ANALYSIS & PDE

Volume 10 No. 8 2017

msp

Analysis & PDE

msp.org/apde

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

The subscription price for 2017 is US $265/year for the electronic version, and $470/year (+$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscriber address should be sent to MSP.

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

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ANALYSIS AND PDE
Vol. 10, No. 8, 2017
dx.doi.org/10.2140/apde.2017.10.1793

KOSZUL COMPLEXES, BIRKHOFF NORMAL FORM AND THE MAGNETIC DIRAC OPERATOR

NIKHIL SAVALE

We consider the semiclassical Dirac operator coupled to a magnetic potential on a large class of manifolds, including all metric contact manifolds. We prove a sharp Weyl law and a bound on its eta invariant. In the absence of a Fourier integral parametrix, the method relies on the use of almost analytic continuations combined with the Birkhoff normal form and local index theory.

1. Introduction

Semiclassical analysis concerns the study of the spectrum of (h-)pseudodifferential operators $P_h : C^\infty(X) \to C^\infty(X)$, $h \in (0, 1]$, in the limit $h \to 0$ and is now the subject of several texts [Dimassi and Sjöstrand 1999; Guillemin and Sternberg 2013; Ivrii 1998; 2017; Maslov and Fedoriuk 1981; Robert 1987; Zworski 2012]. Standard examples of such operators include the Schrödinger operator $P_h = -h^2\Delta_X + V$ on a compact n-dimensional Riemannian manifold X with potential $V \in C^\infty(X)$. The clearest asymptotic result is given by the celebrated Weyl law, see for example [Dimassi and Sjöstrand 1999, Chapter 10], on the asymptotic number of eigenvalues $N[a, b]$ in a fixed interval $[a, b]$. A related result is on the number of eigenvalues $N(-ch, ch)$ of P_h in the finer interval $(-ch, ch)$: assuming 0 is not a critical value of the symbol $\sigma(P) = p(x, \xi) \in C^\infty(T^*X)$, one has

$$N(-ch, ch) = O(h^{-n+1}) \tag{1-1}$$

as $h \to 0$, for all $c > 0$. Similar results also exist in the case where 0 is a Morse–Bott critical level for the symbol; see [Brummelhuis et al. 1995]. In the critical case, the exponent in the Weyl law may drop depending on the codimension of zero energy level $\Sigma_0^P := \{p(x, \xi) = 0\}$ and the signature of the normal Hessian. The Weyl laws thus obtained are sharp and are proved using a parametrix construction for the evolution operator $e^{\frac{it}{h}P_h}$ as a Fourier integral operator.

In the context of nonscalar operators $P_h : C^\infty(X; E) \to C^\infty(X; E)$ acting on sections of a vector bundle E, fewer result are known. The simplest case is when the nonscalar symbol $p(x, \xi) \in C^\infty(T^*X; E)$ is smoothly diagonalizable near the zero energy level $\Sigma_0^P = \{\det(p(x, \xi)) = 0\}$. In this case, similar Fourier integral methods apply; see [Emmrich and Weinstein 1996; Maslov and Fedoriuk 1981] or [Guillemin; Sandoval 1999] for an exposition in the microlocal/classical setting. For nonscalar operators

MSC2010: primary 35P20, 81Q20; secondary 58J40, 58J28.
Keywords: Dirac operator, Weyl law, eta invariant.

another method is provided under the microhyperbolicity condition of Ivrii [1998, Chapters 2 and 3]; see also [Dimassi and Sjöstrand 1999, Chapter 12]. In this paper, we study the particular case of the magnetic Dirac operator where neither diagonalizability nor the microhyperbolicity condition is satisfied.

More precisely, let (X, g^{TX}) be an oriented Riemannian manifold of odd dimension $n = 2m + 1$ equipped with a spin structure. Let S be the corresponding spin bundle and let L be an auxiliary Hermitian line bundle. Fix a unitary connection A_0 on L and let $a \in \Omega^1(X; \mathbb{R})$ be a one-form. This gives a family of unitary connections on L via $\nabla^h = A_0 + \frac{i}{h}a$ and a corresponding family of coupled magnetic Dirac operators

$$D_h := hD_{A_0} + ic(a) \tag{1-2}$$

for $h \in (0, 1]$ and where c stands for the Clifford multiplication endomorphism (see Section 2B).

In order to derive sharp spectral asymptotics, we shall make a couple of restrictive assumptions on the one-form a and the metric g^{TX}. First, the one-form a will be assumed to be a contact one-form (i.e., one satisfying $a \wedge (da)^m > 0$). This gives rise to the contact hyperplane $H = \ker(a) \subset TX$ as well as the Reeb vector field R defined via $i_R da = 0,\ i_R a = 1$.

To state the assumption on the metric, consider the contracted endomorphism $\mathfrak{J} : T_xX \to T_xX$ defined at each point $x \in X$ via

$$da(v_1, v_2) = g^{TX}(v_1, \mathfrak{J}v_2) \quad \forall v_1, v_2 \in T_xX.$$

From the contact assumption, $\mathfrak{J}$ has a one-dimensional kernel spanned by the Reeb vector field R. The endomorphism $\mathfrak{J}$ is clearly antisymmetric with respect to the metric

$$g^{TX}(v_1, \mathfrak{J}v_2) = -g^{TX}(\mathfrak{J}v_1, v_2),$$

and hence its nonzero eigenvalues come in purely imaginary pairs $\pm i\mu,\ \mu > 0$. The assumption on the metric g^{TX} is then as follows.

Definition 1.1. We say that the metric g^{TX} is *suitable* to the contact form a if there exist positive constants $0 < \mu_1 \le \mu_2 \le \cdots \le \mu_m$ (independent of $x \in X$) and a positive real function $\nu(x) > 0$ such that

$$\operatorname{Spec}(\mathfrak{J}_x) = \{0,\ \pm i\mu_1\nu(x),\ \pm i\mu_2\nu(x),\ \ldots,\ \pm i\mu_m\nu(x)\} \tag{1-3}$$

for all $x \in X$.

Before proceeding further, we give two examples of suitable metrics:

(1) In the case that the dimension of the manifold X is 3, any metric g^{TX} is suitable, as $\operatorname{Spec}(\mathfrak{J}_x) = \{0, \pm i|da|\}$ has only two nonzero eigenvalues.

(2) There is a smooth endomorphism $J : TX \to TX$ such that (X^{2m+1}, a, g^{TX}, J) is a metric contact manifold. That is, we have

$$\begin{aligned} J^2 v_1 &= -v_1 + a(v_1)R, \\ g^{TX}(v_1, Jv_2) &= da(v_1, v_2) \quad \forall v_1, v_2 \in T_xX. \end{aligned} \tag{1-4}$$

In this case the nonzero eigenvalues of $\mathfrak{J}_x = J_x$ are $\pm i$ (each with multiplicity m). For any given contact form a there exists an infinite-dimensional space of (g^{TX}, J) satisfying (1-4). This case in particular includes all strictly pseudoconvex CR manifolds.

In addition to the Weyl law we shall also be interested in the asymptotics of the eta invariant $\eta_h = \eta(D_h)$ of the Dirac operator, formally its signature (see Section 2A for a definition). The main result is now stated as follows.

Theorem 1.2. *Under the contact and suitability assumptions on a and g^{TX}, the Weyl counting function and eta invariant of D_h satisfy the sharp asymptotics*

$$N(-ch, ch) = O(h^{-m}), \tag{1-5}$$

$$\eta_h = O(h^{-m}) \tag{1-6}$$

as $h \to 0$.

We note that the exponents above are significantly lower than (1-1). This is again partly attributed to the high codimension of the zero energy level Σ_0^D. In this case $\Sigma_0^D = \{\xi = -a\} \subset T^*X$ is the graph of the contact form a, a submanifold of half-dimension $2m+1$ on which the canonical symplectic form is maximally nondegenerate of rank $2m$.

The proof of the asymptotic result Theorem 1.2 above will be based on a functional trace expansion. To state the trace expansion involved, set $\nu_0 := \mu_1[\min_{x\in X} \nu(x)]$ and choose $f \in C_c^\infty(-\sqrt{2\nu_0}, \sqrt{2\nu_0})$. Pick real numbers $0 < T' < T$ and let $\theta \in C_c^\infty((-T, T); [0, 1])$ such that $\theta(x) = 1$ on $(-T', T')$. Let

$$\mathcal{F}^{-1}\theta(x) := \check{\theta}(x) = \frac{1}{2\pi}\int e^{ix\xi}\theta(\xi)\, d\xi,$$

$$\mathcal{F}_h^{-1}\theta(x) := \frac{1}{h}\check{\theta}\left(\frac{x}{h}\right) = \frac{1}{2\pi h}\int e^{\frac{i}{h}x\xi}\theta(\xi)\, d\xi$$

be its classical and semiclassical inverse Fourier transforms respectively. We now have the following functional trace expansion for the magnetic Dirac operator $D = D_h$ given in (1-2).

Theorem 1.3. *Let a be a contact form, g^{TX} be a suitable metric and f be as above. There exist smooth functions $u_j \in C^\infty(\mathbb{R})$ such that there is a trace expansion*

$$\begin{aligned}\operatorname{tr}\left[f\left(\frac{D}{\sqrt{h}}\right)(\mathcal{F}_h^{-1}\theta)(\lambda\sqrt{h} - D)\right] &= \operatorname{tr}\left[f\left(\frac{D}{\sqrt{h}}\right)\frac{1}{h}\check{\theta}\left(\frac{\lambda\sqrt{h} - D}{h}\right)\right] \\ &= h^{-m-1}\left(f(\lambda)\sum_{j=0}^{N-1} u_j(\lambda)h^{\frac{j}{2}} + O(h^{\frac{N}{2}})\right)\end{aligned} \tag{1-7}$$

for T sufficiently small and for each $N \in \mathbb{N}$, $\lambda \in \mathbb{R}$.

Again, the trace (1-7) should be compared with the wave trace expansions for scalar and microhyperbolic operators [Dimassi and Sjöstrand 1999, Chapters 10 and 12], although a different scale of size $\sqrt{h}$ is being used. In the absence of a Fourier integral parametrix or microhyperbolicity our strategy is to combine the use of almost analytic continuations with local index theory expansions. We first show that the trace is $O(h^\infty)$ in the region $\operatorname{spt}(\theta) \subset \{T > |x| \geq h^\varepsilon\}$, $\varepsilon \in \left(0, \frac{1}{2}\right)$ (see Lemma 3.1). Here the lack of microhyperbolicity for the symbol poses a difficulty in the use of almost analytic continuations [Dimassi and Sjöstrand 1999, Chapter 12]; see also [Dimassi and Sjöstrand 1996]. We however show that this can be overcome with a closer understanding of the total symbol of D via its Birkhoff normal form. It is in deriving the Birkhoff normal form that Koszul complexes are used and the assumptions on a, g^{TX} are required. The local index theory method [Bismut 1987; Ma and Marinescu 2007] finally provides the expansion in the region $\operatorname{spt}(\theta) \subset \{|x| < h^\varepsilon\}$ (see Lemma 3.2).

There is a large recent literature for semiclassical problems in the presence of magnetic fields. In particular the extensive book of Ivrii [2017] specifically considers the case of the magnetic Dirac operator in Chapter 17. The Birkhoff normal form here (5-13) generalizes Proposition 17.2.1 therein. Our use of normal forms should also be compared to their use in scalar cases from [Charles and Vũ Ngọc 2008; Helffer et al. 2016; Raymond and Vũ Ngọc 2015]. We note that some of the spectral literature on Dirac operators treats the massive case (e.g., mass $m = 1$ in [Helffer and Robert 1983]), where the mass term renders the symbol diagonalizable. The geometric Dirac operator considered here corresponds to the odd-dimensional purely massless case.

The asymptotic problem of the eta invariant (1-6) was earlier considered by the author in [Savale 2014], where a nonsharp estimate was proved, under no assumptions on a, g^{TX}, via the use of the heat trace. This asymptotic problem was first considered and applied in [Taubes 2007] in the proof of the three-dimensional Weinstein conjecture using Seiberg–Witten theory. The three-dimensional case has been further explored in [Tsai 2014].

The paper is organized as follows. In Section 2, we begin with preliminary notions used throughout the paper, including basic facts about Clifford representations, Dirac operators and the semiclassical calculus. In Section 2B1 we compute the spectrum of a model magnetic Dirac operator on $\mathbb{R}^m$ using Clifford representations and the harmonic oscillator. In Section 3 we perform certain reductions towards proving Theorem 1.3, including a time scale breakup of the trace into Lemmas 3.1 and 3.2. These reductions are then used in Section 4 to further reduce Lemma 3.1 to the case of a Euclidean magnetic Dirac operator on $\mathbb{R}^n$. In Section 5 we obtain the Birkhoff normal form for the Euclidean magnetic Dirac operator on $\mathbb{R}^n$ from Section 4. It is here in Section 5A that Koszul complexes are employed for the normal form. In Section 6 we show how the normal form is used in proving Lemma 3.1 via the use of almost analytic continuations. In Section 7 we prove Lemma 3.2 using the methods of local index theory. In Section 8 we show how to prove the spectral estimates of Theorem 1.2 via the trace expansion Theorem 1.3. Finally, in the Appendix we prove some spectral estimates useful in Sections 4 and 5.

2. Preliminaries

2A. ***Spectral invariants of the Dirac operator.*** Here we review the basic facts about Dirac operators used throughout the paper, with [Berline et al. 2004] providing a standard reference. Consider a compact, oriented, Riemannian manifold (X, g^{TX}) of odd dimension $n = 2m + 1$. Let X be equipped with spin structure, i.e., a principal $\operatorname{Spin}(n)$ bundle $\operatorname{Spin}(TX) \to \operatorname{SO}(TX)$ with an equivariant double covering of the principal $\mathrm{SO}(n)$-bundle of orthonormal frames $\mathrm{SO}(TX)$. The corresponding spin bundle $S = \mathrm{Spin}(TX) \times_{\mathrm{Spin}(n)} S_{2m}$ is associated to the unique irreducible representation of $\mathrm{Spin}(n)$. Let ∇^{TX} denote the Levi-Civita connection on TX. This lifts to the spin connection ∇^S on the spin bundle S. The Clifford multiplication endomorphism $c : T^*X \to S \otimes S^*$ may be defined (see Section 2B) satisfying

$$c(a)^2 = -|a|^2 \quad \forall a \in T^*X.$$

Let L be a Hermitian line bundle on X. Let A_0 be a fixed unitary connection on L and let $a \in \Omega^1(X;\mathbb{R})$ be a one-form on X. This gives a family $\nabla^h = A_0 + \frac{i}{h}a$ of unitary connections on L. We denote by $\nabla^{S\otimes L} = \nabla^S \otimes 1 + 1 \otimes \nabla^h$ the tensor product connection on $S \otimes L$. Each such connection defines a coupled Dirac operator

$$D_h := hD_{A_0} + ic(a) = hc \circ (\nabla^{S\otimes L}) : C^\infty(X; S \otimes L) \to C^\infty(X; S \otimes L)$$

for $h \in (0, 1]$. Each Dirac operator D_h is elliptic and self-adjoint. It hence possesses a discrete spectrum of eigenvalues.

We define the eta function of D_h by the formula

$$\eta(D_h, s) := \sum_{\substack{\lambda \neq 0 \\ \lambda \in \mathrm{Spec}(D_h)}} \mathrm{sign}(\lambda)|\lambda|^{-s} = \frac{1}{\Gamma\left(\frac{s+1}{2}\right)} \int_0^\infty t^{\frac{s-1}{2}} \,\mathrm{tr}(D_h e^{-tD_h^2})\, dt. \tag{2-1}$$

Here, and in the remainder of the paper, we use the convention that $\mathrm{Spec}(D_h)$ denotes a multiset with each eigenvalue of D_h being counted with its multiplicity. The above series converges for $\mathrm{Re}(s) > n$. It was shown in [Atiyah et al. 1975; 1976] that the eta function possesses a meromorphic continuation to the entire complex s-plane and has no pole at zero. Its value at zero is defined to be the eta invariant of the Dirac operator

$$\eta_h := \eta(D_h, 0).$$

By including the zero eigenvalue in (2-1), with an appropriate convention, we may define a variant, known as the reduced eta invariant, by

$$\bar{\eta}_h := \tfrac{1}{2}\{k_h + \eta_h\},$$

with $k_h = \dim\ker D_h$.

The eta invariant is unchanged under positive scaling:

$$\eta(D_h, 0) = \eta(cD_h, 0) \quad \forall c > 0. \tag{2-2}$$

Let $L_{t,h}$ denote the Schwartz kernel of the operator $D_h e^{-tD_h^2}$ on the product $X \times X$. Throughout the paper all Schwartz kernels will be defined with respect to the Riemannian volume density. Denote by $\mathrm{tr}(L_{t,h}(x,x))$ the pointwise trace of $L_{t,h}$ along the diagonal. We may now analogously define the function

$$\eta(D_h, s, x) = \frac{1}{\Gamma\left(\frac{s+1}{2}\right)} \int_0^\infty t^{\frac{s-1}{2}} \,\mathrm{tr}(L_{t,h}(x,x))\, dt. \tag{2-3}$$

In [Bismut and Freed 1986, Theorem 2.6], it was shown that for $\mathrm{Re}(s) > -2$, the function $\eta(D_h, s, x)$ is holomorphic in s and smooth in x. From (2-3) it is clear that this is equivalent to

$$\mathrm{tr}(L_{t,h}) = O(t^{\frac{1}{2}}) \quad \text{as } t \to 0. \tag{2-4}$$

The eta invariant is then given by the convergent integral

$$\eta_h = \int_0^\infty \frac{1}{\sqrt{\pi t}} \,\mathrm{tr}(D_h e^{-tD_h^2})\, dt. \tag{2-5}$$

2B. ***Clifford algebra and its representations.*** Here we review the construction of the spin representation of the Clifford algebra. The following, being standard, is merely used to set up our conventions and subsequently compute the spectrum of the model magnetic Dirac operator on $\mathbb{R}^m$ in Section 2B1.

Consider a real vector space V of even dimension $2m$ with metric $\langle\cdot,\cdot\rangle$. Recall that its Clifford algebra $\mathrm{Cl}(V)$ is defined as the quotient of the tensor algebra $T(V) := \bigoplus_{j=0}^\infty V^{\otimes j}$ by the ideal generated from the relations $v \otimes v + |v|^2 = 0$. Fix a compatible almost complex structure J and split $V \otimes \mathbb{C} = V^{1,0} \oplus V^{0,1}$ into the $\pm i$ eigenspaces of J. The complexification $V \otimes \mathbb{C}$ carries an induced $\mathbb{C}$-bilinear inner product $\langle\cdot,\cdot\rangle_{\mathbb{C}}$, as well as an induced Hermitian inner product $h^{\mathbb{C}}(\cdot,\cdot)$. Next, define $S_{2m} = \Lambda^* V^{1,0}$. Clearly S_{2m} is a complex vector space of dimension 2^m on which the unique irreducible (spin)-representation of the Clifford algebra $\mathrm{Cl}(V) \otimes \mathbb{C}$ is defined by the rule

$$c_{2m}(v)\omega = \sqrt{2}(v^{1,0} \wedge \omega - \iota_{v^{0,1}}\omega), \qquad v \in V,\ \omega \in S_{2m}.$$

The contraction above is taken with respect to $\langle\cdot,\cdot\rangle_{\mathbb{C}}$. It is clear that $c_{2m}(v) : \Lambda^{\text{even/odd}} \to \Lambda^{\text{odd/even}}$ switches the odd and even factors. For the Clifford algebra $\mathrm{Cl}(W) \otimes \mathbb{C}$ of an odd-dimensional vector space $W = V \oplus \mathbb{R}[e_0]$ there are exactly two irreducible representations. These two (spin)-representations $S^+_{2m+1} = S^-_{2m+1} = \Lambda^* V^{1,0}$ are defined via

$$\begin{gathered} c^{\pm}_{2m+1}(v) = c_{2m}(v), \quad v \in V, \\ c^{+}_{2m+1}(e_0)\omega_{\text{even/odd}} = -c^{-}_{2m+1}(e_0)\omega_{\text{even/odd}} = \pm i\,\omega_{\text{even/odd}}. \end{gathered} \tag{2-6}$$

Throughout the rest of the paper, we stick with the positive convention and use the shorthand $c = c_{2m}$, $c = c^+_{2m+1}$ when the indices $2m$, $2m+1$ are implicitly understood.

Pick an orthonormal basis $e_1, e_2, \dots, e_{2m}$ for V in which the almost complex structure is given by $Je_{2j-1} = e_{2j}$, $1 \le j \le m$. An $h^{\mathbb{C}}$-orthonormal basis for $V^{1,0}$ is now given by $w_j = \frac{1}{\sqrt{2}}(e_{2j} + ie_{2j-1})$, $1 \le j \le m$. A basis for S_{2m} and $S^{\pm}_{2m+1}$ is given by $w_k = w_1^{k_1} \wedge \cdots \wedge w_m^{k_m}$ with $k = (k_1, k_2, \dots, k_m) \in \{0,1\}^m$. Ordering the above chosen bases lexicographically in k, we may define the Clifford matrices, of rank 2^m, via

$$\gamma_j^m = c(e_j), \quad 0 \le j \le 2m,$$

for each m. Again, we often write $\gamma_j^m = \gamma_j$ with the index m implicitly understood. Giving representations of the Clifford algebra, these matrices satisfy the relation

$$\gamma_i\gamma_j + \gamma_j\gamma_i = -2\delta_{ij}. \tag{2-7}$$

Next, one may further define the Clifford quantization map on the exterior algebra

$$\begin{aligned} c : \Lambda^* W \otimes \mathbb{C} &\to \mathrm{End}(S_{2m}), \\ c(e_0^{k_0} \wedge \cdots \wedge e_{2m}^{k_{2m}}) &= c(e_0)^k \dots c(e_{2m})^{k_{2m}}. \end{aligned} \tag{2-8}$$

An easy computation yields

$$c(e_0 \wedge \cdots \wedge e_{2m}) = i^{m+1}.$$

Furthermore, if $e_0 \wedge \cdots \wedge e_{2m}$ is designated to give a positive orientation for W then for $\omega \in \Lambda^k W$ we have

$$c(*\omega) = i^{m+1}(-1)^{\frac{k(k+1)}{2}} c(\omega), \tag{2-9}$$

$$c(\omega)^* = (-1)^{\frac{k(k+1)}{2}} c(\omega) \tag{2-10}$$

under the Hodge star and $h^{\mathbb{C}}$-adjoint. The Clifford quantization map (2-8) is a linear surjection with kernel spanned by elements of the form $*\omega - i^{m+1}(-1)^{\frac{k(k+1)}{2}}\omega$. Thus, in particular one has linear isomorphisms

$$c : \Lambda^{\text{even/odd}} W \otimes \mathbb{C} \to \mathrm{End}(S_{2m}). \tag{2-11}$$

Next, given $(r_1, \dots, r_m) \in \mathbb{R}^m \setminus 0$, we define

$$I_r := \{ j \mid r_j \neq 0 \} \subset \{1, 2, \dots, m\}, \tag{2-12}$$

$$Z_r := |I_r|, \tag{2-13}$$

$$V_r := \bigoplus_{j \in I_r} \mathbb{C}[w_j] \subset V^{1,0}, \tag{2-14}$$

$$w_r := \sum_{j=1}^{m} r_j w_j \in V_r. \tag{2-15}$$

Clearly, $\|w_r\| = |r|$. Denoting by $w_r^{\perp}$ the $h^{\mathbb{C}}$-orthogonal complement of $w_r \subset V_r$, one clearly has $V_r = \mathbb{C}[w_r] \oplus w_r^{\perp}$. Hence

$$\begin{aligned} \Lambda^{\text{even}} V_r &= (\Lambda^{\text{even}} w_r^{\perp}) \oplus \frac{w_r}{|r|} \wedge (\Lambda^{\text{odd}} w_r^{\perp}), \\ \Lambda^{\text{odd}} V_r &= (\Lambda^{\text{odd}} w_r^{\perp}) \oplus \frac{w_r}{|r|} \wedge (\Lambda^{\text{even}} w_r^{\perp}). \end{aligned} \tag{2-16}$$

Next, we define

$$\mathtt{i}_r : \Lambda^* V_r \to \Lambda^* V_r \qquad \text{via} \qquad \mathtt{i}_r(\omega) := \frac{w_r}{|r|} \wedge \omega, \quad \mathtt{i}_r\left(\frac{w_r}{|r|} \wedge \omega\right) := \omega \tag{2-17}$$

for $\omega \in \Lambda^* w_r^{\perp}$. Clearly, $\mathtt{i}_r^2 = 1$ with the decomposition (2-16) implying that

$$\begin{aligned} \mathtt{i}_r &: \Lambda^{\text{even}} V_r \to \Lambda^{\text{odd}} V_r, \\ \mathtt{i}_r &: \Lambda^{\text{odd}} V_r \to \Lambda^{\text{even}} V_r \end{aligned}$$

are linear isomorphisms. Next, the endomorphism

$$c\left(\frac{w_r-\bar{w}_r}{\sqrt{2}}\right)=(w_r\wedge+\iota_{\bar{w}_r}):\Lambda^*V_r\to\Lambda^*V_r \tag{2-18}$$

has the form

$$c\left(\frac{w_r-\bar{w}_r}{\sqrt{2}}\right)=\begin{bmatrix} & |r|\mathtt{i}_r\\ |r|\mathtt{i}_r & \end{bmatrix} \tag{2-19}$$

with respect to the decomposition $\Lambda^*V_r=\Lambda^{\mathrm{odd}}V_r\oplus\Lambda^{\mathrm{even}}V_r$. This finally allows us to write the eigenspaces of (2-18) as

$$V_r^{\pm}=(1\pm\mathtt{i}_r)(\Lambda^{\mathrm{even}}V_r) \tag{2-20}$$

with eigenvalues $\pm|r|$ respectively.

2B1. *Magnetic Dirac operator on* $\mathbb{R}^m$. We now define the magnetic Dirac operator on $\mathbb{R}^m$ via

$$D_{\mathbb{R}^m}=\sum_{j=1}^{m}\left(\frac{\mu_j}{2}\right)^{\frac{1}{2}}[\gamma_{2j}(h\partial_{x_j})+i\gamma_{2j-1}x_j]\in\Psi^1_{\mathrm{cl}}(\mathbb{R}^m;\mathbb{C}^{2^m}). \tag{2-21}$$

Its square is computed in terms of the harmonic oscillator

$$D^2_{\mathbb{R}^m}=\mathtt{H}_2-ih\mathtt{R}_{2m+1}, \tag{2-22}$$

with

$$\mathtt{H}_2=\frac{1}{2}\sum_{j=1}^{m}\mu_j[-(h\partial_{x_j})^2+x_j^2],\quad \mathtt{R}_{2m+1}=\frac{1}{2}\sum_{j=1}^{m}\mu_j[\gamma_{2j-1}\gamma_{2j}]. \tag{2-23}$$

It is an easy exercise to show that

$$\mathtt{R}_{2m+1}w_k=\frac{i}{2}\left[\sum_{j=1}^{m}(-1)^{k_j-1}\mu_j\right]w_k. \tag{2-24}$$

Next, define the lowering and raising operators $A_j=h\partial_{x_j}+x_j$ and $A_j^*=-h\partial_{x_j}+x_j$ for $1\le j\le m$, and the Hermite functions

$$\begin{aligned}\psi_{\tau,k}(x)&:=\psi_\tau(x)\otimes w_k,\\ \psi_\tau(x)&:=\frac{1}{(\pi h)^{\frac{m}{4}}(2h)^{\frac{|\tau|}{2}}\sqrt{\tau!}}\left[\prod_{j=1}^{m}(A_j^*)^{\tau_j}\right]e^{-\frac{|x|^2}{2h}}\quad\text{for } \tau=(\tau_1,\tau_2,\dots,\tau_m)\in\mathbb{N}_0^m.\end{aligned} \tag{2-25}$$

It is well known that $\psi_{\tau,k}(x)$ form an orthonormal basis for $L^2(\mathbb{R}^m;\mathbb{C}^{2^m})$. Furthermore we have the standard relations

$$[A_j,A_j^*]=2h,\quad \mathtt{H}_2=\frac{1}{2}\sum_{j=1}^{m}\mu_j(A_jA_j^*-1). \tag{2-26}$$

It is clear from (2-22), (2-24) and (2-26) that each $\psi_{\tau,k}(x)$ is an eigenvector of $D^2_{\mathbb{R}^m}$ with eigenvalue

$$\lambda_{\tau,k} = h\sum_{j=1}^{m}(2\tau_j + 1 + (-1)^{k_j-1})\frac{\mu_j}{2}.$$

Hence, clearly the kernel of $D_{\mathbb{R}^m}$ is one-dimensional and spanned by $\psi_{0,0} = e^{-\frac{|x|^2}{2h}}$. We now find a decomposition of $L^2(\mathbb{R}^m;\mathbb{C}^{2^m})$ into eigenspaces of $D_{\mathbb{R}^m}$. First, if we define

$$\bar{\partial} = \frac{1}{2}\sum_{j=1}^{m}\left(\frac{\mu_j}{2}\right)^{\frac{1}{2}} c(w_j)A_j, \tag{2-27}$$

then one quickly computes

$$\bar{\partial}^* = -\frac{1}{2}\sum_{j=1}^{m}\left(\frac{\mu_j}{2}\right)^{\frac{1}{2}} c(\bar{w}_j)A_j^* \tag{2-28}$$

and

$$D_{\mathbb{R}^m} = \sqrt{2}(\bar{\partial} + \bar{\partial}^*). \tag{2-29}$$

For each $\tau \in \mathbb{N}_0^m \setminus 0$, we define I_τ, V_τ as in (2-12), (2-14) and set

$$E_\tau := \bigoplus_{b\in\{0,1\}^{I_\tau}} \mathbb{C}\left[\prod_{j\in I_\tau}\left(\frac{c(w_j)A_j}{\sqrt{2\tau_j h}}\right)^{b_j}\psi_{\tau,0}\right].$$

It is clear that we have an orthogonal decomposition

$$L^2(\mathbb{R}^m;\mathbb{C}^{2^m}) = \mathbb{C}[\psi_{0,0}] \oplus \bigoplus_{\tau\in\mathbb{N}_0^m\setminus 0} E_\tau.$$

Furthermore, we have the isomorphism

$$\mathscr{I}_\tau : \Lambda^* V_\tau \to E_\tau,$$
$$\mathscr{I}_\tau\left(\bigwedge_{j\in I_\tau} w_j^{b_j}\right) := \prod_{j\in I_\tau}\left(\frac{c(w_j)A_j}{\sqrt{2\tau_j h}}\right)^{b_j}\psi_{\tau,0}.$$

Each E_τ hence has dimension 2^{Z_τ} and is closed under $c(w_j)A_j$ and $c(\bar{w}_j)A_j^*$ for $1 \le j \le m$. We again have

$$E_\tau = E_\tau^{\text{even}} \oplus E_\tau^{\text{odd}}, \quad \text{where } E_\tau^{\text{even/odd}} := \mathscr{I}_\tau(\Lambda^{\text{even/odd}} V_\tau), \tag{2-30}$$

thus giving the Landau decomposition

$$L^2(\mathbb{R}^m;\mathbb{C}^{2^m}) = \mathbb{C}[\psi_{0,0}] \oplus \bigoplus_{\tau\in\mathbb{N}_0^m\setminus 0}(E_\tau^{\text{even}} \oplus E_\tau^{\text{odd}}). \tag{2-31}$$

The Dirac operator $D_{\mathbb{R}^m}$ by virtue of (2-27)–(2-29) preserves and acts on E_τ via

$$c\left(\frac{w_{r_\tau} + \bar{w}_{r_\tau}}{\sqrt{2}}\right) = (w_{r_\tau} \wedge + \iota_{\bar{w}_{r_\tau}}),$$

under the isomorphism $\mathscr{I}_\tau$, where $r_\tau := (\sqrt{\tau_1\mu_1 h}, \dots, \sqrt{\tau_m\mu_m h})$ and w_{r_τ} is as in (2-15). Hence, if we define $\mathtt{i}_\tau := \mathscr{I}_\tau \mathtt{i}_{r_\tau} \mathscr{I}_\tau^{-1} : E_\tau^{\text{even/odd}} \to E_\tau^{\text{odd/even}}$, we have that the restriction of $D_{\mathbb{R}^m}$ to E_τ is of the form

$$D_{\mathbb{R}^m} = \begin{bmatrix} & |r_\tau|\mathtt{i}_\tau \\ |r_\tau|\mathtt{i}_\tau & \end{bmatrix} \tag{2-32}$$

via (2-19). Also note that since $E_\tau^{\text{even/odd}} \subset \mathscr{I}_\tau(C^\infty(\mathbb{R}^m) \otimes \Lambda^{\text{even/odd}} V^{1,0})$ respectively, one has

$$c(e_0) E_\tau^{\text{even/odd}} = \pm i E_\tau^{\text{even/odd}} \tag{2-33}$$

using (2-6). The eigenspaces for $D_{\mathbb{R}^m}$ are now given by

$$E_\tau^\pm = \mathscr{I}_\tau(V_\tau^\pm), \tag{2-34}$$

via (2-20) with eigenvalues $\pm|r_\tau| = \pm\sqrt{\mu.\tau h}$ respectively. We now summarize.

Proposition 2.1. *An orthogonal decomposition of $L^2(\mathbb{R}^m; \mathbb{C}^{2^m})$ consisting of eigenspaces of the magnetic Dirac operator $D_{\mathbb{R}^m}$ (2-21) is given by*

$$L^2(\mathbb{R}^m; \mathbb{C}^{2^m}) = \mathbb{C}[\psi_{0,0}] \oplus \bigoplus_{\tau \in \mathbb{N}_0^m \setminus 0} (E_\tau^+ \oplus E_\tau^-).$$

Here $E_\tau^\pm$, as in (2-34), have dimension $2^{Z_\tau - 1}$ and correspond to the eigenvalues $\pm\sqrt{\mu.\tau h}$ respectively.

2C. ***The semiclassical calculus.*** Finally, here we review the semiclassical pseudodifferential calculus used throughout the paper, with [Guillemin and Sternberg 2013; Zworski 2012] being the detailed references. Let $\mathfrak{gl}(l)$ denote the space of all $l \times l$ complex matrices. For $A = (a_{ij}) \in \mathfrak{gl}(l)$ we define $|A| = \max_{ij} |a_{ij}|$. Denote by $\mathcal{S}(\mathbb{R}^n; \mathbb{C}^l)$ the space of Schwartz maps $f : \mathbb{R}^n \to \mathbb{C}^l$. We define the symbol space $S^m(\mathbb{R}^{2n}; \mathbb{C}^l)$ as the space of maps $a : (0, 1]_h \to C^\infty(\mathbb{R}^{2n}_{x,\xi}; \mathfrak{gl}(l))$ such that each of the seminorms

$$\|a\|_{\alpha,\beta} := \sup_{x,\xi,h} \langle\xi\rangle^{-m+|\beta|} |\partial_x^\alpha \partial_\xi^\beta a(x, \xi; h)|$$

is finite for all $\alpha, \beta \in \mathbb{N}_0^n$. Such a symbol is said to lie in the more refined class $a \in S^m_{\text{cl}}(\mathbb{R}^{2n}; \mathbb{C}^l)$ if there exists an h-independent sequence a_k, $k = 0, 1, \dots$ of symbols such that

$$a - \left(\sum_{k=0}^N h^k a_k\right) \in h^{N+1} S^m(\mathbb{R}^{2n}; \mathbb{C}^l) \quad \forall N. \tag{2-35}$$

Symbols as above can be Weyl quantized to define one-parameter families of operators $a^W : \mathcal{S}(\mathbb{R}^n; \mathbb{C}^l) \to \mathcal{S}(\mathbb{R}^n; \mathbb{C}^l)$ with Schwartz kernels given by

$$a^W := \frac{1}{(2\pi h)^n} \int e^{i(x-y).\frac{\xi}{h}} a\left(\frac{x+y}{2}, \xi; h\right) d\xi.$$

We denote by $\Psi^m_{\text{cl}}(\mathbb{R}^n; \mathbb{C}^l)$ the class of operators thus obtained by quantizing $S^m_{\text{cl}}(\mathbb{R}^{2n}; \mathbb{C}^l)$. This class of operators is closed under the standard operations of composition and formal-adjoint. Indeed, the Weyl symbols of the composition and adjoint satisfy

$$a^W \circ b^W = (a * b)^W := \left[e^{\frac{ih}{2}(\partial_{r_1}\partial_{s_2}-\partial_{r_2}\partial_{s_1})}(a(s_1,r_1;h)b(s_2,r_2;h))\right]^W_{x=s_1=s_2,\xi=r_1=r_2}, \\ (a^W)^* = (a^*)^W. \tag{2-36}$$

Furthermore the class is invariant under changes of coordinates and basis for $\mathbb{C}^l$. This allows one to define an invariant class of operators $\Psi^m_{\mathrm{cl}}(X;E)$ on $C^\infty(X;E)$ associated to any complex vector bundle on a smooth compact manifold X. These define uniformly in h bounded operators between the Sobolev spaces $H^s(X;E)$ and $H^{s-m}(X;E)$ with the h-dependent norm on each Sobolev space defined via

$$\|u\|_{H^s(X)} := \left\|(1+h^2\nabla^{E*}\nabla^E)^{\frac{s}{2}}u\right\|_{L^2}, \quad s\in\mathbb{R},$$

with respect to any metric g^{TX}, h^E on X, E and unitary connection ∇^E.

For $A\in\Psi^m_{\mathrm{cl}}(X;E)$, its principal symbol is well defined as an element in $\sigma(A)\in S^m(X;\mathrm{End}(E))\subset C^\infty(X;\mathrm{End}(E))$. One has that $\sigma(A)=0$ if and only if $A\in h\Psi^m_{\mathrm{cl}}(X;E)$. We remark that $\sigma(A)$ is the restriction of standard symbol in [Zworski 2012] to the refined class $\Psi^m_{\mathrm{cl}}(X;E)$ and is locally given by the first coefficient a_0 in the expansion of its Weyl symbol. The principal symbol satisfies the basic relations $\sigma(AB)=\sigma(A)\sigma(B)$ and $\sigma(A^*)=\sigma(A)^*$ with the formal adjoints being defined with respect to the same Hermitian metric h^E. The principal symbol map has an inverse given by the quantization map $\mathrm{Op}: S^m(X;\mathrm{End}(E))\to\Psi^m_{\mathrm{cl}}(X;E)$ satisfying $\sigma(\mathrm{Op}(a))=a\in S^m(X;\mathrm{End}(E))$. We often use the alternate notation $\mathrm{Op}(a)=a^W$. For a scalar function $b\in S^m(X)$, it is clear from the multiplicative property of the symbol that $[a^W,b^W]\in h\Psi^m_{\mathrm{cl}}(X;E)$ and we define $H_b(a):=\frac{i}{h}\sigma([a^W,b^W])\in S^m(X;\mathrm{End}(E))$. If a is self adjoint and b real, then it is easy to see that $H_b(a)$ is self-adjoint. We then define $|H_b(a)|=\max_{\lambda\in\mathrm{Spec}\,H_b(a)}|\lambda|$.

The wavefront set of an operator $A\in\Psi^m_{\mathrm{cl}}(X;E)$ can be defined invariantly as a subset $\mathrm{WF}(A)\subset\overline{T^*X}$ of the fiberwise radial compactification of its cotangent bundle. If the local Weyl symbol of A is given by a then $(x_0,\xi_0)\notin\mathrm{WF}(A)$ if and only if there exists an open neighborhood $(x_0,\xi_0;0)\in U\subset\overline{T^*X}\times(0,1]_h$ such that $a\in h^\infty\langle\xi\rangle^{-\infty}C^k(U;\mathbb{C}^l)$ for all k. The wavefront set satisfies the basic properties

$$\mathrm{WF}(A+B)\subset\mathrm{WF}(A)\cup\mathrm{WF}(B),\quad \mathrm{WF}(AB)\subset\mathrm{WF}(A)\cap\mathrm{WF}(B)\quad\text{and}\quad \mathrm{WF}(A^*)=\mathrm{WF}(A).$$

The wavefront set $\mathrm{WF}(A)=\varnothing$ is empty if and only if $A\in h^\infty\Psi^{-\infty}(X;E)$. We say that two operators A and B are equal microlocally on $U\subset\overline{T^*X}$ if $\mathrm{WF}(A-B)\cap U=\varnothing$. We also define by $\Psi^c_{\mathrm{cl}}(X;E)$ the class of pseudodifferential operators A with wavefront set $\mathrm{WF}(A)\Subset T^*X$ compactly contained in the cotangent bundle. It is clear that $\Psi^c_{\mathrm{cl}}(X;E)\subset\Psi^{-\infty}_{\mathrm{cl}}(X;E)$.

An operator $A\in\Psi^m_{\mathrm{cl}}(X;E)$ is said to be elliptic if $\langle\xi\rangle^m\sigma(A)^{-1}$ exists and is uniformly bounded on T^*X. If $A\in\Psi^m_{\mathrm{cl}}(X;E)$, $m>0$, is formally self-adjoint such that $A+i$ is elliptic then it is essentially self-adjoint (with domain $C^\infty_c(X;E)$) as an unbounded operator on $L^2(X;E)$. Its resolvent $(A-z)^{-1}\in\Psi^{-m}_{\mathrm{cl}}(X;E)$, $z\in\mathbb{C}$, $\mathrm{Im}\,z\neq 0$, now exists and is pseudodifferential by an application of Beals's lemma. The resolvent furthermore has an expansion $(A-z)^{-1}\sim\sum_{j=0}^\infty h^j\,\mathrm{Op}(a^z_j)$ in $\Psi^{-m}_{\mathrm{cl}}(X;E)$. Here each symbol appearing in the expansion has the form

$$a_j^z = (\sigma(A)-z)^{-1}a_{j,1}^z(\sigma(A)-z)^{-1}\cdots(\sigma(A)-z)^{-1}a_{j,2j}^z(\sigma(A)-z)^{-1} \in S^{-m}(X;\mathrm{End}(E)),$$

where $a_{j,k}^z$ is a polynomial in z symbols for $k=1,\dots,2j$. Given a Schwartz function $f\in\mathcal{S}(\mathbb{R})$, the Helffer–Sjöstrand formula now expresses the function $f(A)$ of such an operator in terms of its resolvent and an almost analytic continuation $\tilde{f}$ via

$$f(A) = \frac{1}{2\pi}\int_{\mathbb{C}} \bar{\partial}\tilde{f}(z)(A-z)^{-1}\,dz\,d\bar{z}.$$

Plugging the resolvent expansion into the above formula then shows that the above lies in and has an expansion $f(A)\sim\sum_{j=0}^{\infty} h^j A_j^f$ in $\Psi_{\mathrm{cl}}^{-\infty}(X;E)$. Finally, one defines the classical λ-energy level of A via

$$\Sigma_\lambda^A = \bigl\{(x,\xi)\in T^*X \bigm| \det\bigl(\sigma(A)(x,\xi)-\lambda I\bigr)=0\bigr\}.$$

Now, the form for the coefficients of the resolvent expansion also shows

$$\mathrm{WF}(f(A)) \subset \Sigma_{\mathrm{spt}(f)}^A := \bigcup_{\lambda\in\mathrm{spt}(f)} \Sigma_\lambda^A.$$

2C1. *The class $\Psi_\delta^m(X;E)$.* In Section 3 we shall need the more exotic class of symbols $S_\delta^m(\mathbb{R}^{2n};\mathbb{C})$ defined for each $0<\delta<\frac{1}{2}$. A function $a:(0,1]_h\to C^\infty(\mathbb{R}^{2n}_{x,\xi};\mathbb{C})$ is said to be in this class if and only if

$$\|a\|_{\alpha,\beta} := \sup_{x,\xi,h}\langle\xi\rangle^{-m+|\beta|}h^{(|\alpha|+|\beta|)\delta}|\partial_x^\alpha\partial_\xi^\beta a(x,\xi;h)| \tag{2-37}$$

is finite for all $\alpha,\beta\in\mathbb{N}_0^n$. This class of operators is closed under the standard operations of composition, adjoint and changes of coordinates allowing the definition of the exotic pseudodifferential algebra $\Psi_\delta^m(X)$ on a compact manifold. The class $S_\delta^m(X)$ is a family of functions $a:(0,1]_h\to C^\infty(T^*X;\mathbb{C})$ satisfying the estimates (2-37) in every coordinate chart and induced trivialization. Such a family can be quantized to $a^W\in\Psi_\delta^m(X)$ satisfying $a^Wb^W=(ab)^W+h^{1-2\delta}\Psi_\delta^{m+m'-1}(X)$ for another $b\in S_\delta^{m'}(X)$. The operators in $\Psi_\delta^0(X)$ are uniformly bounded on $L^2(X)$. Finally, the wavefront of an operator $A\in\Psi_\delta^m(X;E)$ is similarly defined and satisfies the same basic properties as before.

2C2. *Fourier integral operators.* We shall also need the local theory of Fourier integral operators. Let $\kappa: U\to V$ be an exact symplectomorphism between two open subsets $U\subset T^*X$ and $V\subset T^*Y$ inside cotangent spaces of manifolds of same dimension n. Assume that there exist local coordinates $(x_1,\dots,x_n)$,$(y_1,\dots y_n)$ on $\pi(U),\pi(V)$ respectively with induced canonical coordinates (x,ξ), (y,η) on U,V. A function $S(x,\eta)\in C^\infty(\Omega)$ on an open subset $\Omega\subset\mathbb{R}^{2n}_{x,\eta}$ is said to be a generating function for the graph of κ if the Lagrangian submanifolds

$$(T^*X)\times(T^*Y)^-\supset\Lambda_\kappa := \bigl\{((x,\xi);\kappa(x,\xi))\bigm|(x,\xi)\in U\bigr\}\quad\text{and}\quad\bigl\{(x,\partial_xS;\partial_\eta S,\eta)\bigm|(x,\eta)\in\Omega\bigr\}$$

are equal. Here $(T^*Y)^-$ denotes the cotangent bundle with the negative canonical symplectic form. A generating function S always exists locally near any point on Λ_κ. Letting $a:(0,1]_h\to C_c^\infty(\Omega\times\pi(V);\mathbb{C})$, which admits an expansion $a(x,y,\eta;h) \sim \sum_{k=0}^{\infty} h^k a_k(x,y,\eta)$, one may now define a Fourier integral operator associated to κ via

$$A : L^2(Y) \to L^2(X),$$
$$(Af)(x) = \frac{1}{(2\pi h)^n} \int_{\mathbb{R}^{2n}} e^{\frac{i}{h}(S(x,\eta)-y.\eta)} a(x,y,\eta;h) f(y)\, dy\, d\eta.$$

The symbol of $\sigma(A) \in C_c^\infty(\Lambda_\kappa;\mathbb{C})$ is defined using the generating function via $\sigma(A)(x,\eta) = a_0(x,\partial_x S,\eta)$. The adjoint A^* is again a Fourier integral operator associated to the symplectomorphism κ^{-1}. The wavefront set of A maybe defined as a subset $\mathrm{WF}(A) \subset \overline{T^*X} \times \overline{T^*Y}$. A point $(x,\xi;y,\eta)$ is not in $\mathrm{WF}(A)$ if and only if there exist pseudodifferential operators $B \in \Psi_{\mathrm{cl}}^m(X)$, $C \in \Psi_{\mathrm{cl}}^{m'}(Y)$ with $(x,\xi;y,\eta) \in \mathrm{WF}(B) \times \mathrm{WF}(C)$ such that $\|BAC\|_{H^s(Y)\to H^{s'}(X)} = O(h^\infty)$ for each $s, s' \in \mathbb{R}$. It can be shown that the wavefront set is in fact a compact subset $\mathrm{WF}(A) \subset \Lambda_\kappa$. Given a pseudodifferential operator $B \in \Psi_{\mathrm{cl}}^m(X)$, Egorov's theorem says that the composite is a pseudodifferential operator $A^*BA \in \Psi_{\mathrm{cl}}^m(Y)$. Moreover its principal symbol is given via $\sigma(A^*BA) = (\kappa^{-1})^* |\sigma(A)|^2 \sigma(B) \in C_c^\infty(V)$, where we have again used the identification of V with Λ_κ given by the generating function. Finally one has the wavefront relation $\mathrm{WF}(A^*BA) \subset \mathrm{WF}(A) \cap \mathrm{WF}(B)$, again using the identifications of U, V and Λ_κ.

An important special case arises when $\kappa = e^{tH_g}$ is the time t flow of a Hamiltonian $g \in S^m(T^*X)$. The operator $e^{\frac{it}{h} g^W}$, defined as a unitary operator via Stone's theorem, is now a Fourier integral operator associated to κ. Egorov's theorem now gives that the conjugation $e^{\frac{it}{h} g^W} A e^{-\frac{it}{h} g^W} \in \Psi_{\mathrm{cl}}^{m'}(X)$ is pseudodifferential for each $A \in \Psi_{\mathrm{cl}}^{m'}(X)$ with principal symbol $\sigma(e^{\frac{it}{h} g^W} A e^{-\frac{it}{h} g^W}) = (e^{tH_g})^* \sigma(A)$.

3. First reductions

The trace expansion theorem, Theorem 1.3, will be proved in two steps based on the following two lemmas. Below, τ, T, T', f, θ and D are the same as in Section 1.

Lemma 3.1. *Let $\varepsilon \in \left(0, \frac{1}{2}\right)$ and $\vartheta \in C_c^\infty((T'h^\varepsilon, T); [-1,1])$. Then*

$$\mathrm{tr}\left[f\left(\frac{D}{\sqrt{h}}\right) (\mathcal{F}_h^{-1}\vartheta)(\lambda\sqrt{h} - D)\right] = \mathrm{tr}\left[f\left(\frac{D}{\sqrt{h}}\right) \frac{1}{h} \check{\vartheta}\left(\frac{\lambda\sqrt{h} - D}{h}\right)\right] = O(h^\infty)$$

for all $\lambda \in \mathbb{R}$.

We note that in the above lemma the function ϑ is allowed to depend on h, while its support and range are contained in h-independent intervals.

Lemma 3.2. *There exist smooth functions $u_j \in C^\infty(\mathbb{R})$ such that for each $\lambda \in \mathbb{R}$ and $\varepsilon \in \left(0, \frac{1}{2}\right)$ one has a trace expansion*

$$\mathrm{tr}\left[f\left(\frac{D}{\sqrt{h}}\right) (\mathcal{F}_h^{-1}\theta_\varepsilon)(\lambda\sqrt{h} - D)\right] = \mathrm{tr}\left[f\left(\frac{D}{\sqrt{h}}\right) \frac{1}{h^{1-\varepsilon}} \check{\theta}\left(\frac{\lambda\sqrt{h} - D}{h^{1-\varepsilon}}\right)\right] = h^{-m-1}\left(\sum_{j=0}^{N-1} c_j h^{\frac{j}{2}} + O(h^{\frac{N}{2}})\right),$$

where $\theta_\varepsilon(x) := \theta(x/h^\varepsilon)$.

We note that the trace expansion theorem, Theorem 1.3, follows from the above two lemmas by simply splitting

$$\theta(x) = \theta_\varepsilon(x) + \underbrace{[\theta(x) - \theta_\varepsilon(x)]}_{\vartheta(x)}$$

and applying Lemmas 3.2 and 3.1 to the first and second summands respectively. Lemma 3.2 is a relatively classical expansion proved via local index theory and will be deferred to Section 7. Our main occupation until then is in proving Lemma 3.1.

As a first step, for $\tau > 0$ fixed one chooses a microlocal partition of unity $A_\alpha \in \Psi^0_{\mathrm{cl}}(X)$, $0 \le \alpha \le N$, satisfying

$$\sum_{\alpha=0}^{N} A_\alpha = 1, \qquad \mathrm{WF}(A_0) \subset U_0 \subset \overline{T^*X} \setminus \Sigma^D_{(-\tau,\tau)}, \qquad \mathrm{WF}(A_\alpha) \Subset U_\alpha \subset \Sigma^D_{(-2\tau,2\tau)}, \quad 1 \le \alpha \le N, \tag{3-1}$$

subordinate to an open cover $\{U_\alpha\}_{\alpha=0}^N$ of T^*X. Clearly, it suffices to prove

$$\mathrm{tr}\left[A_\alpha f\left(\frac{D}{\sqrt{h}}\right)\check{\vartheta}\left(\frac{\lambda\sqrt{h} - D}{h}\right)A_\beta\right] = O(h^\infty) \tag{3-2}$$

for $1 \le \alpha, \beta \le N$ with $\mathrm{WF}(A_\alpha) \cap \mathrm{WF}(A_\beta) \neq \varnothing$.

By the Helffer–Sjöstrand formula we have the trace above is given by

$$\mathcal{T}^{\vartheta}_{\alpha\beta}(D) := \frac{1}{2\pi}\int_{\mathbb{C}} \bar\partial \tilde f(z)\check{\vartheta}\left(\frac{\lambda - z}{\sqrt{h}}\right)\mathrm{tr}\left[A_\alpha\left(\frac{1}{\sqrt{h}}D - z\right)^{-1}A_\beta\right] dz\, d\bar z \tag{3-3}$$

for $\tilde f$ an almost analytic extension of f. We note that the resolvent, the above trace, and the left-hand side of (3-2) are well defined for any essentially self-adjoint pseudodifferential operator in place of D. The next reduction step attempts to modify D without affecting the asymptotics of $\mathcal{T}^{\vartheta}_{\alpha\beta}(D)$. To this end, choose open subsets $U_{\alpha\beta}$, $V_{\alpha\beta}$ such that

$$\begin{aligned} \mathrm{WF}(A_\alpha) \cap \mathrm{WF}(A_\beta) \subset\, & U_{\alpha\beta} \\ & \cap \\ \mathrm{WF}(A_\alpha) \cup \mathrm{WF}(A_\beta) \subset\, & V_{\alpha\beta} \Subset T^*X \end{aligned} \tag{3-4}$$

for each such pair α, β with $\mathrm{WF}(A_\alpha) \cap \mathrm{WF}(A_\beta) \neq \varnothing$. With $d = \sigma(D) \in C^\infty(X; i\mathfrak{u}(S))$, define the required exit time

$$T_{\alpha\beta} := \frac{1}{\inf_{g \in \mathcal{G}_{\alpha\beta}} |H_g d|}, \quad \text{where } \mathcal{G}_{\alpha\beta} := \left\{g \in C^\infty(T^*X; [0,1]) \;\middle|\; g|_{U_{\alpha\beta}} = 1,\ g|_{V^c_{\alpha\beta}} = 0\right\}. \tag{3-5}$$

If one were to use a scalar symbol $d \in C^\infty(X)$ instead in (3-5), the required exit time $T_{\alpha\beta}$ would have the following significance: any Hamiltonian trajectory $\gamma(t) = e^{tH_d}$ with $\gamma(0) \in U_{\alpha\beta}$ and $\gamma(T) \in V^c_{\alpha\beta}$ would have length $T \ge T_{\alpha\beta}$ at least the required exit time. We now have the following.

Lemma 3.3. *Let $D' \in \Psi^1_{\mathrm{cl}}(X;E)$ be essentially self-adjoint such that $D = D'$ microlocally on $V_{\alpha\beta}$. Then for $\vartheta \in C_c^\infty((T'_{\alpha\beta}h^\varepsilon, T_{\alpha\beta});[0,1])$, $0 < T'_{\alpha\beta} < T_{\alpha\beta}$, one has*

$$\mathcal{T}^{\vartheta}_{\alpha\beta}(D) = \mathcal{T}^{\vartheta}_{\alpha\beta}(D') \bmod h^\infty.$$

Proof. Let $B \in \Psi^0_{\mathrm{cl}}(X)$ be a microlocal cutoff such that $B = 0$ on $\mathrm{WF}(D - D')$ and $B = 1$ on $V_{\alpha\beta}$. Then $(1 - B)A_\beta = 0$ microlocally implies

$$\begin{aligned}\Big(z - \frac{1}{\sqrt{h}}D\Big)B\Big(z - \frac{1}{\sqrt{h}}D'\Big)^{-1}A_\beta&\\ = A_\beta - \Big[\frac{1}{\sqrt{h}}D, B\Big](z - D')^{-1}A_\beta + B\Big(\frac{1}{\sqrt{h}}D' - \frac{1}{\sqrt{h}}D\Big)\Big(z - \frac{1}{\sqrt{h}}D'\Big)^{-1}A_\beta \ (\mathrm{mod}\, h^\infty)&\end{aligned} \tag{3-6}$$

in trace norm. Next, multiplying through by $A_\alpha\big(z - \frac{1}{\sqrt{h}}D\big)^{-1}$and using $A_\alpha B = A_\alpha$ microlocally gives

$$\begin{aligned}A_\alpha\Big(z-\frac{1}{\sqrt{h}}D'\Big)^{-1}A_\beta - A_\alpha\Big(z-\frac{1}{\sqrt{h}}D\Big)^{-1}A_\beta = A_\alpha\Big(z-\frac{1}{\sqrt{h}}D\Big)^{-1}B\Big(\frac{1}{\sqrt{h}}D'-\frac{1}{\sqrt{h}}D\Big)\Big(z-\frac{1}{\sqrt{h}}D'\Big)^{-1}A_\beta&\\ -A_\alpha\Big(z-\frac{1}{\sqrt{h}}D\Big)^{-1}\Big[\frac{1}{\sqrt{h}}D, B\Big]\Big(z-\frac{1}{\sqrt{h}}D'\Big)^{-1}A_\beta&\\ +O(|\mathrm{Im}\, z|^{-1}h^\infty)&\end{aligned} \tag{3-7}$$

in trace norm. Now $B = 0$ on $\mathrm{WF}(D - D')$ gives that the first term on the right-hand side above is $O(|\mathrm{Im}\, z|^{-2}h^\infty)$.

We now estimate the second term. Let $S_{\alpha\beta} < S''_{\alpha\beta} < S'''_{\alpha\beta} < T_{\alpha\beta}$ and $S'_{\alpha\beta} > T'_{\alpha\beta}$ be such that $\vartheta \in C_c^\infty([S'_{\alpha\beta}h^\varepsilon, S_{\alpha\beta}];[0,1])$. Let $g_0 \in \mathcal{G}_{\alpha\beta}$ with $|H_{g_0}(d)| \le 1/S'''_{\alpha\beta}$. Set $g = \alpha_z g_0$, where

$$\alpha_z = \min\Big(\frac{S''_{\alpha\beta}\,\mathrm{Im}\, z}{\sqrt{h}\log(1/h)}, N\Big),$$

with the constant $N > 0$ to be specified later. We note that

$$G = (e^{g\log\frac{1}{h}})^W \in h^{-N}\Psi^0_\delta(X)$$

for each $0 < \delta < \frac{1}{2}$. Since it has an elliptic symbol, we may construct its inverse by symbolic calculus $G^{-1} \in h^N\Psi^0_\delta(X)$. Moreover

$$G\Big(z - \frac{1}{\sqrt{h}}D_h\Big)G^{-1} = \Big(z - \frac{1}{\sqrt{h}}D_h\Big) + i\Big(\alpha_z\sqrt{h}\log\frac{1}{h}\Big)(H_{g_0}(d))^W, \tag{3-8}$$

with

$$R = O\Big(h^{\frac{3}{2}}\alpha_z\log\frac{1}{h}\Big) \quad \text{in } S^0_\delta(X). \tag{3-9}$$

Now, since

$$\Big|\Big(\alpha_z\sqrt{h}\log\frac{1}{h}\Big)H_{g_0}(d)\Big| \le \frac{S''_{\alpha\beta}}{S'''_{\alpha\beta}}|\mathrm{Im}\, z| < |\mathrm{Im}\, z|,$$

the inverse $G\big(z-\frac{1}{\sqrt{h}}D_h\big)^{-1}G^{-1}$ of the above exists and is $O(|\operatorname{Im} z|^{-1})$ in operator norm for $\operatorname{Im} z \neq 0$ and h sufficiently small.

Next, pick $C \in \Psi^0_{\mathrm{cl}}(X)$ such that $\mathrm{WF}(C) \subset U_{\alpha\beta}$ and $C = 1$ on $\mathrm{WF}(A_\alpha) \cap \mathrm{WF}(A_\beta)$. Now $G = e^{\alpha_z \log \frac{1}{h}}$ on $\mathrm{WF}(CA_\alpha)$, $G = G^{-1} = I$ on $\mathrm{WF}(B) \setminus V_{\alpha\beta}$ and $[D_h, B] = 0$ on $V_{\alpha\beta}$ imply

$$e^{\alpha_z \log \frac{1}{h}} CA_\alpha\left(z-\frac{1}{\sqrt{h}}D_h\right)^{-1}\left[\frac{1}{\sqrt{h}}D_h, B\right] = CA_\alpha G\left(z-\frac{1}{\sqrt{h}}D_h\right)^{-1}G^{-1}\left[\frac{1}{\sqrt{h}}D_h, B\right] + O(|\operatorname{Im} z|^{-1}h^\infty)$$

in trace norm. The above is now $O(|\operatorname{Im} z|^{-1}h^{-n})$ in trace norm. Hence

$$CA_\alpha\left(z-\frac{1}{\sqrt{h}}D_h\right)^{-1}\left[\frac{1}{\sqrt{h}}D_h, B\right] = O\big(|\operatorname{Im} z|^{-1}h^{-n}\max(h^N, e^{-\frac{S''_{\alpha\beta}\operatorname{Im} z}{\sqrt{h}}})\big)$$

in trace norm. This and $CA_\alpha A_\beta = A_\alpha A_\beta$ now estimate the second term of (3-7) to give

$$A_\alpha\left(z-\frac{1}{\sqrt{h}}D'_h\right)^{-1}A_\beta - A_\alpha\left(z-\frac{1}{\sqrt{h}}D_h\right)^{-1}A_\beta = O\big(|\operatorname{Im} z|^{-2}h^{-n}\max(h^N, e^{-\frac{S''_{\alpha\beta}\operatorname{Im} z}{\sqrt{h}}})\big) \tag{3-10}$$

in trace norm.

Next, we have the Paley–Wiener estimate

$$\check{\vartheta}\left(\frac{\lambda-z}{\sqrt{h}}\right) = \begin{cases} O(e^{\frac{S_{\alpha\beta}(\operatorname{Im} z)}{\sqrt{h}}}), & \operatorname{Im} z > 0, \\ O(e^{\frac{S'_{\alpha\beta}(\operatorname{Im} z)}{h^{1/2-\varepsilon}}}), & \operatorname{Im} z < 0. \end{cases} \tag{3-11}$$

Introduce $\psi \in C^\infty(\mathbb{R}; [0,1])$ such that

$$\psi(x) = \begin{cases} 1, & x \le 1, \\ 0, & x \ge 2. \end{cases}$$

Setting

$$\psi_M(z) = \psi\left(\frac{\operatorname{Im} z}{M\sqrt{h}\log(1/h)}\right)$$

for another constant $M > 1$ yet to be chosen, we have the estimate

$$\bar{\partial}(\psi_M \tilde{f}) = \begin{cases} O\left(\psi_M|\operatorname{Im} z|^N + \dfrac{1}{M\sqrt{h}\log(1/h)}\,1_{[1,2]}\left(\dfrac{\operatorname{Im} z}{M\sqrt{h}\log(1/h)}\right)\right), & \operatorname{Im} z > 0, \\ O(|\operatorname{Im} z|^N), & \operatorname{Im} z < 0. \end{cases} \tag{3-12}$$

Finally, (3-10)–(3-12), along with the observation

$$\psi_M|\operatorname{Im} z|^N = O\left(\left(M\sqrt{h}\log\frac{1}{h}\right)^N\right),$$

give

$$
\begin{aligned}
&\mathcal{T}^{\vartheta}_{\alpha\beta}(D') - \mathcal{T}^{\vartheta}_{\alpha\beta}(D) \\
&\quad = \frac{1}{\pi}\int_{\mathbb{C}} \bar{\partial}(\psi_M \tilde{f})\check{\vartheta}\left(\frac{\lambda - z}{\sqrt{h}}\right)\left[A_\alpha\left(z - \frac{1}{\sqrt{h}}D'_h\right)^{-1}A_\beta - A_\alpha\left(z - \frac{1}{\sqrt{h}}D_h\right)^{-1}A_\beta\right]dz\, d\bar{z} \\
&\quad = O(h^\infty) + O\left[\int_{\{M\sqrt{h}\log\frac{1}{h} \le \operatorname{Im} z \le 2M\sqrt{h}\log\frac{1}{h}\}} \frac{h^{-n}}{\sqrt{h}\log\frac{1}{h}} \max(h^N e^{\frac{S_{\alpha\beta}(\operatorname{Im} z)}{\sqrt{h}}}, e^{-\frac{(S''_{\alpha\beta} - S_{\alpha\beta})\operatorname{Im} z}{\sqrt{h}}})\right] \\
&\quad = O\left[\max(h^{N - 2MS_{\alpha\beta} - n}, h^{M(S''_{\alpha\beta} - S_{\alpha\beta}) - n})\right].
\end{aligned}
$$

Choosing $M \gg n/(S''_{\alpha\beta} - S_{\alpha\beta})$ and furthermore $N \gg 2MS_{\alpha\beta} + n$ gives the result. □

In the proof above we have closely followed [Dimassi and Sjöstrand 1999, Lemma 12.7]. Again, the proof above avoids the use of an unknown parametrix for $e^{\frac{it}{h}D}$ which, following the significance of the required exit time $T_{\alpha\beta}$ noted before, maybe used to give an alternate proof in the case when d is scalar.

4. Reduction to $\mathbb{R}^n$

In this section we shall further reduce to the case of a Dirac operator on $\mathbb{R}^n$. First we cover X by a finite set of Darboux charts $\{\varphi_s : \Omega_s \to \Omega^0_s \subset \mathbb{R}^n\}_{s\in S}$ for the contact form a, centered at points $\{x_s\}_{s\in S} \in X$. By shrinking the partition of unity (3-1) we may assume that for each pair α, β, with $\mathrm{WF}(A_\alpha) \cap \mathrm{WF}(A_\beta) \neq \varnothing$, the open sets $V_{\alpha\beta} \subset T^*\Omega_s$ in (3-4) are contained in some Darboux chart. Now consider such a chart Ω_s with coordinates$(x_0, \ldots, x_{2m})$ centered at $x_s \in X$ and an orthonormal frame $\{e_j = w^k_j \partial_{x_k}\}$, $0 \le j \le 2m$, for the tangent bundle on Ω_s. We hence have

$$
w^k_j g_{kl} w^l_r = \delta_{jr}, \tag{4-1}
$$

where g_{kl} is the metric in these coordinates and the Einstein summation convention is being used. Let Γ^l_{jk} be the Christoffel symbols for the Levi-Civita connection in the orthonormal frame e_i satisfying $\nabla_{e_j} e_k = \Gamma^l_{jk} e_l$. This orthonormal frame induces an orthonormal frame u_q, $1 \le q \le 2^m$, for the spin bundle S. We further choose a local orthonormal section $\mathtt{1}(x)$ for the Hermitian line bundle L and define via $\nabla^{A_0}_{e_j}\mathtt{1} = \Upsilon_j(x)\mathtt{1}$, $0 \le j \le 2m$, the Christoffel symbols of the unitary connection A_0 on L. In terms of the induced frame $u_q \otimes \mathtt{1}$, $1 \le q \le 2^m$, for $S \otimes L$ the Dirac operator (1-2) has the form [Berline et al. 2004, Section 3.3]

$$
D = \gamma^j w^k_j P_k + h\left(\tfrac{1}{4}\Gamma^l_{jk}\gamma^j\gamma^k\gamma_l + \Upsilon_j\gamma^j\right), \tag{4-2}
$$

where

$$
P_k = h\partial_{x_k} + i a_k, \tag{4-3}
$$

and

$$
a(x) = a_k\, dx^k = dx_0 + \sum_{j=1}^{m}(x_j\, dx_{j+m} - x_{j+m}\, dx_j) \tag{4-4}
$$

is the standard contact one-form in these coordinates.

The expression in (4-2) is formally self-adjoint with respect to the Riemannian density $e^0\wedge\cdots\wedge e^{2m} = \sqrt{g}\,dx := \sqrt{g}\,dx^0\wedge\cdots\wedge dx^{2m}$ with $g=\det(g_{ij})$. To get an operator self-adjoint with respect to the Euclidean density dx, one expresses the Dirac operator in the framing $g^{\frac14}u_q\otimes 1$, $1\le q\le 2^m$. In this new frame the expression (4-2) for the Dirac operator needs to be conjugated by $g^{\frac14}$ and hence the term $h\gamma^j w_j^k g^{-\frac14}(\partial_{x_k}g^{\frac14})$ needs to be added. Hence, the Dirac operator in the new frame has the form

$$D=[\sigma^j w_j^k(\xi_k+a_k)]^W+hE\in\Psi^1_{\mathrm{cl}}(\Omega_s^0;\mathbb{C}^{2^m}),$$

with $\sigma^j=i\gamma^j$, for some self-adjoint endomorphism $E(x)\in C^\infty(\Omega_s^0;i\mathfrak{u}(\mathbb{C}^{2^m}))$.

The one-form a is extended to all of $\mathbb{R}^n$ by the same formula (4-4). The functions w_j^k are extended such that

$$(w_j^k\partial_{x_k}\otimes dx^j)|_{(K_s^0)^c}=\partial_{x_0}\otimes dx^0+\sum_{j=1}^m\mu_j^{\frac12}(\partial_{x_j}\otimes dx^j+\partial_{x_{j+m}}\otimes dx^{j+m})$$

$\big($and hence $g|_{(K_s^0)^c}=dx_0^2+\sum_{j=1}^m\mu_j(dx_j^2+dx_{j+m}^2)\big)$ outside a compact neighborhood $\Omega_s^0\Subset K_s^0$. These extensions may further be chosen such that the suitability assumption Definition 1.1 holds globally on $\mathbb{R}^n$ and for an extended positive function $\nu\in C_c^\infty(\mathbb{R}^n)$ satisfying

$$\nu_0\le\mu_1(\inf_{\mathbb{R}^n}\nu). \tag{4-5}$$

The endomorphism $E(x)\in C_c^\infty(\mathbb{R}^n;i\mathfrak{u}(\mathbb{C}^{2^m}))$ is extended to an arbitrary self-adjoint endomorphism of compact support. This now gives

$$D_0=[\sigma^j w_j^k(\xi_k+a_k)]^W+hE\in\Psi^1_{\mathrm{cl}}(\mathbb{R}^n;\mathbb{C}^{2^m}), \tag{4-6}$$

as a well defined formally self adjoint operator on $\mathbb{R}^n$. Furthermore, the symbol of D_0+i is elliptic in the class $S^0(m)$ for the order function

$$m=\left(1+\sum_{k=0}^{2m}(\xi_k+a_k)^2\right)^{\frac12},$$

and hence D_0 is essentially self adjoint; see [Dimassi and Sjöstrand 1999, Chapter 8]. Below $\vartheta\in C_c^\infty((T'_{\alpha\beta}h^\varepsilon,T_{\alpha\beta});[0,1])$, $0<T'_{\alpha\beta}<T_{\alpha\beta}$, as before and we set $V_{\alpha\beta}^0:=(d\varphi_s)^*V_{\alpha\beta}\subset T^*\Omega_s^0$.

Proposition 4.1. *There exist $A_\alpha^0, A_\beta^0\in\Psi^0_{\mathrm{cl}}(\mathbb{R}^n)$, with $\mathrm{WF}(A_\alpha^0)\cup\mathrm{WF}(A_\beta^0)\Subset V_{\alpha\beta}^0\subset T^*\tilde{\Omega}_s$, such that*

$$\mathcal{T}_{\alpha\beta}^\vartheta(D)=\underbrace{\mathrm{tr}\left[A_\alpha^0 f\left(\frac{D_0}{\sqrt{h}}\right)\check{\vartheta}\left(\frac{\lambda\sqrt{h}-D_0}{h}\right)A_\beta^0\right]}_{:=\mathcal{T}_{\alpha\beta}^\vartheta(D_0)}\ \mathrm{mod}\ h^\infty.$$

Proof. Let $K'_{\alpha\beta}, K''_{\alpha\beta}$ and $V'_{\alpha\beta}, V''_{\alpha\beta}$ be compact and open subsets respectively satisfying $V_{\alpha\beta}\subset K'_{\alpha\beta}\subset V'_{\alpha\beta}\subset K''_{\alpha\beta}\subset V''_{\alpha\beta}\subset T^*\Omega_s$. Choose $D'\in\Psi^0_{\mathrm{cl}}(X;S)$ self-adjoint such that $D=D'$ microlocally on $K'_{\alpha\beta}$ and

$$\Sigma^{D'}_{(-\infty,2\tau]}\subset V'_{\alpha\beta} \tag{4-7}$$

and set $E = D' - 3\tau \in \Psi^0_{\mathrm{cl}}(X; S)$. Pick a cutoff function $\chi(x; y, \eta) \in C_c^\infty\big(\pi(V''_{\alpha\beta}) \times (d\varphi_s)^* V''_{\alpha\beta}; [0, 1]\big)$ such that $\chi = 1$ on $\pi(K''_{\alpha\beta}) \times (d\varphi_s)^* K''_{\alpha\beta}$. Now define the operator

$$U : L^2(\mathbb{R}^n; \mathbb{C}^{2^m}) \to L^2(X; S),$$
$$(Uf)(x) = \frac{1}{(2\pi h)^n} \int e^{\frac{i}{h}(\varphi_s(x) - y).\eta} \chi(x; y, \eta) f(y)\, dy\, d\eta, \quad x \in X.$$

The above is a semiclassical Fourier integral operator associated to symplectomorphism $\kappa = (d\varphi_s^{-1})^*$ given by the canonical coordinates. Its adjoint $U^* : L^2(X; S) \to L^2(\mathbb{R}^n; \mathbb{C}^{2^m})$ is again a semiclassical Fourier integral operator associated to the symplectomorphism $\kappa^{-1} = (d\varphi_s)^*$. A simple computation gives the following compositions are pseudodifferential with

$$UU^* = I \quad \text{microlocally on } K''_{\alpha\beta}, \tag{4-8}$$
$$U^* U = I \quad \text{microlocally on } \kappa(K''_{\alpha\beta}). \tag{4-9}$$

The composition

$$E' = E_0 := U^* E U \in \Psi^0_{\mathrm{cl}}(\mathbb{R}^n; \mathbb{C}^{2^m})$$

is now a pseudodifferential operator by Egorov's theorem with symbol

$$\sigma(E_0) = (d\varphi_s)^* \chi^2.\sigma(E). \tag{4-10}$$

Similarly, $E'_0 := U E_0 U^* \in \Psi^0_{\mathrm{cl}}(X; S)$ and

$$\sigma(E'_0) = (d\varphi_s)^* \chi^4.\sigma(E_0). \tag{4-11}$$

By (4-7), (4-10) and (4-11) we have $\Sigma^{E_0}_{(-\infty,-\tau]} \subset \kappa(V'_{\alpha\beta})$ and $\Sigma^{E'_0}_{(-\infty,-\tau]} \subset V'_{\alpha\beta}$. Hence by Proposition A.6, E, E', E_0 and E'_0 all have discrete spectrum in $(-\infty, -\tau]$. We now select $g \in C_c^\infty(-5\tau, -\tau)$ such that $g = 1$ on $[-4\tau, -2\tau]$. We have

$$\mathrm{WF}(g(E)) \subset \Sigma^E_{\mathrm{spt}(g)} \subset \Sigma^E_{(-\infty,-\tau]} \subset V'_{\alpha\beta}.$$

Combined with (4-9) this gives $(U^*U - I) g(E) \in h^\infty \Psi^{-\infty}_{\mathrm{cl}}(X; S)$ and hence $\|(U^*U - I) g(E)\| = O(h^\infty)$ as an operator on $L^2(X; S)$. This in turn now gives

$$\big\|(U^*U - I)\Pi^E\big\|(\|E\|\,\|U\| + 1) = O(h^\infty), \tag{4-12}$$

with $\Pi^E = \Pi^E_{[-4\tau,-2\tau]}$ denoting the spectral projector of E onto the interval $[-4\tau, -2\tau]$. Similarly, we get

$$\big\|(UU^* - I)\Pi^{E_0}\big\|(\|E_0\|\,\|U^*\| + 1) = O(h^\infty). \tag{4-13}$$

Another easy computation gives $E = E'_0$ microlocally on $K''_{\alpha\beta}$ and we may similarly estimate

$$\big\|(E - E'_0)\Pi^{E'_0}\big\| = O(h^\infty). \tag{4-14}$$

Next we define $A^0_\alpha := U^* A_\alpha U,\ A^0_\beta := U^* A_\beta U \in \Psi^0_{\mathrm{cl}}(\mathbb{R}^n)$ and again note

$$\begin{aligned} UA^0_\alpha A^0_\beta U^* &= A_\alpha A_\beta \quad \text{microlocally on } K''_{\alpha\beta}, \\ U^* A_\alpha A_\beta U &= A^0_\alpha A^0_\beta \quad \text{microlocally on } \kappa(K''_{\alpha\beta}). \end{aligned} \tag{4-15}$$

This again gives

$$\big\| [UA^0_\alpha A^0_\beta U^* - A_\alpha A_\beta]\Pi^E \big\| = O(h^\infty), \tag{4-16}$$

$$\big\| [U^* A_\alpha A_\beta U - A^0_\alpha A^0_\beta]\Pi^{E_0} \big\| = O(h^\infty). \tag{4-17}$$

Now using (4-12), (4-13), (4-14), (4-16), (4-17) and using the cyclicity of the trace we may apply Proposition A.5 of the Appendix with

$$\rho(x) = f\left(\frac{x+3\tau}{\sqrt{h}}\right)\check{\vartheta}\left(\frac{\lambda\sqrt{h}-3\tau-x}{h}\right)$$

to get

$$\mathrm{tr}\left[A_\alpha f\left(\frac{D'}{\sqrt{h}}\right)\check{\vartheta}\left(\frac{\lambda\sqrt{h}-D'}{h}\right)A_\beta\right] - \mathrm{tr}\left[A^0_\alpha f\left(\frac{D'_0}{\sqrt{h}}\right)\check{\vartheta}\left(\frac{\lambda\sqrt{h}-D'_0}{h}\right)A^0_\beta\right] = O(h^\infty)$$

for $D'_0 := E_0 + 3\tau$. Finally observing $D = D'$ on $V_{\alpha\beta}$, $D_0 = D'_0$ on $V^0_{\alpha\beta}$ and using Lemma 3.3 completes the proof. □

5. Birkhoff normal form for the Dirac operator

In this section we derive a Birkhoff normal form for the Dirac operator (4-6) on $\mathbb{R}^n$. First consider the function

$$f_0 := \left(x_0\xi_0 - \frac{x_0}{(\sqrt{2}-1)}\right)\frac{\ln 4}{\pi} + \sum_{j=1}^m (x_j x_{j+m} + \xi_j \xi_{j+m}).$$

If H_{f_0} and $e^{tH_{f_0}}$ denote the Hamilton vector field and time t flow of f_0 respectively then it is easy to compute

$$\begin{aligned} e^{\frac{\pi}{4}H_{f_0}}(x_0,\xi_0) &= \left(\sqrt{2}x_0, \frac{\xi_0+1}{\sqrt{2}}\right), \\ e^{\frac{\pi}{4}H_{f_0}}(x_j,\xi_j;x_{j+m},\xi_{j+m}) &= \left(\frac{x_j+\xi_{j+m}}{\sqrt{2}}, \frac{-x_{j+m}+\xi_j}{\sqrt{2}}; \frac{x_{j+m}+\xi_j}{\sqrt{2}}, \frac{-x_j+\xi_{j+m}}{\sqrt{2}}\right). \end{aligned}$$

We abbreviate $(x',\xi') = (x_1,\dots,x_m;\xi_1,\dots,\xi_m)$, $(x'',\xi'') = (x_{m+1},\dots,x_{2m};\xi_{m+1},\dots,\xi_{2m})$ and $(x,\xi) = (x_0,x',x'';\xi_0,\xi',\xi'')$. Further, let $o_N \subset S^1_{\mathrm{cl}}(\mathbb{R}^{2n};\mathbb{C}^l)$ denote the subspace of self-adjoint symbols $a : (0,1]_h \to C^\infty(\mathbb{R}^{2n}_{x,\xi}; i\mathfrak{u}(2^m))$ such that each of the coefficients a_k, $k = 0,1,2,\dots$, in its symbolic expansion (2-35) vanishes to order N in (ξ_0,x',ξ') at 0. We also denote by o_N the space of Weyl quantizations of such symbols.

Using Egorov's theorem, the operator (4-6) is conjugated to

$$e^{\frac{i\pi}{4h}f_0^W} D_0 e^{-\frac{i\pi}{4h}f_0^W} = d_0^W, \tag{5-1}$$

with

$$d_0 = \sqrt{2}(\sigma^j w^0_{j,f_0}\xi_0 + \sigma^j w^k_{j,f_0}\xi_k + \sigma^j w^{k+m}_{j,f_0} x_k) + ho_0, \tag{5-2}$$

where

$$w^k_{j,f_0} = (e^{-\frac{\pi}{4}H_{f_0}})^* w^k_j. \tag{5-3}$$

Note that the index k ranges from 1 to m in the Einstein summation above. A Taylor expansion of d_0, given in (5-2), in (ξ_0, x', ξ') now gives $r^0_j \in o_2,\ 0 \le j \le 2m$, such that

$$d_0 = \sqrt{2}\sigma^j(\bar{w}^0_j\xi_0 + \bar{w}^k_j\xi_k + \bar{w}^{k+m}_j x_k) + \sigma^j r^0_j + ho_0$$

and where $\bar{w}^k_j(x_0, x'', \xi'') = w^k_j\big(x_0, -\frac{\xi''}{\sqrt{2}}, \frac{x''}{\sqrt{2}}\big)$. On squaring using (4-1) we obtain

$$(d^W_0)^2 = Q^W_0 + ho_1 + o_3 + h^2 o_0,$$

with

$$Q_0 = \begin{bmatrix} x' & \xi_0 & \xi' \end{bmatrix} \begin{bmatrix} \bar{g}^{(k+m)(l+m)}(x_0, x'', \xi'') & \bar{g}^{(k+m)0}(x_0, x'', \xi'') & \bar{g}^{(k+m)l}(x_0, x'', \xi'') \\ \bar{g}^{0(l+m)}(x_0, x'', \xi'') & \bar{g}^{00}(x_0, x'', \xi'') & \bar{g}^{0l}(x_0, x'', \xi'') \\ \bar{g}^{k(l+m)}(x_0, x'', \xi'') & \bar{g}^{k0}(x_0, x'', \xi'') & \bar{g}^{kl}(x_0, x'', \xi'') \end{bmatrix} \begin{bmatrix} x' \\ \xi_0 \\ \xi' \end{bmatrix}.$$

Here $\bar{g}^{kl}(x_0, x'', \xi'') = 2g^{kl}\big(x_0, -\frac{\xi''}{\sqrt{2}}, \frac{x''}{\sqrt{2}}\big)$ and the g^{kl} are the components of the inverse metric on $T^*\mathbb{R}^n$.

Next we consider another function f_1 of the form

$$f_1 = \tfrac{1}{2}\begin{bmatrix} x' & \xi_0 & \xi' \end{bmatrix} \begin{bmatrix} \alpha_{m\times m}(x_0, x'', \xi'') & \gamma_{m\times m+1}(x_0, x'', \xi'') \\ \gamma^t_{m+1\times m}(x_0, x'', \xi'') & \beta_{m+1\times m+1}(x_0, x'', \xi'') \end{bmatrix} \begin{bmatrix} x' \\ \xi_0 \\ \xi' \end{bmatrix},$$

where α, β and γ are matrix-valued functions of the given orders, with α, β symmetric. An easy computation now shows

$$(e^{H_{f_1}})^* \begin{bmatrix} x' \\ \xi_0 \\ \xi' \end{bmatrix} = e^{\Lambda} \begin{bmatrix} x' \\ \xi_0 \\ \xi' \end{bmatrix} + o_2,$$

with

$$\Lambda(x_0, x'', \xi'') = \begin{bmatrix} 0 & -I_{m+1\times m+1} \\ I_{m\times m} & 0 \end{bmatrix} \begin{bmatrix} \alpha_{m\times m}(x_0, x'', \xi'') & \gamma_{m\times m+1}(x_0, x'', \xi'') \\ \gamma^t_{m+1\times m}(x_0, x'', \xi'') & \beta_{m+1\times m+1}(x_0, x'', \xi'') \end{bmatrix}.$$

From the suitability assumption (1-3), we have that there exist smooth matrix-valued functions α, β and γ such that

$$\begin{aligned} &\begin{bmatrix} x' & \xi_0 & \xi' \end{bmatrix} e^{\Lambda^t} \begin{bmatrix} \bar{g}^{(k+m)(l+m)}(x_0, x'', \xi'') & \bar{g}^{(k+m)0}(x_0, x'', \xi'') & \bar{g}^{(k+m)l}(x_0, x'', \xi'') \\ \bar{g}^{0(l+m)}(x_0, x'', \xi'') & \bar{g}^{00}(x_0, x'', \xi'') & \bar{g}^{0l}(x_0, x'', \xi'') \\ \bar{g}^{k(l+m)}(x_0, x'', \xi'') & \bar{g}^{k0}(x_0, x'', \xi'') & \bar{g}^{kl}(x_0, x'', \xi'') \end{bmatrix} e^{\Lambda} \begin{bmatrix} x' \\ \xi_0 \\ \xi' \end{bmatrix} \\ &\qquad = \xi_0^2 + \bar{v}\Big[\sum_{j=1}^m \mu_j(x_j^2 + \xi_j^2)\Big] + o_3, \end{aligned}$$

where

$$\bar{\nu}(x_0, x'', \xi'') = \nu\left(x_0, -\frac{\xi''}{\sqrt{2}}, \frac{x''}{\sqrt{2}}\right). \tag{5-4}$$

Letting

$$H_2 = \frac{1}{2}\sum_{j=1}^{m} \mu_j (x_j^2 + \xi_j^2),$$

Egorov's theorem now gives

$$e^{\frac{i}{h} f_1^W} d_0^W e^{-\frac{i}{h} f_1^W} = \left(\sum_{j=0}^{2m} \sigma_j b_j\right)^W + h o_0, \tag{5-5}$$

with

$$\sum_{j=0}^{2m} b_j^2 = (\xi_0^2 + 2\bar{\nu} H_2)^W + o_3.$$

Another Taylor expansion in the variables (x', ξ_0, ξ') gives $A = (a_{jk}(x_0, x'', \xi'')) \in C^\infty(\mathbb{R}^n_{(x_0, x'', \xi'')}; \mathfrak{so}(n))$ and $r_j \in o_2,\ j = 0, \ldots, 2m$, such that

$$e^{-A}\begin{bmatrix} b_0 \\ \vdots \\ b_{2m} \end{bmatrix} = \begin{bmatrix} \xi_0 \\ (2\bar{\nu}\mu_1)^{\frac{1}{2}} x_1 \\ (2\bar{\nu}\mu_1)^{\frac{1}{2}} \xi_1 \\ \vdots \\ (2\bar{\nu}\mu_m)^{\frac{1}{2}} x_m \\ (2\bar{\nu}\mu_m)^{\frac{1}{2}} \xi_m \end{bmatrix} + \begin{bmatrix} r_0 \\ \vdots \\ r_{2m} \end{bmatrix}.$$

We may now set $c_A = \frac{1}{i} a_{jk}\sigma^j\sigma^k \in C^\infty(\mathbb{R}^n_{(x_0, x'', \xi'')}; i\,\mathfrak{u}(2^m))$ and compute

$$e^{i c_A^W} e^{\frac{i}{h} f_1^W} d_0^W e^{-\frac{i}{h} f_1^W} e^{-i c_A^W} = d_1^W, \tag{5-6}$$

where

$$d_1 = H_1 + \sigma^j r_j + h o_0, \tag{5-7}$$

$$H_1 := \xi_0\sigma_0 + (2\bar{\nu})^{\frac{1}{2}} \sum_{j=1}^{m} \mu_j^{\frac{1}{2}} (x_j \sigma_{2j-1} + \xi_j \sigma_{2j}). \tag{5-8}$$

5A. ***Weyl product and Koszul complexes.*** We now derive a formal Birkhoff normal form for the symbol d_1 in (5-7). First denote by $R = C^\infty(x_0, x'', \xi'')$ the ring of real-valued functions in the given $2m+1$ variables. Further define

$$S := R[\![x', \xi_0, \xi'; h]\!],$$

the ring of formal power series in the further given $2m+2$ variables with coefficients in R. The ring $S \otimes \mathbb{C}$ is now equipped with the Weyl product

$$a * b := \left[e^{\frac{ih}{2}(\partial_{r_1}\partial_{s_2} - \partial_{r_2}\partial_{s_1})}(a(s_1, r_1; h) b(s_2, r_2; h))\right]_{x = s_1 = s_2, \xi = r_1 = r_2},$$

corresponding to the composition formula (2-36) for pseudodifferential operators, with

$$[a,b] := a * b - b * a$$

being the corresponding Weyl bracket. It is an easy exercise to show that for $a, b \in S$ real-valued, the commutator $i[a,b] \in S$ is real-valued.

Next, we define a filtration on S. Each monomial $h^k \xi_0^a (x')^\alpha (\xi')^\beta$ in S is given the weight $2k + a + |\alpha| + |\beta|$. The ring S is equipped with a decreasing filtration

$$S = O_0 \supset O_1 \supset \cdots \supset O_N \supset \ldots, \qquad \bigcap_N O_N = \{0\},$$

where O_N consists of those power series with monomials of weight N or more. It is an exercise to show that

$$\begin{aligned} O_N * O_M &\subset O_{N+M}, \\ [O_N, O_M] &\subset ihO_{N+M-2}. \end{aligned}$$

The associated grading is given by

$$S = \bigoplus_{N=0}^{\infty} S_N,$$

where S_N consists of those power series with monomials of weight exactly N. We also define the quotient ring $D_N := S/O_{N+1}$ whose elements may be identified with the set of homogeneous polynomials with monomials of weight at most N. The ring D_N is also similarly graded and filtered. In a similar vein, we may also define the ring

$$S(m) = S \otimes \mathfrak{gl}_{\mathbb{C}}(2^m)$$

of $R \otimes \mathfrak{gl}_{\mathbb{C}}(2^m)$-valued formal power series in $(x', \xi_0, \xi'; h)$. The ring $S(m)$ is equipped with an induced product $*$ and decreasing filtration

$$O_0(m) \supset O_1(m) \supset \cdots \supset O_N(m) \supset \cdots, \qquad \bigcap_N O_N(m) = \{0\},$$

where $O_N(m) = O_N \otimes \mathfrak{gl}_{\mathbb{C}}(2^m)$. It is again a straightforward exercise to show that for $a, b \in S \otimes i\mathfrak{u}_{\mathbb{C}}(2^m)$ self-adjoint, the commutator $i[a,b] \in S \otimes i\mathfrak{u}_{\mathbb{C}}(2^m)$ is self-adjoint.

5A1. *Koszul complexes.* Let us now again consider the $2m$ and $(2m+1)$-dimensional real inner product spaces $V = \mathbb{R}[e_1, \ldots, e_{2m}]$ and $W = \mathbb{R}[e_0] \oplus V$ from Section 2B. Considering the chain groups $D_N \otimes \Lambda^k V$, $k = 0, 1, \ldots, n$, one may define four differentials

$$\begin{aligned} w_x^0 &= \sum_{j=1}^m \mu_j^{\frac{1}{2}} (x_j e_{2j-1} \wedge + \xi_j e_{2j} \wedge), & i_x^0 &= \sum_{j=1}^m \mu_j^{\frac{1}{2}} (x_j i_{e_{2j-1}} + \xi_j i_{e_{2j}}), \\ w_\partial^0 &= \sum_{j=1}^m \mu_j^{\frac{1}{2}} (\partial_{x_j} e_{2j-1} \wedge + \partial_{\xi_j} e_{2j} \wedge), & i_\partial^0 &= \sum_{j=1}^m \mu_j^{\frac{1}{2}} (\partial_{x_j} i_{e_{2j-1}} + \partial_{\xi_j} i_{e_{2j}}). \end{aligned}$$

We equip D_N with the $R[\![h]\!]$-valued inner products, where the distinct monomials

$$\frac{1}{\sqrt{a!\alpha!\beta!}}\xi_0^a(x')^\alpha(\xi')^\beta$$

are orthonormal. With these inner products w_x^0, i_∂^0 and w_∂^0, i_x^0 are respectively adjoints. The combinatorial Laplacians $\Delta^0 = w_x^0 i_\partial^0 + i_\partial^0 w_x^0 = w_\partial^0 i_x^0 + i_x^0 w_\partial^0$ are computed to be equal and act on basis elements $\xi_0^a(x')^\alpha(\xi')^\beta\left(\bigwedge e_j^{\gamma_j}\right)$ via multiplication by $\mu.(2(\alpha+\beta)+\gamma)$. It now follows that these have (co-)homology only in degree zero given by $R[\![h]\!]$.

Similarly, we may consider the chain groups $D_N \otimes \Lambda^k W$, $k = 0, 1, \dots, n$; one may define four differentials

$$\begin{aligned} w_x &= \xi_0 e_0 \wedge + (2\bar{\nu})^{\frac{1}{2}} w_x^0, & i_x &= \xi_0 i_{e_0} + (2\bar{\nu})^{\frac{1}{2}} i_x^0, \\ w_\partial &= \partial_{\xi_0} e_0 \wedge + (2\bar{\nu})^{\frac{1}{2}} w_\partial^0, & i_\partial &= \partial_{\xi_0} i_{e_0} + (2\bar{\nu})^{\frac{1}{2}} i_\partial^0. \end{aligned}$$

Again these complexes have cohomology only in degree zero given by $R[\![h]\!]$.

Next, we define twisted Koszul differentials on $D_N \otimes \Lambda^k V$ via

$$\begin{aligned} \tilde{w}_\partial^0 &= \frac{i}{h}\sum_{j=1}^m \mu_j^{\frac{1}{2}}(\mathrm{ad}_{x_j} e_{2j-1}\wedge + \mathrm{ad}_{\xi_j} e_{2j}\wedge) = \sum_{j=1}^m \mu_j^{\frac{1}{2}}(\partial_{x_j} e_{2j}\wedge - \partial_{\xi_j} e_{2j-1}\wedge), \\ \tilde{i}_\partial^0 &= \frac{i}{h}\sum_{j=1}^m \mu_j^{\frac{1}{2}}(\mathrm{ad}_{x_j} i_{e_{2j-1}} + \mathrm{ad}_{\xi_j} i_{e_{2j}}) = \sum_{j=1}^m \mu_j^{\frac{1}{2}}(\partial_{x_j} i_{e_{2j}} - \partial_{\xi_j} i_{e_{2j-1}}). \end{aligned}$$

We note that the above are symplectic adjoints to their untwisted counterparts with respect to the symplectic pairing $\sum_{j=1}^m e_{2j-1}\wedge e_{2j}$ on V.

Similar twisted Koszul differentials on $D_N \otimes \Lambda^k W$ are defined via

$$\begin{aligned} \tilde{w}_\partial &= \frac{i}{h}\mathrm{ad}_{\xi_0} e_0 \wedge + (2\bar{\nu})^{\frac{1}{2}} \tilde{w}_\partial^0 = -\partial_{x_0} e_0 \wedge + (2\bar{\nu})^{\frac{1}{2}} \tilde{w}_\partial^0, \\ \tilde{i}_\partial &= \frac{i}{h} i_{e_0}\mathrm{ad}_{\xi_0} + (2\bar{\nu})^{\frac{1}{2}} \tilde{i}_\partial^0 = -\partial_{x_0} i_{e_0} + (2\bar{\nu})^{\frac{1}{2}} \tilde{i}_\partial^0. \end{aligned}$$

These twisted differentials correspond to the untwisted ones by a mere change of basis in V, W and hence also have (co-)homology only in degree zero given by $R[\![h]\!]$.

We now compute the twisted combinatorial Laplacian to be

$$\tilde{\Delta}^0 = \tilde{w}_\partial^0 i_x^0 + i_x^0 \tilde{w}_\partial^0 = -(w_x^0 \tilde{i}_\partial^0 + \tilde{i}_\partial^0 w_x^0) = \sum_{j=1}^m \mu_j [\xi_j \partial_{x_j} - x_j \partial_{\xi_j} + e_{2j} i_{e_{2j-1}} - e_{2j-1} i_{e_{2j}}].$$

One may similarly define $\tilde{\Delta} = \tilde{w}_\partial i_x + i_x \tilde{w}_\partial$. Next, we define the spaces of twisted $\tilde{\Delta}^0$-harmonic, ξ_0-independent elements

$$\begin{aligned} \mathcal{H}_N^k &= \left\{\omega \in D_N \otimes \Lambda^k W \;\middle|\; \tilde{\Delta}^0\omega = 0,\ \partial_{\xi_0}\omega = 0\right\}, \\ \mathcal{H}^k &= \left\{\omega \in S \otimes \Lambda^k W \;\middle|\; \tilde{\Delta}^0\omega = 0,\ \partial_{\xi_0}\omega = 0\right\}. \end{aligned}$$

We now prove a twisted version of the Hodge decomposition theorem.

Lemma 5.1. *The k-th chain group is spanned by three subspaces*:

$$D_N \otimes \Lambda^k W = \mathbb{R}[\mathrm{Im}(i_x \tilde{w}_\partial), \mathrm{Im}(\tilde{w}_\partial i_x), \mathcal{H}_N^k].$$

Proof. We first compute $\tilde{\Delta}$ in terms of $\tilde{\Delta}^0$ to be

$$\tilde{\Delta} = -\xi_0 \partial_{x_0} + 2\bar{\nu}\tilde{\Delta}^0 - 2(\partial_{x_0}\bar{\nu}^{\frac{1}{2}})e_0 i_x^0.$$

Next, since $\tilde{\Delta}^0$ is skew-adjoint, we may decompose

$$D_N \otimes \Lambda^k W = E_0 \oplus \bigoplus_{\lambda>0}[E_{i\lambda} \oplus E_{-i\lambda}]$$

into its eigenspaces. Following $[\tilde{\Delta}^0, \bar{\nu}] = 0$ we may now invert $\tilde{\Delta}$ on the nonzero eigenspaces of $\tilde{\Delta}^0$ above using the Volterra series:

$$\tilde{\Delta}^{-1} = (2\bar{\nu}\tilde{\Delta}^0)^{-1} \sum_{j=0}^{\infty} \big[(2\bar{\nu}\tilde{\Delta}^0)^{-1}(\xi_0\partial_{x_0} + 2(\partial_{x_0}\bar{\nu}^{\frac{1}{2}})e_0 i_x^0)\big]^j.$$

The sum above is finite since $\xi_0\partial_{x_0} + 2(\partial_{x_0}\bar{\nu}^{\frac{1}{2}})e_0 i_x^0$ is nilpotent on $D_N \otimes \Lambda^k W$. Thus we have

$$\bigoplus_{\lambda>0}[E_{i\lambda} \oplus E_{-i\lambda}] \subset \mathrm{Im}(\tilde{\Delta}) \subset \mathbb{R}\big[\mathrm{Im}(i_x \tilde{w}_\partial), \mathrm{Im}(\tilde{w}_\partial i_x)\big].$$

Finally, we have the decomposition

$$E_0 = \bigoplus_{j=0}^{N} \xi_0^j \mathcal{H}_N^k$$

and we write each $\omega \in \xi_0^j \mathcal{H}_N^k$, $j \geq 1$, as

$$\omega = \omega_0 + \tilde{\Delta}\omega_1,$$

where

$$\omega_0 = \left[-2(\partial_{x_0}\bar{\nu}^{\frac{1}{2}})e_0 i_x^0 \xi_0^{-1}\int_0^{x_0}\right]^j \omega \in \mathcal{H}_N^k,$$

$$\omega_1 = -\left(\xi_0^{-1}\int_0^{x_0}\right)\sum_{l=0}^{j-1}\left[-2(\partial_{x_0}\bar{\nu}^{\frac{1}{2}})e_0 i_x^0 \xi_0^{-1}\int_0^{x_0}\right]^l \omega,$$

to complete the proof. □

5B. ***Formal Birkhoff normal form.*** The importance of the Koszul complexes introduced in the previous subsection is in continuing the Birkhoff normal form procedure for the symbol d_1 in (5-7). The remaining steps in the procedure are formal.

First let us define the Clifford quantization of an element in $a \in S \otimes \Lambda^k W$ using (2-8) as an element in

$$c_0(a) := i^{\frac{k(k+1)}{2}} c(a) \in S(m).$$

It is clear from (2-10) and (2-11) this gives an isomorphism

$$c_0 : S \otimes \Lambda^{\text{odd/even}} W \to S \otimes i\,\mathfrak{u}_{\mathbb{C}}(2^m) \tag{5-9}$$

of real elements of the even or odd exterior algebra with self-adjoint elements in $S(m)$. It is clear from (5-7) that

$$d_1 = H_1 + c_0(r) + hS \otimes i\,\mathfrak{u}_{\mathbb{C}}(2^m) \tag{5-10}$$

for $r := \sum_{j=1}^{n} r_j e_j \in O_2 \otimes W$.

For $a \in \Lambda^k W$, we define $[a] := \left[\frac{k}{2}\right]$. Now for $f \in O_N,\ N \geq 3$, and $a \in O_N \otimes \Lambda^{\text{even}} W,\ N \geq 1$, we may compute the conjugations

$$e^{\frac{i}{h}f} H_1 e^{-\frac{i}{h}f} = H_1 + c_0(\tilde{w}_\partial f) + O_N \otimes i\,\mathfrak{u}_{\mathbb{C}}(2^m), \tag{5-11}$$

$$e^{ic_0(a)} H_1 e^{-ic_0(a)} = H_1 + (-1)^{[a]+1} 2c_0(i_x a) + hc_0(\tilde{w}_\partial a) + O_{N+2} \otimes i\,\mathfrak{u}_{\mathbb{C}}(2^m) \tag{5-12}$$

in terms of the Koszul differentials.

We now come to the formal Birkhoff normal form for the symbol d_1.

Proposition 5.2. *There exist $f \in O_3$, $a \in O_2 \otimes \Lambda^{even} W$ and $\omega \in \mathcal{H}^{odd} \cap O_2$ such that*

$$e^{ic_0(a)} e^{\frac{i}{h}f} d_1 e^{-\frac{i}{h}f} e^{-ic_0(a)} = H_1 + c_0(\omega). \tag{5-13}$$

Proof. We first prove that for each $N \geq 1$, there exist $f_N \in O_3$, $a_N^0 \in O_1 \otimes \Lambda^2 W$, $\omega_N^0 \in \mathcal{H}^1 \cap O_2$ and $r_N^0 \in O_{N+1} \otimes W$ such that

$$\begin{gathered} e^{ic_0(a_N^0)} e^{\frac{i}{h}f_N} d_1 e^{-\frac{i}{h}f_N} e^{-ic_0(a_N^0)} = H_1 + c_0(\omega_N^0) + c_0(r_N^0) + hS \otimes i\,\mathfrak{u}_{\mathbb{C}}(2^m), \\ f_{N+1} - f_N \in O_{N+2}, \quad a_{N+1}^0 - a_N^0 \in O_N, \quad \omega_{N+1}^0 - \omega_N^0 \in O_{N+1}. \end{gathered} \tag{5-14}$$

The base case $N = 1$ is given by (5-10) with $a_1^0 = f_1 = \omega_1^0 = 0$ and $r_1^0 = r$. To complete the induction step we decompose r_N^0 as

$$r_N^0 = \underbrace{u_N^0}_{\in S_{N+1} \otimes W} + \underbrace{r_{N+1}^0}_{\in O_{N+2} \otimes W}. \tag{5-15}$$

Next we use Lemma 5.1 to find $b_N, g_N \in O_{N+1} \otimes W$ and $\upsilon_N^0 \in \mathcal{H}^1 \cap S_{N+1}$ such that

$$u_N^0 = \upsilon_N^0 - i_x \tilde{w}_\partial b_N^0 - \tilde{w}_\partial i_x g_N^0 + O_{N+2}. \tag{5-16}$$

Next, define $f_{N+1} = f_N + i_x g_N^0 \in O_3$, $a_{N+1}^0 = a_N^0 + \frac{1}{2}\tilde{w}_\partial b_N^0 \in O_1 \otimes \Lambda^2 W$ and $\omega_{N+1}^0 = \omega_N^0 + \upsilon_N^0$. We now use (5-11), (5-12), (5-15) and (5-16) to compute

$$\begin{aligned} &e^{ic_0(a_{N+1}^0)} e^{\frac{i}{h}f_{N+1}} d_1 e^{-\frac{i}{h}f_{N+1}} e^{-ic_0(a_{N+1}^0)} \\ &\qquad = e^{ic_0\left(\frac{1}{2}\tilde{w}_\partial b_N^0\right)} e^{\frac{i}{h}i_x g_N^0} H_1 e^{-\frac{i}{h}i_x g_N^0} e^{-ic_0\left(\frac{1}{2}\tilde{w}_\partial b_N^0\right)} + c_0(\omega_N^0) + c_0(r_N^0) + hS \otimes i\,\mathfrak{u}_{\mathbb{C}}(2^m) \\ &\qquad = H_1 + c_0(\omega_{N+1}^0) + c_0(r_{N+1}^0) + hS \otimes i\,\mathfrak{u}_{\mathbb{C}}(2^m), \end{aligned}$$

completing the induction step. Now setting $f=\lim_{N\to\infty} f_N$, $a_0=\lim_{N\to\infty} a_N^0$ and $\omega_0=\lim_{N\to\infty}\omega_N^0$ and letting $N\to\infty$ in (5-14) gives the relation

$$e^{ic_0(a_0)}e^{\frac{i}{h}f}d_1e^{-\frac{i}{h}f}e^{-ic_0(a_0)}=H_1+c_0(\omega_0)+hS\otimes i\mathfrak{u}_{\mathbb{C}}(2^m). \tag{5-17}$$

Next we claim that for each $N\geq 0$, there exist $a_N\in O_1\otimes\Lambda^{\text{even}}W$, $\omega_N\in\mathcal{H}^*\cap O_2$ such that

$$\begin{gathered} e^{ic_0(a_N)}e^{\frac{i}{h}f}d_1e^{-\frac{i}{h}f}e^{-ic_0(a_N)}=H_1+c_0(\omega_N)+hO_N\otimes i\mathfrak{u}_{\mathbb{C}}(2^m),\\ a_{N+1}-a_N\in O_{N+1}\otimes\Lambda^{\text{even}}W,\quad \omega_{N+1}-\omega_N\in\mathcal{H}^{\text{odd}}\cap O_N. \end{gathered} \tag{5-18}$$

The base case $N=0$ is now provided by (5-17). To complete the induction step, we use the isomorphism (5-9) to decompose the remainder term in (5-18) above as

$$c_0(u_N)+ihO_{N+1}\otimes\mathfrak{u}_{\mathbb{C}}(2^m)$$

for $u_N\in S_N\otimes\Lambda^{\text{odd}}W$. Next we use Lemma 5.1 to find $b_N,g_N\in O_N\otimes\Lambda^{\text{odd}}W$ and $\upsilon_N\in\mathcal{H}^{\text{odd}}\cap S_N$ such that

$$u_N=\upsilon_N-i_x\tilde{w}_\partial b_N-\tilde{w}_\partial i_xg_N+O_{N+1}. \tag{5-19}$$

Now define $a_{N+1}=a_N+i_xg_N+\frac{1}{2}h(-1)^{[b_N]}\tilde{w}_\partial b_N\in O_1$ and $\omega_{N+1}=\omega_N+\upsilon_N$. We now use (5-11), (5-12), (5-15) and (5-19) to compute

$$e^{ic_0(a_{N+1})}e^{\frac{i}{h}f}d_1e^{-\frac{i}{h}f}e^{-ic_0(a_{N+1})}=H_1+c_0(\omega_{N+1})+ihO_{N+1}\otimes\mathfrak{u}_{\mathbb{C}}(2^m),$$

completing the induction step. Now setting $a=\lim_{N\to\infty}a_N$ and $\omega=\lim_{N\to\infty}\omega_N$ and letting $N\to\infty$ in (5-18) gives the proposition. $\square$

Finally, we show how the Birkhoff normal form maybe used to perform a further reduction on the trace. First note that we may similarly use (2-8) to define a self-adjoint Clifford–Weyl quantization map

$$c_0^W:=\operatorname{Op}\otimes c_0:S^0_{\text{cl}}(\mathbb{R}^{2n};\mathbb{C})\otimes\Lambda^{\text{odd/even}}W\to\Psi^0_{\text{cl}}(\mathbb{R}^n;\mathbb{C}^{2^m}),$$

which maps real-valued symbols $S^0_{\text{cl}}(\mathbb{R}^{2n};\mathbb{R})\otimes\Lambda^{\text{odd/even}}W$ to self-adjoint operators in $\Psi^0_{\text{cl}}(\mathbb{R}^n;\mathbb{C}^{2^m})$. Similarly we define a space of real-valued, twisted $\tilde{\Delta}^0$-harmonic, ξ_0- independent symbols

$$\mathcal{H}^kS^0_{\text{cl}}:=\left\{\omega\in S^0_{\text{cl}}(\mathbb{R}^{2n};\mathbb{R})\otimes\Lambda^kW\;\middle|\;\tilde{\Delta}^0\omega=0,\ \partial_{\xi_0}\omega=0\right\}.$$

Next, an application of Borel's lemma by virtue of (5-1), (5-6) and (5-13) gives the existence of

$$\begin{aligned} \bar{a}&\sim\sum_{j=0}^{\infty}h^j\bar{a}_j\in S^0_{\text{cl}}(\mathbb{R}^{2n};\mathbb{R})\otimes\Lambda^{\text{odd}}W, & \bar{f}&\sim\sum_{j=0}^{\infty}h^j\bar{f}_j\in S^0_{\text{cl}}(\mathbb{R}^{2n};\mathbb{R}),\\ \bar{r}&\sim\sum_{j=0}^{\infty}h^j\bar{r}_j\in S^0_{\text{cl}}(\mathbb{R}^{2n};\mathbb{R})\otimes\Lambda^{\text{odd}}W, & \bar{\omega}&\sim\sum_{j=0}^{\infty}h^j\bar{\omega}_j\in\mathcal{H}^{\text{odd}}S^0_{\text{cl}} \end{aligned}$$

such that

$$e^{ic_0^W(\bar{a})}e^{\frac{i}{h}\bar{f}^W}d_0^We^{-\frac{i}{h}\bar{f}^W}e^{-ic_0^W(\bar{a})}=\underbrace{H_1^W+c_0^W(\bar{\omega})}_{:=\bar{D}}+c_0^W(\bar{r}) \tag{5-20}$$

on $\overline{V}_{\alpha\beta} := e^{X_{\bar{f}_0}}(V^0_{\alpha\beta})$. Here $\{\bar{r}_j\}_{j\in\mathbb{N}_0}$, $\bar{f}_0$, $\bar{\omega}_0$ vanish to infinite, second and second order respectively along

$$\Sigma_0^{D_0} = \Sigma_0^{\bar{D}} = \Sigma_0^{\bar{D}+c_0^W(\bar{r})} = \{\xi_0 = x' = \xi' = 0\}.$$

Note that on account of (4-5) and (5-4) one again has

$$\nu_0 = \mu_1 \min_{x\in X} \nu(x) \le \mu_1 \inf_{\mathbb{R}^n_{x_0,x'',\xi''}} \bar{\nu}.$$

Furthermore, since $\bar{\omega}_0$ vanishes to second order we may choose $\bar{\omega}_0$ arbitrarily small satisfying the estimate

$$\|\bar{\omega}_0\|_{C^1} < \varepsilon \tag{5-21}$$

for any $\varepsilon > 0$, while still satisfying (5-20).

