zbl 0303.10056

[Tao 2008a] T. Tao, “Product set estimates for non-commutative groups”, *Combinatorica* **28**:5 (2008), 547–594. MR 2010b:11017 Zbl 05494691

[Tao 2008b] T. Tao, *Structure and randomness: pages from year one of a mathematical blog*, American Mathematical Society, Providence, RI, 2008. MR 2010h:00002 Zbl 05380664

[Tao 2010] T. Tao, “Freiman’s theorem for solvable groups”, *Contrib. Discrete Math.* **5**:2 (2010), 137–184. MR 2012e:05063

[Tao and Vu 2006] T. Tao and V. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics **105**, Cambridge University Press, 2006. MR 2008a:11002 Zbl 1127.11002

[Wolf 2009] J. Wolf, “A local inverse theorem in $\mathbb{F}_2^n$”, preprint, 2009.

[Wolf 2010] J. Wolf, “The structure of popular difference sets”, *Israel J. Math.* **179** (2010), 253–278. MR 2011k:05043 Zbl 05823090

Received 4 Nov 2010. Revised 12 Sep 2011. Accepted 9 Oct 2011.

TOM SANDERS: t.sanders@dpmms.cam.ac.uk
Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, United Kingdom

ANALYSIS AND PDE
Vol. 5, No. 3, 2012
dx.doi.org/10.2140/apde.2012.5.657

REAL ANALYTICITY AWAY FROM THE NUCLEUS OF PSEUDORELATIVISTIC HARTREE–FOCK ORBITALS

ANNA DALL'ACQUA, SØREN FOURNAIS,
THOMAS ØSTERGAARD SØRENSEN AND EDGARDO STOCKMEYER

We prove that the Hartree–Fock orbitals of pseudorelativistic atoms, that is, atoms where the kinetic energy of the electrons is given by the pseudorelativistic operator $\sqrt{-\Delta+1}-1$, are real analytic away from the origin. As a consequence, the quantum mechanical ground state of such atoms is never a Hartree–Fock state.

Our proof is inspired by the classical proof of analyticity by nested balls of Morrey and Nirenberg. However, the technique has to be adapted to take care of the nonlocal pseudodifferential operator, the singularity of the potential at the origin, and the nonlinear terms in the equation.

1. Introduction and results

In [Dall'Acqua et al. 2008], three of the present authors studied the Hartree–Fock model for pseudorelativistic atoms, and proved the existence of Hartree–Fock minimizers. Furthermore, they proved that the corresponding Hartree–Fock orbitals (solutions to the associated Euler–Lagrange equation) are smooth away from the nucleus, and that they decay exponentially. In this paper we prove that all of these orbitals are, in fact, real analytic away from the origin. Apart from intrinsic mathematical interest, analyticity of solutions has important consequences. For example, in the nonrelativistic case, the analyticity of the orbitals was used in [Friesecke 2003; Lewin 2004a] to prove that the quantum mechanical ground state is never a Hartree–Fock state (or, more generally, is never a finite linear combination of Slater determinants). A direct consequence of our main regularity result is that this also holds in the pseudorelativistic case. Our proof also shows that any $H^{1/2}$-solution $\varphi:\mathbb{R}^3\to\mathbb{C}$ to the nonlinear equation

$$(\sqrt{-\Delta+1})\varphi-\frac{Z}{|\cdot|}\varphi\pm\big(|\varphi|^2*|\cdot|^{-1}\big)\varphi=\lambda\varphi \tag{1}$$

which is smooth away from $\boldsymbol{x}=0$, is in fact real analytic there. As will be clear from the proof, our method yields the same result for solutions to equations of the form

$$(-\Delta+m)^s\varphi+V\varphi+|\varphi|^k\varphi=\lambda\varphi, \tag{2}$$

Dall'Acqua, Østergaard Sørensen, and Stockmeyer gratefully acknowledge a "Research in Pairs" (RiP) stay at the Mathematisches Forschungsinstitut Oberwolfach, where parts of this research were carried out. Fournais was partially supported by the Lundbeck Foundation and the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 202859. Stockmeyer was partially supported by the DFG (SFB/TR12).
MSC2010: primary 35B65, 35J10, 81Q05, 81V55; secondary 81V70, 35Q40.
Keywords: Hartree–Fock model, pseudorelativistic, regularity of wavefunctions, nonlocal operator, real analyticity.

where V has a finite number of point singularities (but is analytic elsewhere), under certain conditions on m, s, V, and k (see Remark 1.2 below). We believe this result is of independent interest, but stick concretely to the case of pseudorelativistic Hartree–Fock orbitals, since this was the original motivation for the present work.

We consider a model for an atom with N electrons and nuclear charge Z (fixed at the origin), where the kinetic energy of the electrons is described by the expression $\sqrt{(|\boldsymbol{p}|c)^2+(mc^2)^2}-mc^2$. This model takes into account some (kinematic) relativistic effects; in units where $\hbar = e = m = 1$, the Hamiltonian becomes

$$H=\sum_{j=1}^{N}\alpha^{-1}\bigl\{T(-\mathrm{i}\nabla_j)-V(\boldsymbol{x}_j)\bigr\}+\sum_{1\le i<j\le N}\frac{1}{|\boldsymbol{x}_i-\boldsymbol{x}_j|}, \tag{3}$$

with $T(\boldsymbol{p}) = E(\boldsymbol{p}) - \alpha^{-1} = \sqrt{|\boldsymbol{p}|^2+\alpha^{-2}}-\alpha^{-1}$ and $V(\boldsymbol{x}) = Z\alpha/|\boldsymbol{x}|$. Here, α is Sommerfeld's fine structure constant; physically, $\alpha \simeq 1/137$.

The operator H acts on a dense subspace of the N-particle Hilbert space $\mathcal{H}_F = \wedge_{i=1}^{N} L^2(\mathbb{R}^3)$ of antisymmetric functions. (We will not consider spin since it is irrelevant for our discussion.) It is bounded from below on this subspace if and only if $Z\alpha \le 2/\pi$ (see [Lieb and Yau 1988]; for a number of other works on this operator, see [Carmona et al. 1990; Daubechies and Lieb 1983; Fefferman and de la Llave 1986; Herbst 1977; Lewis et al. 1997; Nardini 1986; Weder 1975; Zhislin and Vugalter 2002]).

The (*quantum*) *ground state energy* is the infimum of the quadratic form $\mathfrak{q}$ defined by H, over the subset of elements of norm 1 of the corresponding form domain. Hence, it coincides with the infimum of the spectrum of H considered as an operator acting in $\mathcal{H}_F$. A corresponding minimizer is called a (*quantum*) *ground state* of H.

In the Hartree–Fock approximation, instead of minimizing the quadratic form $\mathfrak{q}$ in the entire N-particle space $\mathcal{H}_F$, one restricts to wavefunctions Ψ which are pure wedge products, also called Slater determinants:

$$\Psi(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N)=\frac{1}{\sqrt{N!}}\det(u_i(\boldsymbol{x}_j))_{i,j=1}^{N}, \tag{4}$$

with $\{u_i\}_{i=1}^N$ orthonormal in $L^2(\mathbb{R}^3)$ (called *orbitals*). Notice that this way, $\Psi \in \mathcal{H}_F$ and $\|\Psi\|_{L^2(\mathbb{R}^{3N})} = 1$.

The *Hartree–Fock ground state energy* is the infimum of the quadratic form $\mathfrak{q}$ defined by H over such Slater determinants:

$$E^{\mathrm{HF}}(N,Z,\alpha):=\inf\{\,\mathfrak{q}(\Psi,\Psi)\mid \Psi \text{ Slater determinant}\,\}. \tag{5}$$

Inserting Ψ of the form in (4) into $\mathfrak{q}$ formally yields

$$\begin{aligned}\mathcal{E}^{\mathrm{HF}}&(u_1,\ldots,u_N):=\mathfrak{q}(\Psi,\Psi)\\ &=\alpha^{-1}\sum_{j=1}^{N}\int_{\mathbb{R}^3}\bigl\{\overline{u_j(\boldsymbol{x})}\,[T(-\mathrm{i}\nabla)u_j](\boldsymbol{x})-V(\boldsymbol{x})|u_j(\boldsymbol{x})|^2\bigr\}\,d\boldsymbol{x}\\ &\quad+\frac12\sum_{1\le i,j\le N}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\frac{|u_i(\boldsymbol{x})|^2|u_j(\boldsymbol{y})|^2}{|\boldsymbol{x}-\boldsymbol{y}|}\,d\boldsymbol{x}\,d\boldsymbol{y}-\frac12\sum_{1\le i,j\le N}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\frac{\overline{u_j(\boldsymbol{x})}u_i(\boldsymbol{x})\overline{u_i(\boldsymbol{y})}u_j(\boldsymbol{y})}{|\boldsymbol{x}-\boldsymbol{y}|}\,d\boldsymbol{x}\,d\boldsymbol{y}.\end{aligned} \tag{6}$$

In fact, $u_i \in H^{1/2}(\mathbb{R}^3)$, $1 \le i \le N$, is needed for this to be well-defined (see Section 3 for a detailed discussion), and so (5)–(6) can be written

$$E^{\mathrm{HF}}(N, Z, \alpha) = \inf\{ \mathscr{E}^{\mathrm{HF}}(u_1, \dots, u_N) \mid (u_1, \dots, u_N) \in \mathcal{M}_N \}, \tag{7}$$

$$\mathcal{M}_N = \big\{ (u_1, \dots, u_N) \in [H^{1/2}(\mathbb{R}^3)]^N \bigm| (u_i, u_j) = \delta_{ij} \big\}. \tag{8}$$

Here $(\ ,\ )$ denotes the scalar product in $L^2(\mathbb{R}^3)$. The existence of minimizers for the problem (7)–(8) was proved in [Dall'Acqua et al. 2008] when $Z > N - 1$ and $Z\alpha < 2/\pi$. (Note that such minimizers are generally not unique since $\mathscr{E}^{\mathrm{HF}}$ is not convex; see [Fournais et al. 2009]). The existence of infinitely many distinct critical points of the functional $\mathscr{E}^{\mathrm{HF}}$ on $\mathcal{M}_N$ was proved recently (under the same conditions) in [Enstedt and Melgaard 2009].

The Euler–Lagrange equations of the problem (7)–(8) are the *Hartree–Fock equations*,

$$\begin{aligned} \big[\big(T(-\mathrm{i}\nabla) - V\big)\varphi_i\big](\boldsymbol{x}) + \alpha\Big(\sum_{j=1}^{N} \int_{\mathbb{R}^3} \frac{|\varphi_j(\boldsymbol{y})|^2}{|\boldsymbol{x}-\boldsymbol{y}|}\, d\boldsymbol{y}\Big)\varphi_i(\boldsymbol{x}) \\ - \alpha \sum_{j=1}^{N} \Big(\int_{\mathbb{R}^3} \frac{\overline{\varphi_j(\boldsymbol{y})}\varphi_i(\boldsymbol{y})}{|\boldsymbol{x}-\boldsymbol{y}|}\, d\boldsymbol{y}\Big)\varphi_j(\boldsymbol{x}) = \varepsilon_i \varphi_i(\boldsymbol{x}), \quad 1 \le i \le N. \end{aligned} \tag{9}$$

Here the ε_i are the Lagrange multipliers of the orthonormality constraints in (8). (The naive Euler–Lagrange equations are more complicated than (9), but can be transformed to (9); see [Fournais et al. 2009].) Note that (9) can be reformulated as

$$h_{\boldsymbol{\varphi}} \varphi_i = \varepsilon_i \varphi_i, \quad 1 \le i \le N, \tag{10}$$

with $h_{\boldsymbol{\varphi}}$ the *Hartree–Fock operator associated to* $\boldsymbol{\varphi} = \{\varphi_1, \dots, \varphi_N\}$, formally given by

$$h_{\boldsymbol{\varphi}} u = [T(-\mathrm{i}\nabla) - V]u + \alpha R_{\boldsymbol{\varphi}} u - \alpha K_{\boldsymbol{\varphi}} u, \tag{11}$$

where $R_{\boldsymbol{\varphi}} u$ is the *direct interaction*, given by the multiplication operator defined by

$$R_{\boldsymbol{\varphi}}(\boldsymbol{x}) := \sum_{j=1}^{N} \int_{\mathbb{R}^3} \frac{|\varphi_j(\boldsymbol{y})|^2}{|\boldsymbol{x}-\boldsymbol{y}|}\, d\boldsymbol{y} \tag{12}$$

and $K_{\boldsymbol{\varphi}} u$ is the *exchange term*, given by the integral operator

$$(K_{\boldsymbol{\varphi}} u)(\boldsymbol{x}) = \sum_{j=1}^{N} \Big(\int_{\mathbb{R}^3} \frac{\overline{\varphi_j(\boldsymbol{y})} u(\boldsymbol{y})}{|\boldsymbol{x}-\boldsymbol{y}|}\, d\boldsymbol{y}\Big)\varphi_j(\boldsymbol{x}). \tag{13}$$

The equations (9) (or equivalently (10)) are called the *self-consistent Hartree–Fock equations*. One has that $\sigma_{\mathrm{ess}}(h_{\boldsymbol{\varphi}}) = [0, \infty)$ and that, when in addition $N < Z$, the operator $h_{\boldsymbol{\varphi}}$ has infinitely many eigenvalues in $[-\alpha^{-1}, 0)$ (see [Dall'Acqua et al. 2008, Lemma 2]; the argument given there holds for any $\boldsymbol{\varphi} = \{\varphi_1, \dots, \varphi_N\}$, $\varphi_i \in H^{1/2}(\mathbb{R}^3)$, as long as $Z\alpha < 2/\pi$). If $(\varphi_1, \dots, \varphi_N) \in \mathcal{M}_N$ is a minimizer for the problem (7)–(8), then the φ_i solve (10) with $\varepsilon_1 \le \varepsilon_2 \le \cdots \le \varepsilon_N < 0$ the N *lowest* eigenvalues of the operator $h_{\boldsymbol{\varphi}}$ [Dall'Acqua et al. 2008].

In [Dall'Acqua et al. 2008] it was proved that solutions $\{\varphi_1, \dots, \varphi_N\}$ to (9) — and, more generally, all eigenfunctions of the corresponding Hartree–Fock operator $h_{\boldsymbol{\varphi}}$ — are smooth away from $\boldsymbol{x} = 0$ (the singularity of V), and that (for the φ_i for which $\varepsilon_i < 0$) they decay exponentially. (The solutions studied in [Dall'Acqua et al. 2008] came from a minimizer of $\mathscr{E}^{\mathrm{HF}}$, but the proof trivially extends to the solutions $\{\boldsymbol{\varphi}_n\}_{n\in\mathbb{N}} = \big\{\{\varphi_1^n, \dots, \varphi_N^n\}\big\}_{n\in\mathbb{N}}$ to (9) found in [Enstedt and Melgaard 2009], and to all the eigenfunctions of the corresponding Hartree–Fock operators mentioned above). The main theorem of this paper is the following, which completely settles the question of regularity away from the origin of solutions to the equations (9).

Theorem 1.1. *Let $Z\alpha < 2/\pi$, and let $N \ge 2$ be a positive integer such that $N < Z + 1$. Let $\boldsymbol{\varphi} = \{\varphi_1, \dots, \varphi_N\}$, $\varphi_i \in H^{1/2}(\mathbb{R}^3)$, $i = 1, \dots, N$, be solutions to the pseudorelativistic Hartree–Fock equations in* (9).

Then, for $i = 1, \dots, N$,

$$\varphi_i \in C^{\omega}(\mathbb{R}^3 \setminus \{0\}), \tag{14}$$

that is, the Hartree–Fock orbitals are real analytic away from the origin in $\mathbb{R}^3$.

Remark 1.2. (i) The restrictions $Z\alpha < 2/\pi$, $N < Z + 1$, and $N \ge 2$ are only made to ensure *existence* of $H^{1/2}$-solutions to (9). In fact, our proof proves analyticity away from $\boldsymbol{x} = 0$ for $H^{1/2}$-solutions to (9) for *any* $Z\alpha$. For the case $N = 1$, (9) reduces to $(T - V)\varphi = \varepsilon\varphi$ and our result also holds for $H^{1/2}$-solutions to this equation (see also (iv) and (v) below about more general V for which the result also holds for the linear equation). More interestingly, the result also holds for $H^{1/2}$-solutions to (1) (which, strictly speaking, cannot be obtained from (9) by any choice of N).

(ii) The statement also holds for any eigenfunction of the associated Hartree–Fock operator given by (11).

(iii) It is obvious from the proof that the theorem holds true if we include spin.

(iv) As will also be clear from the proof, the statement of Theorem 1.1 (appropriately modified) also holds for molecules. More explicitly, for a molecule with K nuclei of charges $Z_1, \dots, Z_K$, fixed at $R_1, \dots, R_K \in \mathbb{R}^3$, replace V in (9) by $\sum_{k=1}^K V_k$ with $V_k(\boldsymbol{x}) = Z_k\alpha/|\boldsymbol{x} - R_k|$, $Z_k\alpha < 2/\pi$. Then, for $N < 1 + \sum_{k=1}^K Z_k$, Hartree–Fock minimizers exist (see [Dall'Acqua et al. 2008, Remark 1(viii)]), and the corresponding Hartree–Fock orbitals are real analytic away from the positions of the nuclei, i.e., belong to $C^{\omega}(\mathbb{R}^3 \setminus \{R_1, \dots, R_K\})$.

(v) Another approximation to the full quantum mechanical problem is the *multiconfiguration self-consistent field method* (MC-SCF). Here one minimizes the quadratic form $\mathfrak{q}$ defined by the operator H given in (3) (or, more generally, with V from (iv)) over the set of *finite* sums of Slater determinants instead of only on single Slater determinants as in Hartree–Fock theory. If minimizers exist they satisfy what is called the *multiconfiguration equations* (MC equations). For more details, see [Fournais et al. 2009; Friesecke 2003; Lewin 2004b]. As will be clear from the proof, the statement of Theorem 1.1 also holds for solutions to these equations.

(vi) In fact, for V we only need the analyticity of V away from finitely many points in $\mathbb{R}^3$, and certain integrability properties of $V\varphi_i$ in the vicinity of each of these points, and at infinity; for more details, see Remark 4.1.

(vii) As will be clear from the proof, the statement of Theorem 1.1 also holds for other nonlinearities than the Hartree–Fock term in (9), namely $|\varphi|^k\varphi$ as in (2) (for k even; for k odd, one needs to take φ^{k+1}). The L^p-space in which one needs to study the problem (see Proposition 2.1 and the description of the proof below for details) needs to be chosen depending on k in this case (the larger the k, the larger the p).

(viii) Also, as will be clear from the proof, the result holds if $T(-\mathrm{i}\nabla) = |\nabla|$ (i.e., $T(\boldsymbol{p}) = |\boldsymbol{p}|$) in (9). In (35) below, $E(\boldsymbol{p})^{-1}$ should then be replaced by $(|\boldsymbol{p}|+1)^{-1}$ (and 1 added to $\alpha^{-1}+\varepsilon_i$). The only properties of $E(\boldsymbol{p})^{-1}$ used are in Lemmas C.1 and C.2, which follow also for $(|\boldsymbol{p}|+1)^{-1}$ from the same methods with minor modifications. Similarly, one can replace $T(\boldsymbol{p})$ with $(-\Delta+\alpha^{-2})^s$, $s \in [1/2, 1]$.

(ix) The result of Theorem 1.1 in the nonrelativistic case ($T(-\mathrm{i}\nabla)$ replaced by $-\alpha\Delta$ in (3)) was proved in [Friesecke 2003; Lewin 2004a]; see also the discussion below. In this case, it is furthermore known [Fournais et al. 2009] that, for $\boldsymbol{x} \in B_r(0)$ for some $r > 0$, $\varphi_i(\boldsymbol{x}) = \varphi_i^{(1)}(\boldsymbol{x}) + |\boldsymbol{x}|\varphi_i^{(2)}(\boldsymbol{x})$ with $\varphi_i^{(1)}, \varphi_i^{(2)} \in C^\omega(B_r(0))$.

Combining the argument in [Friesecke 2003; Lewin 2004a] with the analyticity away from the position of the nucleus of solutions to the MC equations (see Remark 1.2(v)) we readily obtain the following result.

Theorem 1.3. *Let Ψ be a (quantum) ground state of the operator H given in* (3). *Then Ψ is not a* finite *linear combination of Slater determinants.*

Remark 1.4. The same holds with V as in Remark 1.2(iv).

Description of the proof of Theorem 1.1. The proof of Theorem 1.1 is inspired by the standard Morrey–Nirenberg proof of analyticity of solutions to general (linear) elliptic partial differential equations with real analytic coefficients by “nested balls” [Morrey and Nirenberg 1957]. A good presentation of this technique can be found in [Hörmander 1969]. (Other proofs using a complexification of the coordinates also exist and have been applied to both linear and nonlinear equations; see [Morrey 2008] and references therein.)

In [Hörmander 1969] one proves L^2-bounds on derivatives of order k of the solution in a ball B_r (of some radius r) around a given point. These bounds should behave suitably in k in order to make the Taylor series of the solution converge locally, thereby proving analyticity.

The proof of these bounds is inductive. In fact, for some ball B_R with $R > r$, one proves the bounds on all balls B_ρ with $r \le \rho \le R$, with the appropriate (with respect to k) behavior in $R-\rho$. The base of induction is provided by standard elliptic estimates. In the induction step, one has to bound $k+1$ derivatives of the solution in the ball B_ρ. To do so, one divides the difference $B_R \setminus B_\rho$ into $k+1$ nested balls using $k+1$ localization functions with successively larger supports. Commuting m of the k derivatives (in the case of an operator of order m) with these localization functions produces (local) differential operators of order $m-1$, with support in a larger ball. These local commutator terms are controlled by the induction hypothesis, since they contain one derivative less. For the last term — the term where no commutators occur — one then uses the equation.

This approach poses new technical difficulties in our case, due to the nonlocality of the kinetic energy $T(\boldsymbol{p}) = \sqrt{-\Delta + \alpha^{-2}} - \alpha^{-1}$ and the nonlinearity of the terms $R_{\varphi}\varphi_i$ and $K_{\varphi}\varphi_i$.

The nonlocality of the operator $\sqrt{-\Delta + \alpha^{-2}}$ implies that, as opposed to the case of a differential operator, the commutator of the kinetic energy with a localization function is not localized in the support of the localization function. That is, when resorting to proving analyticity by differentiating the equation, the localization argument described above introduces commutators which are (nonlocal) pseudodifferential operators. Now the induction hypothesis does not provide control of these terms. Furthermore, it is far from obvious that the singularity of the potential V outside B_R does not influence the regularity in B_R of the solution through these operators (or rather, through the nonlocality of $\sqrt{-\Delta + \alpha^{-2}}$). Loosely speaking, the singularity of the nuclear potential can be felt everywhere. (Note that if we would not have a (singular) potential V one could proceed as in [Frank and Lenzmann 2010] and prove global analyticity by showing exponential decay of the solutions in Fourier space.)

We overcome this problem by a new localization argument which enable us to capture in more detail the action of high order derivatives on nested balls (manifested in Lemma B.1 in the appendix). This, together with very explicit bounds on the (smoothing) operators $\phi E(\boldsymbol{p})^{-1} D^{\beta} \chi$ for χ and ϕ with disjoint supports (see Lemma C.2), are the main ingredients in solving the problem of nonlocality. The estimates are on $\phi E(\boldsymbol{p})^{-1} D^{\beta} \chi$ (not $\phi E(\boldsymbol{p}) D^{\beta} \chi$), since we invert $E(\boldsymbol{p})$ (turning the equation into an integral operator equation, see (35)). Our method of proof would also work in the nonrelativistic case, since the integral operators $(-\Delta + 1)^{-1}$ and $E(\boldsymbol{p})^{-1}$ enjoy similar properties.

The second major obstacle is the (morally cubic) nonlinearity of the terms $R_{\varphi}\varphi_i$ and $K_{\varphi}\varphi_i$.

To illustrate the problem, we discuss proving analyticity by the above method (local L^2-estimates) for solutions u to the equation $\Delta u = u^3$. When differentiating this equation (and therefore u^3), the application of Leibniz's rule introduces a sum of terms. After using Hölder's inequality on each term (the product of three factors, each a number of derivatives on u), one needs to use a Sobolev inequality to "get back down to L^2" in order to use the induction hypothesis. Summing the many terms, the needed estimate does not come out (in fact, some Gevrey-regularity would follow, but not analyticity).

In the quadratic case this can be done (that is, for the equation $\Delta u = u^2$ this problem does *not* occur), but in the cubic case, one looses too many derivatives.

The second insight of our proof is that this problem of loss of derivatives may be overcome by characterizing analyticity by growth of derivatives in some L^p with $p > 2$. When working in L^p for $p > 2$, the loss of derivatives in the Sobolev inequality mentioned above is less (as seen in Theorem D.1). Choosing p sufficiently large allows us to prove the needed estimate. The operator estimates on $\phi E(\boldsymbol{p})^{-1} D^{\beta} \chi$ mentioned above therefore have to be L^p-estimates. In fact, using $L^p - L^q$ estimates, one can also deal with the problem that the singularity of the nuclear potential V can be felt everywhere.

Note that taking $p = \infty$ would avoid using a Sobolev inequality altogether (L^{∞} being an algebra), but the needed estimates on $\phi E(\boldsymbol{p})^{-1} D^{\beta} \chi$ cannot hold in this case. For local equations an approach to handle the loss of derivatives (due to Sobolev inequalities) exists. This was carried out in [Friedman 1958], where analyticity of solutions to elliptic partial differential equations with general analytic nonlinearities was proved. Friedman works in spaces of continuous functions. In this approach, one needs to have a sufficiently high degree of regularity of the solution beforehand (it is not proved along the way). Also, since the elliptic regularity in spaces of continuous functions have an inherent loss of derivative, one needs to work on a sufficiently small domain in order for the method to work. We prefer to work in Sobolev spaces since this is the natural setting for our equation and since the needed estimates on the resolvent are readily obtained in these spaces.