We note that $\bar{D} \in \Psi^1_{\mathrm{cl}}(\mathbb{R}^n; \mathbb{C}^{2^m})$, with $\bar{D} + i$ having an elliptic symbol in the class $S^0(\langle \xi_0, \xi'\rangle)$, and is hence essentially self-adjoint as an unbounded operator on $L^2(\mathbb{R}^n; \mathbb{C}^{2^m})$. The domain of its unique self-adjoint extension is $H^1(\mathbb{R}_{x_0}) \otimes L^2(\mathbb{R}^{n-1}_{x',x''}; \mathbb{C}^{2^m})$; see [Dimassi and Sjöstrand 1999, Chapter 8]. We now set

$$\bar{A}_\alpha := e^{ic_0^W(\bar{a})} e^{\frac{i}{h}\bar{f}^W} A^0_\alpha e^{-\frac{i}{h}\bar{f}^W} e^{-ic_0^W(\bar{a})}, \tag{5-22}$$

$$\begin{aligned} \mathcal{T}^{\vartheta}_{\alpha\beta}(\bar{D}) &:= \operatorname{tr}\left[\bar{A}_\alpha f\left(\frac{\bar{D}}{\sqrt{h}}\right)\check{\vartheta}\left(\frac{\lambda\sqrt{h}-\bar{D}}{h}\right)\bar{A}_\beta\right] \\ &= \frac{1}{\pi}\int_{\mathbb{C}} \bar{\partial}\tilde{f}(z)\check{\vartheta}\left(\frac{\lambda - z}{\sqrt{h}}\right)\operatorname{tr}\left[\bar{A}_\alpha\left(\frac{1}{\sqrt{h}}\bar{D} - z\right)^{-1}\bar{A}_\beta\right] dz\, d\bar{z}. \end{aligned} \tag{5-23}$$

Proposition 5.3. *We have*

$$\mathcal{T}^{\vartheta}_{\alpha\beta}(D_0) = \mathcal{T}^{\vartheta}_{\alpha\beta}(\bar{D}) \bmod h^\infty.$$

Proof. Since the conjugations in (5-1) and (5-20) are unitary and $\mathrm{WF}(\bar{A}_\alpha), \mathrm{WF}(\bar{A}_\beta) \subset \overline{V}_{\alpha\beta}$, we have

$$\mathcal{T}^{\vartheta}_{\alpha\beta}(D_0) = \frac{1}{\pi}\int_{\mathbb{C}} \bar{\partial}\tilde{f}(z)\check{\vartheta}\left(\frac{\lambda - z}{\sqrt{h}}\right)\operatorname{tr}\left[\bar{A}_\alpha\left(\frac{1}{\sqrt{h}}(\bar{D} + c_0^W(\bar{r})) - z\right)^{-1}\bar{A}_\beta\right] dz\, d\bar{z}.$$

It now remains to do away with the $c_0^W(\bar{r})$ above. Since this term vanishes to infinite order along $\Sigma_0^{\bar{D}} = \Sigma_0^{\bar{D}+c_0^W(\bar{r})}$, we may use symbolic calculus to find $P_N, Q_N \in \Psi^0_{\mathrm{cl}}(\mathbb{R}^n; \mathbb{C}^{2^m})$, for all $N \ge 1$, such that

$$c_0^W(\bar{r}) = P_N(\bar{D} + c_0^W(\bar{r}))^N, \tag{5-24}$$

$$c_0^W(\bar{r}) = Q_N(\bar{D})^N. \tag{5-25}$$

Modifying $\bar{D}$ outside a neighborhood of $\overline{V}_{\alpha\beta}$ using Lemma 3.3 and Proposition A.6 we may assume that $\bar{D}$, $\bar{D}+c_0^W(\bar{r})$ have discrete spectrum in $(-\sqrt{2\nu_0}, \sqrt{2\nu_0})$ and hence

$$\mathcal{T}^{\vartheta}_{\alpha\beta}(\bar{D}) = \operatorname{tr}\left[\bar{A}_\alpha f\left(\frac{\bar{D}}{\sqrt{h}}\right)\check{\vartheta}\left(\frac{\lambda\sqrt{h}-\bar{D}}{h}\right)\bar{A}_\beta\right],$$

$$\mathcal{T}^{\vartheta}_{\alpha\beta}(D_0) = \operatorname{tr}\left[\bar{A}_\alpha f\left(\frac{\bar{D}+c_0^W(\bar{r})}{\sqrt{h}}\right)\check{\vartheta}\left(\frac{\lambda\sqrt{h}-\bar{D}-c_0^W(\bar{r})}{h}\right)\bar{A}_\beta\right].$$

Next, with $\Pi^{\bar{D}} = \Pi^{\bar{D}}_{[-\sqrt{2\nu_0 h},\sqrt{2\nu_0 h}]}$ and $\Pi^{\bar{D}+c_0^W(\bar{r})} = \Pi^{\bar{D}+c_0^W(\bar{r})}_{[-\sqrt{2\nu_0 h},\sqrt{2\nu_0 h}]}$ denoting the spectral projections, (5-24) and (5-25) give

$$\|c_0^W(\bar{r})\Pi^{\bar{D}}\| = O(h^{\frac{N}{2}}), \quad \|c_0^W(\bar{r})\Pi^{\bar{D}+c_0^W(\bar{r})}\| = O(h^{\frac{N}{2}})$$

for each $N \geq 1$. Finally applying Proposition A.5 with

$$\rho(x) = f\left(\frac{x}{\sqrt{h}}\right)\check{\vartheta}\left(\frac{\lambda\sqrt{h}-x}{h}\right)$$

and using the cyclicity of the trace gives $\mathcal{T}^{\vartheta}_{\alpha\beta}(D_0) - \mathcal{T}^{\vartheta}_{\alpha\beta}(\bar{D}) = O(h^{-1}h^{\frac{N}{4096}})$, for all $N \geq 1$, completing the proof. □

6. Extension of a resolvent

In this section we complete the proof of Lemma 3.1. On account of the reductions in Propositions 4.1 and 5.3 in the previous sections, it suffices to now consider the trace $\mathcal{T}^{\vartheta}_{\alpha\beta}(\bar{D})$. First let $\bar{A}_\alpha = a_\alpha^W$, $\bar{A}_\beta = a_\beta^W$ for $a_\alpha, a_\beta \in S^0_{\mathrm{cl}}(\mathbb{R}^{2n})$. The conjugations

$$e^{\frac{it}{h}x_0}\bar{A}_\alpha e^{-\frac{it}{h}x_0} = a^W_{\alpha,t} \quad \text{and} \quad e^{\frac{it}{h}x_0}\bar{A}_\beta e^{-\frac{it}{h}x_0} = a^W_{\beta,t}$$

are easily computed in terms of the one-parameter family of symbols $a_{\alpha,t}(\xi_0,\dots) = a_\alpha(\xi_0+t,\dots)$, $a_{\beta,t} = a_\beta(\xi_0+t,\dots) \in S^0_{\mathrm{cl}}(\mathbb{R}^{2n})$, $t \in \mathbb{R}$, obtained by translating in the ξ_0-direction. One now introduces almost analytic continuations of the symbols $a_{\alpha,t}, a_{\beta,t} \in S^0_{\mathrm{cl}}(\mathbb{R}^{2n})$, defined for $t \in \mathbb{C}$, such that all the Fréchet seminorms of $\bar{\partial}a_{\alpha,t}$, $\bar{\partial}a_{\beta,t}$ are $O(|\operatorname{Im} t|^\infty)$. These may be further chosen to have the property that the wavefront sets of their quantizations have uniform compact support when t is restricted to compact subsets of $\mathbb{C}$. Again one clearly has

$$a^W_{\alpha,t} = e^{\frac{i\operatorname{Re}t}{h}x_0}(a_{\alpha,i\operatorname{Im}t})^W e^{-\frac{i\operatorname{Re}t}{h}x_0}, \tag{6-1}$$

$$a^W_{\beta,t} = e^{\frac{i\operatorname{Re}t}{h}x_0}(a_{\beta,i\operatorname{Im}t})^W e^{-\frac{i\operatorname{Re}t}{h}x_0}. \tag{6-2}$$

In similar vein we may define

$$\bar{D}_t := e^{-\frac{it}{h}x_0}\bar{D}e^{\frac{it}{h}x_0} = H^W_{1,t} + c_0^W(\bar{\omega}), \tag{6-3}$$

$$H_{1,t} = (\xi_0+t)\sigma_0 + (2\bar{\nu})^{\frac{1}{2}}\sum_{j=1}^{m}\mu_j^{\frac{1}{2}}(x_j\sigma_{2j-1}+\xi_j\sigma_{2j}) \in S^1_{\mathrm{cl}}(\mathbb{R}^{2n}) \tag{6-4}$$

for $t \in \mathbb{R}$, on account of the ξ_0-independence of $\bar{\omega}$. An almost analytic continuation of $\bar{D}_t$ is easily introduced by simply allowing $t \in \mathbb{C}$ to be complex in (6-4) above. The resolvent $(\bar{D}_t - z)^{-1} : L^2(\mathbb{R}^n; \mathbb{C}^{2^m}) \to L^2(\mathbb{R}^n; \mathbb{C}^{2^m})$ is well defined and holomorphic in the region $\operatorname{Im} z > |\operatorname{Im} t|$.

In the lemma below we set $t = i\gamma(M, \delta) := i2Mh^{\delta} \log \frac{1}{h}$, for $\delta = 1 - \varepsilon \in \left(\frac{1}{2}, 1\right)$ with ε as in Lemma 3.1 and $M > 1$. We now have the following.

Lemma 6.1. *For h sufficiently small and for all $\varepsilon_0 > 0$, the resolvent*

$$\left(\frac{1}{\sqrt{h}} \bar{D}_{i\gamma} - z\right)^{-1} : L^2(\mathbb{R}^n; \mathbb{C}^{2^m}) \to L^2(\mathbb{R}^n; \mathbb{C}^{2^m})$$

extends holomorphically, and is uniformly $O(h^{-\frac{1}{2}})$, in the region $\operatorname{Im} z > -Mh^{\delta - \frac{1}{2}} \log \frac{1}{h}$, $|\operatorname{Re} z| \leq \sqrt{2\nu_0} - \varepsilon_0$.

Proof. We begin with the orthogonal Landau decomposition (2-31)

$$L^2(\mathbb{R}^n; \mathbb{C}^{2^m}) = L^2(\mathbb{R}^{m+1}_{x_0, x''}) \otimes \underbrace{\left(\mathbb{C}[\psi_{0,0}] \oplus \bigoplus_{\Lambda \in \mu.(\mathbb{N}_0^m \setminus 0)} [E_\Lambda^{\text{even}} \oplus E_\Lambda^{\text{odd}}]\right)}_{= L^2(\mathbb{R}^m_{x'}; \mathbb{C}^{2^m})}, \tag{6-5}$$

where

$$E_\Lambda^{\text{even}} := \bigoplus_{\substack{\tau \in \mathbb{N}_0^m \setminus 0 \\ \Lambda = \mu.\tau}} E_\tau^{\text{even}}, \quad E_\Lambda^{\text{odd}} := \bigoplus_{\substack{\tau \in \mathbb{N}_0^m \setminus 0 \\ \Lambda = \mu.\tau}} E_\tau^{\text{odd}} \tag{6-6}$$

according to the eigenspaces of the squared magnetic Dirac operator $D^2_{\mathbb{R}^m}$ (2-21) on $\mathbb{R}^m$. It is clear from (6-4) that

$$H^W_{1,t} = (\xi_0 + t)^W \sigma_0 + [(2\bar{\nu})^{\frac{1}{2}}]^W \otimes D_{\mathbb{R}^m}$$

in terms of the above decomposition. Furthermore one has the commutation relations

$$\begin{aligned} [\sigma_0, D^2_{\mathbb{R}^m}] &= 0, \\ [c_0^W(\bar{\omega}), D^2_{\mathbb{R}^m}] &= ihc_0^W(\tilde{\Delta}^0 \bar{\omega}) = 0, \end{aligned}$$

since $\bar{\omega}$ is $\tilde{\Delta}^0$-harmonic. The above and (6-3) show that the $\left(\frac{1}{\sqrt{h}} \bar{D}_t - z\right)$ preserves the eigenspaces in the decomposition (6-5) for all $t \in \mathbb{C}$. It hence suffices to consider the restriction of $\left(\frac{1}{\sqrt{h}} \bar{D}_{i\gamma} - z\right)$ to each eigenspace.

Let $E_0 := \mathbb{C}[\psi_{0,0}]$, $E_\Lambda := E_\Lambda^{\text{even}} \oplus E_\Lambda^{\text{odd}}$ and $\mathrm{P}_0, \mathrm{P}_\Lambda$ denote the projection onto the corresponding summands of (6-5). Define the restrictions

$$\begin{aligned} \Omega_0 &:= \mathrm{P}_0 c_0^W(\bar{\omega}) \mathrm{P}_0 : L^2(\mathbb{R}^{m+1}_{x_0, x''}) \to L^2(\mathbb{R}^{m+1}_{x_0, x''}), \\ \Omega_\Lambda &:= \mathrm{P}_\Lambda c_0^W(\bar{\omega}) \mathrm{P}_\Lambda : L^2(\mathbb{R}^{m+1}_{x_0, x''}; E_\Lambda^{\text{even}} \oplus E_\Lambda^{\text{odd}}) \to L^2(\mathbb{R}^{m+1}_{x_0, x''}; E_\Lambda^{\text{even}} \oplus E_\Lambda^{\text{odd}}), \quad \Lambda > 0. \end{aligned}$$

Now $\bar{\omega} \sim \sum_{j=0}^{\infty} h^j \bar{\omega}_j \in \mathcal{H}^{\text{odd}} S^0_{\text{cl}}$ with ξ_0-independent $\bar{\omega}_0$ vanishing to second order along $\Sigma_0^{D_0} = \Sigma_0^{\bar{D}} = \{\xi_0 = x' = \xi' = 0\}$. Hence we may decompose

$$\bar{\omega}_0 = \sum_{i \le j} [a_{ij} z_i z_j + \bar{a}_{ij} \bar{z}_i \bar{z}_j + b_{ij} \bar{z}_i z_j + \bar{b}_{ij} z_i \bar{z}_j]$$

in terms of the complex coordinates $z_j = x_j + i\xi_j$, $\bar{z}_j = x_j - i\xi_j$, $1 \le j \le m$, with $a_{ij}, b_{ij} \in S^0_{\text{cl}}(\mathbb{R}^{2n}; \mathbb{R}) \otimes \Lambda^{\text{odd}} W$. The self-adjoint Clifford–Weyl quantization now yields

$$c_0^W(\bar{\omega}_0) = \sum_{i \le j} \big[c_0^W(a_{ij}) A_i A_j + A_j^* A_i^* c_0^W(\bar{a}_{ij}) + c_0^W(b_{ij}) A_i^* A_j + A_j^* A_i c_0^W(\bar{b}_{ij})\big] + h \Psi^0_{\text{cl}}(\mathbb{R}^n; \mathbb{C}^{2^m})$$

in terms of the raising and lowering operators in (2-26). Since each lowering operator A_j annihilates $\psi_{0,0}$, this leads to the estimate

$$\|\Omega_0\| = O(h). \tag{6-7}$$

Next, on account of (5-21) one may also expand $\bar{\omega}_0 = \sum_{j=1}^{m} [a_j z_j + \bar{a}_j \bar{z}_j]$, with $a_j \in S^0_{\text{cl}}(\mathbb{R}^{2n}; \mathbb{R}) \otimes \Lambda^{\text{odd}} W$, satisfying $\|a_j\|_{C^0} \le \varepsilon < 1$. On self-adjoint quantization this now gives

$$c_0^W(\bar{\omega}_0) = \sum_{j=1}^{m} \big[c_0^W(a_j) A_j + A_j^* c_0^W(\bar{a}_j)\big] + h \Psi^0_{\text{cl}}(\mathbb{R}^n; \mathbb{C}^{2^m}),$$

where

$$\|c_0^W(a_j)\|_{L^2 \to L^2}, \ \|c_0^W(\bar{a}_j)\|_{L^2 \to L^2} = \|a_j\|_{C^0} + O(h) \le \varepsilon + O(h).$$

Knowing the action of the lowering and raising operators A_j, A_j^* on each eigenstate (2-25) of $D^2_{\mathbb{R}^m}$ gives the estimate

$$\|\Omega_\Lambda\| \le \varepsilon \sqrt{\Lambda h} + O(h), \tag{6-8}$$

with the $O(h)$ term above being uniform in Λ.

Next we compute the restriction of $\big(\frac{1}{\sqrt{h}} \bar{D}_{i\gamma} - z\big)$ to the E_0 eigenspace in (6-5) using (2-6) to be

$$D_{i\gamma,0}(z) := \mathrm{P}_0 \bigg(\frac{1}{\sqrt{h}} \bar{D}_{i\gamma} - z \bigg) \mathrm{P}_0 = \frac{1}{\sqrt{h}} [-\xi_0 - i\gamma - z\sqrt{h} + \Omega_0]. \tag{6-9}$$

The above is again understood as a closed unbounded operator on $L^2(\mathbb{R}^{m+1}_{x_0, x''})$ with domain $H^1(\mathbb{R}_{x_0}) \otimes L^2(\mathbb{R}^m_{x''})$. Set $R_{i\gamma,0}(z) = [r_{i\gamma,0}(z)]^W$, with

$$r_{i\gamma,0}(z) = \frac{\sqrt{h}}{-\xi_0 - i\gamma - z\sqrt{h}},$$

which is well defined for $\operatorname{Im} z > -\gamma/(2\sqrt{h}) = -M h^{\delta - \frac{1}{2}} \log \frac{1}{h}$, and compute

$$R_{i\gamma,0}(z) D_{i\gamma,0}(z) = I + O(h^{1-\delta}),$$
$$D_{i\gamma,0}(z) R_{i\gamma,0}(z) = I + O(h^{1-\delta})$$

using (6-7). This shows that the inverse $D_{i\gamma,0}(z)^{-1}$ exists and is $O(R_{i\gamma,0}(z)) = O(h^{\frac{1}{2}-\delta})$.

Next, we compute the restriction of $\left(\frac{1}{\sqrt{h}}\bar{D}_{i\gamma}-z\right)$ to the E_Λ, $\Lambda>0$, eigenspace in (6-5). Using (2-32), (2-33) this has the form

$$D_{i\gamma,\Lambda}(z):=\mathtt{P}_\Lambda\left(\frac{1}{\sqrt{h}}\bar{D}_{i\gamma}-z\right)\mathtt{P}_\Lambda=\frac{1}{\sqrt{h}}\begin{bmatrix}-\xi_0-i\gamma-z\sqrt{h} & (\sqrt{2\bar{v}\Lambda h})^W\\ (\sqrt{2\bar{v}\Lambda h})^W & \xi_0+i\gamma-z\sqrt{h}\end{bmatrix}+\frac{1}{\sqrt{h}}\Omega_\Lambda$$

with respect to the $\mathbb{Z}_2$-grading $E_\Lambda=E_\Lambda^{\text{even}}\oplus E_\Lambda^{\text{odd}}$. Here we leave the identification $\mathtt{i}_\tau$ in (2-32) between the odd and even parts as being understood. Set $R_{i\gamma,\Lambda}(z)=[r_{i\gamma,\Lambda}(z)]^W$, where

$$r_{i\gamma,\Lambda}(z):=\frac{\sqrt{h}}{z^2h-(\xi_0+i\gamma)^2-2\bar{v}\Lambda h}\begin{bmatrix}-\xi_0-i\gamma-z\sqrt{h} & (\sqrt{2\bar{v}\Lambda h})\\ (\sqrt{2\bar{v}\Lambda h}) & \xi_0+i\gamma-z\sqrt{h}\end{bmatrix},$$

which is well defined for $|\operatorname{Re} z|\le\sqrt{2v_0}-\varepsilon_0<\inf_{\mathbb{R}^n}\sqrt{2\bar{v}\Lambda}$, and h sufficiently small. We now compute

$$\|R_{i\gamma,\Lambda}(z)D_{i\gamma,\Lambda}(z)-I\|\le C\varepsilon+O(h),$$
$$\|D_{i\gamma,\Lambda}(z)R_{i\gamma,\Lambda}(z)-I\|\le C\varepsilon+O(h)$$

using (6-8) with the constants above being uniform in Λ. Choosing ε sufficiently small in (5-21) shows that the inverse $D_{i\gamma,\Lambda}(z)^{-1}$ exists and is $O(R_{i\gamma,\Lambda}(z))=O(h^{-\frac{1}{2}})$ uniformly. □

We now finally finish the proof of Lemma 3.1.

Proof of Lemma 3.1. As noted in the beginning of the section, on account of (3-2), (3-3) and the reductions in Propositions 4.1 and 5.3, it suffices to show $\mathcal{T}^{\vartheta}_{\alpha\beta}(\bar{D})=O(h^\infty)$. We now define the trace

$$\tau_{\alpha\beta,t}(z):=\operatorname{tr}\left[a^W_{\alpha,t}\left(\frac{1}{\sqrt{h}}\bar{D}_t-z\right)^{-1}a^W_{\beta,t}\right],\quad \operatorname{Im} z>|\operatorname{Im} t|,\tag{6-10}$$

in terms of the almost analytic continuations. We clearly have

$$\tau_{\alpha\beta,t}(z)=O(h^{-n}|\operatorname{Im} z|^{-1}),$$
$$\frac{\partial}{\partial\bar{t}}\tau_{\alpha\beta,t}(z)=O(h^{-n}|\operatorname{Im} t|^\infty|\operatorname{Im} z|^{-2}).$$

Furthermore, by (6-1)–(6-3) $\tau_{\alpha\beta,t}(z)$ only depends on $\operatorname{Im} t$ and we have

$$\tau_{\alpha\beta,i\operatorname{Im} t}(z)=\tau_{\alpha\beta,0}(z)+O(h^{-n}|\operatorname{Im} t|^\infty|\operatorname{Im} z|^{-2}).\tag{6-11}$$

As before, we again introduce $\psi\in C^\infty(\mathbb{R};[0,1])$ such that

$$\psi(x)=\begin{cases}1, & x\le 1,\\ 0, & x\ge 2,\end{cases}$$

and set

$$\psi_M(z)=\psi\left(\frac{\operatorname{Im} z}{M\sqrt{h}\log\frac{1}{h}}\right).$$

The estimates (3-11), (3-12) along with the observation $\psi_M |\mathrm{Im}\, z|^N = O\big((M\sqrt{h}\log\frac{1}{h})^N\big)$ now give

$$\begin{aligned}\mathcal{T}^{\vartheta}_{\alpha\beta}(\bar{D}) &= \frac{1}{\pi}\int_{\mathbb{C}} \bar{\partial}(\psi_M \tilde{f})\check{\vartheta}\left(\frac{\lambda - z}{\sqrt{h}}\right)\tau_{\alpha\beta,0}(z)\,dz\,d\bar{z}\\ &= O(h^\infty) + \frac{1}{\pi}\int_{\{M\sqrt{h}\log\frac{1}{h} \le \mathrm{Im}\, z \le 2M\sqrt{h}\log\frac{1}{h}\}} \bar{\partial}(\psi_M \tilde{f})\check{\vartheta}\left(\frac{\lambda - z}{\sqrt{h}}\right)\tau_{\alpha\beta,0}(z)\,dz\,d\bar{z}.\end{aligned}$$

Using (6-11) and $\gamma = 2Mh^\delta \log\frac{1}{h}$, $\delta \in \big(\frac{1}{2}, 1\big)$, the above now equals

$$\mathcal{T}^{\vartheta}_{\alpha\beta}(\bar{D}) = O(h^\infty) + \frac{1}{\pi}\int_{\{M\sqrt{h}\log\frac{1}{h} \le \mathrm{Im}\, z \le 2M\sqrt{h}\log\frac{1}{h}\}} \bar{\partial}(\psi_M \tilde{f})\check{\vartheta}\left(\frac{\lambda - z}{\sqrt{h}}\right)\tau_{\alpha\beta,i\gamma}(z)\,dz\,d\bar{z}.$$

Since the resolvent $\big(\frac{1}{\sqrt{h}}\bar{D}_{i\gamma} - z\big)^{-1}$, and hence the trace $\tau_{\alpha\beta,i\gamma}(z)$, extends holomorphically to $\mathrm{Im}\, z > -Mh^{\delta-\frac{1}{2}}\log\frac{1}{h}$, $|\mathrm{Re}\, z| \le \sqrt{2\nu_0} - \varepsilon_0$ by Lemma 6.1 we may replace the integral in the last line above:

$$\begin{aligned}\mathcal{T}^{\vartheta}_{\alpha\beta}(\bar{D}) &= O(h^\infty) + \frac{1}{\pi}\int_{\{-1/2Mh^{\delta-1/2}\log\frac{1}{h} \le \mathrm{Im}\, z \le -\frac{1}{4}Mh^{\delta-1/2}\log\frac{1}{h}\}} \bar{\partial}(\psi_M \tilde{f})\check{\vartheta}\left(\frac{\lambda - z}{\sqrt{h}}\right)\tau_{\alpha\beta,i\gamma}(z)\,dz\,d\bar{z}\\ &= O(h^\infty) + O\left[\int_{\{-1/2Mh^{\delta-1/2}\log\frac{1}{h} \le \mathrm{Im}\, z \le -\frac{1}{4}Mh^{\delta-1/2}\log\frac{1}{h}\}} \frac{h^{-n-1/2}}{\sqrt{h}\log(1/h)} e^{\frac{S'_{\alpha\beta}(\mathrm{Im}\, z)}{h^{1/2-\varepsilon}}}\,dz\,d\bar{z}\right]\\ &= O\big[h^{\frac{M}{4}(S'_{\alpha\beta}) - n - \frac{1}{2}}\big]\end{aligned}$$

using (3-11) and $O(h^{-\frac{1}{2}})$ estimate on the resolvent $\big(\frac{1}{\sqrt{h}}\bar{D}_{i\gamma} - z\big)^{-1}$. Choosing M sufficiently large now gives the result. □

7. Local trace expansion

In this section we prove Lemma 3.2. This is a relatively classical trace expansion. A parametrix construction for the operator $e^{\frac{it}{h}D_h^2}$ may potentially be employed in its proof since the principal symbol of D_h^2 is Morse–Bott critical, as in [Brummelhuis et al. 1995]. However Lemma 3.2 would require an understanding of the large time behavior of parametrix left open in that paper; see [Camus 2004; Khuat-Duy 1997]. Here we prove the expansion using the alternate methods of local index theory. The expansion is analogous to the heat trace expansions arising in the analysis of the Bergman kernel [Bismut 1987; Ma and Marinescu 2007]. Here we adopt a modification of the approach in [Ma and Marinescu 2007, Chapters 1 and 4].

First, fix a point $p \in X$. On account of Definition 1.1 there is an orthonormal basis $e_{0,p} = R_p$, $e_{j,p}$, $e_{j+m,p}$, $j = 1, \dots, m$, of T_pX consisting of eigenvectors of $\mathfrak{J}_p$ with eigenvalues $0, \pm\lambda_{j,p} (:= \pm i\mu_j\nu(p))$, $j = 1, \dots, m$, such that

$$da(p) = \sum_{j=1}^{m} \lambda_j(p) e^*_{j,p} \wedge e^*_{j+m,p}. \tag{7-1}$$

Using the parallel transport from this basis, fix a geodesic coordinate system $(x_0, \dots, x_{2m})$ on an open neighborhood of $p \in \Omega$. Let $e_j = w_j^k \partial_{x_k}$, $0 \le j \le 2m$, be the local orthonormal frame of TX obtained by parallel transport of $e_{j,p} = \partial_{x_j}|_p$, $0 \le j \le 2m$, along geodesics. Hence we again have $w_j^k g_{kl} w_r^l = \delta_{jr}$, $w_j^k|_p = \delta_j^k$, with the g_{kl} being the components of the metric in these coordinates. Choose an orthonormal basis $\{u_q(p)\}_{q=1}^{2^m}$for S_p in which Clifford multiplication

$$c(e_j)|_p = \gamma_j \tag{7-2}$$

is standard. Choose an orthonormal basis $\mathtt{l}_p$ for L_p. Parallel transport the bases $\{u_q(p)\}_{q=1}^{2^m}$, $\mathtt{l}_p$ along geodesics using the spin connection ∇^S and unitary family of connections $\nabla^h = A_0 + \frac{i}{h}a$ to obtain trivializations $\{u_q\}_{q=1}^{2^m}$, $\mathtt{l}$ of S, L on Ω. Since Clifford multiplication is parallel, the relation (7-2) now holds on Ω. The connection $\nabla^{S\otimes L} = \nabla^S \otimes 1 + 1 \otimes \nabla^h$ can be expressed in this frame and these coordinates as

$$\nabla^{S\otimes L} = d + A_j^h\, dx^j + \Gamma_j\, dx^j, \tag{7-3}$$

where each A_j^h is a Christoffel symbol of ∇^h and each Γ_j is a Christoffel symbol of the spin connection ∇^S. Since the section l is obtained via parallel transport along geodesics, the connection coefficient A_j^h may be written in terms of the curvature $F_{jk}^h\, dx^j \wedge dx^k$ of ∇^h via

$$A_j^h(x) = \int_0^1 d\rho(\rho x^k F_{jk}^h(\rho x)). \tag{7-4}$$

The dependence of the curvature coefficients F_{jk}^h on the parameter h is seen to be linear in $\frac{1}{h}$ via

$$F_{jk}^h = F_{jk}^0 + \frac{i}{h}(da)_{jk} \tag{7-5}$$

despite the fact that they are expressed in the h-dependent frame $\mathtt{l}$. This is because a gauge transformation from an h-independent frame into $\mathtt{l}$ changes the curvature coefficient by conjugation. Since L is a line bundle, this is conjugation by a function and hence does not change the coefficient. Furthermore, the coefficients in the Taylor expansion of (7-5) at 0 maybe expressed in terms of the covariant derivatives $(\nabla^{A_0})^l F_{jk}^0$, $(\nabla^{A_0})^l (da)_{jk}$ evaluated at p. Next, using the Taylor expansion

$$(da)_{jk} = (da)_{jk}(0) + x^l a_{jkl}, \tag{7-6}$$

we see that the connection $\nabla^{S\otimes L}$ has the form

$$\nabla^{S\otimes L} = d + \left[\frac{i}{h}\left(\frac{x^k}{2}(da)_{jk}(0) + x^k x^l A_{jkl}\right) + x^k A_{jk}^0 + \Gamma_j\right] dx^j, \tag{7-7}$$

where

$$A_{jk}^0 = \int_0^1 d\rho(\rho F_{jk}^0(\rho x)), \quad A_{jkl} = \int_0^1 d\rho(\rho a_{jkl}(\rho x))$$

and Γ_j are all independent of h. Finally from (7-2) and (7-7) may write down the expression for the Dirac operator (1-2) also given as $D = hc \circ (\nabla^{S\otimes L})$ in terms of the chosen frame and coordinates to be

$$D = \gamma^r w_r^j \big[h\partial_{x_j} + i\tfrac{1}{2}x^k (da)_{jk}(0) + i x^k x^l A_{jkl} + h(x^k A_{jk}^0 + \Gamma_j)\big] \tag{7-8}$$

$$\begin{aligned} &= \gamma^r \big[w_r^j h\partial_{x_j} + i w_r^j \tfrac{1}{2}x^k (da)_{jk}(0) + \tfrac{1}{2}hg^{-\frac{1}{2}}\partial_{x_j}(g^{\frac{1}{2}} w_r^j)\big] \\ &\quad + \gamma^r \big[i w_r^j x^k x^l A_{jkl} + h w_r^j (x^k A_{jk}^0 + \Gamma_j) - \tfrac{1}{2}hg^{-\frac{1}{2}}\partial_{x_j}(g^{\frac{1}{2}} w_r^j)\big] \in \Psi_{\mathrm{cl}}^1(\Omega_s^0; \mathbb{C}^{2^m}). \end{aligned} \tag{7-9}$$

In the second expression above, both square brackets are self-adjoint with respect to the Riemannian density $e^1 \wedge \cdots \wedge e^n = \sqrt{g}\, dx := \sqrt{g}\, dx^1 \wedge \cdots \wedge dx^n$ with $g = \det(g_{ij})$. Again one may obtain an expression self-adjoint with respect to the Euclidean density dx in the framing $g^{\frac{1}{4}}u_q \otimes \mathtt{1}$, $1 \leq q \leq 2^m$, with the result being an addition of the term $h\gamma^j w_j^k g^{-\frac{1}{4}}(\partial_{x_k} g^{\frac{1}{4}})$.

Let i_g be the injectivity radius of g^{TX}. Define the cutoff $\chi \in C_c^\infty(-1,1)$ such that $\chi = 1$ on $\left(-\frac{1}{2}, \frac{1}{2}\right)$. We now modify the functions w_j^k, outside the ball $B_{i_g/2}(p)$, such that $w_j^k = \delta_j^k$ (and hence $g_{jk} = \delta_{jk}$) are standard outside the ball $B_{i_g}(p)$ of radius i_g centered at p. This again gives

$$\mathbb{D} = \gamma^r\left[w_r^j h\partial_{x_j} + i w_r^j \tfrac{1}{2} x^k (da)_{jk}(0) + \tfrac{1}{2} h g^{-\frac{1}{2}} \partial_{x_j}(g^{\frac{1}{2}} w_r^j)\right] \\ + \chi(|x|/i_g)\gamma^r\left[i w_r^j x^k x^l A_{jkl} + h w_r^j (x^k A_{jk}^0 + \Gamma_j) - \tfrac{1}{2} h g^{-\frac{1}{2}} \partial_{x_j}(g^{\frac{1}{2}} w_r^j)\right] \in \Psi_{\mathrm{cl}}^1(\mathbb{R}^n; \mathbb{C}^{2^m}) \quad (7\text{-}10)$$

as a well defined operator on $\mathbb{R}^n$ formally self adjoint with respect to $\sqrt{g}\,dx$. Since $\mathbb{D} + i$ is elliptic in the class $S^0(m)$ for the order function

$$m = \sqrt{1 + g^{jl}\left(\xi_j + \tfrac{1}{2} x^k (da)_{jk}(0)\right)\left(\xi_l + \tfrac{1}{2} x^r (da)_{lr}(0)\right)},$$

the operator $\mathbb{D}$ is essentially self adjoint.

Proposition 7.1. *There exist tempered distributions $u_j \in \mathcal{S}'(\mathbb{R}_s)$, $j = 0, 1, 2, \dots,$ such that one has a trace expansion*

$$\operatorname{tr} \phi\left(\frac{D}{\sqrt{h}}\right) = h^{-\frac{n}{2}}\left(\sum_{j=0}^{N} u_j(\phi) h^{\frac{j}{2}}\right) + h^{\frac{N+1-n}{2}} O\left(\sum_{k=0}^{n+1} \|\langle\xi\rangle^N \hat{\phi}^{(k)}\|_{L^1}\right) \quad (7\text{-}11)$$

for each $N \in \mathbb{N}$, $\phi \in \mathcal{S}(\mathbb{R}_s)$.

Proof. We begin by writing $\phi = \phi_0 + \phi_1$, with

$$\phi_0(s) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{i\xi s}\hat{\phi}(\xi)\chi\left(\frac{2\xi\sqrt{h}}{i_g}\right) d\xi, \quad \phi_1(s) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{i\xi s}\hat{\phi}(\xi)\left[1 - \chi\left(\frac{2\xi\sqrt{h}}{i_g}\right)\right] d\xi$$

via Fourier inversion.

First considering ϕ_1, integration by parts gives the estimate

$$|s^{n+1}\phi_1(s)| \leq C_N h^{\frac{N-1}{2}}\left(\sum_{k=0}^{n+1} \|\xi^N \hat{\phi}^{(k)}\|_{L^1}\right)$$

for all $N \in \mathbb{N}$. Hence,

$$\left\| D^{n+1-a}\phi_1\left(\frac{D}{\sqrt{h}}\right) D^a \right\|_{L^2 \to L^2} = C_N h^{\frac{n+N}{2}}\left(\sum_{k=0}^{n+1} \|\xi^N \hat{\phi}^{(k)}\|_{L^1}\right)$$

for all $N \in \mathbb{N}$, for all $a = 0, \dots, n+1$. The semiclassical elliptic estimate and Sobolev's inequality now give the estimate

$$\left|\phi_1\left(\frac{D}{\sqrt{h}}\right)\right|_{C^0(X \times X)} \leq C_N h^{\frac{n+N}{2}}\left(\sum_{k=0}^{n+1} \|\xi^N \hat{\phi}^{(k)}\|_{L^1}\right), \quad (7\text{-}12)$$

for all $N \in \mathbb{N}$, on the Schwartz kernel.

Next, considering ϕ_0, we first use the change of variables $\alpha = \xi\sqrt{h}$ to write

$$\phi_0\left(\frac{D}{\sqrt{h}}\right) = \frac{1}{2\pi\sqrt{h}}\int_{\mathbb{R}} e^{i\alpha(D_{A_0}+ih^{-1}c(a))}\hat{\phi}\left(\frac{\alpha}{\sqrt{h}}\right)\chi\left(\frac{2\alpha}{i_g}\right)d\alpha.$$

Now since $D = \mathbb{D}$ on $B_{i_g/2}(p)$, we may use the finite propagation speed of the wave operators $e^{i\alpha h^{-1}D}$, $e^{i\alpha h^{-1}\mathbb{D}}$ [Ma and Marinescu 2007, Theorem D.2.1] to conclude

$$\phi_0\left(\frac{D}{\sqrt{h}}\right)(p,\cdot) = \phi_0\left(\frac{\mathbb{D}}{\sqrt{h}}\right)(0,\cdot). \tag{7-13}$$

The right-hand side above is defined using functional calculus of self-adjoint operators, with standard local elliptic regularity arguments implying the smoothness of its Schwartz kernel. By virtue of (7-12), a similar estimate for $\phi_1\big(\frac{\mathbb{D}}{\sqrt{h}}\big)$, and (7-13) it now suffices to consider $\phi\big(\frac{\mathbb{D}}{\sqrt{h}}\big)$.

We now introduce the rescaling operator

$$\mathscr{R} : C^\infty(\mathbb{R}^n;\mathbb{C}^{2^m}) \to C^\infty(\mathbb{R}^n;\mathbb{C}^{2^m}), \quad (\mathscr{R}s)(x) := s\left(\frac{x}{\sqrt{h}}\right).$$

Conjugation by $\mathscr{R}$ amounts to the rescaling of coordinates $x \to x\sqrt{h}$. A Taylor expansion in (7-10) now gives the existence of classical (h-independent) self-adjoint, first-order differential operators $\mathtt{D}_j = a_j^k(x)\partial_{x_k} + b_j(x)$, $j = 0, 1, \ldots,$ with polynomial coefficients (of degree at most $j+1$) as well as h-dependent self-adjoint, first-order differential operators $\mathrm{E}_j = \sum_{|\alpha|=N+1} x^\alpha\big[c_{j,\alpha}^k(x;h)\partial_{x_k} + d_{j,\alpha}(x;h)\big]$, $j = 0, 1, \ldots,$ with uniformly C^∞ bounded coefficients $c_{j,\alpha}^k$, $d_{j,\alpha}$ such that

$$\mathscr{R}\mathbb{D}\mathscr{R}^{-1} = \sqrt{h}\mathtt{D}, \tag{7-14}$$

with

$$\mathtt{D} = \left(\sum_{j=0}^{N} h^{\frac{j}{2}}\mathtt{D}_j\right) + h^{\frac{N+1}{2}}\mathrm{E}_{N+1} \quad \forall N. \tag{7-15}$$

The coefficients of the polynomials $a_j^k(x)$, $b_j(x)$ again involve the covariant derivatives of the curvatures F^{TX}, F^{A_0} and da evaluated at p. Furthermore, the leading term in (7-15) is easily computed as

$$\mathtt{D}_0 = \gamma^j\big[\partial_{x_j} + i\tfrac{1}{2}x^k(da)_{jk}(0)\big] \tag{7-16}$$

$$= \gamma^0\partial_{x_0} + \underbrace{\gamma^j\big[\partial_{x_j} + \tfrac{1}{2}i\lambda_j(p)x_{j+m}\big] + \gamma^{j+m}\big[\partial_{x_{j+m}} - \tfrac{1}{2}i\lambda_j(p)x_j\big]}_{:=\mathtt{D}_{00}} \tag{7-17}$$

using (7-1), (7-6). It is now clear from (7-14) that

$$\phi\left(\frac{\mathbb{D}}{\sqrt{h}}\right)(x,x') = h^{-\frac{n}{2}}\phi(\mathtt{D})\left(\frac{x}{\sqrt{h}},\frac{x'}{\sqrt{h}}\right). \tag{7-18}$$

Next, let $I_j = \big\{k=(k_0,k_1,\ldots)\,\big|\, k_\alpha \in \mathbb{N}, \sum k_\alpha = j\big\}$ denote the set of partitions of the integer j and set

$$\mathtt{C}_j^z = \sum_{k\in I_j}(z-\mathtt{D}_0)^{-1}\big[\Pi_\alpha[\mathtt{D}_{k_\alpha}(z-\mathtt{D}_0)^{-1}]\big]. \tag{7-19}$$

Local elliptic regularity estimates again give

$$(z-\mathtt{D})^{-1} = O_{L^2_{\mathrm{loc}}\to L^2_{\mathrm{loc}}}(|\mathrm{Im}\, z|^{-1}) \qquad \text{and} \qquad \mathtt{C}^z_j = O_{L^2_{\mathrm{loc}}\to L^2_{\mathrm{loc}}}(|\mathrm{Im}\, z|^{-j-1}), \quad j=0,1,\ldots.$$

A straightforward computation using (7-15) then yields

$$(z-\mathtt{D})^{-1} - \left(\sum_{j=0}^{N} h^{\frac{j}{2}} \mathtt{C}^z_j\right) = O_{L^2_{\mathrm{loc}}\to L^2_{\mathrm{loc}}}((|\mathrm{Im}\, z|^{-1} h^{\frac{1}{2}})^{N+1}). \tag{7-20}$$

A similar expansion as (7-15) for the operator $(1+\mathtt{D}^2)^{\frac{n+1}{2}}(z-\mathtt{D})$ also gives the bounds

$$(1+\mathtt{D}^2)^{-\frac{n+1}{2}}(z-\mathtt{D})^{-1} - \left(\sum_{j=0}^{N} h^{\frac{j}{2}} \mathtt{C}^z_{j,n+1}\right) = O_{H^s_{\mathrm{loc}}\to H^{s+n+1}_{\mathrm{loc}}}((|\mathrm{Im}\, z|^{-1} h^{\frac{1}{2}})^{N+1}) \tag{7-21}$$

for all $s\in\mathbb{R}$, for classical (h-independent) Sobolev spaces H^s_{loc}. Here each $\mathtt{C}^z_{j,n+1}$ satisfies

$$\mathtt{C}^z_{j,n+1} = O_{H^s_{\mathrm{loc}}\to H^{s+n+1}_{\mathrm{loc}}}(|\mathrm{Im}\, z|^{-j-1})$$

with leading term

$$\mathtt{C}^z_{0,n+1} = (1+\mathtt{D}_0^2)^{-\frac{n+1}{2}}(z-\mathtt{D}_0)^{-1}.$$

Finally, plugging the expansion (7-21) into the Helffer–Sjöstrand formula

$$\phi(\mathtt{D}) = -\frac{1}{2\pi}\int_{\mathbb{C}} \bar\partial\tilde\rho(z)(1+\mathtt{D}^2)^{-\frac{n+1}{2}}(z-\mathtt{D})^{-1}\, dz\, d\bar z,$$

with $\rho(x) := \langle x\rangle^{n+1}\phi(x)$, gives

$$\phi(\mathtt{D})(0,0) = \left(\sum_{j=0}^{N} h^{\frac{j}{2}} U_{j,p}(\phi)\right) + h^{\frac{N+1}{2}} O\left(\sum_{k=0}^{n+1} \|\langle\xi\rangle^N \hat\phi^{(k)}\|_{L^1}\right) \tag{7-22}$$

using Sobolev's inequality. Here each

$$U_{j,p}(\phi) = -\frac{1}{\pi}\int_{\mathbb{C}} \bar\partial\tilde\rho(z)\mathtt{C}^z_{j,n+1}(0,0)\, dz\, d\bar z \in \mathrm{End}\, S^{TX}_p \tag{7-23}$$

defines a smooth family (in $p\in X$) of distributions U_j and the remainder term in (7-22) comes from the estimate

$$\bar\partial\tilde\rho = O\left(|\mathrm{Im}\, z|^{N+1}\sum_{k=0}^{n+1} \|\langle\xi\rangle^N\hat\phi^{(k)}\|_{L^1}\right)$$

on the almost analytic continuation; see [Zworski 2012, Section 3.1]. Integrating the trace of (7-22) over X and using (7-18) gives (7-11). □

Next we would like to understand the structure of the distributions u_j appearing in (7-11). Clearly,

$$u_j = \int_X u_{j,p}, \quad \text{with } u_{j,p} := \mathrm{tr}\, U_{j,p} \in C^\infty(X; \mathcal{S}'(\mathbb{R}_s)), \tag{7-24}$$

is the smooth family of tempered distributions parametrized by X defined via the pointwise trace of (7-23). Letting $H(s) \in \mathcal{S}'(\mathbb{R}_s)$ denote the Heaviside distribution, we now define the following elementary tempered distributions:

$$v_{a;p}(s) := s^a, \quad a \in \mathbb{N}_0, \tag{7-25}$$

$$v_{a,b,c,\Lambda;p}(s) := \partial_s^a \big[|s| s^b (s^2 - 2\nu_p \Lambda)^{c-\frac{1}{2}} H(s^2 - 2\nu_p \Lambda) \big], \quad (a,b,c;\Lambda) \in \mathbb{N}_0 \times \mathbb{Z} \times \mathbb{N}_0 \times \mu.(\mathbb{N}_0^m \setminus 0). \tag{7-26}$$

Proposition 7.2. *For each j, the distribution (7-24) can be written in terms of (7-25), (7-26):*

$$u_{j,p}(s) = \sum_{a \le 2j+2} c_{j;a}(p) s^a + \sum_{\substack{\Lambda \in \mu.(\mathbb{N}_0^m \setminus 0) \\ a,|b|,c \le 4j+4}} c_{j;a,b,c,\Lambda}(p) v_{a,b,c,\Lambda;p}(s). \tag{7-27}$$

Moreover, the coefficient functions $c_{j;a}, c_{j;a,b,c,\Lambda} \in C^\infty(X)$ above are evaluations at p of polynomials in the covariant derivatives (with respect to $\nabla^{TX} \otimes 1 + 1 \otimes \nabla^{A_0}$) of the curvatures F^{TX}, F^{A_0} of the Levi-Civita connection ∇^{TX}, ∇^{A_0} and da.

Proof. It suffices to consider the restriction of u_j to the interval $(-\sqrt{2\nu M}, \sqrt{2\nu M})$ for each $0 < M \notin \mu.(\mathbb{N}_0^m \setminus 0)$. We begin by finding the spectrum of the operator $\mathtt{D}_{00}$ in (7-17). To this end, define the unitary operator $\mathtt{U}_\lambda : C^\infty(\mathbb{R}^n; \mathbb{C}^{2^m}) \to C^\infty(\mathbb{R}^n; \mathbb{C}^{2^m})$,

$$(\mathtt{U}_\lambda s)(x_0, x_1, x_2, \ldots) = \left(\prod_{j=1}^m \lambda_j \right) s(x_0, \lambda_1^{-\frac{1}{2}} x_1, \lambda_1^{-\frac{1}{2}} x_2, \lambda_2^{-\frac{1}{2}} x_3, \lambda_2^{-\frac{1}{2}} x_4, \ldots)$$

and

$$f = \sum_{j=1}^m (x_j x_{j+m} + \xi_j \xi_{j+m}) \in C^\infty(\mathbb{R}^{2m}).$$

Next, as in (5-1) we compute the conjugate

$$e^{\frac{i\pi}{4} f_0^W} \mathtt{U}_\lambda \mathtt{D}_{00} \mathtt{U}_\lambda^* e^{-\frac{i\pi}{4} f_0^W} = [2\nu(p)]^{\frac{1}{2}} D_{\mathbb{R}^m}|_{h=1}$$

of the operator in (7-17) in terms of the magnetic Dirac operator on $\mathbb{R}^m$ (2-21) evaluated at $h = 1$. Hence the eigenspaces of $\mathtt{D}_{00}$ are

$$\mathtt{U}_\lambda^* e^{-\frac{i\pi}{4} f_0^W} (E_0 \otimes L^2(\mathbb{R}^{m+1}_{x_0,x''})), \quad \mathtt{U}_\lambda^* e^{-\frac{i\pi}{4} f_0^W} (E_\Lambda^\pm \otimes L^2(\mathbb{R}^{m+1}_{x_0,x''})); \Lambda \in \mu.(\mathbb{N}_0^m \setminus 0),$$

with eigenvalues $0, \pm\sqrt{2\nu\Lambda}$ respectively, where

$$E_0 := \mathbb{C}[\psi_{0,0}|_{h=1}], \quad E_\Lambda^\pm = \bigoplus_{\substack{\tau \in \mathbb{N}_0^m \setminus 0 \\ \Lambda = \mu.\tau}} E_\tau^\pm \Big|_{h=1}$$

are as in (6-5). We again let $\mathtt{P}_0$, $\mathtt{P}_\Lambda^\pm$ denote the respective projections onto the eigenspaces of $\mathtt{D}_{00}$ and $\mathtt{P}_\Lambda = \mathtt{P}_\Lambda^+ \oplus \mathtt{P}_\Lambda^-$. We also denote by $\mathtt{P}_{>M} = \oplus_{\Lambda > M} \mathtt{P}_\Lambda$ the projection onto eigenspaces with eigenvalue greater than $\sqrt{2\nu M}$ in absolute value.

Now, since expansions in L^2_{loc} are unique, it suffices to work with the resolvent expansion (7-20) in the computation of u_j. The j-th term in the expansion is of the form

$$\mathtt{C}^z_j=\sum_{k\in I_j}(z-\mathtt{D}_0)^{-1}[\Pi_\alpha \mathtt{D}_{k_\alpha}(z-\mathtt{D}_0)^{-1}],\tag{7-28}$$

where each $\mathtt{D}_{k_\alpha}$ is a differential operator with polynomial coefficients involving the covariant derivatives of the curvatures F^{TX}, F^{A_0} and da. Now using (7-17) we decompose each resolvent term above according to the eigenspaces of $\mathtt{D}_{00}$:

$$\begin{aligned}(z-\mathtt{D}_0)^{-1}&=\mathtt{P}_0\Big(\frac{1}{z-\gamma^0\partial_{x_0}}\Big)\mathtt{P}_0\\&\quad\oplus\bigoplus_{\Lambda\in\mu.\mathbb{N}_0^m\cap(0,M)}\mathtt{P}_\Lambda\Big(\frac{z+\gamma^0\partial_{x_0}+\mathtt{D}_{00}}{z^2+\partial^2_{x_0}-2\nu\Lambda}\Big)\mathtt{P}_\Lambda\oplus\mathtt{P}_{>M}\Big(\frac{z+\gamma^0\partial_{x_0}+\mathtt{D}_{00}}{z^2+\partial^2_{x_0}-\mathtt{D}^2_{00}}\Big)\mathtt{P}_{>M}.\end{aligned}\tag{7-29}$$

Next, we plug (7-29) into (7-28). This gives an expansion for $\mathtt{C}^z_j$ with some of the terms given by

$$T^z[\Pi_\alpha\mathtt{D}_{k_\alpha}T^z],\quad\text{where}\ \ T^z=\mathtt{P}_{>M}\Big(\frac{z+\gamma^0\partial_{x_0}+\mathtt{D}_{00}}{z^2+\partial^2_{x_0}-\mathtt{D}^2_{00}}\Big)\mathtt{P}_{>M},$$

and which are holomorphic for $\operatorname{Re} z\in(-\sqrt{2\nu M},\sqrt{2\nu M})$. For the rest of the terms in $\mathtt{C}^z_j$, we use the commutation relations

$$\begin{gathered}[\gamma^0,\mathtt{P}_0]=[\gamma^0,\mathtt{P}_\Lambda]=[\gamma^0,\mathtt{P}_{>M}]=0,\\ [\partial_{x_0},\mathtt{P}_0]=[\partial_{x_0},\mathtt{P}_\Lambda]=[\partial_{x_0},\mathtt{P}_{>M}]=0,\\ [\partial_{x_0},\mathtt{D}_{00}]=0,\\ [(z^2+\partial^2_{x_0}-2\nu\Lambda)^{-1},x_j]=\delta_{0j}(z^2+\partial^2_{x_0}-2\nu\Lambda)^{-2}\partial_{x_0},\\ [(z^2+\partial^2_{x_0}-2\nu\Lambda)^{-1},\partial_{x_J}]=0,\end{gathered}$$

as well as the Clifford relations (2-7). This now gives a finite sum of terms of the form

$$T^z_0\Big[\prod_{k=1}^K S_kT^z_k\Big]\times\Big[\prod_{\Lambda\in\mu.\mathbb{N}_0^m\cap(0,M)}\frac{1}{(z^2+\partial^2_{x_0}-2\nu\Lambda)^{a_\Lambda}}\Big](z-\gamma^0\partial_{x_0})^{-a_0}z^{b_1}x_0^{b_2}\partial^{b_3}_{x_0},\tag{7-30}$$

$a_0+\Sigma a_\Lambda\le 2j+2$, $b_1,b_2,b_3\le j+1$, where each S_k is a differential operator in $(x'x'')$ (i.e., independent of x_0) with polynomial coefficients and each T^z_k is equal to one of

$$\mathtt{P}_0,\quad \mathtt{P}_\Lambda,\quad \mathtt{P}_\Lambda\mathtt{D}_{00}\mathtt{P}_\Lambda,\quad \mathtt{P}_{>M}\Big(\frac{1}{z^2+\partial^2_{x_0}-\mathtt{D}^2_{00}}\Big)\mathtt{P}_{>M},\quad\text{or}\quad \mathtt{P}_{>M}\Big(\frac{\mathtt{D}_{00}}{z^2+\partial^2_{x_0}-\mathtt{D}^2_{00}}\Big)\mathtt{P}_{>M},\tag{7-31}$$

with at least one occurrence of $\mathtt{P}_0,\mathtt{P}_\Lambda$ or $\mathtt{P}_\Lambda\mathtt{D}_{00}\mathtt{P}_\Lambda$ in (7-30). Now using partial fractions, (7-30) may be written as a sum of terms of the forms

$$T_0^z\left[\prod_{k=1}^{K} S_k T_k^z\right] \times (z-\gamma^0\partial_{x_0})^{-a_0} z^{b_1} x_0^{b_2}\partial_{x_0}^{b_3},$$

$$T_0^z\left[\prod_{k=1}^{K} S_k T_k^z\right] \times (z^2+\partial_{x_0}^2-2\nu\Lambda)^{-a_\Lambda} z^{b_1} x_0^{b_2}\partial_{x_0}^{b_3}, \quad \Lambda\in\mu.\mathbb{N}_0^m\cap(0,M), \tag{7-32}$$

$a_0, a_\Lambda \le 2j+2$, $b_1, b_2, b_3 \le j+1$. Next, we plug (7-32) into the Helffer–Sjöstrand formula and use the analyticity of $\mathtt{P}_{>M}(1/(z^2+\partial_{x_0}^2-\mathtt{D}_{00}^2))\mathtt{P}_{>M}$ and $\mathtt{P}_{>M}(\mathtt{D}_{00}/(z^2+\partial_{x_0}^2-\mathtt{D}_{00}^2))\mathtt{P}_{>M}$ for $\operatorname{Re} z\in(-\sqrt{2\nu M},\sqrt{2\nu M})$. This gives

$$U_{j,p}(\phi) = -\frac{1}{2\pi}\int_{\mathbb{C}} \bar\partial\tilde\phi(z)\mathtt{C}_j^z(0,0)\,dz\,d\bar z$$

for $\phi\in C_c^\infty(-\sqrt{2\nu M},\sqrt{2\nu M})$, as a sum of terms of the form

$$\left(T_0^0\left[\prod_{k=1}^{K} S_k T_k^0\right] \times x_0^{b_2}\partial_{x_0}^{b_3}\phi_0(\gamma^0\partial_{x_0})\right)(0,0),$$

$$\left(T_0^0\left[\prod_{k=1}^{K} S_k T_k^0\right] \times x_0^{b_2}\partial_{x_0}^{b_3}\phi_\Lambda(-\partial_{x_0}^2+2\nu\Lambda)\right)(0,0), \quad \Lambda\in\mu.\mathbb{N}_0^m\cap(0,M), \tag{7-33}$$

where each T_k^0 is equal to one of

$$\mathtt{P}_0,\quad \mathtt{P}_\Lambda,\quad \mathtt{P}_\Lambda\mathtt{D}_{00}\mathtt{P}_\Lambda,\quad \mathtt{P}_{>M}\left(\frac{1}{2\nu\Lambda-\mathtt{D}_{00}^2}\right)\mathtt{P}_{>M},\quad \text{or}\quad \mathtt{P}_{>M}\left(\frac{\mathtt{D}_{00}}{2\nu\Lambda-\mathtt{D}_{00}^2}\right)\mathtt{P}_{>M},$$

and

$$\phi_0(s) = \frac{(-1)^{a_0-1}}{(a_0-1)!}x^{b_1}\phi(s),$$

$$\phi_\Lambda(s^2) = \frac{(-1)^{a_\Lambda-1}}{(a_\Lambda-1)!}\left\{\left[\partial_r^{a_\Lambda-1}\left(\frac{r^{b_1}\phi(r)}{(r-s)^{a_\Lambda}}\right)\right]\Bigg|_{r=-s} - \left[\partial_r^{a_\Lambda-1}\left(\frac{r^{b_1}\phi(r)}{(r+s)^{a_\Lambda}}\right)\right]\Bigg|_{r=s}\right\}.$$

At least one occurrence of $\mathtt{P}_0$,$\mathtt{P}_\Lambda$ and $\mathtt{P}_\Lambda\mathtt{D}_{00}\mathtt{P}_\Lambda$ in (7-33) gives the smoothness of the kernel.

Finally, an elementary computation involving Laplace transforms using the knowledge of the heat kernel

$$e^{t\partial_{x_0}^2}(x_0,y_0) = \frac{1}{\sqrt{4\pi t}}e^{-\frac{|x_0-y_0|^2}{4t}}$$

gives

$$x_0^{b_2}\partial_{x_0}^{b_3}\phi_0(\gamma^0\partial_{x_0})(0,0) = \frac{(-\frac{1}{2})^{\left[\frac{b_3+1}{2}\right]}}{\sqrt{\pi}\,\Gamma\left(\left[\frac{b_3+1}{2}\right]+\frac{1}{2}\right)}\delta_{0b_2}v_{b_3;p}(\phi_0),$$

$$x_0^{b_2}\partial_{x_0}^{b_3}\phi_\Lambda(-\partial_{x_0}^2+2\nu\Lambda)(0,0) = \begin{cases} \dfrac{(-\frac{1}{2})^{\frac{b_3}{2}}}{4\pi\Gamma\left(\frac{b_3}{2}-\frac{1}{2}\right)}\delta_{0b_2}v_{0,0,\frac{b_3}{2},\Lambda;p}(\phi_\Lambda(s^2)), & b_3 \text{ even},\\ 0, & b_3 \text{ odd},\end{cases}$$

completing the proof. □

As an immediate corollary of Proposition 7.2, we have that the distributions u_j are smooth near 0.

Corollary 7.3. *For each j,*

$$\operatorname{sing\,spt}(u_j) \subset \mathbb{R} \setminus (-\sqrt{2\nu_0}, \sqrt{2\nu_0}).$$

Proof. This follows immediately from (7-24)–(7-27) on noting that the distributions $v_{a;p}$ are smooth, while $v_{a,b,c,\Lambda;p} = 0$ on $\mathbb{R} \setminus (-\sqrt{2\nu_0}, \sqrt{2\nu_0})$ for each $p \in X$. □

We next give the exact computation for the first coefficient u_0 of Proposition 7.1. In the computation below, recall that $Z_\tau = |I_\tau|$, as in (2-13), denotes the number of nonzero components of $\tau \in \mathbb{N}_0^m \setminus 0$.

Proposition 7.4. *The first coefficient u_0 of (7-11) is given by*

$$u_{0,p} = c_{0;0} + \sum_{\Lambda \in \mu.(\mathbb{N}_0^m \setminus 0)} c_{0;0,0,0,\Lambda}(p) v_{0,0,0,\Lambda;p}(s), \tag{7-34}$$

where

$$c_{0;0} = \frac{\nu_p^m \left(\prod_{j=1}^m \mu_j\right)}{(4\pi)^{\frac{n}{2}}},$$

$$c_{0;0,0,0,\Lambda}(p) = \frac{\nu_p^m \left(\prod_{j=1}^m \mu_j\right)}{(4\pi)^{\frac{n}{2}}} \dim(E_\Lambda) = \frac{\nu_p^m \left(\prod_{j=1}^m \mu_j\right)}{(4\pi)^{\frac{n}{2}}} \Bigg(\sum_{\substack{\tau \in \mathbb{N}_0^m \setminus 0 \\ \mu.\tau = \Lambda}} 2^{Z_\tau} \Bigg). \tag{7-35}$$

Proof. First note that the square of (7-16) gives the harmonic oscillator

$$\mathrm{D}_0^2 = -\delta^{jk} \partial_{x_j} \partial_{x_k} - i(da)_k^j(0) x^k \partial_{x_j} + \frac{1}{4} x^k x_l (da)_k^j(0)(da)_j^l(0) + \frac{i}{2} \gamma^j \gamma^k (da)_{jk}(0).$$

The heat kernel $e^{-t\mathrm{D}_0^2}$ of the above is given by Mehler's formula [Berline et al. 2004, Section 4.2]

$$e^{-t\mathrm{D}_0^2}(x,y) = \frac{1}{(4\pi t)^m} \frac{e^{-\frac{(x_0-y_0)^2}{4t}}}{\sqrt{4\pi t}} \det{}^{\frac{1}{2}} \left(\frac{itda(0)}{\sinh itda(0)} \right) e^{-tc(ida(0))} \tag{7-36}$$

$$\times \exp\left\{ -\frac{\lambda_j}{4 \tanh \lambda_j t} \left((x_j - y_j)^2 + (x_{j+m} - y_{j+m})^2 \right) + \frac{\lambda_j}{2} \tanh\left(\frac{\lambda_j t}{2} \right) (x_j y_j + x_{j+m} y_{j+m}) \right\}. \tag{7-37}$$

Next, using (7-1) we compute

$$e^{-tc(ida(0))} = \prod_{j=1}^m \left[\cosh(t\lambda_j) - ic(e_j)c(e_{j+m}) \sinh(t\lambda_j) \right]. \tag{7-38}$$

For $I \subset \{2, \dots, m\}$ and $\omega_I = \bigwedge_{j \in I} (e_j \wedge e_{j+m})$, the commutation

$$c(e_1)c(e_{m+1})c(\omega_I) = \tfrac{1}{2}[c(e_1), c(e_{m+1})c(\omega_I)]$$

shows that the only traceless terms in (7-38) are the constants. Hence, Mehler's formula (7-36) gives

$$\begin{aligned}\operatorname{tr} e^{-t\mathrm{D}_0^2}(0,0) &= \frac{1}{(4\pi t)^{\frac{n}{2}}}\det{}^{\frac{1}{2}}\left(\frac{itda(0)}{\tanh itda(0)}\right) = \frac{t^{-\frac{1}{2}}}{(4\pi)^{\frac{n}{2}}}\left(\prod_{j=1}^{m}\frac{\lambda_j}{\tanh t\lambda_j}\right)\\ &= \frac{t^{-\frac{1}{2}}}{(4\pi)^{\frac{n}{2}}}\left[\prod_{j=1}^{m}\lambda_j(1+2e^{-2t\lambda_j}+2e^{-4t\lambda_j}+\cdots)\right]\\ &= \frac{t^{-\frac{1}{2}}}{(4\pi)^{\frac{n}{2}}}\left(\prod_{j=1}^{m}\lambda_j\right)\left(\sum_{\tau\in\mathbb{N}_0^m}2^{Z_\tau}e^{-2t\tau.\lambda}\right)\\ &= \frac{\nu_p^m\left(\prod_{j=1}^m\mu_j\right)}{(4\pi)^{\frac{n}{2}}}\left(t^{-\frac{1}{2}}\sum_{\tau\in\mathbb{N}_0^m}2^{Z_\tau}e^{-2t\tau.\lambda}\right) = u_{0,p}(e^{-ts^2}),\end{aligned} \tag{7-39}$$

with $u_{0,p}$ as in (7-34) and the last line above following from an easy computation of Laplace transforms; see [Savale 2014, Section 4]. Furthermore, differentiating Mehler's formula using (7-16) gives

$$\operatorname{tr}\mathrm{D}_0 e^{-t\mathrm{D}_0^2}(0,0) = 0 = u_{0,p}(se^{-ts^2}) \tag{7-40}$$

since the right-hand side of (7-34) is an even distribution. From (7-39) and (7-40) we have that the evaluations of both sides of (7-34) on e^{-ts^2}, se^{-ts^2} are equal. Differentiating with respect to t and setting $t=1$ gives that the two sides of (7-34) evaluate equally on $s^k e^{-s^2}$ for all $k\in\mathbb{N}_0$. The proposition now follows from the density of this collection in $\mathcal{S}(\mathbb{R}_s)$. □

We now complete the proof of Lemma 3.2.

Proof of Lemma 3.2. We begin by writing

$$\operatorname{tr}\left[f\left(\frac{D}{\sqrt{h}}\right)\frac{1}{h^{1-\varepsilon}}\check{\theta}\left(\frac{\lambda\sqrt{h}-D}{h^{1-\varepsilon}}\right)\right] = \frac{h^{-\frac{1}{2}}}{2\pi}\int dt\,\operatorname{tr}\left[f\left(\frac{D}{\sqrt{h}}\right)e^{it(\lambda-\frac{D}{\sqrt{h}})}\right]\theta(th^{\frac{1}{2}-\varepsilon}). \tag{7-41}$$

Next, the expansion result, Proposition 7.1, with $\phi(x)=f(x)e^{it(\lambda-x)}$, combined with the smoothness of u_j on $\operatorname{spt}(f)\subset(-\sqrt{2\nu_0},\sqrt{2\nu_0})$ from Corollary 7.3 gives

$$\operatorname{tr}\left[f\left(\frac{D}{\sqrt{h}}\right)e^{it(\lambda-\frac{D}{\sqrt{h}})}\right] = e^{it\lambda}h^{-\frac{n}{2}}\left(\sum_{j=0}^{N}h^{\frac{j}{2}}\widehat{fu_j}(t)\right)+h^{\frac{N+1-n}{2}}\underbrace{O\left(\sum_{k=0}^{n+1}\|\langle\xi\rangle^N\hat{\phi}^{(k)}(\xi-t)\|_{L^1}\right)}_{=O(\langle t\rangle^N)}. \tag{7-42}$$

Finally, plugging (7-42) into (7-41) and using $\theta(th^{\frac{1}{2}-\varepsilon})=1+O(h^\infty)$ gives via Fourier inversion

$$\frac{h^{-\frac{1}{2}}}{2\pi}\int dt\,\operatorname{tr}\left[f\left(\frac{D}{\sqrt{h}}\right)e^{it(\lambda-\frac{D}{\sqrt{h}})}\right]\theta(th^{\frac{1}{2}-\varepsilon}) = h^{-m-1}\left(\sum_{j=0}^{N}h^{\frac{j}{2}}f(\lambda)u_j(\lambda)\right)+O(h^{\varepsilon(N+1)-m-1})$$

as required. □

8. Asymptotics of spectral invariants

In this section we prove Theorem 1.2 on the asymptotics of the spectral invariants.

Proof of Theorem 1.2. To prove the Weyl law (1-5), we choose $\theta \in C_c^\infty((-T,T);[0,1])$ such that $\theta(x)=1$ on $(-T',T')$, $T'<T$, $\check{\theta}(\xi)\geq 0$ and $\check{\theta}(\xi)\geq 1$ for $|\xi|\leq c$ in Theorem 1.3. Choosing $f(x)\geq 0$ with $f(0)=1$, the trace expansion (1-7) with $\lambda=0$ now gives

$$\frac{1}{h}N(-ch,ch)(1+O(\sqrt{h}))\leq \operatorname{tr}\left[f\left(\frac{D}{\sqrt{h}}\right)\frac{1}{h}\check{\theta}\left(\frac{-D}{h}\right)\right]=O(h^{-m-1})$$

proving (1-5).

To prove the estimate (1-6) on the eta invariant, we first use its invariance under positive scaling (2-2) and the formula (2-5) to write

$$\begin{aligned}\eta_h=\eta\left(\frac{D}{\sqrt{h}}\right)&=\int_0^\infty dt\,\frac{1}{\sqrt{\pi t}}\operatorname{tr}\left[\frac{D}{\sqrt{h}}e^{-\frac{t}{h}D^2}\right]\\&=\int_0^1 dt\,\frac{1}{\sqrt{\pi t}}\operatorname{tr}\left[\frac{D}{\sqrt{h}}e^{-\frac{t}{h}D^2}\right]+\int_1^\infty dt\,\frac{1}{\sqrt{\pi t}}\operatorname{tr}\left[\frac{D}{\sqrt{h}}e^{-\frac{t}{h}D^2}\right].\end{aligned}\tag{8-1}$$

Next, [Savale 2014, equation 4.5, p. 859] with $r=\frac{1}{h}$ translates to the estimate

$$\operatorname{tr}\left[\frac{D}{\sqrt{h}}e^{-\frac{t}{h}D^2}\right]=O(h^{-m}e^{ct}).\tag{8-2}$$

Plugging, (8-2) into the first integral of (8-1) gives

$$\eta_h=O(h^{-m})+\operatorname{tr}E\left(\frac{D}{\sqrt{h}}\right),\tag{8-3}$$

where

$$E(x)=\operatorname{sign}(x)\operatorname{erfc}(|x|)=\operatorname{sign}(x)\cdot\frac{2}{\sqrt{\pi}}\int_{|x|}^\infty e^{-s^2}\,ds$$

with the convention $\operatorname{sign}(0)=0$. The function $E(x)$ above is rapidly decaying with all derivatives, odd and smooth on $\mathbb{R}_x\setminus 0$. We may hence choose functions $f\in C_c^\infty(-\sqrt{2\nu_0},\sqrt{2\nu_0})$, $g\in C_c^\infty(\mathbb{R}_{<0})$ such that

$$f(x)+g(x)=E(x)\quad\text{for } x\leq 0.$$

Define the spectral measure

$$\mathfrak{M}_f(\lambda'):=\sum_{\lambda\in\operatorname{Spec}\left(\frac{D}{\sqrt{h}}\right)}f(\lambda)\delta(\lambda-\lambda').$$

It is clear that the expansion (1-7) to its first term may be written as

$$\mathfrak{M}_f*(\mathcal{F}_h^{-1}\theta_{\frac{1}{2}})(\lambda)=h^{-m-\frac{1}{2}}(f(\lambda)u_0(\lambda)+O(h^{\frac{1}{2}})),$$

where $\theta_{\frac{1}{2}}(x) = \theta(x/\sqrt{h})$ as before. Since both sides above involve Schwartz functions in λ, the remainder maybe replaced by $O(h^{\frac{1}{2}}/\langle\lambda\rangle^2)$. One may then integrate the equation to obtain

$$\int_{-\infty}^{0} d\lambda \int d\lambda' (\mathcal{F}_h^{-1}\theta_{\frac{1}{2}})(\lambda-\lambda')\mathfrak{M}_f(\lambda') = h^{-m-\frac{1}{2}}\left(\int_{-\infty}^{0} d\lambda f(\lambda)u_0(\lambda) + O(h^{\frac{1}{2}})\right). \tag{8-4}$$

Next we observe

$$\int_{-\infty}^{0} d\lambda (\mathcal{F}_h^{-1}\theta_{\frac{1}{2}})(\lambda-\lambda') = \int_{-\infty}^{0} dt \check{\theta}\left(t - \frac{\lambda'}{\sqrt{h}}\right) = 1_{(-\infty,0]}(\lambda') + O\left(\left\langle\frac{\lambda'}{\sqrt{h}}\right\rangle^{-\infty}\right). \tag{8-5}$$

While the Weyl law yields

$$\int d\lambda' \mathfrak{M}_f(\lambda') O\left(\left\langle\frac{\lambda'}{\sqrt{h}}\right\rangle^{-\infty}\right) = O(h^{-m}). \tag{8-6}$$

Substituting (8-5) and (8-6) into (8-4) gives

$$\sum_{\substack{\lambda\leq 0\\ \lambda\in\operatorname{Spec}\left(\frac{D}{\sqrt{h}}\right)}} f(\lambda) = h^{-m-\frac{1}{2}}\left(\int_{-\infty}^{0} d\lambda f(\lambda)u_0(\lambda)\right) + O(h^{-m}).$$

This combined with

$$\operatorname{tr} g\left(\frac{D}{\sqrt{h}}\right) = h^{-m-\frac{1}{2}}u_0(g) + O(h^{-m})$$

then gives

$$\sum_{\substack{\lambda\leq 0\\ \lambda\in\operatorname{Spec}\left(\frac{D}{\sqrt{h}}\right)}} E(\lambda) = h^{-m-\frac{1}{2}}\left(\int_{-\infty}^{0} d\lambda E(\lambda)u_0(\lambda)\right) + O(h^{-m}),$$

where the integral makes sense from the formula (7-34) for u_0. A similar formula for

$$\sum_{\substack{\lambda\geq 0\\ \lambda\in\operatorname{Spec}\left(\frac{D}{\sqrt{h}}\right)}} E(\lambda)$$

now gives

$$\operatorname{tr} E\left(\frac{D}{\sqrt{h}}\right) = h^{-m-\frac{1}{2}}\left(\int_{-\infty}^{\infty} d\lambda E(\lambda)u_0(\lambda)\right) + O(h^{-m}).$$

Since E is odd and u_0 is even from (7-34), the integral above is zero and hence $\eta_h = \operatorname{tr} E(D/\sqrt{h}) = O(h^{-m})$ from (8-3) as required. □

In the above proof we have used a Tauberian argument, as in [Dimassi and Sjöstrand 1999, Chapter 10]. A similar argument along with the trace expansion theorem, Theorem 1.3, also gives a true Weyl law in $O(\sqrt{h})$-sized intervals: the number of eigenvalues $N(-c\sqrt{h}, c\sqrt{h})$, $0 < c < \sqrt{2\nu_0}$, in the given interval satisfies

$$N(-c\sqrt{h}, c\sqrt{h}) = h^{-m-\frac{1}{2}}\left[\frac{2c}{(4\pi)^m}\int_X \nu^m\left(\prod_{j=1}^{m}\mu_j\right)dx\right] + O(h^{-m}). \tag{8-7}$$

The leading term of the above may possibly be obtained by squaring the Dirac operator and using the spectral estimates on an $O(h)$-sized interval near the critical level for D_h^2, as in [Brummelhuis et al. 1995].

8A. *Sharpness of the result.* Here, we finally show that the result Theorem 1.2 is sharp. The worst case example was already noted in [Savale 2014, Section 5] for η_h. To recall, we let Y be a complex manifold of dimension $2m$ with complex structure J and a Riemannian metric g^{TY}. Fix a positive, holomorphic, Hermitian line bundle $\mathcal{L} \to Y$. The curvature $F^{\mathcal{L}}$ of the Chern connection is thus a positive $(1,1)$-form. Let X be the total space of the unit circle bundle $S^1 \to X \xrightarrow{\pi} Y$ of $\mathcal{L}$. The Chern connection gives a splitting of the tangent bundle

$$TX = TS^1 \oplus \pi^* TY, \tag{8-8}$$

where TS^1 is the vertical tangent space spanned by the generator e of the S^1 action. Define a metric g^{TS^1} on TS^1 via $\|e\|_{g^{TS^1}} = 1$. A metric on X can now be given using the splitting (8-8) via

$$g^{TX} = g^{TS^1} \oplus \varepsilon^{-1} \pi^* g^{TY}$$

for any $\varepsilon > 0$. A spin structure on Y corresponds to a holomorphic, Hermitian square root $\mathcal{K}$ of the canonical line bundle $K_Y = \mathcal{K}^{\otimes 2}$. Fixing such a spin structure as well as the trivial spin structure on TS^1 gives a spin structure on X. Finally $a = e^* \in \Omega^1(X)$, while the auxiliary line bundle is chosen to be trivial $L = \mathbb{C}$ with the family of connections $\nabla^h = d + \frac{i}{h} a$. We now have the required family of Dirac operators D_h (1-2). One may check that (X^{2m+1}, a, g^{TX}, J) here gives a metric contact structure (1-4) and hence the assumption Definition 1.1 is satisfied.

Denote by $\Delta^p_{\bar\partial_k} : \Omega^{0,p}(X; \mathcal{K} \otimes \mathcal{L}^{\otimes k}) \longrightarrow \Omega^{0,p}(X; \mathcal{K} \otimes \mathcal{L}^{\otimes k})$ the Hodge Laplacian acting on $(0, p)$ forms on X. Its null-space is given by the cohomology $H^p(X; \mathcal{K} \otimes \mathcal{L}^{\otimes k})$ of the tensor product via Hodge theory. Let $e^{p,k}_\mu$ denote the dimension of a each positive eigenspace with eigenvalue $\frac{1}{2}\mu^2 \in \mathrm{Spec}^+(\Delta^p_{\bar\partial_k})$. The spectrum of D_h was now computed in Proposition 5.2 of [Savale 2014].

Proposition 8.1. *The spectrum of D_h is given by eigenvalues of the following types*:

- *Type* 1.

$$\lambda = (-1)^p h\left(k + \left(\varepsilon - \frac{m}{2}\right) - \frac{1}{h}\right), \tag{8-9}$$

  $0 \le p \le m,\ k \in \mathbb{Z}$, *with multiplicity* $\dim H^p(X; \mathcal{K} \otimes \mathcal{L}^{\otimes k})$.

- *Type* 2.

$$\lambda = h\left[\frac{1}{2}\left((-1)^{p+1}\varepsilon \pm \sqrt{\left(2k + \varepsilon(2p - m) - \frac{2}{h} + 1\right)^2 + 4\mu^2\varepsilon}\right)\right], \tag{8-10}$$

  $0 \le p \le m,\ k \in \mathbb{Z},\ \frac{1}{2}\mu^2 \in \mathrm{Spec}^+(\Delta^p_{\bar\partial_k})$, *with multiplicity* $d^{p,k}_\mu := e^{p,k}_\mu - e^{p-1,k}_\mu + \cdots + (-1)^p e^{0,k}_\mu$.

As observed in [Savale 2014], by choosing

$$\varepsilon < \inf_{k,p} \left\{\tfrac{1}{2}\mu^2 \in \mathrm{Spec}^+(\Delta^p_{\bar\partial_k})\right\},$$

the eigenvalues of type 2 are either positive or negative depending on the sign appearing in (8-10). Hence the dimension of the kernel k_h of D_h is now given by the eigenvalues of type 1:

$$k_h = \begin{cases} \dim H^*(X; \mathcal{K} \otimes \mathcal{L}^{\otimes k}), & \frac{1}{h} = k + \left(\varepsilon - \frac{m}{2}\right), \\ 0, & \text{otherwise}. \end{cases} \tag{8-11}$$

Now by a combination of Kodaira vanishing and Hirzebruch–Riemann–Roch,

$$\dim H^*(X; \mathcal{K} \otimes \mathcal{L}^{\otimes k}) = \dim H^0(X; \mathcal{K} \otimes \mathcal{L}^{\otimes k}) = \chi(X, \mathcal{K} \otimes \mathcal{L}^{\otimes k}) = \int_X \mathrm{ch}(\mathcal{K} \otimes \mathcal{L}^{\otimes k}) \,\mathrm{td}(X)$$

for $k \gg 0$, where $\chi(X, \mathcal{K} \otimes \mathcal{L}^{\otimes k})$, $\mathrm{ch}(\mathcal{K} \otimes \mathcal{L}^{\otimes k})$ and $\mathrm{td}(X)$ denote Euler characteristic, Chern character and Todd genus respectively. Hence (8-11), (8-12) show that the kernel and hence the counting function are discontinuous of order $O(h^{-m}) = k_h \le N(-ch, ch)$ in this example. A similar discontinuity of the eta invariant of $O(h^{-m})$ was proved in Theorem 5.3 of [Savale 2014].

Appendix: Some spectral estimates

In this appendix we prove some spectral estimates used in Sections 4 and 5; see [Helffer 1988, Section 4.1] for some related estimates.

Let H be a separable Hilbert space. Let $A : H \to H$ be a bounded self-adjoint operator. The resolvent set and the spectrum of A are defined to be

$$R(A) = \{\lambda \in \mathbb{C} \mid A - \lambda I \text{ is invertible}\},$$
$$\mathrm{Spec}(A) = \mathbb{C} \setminus R(A).$$

Since A is self-adjoint, $\mathrm{Spec}(A) \subset \mathbb{R}$. We may now define the following subsets of the spectrum:

$$\mathrm{EssSpec}(A) = \{\lambda \in \mathbb{C} \mid A - \lambda I \text{ is not Fredholm}\},$$
$$\mathrm{DiscSpec}(A) = \mathrm{Spec}(A) \setminus \mathrm{EssSpec}(A).$$

We shall consider $\mathrm{DiscSpec}(A)$ above as a multiset with the multiplicity function $m^A : \mathrm{DiscSpec}(A) \to \mathbb{N}_0$ defined by $m^A(\lambda) = \dim \ker(A)$. We may then find a countable set of orthonormal eigenvectors $v_1^A, v_2^A, v_3^A, \dots,$ with eigenvalues $\lambda_1^A \le \lambda_2^A \le \lambda_3^A \le \cdots$ such that $\mathrm{DiscSpec}(A)$ and $\{\lambda_1^A, \lambda_2^A, \dots\}$ are equal as multisets. Now let $[a,b] \subset \mathbb{R}$ be a finite closed interval such that $\mathrm{EssSpec}(A) \cap [a,b] = \varnothing$ (i.e., A has discrete spectrum in $[a,b]$). Then

$$H^A_{[a,b]} = \bigoplus_{\lambda \in \mathrm{Spec}(A) \cap [a,b]} \ker(A - \lambda)$$

is a finite-dimensional vector subspace of H. We denote by

$$\Pi^A_{[a,b]} : H \to H^A_{[a,b]} \subset H$$

the orthogonal projection onto $H^A_{[a,b]}$ and by $N^A_{[a,b]}$ the dimension of $H^A_{[a,b]}$. The operator $\rho(A) : H \to H$ may now be defined for any function $\rho \in C^0_c([a,b])$ by functional calculus.

Lemma A.1. *Let $v \in H$ and $\lambda \in [a,b]$. Assume there exists $\varepsilon > 0$ such that A has discrete spectrum in $[a-\sqrt{\varepsilon}, b+\sqrt{\varepsilon}]$ and $\|(A-\lambda)v\| \le \varepsilon\|v\|$. Then*

$$\|\Pi^A_{[a-\sqrt{\varepsilon},b+\sqrt{\varepsilon}]}v - v\| \le \sqrt{\varepsilon}\|v\|, \tag{A-1}$$

$$\|(\rho(A)-\rho(\lambda))v\| \le 3\sqrt{\varepsilon}\|\rho\|_{C^{0,1}}\|v\| \tag{A-2}$$

for any Holder continuous function $\rho \in C^{0,1}_c([a,b])$.

Proof. We abbreviate $\Pi = \Pi^A_{[a-\sqrt{\varepsilon},b+\sqrt{\varepsilon}]}$. Let $H_0 := H^A_{[a-\sqrt{\varepsilon},b+\sqrt{\varepsilon}]} = \Pi H$, which by assumption is a finite-dimensional vector space. Let $H_0^\perp$ be the orthogonal complement of H_0. By assumption, $\mathrm{Spec}((A-\lambda)^2|_{H_0^\perp}) \cap [-\varepsilon,\varepsilon] = \varnothing$. Hence by the mini-max principle for self-adjoint operators bounded from below [Dimassi and Sjöstrand 1999, Lemma 4.21], we have $\varepsilon \le (A-\lambda)^2|_{H_0^\perp}$. Hence

$$\begin{aligned}\|\Pi v - v\|^2\varepsilon &\le \|(A-\lambda)(\Pi v - v)\|^2 \\ &\le \|(A-\lambda)(\Pi v - v)\|^2 + \|(A-\lambda)\Pi v\|^2 = \|(A-\lambda)v\|^2 \le \varepsilon^2\|v\|^2\end{aligned}$$

since $(A-\lambda)(\Pi v - v)$ and $(A-\lambda)\Pi v$ are orthogonal. This gives

$$\|\Pi v - v\| < \sqrt{\varepsilon}\|v\|. \tag{A-3}$$

To prove (A-2) first note that $\|\Pi' v - v\| < \sqrt{\varepsilon}\|v\|$, for $\Pi' = \Pi^A_{[\lambda-\sqrt{\varepsilon},\lambda+\sqrt{\varepsilon}]}$, by the same argument. We now have

$$\begin{aligned}\|(\rho(A)-\rho(\lambda))v\| &\le \|(\rho(A)-\rho(\lambda))(\Pi' v - v)\| + \|(\rho(A)-\rho(\lambda))\Pi' v\| \\ &\le 2\sqrt{\varepsilon}\|\rho\|_{C^{0,1}}\|v\| + \sqrt{\varepsilon}\|\rho\|_{C^{0,1}}\|v\|. \qquad \square\end{aligned}$$

Before stating the next lemma we need the following definition.

Definition A.2. Given $0 < \varepsilon < 1$, a set of vectors $w_1, w_2, \dots, w_N \in H$ is called an ε-almost orthonormal set of eigenvectors (ε-AOSE for short) of A if

(1) $\left|\|w_j\|^2 - 1\right| < \varepsilon$ for all j,

(2) $|\langle w_j, w_k\rangle| < \varepsilon$ for all $j \ne k$,

(3) $\|(A-\mu_j)w_j\| < \varepsilon$ for some $\mu_j \in \mathbb{R}$, for all j.

Lemma A.3. *Assume that $H_0 \subset H$ has finite dimension M and is mapped onto itself by A. Let $w_1, w_2, \dots, w_N \in H_0$ be an ε-AOSE of A for some $\varepsilon < 1/(2(M+1))$. Then there exist orthonormal $w'_1, w'_2, \dots, w'_{M-N} \in H_0$ such that $\|(A-\mu'_j)w'_j\| < 4M\varepsilon$ for some $\mu'_j \in \mathbb{R}$, for all j. Furthermore $\langle w_j, w'_k\rangle = 0$ for each j,k.*

Proof. It follows from $\varepsilon < 1/(2(M+1))$ that $w_1, w_2, \dots, w_N$ are linearly independent. Let W denote their span and $W^\perp \subset H_0$ its orthogonal complement. Let $\Pi, \Pi^\perp$ be the orthogonal projections onto $W, W^\perp$ and consider the operator $A_0 := \Pi^\perp A \Pi^\perp : W^\perp \to W^\perp$. Let $w'_1, w'_2, \dots, w'_{M-N} \in W^\perp$ be an orthogonal basis of eigenvectors of A_0. Hence

$$\Pi^\perp A w'_j = \mu'_j w'_j$$

for some $\mu_j' \in \mathbb{R}$, for all j. Also

$$|\langle Aw_j', w_k\rangle| = |\langle w_j', (A-\mu_k)w_k\rangle| < \varepsilon.$$

It then follows that $\|\Pi Aw_j'\| \leq 2M\varepsilon\sqrt{1+\varepsilon} < 4M\varepsilon$ giving the result. $\square$

Lemma A.4. *Given $N \in \mathbb{N}$, let $0 < \varepsilon < (\|A\| + |a| + |b| + N + 1)^{-4}$. Let $w_1, w_2, \ldots, w_N \in H$ be an ε-AOSE for A. Assume that A has discrete spectrum in $[a - \varepsilon^{\frac{1}{8}}, b + \varepsilon^{\frac{1}{8}}]$. Then there exist orthonormal vectors $\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_N \in H$ which span the same subspace of H as $w_1, w_2, \ldots, w_N$. Moreover $\|w_j - \bar{w}_j\| < \sqrt{\varepsilon}$ and $\|(\rho(A) - \rho(\mu_j))\bar{w}_j\| \leq 3\varepsilon^{\frac{1}{8}}\|\rho\|_{C^{0,1}}$ for $1 \leq j \leq N$ and any Holder continuous function $\rho \in C_c^{0,1}([a,b])$.*

Proof. Again it follows easily that the vectors w_j, $1 \leq j \leq N$, are linearly independent. Let $W \subset H$ be their span and choose an orthonormal basis e_i, $1 \leq j \leq N$, for W. We write

$$w_j = \sum_{k=1}^{N} m_{jk}e_k.$$

If we consider the matrix $M = [m_{jk}]$, then assumptions (1) and (2) of Definition A.2 are equivalent to $|M^*M - I| < \varepsilon$. Consider the polar decomposition $M = UP$, where U is unitary and P is a positive semidefinite Hermitian matrix. We have $|P^*P - I| < \varepsilon$ and hence $\|P^*P - I\| < N\varepsilon$. Thus any eigenvalue λ^P of P, being nonnegative, satisfies $|\lambda^P - 1| < \varepsilon$ and we have $\|P - I\| < N\varepsilon$. Thus $\|M - U\| = \|UP - U\| < N\varepsilon$. If we now let $U = [u_{jk}]$ and $\bar{w}_j = \sum_{k=1}^N u_{jk}e_k$, then the $\bar{w}_j$ are clearly orthonormal and satisfy $\|w_j - \bar{w}_j\| < \sqrt{\varepsilon}$. This last inequality along with assumption (3) of Definition A.2 easily gives

$$\|(A - \mu_j)\bar{w}_j\| < \varepsilon^{\frac{1}{4}}.$$

Now Lemma A.1 gives

$$\|\Pi\bar{w}_j - \bar{w}_j\| < \varepsilon^{\frac{1}{8}}, \tag{A-4}$$

$$\|(\rho(A) - \rho(\mu_j))\bar{w}_j\| < 3\varepsilon^{\frac{1}{8}}\|\rho\|_{C^{0,1}}, \tag{A-5}$$

completing the proof. $\square$

Next, let H' be another separable Hilbert space. Let $U : H \to H'$ be a bounded operator. Let $B, D : H' \to H'$ and $C : H \to H$ be bounded self-adjoint operators. Define $A' = UAU^* : H' \to H'$, $B' = U^*BU : H \to H$, $C' = UCU^* : H' \to H'$ and $D' = U^*DU : H \to H$. In the next proposition we assume that there exists $\delta > 0$ such that A, A', B and B' have discrete spectrum in $[a - \delta, b + \delta]$. We also abbreviate $N^A = N^A_{[a-\delta,b+\delta]}$ and $\Pi^A = \Pi^A_{[a-\delta,b+\delta]}$ and similarly define $N^{A'}$, N^B, $N^{B'}$, $\Pi^{A'}$, Π^B, $\Pi^{B'}$.

Proposition A.5. *Suppose there exists $0 < \varepsilon < L^{-2048}$, with*

$$L = 25\big\{\|A\| + \|A'\| + \|B\| + \|B'\| + \|C\| + \|D\| + N^A + N^{A'} + N^B + N^{B'} + |a| + |b| + \delta^{-1} + 1\big\},$$

such that

(1) $\|(U^*U - I)\Pi^A\|\big(\|A\|\,\|U\| + 1\big) < \varepsilon$ *and* $\|(UU^* - I)\Pi^B\|\big(\|B\|\,\|U^*\| + 1\big) < \varepsilon$,

(2) $\|(A'-B)\Pi^{A'}\| < \varepsilon$ *and* $\|(A-B')\Pi^{B'}\| < \varepsilon$,

(3) $\|(C'-D)\Pi^{A}\| < \varepsilon$ *and* $\|(C-D')\Pi^{B}\| < \varepsilon$.

Then we have

$$\left|\mathrm{tr}[C\rho(A)] - \mathrm{tr}[D\rho(B)]\right| \le \varepsilon^{\frac{1}{2048}} \|\rho\|_{C^1}$$

for any $\rho \in C_c^1([a,b])$.

Proof. Let $(\mathrm{DiscSpec}(A), m^A) \cap [a,b] = \{\lambda^A_{a_1}, \lambda^A_{a_2}, \dots, \lambda^A_{a_N}\}$, with $N = N^A_{[a,b]}$, as multisets. Let $\rho^+(x) = \frac{1}{2}(\rho(x) + |\rho(x)|)$ and $\rho^-(x) = \frac{1}{2}(\rho(x) - |\rho(x)|)$. We then have $\rho^+, \rho^- \in C_c^{0,1}([a,b])$ with $\|\rho^+\|_{C^{0,1}} \le \|\rho\|_{C^1}$, $\|\rho^-\|_{C^{0,1}} \le \|\rho\|_{C^1}$. We further decompose $C = C^+ + C^-$, $D = D^+ + D^-$ into their positive and nonpositive parts. Clearly

$$\mathrm{tr}[C^+\rho^+(A)] = \sum_{j=1}^{N} \rho^+(\lambda_{a_j})\langle v_{a_j}, C^+ v_{a_j}\rangle.$$

Next we consider $w_j = U v_{a_j} \in H'$. From assumption (1) we have

$$\left\|(A' - \lambda_{a_j})w_j\right\| = \left\|(UAU^* - \lambda_{a_j})Uv_{a_j}\right\| \le \left\|(U^*U - I)\Pi^A_{[a,b]}\right\| \|A\| \|U\| < \varepsilon.$$

Similar estimates give $\left|\|w_j\|^2 - 1\right| < \varepsilon$ and $|\langle w_j, w_k\rangle| < \varepsilon$ for $j \ne k$. Now by Lemma A.1 we have $\|\Pi w_j - w_j\| < (2\varepsilon)^{\frac{1}{2}}$ with $\Pi = \Pi^{A'}_{[a-\sqrt{2\varepsilon}, b+\sqrt{2\varepsilon}]}$. Following this and using assumption (3) we have

$$\begin{aligned}\left\|(B - \lambda_{a_j})w_j\right\| &\le \left\|(A' - \lambda_{a_j})w_j\right\| + \left\|(B - A')\Pi w_j\right\| + \left\|(B - A')(\Pi w_j - w_j)\right\| \\ &\le \varepsilon + \varepsilon\sqrt{1+\varepsilon} + (2\varepsilon)^{\frac{1}{2}}(\|A'\| + \|B\|) \\ &< \varepsilon^{\frac{1}{4}} \le \varepsilon^{\frac{1}{8}}\|w_j\|.\end{aligned}$$

Next define $w^0_j := \Pi^B_{[a-\varepsilon^{1/16}, b+\varepsilon^{1/16}]} w_j$. By Lemma A.1,

$$\|w^0_j - w_j\| \le \varepsilon^{\frac{1}{16}}\|w_j\|. \tag{A-6}$$

From here it follows immediately that $w^0_1, w^0_2, \dots, w^0_N$ form an $\varepsilon^{\frac{1}{64}}$-AOSE of B. If we let $H_0 = H^B_{[a-\varepsilon^{1/16}, b+\varepsilon^{1/16}]}$, then by Lemma A.4 there exist orthonormal $\bar{w}_1, \bar{w}_2, \dots, \bar{w}_N \in H_0$ which span the same subspace of H_0 as the w^0_j. Furthermore

$$\|w^0_j - \bar{w}_j\| < \varepsilon^{\frac{1}{128}} \tag{A-7}$$

and $\|(\rho^+(B) - \rho^+(\lambda_{a_j}))\bar{w}_j\| \le 3\|\rho\|_{C^1}\varepsilon^{\frac{1}{512}}$. From (A-6) and (A-7) we also have $\|w_j - \bar{w}_j\| < \varepsilon^{\frac{1}{256}}$. From Lemma A.3 there exist orthonormal $w'_1, w'_2, \dots, w'_{M-N}$ with $M = N^B_{[a-\varepsilon^{1/16}, b+\varepsilon^{1/16}]}$ such that $\langle w'_i, \bar{w}_j\rangle = 0$ and $\|(B - \mu'_j)w'_j\| < 4M\varepsilon^{\frac{1}{64}} < \varepsilon^{\frac{1}{128}}$. Hence by Lemma A.1, $\|(\rho^+(B) - \rho^+(\mu'_j))w'_j\| \le$

$3\|\rho\|_{C^1}\varepsilon^{\frac{1}{256}}$. We now have

$$\begin{aligned}
\operatorname{tr}[D^+\rho^+(B)] &= \sum_{j=1}^{N}\langle\bar{w}_j, D^+\rho^+(B)\bar{w}_j\rangle + \sum_{j=1}^{M-N}\langle w_j', D^+\rho^+(B)w_j'\rangle \\
&\geq \sum_{j=1}^{N}\rho^+(\lambda_{a_j})\langle\bar{w}_j, D^+\bar{w}_j\rangle + \sum_{j=1}^{M-N}\rho^+(\mu_j')\langle w_j', D^+w_j'\rangle - 3\varepsilon^{\frac{1}{512}}M\|D\|\|\rho\|_{C^1} \\
&\geq \sum_{j=1}^{N}\rho^+(\lambda_{a_j})\langle\bar{w}_j, D^+\bar{w}_j\rangle - 3\varepsilon^{\frac{1}{512}}M\|D\|\|\rho\|_{C^1} \\
&\geq \sum_{j=1}^{N}\rho^+(\lambda_{a_j})\langle w_j, D^+w_j\rangle - 6\varepsilon^{\frac{1}{512}}M\|D\|\|\rho\|_{C^1} \\
&\geq \sum_{j=1}^{N}\rho^+(\lambda_{a_j})\langle v_{a_j}, C^+v_{a_j}\rangle - 6\varepsilon^{\frac{1}{512}}M(\|D\|+1)\|\rho\|_{C^1} \\
&\geq \operatorname{tr}[C^+\rho^+(A)] - \varepsilon^{\frac{1}{1024}}\|\rho\|_{C^1}.
\end{aligned}$$

Reversing the roles of H and H' gives

$$\left|\operatorname{tr}[D^+\rho^+(B)] - \operatorname{tr}[C^+\rho^+(A)]\right| \leq \varepsilon^{\frac{1}{1024}}\|\rho\|_{C^1}.$$

Similar estimates with $C^+\rho^-(A)$, $C^-\rho^+(A)$ and $C^-\rho^-(A)$ give the result. ☐

Finally, we now give a criterion implying the discreteness of spectrum for pseudodifferential operators required by the preceding propositions in this appendix.

Proposition A.6. *Let $A \in \Psi^m_{\mathrm{cl}}(\mathbb{R}^n;\mathbb{C}^l)$ and $I = [a,b] \subset \mathbb{R}$ be a closed interval such that the I-energy band*

$$\Sigma^A_I := \bigcup_{\lambda\in I}\Sigma^A_\lambda$$

is bounded. Then for $h < h_0$ sufficiently small

$$\operatorname{EssSpec}(A) \cap I = \varnothing.$$

Proof. Let $\sigma(A) = a(x,\xi) \in C^\infty(\mathbb{R}^{2n})$ and $\Sigma_I(a) \subset B_R$ be some open ball of finite radius R around the origin. For $\lambda \in I$ and $(x,\xi) \notin B_R$, we hence have that $a_{-1} := (a(x,\xi)-\lambda)^{-1}$ exists. Let $\chi \in C_c^\infty(-4R,4R)$ such that $\chi(x) = 1$ for $x < 2R$. Set $\phi(x) = 1-\chi(x)$ and define

$$A_{-1} = \left[\phi(|(x,\xi)|)a_{-1}(x,\xi)\right]^W \in \Psi^0_{\mathrm{cl}}(\mathbb{R}^n;\mathbb{C}^l).$$

Then since it has vanishing symbol, we have

$$(A-\lambda)A_{-1} - (I - \chi(|(x,\xi)|)^W) = hR \in h\Psi^0_{\mathrm{cl}}(\mathbb{R}^n;\mathbb{C}^l).$$

Next, we clearly have $I + hR$ is invertible for $h < h_0$ sufficiently small. Also, $\chi(|(x,\xi)|)^W$ is trace class by [Hörmander 1994, Lemma 19.3.2]. Hence if $S := A_{-1}(I + hR)^{-1}$, then $(A-\lambda)S - I$ is trace class. By a similar argument, $S(A-\lambda) - I$ is trace class. Hence by Proposition 19.1.14 of [Hörmander 1994], $A-\lambda$ is Fredholm. $\square$

Acknowledgement

The author would like to thank the referees for constructive comments and suggestions to improve the paper.

References

[Atiyah et al. 1975] M. F. Atiyah, V. K. Patodi, and I. M. Singer, "Spectral asymmetry and Riemannian geometry, I", *Math. Proc. Cambridge Philos. Soc.* **77** (1975), 43–69. MR Zbl

[Atiyah et al. 1976] M. F. Atiyah, V. K. Patodi, and I. M. Singer, "Spectral asymmetry and Riemannian geometry, III", *Math. Proc. Cambridge Philos. Soc.* **79**:1 (1976), 71–99. MR Zbl

[Berline et al. 2004] N. Berline, E. Getzler, and M. Vergne, *Heat kernels and Dirac operators*, Springer, 2004. MR Zbl

[Bismut 1987] J.-M. Bismut, "Demailly's asymptotic Morse inequalities: a heat equation proof", *J. Funct. Anal.* **72**:2 (1987), 263–278. MR Zbl

[Bismut and Freed 1986] J.-M. Bismut and D. S. Freed, "The analysis of elliptic families, II: Dirac operators, eta invariants, and the holonomy theorem", *Comm. Math. Phys.* **107**:1 (1986), 103–163. MR Zbl

[Brummelhuis et al. 1995] R. Brummelhuis, T. Paul, and A. Uribe, "Spectral estimates around a critical level", *Duke Math. J.* **78**:3 (1995), 477–530. MR Zbl

[Camus 2004] B. Camus, "A semi-classical trace formula at a non-degenerate critical level", *J. Funct. Anal.* **208**:2 (2004), 446–481. MR Zbl

[Charles and Vũ Ngọc 2008] L. Charles and S. Vũ Ngọc, "Spectral asymptotics via the semiclassical Birkhoff normal form", *Duke Math. J.* **143**:3 (2008), 463–511. MR Zbl

[Dimassi and Sjöstrand 1996] M. Dimassi and J. Sjöstrand, "Trace asymptotics via almost analytic extensions", pp. 126–142 in *Partial differential equations and mathematical physics* (Copenhagen, 1995; Lund, 1995), edited by L. Hörmander and A. Melin, Progr. Nonlinear Differential Equations Appl. **21**, Birkhäuser, Boston, 1996. MR Zbl

[Dimassi and Sjöstrand 1999] M. Dimassi and J. Sjöstrand, *Spectral asymptotics in the semi-classical limit*, London Mathematical Society Lecture Note Series **268**, Cambridge University Press, 1999. MR Zbl

[Emmrich and Weinstein 1996] C. Emmrich and A. Weinstein, "Geometry of the transport equation in multicomponent WKB approximations", *Comm. Math. Phys.* **176**:3 (1996), 701–711. MR Zbl

[Guillemin] V. Guillemin, "Fourier integral operators for systems", unpublished preprint, available at http://msp.org/extras/Guillemin/V.Guillemin-Fourier_Integral_Operators-1974.pdf.

[Guillemin and Sternberg 2013] V. Guillemin and S. Sternberg, *Semi-classical analysis*, International Press, Boston, 2013. MR Zbl

[Helffer 1988] B. Helffer, *Semi-classical analysis for the Schrödinger operator and applications*, Lecture Notes in Mathematics **1336**, Springer, 1988. MR Zbl

[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 Zbl

[Helffer et al. 2016] B. Helffer, Y. Kordyukov, N. Raymond, and S. Vũ Ngọc, "Magnetic wells in dimension three", *Anal. PDE* **9**:7 (2016), 1575–1608. MR Zbl

[Hörmander 1994] L. Hörmander, *The analysis of linear partial differential operators, III: Pseudo-differential operators*, Grundlehren der Mathematischen Wissenschaften **274**, Springer, 1994. MR

[Ivrii 1998] V. Ivrii, *Microlocal analysis and precise spectral asymptotics*, Springer, 1998. MR Zbl

[Ivrii 2017] V. Ivrii, "Microlocal analysis and sharp spectral asymptotics", book in progress, 2017, available at http://www.math.toronto.edu/ivrii/monsterbook.pdf.

[Khuat-Duy 1997] D. Khuat-Duy, "A semi-classical trace formula for Schrödinger operators in the case of a critical energy level", *J. Funct. Anal.* **146**:2 (1997), 299–351. MR Zbl

[Ma and Marinescu 2007] X. Ma and G. Marinescu, *Holomorphic Morse inequalities and Bergman kernels*, Progress in Mathematics **254**, Birkhäuser, Basel, 2007. MR Zbl

[Maslov and Fedoriuk 1981] V. P. Maslov and M. V. Fedoriuk, *Semiclassical approximation in quantum mechanics*, Mathematical Physics and Applied Mathematics **7**, D. Reidel, Dordrecht, 1981. MR

[Raymond and Vũ Ngọc 2015] N. Raymond and S. Vũ Ngọc, "Geometry and spectrum in 2D magnetic wells", *Ann. Inst. Fourier (Grenoble)* **65**:1 (2015), 137–169. MR Zbl

[Robert 1987] D. Robert, *Autour de l'approximation semi-classique*, Progress in Mathematics **68**, Birkhäuser, Boston, 1987. MR Zbl

[Sandoval 1999] M. R. Sandoval, "Wave-trace asymptotics for operators of Dirac type", *Comm. Partial Differential Equations* **24**:9-10 (1999), 1903–1944. MR Zbl

[Savale 2014] N. Savale, "Asymptotics of the eta invariant", *Comm. Math. Phys.* **332**:2 (2014), 847–884. MR Zbl

[Taubes 2007] C. H. Taubes, "The Seiberg–Witten equations and the Weinstein conjecture", *Geom. Topol.* **11** (2007), 2117–2202. MR Zbl

[Tsai 2014] C.-J. Tsai, "Asymptotic spectral flow for Dirac operations of disjoint Dehn twists", *Asian J. Math.* **18**:4 (2014), 633–685. MR Zbl

[Zworski 2012] M. Zworski, *Semiclassical analysis*, Graduate Studies in Mathematics **138**, Amer. Math. Soc., Providence, RI, 2012. MR Zbl

Received 1 Oct 2015. Revised 19 May 2017. Accepted 20 Jun 2017.

NIKHIL SAVALE: nsavale@nd.edu
Department of Mathematics, University of Notre Dame, Notre Dame, IN, United States

ANALYSIS AND PDE
Vol. 10, No. 8, 2017
dx.doi.org/10.2140/apde.2017.10.1845

INCOMPRESSIBLE IMMISCIBLE MULTIPHASE FLOWS IN POROUS MEDIA A VARIATIONAL APPROACH

CLÉMENT CANCÈS, THOMAS O. GALLOUËT AND LÉONARD MONSAINGEON

We describe the competitive motion of $N+1$ incompressible immiscible phases within a porous medium as the gradient flow of a singular energy in the space of nonnegative measures with prescribed masses, endowed with some tensorial Wasserstein distance. We show the convergence of the approximation obtained by a minimization scheme á la R. Jordan, D. Kinderlehrer and F. Otto (*SIAM J. Math. Anal.* **29**:1 (1998) 1–17). This allows us to obtain a new existence result for a physically well-established system of PDEs consisting of the Darcy–Muskat law for each phase, N capillary pressure relations, and a constraint on the volume occupied by the fluid. Our study does not require the introduction of any global or complementary pressure.

1. Introduction

Equations for multiphase flows in porous media. We consider a convex open bounded set $\Omega \subset \mathbb{R}^d$ representing a porous medium; $N+1$ incompressible and immiscible phases, labeled by subscripts $i \in \{0, \dots, N\}$ are supposed to flow within the pores. Let us present now some classical equations that describe the motion of such a mixture. The physical justification of these equations can be found, for instance, in [Bear and Bachmat 1990, Chapter 5]. Let $T > 0$ be an arbitrary finite time horizon. We denote by $s_i : \Omega \times (0, T) =: Q \to [0, 1]$ the content of the phase i, i.e., the volume ratio of the phase i compared to all the phases and the solid matrix, and by $\boldsymbol{v}_i$ the filtration speed of the phase i. Then the conservation of the volume of each phase can be written as

$$\partial_t s_i + \nabla \cdot (s_i \boldsymbol{v}_i) = 0 \quad \text{in } Q,\ \forall i \in \{0, \dots, N\}. \tag{1}$$

The filtration speed of each phase is assumed to be given by Darcy's law

$$\boldsymbol{v}_i = -\frac{1}{\mu_i}\mathbb{K}(\nabla p_i - \rho_i \boldsymbol{g}) \quad \text{in } Q,\ \forall i \in \{0, \dots, N\}. \tag{2}$$

In the above relation, $\boldsymbol{g}$ is the gravity vector, μ_i denotes the constant viscosity of the phase i, p_i its pressure, and ρ_i its density. The intrinsic permeability tensor $\mathbb{K} : \overline{\Omega} \to \mathbb{R}^{d\times d}$ is supposed to be smooth, symmetric, that is, $\mathbb{K} = \mathbb{K}^T$, and uniformly positive definite: there exist $\kappa_\star, \kappa^\star > 0$ such that

$$\kappa_\star |\boldsymbol{\xi}|^2 \le \mathbb{K}(\boldsymbol{x})\boldsymbol{\xi} \cdot \boldsymbol{\xi} \le \kappa^\star |\boldsymbol{\xi}|^2 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^d,\ \forall \boldsymbol{x} \in \overline{\Omega}. \tag{3}$$

MSC2010: 35K65, 35A15, 49K20, 76S05.
Keywords: multiphase porous media flows, Wasserstein gradient flows, constrained parabolic system, minimizing movement scheme.

The pore volume is supposed to be saturated by the fluid mixture

$$\sigma := \sum_{i=0}^{N} s_i = \omega(\boldsymbol{x}) \quad \text{a.e. in } Q, \tag{4}$$

where the porosity $\omega : \overline{\Omega} \to (0, 1)$ of the surrounding porous matrix is assumed to be smooth. In particular, there exists $0 < \omega_\star \le \omega^\star$ such that $\omega_\star \le \omega(\boldsymbol{x}) \le \omega^\star$ for all $\boldsymbol{x} \in \overline{\Omega}$. In what follows, we set $\boldsymbol{s} = (s_0, \dots, s_N)$,

$$\Delta(\boldsymbol{x}) = \left\{ \boldsymbol{s} \in (\mathbb{R}_+)^{N+1} \;\middle|\; \textstyle\sum_{i=0}^{N} s_i = \omega(\boldsymbol{x}) \right\},$$

and

$$\boldsymbol{\mathcal{X}} = \{ \boldsymbol{s} \in L^1(\Omega; \mathbb{R}_+^{N+1}) \mid \boldsymbol{s}(\boldsymbol{x}) \in \Delta(\boldsymbol{x}) \text{ a.e. in } \Omega \}.$$

There is an obvious one-to-one mapping between the sets $\Delta(\boldsymbol{x})$ and

$$\Delta^*(\boldsymbol{x}) = \left\{ \boldsymbol{s}^* = (s_1, \dots, s_N) \in (\mathbb{R}_+)^N \;\middle|\; \textstyle\sum_{i=1}^{N} s_i \le \omega(\boldsymbol{x}) \right\},$$

and consequently also between $\boldsymbol{\mathcal{X}}$ and

$$\boldsymbol{\mathcal{X}}^* = \{ \boldsymbol{s}^* \in L^1(\Omega; \mathbb{R}_+^{N}) \mid \boldsymbol{s}^*(\boldsymbol{x}) \in \Delta^*(\boldsymbol{x}) \text{ a.e. in } \Omega \}.$$

In what follows, we set $\boldsymbol{\Upsilon} = \bigcup_{\boldsymbol{x} \in \overline{\Omega}} \Delta^*(\boldsymbol{x}) \times \{\boldsymbol{x}\}$.

In order to close the system, we impose N capillary pressure relations

$$p_i - p_0 = \pi_i(\boldsymbol{s}^*, \boldsymbol{x}) \quad \text{a.e. in } Q, \ \forall i \in \{1, \dots, N\}, \tag{5}$$

where the capillary pressure functions $\pi_i : \boldsymbol{\Upsilon} \to \mathbb{R}$ are assumed to be continuously differentiable and to derive from a strictly convex potential $\Pi : \boldsymbol{\Upsilon} \to \mathbb{R}_+$; that is,

$$\pi_i(\boldsymbol{s}^*, \boldsymbol{x}) = \frac{\partial \Pi}{\partial s_i}(\boldsymbol{s}^*, \boldsymbol{x}) \quad \forall i \in \{1, \dots, N\}.$$

We assume that Π is uniformly convex with respect to its first variable. More precisely, we assume that there exist two positive constants $\varpi_\star$ and $\varpi^\star$ such that, for all $\boldsymbol{x} \in \overline{\Omega}$ and all $\boldsymbol{s}^*, \hat{\boldsymbol{s}}^* \in \Delta^*(\boldsymbol{x})$, one has

$$\tfrac{1}{2}\varpi^\star |\hat{\boldsymbol{s}}^* - \boldsymbol{s}^*|^2 \ge \Pi(\hat{\boldsymbol{s}}^*, \boldsymbol{x}) - \Pi(\boldsymbol{s}^*, \boldsymbol{x}) - \boldsymbol{\pi}(\boldsymbol{s}^*, \boldsymbol{x}) \cdot (\hat{\boldsymbol{s}}^* - \boldsymbol{s}^*) \ge \tfrac{1}{2}\varpi_\star |\hat{\boldsymbol{s}}^* - \boldsymbol{s}^*|^2, \tag{6}$$

where we introduced the notation

$$\boldsymbol{\pi} : \boldsymbol{\Upsilon} \to \mathbb{R}^N, \quad (\boldsymbol{s}^*, \boldsymbol{x}) \mapsto \boldsymbol{\pi}(\boldsymbol{s}^*, \boldsymbol{x}) = (\pi_1(\boldsymbol{s}^*, \boldsymbol{x}), \dots, \pi_N(\boldsymbol{s}^*, \boldsymbol{x})).$$

The relation (6) implies that $\boldsymbol{\pi}$ is monotone and injective with respect to its first variable. Denoting by

$$z \mapsto \boldsymbol{\phi}(z, \boldsymbol{x}) = (\phi_1(z, \boldsymbol{x}), \dots, \phi_N(z, \boldsymbol{x})) \in \Delta^*(\boldsymbol{x})$$

the inverse of $\boldsymbol{\pi}(\cdot, \boldsymbol{x})$, it follows from (6) that

$$0 < \frac{1}{\varpi^\star} \le \mathbb{J}_z \boldsymbol{\phi}(z, \boldsymbol{x}) \le \frac{1}{\varpi_\star} \quad \forall \boldsymbol{x} \in \overline{\Omega}, \ \forall z \in \boldsymbol{\pi}(\Delta^*(\boldsymbol{x}), \boldsymbol{x}), \tag{7}$$

where $\mathbb{J}_z$ stands for the Jacobian with respect to z and the above inequality should be understood in the sense of positive definite matrices. Moreover, due to the regularity of $\boldsymbol{\pi}$ with respect to the space variable, there exists $M_{\boldsymbol{\phi}} > 0$ such that

$$|\nabla_{\boldsymbol{x}} \boldsymbol{\phi}(z, \boldsymbol{x})| \leq M_{\boldsymbol{\phi}} \quad \forall \boldsymbol{x} \in \overline{\Omega},\ \forall z \in \boldsymbol{\pi}(\Delta^*(\boldsymbol{x}), \boldsymbol{x}), \tag{8}$$

where $\nabla_{\boldsymbol{x}}$ denotes the gradient with respect to the second variable only.

The problem is complemented with no-flux boundary conditions

$$\boldsymbol{v}_i \cdot \boldsymbol{n} = 0 \quad \text{on } \partial\Omega \times (0, T),\ \forall i \in \{0, \dots, N\}, \tag{9}$$

and by the initial content profile $\boldsymbol{s}^0 = (s_0^0, \dots, s_N^0) \in \boldsymbol{\mathcal{X}}$:

$$s_i(\,\cdot\,, 0) = s_i^0 \quad \forall i \in \{0, \dots, N\}, \text{ with } \sum_{i=0}^{N} s_i^0 = \omega \ \text{ a.e. in } \Omega. \tag{10}$$

Since we did not consider sources, and since we imposed no-flux boundary conditions, the volume of each phase is conserved along time:

$$\int_\Omega s_i(\boldsymbol{x}, t)\,\mathrm{d}\boldsymbol{x} = \int_\Omega s_i^0(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} =: m_i > 0 \quad \forall i \in \{0, \dots, N\}. \tag{11}$$

We can now give a proper definition of what we call a weak solution to the problem (1)–(2), (4)–(5), and (9)–(10).

Definition 1.1 (weak solution). A measurable function $\boldsymbol{s} : Q \to (\mathbb{R}_+)^{N+1}$ is said to be a weak solution if $\boldsymbol{s} \in \Delta$ a.e. in Q, if there exists $\boldsymbol{p} = (p_0, \dots, p_N) \in L^2((0, T); H^1(\Omega))^{N+1}$ such that the relations (5) hold, and such that, for all $\phi \in C_c^\infty(\overline{\Omega} \times [0, T))$ and all $i \in \{0, \dots, N\}$, one has

$$\iint_Q s_i\, \partial_t \phi \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t + \int_\Omega s_i^0 \phi(\,\cdot\,, 0)\,\mathrm{d}\boldsymbol{x} - \iint_Q \frac{s_i}{\mu_i} \mathbb{K}(\nabla p_i - \rho_i \boldsymbol{g}) \cdot \nabla\phi \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t = 0. \tag{12}$$

Wasserstein gradient flow of the energy.

Energy of a configuration. First, we extend the convex function $\Pi : \boldsymbol{\Upsilon} \to [0, +\infty]$, called *capillary energy density*, to a convex function (still denoted by) $\Pi : \mathbb{R}^{N+1} \times \overline{\Omega} \to [0, +\infty]$ by setting

$$\Pi(\boldsymbol{s}, \boldsymbol{x}) = \begin{cases} \Pi\left(\omega \dfrac{\boldsymbol{s}^*}{\sigma}, \boldsymbol{x}\right) = \Pi\left(\omega \dfrac{s_1}{\sigma}, \dots, \omega \dfrac{s_N}{\sigma}, \boldsymbol{x}\right) & \text{if } \boldsymbol{s} \in \mathbb{R}_+^{N+1} \text{ and } \sigma \leq \omega(\boldsymbol{x}), \\ +\infty & \text{otherwise,} \end{cases}$$

σ being defined by (4). The extension of Π by $+\infty$ where $\sigma > \omega$ is natural because of the incompressibility of the fluid mixture. The extension to $\{\sigma < \omega\} \cup \mathbb{R}_+^{N+1}$ is designed so that the energy density only depends on the relative composition of the fluid mixture. However, this extension is somehow arbitrary, and, as it will appear in the sequel, it has no influence on the flow since the solution $\boldsymbol{s}$ remains in $\boldsymbol{\mathcal{X}}$; i.e., $\sum_{i=0}^N s_i = \omega$. In our previous note [Cancès et al. 2015] the appearance of void $\sigma < \omega$ was directly prohibited by a penalization in the energy.

The second part in the energy comes from the gravity. In order to lighten the notation, we introduce the functions

$$\Psi_i : \overline{\Omega} \to \mathbb{R}_+, \quad \boldsymbol{x} \mapsto -\rho_i \boldsymbol{g} \cdot \boldsymbol{x}, \quad \forall i \in \{0, \dots, N\},$$

and

$$\boldsymbol{\Psi} : \overline{\Omega} \to \mathbb{R}_+^{N+1}, \quad \boldsymbol{x} \mapsto (\Psi_0(\boldsymbol{x}), \dots, \Psi_N(\boldsymbol{x})).$$

The fact that Ψ_i can be assumed to be positive comes from the fact that Ω is bounded. Even though the physically relevant potentials are indeed the gravitationals $\Psi_i(\boldsymbol{x}) = -\rho_i \boldsymbol{g} \cdot \boldsymbol{x}$, the subsequent analysis allows for a broader class of external potentials and for the sake of generality we shall therefore consider arbitrary $\Psi_i \in C^1(\overline{\Omega})$ in the sequel.

We can now define the convex energy functional $\mathcal{E} : L^1(\Omega, \mathbb{R}^{N+1}) \to \mathbb{R} \cup \{+\infty\}$ by adding the capillary energy to the gravitational one:

$$\mathcal{E}(\boldsymbol{s}) = \int_\Omega (\Pi(\boldsymbol{s}, \boldsymbol{x}) + \boldsymbol{s} \cdot \boldsymbol{\Psi}) \,\mathrm{d}\boldsymbol{x} \geq 0 \quad \forall \boldsymbol{s} \in L^1(\Omega; \mathbb{R}^{N+1}). \tag{13}$$

Note moreover that $\mathcal{E}(\boldsymbol{s}) < \infty$ if and only if $\boldsymbol{s} \geq 0$ and $\sigma \leq \omega$ a.e. in Ω. It follows from the mass conservation (11) that

$$\int_\Omega \sigma(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} = \sum_{i=0}^{N} m_i = \int_\Omega \omega(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}.$$

Assume that there exists a nonnegligible subset A of Ω such that $\sigma < \omega$ on A; then necessarily, there must be a nonnegligible subset B of Ω such that $\sigma > \omega$ so that the above equation holds, and hence $\mathcal{E}(\boldsymbol{s}) = +\infty$. Therefore,

$$\mathcal{E}(\boldsymbol{s}) < \infty \iff \boldsymbol{s} \in \boldsymbol{\mathcal{X}}. \tag{14}$$

Let $\boldsymbol{p} = (p_0, \dots, p_N) : \Omega \to \mathbb{R}^{N+1}$ be such that $\boldsymbol{p} \in \partial_{\boldsymbol{s}} \Pi(\boldsymbol{s}, \boldsymbol{x})$ for a.e. $\boldsymbol{x}$ in Ω. Then, defining $h_i = p_i + \Psi_i(\boldsymbol{x})$ for all $i \in \{0, \dots, N\}$ and $\boldsymbol{h} = (h_i)_{0 \leq i \leq N}$, we have $\boldsymbol{h}$ belongs to the subdifferential $\partial_{\boldsymbol{s}} \mathcal{E}(\boldsymbol{s})$ of $\mathcal{E}$ at $\boldsymbol{s}$; i.e.,

$$\mathcal{E}(\hat{\boldsymbol{s}}) \geq \mathcal{E}(\boldsymbol{s}) + \sum_{i=0}^{N} \int_\Omega h_i (\hat{s}_i - s_i) \,\mathrm{d}\boldsymbol{x} \quad \forall \hat{\boldsymbol{s}} \in L^1(\Omega; \mathbb{R}^{N+1}).$$

The reverse inclusion also holds; hence

$$\partial_{\boldsymbol{s}} \mathcal{E}(\boldsymbol{s}) = \{\boldsymbol{h} : \Omega \to \mathbb{R}^{N+1} \mid h_i - \Psi_i(\boldsymbol{x}) \in \partial_{\boldsymbol{s}} \Pi(\boldsymbol{s}, \boldsymbol{x}) \text{ for a.e. } \boldsymbol{x} \in \Omega\}. \tag{15}$$

Thanks to (14), we know that a configuration $\boldsymbol{s}$ has finite energy if and only if $\boldsymbol{s} \in \boldsymbol{\mathcal{X}}$. Since we are interested in finite energy configurations, it is relevant to consider the restriction of $\mathcal{E}$ to $\boldsymbol{\mathcal{X}}$. Then using the one-to-one mapping between $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{X}}^*$, we define the energy of a configuration $\boldsymbol{s}^* \in \boldsymbol{\mathcal{X}}^*$, which we denote by $\mathcal{E}(\boldsymbol{s}^*)$, by setting $\mathcal{E}(\boldsymbol{s}^*) = \mathcal{E}(\boldsymbol{s})$, where $\boldsymbol{s}$ is the unique element of $\boldsymbol{\mathcal{X}}$ corresponding to $\boldsymbol{s}^* \in \boldsymbol{\mathcal{X}}^*$.

Geometry of Ω and Wasserstein distance. Inspired by [Lisini 2009], where heterogeneous anisotropic degenerate parabolic equations are studied from a variational point of view, we introduce $N+1$ distances on Ω that take into account the permeability of the porous medium and the phase viscosities. Given two points $\boldsymbol{x}$, $\boldsymbol{y}$ in Ω, we denote by

$$P(\boldsymbol{x},\boldsymbol{y}) = \{\boldsymbol{\gamma} \in C^1([0,1];\Omega) \mid \boldsymbol{\gamma}(0)=\boldsymbol{x} \text{ and } \boldsymbol{\gamma}(1)=\boldsymbol{y}\}$$

the set of the smooth paths joining $\boldsymbol{x}$ to $\boldsymbol{y}$, and we introduce distances d_i, $i \in \{0,\dots,N\}$, between elements on Ω by setting

$$d_i(\boldsymbol{x},\boldsymbol{y}) = \inf_{\boldsymbol{\gamma}\in P(\boldsymbol{x},\boldsymbol{y})} \left(\int_0^1 \mu_i \mathbb{K}^{-1}(\boldsymbol{\gamma}(\tau))\boldsymbol{\gamma}'(\tau)\cdot\boldsymbol{\gamma}'(\tau)\,\mathrm{d}\tau \right)^{1/2} \quad \forall (\boldsymbol{x},\boldsymbol{y}) \in \overline{\Omega}. \tag{16}$$

It follows from (3) that

$$\sqrt{\frac{\mu_i}{\kappa^\star}}|\boldsymbol{x}-\boldsymbol{y}| \le d_i(\boldsymbol{x},\boldsymbol{y}) \le \sqrt{\frac{\mu_i}{\kappa_\star}}|\boldsymbol{x}-\boldsymbol{y}| \quad \forall (\boldsymbol{x},\boldsymbol{y}) \in \overline{\Omega}^2. \tag{17}$$

For $i \in \{0,\dots,N\}$ we define

$$\mathcal{A}_i = \left\{ s_i \in L^1(\Omega;\mathbb{R}_+) \;\middle|\; \textstyle\int_\Omega s_i\,\mathrm{d}\boldsymbol{x} = m_i \right\}.$$

Given $s_i, \hat{s}_i \in \mathcal{A}_i$, the set of admissible transport plans between s_i and $\hat{s}_i$ is given by

$$\Gamma_i(s_i,\hat{s}_i) = \left\{ \theta_i \in \mathcal{M}_+(\Omega\times\Omega) \;\middle|\; \theta_i(\Omega\times\Omega)=m_i,\ \theta_i^{(1)}=s_i \text{ and } \theta_i^{(2)}=\hat{s}_i \right\},$$

where $\mathcal{M}_+(\Omega\times\Omega)$ stands for the set of Borel measures on $\Omega\times\Omega$ and $\theta_i^{(k)}$ is the k-th marginal of the measure θ_i. We define the quadratic Wasserstein distance W_i on $\mathcal{A}_i$ by setting

$$W_i(s_i,\hat{s}_i) = \left(\inf_{\theta_i\in\Gamma(s_i,\hat{s}_i)} \iint_{\Omega\times\Omega} d_i(\boldsymbol{x},\boldsymbol{y})^2\,\mathrm{d}\theta_i(\boldsymbol{x},\boldsymbol{y}) \right)^{1/2}. \tag{18}$$

Due to the permeability tensor $\mathbb{K}(\boldsymbol{x})$, the porous medium Ω might be heterogeneous and anisotropic. Therefore, some directions and areas might be privileged by the fluid motions. This is encoded in the distances d_i we put on Ω. Moreover, the more viscous the phase is, the more costly are its displacements, hence the μ_i in the definition (16) of d_i. But it follows from (17) that

$$\sqrt{\frac{\mu_i}{\kappa^\star}} W_{\mathrm{ref}}(s_i,\hat{s}_i) \le W_i(s_i,\hat{s}_i) \le \sqrt{\frac{\mu_i}{\kappa_\star}} W_{\mathrm{ref}}(s_i,\hat{s}_i) \quad \forall s_i,\hat{s}_i \in \mathcal{A}_i, \tag{19}$$

where W_{ref} denotes the classical quadratic Wasserstein distance defined by

$$W_{\mathrm{ref}}(s_i,\hat{s}_i) = \left(\inf_{\theta_i\in\Gamma(s_i,\hat{s}_i)} \iint_{\Omega\times\Omega} |\boldsymbol{x}-\boldsymbol{y}|^2\,\mathrm{d}\theta_i(\boldsymbol{x},\boldsymbol{y}) \right)^{1/2}. \tag{20}$$

With the phase Wasserstein distances $(W_i)_{0\le i\le N}$ at hand, we can define the global Wasserstein distance $\boldsymbol{W}$ on $\boldsymbol{\mathcal{A}} := \mathcal{A}_0 \times \cdots \times \mathcal{A}_N$ by setting

$$\boldsymbol{W}(\boldsymbol{s},\hat{\boldsymbol{s}}) = \left(\sum_{i=0}^N W_i(s_i,\hat{s}_i)^2 \right)^{1/2} \quad \forall \boldsymbol{s},\hat{\boldsymbol{s}} \in \boldsymbol{\mathcal{A}}.$$

Finally for technical reasons we also assume that there exists a smooth extension $\widetilde{\mathbb{K}}$ to $\mathbb{R}^d$ of the permeability tensor such that (3) holds on $\mathbb{R}^d$. This allows us to define distances $\tilde{d}_i$ on the whole $\mathbb{R}^d$ by

$$\tilde{d}_i(\boldsymbol{x},\boldsymbol{y}) = \inf_{\boldsymbol{\gamma}\in\widetilde{P}(\boldsymbol{x},\boldsymbol{y})}\left(\int_0^1 \mu_i \widetilde{\mathbb{K}}^{-1}(\boldsymbol{\gamma}(\tau))\boldsymbol{\gamma}'(\tau)\cdot\boldsymbol{\gamma}'(\tau)\,\mathrm{d}\tau\right)^{1/2} \quad \forall \boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d, \tag{21}$$

where $\widetilde{P}(\boldsymbol{x},\boldsymbol{y}) = \{\boldsymbol{\gamma}\in C^1([0,1];\mathbb{R}^d) \mid \boldsymbol{\gamma}(0)=\boldsymbol{x} \text{ and } \boldsymbol{\gamma}(1)=\boldsymbol{y}\}$. In the sequel, we assume that the extension $\widetilde{\mathbb{K}}$ of $\mathbb{K}$ is such that

$$\Omega \text{ is geodesically convex in } \mathcal{M}_i = (\mathbb{R}^d, \tilde{d}_i) \text{ for all } i. \tag{22}$$

In particular $\tilde{d}_i = d_i$ on $\Omega\times\Omega$. Since $\widetilde{\mathbb{K}}^{-1}$ is smooth, at least $C_b^2(\mathbb{R}^d)$, the Ricci curvature of the smooth complete Riemannian manifold $\mathcal{M}_i$ is uniformly bounded; i.e., there exists C depending only on $(\mu_i)_{0\le i\le N}$ and $\widetilde{\mathbb{K}}$ such that

$$|\mathrm{Ric}_{\mathcal{M}_i,\boldsymbol{x}}(\boldsymbol{v})| \le C\mu_i\mathbb{K}^{-1}\boldsymbol{v}\cdot\boldsymbol{v} \quad \forall\boldsymbol{x}\in\mathbb{R}^d,\ \forall\boldsymbol{v}\in\mathbb{R}^d. \tag{23}$$

We deduce from the lower bound on the Ricci curvature and on the geodesic convexity of Ω that the Boltzmann relative entropy $\mathcal{H}_\omega$ with respect to ω_i, defined by

$$\mathcal{H}_\omega(s) = \int_{\mathbb{R}^d} s\log\left(\frac{s}{\omega}\right)\mathrm{d}\boldsymbol{x} \quad \text{for all measurable } s:\Omega\to\mathbb{R}_+, \tag{24}$$

is λ_i-displacement convex on $\mathcal{P}^{\mathrm{ac}}(\Omega)$ for some $\lambda_i\in\mathbb{R}$. Here, $\mathcal{P}^{\mathrm{ac}}(\Omega)$ denotes the set of probability measures on Ω that are absolutely continuous with respect to the Lebesgue measure. Then mass scaling implies that $\mathcal{H}_\omega$ is also λ_i-displacement convex on $(\mathcal{A}_i, W_i)$. We refer to [Villani 2009, Chapters 14 and 17] for further details on the Ricci curvature and its links with optimal transportation.

In the homogeneous and isotropic case $\mathbb{K}(\boldsymbol{x}) = \mathrm{Id}$, condition (22) simply amounts to assuming that Ω is convex. A simple sufficient condition implying (22) is given in Appendix A in the isotropic but heterogeneous case $\mathbb{K}(\boldsymbol{x}) = \kappa(\boldsymbol{x})\mathbb{I}_d$.

Gradient flow of the energy. The content of this section is formal. Our aim is to write the problem as a gradient flow, i.e.,

$$\frac{\mathrm{d}\boldsymbol{s}}{\mathrm{d}t} \in -\,\mathbf{grad}_{\boldsymbol{W}}\,\mathcal{E}(\boldsymbol{s}) = -\big(\mathrm{grad}_{W_0}\,\mathcal{E}(\boldsymbol{s}),\dots,\mathrm{grad}_{W_N}\,\mathcal{E}(\boldsymbol{s})\big), \tag{25}$$

where $\mathbf{grad}_{\boldsymbol{W}}\,\mathcal{E}(\boldsymbol{s})$ denotes the full Wasserstein gradient of $\mathcal{E}(\boldsymbol{s})$, and $\mathrm{grad}_{W_i}\,\mathcal{E}(\boldsymbol{s})$ stands for the partial gradient of $s_i\mapsto\mathcal{E}(\boldsymbol{s})$ with respect to the Wasserstein distance W_i. The Wasserstein distance W_i was built so that $\dot{\boldsymbol{s}} = (\dot{s}_i)_i \in \mathbf{grad}_{\boldsymbol{W}}\,\mathcal{E}(\boldsymbol{s})$ if and only if there exists $\boldsymbol{h}\in\partial_{\boldsymbol{s}}\mathcal{E}(\boldsymbol{s})$ such that

$$\partial_t s_i = -\boldsymbol{\nabla}\cdot\left(s_i\frac{\mathbb{K}}{\mu_i}\boldsymbol{\nabla}h_i\right) \quad \forall i\in\{0,\dots,N\}.$$

Such a construction was already performed by Lisini in the case of a single equation. Owing to the definitions (13) and (15) of the energy $\mathcal{E}(\boldsymbol{s})$ and its subdifferential $\partial_{\boldsymbol{s}}\mathcal{E}(\boldsymbol{s})$, the partial differential equations can be (at least formally) recovered. This was, roughly speaking, the purpose of our note [Cancès et al. 2015].

In order to define rigorously the gradient $\mathbf{grad}_W\,\mathcal{E}$ in (25), $\boldsymbol{\mathcal{A}}$ has to be a Riemannian manifold. The so-called Otto's calculus [2001], see also [Villani 2009, Chapter 15], allows to put a formal Riemannian structure on $\boldsymbol{\mathcal{A}}$. But as far as we know, this structure cannot be made rigorous and $\boldsymbol{\mathcal{A}}$ is a mere metric space. This leads us to consider generalized gradient flows in metric spaces; see [Ambrosio et al. 2008]. We won't go deep into details in this direction, but we will prove that weak solutions can be obtained as limits of a minimizing movement scheme presented in the next section. This characterizes the gradient flow structure of the problem.

Minimizing movement scheme and main result.

The scheme and existence of a solution. For a fixed time-step $\tau > 0$, the so-called minimizing movement scheme [De Giorgi 1993; Ambrosio et al. 2008] or JKO scheme [Jordan et al. 1998] consists in computing recursively $(\boldsymbol{s}^n)_{n\geq 1}$ as the solution to the minimization problem

$$\boldsymbol{s}^n = \underset{\boldsymbol{s}\in\boldsymbol{\mathcal{A}}}{\operatorname{Argmin}}\left(\frac{\boldsymbol{W}(\boldsymbol{s},\boldsymbol{s}^{n-1})^2}{2\tau} + \mathcal{E}(\boldsymbol{s})\right), \tag{26}$$

the initial data $\boldsymbol{s}^0$ being given in (10).

Approximate solution and main result. Anticipating that the JKO scheme (26) is well-posed (this is the purpose of Proposition 2.1 below), we can now define the piecewise constant interpolation $\boldsymbol{s}^\tau \in L^\infty((0,T);\boldsymbol{\mathcal{X}}\cap\boldsymbol{\mathcal{A}})$ by

$$\boldsymbol{s}^\tau(0,\cdot) = \boldsymbol{s}^0 \qquad \text{and} \qquad \boldsymbol{s}^\tau(t,\cdot) = \boldsymbol{s}^n \quad \forall t\in((n-1)\tau, n\tau],\ \forall n\geq 1. \tag{27}$$

The main result of our paper is the following.

Theorem 1.2. *Let $(\tau_k)_{k\geq 1}$ be a sequence of time steps tending to 0. Then there exists one weak solution $\boldsymbol{s}$ in the sense of Definition 1.1 such that, up to an unlabeled subsequence, $(\boldsymbol{s}^{\tau_k})_{k\geq 1}$ converges a.e. in Q towards $\boldsymbol{s}$ as k tends to ∞.*

As a direct by-product of Theorem 1.2, the continuous problem admits (at least) one solution in the sense of Definition 1.1. As far as we know, this existence result is new.

Remark 1.3. It is worth stressing that our final solution will satisfy a posteriori $\partial_t s_i \in L^2((0,T);H^1(\Omega)')$, $s_i\in L^2((0,T);H^1(\Omega))$, and thus $s_i \in \mathcal{C}([0,T];L^2(\Omega))$. This regularity is enough to retrieve the so-called *energy-dissipation equality*

$$\frac{d}{dt}\mathcal{E}(\boldsymbol{s}(t)) = -\sum_{i=0}^{N}\int_\Omega \mathbb{K}\frac{s_i(t)}{\mu_i}\nabla(p_i(t)+\Psi_i)\cdot\nabla(p_i(t)+\Psi_i)\,\mathrm{d}\boldsymbol{x} \leq 0 \quad \text{for a.e. } t\in(0,T),$$

which is another admissible formulation of gradient flows in metric spaces [Ambrosio et al. 2008].

Goal and positioning of the paper. The aims of the paper are twofold. First, we aim to provide a rigorous foundation to the formal variational approach introduced in the authors' recent note [Cancès et al. 2015]. This gives new insights into the modeling of complex porous media flows and their numerical approximation. Our approach appears to be very natural since only physically motivated quantities appear in the study. Indeed, we manage to avoid the introduction of the so-called Kirchhoff transform and global pressure, which classically appear in the mathematical study of multiphase flows in porous media; see, for instance, [Chavent 1976; 2009; Antoncev and Monahov 1978; Chavent and Jaffré 1986; Fabrie and Saad 1993; Gagneux and Madaune-Tort 1996; Chen 2001; Amaziane et al. 2012; 2014].

Second, the existence result that we deduce from the convergence of the variational scheme is new as soon as there are at least three phases ($N \geq 2$). Indeed, since our study does not require the introduction of any global pressure, we get rid of many structural assumptions on the data, among which is the so-called *total differentiability condition*; see, for instance, Assumption (H3) in [Fabrie and Saad 1993]. This structural condition is not naturally satisfied by the models, and suitable algorithms have to be employed in order to adapt the data to this constraint [Chavent and Salzano 1985]. However, our approach suffers from another technical difficulty: we are limited to the case of linear relative permeabilities. The extension to the case of nonlinear concave relative permeabilities, i.e., where (1) is replaced by

$$\partial_t s_i + \nabla \cdot (k_i(s_i)\boldsymbol{v}_i) = 0,$$

may be reachable thanks to the contributions of Dolbeault, Nazaret, and Savaré [Dolbeault et al. 2009], see also [Zinsl and Matthes 2015b], but we did not push in this direction since the relative permeabilities k_i are in general supposed to be convex in models coming from engineering.

Since the seminal paper of Jordan, Kinderlehrer, and Otto [Jordan et al. 1998], gradient flows in metric spaces (and particularly in the space of probability measures endowed with the quadratic Wasserstein distance) were the object of many studies. Let us for instance refer to the monograph of Ambrosio, Gigli, and Savaré [Ambrosio et al. 2008] and to Villani's book [2009, Part II] for a complete overview. Applications are numerous. We refer for instance to [Otto 1998] for an application to magnetic fluids, to [Sandier and Serfaty 2004; Ambrosio and Serfaty 2008; Ambrosio et al. 2011] for applications to superconductivity to [Blanchet et al. 2008; Blanchet 2013; Zinsl and Matthes 2015a] for applications to chemotaxis, to [Lisini et al. 2012] for phase field models, to [Maury et al. 2010] for a macroscopic model of crowd motion, to [Bolley et al. 2013] for an application to granular media, to [Carrillo et al. 2011] for aggregation equations, and to [Kinderlehrer et al. 2017] for a model of ionic transport that applies in semiconductors. In the context of porous media flows, this framework has been used by Otto [2001] to study the asymptotic behavior of the porous medium equation, which is a simplified model for the filtration of a gas in a porous medium. The gradient flow approach in Wasserstein metric spaces was used more recently by Laurençot and Matioc [2013] on a thin film approximation model for two-phase flows in porous media. Finally, let us mention that similar ideas were successfully applied for multicomponent systems; see, e.g., [Carlier and Laborde 2015; Laborde 2016; Zinsl and Matthes 2015b; Zinsl 2014].

The variational structure of the system governing incompressible immiscible two-phase flows in porous media was recently depicted by the authors in their note [Cancès et al. 2015]. Whereas the purpose of that paper is formal, our goal is here to give a rigorous foundation to the variational approach for complex flows in porous media. Finally, let us mention the work of Gigli and Otto [2013], where it was noticed that multiphase linear transportation with saturation constraint, as we have here thanks to (1) and (4), yields nonlinear transport with mobilities that appear naturally in the two-phase flow context.

The paper is organized as follows. In Section 2, we derive estimates on the solution $\boldsymbol{s}^\tau$ for a fixed τ. Beyond the classical energy and distance estimates detailed in the first subsection, in the second subsection we obtain enhanced regularity estimates thanks to an adaptation of the so-called *flow interchange* technique of Matthes, McCann, and Savaré [Matthes et al. 2009] to our inhomogeneous context. Because of the constraint on the pore volume (4), the auxiliary flow we use is no longer the heat flow, and a drift term has to be added. An important effort is then done in Section 3 to derive the Euler–Lagrange equations that follow from the optimality of $\boldsymbol{s}^n$. Our proof is inspired by the work of Maury, Roudneff-Chupin, and Santambrogio [Maury et al. 2010]. It relies on an intensive use of the dual characterization of the optimal transportation problem and the corresponding Kantorovich potentials. However, additional difficulties arise from the multiphase aspect of our problem, in particular when there are at least three phases (i.e., $N \geq 2$). These are bypassed using a generalized multicomponent bathtub principle (Theorem B.1 in Appendix B) and computing the associated Lagrange multipliers in the first subsection. This key step then allows to define the notion of discrete phase and capillary pressures in the second subsection. Then Section 4 is devoted to the convergence of the approximate solutions $(\boldsymbol{s}^{\tau_k})_k$ towards a weak solution $\boldsymbol{s}$ as τ_k tends to 0. The estimates we obtained in Section 2 are integrated with respect to time in the first subsection. In the second subsection, we show that these estimates are sufficient to enforce the relative compactness of $(\boldsymbol{s}^{\tau_k})_k$ in the strong $L^1(Q)^{N+1}$ topology. Finally, it is shown in the third subsection that any limit $\boldsymbol{s}$ of $(\boldsymbol{s}^{\tau_k})_k$ is a weak solution in the sense of Definition 1.1.

2. One-step regularity estimates

The first thing to do is to show that the JKO scheme (26) is well-posed. This is the purpose of the following proposition.

Proposition 2.1. *Let $n \geq 1$ and $\boldsymbol{s}^{n-1} \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$. Then there exists a unique solution $\boldsymbol{s}^n$ to the scheme (26). Moreover, one has $\boldsymbol{s}^n \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$.*

Proof. Any $\boldsymbol{s}^{n-1} \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$ has finite energy thanks to (14). Let $(\boldsymbol{s}^{n,k})_k \subset \boldsymbol{\mathcal{A}}$ be a minimizing sequence in (26). Plugging $\boldsymbol{s}^{n-1}$ into (26), it is easy to see that $\mathcal{E}(\boldsymbol{s}^{n,k}) \leq \mathcal{E}(\boldsymbol{s}^{n-1}) < \infty$ for large k; thus $(\boldsymbol{s}^{n,k})_k \subset \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$ thanks to (14). Hence, one has $0 \leq s_i^{n,k}(\boldsymbol{x}) \leq \omega(\boldsymbol{x})$ for all k. By the Dunford–Pettis theorem, we can therefore assume that $s_i^{n,k} \rightharpoonup s_i^n$ weakly in $L^1(\Omega)$. It is then easy to check that the limit $\boldsymbol{s}^n$ of $\boldsymbol{s}^{n,k}$ belongs to $\boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$. The lower semicontinuity of the Wasserstein distance with respect to weak L^1 convergence is well known, see, e.g., [Santambrogio 2015, Proposition 7.4], and since the energy functional is convex and thus lower semicontinuous, we conclude that $\boldsymbol{s}^n$ is indeed a minimizer. Uniqueness follows from the strict convexity of the energy as well as from the convexity of the Wasserstein distances (with respect to linear interpolation $\boldsymbol{s}_\theta = (1-\theta)\boldsymbol{s}_0 + \theta \boldsymbol{s}_1$). □

The rest of this section is devoted to improving the regularity of the successive minimizers.

Energy and distance estimates. Plugging $\boldsymbol{s} = \boldsymbol{s}^{n-1}$ into (26) we obtain

$$\frac{\boldsymbol{W}(\boldsymbol{s}^n, \boldsymbol{s}^{n-1})^2}{2\tau} + \mathcal{E}(\boldsymbol{s}^n) \leq \mathcal{E}(\boldsymbol{s}^{n-1}). \tag{28}$$

As a consequence we have the monotonicity

$$\cdots \le \mathcal{E}(\boldsymbol{s}^n) \le \mathcal{E}(\boldsymbol{s}^{n-1}) \le \cdots \le \mathcal{E}(\boldsymbol{s}^0) < \infty$$

at the discrete level; thus $\boldsymbol{s}^n \in \boldsymbol{\mathcal{X}}$ for all $n \ge 0$ thanks to (14). Summing (28) over n we also obtain the classical *total square distance* estimate

$$\frac{1}{\tau}\sum_{n\ge 0} \boldsymbol{W}^2(\boldsymbol{s}^{n+1}, \boldsymbol{s}^n) \le 2\mathcal{E}(\boldsymbol{s}^0) \le C(\Omega, \Pi, \boldsymbol{\Psi}), \tag{29}$$

where the last inequality comes from the fact that $\boldsymbol{s}^0$ is uniformly bounded since it belongs to $\boldsymbol{\mathcal{X}}$, and thus so is $\mathcal{E}(\boldsymbol{s}^0)$. This readily gives the approximate $\frac{1}{2}$-Hölder estimate

$$\boldsymbol{W}(\boldsymbol{s}^{n_1}, \boldsymbol{s}^{n_2}) \le C\sqrt{|n_2 - n_1|\tau}. \tag{30}$$

Flow interchange, entropy estimate and enhanced regularity. The goal of this section is to obtain some additional Sobolev regularity on the capillary pressure field $\boldsymbol{\pi}(\boldsymbol{s}^{n*}, \boldsymbol{x})$, where $\boldsymbol{s}^{n*} = (s_1^n, \dots, s_N^n)$ is the unique element of $\boldsymbol{\mathcal{X}}^*$ corresponding to the minimizer $\boldsymbol{s}^n$ of (26). In what follows, we set

$$\pi_i^n : \Omega \to \mathbb{R}, \quad \boldsymbol{x} \mapsto \pi_i(\boldsymbol{s}^{n*}(\boldsymbol{x}), \boldsymbol{x}), \quad \forall i \in \{1, \dots, N\}$$

and $\boldsymbol{\pi}^n = (\pi_1^n, \dots, \pi_N^n)$. Bearing in mind that $\omega(\boldsymbol{x}) \ge \omega_\star > 0$ in $\overline{\Omega}$, we can define the relative Boltzmann entropy $\mathcal{H}_\omega$ with respect to ω by (24).

Lemma 2.2. *There exists C depending only on Ω, Π, ω, $\mathbb{K}$, $(\mu_i)_i$, and $\boldsymbol{\Psi}$ such that, for all $n \ge 1$ and all $\tau > 0$, one has*

$$\sum_{i=0}^{N} \|\nabla \pi_i^n\|^2_{L^2(\Omega)} \le C\left(1 + \frac{\boldsymbol{W}^2(\boldsymbol{s}^n, \boldsymbol{s}^{n-1})}{\tau} + \sum_{i=0}^{N} \frac{\mathcal{H}_\omega(s_i^{n-1}) - \mathcal{H}_\omega(s_i^n)}{\tau}\right). \tag{31}$$

Proof. The argument relies on the *flow interchange* technique introduced by Matthes, McCann, and Savaré [Matthes et al. 2009]. Throughout the proof, C denotes a fluctuating constant that depends on the prescribed data Ω, Π, ω, $\mathbb{K}$, $(\mu_i)_i$, and $\boldsymbol{\Psi}$, but neither on t, τ, nor on n. For $i = 0, \dots, N$ consider the auxiliary flows

$$\begin{cases} \partial_t \check{s}_i = \operatorname{div}(\mathbb{K}\nabla \check{s}_i - \check{s}_i \mathbb{K}\nabla \log \omega), & t > 0,\ \boldsymbol{x} \in \Omega, \\ \mathbb{K}(\nabla \check{s}_i - \check{s}_i \nabla \log \omega) \cdot \nu = 0, & t > 0,\ \boldsymbol{x} \in \partial\Omega, \\ \check{s}_i|_{t=0} = s_i^n, & \boldsymbol{x} \in \Omega \end{cases} \tag{32}$$

for each $i \in \{0, \dots, N\}$. By standard parabolic theory, see for instance [Ladyženskaja et al. 1968, Chapter III, Theorem 12.2], these initial-boundary value problems are well-posed, and their solutions $\check{s}_i(\boldsymbol{x})$ belong to $\mathcal{C}^{1,2}((0, 1] \times \overline{\Omega}) \cap \mathcal{C}([0, 1]; L^p(\Omega))$ for all $p \in (1, \infty)$ if $\omega \in \mathcal{C}^{2,\alpha}(\overline{\Omega})$ and $\mathbb{K} \in \mathcal{C}^{1,\alpha}(\overline{\Omega})$ for some $\alpha > 0$. Therefore, $t \mapsto \check{s}_i(\cdot, t)$ is absolutely continuous in $L^1(\Omega)$, and thus in $\mathcal{A}_i$ endowed with the usual quadratic distance W_{ref} (20) thanks to [Santambrogio 2015, Proposition 7.4]. Because of (19), the curve $t \mapsto \check{s}_i(\cdot, t)$ is also absolutely continuous in $\mathcal{A}_i$ endowed with W_i.

From Lisini's results [2009], we know that the evolution $t \mapsto \check{s}_i(\cdot, t)$ can be interpreted as the gradient flow of the relative Boltzmann functional $(1/\mu_i)\mathcal{H}_\omega$ with respect to the metric W_i, the scaling factor $1/\mu_i$ appearing due to the definition (18) of the distance W_i. As a consequence of (23), The Ricci curvature of (Ω, d_i) is bounded, and hence bounded from below. Since $\omega \in \mathcal{C}^2(\overline{\Omega})$, and with our assumption (22), we also have that $(1/\mu_i)\mathcal{H}_\omega$ is λ_i-displacement convex with respect to W_i for some $\lambda_i \in \mathbb{R}$ depending on ω and the geometry of (Ω, d_i); see [Villani 2009, Chapter 14]. Therefore, we can use the so-called *evolution variational inequality* characterization of gradient flows, see for instance [Ambrosio and Gigli 2013, Definition 4.5], centered at s_i^{n-1}, namely

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} W_i^2(\check{s}_i(t), s_i^{n-1}) + \frac{\lambda_i}{2} W_i^2(\check{s}_i(t), s_i^{n-1}) \le \frac{1}{\mu_i}\mathcal{H}_\omega(s_i^{n-1}) - \frac{1}{\mu_i}\mathcal{H}_\omega(\check{s}_i(t)).$$

Define $\check{\boldsymbol{s}} = (\check{s}_0, \dots, \check{s}_N)$ and $\check{\boldsymbol{s}}^* = (\check{s}_1, \dots, \check{s}_N)$. Summing the previous inequality over $i \in \{0, \dots, N\}$ leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2\tau}\boldsymbol{W}^2(\check{\boldsymbol{s}}(t), \boldsymbol{s}^{n-1})\right) \le C\left(\frac{\boldsymbol{W}^2(\check{\boldsymbol{s}}(t), \boldsymbol{s}^{n-1})}{\tau} + \sum_{i=0}^N \frac{\mathcal{H}_\omega(s_i^{n-1}) - \mathcal{H}_\omega(\check{s}_i(t))}{\tau}\right). \tag{33}$$

In order to estimate the internal energy contribution in (26), we first note that $\sum s_i^n(\boldsymbol{x}) = \omega(\boldsymbol{x})$ for all $\boldsymbol{x} \in \overline{\Omega}$; thus by the linearity of (32) and since ω is a stationary solution we have $\sum \check{s}_i(\boldsymbol{x}, t) = \omega(\boldsymbol{x})$ as well. Moreover, the problem (32) is monotone, thus order preserving, and admits 0 as a subsolution. Hence $\check{s}_i(\boldsymbol{x}, t) \ge 0$, so that $\check{\boldsymbol{s}}(t) \in \boldsymbol{\mathcal{A}} \cap \boldsymbol{\mathcal{X}}$ is an admissible competitor in (26) for all $t > 0$. The smoothness of $\check{\boldsymbol{s}}$ for $t > 0$ allows us to write

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_\Omega \Pi(\check{\boldsymbol{s}}^*(\boldsymbol{x}, t), \boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right) = \sum_{i=1}^N \int_\Omega \check{\pi}_i(\boldsymbol{x}, t)\,\partial_t \check{s}_i(\boldsymbol{x}, t)\,\mathrm{d}\boldsymbol{x} = I_1(t) + I_2(t), \tag{34}$$

where $\check{\pi}_i := \pi_i(\check{\boldsymbol{s}}^*, \cdot)$, and where, for all $t > 0$, we have set

$$I_1(t) = -\sum_{i=1}^N \int_\Omega \nabla\check{\pi}_i(t) \cdot \mathbb{K}\nabla\check{s}_i(t)\,\mathrm{d}\boldsymbol{x}, \qquad I_2(t) = -\sum_{i=1}^N \int_\Omega \frac{\check{s}_i(t)}{\omega}\nabla\check{\pi}_i(t) \cdot \mathbb{K}\nabla\omega\,\mathrm{d}\boldsymbol{x}.$$

To estimate I_1, we first use the invertibility of $\boldsymbol{\pi}$ to write

$$\check{\boldsymbol{s}}(\boldsymbol{x}, t) = \boldsymbol{\phi}(\check{\boldsymbol{\pi}}(\boldsymbol{x}, t), \boldsymbol{x}) =: \check{\boldsymbol{\phi}}(\boldsymbol{x}, t),$$

yielding

$$\nabla\check{\boldsymbol{s}}(\boldsymbol{x}, t) = \mathbb{J}_z\boldsymbol{\phi}(\check{\boldsymbol{\pi}}(\boldsymbol{x}, t), \boldsymbol{x})\,\nabla\check{\boldsymbol{\pi}}(\boldsymbol{x}, t) + \nabla_{\boldsymbol{x}}\boldsymbol{\phi}(\check{\boldsymbol{\pi}}(\boldsymbol{x}, t), \boldsymbol{x}). \tag{35}$$

Combining (3), (7), (8) and the elementary inequality

$$ab \le \delta\frac{a^2}{2} + \frac{b^2}{2\delta} \quad \text{with } \delta > 0 \text{ arbitrary}, \tag{36}$$

we get that for all $t > 0$,

$$I_1(t) \le -\frac{\kappa_\star}{\varpi^\star}\int_\Omega |\nabla\check{\boldsymbol{\pi}}(t)|^2\,\mathrm{d}\boldsymbol{x} + \kappa^\star\left(\delta\int_\Omega |\nabla\check{\boldsymbol{\pi}}(t)|^2\,\mathrm{d}\boldsymbol{x} + \frac{1}{\delta}\int_\Omega |\nabla_{\boldsymbol{x}}\boldsymbol{\phi}(\check{\boldsymbol{\pi}}(t))|^2\,\mathrm{d}\boldsymbol{x}\right).$$

Choosing $\delta = \kappa_\star/(4\kappa^\star\varpi^\star)$, we get that

$$I_1(t) \le -\frac{3\kappa_\star}{4\varpi^\star}\int_\Omega |\nabla\check{\boldsymbol{\pi}}(t)|^2\,\mathrm{d}\boldsymbol{x} + C \quad \forall t > 0. \tag{37}$$

In order to estimate I_2, we use that $\check{\boldsymbol{s}}(t) \in \boldsymbol{\mathcal{X}}$ for all $t > 0$, so that $0 \le \check{s}_i(\boldsymbol{x}, t) \le \omega(\boldsymbol{x})$; hence we deduce that $\sum_{i=1}^N (\check{s}_i/\omega)^2 \le 1$. Therefore, using (36) again, we get

$$I_2(t) \le \delta\kappa^\star \int_\Omega |\nabla\check{\boldsymbol{\pi}}(t)|^2\,\mathrm{d}\boldsymbol{x} + \frac{\kappa^\star}{\delta}\int_\Omega |\nabla\omega|^2\,\mathrm{d}\boldsymbol{x}.$$

Choosing again $\delta = \kappa_\star/(4\kappa^\star\varpi^\star)$ yields

$$I_2(t) \le \frac{\kappa_\star}{4\varpi^\star}\int_\Omega |\nabla\check{\boldsymbol{\pi}}(t)|^2\,\mathrm{d}\boldsymbol{x} + C. \tag{38}$$

Taking (37)–(38) into account in (34) provides

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_\Omega \Pi(\check{\boldsymbol{s}}^*(\boldsymbol{x}, t), \boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right) \le -\frac{\kappa_\star}{2\varpi^\star}\int_\Omega |\nabla\check{\boldsymbol{\pi}}(t)|^2\,\mathrm{d}\boldsymbol{x} + C \quad \forall t > 0. \tag{39}$$

Let us now focus on the potential (gravitational) energy. Since $\check{\boldsymbol{s}}(t)$ belongs to $\boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$ for all $t > 0$, we can make use of the relation

$$\check{s}_0(\boldsymbol{x}, t) = \omega(\boldsymbol{x}) - \sum_{i=1}^N \check{s}_i(\boldsymbol{x}, t) \quad \forall (\boldsymbol{x}, t) \in \Omega \times \mathbb{R}_+,$$

to write: for all $t > 0$,

$$\sum_{i=0}^N \int_\Omega \check{s}_i(\boldsymbol{x}, t)\Psi_i(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \sum_{i=1}^N \int_\Omega \check{s}_i(\boldsymbol{x}, t)(\Psi_i - \Psi_0)(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \int_\Omega \omega(\boldsymbol{x})\Psi_0(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.$$

This leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{i=0}^N \int_\Omega \check{s}_i(t)\Psi_i\,\mathrm{d}\boldsymbol{x}\right) = \sum_{i=1}^N \int_\Omega (\Psi_i(\boldsymbol{x}) - \Psi_0(\boldsymbol{x}))\,\partial_t s_i(\boldsymbol{x}, t)\,\mathrm{d}\boldsymbol{x} = J_1(t) + J_2(t), \tag{40}$$

where, using the equations (32), we have set

$$J_1(t) = -\sum_{i=1}^N \int_\Omega \nabla(\Psi_i - \Psi_0)\cdot\mathbb{K}\nabla\check{s}_i(t)\,\mathrm{d}\boldsymbol{x}, \quad J_2(t) = \sum_{i=1}^N \int_\Omega \frac{\check{s}_i(t)}{\omega}\nabla(\Psi_i - \Psi_0)\cdot\mathbb{K}\nabla\omega\,\mathrm{d}\boldsymbol{x}.$$

The term J_1 can be estimated using (36). More precisely, for all $\delta > 0$, we have

$$J_1(t) \le \kappa^\star\left(\delta\|\nabla\check{\boldsymbol{s}}^*(t)\|_{L^2}^2 + \frac{1}{\delta}\sum_{i=1}^N \|\nabla(\Psi_i - \Psi_0)\|_{L^2}^2\right). \tag{41}$$

Using (35) together with (7)–(8), we get that

$$\|\nabla\check{\boldsymbol{s}}^*\|_{L^2}^2 \le \left(\frac{1}{\varpi_\star}\|\nabla\check{\boldsymbol{\pi}}\|_{L^2} + |\Omega|M_{\boldsymbol{\phi}}\right)^2 \le \frac{2}{(\varpi_\star)^2}\|\nabla\check{\boldsymbol{\pi}}\|_{L^2}^2 + 2(|\Omega|M_{\boldsymbol{\phi}})^2.$$

Therefore, choosing $\delta = (\varpi_\star)^2\kappa_\star/(8\kappa^\star\varpi^\star)$ in (41), we infer from the regularity of $\boldsymbol{\Psi}$ that

$$J_1(t) \leq \frac{\kappa_\star}{4\varpi^\star}\int_\Omega |\nabla\check{\boldsymbol{\pi}}(t)|^2\,\mathrm{d}\boldsymbol{x} + C \quad \forall t > 0. \tag{42}$$

Finally, it follows from the fact that $\sum_{i=1}^N \check{s}_i \leq \omega$, from the Cauchy–Schwarz inequality, and from the regularity of $\boldsymbol{\Psi}$, ω that

$$J_2(t) \geq -\kappa^\star \sum_{i=1}^N \|\nabla\Psi_i - \nabla\Psi_0\|_{L^2}\,\|\nabla\omega\|_{L^2} = C. \tag{43}$$

Combining (40), (42), and (43) with (39), we get that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(\check{\boldsymbol{s}}(t)) \leq -\frac{\kappa_\star}{4\varpi^\star}\int_\Omega |\nabla\check{\boldsymbol{\pi}}(t)|^2\,\mathrm{d}\boldsymbol{x} + C \quad \forall t > 0. \tag{44}$$

Denote by

$$\mathcal{F}_\tau^n(\boldsymbol{s}) := \frac{1}{2\tau}\boldsymbol{W}^2(\boldsymbol{s}, \boldsymbol{s}^{n-1}) + \mathcal{E}(\boldsymbol{s}) \tag{45}$$

the functional to be minimized in (26); then combining (33) and (44) provides

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}_\tau^n(\check{\boldsymbol{s}}(t)) + \frac{\kappa_\star}{4\varpi^\star}\|\nabla\check{\boldsymbol{\pi}}\|_{L^2}^2 \leq C\left(1 + \frac{\boldsymbol{W}^2(\check{\boldsymbol{s}}(t), \boldsymbol{s}^{n-1})}{\tau} + \sum_{i=0}^N \frac{\mathcal{H}_\omega(s_i^{n-1}) - \mathcal{H}_\omega(\check{s}_i(t))}{\tau}\right) \quad \forall t > 0.$$

Since $\check{\boldsymbol{s}}(0) = \boldsymbol{s}^n$ is a minimizer of (26), we must have

$$0 \leq \limsup_{t\to 0^+}\left(\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}_\tau^n(\check{\boldsymbol{s}}(t))\right),$$

otherwise $\check{\boldsymbol{s}}(t)$ would be a strictly better competitor than $\boldsymbol{s}^n$ for small $t > 0$. As a consequence, we get

$$\liminf_{t\to 0^+}\|\nabla\check{\boldsymbol{\pi}}(t)\|_{L^2}^2 \leq C\limsup_{t\to 0^+}\left(1 + \frac{\boldsymbol{W}^2(\check{\boldsymbol{s}}(t), \boldsymbol{s}^{n-1})}{\tau} + \sum_{i=0}^N \frac{\mathcal{H}_\omega(s_i^{n-1}) - \mathcal{H}_\omega(\check{s}_i(t))}{\tau}\right).$$

Since $\check{s}_i$ belongs to $C([0,1]; L^p(\Omega))$ for all $p \in [1,\infty)$, see for instance [Cancès and Gallouët 2011], the continuity of the Wasserstein distance and of the Boltzmann entropy with respect to strong L^p-convergence imply that

$$\boldsymbol{W}^2(\check{\boldsymbol{s}}(t), \boldsymbol{s}^{n-1}) \xrightarrow{t\to 0^+} \boldsymbol{W}^2(\boldsymbol{s}^n, \boldsymbol{s}^{n-1}) \quad \text{and} \quad \mathcal{H}_\omega(\check{s}_i(t)) \xrightarrow{t\to 0^+} \mathcal{H}_\omega(s_i^n).$$

Therefore, we obtain that

$$\liminf_{t\to 0^+}\|\nabla\check{\boldsymbol{\pi}}(t)\|_{L^2}^2 \leq C\left(1 + \frac{\boldsymbol{W}^2(\boldsymbol{s}^n, \boldsymbol{s}^{n-1})}{\tau} + \sum_{i=0}^N \frac{\mathcal{H}_\omega(s_i^{n-1}) - \mathcal{H}_\omega(s_i^n)}{\tau}\right). \tag{46}$$

It follows from the regularity of $\boldsymbol{\pi}$ that

$$\boldsymbol{\pi}(\check{\boldsymbol{s}}^*(t), \boldsymbol{x}) = \check{\boldsymbol{\pi}}(t) \xrightarrow{t\to 0^+} \boldsymbol{\pi}^n = \boldsymbol{\pi}(\boldsymbol{s}^{n*}, \boldsymbol{x}) \quad \text{in } L^p(\Omega).$$

Finally, let $(t_\ell)_{\ell\ge 1}$ be a decreasing sequence tending to 0 realizing the lim inf in (46); then the sequence $(\nabla\check{\boldsymbol{\pi}}(t_\ell))_{\ell\ge 1}$ converges weakly in $L^2(\Omega)^{N\times d}$ towards $\nabla\boldsymbol{\pi}^n$. The lower semicontinuity of the norm with respect to the weak convergence leads to

$$\begin{aligned}\sum_{i=1}^{N}\|\nabla\pi_i^n\|_{L^2}^2 &\le \lim_{\ell\to\infty}\|\nabla\check{\boldsymbol{\pi}}(t_\ell)\|_{L^2}^2\\ &= \liminf_{t\to 0^+}\|\nabla\check{\boldsymbol{\pi}}(t)\|_{L^2}^2 \le C\left(1+\frac{\boldsymbol{W}^2(\boldsymbol{s}^n,\boldsymbol{s}^{n-1})}{\tau}+\sum_{i=0}^{N}\frac{\mathcal{H}_\omega(s_i^{n-1})-\mathcal{H}_\omega(s_i^n)}{\tau}\right). \qquad \square\end{aligned}$$

3. The Euler–Lagrange equations and pressure bounds

The goal of this section is to extract information coming from the optimality of $\boldsymbol{s}^n$ in the JKO minimization (26). The main difficulty consists in constructing the phase and capillary pressures from this optimality condition. Our proof is inspired by [Maury et al. 2010] and makes extensive use of the Kantorovich potentials. Therefore, we first recall their definition and some useful properties. We refer to [Santambrogio 2015, §1.2; Villani 2009, Chapter 5] for details.

Let $(\nu_1,\nu_2)\in\mathcal{M}_+(\Omega)^2$ be two nonnegative measures with same total mass. A pair of Kantorovich potentials $(\varphi_i,\psi_i)\in L^1(\nu_1)\times L^1(\nu_2)$ associated to the measures ν_1 and ν_2 and to the cost function $\frac{1}{2}d_i^2$ defined by (16), $i\in\{0,\dots,N\}$, is a solution of the Kantorovich *dual problem*

$$\mathrm{DP}_i(\nu_1,\nu_2)=\max_{\substack{(\varphi_i,\psi_i)\in L^1(\nu_1)\times L^1(\nu_2)\\ \varphi_i(\boldsymbol{x})+\psi_i(\boldsymbol{y})\le\frac{1}{2}d_i^2(\boldsymbol{x},\boldsymbol{y})}}\int_\Omega\varphi_i(\boldsymbol{x})\nu_1(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}+\int_\Omega\psi_i(\boldsymbol{y})\nu_2(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}.$$

We will use the three following important properties of the Kantorovich potentials:

(a) There is always duality; that is,

$$\mathrm{DP}_i(\nu_1,\nu_2)=\tfrac{1}{2}W_i^2(\nu_1,\nu_2)\quad\forall i\in\{0,\dots,N\}.$$

(b) A pair of Kantorovich potentials (φ_i,ψ_i) is $\mathrm{d}\nu_1\otimes\mathrm{d}\nu_2$ unique, up to additive constants.

(c) The Kantorovich potentials φ_i and ψ_i are $\frac{1}{2}d_i^2$-conjugate; that is,

$$\begin{aligned}\varphi_i(\boldsymbol{x})&=\inf_{\boldsymbol{y}\in\Omega}\tfrac{1}{2}d_i^2(\boldsymbol{x},\boldsymbol{y})-\psi_i(\boldsymbol{y})\quad\forall\,\boldsymbol{x}\in\Omega,\\ \psi_i(\boldsymbol{y})&=\inf_{\boldsymbol{x}\in\Omega}\tfrac{1}{2}d_i^2(\boldsymbol{x},\boldsymbol{y})-\varphi_i(\boldsymbol{x})\quad\forall\,\boldsymbol{y}\in\Omega.\end{aligned}$$

Remark 3.1. Since Ω is bounded, the cost functions $(\boldsymbol{x},\boldsymbol{y})\mapsto\frac{1}{2}d_i^2(\boldsymbol{x},\boldsymbol{y})$, $i\in\{1,\dots,N\}$, are globally Lipschitz continuous; see (17). Thus item (c) shows that φ_i and ψ_i are also Lipschitz continuous.

A decomposition result. The next lemma is an adaptation of [Maury et al. 2010, Lemma 3.1] to our framework. It essentially states that, since $\boldsymbol{s}^n$ is a minimizer of (26), it is also a minimizer of the linearized problem.

Lemma 3.2. *For $n \geq 1$ and $i = 0, \dots, N$ there exist some (backward, optimal) Kantorovich potentials φ_i^n from s_i^n to s_i^{n-1} such that, using the convention $\pi_0^n = (\partial \Pi / \partial s_0)(s_1^n, \dots, s_N^n, \boldsymbol{x}) = 0$, setting*

$$F_i^n := \frac{\varphi_i^n}{\tau} + \pi_i^n + \Psi_i, \quad \forall i \in \{0, \dots, N\}, \tag{47}$$

and defining $\boldsymbol{F}^n = (F_i^n)_{0 \leq i \leq N}$, we have

$$\boldsymbol{s}^n \in \underset{\boldsymbol{s} \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}}{\operatorname{Argmin}} \int_\Omega \boldsymbol{F}^n(\boldsymbol{x}) \cdot \boldsymbol{s}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}. \tag{48}$$

Moreover, $F_i^n \in L^\infty \cap H^1(\Omega)$ for all $i \in \{0, \dots, N\}$.

Proof. We assume first that $s_i^{n-1}(\boldsymbol{x}) > 0$ everywhere in Ω for all $i \in \{1, \dots, N\}$, so that the Kantorovich potentials (φ_i^n, ψ_i^n) from s_i^n to s_i^{n-1} are uniquely determined after normalizing $\varphi_i^n(\boldsymbol{x}_{\mathrm{ref}}) = 0$ for some arbitrary point $\boldsymbol{x}_{\mathrm{ref}} \in \Omega$; see [Santambrogio 2015, Proposition 7.18]. Given any $\boldsymbol{s} = (s_i)_{1 \leq 0 \leq N} \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$ and $\varepsilon \in (0, 1)$ we define the perturbation

$$\boldsymbol{s}^\varepsilon := (1 - \varepsilon)\boldsymbol{s}^n + \varepsilon \boldsymbol{s}.$$

Note that $\boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$ is convex; thus $\boldsymbol{s}^\varepsilon$ is an admissible competitor for all $\varepsilon \in (0, 1)$. Let $(\varphi_i^\varepsilon, \psi_i^\varepsilon)$ be the unique Kantorovich potentials from s_i^ε to s_i^{n-1}, similarly normalized as $\varphi_i^\varepsilon(\boldsymbol{x}_{\mathrm{ref}}) = 0$. Then by characterization of the squared Wasserstein distance in terms of the dual Kantorovich problem we have

$$\begin{cases} \frac{1}{2} W_i^2(s_i^\varepsilon, s_i^{n-1}) = \displaystyle\int_\Omega \varphi_i^\varepsilon(\boldsymbol{x}) s_i^\varepsilon(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} + \int_\Omega \psi_i^\varepsilon(\boldsymbol{y}) s_i^{n-1}(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}, \\ \frac{1}{2} W_i^2(s_i^n, s_i^{n-1}) \geq \displaystyle\int_\Omega \varphi_i^\varepsilon(\boldsymbol{x}) s_i^n(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} + \int_\Omega \psi_i^\varepsilon(\boldsymbol{y}) s_i^{n-1}(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}. \end{cases}$$

By definition of the perturbation $\boldsymbol{s}^\varepsilon$ it is easy to check that $s_i^\varepsilon - s_i^n = \varepsilon(s_i - s_i^n)$. Subtracting the previous inequalities we get

$$\frac{W_i^2(s_i^\varepsilon, s_i^{n-1}) - W_i^2(s_i^n, s_i^{n-1})}{2\tau} \leq \frac{\varepsilon}{\tau} \int_\Omega \varphi_i^\varepsilon (s_i - s_i^n) \, \mathrm{d}\boldsymbol{x}. \tag{49}$$

Define $\boldsymbol{s}^{\varepsilon *} = (s_1^\varepsilon, \dots, s_N^\varepsilon)$, $\boldsymbol{\pi}^\varepsilon = \boldsymbol{\pi}(\boldsymbol{s}^{\varepsilon *}, \cdot)$, and extend to the zeroth component $\bar{\boldsymbol{\pi}}^\varepsilon = (0, \boldsymbol{\pi}^\varepsilon)$. The convexity of Π as a function of $s_1, \dots, s_N$ implies

$$\int_\Omega \big(\Pi(\boldsymbol{s}^{n*}, \boldsymbol{x}) - \Pi(\boldsymbol{s}^{\varepsilon *}, \boldsymbol{x})\big) \, \mathrm{d}\boldsymbol{x} \geq \int_\Omega \boldsymbol{\pi}^\varepsilon \cdot (\boldsymbol{s}^{n*} - \boldsymbol{s}^{\varepsilon *}) \, \mathrm{d}\boldsymbol{x} = \int_\Omega \bar{\boldsymbol{\pi}}^\varepsilon \cdot (\boldsymbol{s}^n - \boldsymbol{s}^\varepsilon) \, \mathrm{d}\boldsymbol{x} = -\varepsilon \int_\Omega \bar{\boldsymbol{\pi}}^\varepsilon \cdot (\boldsymbol{s} - \boldsymbol{s}^n) \, \mathrm{d}\boldsymbol{x}. \tag{50}$$

For the potential energy, we obtain by linearity that

$$\int_\Omega (\boldsymbol{s}^\varepsilon - \boldsymbol{s}^n) \cdot \boldsymbol{\Psi} \, \mathrm{d}\boldsymbol{x} = \varepsilon \int_\Omega (\boldsymbol{s} - \boldsymbol{s}^n) \cdot \boldsymbol{\Psi} \, \mathrm{d}\boldsymbol{x}. \tag{51}$$

Summing (49)–(51), dividing by ε, and recalling that $\boldsymbol{s}^n$ minimizes the functional $\mathcal{F}_\tau^n$ defined by (45), we obtain

$$0 \leq \frac{\mathcal{F}_\tau^n(\boldsymbol{s}^\varepsilon) - \mathcal{F}_\tau^n(\boldsymbol{s}^n)}{\varepsilon} \leq \sum_{i=0}^N \int_\Omega \left(\frac{\varphi_i^\varepsilon}{\tau} + \bar{\pi}_i^\varepsilon + \Psi_i \right) (s_i - s_i^n) \, \mathrm{d}\boldsymbol{x} \tag{52}$$

for all $\boldsymbol{s} \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$ and all $\varepsilon \in (0, 1)$. Because Ω is bounded, any Kantorovich potential is globally Lipschitz with bounds uniform in ε; see, for instance, the proof of [Santambrogio 2015, Theorem 1.17]. Since $\boldsymbol{s}^\varepsilon$ converges uniformly towards $\boldsymbol{s}^n$ when ε tends to 0, we infer from Theorem 1.52 of the same paper that φ_i^ε converges uniformly towards φ_i^n as ε tends to 0, where φ_i^n is a Kantorovich potential from s_i^n to s_i^{n-1}. Moreover, since $\boldsymbol{\pi}$ is uniformly continuous in $\boldsymbol{s}$, we also know that $\boldsymbol{\pi}^\varepsilon$ converges uniformly towards $\boldsymbol{\pi}^n$ and thus $\bar{\boldsymbol{\pi}}^\varepsilon \to \bar{\boldsymbol{\pi}}^n = (0, \boldsymbol{\pi}^n)$ as well. Then we can pass to the limit in (52) and infer that

$$0 \le \int_\Omega \boldsymbol{F}^n \cdot (\boldsymbol{s} - \boldsymbol{s}^n)\,\mathrm{d}\boldsymbol{x} \quad \forall \boldsymbol{s} \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}} \tag{53}$$

and (48) holds.

If $s_i^{n-1} > 0$ does not hold everywhere, we argue by approximation. Running the flow (32) for a short time $\delta > 0$ starting from $\boldsymbol{s}^{n-1}$, we construct an approximation $\boldsymbol{s}^{n-1,\delta} = (s_0^{n-1,\delta}, \dots, s_N^{n-1,\delta})$ converging to $\boldsymbol{s}^{n-1} = (s_0^{n-1}, \dots, s_N^{n-1})$ in $L^1(\Omega)$ as δ tends to 0. By construction $\boldsymbol{s}^{n-1,\delta} \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$, and it follows from the strong maximum principle that $s_i^{n-1,\delta} > 0$ in $\bar{\Omega}$ for all $\delta > 0$. By Proposition 2.1 there exists a unique minimizer $\boldsymbol{s}^{n,\delta}$ to the functional

$$\mathcal{F}_\tau^{n,\delta} : \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}} \to \mathbb{R}_+, \quad \boldsymbol{s} \mapsto \frac{1}{2\tau} \boldsymbol{W}^2(\boldsymbol{s}, \boldsymbol{s}^{n-1,\delta}) + \mathcal{E}(\boldsymbol{s}).$$

Since $\boldsymbol{s}^{n-1,\delta} > 0$, there exist unique Kantorovich potentials $(\varphi_i^{n,\delta}, \psi_i^{n,\delta})$ from $s_i^{n,\delta}$ to $s_i^{n-1,\delta}$. This allows us to construct $\boldsymbol{F}^{n,\delta}$ using (47), where φ_i^n and π_i^n have been replaced by $\varphi_i^{n,\delta}$ and $\pi_i^{n,\delta}$. Thanks to the above discussion,

$$0 \le \int_\Omega \boldsymbol{F}^{n,\delta *} \cdot (\boldsymbol{s}^* - \boldsymbol{s}^{n,\delta *})\,\mathrm{d}\boldsymbol{x} \quad \forall \boldsymbol{s}^* \in \boldsymbol{\mathcal{X}}^* \cap \boldsymbol{\mathcal{A}}^*. \tag{54}$$

We can now let δ tend to 0. Because of the time continuity of the solutions to (32), we know that $\boldsymbol{s}^{n-1,\delta}$ converges towards $\boldsymbol{s}^{n-1}$ in $L^1(\Omega)$. On the other hand, from the definition of $\boldsymbol{s}^{n,\delta}$ and Lemma 2.2 (in particular (31) with $s^{n-1,\delta}$, $s^{n,\delta}$, $\boldsymbol{\pi}^{n,\delta}$ instead of s^{n-1}, s^n, $\boldsymbol{\pi}^n$) we see that $\boldsymbol{\pi}^{n,\delta}$ is bounded in $H^1(\Omega)^{N+1}$ uniformly in $\delta > 0$. Using next the Lipschitz continuity (8) of $\boldsymbol{\phi}$, one deduces that $\boldsymbol{s}^{n,\delta}$ is uniformly bounded in $H^1(\Omega)^{N+1}$. Then, thanks to Rellich's compactness theorem, we can assume that $\boldsymbol{s}^{n,\delta}$ converges strongly in $L^2(\Omega)^{N+1}$ as δ tends to 0. By the strong convergence $\boldsymbol{s}^{n-1,\delta} \to \boldsymbol{s}^{n-1}$ and standard properties of the squared Wasserstein distance, one readily checks that $\mathcal{F}_\tau^{n,\delta}$ Γ-converges towards $\mathcal{F}_\tau^n$, and we can therefore identify the limit of $\boldsymbol{s}^{n,\delta}$ as the unique minimizer $\boldsymbol{s}^n$ of $\mathcal{F}_\tau^n$. Thanks to Lebesgue's dominated convergence theorem, we also infer that $\pi_i^{n,\delta}$ converges in $L^2(\Omega)$ towards π_i^n. Using once again the stability of the Kantorovich potentials [Santambrogio 2015, Theorem 1.52], we know that $\varphi_i^{n,\delta}$ converges uniformly towards some Kantorovich potential φ_i^n. Then we can pass to the limit in (54) and claim that (53) is satisfied even when some coordinates of $\boldsymbol{s}^{n-1}$ vanish on some parts of Ω.

Finally, note that since the Kantorovich potentials φ_i^n are Lipschitz continuous and because $\pi_i^n \in H^1$ (see Lemma 2.2) and $\boldsymbol{\Psi}$ is smooth, we have $F_i^n \in H^1$. Since the phases are bounded $0 \le s_i^n(\boldsymbol{x}) \le \omega(\boldsymbol{x})$ and $\boldsymbol{\pi}$ is continuous we have $\boldsymbol{\pi}^n \in L^\infty$; thus $F_i^n \in L^\infty$ as well and the proof is complete. □

We can now suitably decompose the vector field $\boldsymbol{F}^n = (F_i^n)_{0 \le i \le N}$ defined by (47).

Corollary 3.3. *Let $\boldsymbol{F}^n = (F_0^n, \dots, F_N^n)$ be as in Lemma 3.2. There exists $\boldsymbol{\alpha}^n \in \mathbb{R}^{N+1}$ such that, setting $\lambda^n(\boldsymbol{x}) := \min_j (F_j^n(\boldsymbol{x}) + \alpha_j^n)$, we have $\lambda^n \in H^1(\Omega)$ and*

$$F_i^n + \alpha_i^n = \lambda^n \qquad \mathrm{d}s_i^n\text{-a.e. in } \Omega,\ \forall i \in \{0, \dots, N\}, \tag{55}$$

$$\nabla F_i^n = \nabla \lambda^n \quad \mathrm{d}s_i^n\text{-a.e. in } \Omega,\ \forall i \in \{0, \dots, N\}. \tag{56}$$

Proof. By Lemma 3.2 we know that $\boldsymbol{s}^n$ minimizes $\boldsymbol{s} \mapsto \int \boldsymbol{F}^n \cdot \boldsymbol{s}$ among all admissible $\boldsymbol{s} \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$. Applying the multicomponent bathtub principle, Theorem B.1 in Appendix B, we infer that there exists $\boldsymbol{\alpha}^n = (\alpha_0^n, \dots, \alpha_N^n) \in \mathbb{R}^{N+1}$ such that $F_i^n + \alpha_i^n = \lambda^n$ for $\mathrm{d}s_i^n$-a.e. $\boldsymbol{x} \in \Omega$ and $\lambda^n = \min_j (F_j^n + \alpha_j^n)$ as in our statement. Note first that $\lambda^n \in H^1(\Omega)$ as the minimum of finitely many H^1 functions $F_0, \dots, F_N \in H^1(\Omega)$. From the usual Serrin's chain rule we have moreover that

$$\nabla \lambda^n = \nabla \min_j (F_j^n + \alpha_j^n) = \nabla F_i . \chi_{[F_i^n + \alpha_i^n = \lambda^n]},$$

and since $s_i^n = 0$ inside $[F_i^n + \alpha_i^n \neq \lambda^n]$, the proof is complete. $\square$

The discrete capillary pressure law and pressure estimates. In this section, some calculations in the Riemannian settings (Ω, d_i) will be carried out. In order to make them as readable as possible, we have to introduce a few basics. We refer to [Villani 2009, Chapter 14] for a more detailed presentation.

Let $i \in \{0, \dots, N\}$; then consider the Riemannian geometry (Ω, d_i), and let $\boldsymbol{x} \in \Omega$. We denote by $g_{i,\boldsymbol{x}} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ the local metric tensor defined by

$$g_{i,\boldsymbol{x}}(\boldsymbol{v}, \boldsymbol{v}) = \mu_i \mathbb{K}^{-1}(\boldsymbol{x}) \boldsymbol{v} \cdot \boldsymbol{v} = \mathbb{G}_i(\boldsymbol{x}) \boldsymbol{v} \cdot \boldsymbol{v} \quad \forall \boldsymbol{v} \in \mathbb{R}^d.$$

In this framework, the gradient $\nabla_{g_i} \varphi$ of a function $\varphi \in \mathcal{C}^1(\Omega)$ is defined by

$$\varphi(\boldsymbol{x} + h\boldsymbol{v}) = \varphi(\boldsymbol{x}) + h g_{i,\boldsymbol{x}}(\nabla_{g_{i,x}} \varphi(\boldsymbol{x}), \boldsymbol{v}) + o(h) \quad \forall \boldsymbol{v} \in \mathbb{S}^{d-1},\ \forall \boldsymbol{x} \in \Omega.$$

It is easy to check that this leads to the formula

$$\nabla_{g_i} \varphi = \frac{1}{\mu_i} \mathbb{K} \nabla \varphi, \tag{57}$$

where $\nabla \varphi$ stands for the usual (euclidean) gradient. The formula (57) can be extended to Lipschitz continuous functions φ thanks to Rademacher's theorem.

For φ belonging to $\mathcal{C}^2$, we can also define the Hessian $D_{g_i}^2 \varphi$ of φ in the Riemannian setting by

$$g_{i,\boldsymbol{x}}(D_{g_i}^2 \varphi(\boldsymbol{x}) \cdot \boldsymbol{v}, \boldsymbol{v}) = \frac{d^2}{dt^2} \varphi(\boldsymbol{\gamma}_t) \Big|_{t=0}$$

for any geodesic $\boldsymbol{\gamma}_t = \exp_{i,\boldsymbol{x}}(t\boldsymbol{v})$ starting from $\boldsymbol{x}$ with initial speed $\boldsymbol{v} \in T_{i,\boldsymbol{x}} \Omega$.

Denote by φ_i^n the backward Kantorovich potential sending s_i^n to s_i^{n-1} associated to the cost $\frac{1}{2} d_i^2$. By the usual definition of the Wasserstein distance through the Monge problem, one has

$$W_i^2(s_i^n, s_i^{n-1}) = \int_\Omega d_i^2(\boldsymbol{x}, \boldsymbol{t}_i^n(\boldsymbol{x})) s_i^n(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x},$$

where $\boldsymbol{t}_i^n$ denotes the optimal map sending s_i^n to s_i^{n-1}. It follows from [Villani 2009, Theorem 10.41] that

$$\boldsymbol{t}_i^n(\boldsymbol{x}) = \exp_{i,\boldsymbol{x}}(-\nabla_{g_i}\varphi_i^n(\boldsymbol{x})) \quad \forall \boldsymbol{x} \in \Omega. \tag{58}$$

Moreover, using the definition of the exponential and the relation (57), one gets that

$$d_i^2\big(\boldsymbol{x}, \exp_{i,\boldsymbol{x}}(-\nabla_{g_i}\varphi_i^n(\boldsymbol{x})\big) = g_{i,\boldsymbol{x}}\big(\nabla_{g_i}\varphi_i^n(\boldsymbol{x}), \nabla_{g_i}\varphi_i^n(\boldsymbol{x})\big) = \frac{1}{\mu_i}\mathbb{K}(\boldsymbol{x})\nabla\varphi_i^n(\boldsymbol{x}) \cdot \nabla\varphi_i^n(\boldsymbol{x}).$$

This yields the formula

$$W_i^2(s_i^n, s_i^{n-1}) = \int_\Omega \frac{s_i^n}{\mu_i}\mathbb{K}\nabla\varphi_i^n \cdot \nabla\varphi_i^n \,\mathrm{d}\boldsymbol{x} \quad \forall i \in \{0, \dots, N\}. \tag{59}$$

We have now introduced the necessary material in order to reconstruct the phase and capillary pressures. This is the purpose of the following Proposition 3.4 and then of Corollary 3.5.

Proposition 3.4. *For $n \ge 1$ let $\varphi_i^n : s_i^n \to s_i^{n-1}$ be the (backward) Kantorovich potentials from Lemma 3.2. There exists $\boldsymbol{h} = (h_0^n, \dots, h_N^n) \in H^1(\Omega)^{N+1}$ such that*

(i) $\nabla h_i^n = -\nabla\varphi_i^n/\tau$ *for* $\mathrm{d}s_i^n$*-a.e.* $\boldsymbol{x} \in \Omega$,

(ii) $h_i^n(\boldsymbol{x}) - h_0^n(\boldsymbol{x}) = \pi_i^n(\boldsymbol{x}) + \Psi_i(\boldsymbol{x}) - \Psi_0(\boldsymbol{x})$ *for* $\mathrm{d}\boldsymbol{x}$*-a.e.* $\boldsymbol{x} \in \Omega$, $i \in \{1, \dots, N\}$,

(iii) *there exists C depending only on Ω, Π, ω, $\mathbb{K}$, $(\mu_i)_i$, and $\boldsymbol{\Psi}$ such that, for all $n \ge 1$ and all $\tau > 0$, one has*
$$\|\boldsymbol{h}^n\|^2_{H^1(\Omega)^{N+1}} \le C\left(1 + \frac{\boldsymbol{W}^2(\boldsymbol{s}^n, \boldsymbol{s}^{n-1})}{\tau^2} + \sum_{i=0}^N \frac{\mathcal{H}_\omega(s_i^{n-1}) - \mathcal{H}_\omega(s_i^n)}{\tau}\right).$$

Proof. Let φ_i^n be the Kantorovich potentials from Lemma 3.2 and $F_i^n \in L^\infty \cap H^1(\Omega)$ as in (47), as well as $\boldsymbol{\alpha}^n \in \mathbb{R}^{N+1}$ and $\lambda^n = \min_j(F_j^n + \alpha_j^n) \in L^\infty \cap H^1(\Omega)$ as in Corollary 3.3. Setting

$$h_i^n := -\frac{\varphi_i^n}{\tau} + F_i^n - \lambda^n \quad \forall i \in \{0, \dots, N\},$$

we have $h_i^n \in H^1(\Omega)$ as the sum of Lipschitz functions (the Kantorovich potentials φ_i^n) and H^1 functions F_i^n, λ^n. Recalling that we use the notation $\pi_0 = \partial\Pi/\partial s_0 = 0$, we see from the definition (47) of F_i^n that

$$h_i^n - h_0^n = \left(F_i^n - \frac{\varphi_i^n}{\tau}\right) - \left(F_0^n - \frac{\varphi_0^n}{\tau}\right) = (\pi_i^n + \Psi_i) - (\pi_0^n + \Psi_0) = \pi_i^n + \Psi_i - \Psi_0 \tag{60}$$

for all $i \in \{1, \dots, N\}$ and $\mathrm{d}\boldsymbol{x}$-a.e. x, which is exactly our statement (ii).

For (i), we simply use (56) to compute

$$\nabla h_i^n = -\frac{\nabla\varphi_i^n}{\tau} + \nabla(F_i^n - \lambda_i^n) = -\frac{\nabla\varphi_i^n}{\tau} \quad \text{for } \mathrm{d}s_i^n\text{-a.e. } \boldsymbol{x} \in \Omega, \ \forall i \in \{0, \dots, N\}. \tag{61}$$

In order to establish now the H^1 estimate (iii), let us define

$$\mathcal{U}_i = \{\boldsymbol{x} \in \Omega \mid s_i^n(\boldsymbol{x}) \ge \omega_\star/(N+1)\}.$$

Then since $\sum s_i^n(\boldsymbol{x}) = \omega(\boldsymbol{x}) \geq \omega_\star > 0$, one gets that, up to a negligible set,

$$\bigcup_{i=0}^{N} \mathcal{U}_i = \Omega, \quad \text{hence} \quad (\mathcal{U}_i)^c \subset \bigcup_{j \neq i} \mathcal{U}_j. \tag{62}$$

We first estimate ∇h_0^n. To this end, we write

$$\|\nabla h_0^n\|_{L^2}^2 \leq \frac{1}{\kappa_\star} \int_\Omega \mathbb{K} \nabla h_0^n \cdot \nabla h_0^n \, \mathrm{d}\boldsymbol{x} \leq A + B, \tag{63}$$

where we have set

$$A = \frac{1}{\kappa_\star} \int_{\mathcal{U}_0} \mathbb{K} \nabla h_0^n \cdot \nabla h_0^n \, \mathrm{d}\boldsymbol{x}, \quad B = \frac{1}{\kappa_\star} \int_{(\mathcal{U}_0)^c} \mathbb{K} \nabla h_0^n \cdot \nabla h_0^n \, \mathrm{d}\boldsymbol{x}.$$

Owing to (61) one has $\nabla h_0^n = -\nabla \varphi_0 / \tau$ on $\mathcal{U}_0 \subset \Omega$, where $s_0^n \geq \omega_\star/(N+1)$. Therefore,

$$A \leq \frac{(N+1)\mu_0}{\omega_\star \kappa_\star} \int_{\mathcal{U}_0} \frac{s_0^n}{\mu_0} \mathbb{K} \nabla h_0^n \cdot \nabla h_0^n \, \mathrm{d}\boldsymbol{x} \leq \frac{(N+1)\mu_0}{\tau^2 \omega_\star \kappa_\star} \int_\Omega \frac{s_0^n}{\mu_0} \mathbb{K} \nabla \varphi_0^n \cdot \nabla \varphi_0^n \, \mathrm{d}\boldsymbol{x}.$$

Then it results from formula (59) that

$$A \leq \frac{C}{\tau^2} W_0^2(s_0^n, s_0^{n-1}), \tag{64}$$

where C depends neither on n nor on τ. Combining (62) and (60), we infer

$$B \leq \frac{1}{\kappa_\star} \sum_{i=1}^{N} \int_{\mathcal{U}_i} \mathbb{K} \nabla [h_i^n - (\pi_i^n + \Psi_i - \Psi_0)] \cdot \nabla [h_i^n - (\pi_i^n + \Psi_i - \Psi_0)] \, \mathrm{d}\boldsymbol{x}.$$

Using $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$ and (3), we get that

$$B \leq \frac{3}{\kappa_\star} \sum_{i=1}^{N} \int_{\mathcal{U}_i} \mathbb{K} \nabla h_i \cdot \nabla h_i \, \mathrm{d}\boldsymbol{x} + \frac{3\kappa^\star}{\kappa_\star} \sum_{i=1}^{N} \big(\|\nabla \pi_i^n\|_{L^2}^2 + \|\nabla(\Psi_i - \Psi_0)\|_{L^2}^2 \big). \tag{65}$$

Similar calculations to those carried out to estimate A yield

$$\int_{\mathcal{U}_i} \mathbb{K} \nabla h_i \cdot \nabla h_i \, \mathrm{d}\boldsymbol{x} \leq \frac{C}{\tau^2} W_i^2(s_i^n, s_i^{n-1})$$

for some C depending neither on n, i nor on τ. Combining this inequality with Lemma 2.2 and the regularity of $\boldsymbol{\Psi}$, we get from (65) that

$$B \leq C \left(1 + \frac{\boldsymbol{W}^2(\boldsymbol{s}^n, \boldsymbol{s}^{n-1})}{\tau^2} + \sum_{i=0}^{N} \frac{\mathcal{H}_\omega(s_i^{n-1}) - \mathcal{H}_\omega(s_i^n)}{\tau} \right) \tag{66}$$

for some C not depending on n and τ (here we also used $1/\tau \leq 1/\tau^2$ for small τ in the W^2 terms). Gathering (64) and (66) in (63) provides

$$\|\nabla h_0^n\|_{L^2}^2 \leq C \left(1 + \frac{\boldsymbol{W}^2(\boldsymbol{s}^n, \boldsymbol{s}^{n-1})}{\tau^2} + \sum_{i=0}^{N} \frac{\mathcal{H}_\omega(s_i^{n-1}) - \mathcal{H}_\omega(s_i^n)}{\tau} \right).$$

Note that (i) and (ii) remain invariant under subtraction of the same constant, that is, $h_0^n, h_i^n \rightsquigarrow h_0^n - C, h_i^n - C$, as the gradients remain unchanged in (i) and only the differences $h_i^n - h_0^n$ appear in (ii) for $i \in \{1, \dots, N\}$. We can therefore assume without loss of generality that $\int_\Omega h_0^n \,\mathrm{d}\boldsymbol{x} = 0$. Hence by the Poincaré–Wirtinger inequality, we get that

$$\|h_0^n\|_{H^1}^2 \le C\|\nabla h_0^n\|_{L^2}^2 \le C\left(1 + \frac{\boldsymbol{W}^2(\boldsymbol{s}^n, \boldsymbol{s}^{n-1})}{\tau^2} + \sum_{i=0}^N \frac{\mathcal{H}_\omega(s_i^{n-1}) - \mathcal{H}_\omega(s_i^n)}{\tau}\right).$$

Finally, from (ii) $h_i^n = h_0^n + \pi_i^n + \Psi_i - \Psi_0$, the smoothness of $\boldsymbol{\Psi}$, and using again the estimate (31) for $\|\nabla \boldsymbol{\pi}^n\|_{L^2}^2$ we finally get that for all $i \in \{1, \dots, N\}$, one has

$$\|h_i^n\|_{H^1}^2 \le C\big(\|h_0^n\|_{H^1}^2 + \|\pi_i^n\|_{H^1}^2 + \|\Psi_i\|_{H^1}^2 + \|\Psi_0\|_{H^1}^2\big) \le C\left(1 + \frac{\boldsymbol{W}^2(\boldsymbol{s}^n, \boldsymbol{s}^{n-1})}{\tau^2} + \sum_{i=0}^N \frac{\mathcal{H}_\omega(s_i^{n-1}) - \mathcal{H}_\omega(s_i^n)}{\tau}\right).$$

and the proof of Proposition 3.4 is complete. □

We can now define the phase pressures $(p_i^n)_{i=0,\dots,N}$ by setting

$$p_i^n := h_i^n - \Psi_i \quad \forall i \in \{0, \dots, N\}. \tag{67}$$

The following corollary is a straightforward consequence of Proposition 3.4 and of the regularity of Ψ_i.

Corollary 3.5. *The phase pressures $\boldsymbol{p}^n = (p_i^n)_{0\le i\le N} \in H^1(\Omega)^{N+1}$ satisfy*

$$\|\boldsymbol{p}^n\|_{H^1(\Omega)}^2 \le C\left(1 + \frac{\boldsymbol{W}^2(\boldsymbol{s}^n, \boldsymbol{s}^{n-1})}{\tau^2} + \sum_{i=0}^N \frac{\mathcal{H}_\omega(s_i^{n-1}) - \mathcal{H}_\omega(s_i^n)}{\tau}\right) \tag{68}$$

for some C depending only on Ω, Π, ω, $\mathbb{K}$, $(\mu_i)_i$, and $\boldsymbol{\Psi}$ (but neither on n nor on τ), and the capillary pressure relations are fulfilled:

$$p_i^n - p_0^n = \pi_i^n \quad \forall i \in \{1, \dots, N\}. \tag{69}$$

Our next result is a first step towards the recovery of the PDEs.

Lemma 3.6. *There exists C depending only on Ω, Π, ω, $\mathbb{K}$, $(\mu_i)_i$, and $\boldsymbol{\Psi}$ (but neither on n nor on τ) such that, for all $i \in \{0, \dots, N\}$ and all $\xi \in \mathcal{C}^2(\overline{\Omega})$, one has*

$$\left|\int_\Omega (s_i^n - s_i^{n-1})\xi \,\mathrm{d}\boldsymbol{x} + \tau \int_\Omega s_i^n \frac{\mathbb{K}}{\mu_i}\nabla(p_i^n + \Psi_i)\cdot\nabla\xi \,\mathrm{d}\boldsymbol{x}\right| \le C\, W_i^2(s_i^n, s_i^{n-1})\|D_{g_i}^2\xi\|_\infty. \tag{70}$$

This is of course a discrete approximation to the continuity equation $\partial_t s_i = \nabla\cdot(s_i(\mathbb{K}/\mu_i)\nabla(p_i + \Psi_i))$.

Proof. Let φ_i^n denote the (backward) optimal Kantorovich potential from Lemma 3.2 sending s_i^n to s_i^{n-1}, and let $\boldsymbol{t}_i^n$ be the corresponding optimal map as in (58). For fixed $\xi \in \mathcal{C}^2(\overline{\Omega})$ let us first Taylor expand (in the g_i Riemannian framework)

$$\left|\xi(\boldsymbol{t}_i^n(\boldsymbol{x})) - \xi(\boldsymbol{x}) + \frac{1}{\mu_i}\mathbb{K}(\boldsymbol{x})\nabla\xi(\boldsymbol{x})\cdot\nabla\varphi_i^n(\boldsymbol{x})\right| \le \tfrac{1}{2}\|D_{g_i}^2\xi\|_\infty d_i^2(\boldsymbol{x}, \boldsymbol{t}_i^n(\boldsymbol{x})).$$

Using the definition of the pushforward $s_i^{n-1} = \boldsymbol{t}_i^n \# s_i^n$, we then compute

$$\begin{aligned}\left|\int_\Omega (s_i^n(\boldsymbol{x}) - s_i^{n-1}(\boldsymbol{x}))\xi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_\Omega \frac{\mathbb{K}(\boldsymbol{x})}{\mu_i}\nabla\xi(x)\cdot\nabla\varphi_i^n(\boldsymbol{x})s_i^n(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right| \\ = \left|\int_\Omega (\xi(\boldsymbol{x}) - \xi(\boldsymbol{t}_i^n(\boldsymbol{x}))s_i^n(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_\Omega \frac{\mathbb{K}(\boldsymbol{x})}{\mu_i}\nabla\xi(x)\cdot\nabla\varphi_i^n(\boldsymbol{x})s_i^n(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right| \\ \le \int_\Omega \tfrac{1}{2}\|D_{g_i}^2\xi\|_\infty d_i^2(\boldsymbol{x}, \boldsymbol{t}_i^n(\boldsymbol{x}))s_i^n(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \tfrac{1}{2}\|D_{g_i}^2\xi\|_\infty W_i^2(s_i^n, s_i^{n-1}).\end{aligned}$$

From Proposition 3.4(i) we have $\nabla\varphi_i^n = -\tau\nabla h_i^n$ for $\mathrm{d}s_i^n$-a.e. $\boldsymbol{x} \in \Omega$; thus by the definition (67) of p_i^n, we get $\nabla\varphi^n = -\tau\nabla(p_i^n + \Psi_i)$. Substituting in the second integral of the left-hand side gives exactly (70). □

4. Convergence towards a weak solution

The goal is now to prove the convergence of the piecewise constant interpolated solutions $\boldsymbol{s}^\tau$, defined by (27), towards a weak solution $\boldsymbol{s}$ as $\tau \to 0$. Similarly, the τ superscript denotes the piecewise constant interpolation of any previous discrete quantity (e.g., $p_i^\tau(t)$ stands for the piecewise constant time interpolation of the discrete pressures p_i^n). In what follows, we will also use the notation $\boldsymbol{s}^{\tau *} = (s_1^\tau, \dots, s_N^\tau) \in L^\infty((0,T); \boldsymbol{\mathcal{X}}^*)$ and $\boldsymbol{\pi}^\tau = \boldsymbol{\pi}(\boldsymbol{s}^{\tau *}, \boldsymbol{x})$.

Time integrated estimates. We immediately deduce from (30) that

$$\boldsymbol{W}(\boldsymbol{s}^\tau(t_2), \boldsymbol{s}^\tau(t_1)) \le C|t_2 - t_1 + \tau|^{1/2} \quad \forall\, 0 \le t_1 \le t_2 \le T. \tag{71}$$

From the total saturation $\sum_{i=0}^N s_i^n(\boldsymbol{x}) = \omega(\boldsymbol{x}) \le \omega^\star$ and $s_i^\tau \ge 0$, we have the L^∞-estimates

$$0 \le s_i^\tau(\boldsymbol{x}, t) \le \omega^\star \quad \text{a.e. in } Q \text{ for all } i \in \{0, \dots, N\}. \tag{72}$$

Lemma 4.1. *There exists C depending only on Ω, T, Π, ω, $\mathbb{K}$, $(\mu_i)_i$, and $\boldsymbol{\Psi}$ such that*

$$\|\boldsymbol{p}^\tau\|^2_{L^2((0,T);H^1(\Omega)^{N+1})} + \|\boldsymbol{\pi}^\tau\|^2_{L^2((0,T);H^1(\Omega)^N)} \le C. \tag{73}$$

Proof. Summing (68) from $n = 1$ to $n = N_\tau := \lceil T/\tau \rceil$, we get

$$\begin{aligned}\|\boldsymbol{p}^\tau\|^2_{L^2(H^1)} = \sum_{n=1}^{N_\tau} \tau\|\boldsymbol{p}^n\|^2_{H^1} \le C\sum_{n=1}^{N_\tau}\tau\left(1 + \frac{\boldsymbol{W}^2(\boldsymbol{s}^n, \boldsymbol{s}^{n-1})}{\tau^2} + \sum_{i=0}^{N_\tau}\frac{\mathcal{H}_\omega(s_i^{n-1}) - \mathcal{H}_\omega(s_i^n)}{\tau}\right) \\ \le C\left((T+1) + \sum_{n=1}^{N_\tau}\frac{\boldsymbol{W}^2(\boldsymbol{s}^n, \boldsymbol{s}^{n-1})}{\tau} + \sum_{i=0}^{N}\big(\mathcal{H}_\omega(s_i^0) - \mathcal{H}_\omega(s_i^{N_\tau})\big)\right).\end{aligned}$$

We use that

$$0 \ge \mathcal{H}_\omega(s) \ge -\frac{1}{e}\|\omega\|_{L^1} \ge -\frac{|\Omega|}{e} \quad \forall s \in L^\infty(\Omega) \text{ with } 0 \le s \le \omega$$

together with the total square distance estimate (29) to infer that $\|\boldsymbol{p}\|^2_{L^2(H^1)} \le C$. The proof is identical for the capillary pressure $\boldsymbol{\pi}^\tau$ (simply summing the one-step estimate from Lemma 2.2). □

Compactness of approximate solutions. We define $H' = H^1(\Omega)'$.

Lemma 4.2. *For each $i \in \{0, \dots, N\}$, there exists C depending only on Ω, Π, $\boldsymbol{\Psi}$, $\mathbb{K}$, and μ_i (but not on τ) such that*

$$\|s_i^\tau(t_2) - s_i^\tau(t_1)\|_{H'} \le C|t_2 - t_1 + \tau|^{1/2} \quad \forall\, 0 \le t_1 \le t_2 \le T.$$

Proof. Thanks to (72), we can apply [Maury et al. 2010, Lemma 3.4] to get

$$\left|\int_\Omega f\{s_i^\tau(t_2) - s_i^\tau(t_1)\}\,\mathrm{d}\boldsymbol{x}\right| \le \|\nabla f\|_{L^2(\Omega)} W_{\mathrm{ref}}(s_i^\tau(t_1), s_i^\tau(t_2)) \quad \forall f \in H^1(\Omega).$$

Thus by duality and thanks to the distance estimate (71) and to the lower bound in (19), we obtain that

$$\|s_i^\tau(t_2) - s_i^\tau(t_1)\|_{H'} \le W_{\mathrm{ref}}(s_i^\tau(t_1), s_i^\tau(t_2)) \le C W_i(s_i^\tau(t_1), s_i^\tau(t_2)) \le C|t_2 - t_1 + \tau|^{1/2}$$

for some C depending only on Ω, Π, $(\rho_i)_i$, $\boldsymbol{g}$, $(\mu_i)_i$, $\mathbb{K}$. □

From the previous equicontinuity in time, we deduce full compactness of the capillary pressure:

Lemma 4.3. *The family $(\boldsymbol{\pi}^\tau)_{\tau>0}$ is sequentially relatively compact in $L^2(Q)^N$.*

Proof. We use Alt and Luckhaus' trick [1983] (an alternate solution would consist in slightly adapting the nonlinear time compactness results [Moussa 2016; Andreianov et al. 2015] to our context). Let $h > 0$ be a small time shift; then by monotonicity and Lipschitz continuity of the capillary pressure function $\boldsymbol{\pi}(\,\cdot\,, \boldsymbol{x})$,

$$\begin{aligned}\|\boldsymbol{\pi}^\tau(\cdot+h)-\boldsymbol{\pi}^\tau(\cdot)\|^2_{L^2((0,T-h);L^2(\Omega)^N)} &\le \frac{1}{\kappa_\star}\int_0^{T-h}\int_\Omega\big(\boldsymbol{\pi}^\tau(t+h,\boldsymbol{x})-\boldsymbol{\pi}^\tau(t,\boldsymbol{x})\big)\cdot\big(\boldsymbol{s}^{\tau*}(t+h,\boldsymbol{x})-\boldsymbol{s}^{\tau*}(t,\boldsymbol{x})\big)\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t\\ &\le \frac{2\sqrt{T}}{\kappa_\star}\|\boldsymbol{\pi}^\tau\|_{L^2((0,T);H^1(\Omega)^N)}\|\boldsymbol{s}^{\tau*}(\cdot+h,\cdot)-\boldsymbol{s}^{\tau*}\|_{L^\infty((0,T-h);H')^N}.\end{aligned}$$

Then it follows from Lemmas 4.1 and 4.2 that there exists $C > 0$, depending neither on h nor on τ, such that

$$\|\boldsymbol{\pi}^\tau(\cdot\,+h,\cdot\,) - \boldsymbol{\pi}^\tau\|_{L^2((0,T-h);L^2(\Omega)^N)} \le C|h+\tau|^{1/2}.$$

On the other hand, the (uniform with respect to τ) $L^2((0,T);H^1(\Omega)^N)$- and $L^\infty(Q)^N$-estimates on $\boldsymbol{\pi}^\tau$ ensure that

$$\|\boldsymbol{\pi}^\tau(\,\cdot\,,\,\cdot\,+\boldsymbol{y})) - \boldsymbol{\pi}^\tau\|_{L^2(0,T;L^2)} \le C\sqrt{|\boldsymbol{y}|}(1+\sqrt{|\boldsymbol{y}|}) \quad \forall\, \boldsymbol{y} \in \mathbb{R}^d,$$

where $\boldsymbol{\pi}^\tau$ is extended by 0 outside Ω. This allows us to apply Kolmogorov's compactness theorem, see, for instance, [Hanche-Olsen and Holden 2010], and gives the desired relative compactness. □

Identification of the limit. In this section we prove our main result, Theorem 1.2, and the proof goes in two steps: we first retrieve strong convergence of the phase contents $\boldsymbol{s}^\tau \to \boldsymbol{s}$ and weak convergence of the pressures $\boldsymbol{p}^\tau \rightharpoonup \boldsymbol{p}$, and then use the strong-weak limit of products to show that the limit is a weak solution. Throughout this section, $(\tau_k)_{k\ge1}$ denotes a sequence of times steps tending to 0 as $k \to \infty$.

Lemma 4.4. *There exist $\boldsymbol{p} \in L^2((0,T); H^1(\Omega)^{N+1})$ and $\boldsymbol{s} \in L^\infty(Q)^{N+1}$ with $\boldsymbol{s}(\cdot, t) \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$ for a.e. $t \in (0,T)$ such that, up to an unlabeled subsequence, the following convergence properties hold:*

$$\boldsymbol{s}^{\tau_k} \xrightarrow{k\to\infty} \boldsymbol{s} \qquad \text{a.e. in } Q, \tag{74}$$

$$\boldsymbol{\pi}^{\tau_k} \xrightarrow{k\to\infty} \boldsymbol{\pi}(\boldsymbol{s}^*, \cdot) \qquad \text{weakly in } L^2((0,T); H^1(\Omega)^N), \tag{75}$$

$$\boldsymbol{p}^{\tau_k} \xrightarrow{k\to\infty} \boldsymbol{p} \qquad \text{weakly in } L^2((0,T); H^1(\Omega)^{N+1}). \tag{76}$$

Moreover, the capillary pressure relations (5) hold.