For an alternative method of proof (one *fixed* localization function, to the power k, and estimating in a higher order Sobolev space (instead of in L^2) which is also an algebra), see [Kato 1996] (for the equation $\Delta u = u^2$) and [Hashimoto 2006] (for general second order nonlinear analytic PDEs).

Additional technical difficulties occur due to the fact that the cubic terms, $R_{\boldsymbol{\varphi}}\varphi_i$ and $K_{\boldsymbol{\varphi}}\varphi_i$, are actually nonlocal.

Note that in the proof that *nonrelativistic* Hartree–Fock orbitals are analytic away from the positions of the nuclei (see [Friesecke 2003; Lewin 2004b]), the nonlinearities are dealt with by cleverly rewriting the Hartree–Fock equations as a system. One introduces new functions $\phi_{i,j} = [\varphi_i\overline{\varphi_j}] * |\cdot|^{-1}$, which satisfy $-\Delta\phi_{i,j} = 4\pi\varphi_i\overline{\varphi_j}$. This eliminates the terms $R_{\boldsymbol{\varphi}}\varphi_i$, $K_{\boldsymbol{\varphi}}\varphi_i$, turning these into quadratic products in the functions φ_i, $\phi_{i,j}$, hence one obtains a (quadratic and local) nonlinear system of elliptic second order equations with coefficients analytic away from the positions of the nuclei. The result now follows from the results cited above [Kato 1996; Morrey 2008]. (In fact, this argument extends to solutions of the more general multiconfiguration self-consistent field equations, see [Friesecke 2003; Lewin 2004b].)

This idea cannot readily be extended to our case. The operator $E(\boldsymbol{p})$ is a pseudodifferential operator of first order, so when rewriting the Hartree–Fock equations as described above, one obtains a system of pseudodifferential equations. This system is, as before, of second (differential) order in the auxiliary functions $\phi_{i,j}$, but only of first (pseudodifferential) order in the original functions φ_i. Hence, the leading (second) order matrix is singular elliptic. Hence (even if we ignore the fact that the square root is nonlocal) the above argument does not apply.

To summarize, our approach is as follows. We invert the kinetic energy in the equation for the orbitals thereby obtaining an integral equation to which we apply successive differentiations. The localization argument of Lemma B.1 together with the smoothing estimates on $\phi E(\boldsymbol{p})^{-1} D^\beta \chi$ handle the nonlocality of this equation. By working in L^p for suitably large p one can afford the necessary loss of derivatives from using Sobolev inequalities when treating the nonlinear terms.

2. Proof of analyticity

In order to prove that the φ_i are real analytic in $\mathbb{R}^3 \setminus \{0\}$ it is sufficient, by [Krantz and Parks 2002, Proposition 2.2.10], to prove that for every $\boldsymbol{x}_0 \in \mathbb{R}^3 \setminus \{0\}$ there exists an open set $U \subseteq \mathbb{R}^3 \setminus \{0\}$ containing $\boldsymbol{x}_0$, and constants $\mathscr{C}, \mathscr{R} > 0$, such that (with $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$)

$$|\partial^\beta \varphi_i(\boldsymbol{x})| \le \mathscr{C}\,\frac{\beta!}{\mathscr{R}^{|\beta|}} \quad \text{for all } \boldsymbol{x} \in U \text{ and all } \beta \in \mathbb{N}_0^3. \tag{15}$$

Let $\boldsymbol{x}_0 \in \mathbb{R}^3 \setminus \{0\}$, and let ω be the ball $B_R(\boldsymbol{x}_0)$ with center $\boldsymbol{x}_0$ and radius $R := \min\{1, |\boldsymbol{x}_0|/4\}$. For $\delta > 0$ we denote by ω_δ the set of points in ω at distance larger than δ from $\partial\omega$, i.e.,

$$\omega_\delta := \{\boldsymbol{x} \in \omega \mid d(\boldsymbol{x}, \partial\omega) > \delta\}. \tag{16}$$

By our choice of ω we have $\omega_\delta = B_{R-\delta}(\boldsymbol{x}_0)$. Therefore $\omega_\delta = \varnothing$ for $\delta \geq R$. In particular, by our choice of R,

$$\omega_\delta = \varnothing \quad \text{for } \delta \geq 1. \tag{17}$$

For $\Omega \subseteq \mathbb{R}^n$ and $p \geq 1$ we let $L^p(\Omega)$ denote the usual L^p-space with norm $\|f\|_{L^p(\Omega)} = \big(\int_\Omega |f(\boldsymbol{x})|^p\, d\boldsymbol{x}\big)^{1/p}$. We write $\|f\|_p \equiv \|f\|_{L^p(\mathbb{R}^3)}$. In the following we equip the Sobolev space $W^{m,p}(\Omega)$, $\Omega \subseteq \mathbb{R}^n$, $m \in \mathbb{N}$ and $p \in [1, \infty)$, with the norm

$$\|u\|_{W^{m,p}(\Omega)} := \sum_{|\sigma| \leq m} \|D^\sigma u\|_{L^p(\Omega)}. \tag{18}$$

Theorem 1.1 follows from the following proposition.

Proposition 2.1. *Let $Z\alpha < 2/\pi$, and let $N \geq 2$ be a positive integer such that $N < Z + 1$. Let $\boldsymbol{\varphi} = \{\varphi_1, \ldots, \varphi_N\}$, $\varphi_i \in H^{1/2}(\mathbb{R}^3)$, $i = 1, \ldots, N$, be solutions to the pseudorelativistic Hartree–Fock equations in* (9)*. Let $\boldsymbol{x}_0 \in \mathbb{R}^3 \setminus \{0\}$, $R = \min\{1, |\boldsymbol{x}_0|/4\}$, and $\omega = B_R(\boldsymbol{x}_0)$. Define $\omega_\delta = B_{R-\delta}(\boldsymbol{x}_0)$ for $\delta > 0$.*

Then for all $p \geq 5$ there exist constants $C, B > 1$ such that for all $j \in \mathbb{N}$, for all $\epsilon > 0$ such that $\epsilon j \leq R/2$, and for all $i \in \{1, \ldots, N\}$ we have

$$\epsilon^{|\beta|} \|D^\beta \varphi_i\|_{L^p(\omega_{\epsilon j})} \leq C B^{|\beta|} \quad \textit{for all } \beta \in \mathbb{N}_0^3 \textit{ with } |\beta| \leq j. \tag{19}$$

Given Proposition 2.1, the proof that the φ_i are real analytic is standard, using Sobolev embedding. We give the argument here for completeness. We then give the proof of Proposition 2.1 in the next section.

Let $U = B_{R/2}(\boldsymbol{x}_0) = \omega_{R/2} \subseteq \omega$. Using Theorem D.5 and (19) we have $\varphi_i \in C(\overline{U})$. Therefore it suffices to prove (15) for $|\beta| \geq 1$. Fix $i \in \{1, \ldots, N\}$ and consider $\beta \in \mathbb{N}_0^3 \setminus \{0\}$ an arbitrary multiindex. Setting $j = |\beta|$ and $\epsilon = (R/2)/j$ it follows from Proposition 2.1 (since $\epsilon j = R/2$) that there exist constants $C, B > 1$ such that

$$\|D^\beta \varphi_i\|_{L^p(\omega_{R/2})} \leq C\Big(\frac{B}{\epsilon}\Big)^{|\beta|} = C\Big(\frac{2B}{R}\Big)^{|\beta|} |\beta|^{|\beta|}, \tag{20}$$

with C, B independent of the choice of β. By Theorem D.5 (see also Remark D.6) there exists a constant $K_4 = K_4(p, \boldsymbol{x}_0)$ such that, for all $\beta' \in \mathbb{N}_0^3 \setminus \{0\}$,

$$\sup_{\boldsymbol{x} \in U} |D^{\beta'} \varphi_i(\boldsymbol{x})| \leq K_4 \sum_{|\sigma| \leq 1} \|D^{\beta'+\sigma} \varphi_i\|_{L^p(\omega_{R/2})} \leq K_4 \sum_{|\sigma| \leq 1} C\Big(\frac{2B}{R}\Big)^{|\sigma|+|\beta'|} \big(|\sigma| + |\beta'|\big)^{|\sigma|+|\beta'|},$$

using (20). Using that $R \leq 1 \leq B$, that $\#\{\sigma \in \mathbb{N}_0^3 \mid |\sigma| = 1\} = 3$, and that, from (A.7),

$$\big(1 + |\beta'|\big)^{1+|\beta'|} \leq \frac{\mathrm{e}}{\sqrt{2\pi}} \mathrm{e}^{2|\beta'|} |\beta'|!,$$

this implies that for all $\beta' \in \mathbb{N}_0^3 \setminus \{0\}$,

$$\sup_{x\in U} |D^{\beta'}\varphi_i(x)| \le \Big(\frac{8\mathrm{e}K_4CB}{\sqrt{2\pi}R}\Big)\Big(\frac{2\mathrm{e}^2B}{R}\Big)^{|\beta'|}|\beta'|!. \tag{21}$$

Since $|\sigma|! \le 3^{|\sigma|}\sigma!$ for all $\sigma \in \mathbb{N}_0^3$ (see (A.4) in the appendix), this implies that

$$\sup_{x\in U} |D^{\beta'}\varphi_i(x)| \le \mathscr{C}\,\frac{\beta'!}{\mathscr{R}^{|\beta'|}}, \tag{22}$$

for some $\mathscr{C}, \mathscr{R} > 0$. This proves (15). Hence φ_i is real analytic in $\mathbb{R}^3 \setminus \{0\}$. This finishes the proof of Theorem 1.1.

It therefore remains to prove Proposition 2.1.

Remark 2.2. We here give explicit choices for the constants C and B in Proposition 2.1.

Let

$$C_1 := \max_{1\le a,b\le N} \Big\| \int_{\mathbb{R}^3} \frac{|\varphi_a(y)\varphi_b(y)|}{|\cdot - y|}\,dy\Big\|_\infty. \tag{23}$$

Note that by (29) below, this is finite since $\varphi_i \in H^{1/2}(\mathbb{R}^3)$, $i = 1, \dots, N$.

Furthermore, let $A = A(x_0) \ge 1$ be such that, for all $\sigma \in \mathbb{N}_0^3$,

$$\sup_{x\in\omega} |D^\sigma V(x)| \le A^{|\sigma|+1}|\sigma|!. \tag{24}$$

The existence of A follows from the real analyticity in $\omega = B_R(x_0)$ (recall that $R = \min\{1, |x_0|/4\}$) of $V = Z\alpha|\cdot|^{-1}$ (see e.g. [Krantz and Parks 2002, Proposition 2.2.10]). Assume without restriction that $A \ge \alpha^{-1} + \max_{1\le i\le N} |\varepsilon_i|$.

Let $K_1 = K_1(p)$, $K_2 = K_2(p)$, and $K_3 = K_3(p)$ be the constants in Lemma C.1, Corollary D.2, and Corollary D.4, respectively (see Appendices C and D below). Then let

$$C_2 = \max\big\{K_1, 256\sqrt{2}/\pi\big\}, \tag{25}$$

$$C_3 = \max\big\{4\pi(1+2C_1/R^2)K_3, 160\pi K_2^2K_3\big\}. \tag{26}$$

Choose

$$C > \max_{i\in\{1,\dots,N\}} \Big\{1, \|\varphi_i\|_{W^{1,p}(\omega)}, \|\varphi_i\|_{L^{3p}(B_{2R}(x_0))}, \frac{768}{\pi}|x_0|^{3(2-p)/(2p)}\|\varphi_i\|_2, \\ \Big[\frac{48\sqrt{2}}{\pi}A + 48\sqrt{2}C_1\frac{N}{Z\pi} + \frac{1536\sqrt{2}}{\pi^2|x_0|}\Big]\|\varphi_i\|_3\Big\}. \tag{27}$$

That $C < \infty$ follows from the smoothness away from $x = 0$ of the φ_i [Dall'Acqua et al. 2008, Theorem 1(ii)] and the fact that, since $\varphi_i \in H^{1/2}(\mathbb{R}^3)$, $1 \le i \le N$, we have $\varphi_i \in L^3(\mathbb{R}^3)$, $1 \le i \le N$, by Sobolev's inequality. Then choose

$$B > \max\Big\{48AC_2, C_*, \frac{16}{|x_0|}, 4C_1^2, (160C^2K_2C_3)^2, (24NC_2/Z)^2, 16K_3\Big\}, \tag{28}$$

where C_* is the constant (related to a smooth partition of unity) introduced in (B.3). In particular, $B > 48$. We will prove Proposition 2.1 with these choices of C and B.

3. Proof of the main estimate

We first make (6) more precise, thereby also explaining the choice of $\mathcal{M}_N$ in (8). By Kato's inequality [Kato 1995, (5.33) p. 307],

$$\int_{\mathbb{R}^3} \frac{|f(\boldsymbol{x})|^2}{|\boldsymbol{x}|}\, d\boldsymbol{x} \le \frac{\pi}{2} \int_{\mathbb{R}^3} |\boldsymbol{p}||\hat{f}(\boldsymbol{p})|^2\, d\boldsymbol{p} \quad \text{for } f \in H^{1/2}(\mathbb{R}^3) \tag{29}$$

(where $\hat{f}(\boldsymbol{p}) = (2\pi)^{-3/2} \int_{\mathbb{R}^3} \mathrm{e}^{-\mathrm{i}\boldsymbol{x}\cdot\boldsymbol{p}} f(\boldsymbol{x})\, d\boldsymbol{x}$ denotes the Fourier transform of f), and the KLMN theorem [Reed and Simon 1975, Theorem X.17] the operator h_0 given as

$$h_0 = T(-\mathrm{i}\nabla) - V \tag{30}$$

is well-defined on $H^{1/2}(\mathbb{R}^3)$ (and bounded below by $-\alpha^{-1}$) as a form sum when $Z\alpha < 2/\pi$, that is,

$$(u, h_0 v) = (E(\boldsymbol{p})^{1/2}u, E(\boldsymbol{p})^{1/2}v) - \alpha^{-1}(u, v) - (V^{1/2}u, V^{1/2}v) \quad \text{for } u, v \in H^{1/2}(\mathbb{R}^3). \tag{31}$$

By abuse of notation, we write $E(\boldsymbol{p})$ for the (strictly positive) operator $E(-\mathrm{i}\nabla) = \sqrt{-\Delta + \alpha^{-2}}$. For $(\varphi_1, \dots, \varphi_N) \in \mathcal{M}_N$, the function $R_{\boldsymbol{\varphi}}$ given in (12) belongs to $L^\infty(\mathbb{R}^3)$ (using Kato's inequality above), and the operator $K_{\boldsymbol{\varphi}}$ given in (13) is Hilbert–Schmidt (see [Dall'Acqua et al. 2008, Lemma 2]). As a consequence, when $Z\alpha < 2/\pi$, the operator $h_{\boldsymbol{\varphi}}$ in (11) is a well-defined self-adjoint operator with quadratic form domain $H^{1/2}(\mathbb{R}^3)$ such that

$$(u, h_{\boldsymbol{\varphi}} v) = (u, h_0 v) + \alpha(u, R_{\boldsymbol{\varphi}} v) - \alpha(u, K_{\boldsymbol{\varphi}} v) \quad \text{for } u, v \in H^{1/2}(\mathbb{R}^3). \tag{32}$$

Since $(u, R_{\boldsymbol{\varphi}} u) - (u, K_{\boldsymbol{\varphi}} u) \ge 0$ for any $u \in L^2(\mathbb{R}^3)$, also $h_{\boldsymbol{\varphi}}$ is bounded from below by $-\alpha^{-1}$.

Then, for $(u_1, \dots, u_N) \in \mathcal{M}_N$, the precise version of (6) becomes

$$\begin{aligned} \mathscr{E}^{\mathrm{HF}}(u_1, \dots, u_N) &= \sum_{j=1}^N \alpha^{-1}(u_j, h_0 u_j) \\ &+ \frac{1}{2} \sum_{1\le i,j\le N} \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{|u_i(\boldsymbol{x})|^2 |u_j(\boldsymbol{y})|^2}{|\boldsymbol{x}-\boldsymbol{y}|}\, d\boldsymbol{x}\, d\boldsymbol{y} - \frac{1}{2} \sum_{1\le i,j\le N} \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{\overline{u_j(\boldsymbol{x})} u_i(\boldsymbol{x}) \overline{u_i(\boldsymbol{y})} u_j(\boldsymbol{y})}{|\boldsymbol{x}-\boldsymbol{y}|}\, d\boldsymbol{x}\, d\boldsymbol{y}. \end{aligned} \tag{33}$$

The considerations on $R_{\boldsymbol{\varphi}}$ and $K_{\boldsymbol{\varphi}}$ above imply that also the nonlinear terms in (33) are finite for $u_i \in H^{1/2}(\mathbb{R}^3)$, $1 \le i \le N$.

If $(\varphi_1, \dots, \varphi_N) \in \mathcal{M}_N$ is a critical point of $\mathscr{E}^{\mathrm{HF}}$ in (33), then $\boldsymbol{\varphi} = \{\varphi_1, \dots, \varphi_N\}$ satisfies the self-consistent HF-equations (10) with the operator $h_{\boldsymbol{\varphi}}$ defined above.

Note that $E(\boldsymbol{p})$ is a bounded operator from $H^{1/2}(\mathbb{R}^3)$ to $H^{-1/2}(\mathbb{R}^3)$, and recall that (29) shows that V also defines a bounded operator from $H^{1/2}(\mathbb{R}^3)$ to $H^{-1/2}(\mathbb{R}^3)$ (for any $Z\alpha$). As noted above, both $R_{\boldsymbol{\varphi}}$ and $K_{\boldsymbol{\varphi}}$ are bounded operators on $L^2(\mathbb{R}^3)$ when $(\varphi_1, \dots, \varphi_N) \in \mathcal{M}_N$. In particular, this shows that if $(\varphi_1, \dots, \varphi_N) \in \mathcal{M}_N$ solves (10), then

$$E(\boldsymbol{p})\varphi_i - \alpha^{-1}\varphi_i - V\varphi_i + \alpha R_{\boldsymbol{\varphi}}\varphi_i - \alpha K_{\boldsymbol{\varphi}}\varphi_i = \varepsilon_i \varphi_i, \quad 1 \le i \le N, \tag{34}$$

hold as equations in $H^{-1/2}(\mathbb{R}^3)$. Using that $E(\boldsymbol{p})^{-1}$ is a bounded operator from $H^{-1/2}(\mathbb{R}^3)$ to $H^{1/2}(\mathbb{R}^3)$, this implies that, as equalities in $H^{1/2}(\mathbb{R}^3)$ (and therefore, in particular, in $L^2(\mathbb{R}^3)$),

$$\varphi_i = E(\boldsymbol{p})^{-1}V\varphi_i - \alpha E(\boldsymbol{p})^{-1}R_{\boldsymbol{\varphi}}\varphi_i + \alpha E(\boldsymbol{p})^{-1}K_{\boldsymbol{\varphi}}\varphi_i + (\alpha^{-1}+\varepsilon_i)E(\boldsymbol{p})^{-1}\varphi_i, \quad 1 \le i \le N. \tag{35}$$

Proof of Proposition 2.1. The proof is by induction on $j \in \mathbb{N}_0$. More precisely:

Definition 3.1. For $p \ge 1$ and $j \in \mathbb{N}_0$, let $\mathcal{P}(p, j)$ be the statement:
For all $\epsilon > 0$ with $\epsilon j \le R/2$, and all $i \in \{1, \dots, N\}$ we have

$$\epsilon^{|\beta|}\|D^\beta\varphi_i\|_{L^p(\omega_{\epsilon j})} \le C\, B^{|\beta|} \quad \text{for all } \beta \in \mathbb{N}_0^3 \text{ with } |\beta| \le j, \tag{36}$$

where $C, B > 1$ are the constants in Remark 2.2.

Then Proposition 2.1 is equivalent to the statement: For all $p \ge 5$, $\mathcal{P}(p, j)$ holds for all $j \in \mathbb{N}_0$. This is the statement we will prove by induction on $j \in \mathbb{N}_0$.

Base of induction. For convenience, we prove $\mathcal{P}(p, j)$ for both $j = 0$ and $j = 1$. Note that $\mathcal{P}(p, 0)$ trivially holds since (see Remark 2.2)

$$C = C(p) > \max_{1\le i\le N} \|\varphi_i\|_{L^p(\omega)}. \tag{37}$$

Also $\mathcal{P}(p, 1)$ holds by the choice of C, since

$$C = C(p) > \max_{\substack{1\le i\le N,\\ \nu\in\{1,2,3\}}} \|D_\nu\varphi_i\|_{L^p(\omega)}. \tag{38}$$

Namely, since $\omega_\epsilon \subseteq \omega$, (36) holds for $|\beta| = 0$ (and *all* $\epsilon > 0$) using (37). For $\beta \in \mathbb{N}_0$ with $|\beta| = 1 = j$ (i.e., $\beta = e_\nu$ for some $\nu \in \{1, 2, 3\}$), and all $\epsilon > 0$ with $\epsilon = \epsilon j \le R/2 < 1$,

$$\epsilon^{|\beta|}\|D^\beta\varphi_i\|_{L^p(\omega_{\epsilon j})} = \epsilon\|D_\nu\varphi_i\|_{L^p(\omega_\epsilon)} \le \|D_\nu\varphi_i\|_{L^p(\omega)} \le C \le CB = CB^{|\beta|}. \tag{39}$$

Here we again used that $\omega_\epsilon \subseteq \omega$, (38), and that $B > 1$ (see Remark 2.2).

Induction hypothesis:

$$\text{Let } p \ge 5 \text{ and } j \in \mathbb{N}_0,\ j \ge 1. \text{ Then } \mathcal{P}(p, \tilde{j}) \text{ holds for all } \tilde{j} \le j. \tag{40}$$

We now prove that $\mathcal{P}(p, j+1)$ holds. Note that to prove this, it suffices to study $\beta \in \mathbb{N}_0^3$ with $|\beta| = j+1$. Namely, assume $\epsilon > 0$ is such that $\epsilon(j+1) \le R/2$ and let $\beta \in \mathbb{N}_0^3$ with $|\beta| < j+1$. Then $|\beta| \le j$ and $\epsilon j \le R/2$ so, by the definition of ω_δ and the induction hypothesis,

$$\epsilon^{|\beta|}\|D^\beta\varphi_i\|_{L^p(\omega_{\epsilon(j+1)})} \le \epsilon^{|\beta|}\|D^\beta\varphi_i\|_{L^p(\omega_{\epsilon j})} \le CB^{|\beta|}. \tag{41}$$

It therefore remains to prove that

$$\epsilon^{|\beta|}\|D^\beta\varphi_i\|_{L^p(\omega_{\epsilon(j+1)})} \le C\, B^{|\beta|} \quad \text{for all } \epsilon > 0 \text{ with } \epsilon(j+1) \le R/2 \text{ and all } \beta \in \mathbb{N}_0^3 \text{ with } |\beta| = j+1. \tag{42}$$

Remark 3.2. To use the induction hypothesis in its full strength, it is convenient to write, for $\ell > 0$, $\epsilon > 0$ such that $\epsilon\ell \leq R/2$, and $\sigma \in \mathbb{N}_0^3$ with $0 < |\sigma| \leq j$,

$$\|D^\sigma \varphi_i\|_{L^p(\omega_{\epsilon\ell})} = \|D^\sigma \varphi_i\|_{L^p(\omega_{\tilde{\epsilon}\tilde{j}})} \quad \text{with } \tilde{\epsilon} = \frac{\epsilon\ell}{|\sigma|},\ \tilde{j} = |\sigma|,$$

so that, by the induction hypothesis (applied on the term with $\tilde{\epsilon}$ and $\tilde{j}$) we get that

$$\|D^\sigma \varphi_i\|_{L^p(\omega_{\epsilon\ell})} \leq C\Big(\frac{B}{\tilde{\epsilon}}\Big)^{|\sigma|} = C\Big(\frac{|\sigma|}{\ell}\Big)^{|\sigma|}\Big(\frac{B}{\epsilon}\Big)^{|\sigma|}. \tag{43}$$

Compare this with (36). With the convention that $0^0 = 1$, (43) also holds for $|\sigma| = 0$.

We choose a function Φ (depending on j) satisfying

$$\Phi \in C_0^\infty(\omega_{\epsilon(j+3/4)}), \quad 0 \leq \Phi \leq 1, \quad \text{with } \Phi \equiv 1 \text{ on } \omega_{\epsilon(j+1)}. \tag{44}$$

Then

$$\|D^\beta \varphi_i\|_{L^p(\omega_{\epsilon(j+1)})} \leq \|\Phi D^\beta \varphi_i\|_p. \tag{45}$$

The estimate (42) — and hence, by induction, the proof of Proposition 2.1 — now follows from the equations (35) for the φ_i, (45) and the following two lemmas.

Lemma 3.3. *Assume the induction hypothesis* (40) *holds. Let Φ be as in* (44)*. Then for all $i \in \{1, \dots, N\}$, all $\epsilon > 0$ with $\epsilon(j+1) \leq R/2$, and all $\beta \in \mathbb{N}_0^3$ with $|\beta| = j+1$, both $\Phi D^\beta E(\boldsymbol{p})^{-1} V\varphi_i$ and $\Phi D^\beta E(\boldsymbol{p})^{-1}\varphi_i$ belong to $L^p(\mathbb{R}^3)$, and*

$$\|\Phi D^\beta E(\boldsymbol{p})^{-1} V\varphi_i\|_p \leq \frac{C}{4}\Big(\frac{B}{\epsilon}\Big)^{|\beta|}, \tag{46}$$

$$\|(\alpha^{-1} + \varepsilon_i)\Phi D^\beta E(\boldsymbol{p})^{-1}\varphi_i\|_p \leq \frac{C}{4}\Big(\frac{B}{\epsilon}\Big)^{|\beta|}, \tag{47}$$

where $C, B > 1$ are the constants in (36) (see also Remark 2.2).