Proof. From Lemma 4.3, we can assume that $\boldsymbol{\pi}^{\tau_k} \to \boldsymbol{z}$ strongly in $L^2(Q)^N$ for some limit $\boldsymbol{z}$, thus a.e. up to the extraction of an additional subsequence. Since $\boldsymbol{z} \mapsto \boldsymbol{\phi}(\boldsymbol{z}, \boldsymbol{x}) = \boldsymbol{\pi}^{-1}(\boldsymbol{z}, \boldsymbol{x})$ is continuous, we have

$$\boldsymbol{s}^{\tau_k *} = \boldsymbol{\phi}(\boldsymbol{\pi}^{\tau_k}, \boldsymbol{x}) \xrightarrow{k\to\infty} \boldsymbol{\phi}(\boldsymbol{\pi}, \boldsymbol{x}) =: \boldsymbol{s}^* \quad \text{a.e. in } Q.$$

In particular, this yields $\boldsymbol{\pi}^{\tau_k} \xrightarrow{k\to\infty} \boldsymbol{\pi}(\boldsymbol{s}^*, \cdot)$ a.e. in Q. Since we have the total saturation $\sum_{i=0}^N s_i^{\tau_k}(t, \boldsymbol{x}) = \omega(\boldsymbol{x})$, we conclude that the first component $i = 0$ converges pointwise as well. Therefore, (74) holds. Thanks to Lebesgue's dominated convergence theorem, it is easy to check that $\boldsymbol{s}(\cdot, t) \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$ for a.e. $t \in (0,T)$. The convergences (75) and (76) are straightforward consequences of Lemma 4.1. Lastly, it follows from (69) that

$$p_i^{\tau_k} - p_0^{\tau_k} = \boldsymbol{\pi}_i(\boldsymbol{s}^{\tau_k *}, \cdot) \quad \forall i \in \{1, \dots, N\},\ \forall k \geq 1.$$

We can finally pass to the limit $k \to \infty$ in the above relation thanks to (75)–(76) and infer

$$p_i - p_0 = \boldsymbol{\pi}_i(\boldsymbol{s}^*, \boldsymbol{x}) \quad \text{in } L^2((0,T); H^1(\Omega)),\ \forall i \in \{1, \dots, N\},$$

which immediately implies (5) as claimed. □

Lemma 4.5. *Up to the extraction of an additional subsequence, the limit $\boldsymbol{s}$ of $(\boldsymbol{s}^{\tau_k})_{k\geq 1}$ belongs to $C([0,T]; \boldsymbol{\mathcal{A}})$, where $\boldsymbol{\mathcal{A}}$ is equipped with the metric $\boldsymbol{W}$. Moreover, $\boldsymbol{W}(\boldsymbol{s}^{\tau_k}(t), \boldsymbol{s}(t)) \xrightarrow{k\to\infty} 0$ for all $t \in [0,T]$.*

Proof. It follows from the bounds (72) on s_i that for all $t \in [0,T]$, the sequence $(s_i^{\tau_k})_k$ is weakly compact in $L^1(\Omega)$. It is also compact in $\mathcal{A}_i$ equipped with the metric W_i due to the continuity of W_i with respect to the weak convergence in $L^1(\Omega)$; this is, for instance, a consequence of [Santambrogio 2015, Theorem 5.10] together with the equivalence of W_i with W_{ref} stated in (19). Thanks to (71), one has

$$\limsup_{k\to\infty} W_i(s_i^{\tau_k}(t_2), s_i^{\tau_k}(t_1)) \leq |t_2 - t_1|^{1/2} \quad \forall t_1, t_2 \in [0,T].$$

Applying a refined version of the Arzelà–Ascoli theorem [Ambrosio et al. 2008, Proposition 3.3.1] then provides the desired result. □

In order to conclude the proof of Theorem 1.2, it only remains to show that $\boldsymbol{s} = \lim \boldsymbol{s}^{\tau_k}$ and $\boldsymbol{p} = \lim \boldsymbol{p}^{\tau_k}$ satisfy the weak formulation (12):

Proposition 4.6. *Let $(\tau_k)_{k\geq 1}$ be a sequence such that the convergences in Lemmas 4.4 and 4.5 hold. Then the limit $\boldsymbol{s}$ of $(\boldsymbol{s}^{\tau_k})_{k\geq 1}$ is a weak solution in the sense of Definition 1.1 (with $-\rho_i \boldsymbol{g}$ replaced by $+\nabla\Psi_i$ in the general case).*

Proof. Let $0 \leq t_1 \leq t_2 \leq T$, and define $n_{j,k} = \lceil t_j/\tau_k \rceil$ and $\tilde{t}_j = n_{j,k}\tau_k$ for $j \in \{1, 2\}$. Fixing an arbitrary $\xi \in \mathcal{C}^2(\overline{\Omega})$ and summing (70) from $n = n_{1,k} + 1$ to $n = n_{2,k}$ yields

$$\int_\Omega (s_i^{\tau_k}(t_2) - s_i^{\tau_k}(t_1))\xi \,\mathrm{d}\boldsymbol{x} = \sum_{n=n_{1,k}+1}^{n_{2,k}} \int_\Omega (s_i^n - s_i^{n-1})\xi \,\mathrm{d}\boldsymbol{x}$$
$$= -\int_{\tilde{t}_1}^{\tilde{t}_2}\int_\Omega \frac{s_i^{\tau_k}}{\mu_i}\mathbb{K}\nabla(p_i^{\tau_k} + \Psi_i)\cdot\nabla\xi \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t + \mathcal{O}\bigg(\sum_{n=n_{1,k}+1}^{n_{2,k}} W_i^2(s_i^n, s_i^{n-1})\bigg). \tag{77}$$

Since $0 \leq \tilde{t}_j - t_j \leq \tau_k$ and $(s_i^{\tau_k}/\mu_i)\mathbb{K}\nabla(p_i^{\tau_k} + \Psi_i)\cdot\nabla\xi$ is uniformly bounded in $L^2(Q)$, one has

$$\int_{\tilde{t}_1}^{\tilde{t}_2}\int_\Omega \frac{s_i^{\tau_k}}{\mu_i}\mathbb{K}\nabla(p_i^{\tau_k} + \Psi_i)\cdot\nabla\xi \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t = \int_{t_1}^{t_2}\int_\Omega \frac{s_i^{\tau_k}}{\mu_i}\mathbb{K}\nabla(p_i^{\tau_k} + \Psi_i)\cdot\nabla\xi \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t + \mathcal{O}(\sqrt{\tau_k}).$$

Combining the above estimate with the total square distance estimate (29) in (77), we obtain

$$\int_\Omega (s_i^{\tau_k}(t_2) - s_i^{\tau_k}(t_1))\xi \,\mathrm{d}\boldsymbol{x} + \int_{t_1}^{t_2}\int_\Omega \frac{s_i^{\tau_k}}{\mu_i}\mathbb{K}\nabla(p_i^{\tau_k} + \Psi_i)\cdot\nabla\xi \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t = \mathcal{O}(\sqrt{\tau_k}). \tag{78}$$

Thanks to Lemma 4.5, and since the convergence in $(\mathcal{A}_i, W_i)$ is equivalent to the narrow convergence of measures (i.e., the convergence in $\mathcal{C}(\overline{\Omega})'$, see for instance [Santambrogio 2015, Theorem 5.10]), we get

$$\int_\Omega (s_i^{\tau_k}(t_2) - s_i^{\tau_k}(t_1))\xi \,\mathrm{d}\boldsymbol{x} \xrightarrow{k\to\infty} \int_\Omega (s_i(t_2) - s_i(t_1))\xi \,\mathrm{d}\boldsymbol{x}. \tag{79}$$

Moreover, thanks to Lemma 4.4, one has

$$\int_{t_1}^{t_2}\int_\Omega \frac{s_i^{\tau_k}}{\mu_i}\mathbb{K}\nabla(p_i^{\tau_k} + \Psi_i)\cdot\nabla\xi \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \xrightarrow{k\to\infty} \int_{t_1}^{t_2}\int_\Omega \frac{s_i}{\mu_i}\mathbb{K}\nabla(p_i + \Psi_i)\cdot\nabla\xi \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t. \tag{80}$$

Combining (78)–(80) yields, for all $\xi \in \mathcal{C}^2(\overline{\Omega})$ and all $0 \leq t_1 \leq t_2 \leq T$,

$$\int_\Omega (s_i(t_2) - s_i(t_1))\xi \,\mathrm{d}\boldsymbol{x} + \int_{t_1}^{t_2}\int_\Omega \frac{s_i}{\mu_i}\mathbb{K}\nabla(p_i + \Psi_i)\cdot\nabla\xi \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t = 0. \tag{81}$$

In order to conclude the proof, it remains to check that the formulation (81) is stronger the formulation (12). Let $\varepsilon > 0$ be a time step, unrelated to that appearing in the minimization scheme (26), and set $L_\varepsilon = \lfloor T/\varepsilon \rfloor$. Let $\phi \in \mathcal{C}_c^\infty(\overline{\Omega} \times [0, T))$, and set $\phi_\ell = \phi(\,\cdot\,, \ell\varepsilon)$ for $\ell \in \{0, \ldots, L_\varepsilon\}$. Since $t \mapsto \phi(\,\cdot\,, t)$ is compactly supported in $[0, T)$, there exists $\varepsilon^\star > 0$ such that $\phi_{L_\varepsilon} \equiv 0$ for all $\varepsilon \in (0, \varepsilon^\star]$. Then define

$$\phi^\varepsilon : \overline{\Omega} \times [0, T] \to \mathbb{R}, \qquad (\boldsymbol{x}, t) \mapsto \phi_\ell(\boldsymbol{x}) \quad \text{if } t \in [\ell\varepsilon, (\ell + 1)\varepsilon).$$

Choose $t_1 = \ell\varepsilon,\ t_2 = (\ell + 1)\varepsilon,\ \xi = \phi_\ell$ in (81) and sum over $\ell \in \{0, \ldots, L_\varepsilon - 1\}$. This provides

$$A(\varepsilon) + B(\varepsilon) = 0 \quad \forall \varepsilon > 0, \tag{82}$$

where

$$A(\varepsilon) = \sum_{\ell=0}^{L_\varepsilon - 1} \int_\Omega \big(s_i((\ell+1)\varepsilon) - s_i(\ell\varepsilon)\big)\phi^\ell \,\mathrm{d}\boldsymbol{x}, \quad B(\varepsilon) = \iint_Q \frac{s_i}{\mu_i} \mathbb{K}\nabla(p_i + \Psi_i) \cdot \nabla\phi^\varepsilon \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t.$$

Due to the regularity of ϕ, we know $\nabla\phi^\varepsilon$ converges uniformly towards ϕ as ε tends to 0, so that

$$B(\varepsilon) \xrightarrow{\varepsilon\to 0} \iint_Q \frac{s_i}{\mu_i} \mathbb{K}\nabla(p_i + \Psi_i) \cdot \nabla\phi \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t. \tag{83}$$

Reorganizing the first term and using that $\phi_{L_\varepsilon} \equiv 0$, we get

$$A(\varepsilon) = -\sum_{\ell=1}^{L_\varepsilon} \varepsilon \int_\Omega s_i(\ell\varepsilon) \frac{\phi_\ell - \phi_{\ell-1}}{\varepsilon} \,\mathrm{d}\boldsymbol{x} - \int_\Omega s_i^0 \phi(\,\cdot\,, 0) \,\mathrm{d}\boldsymbol{x}.$$

It follows from the continuity of $t \mapsto s_i(\,\cdot\,, t)$ in $\mathcal{A}_i$ equipped with W_i and from the uniform convergence of

$$(\boldsymbol{x}, t) \mapsto \frac{\phi_\ell(\boldsymbol{x}) - \phi_{\ell-1}(\boldsymbol{x})}{\varepsilon} \quad \text{if } t \in [(\ell-1)\varepsilon, \ell\varepsilon)$$

towards $\partial_t\phi$ that

$$A(\varepsilon) \xrightarrow{\varepsilon\to 0} -\iint_Q s_i\, \partial_t\phi \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t - \int_\Omega s_i^0 \phi(\,\cdot\,, 0) \,\mathrm{d}\boldsymbol{x}. \tag{84}$$

Combining (82)–(84) shows that the weak formulation (12) is fulfilled. □

Appendix A: A simple condition for the geodesic convexity of $(\boldsymbol{\Omega}, \boldsymbol{d_i})$

The goal of this appendix is to provide a simple condition on the permeability tensor in order to ensure that condition (22) is fulfilled. For the sake of simplicity, we only consider here the case of isotropic permeability tensors

$$\mathbb{K}(\boldsymbol{x}) = \kappa(\boldsymbol{x})\mathbb{I}_d \quad \forall \boldsymbol{x} \in \overline{\Omega} \tag{85}$$

with $\kappa_\star \le \kappa(\boldsymbol{x}) \le \kappa^\star$ for all $\boldsymbol{x} \in \overline{\Omega}$. Let us stress that the condition we provide is not optimal.

As in the core of the paper, Ω denotes a convex open subset of $\mathbb{R}^d$ with C^2 boundary $\partial\Omega$. For $\bar{\boldsymbol{x}} \in \partial\Omega$, we denote by $\boldsymbol{n}(\bar{\boldsymbol{x}})$ the outward-pointing normal. Since $\partial\Omega$ is smooth, there exists $\ell_0 > 0$ such that, for all $\boldsymbol{x} \in \Omega$ such that $\mathrm{dist}(\boldsymbol{x}, \partial\Omega) < \ell_0$, there exists a unique $\bar{\boldsymbol{x}} \in \partial\Omega$ such that $\mathrm{dist}(\boldsymbol{x}, \partial\Omega) = |\boldsymbol{x} - \bar{\boldsymbol{x}}|$ (here dist denotes the usual euclidean distance between sets in $\mathbb{R}^d$). As a consequence, one can rewrite $\boldsymbol{x} = \bar{\boldsymbol{x}} - \ell\boldsymbol{n}(\bar{\boldsymbol{x}})$ for some $\ell \in (0, \ell_0)$.

In what follows, a function $f : \overline{\Omega} \to \mathbb{R}$ is said to be normally nondecreasing (resp. nonincreasing) on a neighborhood of $\partial\Omega$ if there exists $\ell_1 \in (0, \ell_0]$ such that $\ell \mapsto f(\bar{\boldsymbol{x}} - \ell\boldsymbol{n}(\bar{\boldsymbol{x}}))$ is nonincreasing (resp. nondecreasing) on $[0, \ell_1]$.

Proposition A.1. *Assume that*

(i) *the permeability field $\boldsymbol{x} \mapsto \kappa(\boldsymbol{x})$ is normally nonincreasing in a neighborhood of $\partial\Omega$;*

(ii) *for all $\bar{\boldsymbol{x}} \in \partial\Omega$, either $\nabla\kappa(\bar{\boldsymbol{x}}) \cdot \boldsymbol{n}(\bar{\boldsymbol{x}}) < 0$, or $\nabla\kappa(\bar{\boldsymbol{x}}) \cdot \boldsymbol{n}(\bar{\boldsymbol{x}}) = 0$ and $D^2\kappa(\bar{\boldsymbol{x}})\boldsymbol{n}(\boldsymbol{x}) \cdot \boldsymbol{n}(\boldsymbol{x}) = 0$.*

Then there exists a C^2 extension $\tilde{\kappa} : \mathbb{R}^d \to \left[\frac{1}{2}\kappa_\star, \kappa^\star\right]$ of κ and a Riemannian metric

$$\tilde{\delta}(\boldsymbol{x}, \boldsymbol{y}) = \inf_{\boldsymbol{\gamma} \in \widetilde{P}(\boldsymbol{x},\boldsymbol{y})} \left(\int_0^1 \frac{1}{\tilde{\kappa}(\boldsymbol{\gamma}(\tau))} |\boldsymbol{\gamma}'(\tau)|^2 \, \mathrm{d}\tau \right)^{1/2} \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d \tag{86}$$

with $\widetilde{P}(\boldsymbol{x}, \boldsymbol{y}) = \{\boldsymbol{\gamma} \in C^1([0, 1]; \mathbb{R}^d) \mid \boldsymbol{\gamma}(0) = \boldsymbol{x}$ and $\boldsymbol{\gamma}(1) = \boldsymbol{y}\}$, such that $(\Omega, \tilde{\delta})$ is geodesically convex.

Proof. Since Ω is convex, for all $\boldsymbol{x} \in \mathbb{R}^d \setminus \Omega$ there exists a unique $\bar{\boldsymbol{x}} \in \partial\Omega$ such that $\operatorname{dist}(\boldsymbol{x}, \Omega) = |\boldsymbol{x} - \bar{\boldsymbol{x}}|$. Then one can extend κ in a C^2 way into the whole $\mathbb{R}^d$ by defining

$$\kappa(\boldsymbol{x}) = \kappa(\bar{\boldsymbol{x}}) + |\boldsymbol{x} - \bar{\boldsymbol{x}}| \nabla\kappa(\bar{\boldsymbol{x}}) \cdot \boldsymbol{n}(\bar{\boldsymbol{x}}) + \tfrac{1}{2}|\boldsymbol{x} - \bar{\boldsymbol{x}}|^2 D^2\kappa(\bar{\boldsymbol{x}})\boldsymbol{n}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{n}(\bar{\boldsymbol{x}}), \quad \forall \boldsymbol{x} \in \mathbb{R}^d \setminus \Omega.$$

Thanks to assumptions (i) and (ii), the function $\ell \mapsto \kappa(\bar{\boldsymbol{x}} - \ell\boldsymbol{n}(\bar{\boldsymbol{x}}))$ is nondecreasing on $(-\infty, \ell_1]$ for all $\bar{\boldsymbol{x}} \in \partial\Omega$. Since $\partial\Omega$ is compact, there exists $\ell_2 > 0$ such that

$$\kappa(\bar{\boldsymbol{x}} - \ell\boldsymbol{n}(\bar{\boldsymbol{x}})) \geq \tfrac{1}{2}\kappa_\star \quad \forall \ell \in (-\ell_2, 0].$$

Let $\rho : \mathbb{R}_+ \to \mathbb{R}$ be a nondecreasing C^2 function such that $\rho(0) = 1$, $\rho'(0) = \rho''(0) = 0$ and $\rho(\ell) = 0$ for all $\ell \geq \ell_2$. Then define

$$\tilde{\kappa}(\boldsymbol{x}) = \rho(\operatorname{dist}(\boldsymbol{x}, \Omega))\kappa(\boldsymbol{x}) + (1 - \rho(\operatorname{dist}(\boldsymbol{x}, \Omega)))\tfrac{1}{2}\kappa_\star \quad \forall \boldsymbol{x} \in \mathbb{R}^d,$$

so that the function $\ell \mapsto \tilde{\kappa}(\bar{\boldsymbol{x}} - \ell\boldsymbol{n}(\bar{\boldsymbol{x}}))$ is nonincreasing on $(-\infty, \ell_1)$ and bounded from below by $\frac{1}{2}\kappa_\star$.

Let $\boldsymbol{x}, \boldsymbol{y} \in \Omega$; then there exists $\varepsilon > 0$ such that $\operatorname{dist}(\boldsymbol{x}, \partial\Omega) \geq \varepsilon$, $\operatorname{dist}(\boldsymbol{y}, \partial\Omega) \geq \varepsilon$, and κ is normally nonincreasing on $\partial\Omega_\varepsilon := \{\boldsymbol{x} \in \overline{\Omega} \mid \operatorname{dist}(\boldsymbol{x}, \partial\Omega) < \varepsilon\}$. A sufficient condition for $(\Omega, \tilde{\delta})$ to be geodesic is that the geodesic $\boldsymbol{\gamma}^{\mathrm{opt}}_{\boldsymbol{x},\boldsymbol{y}}$ from $\boldsymbol{x}$ to $\boldsymbol{y}$ is such that

$$\operatorname{dist}(\boldsymbol{\gamma}^{\mathrm{opt}}_{\boldsymbol{x},\boldsymbol{y}}(t), \partial\Omega) \geq \varepsilon, \quad \forall t \in [0, 1]. \tag{87}$$

In order to ease the reading, we denote by $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{\mathrm{opt}}_{\boldsymbol{x},\boldsymbol{y}}$ any geodesic such that

$$\tilde{\delta}^2(\boldsymbol{x}, \boldsymbol{y}) = \int_0^1 \frac{1}{\tilde{\kappa}(\boldsymbol{\gamma}(\tau))} |\boldsymbol{\gamma}'(\tau)|^2 \, \mathrm{d}\tau. \tag{88}$$

We define the continuous and piecewise C^1 path $\boldsymbol{\gamma}_\varepsilon$ from $\boldsymbol{x}$ to $\boldsymbol{y}$ by setting

$$\boldsymbol{\gamma}_\varepsilon(t) = \operatorname{proj}_{\overline{\Omega}_\varepsilon}(\boldsymbol{\gamma}(t)) \quad \forall t \in [0, 1], \tag{89}$$

where $\overline{\Omega}_\varepsilon := \{\boldsymbol{x} \in \Omega \mid \operatorname{dist}(\boldsymbol{x}, \partial\Omega) \geq \varepsilon\}$ is convex, and the orthogonal (with respect to the euclidean distance dist) projection $\operatorname{proj}_{\overline{\Omega}_\varepsilon}$ onto $\overline{\Omega}_\varepsilon$ is therefore uniquely defined.

Assume that condition (87) is violated. Then by continuity there exists a nonempty interval $[a, b] \subset [0, 1]$ such that

$$\operatorname{dist}(\boldsymbol{\gamma}(t), \partial\Omega) < \varepsilon \quad \forall t \in (a, b);$$

that is, the geodesic between $\gamma(a)$ and $\gamma(b)$ coincides with the part of the geodesic between $\boldsymbol{x}$ and $\boldsymbol{y}$. Then, changing $\boldsymbol{x}$ into $\boldsymbol{\gamma}(a)$ and $\boldsymbol{y}$ into $\boldsymbol{\gamma}(b)$, we can assume without loss of generality that

$$\operatorname{dist}(\boldsymbol{\gamma}(t), \partial\Omega) < \varepsilon \quad \forall t \in (0, 1).$$

It is easy to verify that

$$|\boldsymbol{\gamma}_\varepsilon'(t)| \le |\boldsymbol{\gamma}'(t)| \quad \forall t \in [0,1] \qquad \text{and} \quad |\boldsymbol{\gamma}_\varepsilon'(t)| < |\boldsymbol{\gamma}'(t)| \quad \text{on } (a,b) \tag{90}$$

for some nonempty interval $(a,b) \subset [0,1]$. It follows from (86) that

$$\tilde{\delta}^2(\boldsymbol{x},\boldsymbol{y}) \le \int_0^1 \frac{1}{\tilde{\kappa}(\boldsymbol{\gamma}_\varepsilon(\tau))} |\boldsymbol{\gamma}_\varepsilon'(\tau)|^2 \,\mathrm{d}\tau.$$

Since κ is normally nonincreasing, one has

$$\tilde{\delta}^2(\boldsymbol{x},\boldsymbol{y}) \le \int_0^1 \frac{1}{\tilde{\kappa}(\boldsymbol{\gamma}(\tau))} |\boldsymbol{\gamma}_\varepsilon'(\tau)|^2 \,\mathrm{d}\tau.$$

Thanks to (90), one obtains that

$$\tilde{\delta}^2(\boldsymbol{x},\boldsymbol{y}) < \int_0^1 \frac{1}{\tilde{\kappa}(\boldsymbol{\gamma}(\tau))} |\boldsymbol{\gamma}'(\tau)|^2 \,\mathrm{d}\tau,$$

providing a contradiction with the optimality (88) of $\boldsymbol{\gamma}$. Thus condition (87) holds; hence (Ω,δ) is a geodesic space. □

Appendix B: A multicomponent bathtub principle

The following theorem can be seen as a generalization of the classical scalar bathtub principle; see, for instance, [Lieb and Loss 2001, Theorem 1.14]. In what follows, N is a positive integer and Ω denotes an arbitrary measurable subset of $\mathbb{R}^d$.

Theorem B.1. *Let $\omega \in L^1_+(\Omega)$, and let $\boldsymbol{m} = (m_0,\dots,m_N) \in (\mathbb{R}^*_+)^{N+1}$ be such that $\sum_{i=0}^N m_i = \int_\Omega \omega \,\mathrm{d}\boldsymbol{x}$. We define*

$$\boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}} = \left\{ \boldsymbol{s} = (s_0,\dots,s_N) \in L^1_+(\Omega)^{N+1} \;\middle|\; \int_\Omega s_i \,\mathrm{d}\boldsymbol{x} = m_i \text{ and } \sum_{i=0}^N s_i = \omega \text{ a.e. in } \Omega \right\}.$$

Then for any $\boldsymbol{F} = (F_0,\dots,F_N) \in (L^\infty(\Omega))^{N+1}$, the functional

$$\mathcal{F} : \boldsymbol{s} \mapsto \int_\Omega \boldsymbol{F} \cdot \boldsymbol{s} \,\mathrm{d}\boldsymbol{x}$$

has a minimizer in $\boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$. Moreover, there exists $\boldsymbol{\alpha} = (\alpha_0,\dots,\alpha_N) \in \mathbb{R}^{N+1}$ such that, defining

$$\lambda(\boldsymbol{x}) := \min_{0 \le j \le N} \{F_j(\boldsymbol{x}) + \alpha_j\}, \qquad \boldsymbol{x} \in \Omega,$$

any minimizer $\underline{\boldsymbol{s}} = (\underline{s}_0,\dots,\underline{s}_N)$ satisfies

$$F_i + \alpha_i = \lambda \quad \mathrm{d}\underline{s}_i\text{-a.e. in } \Omega, \ \forall i \in \{0,\dots,N\}.$$

One can think of this as: $\underline{s}_i = 0$ in $\{F_i + \alpha_i > \lambda\}$ and $F_i + \alpha_i \ge \lambda$ everywhere; i.e., $\underline{s}_i > 0$ can only occur in the "contact set" $\{\boldsymbol{x} \mid F_i(\boldsymbol{x}) + \alpha_i = \min_j(F_j(\boldsymbol{x}) + \alpha_j)\}$.

Proof. For the existence part, note that $\mathcal{F}$ is continuous for the weak L^1 convergence, and that $\boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$ is weakly closed. Since $\sum s_i = \omega$ and $s_i \geq 0$, we have in particular $0 \leq s_i \leq \omega \in L^1$ for all i and $\boldsymbol{s} \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$. This implies that $\boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$ is uniformly integrable, and since the mass $\|s_i\|_{L^1} = \int s_i = m_i$ is prescribed, the Dunford–Pettis theorem shows that $\boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$ is L^1-weakly relatively compact. Hence from any minimizing sequence we can extract a weakly-L^1 converging subsequence, and by weak L^1 continuity the weak limit is a minimizer.

Let us now introduce a dual problem: for fixed $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_N) \in \mathbb{R}^{N+1}$ we set

$$\lambda_{\boldsymbol{\alpha}}(\boldsymbol{x}) := \min_i \{F_i(\boldsymbol{x}) + \alpha_i\} \tag{91}$$

and define

$$J(\boldsymbol{\alpha}) := \int_\Omega \lambda_{\boldsymbol{\alpha}}(\boldsymbol{x})\,\omega(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \sum_{i=0}^N \alpha_i m_i.$$

We shall prove below that

(i) $\sup_{\boldsymbol{\alpha} \in \mathbb{R}^{N+1}} J(\boldsymbol{\alpha}) = \max_{\boldsymbol{\alpha} \in \mathbb{R}^{N+1}} J(\boldsymbol{\alpha})$ is achieved,

(ii) $\min_{\boldsymbol{s} \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}} \mathcal{F}(\boldsymbol{s}) = \max_{\boldsymbol{\alpha} \in \mathbb{R}^{N+1}} J(\boldsymbol{\alpha})$.

The desired decomposition will then follow from equality conditions in (ii), and $\lambda(\boldsymbol{x}) = \lambda_{\bar{\boldsymbol{\alpha}}}(\boldsymbol{x})$ will be retrieved from any maximizer $\bar{\boldsymbol{\alpha}} \in \operatorname{Argmax} J$.

Remark B.2. The above dual problem can be guessed by introducing suitable Lagrange multipliers $\lambda(\boldsymbol{x}), \boldsymbol{\alpha}$ for the total saturation and mass constraints, respectively, and writing the convex indicator of the constraints as a supremum over these multipliers. Formally exchanging inf sup and sup inf and computing the optimality conditions in the rightmost infimum relates λ to $\boldsymbol{\alpha}$ as in (91), which in turn yields exactly the duality $\inf_{\boldsymbol{s}} \mathcal{F} = \max_{\boldsymbol{\alpha}} J$.

Let us first establish property (i). For all $\boldsymbol{\alpha} \in \mathbb{R}^{N+1}$ and all $\boldsymbol{s} \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$, we first observe that

$$\begin{aligned} J(\boldsymbol{\alpha}) &= \int_\Omega \min_j \{F_j(\boldsymbol{x}) + \alpha_j\}\,\omega(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \sum_{i=0}^N \alpha_i m_i \\ &= \int_\Omega \min_j \{F_j(\boldsymbol{x}) + \alpha_j\} \sum_{i=0}^N s_i(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \sum_{i=0}^N \alpha_i \int_\Omega s_i(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\ &= \sum_{i=0}^N \int_\Omega \big(\min_j \{F_j(\boldsymbol{x}) + \alpha_j\} - \alpha_i\big) s_i(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \leq \int_\Omega \boldsymbol{F} \cdot \boldsymbol{s}\,\mathrm{d}\boldsymbol{x} = \mathcal{F}(\boldsymbol{s}). \end{aligned}$$

In particular J is bounded from above and

$$\sup_{\boldsymbol{\alpha} \in \mathbb{R}^{N+1}} J(\boldsymbol{\alpha}) \leq \min_{\boldsymbol{s} \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}} \mathcal{F}(\boldsymbol{s}). \tag{92}$$

Since $\int \omega\,\mathrm{d}\boldsymbol{x} = \sum m_i$, the function J is invariant under diagonal shifts, i.e., $J(\boldsymbol{\alpha} + c\mathbf{1}) = J(\boldsymbol{\alpha})$ for any constant $c \in \mathbb{R}$. As a consequence we can choose a maximizing sequence $\{\boldsymbol{\alpha}^k\}_{k \geq 1}$ such that $\min_j \alpha_j^k = 0$

for all $k \geq 0$. Let $j(k)$ be an index such that $\alpha^k_{j(k)} = \min_j \alpha^k_j = 0$. Then, since $\boldsymbol{\alpha}^k$ is maximizing and $\omega(\boldsymbol{x}) \geq 0$, we get, for k large enough,

$$\begin{aligned}\sup J - 1 \leq J(\boldsymbol{\alpha}^k) &= \int_\Omega \min_j \{F_j(\boldsymbol{x}) + \alpha^k_j\}\omega(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \sum \alpha^k_i m_i \\ &\leq \int_\Omega \big(F_{j(k)}(\boldsymbol{x}) + \underbrace{\alpha^k_{j(k)}}_{=0}\big)\omega(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \sum \alpha^k_i m_i \leq \|\boldsymbol{F}\|_{L^\infty}\|\omega\|_{L^1} - \sum \alpha^k_i m_i.\end{aligned}$$

Thus $\sum \alpha^k_i m_i \leq C$, and since $\alpha^k_i \geq 0$ and $m_i > 0$ we deduce that $(\boldsymbol{\alpha}^k)_k$ is bounded. Hence, up to extraction of a nonrelabeled subsequence, we can assume that $\boldsymbol{\alpha}^k$ converges towards some $\bar{\boldsymbol{\alpha}} \in \mathbb{R}^{N+1}_+$. The map J is continuous; hence $\bar{\boldsymbol{\alpha}}$ is a maximizer.

Let us now focus on property (ii). Note from (92) and (i) it suffices to prove the reverse inequality

$$\max_{\boldsymbol{\alpha}\in\mathbb{R}^{N+1}} J(\boldsymbol{\alpha}) \geq \min_{\boldsymbol{s}\in\boldsymbol{\mathcal{X}}\cap\boldsymbol{\mathcal{A}}} \mathcal{F}(\boldsymbol{s}).$$

We show below that, for any maximizer $\bar{\boldsymbol{\alpha}}$ of J, we can always construct a suitable $\boldsymbol{s} \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$ such that $\mathcal{F}(\boldsymbol{s}) = J(\bar{\boldsymbol{\alpha}})$. This will immediately imply the reverse inequality and thus our claim (ii). In order to do so, we first observe that J is concave; thus the optimality condition at $\bar{\boldsymbol{\alpha}}$ can be written in terms of superdifferentials as $\mathbf{0}_{\mathbb{R}^{N+1}} \in \partial J(\bar{\boldsymbol{\alpha}})$. Denoting by

$$\Lambda(\boldsymbol{\alpha}) = \int_\Omega \lambda_{\boldsymbol{\alpha}}\,\omega\,\mathrm{d}\boldsymbol{x} = \int_\Omega \min_j \{F_j(x) + \alpha_j\}\omega(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

the first contribution in J, this optimality can be recast as

$$\boldsymbol{m} \in \partial\Lambda(\bar{\boldsymbol{\alpha}}). \tag{93}$$

For fixed $\boldsymbol{x} \in \Omega$ and by usual properties of the min function, the superdifferential $\partial\lambda_{\boldsymbol{\alpha}}(\boldsymbol{x})$ of the concave map $\boldsymbol{\alpha} \mapsto \lambda_{\boldsymbol{\alpha}}(\boldsymbol{x})$ at $\boldsymbol{\alpha} \in \mathbb{R}^{N+1}$ is characterized by

$$\partial\lambda_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \big\{\boldsymbol{\theta} \in \mathbb{R}^{N+1}_+ \;\big|\; \textstyle\sum_{i=0}^N \theta_i = 1 \text{ and } \theta_i = 0 \text{ if } F_i(\boldsymbol{x}) + \alpha_i > \lambda_{\boldsymbol{\alpha}}(\boldsymbol{x})\big\}.$$

Therefore, it follows from the extension of the formula of differentiation under the integral to the nonsmooth case, see [Clarke 1990, Theorem 2.7.2], that

$$\partial\Lambda(\boldsymbol{\alpha}) = \big\{\boldsymbol{w} \in \mathbb{R}^{N+1}_+ \;\big|\; \boldsymbol{w} = \textstyle\int_\Omega \boldsymbol{\theta}(\boldsymbol{x})\omega(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \text{ for some } \boldsymbol{\theta}(\boldsymbol{x}) \in \partial\lambda_{\boldsymbol{\alpha}}(\boldsymbol{x}) \text{ a.e. in } \Omega\big\}. \tag{94}$$

The optimality criterion (93) at any maximizer $\bar{\boldsymbol{\alpha}}$ gives the existence of some function $\boldsymbol{\theta}$ as in (94) such that

$$m_i = \int_\Omega \theta_i(\boldsymbol{x})\omega(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \quad \forall i \in \{0, \dots, N\}.$$

Defining

$$s_i(\boldsymbol{x}) := \theta_i(\boldsymbol{x})\omega(\boldsymbol{x}) \quad \forall i \in \{0, \dots, N\}, \tag{95}$$

we have by construction that $s_i \geq 0$, $\int s_i = m_i$, and $\sum s_i = \big(\sum_i \theta_i\big)\omega = \omega$ a.e.; thus $\boldsymbol{s} \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$. Exploiting again $\sum s_i = \omega$ as well as the crucial property that $\theta_i = 0$ a.e. in $\{\boldsymbol{x} \mid F_i + \bar{\alpha}_i > \lambda_{\bar{\boldsymbol{\alpha}}}\}$, or in other words that $F_i + \bar{\alpha}_i = \lambda_{\bar{\boldsymbol{\alpha}}}$ for $\mathrm{d}s_i$-a.e $\boldsymbol{x} \in \Omega$, we get

$$J(\bar{\boldsymbol{\alpha}}) = \int_\Omega \lambda_{\bar{\boldsymbol{\alpha}}}\, \omega \,\mathrm{d}\boldsymbol{x} - \sum_{i=0}^N \bar{\alpha}_i m_i = \sum_{i=0}^N \int_\Omega \lambda_{\bar{\boldsymbol{\alpha}}} s_i \,\mathrm{d}\boldsymbol{x} - \sum_{i=0}^N \bar{\alpha}_i m_i = \sum_{i=0}^N \int_\Omega (F_i + \bar{\alpha}_i) s_i \,\mathrm{d}\boldsymbol{x} - \sum_{i=0}^N \bar{\alpha}_i m_i = \mathcal{F}(\boldsymbol{s})$$

as claimed. Therefore $\boldsymbol{s}$ constructed by (95) is a minimizer of $\mathcal{F}$ and

$$J(\bar{\boldsymbol{\alpha}}) = \mathcal{F}(\underline{\boldsymbol{s}}). \tag{96}$$

In order to finally retrieve the desired decomposition, choose any minimizer $\underline{\boldsymbol{s}} \in \boldsymbol{\mathcal{X}} \cap \boldsymbol{\mathcal{A}}$ of $\mathcal{F}$ and any maximizer $\bar{\boldsymbol{\alpha}} \in \mathbb{R}^{N+1}$ of J. Then it follows from (96) that

$$0 = \mathcal{F}(\underline{\boldsymbol{s}}) - J(\bar{\boldsymbol{\alpha}}) = \sum_{i=0}^N \int_\Omega F_i \underline{s}_i \,\mathrm{d}\boldsymbol{x} - \int_\Omega \lambda_{\bar{\boldsymbol{\alpha}}}\, \omega \,\mathrm{d}\boldsymbol{x} + \sum_{i=0}^N \bar{\alpha}_i m_i .$$

Using once again that $\int \underline{s}_i = m_i$ and $\sum_i \underline{s}_i = \omega$, we get that

$$\sum_{i=0}^N \int_\Omega (F_i + \bar{\alpha}_i - \lambda_{\bar{\boldsymbol{\alpha}}}) \underline{s}_i \,\mathrm{d}\boldsymbol{x} = 0.$$

By the definition of $\lambda_{\bar{\boldsymbol{\alpha}}}$, the above integrand is nonnegative; hence $F_i + \bar{\alpha}_i = \lambda_{\bar{\boldsymbol{\alpha}}}$ a.e. in $\{\underline{s}_i > 0\}$. ☐

Acknowledgements

This project was supported by the ANR GEOPOR project (ANR-13-JS01-0007-01). Cancès also acknowledges the support of Labex CEMPI (ANR-11-LABX-0007-01). Gallouët was supported by the ANR ISOTACE project (ANR-12-MONU-0013) and by the Fonds de la Recherche Scientifique - FNRS under Grant MIS F.4539.16. Monsaingeon was supported by the Portuguese Science Foundation through FCT fellowship SFRH/BPD/88207/2012 and the UT Austin | Portugal CoLab project. Part of this work was carried out during the stay of Cancès and Gallouët at CAMGSD, Instituto Superior Técnico, Universidade de Lisboa. The authors wish to thank Quentin Mérigot for fruitful discussion.

References

[Alt and Luckhaus 1983] H. W. Alt and S. Luckhaus, "Quasilinear elliptic-parabolic differential equations", *Math. Z.* **183**:3 (1983), 311–341. MR Zbl

[Amaziane et al. 2012] B. Amaziane, M. Jurak, and A. Vrbaški, "Existence for a global pressure formulation of water-gas flow in porous media", *Electron. J. Differential Equations* **2012**:102 (2012), 1–22. MR Zbl

[Amaziane et al. 2014] B. Amaziane, M. Jurak, and A. Žgaljić Keko, "Modeling compositional compressible two-phase flow in porous media by the concept of the global pressure", *Comput. Geosci.* **18**:3-4 (2014), 297–309. MR

[Ambrosio and Gigli 2013] L. Ambrosio and N. Gigli, "A user's guide to optimal transport", pp. 1–155 in *Modelling and optimisation of flows on networks*, Lecture Notes in Math. **2062**, Springer, 2013. MR

[Ambrosio and Serfaty 2008] L. Ambrosio and S. Serfaty, "A gradient flow approach to an evolution problem arising in superconductivity", *Comm. Pure Appl. Math.* **61**:11 (2008), 1495–1539. MR Zbl

[Ambrosio et al. 2008] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, 2nd ed., Birkhäuser, Basel, 2008. MR Zbl

[Ambrosio et al. 2011] L. Ambrosio, E. Mainini, and S. Serfaty, "Gradient flow of the Chapman–Rubinstein–Schatzman model for signed vortices", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **28**:2 (2011), 217–246. MR Zbl

[Andreianov et al. 2015] B. Andreianov, C. Cancès, and A. Moussa, "A nonlinear time compactness result and applications to discretization of degenerate parabolic-elliptic PDEs", preprint, 2015, available at https://hal.archives-ouvertes.fr/hal-01142499/document.

[Antoncev and Monahov 1978] S. N. Antoncev and V. N. Monahov, "Three-dimensional problems of transient two-phase filtration in inhomogeneous anisotropic porous media", *Dokl. Akad. Nauk SSSR* **243**:3 (1978), 553–556. In Russian; translated in *Soviet Math., Dokl.* **19** (1978), 1354–1358. MR Zbl

[Bear and Bachmat 1990] J. Bear and Y. Bachmat, *Introduction to modeling of transport phenomena in porous media*, Springer, 1990. Zbl

[Blanchet 2013] A. Blanchet, "A gradient flow approach to the Keller–Segel systems", *RIMS Kôkyûroku* **1837** (2013), 52–73.

[Blanchet et al. 2008] A. Blanchet, V. Calvez, and J. A. Carrillo, "Convergence of the mass-transport steepest descent scheme for the subcritical Patlak–Keller–Segel model", *SIAM J. Numer. Anal.* **46**:2 (2008), 691–721. MR Zbl

[Bolley et al. 2013] F. Bolley, I. Gentil, and A. Guillin, "Uniform convergence to equilibrium for granular media", *Arch. Ration. Mech. Anal.* **208**:2 (2013), 429–445. MR Zbl

[Cancès and Gallouët 2011] C. Cancès and T. Gallouët, "On the time continuity of entropy solutions", *J. Evol. Equ.* **11**:1 (2011), 43–55. MR Zbl

[Cancès et al. 2015] C. Cancès, T. O. Gallouët, and L. Monsaingeon, "The gradient flow structure for incompressible immiscible two-phase flows in porous media", *C. R. Math. Acad. Sci. Paris* **353**:11 (2015), 985–989. MR Zbl

[Carlier and Laborde 2015] G. Carlier and M. Laborde, "On systems of continuity equations with nonlinear diffusion and nonlocal drifts", preprint, 2015. arXiv

[Carrillo et al. 2011] J. A. Carrillo, M. DiFrancesco, A. Figalli, T. Laurent, and D. Slepčev, "Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations", *Duke Math. J.* **156**:2 (2011), 229–271. MR Zbl

[Chavent 1976] G. Chavent, "A new formulation of diphasic incompressible flows in porous media", pp. 258–270 in *Applications of methods of functional analysis to problems in mechanics* (Marseille, 1975), Lecture Notes in Math. **503**, Springer, 1976. MR Zbl

[Chavent 2009] G. Chavent, "A fully equivalent global pressure formulation for three-phases compressible flows", *Appl. Anal.* **88**:10-11 (2009), 1527–1541. MR Zbl

[Chavent and Jaffré 1986] G. Chavent and J. Jaffré, *Mathematical models and finite elements for reservoir simulation*, Studies in Mathematics and its Applications **17**, North Holland, Amsterdam, 1986. Zbl

[Chavent and Salzano 1985] G. Chavent and G. Salzano, "Un algorithme pour la détermination de perméabilités relatives triphasiques satisfaisant une condition de différentielle totale", INRIA Technical Report 335, 1985, available at https://hal.inria.fr/inria-00076202v1.

[Chen 2001] Z. Chen, "Degenerate two-phase incompressible flow, I: Existence, uniqueness and regularity of a weak solution", *J. Differential Equations* **171**:2 (2001), 203–232. MR Zbl

[Clarke 1990] F. H. Clarke, *Optimization and nonsmooth analysis*, 2nd ed., Classics in Applied Mathematics **5**, SIAM, Philadelphia, PA, 1990. MR Zbl

[De Giorgi 1993] E. De Giorgi, "New problems on minimizing movements", pp. 81–98 in *Boundary value problems for partial differential equations and applications*, edited by J.-L. Lions and C. Baiocchi, RMA Res. Notes Appl. Math. **29**, Masson, Paris, 1993. MR Zbl

[Dolbeault et al. 2009] J. Dolbeault, B. Nazaret, and G. Savaré, "A new class of transport distances between measures", *Calc. Var. Partial Differential Equations* **34**:2 (2009), 193–231. MR Zbl

[Fabrie and Saad 1993] P. Fabrie and M. Saad, "Existence de solutions faibles pour un modèle d'écoulement triphasique en milieu poreux", *Ann. Fac. Sci. Toulouse Math.* (6) **2**:3 (1993), 337–373. MR Zbl

[Gagneux and Madaune-Tort 1996] G. Gagneux and M. Madaune-Tort, *Analyse mathématique de modèles non linéaires de l'ingénierie pétrolière*, Mathématiques & Applications (Berlin) **22**, Springer, 1996. MR Zbl

[Gigli and Otto 2013] N. Gigli and F. Otto, "Entropic Burgers' equation via a minimizing movement scheme based on the Wasserstein metric", *Calc. Var. Partial Differential Equations* **47**:1-2 (2013), 181–206. MR Zbl

[Hanche-Olsen and Holden 2010] H. Hanche-Olsen and H. Holden, "The Kolmogorov–Riesz compactness theorem", *Expo. Math.* **28**:4 (2010), 385–394. MR Zbl

[Jordan et al. 1998] R. Jordan, D. Kinderlehrer, and F. Otto, "The variational formulation of the Fokker–Planck equation", *SIAM J. Math. Anal.* **29**:1 (1998), 1–17. MR Zbl

[Kinderlehrer et al. 2017] D. Kinderlehrer, L. Monsaingeon, and X. Xu, "A Wasserstein gradient flow approach to Poisson–Nernst–Planck equations", *ESAIM Control Optim. Calc. Var.* **23**:1 (2017), 137–164. MR Zbl

[Laborde 2016] M. Laborde, *Systèmes de particules en interaction, approche par flot de gradient dans l'espace de Wasserstein*, Ph.D. thesis, Université Paris-Dauphine, 2016, available at https://basepub.dauphine.fr/handle/123456789/16518.

[Ladyženskaja et al. 1968] O. A. Ladyženskaja, V. A. Solonnikov, and N. N. Ural'ceva, *Linear and quasilinear equations of parabolic type*, Translations of Mathematical Monographs **23**, Amer. Math. Soc., Providence, RI, 1968. MR Zbl

[Laurençot and Matioc 2013] P. Laurençot and B.-V. Matioc, "A gradient flow approach to a thin film approximation of the Muskat problem", *Calc. Var. Partial Differential Equations* **47**:1-2 (2013), 319–341. MR Zbl

[Lieb and Loss 2001] E. H. Lieb and M. Loss, *Analysis*, 2nd ed., Graduate Studies in Mathematics **14**, Amer. Math. Soc., Providence, RI, 2001. MR Zbl

[Lisini 2009] S. Lisini, "Nonlinear diffusion equations with variable coefficients as gradient flows in Wasserstein spaces", *ESAIM Control Optim. Calc. Var.* **15**:3 (2009), 712–740. MR Zbl

[Lisini et al. 2012] S. Lisini, D. Matthes, and G. Savaré, "Cahn–Hilliard and thin film equations with nonlinear mobility as gradient flows in weighted-Wasserstein metrics", *J. Differential Equations* **253**:2 (2012), 814–850. MR Zbl

[Matthes et al. 2009] D. Matthes, R. J. McCann, and G. Savaré, "A family of nonlinear fourth order equations of gradient flow type", *Comm. Partial Differential Equations* **34**:10-12 (2009), 1352–1397. MR Zbl

[Maury et al. 2010] B. Maury, A. Roudneff-Chupin, and F. Santambrogio, "A macroscopic crowd motion model of gradient flow type", *Math. Models Methods Appl. Sci.* **20**:10 (2010), 1787–1821. MR Zbl

[Moussa 2016] A. Moussa, "Some variants of the classical Aubin–Lions lemma", *J. Evol. Equ.* **16**:1 (2016), 65–93. MR Zbl

[Otto 1998] F. Otto, "Dynamics of labyrinthine pattern formation in magnetic fluids: a mean-field theory", *Arch. Rational Mech. Anal.* **141**:1 (1998), 63–103. MR Zbl

[Otto 2001] F. Otto, "The geometry of dissipative evolution equations: the porous medium equation", *Comm. Partial Differential Equations* **26**:1-2 (2001), 101–174. MR Zbl

[Sandier and Serfaty 2004] E. Sandier and S. Serfaty, "Gamma-convergence of gradient flows with applications to Ginzburg–Landau", *Comm. Pure Appl. Math.* **57**:12 (2004), 1627–1672. MR Zbl

[Santambrogio 2015] F. Santambrogio, *Optimal transport for applied mathematicians; calculus of variations, PDEs, and modeling*, Progress in Nonlinear Differential Equations and their Applications **87**, Springer, 2015. MR Zbl

[Villani 2009] C. Villani, *Optimal transport: old and new*, Grundlehren der Mathematischen Wissenschaften **338**, Springer, 2009. MR Zbl

[Zinsl 2014] J. Zinsl, "Existence of solutions for a nonlinear system of parabolic equations with gradient flow structure", *Monatsh. Math.* **174**:4 (2014), 653–679. MR Zbl

[Zinsl and Matthes 2015a] J. Zinsl and D. Matthes, "Exponential convergence to equilibrium in a coupled gradient flow system modeling chemotaxis", *Anal. PDE* **8**:2 (2015), 425–466. MR Zbl

[Zinsl and Matthes 2015b] J. Zinsl and D. Matthes, "Transport distances and geodesic convexity for systems of degenerate diffusion equations", *Calc. Var. Partial Differential Equations* **54**:4 (2015), 3397–3438. MR Zbl

Received 13 Jul 2016. Revised 23 May 2017. Accepted 29 Jun 2017.

CLÉMENT CANCÈS: clement.cances@inria.fr
Team RAPSODI, Inria Lille – Nord Europe, Lille, France

THOMAS O. GALLOUËT: thomas.gallouet@ulg.ac.be
Département de Mathématiques, Université de Liège, Liège, Belgium

LÉONARD MONSAINGEON: leonard.monsaingeon@univ-lorraine.fr
Institut Élie Cartan de Lorraine, Université de Lorraine, Nancy, France

mathematical sciences publishers msp

ANALYSIS AND PDE
Vol. 10, No. 8, 2017
dx.doi.org/10.2140/apde.2017.10.1877

RESONANCES FOR SYMMETRIC TENSORS ON ASYMPTOTICALLY HYPERBOLIC SPACES

CHARLES HADFIELD

On manifolds with an even Riemannian conformally compact Einstein metric, the resolvent of the Lichnerowicz Laplacian, acting on trace-free, divergence-free, symmetric 2-tensors is shown to have a meromorphic continuation to the complex plane, defining quantum resonances of this Laplacian. For higher-rank symmetric tensors, a similar result is proven for (convex cocompact) quotients of hyperbolic space.

1. Introduction

This paper studies the meromorphic extension of the resolvent of the Laplacian acting on symmetric tensors above asymptotically hyperbolic manifolds. The geometric setting of asymptotically hyperbolic manifolds, modelled on convex cocompact quotients of hyperbolic space, dates back to [Mazzeo 1988; Mazzeo and Melrose 1987; Fefferman and Graham 1985]. The meromorphic extension with finite-rank poles of the resolvent of the Laplacian on functions is obtained in [Mazzeo and Melrose 1987], excluding certain exceptional points in $\mathbb{C}$. Refining the definition of asymptotically hyperbolic manifolds by introducing a notion of evenness, Guillarmou [2005] provided the meromorphic extension to all of $\mathbb{C}$ and showed that for such an extension, said evenness is essential; see also [Guillopé and Zworski 1995]. By shifting his viewpoint and studying a Fredholm problem, rather than using Melrose's pseudodifferential calculus on manifolds with corners, Vasy [2013a; 2013b] was also able to recover the result of [Guillarmou 2005]. This technique is presented in a very accessible article of Zworski [2016] in a microlocal language (nonsemiclassical). This alternative method is more appropriate when one considers vector bundles, and, for symmetric tensors, is lightly explained later in this introduction. Effectively contained in [Vasy 2013a], the meromorphic extension is explicitly obtained in [Vasy 2017] for the resolvent of the Hodge Laplacian upon restriction to coclosed forms (or excluding top forms, for closed forms). Such a restriction is natural in light of works in a conformal setting [Aubry and Guillarmou 2011; Branson and Gover 2005], i.e., the boundary of the asymptotic space. In fact, from the conformal geometry viewpoint, Vasy's method of placing the asymptotically hyperbolic manifold in an ambient manifold equipped with a Lorentzian metric is very much in the spirit of both the tractor calculus [Bailey et al. 1994], as well as the ambient metric construction [Fefferman and Graham 2012].

We give the theorems (with precise definitions of the objects involved left to the body of the article) and sketch their proofs. Let $\overline{X}$ be a compact manifold with boundary $Y = \partial\overline{X}$. That (X, g) is asymptotically

MSC2010: primary 35P25; secondary 35Q75, 53B21.
Keywords: quantum resonances, asymptotically hyperbolic, meromorphic extension of resolvent.

hyperbolic means that, locally near Y in $\overline{X}$, there exists a chart $[0,\varepsilon)_\rho \times Y$ such that on $(0,\varepsilon)\times Y$, the metric g takes the form

$$g = \frac{d\rho^2 + h}{\rho^2},$$

where h is a family of Riemannian metrics on Y, depending smoothly on $\rho \in [0,\varepsilon)$. That g is even means that h has a Taylor series about $\rho = 0$ in which only even powers of ρ appear. Above X, we consider the set of symmetric cotensors of rank m, denoting this vector bundle by $\mathcal{E}^{(m)} = \operatorname{Sym}^m \mathrm{T}^*X$. On symmetric tensors, there exist two common Laplacians. The (positive) rough Laplacian $\nabla^*\nabla$ and the Lichnerowicz Laplacian Δ, originally defined on 2-cotensors [Lichnerowicz 1961], but easily extendible to arbitrary degree [Heil et al. 2016]. On functions, these two Laplacians coincide; on one-forms, the Lichnerowicz Laplacian agrees with the Hodge Laplacian; and in general, for symmetric m-cotensors, the Lichnerowicz Laplacian differs from the rough Laplacian by a zeroth-order curvature operator

$$\Delta = \nabla^*\nabla + q(\mathrm{R}).$$

We construct the Lorentzian cone $M = \mathbb{R}^+_s \times X$ with metric

$$\eta = -ds \otimes ds + s^2 g$$

(and call s the Lorentzian scale). Pulling $\mathcal{E}^{(m)}$ back to M we naturally see $\mathcal{E}^{(m)}$ as a subbundle of the bundle of all symmetric cotensors of rank m above M; this larger bundle is denoted by $\mathcal{F} = \operatorname{Sym}^m \mathrm{T}^*M$. On $\mathcal{F}$ we consider the Lichnerowicz d'Alembertian $\Box$. Up to symmetric powers of ds/s we may identify $\mathcal{F}$ with the direct sum of $\mathcal{E}^{(k)} = \operatorname{Sym}^k \mathrm{T}^*X$ for all $k \le m$. Indeed by denoting by $\mathcal{E} = \bigoplus_{k=0}^m \mathcal{E}^{(k)}$ the bundle of all symmetric tensors above X of rank not greater than m, we are able to pull back sections of this bundle and see them as sections of $\mathcal{F}$:

$$\pi_s^* : C^\infty(X;\mathcal{E}) \to C^\infty(M;\mathcal{F}).$$

A long calculation gives the structure of the Lichnerowicz d'Alembertian with respect to this identification. It is seen that $s^2\Box$ decomposes as the Lichnerowicz Laplacian Δ acting on each subbundle of $\mathcal{E}^{(k)}$ for $0 \le k \le m$; however, these fibres are coupled via off-diagonal terms consisting of the symmetric differential d and its adjoint, the divergence δ. (There are also less important couplings due to the trace Λ and its adjoint L.) Also present in the diagonal are terms involving $s\,\partial_s$ and $(s\,\partial_s)^2$. By conjugating by $s^{-n/2+m}$ we obtain the operator

$$\boldsymbol{Q} = \nabla^*\nabla + (s\,\partial_s)^2 + \boldsymbol{D} + \boldsymbol{G},$$

where $\boldsymbol{D}$ is of first order consisting of the symmetric differential and the divergence, while $\boldsymbol{G}$ is a smooth endomorphism on $\mathcal{F}$. By appealing to the b-calculus of Melrose [1993], we can push this operator acting on $\mathcal{F}$ above M to a family of operators (holomorphic in the complex variable λ) acting on $\mathcal{E}$ above X of the form

$$\mathcal{Q}_\lambda = \nabla^*\nabla + \lambda^2 + \mathcal{D} + \mathcal{G},$$

where $\mathcal{D}$ is of first order consisting of the symmetric differential and the divergence, while $\mathcal{G}$ is a smooth endomorphism on $\mathcal{E}$. Explicitly, in matrix notation writing

$$u = \begin{bmatrix} u^{(m)} \\ \vdots \\ u^{(0)} \end{bmatrix}, \quad u \in C^\infty(X; \mathcal{E}),\ u^{(k)} \in C^\infty(X; \mathcal{E}^{(k)}),$$

the operator $\mathcal{Q}_\lambda$ takes the form

$$\begin{bmatrix}
\Delta+\lambda^2-c_m-\mathrm{L}\Lambda & 2b_{m-1}\mathrm{d} & -b_{m-2}b_{m-1}\mathrm{L} & & & \mathbf{0} \\
-2b_{m-1}\delta & \ddots & \ddots & \ddots & & \\
-b_{m-2}b_{m-1}\Lambda & \ddots & \ddots & \ddots & & -b_0b_1\mathrm{L} \\
 & \ddots & \ddots & \ddots & & 2b_0\mathrm{d} \\
\mathbf{0} & & & -b_0b_1\Lambda & -2b_0\delta & \Delta+\lambda^2-c_0-\mathrm{L}\Lambda
\end{bmatrix}$$

for constants

$$b_k = \sqrt{m-k}, \quad c_k = \tfrac{1}{4}n^2 + m(n+2k+1) - k(2n+3k-1),$$

and operators Δ the Lichnerowicz Laplacian, δ the divergence, d the symmetric differential, Λ the trace, and L the adjoint of the trace. (The operator $\mathcal{Q}_\lambda$ naively does not appear self-adjoint for $\lambda \in i\mathbb{R}$ since δ is the adjoint of d. The sign discrepancy is due to the Lorentzian signature of η. The operator is indeed self-adjoint for $\lambda \in i\mathbb{R}$ as detailed in Proposition 5.13.) When this family of operators acts on L^2 sections, denoted by $L^2_s(X; \mathcal{E})$ described in (5), it has an inverse for $\operatorname{Re}\lambda \gg 1$. This family of operators has the following meromorphic family of inverses.

Theorem 1.1. *Let (X^{n+1}, g) be even asymptotically hyperbolic. Then the inverse of (Definition 5.11)*

$$\mathcal{Q}_\lambda \text{ acting on } L^2_s(X; \mathcal{E}),$$

written as $\mathcal{Q}_\lambda^{-1}$, has a meromorphic continuation from $\operatorname{Re}\lambda \gg 1$ *to* $\mathbb{C}$,

$$\mathcal{Q}_\lambda^{-1} : C_c^\infty(X; \mathcal{E}) \to \rho^{\lambda+n/2-m} \bigoplus_{k=0}^{m} \rho^{-2k} C^\infty_{\text{even}}(\bar{X}; \mathcal{E}^{(k)})$$

with finite-rank poles.

Consider $u \in C^\infty(X; \mathcal{E})$. Although the trace operator Λ acting on each subbundle $\mathcal{E}^{(k)}$ gives a notion of u being trace-free, it is more natural to consider the ambient trace operator from $\mathcal{F}$, denoted by Λ_η (Section 3B). Pulling u back to M, we have $\pi_s^* u \in C^\infty(M; \mathcal{F})$ and we may consider the condition that $\pi_s^* u \in \ker \Lambda_\eta$. Avoiding extra notation for this subbundle of $\mathcal{E}$ (consisting of symmetric tensors above X which are trace-free with respect to the ambient trace operator Λ_η) we will simply refer to its sections using the notation

$$C^\infty(X; \mathcal{E}) \cap \ker(\Lambda_\eta \circ \pi_s^*).$$

On this subbundle, the operator $\mathcal{Q}_\lambda$ takes the form

$$\begin{bmatrix} \Delta+\lambda^2-c'_m & 2b_{m-1}\mathrm{d} & & \mathbf{0} \\ -2b_{m-1}\delta & \ddots & \ddots & \\ & \ddots & \ddots & 2b_0\mathrm{d} \\ \mathbf{0} & & -2b_0\delta & \Delta+\lambda^2-c'_0 \end{bmatrix}$$

with the modified constants

$$c'_k = c_k - (m-k)(m-k-1).$$

Note that if $u = u^{(m)} \in C^\infty(X; \mathcal{E}^{(m)})$ then $u \in \ker \Lambda$ if and only if $\pi_s^* u \in \ker \Lambda_\eta$. Again, a similar meromorphic extension of the inverse may be obtained.

Theorem 1.2. *Let (X^{n+1}, g) be even asymptotically hyperbolic. Then the inverse of (Definition 5.11)*

$$\mathcal{Q}_\lambda \text{ acting on } L_s^2(X; \mathcal{E}) \cap \ker(\Lambda_\eta \circ \pi_s^*),$$

written as $\mathcal{Q}_\lambda^{-1}$, has a meromorphic continuation from $\operatorname{Re}\lambda \gg 1$ to $\mathbb{C}$,

$$\mathcal{Q}_\lambda^{-1} : C_c^\infty(X; \mathcal{E}) \cap \ker(\Lambda_\eta \circ \pi_s^*) \to \rho^{\lambda+n/2-m} \left(\bigoplus_{k=0}^{m} \rho^{-2k} C_{\mathrm{even}}^\infty(\bar{X}; \mathcal{E}^{(k)}) \right) \cap \ker(\Lambda_\eta \circ \pi_s^*)$$

with finite-rank poles.

In order to uncouple the Lichnerowicz Laplacian acting on $\mathcal{E}^{(m)}$ and obtain the desired meromorphic extension of the resolvent, we need to restrict further from simply trace-free tensors to trace-free, divergence-free tensors. Equivalently, we must be able to commute the Lichnerowicz Laplacian with both the trace operator and the divergence operator. The first commutation is always possible giving the preceding structure of $\mathcal{Q}_\lambda$; however, unlike in the setting of differential forms (where the Hodge Laplacian always commutes with the divergence), such a commutation on symmetric tensors depends on the geometry of (X, g). For $m = 2$ the condition is that the Ricci tensor be parallel, while for $m \geq 3$, the manifold must be locally isomorphic to hyperbolic space.

Theorem 1.3. *Let (X^{n+1}, g) be even asymptotically hyperbolic and Einstein. Then the inverse of*

$$\Delta - \tfrac{1}{4}n(n-8) + \lambda^2 \text{ acting on } L^2(X; \mathcal{E}^{(2)}) \cap \ker\Lambda \cap \ker\delta,$$

written as $\mathcal{R}_\lambda$, has a meromorphic continuation from $\operatorname{Re}\lambda \gg 1$ to $\mathbb{C}$,

$$\mathcal{R}_\lambda : C_c^\infty(X; \mathcal{E}^{(2)}) \cap \ker\Lambda \cap \ker\delta \to \rho^{\lambda+n/2-2} C_{\mathrm{even}}^\infty(\bar{X}; \mathcal{E}^{(2)}) \cap \ker\Lambda \cap \ker\delta$$

with finite-rank poles.

Theorem 1.4. *Let (X^{n+1}, g) be a convex cocompact quotient of $\mathbb{H}^{n+1}$. Then the inverse of*

$$\Delta - \tfrac{1}{4}(n^2 - 4m(n+m-2)) + \lambda^2 \text{ acting on } L^2(X; \mathcal{E}^{(m)}) \cap \ker \Lambda \cap \ker \delta,$$

written as $\mathcal{R}_\lambda$, has a meromorphic continuation from $\operatorname{Re}\lambda \gg 1$ *to* $\mathbb{C}$,

$$\mathcal{R}_\lambda : C_c^\infty(X; \mathcal{E}^{(m)}) \cap \ker \Lambda \cap \ker \delta \to \rho^{\lambda+n/2-m} C^\infty_{\text{even}}(\bar{X}; \mathcal{E}^{(m)}) \cap \ker \Lambda \cap \ker \delta$$

with finite-rank poles.

Note that on $\mathbb{H}^{n+1}$, the difference between the Lichnerowicz Laplacian and the rough Laplacian is $q(\mathrm{R}) = -m(n+m-1)$. Thus by introducing a spectral parameter $s = \lambda + \frac{1}{2}n$ (not to be confused with the Lorentzian scale), the previous operator $\Delta - c_m + \lambda^2$ may be equivalently written as

$$\nabla^*\nabla - s(n-s) - m$$

in the spirit of [Dyatlov et al. 2015].

In order to demonstrate Theorem 1.1, Vasy's technique is to consider a slightly larger manifold X_e, as well as the ambient space $M_e = \mathbb{R}^+ \times X_e$. Using two key tricks near the boundary $Y = \partial \bar{X}$: the evenness property allows us to introduce the coordinate $\mu = \rho^2$ and twisting the Lorentzian scale with the boundary-defining function gives (what is termed the Euclidean scale) $t = s/\rho$, it is seen that the ambient metric η may be extended nondegenerately past $\mathbb{R}^+ \times Y$ to M_e. On $\operatorname{Sym}^m \mathrm{T}^* M_e$ we construct, analogous to $\boldsymbol{Q}$, an operator $\boldsymbol{P}$ replacing appearances of s by t which, on M, is easily related to $\boldsymbol{Q}$. Again the b-calculus provides a family of operators $\mathcal{P}$ on $\bigoplus_{k=0}^m \operatorname{Sym}^k \mathrm{T}^* X_e$ above X_e. Section 7 shows precisely how this family of operators fits into a Fredholm framework giving a meromorphic inverse, and very quickly also provides Theorem 1.1.

Such theorems are desirable for several reasons. Firstly, the quantum/classical correspondence between the spectrum of the Laplacian on a closed hyperbolic surface and Ruelle resonances of the generator of the geodesic flow on the unit tangent bundle [Faure and Tsujii 2013, Proposition 4.1] has been extended to compact hyperbolic manifolds of arbitrary dimension [Dyatlov et al. 2015], at which point the correspondence is between Ruelle resonances and the spectrum of the Laplacian acting on trace-free, divergence-free, symmetric tensors of arbitrary rank. This correspondence is extended in [Guillarmou et al. 2016] to convex cocompact hyperbolic surfaces using the scattering operator [Graham and Zworski 2003], as well as [Dyatlov and Guillarmou 2016], to obtain Ruelle resonances in this open system. Theorem 1.4 has been applied, along with results from [Dyatlov et al. 2015; Dyatlov and Guillarmou 2016], in order to provide such a correspondence in the setting of convex cocompact hyperbolic manifolds of arbitrary dimension [Hadfield 2017]. Secondly, with knowledge of the asymptotics of the resolvent of the Laplacian on functions, it is possible to construct the Poisson operator, the scattering operator, and study in a conformal setting, the GJMS operators and the Q-curvature of Branson [Djadli et al. 2008, Chapters 5–6]. This problem should be particularly interesting on symmetric 2-cotensors above a conformal manifold which, upon extension to a "bulk" Poincaré–Einstein manifold, makes contact with Theorem 1.3. Finally, and again with respect to Theorem 1.3, the Lichnerowicz Laplacian plays a fundamental role in problems involving deformations of metrics and their Ricci tensors [Biquard 2000;

Delay 1999; Graham and Lee 1991], as well as to linearised gravity [Wang 2009]. Spectral analysis of the Lichnerowicz Laplacian [Delay 2002; 2007], as well as the desire to build a scattering operator, emphasise the importance of considering this Laplacian acting on more general spaces than that of L^2 sections. From the viewpoint of gravitational waves, the recent work [Baskin et al. 2015] studies decay rates of solutions to the wave equation (acting on the trivial bundle) on Minkowski space with metrics similar to (1). It is very natural to consider this problem on symmetric 2-cotensors acted upon by the Lichnerowicz d'Alembertian.

Theorem 1.3 requires the global condition that the manifold be Einstein. It is unclear whether such a condition is necessary. Vasy's technique deals with the condition of even asymptotic hyperbolicity near the boundary. Indeed, this is reflected in Theorem 1.2. However to obtain our desired result, uncoupling the Lichnerowicz Laplacian from the operator $\mathcal{Q}$ currently requires a global condition on the base manifold. One should study whether perturbation techniques could provide a more general theorem, giving precise conditions for when such a meromorphic continuation exists.

The paper is structured as follows. Section 2 sets up the geometric side of the problem, introducing the various manifolds of interest as well as the construction of the ambient metric η. This section also includes a digression into the model geometry $X = \mathbb{H}^{n+1}$ to motivate Vasy's construction. Section 3 introduces the algebraic aspects of symmetric tensors, introduces many notational conventions and establishes several relationships between symmetric tensors when working relative to the Lorentzian and Euclidean scales. Section 4 recalls standard notions from microlocal analysis and gives several notions from the b-calculus framework adapted to vector bundles. Section 5 contains the bulk of the calculations of this paper, relating $\boldsymbol{Q}$ and $\mathcal{Q}$ with the Lichnerowicz Laplacian. Sections 6 and 7 introduce the operators $\boldsymbol{P}$ and $\mathcal{P}$ and provide the desired meromorphic inverse. Section 8 establishes the four theorems. Section 9 details the particular case of symmetric cotensors of rank $m = 2$. It is useful to gain insight into this problem via this low-rank setting, and it is hoped that the presentation of this case will aid the reader particularly during Sections 5 and 8. Finally, Section 10 gives the high energy estimates one would obtain if the microlocal analysis performed in Section 7 was performed using semiclassical notions.

2. Geometry

2A. ***Model geometry.*** It is worth mentioning the model geometry which provides a clear geometric motivation for the construction of the ambient space, as well as the Minkowski and Euclidean scales.

Let $\mathbb{R}^{1,n+1}$ be Minkowski space with the Lorentzian metric

$$\eta := -dx_0^2 + \sum_{i=1}^{n+1} dx_i^2$$

and set M_e to be Minkowski space minus the closure of the backward light cone. The metric gives the Minkowski distance function, denoted by η^2, on $\mathbb{R}^{1,n+1}$ from the origin:

$$\eta^2(x) := -x_0^2 + \sum_{i=1}^{n+1} x_i^2.$$

Hyperbolic space $X = \mathbb{H}^{n+1}$ is then identified with the (connected) hypersurface

$$X := \{x \in \mathbb{R}^{1,n+1} \mid \eta^2(x) = -1,\ x_0 > 0\}$$

and is given the metric g induced by the restriction of η. The boundary at infinity of hyperbolic space, i.e., the sphere $Y = \mathbb{S}^n$, is identified with the (connected) submanifold

$$Y := \{x \in \mathbb{R}^{1,n+1} \mid \eta^2(x) = 0,\ x_0 = 1\},$$

which, as an aside, inherits the standard metric, denoted by h, by restriction of η. For completeness we introduce de Sitter space dS^{n+1} as the hypersurface

$$dS^{n+1} := \{x \in \mathbb{R}^{1,n+1} \mid \eta^2(x) = 1\}.$$

We define the forward light cone

$$M := \{x \in \mathbb{R}^{1,n+1} \mid \eta^2(x) < 0,\ x_0 > 0\}$$

and note the decomposition $M = \mathbb{R}_s^+ \times X$ via the identification

$$\mathbb{R}_s^+ \times X \ni (s, x) \mapsto s \cdot x \in X.$$

In these coordinates, the metric η restricted to M takes the form

$$\eta = -ds \otimes ds + s^2 g$$

and we refer to s as the Minkowski scale. We define X_e to be the subset of the $(n+1)$-sphere contained in M_e

$$X_e := \left\{ x \in \mathbb{R}^{1,n+1} \;\middle|\; \sum_{i=0}^{n+1} x_i^2 = 1, x_0 > \frac{-1}{\sqrt{2}} \right\}$$

and note that the ambient space M_e is diffeomorphic to $\mathbb{R}_t^+ \times X_e$ via the identification

$$\mathbb{R}_t^+ \times X_e \ni (t, x) \mapsto t \cdot x \in M_e.$$

We refer to t as the Euclidean scale. The dilations induced by the Euclidean scale allow the identification

$$X_e \simeq X \sqcup Y \sqcup dS^{n+1}.$$

2B. *General setting.* We now properly introduce the geometric setting of the article. Let (X, g) be a Riemannian manifold of dimension $n+1$ which is even asymptotically hyperbolic [Guillarmou 2005, Definition 1.2] with boundary at infinity denoted by Y. We recall the definition of evenness.

Definition 2.1. Let (X, g) be an asymptotically hyperbolic manifold. We say that g is even if there exists a boundary-defining function ρ and a family of tensors $(h_{2i})_{i\in\mathbb{N}_0}$ on $Y = \partial\overline{X}$ such that, for all N, one has the following decomposition of g near Y:

$$\phi^*(\rho^2 g) = dr^2 + \sum_{i=0}^{N} h_{2i} r^{2i} + O(r^{2N+2}),$$

where ϕ is the diffeomorphism induced by the flow ϕ_r of the gradient $\operatorname{grad}_{\rho^2 g}(\rho)$:

$$\begin{aligned}\phi : [0,1) \times Y &\to \phi([0,1)\times Y) \subset \bar{X},\\ (r, y) &\mapsto \phi_r(y).\end{aligned}$$

We define $X^2 := (\bar{X} \sqcup \bar{X})/Y$ to be the topological double of $\bar{X}$. (For a slicker definition, we stray ever so slightly from the model geometry.) From the diffeomorphism ϕ we initially construct a C^∞ atlas on X^2 by noting that $Y \subset X^2$ is contained in an open set $U^2 := (U_- \sqcup U_+)/Y$ with $U_\pm := \phi([0,1)\times Y)$ and we declare this set to be C^∞ diffeomorphic to $(-1,1)\times Y$ via

$$\begin{aligned}(-1,1)\times Y &\simeq U^2,\\ (t, y) &\mapsto \begin{cases}\phi_{-t}(y) \in U_- & \text{if } t \le 0,\\ \phi_{+t}(y) \in U_+ & \text{if } t \ge 0.\end{cases}\end{aligned}$$

Charts on the interior of X in $\bar{X}$ complete the atlas on X^2.

We want to consider the boundary-defining function ρ as a function from X^2 to $[-1,1]$ such that X may be identified with $\{\rho > 0\}$. Using the previous chart for $U^2 \simeq (-1,1)\times Y$ we initially set

$$\begin{aligned}\rho : (-1,1)\times Y &\to (-1,1),\\ (r, y) &\mapsto r,\end{aligned}$$

and extend ρ to a continuous function on X^2 by demanding that ρ be constant on $X^2 \backslash U^2$. In order to ensure smoothness at $\partial \bar{U}^2$ we deform ρ smoothly on the two subsets $(-1, -1+\varepsilon)\times Y$ and $(1-\varepsilon, 1)\times Y$ of U^2. This achieves our goal. We now define the function μ on X^2 by declaring

$$\mu : X^2 \to [-1,1], \qquad \mu = \begin{cases}-\rho^2 & \text{if } \rho \le 0,\\ \rho^2 & \text{if } \rho \ge 0.\end{cases}$$

Remark 2.2. Although we have performed a deformation of ρ near $\partial \bar{U}^2$ we will continue to think of ρ and μ as coordinates for the first factor of $U^2 = (-1,1)\times Y$ (if we wanted to be correct, in what follows we would replace $(-1,1)$ with $(-1+\varepsilon, 1-\varepsilon)$ but this is cumbersome and we prefer to free up the variable ε). Of course, only the coordinates (μ, y) provide a smooth chart for X^2 near Y.

We now weaken the atlas on X^2 near Y. By the previous remark, we may think of μ as coordinates for the first factor of U^2 and we thus demand that the C^∞ atlas is with respect to this coordinate rather than ρ (as was the case for the initial atlas). It is now the case that on X^2, only μ (and not ρ) is a smooth function.

We define the set $C^\infty_{\text{even}}(\bar{X})$ to be the subset of functions in $C^\infty(X)$ which are extensible to $C^\infty(X^2)$ and whose extension is invariant with respect to the natural involution on X^2. (For example, consider the restriction of μ to X. However, such an invariant extension would of course not give the function μ previously constructed due to a sign discrepancy.) We remark that $\dot{C}^\infty(X)$, the subset of functions in $C^\infty(\bar{X})$ which vanish to all orders at Y, injects naturally into $C^\infty(X^2)$ and may be identified with the subset of $C^\infty(X^2)$ whose elements vanish on $\{\rho < 0\}$. Such constructions may also readily be extended to the setting of vector bundles above X by using a local basis near Y of such a vector bundle which smoothly extends across Y.

Definition 2.3. We denote by X_e the extension of X

$$X_e := \{\mu > -1\} \subset X^2$$

by S the hypersurface $\{\mu = -\frac{1}{2}\} \subset X_e$, and by X_{cs} the open submanifold $\{\mu > -\frac{1}{2}\} \subset X_e$ such that $\partial \overline{X}_{cs} = S$.

We construct two product manifolds $M := \mathbb{R}^+_s \times X$ and $M_e := \mathbb{R}^+_t \times X_e$. We supply M with the Lorentzian cone metric

$$\eta := -ds \otimes ds + s^2 g$$

and explain how this structure may be smoothly extended to M_e.

Using the even neighbourhood at infinity $U := (0, 1)_\mu \times Y$, we remark that, on $\mathbb{R}^+_s \times U$, the Lorentzian metric takes the form

$$\eta = -ds \otimes ds + s^2\left(\frac{d\mu \otimes d\mu}{4\mu^2} + \frac{h}{\mu}\right),$$

where h has a smooth Taylor expansion about $\mu = 0$ by the evenness hypothesis. Upon the change of variables $t = s/\rho$ with $t \in \mathbb{R}^+$, the metric on $\mathbb{R}^+_t \times U$ takes the form

$$\eta = -\mu dt \otimes dt - \tfrac{1}{2} t(d\mu \otimes dt + dt \otimes d\mu) + t^2 h$$

or, in a slightly more attractive convention,

$$t^{-2}\eta = -\frac{\mu}{2}\left(\frac{dt}{t}\right)^2 - \frac{1}{2}\frac{dt}{t} \cdot d\mu + h \tag{1}$$

with the convention for the symmetric product $\cdot$ introduced in the following section. From this display we see that, by extending h to a family of Riemannian metrics on Y parametrised smoothly by $\mu \in (-1, 1)$, we can extend η smoothly onto the chart $\mathbb{R}^+_t \times U^2 \subset M_e$. We do this, thus furnishing M_e with a Lorentzian metric. As in the model geometry we refer to s (which is only defined on M) as the Minkowski scale, and to t (which is defined on M_e) as the Euclidean scale.

From (1), the measure associated with $t^{-2}\eta$ on $\mathbb{R}^+_t \times U^2$ is $\dfrac{dt}{t} dx$ where $dx = \frac{1}{2} d\mu\, d\mathrm{vol}_h$. On U, we have $dx = \rho^{n+2} d\mathrm{vol}_g$; hence dx extends smoothly to a measure on X_e, also denoted dx, and agrees with $d\mathrm{vol}_g$ on $X \backslash U$.

3. Symmetric tensors

This section introduces the necessary algebraic aspects of symmetric tensors and establishes conventions, which follow [Heil et al. 2016].

3A. ***A single fibre.*** Let E be a vector space of dimension $n+1$ equipped with an inner product g and let $\{e_i\}_{i=0}^n$ be an orthonormal basis and $\{e^i\}_{i=0}^n$ be the corresponding dual basis for E^*. We denote by $\mathrm{Sym}^k\, E^*$ the k-fold symmetric tensor product of E^*. Elements are symmetrised tensor products

$$u_1 \cdot \cdots \cdot u_k := \sum_{\sigma \in \Pi_k} u_{\sigma(1)} \otimes \cdots \otimes u_{\sigma(k)}, \quad u_i \in E^*,$$

where Π_k is the permutation group of $\{1, \dots, k\}$. By linearity, this extends the operation $\cdot$ to a map from $\mathrm{Sym}^k E^* \times \mathrm{Sym}^{k'} E^*$ to $\mathrm{Sym}^{k+k'} E^*$. Note the inner product takes the form $g = \frac{1}{2}\sum_{i=0}^n e^i \cdot e^i$ and that for $u \in E^*$ we write u^k to denote the symmetric product of k copies of u. The inner product induces an inner product on $\mathrm{Sym}^k E^*$ defined by

$$\langle u_1 \cdot \cdots \cdot u_k, v_1 \cdot \cdots \cdot v_k\rangle := \sum_{\sigma \in \Pi_k} g^{-1}(u_1, v_{\sigma(1)}) \cdots g^{-1}(u_k, v_{\sigma(k)}), \quad u_i, v_i \in E^*.$$

For $u \in E^*$, the metric adjoint of the linear map $u\cdot : \mathrm{Sym}^k E^* \to \mathrm{Sym}^{k+1} E^*$ is the contraction $u \lrcorner : \mathrm{Sym}^{k+1} E^* \to \mathrm{Sym}^k E^*$ defined by

$$(u \lrcorner v)(w_1, \dots, w_k) := v(u^\#, w_1, \dots, w_k), \quad u \in E^*,\ v \in \mathrm{Sym}^k E^*,\ w_i \in E,$$

where $u^\#$ is dual to u relative to the inner product on E. Contraction and multiplication with the metric g define two additional linear maps:

$$\begin{aligned} \Lambda : \mathrm{Sym}^k E^* &\to \mathrm{Sym}^{k-2} E^*, \\ u &\mapsto \textstyle\sum_{i=0}^n e^i \lrcorner e^i \lrcorner u, \end{aligned}$$

and

$$\begin{aligned} \mathrm{L} : \mathrm{Sym}^k E^* &\to \mathrm{Sym}^{k+2} E^*, \\ u &\mapsto \textstyle\sum_{i=0}^n e^i \cdot e^i \cdot u, \end{aligned}$$

which are adjoint to each other. As the notation is motivated by standard notation from complex geometry, we will refer to these two operators as Lefschetz-type operators.

Let F be the vector space $\mathbb{R} \times E$ equipped with the standard Lorentzian inner product $-f \otimes f + g$, where f is the canonical vector in $\mathbb{R}^*$. The previous constructions have obvious counterparts on F which will not be detailed. (For this subsection, we write $\langle \cdot, \cdot \rangle_F$ for the Lorentzian inner product on $\mathrm{Sym}^m F^*$.) The decomposition of F provides a decomposition of $\mathrm{Sym}^m F^*$:

$$\mathrm{Sym}^m F^* = \bigoplus_{k=0}^m a_k f^{m-k} \cdot \mathrm{Sym}^k E^*, \quad a_k = \frac{1}{\sqrt{(m-k)!}}$$

and we write

$$u = \sum_{k=0}^m a_k f^{m-k} \cdot u^{(k)}, \quad u \in \mathrm{Sym}^m F^*,\ u^{(k)} \in \mathrm{Sym}^k E^*.$$

The choice of the normalising constant a_k is chosen so that $\langle u, v\rangle_F = \sum_{k=0}^m (-1)^{m-k} \langle u^{(k)}, v^{(k)}\rangle$. There is a simple relationship between the terms $u^{(k)}$ in this decomposition of u when u is trace-free.

Lemma 3.1. *Let Λ_F and Λ denote the Lefschetz-type trace operators obtained from the inner products on F and E respectively. For $u \in \mathrm{Sym}^m F^*$ in the kernel of Λ_F, we have*

$$\Lambda u^{(k)} = -b_{k-2} b_{k-1} u^{(k-2)},$$

where $u = \sum_{k=0}^m a_k f^{m-k} \cdot u^{(k)}$ for $u^{(k)} \in \mathrm{Sym}^k E^*$ and constants $b_k = \sqrt{m-k}$.

Proof. Beginning with $\Lambda_F f^{m-k} = (m-k)(m-k-1) f^{m-k-2}$ we obtain

$$\Lambda_F(a_k f^{m-k} \cdot u^{(k)}) = a_{k+2}\sqrt{(m-k)(m-k-1)}\, f^{m-k-2} \cdot u^{(k)} + a_k f^{m-k} \cdot \Lambda u^{(k)}.$$

Therefore, as $u \in \ker \Lambda_F$, equating powers of f in the resulting formula for

$$\Lambda_F\Big(\sum_{k=0}^{m} a_k f^{m-k} \cdot u^{(k)}\Big)$$

gives

$$a_k f^{m-k} \Lambda u^{(k)} + a_k \sqrt{(m-k+2)(m-k+1)}\, f^{m-k} u^{(k-2)} = 0. \qquad \square$$

We introduce some notation for finite sequences to simplify the calculations below. Denote by $\mathscr{A}^k$ the space of all sequences $K = k_1 \cdots k_k$ with $0 \le k_r \le n$. We write $\{k_r \to j\}K$ for the result of replacing the r-th element of K by j. If j is not present, this implies we remove the r-th element from K, while if k_r is not present, this implies we add j to K to obtain jK. This notation extends to replacing multiple indices at once. For example, $\{k_p \to,\ k_r \to\}K$ indicates we first remove the r-th element from K and then remove the p-th element from $\{k_r \to\}K$. We set

$$e^K = e^{k_1} \cdot \cdots \cdot e^{k_m} \in \otimes^k E^*, \qquad K = k_1 \cdots k_m \in \mathscr{A}^k.$$

3B. ***Vector bundles.*** These constructions are naturally extended to vector bundles above manifolds. We include this subsection in order to introduce our notation and conventions. Consider M and X (with similar constructions for M_e and X_e). We define

$$\mathcal{F} := \mathrm{Sym}^m \mathrm{T}^* M, \qquad \mathcal{E}^{(k)} := \mathrm{Sym}^k \mathrm{T}^* X, \qquad \mathcal{E} := \bigoplus_{k=0}^{m} \mathcal{E}^{(k)}.$$

If we want to make precise that $\mathcal{F}$ consists of rank-m symmetric cotensors, we will write $\mathcal{F}^{(m)}$. The Minkowski scale gives the decomposition $M = \mathbb{R}^+_s \times X$ and we denote by π the projection onto the second factor $\pi : M \to X$. (Remark that on M this gives the same map as the projection $\pi : M_e \to X_e$ using the Euclidean scale $M_e = \mathbb{R}^+_t \times X_e$.) This enables $\mathcal{E}^{(k)}$ to be pulled back to a bundle over M which we will also denote by $\mathcal{E}^{(k)}$.

Given $u \in C^\infty(M; \mathcal{F})$, we decompose u as

$$u = \sum_{k=0}^{m} a_k \left(\frac{ds}{s}\right)^{m-k} \cdot u^{(k)}, \qquad u^{(k)} \in C^\infty(M; \mathcal{E}^{(k)}), \tag{2}$$

where a_k is the previously introduced constant $((m-k)!)^{-1/2}$. We say that such a decomposition is relative to the Minkowski scale.

For a fixed value of s, say s_0, there is an identification of the corresponding subset of M with X via the map $\pi_{|s=s_0}$. We will thus reuse π for the map

$$\pi_{s=s_0} : C^\infty(M; \mathcal{F}) \to C^\infty(X; \mathcal{E}),$$
$$u = \sum_{k=0}^{m} a_k \Big(\frac{ds}{s}\Big)^{m-k} \cdot u^{(k)} \mapsto \sum_{k=0}^{m} \pi_{|s=s_0} u^{(k)},$$

and in order to map from $C^\infty(X;\mathcal{E})$ to $C^\infty(M;\mathcal{F})$, taking into account the Minkowski scale, we introduce

$$\pi_s^*: C^\infty(X;\mathcal{E}) \to C^\infty(M;\mathcal{F}),$$
$$u = \sum_{k=0}^{m} u^{(k)} \mapsto \sum_{k=0}^{m} a_k \left(\frac{ds}{s}\right)^{m-k} \cdot \pi^* u^{(k)}.$$

On M we have two useful metrics. First, $s^{-2}\eta$, which takes the model form of the metric on F introduced in the previous subsection

$$s^{-2}\eta = -\frac{1}{2}\left(\frac{ds}{s}\right)^2 + g.$$

Second, we have the metric η, which is geometrically advantageous as it gives the Lorentzian cone metric on M. Notationally we will distinguish the two constructions by decorating the Lefschetz-type operators with a subscript of the particular metric used. A similar decoration will be used for the two inner products on $\mathcal{F}$. There are two useful relationships. First,

$$\Lambda_{s^{-2}\eta} u = s^4 \Lambda_\eta u, \quad u \in \mathcal{F}, \tag{3}$$

and second,

$$\langle u, v\rangle_{s^{-2}\eta} = s^{2m} \langle u, v\rangle_\eta, \quad u, v \in \mathcal{F}. \tag{4}$$

On X, when the metric g is used, no such decoration will be added. We can, however, make use of the metric $s^{-2}\eta$ by appealing to π_s^*. We introduce $\langle\,\cdot\,,\,\cdot\,\rangle_s$ on $C^\infty(X;\mathcal{E})$ by declaring

$$\langle u, v\rangle_s := \langle \pi_s^* u, \pi_s^* v\rangle_{s^{-2}\eta}, \quad u, v \in C^\infty(X;\mathcal{E}).$$

Note that such a definition does not depend on the value of $s \in \mathbb{R}^+$ at which point the inner product on $\mathcal{F}$ is applied. With this inner product given, and the measure $d\mathrm{vol}_g$ previously introduced, we obtain the notion of L^2 sections and define

$$L_s^2(X;\mathcal{E}) := L^2(X, d\mathrm{vol}_g; \mathcal{E}, \langle\,\cdot\,,\,\cdot\,\rangle_s), \tag{5}$$

whose inner product is provided by

$$(u, v)_s := \int_X \langle u, v\rangle_s \, d\mathrm{vol}_g, \quad u, v \in C_c^\infty(X;\mathcal{E}).$$

On X_e, we define L^2 sections with respect to the measure dx,

$$L_t^2(X_e;\mathcal{E}) := L^2(X_e, dx; \mathcal{E}, \langle\,\cdot\,,\,\cdot\,\rangle_t).$$

On X, the necessary correspondences between the constructions using the Lorentzian and Euclidean scales are given in the following lemma.

Lemma 3.2. *There exists $J \in C^\infty(X; \mathrm{End}\,\mathcal{E})$ such that*

$$\pi_s^* u = \pi_t^* J u, \quad u \in C^\infty(X;\mathcal{E}),$$

whose entries are homogeneous polynomials of degree at most m in $d\rho/\rho$, upper triangular in the sense that $J(\mathcal{E}^{(k_0)}) \subset \bigoplus_{k=k_0}^{m} \mathcal{E}^{(k)}$, and whose diagonal entries are the identity. Moreover,

$$\langle u, v\rangle_s = \rho^{2m}\langle Ju, Jv\rangle_t, \quad u, v \in C^\infty(X; \mathcal{E}).$$

Finally,

$$L^2_s(X; \mathcal{E}) = \rho^{n/2-m+1} J^{-1} L^2_t(X; \mathcal{E}).$$

Proof. As $t = s/\rho$, the differentials are related by

$$\frac{ds}{s} = \frac{dt}{t} + \frac{d\rho}{\rho},$$

and hence by the binomial expansion

$$a_k\left(\frac{ds}{s}\right)^{m-k} \cdot \pi^* u^{(k)} = \sum_{j=0}^{m-k} a_{k+j}\left(\frac{dt}{t}\right)^{m-k-j} \cdot \binom{m-k}{j}\frac{a_k}{a_{k+j}}\left(\frac{d\rho}{\rho}\right)^j \cdot \pi^* u^{(k)},$$

where $u^{(k)} \in C^\infty(X; \mathcal{E}^{(k)})$. This defines the endomorphism J by declaring

$$Ju^{(k)} = \sum_{j=0}^{m-k}\binom{m-k}{j}\frac{a_k}{a_{k+j}}\left(\frac{d\rho}{\rho}\right)^j \cdot u^{(k)}.$$

The second claim is direct from $s^{-2}\eta = \rho^{-2}t^{-2}\eta$; hence on $\mathcal{F}$, where the inner product requires m applications of the inverse metric, $\langle\cdot,\cdot\rangle_{s^{-2}\eta} = \rho^{2m}\langle\cdot,\cdot\rangle_{t^{-2}\eta}$. The final claim follows from the second claim and the previously mentioned correspondence, $dx = \rho^{n+2} d\mathrm{vol}_g$. □

4. b-calculus and microlocal analysis

This section introduces the necessary b-calculus formalism on symmetric cotensors. The standard reference is [Melrose 1993]; in particular we make much use of Chapters 2 and 5. We also recall some now standard ideas from microlocal analysis.

4A. ***b-calculus.*** For convenience we will only work on $M = \mathbb{R}^+_s \times X$ rather than on both M and M_e. We define $\overline{M}$ to be the closure of M seen as a submanifold of $\mathbb{R}_s \times X$ with its usual topology. Then

$$\overline{M} = M \sqcup X,$$

where X is naturally identified with the boundary $\partial\overline{M} = \{s = 0\}$.

We let $\{e_i\}_{i=0}^n$ denote a (local) holonomic frame for $\mathrm{T}X$ and $\{e^i\}_{i=0}^n$ its dual frame for T^*X. The Lie algebra of b-vector fields consists of smooth vector fields on $\overline{M}$ tangent to the boundary X. It is thus generated by $\{s\,\partial_s, e_i\}$. This provides the smooth vector bundle ${}^{\mathrm{b}}\mathrm{T}\overline{M}$. The dual bundle, ${}^{\mathrm{b}}\mathrm{T}^*\overline{M}$, has basis $\{ds/s, e^i\}$. This dual bundle is used to construct the b-symmetric bundle of m-cotensors, denoted by ${}^{\mathrm{b}}\mathcal{F}$. On the interior of $\overline{M}$, this bundle is canonically isomorphic to $\mathcal{F}$.

An operator $\boldsymbol{Q}$ belongs to $\mathrm{Diff}_{\mathrm{b}}^{p}(\overline{M}; \mathrm{End}\,{}^{\mathrm{b}}\mathcal{F})$ if, relative to a frame generated by $\{ds/s, e^i\}$ the operator $\boldsymbol{Q}$ may be written as a matrix

$$\boldsymbol{Q} = [\boldsymbol{Q}_{i,j}]$$

whose coefficients $\boldsymbol{Q}_{i,j}$ belong to $\mathrm{Diff}_{\mathrm{b}}^{p}(\overline{M})$. That is, each $\boldsymbol{Q}_{i,j}$ may be written as

$$\boldsymbol{Q}_{i,j} = \sum_{k,|\alpha|\leq p} q_{i,j,k,\alpha}(s\,\partial_s)^k \partial_x^\alpha$$

for smooth functions $q_{i,j,k,\alpha} \in C^\infty(\overline{M})$.

Operators in $\mathrm{Diff}_{\mathrm{b}}^{p}(\overline{M}; \mathrm{End}\,{}^{\mathrm{b}}\mathcal{F})$ provide indicial families of operators belonging to $\mathrm{Diff}^p(X; \mathrm{End}\,\mathcal{E})$. In order to define this mapping we recall the operator $\pi_{s=s_0}$ defined in the previous section for $s_0 \in \mathbb{R}^+$. This family of maps clearly has an extension to $\overline{M}$ giving

$$\pi_{s=s_0} : C^\infty(\overline{M}; {}^{\mathrm{b}}\mathcal{F}) \to C^\infty(X; \mathcal{E}),$$

where $s_0 \in [0, \infty)$. The indicial family mapping (with respect to the Minkowski scale s)

$$\mathrm{I}_s : \mathrm{Diff}_{\mathrm{b}}^{p}(\overline{M}; \mathrm{End}\,{}^{\mathrm{b}}\mathcal{F}) \to \mathcal{O}(\mathbb{C}; \mathrm{Diff}^p(X; \mathrm{End}\,\mathcal{E}))$$

is defined by

$$\mathrm{I}_s(\boldsymbol{Q}, \lambda)(u) := \pi_{s=0}(s^\lambda \boldsymbol{Q} s^{-\lambda}(\pi_s^* u)), \quad u \in C^\infty(X; \mathcal{E}).$$

When the scale s is understood, we will use the convention of removing the bold font from such an operator and write

$$\mathcal{Q} := \mathrm{I}_s(\boldsymbol{Q}, \cdot), \quad \mathcal{Q}_\lambda := \mathrm{I}_s(\boldsymbol{Q}, \lambda).$$

Remark 4.1. This definition effectively does three things. First, if $\boldsymbol{Q}$ is written as a matrix, relative to the decomposition established by the Minkowski scale (2), then $\mathcal{Q}$ will take the same form but without the appearances of $a_k(ds/s)^{m-k}\,\cdot$. Next, the functions $q_{i,j,k,\alpha}$ are frozen to their values at $s = 0$. (These two results are due to the appearance of $\pi_{s=0}$.) Finally, due to the conjugation by s^λ, all appearances of $s\,\partial_s$ in $\boldsymbol{Q}$ are replaced by the complex parameter $-\lambda$.

Remark 4.2. The choice to conjugate by $s^{-\lambda}$ is to ensure that the subsequent operators (in particular $\mathcal{P}$) acting on L^2 sections have physical domains corresponding to $\mathrm{Re}\,\lambda \gg 1$. If one is convinced that the convention ought to be conjugation by s^λ rather than $s^{-\lambda}$ one can kill two birds with one stone: Considering the model geometry, which motivates the viewpoint of hyperbolic space "at infinity" inside the forward light cone of compactified Minkowski space, it would be somewhat more natural to introduce the coordinate $\tilde{s} = s^{-1}$ on M, then construct the closure of M as a submanifold of $\mathbb{R}_{\tilde{s}} \times X$. The indicial family would then by constructed via a conjugation of $\tilde{s}^\lambda$ and appearances of $\tilde{s}\partial_{\tilde{s}} = -s\,\partial_s$ would be replaced by λ. For this article, the aesthetics of such a choice are outweighed by the superfluous introduction of two dual variables, one for each of s and t.

The b-operators we consider are somewhat simpler than the previous definition in that the coefficients $q_{i,j,k,\alpha}$ do not depend on s (in the correct basis).

Definition 4.3. A b-operator $\boldsymbol{Q} \in \mathrm{Diff}_{\mathrm{b}}^{p}(\overline{M}; {}^{\mathrm{b}}\mathcal{F})$ is b-trivial if, for all $s_0 \in \mathbb{R}^+$,

$$\mathrm{I}_s(\boldsymbol{Q}, \lambda)(u) = \pi_{s=s_0}(s^{\lambda}\boldsymbol{Q}s^{-\lambda}(\pi_s^* u)), \quad u \in C^{\infty}(X; \mathcal{E}).$$

One advantage of this property is that self-adjointness of $\boldsymbol{Q}$ easily implies self-adjointness of $\mathcal{Q}_\lambda$ for $\lambda \in i\mathbb{R}$.

Lemma 4.4. *Suppose $\boldsymbol{Q}$ is b-trivial and formally self-adjoint relative to the inner product*

$$(u, v)_{s^{-2}\eta} = \int_M \langle u, v\rangle_{s^{-2}\eta} \frac{ds}{s}\, d\mathrm{vol}_g, \quad u, v \in C_c^{\infty}(M; \mathcal{F}).$$

Then, the indicial family $\mathcal{Q}$ is, upon restriction to $\lambda \in i\mathbb{R}$, formally self-adjoint relative to the inner product

$$(u, v)_s = \int_X \langle u, v\rangle_s\, d\mathrm{vol}_g, \quad u, v \in C_c^{\infty}(X; \mathcal{E}).$$

Moreover, for all λ, we have $\mathcal{Q}_\lambda^* = \mathcal{Q}_{-\bar{\lambda}}$.

Proof. We prove only the first claim. That $\mathcal{Q}_\lambda^* = \mathcal{Q}_{-\bar{\lambda}}$ for all λ follows by the same reasoning, making the obvious changes in the second display provided below. Let ψ be a smooth function on $\mathbb{R}_s^+$ with compact support (away from $s = 0$) and with unit mass $\int_{\mathbb{R}^+} \psi\, (ds/s) = 1$. Let $u, v \in C_c^{\infty}(X; \mathcal{E})$. The b-triviality provides

$$(\mathcal{Q}_\lambda u, v)_s = \int_{\mathbb{R}^+} (\mathcal{Q}_\lambda u, v)_s\, \psi\, \frac{ds}{s} = (s^{\lambda}\boldsymbol{Q}s^{-\lambda}\pi_s^* u, \psi\pi_s^* v)_{s^{-2}\eta}.$$

For $\lambda \in i\mathbb{R}$ this develops as

$$\begin{aligned}(\mathcal{Q}_\lambda u, v)_s &= (\pi_s^* u, s^{\lambda}\boldsymbol{Q}s^{-\lambda}\psi\pi_s^* v)_{s^{-2}\eta}\\ &= (\pi_s^* u, \psi s^{\lambda}\boldsymbol{Q}s^{-\lambda}\pi_s^* v)_{s^{-2}\eta} + (\pi_s^* u, [s^{\lambda}\boldsymbol{Q}s^{-\lambda}, \psi]\pi_s^* v)_{s^{-2}\eta}\\ &= (u, \mathcal{Q}_\lambda v)_s + (\pi_s^* u, [s^{\lambda}\boldsymbol{Q}s^{-\lambda}, \psi]\pi_s^* v)_{s^{-2}\eta},\end{aligned}$$

where the last line has again used the b-triviality. Thus we require

$$(\pi_s^* u, [s^{\lambda}\boldsymbol{Q}s^{-\lambda}, \psi]\pi_s^* v)_{s^{-2}\eta} = 0. \tag{6}$$

Consider $\boldsymbol{Q}$ as a matrix $\boldsymbol{Q} = [\boldsymbol{Q}_{i,j}]$ with respect to a basis in which

$$\boldsymbol{Q}_{i,j} = \sum_{k,|\alpha|\le p} q_{i,j,k,\alpha}(s\,\partial_s)^k \partial_x^{\alpha}$$

for $q_{i,j,k,\alpha} \in C^{\infty}(X)$. The key is to note that we may write

$$[s^{\lambda}\boldsymbol{Q}_{i,j}s^{-\lambda}, \psi] = \sum_{k,|\alpha|\le p-1} \kappa_{i,j,k,\alpha}(s\,\partial_s)^k \partial_x^{\alpha} \tag{7}$$

for smooth functions (which depend on λ) $\kappa_{i,j,k,\alpha} \in C^{\infty}(X)$ such that every term in each $\kappa_{i,j,k,\alpha}$ is smoothly divisible by some nonzero integer $(s\,\partial_s)$-derivative of ψ. Factoring out these appearances and integrating over $\mathbb{R}^+$ in (6) causes, by the fundamental theorem of calculus, the problematic term to vanish.

The factorisation claim involving the functions $\kappa_{i,j,k,\alpha}$ follows directly from the following calculation. First

$$[s^\lambda \boldsymbol{Q}_{i,j} s^{-\lambda}, \psi] = \sum_{k,|\alpha|\le p} q_{i,j,k,\alpha}[(s\,\partial_s - \lambda)^k \partial_x^\alpha, \psi] = \sum_{\substack{k,|\alpha|\le p\\ k\ge 1}} q_{i,j,k,\alpha}[(s\,\partial_s - \lambda)^k, \psi]\partial_x^\alpha,$$

and for $k > 1$,

$$[(s\,\partial_s - \lambda)^k, \psi] = \sum_{\ell=1}^{k} \binom{k}{\ell}(-\lambda)^{k-\ell}[(s\,\partial_s)^\ell, \psi] = \sum_{\ell=1}^{k}\sum_{m=1}^{\ell} \binom{k}{\ell}(-\lambda)^{k-\ell}\binom{\ell}{m}((s\,\partial_s)^m\psi)(s\,\partial_s)^{\ell-m},$$

which, due to the appearance of $(s\,\partial_s)^m\psi$ gives (7) with the desired structure. □

Remark 4.5. The use of $d\mathrm{vol}_g$ is unimportant; the result holds for any measure on X given such a measure also appears as $d\mathrm{vol}_g$ does in the inner product on M.

We finish this subsection by remarking on the effect that the scale (Minkowski or Euclidean) has on the indicial family.

Lemma 4.6. *For $\boldsymbol{Q} \in \mathrm{Diff}_{\mathrm{b}}^p(\overline{M}; {}^{\mathrm{b}}\mathcal{F})$, the indicial families obtained using the scales s and t are related by*

$$\mathrm{I}_s(\boldsymbol{Q}, \lambda) = \rho^\lambda J^{-1} \mathrm{I}_t(\boldsymbol{Q}, \lambda) J \rho^{-\lambda}$$

with J presented in Lemma 3.2.

Proof. Lemma 3.2 provides $\pi_s^* = \pi_t^* \circ J$. Dual to this equation, $\pi_{s=0} = J^{-1} \circ \pi_{t=0}$. Combining these observations gives the result

$$\begin{aligned}\mathrm{I}_s(\boldsymbol{Q}, \lambda)(u) &= \pi_{s=0}(s^\lambda \boldsymbol{Q} s^{-\lambda}(\pi_s^* u))\\ &= J^{-1}\pi_{t=0}\big(\rho^\lambda t^\lambda \boldsymbol{Q} t^{-\lambda}\rho^{-\lambda}(\pi_t^* J u)\big)\\ &= \rho^\lambda J^{-1} \mathrm{I}_t(\boldsymbol{Q}, \lambda)(J\rho^{-\lambda}u).\end{aligned}$$

□

4B. *Microlocal analysis.* We recall standard objects in microlocal analysis (the necessary information is given in [Zworski 2016] for pseudodifferential operators acting on the trivial bundle; here we merely indicate the small changes that occur when acting on a vector bundle). Recall the open submanifold $X_{cs} = \{\mu > -\frac{1}{2}\} \subset X_e$ from Definition 2.3. We will assume that $L_t^2(X_e; \mathcal{E})$ provides a notion of sections above X_{cs} with Sobolev regularity s, denoted by $H^s(X_{cs}; \mathcal{E})$, with norm $\|\cdot\|_{H^s}$ (see Section 7A for subtleties arising due to the boundary S). Let ζ denote the coefficients of a covector relative to some local base for $\mathrm{T}^* X_{cs}$ such that we may define the Japanese bracket $\langle\zeta\rangle$. We denote by

$$\Psi_{\mathrm{scal}}^p(X_{cs}; \mathrm{End}\,\mathcal{E}) \subset \Psi^p(X_{cs}; \mathrm{End}\,\mathcal{E})$$

the space of properly supported pseudodifferential operators of order p acting on $\mathcal{E}$ and which have scalar principal symbol. For $A \in \Psi_{\mathrm{scal}}^a(X_{cs}; \mathrm{End}\,\mathcal{E})$ such a symbol is written as

$$\sigma(A) \in S^a(\mathrm{T}^* X_{cs} \backslash 0; \mathrm{End}\,\mathcal{E}) / S^{a-1}(\mathrm{T}^* X_{cs} \backslash 0; \mathrm{End}\,\mathcal{E})$$

and is scalar. For such operators, it continues to hold that, for $B \in \Psi^b_{\text{scal}}(X_{cs}; \operatorname{End}\mathcal{E})$, the principal symbol of the composition

$$\sigma(AB) = \sigma(A)\sigma(B) \in S^{a+b}(\mathrm{T}^*X_{cs}\backslash 0; \operatorname{End}\mathcal{E})/S^{a+b-1}(\mathrm{T}^*X_{cs}\backslash 0; \operatorname{End}\mathcal{E})$$

remains scalar. However now, as lower-order terms are not required to be diagonal, the commutator has principal symbol

$$\sigma([A, B]) \in S^{a+b-1}(\mathrm{T}^*X_{cs}\backslash 0; \operatorname{End}\mathcal{E})/S^{a+b-2}(\mathrm{T}^*X_{cs}\backslash 0; \operatorname{End}\mathcal{E}),$$

which, in general, is not scalar. In the case that $A \in \Psi^a(X_{cs}) \subset \Psi^a_{\text{scal}}(X_{cs}; \operatorname{End}\mathcal{E})$ we get

$$\sigma\Big(\frac{1}{2i}[A, B]\Big) = \frac{1}{2} H_{\sigma(B)}(\sigma(A)),$$

where $H_{\sigma(B)}$ is the Hamiltonian vector field associated with $\sigma(B)$. Exactly as in the case that $\mathcal{E}$ is the trivial bundle, associated with the operator A are the notions of the wave front set $\mathrm{WF}(A)$ and the characteristic variety $\mathrm{Char}(A)$.