Lemma 3.4. *Assume the induction hypothesis* (40) *holds. Let Φ be as in* (44)*. Then for all $i \in \{1, \dots, N\}$, all $\epsilon > 0$ with $\epsilon(j+1) \leq R/2$, and all $\beta \in \mathbb{N}_0^3$ with $|\beta| = j+1$, both $\Phi D^\beta E(\boldsymbol{p})^{-1} R_{\boldsymbol{\varphi}}\varphi_i$ and $\Phi D^\beta E(\boldsymbol{p})^{-1} K_{\boldsymbol{\varphi}}\varphi_i$ belong to $L^p(\mathbb{R}^3)$, and*

$$\|\alpha\, \Phi D^\beta E(\boldsymbol{p})^{-1} R_{\boldsymbol{\varphi}}\varphi_i\|_p \leq \frac{C}{4}\Big(\frac{B}{\epsilon}\Big)^{|\beta|},$$

$$\|\alpha\, \Phi D^\beta E(\boldsymbol{p})^{-1} K_{\boldsymbol{\varphi}}\varphi_i\|_p \leq \frac{C}{4}\Big(\frac{B}{\epsilon}\Big)^{|\beta|},$$

where $C, B > 1$ are the constants in (36) (see also Remark 2.2).

Remark 3.5. For $a, b \in \{1, \dots, N\}$, let $U_{a,b}$ denote the function

$$U_{a,b}(\boldsymbol{x}) = \int_{\mathbb{R}^3} \frac{\varphi_a(\boldsymbol{y})\overline{\varphi_b(\boldsymbol{y})}}{|\boldsymbol{x} - \boldsymbol{y}|}\, d\boldsymbol{y}, \quad \boldsymbol{x} \in \mathbb{R}^3. \tag{48}$$

In particular, $\|U_{a,b}\|_\infty \le C_1$ for all $a, b \in \{1, \dots, N\}$ (see (23)). Using (12) and (13), we can write

$$R_{\boldsymbol{\varphi}}\varphi_i = \sum_{\ell=1}^{N} U_{\ell,\ell}\varphi_i, \quad K_{\boldsymbol{\varphi}}\varphi_i = \sum_{\ell=1}^{N} U_{i,\ell}\varphi_\ell. \tag{49}$$

Hence Lemma 3.4 follows from the following lemma and the fact that $Z\alpha < 2/\pi < 1$.

Lemma 3.6. *Assume the induction hypothesis* (40) *holds. Let* Φ *be as in* (44). *For* $a, b \in \{1, \dots, N\}$, *let* $U_{a,b}$ *be given by* (48). *Then for all* $a, b, i \in \{1, \dots, N\}$, *all* $\epsilon > 0$ *with* $\epsilon(j+1) \le R/2$, *and all* $\beta \in \mathbb{N}_0^3$ *with* $|\beta| = j+1$, $\Phi D^\beta E(\boldsymbol{p})^{-1} U_{a,b}\varphi_i$ *belong to* $L^p(\mathbb{R}^3)$, *and*

$$\|\Phi D^\beta E(\boldsymbol{p})^{-1} U_{a,b}\varphi_i\|_p \le \frac{CZ}{4N}\Big(\frac{B}{\epsilon}\Big)^{|\beta|}, \tag{50}$$

where $C, B > 1$ *are the constants in* (36) (see also Remark 2.2).

It therefore remains to prove Lemmas 3.3 and 3.6. This will be done in the two following sections. □

4. Proof of Lemma 3.3

We prove Lemma 3.3 by proving (46) and (47) separately.

Proof of (46). Let $\sigma \in \mathbb{N}_0^3$ and $\nu \in \{1, 2, 3\}$ be such that $\beta = \sigma + e_\nu$, so that $D^\beta = D_\nu D^\sigma$. Notice that $|\sigma| = j$. Choose localization functions $\{\chi_k\}_{k=0}^{j}$ and $\{\eta_k\}_{k=0}^{j}$ as in Appendix B. Since $V\varphi_i \in H^{-1/2}(\mathbb{R}^3)$, and $E(\boldsymbol{p})^{-1}$ maps $H^s(\mathbb{R}^3)$ to $H^{s+1}(\mathbb{R}^3)$ for all $s \in \mathbb{R}$, Lemma B.1 (with $\ell = j$) implies that

$$\begin{aligned}\Phi D^\beta E(\boldsymbol{p})^{-1}[V\varphi_i] = \sum_{k=0}^{j} \Phi D_\nu E(\boldsymbol{p})^{-1} D^{\beta_k}\chi_k D^{\sigma-\beta_k}[V\varphi_i] + \sum_{k=0}^{j-1} \Phi D_\nu E(\boldsymbol{p})^{-1} D^{\beta_k}[\eta_k, D^{\mu_k}] D^{\sigma-\beta_{k+1}}[V\varphi_i] \\ + \Phi D_\nu E(\boldsymbol{p})^{-1} D^\sigma[\eta_j V\varphi_i],\end{aligned} \tag{51}$$

as an identity in $H^{-|\beta|+1/2}(\mathbb{R}^3)$ (we have also used that $E(\boldsymbol{p})^{-1}$ commutes with derivatives on any $H^s(\mathbb{R}^3)$). Here, $[\,\cdot, \cdot\,]$ denotes the commutator. Also, $|\beta_k| = k$, $|\mu_k| = 1$, and $0 \le \eta_k, \chi_k \le 1$. (For the support properties of η_k, χ_k, see Appendix B.) We will prove that each term on the right side of (51) belong to $L^p(\mathbb{R}^3)$, and bound their norms. The proof of (46) will follow by summing these bounds.

The first sum in (51). Let θ_k be the characteristic function of the support of χ_k (which is contained in ω). Since V is smooth on the closure of ω it follows from the induction hypothesis that the $D^{\sigma-\beta_k}[V\varphi_i]$ belong to $L^p(\omega')$ for any $\omega' \Subset \omega$. Also, the operator $\Phi D_\nu E(\boldsymbol{p})^{-1} D^{\beta_k}\chi_k$ is bounded on $L^p(\mathbb{R}^3)$ (as we will observe below). Therefore we can estimate, for $k \in \{0, \dots, j\}$,

$$\begin{aligned}\|\Phi D_\nu E(\boldsymbol{p})^{-1} D^{\beta_k}\chi_k D^{\sigma-\beta_k}[V\varphi_i]\|_p &= \|(\Phi E(\boldsymbol{p})^{-1} D_\nu D^{\beta_k}\chi_k)\theta_k D^{\sigma-\beta_k}[V\varphi_i]\|_p \\ &\le \|\Phi E(\boldsymbol{p})^{-1} D_\nu D^{\beta_k}\chi_k\|_{\mathcal{B}_p} \|\theta_k D^{\sigma-\beta_k}[V\varphi_i]\|_p.\end{aligned} \tag{52}$$

Here, $\|\cdot\|_{\mathcal{B}_p}$ is the operator norm on $\mathcal{B}_p := B(L^p(\mathbb{R}^3))$, the bounded operators on $L^p(\mathbb{R}^3)$.

For $k = 0$, the first factor on the right side of (52) can be estimated using Lemma C.1 (since $|\beta_0| = 0$). This way, since $\|\chi_0\|_\infty = \|\Phi\|_\infty = 1$,

$$\|\Phi E(\boldsymbol{p})^{-1} D_\nu \chi_0\|_{\mathcal{B}_p} \le K_1, \tag{53}$$

with $K_1 = K_1(p)$ the constant in (C.1).

For $k > 0$, the first factor on the right side of (52) can be estimated using (C.4) in Lemma C.2 (with $\mathfrak{r} = 1$, $\mathfrak{q}^* = \mathfrak{p} = p$). Since

$$\mathrm{dist}(\mathrm{supp}\, \chi_k, \mathrm{supp}\, \Phi) \geq \epsilon(k - 1 + 1/4)$$

and $\|\chi_k\|_\infty = \|\Phi\|_\infty = 1$, this gives (since $(\beta_k + e_\nu)! \leq (|\beta_k| + 1)! = (k+1)!$) that

$$\|\Phi E(\boldsymbol{p})^{-1} D_\nu D^{\beta_k} \chi_k\|_{\mathcal{B}_p} \leq \frac{32\sqrt{2}}{\pi} \frac{(k+1)!}{k} \Big(\frac{8}{\epsilon(k-1+1/4)}\Big)^k \leq \frac{256\sqrt{2}}{\pi} \Big(\frac{8}{\epsilon}\Big)^k. \tag{54}$$

It follows from (53) and (54) that, for all $k \in \{0, \dots, j\}$, $\nu \in \{1, 2, 3\}$,

$$\|\Phi E(\boldsymbol{p})^{-1} D_\nu D^{\beta_k} \chi_k\|_{\mathcal{B}_p} \leq C_2 \Big(\frac{8}{\epsilon}\Big)^k, \tag{55}$$

with C_2 as defined in (25).

It remains to estimate the second factor in (52). Recall the definition of the constant A in (24). It follows from (24) and (17) that, for all $\epsilon > 0$, $\ell \in \mathbb{N}_0$, and $\sigma \in \mathbb{N}_0^3$,

$$\epsilon^{|\sigma|} \sup_{\boldsymbol{x} \in \omega_{\epsilon\ell}} |D^\sigma V(\boldsymbol{x})| \leq A^{|\sigma|+1} |\sigma|!\, \ell^{-|\sigma|}, \tag{56}$$

with $\omega_{\epsilon\ell} \subseteq \omega$ as in defined in (16).

For $k = j$, since $\beta_j = \sigma$, we find, by (56) and the choice of C (see Remark 2.2), that

$$\|\theta_j V \varphi_i\|_p \leq \|V\|_{L^\infty(\omega)} \|\varphi_i\|_{L^p(\omega)} \leq CA. \tag{57}$$

The estimate for $k \in \{0, \dots, j-1\}$ is a bit more involved. We get, by Leibniz's rule, that

$$\|\theta_k D^{\sigma-\beta_k}[V\varphi_i]\|_p \leq \sum_{\mu \leq \sigma - \beta_k} \binom{\sigma-\beta_k}{\mu} \|\theta_k D^\mu V\|_\infty \|\theta_k D^{\sigma-\beta_k-\mu} \varphi_i\|_p. \tag{58}$$

Now, $\mathrm{supp}\, \theta_k = \mathrm{supp}\, \chi_k \subseteq \omega_{\epsilon(j-k+1/4)}$, so by (56), for all $\mu \leq \sigma - \beta_k$,

$$\|\theta_k D^\mu V\|_\infty \leq \sup_{\boldsymbol{x} \in \omega_{\epsilon(j-k+1/4)}} |D^\mu V(\boldsymbol{x})| \leq \epsilon^{-|\mu|} A^{|\mu|+1} |\mu|!\, (j-k)^{-|\mu|}. \tag{59}$$

By the induction hypothesis (in the form discussed in Remark 3.2),

$$\|\theta_k D^{\sigma-\beta_k-\mu} \varphi_i\|_p \leq \|D^{\sigma-\beta_k-\mu} \varphi_i\|_{L^p(\omega_{\epsilon(j-k)})} \leq C \Big(\frac{|\sigma-\beta_k-\mu|}{j-k}\Big)^{|\sigma-\beta_k-\mu|} \Big(\frac{B}{\epsilon}\Big)^{|\sigma-\beta_k-\mu|}. \tag{60}$$

It follows from (58), (59), and (60) that (using that $|\sigma| = j$, $|\beta_k| = k$, and (A.6), summing over $m = |\mu|$)

$$\|\theta_k D^{\sigma-\beta_k}[V\varphi_i]\|_p \leq CA \Big(\frac{B}{\epsilon}\Big)^{j-k} \sum_{m=0}^{j-k} \binom{j-k}{m} \frac{m!\,(j-k-m)^{j-k-m}}{(j-k)^{j-k}} \Big(\frac{A}{B}\Big)^m. \tag{61}$$

Note that, by (A.7), for $0 < m < j-k$,

$$\binom{j-k}{m} \frac{m!\,(j-k-m)^{j-k-m}}{(j-k)^{j-k}} \leq \frac{\mathrm{e}^{1/12}\sqrt{j-k}}{\sqrt{j-k-m}\,\mathrm{e}^m} \leq 1. \tag{62}$$

To see the last inequality, look at the cases $0 < m \le (j-k)/2$ and $j-k > m \ge (j-k)/2$ separately.

Hence (since $B > 2A$, see Remark 2.2), for any $k \in \{0, \dots, j-1\}$,

$$\|\theta_k D^{\sigma-\beta_k}[V\varphi_i]\|_p \le CA\Big(\frac{B}{\epsilon}\Big)^{j-k} \sum_{m=0}^{j-k} \Big(\frac{A}{B}\Big)^m \le 2CA\Big(\frac{B}{\epsilon}\Big)^{j-k}. \tag{63}$$

Note that, by (57), the same estimate holds true if $k = j$.

So, from (52), (55), (63), the fact that $\epsilon \le 1$ (since $\epsilon(j+1) \le R/2 \le 1/2$), and the choice of B (in particular, $B > 16$; see Remark 2.2), it follows that

$$\begin{aligned}\Big\|\sum_{k=0}^{j} \Phi D_\nu E(\boldsymbol{p})^{-1} D^{\beta_k}\chi_k D^{\sigma-\beta_k}[V\varphi_i]\Big\|_p &\le 2CAC_2\Big(\frac{B}{\epsilon}\Big)^j \sum_{k=0}^{j}\Big(\frac{8}{B}\Big)^k \\ &\le C(4AC_2)\Big(\frac{B}{\epsilon}\Big)^j \le \frac{C}{12}\Big(\frac{B}{\epsilon}\Big)^{j+1}.\end{aligned} \tag{64}$$

The second sum in (51). Note first that $[\eta_k, D^{\mu_k}] = -(D^{\mu_k}\eta_k)$ (recall that $|\mu_k| = 1$; see Lemma B.1).

Comparing the second sum in (51) with the first sum in (51), one sees that the second sum is the first one with j replaced by $j-1$ and χ_k replaced by $-D^{\mu_k}\eta_k$. Having now a derivative on the localization functions we have one derivative less falling on the term $V\varphi_i$. More precisely, the operator $D^{\sigma-\beta_{k+1}}$ contains $|\sigma-\beta_{k+1}| = j-(k+1) = (j-1)-k$ derivatives instead of $|\sigma-\beta_k| = j-k$ in $D^{\sigma-\beta_k}$. Then, to control $D^{\sigma-\beta_{k+1}}[V\varphi_i]$ (with the same method used above for $D^{\sigma-\beta_k}[V\varphi_i]$) we need that $\operatorname{supp} D^{\mu_k}\eta_k$ is contained in $\omega_{\epsilon((j-1)-k+1/4)}$. Indeed we have much more: as for χ_k we have $\operatorname{supp} D^{\mu_k}\eta_k \subseteq \omega_{\epsilon(j-k+1/4)} \subseteq \omega_{\epsilon((j-1)-k+1/4)}$. Finally, $\|D^{\mu_k}\eta_k\|_\infty \le C_*/\epsilon$, with $C_* > 0$ the constant in (B.3) in the appendix.

It follows that the second sum in (51) can be estimated as the first one, up to *one* extra factor of C_*/ϵ *and* up to replacing j by $j-1$ in the estimate (64). Hence, using that $\epsilon \le 1$, and the choice of B (see Remark 2.2), we get that

$$\begin{aligned}\Big\|\sum_{k=0}^{j-1} \Phi D_\nu E(\boldsymbol{p})^{-1} D^{\beta_k}[\eta_k, D^{\mu_k}] D^{\sigma-\beta_{k+1}}[V\varphi_i]\Big\|_p &\le \frac{C_*}{\epsilon} C(4AC_2)\Big(\frac{B}{\epsilon}\Big)^{j-1} \\ &\le C(4AC_2)\Big(\frac{B}{\epsilon}\Big)^j \le \frac{C}{12}\Big(\frac{B}{\epsilon}\Big)^{j+1}.\end{aligned} \tag{65}$$

The last term in (51). It remains to study

$$\Phi D^\beta E(\boldsymbol{p})^{-1}[\eta_j V\varphi_i]. \tag{66}$$

We split V in two parts, one supported around $\boldsymbol{x} = 0$, and one supported away from $\boldsymbol{x} = 0$, and study the two terms separately. We will prove below that this way, $\eta_j V\varphi_i$ is actually a function in $L^1(\mathbb{R}^3) + L^3(\mathbb{R}^3)$. Upon using suitable operator bounds on $\Phi D^\beta E(\boldsymbol{p})^{-1}\chi$ (for some suitable smooth χ's), combined with bounds on the norms of the two parts of $\eta_j V\varphi_i$, we will finish the proof.

Let $\rho = |\boldsymbol{x}_0|/4$, and let θ_ρ and $\theta_{\rho/2}$ be the characteristic functions of the balls $B_\rho(0)$ and $B_{\rho/2}(0)$, respectively. Choose $\widetilde{\chi}_\rho \in C_0^\infty(\mathbb{R}^3)$ with $\operatorname{supp}\widetilde{\chi}_\rho \subseteq B_\rho(0)$, $0 \le \widetilde{\chi}_\rho \le 1$, and $\widetilde{\chi}_\rho = 1$ on $B_{\rho/2}(0)$. Note that then

$$\operatorname{dist}(\operatorname{supp}\Phi, \operatorname{supp}\widetilde{\chi}_\rho) \geq \frac{|\boldsymbol{x}_0|}{2} = 2\rho, \tag{67}$$

by the choice of $\omega = B_R(\boldsymbol{x}_0)$, $R = \min\{1, |\boldsymbol{x}_0|/4\}$, since $\operatorname{supp}\Phi \subseteq \omega_{\epsilon(j+1)} \subseteq \omega$.

Now,

$$\Phi D^\beta E(\boldsymbol{p})^{-1}[\eta_j V\varphi_i] = \Phi D^\beta E(\boldsymbol{p})^{-1}[\eta_j V\widetilde{\chi}_\rho\varphi_i] + \Phi D^\beta E(\boldsymbol{p})^{-1}[\eta_j V(1-\widetilde{\chi}_\rho)\varphi_i]. \tag{68}$$

For the first term in (68), we use Lemma C.2, with $\mathfrak{p} = 1$, $\mathfrak{q} = p/(p-1)$, and $\mathfrak{r} = p$. Then $\mathfrak{p}, \mathfrak{r} \in [1,\infty)$ and $\mathfrak{q} > 1$, and $\mathfrak{q}^{-1} + p^{-1} = 1$. We get that (recall (67) and that $\widetilde{\chi}_\rho\theta_\rho = \widetilde{\chi}_\rho$),

$$\begin{aligned}\|\Phi D^\beta E(\boldsymbol{p})^{-1}[\eta_j V\widetilde{\chi}_\rho\varphi_i]\|_p &\leq \|\Phi D^\beta E(\boldsymbol{p})^{-1}\widetilde{\chi}_\rho\|_{\mathfrak{B}_{1,p}}\|\eta_j V\theta_\rho\varphi_i\|_1 \\ &\leq \frac{4\sqrt{2}}{\pi}\beta!\Big(\frac{8}{2\rho}\Big)^{|\beta|}(2\rho)^{3/\mathfrak{r}-2}\big(\mathfrak{r}(|\beta|+2)-3\big)^{-1/\mathfrak{r}}\|V\theta_\rho\varphi_i\|_1. \end{aligned} \tag{69}$$

Here we used that $\|\Phi\|_\infty = \|\widetilde{\chi}_\rho\|_\infty = 1$ and that $\eta_j \equiv 1$ where $\theta_\rho \neq 0$. Note that $j+1 \leq \epsilon^{-1}$ (since, by assumption, $\epsilon(j+1) \leq R/2 \leq 1/2$). Therefore,

$$\beta! \leq |\beta|! = (j+1)! \leq (j+1)^{j+1} \leq \epsilon^{-(j+1)} = \epsilon^{-|\beta|}. \tag{70}$$

Note furthermore that since $|\beta| = j+1 \geq 2$ and $\mathfrak{r} \geq 1$,

$$\big(\mathfrak{r}(|\beta|+2)-3\big)^{-1/\mathfrak{r}} \leq 1, \tag{71}$$

independently of β. It follows that

$$\|\Phi D^\beta E(\boldsymbol{p})^{-1}[\eta_j V\widetilde{\chi}_\rho\varphi_i]\|_p \leq \frac{4\sqrt{2}}{\pi}\Big(\frac{|\boldsymbol{x}_0|}{2}\Big)^{(3-2p)/p}\|V\theta_\rho\varphi_i\|_1\Big(\frac{16/|\boldsymbol{x}_0|}{\epsilon}\Big)^{|\beta|}. \tag{72}$$

Using Schwarz's inequality and that $Z\alpha < 2/\pi$,

$$\|V\theta_\rho\varphi_i\|_1 \leq \|V\theta_\rho\|_2\|\varphi_i\|_2 = Z\alpha\sqrt{|\boldsymbol{x}_0|\pi}\,\|\varphi_i\|_2 \leq \frac{2}{\sqrt{\pi}}\sqrt{|\boldsymbol{x}_0|}\|\varphi_i\|_2. \tag{73}$$

(Note that $\|V\theta_\rho\|_t < \infty \Leftrightarrow t < 3$.) It follows from (72), (73), and the choice of B and C (see Remark 2.2) that

$$\big\|\Phi D^\beta E(\boldsymbol{p})^{-1}[\eta_j V\widetilde{\chi}_\rho\varphi_i]\big\|_p \leq \frac{32}{\pi}|\boldsymbol{x}_0|^{3(2-p)/(2p)}\|\varphi_i\|_2\Big(\frac{16/|\boldsymbol{x}_0|}{\epsilon}\Big)^{|\beta|} \leq \frac{C}{24}\Big(\frac{B}{\epsilon}\Big)^{j+1}. \tag{74}$$

We now consider the second term in (68). Recall that Φ is supported in $\omega_{\epsilon(j+1)}$ and

$$\operatorname{dist}(\operatorname{supp}\Phi, \operatorname{supp}\eta_j) \geq \epsilon(j+1/4). \tag{75}$$

Again, we use Lemma C.2, this time with $\mathfrak{p} = 3$, $\mathfrak{q} = p/(p-1)$, and $\mathfrak{r} = 3p/(2p+3)$. Then

$$\mathfrak{p}^{-1} + \mathfrak{q}^{-1} + \mathfrak{r}^{-1} = 2, \quad \mathfrak{p} \in [1,\infty), \quad \mathfrak{q} > 1, \quad \mathfrak{r} \in [1, 3/2)$$

(since $p > 3$), and $\mathfrak{q}^{-1} + p^{-1} = 1$. This gives

$$
\begin{aligned}
&\big\|\Phi D^{\beta}E(\boldsymbol{p})^{-1}[\eta_j V(1-\widetilde{\chi}_\rho)\varphi_i]\big\|_p \\
&\qquad \le \big\|\Phi D^{\beta}E(\boldsymbol{p})^{-1}\eta_j\big\|_{\mathcal{B}_{3,p}}\,\big\|V(1-\widetilde{\chi}_\rho)\varphi_i\big\|_3 \\
&\qquad \le \frac{4\sqrt{2}}{\pi}\beta!\Big(\frac{8}{\epsilon(j+1/4)}\Big)^{|\beta|}\big(\epsilon(j+1/4)\big)^{3/\mathfrak{r}-2}\big(\mathfrak{r}(|\beta|+2)-3\big)^{-1/\mathfrak{r}}\|V(1-\widetilde{\chi}_\rho)\|_\infty\|\varphi_i\|_3.
\end{aligned}
$$

As before, we used that $\|\Phi\|_\infty=\|\eta_j\|_\infty=1$. Note that

$$
\beta!\Big(\frac{8}{j+1/4}\Big)^{|\beta|}\le 32^{|\beta|}\frac{|\beta|!}{(j+1)^{|\beta|}}=32^{|\beta|}\frac{(j+1)!}{(j+1)^{j+1}}\le 32^{|\beta|}. \tag{76}
$$

Since $\epsilon(j+1)\le R/2<1$ and $\mathfrak{r}<3/2$ it follows that $(\epsilon(j+1/4))^{3/\mathfrak{r}-2}\le 1$. Also, by the choice of ρ, the definition of V, and since $Z\alpha<2/\pi$,

$$
\big|((1-\theta_{\rho/2})V)(\boldsymbol{x})\big|\le\frac{8Z\alpha}{|\boldsymbol{x}_0|}\le\frac{16}{\pi|\boldsymbol{x}_0|},\quad \boldsymbol{x}\in\mathbb{R}^3. \tag{77}
$$

It follows from (77) (and that $0\le 1-\widetilde{\chi}_\rho\le 1-\theta_{\rho/2}$), (71), (76), and the choice of C and B (see Remark 2.2), that for all $i=1,\dots,N$ (recall that $|\beta|=j+1$)

$$
\big\|\Phi D^{\beta}E(\boldsymbol{p})^{-1}[\eta_j V(1-\widetilde{\chi}_\rho)\varphi_i]\big\|_p\le\frac{4\sqrt{2}}{\pi}\frac{16}{\pi|\boldsymbol{x}_0|}\|\varphi_i\|_3\Big(\frac{32}{\epsilon}\Big)^{|\beta|}\le\frac{C}{24}\Big(\frac{B}{\epsilon}\Big)^{j+1}. \tag{78}
$$

It follows from (68), (74), and (78) that

$$
\big\|\Phi D^{\beta}E(\boldsymbol{p})^{-1}[\eta_j V\varphi_i]\big\|_p\le\frac{C}{12}\Big(\frac{B}{\epsilon}\Big)^{j+1}. \tag{79}
$$

The estimate (46) now follows from (51) and the estimates (64), (65), and (79). □

Proof of (47). The constant functions $W_i(\boldsymbol{x})=\alpha^{-1}+\varepsilon_i$ trivially satisfy the conditions on V $(=Z\alpha|\cdot|^{-1})$ needed in the proof above. In fact, having assumed $A\ge\alpha^{-1}+\max_{1\le i\le N}|\varepsilon_i|$ (see Remark 2.2), (24) (and therefore (56)) trivially holds for W_i. Also, for the term $\Phi D^{\beta}E(\boldsymbol{p})^{-1}[\eta_j W_i\varphi_i]$ we proceed directly as for the term $\Phi D^{\beta}E(\boldsymbol{p})^{-1}[\eta_j V(1-\widetilde{\chi}_\rho)\varphi_i]$ above (but without any splitting in $\widetilde{\chi}_\rho$ and $1-\widetilde{\chi}_\rho$), using that $|W_i(\boldsymbol{x})|\le A$, $\boldsymbol{x}\in\mathbb{R}^3$. The proof of (47) therefore follows from the proof of (46) above, by the choice of C and B (see Remark 2.2).