There are two radial estimates used in the analysis of $\mathcal{P}$ (the family of operators introduced in Section 6) in order to prove Proposition 7.3. The analysis is performed in [Vasy 2013a, Section 2.4] for functions with an alternative description given in [Dyatlov and Zworski 2017, Section E.5.2]. We will follow the second approach and translate the results into a (nonsemiclassical) setting adapted to vector bundles. For this, and to follow closely the referenced works, we introduce [Dyatlov and Zworski 2017, Section E.1.2] the radially compactified cotangent bundle $\overline{\mathrm{T}}^*X_{cs}$ and projection map $\kappa : \mathrm{T}^*X_{cs}\backslash 0 \to \partial\overline{\mathrm{T}}^*X_{cs}$. Consider $P \in \Psi^p_{\text{scal}}(X_{cs}; \operatorname{End}\mathcal{E})$ with real principal symbol $\sigma(P)$ and Hamiltonian vector field $H_{\sigma(P)}$. Write P as

$$P = \operatorname{Re} P + i \operatorname{Im} P$$

for

$$\operatorname{Re} P = \frac{P + P^*}{2} \in \Psi^p_{\text{scal}}(X_{cs}; \operatorname{End}\mathcal{E}), \quad \operatorname{Im} P = \frac{P - P^*}{2i} \in \Psi^{p-1}(X_{cs}; \operatorname{End}\mathcal{E}).$$

In the sense of [Dyatlov and Zworski 2017, Definition E.52], let Γ_+ and Γ_- be a source and a sink of $\sigma(P)$ respectively. Suppose that $\langle\zeta\rangle^{1-p} H_{\sigma(P)}$ vanishes on $\Gamma_\pm$.

Lemma 4.7. *Let s satisfy the following threshold condition on Γ_+:*

$$\langle\zeta\rangle^{1-p}\Big(\sigma(\operatorname{Im} P) + \Big(s + \frac{1-p}{2}\Big) H_{\sigma(P)} \log\langle\zeta\rangle\Big) \quad \textit{is negative definite.}$$

Then for all $B_1 \in \Psi^0(X_{cs})$ with $\mathrm{WF}(I - B_1) \cap \Gamma_+ = \varnothing$, there exists $A \in \Psi^0(X_{cs})$ with $\mathrm{Char}(A) \cap \Gamma_+ = \varnothing$ such that for any $u \in C^\infty_c(X_{cs}; \mathcal{E})$ (and any N large enough)

$$\|Au\|_{H^s} \le C(\|B_1 P u\|_{H^{s-p+1}} + \|u\|_{H^{-N}}).$$

Lemma 4.8. *Let s satisfy the following threshold condition on Γ_-:*

$$\langle\zeta\rangle^{1-p}\Big(\sigma(\operatorname{Im} P) + \Big(s + \frac{1-p}{2}\Big) H_{\sigma(P)} \log\langle\zeta\rangle\Big) \quad \textit{is negative definite.}$$

Then for all $B_1 \in \Psi^0(X_{cs})$ *with* $\mathrm{WF}(I-B_1)\cap\Gamma_- = \varnothing$, *there exists* $A, B \in \Psi^0(X_{cs})$ *with* $\mathrm{Char}(A)\cap\Gamma_- = \varnothing$ *and* $\mathrm{WF}(B)\cap\Gamma_- = \varnothing$ *such that for any* $u \in C_c^\infty(X_{cs}; \mathcal{E})$ (*and any* N *large enough*)

$$\|Au\|_{H^s} \le C(\|Bu\|_{H^s} + \|B_1 Pu\|_{H^{s-p+1}} + \|u\|_{H^{-N}}).$$

Remark 4.9. There are two trivial but important points to make. First, a source for P is a sink for $-P$ (and similarly a sink for P is a source for $-P$). Second, we have assumed P has real principal symbol; therefore, when considering its adjoint P^*, we have $H_{\sigma(P^*)} = H_{\sigma(P)}$. Less trivially, by approximation [Dyatlov and Zworski 2017, Lemma E.47], these results do not need to assume $u \in C_c^\infty(X_{cs}; \mathcal{E})$. In Lemma 4.7, if $s > \tilde{s}$ with $\tilde{s}$ satisfying the threshold condition and $u \in H^{\tilde{s}}(X_{cs}; \mathcal{E})$ then the inequality holds (on the condition that the right-hand side is finite). Similarly in Lemma 4.8, if u is a distribution such that the right-hand side of the inequality is well defined, then so too is the left-hand side, and the inequality holds.

5. The Laplacian, the d'Alembertian and the operator Q

This section shows the relationship between the Laplacian on (X, g) and the d'Alembertian on (M, η). We first introduce several differential operators on X using the Levi-Civita connection ∇ of g extended to all associated vector bundles associated with the principal orthonormal frame bundle. Let $\{e_i\}_{i=0}^n$ be a local orthonormal frame for $\mathrm{T}X$ and $\{e^i\}_{i=0}^n$ be the corresponding dual frame for T^*X. We define two first-order differential operators. Let the symmetrisation of the covariant derivative, called the symmetric differential, be denoted by d:

$$\begin{aligned} \mathrm{d} : C^\infty(X; \mathcal{E}^{(k)}) &\to C^\infty(X; \mathcal{E}^{(k+1)}), \\ u &\mapsto \textstyle\sum_{i=0}^n e^i \cdot \nabla_{e_i} u. \end{aligned}$$

Denote by δ its formal adjoint called the divergence:

$$\begin{aligned} \delta : C^\infty(X; \mathcal{E}^{(k)}) &\to C^\infty(X; \mathcal{E}^{(k-1)}), \\ u &\mapsto -\textstyle\sum_{i=0}^n e^i \lrcorner \nabla_{e_i} u. \end{aligned}$$

The two first-order operators behave nicely with L and Λ, giving the following commutation relations [Heil et al. 2016, Equation (8)]:

$$[\Lambda, \delta] = 0 = [\mathrm{L}, \mathrm{d}], \quad [\Lambda, \mathrm{d}] = -2\delta, \quad [\mathrm{L}, \delta] = 2\mathrm{d}. \tag{8}$$

The rough Laplacian on this space will be denoted by $\nabla^*\nabla$:

$$\begin{aligned} \nabla^*\nabla : C^\infty(X; \mathcal{E}^{(k)}) &\to C^\infty(X; \mathcal{E}^{(k)}), \\ u &\mapsto \nabla^*\nabla u, \end{aligned}$$

where ∇^* is the formal adjoint of $\nabla : C^\infty(X; \mathcal{E}^{(k)}) \to C^\infty(X; \mathrm{T}^*X \otimes \mathcal{E}^{(k)})$. Equivalently

$$\nabla^*\nabla\, u = (-\operatorname{tr} \circ \nabla \circ \nabla)(u), \quad u \in C^\infty(X; \mathcal{E}^{(k)}),$$

where $\mathrm{tr} : \mathrm{T}^*X \otimes \mathrm{T}^*X \to \mathbb{R}$ is the trace operator obtained from g and is extended to

$$\mathrm{tr} : \mathrm{T}^*X \otimes \mathrm{T}^*X \otimes \mathcal{E}^{(k)} \to \mathcal{E}^{(k)}.$$

For the Lichnerowicz Laplacian, we introduce the Riemann curvature tensor, which will be denoted by R:

$$\mathrm{R}_{u,v} w = [\nabla_u, \nabla_v] w - \nabla_{[u,v]} w, \quad u, v, w \in C^\infty(X; TX).$$

It is extended to all tensor bundles as a derivation. On symmetric k-cotensors we introduce the curvature endomorphism which will be denoted by $q(\mathrm{R})$:

$$q(\mathrm{R})\, u = \sum_{i,j=0}^{n} e^j \cdot e^i \lrcorner \mathrm{R}_{e_i,e_j} u, \quad u \in \mathcal{E}^{(k)}.$$

The Lichnerowicz Laplacian, hereafter simply referred to as the Laplacian, will be denoted by Δ:

$$\begin{aligned} \Delta : C^\infty(X; \mathcal{E}^{(k)}) &\to C^\infty(X; \mathcal{E}^{(k)}), \\ u &\mapsto (\nabla^*\nabla + q(\mathrm{R}))u. \end{aligned}$$

We decompose symmetric k-cotensors using the symmetrised basis elements:

$$u = \sum_{K \in \mathscr{A}^k} u_K e^K, \quad u \in C^\infty(X; \mathcal{E}^{(k)}),\ u_K \in C^\infty(X).$$

Useful formulae for the preceding operators thus far introduced are given in the following lemma. Recall the notation for finite sequences $\mathscr{A}^k$ introduced in the final paragraph of Section 3A.

Lemma 5.1. *Let $u \in C^\infty(X; \mathcal{E}^{(k)})$. At a point in X about which $\{e_i\}$ are normal coordinates, the trace is*

$$\Lambda u = \sum_{K \in \mathscr{A}^k} \sum_{k_r \in K} \sum_{k_p \in \{k_r \to\}K} g^{k_r k_p} u_K e^{\{k_p \to,\, k_r \to\}K},$$

the symmetric differential is

$$\mathrm{d}u = \sum_{K \in \mathscr{A}^k} \sum_{i=0}^{n} (e_i u_K) e^{\{\to i\}K},$$

the divergence is

$$\delta u = -\sum_{K \in \mathscr{A}^k} \sum_{k_r \in K} \sum_{i=0}^{n} g^{i k_r} (e_i u_K) e^{\{k_r \to\}K},$$

and the rough Laplacian is

$$\nabla^*\nabla\, u = \sum_{K \in \mathscr{A}^k} \sum_{i,j=0}^{n} \Biggl(-g^{ij} (e_i e_j u_K) e^K + \sum_{k_r \in K} \sum_{\ell=0}^{n} g^{i\ell} u_K (e_\ell \Gamma^{k_r}_{ij}) e^{\{k_r \to j\}K} \Biggr),$$

where the connection coefficients are given locally by $\nabla_{e_i} e^k = -\sum_{j=0}^{n} \Gamma^k_{ij} e^j$. Finally, (at a point using normal coordinates), the Riemann curvature takes the form

$$\mathrm{R}_{e_i,e_j} e^\ell = -\sum_{k=0}^{n} \mathrm{R}_{ij}{}^{\ell}{}_{k} e^k, \quad \mathrm{R}_{ij}{}^{\ell}{}_{k} = e_i \Gamma^\ell_{jk} - e_j \Gamma^\ell_{ik}.$$

In a similar vein to (8) we have the following two useful results, the second of which originates from [Lichnerowicz 1961, Section 10].

Lemma 5.2. *Let $u \in C^\infty(X, \mathcal{E}^{(k)})$. The Laplacian commutes with the Lefschetz-type trace operator*

$$[\Lambda, \Delta]u = 0$$

and commutes with the divergence under the following conditions:

$$[\delta, \Delta]u = 0 \quad \textit{if} \begin{cases} k = 0, 1, \\ k = 2 \textit{ and } X \textit{ is Ricci parallel}, \\ k \geq 3 \textit{ and } X \textit{ is locally isomorphic to } \mathbb{H}^{n+1}. \end{cases}$$

Proof. The first result is very standard. As the metric is parallel, the Riemann curvature tensor (acting as a derivation on $\mathcal{E}^{(k)}$) commutes with L; hence

$$[\mathrm{L}, q(\mathrm{R})]u = \sum_{i,j=0}^{n} (\mathrm{L}e^j \lrcorner\, e^i \lrcorner - e^j \lrcorner\, e^i \lrcorner\, \mathrm{L})\, \mathrm{R}_{e_i,e_j} u$$

and developing the second term with the aid of the commutation formula $[e^i \lrcorner, \mathrm{L}] = 2e^i \cdot$ provides

$$[\mathrm{L}, q(\mathrm{R})]u = \sum_{i,j=0}^{n} -2(e^j \cdot e^i \lrcorner + e^j \lrcorner\, e^i \cdot)\, \mathrm{R}_{e_i,e_j} u = \sum_{i,j=0}^{n} -2(e^j \cdot e^i \lrcorner + \delta^{ij} + e^i \cdot e^j \lrcorner)\, \mathrm{R}_{e_i,e_j} u,$$

which vanishes due to the skew-symmetry of the Riemann curvature tensor. By duality, $[\Lambda, q(\mathrm{R})] = 0$. Now using the commutation relations (8) and the characterisation of the Laplacian [Heil et al. 2016, Proposition 6.2]

$$\Delta = \delta \mathrm{d} - \mathrm{d}\delta + 2q(\mathrm{R})$$

provides the commutation of Λ with Δ.

The second result is more involved as a demonstration via a direct calculation (however, as these statements are well known, we only sketch said calculations). For $k = 0, 1$ the Laplacian and divergence agree with Hodge Laplacian and the adjoint of the exterior derivative. We will thus assume X is Ricci parallel (and $k \geq 2$). We break the calculation into two parts studying $[\delta, \nabla^*\nabla]$ and $[\delta, q(\mathrm{R})]$. As usual, we use a frame $\{e_i\}_{i=0}^n$ for TX with dual frame $\{e^i\}_{i=0}^n$ and calculate at a point about which the connection coefficients vanish. We act on $u = u_K e^K \in C^\infty(X; \mathcal{E}^{(k)})$. That the Ricci tensor is parallel implies, by the (second) Bianchi identity, $\sum_\ell \nabla_{e_\ell} \mathrm{R}_{ij}{}^\ell{}_k = 0$. This observation is repeatedly used. Also, the Ricci endomorphism may be written as $\sum_{i,j} \mathrm{Ric}_i^j\, e^i \otimes e_j$ with $\mathrm{Ric}_i^j = \sum_{k,\ell} g^{k\ell}(\nabla_{e_i}\Gamma_{k\ell}^j - \nabla_{e_k}\Gamma_{\ell i}^j)$.

Consider $[\delta, \nabla^*\nabla]$. Calculating simply $\delta\nabla^*\nabla$ gives

$$\begin{aligned} \delta\nabla^*\nabla &= -\sum_k e^k \lrcorner\, \nabla_{e_k}\Big(-\operatorname{tr}\sum_{i,j} e^i \otimes \nabla_{e_i}(e^j \otimes \nabla_{e_j})\Big) \\ &= \sum_{i,j,k} g^{ij} e^k \lrcorner\, \nabla_{e_k}\nabla_{e_i}\nabla_{e_j} - \sum_{i,j,k,\ell} g^{i\ell}(\nabla_{e_k}\Gamma_{i\ell}^j)\, e^k \lrcorner\, \nabla_{e_j}, \end{aligned}$$

with a similar calculation for $\nabla^*\nabla\delta$. Combining these results and commuting ∇_{e_k} with $\nabla_{e_i}\nabla_{e_j}$ gives

$$[\delta,\, \nabla^*\nabla] = \sum_{i,j,k} g^{ij} e^k \lrcorner [\nabla_{e_k},\, \nabla_{e_i}\nabla_{e_j}] - \sum_i (\operatorname{Ric} e^i) \lrcorner \nabla_{e_i}$$
$$= -\sum_{i,j,k} g^{ij} e^k \lrcorner \{\nabla_{e_i},\, \mathrm{R}_{e_j,e_k}\} - \sum_i (\operatorname{Ric} e^i) \lrcorner \nabla_{e_i},$$

where $\{\cdot\,,\cdot\}$ is the anticommutator. After a tedious calculation, we obtain

$$[\delta,\, \nabla^*\nabla]u = \sum_i (\operatorname{Ric} e^i) \lrcorner \nabla_{e_i} u + 2(\mathrm{R},\, \nabla,\, u), \tag{9}$$

where $(\mathrm{R},\, \nabla,\, u)$ is shorthand for the unwieldy term

$$(\mathrm{R},\, \nabla,\, u) = \sum_{i,j} \sum_{k_r \in K} \sum_{k_p \in \{k_r \to\}K} \mathrm{R}^{i k_r k_p}{}_j (\nabla_{e_i} u_K) e^{\{k_p \to j,\, k_r \to\}K}.$$

For completeness we outline this calculation:

$$-\sum_{i,j,\ell} g^{ij} e^\ell \lrcorner \{\nabla_{e_i},\, \mathrm{R}_{e_j,e_\ell}\} u = -\sum_{i,j,\ell} \sum_{k_r \in K} (\{\nabla_{e_i},\, \mathrm{R}_\ell{}^{i k_r}{}_j\} u_K) e^\ell \lrcorner e^{\{k_r \to j\}K}$$
$$= -2 \sum_{i,j,\ell} \sum_{k_r \in K} \mathrm{R}_\ell{}^{i k_r}{}_j (\nabla_{e_i} u_K) e^\ell \lrcorner e^{\{k_r \to j\}K},$$

where the anticommutator has been removed using $\sum_\ell \nabla_{e_\ell} \mathrm{R}_{ij}{}^\ell{}_k = 0$. Developing the final term in the preceding display gives

$$e^\ell \lrcorner e^{\{k_r \to j\}K} = g^{j\ell} e^{\{k_r \to\}K} + \sum_{k_p \in \{k_r \to\}K} g^{k_p \ell} e^{\{k_p \to j,\, k_r \to\}K},$$

which after a little rearrangement of dummy indices and using the algebraic symmetries of the Riemann curvature tensor gives

$$-\sum_{i,j,\ell} g^{ij} e^\ell \lrcorner \{\nabla_{e_i},\, \mathrm{R}_{e_j,e_\ell}\} u = 2 \sum_i (\operatorname{Ric} e^i) \lrcorner \nabla_{e_i} u + 2(\mathrm{R},\, \nabla,\, u).$$

Upon subtraction of $\sum_i (\operatorname{Ric} e^i) \lrcorner \nabla_{e_i} u$, this provides (9).

Consider $[\delta,\, q(\mathrm{R})]$. Similar to the previous calculations we obtain

$$[\delta,\, q(\mathrm{R})] = \sum_{i,j,k} -e^k \lrcorner e^j \cdot e^i \lrcorner \nabla_{e_k} \mathrm{R}_{e_i,e_j} + e^j \cdot e^i \lrcorner \mathrm{R}_{e_i,e_j} (e^k \lrcorner \nabla_{e_k})$$
$$= \sum_{i,j,k} e^j \cdot e^i \lrcorner e^k \lrcorner [\mathrm{R}_{e_i,e_j},\, \nabla_{e_k}] - g^{jk} e^i \lrcorner \nabla_{e_k} \mathrm{R}_{e_i,e_j} + e^j \cdot e^i \lrcorner (\mathrm{R}_{e_i,e_j} e^k) \lrcorner \nabla_{e_k}.$$

After an even more tedious calculation treating each of the three terms in the previous display, we obtain

$$[\delta,\, q(\mathrm{R})]u = -[\delta,\, \nabla^*\nabla]u - (\nabla,\, \mathrm{R},\, u), \tag{10}$$

where (∇, R, u) represents the even more unwieldy term

$$(\nabla, \mathrm{R}, u) = \sum_{i,j,\ell} \sum_{\substack{k_r \in K \\ k_p \in \{k_r \to\} K \\ k_s \in \{k_p \to, \, k_r \to\} K}} g^{\ell k_s} (\nabla_{e_\ell} \mathrm{R}_i{}^{k_r k_p}{}_j) u_K e^{\{k_s \to i, \, k_p \to j, \, k_r \to\} K}.$$

Again, we sketch the calculation. One of the three terms is easy to calculate directly, giving

$$\sum_{i,j,k} e^j \cdot e^i \lrcorner (\mathrm{R}_{e_i, e_j} e^k) \lrcorner \nabla_{e_k} u = -(\mathrm{R}, \nabla, u).$$

Another term is also relatively easy, again using the trick that $\sum_\ell \nabla_{e_\ell} \mathrm{R}_{ij}{}^\ell{}_k = 0$:

$$-\sum_{i,j,k} g^{jk} e^i \lrcorner \nabla_{e_k} \mathrm{R}_{e_i, e_j} u = -\sum_i (\operatorname{Ric} e^i) \lrcorner \nabla_{e_i} u - (\mathrm{R}, \nabla, u).$$

The involved step is treating $\sum_{i,j,k} e^j \cdot e^i \lrcorner e^k \lrcorner [\mathrm{R}_{e_i, e_j}, \nabla_{e_k}]$. We first obtain

$$\sum_{i,j,\ell} e^j \cdot e^i \lrcorner e^\ell \lrcorner [\mathrm{R}_{e_i, e_j}, \nabla_{e_\ell}] u = \sum_{i,k,\ell,m} \sum_{k_r \in K} ([\mathrm{R}_{ji}{}^{k_r}{}_m, \nabla_{e_\ell}] u_K) \, e^j \cdot e^i \lrcorner e^\ell \lrcorner e^{\{k_r \to m\} K},$$

and it is important to realise that whenever the index ℓ contracts with m (or i or j), the resulting sum vanishes $\big($as $\sum_\ell \nabla_{e_\ell} \mathrm{R}_{ij}{}^\ell{}_k = 0\big)$. Similarly, if i and m are contracted then, as Ricci is parallel, the resulting sum vanishes. Expanding the final part of the previous display (and letting terms$(g^{\ell m}, g^{im})$ denote any terms involving $g^{\ell m}$ or g^{im}) gives

$$\begin{aligned} e^j \cdot e^i \lrcorner e^\ell \lrcorner e^{\{k_r \to m\} K} &= \sum_{k_p \in \{k_r \to\} K} g^{\ell k_p} e^j \cdot e^i \lrcorner e^{\{k_p \to m, \, k_r \to\} K} + \operatorname{terms}(g^{\ell m}) \\ &= \sum_{\substack{k_p \in \{k_r \to\} K \\ k_s \in \{k_p \to, \, k_r \to\} K}} g^{\ell k_p} g^{i k_s} e^{\{k_s \to j, \, k_p \to m, \, k_r \to\} K} + \operatorname{terms}(g^{\ell m}, g^{im}) \end{aligned}$$

and after a little rearrangement of dummy indices, this gives

$$\sum_{i,j,\ell} e^j \cdot e^i \lrcorner e^\ell \lrcorner [\mathrm{R}_{e_i, e_j}, \nabla_{e_\ell}] u = -(\nabla, \mathrm{R}, u),$$

whence (10) is obtained.

Combining (9) with (10) gives $[\delta, \Delta] u = -(\nabla, \mathrm{R}, u)$. For symmetric tensors of rank 2, such a summation (over k_r, k_p, k_s) does not arrive, so such a term instantly vanishes and the result follows. For tensors of higher rank, one needs the Riemann curvature to be parallel. This is assured in the constant curvature setting of $\mathbb{H}^{n+1}$. □

The objects thus far introduced in this section all have natural analogues in the Lorentzian setting on (M, η). We denote by ${}^M\nabla$ the Levi-Civita connection of η extended to all associated vector bundles and ${}^M\mathrm{R}$ the Riemann curvature tensor of η. We let d_η and δ_η denote the symmetric differential and the divergence with respect to η. Finally we let ${}^M\nabla^* {}^M\nabla$ denote the rough d'Alembertian and $\square$ the (Lichnerowicz) d'Alembertian, both constructed with respect to the metric η.

5A. ***Minkowski scale and the operator $\boldsymbol{Q}$.*** We define the first of our two main operators.

Definition 5.3. The second-order differential operator $\boldsymbol{Q} \in \mathrm{Diff}^2(M; \operatorname{End} \mathcal{F})$ is the following conjugation of the d'Alembertian:

$$\begin{aligned}\boldsymbol{Q} : C^\infty(M; \mathcal{F}) &\to C^\infty(M; \mathcal{F}),\\ u &\mapsto s^{n/2-m+2} \Box s^{-n/2+m} u.\end{aligned}$$

Lemma 5.4. *The differential operator $\boldsymbol{Q}$ is formally self-adjoint with respect to the inner product*

$$(u, v)_{s^{-2}\eta} = \int_M \langle u, v\rangle_{s^{-2}\eta} \frac{ds}{s}\, d\mathrm{vol}_g, \quad u, v \in C_c^\infty(M; \mathcal{F}).$$

Proof. The d'Alembertian is self-adjoint with respect to the inner product

$$(u, v)_\eta = \int_M \langle u, v\rangle_\eta \, d\mathrm{vol}_\eta, \quad u, v \in C_c^\infty(M; \mathcal{F}).$$

The two inner products on $\mathcal{F}$ are related via (4). Tracking the effects of the conjugations by powers of s on $\Box$, as well as the multiplication by s^2, in order to obtain $\boldsymbol{Q}$ implies self-adjointness when using the inner product $\langle\,\cdot\,,\cdot\,\rangle_{s^{-2}\eta}$ with the measure $s^{-(n+2)}\, d\mathrm{vol}_\eta$, which gives the result as $d\mathrm{vol}_\eta = s^{n+2}(ds/s)\, d\mathrm{vol}_g$. □

Lemma 5.5. *The operator $\boldsymbol{Q}$ commutes with the Lefschetz-type trace operator $s^{-2}\Lambda_{s^{-2}\eta}$:*

$$[s^{-2}\Lambda_{s^{-2}\eta}, \boldsymbol{Q}]u = 0, \quad u \in C^\infty(M; \mathcal{F}).$$

Proof. The Lorentzian analogue of Lemma 5.2 is that the d'Alembertian commutes with Λ_η:

$$[\Lambda_\eta, \Box] = 0.$$

This operator is related to our standard Lefschetz-type operator $\Lambda_{s^{-2}\eta}$ via (3). The result is now a direct calculation. For clarity we denote differential operators with a superscript (m) to indicate that they act on symmetric cotensors of rank m. In particular, on $C^\infty(M; \mathcal{F})$ we have

$$\begin{aligned}s^{-2}\Lambda_{s^{-2}\eta}\boldsymbol{Q}^{(m)} &= s^2\Lambda_\eta s^{n/2-m+2}\Box^{(m)} s^{-n/2+m} = s^2 s^{n/2-m+2}\Box^{(m-2)} s^{-n/2+m}\Lambda_\eta\\ &= s^{n/2-(m-2)+2}\Box^{(m-2)} s^{-n/2+(m-2)} s^2\Lambda_\eta = \boldsymbol{Q}^{(m-2)} s^{-2}\Lambda_{s^{-2}\eta}.\end{aligned}$$
□

The rest of this subsection is dedicated to proving:

Proposition 5.6. *For $u \in C^\infty(M; \mathcal{F})$ decomposed relative to the Minkowski scale (2), the conjugated d'Alembertian $\boldsymbol{Q}$ is given by*

$$\begin{aligned}\boldsymbol{Q} a_k \left(\frac{ds}{s}\right)^{m-k} \cdot u^{(k)} \quad = \quad & a_{k+2}\left(\frac{ds}{s}\right)^{m-k-2} \cdot (-b_k b_{k+1}\mathrm{L})u^{(k)}\\ &+ a_{k+1}\left(\frac{ds}{s}\right)^{m-k-1} \cdot (2b_k \mathrm{d})u^{(k)}\\ &+ a_k \left(\frac{ds}{s}\right)^{m-k} \cdot (\Delta + (s\,\partial_s)^2 - c_k - \mathrm{L}\Lambda)u^{(k)}\\ &+ a_{k-1}\left(\frac{ds}{s}\right)^{m-k+1} \cdot (-2b_{k-1}\delta)u^{(k)}\\ &+ a_{k-2}\left(\frac{ds}{s}\right)^{m-k+2} \cdot (-b_{k-2}b_{k-1}\Lambda)u^{(k)}\end{aligned}$$

with constants

$$\begin{aligned} a_k &= ((m-k)!)^{-1/2}, \\ b_k &= \sqrt{m-k}, \\ c_k &= \tfrac{1}{4}n^2 + m(n+2k+1) - k(2n+3k-1). \end{aligned}$$

Consequently, relative to this scale, there exist $\boldsymbol{D} \in \operatorname{Diff}^1(M; \operatorname{End}\mathcal{F})$ *and* $\boldsymbol{G} \in C^\infty(M; \operatorname{End}\mathcal{F})$ *independent of* s *such that*

$$\boldsymbol{Q} = \nabla^*\nabla + (s\,\partial_s)^2 + \boldsymbol{D} + \boldsymbol{G}.$$

Proof. The result will follow from Lemmas 5.8 and 5.9. The conjugation by $s^{-n/2+m}$ is chosen so that the term $\left(s\,\partial_s + \frac{n}{2} - m\right)^2$ in Lemma 5.8 becomes simply $(s\,\partial_s)^2$. □

Proposition 5.6 is a direct calculation which we present in the rest of this subsection. To begin we state the following lemma whose proof need not be detailed.

Lemma 5.7. *In the Minkowski scale, with* $\{e_i\}_{i=0}^n$ *a local holonomic frame on* (X, g) *with dual frame* $\{e^i\}_{i=0}^n$ *such that* $g = \sum_{i,j} g_{ij} e^i \otimes e^j$, *the connection* ${}^M\nabla$ *acts in the following manner*:

$$\begin{aligned} {}^M\nabla_{s\,\partial_s} \frac{ds}{s} &= -\frac{ds}{s}, & {}^M\nabla_{e_i} \frac{ds}{s} &= -\sum_{j=0}^n g_{ij} e^j, \\ {}^M\nabla_{s\,\partial_s} e^i &= -e^i, & {}^M\nabla_{e_i} e^j &= \delta_i^j \frac{ds}{s} + \nabla_{e_i} e^j. \end{aligned}$$

This lemma provides the following two important formulae for the symmetrised basis:

$$ {}^M\nabla_{s\,\partial_s} \left(\frac{ds}{s}\right)^{m-k} \cdot e^K = -m \left(\frac{ds}{s}\right)^{m-k} \cdot e^K \tag{11}$$

and

$$\begin{aligned} {}^M\nabla_{e_i} \left(\frac{ds}{s}\right)^{m-k} \cdot e^K = {} & \left(\frac{ds}{s}\right)^{m-k-1} \cdot \left(-(m-k) g_{ij} e^{\{\to j\}K}\right) \\ & + \left(\frac{ds}{s}\right)^{m-k} \cdot \Bigl(-\sum_{k_r \in K} \Gamma_{ij}^{k_r} e^{\{k_r \to j\}K}\Bigr) \\ & + \left(\frac{ds}{s}\right)^{m-k+1} \cdot \Bigl(\sum_{k_r \in K} \delta_i^{k_r} e^{\{k_r \to\}K}\Bigr), \end{aligned} \tag{12}$$

where the second result is a consequence of

$$ {}^M\nabla_{e_i} e^K = \sum_{k_r \in K} \delta_i^{k_r} \frac{ds}{s} \cdot e^{\{k_r \to\}K} + \nabla_{e_j} e^K$$

and we recall that the connection coefficients were introduced in Lemma 5.1. We split the calculation of the d'Alembertian into two calculations, treating the rough d'Alembertian separately from the curvature endomorphism.

Lemma 5.8. *For $u \in C^\infty(M; \mathcal{F})$ decomposed relative to the Minkowski scale (2), the rough d'Alembertian is given by*

$$\begin{aligned} s^{2\,M}\nabla^{*M}\nabla\, a_k\Big(\frac{ds}{s}\Big)^{m-k} \cdot u^{(k)} \quad = \quad & a_{k+2}\Big(\frac{ds}{s}\Big)^{m-k-2} \cdot (-b_k b_{k+1}\mathrm{L})u^{(k)} \\ & + a_{k+1}\Big(\frac{ds}{s}\Big)^{m-k-1} \cdot (2b_k \mathrm{d})u^{(k)} \\ & + a_k \quad \Big(\frac{ds}{s}\Big)^{m-k} \quad \cdot \big(\nabla^*\nabla + \big(s\,\partial_s + \tfrac{1}{2}n - m\big)^2 - \tilde{c}_k\big)u^{(k)} \\ & + a_{k-1}\Big(\frac{ds}{s}\Big)^{m-k+1} \cdot (-2b_{k-1}\delta)u^{(k)} \\ & + a_{k-2}\Big(\frac{ds}{s}\Big)^{m-k+2} \cdot (-b_{k-2}b_{k-1}\Lambda)u^{(k)} \end{aligned}$$

with modified constants

$$\tilde{c}_k = \tfrac{1}{4}n^2 + m(n+2k+1) - k(n+2k).$$

Proof. It suffices to consider a single term $u_K(ds/s)^{m-k} \cdot e^K$ and we will ignore the normalising constants a_k until the final step. Upon a first application of ${}^M\nabla$ we obtain a section of $\mathrm{T}^*M \otimes \mathcal{F}$:

$$\begin{aligned} {}^M\nabla u_K\Big(\frac{ds}{s}\Big)^{m-k} \cdot e^K = s\,\partial_s u_K \frac{ds}{s} \otimes \Big(\frac{ds}{s}\Big)^{m-k} \cdot e^K + u_K \frac{ds}{s} \otimes {}^M\nabla_{s\,\partial_s}\Big(\Big(\frac{ds}{s}\Big)^{m-k} \cdot e^K\Big) \\ + \sum_i e_i u_K e^i \otimes \Big(\frac{ds}{s}\Big)^{m-k} \cdot e^K + \sum_i u_K e^i \otimes {}^M\nabla_{e_i}\Big(\Big(\frac{ds}{s}\Big)^{m-k} \cdot e^K\Big). \end{aligned}$$

Using (11) and (12) to develop the terms involving ${}^M\nabla_{s\,\partial_s}$ and ${}^M\nabla_{e_i}$ we group the result in terms of symmetric powers of ds/s. In order to handle the equations we write

$$ {}^M\nabla u_K\Big(\frac{ds}{s}\Big)^{m-k} \cdot e^K = \boxed{1} + \boxed{2} + \boxed{3} + \boxed{4}, \tag{13}$$

where

$$\boxed{1} = -(m-k)\sum_{i,j} u_K g_{ij} e^i \otimes \Big(\frac{ds}{s}\Big)^{m-k-1} \cdot e^{\{\to j\}K},$$

$$\boxed{2} = (s\,\partial_s - m)u_K \frac{ds}{s} \otimes \Big(\frac{ds}{s}\Big)^{m-k} \cdot e^K,$$

$$\boxed{3} = \sum_i e_i u_K e^i \otimes \Big(\frac{ds}{s}\Big)^{m-k} \cdot e^K - \sum_{i,j} u_K e^i \otimes \Big(\frac{ds}{s}\Big)^{m-k} \cdot \sum_{k_r \in K} \Gamma^{k_r}_{ij} e^{\{k_r \to j\}K},$$

$$\boxed{4} = -\sum_i u_K e^i \otimes \Big(\frac{ds}{s}\Big)^{m-k+1} \cdot \sum_{k_r \in K} \delta^{k_r}_i e^{\{k_r \to\}K}.$$

Taking the second derivative, we calculate at a point about which $\{e_i\}$ are normal coordinates. Of course, we only need to keep track of terms which are not subsequently killed upon applying the trace tr_η (which, as the notation suggests, is the trace map from $\mathrm{T}^*M \otimes \mathrm{T}^*M \to \mathbb{R}$ built using the metric η).

$\boxed{1}$ Considering the first term in (13), applying ${}^M\nabla_{s\,\partial_s}$ provides only terms in the kernel of tr_η and applying ${}^M\nabla_{e_i}$ gives

$$\sum_\ell e^\ell \otimes {}^M\nabla_{e_\ell}\boxed{1} = -(m-k)\sum_{i,j,\ell} e_\ell u_K g_{ij} e^\ell \otimes e^i \otimes \Big(\frac{ds}{s}\Big)^{m-k-1} \cdot e^{\{\to j\}K} \\ -(m-k)\sum_{i,j,\ell} u_K g_{ij} e^\ell \otimes e^i \otimes {}^M\nabla_{e_\ell}\Big(\Big(\frac{ds}{s}\Big)^{m-k-1} \cdot e^{\{\to j\}K}\Big) + \ker \mathrm{tr}_\eta$$

and we immediately apply tr_η to get

$$-s^2(\mathrm{tr}_\eta \circ {}^M\nabla)\boxed{1} = (m-k)\Big(\frac{ds}{s}\Big)^{m-k-1} \cdot \Big(\sum_i e_i u_K e^{\{\to i\}K}\Big) + (m-k)u_K \sum_j {}^M\nabla_{e_j}\Big(\Big(\frac{ds}{s}\Big)^{m-k-1} \cdot e^{\{\to j\}K}\Big).$$

The first term of the preceding display reduces to the symmetric differential $(m-k)(ds/s)^{m-k-1} \cdot \mathrm{d}(u_K e^K)$ by Lemma 5.1. The second term of the preceding display is calculated with the aid of (12) and remembering that the connection coefficients cancel at the point of interest. Specifically

$$ {}^M\nabla_{e_j}\Big(\Big(\frac{ds}{s}\Big)^{m-k-1} \cdot e^{\{\to j\}K}\Big) = \Big(\frac{ds}{s}\Big)^{m-k-2} \cdot \Big(-\sum_i (m-k-1) g_{ij} e^{\{\to i, \to j\}K}\Big) \\ + \Big(\frac{ds}{s}\Big)^{m-k} \cdot \Big(\delta_j^j e^K + \sum_{k_r \in K} \delta_j^{k_r} e^{\{\to j, k_r \to\}K}\Big).$$

Observe that $\sum_j \sum_{k_r\in K} \delta_j^{k_r} e^{\{\to j, k_r\to\}K} = k e^K$. Using Lemma 5.1 again this time to recover L, the result is

$$-s^2(\mathrm{tr}_\eta \circ {}^M\nabla)\boxed{1} = \Big(\frac{ds}{s}\Big)^{m-k-2} \cdot \big(-(m-k)(m-k-1)\mathrm{L}\big)u_K e^K \\ + \Big(\frac{ds}{s}\Big)^{m-k-1} \cdot ((m-k)\mathrm{d})u_K e^K \\ + \Big(\frac{ds}{s}\Big)^{m-k} \cdot \big(-(m-k)(n+1+k)\big)u_K e^K.$$

$\boxed{2}$ Considering the second term in (13) is much simpler. A second application of ${}^M\nabla$ provides

$$ {}^M\nabla\boxed{2} = (s\,\partial_s - m - 1)(s\,\partial_s - m)u_K \frac{ds}{s} \otimes \frac{ds}{s} \otimes \Big(\frac{ds}{s}\Big)^{m-k} \cdot e^K \\ -(s\,\partial_s - m)u_K \sum_{i,j} g_{ij} e^i \otimes e^j \otimes \Big(\frac{ds}{s}\Big)^{m-k} \cdot e^K + \ker \mathrm{tr}_\eta$$

and the desired result is

$$-s^2(\mathrm{tr}_\eta \circ {}^M\nabla)\boxed{2} = \Big(\frac{ds}{s}\Big)^{m-k} \cdot \big((s\,\partial_s - m + n)(s\,\partial_s - m)\big)u_K e^K.$$

$\boxed{3}$ Considering the third term in (13) is somewhat similar to the first term in that ${}^M\nabla_{s\,\partial_s}$ provides only terms in the kernel of tr_η. Remembering that at the point of interest, the connection coefficients vanish, applying ${}^M\nabla_{e_j}$ gives

$$\begin{aligned}\sum_j e^j \otimes {}^M\nabla_{e_j}\boxed{3} &= \sum_{i,j} e_j e_i u_K e^j \otimes e^i \otimes \Big(\frac{ds}{s}\Big)^{m-k} \cdot e^K \\ &\quad + \sum_{i,j} e_i u_K e^j \otimes e^i \otimes {}^M\nabla_{e_j}\Big(\Big(\frac{ds}{s}\Big)^{m-k} \cdot e^K\Big) \\ &\quad - \sum_{i,j,\ell} u_K e^\ell \otimes e^i \otimes \Big(\frac{ds}{s}\Big)^{m-k} \cdot \sum_{k_r\in K} (\nabla_{e_\ell}\Gamma^{k_r}_{ij}) e^{\{k_r\to j\}K} + \ker \operatorname{tr}_\eta,\end{aligned}$$

and we immediately apply tr_η to recover the rough Laplacian from the first and third terms in the previous display

$$-s^2(\operatorname{tr}_\eta \circ {}^M\nabla)\boxed{3} = \Big(\frac{ds}{s}\Big)^{m-k} \cdot \nabla^*\nabla(u_K e^K) - \sum_{i,j} g^{ij} e_i u_K {}^M\nabla_{e_j}\Big(\Big(\frac{ds}{s}\Big)^{m-k} \cdot e^K\Big),$$

while the second term in the previous display is first treated using (12) and then Lemma 5.1 to recover the symmetric differential and the divergence:

$$\begin{aligned}&-\sum_{i,j} g^{ij} e_i u_K {}^M\nabla_{e_j}\Big(\Big(\frac{ds}{s}\Big)^{m-k} \cdot e^K\Big) \\ &\qquad = \sum_{i,j,\ell} g^{ij} e_i u_K (m-k)\Big(\frac{ds}{s}\Big)^{m-k-1} \cdot g_{\ell j} e^{\{\to\ell\}K} + \sum_{i,j,\ell} g^{ij} e_i u_K \Big(\frac{ds}{s}\Big)^{m-k+1} \cdot \sum_{k_r\in K} \delta^{k_r}_j e^{\{k_r\to\}K} \\ &\qquad = (m-k)\Big(\frac{ds}{s}\Big)^{m-k-1} \cdot \mathrm{d}(u_K e^K) \quad - \Big(\frac{ds}{s}\Big)^{m-k+1} \cdot \delta(u_K e^K).\end{aligned}$$

The result is

$$\begin{aligned}-s^2(\operatorname{tr}_\eta \circ {}^M\nabla)\boxed{3} &= \Big(\frac{ds}{s}\Big)^{m-k-1} \cdot ((m-k)\mathrm{d}) u_K e^K \\ &\quad + \Big(\frac{ds}{s}\Big)^{m-k} \cdot (\nabla^*\nabla) u_K e^K \\ &\quad + \Big(\frac{ds}{s}\Big)^{m-k+1} \cdot (-\delta) u_K e^K.\end{aligned}$$

$\boxed{4}$ Considering finally the fourth term in (13) we immediately remove the sum over i using the Kronecker delta. Again ${}^M\nabla_{s\,\partial_s}$ provides only terms in the kernel of tr_η and applying ${}^M\nabla_{e_i}$ gives

$$\begin{aligned}\sum_i e^i \otimes {}^M\nabla_{e_i}\boxed{4} &= -\sum_i \sum_{k_r\in K} e_i u_K e^i \otimes e^{k_r} \otimes \Big(\frac{ds}{s}\Big)^{m-k+1} \cdot e^{\{k_r\to\}K} \\ &\quad - \sum_i \sum_{k_r\in K} u_K e^i \otimes e^{k_r} \otimes {}^M\nabla_{e_i}\Big(\Big(\frac{ds}{s}\Big)^{m-k+1} \cdot e^{\{k_r\to\}K}\Big) + \ker \operatorname{tr}_\eta\end{aligned}$$

and we immediately apply tr_η to get

$$-s^2(\operatorname{tr}_\eta \circ {}^M\nabla)\boxed{4} = \Big(\frac{ds}{s}\Big)^{m-k+1} \cdot \Big(\sum_i \sum_{k_r\in K} g^{ik_r} e_i u_K e^{\{k_r\to\}K}\Big) + \sum_i \sum_{k_r\in K} g^{ik_r} u_K {}^M\nabla_{e_i}\Big(\Big(\frac{ds}{s}\Big)^{m-k+1} \cdot e^{\{k_r\to\}K}\Big).$$

The first term provides the divergence $-(ds/s)^{m-k+1}\cdot\delta(u_K e^K)$, while the second term is treated using (12) and then Lemma 5.1 to recover a multiple of $u_K e^K$ and a term involving Λ:

$$\begin{aligned}
&\sum_i \sum_{k_r\in K} g^{ik_r} u_K {}^M\nabla_{e_i}\Big(\Big(\frac{ds}{s}\Big)^{m-k+1}\cdot e^{\{k_r\to\}K}\Big)\\
&\quad = -(m-k+1)\Big(\frac{ds}{s}\Big)^{m-k}\cdot\sum_{i,j}\sum_{k_r\in K} g^{ik_r} g_{ij} e^{\{\to j,\,k_r\to\}K} - \Big(\frac{ds}{s}\Big)^{m-k+2}\cdot\sum_{k_r\in K}\ \sum_{k_p\in\{k_r\to\}K} g^{k_r k_p} e^{\{k_p\to,\,k_r\to\}K}\\
&\quad = -k(m-k+1)\Big(\frac{ds}{s}\Big)^{m-k}\cdot u_K e^K - \Big(\frac{ds}{s}\Big)^{m-k+2}\cdot\Lambda(u_K e^K).
\end{aligned}$$

The result is

$$\begin{aligned}
-s^2(\mathrm{tr}_\eta\circ{}^M\nabla)\boxed{4} &= \Big(\frac{ds}{s}\Big)^{m-k}\cdot(-k(m-k+1))u_K e^K\\
&\quad + \Big(\frac{ds}{s}\Big)^{m-k+1}\cdot(-\delta)u_K e^K\\
&\quad + \Big(\frac{ds}{s}\Big)^{m-k+2}\cdot(-\Lambda)u_K e^K.
\end{aligned}$$

Upon summation of these four terms coming from (13) we obtain

$$\begin{aligned}
s^{2\,M}\nabla^{*M}\nabla\Big(\frac{ds}{s}\Big)^{m-k}\cdot u^{(k)} &= \Big(\frac{ds}{s}\Big)^{m-k-2}\cdot(-(m-k)(m-k-1)\mathrm{L})u^{(k)}\\
&\quad + \Big(\frac{ds}{s}\Big)^{m-k-1}\cdot(2(m-k)\mathrm{d})u^{(k)}\\
&\quad + \Big(\frac{ds}{s}\Big)^{m-k}\cdot\big(\nabla^*\nabla + \big(s\,\partial_s + \tfrac{1}{2}n - m\big)^2 - \tilde{c}_k\big)u^{(k)}\\
&\quad + \Big(\frac{ds}{s}\Big)^{m-k+1}\cdot(-2k\delta)u^{(k)}\\
&\quad + \Big(\frac{ds}{s}\Big)^{m-k+2}\cdot(-k(k-1)\Lambda)u^{(k)}
\end{aligned}$$

with constant $\tilde{c}_k$ as given in the proposition. The final step is to reintroduce the normalisation constants a_k. Treating, for example, the term containing $(ds/s)^{m-k-1}$ amounts to observing

$$a_{k+1}^{-1}(m-k)a_k = \sqrt{(m-k)} \qquad \square$$

Lemma 5.9. *For $u\in C^\infty(M;\mathcal{F})$ decomposed relative to the Minkowski scale (2), the curvature endomorphism acts diagonally with respect to the Minkowski scale and is given by*

$$s^2 q({}^M\mathrm{R})u^{(k)} = (q(\mathrm{R}) + k(n+k-1) - \mathrm{L}\Lambda)u^{(k)}.$$

Proof. We need only concern ourselves with the effect of $q({}^M\mathrm{R})$ on $(ds/s)^{m-k}\cdot e^K$. It is easy to see from Lemma 5.7 that ${}^M\mathrm{R}_{s\,\partial_s,e_i}$ is the zero endomorphism, that ${}^M\mathrm{R}_{e_i,e_j}(ds/s)=0$, and that $\eta({}^M\mathrm{R}_{e_i,e_j}e_k^*, ds/s)=0$. Therefore we need only calculate the effect of $q({}^M\mathrm{R})$ on e^K. The nontrivial information of ${}^M\mathrm{R}$ is encoded in the equation

$${}^M\mathrm{R}_{ij}{}^k{}_\ell = g_{j\ell}\delta_i^k - g_{i\ell}\delta_j^k + \mathrm{R}_\ell{}^{ik_r}{}_j.$$

We extend ${}^M\mathrm{R}_{e_i,e_j}$ to $\mathcal{E}^{(k)}$, giving

$$ {}^M\mathrm{R}_{e_i,e_j} e^K = \mathrm{R}_{e_i,e_j} e^K + \sum_{k_r \in K} (\delta_j^{k_r} g_{i\ell} - \delta_i^{k_r} g_{j\ell}) e^{\{k_r \to \ell\}K}. $$

Calculating the interior product requires the metric; in particular,

$$ s^2 e^i \lrcorner_\eta {}^M\mathrm{R}_{e_i,e_j} = e^i \lrcorner {}^M\mathrm{R}_{e_i,e_j}, $$

where $\lrcorner_\eta$ uses the metric η to identify $\mathrm{T}M$ with T^*M. Consequently calculating

$$ \sum_i (s^2 e^i \lrcorner_\eta {}^M\mathrm{R}_{e_i,e_j} e^K - e^i \lrcorner \mathrm{R}_{e_i,e_j} e^K) $$

gives

$$ \sum_i \sum_{k_r \in K} (\delta_j^{k_r} g_{i\ell} - \delta_i^{k_r} g_{j\ell}) \Big(g^{i\ell} e^{\{k_r \to\}K} + \sum_{k_p \in \{k_r \to\}K} g^{ik_p} e^{\{k_p \to,\, k_r \to \ell\}K} \Big). $$

Applying $\sum_j e^j \cdot$ to the preceding display provides $s^2 q({}^M\mathrm{R}) - q(\mathrm{R})$. Splitting the calculation into four terms, the results are

$$ \begin{aligned} \sum_{i,j} \sum_{k_r \in K} e^j \cdot \delta_j^{k_r} g_{i\ell} g^{i\ell} e^{\{k_r \to\}K} &= k(n+1) e^K, \\ -\sum_{i,j} \sum_{k_r \in K} e^j \cdot \delta_i^{k_r} g_{j\ell} g^{i\ell} e^{\{k_r \to\}K} &= -k e^K, \\ \sum_{i,j} \sum_{k_r \in K} \sum_{k_p \in \{k_r \to\}K} e^j \cdot \delta_j^{k_r} g_{i\ell} g^{ik_p} e^{\{k_p \to,\, k_r \to \ell\}K} &= k(k-1) e^K, \\ -\sum_{i,j} \sum_{k_r \in K} \sum_{k_p \in \{k_r \to\}K} e^j \cdot \delta_i^{k_r} g_{j\ell} g^{ik_p} e^{\{k_p \to,\, k_r \to \ell\}K} &= -\mathrm{L}\Lambda e^K. \end{aligned} $$

Upon summation of these four terms, the proof is complete. □

Proposition 5.10. *Suppose $u \in C^\infty(M; \mathcal{F})$, decomposed relative to the Minkowski scale (2), is trace-free with respect to the trace operator $\Lambda_{s^{-2}\eta}$. Then the conjugated d'Alembertian $\boldsymbol{Q}$ is given by*

$$ \begin{aligned} \boldsymbol{Q} a_k \left(\frac{ds}{s}\right)^{m-k} \cdot u^{(k)} = {} & a_{k+1} \left(\frac{ds}{s}\right)^{m-k-1} \cdot (2b_k \mathrm{d}) u^{(k)} \\ & + a_k \left(\frac{ds}{s}\right)^{m-k} \cdot (\Delta + (s\,\partial_s)^2 - c_k')u^{(k)} \\ & + a_{k-1} \left(\frac{ds}{s}\right)^{m-k+1} \cdot (-2b_{k-1}\delta) u^{(k)} \end{aligned} $$

with constants a_k, b_k given in Proposition 5.6 and the modified constants

$$ c_k' = c_k - (m-k)(m-k-1). $$

Proof. This follows directly from the structure of $\boldsymbol{Q}$ given in Proposition 5.6 and the condition that $\Lambda u^{(k)} = -b_{k-2} b_{k-1} u^{(k-2)}$, coming from Lemma 3.1. □

5B. ***The indicial family of Q.***

Definition 5.11. Denote by $\mathcal{Q}$ the indicial family of the operator $\boldsymbol{Q} \in \mathrm{Diff}_{\mathrm{b}}^2(\overline{M}; \mathcal{F})$ relative to the Minkowski scale s:

$$\mathcal{Q} = \mathrm{I}_s(\boldsymbol{Q}; \lambda) \in \mathrm{Diff}^2(X; \mathcal{E}).$$

The previous section introduced $\boldsymbol{Q}$ as a differential operator on $\mathcal{F}$ above M; however, from the structure of $\boldsymbol{Q}$ given as in Proposition 5.6, it is clear that the operator extends to $\overline{M}$. Moreover by the same proposition we immediately get the structure of $\mathcal{Q}$.

Proposition 5.12. *For $u = \sum_{k=0}^m u^{(k)} \in C^\infty(X; \mathcal{E})$, the operator $\mathcal{Q}$ is given by*

$$\mathcal{Q}_\lambda u^{(k)} = (-b_k b_{k+1} \mathrm{L}) u^{(k)} + (2b_k \mathrm{d}) u^{(k)} + (\Delta + \lambda^2 - c_k - \mathrm{L}\Lambda) u^{(k)} + (-2b_{k-1}\delta) u^{(k)} + (-b_{k-2} b_{k-1} \Lambda) u^{(k)}$$

with constants

$$b_k = \sqrt{m-k},$$
$$c_k = \tfrac{1}{4} n^2 + m(n + 2k + 1) - k(2n + 3k - 1).$$

Consequently, there exist $\mathcal{D} \in \mathrm{Diff}^1(X; \mathrm{End}\, \mathcal{E})$ and $\mathcal{G} \in C^\infty(X; \mathrm{End}\, \mathcal{E})$ independent of λ such that

$$\mathcal{Q}_\lambda = \nabla^* \nabla + \lambda^2 + \mathcal{D} + \mathcal{G}.$$

Proposition 5.13. *The family of differential operators $\mathcal{Q}$ is, upon restriction to $\lambda \in i\mathbb{R}$, a family of formally self-adjoint operators with respect to the inner product*

$$(u, v)_s = \int_X \sum_{k=0}^m (-1)^{m-k} \langle u^{(k)}, v^{(k)} \rangle \, d\mathrm{vol}_g,$$

where $u = \sum_{k=0}^m u^{(k)}$, $v = \sum_{k=0}^m v^{(k)}$ for $u^{(k)}, v^{(k)} \in C_c^\infty(X; \mathcal{E}^{(k)})$. Moreover, for all λ, we have $\mathcal{Q}_\lambda^* = \mathcal{Q}_{-\bar{\lambda}}$.

Proof. This follows from Lemmas 4.4 and 5.4. □

The operator $\boldsymbol{Q}$ preserves the subbundle $\mathcal{F} \cap \ker \Lambda_{s^{-2}\eta}$ by Lemma 5.5. As π_s^* is algebraic, we may consider it as a map from $\mathcal{E}$ over X to $\mathcal{F}$ over M. We thus obtain the subbundle $\mathcal{E} \cap \ker(\Lambda_{s^{-2}\eta} \circ \pi_s^*)$ over X, that is, symmetric tensors above X which are trace-free with respect to the ambient trace operator $\Lambda_{s^{-2}\eta}$. It thus follows that $\mathcal{Q}$ may also be considered a family of differential operators on this subbundle and we obtain:

Proposition 5.14. *For $u = \sum_{k=0}^m u^{(k)} \in C^\infty(X; \mathcal{E}) \cap \ker(\Lambda_{s^{-2}\eta} \circ \pi_s^*)$, the operator $\mathcal{Q}$ is given by*

$$(2b_k \mathrm{d}) u^{(k)} + \mathcal{Q}_\lambda u^{(k)} = (\Delta + \lambda^2 - c_k') u^{(k)} + (-2b_{k-1}\delta) u^{(k)}.$$

6. The operator P and its indicial family

This section introduces the operator $\boldsymbol{P}$ on M_e and its indicial family $\mathcal{P}$ on X_e and similar results to those presented for $\boldsymbol{Q}$ and $\mathcal{Q}$ are given. The relationship between these two constructions is also detailed.

6A. *Euclidean scale.* The manifold $M_e = \mathbb{R}_t^+ \times X_e$ has been equipped with the Lorentzian metric η which agrees with the Lorentzian cone metric put on M. Recalling the smooth chart $U = (0, 1)_\mu \times Y \subset X \subset X_e$, the metric on $\mathbb{R}_t^+ \times U$ takes the form of (1) and we may assume that this is the form of η on the larger chart $\mathbb{R}_t^+ \times U^2$, where $U^2 = (-1, 1)_\mu \times Y$. For later use we record the behaviour of ${}^{M_e}\nabla$.

Lemma 6.1. *On the chart $\mathbb{R}_t^+ \times (-1, 1)_\mu \times Y$ with $\{e_i\}_{i=1}^n$ a local holonomic frame on Y with dual frame $\{e^i\}_{i=1}^n$ such that $h = \sum_{i,j} h_{ij} e^i \otimes e^j$, the connection ${}^{M_e}\nabla$ acts in the following manner*:

$$\begin{aligned} {}^{M_e}\nabla_{t\,\partial_t} \frac{dt}{t} &= 0, & {}^{M_e}\nabla_{\partial_\mu} \frac{dt}{t} &= 0, \\ {}^{M_e}\nabla_{t\,\partial_t} d\mu &= -d\mu, & {}^{M_e}\nabla_{\partial_\mu} d\mu &= -\frac{dt}{t}, \\ {}^{M_e}\nabla_{t\,\partial_t} e^i &= -e^i, & {}^{M_e}\nabla_{\partial_\mu} e^i &= -\tfrac{1}{2} h^{ij} (\partial_\mu h_{jk}) e^k, \end{aligned}$$

and

$$\begin{aligned} {}^{M_e}\nabla_{e_i} \frac{dt}{t} &= -(\partial_\mu h_{ij}) e^j, \\ {}^{M_e}\nabla_{e_i} d\mu &= -2((1 - \mu\partial_\mu) h_{ij}) e^j, \\ {}^{M_e}\nabla_{e_i} e^j &= -\delta_i^j \frac{dt}{t} - \tfrac{1}{2} h^{jk} (\partial_\mu h_{ik}) d\mu + {}^{Y}\nabla_{e_i} e^j. \end{aligned}$$

Motivated by the structure of $\boldsymbol{Q}$ from the previous section we define the second of a our two main operators.

Definition 6.2. The second-order differential operator $\boldsymbol{P} \in \operatorname{Diff}^2(M_e; \mathcal{F})$ is the following conjugation of the d'Alembertian:

$$\boldsymbol{P} : C^\infty(M; \mathcal{F}) \to C^\infty(M_e; \mathcal{F}), \qquad u \mapsto t^{n/2-m+2} \Box t^{-n/2+m}\, u.$$

Note that on $M \subset M_e$ there is a trivial correspondence between $\boldsymbol{P}$ and $\boldsymbol{Q}$,

$$\boldsymbol{P} = \rho^{-n/2+m-2} \boldsymbol{Q} \rho^{n/2-m},$$

and that, since $\rho = 1$ on $X \backslash U$, we have equality $\boldsymbol{P} = \boldsymbol{Q}$ on $M \backslash (\mathbb{R}^+ \times U)$.

Lemma 6.3. *The operator $\boldsymbol{P} \in \operatorname{Diff}^2(M_e; \mathcal{F})$ naturally extends to an operator $\boldsymbol{P} \in \operatorname{Diff}^2_{\mathrm{b}}(\overline{M}_e; {}^{\mathrm{b}}\mathcal{F})$ and is b-trivial.*

Proof. The important point is to verify that at $\mu = 0$, $\boldsymbol{P}$ fits into the b-calculus framework. This is reasonably clear from Lemma 6.1. Indeed, the Lie algebra of b-vector fields is generated by $\{t\,\partial_t, \partial_\mu, e_i\}$, where $\{e_i\}_{i=1}^n$ is a local holonomic frame on Y, while the b-cotangent bundle has basis $\{dt/t, d\mu, e^i\}$ with $\{e^i\}_{i=1}^n$ the dual frame on T^*Y. Lemma 6.1 thus shows that ${}^{M_e}\nabla$ is a b-connection. Taking the trace using η and then multiplying by t^2 is equivalent to taking the trace with $t^{-2}\eta$, whose structure (1) indicates it is a b-metric. Therefore $t^2\, {}^{M_e}\nabla^* {}^{M_e}\nabla$ is a b-differential operator. That $t^2\, {}^{M_e}\nabla^* {}^{M_e}\nabla$ is b-trivial is also immediate from Lemma 6.1 and the structure of $t^{-2}\eta$. A similar line of reasoning for $q({}^{M_e}\mathrm{R})$ (which uses one application of the inverse of the metric η) shows that $t^2\Box$ is also a b-differential operator. The final conjugation by powers of t preserves the b-structure (and its b-triviality) as it merely conjugates appearances of $t\,\partial_t$. This implies the result. $\square$

Lemma 6.4. *The differential operator $\boldsymbol{P}$ is formally self-adjoint with respect to the inner product*

$$(u,v)_{t^{-2}\eta} = \int_{M_e} \langle u,v\rangle_{t^{-2}\eta} \frac{dt}{t} dx, \quad u,v \in C_c^\infty(M_e;\mathcal{F}).$$

Proof. By the correspondence between $\boldsymbol{P}$ and $\boldsymbol{Q}$ on $M\backslash(\mathbb{R}^+ \times U)$ and Lemma 5.4, it suffices to verify this claim when u, v are supported on $\mathbb{R}_t^+ \times U^2$. The d'Alembertian is self-adjoint with respect to the inner product

$$(u,v)_\eta = \int_{M_e} \langle u,v\rangle_\eta \, d\mathrm{vol}_\eta, \quad u,v \in C_c^\infty(\mathbb{R}_t^+ \times U^2;\mathcal{F}).$$

The two inner products on the fibres of $\mathcal{F}$ are related via the Euclidean scale analogue of (4). Tracking the effects of the conjugations by powers of t on $\Box$, as well as the multiplication by t^2, in order to obtain $\boldsymbol{P}$ implies self-adjointness when using the inner product $\langle\cdot,\cdot\rangle_{t^{-2}\eta}$ with the measure $t^{-n-2} d\mathrm{vol}_\eta$. As $\det\eta = -\frac{1}{4}t^{2n+2}\det h$, we have

$$t^{-n-2} d\mathrm{vol}_\eta = \frac{1}{2}\frac{dt}{t} d\mu \, d\mathrm{vol}_h. \qquad \Box$$

6B. *The indicial family of $\boldsymbol{P}$.*

Definition 6.5. Denote by $\mathcal{P}$ the indicial family of the operator $\boldsymbol{P} \in \mathrm{Diff}_\mathrm{b}^2(\overline{M}_e; {}^\mathrm{b}\mathcal{F})$ relative to the Euclidean scale t:

$$\mathcal{P}_\lambda = \mathrm{I}_t(\boldsymbol{P};\lambda) \in \mathrm{Diff}^2(X_e;\mathcal{E}).$$

Lemma 4.6 gives the following proposition (whose final statement follows as ρ is constant on $X\backslash U$).

Proposition 6.6. *On $X \subset X_e$ the indicial family operators $\mathcal{P}$ and $\mathcal{Q}$ are related by*

$$\mathcal{P}_\lambda = \rho^{-\lambda-n/2+m-2} J \mathcal{Q}_\lambda J^{-1} \rho^{\lambda+n/2-m}$$

with J presented in Lemma 3.2. Moreover, on $X\backslash U$, we have $\mathcal{P} = \mathcal{Q}$.

Proposition 6.7. *The family of differential operators $\mathcal{P}$ is, upon restriction to $\lambda \in i\mathbb{R}$, a family of formally self-adjoint operators with respect to the inner product*

$$(u,v)_t = \int_{X_e} \langle u,v\rangle_t \, dx, \quad u,v \in C_c^\infty(X_e;\mathcal{E}).$$

Moreover, for all λ, we have $\mathcal{P}_\lambda^* = \mathcal{P}_{-\bar{\lambda}}$.

7. Microlocal analysis

This section constructs an inverse to the family $\mathcal{P}$ introduced in the preceding section. This is done by first showing that the family is a family of Fredholm operators and then by considering a Cauchy problem which provides an inverse for $\mathrm{Re}\,\lambda \gg 1$. In [Vasy 2013b; Zworski 2016], the procedure is described for functions, rather than symmetric tensors. We are required to alter only minor details in order to apply the technique to symmetric tensors.

7A. ***Function spaces.*** From Section 3B, we have the space of L^2 sections $L^2_t(X_e;\mathcal{E})$. This defines $H^s_{\mathrm{loc}}(X_e;\mathcal{E})$, the space of (locally) H^s sections for $s\in\mathbb{R}$. For all notions of Sobolev regularity, we will only use the Euclidean scale; we thus need not decorate these spaces with a subscript t.

As is standard, we denote by $\dot{C}^\infty(X_{cs};\mathcal{E})$ the set of smooth sections which are extensible to smooth sections over X_e and whose support is contained in $\overline{X}_{cs}$, and by $C^\infty(\overline{X}_{cs};\mathcal{E})$ all smooth sections which are smoothly extensible to X_e.

Following [Hörmander 1994, Appendix B.2] we obtain, for $s\in\mathbb{R}$, the Sobolev spaces

$$\dot{H}^s(\overline{X}_{cs};\mathcal{E})\quad\text{and}\quad \overline{H}^s(X_{cs};\mathcal{E}),$$

which are, respectively, the set of elements in $H^s_{\mathrm{loc}}(X_e;\mathcal{E})$ supported by $\overline{X}_{cs}$ and the space of restrictions to X_{cs} of $H^s_{\mathrm{loc}}(X_e;\mathcal{E})$. Then $\dot{H}^s(\overline{X}_{cs};\mathcal{E})$ gets its norm directly from that of $H^s_{\mathrm{loc}}(X_e;\mathcal{E})$, while the norm of an element in $\overline{H}^s(X_{cs};\mathcal{E})$ is that obtained by taking the infimum of the norms of all permissible extensions of the element which have compact support in X_e. (Such norms will be denoted, for simplicity, by $\|\cdot\|_{\dot{H}^s}$ and $\|\cdot\|_{\overline{H}^s}$. Furthermore, if an object is supported away from S, these norms correspond and we may simply write $\|\cdot\|_{H^s}$.)

The inner product $\langle\cdot,\cdot\rangle_t$ gives the L^2 pairing

$$(\cdot,\cdot)_t:\dot{C}^\infty(X_{cs};\mathcal{E})\times C^\infty(\overline{X}_{cs};\mathcal{E})\to\mathbb{C},$$

which extends by density [Hörmander 1994, Theorem B.2.1] to a pairing between the spaces $\dot{H}^{-s}(\overline{X}_{cs};\mathcal{E})$ and $\overline{H}^s(X_{cs};\mathcal{E})$, providing the identification of dual spaces

$$(\overline{H}^s(X_{cs};\mathcal{E}))^*\simeq\dot{H}^{-s}(\overline{X}_{cs};\mathcal{E}),\qquad s\in\mathbb{R}. \tag{14}$$

Definition 7.1. For $s\in\mathbb{R}$, let $\mathcal{X}^s$ and $\mathcal{Y}^s$ be the spaces

$$\begin{aligned}\mathcal{Y}^s&=\overline{H}^s(X_{cs};\mathcal{E}),\\ \mathcal{X}^s&=\{u:u\in\mathcal{Y}^s,\ \mathcal{P}u\in\mathcal{Y}^{s-1}\}.\end{aligned}$$

These spaces come with the standard norms, in particular,

$$\|u\|_{\mathcal{X}^s}=\|u\|_{\mathcal{Y}^s}+\|\mathcal{P}u\|_{\mathcal{Y}^{s-1}},\qquad u\in\mathcal{X}^s.$$

Remark 7.2. It will be seen that λ does not appear in the principal symbol of $\mathcal{P}$; it is thus unimportant to state with respect to what value of λ the preceding norm is taken, as all such norms are equivalent.

When restricting to $U^2\subset X_e$, we will let $\{e_i\}_{i=1}^n$ denote an orthonormal frame for (Y,h), which depends on $\mu\in(-1,1)$, and by $\{e^i\}_{i=1}^n$ its dual frame. The frames are completed to frames for TU^2 and T^*U^2 by including ∂_μ and $d\mu$ respectively. A dual vector will be denoted by

$$\xi\,d\mu+\sum_{i=0}^n\eta_ie^i\in\mathrm{T}^*U^2. \tag{15}$$

The next subsection proves the following two propositions.

Proposition 7.3. *For fixed s, the family of operators*

$$\mathcal{P} : \mathcal{X}^s \to \mathcal{Y}^{s-1}$$

is Fredholm for $\operatorname{Re}\lambda > \frac{1}{2} - s$.

Proof. See Lemmas 7.6 and 7.7. □

Proposition 7.4. *For fixed s, the Fredholm operator $\mathcal{P}_\lambda : \mathcal{X}^s \to \mathcal{Y}^{s-1}$ is Fredholm of index 0 for* $\operatorname{Re}\lambda > m + \frac{1}{2} - s$ *and it has a meromorphic inverse*

$$\mathcal{P}^{-1} : \mathcal{Y}^{s-1} \to \mathcal{X}^s$$

with poles of finite rank.

Proof. See Lemmas 7.8 and 7.9. □

7B. ***Proofs of Propositions 7.3 and 7.4.*** On $\mathbb{R}^+_t \times U^2$, the inverse of the metric η takes the form

$$t^2\eta^{-1} = -2t\,\partial_t \cdot \partial_\mu + 2\mu\partial_\mu \cdot \partial_\mu + h^{-1},$$

which implies to highest order for $t^2\, {}^{M_e}\nabla^{*\,M_e}\nabla$ that

$$t^2\, {}^{M_e}\nabla^{*\,M_e}\nabla = -4\mu\partial_\mu^2 + 4t\,\partial_t\partial_\mu + \Delta_h + \operatorname{Diff}^1(\mathbb{R}^+_t \times U^2; \operatorname{End}\mathcal{F}),$$

where Δ_h may be considered the rough Laplacian on (Y, h). Considering $\mathcal{P}$, conjugation by $t^{-n/2+m}$ replaces $t\,\partial_t$ by $\left(t\,\partial_t - \frac{1}{2}n + m\right)$ and we can absorb the newly created term $4\left(-\frac{1}{2}n + m\right)\partial_\mu$ into $\operatorname{Diff}^1(\mathbb{R}^+_t \times U^2; \operatorname{End}\mathcal{F})$. Also, the curvature term is of order zero so

$$\boldsymbol{P} = -4\mu\partial_\mu^2 + 4t\,\partial_t\partial_\mu + \Delta_h + \boldsymbol{A}$$

for some $\boldsymbol{A} \in \operatorname{Diff}^1(\mathbb{R}^+_t \times U^2; \operatorname{End}\mathcal{F})$. This structure of $\boldsymbol{P}$ immediately gives the structure of $\mathcal{P}$ to highest order. Keeping track of the term $4t\,\partial_t\partial_\mu$ for the moment, we write

$$\mathcal{P}_\lambda = -4\mu\partial_\mu^2 - 4\lambda\partial_\mu + \Delta_h + \mathcal{A}_\lambda, \tag{16}$$

where $\mathcal{A}_\lambda \in \operatorname{Diff}^1(U^2; \operatorname{End}\mathcal{E})$ is the indicial family of $\boldsymbol{A}$. The most obvious conclusion we draw from such a presentation of $\mathcal{P}$ is that $\mathcal{P}$ is a family of elliptic operators on $U^2 \cap \{\mu > 0\}$ and a family of strictly hyperbolic operators for $\{\mu < 0\}$ (with respect to the level sets $\{\mu = \text{constant}\}$). Of course the ellipticity extends to all of X. The principal symbol on U^2 is also immediately recognisable as

$$\sigma(\mathcal{P}) = 4\mu\xi^2 + |\eta|^2$$

using the notation from (15) and $|\eta|^2 = \sum_{i=1}^n \eta_i^2$. And on U^2, the Hamiltonian vector field associated with $\sigma(\mathcal{P})$ is

$$H_{\sigma(\mathcal{P})} = 8\mu\xi\,\partial_\mu - 4\xi^2\partial_\xi + H_{|\eta|^2}.$$

The strategy to obtain a Fredholm problem is to combine standard results for elliptic and hyperbolic operators with some analysis performed at the junction $Y = \{\mu = 0\}$. The analysis was first presented in [Vasy 2013a, Section 4.4]. It turns out the dynamics of interest are those of radial sources and sinks [Dyatlov and Zworski 2017, Definition E.52]. The original radial estimates of Melrose [1994] on asymptotically Euclidean spaces have been adapted to functions on asymptotically hyperbolic spaces by Vasy [2013a]. Indeed, to see that such dynamics are relevant for $\mathcal{P}$, consider $\sigma(\mathcal{P})$ and $H_{\sigma(\mathcal{P})}$ given in the preceding displays. Define the characteristic variety $\Sigma \subset \mathrm{T}^*X_{cs}\backslash 0$ which is contained in T^*U. As $(\mu, y, 0, \eta) \notin \Sigma$, we may split Σ as $\Sigma = \Sigma_+ \sqcup \Sigma_-$, given by $\Sigma_\pm = \Sigma \cap \{\pm\xi > 0\}$. At Y remark that

$$\Sigma \cap \mathrm{T}^*_Y U = \{(0, y, \xi, 0) : \xi \neq 0\} \subset N^*Y$$

and, recalling the projection $\kappa : \mathrm{T}^*U\backslash 0 \to \partial\overline{\mathrm{T}}^*U$, define

$$\Gamma_+ = \kappa(\Sigma_+ \cap Y), \quad \Gamma_- = \kappa(\Sigma_- \cap Y).$$

In [Vasy 2013b, Section 3.2], it is shown that $\Gamma_\pm$ are respectively a source and a sink for $\sigma(\mathcal{P})$. In order to apply Lemmas 4.7 and 4.8, we introduce the principal symbol of the imaginary part of $\mathcal{P}$. By Remark 4.9, $H_{\sigma(\mathcal{P})} = H_{\sigma(\mathcal{P}^*)}$ and by Proposition 6.7, $\mathcal{P}^*_\lambda = \mathcal{P}_{-\bar{\lambda}}$, hence $\sigma(\operatorname{Im}\mathcal{P}) = -\sigma(\operatorname{Im}\mathcal{P}^*)$. Also, by a direct calculation using the structure of $H_{\sigma(\mathcal{P})}$,

$$\langle\xi + \eta\rangle^{-1} H_{\sigma(\mathcal{P})} \log\langle\xi + \eta\rangle = \mp 4 \quad \text{on } \Gamma_\pm. \tag{17}$$

In fact Proposition 6.7 along with (16) gives more precisely

$$\operatorname{Im}\mathcal{P}_\lambda = \frac{\mathcal{P}_\lambda - \mathcal{P}^*_\lambda}{2i} = 4i(\operatorname{Re}\lambda)\partial_\mu + \frac{\mathcal{A}_\lambda - \mathcal{A}_{-\bar{\lambda}}}{2i}.$$

However, as $\boldsymbol{A}$ is of first order, $\mathcal{A}_\lambda$ may be written as the sum of a first-order operator independent of λ and a zeroth-order operator (which may depend on λ). Therefore

$$\sigma(\operatorname{Im}\mathcal{P}_\lambda) = -4\operatorname{Re}\lambda\,\xi. \tag{18}$$

Bringing this all together in preparation for the proof of Proposition 7.3 we have:

Lemma 7.5. *For $\mathcal{P}$, we have Γ_+ is a source, while Γ_- is a source for $-\mathcal{P}$. In both situations, the threshold condition, when working on $H^s(X_{cs}; \mathcal{E})$, is satisfied if*

$$s > -\operatorname{Re}\lambda + \tfrac{1}{2}.$$

For $\mathcal{P}^*$, we have Γ_- is a sink, while Γ_+ is a sink for $-\mathcal{P}^*$. In both situations, the threshold condition, when working on $H^{\tilde{s}}(X_{cs}; \mathcal{E})$, is satisfied if

$$\tilde{s} < \operatorname{Re}\lambda + \tfrac{1}{2}.$$

Proof. We explain the first result, all others are similar after taking into account Remark 4.9. On Γ_+, by (17) and (18),

$$\langle\xi + \eta\rangle^{-1}\big(\sigma(\operatorname{Im}\mathcal{P}) + \big(s - \tfrac{1}{2}\big)H_{\sigma(\mathcal{P})}\log\langle\xi + \eta\rangle\big) = -4\big(\operatorname{Re}\lambda + s - \tfrac{1}{2}\big).$$

For this to be negative definite requires precisely that $s > -\operatorname{Re}\lambda + \frac{1}{2}$. □

Lemma 7.6. *Restricting to $s > -\operatorname{Re}\lambda + \frac{1}{2}$, the operators $\mathcal{P}_\lambda : \mathcal{X}^s \to \mathcal{Y}^{s-1}$ have finite-dimensional kernels.*

Proof. It suffices to obtain an estimate, for $u \in \mathcal{X}^s$, of the form

$$\|u\|_{\bar{H}^s} \leq C(\|\mathcal{P}_\lambda u\|_{\bar{H}^{s-1}} + \|\psi u\|_{H^{-N}})$$

for some ψ supported on $\{\mu > -\frac{1}{2}\}$ and such that $\psi = 1$ near $\{\mu > -\frac{1}{2} + \varepsilon\}$. This is done by writing $u = (\psi_- + \psi_0 + \psi_+)u$ with the supports of ψ_-, ψ_0, ψ_+ respectively contained in $\{\mu < -\varepsilon\}$, $\{|\mu| < 2\varepsilon\}$, $\{\mu > \varepsilon\}$. The estimate for $\psi_+ u$ is due to ellipticity of $\mathcal{P}$. The estimate for $\psi_- u$ is due to hyperbolicity, which allows us to reduce to the estimate for $\psi_0 u$:

$$\|\psi_- u\|_{\bar{H}^s} \leq C(\|\mathcal{P}_\lambda u\|_{\bar{H}^{s-1}} + \|\psi_0 u\|_{H^s}).$$

The estimate for $\psi_0 u$ is obtained by microlocalising. Away from Σ, ellipticity gives the result, while near Σ, propagation of singularities implies that the norms can be controlled by $\Gamma_\pm$. The high regularity results for Γ_+ and Γ_- from Lemma 4.7 are applicable as these are sources for $\mathcal{P}$ and $-\mathcal{P}$ respectively. Lemma 7.5 ensures that the threshold conditions are satisfied (by hypothesis of this proposition). The desired estimate is obtained. □

Lemma 7.7. *Restricting to $s > -\operatorname{Re}\lambda + \frac{1}{2}$, the operators $\mathcal{P}_\lambda : \mathcal{X}^s \to \mathcal{Y}^{s-1}$ have finite-dimensional cokernels.*

Proof. To show that the range is of finite codimension we study the adjoint operator $\mathcal{P}^*$. By (14) the dual space of $\bar{H}^{s-1}(X_{cs}; \mathcal{E})$ is $\dot{H}^{1-s}(\bar{X}_{cs}; \mathcal{E})$ and the dimension of the kernel of $\mathcal{P}^*$ equals the dimension of the cokernel of $\mathcal{P}$. It suffices to obtain an estimate of the form

$$v \in \dot{H}^{1-s}(\bar{X}_{cs}; \mathcal{E}) \cap \ker \mathcal{P}^* \quad \Longrightarrow \quad \|v\|_{\dot{H}^{1-s}} \leq C\|\psi v\|_{H^{-N}}$$

with ψ as defined in the previous proof. Again, we use the partition $v = (\psi_- + \psi_0 + \psi_+)v$. Again, the estimate for $\psi_+ v$ is due to ellipticity of $\mathcal{P}^*$. This time, the estimates for $\psi_- v$ are immediate due to hyperbolicity and the requirement at S that v vanish to all orders, which implies that $v = 0$ on $\{\mu < 0\}$. The estimate for $\psi_0 v$ is obtained by microlocalising. (Away from $\operatorname{Char}(P)$, the result is obtained by ellipticity.) The low regularity results for Γ_- and Γ_+ from Lemma 4.8 are applicable as these are sinks for $\mathcal{P}^*$ and $-\mathcal{P}^*$ respectively. Lemma 7.5 ensures that the threshold conditions are satisfied. Therefore there exist $A, B \in \Psi^0(X_{cs})$ with $\operatorname{Char}(A) \cap \Gamma_\pm = \varnothing$ and $\operatorname{WF}(B) \cap \Gamma_\pm = \varnothing$ such that $\|A\psi_0 v\|_{H^{1-s}} \leq C(\|B\psi_0 v\|_{H^{1-s}} + \|\psi v\|_{H^{-N}})$. As $v = 0$ on $\{\mu < 0\}$ and v is smooth (by ellipticity of $\mathcal{P}^*$) on $\{\mu > 0\}$, we have $\operatorname{WF}(B\psi_0 v) \cap \operatorname{Char}(\mathcal{P}^*) = \varnothing$ so microellipticity gives $\|B\psi_0 v\|_{H^{1-s}} \leq C\|\psi v\|_{H^{-N}}$. The desired estimate is obtained. □

Lemma 7.8. *For $\mathcal{P}_\lambda$ with $\lambda \in \mathbb{R}$ acting on $\bar{H}^s(X_{cs}; \mathcal{E})$, the kernel of $\mathcal{P}_\lambda$ is trivial for $\lambda \gg 1$.*

Proof. Consider $u \in \ker \mathcal{P}_\lambda$. By the estimate obtained in Lemma 7.6, $u \in C^\infty(\bar{X}_{cs}; \mathcal{E})$. Restricting our attention to $\{\mu > 0\}$, Proposition 6.6 gives

$$\rho^{-\lambda-n/2+m-2} J \mathcal{Q}_\lambda J^{-1} \rho^{\lambda+n/2-m} u = 0,$$

so defining $\tilde{u} = J^{-1}\rho^{\lambda+n/2-m}u$, we get $\mathcal{Q}_\lambda \tilde{u} = 0$, or by Proposition 5.12,

$$(\nabla^*\nabla + \lambda^2 + \mathcal{D} + \mathcal{G})\tilde{u} = 0.$$

Now $\mathcal{D}$ may be bounded (up to a constant) by ∇ (and $\mathcal{G}$ by a constant as the curvature is bounded on X) so we can find C independent of λ such that

$$|(\mathcal{Q}_\lambda \tilde{u}, \tilde{u})_s| \geq C^{-1}\|\nabla \tilde{u}\|_s^2 + (\lambda^2 - C)\|\tilde{u}\|_s^2,$$

and taking $\lambda \gg \sqrt{C}$ shows $\tilde{u} = 0$ on $\{\rho > 0\}$. By smoothness, u vanishes on $\{\mu \geq 0\}$ (and so too do all its derivatives on Y). Standard hyperbolic estimates give the desired result $u = 0$ if we can show a type of unique continuation result that $u = 0$ on $\{\mu > -\varepsilon\}$.

To this end we work on U^2 and consider $\boldsymbol{P}$ written in the form

$$\mathcal{P}_\lambda = -\mu\partial_\mu^2 + \Delta_h + \mathcal{B}_\lambda$$

for $\mathcal{B}_\lambda = -4\lambda\partial_\mu + \mathcal{A}_\lambda \in \mathrm{Diff}^1(U^2; \mathrm{End}\,\mathcal{E})$. Let $\langle\,\cdot\,,\,\cdot\,\rangle_{h,t}$ on $\mathrm{T}^*Y \otimes \mathcal{E}$ denote the coupling of the metrics h on T^*Y with $\langle\,\cdot\,,\,\cdot\,\rangle_t$ on $\mathcal{E}$. For ease of presentation, we will assume throughout this proof that all objects are real-valued. Consider $u, v \in C_c^\infty(U^2, \mathcal{E})$ (and we may assume supp $u \subset (-1, 0] \times Y$). Then we have the formula

$$\langle {}^Y\nabla u, {}^Y\nabla v\rangle_{h,t} = \langle \Delta_h u, v\rangle_t + \mathrm{div},$$

where div denotes any term which is of divergence nature on Y, and hence vanishes upon integrating over Y (using $d\mathrm{vol}_h$). Indeed such an equation is obtained by considering $f \in C^\infty(Y)$ and calculating, at some value μ,

$$\begin{aligned}\int_Y \langle {}^Y\nabla u, {}^Y\nabla v\rangle_{h,t} f\, d\mathrm{vol}_h &= \int_Y \langle {}^Y\nabla u, {}^Y\nabla(fv)\rangle_{h,t} - \langle {}^Y\nabla u, {}^Y\nabla f \otimes v\rangle_{h,t}\, d\mathrm{vol}_h \\ &= \int_Y (\langle \Delta_h u, v\rangle_t + \mathrm{div}) f\, d\mathrm{vol}_h,\end{aligned}$$

where the second term was dealt with in the following way:

$$\begin{aligned}\int_Y \langle {}^Y\nabla u, {}^Y\nabla f \otimes v\rangle_{h,t}\, d\mathrm{vol}_h &= \int_Y \sum_i \langle {}^Y\nabla_{e_i} u, v\rangle_t\, \mathrm{tr}_h(e^i \otimes {}^Y\nabla f)\, d\mathrm{vol}_h \\ &= \int_Y {}^Y\nabla^*\Big(\sum_i \langle {}^Y\nabla_{e_i} u, v\rangle_t e^i\Big) f\, d\mathrm{vol}_h.\end{aligned}$$

With this formula established we define, for given u,

$$\mathcal{H}(\mu) = |\mu|\langle \partial_\mu u, \partial_\mu u\rangle_t + \langle {}^Y\nabla u, {}^Y\nabla u\rangle_{h,t} + \langle u, u\rangle_t$$

and on $\{\mu < 0\}$ (using $v = \partial_\mu u$ in the previously established formula)

$$-\partial_\mu \mathcal{H} = -2\langle \mathcal{P}u, \partial_\mu u\rangle_t + \langle (2\mathcal{B}_\lambda - \partial_\mu)u, \partial_\mu u\rangle_t + \mathrm{div} - \widetilde{\mathcal{H}},$$

where $\widetilde{\mathcal{H}}$ has the same structure as $\mathcal{H}$ but with appearances of h (used to construct the various inner products) replaced by its Lie derivative, $\mathcal{L}_{\partial_\mu} h$. Recall that $\mathrm{supp}\, u \subset (-1, 0] \times Y$ and u is smooth, and hence $\partial_\mu^N u = 0$ at $\{\mu = 0\}$ for all N. Continuing to work on $\{\mu < 0\}$,

$$-\partial_\mu(|\mu|^{-N}\mathcal{H}) + |\mu|^{-N}\mathrm{div}$$
$$= -N|\mu|^{-N-1}\mathcal{H} - 2|\mu|^{-N}\operatorname{Re}\langle\mathcal{P}_\lambda u, \partial_\mu u\rangle_t + |\mu|^{-N}\langle(2\mathcal{B}_\lambda - \partial_\lambda)u, \partial_\mu u\rangle_t - |\mu|^{-N}\widetilde{\mathcal{H}}.$$

Now suppose that $u \in \ker\mathcal{P}_\lambda$. Fix $\delta > 0$ small and let $0 < \varepsilon < \delta$. We take the previous display and insert it into the operator $\int_{-\delta}^{-\varepsilon}\int_Y \cdots\, d\mu\, d\mathrm{vol}_h$. The first term on the left-hand side of the previous display is treated with the fundamental theorem of calculus, and the second term vanishes due to the appearance of $\int_Y \mathrm{div}\, d\mathrm{vol}_h$. We claim the right-hand side is negative for large N. Indeed the second term vanishes as u is assumed in the kernel of $\mathcal{P}_\lambda$. Considering the third term, $\langle(2\mathcal{B}_\lambda - \partial_\lambda)u, \partial_\mu u\rangle_t$ is quadratic in u, ${}^Y\nabla u$, and $\partial_\mu u$; hence for N large enough, it may be bounded by $N|\mu|^{-1}\mathcal{H}$. Thus the third term's potential positivity may be absorbed by the negativity of the first term. The fourth term may be treated in a similar manner upon consideration of the Taylor expansion of h at Y. We obtain

$$\delta^{-N}\int_Y \mathcal{H}(-\delta)\, d\mathrm{vol}_h \le \varepsilon^{-N}\int_Y \mathcal{H}(-\varepsilon)\, d\mathrm{vol}_h.$$

As u is smooth and vanishes to all orders at $\mu = 0$, we may bound $\int_Y \mathcal{H}(-\varepsilon)\, d\mathrm{vol}_h$ by $C|\mu|^K$ on $[-\varepsilon, 0]$ for arbitrarily large K. We can obtain a similar bound for $\int_Y \mathcal{H}(-\varepsilon)\, d\mathrm{vol}_h$, in particular, for $K > N$. This produces

$$\delta^{-N}\int_Y \mathcal{H}(-\delta)\, d\mathrm{vol}_h \le C\varepsilon^{-N+K}$$

and letting $\varepsilon \to 0^+$ shows $\int_Y \mathcal{H}(-\delta)\, \mathrm{dvol}_h = 0$; hence $\mathcal{H}(-\delta) = 0$. Doing this for all δ less than the original δ gives $\mathcal{H} = 0$ near 0. Hence $\partial_\mu u$ and $\nabla^Y u$ vanish and $u = 0$ near 0. This suffices to conclude the proof. □

Lemma 7.9. *For $\mathcal{P}_\lambda^*$ with $\lambda \in \mathbb{R}$ acting on $\dot{H}^{1-s}(\bar{X}_{cs}; \mathcal{E})$, the kernel of $\mathcal{P}_\lambda^*$ is trivial for $\lambda \gg 1$.*

Proof. Take λ satisfying the threshold condition and consider $v \in \ker\mathcal{P}_\lambda^*$. Hyperbolicity, as used in Lemma 7.7, implies $v = 0$ on $\{\mu \le 0\}$, and that v is smooth on X due to ellipticity. The strategy given in Lemma 7.7 implies $v \in \dot{H}^{\tilde{s}}(\bar{X}_{cs}; \mathcal{E})$ for all $\tilde{s} < \lambda + \frac{1}{2}$, which with $\lambda \gg n$ implies v is continuous. By the same logic, again by taking λ sufficiently large, we may assume v is regular enough to conclude $\partial_\mu^N v_{|Y} = 0$ for $N \le \frac{1}{2}\lambda$. Equivalently, $v_{|X} \in \rho^{2N} C^\infty_{\mathrm{even}}(\bar{X}; \mathcal{E})$. Meanwhile, direct calculations on $C^\infty(X; \mathcal{E})$ give

$$\begin{aligned}
\rho^N\nabla^*\nabla\rho^{-N} &= \nabla^*\nabla - N^2 - N(\Delta\log\rho) + 2N\nabla_{\rho\,\partial_\rho},\\
\rho^N \mathrm{d}\rho^{-N} &= d - N\frac{d\rho}{\rho}\,\cdot\,,\\
\rho^N\delta\rho^{-N} &= \delta + N\frac{d\rho}{\rho}\,\lrcorner\,,
\end{aligned}$$

where $\Delta\log\rho = n - \left(\frac{1}{2}\sum_{ij} h^{ij}\rho\,\partial_\rho h_{ij}\right) \in n - \rho^2 C^\infty_{\mathrm{even}}(\bar{X}; \mathcal{E})$. Also for $\tilde{u} \in C_c^\infty(X; \mathcal{E})$ we have

$$|(2N\nabla_{\rho\,\partial_\rho}\tilde{u}, \tilde{u})_s| = \left|N\int_X \|u\|_s^2\partial_\rho\left(\frac{d\rho\, d\mathrm{vol}_h}{\rho^n}\right)\right| \le CN\|u\|_s^2.$$

So consider the difference operator $(\mathcal{Q}_\lambda - N^2 + 2N\nabla_{\rho\,\partial_\rho}) - \rho^N \mathcal{Q}_\lambda \rho^{-N}$ acting on $\tilde{u} \in C_c^\infty(X; \mathcal{E})$. All terms are of order N and of differential order 0. Similar to the previous proof (and using the preceding remark in order to treat the term involving $N\nabla_{\rho\,\partial_\rho}$) we may obtain

$$|(\rho^N \mathcal{Q}_\lambda \rho^{-N} \tilde{u}, \tilde{u})_s| \geq C^{-1} \|\nabla \tilde{u}\|_s^2 + (\lambda^2 - N^2 - CN)\|\tilde{u}\|_s^2,$$

and provided $N \gg C$, the final term in the preceding display may be written with coefficient $\lambda^2 - 2N^2$. Set $N = \lfloor \frac{1}{2}\lambda \rfloor$ with $\lambda \gg 2C$ so that

$$|(\rho^N \mathcal{Q}_\lambda \rho^{-N} \tilde{u}, \tilde{u})_s| \geq C^{-1} \|\nabla \tilde{u}\|_s^2 + \tfrac{1}{2}\lambda^2 \|\tilde{u}\|_s^2.$$

Considering the Hilbert space $\{w \in L_s^2(X; \mathcal{E}) : B(w, w) < \infty\}$ with $B(w, w) = \|\rho^N \mathcal{Q}_\lambda \rho^{-N} w\|_s^2 < \infty$, the previous inequality shows that $w \mapsto (w, \tilde{f})_s$ is a linear functional for $\tilde{f} \in L_s^2(X; \mathcal{E})$ so by the Riesz representation theorem, there exists $\tilde{u} \in L_s^2(X; \mathcal{E})$ with $(\rho^N \mathcal{Q}_\lambda \rho^{-N} w, \tilde{u})_s = (w, \tilde{f})_s$ for all w. To show v vanishes on X, it suffices to show $(f, v)_t = 0$ for all $f \in C_c^\infty(X; \mathcal{E})$. Let $f \in C_c^\infty(X; \mathcal{E})$ and

$$\tilde{f} = \rho^{\lambda + n/2 - m + 2} J^{-1} \rho^{-N} f \in C_c^\infty(X; \mathcal{E}).$$

Then the preceding argument gives $\tilde{u} \in L_s^2(X; \mathcal{E})$ such that $\rho^{-N} \mathcal{Q}_\lambda \rho^N \tilde{u} = \tilde{f}$; hence $\mathcal{P}_\lambda u = f$, where

$$u = J \rho^{-\lambda - n/2 + m} \rho^N \tilde{u} \in \rho^{-(1/2)\lambda + 1} L_t^2(X; \mathcal{E})$$

(the inclusion is a consequence of Lemma 3.2). This gives u enough regularity to perform the following pairing which provides the desired result:

$$(f, v)_t = (\mathcal{P}_\lambda u, v)_t = (u, \mathcal{P}_\lambda^* v)_t = (u, 0)_t = 0. \qquad \square$$

8. Proofs of theorems

8A. ***Proof of Theorem 1.1.*** Proposition 6.6 gives

$$\mathcal{Q}_\lambda = J^{-1} \rho^{\lambda + n/2 - m + 2} \mathcal{P}_\lambda \rho^{-\lambda - n/2 + m} J$$

By Propositions 7.3 and 7.4, there is a meromorphic family $\mathcal{P}^{-1}$ on $\mathbb{C}$ mapping $\dot{C}^\infty(X; \mathcal{E})$ to $C^\infty(X_{cs}; \mathcal{E})$. Hence an extension of $\mathcal{Q}^{-1}$ from $\operatorname{Re}\lambda \gg 1$ to all of $\mathbb{C}$ as a meromorphic family is given by

$$\mathcal{Q}_\lambda^{-1} = J^{-1} \rho^{\lambda + n/2 - m} r_X \mathcal{P}_\lambda^{-1} \rho^{-\lambda - n/2 + m - 2} J,$$

where r_X is the restriction of sections above X_{cs} to sections above X. The previous display implies

$$\mathcal{Q}_\lambda^{-1} : \dot{C}^\infty(X; \mathcal{E}) \to \rho^{\lambda + n/2 - m} J^{-1} C_{\text{even}}^\infty(\overline{X}; \mathcal{E})$$

and for $f \in \dot{C}^\infty(X; \mathcal{E})$, we may write near $\partial \overline{X}$

$$\mathcal{Q}_\lambda^{-1} f_{|U} = \mu^{\lambda/2 + n/4 - m/2} J^{-1} \sum_{k=0}^{m} \sum_{\ell=0}^{k} (d\mu)^{k-\ell} \cdot \tilde{u}^{(\ell)}, \quad \tilde{u}^{(\ell)} \in C_{\text{even}}^\infty([0, 1) \times Y; \operatorname{Sym}^\ell \mathrm{T}^* Y).$$

The proof of Lemma 3.2 shows that the part of J (or J^{-1}) which sends $\mathcal{E}^{(k)}$ to $\mathcal{E}^{(k+p)}$ for $0 \le p \le m-k$ is, up to a constant, $(d\mu/\mu)^p$. Therefore,

$$\mathcal{Q}_\lambda^{-1} f_{|U} \in \mu^{\lambda/2+n/4-m/2} \bigoplus_{k=0}^{m} \bigoplus_{p=0}^{m-k} \Big(\frac{d\mu}{\mu}\Big)^p \cdot \bigoplus_{\ell=0}^{k} (d\mu)^{k-\ell} \cdot C^\infty_{\mathrm{even}}([0,1) \times Y; \mathrm{Sym}^\ell \mathrm{T}^* Y).$$

Hence on X,

$$\mathcal{Q}_\lambda^{-1} f \in \rho^{\lambda+n/2-m} \bigoplus_{k=0}^{m} \bigoplus_{p=0}^{m-k} \rho^{-2p} C^\infty_{\mathrm{even}}(\bar{X}; \mathcal{E}^{(k+p)}),$$

which is contained in $\rho^{\lambda+n/2-m} \bigoplus_{k=0}^{m} \rho^{-2k} C^\infty_{\mathrm{even}}(\bar{X}; \mathcal{E}^{(k)})$.

Remark 8.1. Suppose that, for $f \in \dot{C}^\infty(X; \mathcal{E})$, it were possible to write in the preceding proof that near $\partial \bar{X}$

$$\mathcal{Q}_\lambda^{-1} f_{|U} = \rho^{\lambda+n/2-m} J^{-1} \tilde{u}^{(m)}, \quad \tilde{u}^{(m)} \in C^\infty_{\mathrm{even}}(\bar{U}; \mathcal{E}^{(m)}).$$

Then as J^{-1} acts as the identity upon restriction to $\mathcal{E}^{(m)}$, we would obtain

$$\mathcal{Q}_\lambda^{-1} f \in \rho^{\lambda+n/2-m} C^\infty_{\mathrm{even}}(\bar{X}; \mathcal{E}^{(m)}).$$

This will be useful for the asymptotics given in Theorems 1.3 and 1.4.

8B. *Proof of Theorem 1.2.* The meromorphic inverse of $\mathcal{Q}_\lambda$ is precisely that given in the preceding proof:

$$\mathcal{Q}_\lambda^{-1} = J^{-1} \rho^{\lambda+n/2-m} r_X \mathcal{P}_\lambda^{-1} \rho^{-\lambda-n/2+m-2} J.$$

All we must check is that, given $f \in \dot{C}^\infty(X; \mathcal{E}) \cap \ker(\Lambda_{s^{-2}\eta} \circ \pi_s^*)$, the resulting section $u = \mathcal{Q}_\lambda^{-1} f$ is indeed trace-free with respect to the ambient trace operator. To this end, we first lift the equation $\mathcal{Q}_\lambda u = f$ to an equation on M involving $\boldsymbol{Q}$, giving

$$s^\lambda \boldsymbol{Q} s^{-\lambda} (\pi_s^* u) = \pi_s^* f.$$

We apply $\Lambda_{s^{-2}\eta}$ to obtain an equation on $\mathcal{F}^{(m-2)}$. Using the hypothesis $\Lambda_{s^{-2}\eta} \pi_s^* f = 0$ and Lemma 5.5 to commute $s^{-2} \Lambda_{s^{-2}\eta}$ with $\boldsymbol{Q}$ gives

$$s^2 s^\lambda \boldsymbol{Q} s^{-\lambda} s^{-2} \Lambda_{s^{-2}\eta} (\pi_s^* u) = 0.$$

Freezing this differential equation at $s = 0$ with $\pi_{s=0}$ to obtain the indicial family of $\boldsymbol{Q}$ provides the equation

$$\mathrm{I}_s(\boldsymbol{Q}, \lambda+2) \pi_{s=0} \Lambda_{s^{-2}\eta} (\pi_s^* u) = 0.$$

Section 7 ensures that, for $\operatorname{Re}\lambda \gg 1$, this operator has trivial kernel; hence

$$\pi_{s=0} \Lambda_{s^{-2}\eta} (\pi_s^* u) = 0$$

and $u \in \ker(\Lambda_{s^{-2}\eta} \circ \pi_s^*)$ as required.

8C. ***Proof of Theorems 1.3 and 1.4.*** We are finally in a position to consider the original problem of proving Theorems 1.3 and 1.4. Let

$$f \in \dot{C}^\infty(X; \mathcal{E}^{(m)}) \cap \ker \Lambda \cap \ker \delta$$

and define, using Theorem 1.1,

$$u = \sum_{k=0}^{m} u^{(k)} = \mathcal{Q}_\lambda^{-1} f, \quad u^{(k)} \in \rho^{\lambda+n/2-m-2k} C^\infty_{\mathrm{even}}(\bar{X}; \mathcal{E}^{(k)}).$$

Note that the growth near $\partial \bar{X}$ of $u^{(k)}$ and $\delta u^{(k)}$ may be controlled by the size of $\operatorname{Re}\lambda$; hence for $\operatorname{Re}\lambda \gg 1$ we may assume that they are sections of $L^2_s(X; \mathcal{E}^{(k)})$ and $L^2_s(X; \mathcal{E}^{(k-1)})$ respectively. We claim, for $\operatorname{Re}\lambda \gg 1$ and $|\operatorname{Im}\lambda| \ll 1$, that

$$u = u^{(m)} \in \rho^{\lambda+n/2-m} C^\infty_{\mathrm{even}}(\bar{X}; \mathcal{E}^{(k)}) \cap \ker \Lambda \cap \ker \delta,$$

at which point the equation $\mathcal{Q}_\lambda u = f$ decouples giving

$$(\Delta + \lambda^2 - c_m) u = f,$$

and by uniqueness of the L^2 inverse of the Laplacian, we have the formula, for $\operatorname{Re}\lambda \gg 1$ and $|\operatorname{Im}\lambda| \ll 1$,

$$(\Delta + \lambda^2 - c_m)^{-1} = J^{-1} \rho^{\lambda+n/2-m} r_X \mathcal{P}_\lambda^{-1} \rho^{-\lambda-n/2+m-2} J,$$

with the right-hand side giving the meromorphic extension of the resolvent stated in the theorems.

To this end take $\operatorname{Re}\lambda \gg 1$ and $|\operatorname{Im}\lambda| \ll 1$. By Theorem 1.2, we deduce u is trace-free with respect to the ambient trace operator; thus $\mathcal{Q}_\lambda$ takes the form detailed in Proposition 5.14. We begin by remarking that while working on $L^2_s(X; \mathcal{E}^{(k)})$, if $\mathcal{R}_\lambda^{(k)}$ is any operator of the form $(\Delta + \lambda^2 + O(1))^{-1}$ (which has order $O(|\lambda|^{-2})$, then the operator $\mathrm{d}\mathcal{R}_\lambda^{(k)}\delta$ has norm of order $O(1)$. We define $\mathcal{R}_\lambda^{(0)} = (\Delta + \lambda^2 - c_0')^{-1}$ and for $0 < k < m$,

$$\mathcal{R}_\lambda^{(k)} = \left(\Delta + \lambda^2 - c_k' + 4(m-k+1)\mathrm{d}\mathcal{R}_\lambda^{(k-1)}\delta\right)^{-1}.$$

The component of $\mathcal{Q}_\lambda u = f$ in $\mathcal{E}^{(0)}$ is

$$(\Delta + \lambda^2 - c_0') u^{(0)} = 2\sqrt{m}\,\delta u^{(1)};$$

hence $u^{(0)} = 2\sqrt{m}\mathcal{R}_\lambda^{(0)} \delta u^{(1)}$. The component of $\mathcal{Q}_\lambda u = f$ in $\mathcal{E}^{(1)}$ now reads as

$$(\Delta + \lambda^2 - c_1' + 4m\mathrm{d}\mathcal{R}_\lambda^{(0)}\delta) u^{(1)} = 2\sqrt{m-1}\,\delta u^{(2)};$$

hence $u^{(1)} = 2\sqrt{m-1}\mathcal{R}_\lambda^{(1)} \delta u^{(2)}$. Continuing, we obtain on $\mathcal{E}^{(m)}$,

$$(\Delta + \lambda^2 - c_m + 4\mathrm{d}\mathcal{R}_\lambda^{(m-1)}\delta) u^{(m)} = f.$$

Applying the divergence, we recall Lemma 5.2. For this, we must assume that if $m = 2$ then X has parallel Ricci curvature, and if $m \geq 3$ then X is locally isomorphic to $\mathbb{H}^{n+1}$. We obtain

$$(\Delta + \lambda^2 - c_m + 4\delta\mathrm{d}\mathcal{R}_\lambda^{(m-1)}) \delta u^{(m)} = 0.$$

Again, $\delta \mathrm{d}\mathcal{R}_\lambda^{(m-1)}$ has norm of order $O(1)$ so we may invert this equation and deduce that $\delta u^{(m)} = 0$. This implies, for all $k < m$,

$$u^{(k)} = 2\sqrt{m-k}\,\mathcal{R}_\lambda^{(k)} \delta u^{k+1} = 0.$$

Therefore $u = u^{(m)}$. By Remark 8.1, $u \in \rho^{\lambda+n/2-m} C^\infty_{\text{even}}(\bar{X}; \mathcal{E}^{(m)})$. By Theorem 1.2, $u \in \ker \Lambda$, and as previously mentioned $u \in \ker \delta$. This completes the proof.

9. Symmetric cotensors of rank 2

This section details the results stated in Sections 5 and 8 for rank-2 symmetric cotensors. In this low rank, writing the action of the d'Alembertian, or its conjugation $\boldsymbol{Q}$, on $\mathcal{F} = \operatorname{Sym}^2 \mathrm{T}^* M$ is tractable.

9A. ***The operator Q for 2-cotensors.*** Using the decomposition given by the Minkowski scale, we write

$$u = \begin{bmatrix} 1 & \frac{ds}{s}\cdot & \frac{1}{\sqrt{2}}\left(\frac{ds}{s}\right)^2 \end{bmatrix} \begin{bmatrix} u^{(2)} \\ u^{(1)} \\ u^{(0)} \end{bmatrix}, \qquad u \in C^\infty(M; \mathcal{F}),\ u^{(k)} \in C^\infty(M; \mathcal{E}^{(k)}).$$

The change of basis matrix J takes the form

$$J = \begin{bmatrix} 1 & \frac{d\rho}{\rho}\cdot & \frac{1}{\sqrt{2}}\left(\frac{d\rho}{\rho}\right)^2 \\ 0 & 1 & \sqrt{2}\frac{d\rho}{\rho}\cdot \\ 0 & 0 & 1 \end{bmatrix}.$$

Propositions 5.6 and 5.10 become:

Proposition 9.1. *For $u \in C^\infty(M; \mathcal{F})$ decomposed relative to the Minkowski scale (2), the conjugated d'Alembertian $\boldsymbol{Q}$ is given by*

$$\boldsymbol{Q}u = \begin{bmatrix} 1 & \frac{ds}{s}\cdot & \frac{1}{\sqrt{2}}\left(\frac{ds}{s}\right)^2 \end{bmatrix} \begin{bmatrix} \Delta + (s\,\partial_s)^2 - c_2 - \mathrm{L}\Lambda & 2\mathrm{d} & -\sqrt{2}\mathrm{L} \\ -2\delta & \Delta + (s\,\partial_s)^2 - c_1 & 2\sqrt{2}\mathrm{d} \\ -\sqrt{2}\Lambda & -2\sqrt{2}\delta & \Delta + (s\,\partial_s)^2 - c_0 \end{bmatrix} \begin{bmatrix} u^{(2)} \\ u^{(1)} \\ u^{(0)} \end{bmatrix}$$

with constants

$$c_2 = \tfrac{1}{4}n(n-8), \qquad c_1 = \tfrac{1}{4}(n^2+16), \qquad c_0 = \tfrac{1}{4}(n^2+8n+8).$$

If, furthermore, u is trace-free with respect to the trace operator $\Lambda_{s^{-2}\eta}$, then $\Lambda u^{(2)} = -\sqrt{2}u^{(0)}$, and

$$\boldsymbol{Q}u = \begin{bmatrix} 1 & \frac{ds}{s}\cdot & \frac{1}{\sqrt{2}}\left(\frac{ds}{s}\right)^2 \end{bmatrix} \begin{bmatrix} \Delta + (s\,\partial_s)^2 - c_2' & 2\mathrm{d} & 0 \\ -2\delta & \Delta + (s\,\partial_s)^2 - c_1' & 2\sqrt{2}\mathrm{d} \\ 0 & -2\sqrt{2}\delta & \Delta + (s\,\partial_s)^2 - c_0' \end{bmatrix} \begin{bmatrix} u^{(2)} \\ u^{(1)} \\ u^{(0)} \end{bmatrix}$$

with modified constants

$$c_2' = c_2, \qquad c_1' = c_1, \qquad c_0' = \tfrac{1}{4}(n^2+8n).$$

9B. *The indicial family of Q for 2-cotensors.* Propositions 5.12 and 5.14 become:

Proposition 9.2. *For $u = \sum_{k=0}^{2} u^{(k)} \in C^\infty(X; \mathcal{E})$ the operator $\mathcal{Q}$ is given by*

$$\mathcal{Q}_\lambda u = \begin{bmatrix} \Delta+\lambda^2-c_2-\mathrm{L}\Lambda & 2\mathrm{d} & -\sqrt{2}\mathrm{L} \\ -2\delta & \Delta+\lambda^2-c_1 & 2\sqrt{2}\mathrm{d} \\ -\sqrt{2}\Lambda & -2\sqrt{2}\delta & \Delta+\lambda^2-c_0 \end{bmatrix} \begin{bmatrix} u^{(2)} \\ u^{(1)} \\ u^{(0)} \end{bmatrix}$$

and if, furthermore, $u \in \ker(\Lambda_{s^{-2}\eta} \circ \pi_s^*)$ then

$$\mathcal{Q}_\lambda u = \begin{bmatrix} \Delta+\lambda^2-c_2' & 2\mathrm{d} & 0 \\ -2\delta & \Delta+\lambda^2-c_1' & 2\sqrt{2}\mathrm{d} \\ 0 & -2\sqrt{2}\delta & \Delta+\lambda^2-c_0' \end{bmatrix} \begin{bmatrix} u^{(2)} \\ u^{(1)} \\ u^{(0)} \end{bmatrix}$$

with previously given constants.

9C. *Illustration of proof for 2-cotensors.* Let $f \in \dot{C}^\infty(X; \mathcal{E}^{(2)}) \cap \ker \Lambda \cap \ker \delta$ and define

$$\begin{bmatrix} u^{(2)} \\ u^{(1)} \\ u^{(0)} \end{bmatrix} = J^{-1} \rho^{\lambda+n/2-2} r_X \mathcal{P}^{-1} \rho^{-\lambda-n/2} J \begin{bmatrix} f \\ 0 \\ 0 \end{bmatrix}.$$

Take $\operatorname{Re}\lambda \gg 1$ and $|\operatorname{Im}\lambda| \ll 1$. By Theorem 1.1,

$$u^{(k)} \in \rho^{\lambda+n/2-2-2k} C^\infty_{\mathrm{even}}(\bar{X}; \mathcal{E}^{(k)}),$$

and by Proposition 6.6, $\mathcal{Q}_\lambda u = f$. Theorem 1.2 forces

$$\Lambda_{s^{-2}\eta}\Big(u^{(2)} + \frac{ds}{s}\cdot u^{(1)} + \frac{1}{\sqrt{2}}\Big(\frac{ds}{s}\Big)^2 \cdot u^{(0)}\Big) = 0;$$

hence $\Lambda u^{(2)} = -\sqrt{2}u^{(0)}$, and $\mathcal{Q}_\lambda u = f$ reads explicitly as

$$\begin{bmatrix} \Delta+\lambda^2-c_2 & 2\mathrm{d} & 0 \\ -2\delta & \Delta+\lambda^2-c_1 & 2\sqrt{2}\mathrm{d} \\ 0 & -2\sqrt{2}\delta & \Delta+\lambda^2-c_0' \end{bmatrix} \begin{bmatrix} u^{(2)} \\ u^{(1)} \\ u^{(0)} \end{bmatrix} = \begin{bmatrix} f \\ 0 \\ 0 \end{bmatrix}.$$

Introducing the resolvents $\mathcal{R}_\lambda^{(0)}$ and $\mathcal{R}_\lambda^{(1)}$ provides

$$\begin{bmatrix} \Delta+\lambda^2-c_2+4\mathrm{d}\mathcal{R}_\lambda^{(1)}\delta & 0 & 0 \\ -2\delta & \Delta+\lambda^2-c_1+8\mathrm{d}\mathcal{R}_\lambda^{(0)}\delta & 0 \\ 0 & -2\sqrt{2}\delta & \Delta+\lambda^2-c_0' \end{bmatrix} \begin{bmatrix} u^{(2)} \\ u^{(1)} \\ u^{(0)} \end{bmatrix} = \begin{bmatrix} f \\ 0 \\ 0 \end{bmatrix}$$

and applying δ, assuming that X is Einstein, provides the homogeneous equation

$$\begin{bmatrix} \Delta+\lambda^2-c_2+4\delta\mathrm{d}\mathcal{R}_\lambda^{(1)} & 0 & 0 \\ -2\delta & \Delta+\lambda^2-c_1+8\delta\mathrm{d}\mathcal{R}_\lambda^{(0)} & 0 \\ 0 & -2\sqrt{2}\delta & \Delta+\lambda^2-c_0' \end{bmatrix} \begin{bmatrix} \delta u^{(2)} \\ \delta u^{(1)} \\ \delta u^{(0)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

The lower triangular nature of this system implies $\delta u^{(k)} = 0$ for all k. Hence the system $\mathcal{Q}_\lambda u = f$ collapses. So $u^{(0)}$ and $u^{(1)}$ vanish and by Remark 8.1,

$$u = u^{(2)} \in \rho^{\lambda+n/2-2} C^\infty_{\text{even}}(\bar{X}; \mathcal{E}^{(k)}),$$

giving $(\Delta + \lambda^2 - c_2)u = f$.

10. High energy estimates via semiclassical analysis

This article shows the meromorphic continuation of the resolvent of the Laplacian on symmetric tensors using microlocal techniques. This direction means one does not talk about introducing complex absorbers but rather studies the problem on a manifold with boundary. If one were to follow more closely the track established by Vasy, one obtains semiclassical estimates. We state these estimates.

On X, whose smooth structure at infinity is the even structure given by μ rather than ρ, we have the semiclassical spaces $H^s_{|\lambda|^{-1}}(X; \mathcal{E})$.

Theorem 10.1. *Suppose that X is an even asymptotically hyperbolic manifold which is nontrapping. Then the meromorphic continuation, written as $\mathcal{Q}_\lambda^{-1}$, of the inverse of $\mathcal{Q}_\lambda$ initially acting on $L^2_s(X; \mathcal{E})$ has nontrapping estimates holding in every strip* $|\operatorname{Re}\lambda| < C$, $|\operatorname{Im}\lambda| \gg 0$: *for* $s > \frac{1}{2} + C$,

$$\|\rho^{-\lambda-n/2+m} \mathcal{Q}_\lambda^{-1} f\|_{H^s_{|\lambda|^{-1}}(X;\mathcal{E})} \le C|\lambda|^{-1} \|\rho^{-\lambda-n/2+m-2} f\|_{H^{s-1}_{|\lambda|^{-1}}(X;\mathcal{E})}.$$

If X is furthermore Einstein, then restricting to symmetric 2-cotensors, the meromorphic continuation $\mathcal{R}_\lambda$ of the inverse of

$$\Delta - \tfrac{1}{4}n(n-8) + \lambda^2$$

initially acting on $L^2(X; \mathcal{E}^{(2)}) \cap \ker\Lambda \cap \ker\delta$ *has nontrapping estimates holding in every strip* $|\operatorname{Re}\lambda| < C$, $|\operatorname{Im}\lambda| \gg 0$: *for* $s > \frac{1}{2} + C$,

$$\|\rho^{-\lambda-n/2+2} \mathcal{R}_\lambda f\|_{H^s_{|\lambda|^{-1}}(X;\mathcal{E}^{(2)})} \le C|\lambda|^{-1} \|\rho^{-\lambda-n/2} f\|_{H^{s-1}_{|\lambda|^{-1}}(X;\mathcal{E}^{(2)})}.$$

Acknowledgements

I would like to thank Colin Guillarmou for his guidance and for the many discussions placing this work in a wider context, and also Andrei Moroianu for his continued support and several useful comments on Section 5. Finally, I thank Rod Gover for originally sparking my interest in all things conformal.

References

[Aubry and Guillarmou 2011] E. Aubry and C. Guillarmou, "Conformal harmonic forms, Branson–Gover operators and Dirichlet problem at infinity", *J. Eur. Math. Soc.* (*JEMS*) **13**:4 (2011), 911–957. MR Zbl

[Bailey et al. 1994] T. N. Bailey, M. G. Eastwood, and A. R. Gover, "Thomas's structure bundle for conformal, projective and related structures", *Rocky Mountain J. Math.* **24**:4 (1994), 1191–1217. MR Zbl

[Baskin et al. 2015] D. Baskin, A. Vasy, and J. Wunsch, "Asymptotics of radiation fields in asymptotically Minkowski space", *Amer. J. Math.* **137**:5 (2015), 1293–1364. MR Zbl

[Biquard 2000] O. Biquard, *Métriques d'Einstein asymptotiquement symétriques*, Astérisque **265**, Société Mathématique de France, Paris, 2000. MR Zbl

[Branson and Gover 2005] T. Branson and A. R. Gover, "Conformally invariant operators, differential forms, cohomology and a generalisation of Q-curvature", *Comm. Partial Differential Equations* **30**:10-12 (2005), 1611–1669. MR Zbl

[Delay 1999] E. Delay, "Étude locale d'opérateurs de courbure sur l'espace hyperbolique", *J. Math. Pures Appl.* (9) **78**:4 (1999), 389–430. MR Zbl

[Delay 2002] E. Delay, "Essential spectrum of the Lichnerowicz Laplacian on two tensors on asymptotically hyperbolic manifolds", *J. Geom. Phys.* **43**:1 (2002), 33–44. MR Zbl

[Delay 2007] E. Delay, "TT-eigentensors for the Lichnerowicz Laplacian on some asymptotically hyperbolic manifolds with warped products metrics", *Manuscripta Math.* **123**:2 (2007), 147–165. MR Zbl

[Djadli et al. 2008] Z. Djadli, C. Guillarmou, and M. Herzlich, *Opérateurs géométriques, invariants conformes et variétés asymptotiquement hyperboliques*, Panoramas et Synthèses **26**, Société Mathématique de France, Paris, 2008. MR Zbl

[Dyatlov and Guillarmou 2016] S. Dyatlov and C. Guillarmou, "Pollicott–Ruelle resonances for open systems", *Ann. Henri Poincaré* **17**:11 (2016), 3089–3146. MR Zbl

[Dyatlov and Zworski 2017] S. Dyatlov and M. Zworski, "Mathematical theory of scattering resonances", book in progress, 2017, available at http://math.mit.edu/~dyatlov/res/res_20170323.pdf. Version 0.1 (March 23, 2017).

[Dyatlov et al. 2015] S. Dyatlov, F. Faure, and C. Guillarmou, "Power spectrum of the geodesic flow on hyperbolic manifolds", *Anal. PDE* **8**:4 (2015), 923–1000. MR Zbl

[Faure and Tsujii 2013] F. Faure and M. Tsujii, "Band structure of the Ruelle spectrum of contact Anosov flows", *C. R. Math. Acad. Sci. Paris* **351**:9-10 (2013), 385–391. MR Zbl

[Fefferman and Graham 1985] C. Fefferman and C. R. Graham, "Conformal invariants", pp. 95–116 in *The mathematical heritage of Élie Cartan* (Lyon, 1984), Astérisque (numéro hors série), Société Mathématique de France, Paris, 1985. MR Zbl

[Fefferman and Graham 2012] C. Fefferman and C. R. Graham, *The ambient metric*, Annals of Mathematics Studies **178**, Princeton University Press, 2012. MR Zbl

[Graham and Lee 1991] C. R. Graham and J. M. Lee, "Einstein metrics with prescribed conformal infinity on the ball", *Adv. Math.* **87**:2 (1991), 186–225. MR Zbl

[Graham and Zworski 2003] C. R. Graham and M. Zworski, "Scattering matrix in conformal geometry", *Invent. Math.* **152**:1 (2003), 89–118. MR Zbl

[Guillarmou 2005] C. Guillarmou, "Meromorphic properties of the resolvent on asymptotically hyperbolic manifolds", *Duke Math. J.* **129**:1 (2005), 1–37. MR Zbl

[Guillarmou et al. 2016] C. Guillarmou, J. Hilgert, and T. Weich, "Classical and quantum resonances for hyperbolic surfaces", preprint, 2016. To appear in *Math. Ann.* arXiv

[Guillopé and Zworski 1995] L. Guillopé and M. Zworski, "Polynomial bounds on the number of resonances for some complete spaces of constant negative curvature near infinity", *Asymptotic Anal.* **11**:1 (1995), 1–22. MR Zbl

[Hadfield 2017] C. Hadfield, "Ruelle and quantum resonances for open hyperbolic manifolds", preprint, 2017. arXiv

[Heil et al. 2016] K. Heil, A. Moroianu, and U. Semmelmann, "Killing and conformal Killing tensors", *J. Geom. Phys.* **106** (2016), 383–400. MR Zbl

[Hörmander 1994] L. Hörmander, *The analysis of linear partial differential operators, III: Pseudo-differential operators*, vol. 274, Grundlehren der Mathematischen Wissenschaften, Springer, 1994. MR

[Lichnerowicz 1961] A. Lichnerowicz, "Propagateurs et commutateurs en relativité générale", *Inst. Hautes Études Sci. Publ. Math.* 10 (1961), 56. MR Zbl

[Mazzeo 1988] R. Mazzeo, "The Hodge cohomology of a conformally compact metric", *J. Differential Geom.* **28**:2 (1988), 309–339. MR Zbl

[Mazzeo and Melrose 1987] R. R. Mazzeo and R. B. Melrose, "Meromorphic extension of the resolvent on complete spaces with asymptotically constant negative curvature", *J. Funct. Anal.* **75**:2 (1987), 260–310. MR Zbl

[Melrose 1993] R. B. Melrose, *The Atiyah–Patodi–Singer index theorem*, Research Notes in Mathematics **4**, A K Peters, Wellesley, MA, 1993. MR Zbl

[Melrose 1994] R. B. Melrose, "Spectral and scattering theory for the Laplacian on asymptotically Euclidian spaces", pp. 85–130 in *Spectral and scattering theory* (Sanda, 1992), edited by M. Ikawa, Lecture Notes in Pure and Appl. Math. **161**, Dekker, New York, 1994. MR Zbl

[Vasy 2013a] A. Vasy, "Microlocal analysis of asymptotically hyperbolic and Kerr–de Sitter spaces", *Invent. Math.* **194**:2 (2013), 381–513. MR Zbl

[Vasy 2013b] A. Vasy, "Microlocal analysis of asymptotically hyperbolic spaces and high-energy resolvent estimates", pp. 487–528 in *Inverse problems and applications: inside out, II*, edited by G. Uhlmann, Math. Sci. Res. Inst. Publ. **60**, Cambridge University Press, 2013. MR Zbl

[Vasy 2017] A. Vasy, "Analytic continuation and high energy estimates for the resolvent of the Laplacian on forms on asymptotically hyperbolic spaces", *Adv. Math.* **306** (2017), 1019–1045. MR Zbl

[Wang 2009] F. Wang, "Dirichlet-to-Neumann map for Poincaré–Einstein metrics in even dimensions", preprint, 2009. arXiv

[Zworski 2016] M. Zworski, "Resonances for asymptotically hyperbolic manifolds: Vasy's method revisited", *J. Spectr. Theory* **6**:4 (2016), 1087–1114. MR Zbl

Received 24 Oct 2016. Accepted 12 Jul 2017.

CHARLES HADFIELD: charles.hadfield@ens.fr
DMA, École Normale Supérieure, Paris, France

ANALYSIS AND PDE
Vol. 10, No. 8, 2017
dx.doi.org/10.2140/apde.2017.10.1923

CONSTRUCTION OF TWO-BUBBLE SOLUTIONS FOR THE ENERGY-CRITICAL NLS

JACEK JENDREJ

We construct pure two-bubbles for the energy-critical focusing nonlinear Schrödinger equation in space dimension $N \geq 7$. The constructed solution is global in (at least) one time direction and approaches a superposition of two stationary states both centered at the origin, with the ratio of their length scales converging to 0. One of the bubbles develops at scale 1, whereas the length scale of the other converges to 0 at rate $|t|^{-\frac{2}{N-6}}$. The phases of the two bubbles form the right angle.

1. Introduction

Setting of the problem. We consider the Schrödinger equation with the focusing energy-critical power nonlinearity given by

$$i\partial_t u(t,x) + \Delta u(t,x) + f(u(t,x)) = 0, \quad f(z) := |z|^{\frac{4}{N-2}} z, \quad t \in \mathbb{R},\ x \in \mathbb{R}^N. \tag{1-1}$$

This equation can be studied in space dimension $N \geq 3$, but we will restrict our attention to the case $N \geq 7$.

The *energy functional* associated with this equation is defined for $u_0 \in \dot{H}^1(\mathbb{R}^N; \mathbb{C})$ by the formula

$$E(u_0) := \int_{\mathbb{R}^N} \tfrac{1}{2}|\nabla u_0(x)|^2 - F(u_0(x))\,\mathrm{d}x,$$

where

$$F(z) := \frac{N-2}{2N}|z|^{\frac{2N}{N-2}}.$$

Note that $E(u_0)$ is well-defined due to the Sobolev embedding theorem. The differential of E is $\mathrm{D}E(u_0) = -\Delta u_0 - f(u_0)$; hence we have the following Hamiltonian form of (1-1):

$$\partial_t u(t) = -i\,\mathrm{D}E(u(t)).$$

Equation (1-1) is locally well-posed in the space $\dot{H}^1(\mathbb{R}^N)$, as was proved by Cazenave and Weissler [1990]; see also a complete review of Cauchy theory in [Kenig and Merle 2006] for $N \in \{3,4,5\}$ and [Killip and Visan 2010] for $N \geq 6$. By "well-posed" we mean that for any initial data $u_0 \in \dot{H}^1(\mathbb{R}^N)$ there exists $\tau > 0$ and a linear subspace $S \subset C([t_0-\tau, t_0+\tau]; \dot{H}^1(\mathbb{R}^N))$ such that there exists a unique weak solution $u(t) \in S$ of (1-1) satisfying $u(t_0) = u_0$, and that this solution is continuous with respect to the initial data. By standard arguments, there exists a maximal time of existence (T_-, T_+), $-\infty \leq T_- < t_0 < T_+ \leq +\infty$, and a unique solution $u \in C((T_-, T_+); \dot{H}^1(\mathbb{R}^N))$. Moreover, if $u_0 \in X^1 :=$

MSC2010: 35B40.
Keywords: nonlinear Schrödinger equation, energy-critical, multisoliton.

$\dot H^2(\mathbb{R}^N)\cap \dot H^1(\mathbb{R}^N)$, then $u\in C((T_-,T_+);X^1)$. If $T_+<+\infty$, then $u(t)$ leaves every compact subset of $\dot H^1(\mathbb{R}^N)$ as t approaches T_+. A crucial property of the solutions of (1-1) is that the energy E is a conservation law. If $u_0\in L^2$, then the mass $\|u(t)\|_{L^2}^2$ is another conservation law, but we will never use this fact.

In this paper, we always assume that the initial data are radially symmetric. This symmetry is preserved by the flow. We denote by $\mathcal{E}$ the space of radially symmetric functions in $\dot H^1(\mathbb{R}^N;\mathbb{C})$.

For a function $v\in\mathcal{E}$, we define

$$v_\lambda(x) := \frac{1}{\lambda^{\frac{N-2}{2N}}} v\Big(\frac{x}{\lambda}\Big).$$

A change of variables shows that

$$E((u_0)_\lambda)=E(u_0).$$

Equation (1-1) is invariant under the same scaling: if $u(t)$ is a solution of (1-1) and $\lambda>0$, then $t\mapsto u(t_0+\lambda^{-2}t)_\lambda$ is also a solution with initial data $(u_0)_\lambda$ at time $t=0$. This is why (1-1) is called *energy-critical*.

The solutions of the corresponding *defocusing* equation exist globally and scatter. This was proved by Bourgain [1999] and Tao [2005] for radial solutions, and by Colliander, Keel, Staffilani, Takaoka and Tao [Colliander et al. 2008], Ryckman and Visan [2007], and Visan [2007] for nonradial data.

The study of the dynamical behavior of solutions of the focusing equation (1-1) for large initial data was initiated by Kenig and Merle [2006]. In this case, an important role is played by the family of stationary solutions $u(t)\equiv \mathrm{e}^{i\theta}W_\lambda$, where

$$W(x)=\Big(1+\frac{|x|^2}{N(N-2)}\Big)^{-\frac{N-2}{2}}.$$

The functions $\mathrm{e}^{i\theta}W_\lambda$ are called *ground states* or *bubbles* (of energy). They are the only radially symmetric solutions of the critical elliptic problem

$$-\Delta u - f(u)=0.$$

The ground states achieve the optimal constant in the critical Sobolev inequality, which was proved by Aubin [1976] and Talenti [1976]. They are the "mountain passes" for the potential energy.

Kenig and Merle [2006] exhibited the special role of the ground states $\mathrm{e}^{i\theta}W_\lambda$ as the *threshold elements* for nonlinear dynamics of the solutions of (1-1) in space dimensions $N=3,4,5$ for radial data. They proved the so-called *threshold conjecture* by completely classifying the dynamical behavior of solutions $u(t)$ of (1-1) such that $E(u(t))<E(W)$. An analogous result in higher dimensions, for nonradial data, was obtained by Killip and Visan [2010].

A much stronger statement about the dynamics of solutions is the *soliton resolution conjecture*, which predicts that a bounded (in an appropriate sense) solution decomposes asymptotically into a sum of energy bubbles at different scales and a radiation term (a solution of the linear Schrödinger equation). This was proved for the radial energy-critical wave equation in dimension $N=3$ by Duyckaerts, Kenig and Merle [Duyckaerts et al. 2013]; see also [Duyckaerts et al. 2017] for the nonradial case. For (1-1) this problem is completely open.

Solutions slightly above the ground state energy threshold were studied by Ortoleva and Perelman [2013] in dimension $N = 3$; see also [Perelman 2014] for the closely related critical equivariant Schrödinger map equation with values in the sphere. They constructed global solutions which stay close to $e^{i\theta}W_\lambda$ in the energy space, with λ converging to 0 as time t goes to $+\infty$. These solutions decompose into a concentrating bubble and a radiation term, in accordance with the soliton resolution conjecture. The works of Ortoleva and Perelman follow the approach developed by Krieger, Schlag and Tataru [Krieger et al. 2008; 2009] for wave equations. For the Schrödinger maps, following a different approach, Merle, Rodnianski and Raphaël [Merle et al. 2013] obtained blow-up solutions which are stable relative to a set of finite codimension in some space which contains the bubble.

On the classification side, it is unknown whether the soliton resolution conjecture holds even with an additional assumption that the solution remains close to the family of the ground states. In the mass-critical case and for a solution blowing up in finite time, this was proved by Merle and Raphaël [2004; 2005]; see also [Fan 2016].

Main results. In view of the soliton resolution conjecture, solutions which exhibit no dispersion in one or both time directions play a distinguished role. One obvious example of such solutions are the static solutions $e^{i\theta}W_\lambda$. In this paper, we consider the simplest nontrivial case; namely we construct global radial solutions which approach, in the energy space, a sum of two bubbles. The ratio of the scales at which these bubbles develop tends to 0.

Theorem 1. *There exists a solution $u : (-\infty, T_0] \to \mathcal{E}$ of (1-1) such that*

$$\lim_{t\to-\infty}\left\|u(t) - \left(-iW + W_{\frac{1}{\kappa}(\kappa|t|)^{-2/(N-6)}}\right)\right\|_{\mathcal{E}} = 0,$$

where κ is an explicit constant.

Remark 1.1. For the value of κ, see (3-4).

Remark 1.2. More precisely, we will prove that

$$\left\|u(t) - \left(-iW + W_{\frac{1}{\kappa}(\kappa|t|)^{-2/(N-6)}}\right)\right\|_{\mathcal{E}} \le C_1|t|^{-\frac{1}{2(N-6)}}$$

for some constant $C_1 > 0$.

Remark 1.3. We construct here *pure* two-bubbles; that is, the solution approaches a superposition of two stationary states, with no energy transformed into radiation. By the conservation of energy and the decoupling of the two bubbles, we necessarily have $E(u(t)) = 2E(W)$. Pure one-bubbles cannot concentrate and are completely classified; see [Duyckaerts and Merle 2009].

Remark 1.4. For energy-critical wave equations, similar objects were constructed in [Jendrej 2016].

Remark 1.5. In dimension $N = 6$ one can expect an analogous result, with an exponential concentration rate.

Remark 1.6. In higher dimension, fast dispersion or dissipation sometimes excludes the possibility of a concentration of a bubble of energy for solutions which belong to a small neighborhood of a bubble. This was proved in [Collot et al. 2017] in the case of the critical heat equation; Perelman addressed the case for the Schrödinger equation in a lecture given at an IHES seminar in July 2016. We prove here that once we leave a small neighborhood of a bubble, concentration of a bubble of energy is possible in arbitrarily high dimension.

A similar phenomenon was observed by Martel and Raphaël [2015] for the mass-critical NLS.

Remark 1.7. We expect that the phases of the two bubbles forming the right angle is the only configuration in which a two-bubble can form.

Outline of the proof. The overall structure is similar to the earlier work of the author on the critical wave equations [Jendrej 2016]. We build a sequence $u_n : [T_n, T_0] \to \mathcal{E}$ of solutions of (1-1) with $T_n \to -\infty$ and $u_n(t)$ close to a two-bubble solution for $t \in [T_n, T_0]$. Taking a weak limit finishes the proof. This type of argument goes back to the works of Merle [1990] and Martel [2005]. The heart of the analysis is to obtain uniform energy bounds for the sequence u_n. This is achieved by means of a bootstrap argument, which can be resumed as follows.

We study solutions of (1-1) close to a sum of two bubbles:

$$u(t) = \mathrm{e}^{i\zeta(t)} W_\mu + \mathrm{e}^{i\theta(t)} W_{\lambda(t)} + g(t).$$

One should think of $\zeta(t)$ as being close to $-\frac{\pi}{2}$, $\mu(t) \simeq 1$, $\theta(t) \sim 0$, $\lambda(t) \ll 1$ and $\|g(t)\|_{\mathcal{E}} \ll 1$. In order to specify the values of the modulation parameters, we impose the orthogonality conditions, which make disappear terms linear in g in the modulation equations. There is essentially a unique choice of such orthogonality conditions. In Lemma 3.1 we establish bounds on the evolution of the modulation parameters under some bootstrap assumptions. The goal is to improve these bounds, thus closing the bootstrap. The essential point is to improve the estimate of g, which is the infinite-dimensional part. The novelty of this paper is to use the energy conservation to deal with this. Namely, the energy of the initial data is chosen close to $2E(W)$ and is conserved by the flow. It turns out that if we control the modulation parameters sufficiently well, we can improve the bound on $\|g\|_{\mathcal{E}}$ by simply expanding the formula for $E(u)$ and using coercivity of the energy near a ground state; see Step 3 of the proof of Proposition 4.4.

It remains to control the modulation parameters. Note that the interaction between the two bubbles appears explicitly in the modulation equation for $\lambda'(t)$; see (3-11). In fact, the configuration of the two bubbles (phases forming the right angle) is chosen so as to *maximize* the size of the term appearing in (3-11) and leading to the growth of the parameter λ. The critical part of the proof consists in improving the bound (3-7) on $\theta(t)$. To this end, we add a localized virial correction to $\theta(t)$ to cancel the main quadratic, which is $K(t)$ in the modulation equation (3-12). Note that the size of the term

$$\frac{K(t)}{\lambda(t)^2 \|W\|_{L^2}^2}$$

in (3-12) is $O(|t|^{-\frac{N-5}{N-6}})$. Adding the virial correction allows us to gain a small constant on the right-hand side of (3-12), which is decisive for closing the bootstrap.

Finally, in order to deal with the linear instabilities of the flow, we use a classical topological argument based on the Brouwer fixed point theorem.

Notation. For $z = x + iy \in \mathbb{C}$ we define $\Re(z) = x$ and $\Im(z) = y$. For two functions $v, w \in L^2(\mathbb{R}^N, \mathbb{C})$ we define

$$\langle v, w\rangle := \Re \int_{\mathbb{R}^N} \overline{v(x)} \cdot w(x)\,\mathrm{d}x.$$

In this paper all the functions are radially symmetric. We write $L^2 := L^2_{\mathrm{rad}}(\mathbb{R}^N;\mathbb{C})$ and $\mathcal{E} := \dot H^1_{\mathrm{rad}}(\mathbb{R}^N;\mathbb{C})$. We will think of them as of *real* vector spaces. We define $X^1 := \mathcal{E} \cap \dot H^2(\mathbb{R}^N)$.

2. Variational estimates

Linearization near a ground state. Recall that for $z \in \mathbb{C}$ we define

$$f(z) := |z|^{\frac{4}{N-2}} z \quad \text{and} \quad F(z) := \frac{N-2}{2N}|z|^{\frac{2N}{N-2}}.$$

For $z \in \mathbb{C}$ we define the $\mathbb{R}$-linear function $f'(z) : \mathbb{C} \to \mathbb{C}$ by

$$f'(z) z_1 := |z|^{\frac{4}{N-2}} \Big(z_1 + \frac{4}{N-2} z\, \Re(z^{-1} z_1)\Big)$$

(with the convention $f'(0) z_1 = 0$). It is easy to check that for any $z_1, z_2, z \in \mathbb{C}$,

$$\Re\big(\bar z_2 (f'(z) z_1)\big) = \Re\big(\bar z_1 (f'(z) z_2)\big) = \Re\big((\overline{f'(z) z_2}) z_1\big). \tag{2-1}$$

Integrating this identity on $\mathbb{R}^N$ we see that for a complex-valued function $u(x)$ the operator $g \mapsto f'(u) g$ is symmetric with respect to the real L^2 scalar product. We define

$$|f'(z)| := \frac{N+2}{N-2}|z|^{\frac{4}{N-2}},$$

which is the norm of $f'(z)$ as a linear map up to a constant. For $u : \mathbb{R}^N \to \mathbb{C}$ we define $\|f'(u)\|_{L^p} := \big(\int_{\mathbb{R}^N} |f'(u(x))|^p \,\mathrm{d}x\big)^{\frac{1}{p}}$ for $1 \le p < +\infty$ and $\|f'(u)\|_{L^\infty} := \sup_{x\in\mathbb{R}^N} |f'(u(x))|$.

Lemma 2.1. *Let $N \ge 7$. For $z_1, z_2, z_3 \in \mathbb{C}$,*

$$|f'(z_1 + z_2) - f'(z_1)| \lesssim |f'(z_2)|, \quad |f'(z_1 + z_2) - f'(z_1)| \lesssim |z_1|^{-\frac{N-6}{N-2}} |z_2| \quad \textit{if } z_1 \neq 0, \tag{2-2}$$

$$|f(z_1 + z_2) - f(z_1)| \lesssim |f'(z_1)|\,|z_2| + |f(z_2)|, \tag{2-3}$$

$$\begin{aligned} &|f(z_1 + z_2) - f(z_1) - f'(z_1) z_2| \lesssim f(|z_2|), \\ &|f(z_1 + z_2) - f(z_1) - f'(z_1) z_2| \lesssim |z_1|^{-\frac{N-6}{N-2}} |z_2|^2 \quad \textit{if } z_1 \neq 0, \end{aligned} \tag{2-4}$$

$$|F(z_1 + z_2) - F(z_1) - \Re(\overline{f(z_1)} z_2)| \lesssim |f'(z_1)|\,|z_2|^2 + F(z_2), \tag{2-5}$$

$$|F(z_1 + z_2) - F(z_1) - \Re(\overline{f(z_1)} z_2) - \Re(\overline{f'(z_1) z_2} z_2)| \lesssim F(z_2). \tag{2-6}$$

Remark 2.2. In (2-2), $|f'(z_1+z_2)-f'(z_1)|$ denotes the norm of $f'(z_1+z_2)-f'(z_1)$ as an $\mathbb{R}$-linear map.

Remark 2.3. Note that (2-4) implies $f'(z)$ is the derivative (in the real sense) of f at z; in particular, f is a C^1 function.

Proof. All the bounds are immediate if $|z_2| \ge \frac{1}{2}|z_1|$; hence we can assume $|z_2| < \frac{1}{2}|z_1|$, in particular $z_1 \neq 0$.

The formulas $f'(z_1)z_2 = f(z_1)f'(1)(z_1^{-1}z_2)$ and $f'(z_1+z_2)z_3 = f(z_1)f'(1+z_1^{-1}z_2)(z_1^{-1}z_3)$ allow us to reduce the proof to the case $z_1 = 1$. For $|z| < \frac{1}{2}$, the mappings $F(1+z)$, $f(1+z)$ and $f'(1+z)$ are real-analytic with respect to z and the required bounds follow by writing standard asymptotic expansions. □

We denote by $Z_{\theta,\lambda} := i\Delta + if'(\mathrm{e}^{i\theta}W_\lambda)$ the linearization of $i\Delta u + if(u)$ near $u = \mathrm{e}^{i\theta}W_\lambda$. In order to express $Z_{\theta,\lambda}$ in a more explicit way, we introduce the following notation:

$$V^+ := -\frac{N+2}{N-2}W^{\frac{4}{N-2}}, \quad V^- := -W^{\frac{4}{N-2}}, \quad L^+ := -\Delta + V^+, \quad L^- := -\Delta + V^-.$$

We also introduce the generators of the $\dot{H}^1$-critical and the L^2-critical scaling. For a function $v : \mathbb{R}^N \to \mathbb{C}$ we define

$$\Lambda v := -\frac{\partial}{\partial\lambda}\Big|_{\lambda=1}(v_\lambda) = \Big(\frac{N-2}{2} + x\cdot\nabla\Big)v,$$
$$\Lambda_0 v := -\frac{\partial}{\partial\lambda}\Big|_{\lambda=1}\Big(\frac{1}{\lambda}v_\lambda\Big) = \Big(\frac{N}{2} + x\cdot\nabla\Big)v.$$

It is known that for all $g \in \mathcal{E}$ we have $\langle g, L^-g\rangle \ge 0$ and $\ker L^- = \operatorname{span}(W)$. The operator L^+ has one simple strictly negative eigenvalue and, restricting to radially symmetric functions, $\ker L^+ = \operatorname{span}(\Lambda W)$; see for instance [Nakanishi and Roy 2016].

For future reference, we provide here the values of some integrals involving W and ΛW:

$$\int_{\mathbb{R}^N} W^2\,\mathrm{d}x = \tfrac{1}{2}(N(N-2))^{\frac{N}{2}}B\Big(\frac{N-4}{2}, \frac{N}{2}\Big), \tag{2-7}$$
$$\int_{\mathbb{R}^N} W^{\frac{N+2}{N-2}}\,\mathrm{d}x = \frac{1}{N}(N(N-2))^{\frac{N}{2}}, \tag{2-8}$$
$$-\frac{N+2}{N-2}\int_{\mathbb{R}^N} W^{\frac{4}{N-2}}\Lambda W\,\mathrm{d}x = \frac{N-2}{2N}(N(N-2))^{\frac{N}{2}}. \tag{2-9}$$

For the first integral, we use the formula $B(x,y) = \int_0^{+\infty} t^{x-1}(1+t)^{-x-y}\,\mathrm{d}t$. For the second, we write $W^{\frac{N+2}{N-2}} = -\Delta W$ and we integrate by parts. For the last integral, we write

$$-\frac{N+2}{N-2}W^{\frac{4}{N-2}}\Lambda W = V^+\Lambda W = \Delta\Lambda W$$

and we integrate by parts.

Using the definition of f', one can check that if $g_1 = \Re g$ and $g_2 = \Im g$, then

$$Z_{\theta,\lambda}(\mathrm{e}^{i\theta}g_\lambda) = \frac{\mathrm{e}^{i\theta}}{\lambda^2}(L^-g_2 - iL^+g_1)_\lambda.$$

In particular, we obtain

$$Z_{\theta,\lambda}(i\mathrm{e}^{i\theta}W_\lambda) = \frac{\mathrm{e}^{i\theta}}{\lambda^2}(L^-W)_\lambda = 0,$$
$$Z_{\theta,\lambda}(\mathrm{e}^{i\theta}\Lambda W_\lambda) = \frac{\mathrm{e}^{i\theta}}{\lambda^2}(-iL^+\Lambda W)_\lambda = 0.$$

This can also be seen by differentiating $i\Delta(\mathrm{e}^{i\theta}W_\lambda) + if(\mathrm{e}^{i\theta}W_\lambda)$ with respect to θ and λ.

Consider now the operator $Z_{\theta,\lambda}^*$. We claim that $\{\mathrm{e}^{i\theta}W_\lambda, i\mathrm{e}^{i\theta}\Lambda W_\lambda\} \subset \ker Z_{\theta,\lambda}^*$. Indeed, we have

$$\begin{aligned}\langle \mathrm{e}^{i\theta}W_\lambda, Z_{\theta,\lambda}(\mathrm{e}^{i\theta}g_\lambda)\rangle &= \left\langle \mathrm{e}^{i\theta}W_\lambda, \frac{\mathrm{e}^{i\theta}}{\lambda^2}(L^-g_2 - iL^+g_1)_\lambda\right\rangle \\ &= \langle W, L^-g_2\rangle = \langle L^-W, g_2\rangle = 0, \end{aligned} \tag{2-10}$$

$$\begin{aligned}\langle i\mathrm{e}^{i\theta}\Lambda W_\lambda, Z_{\theta,\lambda}(\mathrm{e}^{i\theta}g_\lambda)\rangle &= \left\langle i\mathrm{e}^{i\theta}\Lambda W_\lambda, \frac{\mathrm{e}^{i\theta}}{\lambda^2}(L^-g_2 - iL^+g_1)_\lambda\right\rangle \\ &= -\langle \Lambda W, L^+g_1\rangle = -\langle L^+\Lambda W, g_1\rangle = 0. \end{aligned} \tag{2-11}$$

One can show that there exist real functions $\mathcal{Y}^{(1)}, \mathcal{Y}^{(2)} \in \mathcal{S}$ and a real number $\nu > 0$ such that

$$L^+\mathcal{Y}^{(1)} = -\nu\mathcal{Y}^{(2)}, \quad L^-\mathcal{Y}^{(2)} = \nu\mathcal{Y}^{(1)} \tag{2-12}$$

(the proof given in [Duyckaerts and Merle 2009, Section 7] for $N = 5$ works in any dimension $N \geq 5$). We can assume that $\|\mathcal{Y}^{(1)}\|_{L^2} = \|\mathcal{Y}^{(2)}\|_{L^2} = 1$. We define

$$\alpha_{\theta,\lambda}^+ := \frac{\mathrm{e}^{i\theta}}{\lambda^2}(\mathcal{Y}_\lambda^{(2)} + i\mathcal{Y}_\lambda^{(1)}), \quad \alpha_{\theta,\lambda}^- := \frac{\mathrm{e}^{i\theta}}{\lambda^2}(\mathcal{Y}_\lambda^{(2)} - i\mathcal{Y}_\lambda^{(1)}). \tag{2-13}$$

For $g = g_1 + ig_2$ we have

$$\langle \alpha_{\theta,\lambda}^+, \mathrm{e}^{i\theta}g_\lambda\rangle = \langle \mathcal{Y}^{(2)}, g_1\rangle + \langle \mathcal{Y}^{(1)}, g_2\rangle \quad \text{and} \quad \langle \alpha_{\theta,\lambda}^-, \mathrm{e}^{i\theta}g_\lambda\rangle = \langle \mathcal{Y}^{(2)}, g_1\rangle - \langle \mathcal{Y}^{(1)}, g_2\rangle.$$

Note that

$$\begin{aligned}\langle W, \mathcal{Y}^{(1)}\rangle &= \frac{1}{\nu}\langle W, L^-\mathcal{Y}^{(2)}\rangle = \frac{1}{\nu}\langle L^-W, \mathcal{Y}^{(2)}\rangle = 0, \\ \langle \Lambda W, \mathcal{Y}^{(2)}\rangle &= -\frac{1}{\nu}\langle \Lambda W, L^+\mathcal{Y}^{(1)}\rangle = -\frac{1}{\nu}\langle L^+(\Lambda W), \mathcal{Y}^{(1)}\rangle = 0.\end{aligned}$$

It follows that

$$\langle \alpha_{\theta,\lambda}^+, i\mathrm{e}^{i\theta}W_\lambda\rangle = \langle \alpha_{\theta,\lambda}^-, i\mathrm{e}^{i\theta}W_\lambda\rangle = 0, \tag{2-14}$$

$$\langle \alpha_{\theta,\lambda}^+, \mathrm{e}^{i\theta}\Lambda W_\lambda\rangle = \langle \alpha_{\theta,\lambda}^-, \mathrm{e}^{i\theta}\Lambda W_\lambda\rangle = 0. \tag{2-15}$$

Since $\mathcal{Y}^{(2)} \neq W$, we also have

$$\langle \mathcal{Y}^{(1)}, \mathcal{Y}^{(2)}\rangle = \frac{1}{\nu}\langle \mathcal{Y}^{(2)}, L^-\mathcal{Y}^{(2)}\rangle > 0. \tag{2-16}$$

We claim that $\alpha_{\theta,\lambda}^+$ and $\alpha_{\theta,\lambda}^-$ are eigenfunctions of $Z_{\theta,\lambda}^*$, with eigenvalues $\frac{\nu}{\lambda^2}$ and $-\frac{\nu}{\lambda^2}$ respectively. Indeed, we have

$$\begin{aligned}\langle \alpha_{\theta,\lambda}^+, Z_{\theta,\lambda}(\mathrm{e}^{i\theta}g_\lambda)\rangle &= \left\langle \alpha_{\theta,\lambda}^+, \frac{\mathrm{e}^{i\theta}}{\lambda^2}(L^-g_2 - iL^+g_1)_\lambda\right\rangle \\ &= \frac{1}{\lambda^2}\big(\langle \mathcal{Y}^{(2)}, L^-g_2\rangle - \langle \mathcal{Y}^{(1)}, L^+g_1\rangle\big) = \frac{1}{\lambda^2}\big(\langle L^-\mathcal{Y}^{(2)}, g_2\rangle - \langle L^+\mathcal{Y}^{(1)}, g_1\rangle\big) \\ &= \frac{\nu}{\lambda^2}\big(\langle \mathcal{Y}^{(1)}, g_2\rangle + \langle \mathcal{Y}^{(2)}, g_1\rangle\big) = \frac{\nu}{\lambda^2}\langle \alpha_{\theta,\lambda}^+, \mathrm{e}^{i\theta}g_\lambda\rangle. \end{aligned} \tag{2-17}$$

Similarly, $\langle \alpha_{\theta,\lambda}^-, Z_{\theta,\lambda}(\mathrm{e}^{i\theta}g_\lambda)\rangle = -\frac{\nu}{\lambda^2}\langle \alpha_{\theta,\lambda}^-, \mathrm{e}^{i\theta}g_\lambda\rangle$.

Coercivity of the energy near a two-bubble. We consider $u \in \mathcal{E}$ of the form $u = \mathrm{e}^{i\zeta} W_\mu + \mathrm{e}^{i\theta} W_\lambda + g$ with

$$\left|\zeta + \tfrac{\pi}{2}\right| + |\mu - 1| + |\theta| + \lambda + \|g\|_{\mathcal{E}} \ll 1.$$

Moreover, we will assume that g satisfies

$$\langle i\mathrm{e}^{i\zeta}\Lambda W_\mu, g\rangle = \langle -\mathrm{e}^{i\zeta} W_\mu, g\rangle = \langle i\mathrm{e}^{i\theta}\Lambda W_\lambda, g\rangle = \langle -\mathrm{e}^{i\theta} W_\lambda, g\rangle = 0. \tag{2-18}$$

This choice of the orthogonality conditions is dictated by the kernel of $Z^*_{\theta,\lambda}$; see (2-10) and (2-11). In this section this has little importance, but will be crucial in the sequel.

When ζ, μ, θ, λ and g are known from the context, we define

$$a_1^+ := \langle \alpha^+_{\zeta,\mu}, g\rangle, \quad a_1^- := \langle \alpha^-_{\zeta,\mu}, g\rangle, \quad a_2^+ := \langle \alpha^+_{\theta,\lambda}, g\rangle, \quad a_2^- := \langle \alpha^-_{\theta,\lambda}, g\rangle.$$

Our objective to prove the following result.

Proposition 2.4. *There exist constants $\eta, C_0, C > 0$ depending only on N such that for all $u \in \mathcal{E}$ of the form $u = \mathrm{e}^{i\zeta} W_\mu + \mathrm{e}^{i\theta} W_\lambda + g$, with $\left|\zeta + \frac{\pi}{2}\right| + |\mu - 1| + |\theta| + \lambda + \|g\|_{\mathcal{E}} \le \eta$ and g verifying (2-18), we have*

$$|E(u) - 2E(W)| \le C\Big(\big(\left|\zeta+\tfrac{\pi}{2}\right| + |\mu-1| + |\theta| + \lambda\big)\lambda^{\frac{N-2}{2}} + \|g\|^2_{\mathcal{E}}\Big), \tag{2-19}$$

$$\|g\|^2_{\mathcal{E}} + C_0\theta\lambda^{\frac{N-2}{2}} \le C\bigg(\lambda^{\frac{N-2}{2}}\big(\left|\zeta+\tfrac{\pi}{2}\right| + |\mu-1| + |\theta|^3 + \lambda\big) + E(u) - 2E(W) + \sum_{j=1,2}\big((a_j^+)^2 + (a_j^-)^2\big)\bigg). \tag{2-20}$$

The scheme of the proof is the following. Inequality (2-6) yields the Taylor expansion of the energy:

$$\left|E(u) - E(\mathrm{e}^{i\zeta} W_\mu + \mathrm{e}^{i\theta} W_\lambda) - \langle \mathrm{D}E(\mathrm{e}^{i\zeta} W_\mu + \mathrm{e}^{i\theta} W_\lambda), g\rangle - \tfrac{1}{2}\langle \mathrm{D}^2 E(\mathrm{e}^{i\zeta} W_\mu + \mathrm{e}^{i\theta} W_\lambda) g, g\rangle\right| \lesssim \|g\|_{\mathcal{E}}^{\frac{2N}{N-2}}. \tag{2-21}$$

We just have to compute all the terms with a sufficiently high precision. We split this computation into a few lemmas.

Lemma 2.5. *Let ζ, μ, θ, λ be as in Proposition 2.4. Then*

$$\left|E(\mathrm{e}^{i\zeta} W_\mu + \mathrm{e}^{i\theta} W_\lambda) - 2E(W) - \tfrac{1}{N}(N(N-2))^{\frac{N}{2}}\theta\lambda^{\frac{N-2}{2}}\right| \le C\lambda^{\frac{N-2}{2}}\big(\left|\zeta + \tfrac{\pi}{2}\right| + |\mu - 1| + |\theta|^3 + \lambda\big), \tag{2-22}$$

with a constant C depending only on N.

Proof. Expanding the energy we find

$$\begin{aligned} E(\mathrm{e}^{i\zeta} W_\mu + \mathrm{e}^{i\theta} W_\lambda) = E(\mathrm{e}^{i\zeta} W_\mu) + E(\mathrm{e}^{i\theta} W_\lambda) + \Re\int_{\mathbb{R}^N} \mathrm{e}^{i(\zeta-\theta)}\nabla(W_\mu)\cdot\nabla(W_\lambda)\,\mathrm{d}x \\ - \int_{\mathbb{R}^N}\big(F(\mathrm{e}^{i\zeta} W_\mu + \mathrm{e}^{i\theta} W_\lambda) - F(\mathrm{e}^{i\zeta} W_\mu) - F(\mathrm{e}^{i\theta} W_\lambda)\big)\,\mathrm{d}x. \end{aligned} \tag{2-23}$$

By scaling invariance, $E(\mathrm{e}^{i\zeta} W_\mu) + E(\mathrm{e}^{i\theta} W_\lambda) = 2E(W)$. Integrating by parts we get

$$\Re\int_{\mathbb{R}^N} \mathrm{e}^{i(\zeta-\theta)}\nabla(W_\mu)\cdot\nabla(W_\lambda)\,\mathrm{d}x = -\Re\int_{\mathbb{R}^N} \overline{\mathrm{e}^{i\theta} W_\lambda}\,\Delta(\mathrm{e}^{i\zeta} W_\mu)\,\mathrm{d}x = \Re\int_{\mathbb{R}^N} \overline{\mathrm{e}^{i\theta} W_\lambda}\cdot f(\mathrm{e}^{i\zeta} W_\mu)\,\mathrm{d}x.$$

Hence (2-23) yields

$$E(e^{i\zeta}W_\mu + e^{i\theta}W_\lambda) \\ = 2E(W) - \int_{\mathbb{R}^N}\big(F(e^{i\zeta}W_\mu + e^{i\theta}W_\lambda) - F(e^{i\zeta}W_\mu) - F(e^{i\theta}W_\lambda) - \Re(\overline{e^{i\theta}W_\lambda}\cdot f(e^{i\zeta}W_\mu))\big)\,dx. \quad (2\text{-}24)$$

In the region $|x| \geq \sqrt{\lambda}$, using (2-5) with $z_1 = e^{i\zeta}W_\mu$ and $z_2 = e^{i\theta}W_\lambda$, we obtain

$$\big|F(e^{i\zeta}W_\mu + e^{i\theta}W_\lambda) - F(e^{i\zeta}W_\mu) - F(e^{i\theta}W_\lambda) - \Re(\overline{e^{i\theta}W_\lambda}\cdot f(e^{i\zeta}W_\mu))\big| \lesssim W_\lambda^2,$$

and we see that

$$\int_{|x|\geq\sqrt{\lambda}} W_\lambda^2 = \lambda^2\int_{|x|\geq 1/\sqrt{\lambda}} W^2\,dx \lesssim \lambda^2\int_{1/\sqrt{\lambda}}^{+\infty} r^{-2N+4}r^{N-1}\,dr = \lambda^{2+\frac{N-4}{2}} = \lambda^{\frac{N}{2}}.$$

In the region $|x| \leq \sqrt{\lambda}$ the last term in (2-24) is negligible, because $|\Re(\overline{e^{i\theta}W_\lambda}\cdot f(e^{i\zeta}W_\mu))| \lesssim W_\lambda$ and $\int_{|x|\leq\sqrt{\lambda}} W_\lambda\,dx \lesssim \lambda^{\frac{N+2}{2}}\int_0^{1/\sqrt{\lambda}} r^{-N+2}r^{N-1}\,dr \sim \lambda^{\frac{N}{2}}$. Similarly, the term $F(e^{i\zeta}W_\mu)$ is negligible. Using (2-5) with $z_1 = e^{i\theta}W_\lambda$ and $z_2 = e^{i\zeta}W_\mu$, we obtain

$$\big|F(e^{i\zeta}W_\mu + e^{i\theta}W_\lambda) - F(e^{i\theta}W_\lambda) - \Re(\overline{e^{i\zeta}W_\mu}\cdot f(e^{i\theta}W_\lambda))\big| \lesssim W_\lambda^{\frac{4}{N-2}},$$

and we see that

$$\int_{|x|\leq\sqrt{\lambda}} W_\lambda^{\frac{4}{N-2}}\,dx = \lambda^{N-2}\int_{|x|\leq 1/\sqrt{\lambda}} W^{\frac{4}{N-2}}\,dx \lesssim \lambda^{N-2}\int_{1/\sqrt{\lambda}}^{+\infty} r^{-4}r^{N-1}\,dr = \lambda^{N-2-\frac{N-4}{2}} = \lambda^{\frac{N}{2}}.$$

In order to complete the proof of (2-22), we thus need to check that

$$\Big|-\int_{|x|\leq\sqrt{\lambda}} \Re(\overline{e^{i\zeta}W_\mu}\cdot f(e^{i\theta}W_\lambda))\,dx - \frac{1}{N}(N(N-2))^{\frac{N}{2}}\theta\lambda^{\frac{N-2}{2}}\Big| \\ \lesssim \lambda^{\frac{N-2}{2}}\big(\big|\zeta + \tfrac{\pi}{2}\big| + |\mu - 1| + |\theta|^3 + \lambda\big). \quad (2\text{-}25)$$

The following holds:

$$\Big|\int_{|x|\leq\sqrt{\lambda}} \Re(\overline{e^{i\zeta}W_\mu}\cdot f(e^{i\theta}W_\lambda))\,dx - \Re(e^{i(\zeta-\theta)})\int_{\mathbb{R}^N} W_\lambda^{\frac{N+2}{N-2}}\,dx\Big| \\ \lesssim \int_{|x|\leq\sqrt{\lambda}} |W_\mu - 1|W_\lambda^{\frac{N+2}{N-2}}\,dx + \int_{|x|\geq\sqrt{\lambda}} W_\lambda^{\frac{N+2}{N-2}}\,dx \\ \lesssim (|\mu-1|+\lambda)\int_{|x|\leq\sqrt{\lambda}} W_\lambda^{\frac{N+2}{N-2}} + \int_{|x|\geq\sqrt{\lambda}} W_\lambda^{\frac{N+2}{N-2}}\,dx \\ \lesssim (|\mu-1|+\lambda)\lambda^{\frac{N-2}{2}} + \lambda^{\frac{N-2}{2}}\int_{|x|\geq 1/\sqrt{\lambda}} W^{\frac{N+2}{N-2}}\,dx \lesssim (|\mu-1|+\lambda)\lambda^{\frac{N-2}{2}}$$

and

$$\int_{\mathbb{R}^N} W_\lambda^{\frac{N+2}{N-2}}\,dx = \lambda^{\frac{N-2}{2}}\int_{\mathbb{R}^N} W^{\frac{N+2}{N-2}}\,dx = \frac{1}{N}(N(N-2))^{\frac{N}{2}}\lambda^{\frac{N-2}{2}}.$$

We have

$$|\Re(-i\mathrm{e}^{-i\theta})+\theta| = |\Im(\mathrm{e}^{-i\theta})+\theta| \lesssim |\theta|^3$$

and, using (2-8),

$$|\mathrm{e}^{i(\zeta-\theta)}+i\mathrm{e}^{-i\theta}| = |\mathrm{e}^{i\zeta}+i| \le \left|\zeta+\tfrac{\pi}{2}\right|;$$

hence

$$|\Re(\mathrm{e}^{i(\zeta-\theta)})+\theta| \lesssim |\theta|^3 + \left|\zeta+\tfrac{\pi}{2}\right| \lesssim |t|^{-\frac{3}{N-6}}. \tag{2-26}$$

The bound (2-25) follows now from (2-26), which finishes the proof. □

Lemma 2.6. *Under the assumptions of Proposition 2.4, we have*

$$\left|\left\langle \mathrm{D}E(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda), g\right\rangle\right| \lesssim \|g\|_{\mathcal{E}}\cdot\lambda^{\frac{N+2}{4}}. \tag{2-27}$$

Proof. Using the fact that $\mathrm{D}E(\mathrm{e}^{i\zeta}W_\mu) = \mathrm{D}E(\mathrm{e}^{i\theta}W_\lambda) = 0$, (2-27) is seen to be equivalent to

$$\left|\left\langle f(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda), g\right\rangle\right| \lesssim \|g\|_{\mathcal{E}}\cdot\lambda^{\frac{N+2}{4}}.$$

By the Sobolev inequality, it suffices to check that

$$\left\| f(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\right\|_{L^{2N/(N+2)}} \lesssim \lambda^{\frac{N+2}{4}}.$$

As usual, we consider separately the regions $|x|\le\sqrt{\lambda}$ and $|x|\ge\sqrt{\lambda}$. In the first region we have $W_\mu \lesssim W_\lambda$; hence (2-3) with $z_1 = W_\lambda$ and $z_2 = W_\mu$ yields

$$\left| f(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\right| \lesssim W_\lambda^{\frac{4}{N-2}}W_\mu + W_\mu^{\frac{N+2}{N-2}} \lesssim W_\lambda^{\frac{4}{N-2}}W_\mu \lesssim W_\lambda^{\frac{4}{N-2}}.$$

By a change of variable we obtain

$$\begin{aligned}\|W_\lambda^{\frac{4}{N-2}}\|_{L^{2N/(N+2)}(|x|\le\sqrt{\lambda})} &= \lambda^{N\cdot\frac{N+2}{2N}-\frac{N-2}{2}\cdot\frac{4}{N-2}}\|W^{\frac{4}{N-2}}\|_{L^{2N/(N+2)}(|x|\le 1/\sqrt{\lambda})} \\ &\lesssim \lambda^{\frac{N-2}{2}}\left(\int_0^{1/\sqrt{\lambda}} r^{-4\frac{2N}{N+2}}r^{N-1}\,\mathrm{d}r\right)^{\frac{N+2}{2N}} \\ &\sim \lambda^{\frac{N-2}{2}-\frac{(N-6)N}{2(N+2)}\cdot\frac{N+2}{2N}} = \lambda^{\frac{N+2}{4}}.\end{aligned}$$

In the region $|x|\ge\sqrt{\lambda}$ we have $W_\lambda \lesssim W_\mu$; hence (2-3) with $z_1 = W_\mu$ and $z_2 = W_\lambda$ yields

$$\left| f(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\right| \lesssim W_\mu^{\frac{4}{N-2}}W_\lambda + W_\lambda^{\frac{N+2}{N-2}} \lesssim W_\mu^{\frac{4}{N-2}}W_\lambda \lesssim W_\lambda,$$

and we have

$$\begin{aligned}\|W_\lambda\|_{L^{2N/(N+2)}(|x|\ge\sqrt{\lambda})} &= \lambda^2\|W\|_{L^{2N/(N+2)}(|x|\ge 1/\sqrt{\lambda})} \\ &\lesssim \lambda^2\left(\int_{1/\sqrt{\lambda}}^{+\infty} r^{-(N-2)\cdot\frac{2N}{N+2}}r^{N-1}\,\mathrm{d}r\right)^{\frac{N+2}{2N}} \sim \lambda^{2+\frac{(N-6)N}{2(N+2)}\cdot\frac{N+2}{2N}} = \lambda^{\frac{N+2}{4}}.\end{aligned}$$ □

We now examine coercivity of the quadratic part in (2-21).

Lemma 2.7. *There exist constants $c, C > 0$ such that*

- *for any real-valued radial $g \in \mathcal{E}$,*

$$\langle g, L^+ g\rangle \geq c\int_{\mathbb{R}^N} |\nabla g|^2\,\mathrm{d}x - C\big(\langle W, g\rangle^2 + \langle \mathcal{Y}^{(2)}, g\rangle^2\big), \tag{2-28}$$

$$\langle g, L^- g\rangle \geq c\int_{\mathbb{R}^N} |\nabla g|^2\,\mathrm{d}x - C\,\langle \Lambda W, g\rangle^2, \tag{2-29}$$

- *if $r_1 > 0$ is large enough, then for any real-valued radial $g \in \mathcal{E}$,*

$$(1-2c)\int_{|x|\leq r_1} |\nabla g|^2\,\mathrm{d}x + c\int_{|x|\geq r_1} |\nabla g|^2\,\mathrm{d}x + \int_{\mathbb{R}^N} V^+|g|^2\,\mathrm{d}x \geq -C\big(\langle W, g\rangle^2 + \langle \mathcal{Y}^{(2)}, g\rangle^2\big), \tag{2-30}$$

$$(1-2c)\int_{|x|\leq r_1} |\nabla g|^2\,\mathrm{d}x + c\int_{|x|\geq r_1} |\nabla g|^2\,\mathrm{d}x + \int_{\mathbb{R}^N} V^-|g|^2\,\mathrm{d}x \geq -C\,\langle \Lambda W, g\rangle^2, \tag{2-31}$$

- *if $r_2 > 0$ is small enough, then for any real-valued radial $g \in \mathcal{E}$,*

$$(1-2c)\int_{|x|\geq r_2} |\nabla g|^2\,\mathrm{d}x + c\int_{|x|\leq r_2} |\nabla g|^2\,\mathrm{d}x + \int_{\mathbb{R}^N} V^+|g|^2\,\mathrm{d}x \geq -C\big(\langle W, g\rangle^2 + \langle \mathcal{Y}^{(2)}, g\rangle^2\big), \tag{2-32}$$

$$(1-2c)\int_{|x|\geq r_2} |\nabla g|^2\,\mathrm{d}x + c\int_{|x|\leq r_2} |\nabla g|^2\,\mathrm{d}x + \int_{\mathbb{R}^N} V^-|g|^2\,\mathrm{d}x \geq -C\,\langle \Lambda W, g\rangle^2. \tag{2-33}$$

Proof. In the proofs of (2-28) and (2-29) we repeat with minor modifications the arguments of Nakanishi and Roy [2016]. We include them for the reader's convenience.

Let us show that

$$g \in \mathcal{E},\quad \langle \mathcal{Y}^{(2)}, g\rangle = 0 \quad\Longrightarrow\quad \langle g, L^+, g\rangle \geq 0. \tag{2-34}$$

Suppose the contrary. Let $(a, b) \in \mathbb{R}^2 \setminus \{(0, 0)\}$ and consider $ag + b\mathcal{Y}^{(1)} \in \mathcal{E}$. Since $\mathcal{Y}^{(2)} \neq W$, (2-12) yields

$$\langle \mathcal{Y}^{(1)}, L^+\mathcal{Y}^{(1)}\rangle = -\nu\langle \mathcal{Y}^{(1)}, \mathcal{Y}^{(2)}\rangle = -\langle L^-\mathcal{Y}^{(2)}, \mathcal{Y}^{(2)}\rangle < 0, \tag{2-35}$$

so we obtain

$$\begin{aligned}\langle ag + bY^{(1)}, L^+(ag + b\mathcal{Y}^{(1)})\rangle &= a^2\langle g, L^+g\rangle + 2ab\langle g, L^+\mathcal{Y}^{(1)}\rangle + b^2\langle \mathcal{Y}^{(1)}, L^+\mathcal{Y}^{(1)}\rangle \\ &= a^2\langle g, L^+g\rangle - 2ab\nu\langle g, \mathcal{Y}^{(2)}\rangle + b^2\langle \mathcal{Y}^{(1)}, L^+\mathcal{Y}^{(1)}\rangle < 0.\end{aligned}$$

This is impossible, because L^+ has only one negative direction. This proves (2-34).

Suppose (2-28) fails. Then there exists a sequence $g_n \in \mathcal{E}$ such that $\|g_n\|_{\mathcal{E}} = 1$ and

$$\langle g_n, L^+g_n\rangle \leq c_n - C_n\big(\langle W, g_n\rangle^2 + \langle \mathcal{Y}^{(2)}, g\rangle^2\big), \quad c_n \to 0,\ \ C_n \to +\infty. \tag{2-36}$$

Upon extracting a subsequence, we can assume that $g_n \rightharpoonup g \in \mathcal{E}$. Since $|\langle g_n, L^+g_n\rangle| \lesssim \|g_n\|_{\mathcal{E}}^2 = 1$, from (2-36) we immediately get $\langle W, g\rangle = \langle \mathcal{Y}^{(2)}, g\rangle = 0$. Also, by standard arguments $\langle g_n, V^+g_n\rangle \to \langle g, V^+g\rangle$. Hence by the Fatou property

$$\langle g, L^+g\rangle \leq \liminf_n \langle g_n, L^+, g_n\rangle \leq \liminf_n c_n = 0.$$

Thus g is a minimizer for the quadratic form associated with L^+ on the hyperplane orthogonal to $\mathcal{Y}^{(2)}$. This implies $\langle h, L^+ g\rangle = 0$ for all $h \in \mathcal{E}$ such that $\langle \mathcal{Y}^{(2)}, h\rangle = 0$. But we also have $\langle \mathcal{Y}^{(1)}, L^+ g\rangle = \langle L^+\mathcal{Y}^{(1)}, g\rangle = -\nu\langle \mathcal{Y}^{(2)}, g\rangle = 0$ and $\langle \mathcal{Y}^{(1)}, \mathcal{Y}^{(2)}\rangle \neq 0$, see (2-35), so we obtain $\langle h, L^+ g\rangle = 0$ for all $h \in \mathcal{E}$. Hence $g = \Lambda W$. But $\langle W, \Lambda W\rangle = -\|W\|_{L^2}^2 \neq 0$, so we get a contradiction. This proves (2-28).

The proof of (2-29) is similar. We obtain that the weak limit g is a minimizer for the quadratic form associated with L^- (without constraints); hence $g = W$, which is incompatible with the orthogonality condition.

Once we have (2-28) and (2-29), the bounds (2-30)–(2-33) follow by repeating the proof of Lemma 2.1 in [Jendrej 2015]. □

We now use this lemma to study the linearization around $\mathrm{e}^{i\theta} W_\lambda$ for a complex-valued perturbation g.

Proposition 2.8. *There exist constants $c, C > 0$ such that for any $\theta \in \mathbb{R}$ and $\lambda > 0$,*

- *for any complex-valued radial $g \in \mathcal{E}$,*

$$\int_{\mathbb{R}^N} |\nabla g|^2\,\mathrm{d}x - \Re\int_{\mathbb{R}^N} \bar{g}\cdot f'(\mathrm{e}^{i\theta}W_\lambda)g\,\mathrm{d}x \\ \geq c\int_{\mathbb{R}^N}|\nabla g|^2\,\mathrm{d}x - C\big(\langle\lambda^{-2}\mathrm{e}^{i\theta}W_\lambda, g\rangle^2 + \langle\lambda^{-2}i\mathrm{e}^{i\theta}\Lambda W_\lambda, g\rangle^2 + \langle\alpha_{\theta,\lambda}^+, g\rangle^2 + \langle\alpha_{\theta,\lambda}^-, g\rangle^2\big), \tag{2-37}$$

- *if $r_1 > 0$ is large enough, then for any complex-valued radial $g \in \mathcal{E}$,*

$$(1-2c)\int_{|x|\leq r_1}|\nabla g|^2\,\mathrm{d}x + c\int_{|x|\geq r_1}|\nabla g|^2\,\mathrm{d}x - \Re\int_{\mathbb{R}^N}\bar{g}\cdot f'(\mathrm{e}^{i\theta}W_\lambda)g\,\mathrm{d}x \\ \geq -C\big(\langle\lambda^{-2}\mathrm{e}^{i\theta}W_\lambda, g\rangle^2 + \langle\lambda^{-2}i\mathrm{e}^{i\theta}\Lambda W_\lambda, g\rangle^2 + \langle\alpha_{\theta,\lambda}^+, g\rangle^2 + \langle\alpha_{\theta,\lambda}^-, g\rangle^2\big), \tag{2-38}$$

- *if $r_2 > 0$ is small enough, then for any complex-valued radial $g \in \mathcal{E}$,*

$$(1-2c)\int_{|x|\geq r_2}|\nabla g|^2\,\mathrm{d}x + c\int_{|x|\leq r_2}|\nabla g|^2\,\mathrm{d}x - \Re\int_{\mathbb{R}^N}\bar{g}\cdot f'(\mathrm{e}^{i\theta}W_\lambda)g\,\mathrm{d}x \\ \geq -C\big(\langle\lambda^{-2}\mathrm{e}^{i\theta}W_\lambda, g\rangle^2 + \langle\lambda^{-2}i\mathrm{e}^{i\theta}\Lambda W_\lambda, g\rangle^2 + \langle\alpha_{\theta,\lambda}^+, g\rangle^2 + \langle\alpha_{\theta,\lambda}^-, g\rangle^2\big). \tag{2-39}$$

Remark 2.9. Note that the scalar products on the right-hand side of these estimates are the ones which appear in the orthogonality conditions (2-18). For the definition of $\alpha_{\theta,\lambda}^\pm$, see (2-13).

Proof. Without loss of generality we can assume that $\theta = 0$ and $\lambda = 1$. Let $g = g_1 + ig_2$. Observe that

$$-f'(W)(g_1 + ig_2) = -W^{\frac{4}{N-2}}(g_1 + ig_2) - \frac{4}{N-2}W^{\frac{4}{N-2}}g_1 = V^+g_1 + iV^-g_2,$$

which gives

$$-\Re\int_{\mathbb{R}^N}\bar{g}\cdot f'(W)g\,\mathrm{d}x = \int_{\mathbb{R}^N}V^+g_1^2\,\mathrm{d}x + \int_{\mathbb{R}^N}V^-g_2^2\,\mathrm{d}x.$$

Also, $\langle W, g\rangle = \langle W, g_1\rangle$ and $\langle i\Lambda W, g\rangle = \langle\Lambda W, g_2\rangle$. We have $\mathcal{Y}^{(2)} = \frac{1}{2}(\alpha_{0,1}^+ + \alpha_{0,1}^-)$, so

$$\langle\mathcal{Y}^{(2)}, g_1\rangle^2 = \langle\mathcal{Y}^{(2)}, g\rangle^2 \leq \tfrac{1}{2}\big(\langle\alpha_{0,1}^+, g\rangle^2 + \langle\alpha_{0,1}^-, g\rangle^2\big).$$

Applying (2-28) with $g = g_1$ and (2-29) with $g = g_2$ we obtain (2-37). The proofs of (2-38) and (2-39) are similar. □

One consequence of the last proposition is the coercivity near a sum of two bubbles at different scales:

Lemma 2.10. *There exist $\eta, C > 0$ such that if $\lambda \le \eta\mu$, then for all $g \in \mathcal{E}$ satisfying (2-18),*

$$\tfrac{1}{C}\|g\|_{\mathcal{E}}^2 \le \tfrac{1}{2}\langle \mathrm{D}^2 E(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda)g, g\rangle + 2\big((a_1^+)^2 + (a_1^-)^2 + (a_2^+)^2 + (a_2^-)^2\big) \le C\|g\|_{\mathcal{E}}^2.$$

Proof. It is essentially the same as the proof of [Jendrej 2015, Lemma 3.5]. □

Proof of Proposition 2.4. Bound (2-19) follows immediately from (2-21), Lemmas 2.5, 2.6 and 2.10 and the triangle inequality.

For any $c > 0$ we have $\|g\|_{\mathcal{E}}^{\frac{2N}{N-2}} \le c\|g\|_{\mathcal{E}}^2$ if η is chosen small enough; hence (2-21) and Lemmas 2.5 and 2.6 yield

$$\Big|E(u) - 2E(W) - \tfrac{1}{N}(N(N-2))^{\frac{N}{2}}\theta\lambda^{\frac{N-2}{2}} - \tfrac{1}{2}\langle \mathrm{D}^2E(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda)g, g\rangle\Big| \\ \le C\big(\big|\zeta + \tfrac{\pi}{2}\big| + |\mu - 1| + |\theta|^3 + \lambda\big)\lambda^{\frac{N-2}{2}} + c\|g\|_{\mathcal{E}}^2;$$

hence

$$\tfrac{1}{N}(N(N-2))^{\frac{N}{2}}\theta\lambda^{\frac{N-2}{2}} + \tfrac{1}{2}\langle \mathrm{D}^2E(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda)g, g\rangle \\ \le E(u) - 2E(W) + C\big(\big|\zeta + \tfrac{\pi}{2}\big| + |\mu - 1| + |\theta|^3 + \lambda\big)\lambda^{\frac{N-2}{2}} + c\|g\|_{\mathcal{E}}^2.$$

Choosing c small enough and invoking Lemma 2.10 finishes the proof of (2-20). □

3. Modulation

Bounds on the modulation parameters. We study solutions of the form

$$u(t) = \mathrm{e}^{i\zeta(t)}W_{\mu(t)} + \mathrm{e}^{i\theta(t)}W_{\lambda(t)} + g(t), \tag{3-1}$$

with

$$|\mu(t) - 1| \ll 1, \quad \big|\zeta(t) + \tfrac{\pi}{2}\big| \ll 1, \quad \lambda(t) \ll 1, \quad |\theta(t)| \ll 1 \quad \text{and} \quad \|g\|_{\mathcal{E}} \ll 1. \tag{3-2}$$

We will often omit the time variable and write ζ for $\zeta(t)$ etc.

Differentiating (3-1) in time we obtain

$$\partial_t u = \zeta' i\mathrm{e}^{i\zeta}W_\mu - \frac{\mu'}{\mu}\mathrm{e}^{i\zeta}\Lambda W_\mu + \theta' i\mathrm{e}^{i\theta}W_\lambda - \frac{\lambda'}{\lambda}\Lambda W_\lambda + \partial_t g.$$

On the other hand, using $\Delta(W_\mu) + f(W_\mu) = \Delta(W_\lambda) + f(W_\lambda) = 0$ we get

$$i\Delta u + if(u) = i\Delta g + i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\big);$$

hence (1-1) yields

$$\partial_t g = i\Delta g + i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\big) \\ - \zeta' i\mathrm{e}^{i\zeta}W_\mu + \frac{\mu'}{\mu}\mathrm{e}^{i\zeta}\Lambda W_\mu - \theta' i\mathrm{e}^{i\theta}W_\lambda + \frac{\lambda'}{\lambda}\mathrm{e}^{i\theta}\Lambda W_\lambda. \tag{3-3}$$

Since we work with nonclassical solutions, it is worth pointing out that the equation above should be understood as a notational simplification. Any computation involving $g(t)$ could be rewritten in terms of

$u(t)$ and the modulation parameters ζ, μ, θ, λ. Most of the time we only use the fact that (3-3) holds in the weak sense, but later we will also need to compute the time derivative of a quadratic form in $g(t)$, in which case the rigorous meaning of the computation is less clear.

We impose the orthogonality conditions (2-18). By standard arguments using the implicit function theorem, they uniquely determine the modulation parameters.

We need precise bootstrap assumptions about the parameters quantifying (3-2). In order to formulate them, set

$$\kappa := \left(\frac{N-6}{N\cdot B\big(\frac{N-4}{2},\frac{N}{2}\big)}\right)^{\frac{2}{N-4}}. \tag{3-4}$$

Lemma 3.1. *Let $c>0$ be an arbitrarily small constant. Let $T_0<0$ with $|T_0|$ large enough (depending on c) and $T<T_1\le T_0$. Suppose that for $T\le t\le T_1$ we have*

$$\big|\zeta(t)+\tfrac{\pi}{2}\big| \le |t|^{-\frac{3}{N-6}}, \tag{3-5}$$

$$|\mu(t)-1| \le |t|^{-\frac{3}{N-6}}, \tag{3-6}$$

$$|\theta(t)| \le |t|^{-\frac{1}{N-6}}, \tag{3-7}$$

$$\big|\lambda(t)-\tfrac{1}{\kappa}(\kappa|t|)^{-\frac{2}{N-6}}\big| \le |t|^{-\frac{5}{2(N-6)}}, \tag{3-8}$$

$$\|g\|_{\mathcal{E}} \le |t|^{-\frac{N-1}{2(N-6)}}. \tag{3-9}$$

Then

$$|\zeta'(t)| \le c|t|^{-\frac{N-3}{N-6}}, \tag{3-10}$$

$$|\mu'(t)| \le c|t|^{-\frac{N-3}{N-6}},$$

$$\left|\lambda'(t)-\frac{2\kappa^{\frac{N-4}{2}}}{N-6}\lambda(t)^{\frac{N-4}{2}}\right| \le c|t|^{-\frac{2N-7}{2(N-6)}}, \tag{3-11}$$

$$\left|\theta'(t)+\frac{(N-2)\kappa^{\frac{N-4}{2}}}{N-6}\theta(t)\lambda(t)^{\frac{N-6}{2}}-\frac{K(t)}{\lambda(t)^2\|W\|_{L^2}^2}\right| \le c|t|^{-\frac{N-5}{N-6}} \tag{3-12}$$

for $T\le t\le T_1$, where

$$K := -\big\langle e^{i\theta}\Lambda W_\lambda, f(e^{i\zeta}W_\mu+e^{i\theta}W_\lambda+g)-f(e^{i\zeta}W_\mu+e^{i\theta}W_\lambda)-f'(e^{i\zeta}W_\mu+e^{i\theta}W_\lambda)g\big\rangle. \tag{3-13}$$

Remark 3.2. We will not really use (3-8), but only the fact that $\lambda(t)\sim|t|^{-\frac{2}{N-6}}$.

Proof. We use the usual method of differentiating the orthogonality conditions in time, which will yield a linear system of the form

$$\begin{pmatrix} M_{11} & M_{12} & M_{13} & M_{14}\\ M_{21} & M_{22} & M_{23} & M_{24}\\ M_{31} & M_{32} & M_{33} & M_{34}\\ M_{41} & M_{42} & M_{43} & M_{44}\end{pmatrix}\begin{pmatrix}\mu^2\zeta'\\ \mu\mu'\\ \lambda^2\theta'\\ \lambda\lambda'\end{pmatrix}=\begin{pmatrix}B_1\\ B_2\\ B_3\\ B_4\end{pmatrix}.$$

Here, the coefficients M_{ij} and B_i depend on g, ζ, μ, θ and λ. We will now compute all these coefficients and prove appropriate bounds.

First row. Differentiating $\langle i\mathrm{e}^{i\zeta}\Lambda W_\mu, g\rangle = 0$ and using (3-3) we obtain

$$\begin{aligned}0 &= \frac{\mathrm{d}}{\mathrm{d}t}\langle i\mathrm{e}^{i\zeta}\Lambda W_\mu, g\rangle = -\zeta'\langle \mathrm{e}^{i\zeta}\Lambda W_\mu, g\rangle - \frac{\mu'}{\mu}\langle i\mathrm{e}^{i\zeta}\Lambda\Lambda W_\mu, g\rangle + \langle i\mathrm{e}^{i\zeta}\Lambda W_\mu, \partial_t g\rangle \\ &= \zeta'\big(-\langle i\mathrm{e}^{i\zeta}\Lambda W_\mu, i\mathrm{e}^{i\zeta}W_\mu\rangle - \langle \mathrm{e}^{i\zeta}\Lambda W_\mu, g\rangle\big) + \frac{\mu'}{\mu}\big(\langle i\mathrm{e}^{i\zeta}\Lambda W_\mu, \mathrm{e}^{i\zeta}\Lambda W_\mu\rangle - \langle i\mathrm{e}^{i\zeta}\Lambda\Lambda W_\mu, g\rangle\big) \\ &\quad + \theta'\langle i\mathrm{e}^{i\zeta}\Lambda W_\mu, -i\mathrm{e}^{i\theta}W_\lambda\rangle + \frac{\lambda'}{\lambda}\langle i\mathrm{e}^{i\zeta}\Lambda W_\mu, \mathrm{e}^{i\theta}\Lambda W_\lambda\rangle \\ &\qquad + \big\langle i\mathrm{e}^{i\zeta}\Lambda W_\mu, i\Delta g + i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\big)\big\rangle.\end{aligned}$$

Note that $\langle -\Lambda W_\mu, W_\mu\rangle = \|W_\mu\|_{L^2}^2 = \mu^2\|W\|_{L^2}^2$; hence we get

$$\begin{aligned}M_{11} &= \mu^{-2}\big(-\langle i\mathrm{e}^{i\zeta}\Lambda W_\mu, i\mathrm{e}^{i\zeta}W_\mu\rangle - \langle \mathrm{e}^{i\zeta}\Lambda W_\mu, g\rangle\big) = \|W\|_{L^2}^2 + O(\|g\|_{\mathcal{E}}) = \|W\|_{L^2}^2 + O(|t|^{-\frac{N-1}{2(N-6)}}), \\ M_{12} &= \mu^{-2}\big(\langle i\mathrm{e}^{i\zeta}\Lambda W_\mu, \mathrm{e}^{i\zeta}\Lambda W_\mu\rangle - \langle i\mathrm{e}^{i\zeta}\Lambda\Lambda W_\mu, g\rangle\big) = O(\|g\|_{\mathcal{E}}) = O(|t|^{-\frac{N-1}{2(N-6)}}), \\ M_{13} &= \lambda^{-2}\langle i\mathrm{e}^{i\zeta}\Lambda W_\mu, -i\mathrm{e}^{i\theta}W_\lambda\rangle = O(1), \\ M_{14} &= \lambda^{-2}\langle i\mathrm{e}^{i\zeta}\Lambda W_\mu, \mathrm{e}^{i\theta}\Lambda W_\lambda\rangle = O(1).\end{aligned}$$

Let us consider the term

$$B_1 = -\big\langle i\mathrm{e}^{i\zeta}\Lambda W_\mu, i\Delta g + i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\big)\big\rangle.$$

From (2-11) (with θ replaced by ζ and λ replaced by μ) we obtain

$$\begin{aligned}B_1 &= -\mu^{-2}\big\langle i\mathrm{e}^{i\zeta}\Lambda W_\mu, i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\zeta}W_\mu)g\big)\big\rangle \\ &= -\mu^{-2}\big\langle \mathrm{e}^{i\zeta}\Lambda W_\mu, \big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\zeta}W_\mu)g\big)\big\rangle.\end{aligned}$$

First we show that

$$\big|\big\langle \mathrm{e}^{i\zeta}\Lambda W_\mu, f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda)g\big\rangle\big| \lesssim \|g\|_{\mathcal{E}}^2. \tag{3-14}$$

Note that (3-5) and (3-7) imply $|\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda| \gtrsim W_\mu$; hence (2-4) with $z_1 = \mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda$ and $z_2 = g$ yields

$$\big|f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda)g\big| \lesssim W_\mu^{-\frac{N-6}{N-2}}|g|^2.$$

Using the fact that $|\Lambda W| \lesssim W$ and the Hölder inequality, we arrive at (3-14).

Next we show that

$$\big|\big\langle \mathrm{e}^{i\zeta}\Lambda W_\mu, f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\big\rangle\big| \lesssim \lambda^{\frac{N-2}{2}}. \tag{3-15}$$

Using (2-3) we get

$$\big|f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\big| \lesssim |W_\mu|^{\frac{4}{N-2}}W_\lambda + |f(W_\lambda)|.$$

The second term is easy. We have $f(W) \in L^1$ and we check that $\|f(W_\lambda)\|_{L^1} \sim \lambda^{\frac{N-2}{2}}$ by a change of variable. Consider the first term. In the region $|x| \le 1$ we write

$$\|W_\lambda\|_{L^1(|x|\le 1)} = \lambda^{\frac{N+2}{2}} \|W\|_{L^1(|x|\le\lambda^{-1})} \lesssim \lambda^{\frac{N+2}{2}} \int_0^{\lambda^{-1}} r^{-N+2} r^{N-1}\,\mathrm{d}r \sim \lambda^{\frac{N-2}{2}}.$$

As for $|x| \ge 1$, we notice that $\|W_\lambda\|_{L^\infty(|x|\ge 1)} \lesssim \lambda^{\frac{N-2}{2}}$ and $|\Lambda W_\mu||W_\mu|^{\frac{4}{N-2}}$ is bounded in L^1.

Finally, we show that

$$\big|\langle \mathrm{e}^{i\zeta}\Lambda W_\mu, (f'(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\zeta}W_\mu))g\rangle\big| \lesssim \lambda^{\frac{N-2}{4}}\|g\|_{\mathcal{E}}. \tag{3-16}$$

In the region $|x| \le \sqrt{\lambda}$ it suffices to use the bound

$$|f'(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\zeta}W_\mu)| \lesssim W_\lambda^{\frac{4}{N-2}}$$

and the fact that

$$\|W_\lambda^{\frac{4}{N-2}}\|_{L^{2N/(N+2)}(|x|\le\sqrt{\lambda})} = \lambda^{\frac{N-2}{2}}\|W^{\frac{4}{N-2}}\|_{L^{2N/(N+2)}(|x|\le\lambda^{-1/2})} \lesssim \lambda^{\frac{N+2}{4}},$$

where the last inequality follows from $W^{\frac{4}{N-2}}(x) \lesssim |x|^{-4}$. In the region $|x| \ge \sqrt{\lambda}$ we use Hölder and the fact that

$$\|W_\lambda\|_{L^{2N/N-2}(|x|\ge\sqrt{\lambda})} = \|W\|_{L^{2N/(N-2)}(|x|\ge\lambda^{-1/2})} \lesssim \left(\int_{\lambda^{-1/2}}^{+\infty} r^{-2N} r^{N-1}\,\mathrm{d}r\right)^{\frac{N-2}{2N}} \lesssim \lambda^{\frac{N-2}{4}}.$$

Taking the sum of (3-14), (3-15), (3-16) and using (3-8), (3-9) we obtain

$$|B_1| \lesssim |t|^{-\frac{N-2}{N-6}}. \tag{3-17}$$

Second row. Differentiating $\langle -\mathrm{e}^{i\zeta}W_\mu, g\rangle = 0$, we obtain

$$\begin{aligned}
0 = \frac{\mathrm{d}}{\mathrm{d}t}\langle -\mathrm{e}^{i\zeta}W_\mu, g\rangle &= -\zeta'\langle i\mathrm{e}^{i\zeta}W_\mu, g\rangle + \frac{\mu'}{\mu}\langle \mathrm{e}^{i\zeta}\Lambda W_\mu, g\rangle - \langle \mathrm{e}^{i\zeta}W_\mu, \partial_t g\rangle \\
&= \zeta'\big(\langle \mathrm{e}^{i\zeta}W_\mu, i\mathrm{e}^{i\zeta}W_\mu\rangle - \langle i\mathrm{e}^{i\zeta}\Lambda W_\mu, g\rangle\big) + \frac{\mu'}{\mu}\big(-\langle \mathrm{e}^{i\zeta}W_\mu, \mathrm{e}^{i\zeta}\Lambda W_\mu\rangle + \langle \mathrm{e}^{i\zeta}\Lambda W_\mu, g\rangle\big) \\
&\quad + \theta'\langle \mathrm{e}^{i\zeta}W_\mu, i\mathrm{e}^{i\theta}W_\lambda\rangle + \frac{\lambda'}{\lambda}\langle -\mathrm{e}^{i\zeta}W_\mu, \mathrm{e}^{i\theta}\Lambda W_\lambda\rangle \\
&\qquad - \big\langle \mathrm{e}^{i\zeta}W_\mu, i\Delta g + i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\big)\big\rangle,
\end{aligned}$$

which yields

$$\begin{aligned}
M_{21} &= \mu^{-2}\big(\langle \mathrm{e}^{i\zeta}W_\mu, i\mathrm{e}^{i\zeta}W_\mu\rangle - \langle i\mathrm{e}^{i\zeta}W_\mu, g\rangle\big) = O(\|g\|_{\mathcal{E}}), \\
M_{22} &= \mu^{-2}\big(-\langle \mathrm{e}^{i\zeta}W_\mu, \mathrm{e}^{i\zeta}\Lambda W_\mu\rangle + \langle \mathrm{e}^{i\zeta}\Lambda W_\mu, g\rangle\big) = \|W\|_{L^2}^2 + O(\|g\|_{\mathcal{E}}), \\
M_{23} &= \lambda^{-2}\langle \mathrm{e}^{i\zeta}W_\mu, i\mathrm{e}^{i\theta}W_\lambda\rangle = O(1), \\
M_{24} &= \lambda^{-2}\langle -\mathrm{e}^{i\zeta}W_\mu, \mathrm{e}^{i\theta}\Lambda W_\lambda\rangle = O(1).
\end{aligned}$$

Consider now the term

$$\begin{aligned}
B_2 &= \big\langle \mathrm{e}^{i\zeta}W_\mu, i\Delta g + i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\big)\big\rangle \\
&= \big\langle \mathrm{e}^{i\zeta}W_\mu, i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\zeta}W_\mu)g\big)\big\rangle,
\end{aligned}$$

where the second equality follows from (2-10). The proof of (3-17) yields

$$|B_2| \lesssim |t|^{-\frac{N-2}{N-6}}.$$

Third row. Differentiating $\langle i\mathrm{e}^{i\theta}\Lambda W_\lambda, g\rangle = 0$ we obtain

$$\begin{aligned}
0 = \frac{\mathrm{d}}{\mathrm{d}t}\langle i\mathrm{e}^{i\theta}\Lambda W_\lambda, g\rangle &= -\theta'\langle \mathrm{e}^{i\theta}\Lambda W_\lambda, g\rangle - \frac{\lambda'}{\lambda}\langle i\mathrm{e}^{i\theta}\Lambda\Lambda W_\lambda, g\rangle + \langle i\mathrm{e}^{i\theta}\Lambda W_\lambda, \partial_t g\rangle \\
&= \zeta'\langle i\mathrm{e}^{i\theta}\Lambda W_\lambda, -i\mathrm{e}^{i\zeta}W_\mu\rangle + \frac{\mu'}{\mu}\langle i\mathrm{e}^{i\theta}\Lambda W_\lambda, \mathrm{e}^{i\zeta}\Lambda W_\mu\rangle \\
&\quad + \theta'\big(\langle i\mathrm{e}^{i\theta}\Lambda W_\lambda, -i\mathrm{e}^{i\theta}W_\lambda\rangle - \langle \mathrm{e}^{i\theta}\Lambda W_\lambda, g\rangle\big) + \frac{\lambda'}{\lambda}\big(\langle i\mathrm{e}^{i\theta}\Lambda W_\lambda, \mathrm{e}^{i\theta}\Lambda W_\lambda\rangle - \langle i\mathrm{e}^{i\theta}\Lambda\Lambda W_\lambda, g\rangle\big) \\
&\quad + \big\langle i\mathrm{e}^{i\theta}\Lambda W_\lambda, i\Delta g + i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\big)\big\rangle,
\end{aligned}$$

which yields

$$\begin{aligned}
M_{31} &= \mu^{-2}\langle i\mathrm{e}^{i\theta}\Lambda W_\lambda, -i\mathrm{e}^{i\zeta}W_\mu\rangle = O(\lambda^2) = O(|t|^{-\frac{4}{N-6}}), \\
M_{32} &= \mu^{-2}\langle i\mathrm{e}^{i\theta}\Lambda W_\lambda, \mathrm{e}^{i\zeta}\Lambda W_\mu\rangle = O(\lambda^2) = O(|t|^{-\frac{4}{N-6}}), \\
M_{33} &= \lambda^{-2}\big(\langle i\mathrm{e}^{i\theta}\Lambda W_\lambda, -i\mathrm{e}^{i\theta}W_\lambda\rangle - \langle \mathrm{e}^{i\theta}\Lambda W_\lambda, g\rangle\big) = \|W\|_{L^2}^2 + O(\|g\|_{\mathcal{E}}) = \|W\|_{L^2}^2 + O(|t|^{-\frac{N-1}{2(N-6)}}), \\
M_{34} &= \lambda^{-2}\big(\langle i\mathrm{e}^{i\theta}\Lambda W_\lambda, \mathrm{e}^{i\theta}\Lambda W_\lambda\rangle - \langle i\mathrm{e}^{i\theta}\Lambda\Lambda W_\lambda, g\rangle\big) = O(\|g\|_{\mathcal{E}}) = O(|t|^{-\frac{N-1}{2(N-6)}}).
\end{aligned}$$

Let us consider the term

$$\begin{aligned}
B_3 &= -\big\langle i\mathrm{e}^{i\theta}\Lambda W_\lambda, i\Delta g + i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\big)\big\rangle \\
&= -\big\langle i\mathrm{e}^{i\theta}\Lambda W_\lambda, i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\theta}W_\lambda)g\big)\big\rangle \\
&= -\big\langle \mathrm{e}^{i\theta}\Lambda W_\lambda, f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\theta}W_\lambda)g\big\rangle,
\end{aligned}$$

where the second equality follows from (2-11). Comparing this formula with (3-13) we obtain

$$\begin{aligned}
B_3 - K = -\big\langle \mathrm{e}^{i\theta}\Lambda W_\lambda, f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\big\rangle& \\
-\big\langle \mathrm{e}^{i\theta}\Lambda W_\lambda, \big(f'(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\theta}W_\lambda)\big)g\big\rangle&. \qquad (3\text{-}18)
\end{aligned}$$

First we treat the second line by showing that

$$\big|\big\langle \mathrm{e}^{i\theta}\Lambda W_\lambda, \big(f'(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\theta}W_\lambda)\big)g\big\rangle\big| \lesssim \lambda^{\frac{N}{4}}\|g\|_{\mathcal{E}}. \qquad (3\text{-}19)$$

We consider separately $|x| \le \lambda^\gamma$ and $|x| \ge \lambda^\gamma$ with $\gamma = \frac{N-4}{2(N-2)}$. In the region $|x| \le \lambda^\gamma$ we use the bound

$$|f'(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\theta}W_\lambda)| \lesssim W_\lambda^{-\frac{N-6}{N-2}}W_\mu.$$

It implies

$$|\mathrm{e}^{i\theta}\Lambda W_\lambda|\big|\big(f'(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\theta}W_\lambda)\big)g\big| \lesssim W_\lambda^{\frac{4}{N-2}}|g|$$

pointwise and it suffices to see that

$$\|W_\lambda^{\frac{4}{N-2}}\|_{L^{2N/(N+2)}(|x|\le\lambda^\gamma)} = \lambda^{\frac{N-2}{2}}\|W^{\frac{4}{N-2}}\|_{L^{2N/(N+2)}(|x|\le\lambda^{\gamma-1})} \lesssim \lambda^{\frac{N-2}{2}}\Big(\int_0^{\lambda^{\gamma-1}} r^{-4\frac{2N}{N+2}} r^{N-1}\,\mathrm{d}r\Big)^{\frac{N+2}{2N}}$$
$$\lesssim \lambda^{\frac{N-2}{2}}\lambda^{(\gamma-1)\frac{N(N-6)}{N+2}\cdot\frac{N+2}{2N}} = \lambda^{\frac{N-2}{2}-\frac{N(N-6)}{4(N-2)}} = \lambda^{\frac{N^2-2N+8}{4(N-2)}} \ll \lambda^{\frac{N}{4}}.$$

In the region $|x|\ge\lambda^\gamma$ we have

$$\|\Lambda W_\lambda\|_{L^{2N/(N-2)}(|x|\ge\lambda^\gamma)} \lesssim \|W_\lambda\|_{L^{2N/(N-2)}(|x|\ge\lambda^\gamma)} = \|W\|_{L^{2N/(N-2)}(|x|\ge\lambda^{\gamma-1})}$$
$$\lesssim \Big(\int_{\lambda^{\gamma-1}}^{+\infty} r^{-2N} r^{N-1}\,\mathrm{d}r\Big)^{\frac{N-2}{2N}} \lesssim \lambda^{(1-\gamma)N\frac{N-2}{2N}} = \lambda^{\frac{N}{4}},$$

which yields the required bound by Hölder.

We are left with the first line in (3-18). We will prove that

$$\Big|\langle e^{i\theta}\Lambda W_\lambda, f(e^{i\zeta}W_\mu+e^{i\theta}W_\lambda)-f(e^{i\zeta}W_\mu)-f(e^{i\theta}W_\lambda)\rangle - \frac{(N-2)\kappa^{\frac{N-4}{2}}\|W\|_{L^2}^2}{N-6}\theta\lambda^{\frac{N-2}{2}}\Big| \lesssim |t|^{-\frac{N}{N-6}}. \tag{3-20}$$

For this, we first check that

$$\big|\langle e^{i\theta}\Lambda W_\lambda, f(e^{i\zeta}W_\mu+e^{i\theta}W_\lambda)-f(e^{i\zeta}W_\mu)-f(e^{i\theta}W_\lambda)-f'(e^{i\theta}W_\lambda)(e^{i\zeta}W_\mu)\rangle\big| \lesssim \lambda^{\frac{N}{2}}. \tag{3-21}$$

In the region $|x|\ge\sqrt{\lambda}$ we have $W_\lambda\lesssim W_\mu$, which implies

$$\big|f(e^{i\zeta}W_\mu+e^{i\theta}W_\lambda)-f(e^{i\zeta}W_\mu)-f(e^{i\theta}W_\lambda)-f'(e^{i\theta}W_\lambda)(e^{i\zeta}W_\mu)\big| \lesssim W_\lambda^{\frac{4}{N-2}}W_\mu;$$

hence the required bound follows from $|\Lambda W|\lesssim W$ and

$$\|W_\lambda^{\frac{N+2}{N-2}}\|_{L^1(|x|\ge\sqrt{\lambda})} \lesssim \lambda^{\frac{N-2}{2}}\int_{\lambda^{-1/2}}^{+\infty} r^{-N-2}r^{N-1}\,\mathrm{d}r \sim \lambda^{\frac{N}{2}}.$$

In the region $|x|\le\sqrt{\lambda}$ we have $W_\mu\lesssim W_\lambda$, which implies

$$\big|f(e^{i\zeta}W_\mu+e^{i\theta}W_\lambda)-f(e^{i\zeta}W_\mu)-f(e^{i\theta}W_\lambda)-f'(e^{i\theta}W_\lambda)(e^{i\zeta}W_\mu)\big| \lesssim W_\mu^{\frac{N+2}{N-2}};$$

hence the required bound follows from

$$\|W_\lambda\|_{L^1(|x|\le\sqrt{\lambda})} \lesssim \lambda^{\frac{N+2}{2}}\int_0^{\lambda^{-1/2}} r^{-N+2}r^{N-1}\,\mathrm{d}r \sim \lambda^{\frac{N}{2}}.$$

Finally, we need to check that

$$\Big|\langle e^{i\theta}\Lambda W_\lambda, f'(e^{i\theta}W_\lambda)(e^{i\zeta}W_\mu)\rangle - \frac{(N-2)\kappa^{\frac{N-4}{2}}\|W\|_{L^2}^2}{N-6}\theta\lambda^{\frac{N-2}{2}}\Big| \lesssim |t|^{-\frac{N}{N-6}}. \tag{3-22}$$

The definition of $f'(z)$ yields

$$f'(e^{i\theta}W_\lambda)(e^{i\zeta}W_\mu) = W_\mu W_\lambda^{\frac{4}{N-2}}\Big(e^{i\zeta}+\frac{4}{N-2}e^{i\theta}\Re(e^{i(\zeta-\theta)})\Big); \tag{3-23}$$

hence

$$\langle \mathrm{e}^{i\theta}\Lambda W_\lambda, f'(\mathrm{e}^{i\theta}W_\lambda)(\mathrm{e}^{i\zeta}W_\mu)\rangle = \frac{N+2}{N-2}\Re(\mathrm{e}^{i(\zeta-\theta)})\int W_\mu W_\lambda^{\frac{4}{N-2}}\Lambda W_\lambda\,\mathrm{d}x. \tag{3-24}$$

Since

$$\left|\int W_\mu W_\lambda^{\frac{4}{N-2}}\Lambda W_\lambda\,\mathrm{d}x\right| \lesssim \lambda^{\frac{N-2}{2}} \lesssim |t|^{-\frac{N-2}{N-6}},$$

we obtain

$$\left|\frac{N+2}{N-2}\Re(\mathrm{e}^{i(\zeta-\theta)})\int W_\mu W_\lambda^{\frac{4}{N-2}}\Lambda W_\lambda\,\mathrm{d}x + \frac{N+2}{N-2}\theta\int W_\mu W_\lambda^{\frac{4}{N-2}}\Lambda W_\lambda\,\mathrm{d}x\right| \lesssim |t|^{-\frac{N+1}{N-6}} \ll |t|^{-\frac{N}{N-6}}. \tag{3-25}$$

Next, we prove that

$$\left|\int W_\lambda^{\frac{4}{N-2}}\Lambda W_\lambda\,\mathrm{d}x - \int W_\mu W_\lambda^{\frac{4}{N-2}}\Lambda W_\lambda\,\mathrm{d}x\right| \lesssim \lambda^{\frac{N}{2}} + |\mu-1|\lambda^{\frac{N-2}{2}} \lesssim |t|^{-\frac{N}{N-6}}. \tag{3-26}$$

Indeed, in the region $|x| \geq \sqrt{\lambda}$ both terms satisfy the bound. In the region $|x| \leq \sqrt{\lambda}$ we have

$$|W_\mu - \mu^{-\frac{N-2}{2}}| \lesssim |x|^2 \lesssim \lambda \quad \text{and} \quad |\mu^{-\frac{N-2}{2}} - 1| \lesssim |\mu - 1|,$$

from which (3-26) follows.

From (2-9) and (2-7) we get

$$\frac{N+2}{N-2}\int W_\lambda^{\frac{4}{N-2}}\Lambda W_\lambda\,\mathrm{d}x = -\frac{(N-2)\kappa^{\frac{N-4}{2}}\|W\|_{L^2}^2}{N-6}\lambda^{\frac{N-2}{2}},$$

and (3-22) follows from (3-24)–(3-26).

From (3-18)–(3-20) and the triangle inequality we infer

$$\left|B_3 - K + \frac{(N-2)\kappa^{\frac{N-4}{2}}\|W\|_{L^2}^2}{N-6}\theta\lambda^{\frac{N-2}{2}}\right| \lesssim |t|^{-\frac{N}{N-6}} + |t|^{-\frac{N}{2(N-6)}}\|g\|_{\mathcal{E}} \lesssim |t|^{-\frac{2N-1}{2(N-6)}}.$$

In particular, since $|\theta\lambda^{\frac{N-2}{2}}| \leq |t|^{-\frac{N-1}{N-6}}$, we have

$$|B_3| \lesssim |t|^{-\frac{N-1}{N-6}} + |t|^{-\frac{N}{N-6}} + |t|^{-\frac{N}{2(N-6)}}\|g\|_{\mathcal{E}} + \|g\|_{\mathcal{E}}^2 \lesssim |t|^{-\frac{N-1}{N-6}} + C_0^2|t|^{-\frac{N}{N-6}} \lesssim |t|^{-\frac{N-1}{N-6}}.$$

Fourth row. Differentiating $\langle -\mathrm{e}^{i\theta}W_\lambda, g\rangle = 0$ we obtain

$$\begin{aligned}
0 = \frac{\mathrm{d}}{\mathrm{d}t}\langle -\mathrm{e}^{i\theta}W_\lambda, g\rangle &= -\theta'\langle i\mathrm{e}^{i\theta}W_\lambda, g\rangle + \frac{\lambda'}{\lambda}\langle \mathrm{e}^{i\theta}\Lambda W_\lambda, g\rangle - \langle \mathrm{e}^{i\theta}W_\lambda, \partial_t g\rangle \\
&= \zeta'\langle i\mathrm{e}^{i\theta}W_\lambda, i\mathrm{e}^{i\zeta}W_\mu\rangle - \frac{\mu'}{\mu}\langle \mathrm{e}^{i\theta}W_\lambda, \mathrm{e}^{i\zeta}\Lambda W_\mu\rangle \\
&\quad + \theta'\big(\langle \mathrm{e}^{i\theta}W_\lambda, i\mathrm{e}^{i\theta}W_\lambda\rangle - \langle i\mathrm{e}^{i\theta}W_\lambda, g\rangle\big) + \frac{\lambda'}{\lambda}\big(\langle -\mathrm{e}^{i\theta}W_\lambda, \mathrm{e}^{i\theta}\Lambda W_\lambda\rangle + \langle \mathrm{e}^{i\theta}\Lambda W_\lambda, g\rangle\big) \\
&\quad - \big\langle \mathrm{e}^{i\theta}W_\lambda, i\Delta g + i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\big)\big\rangle,
\end{aligned}$$

which yields

$$M_{41} = \mu^{-2}\langle i\,\mathrm{e}^{i\theta}W_\lambda, i\,\mathrm{e}^{i\zeta}W_\mu\rangle = O(\lambda^2) = O(|t|^{-\frac{4}{N-6}}),$$
$$M_{42} = \mu^{-2}\langle \mathrm{e}^{i\theta}W_\lambda, \mathrm{e}^{i\zeta}\Lambda W_\mu\rangle = O(\lambda^2) = O(|t|^{-\frac{4}{N-6}}),$$
$$M_{43} = \lambda^{-2}\big(\langle \mathrm{e}^{i\theta}W_\lambda, i\,\mathrm{e}^{i\theta}W_\lambda\rangle - \langle i\,\mathrm{e}^{i\theta}W_\lambda, g\rangle\big) = O(\|g\|_{\mathcal{E}}) = O(|t|^{-\frac{N-1}{2(N-6)}}),$$
$$M_{44} = \lambda^{-2}\big(\langle -\mathrm{e}^{i\theta}W_\lambda, \mathrm{e}^{i\theta}\Lambda W_\lambda\rangle + \langle \mathrm{e}^{i\theta}\Lambda W_\lambda, g\rangle\big) = \|W\|_{L^2}^2 + O(\|g\|_{\mathcal{E}}) = \|W\|_{L^2}^2 + O(|t|^{-\frac{N-1}{2(N-6)}}).$$

Let us consider the term

$$\begin{aligned}B_4 &= \big\langle \mathrm{e}^{i\theta}W_\lambda, i\Delta g + i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\big)\big\rangle \\ &= \big\langle \mathrm{e}^{i\theta}W_\lambda, i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\theta}W_\lambda)g\big)\big\rangle,\end{aligned}$$

where the last equality follows from (2-10).

First we show that

$$\big|\big\langle \mathrm{e}^{i\theta}W_\lambda, i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda)g\big)\big\rangle\big| \lesssim \|g\|_{\mathcal{E}}^2. \tag{3-27}$$

Note that (3-5) and (3-7) imply $|\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda| \gtrsim W_\lambda$; hence (2-4) with $z_1 = \mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda$ and $z_2 = g$ yields

$$\big|f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda)g\big| \lesssim W_\lambda^{-\frac{N-6}{N-2}}|g|^2.$$

Using the fact that $|\Lambda W| \lesssim W$ and the Hölder inequality we arrive at (3-27).

The proof of (3-19) yields

$$\big|\big\langle \mathrm{e}^{i\theta}W_\lambda, i\big(f'(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\theta}W_\lambda)\big)g\big\rangle\big| \lesssim \lambda^{\frac{N}{4}}\|g\|_{\mathcal{E}}. \tag{3-28}$$

The proof of (3-21) yields

$$\big|\big\langle \mathrm{e}^{i\theta}W_\lambda, i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\theta}W_\lambda)(\mathrm{e}^{i\zeta}W_\mu)\big)\big\rangle\big| \lesssim \lambda^{\frac{N}{2}}. \tag{3-29}$$

Finally, we show that

$$\left|\big\langle \mathrm{e}^{i\theta}W_\lambda, if'(\mathrm{e}^{i\theta}W_\lambda)(\mathrm{e}^{i\zeta}W_\mu)\big\rangle - \frac{2\kappa^{\frac{N-4}{2}}\|W\|_{L^2}^2}{N-6}\lambda^{\frac{N-2}{2}}\right| \lesssim |t|^{-\frac{N}{N-6}}. \tag{3-30}$$

Using again (3-23) we get

$$\big\langle \mathrm{e}^{i\theta}W_\lambda, if'(\mathrm{e}^{i\theta}W_\lambda)(\mathrm{e}^{i\zeta}W_\mu)\big\rangle = \Re(i\,\mathrm{e}^{i(\zeta-\theta)})\int W_\mu W_\lambda^{\frac{N+2}{N-2}}\,\mathrm{d}x. \tag{3-31}$$

We have $|\Re(\mathrm{e}^{-i\theta}) - 1| \lesssim |\theta|^2 \le |t|^{-\frac{2}{N-6}}$ and $|i\,\mathrm{e}^{i(\zeta-\theta)} - \mathrm{e}^{-i\theta}| = |\mathrm{e}^{i\zeta} + i| \lesssim |\zeta| \le |t|^{-\frac{3}{N-6}}$; hence

$$|\Re(i\,\mathrm{e}^{i(\zeta-\theta)}) - 1| \lesssim |t|^{-\frac{2}{N-6}}.$$

Since $\left|\int W_\mu W_\lambda^{\frac{N+2}{N-2}}\,\mathrm{d}x\right| \lesssim \lambda^{\frac{N-2}{2}} \lesssim |t|^{-\frac{N-2}{N-6}}$, we obtain

$$\left|\Re(i\,\mathrm{e}^{i(\zeta-\theta)})\int W_\mu W_\lambda^{\frac{N+2}{N-2}}\,\mathrm{d}x - \int W_\mu W_\lambda^{\frac{N+2}{N-2}}\,\mathrm{d}x\right| \lesssim |t|^{-\frac{N}{N-6}}. \tag{3-32}$$

The proof of (3-26) yields

$$\left|\int W_\lambda^{\frac{N+2}{N-2}}\,\mathrm{d}x - \int W_\mu W_\lambda^{\frac{N+2}{N-2}}\,\mathrm{d}x\right| \lesssim \lambda^{\frac{N}{2}} + |\mu-1|\lambda^{\frac{N-2}{2}} \lesssim |t|^{-\frac{N}{N-6}}. \tag{3-33}$$

From (2-8) we get

$$\int W_\lambda^{\frac{N+2}{N-2}}\,\mathrm{d}x = \frac{2\kappa^{\frac{N-4}{2}}\|W\|_{L^2}^2}{N-6}\lambda^{\frac{N-2}{2}};$$

hence (3-30) follows from (3-31)–(3-33).

From (3-27)–(3-30) and the triangle inequality we obtain

$$\left|B_4 - \frac{2\kappa^{\frac{N-4}{2}}\|W\|_{L^2}^2}{N-6}\lambda(t)^{\frac{N-2}{2}}\right| \lesssim |t|^{-\frac{N}{N-6}} + \|g\|_{\mathcal{E}}^2,$$

in particular

$$|B_4| \lesssim |t|^{-\frac{N-2}{N-6}} + \|g\|_{\mathcal{E}}^2 \lesssim |t|^{-\frac{N-2}{N-6}}.$$

□

Remark 3.3. A computation similar to the proof of (3-14) shows that $|K| \lesssim \|g\|_{\mathcal{E}}^2 \leq |t|^{-\frac{N-1}{N-6}}$, so we obtain the following simple consequence of Lemma 3.1:

$$|\zeta'(t)| + \left|\frac{\mu'(t)}{\mu(t)}\right| + |\theta'(t)| + \left|\frac{\lambda'(t)}{\lambda(t)}\right| \lesssim |t|^{-1} \tag{3-34}$$

(for the last term, this bound is sharp).