This finishes the proof of Lemma 3.3. □

Remark 4.1. In fact, with a simple modification the arguments above (the local L^p-bound on the two terms in (68)) can be made to work just assuming that, for all $s>0$,

$$
V\varphi_i\in L^1(B_s(0)),\qquad V\varphi_i\in L^3(\mathbb{R}^3\setminus B_s(0)). \tag{80}
$$

5. Proof of Lemma 3.6

Proof of (50). Similarly to the case of the term with V in Lemma 3.3, we here use the localization functions introduced in Appendix B. With the notation as in the previous section (in particular, $\beta=\sigma+e_\nu$

with $|\sigma| = j$), Lemma B.1 (with $\ell = j$) implies that

$$\begin{aligned}
&\Phi D^{\beta} E(\boldsymbol{p})^{-1}[U_{a,b}\varphi_i] \\
&\quad = \sum_{k=0}^{j} \Phi D_\nu E(\boldsymbol{p})^{-1} D^{\beta_k} \chi_k D^{\sigma-\beta_k}[U_{a,b}\varphi_i] + \sum_{k=0}^{j-1} \Phi D_\nu E(\boldsymbol{p})^{-1} D^{\beta_k}[\eta_k, D^{\mu_k}] D^{\sigma-\beta_{k+1}}[U_{a,b}\varphi_i] \\
&\qquad\qquad + \Phi D_\nu E(\boldsymbol{p})^{-1} D^{\sigma}[\eta_j U_{a,b}\varphi_i], \qquad (81)
\end{aligned}$$

as an identity in $H^{-|\beta|}(\mathbb{R}^3)$. As in the proof of Lemma 3.3, $[\,\cdot\,,\cdot\,]$ denotes the commutator, $|\beta_k| = k$, $|\mu_k| = 1$, and $0 \le \eta_k, \chi_k \le 1$. (For the support properties of η_k, χ_k, see Appendix B.) As in the previous section, we will prove that each term on the right side of (81) belong to $L^p(\mathbb{R}^3)$, and bound their norms. The claim of the lemma will follow by summing these bounds.

The first sum in (81). We first proceed like for the similar sum in the proof of Lemma 3.3 (see (52), and after). Let θ_k be the characteristic function of the support of χ_k. It follows from the induction hypothesis, using that $-\Delta U_{a,b} = 4\pi \varphi_a \overline{\varphi_b}$, and Theorems D.5 and D.3, that the $D^{\sigma-\beta_k}[U_{a,b}\varphi_i]$ belong to $L^p(\omega')$ for any $\omega' \Subset \omega$. As before, the operator $\Phi D_\nu E(\boldsymbol{p})^{-1} D^{\beta_k} \chi_k$ is bounded on $L^p(\mathbb{R}^3)$. Then, for $k \in \{0, \dots, j\}$,

$$\begin{aligned}
\big\|\Phi D_\nu E(\boldsymbol{p})^{-1} D^{\beta_k} \chi_k D^{\sigma-\beta_k}[U_{a,b}\varphi_i]\big\|_p &= \big\|(\Phi E(\boldsymbol{p})^{-1} D_\nu D^{\beta_k} \chi_k)\theta_k D^{\sigma-\beta_k}[U_{a,b}\varphi_i]\big\|_p \\
&\le \big\|\Phi E(\boldsymbol{p})^{-1} D_\nu D^{\beta_k} \chi_k\|_{\mathcal{B}_p} \|\theta_k D^{\sigma-\beta_k}[U_{a,b}\varphi_i]\big\|_p. \qquad (82)
\end{aligned}$$

The first factor on the right side of (82) was estimated in the proof of Lemma 3.3 (see (55)): For all $k \in \{0, \dots, j\}$, $\nu \in \{1, 2, 3\}$,

$$\|\Phi E(\boldsymbol{p})^{-1} D_\nu D^{\beta_k} \chi_k\|_{\mathcal{B}_p} \le C_2 \Big(\frac{8}{\epsilon}\Big)^k, \qquad (83)$$

with C_2 the constant in (25).

It remains to estimate the second factor in (82). For $k = j$, since $\beta_j = \sigma$, we find that, by (23) and the choice of C and B (see Remark 2.2),

$$\|\theta_j U_{a,b}\varphi_i\|_p \le \|U_{a,b}\|_\infty \|\varphi_i\|_{L^p(\omega)} \le C_1 C \le C \Big(\frac{B}{\epsilon}\Big)^{1/2}. \qquad (84)$$

In the last inequality we also used that $\epsilon \le 1$ (since $\epsilon(j+1) \le R/2 < 1$).

The estimate for $k \in \{0, \dots, j-1\}$ is more involved. We get, by Leibniz's rule, that

$$\|\theta_k D^{\sigma-\beta_k}[U_{a,b}\varphi_i]\|_p \le \sum_{\mu \le \sigma-\beta_k} \binom{\sigma-\beta_k}{\mu} \big\|\theta_k (D^\mu U_{a,b})(D^{\sigma-\beta_k-\mu}\varphi_i)\big\|_p. \qquad (85)$$

We estimate separately each term on the right side of (85).

We separate into two cases.

If $\mu = 0$ then, using the induction hypothesis (i.e., $\mathcal{P}(p, j-k)$; recall that $\operatorname{supp}\theta_k \subseteq \omega_{\epsilon(j-k)}$) and (23),

$$\|\theta_k U_{a,b} D^{\sigma-\beta_k}\varphi_i\|_p \le C_1 C \Big(\frac{B}{\epsilon}\Big)^{j-k} \le \frac{C}{2}\Big(\frac{B}{\epsilon}\Big)^{j-k+1/2}. \qquad (86)$$

In the last inequality we used the choice of B (see Remark 2.2) and that $\epsilon \le 1$.

If $0 < \mu \le \sigma - \beta_k$, then (since $\operatorname{supp}\chi_k \subseteq \omega_{\epsilon(j-k+1/4)}$) Hölder's inequality (with $1/p = 1/(3p) + 2/(3p)$) and Corollary D.2 give

$$\begin{aligned}\|\theta_k(D^\mu U_{a,b})(D^{\sigma-\beta_k-\mu}\varphi_i)\|_p &\le \|\theta_k D^\mu U_{a,b}\|_{3p/2}\,\|\theta_k D^{\sigma-\beta_k-\mu}\varphi_i\|_{3p} \\ &\le K_2\|D^\mu U_{a,b}\|_{L^{3p/2}(\omega_{\epsilon(j-k+1/4)})}\,\|D^{\sigma-\beta_k-\mu}\varphi_i\|^{\theta}_{W^{1,p}(\omega_{\epsilon(j-k+1/4)})}\,\|D^{\sigma-\beta_k-\mu}\varphi_i\|^{1-\theta}_{L^p(\omega_{\epsilon(j-k+1/4)})}. \end{aligned} \quad (87)$$

Here, K_2 is the constant in Corollary D.2, and $\theta = 2/p < 1$. Note that $\omega_{\epsilon(j-k+1/4)} = B_r(\boldsymbol{x}_0)$ with $r \in [R/2, 1]$, since $\epsilon(j+1) \le R/2$ and $R = \min\{1, |\boldsymbol{x}_0|/4\}$

We will use Lemma 5.3 below to bound the first factor in (87). The last two factors we now bound using the induction hypothesis.

If $\mu \in \mathbb{N}_0^3$ is such that $0 < \mu \le \sigma - \beta_k$, then the induction hypothesis (in the form discussed in Remark 3.2) gives (recall here (18) and that $|\sigma| = j$, $|\beta_k| = k$) that for the last two factors in (87) we have

$$\|D^{\sigma-\beta_k-\mu}\varphi_i\|^{1-\theta}_{L^p(\omega_{\epsilon(j-k+1/4)})} \le \left[C\left(\frac{j-k-|\mu|}{j-k+1/4}\right)^{j-k-|\mu|}\left(\frac{B}{\epsilon}\right)^{j-k-|\mu|}\right]^{1-\theta} \quad (88)$$

and (using that $B > 1$ (see Remark 2.2) and $\epsilon(j-k+1/4) \le \epsilon(j+1) \le R/2 < 1$)

$$\begin{aligned}&\|D^{\sigma-\beta_k-\mu}\varphi_i\|^{\theta}_{W^{1,p}(\omega_{\epsilon(j-k+1/4)})} \\ &\quad\le \left[C\left(\frac{j-k-|\mu|}{j-k+1/4}\right)^{j-k-|\mu|}\left(\frac{B}{\epsilon}\right)^{j-k-|\mu|} + 3C\left(\frac{j-k-|\mu|+1}{j-k+1/4}\right)^{j-k-|\mu|+1}\left(\frac{B}{\epsilon}\right)^{j-k-|\mu|+1}\right]^{\theta} \\ &\quad\le \left[4C\left(\frac{j-k-|\mu|+1}{j-k+1/4}\right)^{j-k-|\mu|+1}\left(\frac{B}{\epsilon}\right)^{j-k-|\mu|+1}\right]^{\theta}. \end{aligned} \quad (89)$$

It follows from (88) and (89) that for all $\mu \in \mathbb{N}_0^3$ with $0 < \mu \le \sigma - \beta_k$,

$$\|D^{\sigma-\beta_k-\mu}\varphi_i\|^{\theta}_{W^{1,p}(\omega_{\epsilon(j-k+1/4)})}\,\|D^{\sigma-\beta_k-\mu}\varphi_i\|^{1-\theta}_{L^p(\omega_{\epsilon(j-k+1/4)})} \le C4^{\theta}\left(\frac{B}{\epsilon}\right)^{j-k-|\mu|+\theta}\left(\frac{j-k-|\mu|+1}{j-k+1/4}\right)^{j-k-|\mu|+\theta}. \quad (90)$$

From (87), Lemma 5.3, and (90) (using (A.6) in the appendix, summing over $m = |\mu|$), it follows that

$$\begin{aligned}&\sum_{0<\mu\le\sigma-\beta_k}\binom{\sigma-\beta_k}{\mu}\big\|\theta_k(D^\mu U_{a,b})(D^{\sigma-\beta_k-\mu}\varphi_i)\big\|_p \\ &\quad\le C^3C_3K_2\left(\frac{B}{\epsilon}\right)^{j-k+\theta}\sum_{m=1}^{j-k}4^{\theta}\binom{j-k}{m}\frac{(j-k-m+1)^{j-k-m+\theta}(m+1/4)^m}{(j-k+1/4)^{j-k+\theta}} \\ &\qquad\times\left[\left(\frac{1}{\sqrt{B}}\right)^m + \sqrt{m}\left(\frac{B(m+1/4)}{\epsilon(j-k+1/4)}\right)^{2\theta-2}\right]. \end{aligned} \quad (91)$$

Here, C_3 is the constant from (26). Recall also that $\theta = 2/p$.

We prove that for $m \in \{1, \dots, j-k\}$,

$$4^{\theta}\binom{j-k}{m}\frac{(j-k-m+1)^{j-k-m+\theta}(m+1/4)^m}{(j-k+1/4)^{j-k+\theta}} \le 10\epsilon^{-1/2+\theta}\frac{1}{\sqrt{m}}. \quad (92)$$

Note first that, since $\epsilon(j-k+1/4) \le \epsilon(j+1) \le 1$,

$$(j-k+1/4)^{1/2-\theta} \le \epsilon^{-1/2+\theta}. \tag{93}$$

This shows that the inequality in (92) is true for $m = j-k > 0$, since $\theta < 1$. For $m < j-k$, we use (A.8) in the appendix, and (93), to get that (since $(1+1/n)^n \le \mathrm{e}$)

$$\binom{j-k}{m}\frac{(j-k-m+1)^{j-k-m+\theta}(m+1/4)^m}{(j-k+1/4)^{j-k+\theta}} \le \frac{\mathrm{e}^{25/12}}{\sqrt{2\pi}}\frac{(j-k-m+1)^{\theta}}{(j-k-m)^{1/2}}\epsilon^{-1/2+\theta}\frac{1}{\sqrt{m}}. \tag{94}$$

Since $\theta < 1/2$ and $m \le j-k-1$, we have that

$$\frac{(j-k-m+1)^{\theta}}{(j-k-m)^{1/2}} \le 2^{\theta} \le \sqrt{2}. \tag{95}$$

The estimate (92) for $m \in \{1, \dots, j-k-1\}$ now follows from (94)–(95) (since $4^{\theta}\mathrm{e}^{25/12}/\sqrt{\pi} \le 10$).

Inserting (92) in (91) (and using again $\epsilon(j-k+1/4) \le 1$ and $2\theta - 2 < 0$) we find that

$$\begin{aligned}\sum_{0<\mu\le\sigma-\beta_k}\binom{\sigma-\beta_k}{\mu}\big\|\theta_k(D^{\mu}U_{a,b})(D^{\sigma-\beta_k-\mu}\varphi_i)\big\|_p &\\ \le 10C^3C_3K_2\Big(\frac{B}{\epsilon}\Big)^{j-k+\theta}\epsilon^{-1/2+\theta}\sum_{m=1}^{j-k}\bigg[\Big(\frac{1}{\sqrt{B}}\Big)^m + \frac{1}{B^{2-2\theta}}\frac{1}{m^{2-2\theta}}\bigg] &\\ \le 10C^3C_3K_2\Big(\frac{B}{\epsilon}\Big)^{j-k+1/2}\frac{1}{\sqrt{B}}(2+6), &\end{aligned} \tag{96}$$

where we used that $\theta \le 2/5$, $B \ge 4$ (see Remark 2.2), and $\sum_{m=1}^{\infty} m^{-6/5} \le 1 + \int_1^{\infty} x^{-6/5}\,dx = 6$ to estimate

$$\sum_{m=1}^{\infty}\Big(\frac{1}{\sqrt{B}}\Big)^m \le \frac{2}{\sqrt{B}}, \quad \frac{1}{B^{2-2\theta}}\sum_{m=1}^{\infty}\frac{1}{m^{2-2\theta}} \le \frac{6}{\sqrt{B}}. \tag{97}$$

This is the very essential reason for needing $p \ge 5$.

By the choice of B (see Remark 2.2) it follows that

$$\sum_{0<\mu\le\sigma-\beta_k}\binom{\sigma-\beta_k}{\mu}\big\|\theta_k(D^{\mu}U_{a,b})(D^{\sigma-\beta_k-\mu}\varphi_i)\big\|_p \le \frac{C}{2}\Big(\frac{B}{\epsilon}\Big)^{j-k+1/2}. \tag{98}$$

From (85), (86), and (98) it follows that for all $k \in \{0, \dots, j-1\}$,

$$\|\theta_k D^{\sigma-\beta_k}[U_{a,b}\varphi_i]\|_p \le C\Big(\frac{B}{\epsilon}\Big)^{j-k+1/2}. \tag{99}$$

Using (82), (83), (84), and (99) it follows for the first sum in (81) that

$$\begin{aligned}\Big\|\sum_{k=0}^{j}\Phi D_\nu E(\boldsymbol{p})^{-1}D^{\beta_k}\chi_k D^{\sigma-\beta_k}[U_{a,b}\varphi_i]\Big\|_p &\le C_2\sum_{k=0}^{j}8^k\epsilon^{-k}\|\theta_k D^{\sigma-\beta_k}[U_{a,b}\varphi_i]\|_p\\ &\le C_2C\Big(\frac{B}{\epsilon}\Big)^{j+1/2}\sum_{k=0}^{j}\Big(\frac{8}{B}\Big)^k. \end{aligned}\tag{100}$$

Since $B > 16$ (see Remark 2.2) the last sum is less than 2 and so for the first term in (81) we finally get, by the choice of B (see Remark 2.2) that

$$\Big\|\sum_{k=0}^{j}\Phi D_\nu E(\boldsymbol{p})^{-1}D^{\beta_k}\chi_k D^{\sigma-\beta_k}[U_{a,b}\varphi_i]\Big\|_p \le 2C_2C\Big(\frac{B}{\epsilon}\Big)^{j+1/2}\le \frac{CZ}{12N}\Big(\frac{B}{\epsilon}\Big)^{j+1}. \tag{101}$$

The second sum in (81). By the same arguments as for the second sum in (51) (see after (64)), it follows that the second sum in (81) can be estimated as the first one, up to *one* extra factor of C_*/ϵ (with $C_*>0$ the constant in (B.3) in the appendix) *and* up to replacing j by $j-1$ in the estimate (101). Hence, by the choice of B (see Remark 2.2)

$$\Big\|\sum_{k=0}^{j-1}\Phi D_\nu E(\boldsymbol{p})^{-1}D^{\beta_k}[\eta_k,\, D^{\mu_k}]D^{\sigma-\beta_{k+1}}[U_{a,b}\varphi_i]\Big\|_p \le \frac{C_*}{\epsilon}\frac{CZ}{12N}\Big(\frac{B}{\epsilon}\Big)^{j}\le \frac{CZ}{12N}\Big(\frac{B}{\epsilon}\Big)^{j+1}. \tag{102}$$

The last term in (81). Since $\sigma+e_\nu=\beta$, the last term in (81) equals

$$\Phi D^\beta E(\boldsymbol{p})^{-1}[\eta_j U_{a,b}\varphi_i].$$

We proceed exactly as for the term $\Phi D^\beta E(\boldsymbol{p})^{-1}[\eta_j V(1-\widetilde{\chi}_\rho)\varphi_i]$ in (68) (but without any splitting in $\widetilde{\chi}_\rho$ and $1-\widetilde{\chi}_\rho$), except that the estimate in (77) is replaced by $\|U_{a,b}\|_\infty \le C_1$ (see (23)). It follows, from the choice of B and C (see Remark 2.2) that (recall that $|\beta| = j+1$)

$$\begin{aligned}\|\Phi D^\beta E(\boldsymbol{p})^{-1}[\eta_j U_{a,b}\varphi_i]\|_p &\le \|\Phi D^\beta E(\boldsymbol{p})^{-1}\eta_j\|_{\mathcal{B}_{3,p}}\,\|U_{a,b}\varphi_i\|_3\\ &\le \frac{4\sqrt{2}}{\pi}C_1\|\varphi_i\|_3\Big(\frac{32}{\epsilon}\Big)^{|\beta|}\le \frac{CZ}{12N}\Big(\frac{B}{\epsilon}\Big)^{j+1}.\end{aligned}\tag{103}$$

The estimate (50) now follows from (81) and the estimates (101), (102), and (103).

This finishes the proof of Lemma 3.6. □

It remains to prove Lemma 5.3 below ($L^{3p/2}$-bound on derivatives of the Newton potential $U_{a,b}$ of products of orbitals, $\varphi_a\varphi_b$).

In the next lemma we first give an $L^{3p/2}$-estimate on the derivatives of the product of the orbitals φ_i, needed for the proof of the bound in Lemma 5.3 below.

Lemma 5.1. *Assume the induction hypothesis* (40) *holds. Then, for all* $a,b\in\{1,\dots,N\}$, *all* $\beta\in\mathbb{N}_0^3$ *with* $|\beta|\le j-1$, *and all* $\epsilon>0$ *with* $\epsilon(|\beta|+1)\le R/2$,

$$\|D^\beta(\varphi_a\overline{\varphi_b})\|_{L^{3p/2}(\omega_{\epsilon(|\beta|+1)})}\le 10K_2^2C^2(1+\sqrt{|\beta|})\Big(\frac{B}{\epsilon}\Big)^{|\beta|+2\theta}, \tag{104}$$

with K_2 *from Corollary D.2*, C *from Remark 2.2, and* $\theta=\theta(p)=2/p$.

Proof. By Leibniz's rule and Schwarz's inequality we get

$$\|D^{\beta}(\varphi_a\overline{\varphi_b})\|_{L^{3p/2}(\omega_{\epsilon(|\beta|+1)})} \le \sum_{\mu\le\beta}\binom{\beta}{\mu}\|D^{\mu}\varphi_a\|_{L^{3p}(\omega_{\epsilon(|\beta|+1)})}\|D^{\beta-\mu}\varphi_b\|_{L^{3p}(\omega_{\epsilon(|\beta|+1)})}.$$

We use Corollary D.2 (with $\omega_{\epsilon(|\beta|+1)} = B_r(\boldsymbol{x}_0)$, $r = R-\epsilon(|\beta|+1)$; note that $r\in[R/2,1]$, since $\epsilon(|\beta|+1)\le R/2$ and $R=\min\{1,|\boldsymbol{x}_0|/4\}$). This gives, with K_2 from Corollary D.2 and $\theta = 2/p$,

$$\begin{aligned}\|D^{\beta}(\varphi_a\overline{\varphi_b})\|_{L^{3p/2}(\omega_{\epsilon(|\beta|+1)})} \le K_2^2\sum_{\mu\le\beta}\binom{\beta}{\mu}\|D^{\mu}\varphi_a\|^{\theta}_{W^{1,p}(\omega_{\epsilon(|\beta|+1)})}\,\|D^{\mu}\varphi_a\|^{1-\theta}_{L^{p}(\omega_{\epsilon(|\beta|+1)})}&\\ \times\|D^{\beta-\mu}\varphi_b\|^{\theta}_{W^{1,p}(\omega_{\epsilon(|\beta|+1)})}\,\|D^{\beta-\mu}\varphi_b\|^{1-\theta}_{L^{p}(\omega_{\epsilon(|\beta|+1)})}.&\end{aligned}\tag{105}$$

We now use the induction hypothesis (in the form discussed in Remark 3.2) on each of the four factors in the sum on the right side of (105). Note that, by assumption, $\epsilon(|\beta|+1)\le\epsilon j\le R/2$ and $|\mu|<|\mu|+1\le|\beta|+1\le j$ (similarly, $|\beta-\mu|<|\beta-\mu|+1\le j$). Recalling (18), we therefore get, for all $\mu\in\mathbb{N}_0^3$ such that $\mu\le\beta$,

$$\begin{aligned}&\|D^{\mu}\varphi_a\|^{\theta}_{W^{1,p}(\omega_{\epsilon(|\beta|+1)})}\|D^{\mu}\varphi_a\|^{1-\theta}_{L^{p}(\omega_{\epsilon(|\beta|+1)})}\\ &\quad\le\Big[C\Big(\frac{|\mu|}{|\beta|+1}\Big)^{|\mu|}\Big(\frac{B}{\epsilon}\Big)^{|\mu|}\Big]^{1-\theta}\Big[C\Big(\frac{|\mu|}{|\beta|+1}\Big)^{|\mu|}\Big(\frac{B}{\epsilon}\Big)^{|\mu|}+3C\Big(\frac{|\mu|+1}{|\beta|+1}\Big)^{|\mu|+1}\Big(\frac{B}{\epsilon}\Big)^{|\mu|+1}\Big]^{\theta}\\ &\quad\le 4^{\theta}C\Big(\frac{B}{\epsilon}\Big)^{|\mu|+\theta}\frac{(|\mu|+1)^{\theta(|\mu|+1)}|\mu|^{|\mu|(1-\theta)}}{(|\beta|+1)^{|\mu|+\theta}},\end{aligned}$$

since (recall that $\epsilon(|\beta|+1)\le R/2<1$ and $B>1$)

$$\frac{|\mu|^{|\mu|}}{(|\mu|+1)^{|\mu|+1}}\,\epsilon(|\beta|+1)B^{-1}\le 1.$$

Proceeding similarly for the other two factors in (105), we get (using (A.6) in the appendix and summing over $m=|\mu|$) that

$$\begin{aligned}&\sum_{\mu\le\beta}\binom{\beta}{\mu}\|D^{\mu}\varphi_a\|_{L^{3p}(\omega_{\epsilon(|\beta|+1)})}\|D^{\beta-\mu}\varphi_b\|_{L^{3p}(\omega_{\epsilon(|\beta|+1)})}\\ &\quad\le 16^{\theta}(CK_2)^2\Big(\frac{B}{\epsilon}\Big)^{|\beta|+2\theta}\sum_{m=0}^{|\beta|}\binom{|\beta|}{m}\frac{\big[(m+1)^{m+1}(|\beta|-m+1)^{|\beta|-m+1}\big]^{\theta}\big[m^m(|\beta|-m)^{|\beta|-m}\big]^{1-\theta}}{(|\beta|+1)^{|\beta|+2\theta}}.\end{aligned}\tag{106}$$

We simplify the sum in m. Note that for $m=0$ and $m=|\beta|$, the summand is bounded by 1. Therefore, for $|\beta|\le 1$ the estimate (104) follows from (106), since $2\cdot 16^{\theta}\le 7$. It remains to consider $|\beta|\ge 2$. For $m\ge 1$, $m<|\beta|$, we can use (A.8) in the appendix to get (since $(1+1/n)^n\le \mathrm{e}$) that

$$\begin{aligned}&\sum_{0<\mu<\beta}\binom{\beta}{\mu}\|D^{\mu}\varphi_a\|_{L^{3p}(\omega_{\epsilon(|\beta|+1)})}\|D^{\beta-\mu}\varphi_b\|_{L^{3p}(\omega_{\epsilon(|\beta|+1)})}\\ &\qquad\le\frac{\mathrm{e}^{1/12}}{\sqrt{2\pi}}(CK_2)^2(16\mathrm{e}^2)^{\theta}\Big(\frac{B}{\epsilon}\Big)^{|\beta|+2\theta}\frac{|\beta|^{|\beta|+1/2}}{(|\beta|+1)^{|\beta|+2\theta}}\sum_{m=1}^{|\beta|-1}\frac{\big[(m+1)(|\beta|-m+1)\big]^{\theta}}{\sqrt{m}\sqrt{|\beta|-m}}.\end{aligned}$$

Since the function

$$f(x) = (x+1)(|\beta| - x + 1), \quad x \in [1, |\beta| - 1],$$

has its maximum (which is $(|\beta|/2+1)^2$) at $x = |\beta|/2$, and since

$$\sum_{m=1}^{|\beta|-1} \frac{1}{\sqrt{m}\sqrt{|\beta|-m}} \leq \int_0^{|\beta|} \frac{1}{\sqrt{x}\sqrt{|\beta|-x}}\,dx = \pi,$$

we get

$$\sum_{0<\mu<\beta} \binom{\beta}{\mu} \|D^\mu \varphi_a\|_{L^{3p}(\omega_{\epsilon(|\beta|+1)})} \|D^{\beta-\mu}\varphi_b\|_{L^{3p}(\omega_{\epsilon(|\beta|+1)})} \leq \mathrm{e}^{1/12}(16\mathrm{e}^2)^\theta \sqrt{\frac{\pi}{2}}(CK_2)^2\sqrt{|\beta|}\Big(\frac{B}{\epsilon}\Big)^{|\beta|+2\theta}. \quad (107)$$

The estimate (104) now follows from (105), (106), and (107), since $\mathrm{e}^{1/12}(16\mathrm{e}^2)^\theta\sqrt{\pi/2} \leq 10$ and $2\cdot 16^\theta \leq 7$ (recall that $p \geq 5$). This finishes the proof of Lemma 5.1. □

The next two lemmas, used in the proof above of Lemma 3.6, control the $L^{3p/2}$-norm of derivatives of $U_{a,b}$.

Lemma 5.2. *Define $U_{a,b}$ by* (48). *Then for all $a, b \in \{1, \dots, N\}$, and all $\mu \in \mathbb{N}_0^3$ with $|\mu| \leq 2$,*

$$\|D^\mu U_{a,b}\|_{L^{3p/2}(\omega)} \leq 4\pi K_3(C^2 + 2C_1/R^2), \quad (108)$$

with K_3 from Corollary D.4, C from Remark 2.2, C_1 from (23), *and* $R = \min\{1, |\boldsymbol{x}_0|/4\}$.

Proof. Recall that $\omega = B_R(\boldsymbol{x}_0)$, $R = \min\{1, |\boldsymbol{x}_0|/4\}$. Using (18), and Corollary D.4, we get, for all $\mu \in \mathbb{N}_0^3$ with $|\mu| \leq 2$,

$$\|D^\mu U_{a,b}\|_{L^{3p/2}(\omega)} \leq \|U_{a,b}\|_{W^{2,3p/2}(B_R(\boldsymbol{x}_0))} \leq K_3\Big\{\|\Delta U_{a,b}\|_{L^{3p/2}(B_{2R}(\boldsymbol{x}_0))} + \frac{1}{R^2}\|U_{a,b}\|_{L^{3p/2}(B_{2R}(\boldsymbol{x}_0))}\Big\}. \quad (109)$$

By the definition of $U_{a,b}$ (see (48)) we have

$$-\Delta U_{a,b}(\boldsymbol{x}) = 4\pi\, \varphi_a(\boldsymbol{x})\overline{\varphi_b}(\boldsymbol{x}) \quad \text{for } \boldsymbol{x} \in \mathbb{R}^3, \quad (110)$$

and $\|U_{a,b}\|_\infty \leq C_1$ (see (23)). Hence, from (109), Hölder's inequality, and the choice of C (see Remark 2.2; recall also that $p \geq 5$)

$$\begin{aligned}\|D^\mu U_{a,b}\|_{L^{3p/2}(\omega)} &\leq 4\pi K_3\Big\{\|\varphi_a\|_{L^{3p}(B_{2R}(\boldsymbol{x}_0))}\|\varphi_b\|_{L^{3p}(B_{2R}(\boldsymbol{x}_0))} + \frac{1}{R^2}\|U_{a,b}\|_\infty |B_{2R}(\boldsymbol{x}_0)|^{2/3p}\Big\} \\ &\leq 4\pi K_3(C^2 + 2C_1/R^2).\end{aligned}$$

□

Lemma 5.3. *Assume the induction hypothesis* (40) *holds, and define $U_{a,b}$ by* (48). *Then for all $a, b \in \{1, \dots, N\}$, all $k \in \{0, \dots, j-1\}$, all $\mu \in \mathbb{N}_0^3$ with $|\mu| \leq j-k$, and all $\epsilon > 0$ with $\epsilon(j+1) \leq R/2$,*

$$\begin{aligned}&\|D^\mu U_{a,b}\|_{L^{3p/2}(\omega_{\epsilon(j-k+1/4)})} \\ &\quad \leq C_3C^2\Big(\frac{\sqrt{B}}{\epsilon}\Big)^{|\mu|}\Big(\frac{|\mu|+1/4}{j-k+1/4}\Big)^{|\mu|} + C_3C^2\sqrt{|\mu|}\Big(\frac{B}{\epsilon}\Big)^{|\mu|+2\theta-2}\Big(\frac{|\mu|+1/4}{j-k+1/4}\Big)^{|\mu|+2\theta-2}, \quad (111)\end{aligned}$$

with $\theta = \theta(p) = 2/p$, C and B from Remark 2.2, and C_3 the constant in (26).

Proof. If $m := |\mu| \le 2$, (111) follows from Lemma 5.2 and the definition of C_3 in (26), since $\epsilon(j-k+1/4) \le \epsilon(j+1) \le R/2 < 1$, and $C, B > 1$ (see Remark 2.2).

If $m := |\mu| \ge 3$ then we write $\mu = \mu_{m-2} + e_{\nu_1} + e_{\nu_2}$ with $\nu_i \in \{1,2,3\}$, $i = 1, 2$, $|\mu_{m-2}| = m-2$. Then by the definition of the $W^{2,3p/2}$-norm (recall (18)) we find that

$$\begin{aligned} \|D^{\mu} U_{a,b}\|_{L^{3p/2}(\omega_{\epsilon(j-k+1/4)})} &\le \|D^{\mu_{m-2}} U_{a,b}\|_{W^{2,3p/2}(\omega_{\epsilon(j-k+1/4)})} \\ &= \|D^{\mu_{m-2}} U_{a,b}\|_{W^{2,3p/2}(\omega_{\tilde{\epsilon}_1(m-1+1/4)})}, \end{aligned} \tag{112}$$

with $\tilde{\epsilon}_1$ such that

$$\tilde{\epsilon}_1(m-1+1/4) = \epsilon(j-k+1/4). \tag{113}$$

To estimate the norm in (112) we will again use that $U_{a,b}$ satisfies (110). Applying $D^{\mu_{m-2}}$ to (110) and using the elliptic *a priori* estimate in Corollary D.4 (with $r = r_1 = R - \tilde{\epsilon}_1(m-1+1/4)$ and $\delta = \delta_1 = \tilde{\epsilon}_1/4$; recall that $\omega_\rho = B_{R-\rho}(\boldsymbol{x}_0)$) we get

$$\|D^{\mu} U_{a,b}\|_{L^{3p/2}(\omega_{\epsilon(j-k+1/4)})} \le 4\pi K_3 \|D^{\mu_{m-2}}(\varphi_a \overline{\varphi_b})\|_{L^{3p/2}(\omega_{\tilde{\epsilon}_1(m-1)})} + \frac{16K_3}{\tilde{\epsilon}_1^2} \|D^{\mu_{m-2}} U_{a,b}\|_{L^{3p/2}(\omega_{\tilde{\epsilon}_1(m-1)})}, \tag{114}$$

with $K_3 = K_3(p)$ the constant in (D.9). Notice that for this estimate we needed to enlarge the domain, taking the ball with a radius $\tilde{\epsilon}_1/4$ larger.

We now iterate the procedure (on the second term on the right side of (114)), with $\tilde{\epsilon}_i$ ($i = 2, \dots, \lfloor \frac{m}{2} \rfloor$) such that

$$\tilde{\epsilon}_i(m-2i+1+1/4) = \tilde{\epsilon}_{i-1}(m-2(i-1)+1), \tag{115}$$

and with $r = r_i = R - \tilde{\epsilon}_i(m-2i+1+1/4)$ and $\delta = \delta_i = \tilde{\epsilon}_i/4$. Note that (113) and (115) imply that

$$\tilde{\epsilon}_i \ge \tilde{\epsilon}_{i-1} \ge \cdots \ge \tilde{\epsilon}_1 = \epsilon\, \frac{j-k+1/4}{m-1+1/4} \quad \text{for } i = 2, \dots, \left\lfloor \frac{m}{2} \right\rfloor \tag{116}$$

and

$$\tilde{\epsilon}_i(m-2i+1) \le \tilde{\epsilon}_{i-1}(m-2(i-1)+1) \le \ldots \le \tilde{\epsilon}_1(m-1) \le \epsilon(j-k+1/4). \tag{117}$$

We get (with $\prod_{\ell=1}^{0} \equiv 1$ and $|\mu_{m-2i}| = m-2i$),

$$\begin{aligned} \|D^{\mu} U_{a,b}\|_{L^{3p/2}(\omega_{\epsilon(j-k+1/4)})} \le 4\pi K_3 \sum_{i=1}^{\lfloor \frac{m}{2} \rfloor} \biggl(\|D^{\mu_{m-2i}}(\varphi_a \overline{\varphi_b})\|_{L^{3p/2}(\omega_{\tilde{\epsilon}_i(m-2i+1)})} \prod_{\ell=1}^{i-1} \frac{16K_3}{\tilde{\epsilon}_\ell^2} \biggr) \\ + \biggl(\prod_{\ell=1}^{\lfloor \frac{m}{2} \rfloor} \frac{16K_3}{\tilde{\epsilon}_\ell^2} \biggr) \|D^{\mu_{m-2\lfloor \frac{m}{2} \rfloor}} U_{a,b}\|_{L^{3p/2}(\omega_{\tilde{\epsilon}_{\lfloor \frac{m}{2} \rfloor}(m-2\lfloor \frac{m}{2} \rfloor+1)})}. \end{aligned} \tag{118}$$

Using (116), and Lemma 5.1 for each $i = 1, \dots, \lfloor \frac{m}{2} \rfloor$ fixed (note that $\tilde{\epsilon}_i(m-2i+1) \le R/2$ by (117) since $\epsilon(j+1) \le R/2$) we get that

$$\|D^{\mu_{m-2i}}(\varphi_a \overline{\varphi_b})\|_{L^{3p/2}(\omega_{\tilde{\epsilon}_i(m-2i+1)})} \prod_{\ell=1}^{i-1} \frac{16K_3}{\tilde{\epsilon}_\ell^2} \le 20K_2^2 C^2 \sqrt{m} \Bigl(\frac{B}{\epsilon}\Bigr)^{m+2\theta-2} \Bigl(\frac{m-1+1/4}{j-k+1/4}\Bigr)^{m+2\theta-2} \Bigl(\frac{16K_3}{B^2}\Bigr)^{i-1}, \tag{119}$$

with K_2 from Corollary D.2, and $\theta = \theta(p) = 2/p$. Here we also used that $1 + \sqrt{m-2i} \le 2\sqrt{m}$. Note that $\sum_{i=1}^{\lfloor \frac{m}{2} \rfloor} (16K_3/B^2)^{i-1} < 2$ since $B^2 > 32K_3$ (see Remark 2.2). It follows that

$$4\pi K_3 \sum_{i=1}^{\lfloor \frac{m}{2} \rfloor} \biggl(\| D^{\mu_{m-2i}} (\varphi_a \overline{\varphi_b}) \|_{L^{3p/2}(\omega_{\tilde{\epsilon}_i (m-2i+1)})} \prod_{\ell=1}^{i-1} \frac{16K_3}{\tilde{\epsilon}_\ell^2} \biggr) \\ \le 160\pi K_2^2 K_3 C^2 \sqrt{m} \Bigl(\frac{B}{\epsilon} \Bigr)^{m+2\theta-2} \Bigl(\frac{m+1/4}{j-k+1/4} \Bigr)^{m+2\theta-2}. \quad (120)$$

We now estimate the last term in (118). Let $\delta = m - 2\lfloor \frac{m}{2} \rfloor \in \{0, 1\}$ (depending on whether m is even or odd). Then, using (116) and Lemma 5.2, we get that

$$\Biggl(\prod_{\ell=1}^{\lfloor \frac{m}{2} \rfloor} \frac{16K_3}{\tilde{\epsilon}_\ell^2} \Biggr) \| D^{\mu_{m-2\lfloor \frac{m}{2} \rfloor}} U_{a,b} \|_{L^{3p/2}(\omega_{\tilde{\epsilon}_{\lfloor \frac{m}{2} \rfloor}(m-2\lfloor \frac{m}{2} \rfloor+1)})} \\ \le 4\pi K_3 (C^2 + 2C_1/R^2) \Bigl(\frac{\sqrt{16K_3}}{\epsilon} \Bigr)^m \Bigl(\frac{m-1+1/4}{j-k+1/4} \Bigr)^m \Bigl(\frac{\epsilon(j-k+1/4)}{m-1+1/4} \Bigr)^\delta \\ \le 4\pi K_3 (1 + 2C_1/R^2) C^2 \Bigl(\frac{\sqrt{B}}{\epsilon} \Bigr)^m \Bigl(\frac{m+1/4}{j-k+1/4} \Bigr)^m. \quad (121)$$

Here we also used that $m \ge 3$ and $K_3 \ge 1$ (see Corollary D.4), that $C > 1$ and $B > 16K_3$ (see Remark 2.2), and that $\epsilon(j-k+1/4) \le 1$.

Combining (118), (120), and (121) finishes the proof of (111) in the case $m = |\mu| \ge 3$.

This finishes the proof of Lemma 5.3. $\square$

Appendix A: Multiindices and Stirling's formula

For $\sigma = (\sigma_1, \sigma_2, \sigma_3) \in \mathbb{N}_0^3$ we let $|\sigma| := \sigma_1 + \sigma_2 + \sigma_3$, and

$$D^\sigma := D_1^{\sigma_1} D_2^{\sigma_2} D_3^{\sigma_3}, \quad D_\nu := -\mathrm{i} \frac{\partial}{\partial x_\nu} =: -\mathrm{i}\, \partial_\nu, \quad \nu = 1, 2, 3. \quad (A.1)$$

This way,

$$\partial^\sigma := \frac{\partial^{|\sigma|}}{\partial \boldsymbol{x}^\sigma} := \frac{\partial^{|\sigma|}}{\partial x_1^{\sigma_1} x_2^{\sigma_2} x_3^{\sigma_3}} = (-\mathrm{i})^{|\sigma|} D^\sigma.$$

We let $\sigma! := \sigma_1!\, \sigma_2!\, \sigma_3!$, and, for $n \in \mathbb{N}_0$,

$$\binom{n}{\sigma} := \frac{n!}{\sigma!} = \frac{n!}{\sigma_1!\, \sigma_2!\, \sigma_3!}. \quad (A.2)$$

With this notation we have the multinomial formula, for $\boldsymbol{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$ and $n \in \mathbb{N}_0$,

$$(x_1 + x_2 + x_3)^n = \sum_{\substack{\mu \in \mathbb{N}_0^3 \\ |\mu| = n}} \binom{n}{\mu} \boldsymbol{x}^\mu. \quad (A.3)$$

Here, $\boldsymbol{x}^{\mu} := x_1^{\mu_1} x_2^{\mu_2} x_3^{\mu_3}$. It follows that

$$|\sigma|! \leq 3^{|\sigma|}\sigma! \quad \text{for all } \sigma \in \mathbb{N}_0^3, \tag{A.4}$$

since, using (A.2), that $(1, 1, 1)^{\mu} = 1$ for all $\mu \in \mathbb{N}_0^3$, and (A.3),

$$\frac{|\sigma|!}{\sigma!} = \binom{|\sigma|}{\sigma} \leq \sum_{\substack{\mu \in \mathbb{N}_0^3 \\ |\mu| = |\sigma|}} \binom{|\sigma|}{\mu} (1, 1, 1)^{\mu} = (1 + 1 + 1)^{|\sigma|} = 3^{|\sigma|}.$$

We also define

$$\binom{\sigma}{\mu} := \frac{\sigma!}{\mu!\,(\sigma - \mu)!} \tag{A.5}$$

for $\sigma, \mu \in \mathbb{N}_0^3$ with $\mu \leq \sigma$, that is, $\mu_\nu \leq \sigma_\nu$, $\nu = 1, 2, 3$. Note that for all $\sigma \in \mathbb{N}_0^3$ and $k \in \mathbb{N}_0$ (see [Kato 1996, Proposition 2.1]),

$$\sum_{\mu \leq \sigma, |\mu| = k} \binom{\sigma}{\mu} = \binom{|\sigma|}{k}. \tag{A.6}$$

Finally, by [Abramowitz and Stegun 1992, 6.1.38], we have the following generalization of Stirling's formula: For $m \in \mathbb{N}$,

$$m! = \sqrt{2\pi}\, m^{m+\frac{1}{2}} \exp\Bigl(-m + \frac{\vartheta}{12m}\Bigr) \quad \text{for some } \vartheta = \vartheta(m) \in (0, 1), \tag{A.7}$$

and so for $n, m \in \mathbb{N}$, $m < n$,

$$\begin{aligned} \binom{n}{m} &= \frac{1}{\sqrt{2\pi}} \frac{n^{n+1/2}}{m^{m+1/2}(n-m)^{n-m+1/2}} \exp\Bigl(\frac{\vartheta(n)}{12n} - \frac{\vartheta(m)}{12m} - \frac{\vartheta(n-m)}{12(n-m)}\Bigr) \\ &\leq \frac{\mathrm{e}^{1/12}}{\sqrt{2\pi}} \frac{n^{n+1/2}}{m^{m+1/2}(n-m)^{n-m+1/2}}. \end{aligned} \tag{A.8}$$

Appendix B: Choice of the localization

Recall that, for $\boldsymbol{x}_0 \in \mathbb{R}^3 \setminus \{0\}$ and $R = \min\{1, |\boldsymbol{x}_0|/4\}$, we have defined $\omega = B_R(\boldsymbol{x}_0)$, $\omega_\delta = B_{R-\delta}(\boldsymbol{x}_0)$, and that $\epsilon > 0$ is such that $\epsilon(j + 1) \leq R/2$. Also, recall (see (44)) that we have chosen a function Φ (depending on j) satisfying

$$\Phi \in C_0^\infty(\omega_{\epsilon(j+3/4)}), \quad 0 \leq \Phi \leq 1, \quad \text{with } \Phi \equiv 1 \text{ on } \omega_{\epsilon(j+1)}. \tag{B.1}$$

For $j \in \mathbb{N}$ we choose functions $\{\chi_k\}_{k=0}^{j}$, and $\{\eta_k\}_{k=0}^{j}$ (all depending on j) with the following properties (for an illustration, see Figures 1 and 2). The functions $\{\chi_k\}_{k=0}^{j}$ are such that

$$\chi_0 \in C_0^\infty(\omega_{\epsilon(j+1/4)}) \quad \text{with} \quad \chi_0 \equiv 1 \quad \text{on } \omega_{\epsilon(j+1/2)},$$

and, for $k = 1, \ldots, j$,

$$\chi_k \in C_0^\infty(\omega_{\epsilon(j-k+1/4)}) \text{ with } \begin{cases} \chi_k \equiv 1 & \text{on } \omega_{\epsilon(j-k+1/2)} \setminus \omega_{\epsilon(j-k+1+1/4)}, \\ \chi_k \equiv 0 & \text{on } \mathbb{R}^3 \setminus (\omega_{\epsilon(j-k+1/4)} \setminus \omega_{\epsilon(j-k+1+1/2)}). \end{cases}$$

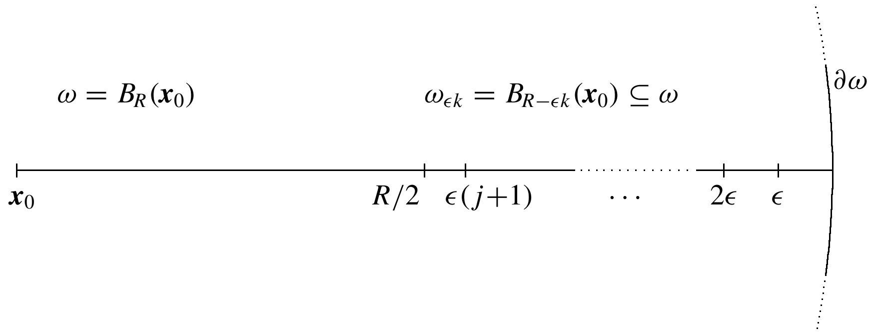

Figure 1. The geometry of $\omega = B_R(\boldsymbol{x}_0)$ and the $\omega_{\epsilon k} = B_{R-\epsilon k}(\boldsymbol{x}_0)$.

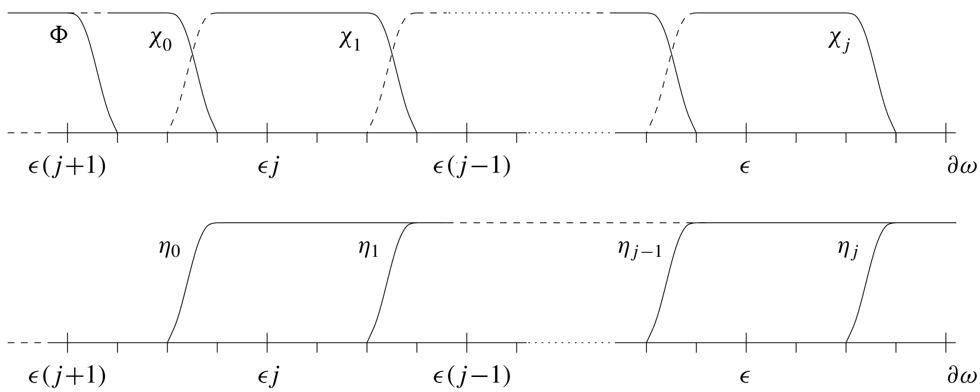

Figure 2. The localization functions.

Finally, the functions $\{\eta_k\}_{k=0}^{j}$ are such that for $k = 0, \dots, j$,

$$\eta_k \in C^\infty(\mathbb{R}^3) \ \text{ with } \begin{cases} \eta_k \equiv 1 & \text{on } \ \mathbb{R}^3 \setminus \omega_{\epsilon(j-k+1/4)}, \\ \eta_k \equiv 0 & \text{on } \ \omega_{\epsilon(j-k+1/2)}. \end{cases}$$

Moreover we ask that

$$\begin{aligned} &\chi_0 + \eta_0 \equiv 1 && \text{on } \ \mathbb{R}^3, \\ &\chi_k + \eta_k \equiv 1 && \text{on } \ \mathbb{R}^3 \setminus \omega_{\epsilon(j-k+1+1/4)} \ \text{ for } k = 1, \dots, j, \\ &\eta_k \equiv \chi_{k+1} + \eta_{k+1} && \text{on } \ \mathbb{R}^3 \ \text{ for } \ k = 0, \dots, j-1. \end{aligned} \tag{B.2}$$

Furthermore, we choose these localization functions such that, for a constant $C_* > 0$ (independent of ϵ, k, j, β) and for all $\beta \in \mathbb{N}_0^3$ with $|\beta| = 1$, we have that

$$|D^\beta \chi_k(\boldsymbol{x})| \le \frac{C_*}{\epsilon} \ \text{ and } \ |D^\beta \eta_k(\boldsymbol{x})| \le \frac{C_*}{\epsilon}, \tag{B.3}$$

for $k = 0, \dots, j$, and all $\boldsymbol{x} \in \mathbb{R}^3$.

The next lemma shows how to use these localization functions.

Lemma B.1. *For $j \in \mathbb{N}$ fixed, choose functions $\{\chi_k\}_{k=0}^{j}$, and $\{\eta_k\}_{k=0}^{j}$ as above, and let $\sigma \in \mathbb{N}_0^3$ with $|\sigma| = j$. For $\ell \in \mathbb{N}$ with $\ell \le j$, choose multiindices $\{\beta_k\}_{k=0}^{\ell}$ such that*

$$|\beta_k| = k \ \text{for}\ k = 0, \dots, \ell, \quad \beta_{k-1} < \beta_k \ \text{for}\ k = 1, \dots, \ell, \ \text{and}\ \beta_\ell \le \sigma.$$

Then for all $g \in \mathcal{S}'(\mathbb{R}^3)$,

$$D^\sigma g = \sum_{k=0}^{\ell} D^{\beta_k} \chi_k D^{\sigma-\beta_k} g + \sum_{k=0}^{\ell-1} D^{\beta_k}[\eta_k, D^{\mu_k}] D^{\sigma-\beta_{k+1}} g + D^{\beta_\ell} \eta_\ell D^{\sigma-\beta_\ell} g, \tag{B.4}$$

with $\mu_k = \beta_{k+1} - \beta_k$ for $k = 0, \dots, \ell-1$ (hence, $|\mu_k| = 1$).

Proof. We use induction on ℓ from $\ell = 1$ to $\ell = j$. We start by proving the claim for $\ell = 1$. By using property (B.2) of the localization functions and that $\beta_1 = \beta_0 + \mu_0 = \mu_0$ (since $\beta_0 = 0$) we find that

$$D^\sigma g = \chi_0 D^\sigma g + \eta_0 D^\sigma g = \chi_0 D^\sigma g + \eta_0 D^{\sigma-\beta_1+\mu_0} g. \tag{B.5}$$

The first term on the right side of (B.5) is the term corresponding to $k = 0$ in the first sum in (B.4). In the second term in (B.5), commuting the derivative through η_0, we find that

$$\eta_0 D^{\sigma-\beta_1+\mu_0} g = D^{\mu_0} \eta_0 D^{\sigma-\beta_1} g + [\eta_0, D^{\mu_0}] D^{\sigma-\beta_1} g.$$

Since $\eta_0 = \chi_1 + \eta_1$ by property (B.2), this implies that

$$\eta_0 D^{\sigma-\beta_1+\mu_0} g = D^{\beta_1} \chi_1 D^{\sigma-\beta_1} g + D^{\beta_1} \eta_1 D^{\sigma-\beta_1} g + [\eta_0, D^{\mu_0}] D^{\sigma-\beta_1} g. \tag{B.6}$$

The identity (B.4) for $\ell = 1$ follows from (B.5) and (B.6).