Control of the stable and unstable component. An important step is to control the stable and unstable components $a_1^{\pm}(t) = \langle \alpha^{\pm}_{\zeta(t),\mu(t)}, g(t)\rangle$ and $a_2^{\pm}(t) = \langle \alpha^{\pm}_{\theta(t),\lambda(t)}, g(t)\rangle$. Recall that $\nu > 0$ is the positive eigenvalue of the linearized flow; see (2-12).

Lemma 3.4. *Under assumptions of Lemma 3.1, for $t \in [T, T_1]$ we have*

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}a_1^+(t) - \frac{\nu}{\mu(t)^2}a_1^+(t)\right| \leq \frac{c}{\mu(t)^2}|t|^{-\frac{N}{2(N-6)}}, \tag{3-35}$$

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}a_1^-(t) + \frac{\nu}{\mu(t)^2}a_1^-(t)\right| \leq \frac{c}{\mu(t)^2}|t|^{-\frac{N}{2(N-6)}},$$

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}a_2^+(t) - \frac{\nu}{\lambda(t)^2}a_2^+(t)\right| \leq \frac{c}{\lambda(t)^2}|t|^{-\frac{N}{2(N-6)}}, \tag{3-36}$$

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}a_2^-(t) + \frac{\nu}{\lambda(t)^2}a_2^-(t)\right| \leq \frac{c}{\lambda(t)^2}|t|^{-\frac{N}{2(N-6)}}, \tag{3-37}$$

with $c \to 0$ as $|T_0| \to +\infty$.

Proof. We will give a proof of (3-35) and (3-36), the other two inequalities being analogous.

Applying the chain rule to the formula $a_1^+(t) = \langle \alpha^+_{\zeta(t),\mu(t)}, g(t)\rangle$ and using the definition of $\alpha^+_{\zeta,\mu}$ we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}a_1^+ = -\frac{\mu'}{\mu}\Big\langle \frac{\mathrm{e}^{i\zeta}}{\mu^2}\big(\Lambda_{-1}\mathcal{Y}^{(2)}_\mu + i\Lambda_{-1}\mathcal{Y}^{(1)}_\mu\big), g\Big\rangle + \zeta'\Big\langle \frac{\mathrm{e}^{i\zeta}}{\mu^2}\big(i\mathcal{Y}^{(2)}_\mu - \mathcal{Y}^{(1)}_\mu\big), g\Big\rangle + \langle \alpha^+_{\zeta,\mu}, \partial_t g\rangle.$$

Thanks to (3-34) and (3-9), the size of the first two terms is $\lesssim |t|^{-1}|t|^{-\frac{N-1}{2(N-6)}} = |t|^{-\frac{3N-13}{2(N-6)}} \ll |t|^{-\frac{N}{2(N-6)}}$. We are left with the third term, and we expand $\partial_t g$ according to (3-3).

Let us consider, one by one, the contributions of the four terms in the second line of (3-3):

(1) The term $\langle \alpha^+_{\zeta,\mu}, -\zeta' i\mathrm{e}^{i\zeta}W_\mu\rangle$ is equal to 0 thanks to (2-14).

(2) The term $\big\langle \alpha^+_{\zeta,\mu}, \frac{\mu'}{\mu}\mathrm{e}^{i\zeta}\Lambda W_\mu\big\rangle$ is equal to 0 thanks to (2-15).

(3) Consider the term $\langle \alpha^+_{\zeta,\mu}, -\theta' i\mathrm{e}^{i\theta}W_\lambda\rangle$. We have $\|\alpha^+_{\zeta,\mu}\|_{\dot H^1} \lesssim 1$; hence

$$|\langle \alpha^+_{\zeta,\mu}, -\theta' i\mathrm{e}^{i\theta}W_\lambda\rangle| \lesssim |\theta'|\,\|\alpha^+_{\zeta,\mu}\|_{\dot H^1}\|W_\lambda\|_{\dot H^{-1}} \lesssim |\theta'|\lambda^2,$$

and (3-34) yields $|\theta'|\lambda^2 \lesssim |t|^{-1}|t|^{-\frac{4}{N-6}} = |t|^{-\frac{N-2}{N-6}} \ll |t|^{-\frac{N}{2(N-6)}}$.

(4) The term $\big\langle \alpha^+_{\zeta,\mu}, \frac{\lambda'}{\lambda}\mathrm{e}^{i\theta}\Lambda W_\lambda\big\rangle$ is treated as the previous one, using $\big|\frac{\lambda'}{\lambda}\big| \lesssim |t|^{-1}$ instead of $|\theta'| \lesssim |t|^{-1}$.

Let us finally consider the contribution of the first line of (3-3). We have

$$\begin{aligned} i\Delta g + i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\big)& \\ = Z_{\zeta,\mu}g + i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) &- f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\zeta}W_\mu)g\big). \end{aligned}$$

From (2-17) we obtain $\langle \alpha^+_{\zeta,\mu}, Z_{\zeta,\mu}g\rangle = \frac{\nu}{\mu^2}a_1^+$; hence we need to show that

$$\big|\big\langle \alpha^+_{\zeta,\mu}, i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\zeta}W_\mu)g\big)\big\rangle\big| \ll |t|^{-\frac{N}{2(N-6)}}.$$

The proof of (3-17) yields the bound $|t|^{-\frac{N-2}{N-6}} \ll |t|^{-\frac{N}{2(N-6)}}$.

We turn to the proof of (3-36). Applying the chain rule to the formula $a_2^+(t) = \langle \alpha^+_{\zeta(t),\mu(t)}, g(t)\rangle$ and using the definition of $\alpha^+_{\theta,\lambda}$ we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}a_2^+ = -\frac{\lambda'}{\lambda}\Big\langle \frac{\mathrm{e}^{i\theta}}{\lambda^2}\big(\Lambda_{-1}\mathcal{Y}^{(2)}_\lambda + i\Lambda_{-1}\mathcal{Y}^{(1)}_\lambda\big), g\Big\rangle + \theta'\Big\langle \frac{\mathrm{e}^{i\theta}}{\lambda^2}\big(i\mathcal{Y}^{(2)}_\lambda - \mathcal{Y}^{(1)}_\lambda\big), g\Big\rangle + \langle \alpha^+_{\theta,\lambda}, \partial_t g\rangle.$$

The first two terms are treated as in the case of a_1^+. In the third term, we expand $\partial_t g$ using (3-3). Let us consider, one by one, the contributions of the four terms in the second line of (3-3):

(1) In order to bound the term $\langle \alpha^+_{\theta,\lambda}, -\zeta' i\mathrm{e}^{i\zeta}W_\mu\rangle$, notice that

$$\|\alpha^+_{\theta,\lambda}\|_{L^1} \lesssim \int_{\mathbb{R}^N}\frac{1}{\lambda^2}\big(|\mathcal{Y}^{(1)}_\lambda| + |\mathcal{Y}^{(2)}_\lambda|\big)\,\mathrm{d}x \lesssim \lambda^{\frac{N-2}{2}} \lesssim |t|^{-\frac{N-2}{N-6}} \ll |t|^{-\frac{N}{2(N-6)}}.$$

This is sufficient since $\|-\zeta' i\mathrm{e}^{i\zeta}W_\mu\|_{L^\infty} \lesssim 1$.

(2) The term $\big\langle \alpha^+_{\theta,\lambda}, \frac{\mu'}{\mu}\mathrm{e}^{i\zeta}\Lambda W_\mu\big\rangle$ is analogous.

(3) The term $\langle\alpha^+_{\theta,\lambda}, -\theta' i \mathrm{e}^{i\theta}W_\lambda\rangle$ is equal to 0 thanks to (2-14).

(4) The term $\big\langle\alpha^+_{\theta,\lambda}, \frac{\lambda'}{\lambda}\mathrm{e}^{i\theta}\Lambda W_\lambda\big\rangle$ is equal to 0 thanks to (2-15).

Let us finally consider the contribution of the first line of (3-3). We have

$$\begin{aligned} &i\Delta g + i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\big) \\ &\qquad = Z_{\theta,\lambda}g + i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\theta}W_\lambda)g\big). \end{aligned}$$

From (2-17) we obtain $\langle\alpha^+_{\theta,\lambda}, Z_{\theta,\lambda}g\rangle = \frac{\nu}{\lambda^2}a_2^+$; hence we need to show that

$$\lambda^2\big|\big\langle\alpha^+_{\theta,\lambda}, i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\theta}W_\lambda)g\big)\big\rangle\big| \ll |t|^{-\frac{N}{2(N-6)}}. \tag{3-38}$$

The proof of (3-19) yields

$$\lambda^2\big|\big\langle\alpha^+_{\theta,\lambda}, i\big(f'(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\theta}W_\lambda)\big)g\big\rangle\big| \lesssim \lambda^{\frac{N}{4}}\|g\|_{\mathcal{E}} \lesssim |t|^{-\frac{N}{2(N-6)}-\frac{N-1}{2(N-6)}} \ll |t|^{-\frac{N}{2(N-6)}}. \tag{3-39}$$

The proof of (3-27) yields

$$\begin{aligned} &\lambda^2\big|\big\langle\alpha^+_{\theta,\lambda}, i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda)g\big)\big\rangle\big| \\ &\qquad \lesssim \|g\|_{\mathcal{E}}^2 \ll |t|^{-\frac{N}{2(N-6)}}. \end{aligned} \tag{3-40}$$

Using (2-3) we get

$$\big\|f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\big\|_{L^\infty} \lesssim \|W_\lambda^{\frac{4}{N-2}}W_\mu\|_{L^\infty} \lesssim \frac{1}{\lambda^2}.$$

By a change of variable, $\|\lambda^2\alpha^+_{\theta,\lambda}\|_{L^1} \lesssim \lambda^{\frac{N+2}{2}}$; hence

$$\lambda^2\big|\big\langle\alpha^+_{\theta,\lambda}, i\big(f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)\big)\big\rangle\big| \lesssim \lambda^{\frac{N-2}{2}} \lesssim |t|^{-\frac{N-2}{N-6}} \ll |t|^{-\frac{N}{2(N-6)}}. \tag{3-41}$$

Taking the sum of (3-39)–(3-41) and using the triangle inequality, we obtain (3-38). □

4. Bootstrap

We turn to the heart of the proof, which consists in establishing bootstrap estimates. We consider a solution $u(t)$, decomposed according to (3-1), (3-2) and (2-18). The initial data at time $T \le T_0$ is chosen as follows.

Lemma 4.1. *There exists $T_0 < 0$ such that for all $T \le T_0$ and for all λ^0, a_1^0, a_2^0 satisfying*

$$\Big|\lambda^0 - \tfrac{1}{\kappa}(\kappa|T|)^{-\frac{2}{N-6}}\Big| \le \tfrac12|T|^{-\frac{5}{2(N-6)}}, \quad |a_1^0| \le \tfrac12|T|^{-\frac{N}{2(N-6)}}, \quad |a_2^0| \le \tfrac12|T|^{-\frac{N}{2(N-6)}}, \tag{4-1}$$

there exists $g^0 \in X^1$ satisfying

$$\langle\Lambda W, g^0\rangle = \langle iW, g^0\rangle = \langle i\Lambda W_{\lambda^0}, g^0\rangle = \langle -W_{\lambda^0}, g^0\rangle = 0, \tag{4-2}$$

$$\langle\alpha^-_{-\frac{\pi}{2},1}, g^0\rangle = 0, \quad \langle\alpha^+_{-\frac{\pi}{2},1}, g^0\rangle = a_1^0, \quad \langle\alpha^-_{0,\lambda^0}, g^0\rangle = 0, \quad \langle\alpha^+_{0,\lambda^0}, g^0\rangle = a_2^0, \tag{4-3}$$

$$\|g^0\|_{\mathcal{E}} \lesssim |T|^{-\frac{N}{2(N-6)}}. \tag{4-4}$$

This g^0 is continuous for the X^1 topology with respect to λ^0, a_1^0 and a_2^0.

Remark 4.2. For the continuity, we just claim that the function g^0 constructed in the proof is continuous with respect to λ^0, a_1^0 and a_2^0. Clearly, g^0 is not uniquely determined by (4-2)–(4-4).

Remark 4.3. Condition (4-2) is exactly (2-18) with $\big(\zeta,\mu,\theta,\lambda\big)=\big(-\frac{\pi}{2},1,0,\lambda^0\big)$. Hence, if we consider the solution $u(t)$ of (1-1) with initial data $u(T)=-iW+W_{\lambda^0}+g^0$ and decompose it according to (3-1), then $g(T)=g^0$ and the initial values of the modulation parameters are $\big(\zeta(T),\mu(T),\theta(T),\lambda(T)\big)=\big(-\frac{\pi}{2},1,0,\lambda^0\big)$.

Proof. We consider functions of the form

$$g^0=a_1^+i\alpha^-_{-\frac{\pi}{2},1}-a_1^-i\alpha^+_{-\frac{\pi}{2},1}+b_1W+c_1(-i\Lambda W)+a_2^+(\lambda^0)^2i\alpha^-_{0,\lambda^0}-a_2^-(\lambda^0)^2i\alpha^+_{0,\lambda^0}+b_2iW_{\lambda^0}+c_2\Lambda W_{\lambda^0},$$

with a_1^+, a_1^-, b_1, c_1, a_2^+, a_2^-, b_2, c_2 being real numbers. Let $\Phi:\mathbb{R}^8\to\mathbb{R}^8$ be the linear map defined as

$$\begin{aligned}\Phi(a_1^+,a_1^-,b_1,c_1,a_2^+,a_2^-,b_2,c_2):=\big(&\langle\alpha^+_{-\frac{\pi}{2},1},g^0\rangle,\langle\alpha^-_{-\frac{\pi}{2},1},g^0\rangle,\langle\Lambda W,g^0\rangle,\langle iW,g^0\rangle,\\ &\langle\alpha^+_{0,\lambda^0},g^0\rangle,\langle\alpha^-_{0,\lambda^0},g^0\rangle,\big\langle(\lambda^0)^{-2}i\Lambda W_{\lambda^0},g^0\big\rangle,\big\langle-(\lambda^0)^{-2}W_{\lambda^0},g^0\big\rangle\big).\end{aligned}$$

Using (2-14)–(2-16) and the fact that λ^0 is small we obtain that the matrix of Φ is strictly diagonally dominant, which implies the result. □

In the remaining part of this section, we will analyze solutions $u(t)$ of (1-1) with the initial data $u(T)=-iW+W_{\lambda^0}+g^0$, where g^0 is given by the previous lemma.

Proposition 4.4. *There exists $T_0<0$ with the following property. Let $T<T_1<T_0$ and let λ^0,a_1^0,a_2^0 satisfy (4-1). Let $g^0\in X^1$ be given by Lemma 4.1 and consider the solution $u(t)$ of (1-1) with the initial data $u(T)=-iW+W_{\lambda^0}+g^0$. Suppose that $u(t)$ exists on the time interval $[T,T_1]$, that for $t\in[T,T_1]$ conditions (3-5)–(3-9) hold, and moreover that*

$$|a_1^+(t)|\le|t|^{-\frac{N}{2(N-6)}},\quad |a_2^+(t)|\le|t|^{-\frac{N}{2(N-6)}}.\tag{4-5}$$

Then for $t\in[T,T_1]$,

$$\big|\zeta(t)+\tfrac{\pi}{2}\big|\le\tfrac12|t|^{-\frac{3}{N-6}},\tag{4-6}$$

$$|\mu(t)-1|\le\tfrac12|t|^{-\frac{3}{N-6}},\tag{4-7}$$

$$|\theta(t)|\le\tfrac12|t|^{-\frac{1}{N-6}},\tag{4-8}$$

$$\|g(t)\|_{\mathcal{E}}\le\tfrac12|t|^{-\frac{N-1}{2(N-6)}}.\tag{4-9}$$

Before we give a proof, we need a little preparation.

A virial-type correction. The delicate part of the proof of Proposition 4.4 will be to control $\theta(t)$. For this, we will need to use a virial functional, which we now define.

Lemma 4.5. *For any $c>0$ and $R>0$ there exists a radial function $q(x)=q_{c,R}(x)\in C^{3,1}(\mathbb{R}^N)$ with the following properties*:

(P1) $q(x)=\frac12|x|^2$ *for* $|x|\le R$.

(P2) *There exists* $\widetilde{R} > 0$ (*depending on* c *and* R) *such that* $q(x) \equiv \text{const}$ *for* $|x| \geq \widetilde{R}$.

(P3) $|\nabla q(x)| \lesssim |x|$ *and* $|\Delta q(x)| \lesssim 1$ *for all* $x \in \mathbb{R}^N$, *with constants independent of* c *and* R.

(P4) $\sum_{1\leq j,k\leq N}(\partial_{x_j x_k} q(x))\bar{v}_j v_k \geq -c\sum_{j=1}^N |v_j|^2$ *for all* $x \in \mathbb{R}^N$, $v_j \in \mathbb{C}$.

(P5) $\Delta^2 q(x) \leq c\cdot|x|^{-2}$, *for all* $x \in \mathbb{R}^N$.

Remark 4.6. We require $C^{3,1}$ regularity in order not to worry about boundary terms in Pohozaev identities; see the proof of (4-13).

Proof. It suffices to prove the result for $R = 1$ since the function $q_R(x) := R^2 q\big(\frac{x}{R}\big)$ satisfies the listed properties if and only if $q(x)$ does.

Let r denote the radial coordinate. Define $q_0(x)$ by the formula

$$q_0(r) := \begin{cases} \frac{1}{2}r^2, & r \leq 1,\\ \dfrac{N(N-2)r}{(N-1)(N-3)} - \dfrac{N}{2(N-4)} + \dfrac{N}{2(N-3)(N-4)r^{N-4}} - \dfrac{1}{2(N-1)r^{N-2}}, & r \geq 1.\end{cases}$$

A direct computation shows that for $r > 1$ we have

$$\begin{aligned} q_0'(r) &= \frac{N(N-2)}{(N-1)(N-3)} - \frac{N}{2(N-3)r^{N-3}} + \frac{N-2}{2(N-1)r^{N-1}},\\ q_0''(r) &= \frac{N}{2r^{N-2}} - \frac{N-2}{2r^N} > 0 \quad (\text{so } q_0(x) \text{ is convex}),\\ q_0'''(r) &= \frac{N(N-2)}{2}\Big(-\frac{1}{r^{N-1}} + \frac{1}{r^{N+1}}\Big),\\ \Delta^2 q_0(r) &= -N(N-2)r^{-3} < 0.\end{aligned}$$

In particular,

$$\lim_{r\to 1^+}\big(q_0(r), q_0'(r), q_0''(r), q_0'''(r)\big) = \big(\tfrac{1}{2}, 1, 1, 0\big).$$

Hence $q_0 \in C^{3,1}$ and it satisfies all the listed properties except for (P2). We correct it as follows.

Let $e_j(r) := (1/j!)r^j\cdot\chi(r)$ for $j \in \{1,2,3\}$, where $\chi(r)$ is the standard cut-off function:

$$\chi \in C^\infty((0,+\infty),\mathbb{R}), \qquad \chi(r) = 1 \quad \text{for } r \leq 1, \qquad \chi(r) = 0 \quad \text{for } r \geq 2.$$

Let $R_0 \gg 1$. We define

$$q(r) := \begin{cases} q_0(r), & r \leq R_0,\\ q_0(R_0) + \sum_{j=1}^3 q_0^{(j)}(R_0)\cdot R_0^j\cdot e_j(-1+R_0^{-1}r), & r \geq R_0.\end{cases}$$

Note that $q_0'(R_0) \sim 1$, $q_0''(R_0) \sim R_0^{-N+2}$ and $q_0'''(R_0) \sim R_0^{-N+1}$. It is clear that $q(x) \in C^{3,1}(\mathbb{R}^N)$. Property (P1) holds since $R_0 > 1$. By the definition of the functions e_j we have $q(r) = q_0(R_0) = \text{const}$ for $r \geq 3R_0$; hence (P2) holds with $\widetilde{R} = 3R_0$. From the definition of $q(r)$ we get $|q'(r)| \lesssim |q_0'(R_0)| \lesssim r$ and $|q''(r)| \lesssim |q_0''(R_0)| \lesssim R_0^{-N+2} \lesssim 1$ for $r \geq R_0$, with a constant independent of R_0, which implies (P3). Similarly, $|\partial_{x_i x_j} q(x)| \lesssim R_0^{-1}$ for $|x| \geq R_0$, which implies (P4) if R_0 is large enough. Finally $|\Delta^2 q(x)| \lesssim R_0^{-3}$ for $|x| \geq R_0$ and $\Delta^2 q(x) = 0$ for $|x| \geq 3R_0$. This proves (P5) if R_0 is large enough. ☐

In the sequel $q(x)$ always denotes a function of class $C^{3,1}(\mathbb{R}^N)$ verifying (P1)–(P5) with sufficiently small c and sufficiently large R.

For $\lambda > 0$ we define the operators $A(\lambda)$ and $A_0(\lambda)$ as

$$[A(\lambda)h](x) := \frac{N-2}{2N\lambda^2}\Delta q\Big(\frac{x}{\lambda}\Big)h(x) + \frac{1}{\lambda}\nabla q\Big(\frac{x}{\lambda}\Big)\cdot\nabla h(x),$$
$$[A_0(\lambda)h](x) := \frac{1}{2\lambda^2}\Delta q\Big(\frac{x}{\lambda}\Big)h(x) + \frac{1}{\lambda}\nabla q\Big(\frac{x}{\lambda}\Big)\cdot\nabla h(x).$$

Combining these definitions with the fact that $q(x)$ is an approximation of $\frac{1}{2}|x|^2$ we see that $A(\lambda)$ and $A_0(\lambda)$ are approximations (in a sense not yet made precise) of $\lambda^{-2}\Lambda$ and $\lambda^{-2}\Lambda_0$ respectively. We will write A and A_0 instead of $A(1)$ and $A_0(1)$ respectively. Note the following scale-change formulas, which follow directly from the definitions:

$$\text{for all } h\in\mathcal{E},\quad A(\lambda)(h_\lambda) = \lambda^{-2}(Ah)_\lambda,\quad A_0(\lambda)(h_\lambda) = \lambda^{-2}(A_0h)_\lambda. \tag{4-10}$$

Lemma 4.7. *The operators $A(\lambda)$ and $A_0(\lambda)$ have the following properties*:

- *For $\lambda > 0$, the families $\{A(\lambda)\}$, $\{A_0(\lambda)\}$, $\{\lambda\partial_\lambda A(\lambda)\}$, $\{\lambda\partial_\lambda A_0(\lambda)\}$ are bounded in $\mathscr{L}(\mathcal{E};\dot H^{-1})$ and the families $\{\lambda A(\lambda)\}$, $\{\lambda A_0(\lambda)\}$ are bounded in $\mathscr{L}(\mathcal{E};L^2)$, with the bound depending on the choice of the function $q(x)$,*
- *For all complex-valued $h_1,h_2\in X^1(\mathbb{R}^N)$ and $\lambda>0$,*

$$\langle A(\lambda)h_1, f(h_1+h_2)-f(h_1)-f'(h_1)h_2\rangle = -\langle A(\lambda)h_2, f(h_1+h_2)-f(h_1)\rangle, \tag{4-11}$$
$$\langle h_1, A_0(\lambda)h_2\rangle = -\langle A_0(\lambda)h_1, h_2\rangle,\quad \textit{and hence } iA_0(\lambda) \textit{ is a symmetric operator.} \tag{4-12}$$

- *For any $c_0>0$, if we choose c in Lemma 4.5 small enough, then for all $h\in X^1$,*

$$\langle A_0(\lambda)h, \Delta h\rangle \le \frac{c_0}{\lambda^2}\|h\|_{\mathcal{E}}^2 - \frac{1}{\lambda^2}\int_{|x|\le R\lambda}|\nabla h(x)|^2\,\mathrm{d}x. \tag{4-13}$$

In dimension $N=6$ and for real-valued functions, this was proved in [Jendrej 2016, Lemma 3.12]. Most arguments apply without change, but we provide here a full computation for the reader's convenience.

Proof. Since $\nabla q(x)$ and $\nabla^2 q(x)$ are continuous and of compact support, it is clear that A and A_0 are bounded operators $\mathcal{E}\to\dot H^{-1}$. From the invariance (4-10) we see that $A(\lambda)$ and $A_0(\lambda)$ have the same norms as A and A_0 respectively. For $\lambda A(\lambda)$, $\lambda A_0(\lambda)$, $\lambda\partial_\lambda A(\lambda)$ and $\lambda\partial_\lambda A_0(\lambda)$ the proof is similar. We compute

$$\partial_\lambda A(\lambda) = -\frac{N-2}{N\lambda^3}\Delta q\Big(\frac{x}{\lambda}\Big) - \frac{N-2}{2N\lambda^4}x\cdot\nabla\Delta q\Big(\frac{x}{\lambda}\Big) - \frac{1}{\lambda^3}x\cdot\nabla^2 q\Big(\frac{x}{\lambda}\Big)\cdot\nabla.$$

Since $\nabla q(x)$, $\nabla^2 q(x)$ and $\nabla^3 q(x)$ are continuous and of compact support, we get boundedness of $\partial_\lambda A(1)$, and boundedness $\{\lambda\partial_\lambda A(\lambda)\}$ follows by the scaling invariance.

In (4-11), we may assume without loss of generality that $\lambda = 1$. Notice that both sides are continuous with respect to the topology $\|h_1\|_{X^1}+\|h_2\|_{X^1}$. Indeed, A is continuous from X^1 to $\mathcal{E}$ and $(h_1,h_2)\mapsto\big(f(h_1+h_2)-f(h_1)-f'(h_1)h_2, f(h_1+h_2)-f(h_1)\big)$ is continuous from $\mathcal{E}$ to $\dot H^{-1}$ by

Sobolev and dual Sobolev. We may therefore assume that $h_1, h_2 \in C_0^\infty$. Observe that for any $h \in C_0^\infty$ we have $f(h)\bar h = \frac{2N}{N-2}F(h)$ and $\Re(f(h)\nabla \bar h) = \nabla(F(h))$; hence

$$\langle Ah, f(h)\rangle = \Re\int_{\mathbb{R}^N}\Big(\frac{N-2}{2N}\Delta q\bar h + \nabla q\cdot\nabla\bar h\Big)f(h)\,\mathrm{d}x = \int_{\mathbb{R}^N}\Delta q\cdot F(h) + \nabla q\cdot\nabla\big(F(h)\big)\,\mathrm{d}x = 0.$$

Using this for $h = h_1 + h_2$ and for $h = h_1$, (4-11) is seen to be equivalent to

$$\langle Ah_2, f(h_1)\rangle + \langle Ah_1, f'(h_1)h_2\rangle = 0. \tag{4-14}$$

Expanding the left side using the definition of A we obtain

$$\begin{aligned}\langle Ah_2, f(h_1)\rangle+\langle Ah_1, f'(h_1)h_2\rangle &= \Re\int_{\mathbb{R}^N}\frac{N-2}{2N}\Delta q\cdot\bar h_2\cdot f(h_1)+\nabla q\cdot\nabla\bar h_2\cdot f(h_1)\,\mathrm{d}x\\ &\quad+\Re\int_{\mathbb{R}^N}\frac{N-2}{2N}\Delta q\cdot\bar h_1\cdot f'(h_1)h_2+\nabla q\cdot\nabla\bar h_1\cdot f'(h_1)h_2\,\mathrm{d}x.\end{aligned}\tag{4-15}$$

We have

$$\Re\int_{\mathbb{R}^N}\nabla q\cdot\nabla\bar h_2\cdot f(h_1)\,\mathrm{d}x = -\Re\int_{\mathbb{R}^N}\bar h_2\cdot\Delta q\cdot f(h_1)\,\mathrm{d}x - \Re\int_{\mathbb{R}^N}\bar h_2\cdot\nabla q\cdot\nabla f(h_1)\,\mathrm{d}x.$$

Using (2-1) and the fact that $f'(h_1)h_1 = \frac{N+2}{N-2}f(h_1)$ we get

$$\Re\int_{\mathbb{R}^N}\frac{N-2}{2N}\Delta q\cdot\bar h_1\cdot f'(h_1)h_2\,\mathrm{d}x = \Re\int_{\mathbb{R}^N}\frac{N-2}{2N}\bar h_2\cdot\Delta q\cdot f'(h_1)h_1\,\mathrm{d}x = \Re\int_{\mathbb{R}^N}\bar h_2\cdot\frac{N+2}{2N}\Delta q\cdot f(h_1)\,\mathrm{d}x.$$

Using (2-1) and the fact that $f'(h_1)\nabla h_1 = \nabla(f(h_1))$ we get

$$\Re\int_{\mathbb{R}^N}\nabla q\cdot\nabla\bar h_1\cdot f'(h_1)h_2\,\mathrm{d}x = \Re\int_{\mathbb{R}^N}\bar h_2\cdot\nabla q\cdot f'(h_1)\nabla h_1\,\mathrm{d}x = \Re\int_{\mathbb{R}^N}\bar h_2\cdot\nabla q\cdot\nabla(f(h_1))\,\mathrm{d}x.$$

Plugging the last three formulas into (4-15) we obtain

$$\begin{aligned}&\langle Ah_2, f(h_1)\rangle+\langle Ah_1, f'(h_1)h_2\rangle\\ &= \Big\langle h_2, \frac{N-2}{2N}\Delta q\cdot f(h_1) - \Delta q\cdot f(h_1) - \nabla q\cdot\nabla(f(h_1)) + \frac{N+2}{2N}\Delta q\cdot f(h_1) + \nabla q\cdot\nabla(f(h_1))\Big\rangle = \langle h_2, 0\rangle = 0,\end{aligned}$$

which proves (4-14).

Identity (4-12) follows by an integration by parts.

In (4-13), we can again assume that $\lambda = 1$ and $h \in C_0^\infty$ (we use the fact that $q \in C^{3,1}$, and hence $\Delta^2 q$ is bounded and of compact support). Inequality (4-13) follows easily from (P1), (P4) and (P5), once we check the following identity:

$$\Re\int_{\mathbb{R}^N}\Delta h\cdot\Big(\frac12\Delta q\cdot\bar h + \nabla q\cdot\nabla\bar h\Big)\,\mathrm{d}x = \frac14\int_{\mathbb{R}^N}(\Delta^2 q)|h|^2\,\mathrm{d}x - \int_{\mathbb{R}^N}\sum_{i,j=1}^N \partial_{ij}q\,\partial_i\bar h\,\partial_j h\,\mathrm{d}x. \tag{4-16}$$

We can assume that $q \in C_0^\infty$, and (4-16) follows from integration by parts:

$$\Re\int_{\mathbb{R}^N} \tfrac{1}{2}\Delta h\cdot\Delta q\cdot\bar{h}+\Delta h\cdot\nabla q\cdot\nabla\bar{h}\,\mathrm{d}x$$

$$=\Re\int_{\mathbb{R}^N}\sum_{j,k=1}^{N}\big(\tfrac{1}{2}\partial_{jj}h\cdot\partial_{kk}q\cdot\bar{h}+\partial_{jj}h\cdot\partial_k q\cdot\partial_k\bar{h}\big)\,\mathrm{d}x$$

$$=\Re\int_{\mathbb{R}^N}-\tfrac{1}{2}\sum_{j,k}\partial_j h(\partial_{kk}q\,\partial_j\bar{h}+\partial_{jkk}q\cdot\bar{h})+\sum_j\tfrac{1}{2}\partial_j(|\partial_j h|^2)\partial_j q+\sum_{j\neq k}\big(-\tfrac{1}{2}\partial_k|\partial_j h|^2\partial_k q-\partial_{jk}q\,\partial_j\bar{h}\,\partial_k h\big)\,\mathrm{d}x$$

$$=\Re\int_{\mathbb{R}^N}-\tfrac{1}{2}\sum_{j,k}\big(\partial_{kk}q|\partial_j h|^2-\tfrac{1}{2}\partial_{jjkk}q\cdot|h|^2\big)-\tfrac{1}{2}\sum_j\partial_{jj}q|\partial_j h|^2+\tfrac{1}{2}\sum_{j\neq k}\partial_{kk}q|\partial_j h|^2-\sum_{j\neq k}\partial_{jk}q\,\partial_j\bar{h}\,\partial_k h\,\mathrm{d}x$$

$$=\int_{\mathbb{R}^N}\tfrac{1}{4}\sum_{j,k}\partial_{jjkk}q\cdot|h|^2-\sum_{j,k}\partial_{jk}q\,\partial_j\bar{h}\,\partial_k h\,\mathrm{d}x. \qquad\square$$

Closing the bootstrap.

Proof of Proposition 4.4. We split the proof into three steps. First we prove (4-6) and (4-7). Then we use the virial functional and variational estimates to prove (4-8), with $\frac{1}{2}$ replaced by any strictly positive constant. To do this, we have to deal somehow with the term $\|W\|_{L^2}^{-2}K$ in the modulation equation (3-12). It involves terms quadratic in g, which is the critical size. However, it turns out that we can use a virial functional to absorb the essential part of K. Proving (4-8) is the most difficult step. Finally, (4-9) will follow from variational estimates.

Step 1. Integrating (3-10) on $[T,t]$ and using the fact that $\zeta(T)=-\frac{\pi}{2}$ we get

$$\big|\zeta(t)+\tfrac{\pi}{2}\big|=\big|\zeta(t)-\zeta(T)\big|=\left|\int_T^t\zeta'(\tau)\,\mathrm{d}\tau\right|\le c\int_T^t|\tau|^{-\frac{N-3}{N-6}}\,\mathrm{d}\tau\le c\cdot\frac{N-6}{3}|t|^{-\frac{3}{N-6}}\le\tfrac{1}{2}|t|^{-\frac{3}{N-6}},$$

provided that $c\le\frac{3}{2(N-6)}$. Recall that $c>0$ can be made arbitrarily small by choosing $|T_0|$ large enough, in particular smaller than $\frac{3}{2(N-6)}$. The proof of (4-7) is similar.

Step 2. First, let us show that for $t\in[T,T_1]$ we have

$$|a_1^-(t)|<|t|^{-\frac{N}{2(N-6)}},\qquad |a_2^-(t)|<|t|^{-\frac{N}{2(N-6)}}. \tag{4-17}$$

This is verified initially; see (4-3). Suppose that $T_2\in(T,T_1)$ is the last time for which (4-17) holds for $t\in[T,T_2)$. Let for example $a_1^-(T_2)=|T_2|^{-\frac{N}{2(N-6)}}$. But since $\|g(T_2)\|_{\mathcal{E}}^2\lesssim|T_2|^{-\frac{N-1}{N-6}}\ll|T_2|^{-\frac{N}{2(N-6)}}$, (3-37) implies $\frac{\mathrm{d}}{\mathrm{d}t}a_1^-(T_2)<0$, which contradicts the assumption that $a_1^-(t)<|T_2|^{-\frac{N}{2(N-6)}}$ for $t<T_2$. The proof of the other inequality is similar.

Let $c_0>0$. We will prove that if T_0 is chosen large enough (depending on c_0), then

$$|\theta(t)|\le c_0|t|^{-\frac{1}{N-6}},\quad \text{for } t\in[T,T_1]. \tag{4-18}$$

By the conservation of energy, (2-19) and (4-4) we have

$$|E(u)-2E(W)|=|E(u(T))-2E(W)|\lesssim|T|^{-\frac{N}{N-6}}\le|t|^{-\frac{N}{N-6}};$$

hence (2-20) yields

$$\theta\lambda^{\frac{N-2}{2}} \lesssim |t|^{-\frac{N}{N-6}} \quad \Longrightarrow \quad \theta \lesssim |t|^{-\frac{N}{N-6}+\frac{N-2}{N-6}} = |t|^{-\frac{2}{N-6}} \ll |t|^{-\frac{1}{N-6}}. \tag{4-19}$$

It remains to prove that

$$\theta \geq -c_0|t|^{-\frac{1}{N-6}}. \tag{4-20}$$

To this end, we consider the real scalar function

$$\psi(t) := \theta(t) - \frac{1}{2\|W\|_{L^2}^2}\langle g(t), iA_0(\lambda(t))g(t)\rangle.$$

We will show that for $t \in [T, T_1]$ we have

$$\psi'(t) \geq -c_1|t|^{-\frac{N-5}{N-6}}, \tag{4-21}$$

with $c_1 > 0$ as small as we like, by eventually enlarging $|T_0|$.

From (4-19) we get $\theta\lambda^{-\frac{N-6}{2}} \ll |t|^{-\frac{N-5}{N-6}}$; hence, taking in Lemma 3.1, say, $c = \frac{1}{4}c_1$ and choosing $|T_0|$ large enough, (3-12) yields

$$\begin{aligned} \psi' &\geq -\frac{(N-2)\kappa^{\frac{N-4}{2}}}{N-6}\theta\lambda^{\frac{N-6}{2}} + \frac{K}{\lambda^2\|W\|_{L^2}^2} - \frac{c_1}{4}|t|^{-\frac{N-5}{N-6}} + \frac{1}{2\|W\|_{L^2}^2}\frac{\mathrm{d}}{\mathrm{d}t}\langle g, iA_0(\lambda)g\rangle \\ &\geq \frac{1}{\|W\|_{L^2}^2}\left(\frac{1}{\lambda^2}K - \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\langle g, iA_0(\lambda)g\rangle\right) - \frac{c_1}{2}|t|^{-\frac{N-5}{N-6}}, \end{aligned} \tag{4-22}$$

so we need to compute $\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\langle g, iA_0(\lambda)g\rangle$, up to terms of order $\ll |t|^{-\frac{N-5}{N-6}}$. In this proof, the sign $\simeq$ will mean "up to terms of order $\ll |t|^{-\frac{N-5}{N-6}}$ as $|T_0| \to +\infty$".

Since $iA_0(\lambda)$ is symmetric, we have

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\langle g, iA_0(\lambda)g\rangle = \frac{1}{2}\lambda'\langle g, i\partial_\lambda A_0(\lambda)g\rangle + \langle \partial_t g, iA_0(\lambda)g\rangle. \tag{4-23}$$

The first term is of size $\lesssim \left|\frac{\lambda'}{\lambda}\right| \cdot \|g\|_{\mathcal{E}}^2 \ll |t|^{-\frac{N-5}{N-6}}$, and hence is negligible. We expand $\partial_t g$ according to (3-3). Consider the terms in the second line of (3-3). It follows from (3-34) and the fact that $\|A_0(\lambda)g\|_{\dot{H}^{-1}} \lesssim \|g\|_{\mathcal{E}}$ that their contribution is $\lesssim |t|^{-1}\|g\|_{\mathcal{E}} \leq |t|^{-\frac{3N-13}{2(N-6)}} \ll |t|^{-\frac{N-5}{N-6}}$, and hence is negligible, so we can write

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\langle g, iA_0(\lambda)g\rangle \simeq \left\langle \Delta g + f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda), A_0(\lambda)g\right\rangle. \tag{4-24}$$

We now check that

$$\left|\left\langle f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda), A_0(\lambda)g\right\rangle\right| \ll |t|^{-\frac{N-5}{N-6}}. \tag{4-25}$$

The function $A_0(\lambda)g$ is supported in the ball of radius $\widetilde{R}\lambda$. In this region we have $W_\lambda \ll W_\mu$; hence (2-3) yields $|f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f(\mathrm{e}^{i\zeta}W_\mu) - f(\mathrm{e}^{i\theta}W_\lambda)| \lesssim |W_\lambda|^{\frac{4}{N-2}}$. By a change of variable we obtain

$$\|W_\lambda^{\frac{4}{N-2}}\|_{L^2(|x|\leq\widetilde{R}\lambda)} = \lambda^{\frac{N-2}{2}}\|W^{\frac{4}{N-2}}\|_{L^2(|x|\leq\widetilde{R})} \lesssim |t|^{-\frac{N-2}{N-6}}.$$

By the first property in Lemma 4.7, $\|A_0(\lambda)g\|_{L^2} \lesssim \lambda^{-1}\|g\|_{\mathcal{E}} \lesssim |t|^{-\frac{N-5}{2(N-6)}}$; hence the Cauchy–Schwarz inequality implies (4-25) (with a large margin). By the triangle inequality, (4-24) and (4-25) yield

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\langle g, iA_0(\lambda)g\rangle \simeq \big\langle \Delta g + f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda), A_0(\lambda)g\big\rangle.$$

We transform the right-hand side using (4-11), (4-13) and the fact that $A_0(\lambda)g = \frac{1}{N\lambda^2}\Delta q\big(\frac{\cdot}{\lambda}\big)g + A(\lambda)g$. Note that for any $c_2 > 0$ we have

$$\frac{c_0}{\lambda^2}\|g\|_{\mathcal{E}}^2 \le \frac{c_2}{2}|t|^{-\frac{N-5}{N-6}}$$

if we choose c_0 small enough; thus

$$\begin{aligned}
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\langle g, iA_0(\lambda)g\rangle \\
&\le c_2|t|^{-\frac{N-5}{N-6}} - \frac{1}{\lambda^2}\bigg(\int_{|x|\le R\lambda}|\nabla g|^2\,\mathrm{d}x - \big\langle f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda), \tfrac{1}{N}\Delta q\big(\tfrac{\cdot}{\lambda}\big)g\big\rangle\bigg) \\
&\quad - \big\langle A(\lambda)(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda), f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda)g\big\rangle, \quad (4\text{-}26)
\end{aligned}$$

where c_2 can be made arbitrarily small. Consider the second line. We will check that

$$\Big|\big\langle f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda), \tfrac{1}{N}\Delta q\big(\tfrac{\cdot}{\lambda}\big)g\big\rangle - \langle f'(\mathrm{e}^{i\theta}W_\lambda)g, g\rangle\Big| \ll |t|^{-\frac{N-1}{N-6}}. \quad (4\text{-}27)$$

Indeed, Δq is bounded; hence $\big\|\frac{1}{N}\Delta q\big(\frac{\cdot}{\lambda}\big)g\big\|_{L^{\frac{2N}{N-2}}} \lesssim \|g\|_{\mathcal{E}}$. By (2-4) we have

$$\|f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda)g\|_{L^{2N/(N+2)}} \lesssim \|g\|_{\mathcal{E}}^{\frac{N+2}{N-2}} \ll \|g\|_{\mathcal{E}}.$$

Now from (2-2) we obtain

$$\big\|\big(f'(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda) - f'(\mathrm{e}^{i\theta}W_\lambda)\big)g\big\|_{L^{2N/(N+2)}(|x|\le\widetilde{R}\lambda)} \lesssim \|f'(\mathrm{e}^{i\zeta}W_\mu)\|_{L^{N/2}(|x|\le\widetilde{R}\lambda)}\|g\|_{\mathcal{E}} \ll \|g\|_{\mathcal{E}}.$$

We have obtained

$$\Big|\big\langle f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda + g) - f(\mathrm{e}^{i\zeta}W_\mu + \mathrm{e}^{i\theta}W_\lambda), \tfrac{1}{N}\Delta q\big(\tfrac{\cdot}{\lambda}\big)g\big\rangle - \big\langle f'(\mathrm{e}^{i\theta}W_\lambda)g, \tfrac{1}{N}\Delta q\big(\tfrac{\cdot}{\lambda}\big)g\big\rangle\Big| \ll |t|^{-\frac{N-1}{N-6}}.$$

But $\frac{1}{N}\Delta q\big(\frac{x}{\lambda}\big) = 1$ for $|x| \le R\lambda$ and $\|f'(\mathrm{e}^{i\theta}W_\lambda)\|_{L^{N/2}(|x|\ge R\lambda)} \ll 1$ for R large. This proves (4-27).

The bounds (4-5) and (4-17) together with (2-38) imply

$$\int_{|x|\le R\lambda}|\nabla g|^2\,\mathrm{d}x - \langle f'(\mathrm{e}^{i\theta}W_\lambda))g, g\rangle \ge -c_3\|g\|_{\mathcal{E}}^2,$$

with c_3 as small as we like by enlarging R. Thus, we have obtained that the second line in (4-26) is $\le c_2|t|^{-\frac{N-5}{N-6}}$, with c_2 which can be made arbitrarily small.

We are left with the third line of (4-26). We will show that it equals $\frac{1}{\lambda^2}K$ up to negligible terms. The support of $A(\lambda)(\mathrm{e}^{i\zeta}W_\mu)$ is contained in $|x| \le \widetilde{R}\lambda$ and $\|A(\lambda)(\mathrm{e}^{i\zeta}W_\mu)\|_{L^\infty} \lesssim \lambda^{-2}$; hence

$$\|A(\lambda)(\mathrm{e}^{i\zeta}W_\mu)\|_{L^{2N/N-2}} \lesssim \big(\lambda^N\lambda^{-\frac{4N}{N-2}}\big)^{\frac{N-2}{2N}} = \lambda^{\frac{N-6}{2}} \sim |t|^{-1}.$$

From (2-4) and Hölder we have

$$\|f(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda+g)-f(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda)-f'(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda)g\|_{L^{2N/(N+2)}}\lesssim\|g\|_{\mathcal{E}}^{\frac{N+2}{N-2}}\ll|t|^{-\frac{1}{N-6}}.$$

Thus, in the third line of (4-26) we can replace $A(\lambda)(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda)$ by $A(\lambda)(\mathrm{e}^{i\theta}W_\lambda)$. Property (P3) implies $|AW-\Lambda W|\lesssim W$ pointwise, with a constant independent of c and R used in the definition of the function q. After rescaling and phase change we obtain $\big|A(\lambda)(\mathrm{e}^{i\theta}W_\lambda)-\frac{1}{\lambda^2}\mathrm{e}^{i\theta}\Lambda W_\lambda\big|\lesssim\frac{1}{\lambda^2}W_\lambda$. But $A(\lambda)W=\frac{1}{\lambda^2}\Lambda W_\lambda$ for $|x|\le R\lambda$, so we obtain

$$\begin{aligned}\Big|\Big\langle A(\lambda)(\mathrm{e}^{i\theta}W_\lambda)-\frac{1}{\lambda^2}\mathrm{e}^{i\theta}\Lambda W_\lambda,\,f(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda+g)-f(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda)-f'(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda)g\Big\rangle\Big|&\\ \lesssim\frac{1}{\lambda^2}\int_{|x|\ge R\lambda}W_\lambda\cdot\big|f(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda+g)-f(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda)-f'(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda)g\big|\,\mathrm{d}x.&\end{aligned}$$

Since $|\zeta-\theta|\simeq\frac{\pi}{2}$, we have $|\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda|\gtrsim W_\lambda$; hence (2-4) yields

$$W_\lambda\cdot\big|f(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda+g)-f(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda)-f'(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda)g\big|\lesssim W_\lambda^{\frac{4}{N-2}}|g|^2.$$

Integrating over $|x|\ge R\lambda$ and using Hölder we find

$$\begin{aligned}\Big|\Big\langle A(\lambda)(\mathrm{e}^{i\theta}W_\lambda)-\frac{1}{\lambda^2}\mathrm{e}^{i\theta}\Lambda W_\lambda,\,f(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda+g)-f(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda)-f'(\mathrm{e}^{i\zeta}W_\mu+\mathrm{e}^{i\theta}W_\lambda)g\Big\rangle\Big|&\\ \lesssim c_2|t|^{-\frac{N-5}{N-6}},\quad\text{with } c_2 \text{ arbitrarily small as } R\to+\infty.&\end{aligned}$$

Resuming all the computations starting with (4-23), we have shown that

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\langle g,iA_0(\lambda)g\rangle\le\frac{c_1}{2}|t|^{-\frac{N-5}{N-6}}+\frac{1}{\lambda^2}K.$$

Hence (4-22) yields (4-21).

Since $\theta(T)=0$, we have $|\theta(T)|\lesssim\|g(T)\|_{\mathcal{E}}^2\ll|T|^{-\frac{1}{N-6}}$. Integrating (4-21) on $[T,t]$ we get $\psi(t)\gtrsim -c_1|t|^{-\frac{1}{N-6}}$. But

$$|\langle g(t),A_0(\lambda)g(t)\rangle|\lesssim\|g(t)\|_{\mathcal{E}}^2\le|t|^{-\frac{N-1}{N-6}}\ll|t|^{-\frac{1}{N-6}};$$

hence we obtain $\theta(t)\gtrsim -c_1|t|^{-\frac{1}{N-6}}$, which yields (4-20) if c_1 is chosen small enough. This finishes the proof of (4-8).

Step 3. From (2-20) we obtain $\|g\|_{\mathcal{E}}^2+C_0\theta\lambda^{\frac{N-2}{2}}\le C_1|t|^{-\frac{N}{N-6}}$; hence

$$\|g\|_{\mathcal{E}}^2\le -C_0\theta\lambda^{\frac{N-2}{2}}+C_1|t|^{-\frac{N}{N-6}}\le\tfrac{1}{8}|t|^{-\frac{N-1}{N-6}}+C_1|t|^{-\frac{N}{N-6}},$$

provided that c_0 in (4-18) is small enough. This yields (4-9). □

Choice of the initial data by a topological argument. The bootstrap in Proposition 4.4 leaves out the control of $\lambda(t)$, $a_1^+(t)$ and $a_2^+(t)$. We will tackle this problem here.

Proposition 4.8. *Let $|T_0|$ be large enough. For all $T < T_0$ there exist λ^0, a_1^0, a_2^0 satisfying (4-1) such that the solution $u(t)$ with the initial data $u(T) = -iW + W_{\lambda^0} + g^0$ exists on the time interval $[T, T_0]$ and for $t \in [T, T_0]$ the bounds (4-6)–(4-9) and*

$$\left|\lambda(t) - \tfrac{1}{\kappa}(\kappa|t|)^{-\frac{2}{N-6}}\right| \le \tfrac{1}{2}|t|^{-\frac{5}{2(N-6)}}, \tag{4-28}$$

$$|a_1^+(t)| \le \tfrac{1}{2}|t|^{-\frac{N}{2(N-6)}}, \tag{4-29}$$

$$|a_2^+(t)| \le \tfrac{1}{2}|t|^{-\frac{N}{2(N-6)}} \tag{4-30}$$

hold.

The proof will be split into some lemmas. For $t \in [T, T_0]$, $\tilde\lambda > 0$, $\tilde a_1 \in \mathbb{R}$ and $\tilde a_2 \in \mathbb{R}$ we define

$$X_t(\tilde\lambda, \tilde a_1, \tilde a_2) := \left(\tfrac{1}{\kappa}(\kappa|t|)^{-\frac{2}{N-6}} + \tilde\lambda|t|^{-\frac{5}{2(N-6)}}, \tilde a_1|t|^{-\frac{N}{2(N-6)}}, \tilde a_2|t|^{-\frac{N}{2(N-6)}}\right).$$

We see that $\lambda(t)$, $a_1^+(t)$ and $a_2^+(t)$ satisfy (4-28)–(4-30) if and only if

$$X_t^{-1}\left(\lambda(t), a_1^+(t), a_2^+(t)\right) \in Q := \left[-\tfrac{1}{2}, \tfrac{1}{2}\right]^3.$$

Lemma 4.9. *Assume that $\lambda(t)$, $a_1^+(t)$ and $a_2^+(t)$ satisfy (3-11), (3-35) and (3-36) on the time interval $t \in (T_1, T_2)$ and that*

$$(p_0, p_1, p_2) := X_t^{-1}\left(\lambda(t), a_1^+(t), a_2^+(t)\right) \in Q \setminus \partial Q \quad \textit{for all } t \in (T_1, T_2).$$

Then for all $t \in (T_1, T_2)$,

$$\left|p_0'(t) - \frac{2N-13}{2(N-6)}|t|^{-1}p_0(t)\right| \le c|t|^{-1}, \tag{4-31}$$

$$\left|p_1'(t) - \frac{\nu}{\mu(t)}p_1(t)\right| \le \frac{c}{\mu(t)}, \tag{4-32}$$

$$\left|p_2'(t) - \frac{\nu}{\lambda(t)}p_2(t)\right| \le \frac{c}{\lambda(t)}, \tag{4-33}$$

where $c > 0$ can be made arbitrarily small by taking T_0 large enough.

Proof. By the definition of $p_0(t)$ we have

$$\lambda(t) = \tfrac{1}{\kappa}(\kappa|t|)^{-\frac{2}{N-6}} + p_0(t)|t|^{-\frac{5}{2(N-6)}}. \tag{4-34}$$

Differentiating in time we obtain

$$\lambda'(t) = \frac{2}{N-6}(\kappa|t|)^{-\frac{N-4}{N-6}} + \frac{5}{2(N-6)}|t|^{-\frac{2N-7}{2(N-6)}}p_0(t) + |t|^{-\frac{5}{2(N-6)}}p_0'(t).$$

Applying the Newton formula (the binomial expansion with power $\frac{N-4}{2}$) to (4-34) and using the fact that $|p_0| \lesssim 1$ we get

$$\lambda(t)^{\frac{N-4}{2}} = \kappa^{-\frac{N-4}{2}}(\kappa|t|)^{-\frac{N-4}{N-6}} + \frac{N-4}{2}\kappa^{-\frac{N-6}{2}}(\kappa|t|)^{-1}p_0(t)|t|^{-\frac{5}{2(N-6)}} + O(|t|^{-\frac{N-3}{N-6}}).$$

Thus

$$\lambda'(t) - \frac{2\kappa^{\frac{N-4}{2}}}{N-6}\lambda(t)^{\frac{N-4}{2}} = \left(\frac{5}{2(N-6)} - \frac{N-4}{N-6}\right)|t|^{-\frac{2N-7}{2(N-6)}} p_0(t) + |t|^{-\frac{5}{2(N-6)}} p_0'(t) + O(|t|^{-\frac{N-3}{N-6}}).$$

Using (3-11) and multiplying both sides by $|t|^{\frac{5}{2(N-6)}}$ we obtain (4-31).

We have $a_1^+(t) = |t|^{-\frac{N}{2(N-6)}} p_1(t)$, which yields

$$\frac{\mathrm{d}}{\mathrm{d}t}a_1^+ - \frac{\nu}{\mu}a_1^+ = |t|^{-\frac{N}{2(N-6)}}\Big(p_1'(t) - \frac{\nu}{\mu}p_1(t)\Big) + O\big(|t|^{-\frac{N}{2(N-6)}-1}\big),$$

so (3-35) implies (4-32). The proof of (4-33) is similar. □

For $C > 1$, $j \in \{0, 1, 2\}$ and $p \in \mathbb{R}^3$ we define

$$V_j(C, p) := \big\{p + (r_0, r_1, r_2) : \operatorname{sign}(r_j) = \operatorname{sign}(p_j) \text{ and } \max_j |r_j| < C|r_j|\big\}.$$

Lemma 4.10. *Assume that $\lambda(t)$, $a_1^+(t)$ and $a_2^+(t)$ satisfy (3-8), (3-11), (4-5), (3-35) and (3-36) for $t \in (T_1, T_2)$. There exists a constant $C > 0$, depending on T_1 and T_2, such that if for some $T_3 \in (T_1, T_2)$ and $j \in \{0, 1, 2\}$ we have $|p_j(T_3)| \geq \frac{1}{4}$, then for all $t \in (T_3, T_2)$ we have $p(t) \in V_j(C, p(T_3))$.*

Proof. From the previous lemma we infer that there exist strictly positive constants c_1 and C_1, depending on T_1 and T_2, such that $|p_j'(t)| \leq C_1$ and

$$|p_j(t)| \geq \tfrac{1}{4} \quad \Longrightarrow \quad |p_j'(t)| \geq c_1 \text{ and } \operatorname{sign} p_j'(t) = \operatorname{sign} p_j(t).$$

It is sufficient to take $C > \frac{C_1}{c_1}$. □

Proof of Proposition 4.8. The proof proceeds by contradiction. Supposing that the result does not hold, we will construct a continuous retraction $\Phi : Q \to \partial Q$, $\Phi(p) = p$ for $p \in \partial Q$. It is a well-known fact from topology that such a function Φ does not exist.

Let $p^0 \in Q$. Take $(\lambda^0, \tilde{a}_1^0, \tilde{a}_2^0) = X_T(p^0)$ and let g^0 be given by Lemma 4.1. Let $u : [T, T_+) \to \mathcal{E}$ be the solution of (1-1) for the initial data $u(T) = -iW + W_{\lambda^0} + g^0$. We will say that the solution u is associated with $p^0 \in Q$.

Let T_2 be the infimum of the values of $t \in [T, T_+)$ such that (4-6), (4-7), (4-8), (4-9), (4-28), (4-29) or (4-30) does not hold. By our assumption that Proposition 4.8 is false, we have that T_2 exists and $T_2 < T_0$. Indeed, if all the listed conditions were satisfied for $t \in [T, T_+)$, then Corollary A.3 would imply $T_+ > T_0$; hence all the conditions would hold on $[T, T_0]$, which contradicts the assumption.

Set $p^1 := X_{T_2}^{-1}\big(\lambda(T_2), a_1^+(T_2), a_2^+(T_2)\big)$. By continuity $p^1 \in Q$, and we will show that in fact $p^1 \in \partial Q$. Indeed, by continuity of the flow, the assumptions of Proposition 4.4 are satisfied for $T_1 = T_2 + \tau$ for some $\tau > 0$. Hence (4-6)–(4-9) continue to hold on $[T_2, T_2 + \tau]$, so one of the conditions (4-28), (4-29) or (4-30) is violated somewhere on $[T_2, T_2 + \tau]$ for every $\tau > 0$. By continuity of the parameters with respect to time, this yields $p^1 \in \partial Q$.

We set

$$\Phi : Q \to \partial Q, \quad \Phi(p^0) := p^1.$$

It is immediate from the definition that $\Phi(p) = p$ for $p \in \partial Q$, and it remains to show that Φ is continuous.

Let $p^0 \in Q$, $\Phi(p^0) = p^1 \in \partial Q$ and $\varepsilon > 0$. Let C be the constant from Lemma 4.10 for $T_1 = T$ and $T_2 = T_0$. We will consider the case $p_0^1 = \frac{1}{2}$, the other cases being similar. It is clear that for $\delta > 0$ small enough $V_\delta := V_0\big(C, \frac{1}{2} - \delta, p_1^1, p_2^1\big) \cap \partial Q$ is an ε-neighborhood of p^1. Thus, by Lemma 4.10, in order to finish the proof it suffices to show that if $q^0 \in Q$ with $|q^0 - p^0|$ small enough, then the solution associated with q passes through V_δ.

If $p^0 = p^1 \in \partial Q$, this is obvious, since V_δ is in this case a neighborhood of p^0. In the case $p^0 \in Q \setminus \partial Q$, the solution associated with p^0 passes through V_δ before reaching ∂Q. Thus, by the continuous dependence on the initial data, the solution associated with q^0 passes through V_δ if $|q^0 - p^0|$ is small enough. □

Proof of Theorem 1. Let $T_0 < 0$ be given by Proposition 4.8 and let $T_0, T_1, T_2, \dots$ be a decreasing sequence tending to $-\infty$. For $n \geq 1$, let u_n be the solution given by Proposition 4.8. Inequalities (4-6), (4-7), (4-8), (4-28) and (4-9) yield

$$\big\|u_n(t) - (-iW + W_{\frac{1}{\kappa}(\kappa|t|)^{-2/(N-6)}})\big\|_{\mathcal{E}} \lesssim |t|^{-\frac{1}{2(N-6)}} \tag{4-35}$$

for all $t \in [T_n, T_0]$ and with a constant independent of n. Upon passing to a subsequence, we can assume that $u_n(T_0) \rightharpoonup u_0 \in \mathcal{E}$. Let u be the solution of (1-1) with the initial condition $u(T_0) = u_0$. Corollary A.4 implies u exists on the time interval $(-\infty, T_0]$ and for all $t \in (-\infty, T_0]$ we have $u_n(t) \rightharpoonup u(t)$. Passing to the weak limit in (4-35) finishes the proof. □

Appendix: Cauchy theory

Profile decomposition. We recall briefly the profile decomposition method of Bahouri and Gérard [1999] and Merle and Vega [1998]. In the case of the energy-critical defocusing NLS, the corresponding theory was developed by Keraani [2001]. For the focusing NLS in high dimensions, which is the case discussed in this paper, see [Killip and Visan 2010].

Proposition A.1 (Killip, Visan). *Let $u_{0,n}$ be a bounded sequence in $\mathcal{E}$. There exists a subsequence of $u_{0,n}$, still denoted $u_{0,n}$, such that there exist a family of solutions of the linear Schrödinger equation $U_{\mathrm{L}}^j(t) = \mathrm{e}^{it\Delta}U_0^j$ and a family of sequences of parameters t_n^j and λ_n^j satisfying the pseudo-orthogonality condition*

$$j \neq k \implies \lim_{n\to+\infty} \frac{\lambda_n^j}{\lambda_n^k} + \frac{\lambda_n^k}{\lambda_n^j} + \frac{|t_n^j - t_n^k|}{\lambda_n^j} = +\infty$$

such that for all $J \geq 0$

$$u_{0,n} = \sum_{j=1}^{J} U_{\mathrm{L}}^j\Big(\frac{-t_n^j}{\lambda_n^j}\Big)_{\lambda_n^j} + w_n^J, \tag{A-1}$$

with

$$\lim_{J\to+\infty}\limsup_{n\to+\infty} \|\mathrm{e}^{it\Delta}w_n^J\|_{L_{t,x}^{2(N+2)/(N-2)}} = 0.$$

Moreover, for any $J \geq 0$,

$$\lim_{n\to+\infty}\Big|\|u_{0,n}\|_{\mathcal{E}}^2 - \sum_{j=1}^{J}\|U_0^j\|_{\mathcal{E}}^2 - \|w_n^J\|_{\mathcal{E}}^2\Big| = 0. \qquad □$$

Formula (A-1) is called the linear profile decomposition. In the applications, we regard $u_{0,n}$ as a sequence of initial data of solutions u_n of (1-1). In order to approximate the solutions u_n, we introduce nonlinear profiles. The nonlinear profile U^j corresponding to the linear profile U_{L}^j is defined as the solution of (1-1) such that

$$\lim_{n\to+\infty}\left\|U^j\left(\frac{-t_n^j}{\lambda_n^j}\right)-U_{\mathrm{L}}^j\left(\frac{-t_n^j}{\lambda_n^j}\right)\right\|_{\mathcal{E}}=0.$$

The next proposition is a version of the result of Keraani for the focusing NLS. Its statement is very similar to Proposition 2.8 in [Duyckaerts et al. 2011].

Proposition A.2. *Let $u_{0,n}$ be a sequence in $\mathcal{E}$ with a linear profile decomposition (A-1) and let $U^j : (T_-(U^j), T_+(U^j)) \to \mathcal{E}$ be the nonlinear profiles. Let $\tau_n > 0$ be a sequence such that for all j and n*

$$\frac{\tau_n-t_n^j}{(\lambda_n^j)^2}<T_+(U^j),\qquad \limsup_{n\to+\infty}\|U^j\|_{L^{2(N+2)/(N-2)}(I\times\mathbb{R}^N)}<+\infty,\quad \textit{where } I=\left[\frac{-t_n^j}{(\lambda_n^j)^2},\frac{\tau_n-t_n^j}{(\lambda_n^j)^2}\right].$$

Let u_n be the solution of (1-1) with the initial data $u_n(0)=u_{0,n}$. Then, for n large, u_n exists on the time interval $[0,\tau_n]$, $\limsup_{n\to+\infty}\|u_n\|_{L^{2(N+2)/(N-2)}([0,\tau_n]\times\mathbb{R}^N)}<+\infty$ and for all $J\ge 0$,

$$u_n(t)=\sum_{j=1}^{J}U^j\left(\frac{t-t_n^j}{(\lambda_n^j)^2}\right)_{\lambda_n^j}+w_n^J(t)+r_n^J(t),$$

with

$$\lim_{J\to+\infty}\limsup_{n\to+\infty}\big(\|r_n^J\|_{L^{2(N+2)/(N-2)}([0,\tau_n]\times\mathbb{R}^N)}+\sup_{t\in[0,\tau_n]}\|r_n^J\|_{\mathcal{E}}\big)=0.$$

Proof. See [Duyckaerts et al. 2011, proof of Proposition 2.8] and [Killip and Visan 2010, proof of Lemma 3.2]. □

Corollaries.

Corollary A.3. *There exists a constant $\eta>0$ such that the following holds. Let $u:[t_0,T_+)\to\mathcal{E}$ be a maximal solution of (1-1) with $T_+<+\infty$. Then for any compact set $K\subset\mathcal{E}$ there exists $\tau<T_+$ such that* $\operatorname{dist}(u(t),K)>\eta$ *for $t\in[\tau,T_+)$.*

Proof. See [Jendrej 2016, Corollary A.4]. □

Corollary A.4. *There exists a constant $\eta>0$ such that the following holds. Let $K\subset\mathcal{E}$ be a compact set and let $u_n:[T_1,T_2]\to\mathcal{E}$ be a sequence of solutions of (1-1) such that*

$$\operatorname{dist}(u_n(t),K)\le\eta\quad \textit{for all } n\in\mathbb{N} \textit{ and } t\in[T_1,T_2].$$

Suppose that $u_n(T_1)\rightharpoonup u_0\in\mathcal{E}$. Then the solution $u(t)$ of (1-1) with the initial condition $u(T_1)=u_0$ is defined for $t\in[T_1,T_2]$ and

$$u_n(t)\rightharpoonup u(t)\quad \textit{for all } t\in[T_1,T_2].$$

Proof. See [Jendrej 2016, Corollary A.6]. □

Acknowledgments

Part of this work was realized when I was a Ph.D. student at École Polytechnique. I would like to thank my advisors Yvan Martel and Frank Merle for encouraging me to pursue this project. I was partially supported by the ERC grant 291214 BLOWDISOL.

References

[Aubin 1976] T. Aubin, "Équations différentielles non linéaires et problème de Yamabe concernant la courbure scalaire", *J. Math. Pures Appl.* (9) **55**:3 (1976), 269–296. MR Zbl

[Bahouri and Gérard 1999] H. Bahouri and P. Gérard, "High frequency approximation of solutions to critical nonlinear wave equations", *Amer. J. Math.* **121**:1 (1999), 131–175. MR Zbl

[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 Zbl

[Cazenave and Weissler 1990] T. Cazenave and F. B. Weissler, "The Cauchy problem for the critical nonlinear Schrödinger equation in H^s", *Nonlinear Anal.* **14**:10 (1990), 807–836. MR Zbl

[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 Zbl

[Collot et al. 2017] C. Collot, F. Merle, and P. Raphaël, "Dynamics near the ground state for the energy critical nonlinear heat equation in large dimensions", *Comm. Math. Phys.* **352**:1 (2017), 215–285. MR Zbl

[Duyckaerts and Merle 2009] T. Duyckaerts and F. Merle, "Dynamic of threshold solutions for energy-critical NLS", *Geom. Funct. Anal.* **18**:6 (2009), 1787–1840. MR Zbl

[Duyckaerts et al. 2011] T. Duyckaerts, C. Kenig, and F. Merle, "Universality of blow-up profile for small radial type II blow-up solutions of the energy-critical wave equation", *J. Eur. Math. Soc.* (*JEMS*) **13**:3 (2011), 533–599. MR Zbl

[Duyckaerts et al. 2013] T. Duyckaerts, C. Kenig, and F. Merle, "Classification of radial solutions of the focusing, energy-critical wave equation", *Camb. J. Math.* **1**:1 (2013), 75–144. MR Zbl

[Duyckaerts et al. 2017] T. Duyckaerts, H. Jia, C. Kenig, and F. Merle, "Soliton resolution along a sequence of times for the focusing energy critical wave equation", *Geom. Funct. Anal.* **27**:4 (2017), 798–862. MR

[Fan 2016] C. Fan, "The L^2 weak sequential convergence of radial mass critical NLS solutions with mass above the ground state", preprint, 2016. arXiv

[Jendrej 2015] J. Jendrej, "Nonexistence of radial two-bubbles with opposite signs for the energy-critical wave equation", 2015. To appear in *Ann. Sc. Norm. Super. Pisa Cl. Sci.* as "On two-bubble solutions for the energy-critical wave equation: nonexistence in the case of opposite signs". arXiv

[Jendrej 2016] J. Jendrej, "Construction of two-bubble solutions for energy-critical wave equations", preprint, 2016. To appear in *Amer. J. Math.* arXiv

[Kenig and Merle 2006] C. E. Kenig and F. Merle, "Global well-posedness, scattering and blow-up for the energy-critical, focusing, non-linear Schrödinger equation in the radial case", *Invent. Math.* **166**:3 (2006), 645–675. MR Zbl

[Keraani 2001] S. Keraani, "On the defect of compactness for the Strichartz estimates of the Schrödinger equations", *J. Differential Equations* **175**:2 (2001), 353–392. MR Zbl

[Killip and Visan 2010] R. Killip and M. Visan, "The focusing energy-critical nonlinear Schrödinger equation in dimensions five and higher", *Amer. J. Math.* **132**:2 (2010), 361–424. MR Zbl

[Krieger et al. 2008] J. Krieger, W. Schlag, and D. Tataru, "Renormalization and blow up for charge one equivariant critical wave maps", *Invent. Math.* **171**:3 (2008), 543–615. MR Zbl

[Krieger et al. 2009] J. Krieger, W. Schlag, and D. Tataru, "Slow blow-up solutions for the $H^1(\mathbb{R}^3)$ critical focusing semilinear wave equation", *Duke Math. J.* **147**:1 (2009), 1–53. MR Zbl

[Martel 2005] Y. Martel, "Asymptotic N-soliton-like solutions of the subcritical and critical generalized Korteweg-de Vries equations", *Amer. J. Math.* **127**:5 (2005), 1103–1140. MR Zbl

[Martel and Raphaël 2015] Y. Martel and P. Raphaël, "Strongly interacting blow up bubbles for the mass critical NLS", preprint, 2015. To appear in *Ann. Sci. École Norm. Sup.* arXiv

[Merle 1990] F. Merle, "Construction of solutions with exactly k blow-up points for the Schrödinger equation with critical nonlinearity", *Comm. Math. Phys.* **129**:2 (1990), 223–240. MR Zbl

[Merle and Raphaël 2004] F. Merle and P. Raphaël, "On universality of blow-up profile for L^2 critical nonlinear Schrödinger equation", *Invent. Math.* **156**:3 (2004), 565–672. MR Zbl

[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 Zbl

[Merle and Vega 1998] F. Merle and L. Vega, "Compactness at blow-up time for L^2 solutions of the critical nonlinear Schrödinger equation in 2D", *Internat. Math. Res. Notices* **1998**:8 (1998), 399–425. MR Zbl

[Merle et al. 2013] F. Merle, P. Raphaël, and I. Rodnianski, "Blowup dynamics for smooth data equivariant solutions to the critical Schrödinger map problem", *Invent. Math.* **193**:2 (2013), 249–365. MR Zbl

[Nakanishi and Roy 2016] K. Nakanishi and T. Roy, "Global dynamics above the ground state for the energy-critical Schrödinger equation with radial data", *Commun. Pure Appl. Anal.* **15**:6 (2016), 2023–2058. MR Zbl

[Ortoleva and Perelman 2013] C. Ortoleva and G. Perelman, "Nondispersive vanishing and blow up at infinity for the energy critical nonlinear Schrödinger equation in $\mathbb{R}^3$", *Algebra i Analiz* **25**:2 (2013), 162–192. In Russian; translated in *St. Petersburg Math. J.* **25**:2 (2014), 271–294. MR Zbl

[Perelman 2014] G. Perelman, "Blow up dynamics for equivariant critical Schrödinger maps", *Comm. Math. Phys.* **330**:1 (2014), 69–105. MR Zbl

[Ryckman and Visan 2007] E. Ryckman and M. Visan, "Global well-posedness and scattering for the defocusing energy-critical nonlinear Schrödinger equation in $\mathbb{R}^{1+4}$", *Amer. J. Math.* **129**:1 (2007), 1–60. MR Zbl

[Talenti 1976] G. Talenti, "Best constant in Sobolev inequality", *Ann. Mat. Pura Appl.* (4) **110** (1976), 353–372. MR Zbl

[Tao 2005] T. Tao, "Global well-posedness and scattering for the higher-dimensional energy-critical nonlinear Schrödinger equation for radial data", *New York J. Math.* **11** (2005), 57–80. MR Zbl

[Visan 2007] M. Visan, "The defocusing energy-critical nonlinear Schrödinger equation in higher dimensions", *Duke Math. J.* **138**:2 (2007), 281–374. MR Zbl

Received 7 Nov 2016. Revised 8 May 2017. Accepted 28 Jul 2017.

JACEK JENDREJ: jacek@math.uchicago.edu
Department of Mathematics, University of Chicago, Chicago, IL, United States

ANALYSIS AND PDE
Vol. 10, No. 8, 2017
dx.doi.org/10.2140/apde.2017.10.1961

msp

BILINEAR RESTRICTION ESTIMATES FOR SURFACES OF CODIMENSION BIGGER THAN 1

JONG-GUK BAK, JUNGJIN LEE AND SANGHYUK LEE

In connection with the restriction problem in $\mathbb{R}^n$ for hypersurfaces including the sphere and paraboloid, the bilinear (adjoint) restriction estimates have been extensively studied. However, not much is known about such estimates for surfaces with codimension (and dimension) larger than 1. In this paper we show sharp bilinear $L^2 \times L^2 \to L^q$ restriction estimates for general surfaces of higher codimension. In some special cases, we can apply these results to obtain the corresponding linear estimates.

1. Introduction and statement of results

For a smooth hypersurface S such as the sphere or paraboloid in $\mathbb{R}^n$, $n \ge 3$, the L^p-L^q boundedness of the (adjoint) restriction operator (or the extension operator) $\widehat{f\,d\sigma}$ has been extensively studied since the late 1960s. Here $d\sigma$ denotes the induced Lebesgue measure on S. Specifically, when S is the sphere, it was conjectured by E. M. Stein [1993] that $\widehat{f\,d\sigma}$ should map $L^p(S)$ boundedly to $L^q(\mathbb{R}^n)$, precisely when $q \ge p'(n+1)/(n-1)$ and $q > 2n/(n-1)$. Since then, a large amount of literature has been devoted to this problem. Over the last couple of decades, the bilinear and multilinear approaches have proven to be quite effective, and through them substantial progress has been made. We refer the reader to [Bennett et al. 2006; Bourgain and Guth 2011; Guth 2016] for the most recent developments.

On the other hand, when the dimension of the manifold is 1, namely, when the associated surface is a curve, the restriction estimate is by now fairly well understood [Bak et al. 2002; 2009; 2013; Stovall 2016].

However, not much is known about the intermediate cases, namely, when the codimension k of the manifold is between 1 and $n-1$. The restriction problem for quadratic surfaces of codimension $k \ge 2$ was first studied by Christ [1982] and Mockenhaupt [1996]. They also considered the problem in a more general setting and found some necessary conditions on the curvature and codimension of the surface. For some surfaces they also established the optimal $L^2 \to L^q$ linear estimates, which may be regarded as generalizations of the Stein–Tomas restriction theorem; see also [Banner 2002]. Although there are some known cases in which the L^p-L^q boundedness is completely characterized, see for example [Bak and Ham 2014; Bak and Lee 2004; Oberlin 2005], for most surfaces with codimension bigger than 1, the current state of the restriction problem is hardly beyond that of the Stein–Tomas theorem.

J. Lee is supported in part by NRF grant No. 2017R1D1A1B03036053 (Republic of Korea). S. Lee is supported by NRF grant No. 2015R1A2A2A05000956 (Republic of Korea). J. Bak was supported in part by the Basic Science Research Institute (BSRI), POSTECH.
MSC2010: 42B15, 42B20.
Keywords: Fourier transform of measures, complex surfaces, Fourier restriction estimates.

In this paper, we are concerned with restriction estimates for surfaces of codimension $k \geq 2$. To be more specific, let us set $k \geq 1$ and $I = [-1, 1]$. Let $\Phi : I^d \to \mathbb{R}^k$ be a smooth function given by

$$\Phi(\xi) = (\varphi_1(\xi), \varphi_2(\xi), \dots, \varphi_k(\xi)).$$

The adjoint restriction operator (the extension operator) $E = E_\Phi$ for the surface $(\xi, \Phi(\xi)) \in \mathbb{R}^d \times \mathbb{R}^k$ is defined by

$$Ef(x,t) = \int_{I^d} e^{2\pi i (x\cdot\xi + t\cdot\Phi(\xi))} f(\xi)\, d\xi, \quad (x,t) \in \mathbb{R}^d \times \mathbb{R}^k.$$

Specific examples of such operators with $2 \leq k \leq d-2$ can be found in [Bak and Ham 2014; Bak and Lee 2004; Christ 1982; Mockenhaupt 1996; Oberlin 2005]. (Also, see Section 5.)

There are some classes of surfaces for which the optimal L^2-L^q boundedness of E is well understood. In fact, using a Knapp-type example it is easy to see that E may be bounded from L^p to L^q only if $(d+2k)/q \leq d(1-1/p)$. Hence, the best possible L^2-L^q bound is that for $q = 2(d+2k)/d$. Christ [1982] and Mockenhaupt [1996] showed that this is true for a class of surfaces satisfying a suitable curvature condition. In particular, let M be a linear map from $\mathbb{R}^k$ to the space of $d \times d$ symmetric matrices and suppose that $\int_{S^{k-1}} |\det M(t)|^{-\gamma}\, d\sigma(t) < \infty$ for $\gamma = k/d$. Then it was proven in [Mockenhaupt 1996] that the extension operator E defined by $\Phi = \xi^t M(t)\xi$ is bounded from L^2 to $L^{2(d+2k)/d}$.

In order to obtain estimates for some $q < 2(d+2k)/d$ and $p > 2$, it seems necessary to consider methods other than the TT^* argument, which solely relies on the decay estimate for the Fourier transform of the surface measure. For this reason we wish to consider the bilinear restriction estimates for surfaces of codimension greater than 1 and try to obtain the best possible estimates.

Let S_1, S_2 be closed cubes contained in I^d and define

$$E_i f(x,t) = \int_{S_i} e^{2\pi i (x\cdot\xi + t\cdot\Phi(\xi))} f(\xi)\, d\xi, \quad i = 1, 2.$$

Let us consider the estimate

$$\|E_1 f\, E_2 g\|_{L^q(\mathbb{R}^{d+k})} \leq C \|f\|_{L^p(\mathbb{R}^d)} \|g\|_{L^p(\mathbb{R}^d)}. \tag{1-1}$$

For the elliptic surfaces, bilinear estimates can be thought of as a generalization of linear estimates, since a linear restriction estimate follows from the corresponding bilinear one by an argument involving a Whitney decomposition; see, e.g., [Tao et al. 1998]. The advantage of the bilinear estimates is that a wider rage of boundedness is possible than for the linear estimate, provided that a separation condition holds between the supports of the functions f, g. For surfaces with codimension 1, the sharp bilinear (adjoint) restriction estimate for the cone was obtained by Wolff [2001], and for the paraboloid the corresponding estimate was proved by Tao [2003]. The bilinear approach has also been applied to the restriction problem for hyperbolic surfaces: Vargas [2005] used it for the saddle surface in $\mathbb{R}^3$ and, independently, Lee [2006] proved the bilinear estimate by extending Tao's method.[1] From these bilinear restriction estimates the corresponding linear ones have been obtained as well.

[1]For more general negatively curved surfaces in $\mathbb{R}^3$ and higher dimensions, Lee [2006] showed the bilinear restriction estimates. However, in higher dimensions the linear estimate could not be deduced from the bilinear one, because the separation condition needed to prove the bilinear estimate for hyperbolic surfaces was more complex than that for the elliptic surfaces.

In order to state our results, we first introduce some notation. For $\nu_1,\ \nu_2 \in I^d$, we define the $k \times d$ matrix $\boldsymbol{D}(\nu_1, \nu_2)$ by

$$\boldsymbol{D}(\nu_1, \nu_2) = \begin{pmatrix} \nabla\varphi_1(\nu_2) - \nabla\varphi_1(\nu_1) \\ \vdots \\ \nabla\varphi_k(\nu_2) - \nabla\varphi_k(\nu_1) \end{pmatrix}.$$

Here $\nabla\varphi_j$ is a row vector. Let $H\varphi$ denote the Hessian of φ and $\boldsymbol{D}^t(\nu_1, \nu_2)$ be the transpose of $\boldsymbol{D}(\nu_1, \nu_2)$. The following is our main theorem.

Theorem 1.1. *Let* $t = (t_1, \dots, t_k)$, $k \geq 1$. *Suppose that, for* $\nu \in S_1 \cup S_2$ *and* $|t| = 1$,

$$\det\Big(\sum_{i=1}^{k} t_i H\varphi_i(\nu)\Big) \neq 0 \tag{1-2}$$

and, for $\nu_1 \in S_1,\ \nu_2 \in S_2,\ \ |t| = 1$ *and for* $\nu = \nu_1, \nu_2$,

$$\det\Big[\boldsymbol{D}(\nu_1, \nu_2)\Big(\sum_{j=1}^{k} t_j H\varphi_j(\nu)\Big)^{-1} \boldsymbol{D}^t(\nu_1, \nu_2)\Big] \neq 0. \tag{1-3}$$

Then, for

$$q > \frac{d+3k}{d+k} \quad \text{and} \quad \frac{1}{p} + \frac{d+3k}{d+k}\frac{1}{2q} < 1,$$

the estimate (1-1) *holds.*

As special cases of Theorem 1.1, one can deduce the known bilinear restriction theorems for the elliptic surfaces in [Tao 2003] and the negatively curved ones in [Vargas 2005; Lee 2006].

Let us set

$$\boldsymbol{M}(t, \nu_1, \nu_2, \nu) := \begin{pmatrix} 0 & \boldsymbol{D}(\nu_1, \nu_2) \\ \boldsymbol{D}^t(\nu_1, \nu_2) & \sum_{i=1}^{k} t_i H\varphi_i(\nu) \end{pmatrix}.$$

Assuming the condition (1-2), it is easy to see that (1-3) is equivalent to

$$\det \boldsymbol{M}(t, \nu_1, \nu_2, \nu) \neq 0 \tag{1-4}$$

for $\nu_1 \in S_1,\ \nu_2 \in S_2,\ |t| = 1$ and for $\nu = \nu_1, \nu_2$. $\big($One can use the block matrix formula $\det\big(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\big) = \det(D)\det(A - BD^{-1}C)$.$\big)$ The condition (1-4) may seem rather complicated, but such a condition appears naturally when one considers the bilinear $L^2 \times L^2 \to L^2$ estimate. When $k = 1$, it is closely related to the "rotational curvature"; see [Lee 2006] for more details. The necessity of the condition (1-4) will become clear in the course of the proof of Proposition 1.3 below.

From the condition (1-3) it follows that $\boldsymbol{D}(\nu_1, \nu_2)$ has rank k. So, the vectors

$$\{\nabla\varphi_i(\nu_2) - \nabla\varphi_i(\nu_1) : i = 1, \dots, k\}$$

are linearly independent. This means $d \geq k$. If $d = k$, then (1-4) implies (1-3), but otherwise (1-4) may hold without (1-3) being satisfied.

In fact, it is possible to obtain a local version (Theorem 1.2 below) of Theorem 1.1, which holds under a weaker assumption. Let $\boldsymbol{n}_1, \dots \boldsymbol{n}_{d-k}$ be orthonormal vectors (seen as row vectors), which are perpendicular to the span of $\{\nabla\varphi_i(\nu_2) - \nabla\varphi_i(\nu_1) : i = 1, \dots, k\}$ and set

$$N(\nu_2, \nu_1) = \begin{pmatrix} \boldsymbol{n}_1 \\ \vdots \\ \boldsymbol{n}_{d-k} \end{pmatrix}.$$

Then we can replace the condition (1-4) with

$$\det\Big[N(\nu_2,\nu_1)\Big(\sum_{i=1}^{k} t_i H\varphi_i(\nu)\Big)N^t(\nu_2,\nu_1)\Big] \neq 0 \tag{1-5}$$

whenever $\nu_1 \in S_1,\ \nu_2 \in S_2,\ |t| = 1$ and $\nu = \nu_1, \nu_2$. It is easy to see that the value of this determinant is independent of the particular choice of orthonormal vectors $\boldsymbol{n}_1, \dots, \boldsymbol{n}_{d-k}$, and that the condition (1-5) is equivalent to (1-4) under the assumption (1-2).[2] If we have (1-5) instead of (1-3), then we don't need (1-2) to get (1-6) for any $\alpha > 0$. More precisely, we have

Theorem 1.2. *Suppose that, for any $\nu_1 \in S_1,\ \nu_2 \in S_2$, the vectors $\nabla\varphi_i(\nu_2) - \nabla\varphi_i(\nu_1),\ i = 1, \dots, k$, are linearly independent and that (1-5) holds for $\nu_1 \in S_1,\ \nu_2 \in S_2,\ |t| = 1$ and for $\nu = \nu_1, \nu_2$. Then, for any $\alpha > 0$, there is a constant C_α such that*

$$\|E_1 f\ E_2 g\|_{L^{(d+3k)/(d+k)}(Q_R)} \le C_\alpha R^\alpha \|f\|_2\, \|g\|_2, \tag{1-6}$$

where Q_R is a cube of side length $R \gg 1$.

However, to obtain the global estimates $L^2 \times L^2 \to L^q$, for $q > (d+3k)/(d+k)$, we need to impose a decay condition on the Fourier transform of the surface measure, since it is needed to apply the epsilon removal lemma [Bourgain and Guth 2011]. Under the condition (1-2) such a decay estimate follows from the stationary phase method.

For $q \ge 2$, the estimate (1-1) is relatively easier to prove under the conditions (1-2), (1-3). The following may be thought of as a generalization of Theorem 2.3 in [Tao et al. 1998], which is concerned with elliptic hypersurfaces; see also Theorem 4.2 in [Moyua et al. 1999]. A generalization to general hypersurfaces had already been observed in [Lee 2006]. As a byproduct this gives estimates for the endpoint cases of (p, q) satisfying

$$\frac{1}{p} + \frac{d+3k}{d+k}\frac{1}{2q} = 1, \qquad q \ge 2.$$

Proposition 1.3. *Suppose the condition (1-4) holds for $\nu_1 \in S_1,\ \nu_2 \in S_2$ and $|t| = 1$. Then, for $q \ge 2$ and*

$$\frac{1}{p} + \frac{d+3k}{d+k}\frac{1}{2q} \le 1,$$

the estimate (1-1) holds.

[2]Indeed, if H, N and D are matrices of sizes $d \times d$, $(d-k)\times d$ and $k \times d$, respectively, such that $ND^t = 0$, $\det H \neq 0$ and $\operatorname{rank}(N^t\ D^t) = d$, then $\det(NHN^t) \neq 0$ if and only if $\det(DH^{-1}D^t) \neq 0$ because $\binom{NH}{D}(N^t\ D^t) = \begin{pmatrix} NHN^t & NHD^t \\ 0 & DD^t \end{pmatrix}$ and $\binom{N}{DH^{-1}}(N^t\ D^t) = \begin{pmatrix} NN^t & 0 \\ DH^{-1}N^t & DH^{-1}D^t \end{pmatrix}$.

Remark 1.4. In the proof of the above results we may assume that the aforementioned conditions hold uniformly, by breaking up the extension operator by decomposing S_1, S_2 into sufficiently small pieces. That is to say, there is a constant $c > 0$ such that for $\nu \in S_1 \cup S_2$ and $|t| = 1$,

$$\left|\det\left(\sum_{i=1}^{k} t_i H\varphi_i(\nu)\right)\right| \geq c \tag{1-7}$$

and, for $\nu_1, \nu_1' \in S_1$, $\nu_2, \nu_2' \in S_2$, $|t| \sim 1$ and for $\nu \in S_1 \cup S_2$,

$$\left|\det\left[\boldsymbol{D}(\nu_1, \nu_2)\left(\sum_{j=1}^{k} t_j H\varphi_j(\nu)\right)^{-1} \boldsymbol{D}^t(\nu_1', \nu_2')\right]\right| \geq c. \tag{1-8}$$

The same holds also for the conditions (1-4) and (1-5).

***Necessary conditions for* (1-1).** By modifying the examples in [Tao and Vargas 2000] with some specific surfaces we see that (1-1) cannot hold in general, unless

$$q \geq \frac{d+k}{d}, \tag{1-9}$$

$$\frac{1}{p} + \frac{d+3k}{d+k}\frac{1}{2q} \leq 1, \tag{1-10}$$

$$\frac{2(d-k)}{p} + \frac{d+3k}{q} \leq 2d. \tag{1-11}$$

In fact,

(i) (1-9) is necessary for (1-1) to hold under (1-2), and

(ii) so is (1-10) under the assumption that the matrix $\boldsymbol{D}(\nu_1, \nu_2)$ has rank k for $\nu_j \in S_j$, $j = 1, 2$.

However, in general, (1-11) is not necessarily required for (1-1), but as is well known there are various Φ satisfying (1-2) and (1-3) for which (1-1) fails if

$$\frac{2(d-k)}{p} + \frac{d+3k}{q} > 2d.$$

We show (i) and (ii) in the following paragraphs.

(i) By making use of the stationary phase method together with the condition (1-2) it is not difficult to see that, with suitable choice of x_0, there is a cube Q of side length $R \gg 1$ such that $|E_1(e^{-2\pi i x_0 \cdot \xi}\psi)| \sim |E_2\psi(x)| \sim R^{-d/2}$ on Q provided that supports of ψ_1, ψ_2 are small enough. We insert these into (1-1) to see $R^{-d/2}R^{-d/2}R^{(d+k)/q} \lesssim 1$, from which we get (1-9) by letting $R \to \infty$. (This can also be shown by making use of a wave packet decomposition, see Lemma 4.2, and randomization.)

(ii) For $j = 1, 2$, let Σ_j be the surface $\{(\xi, \Phi(\xi)) : \xi \in S_j\}$, and denote by $d\sigma_j$ the induced Lebesgue measure on Σ_j. To see (1-9) it is more convenient to consider $f \to \widehat{f\,d\sigma_j}$, instead of dealing with the operator E_j. Also, let ν_j be the center of cube S_j and let $\zeta_j = (\nu_j, \Phi(\nu_j)) \in \Sigma_j$, $j = 1, 2$. The normal space $\boldsymbol{N}_j$ to Σ_j at ζ_j is spanned by

$$\boldsymbol{n}_{j,i} = (-\nabla\varphi_i(\nu_j), e_i), \quad i = 1, 2, \ldots, k,$$

where $e_i \in \mathbb{R}^k$ is the usual unit vector with its i-th entry being equal to 1. Clearly, these vectors are linearly independent because $\boldsymbol{D}(\nu_1, \nu_2)$ has rank k. Let $\boldsymbol{p}_n$, $n = 1, \ldots, d-k$, be an orthonormal basis of the orthogonal complement of $\operatorname{span}\{\boldsymbol{n}_{j,i} : i = 1, 2, \ldots, k,\ j = 1, 2\}$. Let us set, for $j = 1, 2$,

$$\Lambda_j = \big\{\zeta \in \Sigma_j : |(\zeta - \zeta_j) \cdot \boldsymbol{n}_{3-j,i}| \le \delta,\ |(\zeta - \zeta_j) \cdot \boldsymbol{p}_n| \le \delta^{\frac{1}{2}},\ i = 1, \ldots, k,\ n = 1, \ldots, d-k\big\}.$$

Now, we set $f_j = \chi_{\Lambda_j}$, $j = 1, 2$. Then it is easy to see $|\widehat{f_j\, d\sigma_j}(x, t)| \gtrsim \delta^{(d+k)/2}$, $j = 1, 2$, provided that

$$|(x, t) \cdot \boldsymbol{n}_{\ell,i}| \le c\delta^{-1}, \quad |(x, t) \cdot \boldsymbol{p}_n| \le c\delta^{-\frac{1}{2}}, \quad i = 1, \ldots, k, \quad \ell = 1, 2, \quad n = 1, \ldots, d-k$$

with sufficiently small $c > 0$. (For example, see the proof Lemma 4.2.) Since (1-1) implies

$$\|\widehat{f_1\, d\sigma_1}\ \widehat{f_2\, d\sigma_2}\|_q \lesssim \|f_1\|_p\, \|f_2\|_p,$$

we get $\delta^{d+k-(d+3k)/(2q)} \le C\delta^{(d+k)/p}$ and (1-10) by letting $\delta \to 0$.

Restriction to complex surfaces. Using the above theorem we can obtain a bilinear restriction estimate for complex quadratic surfaces. To define the (Fourier) extension operator for a complex surface we first distinguish the dot product and the inner product for complex variables, and define an auxiliary product $\odot$. For $z, w \in \mathbb{C}^m$, we define $z \cdot w$, $\langle z, w\rangle$, $z \odot w$ by

$$z \cdot w = \sum_{j=1}^{m} z_j w_j, \quad \langle z, w\rangle = \sum_{j=1}^{m} z_j \bar{w}_j, \quad z \odot w = \operatorname{Re}\langle z, w\rangle.$$

Hence, if $z = x + iy$ and $w = u + iv$ for $x, y, u, v \in \mathbb{R}^m$, then $z \odot w = x \cdot u + y \cdot v$. If we identify $\mathbb{C}^m$ with $\mathbb{R}^{2m}$ in the usual way, then $z \odot w$ is just the inner product on $\mathbb{R}^{2m}$.

Let $n \ge 1$ be an integer and let D be a real symmetric invertible matrix. Then we define the complex quadratic surface $\gamma \subset \mathbb{C}^{n+1}$ by

$$\gamma(z) = \big(z, \tfrac{1}{2} z^t D z\big), \quad z \in \mathbb{C}^n. \tag{1-12}$$

Now we define the extension operator $E_\gamma f$ by

$$E_\gamma f(w) = \int_{\mathbb{C}^n} e^{2\pi i [w \odot \gamma(z)]} f(z)\, dz, \quad w \in \mathbb{C}^{n+1},$$

where we have written dz for $dx\, dy$, $z = x + iy$. The operator $E_\gamma f$ is an extension operator for surfaces of codimension 2 in $\mathbb{R}^{2n}$, which is given by $\big(x, y, \frac{1}{2}\operatorname{Re}(x+iy)^t D(x+iy), \frac{1}{2}\Im(x+iy)^t D(x+iy)\big)$, $x, y \in \mathbb{R}^n$. From Theorem 1.1 we can establish the following.

Corollary 1.5. *Let S_1, S_2 be closed cubes in $\mathbb{C}^n$. Suppose that, for any $z_1 \in S_1$ and $z_2 \in S_2$,*

$$|(z_2 - z_1)^t D(z_2 - z_1)| \ne 0. \tag{1-13}$$

Then, whenever f, g are supported on S_1, S_2, respectively, for

$$q > \frac{n+3}{n+1} \quad \text{and} \quad \frac{1}{p} + \frac{n+3}{n+1}\frac{1}{2q} < 1,$$

there is a constant C such that

$$\|E_\gamma f\, E_\gamma g\|_{L^q(\mathbb{C}^{n+1})} \le C\|f\|_{L^p(\mathbb{C}^n)}\|g\|_{L^p(\mathbb{C}^n)}.$$

This theorem can also be stated without using the complex number notation, but its use makes it easier to derive the linear estimates from the bilinear one. The condition (1-13) in $\mathbb{C}^2$ can be contrasted with that in $\mathbb{R}^2$. If $S_1, S_2 \subset \mathbb{R}^2$ and the eigenvalues of D have the same sign, then the condition (1-13) is always valid if $\operatorname{dist}(S_1, S_2) \neq 0$. But, when $S_1, S_2 \subset \mathbb{C}^2$, the condition (1-13) may fail even if the separation condition is satisfied. For instance, if D is the 2×2 identity matrix, the condition (1-13) becomes $|(v_1 - w_1)^2 + (v_2 - w_2)^2| \gtrsim 1$ with $z_1 = (v_1, v_2)$ and $z_2 = (w_1, w_2)$. Since we may factor $(v_1 - w_1)^2 + (v_2 - w_2)^2$ as $[(v_1 - w_1) + i(v_2 - w_2)][(v_1 - w_1) - i(v_2 - w_2)]$, the expression $|(v_1 - w_1)^2 + (v_2 - w_2)^2|$ may vanish even if $\operatorname{dist}(S_1, S_2) \gtrsim 1$. When D has eigenvalues with different signs, this phenomenon may occur even when $S_1, S_2 \subset \mathbb{R}^2$; for instance, if D is the 2×2 diagonal matrix with diagonal entries 1 and -1, then we have

$$x \cdot Dx = x_1^2 - x_2^2 = (x_1 + x_2)(x_1 - x_2).$$

This real-variable case was studied by Lee [2006] and Vargas [2005]. In the special case that the surface is two-dimensional they could deduce a linear estimate from the bilinear one.

By adapting their argument, we can obtain the following linear estimate.

Theorem 1.6. *Let $n = 2$ and γ be given by (1-12) with a nonsingular real symmetric matrix D. Then, for $q > \frac{10}{3}$ and $\frac{1}{p} + \frac{2}{q} < 1$,*

$$\|E_\gamma f\|_{L^q(\mathbb{C}^3)} \le C\|f\|_{L^p(\mathbb{C}^2)} \tag{1-14}$$

whenever f is supported in a bounded set.

By analogy with the corresponding problem for the paraboloid (elliptic or hyperbolic) in $\mathbb{R}^3$, it may be conjectured that (1-14) holds if and only if $q > 3$ and $\frac{1}{p} + \frac{2}{q} \le 1$. Theorem 1.6 extends the known (p, q) range for the operator $E_\gamma f$ when D is a nonsingular real symmetric matrix. This result is an analog of the adjoint Fourier restriction estimates for the hyperbolic paraboloid in $\mathbb{R}^3$, which is known to hold for the same range of p, q. As a special case of the results by Christ [1982, Lemma 4.3] and Mockenhaupt [1996, Theorem 2.11], it was previously known that Ef maps $L^2(\mathbb{R}^4)$ boundedly to $L^4(\mathbb{R}^6)$. Also, the slightly stronger Lorentz space estimate $\|Ef\|_{L^{4,2}(\mathbb{R}^6)} \le C\|f\|_{L^2(\mathbb{R}^4)}$ can be deduced by applying Theorem 1.1 in [Bak and Seeger 2011]. It is quite likely that the multilinear approach will yield further progress on these problems. We hope to return to this problem in the near future.

Notation. We adopt the usual convention to let C or c represent strictly positive constants, whose value may vary from line to line. But these constants will always be independent of f, for instance. We write $A \lesssim B$ or $B \gtrsim A$ to mean $A \le CB$, and $A \sim B$ means both $A \lesssim B$ and $B \lesssim A$.

2. $L^{\frac{4(d+k)}{3d+k}} \times L^{\frac{4(d+k)}{3d+k}} \to L^2$ estimates and proof of Proposition 1.3

In this section we show Proposition 1.3. Our proof here is different from that in [Tao et al. 1998]. Instead of making use of the boundedness of the averaging operator, we directly exploit the oscillatory decay estimate which is concealed in the averaging operator. For this we need the following lemma.

Lemma 2.1 [Greenleaf and Seeger 2002, Section 1.1]. *Let $a \in C_c^\infty(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^N)$ and set*

$$T_\lambda f(x) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^N} e^{i\lambda\phi(x,y,\theta)} a(x,y,\theta)\, d\theta\, f(y)\, dy,$$

where ϕ is a smooth function on the support of a. Suppose

$$\det \begin{pmatrix} \phi''_{\theta\theta} & \phi''_{x\theta} \\ \phi''_{y\theta} & \phi''_{xy} \end{pmatrix} \neq 0$$

on the support of a whenever $\phi'_\theta = 0$. Then, $\|T_\lambda f\|_2 \lesssim \lambda^{-(d+N)/2} \|f\|_2$.

Proof of Proposition 1.3. By interpolation with the trivial $L^1 \times L^1 \to L^\infty$ estimate, it suffices to show

$$\|E_1 f_1\, E_2 f_2\|_2 \lesssim \|f_1\|_{\frac{4(d+k)}{3d+k}} \|f_2\|_{\frac{4(d+k)}{3d+k}}.$$

For fixed ξ_2, set

$$\Phi^{\xi_2}(\xi_1, \eta_1) = \Phi(\xi_1) + \Phi(\xi_2) - \Phi(\eta_1) - \Phi(\xi_1 + \xi_2 - \eta_1)$$

and

$$I^{\xi_2}(f_1, \bar{f}_1) = \iint \delta(\Phi^{\xi_2}(\xi_1, \eta_1))\, f_1(\xi_1)\, \bar{f}_1(\eta_1)\, d\xi_1\, d\eta_1,$$

where δ is the delta function. Its composition is well defined, since the vectors $\nabla\varphi_i(\nu_2) - \nabla\varphi_i(\nu_1)$, $i = 1, \dots, k$, are linearly independent.

By Plancherel's theorem

$$\begin{aligned}\|E_1 f_1\, E_2 f_2\|_2^2 &= \iiiint \delta\big(\xi_1 + \xi_2 - \eta_1 - \eta_2, \Phi(\xi_1) + \Phi(\xi_2) - \Phi(\eta_1) - \Phi(\eta_2)\big) \\ &\qquad\qquad \times f_1(\xi_1)\, f_2(\xi_2)\, \bar{f}_1(\eta_1)\, \bar{f}_2(\eta_2)\, d\xi_1\, d\xi_2\, d\eta_1\, d\eta_2 \\ &= \iiint \delta(\Phi^{\xi_2}(\xi_1, \eta_1))\, f_1(\xi_1)\, \bar{f}_1(\eta_1)\, f_2(\xi_2)\, \bar{f}_2(\xi_1 + \xi_2 - \eta_1)\, d\xi_1\, d\xi_2\, d\eta_1,\end{aligned}$$

where f_1, f_2 are assumed to be supported in S_1, S_2, respectively. We claim that

$$\|E_1 f_1\, E_2 f_2\|_2^2 \lesssim \|f_1\|_{p,1}\, \|f_2\|_1\, \|\bar{f}_1\|_{p,1}\, \|\bar{f}_2\|_\infty, \tag{2-1}$$

where $p = (d+k)/d$. Here $\|f\|_{r,s}$ denotes the norm of Lorentz space $L^{r,s}$. For this we may obviously assume that the functions $f_1, \bar{f}_1, f_2, \bar{f}_2$ are nonnegative. In order to show (2-1) it suffices to prove

$$|I^{\xi_2}(f, g)| \lesssim \|f\|_{p,1}\, \|\bar{g}\|_{p,1}. \tag{2-2}$$

Let ψ be a smooth function with compact Fourier support contained in $B(0,1)$ such that $\hat{\psi} = 1$ on $B\big(0, \frac{1}{2}\big)$. Since $h(0) = \lim_{j\to\infty} 2^{jk} \int_{\mathbb{R}^k} \psi(2^j x)\, h(x)\, dx$ for any Schwartz function h, we have

$\delta = \lim_{j\to\infty} 2^{jk}\psi(2^j x)$. So, we may write

$$\delta = \sum_{j=-\infty}^{\infty} [2^{(j+1)k}\psi(2^{j+1}x) - 2^{jk}\psi(2^j x)] = \sum_{j=-\infty}^{\infty} 2^{jk}\eta(2^j x),$$

where $\eta(x) := 2^k\psi(2x) - \psi(x)$. By the choice of ψ we see that the Fourier support of η is contained in $\{\xi : \frac{1}{2} < |\xi| \le 2\}$. We decompose $I^{\xi_2}(f, g)$ by making use of the above decomposition of δ to get

$$I^{\xi_2}(f, g) = \sum_{j=-\infty}^{\infty} I_j(f, g),$$

where

$$I_j(f, g) := 2^{kj} \iint \eta(2^j \Phi^{\xi_2}(\xi_1, \eta_1))\, f(\xi_1)\, g(\eta_1)\, d\xi_1\, d\eta_1.$$

It should be noted that we are assuming that f, g are supported on S_1 and $\xi_1 + \xi_2 - \eta_1 \in S_2$. Using the Fourier transform we write $I_j(f_1, \bar{f}_2)$ as

$$I_j(f, g) = 2^{kj} \int \left(\iint \hat{\eta}(\tau)\, e^{2^j \tau\cdot\Phi^{\xi_2}(\xi_1,\eta_1)}\, d\tau\ f(\xi_1)\, d\xi_1 \right) g(\eta_1)\, d\eta_1.$$

Now, we will apply Lemma 2.1 to the double integral inside the parentheses. If we set $\phi(\xi_1, \eta_1, \tau) = \tau \cdot \Phi^{\xi_2}(\xi_1, \eta_1)$, then

$$\left|\det\begin{pmatrix} \phi''_{\tau\tau} & \phi''_{\tau\xi_1} \\ \phi''_{\eta_1\tau} & \phi''_{\xi_1\eta_1} \end{pmatrix}\right| = \left|\det\begin{pmatrix} 0 & \boldsymbol{D}(\xi_1, \xi_1+\xi_2-\eta_1) \\ \boldsymbol{D}(\eta_1, \xi_1+\xi_2-\eta_1)^t & \sum_{j=1}^{k} \tau_j H\varphi_j(\xi_1, \xi_1+\xi_2-\eta_1) \end{pmatrix}\right|.$$

So, by the condition (1-4) the last expression does not vanish since $|\tau| \sim 1$. Hence, by Lemma 2.1 it follows that

$$|I_j(f, g)| \lesssim 2^{-j\frac{d-k}{2}} \|f\|_2\, \|g\|_2.$$

On the other hand, we have the trivial bound $|I_j(f, g)| \lesssim 2^{kj}\|f\|_1\, \|g\|_1$. Now we may use a summation method (usually called Bourgain's summation trick) to obtain (2-2).

Considering $(f_1, \bar{f}_1, f_2, \bar{f}_2) \to \|Ef_1\, Ef_2\|_2^2$ as a quadrilinear mapping (replacing $\bar{f}_1$, $\bar{f}_2$ on the left-hand side by $\bar{f}_3$ and $\bar{f}_4$, respectively), we apply Christ's multilinear trick [1985]. By symmetry and interpolation we get the estimates

$$\left| \iint Ef_1\, Ef_2\, \overline{Ef_3\, Ef_4}\, dx\, dt \right| \lesssim \prod_{j=1}^{4} \|f_j\|_{p_j,1}$$

for $\left(\frac{1}{p_1}, \frac{1}{p_2}, \frac{1}{p_3}, \frac{1}{p_4}\right)$ contained in the convex hull of the four points

$$v_1 = \left(\frac{1}{p}, \frac{1}{p}, 1, 0\right), \quad v_2 = \left(\frac{1}{p}, \frac{1}{p}, 0, 1\right), \quad v_3 = \left(1, 0, \frac{1}{p}, \frac{1}{p}\right), \quad v_4 = \left(0, 1, \frac{1}{p}, \frac{1}{p}\right)$$

which is contained in the 3-plane $\Pi = \{u_1 + u_2 + u_3 + u_4 = 1 + \frac{2}{p}\}$. The convex hull has a nonempty interior in Π, because $\det(v_1, v_2, v_3, v_4) \ne 0$ as long as $\frac{1}{p} \ne \frac{1}{2}$. Hence we may apply the multilinear trick to get

$$\|Ef_1\, Ef_2\|_2^2 \lesssim \|f_1\|_{\frac{4(d+k)}{3d+k},4} \|f_2\|_{\frac{4(d+k)}{3d+k},4} \|\bar{f}_1\|_{\frac{4(d+k)}{3d+k},4} \|\bar{f}_2\|_{\frac{4(d+k)}{3d+k},4}. \qquad \square$$

3. Transversality and the curvature conditions

In this section we prove several lemmas that will play crucial roles in proving Theorem 1.1. These lemmas are related to the curvature conditions.

For $R \gg 1$ and $\nu \in S_1 \cup S_2$, we set

$$\pi_\nu = \left\{(x,t) : \left| x + \left(\sum_{j=1}^{k} t_j \nabla\varphi_j(\nu)\right)\right| \le R^{\frac{1}{2}}\right\}, \quad R^\delta \pi_\nu = \pi_\nu + O(R^{\frac{1}{2}+\delta}).$$

Here, for any set $A \subset \mathbb{R}^{d+k}$ and $\rho > 0$, we have $A + O(\rho) = \{u \in R^{d+k} : \mathrm{dist}(u, A) \le C\rho\}$.