We now assume that (B.4) holds for $\ell - 1$ for some $\ell \ge 2$, i.e.,

$$D^\sigma g = \sum_{k=0}^{\ell-1} D^{\beta_k} \chi_k D^{\sigma-\beta_k} g + \sum_{k=0}^{\ell-2} D^{\beta_k}[\eta_k, D^{\mu_k}] D^{\sigma-\beta_{k+1}} g + D^{\beta_{\ell-1}} \eta_{\ell-1} D^{\sigma-\beta_{\ell-1}} g, \tag{B.7}$$

and prove it then holds for ℓ. Since $\beta_{\ell-1} = \beta_\ell - \mu_{\ell-1}$ we can rewrite the last term on the right side of (B.7) as

$$D^{\beta_{\ell-1}} \eta_{\ell-1} D^{\sigma-\beta_{\ell-1}} g = D^{\beta_{\ell-1}} \eta_{\ell-1} D^{\sigma-\beta_\ell+\mu_{\ell-1}} g.$$

Again, commuting the $\mu_{\ell-1}$-derivative through $\eta_{\ell-1}$ this implies that

$$\begin{aligned} D^{\beta_{\ell-1}} \eta_{\ell-1} D^{\sigma-\beta_{\ell-1}} g &= D^{\beta_{\ell-1}+\mu_{\ell-1}} \eta_{\ell-1} D^{\sigma-\beta_\ell} g + D^{\beta_{\ell-1}}[\eta_{\ell-1}, D^{\mu_{\ell-1}}] D^{\sigma-\beta_\ell} g \\ &= D^{\beta_\ell}(\eta_\ell + \chi_\ell) D^{\sigma-\beta_\ell} g + D^{\beta_{\ell-1}}[\eta_{\ell-1}, D^{\mu_{\ell-1}}] D^{\sigma-\beta_\ell} g, \end{aligned} \tag{B.8}$$

using (B.2). Collecting together (B.7) and (B.8) proves that (B.4) holds for ℓ.

The claim of the lemma then follows by induction. □

Appendix C: Norms of some operators on $L^p(\mathbb{R}^3)$

In this section we prove two lemmas on bounds on certain operators involving the operator $E(\boldsymbol{p}) = \sqrt{-\Delta + \alpha^{-2}}$.

Lemma C.1. *Let the operators $S_\nu = E(\boldsymbol{p})^{-1} D_\nu$, $\nu \in \{1, 2, 3\}$, be defined for $f \in \mathcal{S}(\mathbb{R}^3)$ by*

$$(S_\nu f)(\boldsymbol{x}) = (2\pi)^{-3/2} \int_{\mathbb{R}^3} \mathrm{e}^{\mathrm{i}\boldsymbol{x}\cdot\boldsymbol{p}} E(\boldsymbol{p})^{-1} p_\nu \hat{f}(\boldsymbol{p})\, d\boldsymbol{p},$$

with $\hat{f}(\boldsymbol{p}) = (2\pi)^{-3/2} \int_{\mathbb{R}^3} \mathrm{e}^{-\mathrm{i}\boldsymbol{x}\cdot\boldsymbol{p}} f(\boldsymbol{x})\, d\boldsymbol{x}$ the Fourier transform of f. (Here, $\boldsymbol{p} = (p_1, p_2, p_3)$.)

Then, for all $\mathfrak{p} \in (1, \infty)$, the S_ν extend to bounded operators, $S_\nu : L^{\mathfrak{p}}(\mathbb{R}^3) \to L^{\mathfrak{p}}(\mathbb{R}^3)$, $\nu \in \{1, 2, 3\}$. Clearly, $\|S_\nu\|_{\mathcal{B}_{\mathfrak{p}}} = \|S_\mu\|_{\mathcal{B}_{\mathfrak{p}}}$, $\nu \neq \mu$. We let

$$K_1 \equiv K_1(\mathfrak{p}) := \|S_1\|_{\mathcal{B}_{\mathfrak{p}}}. \tag{C.1}$$

Proof. This follows from [Sogge 1993, Theorem 0.2.6] and the *Remarks* right after it. In fact, since (by induction),

$$D_{\boldsymbol{p}}^{\gamma}\bigl(p_\nu E(\boldsymbol{p})^{-1}\bigr) = P_{\gamma,\nu}(\boldsymbol{p}) E(\boldsymbol{p})^{-1-2|\gamma|}, \quad \gamma \in \mathbb{N}_0^3,$$

for some polynomials $P_{\gamma,\nu}$ of degree $|\gamma| + 1$, the functions $m_\nu(\boldsymbol{p}) = p_\nu E(\boldsymbol{p})^{-1}$ are smooth and satisfy the estimates

$$|D_{\boldsymbol{p}}^{\gamma} m_\nu(\boldsymbol{p})| \le C_{\gamma,\nu} |\boldsymbol{p}|^{-|\gamma|}, \quad \gamma \in \mathbb{N}_0^3,$$

for some constants $C_{\gamma,\nu} > 0$, which is what is needed in the reference above. □

For $\mathfrak{p}, \mathfrak{q} \in [1, \infty]$, denote by $\|\cdot\|_{\mathcal{B}_{\mathfrak{p},\mathfrak{q}}}$ the operator norm on bounded operators from $L^{\mathfrak{p}}(\mathbb{R}^3)$ to $L^{\mathfrak{q}}(\mathbb{R}^3)$.

Lemma C.2. *For all $\mathfrak{p}, \mathfrak{r} \in [1, \infty)$, $\mathfrak{q} \in (1, \infty)$, with $\mathfrak{p}^{-1} + \mathfrak{q}^{-1} + \mathfrak{r}^{-1} = 2$, all $\alpha > 0$, all $\beta \in \mathbb{N}_0^3$ (with $|\beta| > 1$ if $\mathfrak{r} = 1$), and all $\Phi, \chi \in C^\infty(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3)$ with*

$$\operatorname{dist}(\operatorname{supp}(\chi), \operatorname{supp}(\Phi)) \ge d, \tag{C.2}$$

the operator $\Phi E(\boldsymbol{p})^{-1} D^\beta \chi$ is bounded from $L^{\mathfrak{p}}(\mathbb{R}^3)$ to $(L^{\mathfrak{q}}(\mathbb{R}^3))' = L^{\mathfrak{q}^*}(\mathbb{R}^3)$ (with $\mathfrak{q}^{-1} + \mathfrak{q}^{*-1} = 1$), and

$$\|\Phi E(\boldsymbol{p})^{-1} D^\beta \chi\|_{\mathcal{B}_{\mathfrak{p},\mathfrak{q}^*}} \le \frac{4\sqrt{2}}{\pi} \beta! \Bigl(\frac{8}{d}\Bigr)^{|\beta|} d^{3/\mathfrak{r}-2} \bigl(\mathfrak{r}(|\beta| + 2) - 3\bigr)^{-1/\mathfrak{r}} \|\Phi\|_\infty \|\chi\|_\infty. \tag{C.3}$$

In particular, (when $\mathfrak{r} = 1$, i.e., $\mathfrak{q}^* = \mathfrak{p}$),

$$\|\Phi E(\boldsymbol{p})^{-1} D^\beta \chi\|_{\mathcal{B}_{\mathfrak{p}}} \le \frac{32\sqrt{2}}{\pi} \frac{\beta!}{|\beta| - 1} \Bigl(\frac{8}{d}\Bigr)^{|\beta|-1} \|\Phi\|_\infty \|\chi\|_\infty, \tag{C.4}$$

for all $\beta \in \mathbb{N}_0^3$ with $|\beta| > 1$.

Proof. We use duality. Let $f, g \in \mathcal{S}(\mathbb{R}^3)$. Note that, since $\Phi f, D^\beta(\chi g) \in L^2(\mathbb{R}^3)$, the spectral theorem, and the formula

$$\frac{1}{\sqrt{x}} = \frac{1}{\pi} \int_0^\infty \frac{1}{x + t} \frac{dt}{\sqrt{t}}, \; x > 0, \tag{C.5}$$

imply that

$$(f,\,\Phi E(\boldsymbol{p})^{-1}D^{\beta}\chi g)=\frac{1}{\pi}\int_0^{\infty}\frac{dt}{\sqrt{t}}\,(f,\,\Phi(-\Delta+\alpha^{-2}+t)^{-1}D^{\beta}\chi g).$$

By using the formula for the kernel of the operator $(-\Delta+\alpha^{-2}+t)^{-1}$ [Reed and Simon 1975, (IX.30)], and integrating by parts, we get

$$\begin{aligned}(f,\,\Phi E(\boldsymbol{p})^{-1}D^{\beta}\chi g)&=\frac{1}{\pi}\int_0^{\infty}\int_{\mathbb{R}^3}\overline{f(\boldsymbol{x})}\Phi(\boldsymbol{x})\int_{\mathbb{R}^3}\frac{\mathrm{e}^{-\sqrt{\alpha^{-2}+t}\,|\boldsymbol{x}-\boldsymbol{y}|}}{4\pi|\boldsymbol{x}-\boldsymbol{y}|}\,[D^{\beta}(\chi g)](\boldsymbol{y})\,d\boldsymbol{x}\,d\boldsymbol{y}\,\frac{dt}{\sqrt{t}}\\&=\frac{(-1)^{|\beta|}}{\pi}\int_0^{\infty}\int_{\mathbb{R}^3}\overline{f(\boldsymbol{x})}\Phi(\boldsymbol{x})\int_{\mathbb{R}^3}\biggl(D_{\boldsymbol{y}}^{\beta}\,\frac{\mathrm{e}^{-\sqrt{\alpha^{-2}+t}\,|\boldsymbol{x}-\boldsymbol{y}|}}{4\pi|\boldsymbol{x}-\boldsymbol{y}|}\biggr)\chi(\boldsymbol{y})g(\boldsymbol{y})\,d\boldsymbol{x}\,d\boldsymbol{y}\,\frac{dt}{\sqrt{t}}.\end{aligned}$$

Notice that the integrand is different from zero only for $|\boldsymbol{x}-\boldsymbol{y}|\geq d$, due to the assumption (C.2). Hence, by Fubini's theorem,

$$(f,\,\Phi E(\boldsymbol{p})^{-1}D^{\beta}\chi g)=\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}F(\boldsymbol{x})H(\boldsymbol{x}-\boldsymbol{y})G(\boldsymbol{y})\,d\boldsymbol{x}\,d\boldsymbol{y},\tag{C.6}$$

with $F(\boldsymbol{x})=\overline{f(\boldsymbol{x})}\Phi(\boldsymbol{x})$, $G(\boldsymbol{y})=\chi(\boldsymbol{y})g(\boldsymbol{y})$, and

$$H(z)\equiv H_{\alpha,\beta,d}(z)=\mathbb{1}_{\{|\cdot|\geq d\}}(z)\frac{(-1)^{|\beta|}}{\pi}\int_0^{\infty}\biggl(D_z^{\beta}\,\frac{\mathrm{e}^{-\sqrt{\alpha^{-2}+t}\,|z|}}{4\pi|z|}\biggr)\frac{dt}{\sqrt{t}}.$$

Now, by (C.8) in Lemma C.3 below, uniformly for $\alpha>0$,

$$|H(z)|\leq\mathbb{1}_{\{|\cdot|\geq d\}}(z)\frac{\sqrt{2}}{4\pi^2}\frac{\beta!}{|z|}\biggl(\frac{8}{|z|}\biggr)^{|\beta|}\int_0^{\infty}\mathrm{e}^{-\sqrt{t}|z|/2}\,\frac{dt}{\sqrt{t}}=\mathbb{1}_{\{|\cdot|\geq d\}}(z)\frac{\sqrt{2}}{\pi^2}\frac{\beta!}{|z|^2}\biggl(\frac{8}{|z|}\biggr)^{|\beta|},$$

and so, for all $\alpha>0$, $\mathfrak{r}\in[1,\infty)$, and all $\beta\in\mathbb{N}_0^3$ (with $|\beta|>1$ if $\mathfrak{r}=1$),

$$\begin{aligned}\|H\|_{\mathfrak{r}}&\leq(4\pi)^{1/\mathfrak{r}}\frac{\sqrt{2}}{\pi^2}\beta!\,8^{|\beta|}\biggl(\int_d^{\infty}\bigl(|z|^{-|\beta|-2}\bigr)^{r}|z|^2\,d|z|\biggr)^{1/\mathfrak{r}}\\&=(4\pi)^{1/\mathfrak{r}}\frac{\sqrt{2}}{\pi^2}\beta!\,\Bigl(\frac{8}{d}\Bigr)^{|\beta|}d^{3/\mathfrak{r}-2}\bigl(\mathfrak{r}(|\beta|+2)-3\bigr)^{-1/\mathfrak{r}}.\end{aligned}$$

From this, (C.6), and Young's inequality [Lieb and Loss 2001, Theorem 4.2] (notice that $C_Y\leq 1$), follows that, with $\mathfrak{p},\mathfrak{q},\mathfrak{r}\in[1,\infty)$, $\mathfrak{p}^{-1}+\mathfrak{q}^{-1}+\mathfrak{r}^{-1}=2$,

$$\begin{aligned}|(f,\,\Phi E(\boldsymbol{p})^{-1}D^{\beta}\chi g)|&\leq\|F\|_{\mathfrak{q}}\|H\|_{\mathfrak{r}}\|G\|_{\mathfrak{p}}\\&\leq(4\pi)^{1/\mathfrak{r}}\frac{\sqrt{2}}{\pi^2}\beta!\,\Bigl(\frac{8}{d}\Bigr)^{|\beta|}d^{3/\mathfrak{r}-2}\bigl(\mathfrak{r}(|\beta|+2)-3\bigr)^{-1/\mathfrak{r}}\|F\|_{\mathfrak{q}}\|G\|_{\mathfrak{p}}\\&\leq\frac{4\sqrt{2}}{\pi}\beta!\,\Bigl(\frac{8}{d}\Bigr)^{|\beta|}d^{3/\mathfrak{r}-2}\bigl(\mathfrak{r}(|\beta|+2)-3\bigr)^{-1/\mathfrak{r}}\|\Phi\|_{\infty}\|\chi\|_{\infty}\|f\|_{\mathfrak{q}}\|g\|_{\mathfrak{p}}.\end{aligned}$$

Since $\mathscr{S}(\mathbb{R}^3)$ is dense in both $L^{\mathfrak{p}}(\mathbb{R}^3)$ and $L^{\mathfrak{q}^*}(\mathbb{R}^3)$, this finishes the proof of the lemma. ☐

Lemma C.3. *For all $s > 0$, $\boldsymbol{x} \in \mathbb{R}^3 \setminus \{0\}$, and $\beta \in \mathbb{N}_0^3$,*

$$\Big|\partial_{\boldsymbol{x}}^{\beta} \frac{1}{|\boldsymbol{x}|}\Big| \le \frac{\sqrt{2}\beta!}{|\boldsymbol{x}|}\Big(\frac{8}{|\boldsymbol{x}|}\Big)^{|\beta|}, \tag{C.7}$$

$$\Big|\partial_{\boldsymbol{x}}^{\beta} \frac{\mathrm{e}^{-s|\boldsymbol{x}|}}{|\boldsymbol{x}|}\Big| \le \frac{\sqrt{2}\beta!}{|\boldsymbol{x}|}\Big(\frac{8}{|\boldsymbol{x}|}\Big)^{|\beta|}\mathrm{e}^{-s|\boldsymbol{x}|/2}. \tag{C.8}$$

Proof. We will use the Cauchy inequalities [Hörmander 1973, Theorem 2.2.7]. To avoid confusion with the Euclidean norm $|\cdot|$ (in $\mathbb{R}^3$ or in $\mathbb{C}^3$), we denote by $|\cdot|_{\mathbb{C}}$ the absolute value in $\mathbb{C}$.

Let, for $\boldsymbol{w} = (w_1, w_2, w_3) \in \mathbb{C}^3$ and $r > 0$,

$$P_r^3(\boldsymbol{w}) = \{\boldsymbol{z} \in \mathbb{C}^3 \mid |z_\nu - w_\nu|_{\mathbb{C}} < r,\ \nu = 1, 2, 3\} \tag{C.9}$$

be the *polydisc* with *polyradius* $\boldsymbol{r} = (r, r, r)$. The Cauchy inequalities then state that if u is analytic in $P_r^3(\boldsymbol{w})$ and if $\sup_{\boldsymbol{z}\in P_r^3(\boldsymbol{w})} |u(\boldsymbol{z})|_{\mathbb{C}} \le M$, then

$$|\partial_{\boldsymbol{z}}^{\beta} u(\boldsymbol{w})|_{\mathbb{C}} \le M\beta!\, r^{-|\beta|} \quad \text{for all } \beta \in \mathbb{N}_0^3. \tag{C.10}$$

We take $\boldsymbol{w} = \boldsymbol{x} \in \mathbb{R}^3 \setminus \{0\} \subseteq \mathbb{C}^3$ and choose $r = |\boldsymbol{x}|/8$. We prove below that then we have (with $\boldsymbol{z}^2 := \sum_{\nu=1}^{3} z_\nu^2 \in \mathbb{C}$)

$$\mathrm{Re}(\boldsymbol{z}^2) \ge \tfrac{1}{2}|\boldsymbol{x}|^2 \quad \text{for } \boldsymbol{z} \in P_r^3(\boldsymbol{x}). \tag{C.11}$$

It follows that $\sqrt{\boldsymbol{z}^2} := \exp(\frac{1}{2}\mathrm{Log}\, \boldsymbol{z}^2)$ is well-defined and analytic on $P_r^3(\boldsymbol{x})$ with Log being the principal branch of the logarithm.

We will also argue below that

$$\mathrm{Re}(\sqrt{\boldsymbol{z}^2}) \ge \tfrac{1}{2}|\boldsymbol{x}| \quad \text{for } \boldsymbol{z} \in P_r^3(\boldsymbol{x}). \tag{C.12}$$

Then (by (C.11)) for all $\boldsymbol{z} \in P_r^3(\boldsymbol{x})$,

$$|\sqrt{\boldsymbol{z}^2}|_{\mathbb{C}} = \sqrt{|\boldsymbol{z}^2|_{\mathbb{C}}} \ge \sqrt{|\mathrm{Re}\, \boldsymbol{z}^2|} \ge |\boldsymbol{x}|/\sqrt{2}, \tag{C.13}$$

and (by (C.12)), for all $s \ge 0$ and all $\boldsymbol{z} \in P_r^3(\boldsymbol{x})$,

$$\big|\exp(-s\sqrt{\boldsymbol{z}^2})\big|_{\mathbb{C}} = \exp(-s\, \mathrm{Re}(\sqrt{\boldsymbol{z}^2})) \le \exp(-s|\boldsymbol{x}|/2). \tag{C.14}$$

Therefore, (C.7) and (C.8) follow from (C.10), (C.13), and (C.14).

It remains to prove (C.11) and (C.12).

For $\boldsymbol{z} \in P_r^3(\boldsymbol{x})$, write $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{a} + \mathrm{i}\boldsymbol{b}$ with $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3$ satisfying $|z_\nu - x_\nu|_{\mathbb{C}}^2 = a_\nu^2 + b_\nu^2 \le (|\boldsymbol{x}|/8)^2$. Then

$$\boldsymbol{z}^2 = |\boldsymbol{x} + \boldsymbol{a}|^2 - |\boldsymbol{b}|^2 + 2\mathrm{i}(\boldsymbol{x} + \boldsymbol{a}) \cdot \boldsymbol{b},$$

so, with $\epsilon = 1/8$,

$$\begin{aligned} \mathrm{Re}(\boldsymbol{z}^2) &= |\boldsymbol{x}|^2 + |\boldsymbol{a}|^2 + 2\,\boldsymbol{x} \cdot \boldsymbol{a} - |\boldsymbol{b}|^2 \\ &\ge (1 - \epsilon)|\boldsymbol{x}|^2 + (2 - \epsilon^{-1})|\boldsymbol{a}|^2 - (|\boldsymbol{a}|^2 + |\boldsymbol{b}|^2) \ge \tfrac{35}{64}|\boldsymbol{x}|^2 > \tfrac{1}{2}|\boldsymbol{x}|^2. \end{aligned}$$

This establishes (C.11) .

It follows from (C.11) that, with Arg the principal branch of the argument,

$$-\frac{\pi}{4} \leq \frac{1}{2}\mathrm{Arg}(z^2) \leq \frac{\pi}{4} \quad \text{for } z \in P_r^3(\boldsymbol{x}). \tag{C.15}$$

Furthermore (still for $z \in P_r^3(\boldsymbol{x})$), because of (C.15),

$$\mathrm{Re}(\sqrt{z^2}) = |z^2|_{\mathbb{C}}^{1/2} \cos(\tfrac{1}{2}\mathrm{Arg}(z^2)) \geq |z^2|_{\mathbb{C}}^{1/2}/\sqrt{2}. \tag{C.16}$$

Combining with (C.11) we get (C.12).

This finishes the proof of the lemma. □

Appendix D: Needed results

In this section we gather some results from the literature which are needed in our proofs.

Theorem D.1 [Adams and Fournier 2003, Theorem 5.8]. *Let Ω be a domain in $\mathbb{R}^n$ satisfying the cone condition. Let $m \in \mathbb{N}$, $\mathfrak{p} \in (1, \infty)$. If $m\mathfrak{p} > n$, let $\mathfrak{p} \leq \mathfrak{q} \leq \infty$; if $m\mathfrak{p} = n$, let $\mathfrak{p} \leq \mathfrak{q} < \infty$; if $m\mathfrak{p} < n$, let $\mathfrak{p} \leq \mathfrak{q} \leq \mathfrak{p}^* = n\mathfrak{p}/(n - m\mathfrak{p})$. Then there exists a constant K depending on $m, n, \mathfrak{p}, \mathfrak{q}$ and the dimensions of the cone C providing the cone condition for Ω, such that for all $u \in W^{m,\mathfrak{p}}(\Omega)$,*

$$\|u\|_{L^{\mathfrak{q}}(\Omega)} \leq K \|u\|_{W^{m,\mathfrak{p}}(\Omega)}^{\theta} \|u\|_{L^{\mathfrak{p}}(\Omega)}^{1-\theta}, \tag{D.1}$$

where $\theta = (n/m\mathfrak{p}) - (n/m\mathfrak{q})$.

We write $K = K(m, n, \mathfrak{p}, \mathfrak{q}, \Omega)$. We always use Theorem D.1 with $n = 3$, $m = 1$, and $\mathfrak{p} = p$, $\mathfrak{q} = 3p$ for some $p > 3$. Hence $m\mathfrak{p} > n$, $\mathfrak{p} \leq \mathfrak{q} \leq \infty$, and $\theta = \theta(p) = 2/p < 1$. Moreover, we always use it with Ω being a ball, whose radius in all cases is bounded from above by 1 and from below by $R/2$ for some $R > 0$ fixed.

Let $K_0 \equiv K_0(p) \equiv K(1, 3, p, 3p, B_1(0))$ with $B_1(0) \subseteq \mathbb{R}^3$ the unit ball (which does satisfy the cone condition). Note that then, by scaling, (D.1) implies that for all $r \leq 1$ and all $\boldsymbol{x}_0 \in \mathbb{R}^3$,

$$\|u\|_{L^{3p}(B_r(\boldsymbol{x}_0))} \leq K_0 r^{-\theta} \|u\|_{W^{1,p}(B_r(\boldsymbol{x}_0))}^{\theta} \|u\|_{L^p(B_r(\boldsymbol{x}_0))}^{1-\theta}, \tag{D.2}$$

with $\theta = 2/p$.

To summarize, we therefore have:

Corollary D.2. *Let $p > 3$ and $R \in (0, 1]$. Then there exists a constant K_2, depending only on p and R, such that for all $r \in [R/2, 1]$, $\boldsymbol{x}_0 \in \mathbb{R}^3$, and all $u \in W^{1,p}(B_r(\boldsymbol{x}_0))$,*

$$\|u\|_{L^{3p}(B_r(\boldsymbol{x}_0))} \leq K_2 \|u\|_{W^{1,p}(B_r(\boldsymbol{x}_0))}^{\theta} \|u\|_{L^p(B_r(\boldsymbol{x}_0))}^{1-\theta}, \tag{D.3}$$

with $\theta = 2/p$.

Here,

$$K_2 \equiv K_2(p, R) = (2/R)^{2/p} K_0(p), \tag{D.4}$$

where $K_0(p) = K(1, 3, p, 3p, B_1(0))$ in Theorem D.1 above.

Theorem D.3 [Chen and Wu 1998, Theorem 4.2]. *Let Ω be a bounded domain in $\mathbb{R}^n$ and let $a^{ij} \in C(\overline{\Omega})$, $b^i, c \in L^\infty(\Omega)$ $i, j \in \{1, \dots, n\}$, with $\lambda, \Lambda > 0$ such that*

$$\sum_{i,j=1}^{n} a^{ij}\xi_i\xi_j \geq \lambda|\xi|^2 \quad \textit{for all } x \in \Omega,\ \xi \in \mathbb{R}^n, \tag{D.5}$$

$$\sum_{i,j=1}^{n} \|a^{ij}\|_{L^\infty(\Omega)} + \sum_{i=1}^{n} \|b^i\|_{L^\infty(\Omega)} + \|c\|_{L^\infty(\Omega)} \leq \Lambda. \tag{D.6}$$

Suppose $u \in W^{2,\mathfrak{p}}_{\text{loc}}(\Omega)$ satisfies

$$Lu = \sum_{i,j=1}^{n} -a^{ij} D_i D_j u + \sum_{i=1}^{n} b^i D_i u + cu = f. \tag{D.7}$$

Then for any $\Omega' \Subset \Omega$,

$$\|u\|_{W^{2,\mathfrak{p}}(\Omega')} \leq C\left\{\frac{1}{\lambda}\|f\|_{L^{\mathfrak{p}}(\Omega)} + \|u\|_{L^{\mathfrak{p}}(\Omega)}\right\}, \tag{D.8}$$

where C depends only on n, $\mathfrak{p}$, Λ/λ, $\operatorname{dist}\{\Omega', \partial\Omega\}$, and the modulus of continuity of the a^{ij}'s.

We use Theorem D.3 in the case where Ω' and Ω are concentric balls (and with $n = 3$, $\mathfrak{p} = 3p/2$, $a^{ij} = \delta_{ij}$, $b^i = c = 0$; hence $\Lambda = \lambda = 1$). Reading the proof of the theorem above with this case in mind (see [Chen and Wu 1998, Lemma 4.1] in particular), one can make the dependence on $\operatorname{dist}\{\Omega', \partial\Omega\}$ explicit. More precisely:

Corollary D.4. *For all $p > 1$ there exists a constant $K_3 = K_3(p) \geq 1$ such that*

$$\|u\|_{W^{2,3p/2}(B_r(\boldsymbol{x}_0))} \leq K_3\big\{\|\Delta u\|_{L^{3p/2}(B_{r+\delta}(\boldsymbol{x}_0))} + \delta^{-2}\|u\|_{L^{3p/2}(B_{r+\delta}(\boldsymbol{x}_0))}\big\}. \tag{D.9}$$

for all $u \in W^{2,3p/2}(B_{r+\delta}(\boldsymbol{x}_0))$ (with $\boldsymbol{x}_0 \in \mathbb{R}^3$, $r, \delta > 0$).

Theorem D.5 [Evans 1998, Theorem 5, Section 5.6.2 (Morrey's inequality)]. *Let Ω be a bounded, open subset in $\mathbb{R}^n$, $n \geq 2$, and suppose $\partial\Omega$ is C^1. Assume $n < \mathfrak{p} < \infty$, and $u \in W^{1,\mathfrak{p}}(\Omega)$. Then u has a version $u^* \in C^{0,\gamma}(\overline{\Omega})$, for $\gamma = 1 - n/\mathfrak{p}$, with the estimate*

$$\|u^*\|_{C^{0,\gamma}(\overline{\Omega})} \leq K_4\|u\|_{W^{1,\mathfrak{p}}(\Omega)}. \tag{D.10}$$

The constant K_4 depends only on $\mathfrak{p}$, n, and Ω.

Here, u^* is a version of the given u if $u = u^*$ a.e. Above,

$$\|u\|_{C^{0,\gamma}(\overline{\Omega})} := \sup_{\boldsymbol{x}\in\Omega} |u(\boldsymbol{x})| + \sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in\Omega \\ \boldsymbol{x}\neq\boldsymbol{y}}} \frac{|u(\boldsymbol{x}) - u(\boldsymbol{y})|}{|\boldsymbol{x} - \boldsymbol{y}|^\gamma}. \tag{D.11}$$

Of course, $\sup_{\boldsymbol{x}\in\Omega} |u(\boldsymbol{x})| \leq \|u\|_{C^{0,\gamma}(\overline{\Omega})}$.

Remark D.6. In [Evans 1998, p. 245] a definition of the $W^{m,\mathfrak{p}}$-norm is used which is slightly different from ours (see (18)), but which is an equivalent norm by the equivalence of norms in finite-dimensional vector spaces. Therefore, (D.10) holds with our definition of the norm, though the constant K_4 is not the same as the one in [Evans 1998, Theorem 5, Section 5.6.2].

Acknowledgement

The authors thank Heinz Siedentop for suggesting the study of this problem.

References

[Abramowitz and Stegun 1992] M. Abramowitz and I. A. Stegun (editors), *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, Dover, New York, 1992. MR 94b:00012 Zbl 0643.33001

[Adams and Fournier 2003] R. A. Adams and J. J. F. Fournier, *Sobolev spaces*, 2nd ed., Pure and Applied Mathematics **140**, Elsevier, Amsterdam, 2003. MR 2009e:46025 Zbl 1098.46001

[Carmona et al. 1990] R. Carmona, W. C. Masters, and B. Simon, "Relativistic Schrödinger operators: asymptotic behavior of the eigenfunctions", *J. Funct. Anal.* **91**:1 (1990), 117–142. MR 91i:35139 Zbl 0716.35006

[Chen and Wu 1998] Y.-Z. Chen and L.-C. Wu, *Second order elliptic equations and elliptic systems*, Translations of Mathematical Monographs **174**, American Mathematical Society, Providence, RI, 1998. MR 99i:35016 Zbl 0902.35003

[Dall'Acqua et al. 2008] A. Dall'Acqua, T. Østergaard Sørensen, and E. Stockmeyer, "Hartree–Fock theory for pseudorelativistic atoms", *Ann. Henri Poincaré* **9**:4 (2008), 711–742. MR 2009k:81258 Zbl 1192.81406

[Daubechies and Lieb 1983] I. Daubechies and E. H. Lieb, "One-electron relativistic molecules with Coulomb interaction", *Comm. Math. Phys.* **90**:4 (1983), 497–510. MR 85j:81007 Zbl 0946.81522

[Enstedt and Melgaard 2009] M. Enstedt and M. Melgaard, "Existence of infinitely many distinct solutions to the quasirelativistic Hartree–Fock equations", *Int. J. Math. Math. Sci.* **2009** (2009), Art. ID 651871. MR 2011d:81094 Zbl 1181.35288

[Evans 1998] L. C. Evans, *Partial differential equations*, Graduate Studies in Mathematics **19**, American Mathematical Society, Providence, RI, 1998. MR 99e:35001 Zbl 0902.35002

[Fefferman and de la Llave 1986] C. Fefferman and R. de la Llave, "Relativistic stability of matter, I", *Rev. Mat. Iberoamericana* **2**:1-2 (1986), 119–213. MR 88g:81155 Zbl 0602.58015

[Fournais et al. 2009] S. Fournais, M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, and T. Østergaard Sørensen, "Analytic structure of solutions to multiconfiguration equations", *J. Phys. A* **42**:31 (2009), Paper #315208. MR 2010m:81398 Zbl 1181.35282

[Frank and Lenzmann 2010] R. L. Frank and E. Lenzmann, "On ground states for the L^2-critical boson star equation", preprint, 2010. arXiv 0910.2721v2

[Friedman 1958] A. Friedman, "On the regularity of the solutions of non-linear elliptic and parabolic systems of partial differential equations", *J. Math. Mech.* **7** (1958), 43–59. MR 22 #9739 Zbl 0078.27702

[Friesecke 2003] G. Friesecke, "On the infinitude of non-zero eigenvalues of the single-electron density matrix for atoms and molecules", *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.* **459**:2029 (2003), 47–52. MR 2004g:81049 Zbl 1046.81029

[Hashimoto 2006] Y. Hashimoto, "A remark on the analyticity of the solutions for non-linear elliptic partial differential equations", *Tokyo J. Math.* **29**:2 (2006), 271–281. MR 2007m:35065 Zbl 1213.35214

[Herbst 1977] I. W. Herbst, "Spectral theory of the operator $(p^2+m^2)^{1/2} - Ze^2/r$", *Comm. Math. Phys.* **53**:3 (1977), 285–294. MR 55 #9790 Zbl 0375.35047

[Hörmander 1969] L. Hörmander, *Linear partial differential operators*, Die Grundlehren der mathematischen Wissenschaften **116**, Springer, New York, 1969. MR 40 #1687 Zbl 0175.39201

[Hörmander 1973] L. Hörmander, *An introduction to complex analysis in several variables*, North-Holland Mathematical Library **7**, North-Holland, Amsterdam, 1973. MR 49 #9246 Zbl 0271.32001

[Kato 1995] T. Kato, *Perturbation theory for linear operators*, Springer, Berlin, 1995. MR 96a:47025 Zbl 0836.47009

[Kato 1996] K. Kato, "New idea for proof of analyticity of solutions to analytic nonlinear elliptic equations", *SUT J. Math.* **32**:2 (1996), 157–161. MR 97k:35064 Zbl 0894.35036

[Krantz and Parks 2002] S. G. Krantz and H. R. Parks, *A primer of real analytic functions*, 2nd ed., Birkhäuser, Boston, 2002. MR 2003f:26045 Zbl 1015.26030

[Lewin 2004a] M. Lewin, *Quelques modèles non linéaires en mécanique quantique (Some nonlinear models in quantum mechanics)*, thesis, Université Paris Dauphine, 21 June 2004, Available at http://tel.archives-ouvertes.fr/tel-00006306.

[Lewin 2004b] M. Lewin, "Solutions of the multiconfiguration equations in quantum chemistry", *Arch. Ration. Mech. Anal.* **171**:1 (2004), 83–114. MR 2005e:81269 Zbl 1063.81102

[Lewis et al. 1997] R. T. Lewis, H. Siedentop, and S. A. Vugalter, "The essential spectrum of relativistic multi-particle operators", *Ann. Inst. H. Poincaré Phys. Théor.* **67**:1 (1997), 1–28. MR 98i:81046 Zbl 0886.35126

[Lieb and Loss 2001] E. H. Lieb and M. Loss, *Analysis*, 2nd ed., Graduate Studies in Mathematics **14**, American Mathematical Society, Providence, RI, 2001. MR 2001i:00001 Zbl 0966.26002

[Lieb and Yau 1988] E. H. Lieb and H.-T. Yau, "The stability and instability of relativistic matter", *Comm. Math. Phys.* **118**:2 (1988), 177–213. MR 90c:81251 Zbl 0686.35099

[Morrey 2008] C. B. Morrey, Jr., *Multiple integrals in the calculus of variations*, Springer, Berlin, 2008. MR 2492985 Zbl 1213.49002

[Morrey and Nirenberg 1957] C. B. Morrey, Jr. and L. Nirenberg, "On the analyticity of the solutions of linear elliptic systems of partial differential equations", *Comm. Pure Appl. Math.* **10** (1957), 271–290. MR 19,654b Zbl 0082.09402

[Nardini 1986] F. Nardini, "Exponential decay for the eigenfunctions of the two-body relativistic Hamiltonian", *J. Analyse Math.* **47** (1986), 87–109. MR 88c:81017 Zbl 0626.35075

[Reed and Simon 1975] M. Reed and B. Simon, *Methods of modern mathematical physics, II: Fourier analysis, self-adjointness*, Academic Press, New York, 1975. MR 58 #12429b Zbl 0308.47002

[Sogge 1993] C. D. Sogge, *Fourier integrals in classical analysis*, Cambridge Tracts in Mathematics **105**, Cambridge University Press, Cambridge, 1993. MR 94c:35178 Zbl 0783.35001

[Weder 1975] R. A. Weder, "Spectral analysis of pseudodifferential operators", *J. Funct. Anal.* **20**:4 (1975), 319–337. MR 53 #6366 Zbl 0317.47035

[Zhislin and Vugalter 2002] G. M. Zhislin and S. A. Vugalter, "On the discrete spectrum of Hamiltonians for pseudo-relativistic electrons", *Izv. Ross. Akad. Nauk Ser. Mat.* **66**:1 (2002), 71–102. In Russian; translated in *Izv. Math.* **66**:1 (2002), 71–102. MR 2003e:81055 Zbl 1039.81022

Received 28 Mar 2011. Accepted 28 Aug 2011.

ANNA DALL'ACQUA: anna.dallacqua@ovgu.de
Institut für Analysis und Numerik, Fakultät für Mathematik, Otto-von-Guericke Universität, Postfach 4120, D-39106 Magdeburg, Germany

SØREN FOURNAIS: fournais@imf.au.dk
Department of Mathematical Sciences, University of Aarhus, Ny Munkegade 118, DK-8000 Aarhus C, Denmark

and

CNRS and Laboratoire de Mathématiques d'Orsay, Université Paris-Sud, Orsay Cedex, F-91405, France

THOMAS ØSTERGAARD SØRENSEN: sorensen@math.lmu.de
Department of Mathematics, Imperial College London, Huxley Building, 180 Queen's Gate, London SW7 2AZ, United Kingdom
Current address: *Mathematisches Institut, Universität München, Theresienstraße 39, D-80333 Munich, Germany*

EDGARDO STOCKMEYER: stock@mathematik.uni-muenchen.de
Mathematisches Institut, Universität München, Theresienstraße 39, D-80333 Munich, Germany

ANALYSIS AND PDE
Vol. 5, No. 3, 2012
dx.doi.org/10.2140/apde.2012.5.693

SEMICLASSICAL TRACE FORMULAS AND HEAT EXPANSIONS

YVES COLIN DE VERDIÈRE

In a recent paper (*J. Phys. A* **43**:47 (2011), 474028), B. Helffer and R. Purice compute the second term of a semiclassical trace formula for a Schrödinger operator with magnetic field. We show how to recover their formula by using the methods developed by Riemannian geometers in the seventies for heat expansions.

Introduction

There is a strong similarity between the expansions of the heat kernel as worked out by people in Riemannian geometry in the seventies, starting with the famous "Can one hear the shape of a drum?" by Mark Kac [1966] and continuing with [Berger 1966; McKean and Singer 1967] (see also the books [Berger et al. 1971; Gilkey 1975]), and the so-called semiclassical trace formulas developed by people in semiclassical analysis, starting with [Helffer and Robert 1983]. In fact, this is not only a similarity, but, as we will prove, each of these expansions, even if they differ when expressed numerically for some example, can be deduced from the other one as formal expressions of the fields.

Let us look first at the *heat expansion* on a smooth closed Riemannian manifold of dimension d, (X, g), with the (negative) Laplacian Δ_g[1]. The heat kernel $e(t, x, y)$, with $t > 0$ and $x, y \in X$, is the Schwartz kernel of $\exp(t\Delta_g)$: the solution of the heat equation $u_t - \Delta_g u = 0$ with initial datum u_0 is given by

$$u(t, x) = \int_X e(t, x, y) u_0(y)\, |dy|_g.$$

The function $e(t, x, x)$ admits, as $t \to 0^+$, the following asymptotic expansion:

$$e(t, x, x) \sim (4\pi t)^{-d/2}\big(1 + a_1(x)t + \cdots + a_l(x)t^l + \cdots\big).$$

The a_l are given explicitly in [Gilkey 2004, p. 201] for $l \leq 3$, and are known for $l \leq 5$ [Avramidi 1990; Ven 1998]. See also the related works [Hitrik 2002; Hitrik and Polterovich 2003a; 2003b; Polterovich 2000]. They are universal polynomials in the components of the curvature tensor and its covariant derivatives. For example, $a_0 = 1$ and $a_1 = \tau_g/6$, where τ_g is the scalar curvature.

The previous asymptotic expansion gives the expansion of the trace by integration over X and has been used as an important tool in spectral geometry:

$$\operatorname{trace}(e^{t\Delta_g}) = \int_X e(t, x, x)\, |dx|_g = \sum_{k=1}^{\infty} e^{\lambda_k t},$$

MSC2010: 35P20, 35S05, 58J35.
Keywords: trace formula, magnetic fields, heat expansion, synchronous gauge.

[1] In this note, we will not follow the usual sign convention of geometers, but the convention of analysts

where $-\lambda_1 = 0 \le -\lambda_2 \le \cdots \le -\lambda_k \le \cdots$ is the sequence of eigenvalues of $-\Delta_g$ with the usual convention about multiplicities. If $d = 2$, this gives

$$\operatorname{trace}(e^{t\Delta_g}) = \frac{1}{4\pi t}\left(\operatorname{Area}(X) + \frac{2\pi\chi(X)}{6}t + O(t^2)\right),$$

where $\chi(X)$ is the Euler characteristic of X.

There is an extension of the previous expansion in the case of Laplace type operators on fiber bundles: the coefficients of the expansion are then polynomials in the covariant derivatives of the curvature of the metric and of the connection on the fiber bundle. The heat expansion can be reinterpreted as an expansion of the Schwartz kernel of $f(-\hbar^2\Delta_g)$ on the diagonal $x = y$ in powers of $\hbar$ with $f(u) = \exp(-u)$ and $t = \hbar^2$. This is a particular case of the semiclassical trace.

Let us describe the semiclassical setting in the flat case: $\widehat{H}_\hbar$ is a self-adjoint $\hbar$-pseudodifferential operator with Weyl symbol $H(x,\xi)$ in some open domain X in $\mathbb{R}^d$, or more generally on a Riemannian manifold. Let $f \in \mathcal{S}(\mathbb{R})$ and look at $f(\widehat{H}_\hbar)$. Under some suitable assumptions (ellipticity at infinity in ξ) on H, $f(\widehat{H}_\hbar)$ is a pseudodifferential operator whose Weyl symbol $f^\star(H)$ is a formal power series in $\hbar$, given, using the Moyal product denoted by $\star$, by the following formula (see [Gracia-Saz 2005] for explicit formulas and Section 4.2 therein for a proof; see also [Charles 2003]) at the point $z_0 \in T^\star X$:

$$f^\star(H)(z_0) = (2\pi\hbar)^{-d}\sum_{l=0}^{\infty}\frac{1}{l!}f^{(l)}(H(z_0))\,(H - H(z_0))^{\star l}\,(z_0). \tag{1}$$

From the previous formula, we see that the symbol of $f(\widehat{H}_\hbar)$ at the point z depends only of the Taylor expansions of H at the point z and of f at the point $H(z)$. Helffer and Purice [2010] have studied the case of the magnetic Schrödinger operator whose Weyl symbol is $H_{a,V}(x,\xi) = \sum_{j=1}^{d}(\xi_j - a_j(x))^2 + V(x)$ and show that the Schwartz kernel of $f(\widehat{H}_{\hbar,a,V})$ at the point (x,x) admits an asymptotic expansion of the form

$$[f(\widehat{H}_{\hbar,a,V})](x,x) = (2\pi\hbar)^{-d}\sum_{j=0}^{\infty}\hbar^{2j}\left(\sum_{l=0}^{k_j}\int_{\mathbb{R}^d} f^{(l)}(\|\xi\|^2 + V(x))\,Q^{a,V}_{j,l}(x,\xi)\,|d\xi|\right)$$

where the $Q^{a,V}_{j,l}(x,\xi)$ are polynomials in ξ calculated from the Taylor expansions of the magnetic field $B = da$ and V at the point x. The proof in [Helffer and Purice 2010] uses a pseudodifferential calculus adapted to the magnetic field.

We will give a simplified version of the expansion replacing the (non-unique) $Q^{a,V}_{j,l}(x,\xi)$ by functions $P^{B,V}_{j,l}(x)$ which are uniquely defined and are given by universal $O(d)$-invariant polynomials of the Taylor expansions of B and V at the point x. We present then two ways to compute the $P^{B,V}_{j,l}$:

- we can first use Weyl's invariant theory (see [Gilkey 2004]) in order to reduce the problem to the determination of a finite number of numerical coefficients; then simple examples, like harmonic oscillator and constant magnetic field, allow to determine (part of) these coefficients.

- The $P_{j,l}^{B,V}$ are related in a very simple way to the coefficients of the heat expansion; it is possible to compute the $P_{j,l}^{B,V}$ from the knowledge of the a_l for $j+1 \le l \le 3j$. This is enough to recompute the coefficient of $\hbar^2$ and also, in principle, the coefficients of $\hbar^4$ in the expansion, because the a_l are known up to $l = 6$ in the case of a flat metric (see [Ven 1998]).

In this note, we will first describe precisely the semiclassical expansion for Schrödinger operators (in the case of an Euclidean metric) and the properties of the functions $P_{j,l}^{B,V}(x)$. Then, we will show how to compute the $P_{j,l}^{B,V}(x)$ using an adaptation of the method used for the heat kernel (Weyl's theorem on invariants and explicit examples). Finally, we will explain how the a_l are related to the $P_{j,l}^{B,V}(x)$. This gives us two proofs of the main formula given in [Helffer and Purice 2010]; this paper was the initial motivation to this work.

1. Semiclassical trace for Schrödinger operators

In what follows, X is an open domain in $\mathbb{R}^d$, equipped with the canonical Euclidean metric, and $\Omega^k(X)$ will denote the space of smooth exterior differential forms in X. Let us give a Schrödinger operator, with a smooth magnetic field $B = \sum_{1\le i<j\le d} b_{ij}\, dx_i \wedge dx_j$ (a closed real 2-form) and a smooth electric potential V (a real-valued smooth function) in X. We assume that V is bounded from below. We will assume also that the 2-form B is exact and can be written $B = da$ and we introduce the Schrödinger operator defined by

$$H_{\hbar,a,V} = \sum_{j=1}^{d} \left(\frac{\hbar}{i} \frac{\partial}{\partial x_j} - a_j(x) \right)^2 + V(x).$$

The Weyl symbol of $H_{\hbar,a,V}$ is $H_{a,V}(x,\xi) = \|\xi - a(x)\|^2 + V(x)$. We denote by $\widehat{H}_{\hbar,a,V}$ a self-adjoint extension of $H_{\hbar,a,V}$ in $L^2(X, |dx|)$. Let us give $f \in \mathcal{S}(\mathbb{R})$ and $\phi \in C_o^\infty(X)$ and consider the trace of $\phi f(\widehat{H}_{\hbar,a,V})$ as a distribution on $X \times \mathbb{R}$ (the density of states):

$$\mathrm{Trace}(\phi f(\widehat{H}_{\hbar,a,V})) = \int_X Z_{\hbar,a,V}(g)(x)\phi(x)\,|dx|,$$

where $Z_{\hbar,a,V}(g)(x)$ is the value at the point (x,x) of the Schwartz kernel of $f(\widehat{H}_{\hbar,a,V})$.

Theorem 1. *We have the following asymptotic expansion in powers of $\hbar$:*

$$Z_{\hbar,a,V}(g)(x) \sim (2\pi\hbar)^{-d} \left[\int_{\mathbb{R}^d} f(\|\xi\|^2 + V(x))\,|d\xi| + \sum_{j=1}^{\infty} \hbar^{2j} \left(\sum_{l=j+1}^{l=3j} P_{j,l}^{B,V}(x) \int_{\mathbb{R}^d} f^{(l)}(\|\xi\|^2 + V(x))\,|d\xi| \right) \right].$$

We have the explicit formulas

$$P_{1,2}^{B,V} = -\tfrac{1}{6}\big(\Delta V + \|B\|^2\big), \qquad P_{1,3}^{B,V} = -\tfrac{1}{12}\|\nabla V\|^2,$$
$$P_{2,3}^{B,V} = -\tfrac{1}{180}\big(8\|\nabla B\|^2 + \|d^\star B\|^2 + 12\langle \Delta B | B\rangle + 3\Delta^2 V\big).$$

Here $\|B\|^2 = \sum_{1\le i<j\le d} b_{ij}^2$, $d^\star : \Omega^2(X) \to \Omega^1(X)$ *is the formal adjoint of* d *used in the definition of the Hodge Laplacian on exterior forms. If* $d = 3$, $\|B\|$ *is the Euclidean norm of the vector field associated to* B.

The $P_{j,l}^{B,V}(x)$ *are polynomials of the derivatives of* B *and* V *at the point* x. *Moreover, if* λ, μ, c *are constants and we define* $\lambda^\star(f)(x) = f(\lambda x)$, *we have the following scaling properties*:

(1) $P_{j,l}^{\lambda.\lambda^\star(B),\lambda^\star(V)}(x) = \lambda^{2j} P_{j,l}^{B,V}(\lambda x)$. *This will be used with* $x = 0$.

(2) $P_{j,l}^{\mu B,\mu^2 V}(x) = \mu^{2(l-j)} P_{j,l}^{B,V}(x)$.

(3) $P_{j,l}^{B,V+c}(x) = P_{j,l}^{B,V}(x)$.

(4) $P_{j,l}^{-B,V}(x) = P_{j,l}^{B,V}(x)$.

(5) *The* $P_{j,l}^{B,V}$ *are invariant by the natural action of the orthogonal group* $O(d)$ *on the Taylor expansions of* B *and* V *at the point* x.

Remark 1. From the statement of the theorem, we see that the expansion of the density of states is independent of the chosen self-adjoint extension.

As a consequence, we can get the following full trace expansion under some more assumptions:

Corollary 1. *Let us assume that* $E_0 = \inf V < E_\infty = \liminf_{x\to\partial X} V(x)$ *and that we have chosen the Dirichlet boundary conditions. Let* $f \in C_o^\infty(]-\infty, E_\infty[)$, *then the trace of* $f(\widehat{H}_{\hbar,a,V})$ *admits the asymptotic expansion*

$$\mathrm{Trace}(f(\widehat{H}_{\hbar,a,V})) \sim (2\pi\hbar)^{-d} \int_X \biggl(\int_{\mathbb{R}^d} f(\|\xi\|^2 + V(x))\,|d\xi| + \cdots \\ \cdots \sum_{j=1}^{\infty} \hbar^{2j} \sum_{l=j+1}^{l=3j} P_{j,l}^{B,V}(x) \int_{\mathbb{R}^d} f^{(l)}(\|\xi\|^2 + V(x))\,|d\xi| \biggr)\,|dx|.$$

The coefficient of $\hbar^2$ *can be written as*

$$-\frac{1}{12} \int_{X\times\mathbb{R}^d} f^{(2)}(\|\xi\|^2 + V(x))\bigl(\Delta V(x) + 2\|B(x)\|^2\bigr)\,|dx\,d\xi|.$$

The expansion follows from [Helffer and Robert 1983]. An integration by part in x gives

$$\int_X f^{(3)}(\|\xi\|^2 + V(x))\|\nabla V(x)\|^2\,|dx| = -\int_X f^{(2)}(\|\xi\|^2 + V(x))\Delta V(x)\,|dx|.$$

2. Existence of the $\hbar$-expansion of $Z_{\hbar,a,V}$

Using Theorem 2 in the Appendix, we can work in $\mathbb{R}^d$ with a and V compactly supported. The existence of the expansion is known in general from [Helffer and Robert 1983] and the calculus of the symbol of

$f(\widehat{H}_{\hbar,a,V})$. We get

$$\int_X Z_{\hbar,a,V}(f)(x)\phi(x)\,|dx| = (2\pi\hbar)^{-d}\sum_{j=0}^{\infty}\hbar^{2j}\sum_{l=0}^{k_j}\int \phi(x) f^{(l)}(H_{a,V}(x,\xi))Q_{j,l}(x,\xi)\,|dx\,d\xi|$$

where the $Q_{j,l}(x,\xi)$ are polynomials in the Taylor expansion of $H_{a,V}$ at the point (x,ξ). The previous expansion is valid for any (admissible) pseudodifferential operator. In the case of Schrödinger operators we can make integrations by part in the integrals $\int f^{(l)}(H_{a,V}(x,\xi))Q_{j,l}(x,\xi)\,|d\xi|$ which reduces to a similar formula where we can replace the $Q_{j,l}(x,\xi)$ by the $P_{j,l}(x)$. This is based on the expansion of $Q_{j,l}$ as a polynomial in ξ in powers of $(\xi - a)$: odd powers give 0 and even powers can be reduced using

$$d_\xi\big((\xi_j - a_j) f^{(l)}(H_{a,V})\iota(\partial_{\xi_j})d\xi\big) = 2\|\xi_j - a_j\|^2 f^{(l+1)}(H_{a,V})d\xi + f^{(l)}(H_{a,V})d\xi.$$

We have only to check that the powers of ξ in $Q_{j,l}(x,\xi)$ are less than l: this is based on Equation (1). The coefficients of the l-th Moyal power of $H_{a,V}(z) - H_{a,V}(z_0)$ are homogeneous polynomials of degree l in the derivatives of $H_{a,V}(z)$. At the point $z = z_0$ only derivatives of order ≥ 1 are involved. They are all of degree ≤ 1 in ξ. Using gauge invariance at the point x (Section 3), we can assume that $a(x) = 0$.

3. Gauge invariance

If $S : X \to \mathbb{R}$ is a smooth function, we have

$$\operatorname{Trace}(\phi e^{-iS(x)/\hbar} f(\widehat{H}_{\hbar,a,V})e^{iS(x)/\hbar}) = \operatorname{Trace}(\phi f(\widehat{H}_{\hbar,a,V}))$$

and

$$e^{-iS(x)/\hbar} f(\widehat{H}_{\hbar,a,V})e^{iS(x)/\hbar} = f(\widehat{H}_{a+dS,V}).$$

Hence, we can chose any local gauge a in order to compute the expansion: using the synchronous gauge (see Section 4), we get the individual terms

$$\int f^{(l)}(H_{0,V})P_{j,l}^{B,V}(x)\,|d\xi|$$

for the expansion, where the $P_{j,l}^{B,V}(x)$ depend only of the Taylor expansions of B and V at the point x.

4. The synchronous gauge

The main idea is to find an appropriate gauge a adapted to the point x_0 where we want to make the symbolic computation. In a geometric language, we use the trivialization of the bundle by parallel transportation along the rays: the potential a vanishes on the radial vector field.[2] Here, this is simply the fact that, for any closed 2-form B on $\mathbb{R}^2$, there exists an unique 1-form $a = \sum_{j=1}^d a_j dx_j$ so that $da = B$ and $\sum_{j=1}^d x_j a_j = 0$.

[2]This gauge is sometimes called the Fock–Schwinger gauge; in [Atiyah et al. 1973], it is called the synchronous framing.

We will do that for the Taylor expansions degree by degree. In what follows we will use a decomposition for 1-forms, but it works also for k-forms.

Let us denote by Ω_N^k the finite dimensional vector space of k-differential forms on $\mathbb{R}^d$ whose coefficients are homogeneous polynomials of degree N and by $W = \sum_{j=1}^d x_j \, \partial/\partial x_j$ the radial vector field. The exterior differential induces a linear map from Ω_N^k into Ω_{N-1}^{k+1} and the inner product $\iota(W)$ a map from Ω_N^k into Ω_{N+1}^{k-1}. They define complexes which are exact except at $k = N = 0$. Moreover, we have a situation similar to Hodge theory:

$$\Omega_N^k = d\Omega_{N+1}^{k-1} \oplus \iota(W)\Omega_{N-1}^{k+1}.$$

This is due to Cartan's formula: the Lie derivative of a form $\omega \in \Omega_N^k$ satisfies, from the direct calculation, $\mathcal{L}_W \omega = (k+N)\omega$, and, by Cartan's formula, $\mathcal{L}_W \omega = d(\iota(W)\omega) + \iota(W)d\omega$. So

$$\omega = \frac{1}{k+N} \left(d(\iota(W)\omega) + \iota(W)d\omega \right).$$

It remains to show that this is a direct sum: if $\omega = d\alpha = \iota(W)\gamma$, we have $\iota(W)\omega = 0$ and $d\omega = 0$; from the previous decomposition, we see that $\omega = 0$. Let us denote by $J^N\omega$, where ω is a differential form of degree k, the form in Ω_N^k which appears in the Taylor expansion of ω.

We get:

Proposition 1. *If $P(J^0a, J^1a, \cdots, J^Na)$ is a polynomial in the Taylor expansion of the 1-form a at some order N which is invariant by $a \to a + dS$, P is independent of J^0a and*

$$P(J^1a, \cdots, J^Na) = P\Big(\frac{1}{2} J^1\iota(W)B, \cdots, \frac{1}{N+1} J^N\iota(W)B\Big)$$

is a polynomial of the Taylor expansion of B to the order $N-1$.

5. Properties of the $P_{j,l}$

5.1. ***Range of l for j fixed.*** From the scaling properties, we deduce that, in a monomial

$$D^{\alpha_1} B_{i_1,j_1} \cdots D^{\alpha_k} B_{i_k,j_k} D^{\beta_1} V \cdots D^{\beta_m} V,$$

belonging to $P_{j,l}$, we have $k + 2m = 2(l-j)$ and $k + |\alpha_1| + \cdots + |\alpha_k| + |\beta_1| + \cdots + |\beta_m| = 2j$. Moreover, for $j \geq 1$, $k + m \geq 1$ and $|\beta_p| \geq 1$. Hence $j + 1 \leq l \leq 3j$. The previous bounds are sharp: take the monomials $\Delta^j V$ and $\|\nabla V\|^{2j}$ which give $l = j+1$ and $l = 3j$.

5.2. ***Invariance properties.***

(1) Let us assume that we look at the point $x = 0$ and consider the operator $D_\mu(f)(x) = f(\mu x)$. We have

$$D_\mu \circ \widehat{H}_{\hbar,A,V} \circ D_{1/\mu} = \widehat{H}_{\hbar/\mu, A \circ D_\mu, V \circ D_\mu}.$$

The same relation is true for any function $f(\widehat{H}_{\hbar,A,V})$ and then we have, looking at the Schwartz kernels and using the Jacobian μ^d of D_μ:

$$P_{j,l}^{B,V}(0)\int_{\mathbb{R}^d} f^{(l)}(\|\xi\|^2+V(0))\,|d\xi| = \mu^{-2j}\,P_{j,l}^{\mu.\mu^\star B,\mu^\star V}(0)\int_{\mathbb{R}^d} f^{(l)}(\|\xi\|^2+V(0))\,|d\xi|.$$

(2) We have

$$\widehat{H}_{\hbar,\mu a,\mu^2 V} = \mu^2\,\widehat{H}_{\frac{\hbar}{\mu},a,V}.$$

(3) Changing V into $V+c$ gives a translation by c in the function f but does not change the $P_{j,l}^{B,V}$.

(4) Changing B into $-B$ gives a complex conjugation in the computations. The final result is real-valued.

(5) Orthogonal invariance is clear: an orthogonal change of coordinates around the point x preserves the density of states.

5.3. ***The case $d = 2$.*** We deduce from the scaling properties and invariance by the orthogonal group, that there exists constants $a_d,\ b_d,\ c_d$ so that $P_{1,2}^{B,V}(x) = a_d\Delta V + b_d\|B\|^2,\ P_{1,3}(x) = c_d\|\nabla V\|^2$.

6. Explicit examples

The calculation for the harmonic oscillators and the constant magnetic fields allows to determine the constants $a_d,\ b_d,\ c_d$.

6.1. ***Harmonic oscillators.*** Let us consider $\Omega = -\hbar^2\dfrac{d^2}{dx^2}+x^2$ with $d=1$. The kernel of $P(t,x,y)$ of $\exp(-t\Omega)$ is given by the Mehler formula:

$$P(t,x,y) = (2\pi\hbar\sinh(2t\hbar))^{-\frac{1}{2}}\exp\left(-\frac{1}{2\hbar\sinh(2t\hbar)}\big(\cosh(2t\hbar)(x^2+y^2)-2xy\big)\right).$$

Hence

$$P(t,x,x)\sim(2\pi\hbar)^{-1}e^{-tx^2}\left(\int_{\mathbb{R}} e^{-t\xi^2}\,d\xi\right)\big(1-\hbar^2(t^2-t^3x^2)/3+O(\hbar^4)\big).$$

Hence $P_{1,2}(x) = -V''(x)/6$ and $P_{1,3}(x) = -V'(x)^2/12$.

Similarly, in dimension $d>1$, we get $P_{1,2}(x) = -\Delta V(x)/6$ and $P_{1,3}(x) = -\|\nabla V\|^2/12$.

6.2. ***Constant magnetic field.*** Let us consider the case of a constant magnetic field B in the plane and denote by $Q(t,x,y)$ the kernel of $\exp(-tH_{B,0})$. We have (see [Avron et al. 1978])

$$Q(t,x,x) = \frac{B}{4\pi\hbar\sinh Bt\hbar}.$$

Hence the asymptotic expansion

$$Q(t,x,x) = (2\pi\hbar)^{-2}\int\exp(-t\|\xi\|^2)\,|d\xi|\big(1-t^2\hbar^2B^2/6+O(\hbar^4)\big);$$

hence $P_{1,2}(x) = -B^2/6$ and $P_{1,3}(x) = 0$.

Using the normal form $B = b_{12}dx_1 \wedge dx_2 + b_{34}dx_3 \wedge dx_4 + \cdots$, we get in dimension $d > 2$ the values $P_{1,2}(x) = -\|B\|^2/6$ and $P_{1,3}(x) = 0$.

7. Heat expansion from the semiclassical expansions

We have $t\widehat{H}_{1,a,V} = \widehat{H}_{\sqrt{t},\sqrt{t}a,tV}$. Using the expansion of Theorem 1 with $f(E) = e^{-E}$, we get easily the point-wise expansion of the heat kernel on the diagonal as $t \to 0^+$:

$$[\exp(-t\widehat{H}_{1,a,V})](x,x) \sim \frac{1}{(4\pi t)^{d/2}} e^{-tV(x)} \sum_{l=0}^{\infty} \Bigg(\sum_{l/3 \le j \le l-1} P_{j,l}^{B,V}(x) \Bigg) (-t)^l.$$

In particular, $a_1(x) = -V(x)$ and the coefficient $a_2(x)$ is given by

$$a_2(x) = \tfrac{1}{2}V(x)^2 - \tfrac{1}{6}\Delta V(x) - \tfrac{1}{6}\|B(x)\|^2.$$

This formula agrees with Equation (3) of Theorem 3.3.1 in [Gilkey 2004].

This gives another way to compute the $P_{j,l}$: if, as power series in t,

$$\sum_{l=0}^{\infty} (-1)^l b_l(x) t^l = e^{tV(x)} \sum_{l=0}^{\infty} a_l(x) t^l,$$

we have

$$\sum_{l/3 \le j \le l-1} P_{j,l}^{B,V}(x) = b_l(x).$$

$P_{j,l}^{B,V}$ is the sum of monomials homogeneous of degree $2(l-j)$ in b_l where B and its derivatives have weights 1 while V and its derivatives have weights 2.

The heat coefficients a_l on flat spaces are known for $l \le 6$ from [Ven 1998]. This is enough to check the term in $\hbar^2$ (uses a_2 and a_3) in [Helffer and Purice 2010] and to compute the term in $\hbar^4$ in the semiclassical expansion (uses the a_l for $3 \le l \le 6$).

We have also a mixed expansion writing $t\widehat{H}_{\hbar,a,V} = \widehat{H}_{\sqrt{t}\hbar,\sqrt{t}a,tV}$, we get a power series expansion in powers of $\hbar$ and t valid in the domain $\hbar^2 t \to 0$ and $0 < t \le t_0$ for the point-wise trace of $\exp(-t\widehat{H}_{\hbar,a,V})$:

$$Z_{t,\hbar}(x) \sim \frac{1}{(4\pi t)^{d/2}} e^{-tV(x)} \Bigg(1 + \sum_{\substack{j \ge 1 \\ j+1 \le l \le 3j}} \hbar^{2j} (-t)^l P_{j,l}^{B,V}(x) \Bigg).$$

This shows that the integrals $\int_X V(x)^k |dx|$ and $\int_X P_{j,l}^{B,V}(x)\,|dx|$ are recoverable from the semiclassical spectrum.

Appendix: functional calculus in domains and self-adjoint extensions (after Johannes Sjöstrand)

The content of this Appendix is due to Johannes Sjöstrand. I thank him very much for this contribution.

Let $X \subset \mathbb{R}^d$ be an open set. We say that a linear operator A is a ΨDO in X, with Weyl symbol a if, for any compact $K \subset X$, A acts on functions supported in K as a ΨDO of Weyl symbol a.

Theorem 2. *Let $H_{\hbar,a,V}$ be a Schrödinger operator with magnetic field given by*

$$H_{\hbar,a,V}=\sum_{j=1}^{d}\left(\frac{\hbar}{i}\frac{\partial}{\partial x_j}-a_j(x)\right)^2+V(x),$$

defined in some open domain $X\subset\mathbb{R}^d$. We assume that a and V are smooth in X and that V is bounded from below, so that $H_{\hbar,a,V}$ admits some self-adjoint extensions on the Hilbert space $L^2(X,|dx|)$. One of them will be denoted by $\widehat{H}_{\hbar,a,V}$. Then, for any $f\in\mathcal{S}(\mathbb{R})$, $f(\widehat{H}_{\hbar,a,V})$, given by the functional calculus, is a semiclassical ΨDO in X whose symbol is given by Equation (1) *and is independent of the chosen extension.*

The proof uses a multicommutator method already used by Helffer and Sjöstrand [1984].

Proof. We introduce, for $s\in\mathbb{R}$, the semiclassical ($\hbar$-dependent) Sobolev spaces

$$\mathcal{H}_\hbar^s:=\{u\in\mathcal{S}'(\mathbb{R}^d)\mid \|\mathrm{Op}_\hbar(1+\|\xi\|^2)^{s/2}u\|_{L^2}<\infty\}$$

with the norm

$$\|u\|_s:=\|\mathrm{Op}_\hbar(1+\|\xi\|^2)^{s/2}u\|_{L^2}.$$

The ($\hbar$-dependent) norm $\|A\|_{s_1,s_2}$ is the norm of A as linear operator from $\mathcal{H}_\hbar^{s_1}$ to $\mathcal{H}_\hbar^{s_2}$. A linear operator K is smoothing if, for all $s_1,\ s_2$, $\|K\|_{s_1,s_2}=O(\hbar^\infty)$. This implies that the Schwartz kernel of K is smooth with all derivatives locally $O(\hbar^\infty)$. We have the

Lemma 1. *Let Y be an open set in $\mathbb{R}^d$. Let $P_j=P_j(\hbar)$, $j=0,1$ be two self-adjoint operators on Hilbert spaces $\mathcal{H}_j=L^2(X_j,|dx|)$ with $Y\Subset X_0\subseteq X_1\subseteq\mathbb{R}^d$ and with domains $\mathcal{D}_j$ so that $C_o^\infty(Y)\subset\mathcal{D}_j\subset\mathcal{H}_j$. Let us assume that, on $C_o^\infty(Y)$, $P_0=P_1=H_{\hbar,a,V}(=P)$.*

Then, for any $f\in C_o^\infty(\mathbb{R})$, $f(P_0)-f(P_1)$ is smoothing on Y. In particular, the densities of states $[f(P_j)](x,x)$, $j=0,1$, coincide in Y modulo $O(\hbar^\infty)$.

Assuming Lemma 1, Theorem 2 follows by extending a and V smoothly outside Y so that they have compact support in $\mathbb{R}^d$. We take $Y\Subset X=X_0\subset\mathbb{R}^d=X_1$. It follows that P_1 is essentially self-adjoint and the functional calculus for P_1 follows then easily from [Helffer and Robert 1983]. The result is valid even for $f\in\mathcal{S}(\mathbb{R})$ because C_o^∞ is dense in $\mathcal{S}$ and the result of [Helffer and Robert 1983] is valid for $f\in\mathcal{S}$ and the resulting formulas for the symbols are continuous w.r. to the topology of $\mathcal{S}$. □

Proof. Proof of Lemma 1 If $\chi\in C_o^\infty(Y)$, then, for $z\notin\mathbb{R}$ and $j,k\in\{0,1\}$, we have on $L^2(Y)$:

$$(P_j-z)^{-1}\circ\chi=\chi\circ(P_k-z)^{-1}-(P_j-z)^{-1}[P,\chi](P_k-z)^{-1} \tag{2}$$

Let $\chi_0\le\chi_1\le\cdots\le\chi_N$ with, for $l=0,\cdots,N$, $\chi_l\in C_o^\infty(Y)$ and, for $l=0,\cdots,N-1$, $\chi_l(1-\chi_{l+1})\equiv 0$. By iterating (2) and using $\chi_{l+1}[P,\chi_l]=[P,\chi_l]$, we find:

$$\begin{aligned}(P_1-z)^{-1}\circ\chi_0&=\chi_1\circ(P_0-z)^{-1}\chi_0-\chi_2\circ(P_0-z)^{-1}[P,\chi_1](P_0-z)^{-1}\chi_0+\cdots\\&\quad\pm\chi_N(P_0-z)^{-1}[P,\chi_{N-1}](P_0-z)^{-1}[P,\chi_{N-2}]\cdots(P_0-z)^{-1}\chi_0\\&\quad\mp(P_1-z)^{-1}[P,\chi_N](P_0-z)^{-1}\cdots(P_0-z)^{-1}\chi_0\end{aligned}$$

Let us give now $\chi_0,\ \psi \in C_o^\infty(Y)$ with disjoints supports. By choosing the χ_l for $l > 0$ with supports disjoint from the support of ψ, we see, using Equation (2), that, for any N,

$$\|\psi(P_1 - z)^{-1}\chi_0\|_{0,2} = O\big(\hbar^N |\Im z|^{-(N+1)}\big).$$

The standard a priori elliptic estimates

$$\|u\|_{s+2,\Omega_1} \le C\left(\|(P - z)u\|_{s,\Omega_2} + \|u\|_{s,\Omega_1}\right)$$

for $z \in K \Subset \mathbb{C}$ and $\Omega_1 \Subset \Omega_2 \Subset \mathbb{R}^d$, allow to prove that, for any N, s, there exists $M(N, s)$ so that

$$\|\psi(P_1 - z)^{-1}\chi_0\|_{s,s+N+2} = O(\hbar^N |\Im z|^{-M(N,s)}) \tag{3}$$

Let $\chi \in C_o^\infty(Y)$ so that $\chi \equiv 1$ on the support of χ_0. Let us apply multiplication by χ_0 to the right and to the left in (2) and choose ψ with support disjoint from χ_0 so that $[P, \chi]\psi = [P, \chi]$. Inserting ψ this way in (2), we get, using (3),

$$\chi_0(P_1 - z)^{-1}\chi_0 - \chi_0(P_0 - z)^{-1}\chi_0 = K,$$

and, for any N, there exists $M(N)$ so that $\|K\|_{-N,N} = O(\hbar^N \Im z^{-M(N)})$. We now apply the formula (known to some people as the "Helffer–Sjöstrand formula", proved for example in [Dimassi and Sjöstrand 1999, p. 94–95]), valid for $f \in C_o^\infty(\mathbb{R})$ and $\tilde{f}$ an almost holomorphic extension of f:

$$f(P_j) = \frac{1}{\pi}\int_{\mathbb{C}} \partial_{\bar{z}}\tilde{f}(z)(P_j - z)^{-1}\,dL(z),$$

where $dL(z)$ is the canonical Lebesgue measure in the complex plane. From this, we see that $f(P_0) - f(P_1)$ is smoothing in Y. □

Acknowledgments

I thank the referee for his careful reading of the paper which forced me to make the statements and proofs more precise. Many thanks also to Johannes Sjöstrand for his help in clarifying some points of semiclassical analysis and for allowing me to present them in the Appendix to this paper.

References

[Atiyah et al. 1973] M. Atiyah, R. Bott, and V. K. Patodi, "On the heat equation and the index theorem", *Invent. Math.* **19** (1973), 279–330. MR 58 #31287 Zbl 0257.58008

[Avramidi 1990] I. G. Avramidi, "The covariant technique for the calculation of the heat kernel asymptotic expansion", *Phys. Lett. B* **238**:1 (1990), 92–97. MR 91f:58092

[Avron et al. 1978] J. Avron, I. Herbst, and B. Simon, "Schrödinger operators with magnetic fields, I: General interactions", *Duke Math. J.* **45**:4 (1978), 847–883. MR 80k:35054 Zbl 0399.35029

[Berger 1966] M. Berger, "Sur le spectre d'une variété Riemannienne", *C. R. Acad. Sci. Paris Sér. A* **263** (1966), 13–16. MR 34 #1964 Zbl 0141.38203

[Berger et al. 1971] M. Berger, P. Gauduchon, and E. Mazet, *Le spectre d'une variété Riemannienne*, Lecture Notes in Mathematics **194**, Springer, Berlin, 1971. MR 43 #8025 Zbl 0223.53034

[Charles 2003] L. Charles, "Berezin–Toeplitz operators: a semi-classical approach", *Comm. Math. Phys.* **239**:1-2 (2003), 1–28. MR 2004m:53158 Zbl 1059.47030

[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, Cambridge, 1999. MR 2001b:35237 Zbl 0926.35002

[Gilkey 1975] P. B. Gilkey, "The spectral geometry of a Riemannian manifold", *J. Differential Geom.* **10**:4 (1975), 601–618. MR 53 #4150 Zbl 0316.53035

[Gilkey 2004] P. B. Gilkey, *Asymptotic formulae in spectral geometry*, CRC, Boca Raton, FL, 2004. MR 2005f:58041 Zbl 1080.58023

[Gracia-Saz 2005] A. Gracia-Saz, "The symbol of a function of a pseudo-differential operator", *Ann. Inst. Fourier* (*Grenoble*) **55**:7 (2005), 2257–2284. MR 2006m:53142 Zbl 1091.53062

[Helffer and Purice 2010] B. Helffer and R. Purice, "Magnetic calculus and semiclassical trace formulas", *J. Phys. A* **43**:47 (2010), 474028. MR 2011m:81089 Zbl 1204.81066 arXiv 1005.1795v1

[Helffer and Robert 1983] B. Helffer and D. Robert, "Calcul fonctionnel par la transformation de Mellin et opérateurs admissibles", *J. Funct. Anal.* **53**:3 (1983), 246–268. MR 85i:47052 Zbl 0524.35103

[Helffer and Sjöstrand 1984] B. Helffer and J. Sjöstrand, "Multiple wells in the semiclassical limit, I", *Comm. Partial Differential Equations* **9**:4 (1984), 337–408. MR 86c:35113 Zbl 0546.35053

[Hitrik 2002] M. Hitrik, "Existence of resonances in magnetic scattering", *J. Comput. Appl. Math.* **148**:1 (2002), 91–97. MR 2003m:81063 Zbl 1010.81084

[Hitrik and Polterovich 2003a] M. Hitrik and I. Polterovich, "Regularized traces and Taylor expansions for the heat semigroup", *J. London Math. Soc.* (2) **68**:2 (2003), 402–418. MR 2004h:35045 Zbl 1167.35335

[Hitrik and Polterovich 2003b] M. Hitrik and I. Polterovich, "Resolvent expansions and trace regularizations for Schrödinger operators", pp. 161–173 in *Advances in differential equations and mathematical physics* (Birmingham, AL, 2002), edited by Y. Karpeshina et al., Contemp. Math. **327**, American Mathematical Society, Providence, RI, 2003. MR 2004g:35056 Zbl 1109.35030

[Kac 1966] M. Kac, "Can one hear the shape of a drum?", *Amer. Math. Monthly* **73**:4, part II (1966), 1–23. MR 34 #1121 Zbl 0139.05603

[McKean and Singer 1967] H. P. McKean, Jr. and I. M. Singer, "Curvature and the eigenvalues of the Laplacian", *J. Differential Geom.* **1**:1 (1967), 43–69. MR 36 #828 Zbl 0198.44301

[Polterovich 2000] I. Polterovich, "Heat invariants of Riemannian manifolds", *Israel J. Math.* **119** (2000), 239–252. MR 2002b:58031 Zbl 0996.58019

[Ven 1998] A. E. M. van de Ven, "Index-free heat kernel coefficients", *Classical Quantum Gravity* **15**:8 (1998), 2311–2344. MR 2000a:58072 Zbl 0937.58018

Received 18 Apr 2011. Revised 11 Jun 2011. Accepted 11 Jul 2011.

YVES COLIN DE VERDIÈRE: yves.colin-de-verdiere@ujf-grenoble.fr
Unité mixte de recherche CNRS-UJF 5582, Institut Fourier, BP 74, 38402-Saint Martin d'Hères Cedex, France

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at msp.berkeley.edu/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 5 No. 3 2012

2157-5045(2012)5:3;1-H