Lemma 3.1. *Suppose that the vectors $\nabla\varphi_j(\nu_2) - \nabla\varphi_j(\nu_1)$, $1 \le j \le k$, are linearly independent for all $\nu_1 \in S_1$ and $\nu_2 \in S_2$. Then, there is a constant C such that*

$$\pi_{\nu_1} \cap \pi_{\nu_2} \subset B(0, CR^{\frac{1}{2}}).$$

Proof. Since the set $\{\nabla\varphi_j(\nu_2) - \nabla\varphi_j(\nu_1)\}_{j=1}^k$ is linearly independent for all $\nu_1 \in S_1$ and $\nu_2 \in S_2$, the map $(t_1, \dots, t_k) \to (t_1, \dots, t_k)^t \boldsymbol{D}(\nu_1, \nu_2)$ is injective. So, by continuity and compactness it follows that there is a constant C such that, for all $\nu_1 \in S_1$ and $\nu_2 \in S_2$,

$$|(t_1, \dots, t_k)^t \boldsymbol{D}(\nu_1, \nu_2)| \ge C|(t_1, \dots, t_k)|.$$

If $(x,t) \in \pi_{\nu_1} \cap \pi_{\nu_2}$, then $\left|x + \left(\sum_{j=1}^k t_j \nabla\varphi_j(\nu_i)\right)\right| \le R^{\frac{1}{2}}$ for $i = 1, 2$. This gives

$$|(t_1, \dots, t_k)^t \boldsymbol{D}(\nu_1, \nu_2)| \le 2R^{\frac{1}{2}}.$$

Hence, the above inequality yields $|(t_1, \dots, t_k)| \le CR^{\frac{1}{2}}$. So, we also get $|x| \le CR^{\frac{1}{2}}$. □

As was already shown in [Lee 2006; Vargas 2005], a simple transversality condition between the two wave packets is not enough to obtain a bilinear estimate beyond the range of the linear $L^2 \to L^q$ estimate. So, we need to consider the Fourier supports of the wave packets to put a restriction on the permissible wave packets. This makes the geometry of the associated wave packets more favorable.

For given $\nu_1 \in S_1$ and $\nu_2' \in S_2$ we define $\Pi_1^{\nu_1,\nu_2'}$ by

$$\Pi_1^{\nu_1,\nu_2'} = \{\nu_1' \in S_1 : \nu_1' + \nu_2' - \nu_1 \in S_2,\ \Phi(\nu_1) + \Phi(\nu_1' + \nu_2' - \nu_1) = \Phi(\nu_1') + \Phi(\nu_2')\}. \tag{3-1}$$

Since $\{\nabla\varphi_j(\nu_2) - \nabla\varphi_j(\nu_1)\}_{j=1}^k$ are linearly independent, by the implicit function theorem we may assume that $\Pi_1^{\nu_1,\nu_2'}$ is a smooth $(d-k)$-dimensional surface.[3] We now set

$$\Gamma_1^{\nu_1,\nu_2'}(R) = \bigcup_{\nu_1' \in \Pi_1^{\nu_1,\nu_2'}} R^\delta \pi_{\nu_1'},$$

which is an $O(R^{\frac{1}{2}+\delta})$ neighborhood of the conical set with k null directions. The transversality between $\Gamma_1^{\nu_1,\nu_2'}$ and the opposite plates π_{ν_2} is important. Such a transversality is made precise in the following (see Figure 1).

[3] We may need to assume that S_1 and S_2 are small enough.

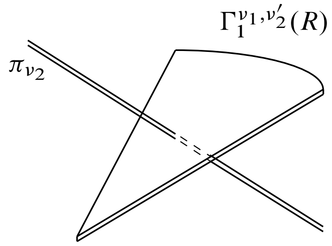

Figure 1. Transversality when $k = 1$ and $d = 2$.

Lemma 3.2. *Let* $0 < \delta \ll 1,\ u \in \mathbb{R}^{d+k}$ *and set*

$$\widetilde{\Gamma}_1^{\nu_1,\nu_2'}(R, R^\delta) = \{(x,t) \in \Gamma_1^{\nu_1,\nu_2'}(R) : R^{1-\delta} \le |(x,t)| \le CR\}.$$

Suppose that the conditions (1-2) *and* (1-3) *hold. Then, if* S_1 *and* S_2 *are sufficiently small, there exist a constant* C, *independent of* ν_1, ν_2', R, *and a vector* $u \in \mathbb{R}^{d+k}$ *such that for some* $u' \in \mathbb{R}^{d+k}$,

$$\widetilde{\Gamma}_1^{\nu_1,\nu_2'}(R, R^\delta) \cap (R^\delta \pi_{\nu_2} + u) \subset B(u', CR^{\frac{1}{2}+C\delta}).$$

Note that the set $\widetilde{\Gamma}_1^{\nu_1,\nu_2'}(R, R^\delta)$ can be represented as an $O(R^{\frac{1}{2}+\delta})$ neighborhood of a surface. Let us define the map $\Phi_1^{\nu_1,\nu_2'} : \Pi_1^{\nu_1,\nu_2'} \times \mathbb{R}^k \to \mathbb{R}^{d+k}$ by

$$\Phi_1^{\nu_1,\nu_2'}(\nu, t) = \Bigl(-\sum_{j=1}^{k} t_j \nabla\varphi_j(\nu), t\Bigr).$$

Then it is easy to see that

$$\widetilde{\Gamma}_1^{\nu_1,\nu_2'}(R, R^\delta) \subset \{\Phi_1^{\nu_1,\nu_2'}(\nu, t) : \nu \in \Pi_1^{\nu_1,\nu_2'},\ cR^{1-\delta} \le |t| \le CR\} + O(R^{\frac{1}{2}+\delta}).$$

Proof. After scaling it is sufficient to show that the intersection of the two sets

$$\Gamma_1 = \{\Phi_1^{\nu_1,\nu_2'}(\nu, t) : \nu \in \Pi_1^{\nu_1,\nu_2'},\ R^{-\delta} \le |t| \le C\} + O(R^{-\frac{1}{2}+\delta})$$

and

$$\mathfrak{C}_2(R^{-\frac{1}{2}+\delta}) = \left\{\Bigl(-\sum_{j=1}^{k} t_j \nabla\varphi_j(\nu_2), t\Bigr) : |t| \le C\right\} + \tilde{u} + O(R^{-\frac{1}{2}+\delta})$$

is contained in a ball of radius $CR^{-\frac{1}{2}+C\delta}$. For $j \ge -C$, let us set

$$\Gamma_1^j(R^{-\frac{1}{2}+\delta}) = \{\Phi_1^{\nu_1,\nu_2'}(\nu, t) : \nu \in \Pi_1^{\nu_1,\nu_2'},\ 2^{-j-1} \le |t| \le 2^{-j}\} + O(R^{-\frac{1}{2}+\delta}).$$

Using homogeneity and a dyadic decomposition in t for Γ_1, the matter can be reduced to the case $2^{-1} \le |t| \le 1$. That is to say,

$$\Gamma_1^0(R^{-\frac{1}{2}+\delta}) \cap \mathfrak{C}_2(R^{-\frac{1}{2}+\delta}) \subset B(u, C_0 R^{-\frac{1}{2}+\delta}) \tag{3-2}$$

for some u and $C_0 > 0$. In fact, applying the scaling change of variables $(x,t) \to 2^{-j}(x,t)$, followed by (3-2) and the reverse change of variables, we see that $\Gamma_1^j(R^{-\frac{1}{2}+\delta}) \cap \mathfrak{C}_2(R^{-\frac{1}{2}+\delta})$ is contained in a ball of radius $C_0 R^{-\frac{1}{2}+\delta}$. Since $\Gamma_1 \subset \bigcup_{2^{-1}R^{-\delta} \le 2^j \le C} \Gamma_1^j$, we know $\Gamma_1 \cap \mathfrak{C}_2(R^{-\frac{1}{2}+\delta})$ is contained in the union of as many as $\sim \log R$ such balls of radius $C_0 R^{-\frac{1}{2}+\delta}$. This union of balls is obviously contained in a ball of radius $CR^{-\frac{1}{2}+C\delta}$ since the set $\Gamma_1 \cap \mathfrak{C}_2(R^{-\frac{1}{2}+\delta})$ is connected.

Since we may assume that S_1 and S_2 are sufficiently small, in order to show (3-2) it is enough to show that the tangent spaces of the surfaces $\Phi_1^{\nu_1,\nu_2'} : \Pi_1^{\nu_1,\nu_2'} \times \{2^{-1} \le |t| \le 1\} \to \mathbb{R}^{d+k}$ and $\{(\sum_{j=1}^k t_j \nabla\varphi_j(\nu_2), t) : |t| \le C\}$ are uniformly transversal to each other. In fact, since all the underlying sets are compact, by continuity it is enough to check this at each point.

Let $u_0 = \Phi_1^{\nu_1,\nu_2'}(\nu_0, t_0)$ for $\nu_0 \in \Pi_1^{\nu_1,\nu_2'}$ and $2^{-1} \le |t_0| \le 1$. Let $\boldsymbol{v}_1, \dots, \boldsymbol{v}_{d-k}$ be orthonormal vectors spanning the tangent space $T_{\nu_0}\Pi_1^{\nu_1,\nu_2'}$. Then the tangent space of the parametrized surface $\Phi_1^{\nu_1,\nu_2'} : \Pi_1^{\nu_1,\nu_2'} \times \{2^{-1} \le |t| \le 1\} \to \mathbb{R}^{d+k}$ at u_0 is spanned by the vectors

$$(\nabla\varphi_1(\nu_0), -1, 0, \dots, 0), \quad (\nabla\varphi_2(\nu_0), 0, -1, 0, \dots, 0), \quad \dots, \quad (\nabla\varphi_k(\nu_0), 0, \dots, 0, -1) \tag{3-3}$$

and

$$\left(\boldsymbol{v}_i\left(\sum_{j=1}^k t_{0,j} H\varphi_j(\nu_0)\right), 0, \dots, 0\right), \quad i = 1, \dots, d-k. \tag{3-4}$$

On the other hand, the k-dimensional plane $\{(-\sum_{j=1}^k t_j \nabla\varphi_j(\nu_2), t) : |t| \le C\}$ is spanned by

$$(\nabla\varphi_1(\nu_2), -1, 0, \dots, 0), \quad (\nabla\varphi_2(\nu_2), 0, -1, 0, \dots, 0), \quad \dots, \quad (\nabla\varphi_k(\nu_2), 0, \dots, 0, -1). \tag{3-5}$$

Hence it suffices to show that these $d+k$ vectors are linearly independent, or equivalently that the determinant of the matrix with these vectors as row vectors is nonzero. After Gaussian elimination it is enough to show

$$\det\begin{pmatrix} \boldsymbol{V}(\sum_{j=1}^k t_{0,j} H\varphi_j(\nu_0)) \\ \boldsymbol{D}(\nu_0, \nu_2) \end{pmatrix} \ne 0, \tag{3-6}$$

where $\boldsymbol{V}$ is the $(d-k) \times d$ matrix having $\boldsymbol{v}_1, \dots, \boldsymbol{v}_{d-k}$ as its row vectors. Now by (3-1) we note that the vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_{d-k}$ are orthogonal to the span of the vectors

$$\nabla\varphi_j(\nu_0 + \nu_2' - \nu_1) - \nabla\varphi_j(\nu_0), \quad j = 1, \dots, k.$$

Assuming S_2 is small enough, we may replace $\boldsymbol{D}(\nu_0, \nu_2)$ by $\boldsymbol{D}(\nu_0, \nu_0+\nu_2'-\nu_1)$. For simplicity we set $\tilde{\nu}_2 = \nu_0+\nu_2'-\nu_1$. (We may assume there is a $c > 0$ such that $\left|\det\left[\boldsymbol{N}(\nu_2, \nu_1)\left(\sum_{i=1}^k t_i H\varphi_i(\nu)\right)\boldsymbol{N}^t(\nu_2, \nu_1)\right]\right| > c$ for $\nu_1 \in S_2$ and $\nu_2 \in S_2$; see Remark 1.4.) Since $(\sum_{j=1}^k t_{0,j} H\varphi_j(\nu_0))$ is invertible, we need only show

$$\det \boldsymbol{A} \ne 0, \quad \text{where } \boldsymbol{A} = \begin{pmatrix} \boldsymbol{V} \\ \boldsymbol{D}(\nu_0, \tilde{\nu}_2)\left(\sum_{j=1}^k t_{0,j} H\varphi_j(\nu_0)\right)^{-1} \end{pmatrix}.$$

Since $\boldsymbol{V}\boldsymbol{D}^t(\nu_0, \tilde{\nu}_2) = 0$, we note that the matrix $\boldsymbol{A}\left(\boldsymbol{V}^t\ \boldsymbol{D}^t(\nu_0, \tilde{\nu}_2)\right)$ equals

$$\begin{pmatrix} I_{d-k} & 0 \\ \boldsymbol{D}(\nu_0, \tilde{\nu}_2)\left(\sum_{j=1}^k t_{0,j} H\varphi_j(\nu_0)\right)^{-1} V^t & \boldsymbol{D}(\nu_0, \tilde{\nu}_2)\left(\sum_{j=1}^k t_{0,j} H\varphi_j(\nu_0)\right)^{-1} \boldsymbol{D}^t(\nu_0, \tilde{\nu}_2) \end{pmatrix}.$$

This matrix is clearly invertible thanks to (1-3). Hence, so is the matrix $\boldsymbol{A}$. □

Below we show that the following version of Lemma 3.2 holds, where we assume (1-5) instead of (1-3), dropping the condition (1-2).

Lemma 3.3. *Suppose that, for any $\nu_1 \in S_1$, $\nu_2 \in S_2$, $\nabla\varphi_i(\nu_2) - \nabla\varphi_i(\nu_1)$, $i = 1, \dots, k$ are linearly independent and (1-5) holds for $\nu_1 \in S_1$, $\nu_2 \in S_2$, $|t| = 1$ and for $\nu = \nu_1, \nu_2$. If S_1 and S_2 are sufficiently small, there is a constant C, independent of ν_1, ν_2', R, and u such that, for some $u' \in \mathbb{R}^{d+1}$,*

$$\widetilde{\Gamma}_1^{\nu_1,\nu_2'}(R, R^\delta) \cap (R^\delta \pi_{\nu_2} + u) \subset B(u', CR^{\frac{1}{2}+C\delta}).$$

Proof. It is sufficient to show that (3-6) holds. As before, under the assumption that S_2 is small enough, we can replace $\boldsymbol{D}(\nu_0, \nu_2)$ with $\boldsymbol{D}(\nu_0, \tilde{\nu}_2)$, where $\tilde{\nu}_2 = \nu_0 + \nu_2' - \nu_1$. We need only show that

$$\det \begin{pmatrix} \boldsymbol{V}\left(\sum_{j=1}^k t_{0,j} H\varphi_j(\nu_0)\right) \\ \boldsymbol{D}(\nu_0, \tilde{\nu}_2) \end{pmatrix} \neq 0.$$

Since vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_{d-k}$ are orthogonal to the row vectors of $\boldsymbol{D}(\nu_0, \tilde{\nu}_2)$, by multiplying the nonsingular matrix $(\boldsymbol{V}^t\ \boldsymbol{D}^t(\nu_0, \tilde{\nu}_2))$ by the matrix inside the determinant from the right, we see that the above is equivalent to

$$\det \begin{pmatrix} \boldsymbol{V}\left(\sum_{j=1}^k t_{0,j} H\varphi_j(\nu_0)\right)\boldsymbol{V}^t & \boldsymbol{V}\left(\sum_{j=1}^k t_{0,j} H\varphi_j(\nu_0)\right)\boldsymbol{D}^t(\nu_0, \tilde{\nu}_2) \\ 0 & \boldsymbol{D}(\nu_0, \tilde{\nu}_2)\boldsymbol{D}^t(\nu_0, \tilde{\nu}_2) \end{pmatrix} \neq 0.$$

Since the matrix $\boldsymbol{D}(\nu_0, \tilde{\nu}_2)\, \boldsymbol{D}^t(\nu_0, \tilde{\nu}_2)$ is nonsingular, it is clear that the above is equivalent to

$$\det\left[\boldsymbol{V}\left(\sum_{j=1}^k t_{0,j} H\varphi_j(\nu_0)\right)\boldsymbol{V}^t\right] \neq 0,$$

which is the condition (1-5). □

4. Proof of Theorem 1.1

In this section we will prove Theorem 1.1. Our proof is similar to that in [Lee 2006]; also see [Tao 2003]. To prove Theorem 1.1, we need only show that, for $p > (d+3k)/(d+k)$,

$$\|E_1 f\ E_2 g\|_p \le C\|f\|_2\,\|g\|_2$$

since we can obtain the desired conclusion by interpolating this estimate with the trivial estimate $\|E_1 f\ E_2 g\|_\infty \le \|f\|_1\,\|g\|_1$. By an ϵ-removal argument [Tao and Vargas 2000; Bourgain and Guth 2011], it is sufficient to show that (1-6) holds for any $\alpha > 0$. In fact, by the assumption that $\sum_{j=1}^k t_j H\varphi_j(\nu)$ is nonsingular for $\nu \in \operatorname{supp} f \cup \operatorname{supp} g$ as long as $|t| = 1$, it follows that

$$|E_\kappa(a_\kappa)(x,t)| \lesssim (|x| + |t|)^{-\frac{d}{2}}, \quad \kappa = 1, 2,$$

where a_1, a_2 are smooth bump functions which vanish on the supports of f and g, respectively. This can be shown by the stationary phase method. Hence, the arguments in the papers mentioned above work here without modification.

Proposition 4.1. *Let $0<\delta\ll 1$. If (1-6) holds, then for any $\epsilon>0$*

$$\|E_1 f\, E_2 g\|_{L^{(d+3k)/(d+k)}(Q_R)} \le C_\epsilon R^{\max(\alpha(1-\delta),C\delta)+\epsilon}\|f\|_2\,\|g\|_2, \tag{4-1}$$

with C independent of δ.

By iterating finitely many times the implication in Proposition 4.1, we can easily obtain the estimate (1-6) for any $\alpha>0$.

Wave packet decomposition. In this section we decompose the function Ef into wave packets. Let $R\gg 1$. We define

$$\mathcal{L}=\mathcal{L}(R):=R^{\frac12}\mathbb{Z}^d,\quad \mathcal{V}=\mathcal{V}(R):=R^{-\frac12}\mathbb{Z}^d.$$

Let ψ be a nonnegative Schwartz function such that $\hat\psi$ is supported on $B(0,1)$ and $\sum_{k\in\mathbb{Z}^d}\psi(\cdot-k)=1$. Also, let ζ be a smooth function supported on $B(0,1)$ and $\sum_{k\in\mathbb{Z}^d}\zeta(\cdot-k)=1$.

For $\ell\in\mathcal{L}$, $\nu\in\mathcal{V}$ we set $\psi_\ell(x):=\psi((x-\ell)/R^{\frac12})$, $\zeta_\nu(\xi)=:\zeta(R^{\frac12}(\xi-\nu))$, and for a given function f, we define $f_{\ell,\nu}$ by

$$f_{\ell,\nu}=\mathcal{F}(\psi_\ell\mathcal{F}^{-1}(\zeta_\nu f)),$$

where $\mathcal{F}$, $\mathcal{F}^{-1}$ denote the Fourier transform and the inverse Fourier transform, respectively. Then, it follows that $f=\sum_{\nu\in\mathcal{V}}\sum_{\ell\in\mathcal{L}}f_{\ell,\nu}$. Hence we may write

$$Ef=\sum_{\nu\in\mathcal{V}}\sum_{\ell\in\mathcal{L}}Ef_{\ell,\nu}. \tag{4-2}$$

Lemma 4.2. *If $|t|\lesssim R$, then*

$$|Ef_{\ell,\nu}(x,t)|\le C_N\Big(1+R^{-\frac12}\Big|x-\ell+\sum_{j=1}^k t_j\nabla\varphi_j(\nu)\Big|\Big)^{-N}M(\mathcal{F}^{-1}(\zeta_\nu f))(\ell) \tag{4-3}$$

for all $N\ge 0$. Here, Mf is the Hardy–Littlewood maximal function of f.

Proof. Since $f_{\ell,\nu}$ is supported in $B(\nu,3R^{-\frac12})$, multiplying by a harmless smooth bump function $\tilde\chi$ supported in $B(0,5)$ and satisfying $\tilde\chi=1$ on $B(0,3)$, we may write

$$Ef_{\ell,\nu}(x,t)=\int K(x-z,t)\psi_\ell(z)\mathcal{F}^{-1}f_\nu(z)\,dz,$$

where $K(x,t)=\int e^{2\pi i(x\cdot\xi+t\cdot\Phi(\xi))}\chi(R^{\frac12}(\xi-\nu))\,d\xi$. Changing variables $\xi\to R^{-\frac12}\xi+\nu$,

$$K(x,t)=R^{-\frac{d}{2}}e^{2\pi i x\cdot\nu}\int e^{2\pi i(R^{-1/2}x\cdot\xi+t\cdot\Phi(R^{-1/2}\xi+\nu))}\chi(\xi)\,d\xi.$$

Since $|t|\lesssim R$, we know $\nabla_\xi(R^{-\frac12}x\cdot\xi+t\cdot\Phi(R^{-\frac12}\xi+\nu))=R^{-\frac12}\big(x+\sum_{j=1}^k t_j\nabla\varphi_j(\nu)\big)+O(1)$. This follows by Taylor's expansion. Hence, by repeated integration by parts we get

$$|K(x,t)|\le C_N R^{-\frac{d}{2}}\Big(1+R^{-\frac12}\Big|x+\sum_{j=1}^k t_j\nabla\varphi_j(\nu)\Big|\Big)^{-N}.$$

Once this is established, (4-3) follows by a standard argument. See [Lee 2006] for the details. ☐

From the above lemma we see that $Ef_{\ell,\nu}$ is essentially supported on

$$\pi_{\ell,\nu} = \pi_\nu + (\ell, 0). \tag{4-4}$$

If $\pi = \pi_{\ell,\nu}$, we define $\nu(\pi) = \nu$, which may be considered as the (generalized) direction of π.

The following is the main lemma of this section.

Lemma 4.3. *Let $R \gg 1$. Then, Ef can be rewritten as*

$$Ef(x,t) = \sum_{(\ell,\nu)\in\mathcal{L}\times\mathcal{V}} c_{\ell,\nu} P_{\ell,\nu}(x,t) \tag{4-5}$$

and $c_{\ell,\nu}$, $P_{\ell,\nu}$ satisfy the following:

(i) *$\mathcal{F}(P_{\ell,\nu}(\cdot,t))$ is supported in the disc $D(\nu, CR^{-\frac{1}{2}})$.*

(ii) *If $|t| \lesssim R$, then for any $N \geq 0$*

$$|P_{\ell,\nu}(x,t)| \leq C_N R^{-\frac{d}{4}} \left(1 + R^{-\frac{1}{2}} \left| x - \ell + \left(\sum_{j=1}^{k} t_j \nabla\varphi_j(\nu)\right)\right|\right)^{-N}.$$

(iii) $\big(\sum_{(\ell,\nu)\in\mathcal{L}\times\mathcal{V}} |c_{\ell,\nu}|^2\big)^{\frac{1}{2}} \lesssim \|f\|_2$.

(iv) *If $|t| \lesssim R$, then $\big\|\sum_{(\ell,\nu)\in\mathcal{W}} P_{\ell,\nu}(\cdot,t)\big\|_2^2 \lesssim \#\mathcal{W}$ for any $\mathcal{W} \subset \mathcal{L}\times\mathcal{V}$.*

Proof. We define $c_{\ell,\nu}$ and $P_{\ell,\nu}$ by

$$c_{\ell,\nu} = R^{\frac{d}{4}} M(\mathcal{F}^{-1} f_\nu)(\ell), \quad P_{\ell,\nu}(x,t) = c_{\ell,\nu}^{-1} E f_{\ell,\nu}(x,t),$$

where M denotes the Hardy–Littlewood maximal function. Then we have (4-5) from (4-2). Since $Ef_{\ell,\nu}(\cdot,y) = \mathcal{F}^{-1}(e^{2\pi i y\Phi} f_{\ell,\nu})$, we know $Ef_{\ell,\nu}(\cdot,y)$ has Fourier support contained in $\operatorname{supp} f_{\ell,\nu}$, which is in turn contained in $D(\nu, CR^{-\frac{1}{2}})$. Thus (i) follows and so does (ii) from Lemma 4.2.

In order to show (iii), note that

$$\sum_{(\ell,\nu)\in\mathcal{L}\times\mathcal{V}} |c_T|^2 = R^{\frac{d}{2}} \sum_{(\ell,\nu)\in\mathcal{L}\times\mathcal{V}} M(\mathcal{F}^{-1}(\zeta_\nu f))(\ell)^2. \tag{4-6}$$

Since $\zeta_\nu f$ is supported on $B(\nu, CR^{\frac{1}{2}})$, we have $M(\mathcal{F}^{-1}(\zeta_\nu f))(x) \sim M(\mathcal{F}^{-1}(\zeta_\nu f))(x')$ if $|x-x'| \lesssim R^{\frac{1}{2}}$. Hence, from the Hardy–Littlewood maximal theorem and Plancherel's theorem we have that, for each ν,

$$R^{\frac{d}{2}} \sum_{(\ell,\nu)\in\mathcal{L}\times\mathcal{V}} |M(\mathcal{F}^{-1}(\zeta_\nu f))(\ell)|^2 \lesssim \int |M(\mathcal{F}^{-1}(\zeta_\nu f))(x)|^2\,dx \lesssim \|\zeta_\nu f\|_2^2.$$

Combining this and (4-6) we obtain $\sum_{(\ell,\nu)\in\mathcal{L}\times\mathcal{V}} |c_{\ell,\nu}|^2 \lesssim \sum_{\nu\in\mathcal{V}} \|\zeta_\nu f\|_2^2 \lesssim \|f\|_2^2$, and (iii).

Finally, we consider (iv). Since $\sum_{\ell:(\ell,\nu)\in\mathcal{W}} P_{\ell,\nu}(\cdot,t)$ is Fourier-supported in $D(\nu, CR^{-\frac{1}{2}})$, which has bounded overlap as ν varies over $\mathcal{V}$, by Plancherel's theorem,

$$\Big\|\sum_{(\ell,\nu)\in\mathcal{W}} P_{\ell,\nu}(\cdot,t)\Big\|_2^2 \lesssim \sum_{\nu\in\mathcal{V}} \Big\|\sum_{\ell:(\ell,\nu)\in\mathcal{W}} P_{\ell,\nu}(\cdot,t)\Big\|_2^2.$$

From (ii) it is easy to see that $\big\|\sum_{\ell:(\ell,\nu)\in\mathcal{W}} P_{\ell,\nu}(\cdot\,,t)\big\|_2^2 \lesssim \#\{\ell:(\ell,\nu)\in\mathcal{W}\}$. Hence, combining this with the above gives (iv). □

Dyadic pigeonholing and reduction. In the remainder of this section we will prove Proposition 4.1. For simplicity we set

$$p_0=\frac{d+3k}{d+k}.$$

By translation invariance we may assume that Q_R is centered at the origin. Let

$$\mathcal{W}_i\subset\big\{(\ell,\nu)\in\mathcal{L}\times\mathcal{V}:\nu\in S_i+O(R^{-\frac12})\big\},\quad i=1,2.$$

By Lemma 4.3 and the standard reduction with pigeonholing, which may only cause a loss of $(\log R)^C$, see [Lee 2006; Tao 2003], the matter is reduced to showing

$$\Big\|\sum_{\omega_1\in\mathcal{W}_1}P_{\omega_1}\sum_{\omega_2\in\mathcal{W}_2}P_{\omega_2}\Big\|_{L^{p_0}(Q_R)}\lessapprox(R^{(1-\delta)\alpha}+R^{C\delta})(\#\mathcal{W}_1\#\mathcal{W}_2)^{\frac12}$$

whenever P_{ω_1}, P_{ω_2} satisfy (i), (ii), (iv) in Lemma 4.3. Here $A\lessapprox B$ means $A\le C_\epsilon R^\epsilon B$ for any $\epsilon>0$.

By a further pigeonholing argument we specify the associated quantities in dyadic scales. Let $\mathcal{Q}$ be a collection of almost disjoint cubes of the same side length $\sim R^{\frac12}$ which cover Q_R. For each $q\in\mathcal{Q}$ we define

$$\mathcal{W}_j(q)=\{\omega_j\in\mathcal{W}_j:\pi_{\omega_j}\cap R^\delta q\neq\varnothing\}.$$

For dyadic numbers ρ_1,ρ_2 with $1\le\rho_1,\rho_2\le R^{100d}$, we define

$$\mathcal{Q}(\rho_1,\rho_2)=\{q\in\mathcal{Q}:\rho_j\le\#\mathcal{W}_j(q)<2\rho_j,\ j=1,2\}.\tag{4-7}$$

For $\omega\in\mathcal{W}_1\cup\mathcal{W}_2$, we set

$$\lambda(\omega;\rho_1,\rho_2)=\#\{q\in\mathcal{Q}(\rho_1,\rho_2):\pi_\omega\cap R^\delta q\neq\varnothing\}.$$

For a dyadic number $1\le\lambda\le R^{100d}$, we define

$$\mathcal{W}_j[\lambda;\rho_1,\rho_2]=\{\omega_j\in\mathcal{W}_j:\lambda\le\lambda(\omega_j;\rho_1,\rho_2)<2\lambda\},\quad j=1,2.\tag{4-8}$$

By a standard pigeonhole argument, it is sufficient to show

$$\Big(\sum_{q\in\mathcal{Q}(\rho_1,\rho_2)}\Big\|\sum_{\omega_1\in\mathcal{W}_1[\lambda_1;\rho_1,\rho_2]}P_{\omega_1}\sum_{\omega_2\in\mathcal{W}_2[\lambda_2;\rho_1,\rho_2]}P_{\omega_2}\Big\|_{L^{p_0}(q)}^{p_0}\Big)^{\frac{1}{p_0}}\lessapprox(R^{(1-\delta)\alpha}+R^{C\delta})\,(\#\mathcal{W}_1\#\mathcal{W}_2)^{\frac12}.\tag{4-9}$$

For the rest of the proof we assume that $q\in\mathcal{Q}(\rho_1,\rho_2)$, $\omega_1\in\mathcal{W}_1[\lambda_1;\rho_1,\rho_2]$ and $\omega_2\in\mathcal{W}_2[\lambda_1;\rho_1,\rho_2]$ if it is not mentioned otherwise. So, the above sums are denoted simply by $\sum_q$, $\sum_{\omega_1}$, and $\sum_{\omega_2}$, respectively.

Induction argument. For brevity let us put

$$\Delta = \bigcup_{q \in \mathcal{Q}(\rho_1,\rho_2)} q.$$

Let $\{B\}$ be a collection of almost disjoint cubes of the same side length $R^{1-\delta}$, which cover Q_R. Then

$$\text{the left-hand side of (4-9)} \le \sum_B \Big\| \sum_{\omega_1} P_{\omega_1} \sum_{\omega_2} P_{\omega_2} \Big\|_{L^{p_0}(\Delta \cap B)}. \tag{4-10}$$

We define a relation $\sim$ between ω_1 (or ω_2) and the cubes in $\{B\}$. For each $\omega \in \mathcal{W}_1[\lambda_1; \rho_1, \rho_2] \cup \mathcal{W}_2[\lambda_2; \rho_1, \rho_2]$, we define $B^*(\omega) \in \{B\}$ to be the cube which maximizes the quantity

$$\#\big\{q \in \mathcal{Q}(\rho_1, \rho_2) : \pi_\omega \cap R^\delta q \ne \varnothing,\ q \cap B \ne \varnothing\big\}. \tag{4-11}$$

Then the relation $\sim$ is defined as

$$\omega \sim B \quad \text{if } B \cap 10B^*(\omega) \ne \varnothing.$$

Here $10B^*(\omega)$ is the cube which has the same center as $B^*(\omega)$ and side length 10 times as large as that of $B^*(\omega)$. Using this relation we divide the sum into three parts to get

$$\begin{aligned}\sum_B \Big\| \sum_{\omega_1} P_{\omega_1} \sum_{\omega_2} P_{\omega_2} \Big\|_{L^{p_0}(\Delta \cap B)} \le{}& \sum_B \Big\| \sum_{\substack{\omega_1: \\ \omega_1 \sim B}} P_{\omega_1} \sum_{\substack{\omega_2: \\ \omega_2 \sim B}} P_{\omega_2} \Big\|_{L^{p_0}(\Delta \cap B)} \\ &+ \sum_B \Big\| \sum_{\substack{\omega_1: \\ \omega_1 \sim B}} P_{\omega_1} \sum_{\substack{\omega_2: \\ \omega_2 \nsim B}} P_{\omega_2} \Big\|_{L^{p_0}(\Delta \cap B)} + \sum_B \Big\| \sum_{\substack{\omega_1: \\ \omega_1 \nsim B}} P_{\omega_1} \sum_{\omega_2} P_{\omega_2} \Big\|_{L^{p_0}(\Delta \cap B)}.\end{aligned}$$

We will first show that

$$\sum_B \Big\| \sum_{\substack{\omega_1: \\ \omega_1 \sim B}} P_{\omega_1} \sum_{\substack{\omega_2: \\ \omega_2 \sim B}} P_{\omega_2} \Big\|_{L^{p_0}(\Delta \cap B)} \lesssim R^{(1-\delta)\alpha} (\#\mathcal{W}_1 \#\mathcal{W}_2)^{\frac{1}{2}}. \tag{4-12}$$

By applying the hypothesis (1-6), (iv) in Lemma 4.3, and the Cauchy–Schwarz inequality,

$$\sum_B \Big\| \sum_{\substack{\omega_1: \\ \omega_1 \sim B}} P_{\omega_1} \sum_{\substack{\omega_2: \\ \omega_2 \sim B}} P_{\omega_2} \Big\|_{L^{p_0}(\Delta \cap B)} \le C R^{(1-\delta)\alpha} \prod_{j=1}^{2} \Big(\sum_B \#\{\omega_j : \omega_j \sim B\} \Big)^{\frac{1}{2}}.$$

From the definition of the relation $\sim$ it is clear that $\#\{B : \omega_j \sim B\} \le C$. Hence, for $j = 1, 2$

$$\sum_B \#\{\omega_j : \omega_j \sim B\} = \sum_{\omega_j} \#\{B : \omega_j \sim B\} \lesssim W_j.$$

By inserting this into the previous inequality, we get (4-12).

Now, to prove (4-9) it is enough to show

$$\Big\| \sum_{\substack{\omega_1:\\ \omega_1\sim B}} P_{\omega_1} \sum_{\substack{\omega_2:\\ \omega_2\nsim B}} P_{\omega_2} \Big\|_{L^{p_0}(\Delta\cap B)} \lessapprox R^{C\delta} (\#\mathcal{W}_1\,\#\mathcal{W}_2)^{\frac12}$$

and

$$\Big\| \sum_{\substack{\omega_1:\\ \omega_1\nsim B}} P_{\omega_1} \sum_{\omega_2} P_{\omega_2} \Big\|_{L^{p_0}(\Delta\cap B)} \lessapprox R^{C\delta} (\#\mathcal{W}_1\,\#\mathcal{W}_2)^{\frac12}. \tag{4-13}$$

The proofs of these two estimates are similar. So, we will only prove (4-13). By Plancherel's theorem, $\|Ef(\cdot,t)\|_2 \le \|f\|_2$ for all $t\in\mathbb{R}^k$. Integration in t gives $\|Ef\|_{L^2(Q_R)} \lesssim R^k/2\|f\|_2$. By the Schwarz inequality it follows that

$$\|E_1 f E_2 g\|_{L^1(Q_R)} \lesssim R^k \|f\|_2\,\|g\|_2.$$

Combining this with (iv) in Lemma 4.3 yields

$$\Big\| \sum_{\substack{\omega_1:\\ \omega_1\nsim B}} P_{\omega_1} \sum_{\omega_2} P_{\omega_2} \Big\|_{L^1(\Delta\cap B)} \lesssim R^k (\#\mathcal{W}_1\#\mathcal{W}_2)^{\frac12}. \tag{4-14}$$

Hence, the (4-13) follows from interpolation between (4-14) and

$$\Big\| \sum_{\substack{\omega_1:\\ \omega_1\nsim B}} P_{\omega_1} \sum_{\omega_2} P_{\omega_2} \Big\|_{L^2(\Delta\cap B)} \lessapprox R^{C\delta} R^{-\frac{d-k}{4}} (\#\mathcal{W}_1\#\mathcal{W}_2)^{\frac12}. \tag{4-15}$$

Now it remains to show the L^2-estimate (4-15).

L^2 estimate. To prove (4-15) it suffices to show

$$\sum_{\substack{q\in\mathcal{Q}(\rho_1,\rho_2)\\ q\subset 2B}} \Big\| \sum_{\substack{\omega_1:\\ \omega_1\nsim B}} P_{\omega_1} \sum_{\omega_2} P_{\omega_2} \Big\|^2_{L^2(q)} \lessapprox R^{C\delta} R^{-\frac{d-k}{2}} \#\mathcal{W}_1\#\mathcal{W}_2. \tag{4-16}$$

For $j=1,2$, let us set

$$\mathcal{W}_j(q) = \{\omega_j \in \mathcal{W}_i[\lambda_j;\rho_1,\rho_2] : \omega_j \cap R^\delta q \neq \varnothing\},\quad \mathcal{W}_j^{\nsim B}(q) = \{\omega_j\in\mathcal{W}_j(q) : \omega_j \nsim B\}.$$

Then by (ii) in Lemma 4.3 we may discard some harmless terms, whose contributions are $O(R^{-C\delta})$. Hence, it suffices to show

$$\sum_{\substack{q\in\mathcal{Q}(\rho_1,\rho_2)\\ q\subset 2B}} \Big\| \sum_{\omega_1\in\mathcal{W}_1^{\nsim B}(q)} P_{\omega_1} \sum_{\omega_2(q)} P_{\omega_2} \Big\|^2_2 \lessapprox R^{C\delta} R^{-\frac{d-k}{2}} \#\mathcal{W}_1\,\#\mathcal{W}_2. \tag{4-17}$$

By using Plancherel's theorem we write

$$\Big\| \sum_{\omega_1\in\mathcal{W}_1^{\nsim B}(q)} P_{\omega_1} \sum_{\omega_2\in\mathcal{W}_2(q)} P_{\omega_2} \Big\|^2_2 = \sum_{\omega_1\in\mathcal{W}_1^{\nsim B}(q)}\ \sum_{\omega_2'\in\mathcal{W}_2(q)}\ \sum_{\omega_1'\in\mathcal{W}_1^{\nsim B}(q)}\ \sum_{\omega_2\in\mathcal{W}_2(q)} \langle \widehat{P}_{\omega_1} * \widehat{P}_{\omega_2}, \widehat{P}_{\omega_1'} * \widehat{P}_{\omega_2'}\rangle.$$

Let us write $\omega_j=(\ell_j,\nu_j)$, $\omega_j'=(\ell_j',\nu_j')$, $j=1,2$. For any $\nu_1\in S_1$, $\nu_2'\in S_2$, we define $\mathcal{W}_1^{\sim B}(q;\nu_1,\nu_2')$ by

$$\mathcal{W}_1^{\sim B}(q;\nu_1,\nu_2')=\{\omega_1'=(\ell_1',\nu_1')\in\mathcal{W}_1^{\sim B}(q):\nu_1'\in\Pi_1^{\nu_1,\nu_2'}+O(R^{-\frac12})\}.$$

Then $\widehat{P}_{\omega_1}*\widehat{P}_{\omega_2}$ is supported on the $O(R^{-\frac12})$-neighborhood of the point $(\nu_1+\nu_2,\Phi(\nu_1)+\Phi(\nu_2))$. So the inner product $\langle\widehat{P}_{\omega_1}*\widehat{P}_{\omega_2},\widehat{P}_{\omega_1'}*\widehat{P}_{\omega_2'}\rangle$ vanishes unless

$$\nu_1+\nu_2=\nu_1'+\nu_2'+O(R^{-\frac12}),\quad \Phi(\nu_1)+\Phi(\nu_2)=\Phi(\nu_1')+\Phi(\nu_2')+O(R^{-\frac12}).$$

Thus, for given ν_1 and ν_2', we see that ν_1' is contained in an $O(R^{-\frac12})$-neighborhood of $\Pi_1^{\nu_1,\nu_2'}$, which is defined by (3-1). Once ν_1, ν_1' and ν_2' are given, then there are only $O(1)$ many ν_2, since ν_2 should be in an $O(R^{-\frac12})$-neighborhood of the point $\nu_1+\nu_1'-\nu_2'$. Therefore,

$$\Big\|\sum_{\omega_1\in\mathcal{W}_1^{\sim B}(q)}P_{\omega_1}\sum_{\omega_2\in\mathcal{W}_2(q)}P_{\omega_2}\Big\|_2^2\lesssim R^{-\frac{d-k}{2}}\sum_{\omega_1\in\mathcal{W}_1^{\sim B}(q)}\;\sum_{\omega_2'\in\mathcal{W}_2(q)}\#\mathcal{W}_1^{\sim B}(q;\nu_1,\nu_2'),$$

where we also used

$$|\langle P_{\omega_1}P_{\omega_2},P_{\omega_1'}P_{\omega_2'}\rangle|\lesssim R^{-\frac{d-k}{2}}.$$

This follows from (ii) in Lemma 4.3 and the transversality between π_{ω_1} ($\pi_{\omega_1'}$) and π_{ω_2} ($\pi_{\omega_2'}$), respectively. Hence, (4-17) follows if we show

$$\max_{q\subset 2B,\,\nu_1,\nu_2'}\#\mathcal{W}_1^{\sim B}(q;\nu_1,\nu_2')\sum_{\substack{q\in\mathcal{Q}(\rho_1,\rho_2)\\ q\subset 2B}}\#\mathcal{W}_1^{\sim B}(q)\#\mathcal{W}_2(q)\lesssim R^{C\delta}\#\mathcal{W}_1\,\#\mathcal{W}_2.\tag{4-18}$$

We will prove (4-18), assuming for the moment that

$$\max_{q\subset 2B,\,\nu_1,\nu_2'}\#\mathcal{W}_1^{\sim B}(q;\nu_1,\nu_2')\lesssim R^{C\delta}\frac{\#\mathcal{W}_2}{\lambda_1\rho_2}.\tag{4-19}$$

To this end it is enough to show

$$\sum_{\substack{q\in\mathcal{Q}(\rho_1,\rho_2)\\ q\subset 2B}}\#\mathcal{W}_1^{\sim B}(q)\#\mathcal{W}_2(q)\lesssim\lambda_1\rho_2\#\mathcal{W}_1.$$

Recalling $\#\mathcal{W}_2(q)\lesssim\rho_2$, we see that the left-hand side is bounded by

$$C\rho_2\sum_{q\in\mathcal{Q}(\rho_1,\rho_2)}\#\mathcal{W}_1(q).$$

Changing the order of summation, we see this in turn is bounded by $C\rho_2\sum_{\omega_1}\#\{q\in\mathcal{Q}(\rho_1,\rho_2):\pi_{w_1}\cap R^\delta q\}$. Since

$$\#\{q\in\mathcal{Q}(\rho_1,\rho_2):\pi_{w_1}\cap R^\delta q\}\lesssim\lambda_1,$$

the desired inequality (4-18) follows.

***Proof of* (4-19).** Fix $q \subset 2B$, $\nu_1 \in S_1$ and $\nu_2' \in S_2$. Let us consider the set

$$\boldsymbol{S} := \big\{(\tilde{q}, \omega_1, \omega_2) \in \mathcal{Q}(\rho_1, \rho_2) \times \mathcal{W}_1^{\nsim B}(q; \nu_1, \nu_2') \times \mathcal{W}_2 \\ : \pi_{\omega_1} \cap R^{\delta}\tilde{q} \neq \varnothing,\ \pi_{\omega_2} \cap R^{\delta}\tilde{q} \neq \varnothing,\ \operatorname{dist}(\tilde{q}, q) \geq R^{1-\delta}\big\}.$$

To prove (4-19) it suffices to show

$$R^{-C\delta}\lambda_1\rho_2 \#\mathcal{W}_1^{\nsim B}(q; \nu_1, \nu_2')) \lesssim \#\boldsymbol{S} \lesssim R^{C\delta}\#\mathcal{W}_2. \tag{4-20}$$

For the lower bound it is enough to show that, for each $\omega_1 \in \mathcal{W}_1^{\nsim B}(q; \nu_1, \nu_2')$,

$$\#\big\{(\tilde{q}, \omega_2) \in \mathcal{Q}(\rho_1, \rho_2) \times \mathcal{W}_2 : \pi_{\omega_1} \cap R^{\delta}\tilde{q} \neq \varnothing,\ \pi_{\omega_2} \cap R^{\delta}\tilde{q} \neq \varnothing,\ \operatorname{dist}(\tilde{q}, q) \geq R^{1-\delta}\big\} \geq R^{-C\delta}\lambda_1\rho_2.$$

By (4-8), ω_1 contains as many as $O(\lambda_1)$ cubes $\tilde{q}$ in $\mathcal{Q}(\rho_1, \rho_2)$. (Recall that we are assume assuming $q \in \mathcal{Q}(\rho_1, \rho_2)$, $\omega_1 \in \mathcal{W}_1[\lambda_1; \rho_1, \rho_2]$ and $\omega_2 \in \mathcal{W}_2[\lambda_1; \rho_1, \rho_2]$.) Let $B^*(\omega_1) \in \mathcal{Q}$ be the cube which maximizes the quantity given by (4-11) with $\omega = \omega_1$. Since $\omega_1 \nsim B$, it follows from the definition of the relation $\sim$ that $\operatorname{dist}(B^*(\omega_1), B) \gtrsim R^{1-\delta}$. Since $\pi_{\omega_1} + O(R^{\frac{1}{2}+\delta})$ can be covered by $R^{C\delta}$ cubes B, by a simple pigeonholing argument we get

$$\#\big\{\tilde{q} \in \mathcal{Q}(\rho_1, \rho_2) : \pi_{\omega_1} \cap R^{\delta}\tilde{q} \neq \varnothing,\ \operatorname{dist}(\tilde{q}, q) \geq R^{1-\delta}\big\} \gtrsim R^{-C\delta}\lambda_1.$$

Next, for the upper bound it suffices to show that, for any $\omega_2 \in \mathcal{W}_2$,

$$\#\big\{(\tilde{q}, \omega_1) \in \mathcal{Q}(\rho_1, \rho_2) \times \mathcal{W}_1^{\nsim B}(q; \nu_1, \nu_2') \\ : \pi_{\omega_1} \cap R^{\delta}\tilde{q} \neq \varnothing,\ \pi_{\omega_2} \cap R^{\delta}\tilde{q} \neq \varnothing,\ \operatorname{dist}(\tilde{q}, q) \gtrsim R^{1-\delta}\big\} \lesssim R^{C\delta}. \tag{4-21}$$

Let $\boldsymbol{z}_0$ be the center of q. Then, by the definition of $\mathcal{W}_1^{\nsim B}(q_0; \nu_1, \nu_2')$, it follows that

$$\bigcup_{\omega_1 \in \mathcal{W}_1^{\nsim B}(q; \nu_1, \nu_2')} \pi_{\omega_1} \subset \Gamma_1^{\nu_1, \nu_2'}(CR^{\frac{1}{2}+\delta}) + \boldsymbol{z}_0.$$

If $\omega_2 \in \mathcal{W}_2$, then it follows from Lemma 3.2 that the intersection

$$\pi_{\omega_2} \cap \Bigg(\bigcup_{\omega_1 \in \mathcal{W}_1^{\nsim B}(q; \nu_1, \nu_2')} \pi_{\omega_1}\Bigg)$$

is contained in a cube of side length $O(R^{\frac{1}{2}+\delta})$. Thus, there are at most $O(R^{C\delta})$ choices of balls $\tilde{q} \in \mathcal{Q}(\rho_1, \rho_2)$ such that $(\tilde{q}, \omega_1)$ is contained in the set in (4-21). On the other hand, since $\operatorname{dist}(\tilde{q}, q) \gtrsim R^{1-C\delta}$, we have

$$\#\big\{w_1 \in \mathcal{W}_1^{\nsim B}(q; \nu_1, \nu_2') : \pi_{\omega_1} \cap R^{\delta}\tilde{q} \neq \varnothing,\ \pi_{\omega_1} \cap R^{\delta}q \neq \varnothing\big\} \lesssim R^{C\delta}. \tag{4-22}$$

To see this, by scaling it is enough to check that the map $S_1 \ni \nu \mapsto \sum_{i=1}^{k} t_j \nabla\varphi_i(\nu)$ is one-to-one whenever $|t| = 1$. But this follows from the condition (1-2) if we take S_1 to be small enough. Thus we obtain the claim (4-21). Hence, we also have (4-9), which finishes the proof of Proposition 4.1. This completes the proof of Theorem 1.1. □

Proof of Theorem 1.2. Thanks to Lemma 3.3, the line of argument in the proof of Theorem 1.1 works without modification except that we need to show (4-22). However, to prove (4-22) we don't need to show $S_1 \ni \nu \mapsto \sum_{i=1}^{k} t_j \nabla\varphi_i(\nu)$ is one-to-one. Instead, as is clear after rescaling, it is enough to show that $\Pi^{\nu_1,\nu_2'} \ni \nu \mapsto \sum_{i=1}^{k} t_j \nabla\varphi_i(\nu)$ is one-to-one. Let $\boldsymbol{t}_1, \dots, \boldsymbol{t}_{d-k}$ be a set of vectors spanning the tangent space of $\Pi^{\nu_1,\nu_2'}$ at ν_0. Then the above follows if we show that the matrix

$$(\boldsymbol{t}_1^t, \dots, \boldsymbol{t}_{d-k}^t)\Big(\sum_{i=1}^{k} t_j H\varphi_i(\nu_0)\Big)$$

has rank $d-k$ for $|t| = 1$. In fact, $\boldsymbol{t}_1, \dots, \boldsymbol{t}_{d-k}$ are almost normal to the span of $\{\nabla\varphi_i(\nu_2) - \nabla\varphi_i(\nu_1) : i = 1, \dots, k\}$. These vectors are close to $\boldsymbol{n}_1, \dots, \boldsymbol{n}_{d-k}$. Hence, assuming that S_1 and S_2 are small enough, the above follows if we show $N(\nu_2, \nu_1)\sum_{i=1}^{k} t_j H\varphi_i(\nu_0)$ has rank $d-k$. This clearly follows from (1-5). $\square$

5. Restriction estimates for complex surfaces

In this section we provide the proofs of Corollary 1.5 and Theorem 1.6. In what follows we set $k = 2$, $d = 2n$.

Proof of Corollary 1.5. Let φ_1, φ_2 be given by $\frac{1}{2} z^t Dz = \varphi_1 + i\varphi_2$ so that

$$\varphi_1(x, y) = (x^t Dx - y^t Dy), \quad \varphi_2(x, y) = x^t Dy, \quad (x, y) \in \mathbb{R}^n \times \mathbb{R}^n.$$

In order to prove Corollary 1.5 we need only to show that the condition (1-13) implies the assumptions in Theorem 1.1.

Let us set $z_j = x_j + iy_j \in \mathbb{C}^n$ for $j = 1, 2$, $\delta_x = x_2 - x_1$, and $\delta_y = y_2 - y_1$. Then a computation shows that the associated matrix $\boldsymbol{M}(t, z_1, z_2, z)$ is given by

$$\boldsymbol{M}(t, z_1, z_2, z) = \begin{pmatrix} 0 & 0 & \delta_x^t D & -\delta_y^t D \\ 0 & 0 & \delta_y^t D & \delta_x^t D \\ D\delta_x & D\delta_y & t_1 D & t_2 D \\ -D\delta_y & D\delta_x & t_2 D & -t_1 D \end{pmatrix}.$$

Note that

$$\sum_{j=1}^{2} t_j H\varphi_j = \begin{pmatrix} t_1 D & t_2 D \\ t_2 D & -t_1 D \end{pmatrix}.$$

Then, it is easy to see that the inverse of $\sum_{j=1}^{2} t_j H\varphi_j$ is

$$(t_1^2 + t_2^2)^{-1} \begin{pmatrix} t_1 D^{-1} & t_2 D^{-1} \\ t_2 D^{-1} & -t_1 D^{-1} \end{pmatrix}.$$

So, the assumption (1-2) holds. Hence, it suffices to show that (1-13) implies (1-3). By the block matrix formula we only need to check

$$\det\left[\begin{pmatrix} \delta_x^t D & -\delta_y^t D \\ \delta_y^t D & \delta_x^t D \end{pmatrix}\begin{pmatrix} t_1 D^{-1} & t_2 D^{-1} \\ t_2 D^{-1} & -t_1 D^{-1} \end{pmatrix}\begin{pmatrix} D\delta_x & D\delta_y \\ -D\delta_y & D\delta_x \end{pmatrix}\right] \neq 0.$$

By a direct computation it is not difficult to see that the left-hand side equals[4]

$$-(t_1^2+t_2^2)\big((\delta_x^t D\delta_x-\delta_y^t D\delta_y)^2+4(\delta_x^t D\delta_y)^2\big).$$

Since $(z_2-z_1)^t D(z_2-z_1)=\delta_x^t D\delta_x-\delta_y^t D\delta_y+2i\delta_x^t D\delta_y$, it is now clear that (1-13) implies (1-3). □

Proof of Theorem 1.6. From the bilinear estimate we can get the linear estimate by adapting the arguments in [Tao et al. 1998; Vargas 2005; Lee 2006]. Since D is nonsingular and symmetric, by making use of linear transforms we may assume that

$$D=\begin{pmatrix}1&0\\0&\pm1\end{pmatrix},$$

and so we have either $\Phi(z_1,z_2)=z_1^2+z_2^2=(z_1+iz_2)(z_1-iz_2)$ or $\Phi(z_1,z_2)=z_1^2-z_2^2=(z_1+z_2)(z_1-z_2)$. By a linear change of variables the problem can be further reduced to showing Theorem 1.6 when $\Phi(z_1,z_2)=z_1z_2$.

The following is an immediate consequence of Theorem 1.1 and the translation invariance of the bilinear estimate.

Lemma 5.1. *Let $\Phi(z_1,z_2)=z_1z_2$ and $Q_1,Q_2\subset\mathbb{C}^2$ be closed cubes. Assume that*

$$2^4\ge|z_1-w_1|\ge2^{-1}\quad\textit{and}\quad 2^4\ge|z_2-w_2|\ge2^{-1}$$

whenever $(z_1,z_2)\in Q_1$ and $(w_1,w_2)\in Q_2$. If $\operatorname{supp}(f)\subset Q_1$ and $\operatorname{supp}(g)\subset Q_2$, then for $q>\frac{10}{3}$ and $\frac{1}{p}+\frac{5}{3q}<1$,

$$\|Ef\,Eg\|_{\frac{q}{2}}\le C_{p,q}\,\|f\|_p\,\|g\|_p.$$

In the next lemma the hypothesis of "nonvanishing rotational curvature" is weakened to the usual separation condition. But then, for the conclusion to hold, the pair $\big(\frac{1}{p},\frac{1}{q}\big)$ needs to satisfy a more restrictive condition. This lemma is an analog of Proposition 4.1 in [Lee 2006].

Lemma 5.2. *Let Q_1, Q_2 be closed cubes in $\mathbb{C}^2$ such that $\operatorname{dist}(Q_1,Q_2)\ge1$. If $\operatorname{supp}(f)\subset Q_1$ and $\operatorname{supp}(g)\subset Q_2$, then there is a constant $C_{p,q}$ such that*

$$\|Ef\,Eg\|_{\frac{q}{2}}\le C_{p,q}\,\|f\|_p\,\|g\|_p\quad\textit{if } \tfrac{1}{p}+\tfrac{2}{q}<1,\ q>\tfrac{10}{3},$$

or

$$\|E\chi_F\,E\chi_G\|_{\frac{q}{2}}\lesssim\|f\|_{p,1}\,\|g\|_{p,1}\quad\textit{if } \tfrac{1}{p}+\tfrac{2}{q}=1,\ q>\tfrac{10}{3}.$$

By translation it is clear that in Lemmas 5.1 and 5.2 the same estimate holds with Q_1, Q_2 replaced by Q_1+a, Q_2+a, respectively, for any $a\in\mathbb{C}^2$. It is possible to prove the strong-type estimate $\|E\chi_F\,E\chi_G\|_{q/2}\lesssim\|f\|_p\,\|g\|_p$ for $\frac{1}{p}+\frac{2}{q}=1$, $q>\frac{10}{3}$ by making use of the asymmetric estimates which are obtained in the course of proof of Proposition 1.3 and the bilinear interpolation; see, e.g., [Bergh and Löfström 1976, Section 3.13, 5(b)]. However, we have decided not to include the details here, because it does not seem to have any consequences for linear estimates.

[4] In fact, the product of the three matrices is equal to $\begin{pmatrix}t_1&-t_2\\t_2&t_1\end{pmatrix}\begin{pmatrix}\delta_x^tD\delta_x-\delta_y^tD\delta_y & 2\delta_x^tD\delta_y\\ 2\delta_x^tD\delta_y & \delta_y^tD\delta_y-\delta_x^tD\delta_x\end{pmatrix}$.

Proof of Lemma 5.2. By interpolation it suffices to consider the case $\frac{10}{3} < q \le 4$ and $p \le q$. By decomposition of the domains, followed by translation and scaling, we may assume that $Q_1 = H_1 \times K$ and $Q_2 = H_2 \times K$, where $\operatorname{dist}(H_1, H_2) \ge 2^{-1}$ and K is the unit cube in $\mathbb{C}$, centered at the origin.

By a Whitney decomposition, we get

$$(K \times K) \setminus D = \bigcup_{j>1} \bigcup_{\substack{(k,k'):\\ I_k^j \sim I_{k'}^j}} I_k^j \times I_{k'}^j$$

where $D = \{(z_2, w_2) : z_2 = w_2\}$, and $\{I_k^j\}_k$ are the dyadic cubes in $\mathbb{C}$ of side length 2^{-j}, and as usual the notation $I_k^j \sim I_{k'}^j$ means that the parent cubes of I_k^j and $I_{k'}^j$ are adjacent, while I_k^j and $I_{k'}^j$ are not.

Let us set

$$f_k^j(z_1, z_2) = \chi_{I_k^j}(z_2) f(z_1, z_2), \quad g_k^j(w_1, w_2) = \chi_{I_k^j}(w_2) g(w_1, w_2).$$

Then, since the cubes $I_k^j \times I_{k'}^j$ are almost disjoint, we may write

$$Ef\, Eg = \sum_j \sum_{\substack{(k,k'):\\ I_k^j \sim I_{k'}^j}} E(f_k^j)\, E(g_{k'}^j).$$

Since $q > \frac{10}{3}$, we get

$$\|Ef\, Eg\|_{\frac{q}{2}} \le \sum_j \Bigg\| \sum_{I_k^j \sim I_{k'}^j} E(f_k^j)\, E(g_{k'}^j) \Bigg\|_{\frac{q}{2}} \lesssim \sum_j \Bigg(\sum_{I_k^j \sim I_{k'}^j} \|E(f_k^j)\, E(g_{k'}^j)\|_{\frac{q}{2}}^{\frac{q}{2}} \Bigg)^{\frac{2}{q}},$$

where the last inequality follows from Lemma 6.1 in [Tao et al. 1998]. Here, we used the fact that for each fixed j the supports of the Fourier transforms of $E(f_k^j)\, E(g_{k'}^j)$ have uniformly bounded overlap as (k, k') varies, provided that $I_k^j \sim I_{k'}^j$. This is a consequence of the Whitney decomposition. We now claim that if $I_k^j \sim I_{k'}^j$, then

$$\|E(f_k^j)\, E(g_{k'}^j)\|_{\frac{q}{2}} \lesssim 2^{4j(\frac{1}{p}+\frac{2}{q}-1)} \|f_k^j\|_p \|g_{k'}^j\|_p \tag{5-1}$$

when $\frac{1}{p} + \frac{5}{3q} < 1$, $q > \frac{10}{3}$. This is an easy consequence of a translated version of Lemma 5.1. Assuming this for the moment, we will finish the proof. Since $q \ge p$, for $\frac{1}{p} + \frac{5}{3q} < 1$, $4 > q > \frac{10}{3}$, we have

$$\begin{aligned}\|Ef\, Eg\|_{\frac{q}{2}} &\le \sum_j 2^{4j(\frac{1}{p}+\frac{2}{q}-1)} \Bigg(\sum_{I_k^j \sim I_{k'}^j} \|f_k^j\|_p^{\frac{q}{2}} \|g_{k'}^j\|_p^{\frac{q}{2}} \Bigg)^{\frac{2}{q}} \\ &\lesssim \sum_j 2^{4j(\frac{1}{p}+\frac{2}{q}-1)} \Bigg(\sum_{I^j} \|f_k^j\|_p^p \Bigg)^{\frac{1}{p}} \Bigg(\sum_{J^j} \|g_{k'}^j\|_p^p \Bigg)^{\frac{1}{p}} \lesssim \sum_j 2^{4j(\frac{1}{p}+\frac{2}{q}-1)} \|f\|_p \|g\|_p.\end{aligned}$$

Now take $f = \chi_F$ and $g = \chi_G$ for measurable sets F, G contained in V_1, V_2, respectively. Fix p, q with $4 > q > \frac{10}{3}$, $\frac{1}{p} + \frac{2}{q} = 1$, and choose p_1 and p_2 such that $\frac{1}{p_j} + \frac{5}{3q} < 1$, $j = 1, 2$, and

$$\frac{1}{p_1} + \frac{2}{q} - 1 = \eta, \quad \frac{1}{p_2} + \frac{2}{q} - 1 = -\eta$$

for some small $\eta > 0$. Then by applying the last estimate for $p = p_1$ and $p = p_2$, we obtain

$$\|E\chi_F\, E\chi_G\|_{\frac{q}{2}} \lesssim \sum_j \min\{2^{4j\eta}|F|^{\eta+1-\frac{2}{q}}|G|^{\eta+1-\frac{2}{q}},\ 2^{-4j\eta}|F|^{-\eta+1-\frac{2}{q}}|G|^{-\eta+1-\frac{2}{q}}\}$$
$$\lesssim |F|^{1-\frac{2}{q}}|G|^{1-\frac{2}{q}} = |F|^{1/p}|G|^{\frac{1}{p}}.$$

This shows the estimate $\|E\chi_F\, E\chi_G\|_{q/2} \lesssim \|f\|_{p,1}\,\|g\|_{p,1}$ for $\frac{1}{p}+\frac{2}{q}=1,\ 4>q>\frac{10}{3}$.

Now it remains to show (5-1). Clearly, I_k^j and $I_{k'}^j$ are contained in a ball of radius 2^{2-j} and $\operatorname{dist}(I_k^j, I_{k'}^j) \ge 2^{-1-j}$. Hence, by a change of variables,

$$E(f_k^j)(w) = 2^{-2j}\, E(f_k^j(\cdot, 2^{-j}\,\cdot))(w_1, 2^{-j}w_2, 2^{-j}w_3),$$
$$E(g_{k'}^j)(w) = 2^{-2j}\, E(g_{k'}^j(\cdot, 2^{-j}\,\cdot))(w_1, 2^{-j}w_2, 2^{-j}w_3).$$

Then we see that $\operatorname{supp} f_k^j(\cdot, 2^{-j}\,\cdot) \subset H_1 \times \tilde I_1$ and $g_{k'}^j(\cdot, 2^{-j}\,\cdot) \subset H_2 \times \tilde I_2$ if $\operatorname{dist}(\tilde I_1, \tilde I_2) \ge 2^{-1}$ and $\tilde I_1, \tilde I_2$ are contained in a ball of radius $\le 2^3$. The assumption of Lemma 5.1 is satisfied with $f = f_k^j(\cdot, 2^{-j}\,\cdot)$ and $g = g_{k'}^j(\cdot, 2^{-j}\,\cdot)$. Hence we may apply it to $E(f_k^j(\cdot, 2^{-j}\,\cdot))E(f_k^j(\cdot, 2^{-j}\,\cdot))$ and get (5-1). $\square$

Once Lemma 5.2 is established, the usual argument in [Tao et al. 1998], used to deduce linear estimates from bilinear ones, works without modification. We omit the details. $\square$

Acknowledgements

We thank Andreas Seeger for bringing the reference [Alvarez 1997] to our attention. We would also like to thank the anonymous referee for the reference [Banner 2002].

References

[Alvarez 1997] D. Alvarez, *Bounds for some Kakeya-type maximal functions*, Ph.D. thesis, University of California, Berkeley, 1997, available at https://search.proquest.com/docview/304343467. MR

[Bak and Ham 2014] J.-G. Bak and S. Ham, "Restriction of the Fourier transform to some complex curves", *J. Math. Anal. Appl.* **409**:2 (2014), 1107–1127. MR Zbl

[Bak and Lee 2004] J.-G. Bak and S. Lee, "Restriction of the Fourier transform to a quadratic surface in $\mathbb{R}^n$", *Math. Z.* **247**:2 (2004), 409–422. MR Zbl

[Bak and Seeger 2011] J.-G. Bak and A. Seeger, "Extensions of the Stein–Tomas theorem", *Math. Res. Lett.* **18**:4 (2011), 767–781. MR Zbl

[Bak et al. 2002] J.-G. Bak, D. M. Oberlin, and A. Seeger, "Two endpoint bounds for generalized Radon transforms in the plane", *Rev. Mat. Iberoamericana* **18**:1 (2002), 231–247. MR Zbl

[Bak et al. 2009] J.-G. Bak, D. M. Oberlin, and A. Seeger, "Restriction of Fourier transforms to curves and related oscillatory integrals", *Amer. J. Math.* **131**:2 (2009), 277–311. MR Zbl

[Bak et al. 2013] J.-G. Bak, D. M. Oberlin, and A. Seeger, "Restriction of Fourier transforms to curves: an endpoint estimate with affine arclength measure", *J. Reine Angew. Math.* **682** (2013), 167–205. MR Zbl

[Banner 2002] A. D. Banner, *Restriction of the Fourier transform to quadratic submanifolds*, Ph.D. thesis, Princeton University, 2002, available at https://search.proquest.com/docview/305532402. MR

[Bennett et al. 2006] J. Bennett, A. Carbery, and T. Tao, "On the multilinear restriction and Kakeya conjectures", *Acta Math.* **196**:2 (2006), 261–302. MR Zbl

[Bergh and Löfström 1976] J. Bergh and J. Löfström, *Interpolation spaces: an introduction*, Grundlehren der Mathematischen Wissenschaften **223**, Springer, 1976. MR Zbl

[Bourgain and Guth 2011] J. Bourgain and L. Guth, "Bounds on oscillatory integral operators based on multilinear estimates", *Geom. Funct. Anal.* **21**:6 (2011), 1239–1295. MR Zbl

[Christ 1982] M. Christ, *Restriction of the Fourier transform to submanifolds of low codimension*, Ph.D. thesis, The University of Chicago, 1982, available at https://search.proquest.com/docview/303090254. MR

[Christ 1985] M. Christ, "On the restriction of the Fourier transform to curves: endpoint results and the degenerate case", *Trans. Amer. Math. Soc.* **287**:1 (1985), 223–238. MR Zbl

[Greenleaf and Seeger 2002] A. Greenleaf and A. Seeger, "Oscillatory and Fourier integral operators with degenerate canonical relations", *Publ. Mat.* extra volume (2002), 93–141. MR Zbl

[Guth 2016] L. Guth, "A restriction estimate using polynomial partitioning", *J. Amer. Math. Soc.* **29**:2 (2016), 371–413. MR Zbl

[Lee 2006] S. Lee, "Bilinear restriction estimates for surfaces with curvatures of different signs", *Trans. Amer. Math. Soc.* **358**:8 (2006), 3511–3533. MR Zbl

[Mockenhaupt 1996] G. Mockenhaupt, *Bounds in Lebesgue spaces of oscillatory integral operators*, Habilitation thesis, Universität Siegen, 1996, available at https://tinyurl.com/Mockenhaupt-thesis. Zbl

[Moyua et al. 1999] A. Moyua, A. Vargas, and L. Vega, "Restriction theorems and maximal operators related to oscillatory integrals in $\mathbb{R}^3$", *Duke Math. J.* **96**:3 (1999), 547–574. MR Zbl

[Oberlin 2005] D. M. Oberlin, "A restriction theorem for a k-surface in $\mathbb{R}^n$", *Canad. Math. Bull.* **48**:2 (2005), 260–266. MR Zbl

[Stein 1993] E. M. Stein, *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, Princeton Mathematical Series **43**, Princeton University Press, 1993. MR Zbl

[Stovall 2016] B. Stovall, "Uniform estimates for Fourier restriction to polynomial curves in $\mathbb{R}^d$", *Amer. J. Math.* **138**:2 (2016), 449–471. MR Zbl

[Tao 2003] T. Tao, "A sharp bilinear restrictions estimate for paraboloids", *Geom. Funct. Anal.* **13**:6 (2003), 1359–1384. MR Zbl

[Tao and Vargas 2000] T. Tao and A. Vargas, "A bilinear approach to cone multipliers, I: Restriction estimates", *Geom. Funct. Anal.* **10**:1 (2000), 185–215. MR Zbl

[Tao et al. 1998] T. Tao, A. Vargas, and L. Vega, "A bilinear approach to the restriction and Kakeya conjectures", *J. Amer. Math. Soc.* **11**:4 (1998), 967–1000. MR Zbl

[Vargas 2005] A. Vargas, "Restriction theorems for a surface with negative curvature", *Math. Z.* **249**:1 (2005), 97–111. MR Zbl

[Wolff 2001] T. Wolff, "A sharp bilinear cone restriction estimate", *Ann. of Math.* (2) **153**:3 (2001), 661–698. MR Zbl

Received 23 Jan 2017. Revised 2 Jun 2017. Accepted 12 Jul 2017.

JONG-GUK BAK: bak@postech.ac.kr
Department of Mathematics, Pohang University of Science and Technology, Pohang, South Korea

JUNGJIN LEE: jungjinlee@unist.ac.kr
Department of Mathematical Sciences, School of Natural Science, Ulsan National Institute of Science and Technology, Ulsan, South Korea

SANGHYUK LEE: shklee@snu.ac.kr
Department of Mathematical Sciences, Seoul National University, Seoul, South Korea

ANALYSIS AND PDE
Vol. 10, No. 8, 2017
dx.doi.org/10.2140/apde.2017.10.1987

COMPLETE EMBEDDED COMPLEX CURVES IN THE BALL OF $\mathbb{C}^2$ CAN HAVE ANY TOPOLOGY

ANTONIO ALARCÓN AND JOSIP GLOBEVNIK

In this paper we prove that the unit ball $\mathbb{B}$ of $\mathbb{C}^2$ admits complete properly embedded complex curves of any given topological type. Moreover, we provide examples containing any given closed discrete subset of $\mathbb{B}$.

1. Introduction

Yang [1977a; 1977b] raised the question of whether there exist complete bounded complex submanifolds of a complex Euclidean space $\mathbb{C}^N$ $(N > 1)$. Recall that an immersed k-dimensional complex submanifold $\psi : M^k \to \mathbb{C}^N$ $(1 \le k < N)$ is said to be *complete* if the Riemannian metric induced on M by the Euclidean metric in $\mathbb{C}^N$ via ψ is complete in the classical sense, or equivalently, if the image by ψ of any divergent path on M has infinite Euclidean length.

The case of main interest to us in this paper is when $k = 1$ and $N = 2$, i.e., complex curves in the complex Euclidean plane $\mathbb{C}^2$. There are many known examples of complete bounded *immersed* complex curves in $\mathbb{C}^2$, which have been obtained by different methods; see the works of Jones [1979] for discs, Martín, Umehara, and Yamada [Martín et al. 2009] for some finite topologies, Alarcón and López [2013a] for examples with arbitrary topology, and Alarcón and Forstnerič [2013] and Alarcón, Drinovec Drnovšek, Forstnerič, and López [Alarcón et al. 2015] for examples normalized by any given bordered Riemann surface. Furthermore, the curves in [Alarcón and López 2013a; Alarcón and Forstnerič 2013; Alarcón et al. 2015] can be chosen to be proper in any given convex domain of $\mathbb{C}^2$, in particular, in the open unit Euclidean ball, which throughout this paper will be denoted by $\mathbb{B}$.

On the other hand, using the techniques developed in the cited sources and taking into account that the general position of complex curves in $\mathbb{C}^N$ is embedded for all $N \ge 3$, it is not very hard to construct complete bounded *embedded* complex curves in $\mathbb{C}^N$ for any such N; see again [Alarcón and Forstnerič 2013; Alarcón et al. 2015]. For submanifolds of higher dimension, Alarcón and Forstnerič [2013] provided examples of complete bounded embedded k-dimensional complex submanifolds of $\mathbb{C}^{3k}$ for any $k \in \mathbb{N}$, whereas Drinovec Drnovšek [2015] proved that every bounded, strictly pseudoconvex, smoothly bounded domain of $\mathbb{C}^k$ admits a complete proper holomorphic embedding into the unit ball of $\mathbb{C}^N$ provided that the codimension $N - k$ is sufficiently large.

However, constructing complete bounded *embedded* complex curves in $\mathbb{C}^2$ (and, more generally, complete bounded embedded complex hypersurfaces of $\mathbb{C}^N$ for $N > 1$) is a much more arduous task; the

MSC2010: 32H02, 32B15, 32C22.
Keywords: complex curves, holomorphic embeddings, complete bounded submanifolds.

main reason why is that *self-intersections of complex curves in* $\mathbb{C}^2$ *are stable under small deformations*. It is therefore not surprising that the first known examples of such curves were only found almost four decades after Yang posed his question; they were given in [Alarcón and López 2016]. Their method, which in fact furnishes complete properly embedded complex curves in any convex domain of $\mathbb{C}^2$, is rather involved and relies, among other things, on a subtle self-intersection removal procedure that does not allow for the inference of any information on the topological type of the examples. A little later Globevnik [2015; 2016b], using a different technique, extended the results in [Alarcón and López 2016] by proving the existence of complete properly embedded complex hypersurfaces in any pseudoconvex domain D of $\mathbb{C}^N$ for any $N > 1$; this settles the *embedded Yang problem* in an optimal way in all dimensions. His examples are given as level sets of highly oscillating holomorphic functions $D \to \mathbb{C}$, and hence, again, no information on their topology is provided. In light of the above, the following questions naturally appear; see [Alarcón and López 2016, Question 1.5; Globevnik 2015, Questions 13.1 and 13.2]:

Problem 1.1. Is there any restriction on the topology of a complete bounded embedded complex hypersurface of $\mathbb{C}^N$? What if $N = 2$? For instance, do there exist complete proper holomorphic embeddings of the unit disk $\mathbb{D} \subset \mathbb{C}$ into the unit ball $\mathbb{B} \subset \mathbb{C}^2$?

The first approach to this problem was recently made in [Alarcón et al. 2016] by Alarcón, Globevnik, and López, who, with a conceptually new method based on the use of holomorphic automorphisms of $\mathbb{C}^N$, constructed complete closed complex hypersurfaces in the unit ball of $\mathbb{C}^N$ for any $N > 1$ with certain restrictions on their topology. In particular, for $N = 2$, they showed that the unit ball of $\mathbb{C}^2$ admits complete properly embedded complex curves with arbitrary *finite* topology, see Corollary 1.2 of that paper, thereby affirmatively answering the third question in Problem 1.1. Going further in this line, Globevnik [2016a] proved the existence of complete proper holomorphic embeddings $\mathbb{D} \hookrightarrow \mathbb{B}$ whose image contains any given closed discrete subset of $\mathbb{B}$. This is reminiscent of an old result by Forstnerič, Globevnik, and Stensønes [Forstnerič et al. 1996] asserting that, given a pseudoconvex Runge domain $D \subset \mathbb{C}^N$ $(N > 1)$ and a closed discrete subset $\Lambda \subset D$, there is a proper holomorphic embedding $\mathbb{D} \hookrightarrow D$ whose image contains Λ. It is nevertheless true that these embeddings are not ensured to be complete in any case.

The aim of this paper is to settle Problem 1.1 for $N = 2$ by proving the existence of complete properly embedded complex curves in $\mathbb{B}$ with *arbitrary topology* (possibly infinite). This completely solves the problem and, in particular, generalizes the existence result [Alarcón et al. 2016, Corollary 1.2], which only deals with finite topological types. Moreover, we provide examples of such curves which contain any given closed discrete subset of $\mathbb{B}$, thereby extending the above-mentioned hitting result by Globevnik [2016a, Theorem 1.1].

The main theorem of this paper can be stated as follows.

Theorem 1.2. *Let* Λ *be a closed discrete subset of the unit ball* $\mathbb{B} \subset \mathbb{C}^2$. *On each open connected orientable smooth surface* M *there exists a complex structure such that the open Riemann surface* M *admits a complete proper holomorphic embedding* $M \hookrightarrow \mathbb{B}$ *whose image contains* Λ.

It is perhaps worth mentioning that, choosing any closed discrete subset $\Lambda \subset \mathbb{B}$ such that $\bar{\Lambda} \setminus \Lambda = b\mathbb{B} = \{\zeta \in \mathbb{C}^2 : |\zeta| = 1\}$, Theorem 1.2 trivially implies the following:

Corollary 1.3. *The unit ball $\mathbb{B} \subset \mathbb{C}^2$ contains complete properly embedded complex curves with any given topology and whose limit set equals* $b\mathbb{B}$.

Although it is not explicitly stated there, Corollary 1.3 in the simply connected case straightforwardly follows from the results by Globevnik [2016a].

Our method of proof exploits some ideas from both [Alarcón et al. 2016] and [Globevnik 2016a] (in particular, our construction technique is based on the use of holomorphic automorphisms of $\mathbb{C}^2$), but also from [Alarcón and López 2013b], where the authors constructed properly embedded complex curves in $\mathbb{C}^2$ with arbitrary topology. The latter contributes to the so-called embedding problem for open Riemann surfaces in $\mathbb{C}^2$, a long-standing open question in Riemann surface theory asking whether every open Riemann surface properly embeds in $\mathbb{C}^2$ as a complex curve [Bell and Narasimhan 1990, Conjecture 3.7, page 20]; for recent advances and a history of this classical problem we refer to [Forstnerič and Wold 2009; 2013]. It is shown in [Forstnerič and Wold 2009] that given a compact bordered Riemann surface $\overline{M} = M \cup bM$ admitting a smooth embedding $f : \overline{M} \hookrightarrow \mathbb{C}^2$ which is holomorphic in M, there is a proper holomorphic embedding $\tilde{f} : M \hookrightarrow \mathbb{C}^2$ which is as close as desired to f uniformly on a given compact subset of M. (A *compact bordered Riemann surface* $\overline{M}$ is a compact Riemann surface with boundary $\varnothing \neq bM \subset \overline{M}$ consisting of finitely many pairwise disjoint smooth Jordan curves; its interior $M = \overline{M} \setminus bM$ is called a *bordered Riemann surface*.) This fact and the arguments in its proof were key in the construction method in [Alarcón and López 2013b], which will be used in the proof of Theorem 1.2.

We strongly expect that the new construction techniques developed in this paper may be adapted to prove the statement of Theorem 1.2 but replacing the ball $\mathbb{B}$ by any convex domain of $\mathbb{C}^2$. The following questions, concerning pseudoconvex domains, remain open and seem to be much more challenging.

Problem 1.4. Let $D \subset \mathbb{C}^2$ be a pseudoconvex Runge domain. Does there exist a complete proper holomorphic embedding $\mathbb{D} \hookrightarrow D$? Given a closed discrete subset $\Lambda \subset D$, do there exist complete properly embedded complex curves in D containing Λ?

As we have already mentioned, every bordered Riemann surface M admits a complete proper holomorphic *immersion* $M \to \mathbb{B}$ [Alarcón and Forstnerič 2013], and if in addition there is a smooth embedding $\overline{M} \to \mathbb{C}^2$ that is holomorphic in M, then M properly holomorphically embeds into $\mathbb{C}^2$ [Forstnerič and Wold 2009]. It is however an open question, likely very difficult, whether every bordered Riemann surface admits a holomorphic embedding in $\mathbb{C}^2$ (even without requiring the embedding to have any global condition such as completeness or properness); see, e.g., the introduction of [Forstnerič and Wold 2009] or Section 8.9 in the monograph [Forstnerič 2011] for more information. Thus, one is also led to ask:

Problem 1.5. Let $\overline{M} = M \cup bM$ be a compact bordered Riemann surface and assume that there is a smooth embedding $\overline{M} \hookrightarrow \mathbb{C}^2$ which is holomorphic in M. Does M admit complete holomorphic embeddings $M \hookrightarrow \mathbb{C}^2$ with bounded image?

We hope to return to these interesting questions in a future work.

Organization of the paper. In Section 2 we set the notation that will be used throughout the paper and, with the aim of making it self-contained, state some already known results which will be used in the proof of Theorem 1.2. In Section 3 we prove an approximation result by properly embedded complex curves in the unit ball $\mathbb{B} \subset \mathbb{C}^2$ (see Theorem 3.1) from which Theorem 1.2 will be easily derived.

2. Preliminaries

We denote by $|\cdot|$ and $\mathbb{B} = \{\zeta \in \mathbb{C}^2 : |\zeta| < 1\}$ the Euclidean norm and the unit Euclidean ball in $\mathbb{C}^2$. For a subset $C \subset \mathbb{C}^2$ we denote by $\bar{C}$, $\mathring{C}$, and $\mathrm{b}C = \bar{C} \setminus \mathring{C}$ the topological closure, interior, and frontier of C in $\mathbb{C}^2$, respectively. Also, given a point $\xi \in \mathbb{C}^2$ and a number $r \in \mathbb{R}_+ = [0, +\infty[$ we write $\xi + rC = \{\xi + r\zeta : \zeta \in C\}$.

Let A be a smoothly bounded compact domain in an open Riemann surface and let $k \in \mathbb{Z}_+ = \{0, 1, 2, \dots\}$. We denote by $\mathscr{A}^k(A)$ the space of functions $A \to \mathbb{C}$ of class $\mathscr{C}^k(A)$ which are holomorphic on the interior $\mathring{A} = \bar{A} \setminus \mathrm{b}A$. If $N \in \mathbb{N}$ we will simply write $\mathscr{A}^k(A)$ instead of

$$\mathscr{A}^k(A)^N = \mathscr{A}^k(A) \times \overset{N \text{ times}}{\cdots} \times \mathscr{A}^k(A)$$

when there is no place for ambiguity. Given an immersion $\psi : A \to \mathbb{C}^N$ $(N > 1)$ of class $\mathscr{C}^1(A)$, we denote by $\mathrm{dist}_\psi : A \times A \to \mathbb{R}_+$ the Riemannian distance in A induced by the Euclidean metric of $\mathbb{C}^N$ via ψ; that is,

$$\mathrm{dist}_\psi(p, q) := \inf\{\mathrm{length}(\psi(\gamma)) : \gamma \subset A \text{ path connecting } p \text{ and } q\}, \qquad p, q \in A,$$

where $\mathrm{length}(\cdot)$ denotes the Euclidean length in $\mathbb{C}^N$.

Tangent balls. Given a point $\zeta \in \mathbb{C}^2 \setminus \{0\}$ and a number $r > 0$, we denote by $\mathscr{T}(\zeta, r)$ the closed ball with center ζ and radius r in the real affine hyperplane H_ζ tangent to the sphere $\mathrm{b}(|\zeta|\mathbb{B})$ at the point ζ; that is,

$$\mathscr{T}(\zeta, r) := \{\xi \in H_\zeta : |\xi - \zeta| \le r\} \subset \mathbb{C}^2.$$

According to [Alarcón et al. 2016, Definition 1.3], the set $\mathscr{T}(\zeta, r)$ above is called *the tangent ball of center ζ and radius r*. A collection $\mathfrak{F} = \{\mathscr{T}(\zeta_j, r_j)\}_{j \in J}$ of tangent balls in $\mathbb{B} \subset \mathbb{C}^2$ will be called *tidy*, see Definition 1.4 of the same paper, if it satisfies the following requirements:

- $\mathscr{T}(\zeta_j, r_j) \subset \mathbb{B}$ for all $j \in J$ and the tangent balls in $\mathfrak{F}$ are pairwise disjoint.
- $t\overline{\mathbb{B}}$ intersects finitely many tangent balls in the family $\mathfrak{F}$ for all $0 < t < 1$.
- If $\mathscr{T}(\zeta, r)$, $\mathscr{T}(\zeta', r') \in \mathfrak{F}$ and $|\zeta| = |\zeta'|$, then $r = r'$.
- If $\mathscr{T}(\zeta, r)$, $\mathscr{T}(\zeta', r') \in \mathfrak{F}$ and $|\zeta| < |\zeta'|$, then $\mathscr{T}(\zeta, r) \subset |\zeta'|\mathbb{B}$.

Notice that a tidy collection $\mathfrak{F} = \{\mathscr{T}(\zeta_j, r_j)\}_{j \in J}$ of tangent balls in $\mathbb{B}$ consists of countably many elements; and so we may assume that $J \subset \mathbb{N} = \{1, 2, 3, \dots\}$. We denote by

$$|\mathfrak{F}| := \bigcup_{j \in J} \mathscr{T}(\zeta_j, r_j)$$

the union of all the tangent balls in a tidy collection. Note that if J is finite then $|\mathfrak{F}|$ is compact, whereas if J is infinite then $|\mathfrak{F}|$ is a proper subset of $\mathbb{B}$.

The following two results, involving tidy collections of tangent balls, are proved in [Alarcón et al. 2016], and will be invoked in our argumentation.

Lemma 2.1 [Alarcón et al. 2016, Lemma 2.4]. *Given numbers $0 < r < r' < 1$ and $\ell > 0$ there is a finite tidy collection $\mathfrak{F}$ of tangent balls in $\mathbb{B}$ such that $|\mathfrak{F}| \subset r'\mathbb{B} \setminus r\overline{\mathbb{B}}$ and the length of any path $\gamma : [0, 1] \to r'\overline{\mathbb{B}} \setminus ((r\mathbb{B}) \cup |\mathfrak{F}|)$ with $|\gamma(0)| = r$ and $|\gamma(1)| = r'$ is at least ℓ.*

The next result is not explicitly stated in [Alarcón et al. 2016] but straightforwardly follows from an inspection of the proof of Theorem 1.6 of that paper as a standard finite recursive application of their Lemma 3.1.

Lemma 2.2. *Let $0 < r < r' < 1$ be numbers and let $\mathfrak{F}$ be a finite tidy collection of tangent balls in $\mathbb{B}$ with $|\mathfrak{F}| \subset r'\mathbb{B} \setminus r\overline{\mathbb{B}}$. Then, given a properly embedded complex curve $Z \subset \mathbb{C}^2$ and a number $\epsilon > 0$, there is a holomorphic automorphism $\Phi : \mathbb{C}^2 \to \mathbb{C}^2$ satisfying the following properties*:

(i) $\Phi(Z) \cap |\mathfrak{F}| = \varnothing$.

(ii) $|\Phi(\zeta) - \zeta| < \epsilon$ *for all* $\zeta \in r\overline{\mathbb{B}}$.

Hitting and approximation lemmas. In this subsection we state two results which will also be used in our construction. The first one, due to Globevnik, will be key in order to achieve the hitting condition in the statement of Theorem 1.2.

Lemma 2.3 [Globevnik 2016a, Lemma 7.2]. *Given a finite subset $\Lambda \subset \mathbb{B}$ there exist numbers $\eta > 0$ and $\mu > 0$ such that the following holds. Given a number $0 < \delta < \eta$ and a map $\varphi : \Lambda \to \mathbb{C}^2$ such that*

$$|\varphi(\zeta) - \zeta| < \delta \quad \textit{for all } \zeta \in \Lambda,$$

there exists a holomorphic automorphism $\Psi : \mathbb{C}^2 \to \mathbb{C}^2$ satisfying the following conditions:

(i) $\Psi(\varphi(\zeta)) = \zeta$ *for all* $\zeta \in \Lambda$,

(ii) $|\Psi(\zeta) - \zeta| < \mu\delta$ *for all* $\zeta \in \overline{\mathbb{B}}$.

The second result, which will help us to increase the topology in the recursive construction, is a particular case of [Alarcón and López 2013b, Theorem 4.5]; it also easily follows from the results in [Forstnerič and Wold 2009].

Lemma 2.4. *Let $\overline{M} = M \cup \mathrm{b}M$ be a compact bordered Riemann surface and let $K \subset M$ be a connected, smoothly bounded, Runge compact domain which is a strong deformation retract of $\overline{M}$. Let $\phi : K \hookrightarrow \mathbb{C}^2$ be an embedding of class $\mathscr{A}^1(K)$ and assume that there exists a number $s > 0$ such that*

$$\phi(bK) \cap s\overline{\mathbb{B}} = \varnothing.$$

Then, given $\epsilon > 0$, there are an open domain $\Omega \subset M$ and a proper holomorphic embedding $\tilde{\phi} : \Omega \hookrightarrow \mathbb{C}^2$ such that:

(i) *$K \subset \Omega$ and Ω is a deformation retract of M and homeomorphic to M.*

(ii) $|\tilde{\phi}(p) - \phi(p)| < \epsilon$ *for all* $p \in K$.

(iii) $\tilde{\phi}(\Omega \setminus \mathring{K}) \cap s\overline{\mathbb{B}} = \varnothing$.

Recall that a compact subset K of an open Riemann surface M is said to be *Runge* or *holomorphically convex* if every continuous function $K \to \mathbb{C}$, holomorphic in $\mathring{K}$, may be approximated uniformly on K by holomorphic functions $M \to \mathbb{C}$. The classical Mergelyan theorem ensures that $K \subset M$ is Runge if and only if $M \setminus K$ has no relatively compact connected components in M; see [Bishop 1958; Runge 1885; Mergelyan 1951].

3. Proof of the main theorem

In this section we prove the following more precise version of Theorem 1.2.

Theorem 3.1. *Let M be an open connected Riemann surface, let $K \subset M$ be a connected, smoothly bounded, Runge compact domain, let $0 < s < r < 1$ be numbers, and assume that there is an embedding $\psi : K \to \mathbb{C}^2$ of class $\mathscr{A}^1(K)$ such that*

$$\psi(\mathsf{b}K) \subset r\mathbb{B} \setminus s\overline{\mathbb{B}}. \tag{3-1}$$

Let $\Lambda \subset \mathbb{B}$ be a closed discrete subset such that

$$\Lambda \cap r\overline{\mathbb{B}} \subset \psi(\mathring{K}) \cap s\mathbb{B}. \tag{3-2}$$

Then, given $\epsilon > 0$, there are a domain $D \subset M$ and a complete proper holomorphic embedding $\tilde{\psi} : D \hookrightarrow \mathbb{B}$ satisfying the following properties:

(i) *$K \subset D$ and D is a deformation retract of M (and hence homeomorphic to M).*

(ii) *$|\tilde{\psi}(\zeta) - \psi(\zeta)| < \epsilon$ for all $\zeta \in K$.*

(iii) *$\Lambda \subset \tilde{\psi}(D)$.*

(iv) *$\tilde{\psi}(D \setminus K) \cap s\overline{\mathbb{B}} = \varnothing$.*

Proof. Pick a number $s < s_0' < r$ such that

$$\psi(\mathsf{b}K) \subset r\mathbb{B} \setminus s_0'\overline{\mathbb{B}}. \tag{3-3}$$

Such an s_0' exists in view of (3-1) by compactness of $\psi(\mathsf{b}K)$. Without loss of generality we assume that Λ is infinite, and hence, since it is closed in $\mathbb{B}$ and discrete, Λ is in bijection with $\mathbb{N}$ and for any ordering $\Lambda = \{p_i\}_{i\in\mathbb{N}}$ of Λ we have $\lim_{i\to\infty} |p_i| = 1$. Thus, there are sequences of numbers $\{r_j\}_{j\in\mathbb{N}}$, with $r_1 = r$, $\{r_j'\}_{j\in\mathbb{N}}$, $\{s_j\}_{j\in\mathbb{N}}$, and $\{s_j'\}_{j\in\mathbb{N}}$, satisfying the following properties:

(a) $s_{j-1}' < r_j < r_j' < s_j < s_j'$ for all $j \in \mathbb{N}$.

(b) $\lim_{j\to\infty} r_j = 1$.

(c) $\Lambda \subset s\mathbb{B} \cup \big(\bigcup_{j\in\mathbb{N}} (s_j'\mathbb{B} \setminus s_j\overline{\mathbb{B}})\big)$, taking into account (3-2).

Write $\Lambda_0 := \Lambda \cap s\mathbb{B} = \Lambda \cap s_0'\mathbb{B}$ and $\Lambda_j = \Lambda \cap (s_j'\mathbb{B} \setminus s_j\overline{\mathbb{B}})$, $j \in \mathbb{N}$, and observe that Λ_j is finite for all $j \in \mathbb{Z}_+$. We also assume without loss of generality that $\Lambda_j \neq \varnothing$ for all $j \in \mathbb{Z}_+$. By (a) and (c) we have

$$\Lambda = \bigcup_{j\in\mathbb{Z}_+} \Lambda_j, \qquad \Lambda_i \cap \Lambda_j = \varnothing \ \text{ for all } i, j \in \mathbb{Z}_+,\ i \neq j. \tag{3-4}$$

Let $\{\epsilon_j\}_{j\in\mathbb{N}} \searrow 0$ be a decreasing sequence of positive numbers such that

$$\sum_{j\in\mathbb{N}} \epsilon_j < \epsilon. \tag{3-5}$$

The precise values of the numbers $\epsilon_j,\ j \in \mathbb{N}$, will be specified later.

Write $M_0 := K$ and let

$$M_0 \Subset M_1 \Subset M_2 \Subset \cdots \Subset M = \bigcup_{j\in\mathbb{Z}_+} M_j \tag{3-6}$$

be an exhaustion of M by connected, smoothly bounded, Runge compact domains such that the Euler characteristic satisfies $\chi(M_j \setminus \mathring{M}_{j-1}) \in \{0, -1\}$ for all $j \in \mathbb{N}$. Such an exhaustion exists by basic topological arguments; see, e.g., [Alarcón and López 2013b, Lemma 4.2] for a simple proof.

Write $D_0 := M_0 = K$ and set $\psi_0 := \psi$. Fix a point $p_0 \in \mathring{K}$. We shall recursively construct a sequence $\Xi_j = \{D_j, \psi_j\},\ j \in \mathbb{N}$, where D_j is a connected, smoothly bounded, Runge compact domain in M and $\psi_j : D_j \hookrightarrow \mathbb{C}^2$ is an embedding of class $\mathscr{A}^1(D_j)$ such that the following hold for all $j \in \mathbb{N}$:

(1_j) $D_{j-1} \Subset D_j \Subset M_j$ and D_j is a strong deformation retract of M_j.

(2_j) $|\psi_j(p) - \psi_{j-1}(p)| < \epsilon_j$ for all $p \in D_{j-1}$.

(3_j) $\mathrm{dist}_{\psi_j}(p_0, bD_j) > j$.

(4_j) $\psi_j(bD_j) \subset r_{j+1}\mathbb{B} \setminus s'_j\overline{\mathbb{B}}$.

(5_j) $\psi_j(D_j \setminus \mathring{D}_{j-1}) \cap s'_{j-1}\overline{\mathbb{B}} = \varnothing$.

(6_j) $\Lambda_i \subset \psi_j(\mathring{D}_i)$ for all $i \in \{0, \dots, j\}$.

Assume for a moment that we have already constructed a sequence $\{\Xi_j\}_{j\in\mathbb{N}}$ enjoying the above conditions. By (1_j) and (3-6), we obtain that

$$D := \bigcup_{j\in\mathbb{Z}_+} D_j$$

is a domain in M satisfying property (i) in the statement of the theorem. We claim that, if the number $\epsilon_j > 0$ is chosen small enough at each step in the recursive construction, the sequence $\{\psi_j\}_{j\in\mathbb{Z}_+}$ converges uniformly on compacta in D to a limit map

$$\tilde{\psi} = \lim_{j\to\infty} \psi_j : D \to \mathbb{C}^2$$

satisfying the conclusion of the theorem. Indeed, by (2_j) and (3-5) the limit map $\tilde{\psi}$ exists and satisfies (ii); recall that $K = D_0$. Since each map ψ_j is an embedding of class $\mathscr{A}^1(D_j)$, choosing the number $\epsilon_j > 0$ sufficiently small at each step in the recursive construction, we have $\tilde{\psi} : D \to \mathbb{C}^2$ is an injective holomorphic immersion. Moreover, (3_j) ensures that $\tilde{\psi}$ is complete, whereas (5_j), (a), and (b) guarantee (iv) and the facts that $\tilde{\psi}(D) \subset \mathbb{B}$ and that $\tilde{\psi} : D \to \mathbb{B}$ is a proper map; recall that $s < s'_0$ and observe that $\lim_{j\to\infty} s'_j = 1$. Thus, $\tilde{\psi} : D \hookrightarrow \mathbb{B}$ is a proper embedding. Finally, (6_j) and (3-4) imply that $\Lambda \subset \tilde{\psi}(D)$, thereby proving (iii).

To conclude the proof, it therefore suffices to construct a sequence $\Xi_j = \{D_j, \psi_j\},\ j \in \mathbb{N}$, satisfying properties (1_j)–(6_j) above. We proceed by induction. The basis is given by the compact domain D_0

and the map ψ_0; observe that (3_0) holds since $p_0 \in \mathring{D}_0$ and ψ_0 is an immersion, $(4_0) = $ (3-3) (recall that $r_1 = r$), (6_0) is implied by (c) and (3-2), and (2_0), (5_0), and the first part of (1_0) are vacuous conditions. Finally, the second part of (1_0) is obvious since $D_0 = M_0$.

For the inductive step, assume that we have $\Xi_{j-1} = \{D_{j-1}, \psi_{j-1}\}$ enjoying the desired properties for some $j \in \mathbb{N}$ and let us construct $\Xi_j = \{D_j, \psi_j\}$. We distinguish two cases.

Case 1: Assume that $\chi(M_j \setminus \mathring{M}_{j-1}) = 0$. In this case M_{j-1} is a strong deformation retract of M_j.

Write $\Lambda' = \bigcup_{i=0}^{j} \Lambda_i$ and let $\eta > 0$ and $\mu > 0$ be the numbers given by Lemma 2.3 applied to the finite subset $\Lambda' \subset \mathbb{B}$.

By (3_{j-1}) and (4_{j-1}) there is another number $\eta' > 0$ with the following property:

(A1) If $\phi : D_{j-1} \to \mathbb{C}^2$ is an immersion of class $\mathscr{A}^1(D_{j-1})$ such that $|\phi(p) - \psi_{j-1}(p)| < \eta'$ for all $p \in D_{j-1}$, then $\operatorname{dist}_\phi(p_0, \mathrm{b}D_{j-1}) > j - 1$ and $\phi(\mathrm{b}D_{j-1}) \subset r_j \mathbb{B} \setminus s'_{j-1}\overline{\mathbb{B}}$.

On the other hand, Lemma 2.1 gives a finite tidy collection $\mathfrak{F}$ of tangent balls in $\mathbb{B}$ satisfying the following conditions:

(A2) $|\mathfrak{F}| \subset r'_j \mathbb{B} \setminus r_j \overline{\mathbb{B}}$.

(A3) The length of any path $\gamma : [0, 1] \to r'_j \overline{\mathbb{B}} \setminus ((r_j \mathbb{B}) \cup |\mathfrak{F}|)$ with $|\gamma(0)| = r_j$ and $|\gamma(1)| = r'_j$ is greater than 2.

Thus, there is a third number $\eta'' > 0$ enjoying the following property:

(A4) If $\alpha : [0, 1] \to r'_j \overline{\mathbb{B}} \setminus r_j \mathbb{B}$ is a path satisfying that there is another path $\gamma : [0, 1] \to r'_j \overline{\mathbb{B}} \setminus ((r_j \mathbb{B}) \cup |\mathfrak{F}|)$ such that $|\gamma(0)| = r_j$, $|\gamma(1)| = r'_j$, and $|\gamma(x) - \alpha(x)| < \eta''$ for all $x \in [0, 1]$, then the length of α is greater than 1.

Next, pick a number t such that $s'_{j-1} < t < r_j$ and

$$\psi_{j-1}(\mathrm{b}D_{j-1}) \subset r_j \mathbb{B} \setminus t\overline{\mathbb{B}}; \tag{3-7}$$

the existence of such number t is ensured by (a) and (4_{j-1}). Finally, choose a number

$$0 < \delta < \min\left\{\eta, \frac{\eta'}{\mu+1}, \frac{\eta''}{\mu+1}, \frac{\epsilon_j}{\mu+1}, \frac{t - s'_{j-1}}{\mu}, \frac{r_{j+1} - s'_j}{2\mu}\right\}. \tag{3-8}$$

Also fix a number $\tau_1 > 0$ which will be specified later.

Taking into account (3-7), Lemma 2.4 furnishes an open domain $\Omega \Subset \mathring{M}_j$ and a proper holomorphic embedding $\phi_1 : \Omega \hookrightarrow \mathbb{C}^2$ such that the following hold:

(B1) $D_{j-1} \subset \Omega$ and Ω is a deformation retract of $\mathring{M}_j$ and is homeomorphic to $\mathring{M}_j$. In particular, the second part of (1_{j-1}) and the fact that M_{j-1} is a strong deformation retract of M_j ensure that $\Omega \setminus D_{j-1}$ consists of a finite collection of pairwise disjoint open annuli.

(B2) $|\phi_1(p) - \psi_{j-1}(p)| < \tau_1$ for all $p \in D_{j-1}$.

(B3) $\phi_1(\Omega \setminus \mathring{D}_{j-1}) \cap t\overline{\mathbb{B}} = \varnothing$.

In view of (3-7) and (B2), and choosing $\tau_1 > 0$ small enough, we also have:

(B4) $\phi_1(\mathrm{b}D_{j-1}) \subset r_j \mathbb{B} \setminus t\overline{\mathbb{B}}$.

Now, given a number $\tau_2 > 0$ which will be specified later and taking into account (A2), Lemma 2.2 provides a holomorphic automorphism $\Phi : \mathbb{C}^2 \to \mathbb{C}^2$ such that

(C1) $\Phi(\phi_1(\Omega)) \cap |\mathfrak{F}| = \varnothing$.

(C2) $|\Phi(\zeta) - \zeta| < \tau_2$ for all $\zeta \in r_j \overline{\mathbb{B}}$.

Write

$$\phi_2 := \Phi \circ \phi_1 : \Omega \hookrightarrow \mathbb{C}^2. \tag{3-9}$$

By (B1), (B3), (B4), (C2), and (a), and choosing $\tau_2 > 0$ sufficiently small, we have:

(D1) $\phi_2(\mathrm{b}D_{j-1}) \subset r_j \mathbb{B} \setminus t\overline{\mathbb{B}}$.

(D2) $\phi_2(\Omega \setminus \mathring{D}_{j-1}) \cap t\overline{\mathbb{B}} = \varnothing$.

and, taking into account (B1), (D1), and the maximum principle for the function $|\phi_2|$, there exists a smoothly bounded compact domain $\Upsilon \subset \Omega$ such that:

(D3) $D_{j-1} \Subset \Upsilon$ and D_{j-1} is a strong deformation retract of Υ.

(D4) $\phi_2(\mathrm{b}\Upsilon) \subset \mathrm{b}(t'\mathbb{B})$ and meets transversely there for some number t' with $r'_j < t' < s_j$.

Next, choose a point $q_0 \in \mathrm{b}\Upsilon$ and a smooth embedded compact path $\lambda \subset \mathbb{C}^2 \setminus t'\mathbb{B}$ having $\phi_2(q_0)$ as an endpoint, meeting $\mathrm{b}(t'\mathbb{B})$ transversely there, and being otherwise disjoint from $t'\overline{\mathbb{B}}$. Assume also that

$$\Lambda_j \subset \lambda \subset s'_j \mathbb{B} \setminus t'\mathbb{B}. \tag{3-10}$$

Also take a smooth embedded compact path $\gamma \subset \Omega \setminus \mathring{\Upsilon}$ having q_0 as an endpoint, meeting $\mathrm{b}\Upsilon$ transversely there, and being otherwise disjoint from Υ. Extend ϕ_2, with the same name, to a smooth embedding $\Upsilon \cup \gamma \hookrightarrow \mathbb{C}^2$ such that

$$\phi_2(\gamma) = \lambda. \tag{3-11}$$

Observe that $\Upsilon \cup \gamma$ is a Runge compact subset of Ω; take into account (B1) and (D3). Thus, given $\tau_3 > 0$ to be specified later, Mergelyan's theorem applied to $\phi_2 : \Upsilon \cup \gamma \hookrightarrow \mathbb{C}^2$ ensures the existence of a smoothly bounded compact domain $\Upsilon' \subset \Omega$ and an embedding $\phi_3 : \Upsilon' \hookrightarrow \mathbb{C}^2$ of class $\mathscr{A}^1(\Upsilon')$ such that:

(E1) $\Upsilon \cup \gamma \subset \mathring{\Upsilon}'$ and Υ' is a strong deformation retract of M_j. In particular, D_{j-1} is a strong deformation retract of Υ'; see (D3).

(E2) $|\phi_3(p) - \phi_2(p)| < \tau_3$ for all $p \in \Upsilon \cup \gamma$.

Furthermore, if we take Υ' close enough to $\Upsilon \cup \gamma$ and if $\tau_3 > 0$ is chosen sufficiently small, we obtain in view of (E2) that:

(E3) $\phi_3(\Upsilon' \setminus \mathring{\Upsilon}) \subset s'_j \mathbb{B} \setminus r'_j \overline{\mathbb{B}}$. See (D4), (a), (3-10), and (3-11).

(E4) $\phi_3(\mathrm{b}D_{j-1}) \subset r_j \mathbb{B} \setminus t\overline{\mathbb{B}}$. See (D1) and (D3).

(E5) $\phi_3(\Upsilon' \setminus \mathring{D}_{j-1}) \cap t\overline{\mathbb{B}} = \varnothing$. See (D2).

(E6) $\phi_3(\Upsilon') \cap |\mathfrak{F}| = \varnothing$. See (3-9), (C1), (E3), and (A2).

Given $\tau_4 > 0$ to be specified later, applying Lemma 2.4 once again, we obtain, in view of (E3), a domain $\Omega' \Subset M_j$ and a proper holomorphic embedding $\phi_4 : \Omega' \hookrightarrow \mathbb{C}^2$ such that:

(F1) $\Upsilon' \subset \Omega'$ and Ω' is a deformation retract of $\mathring{M}_j$ and homeomorphic to $\mathring{M}_j$. In particular, $\Omega' \setminus \Upsilon'$ consists of a finite collection of pairwise disjoint open annuli. See (B1).

(F2) $|\phi_4(p) - \phi_3(p)| < \tau_4$ for all $p \in \Upsilon'$.

(F3) $\phi_4(\Omega' \setminus \mathring{\Upsilon}') \cap r'_j \overline{\mathbb{B}} = \varnothing$.

If $\tau_4 > 0$ is chosen sufficiently small then, in view of (F2), we also have:

(F4) $\phi_4(bD_{j-1}) \subset r_j \mathbb{B} \setminus t\overline{\mathbb{B}}$. See (E4).

(F5) $\phi_4(\Omega' \setminus \mathring{D}_{j-1}) \cap t\overline{\mathbb{B}} = \varnothing$. See (F3) and (E5), and recall that $t < r_j < r'_j$.

(F6) $\phi_4(\Omega') \cap |\mathfrak{F}| = \varnothing$. See (E6), (F3), and (A2).

Assume now that the numbers τ_i, $i = 1, \dots, 4$, are chosen small enough such that $\sum_{i=1}^4 \tau_i < \delta$, where $\delta > 0$ is the number in (3-8). Thus, by (B2), (C2), (3-9), (E2), and (F2), we have:

(F7) $|\phi_4(p) - \psi_{j-1}(p)| < \sum_{i=1}^4 \tau_i < \delta$ for all $p \in D_{j-1}$.

(F8) $|\phi_4(p) - \phi_2(p)| < \tau_3 + \tau_4 < \sum_{i=1}^4 \tau_i < \delta$ for all $p \in \Upsilon \cup \gamma$.

By (6_{j-1}), (c), (3-10), and (3-11), we infer that $\Lambda_i \subset \psi_{j-1}(\mathring{D}_i)$, $i = 0, \dots, j-1$, and $\Lambda_j \subset \phi_2(\gamma)$. Thus, (F7) and (F8) guarantee the existence of an injective map $\varphi : \Lambda' \to \phi_4(\Omega') \subset \mathbb{C}^2$ such that:

(G1) $|\varphi(\zeta) - \zeta| < \delta$ for all $\zeta \in \Lambda'$.

(G2) $\varphi(\Lambda_i) \subset \phi_4(\mathring{D}_i)$ for all $i \in \{0, \dots, j-1\}$.

(G3) $\varphi(\Lambda_j) \subset \phi_4(\gamma) \subset \phi_4(\mathring{\Upsilon}' \setminus \Upsilon)$.

Observe that, since $0 < \delta < \eta$ where η is given by Lemma 2.3 for the subset $\Lambda' \subset \mathbb{B}$, see (3-8), every map $\varphi : \Lambda' \to \mathbb{C}^2$ satisfying (G1) is injective. In view of (3-8) and (G1), Lemma 2.3 provides a holomorphic automorphism $\Psi : \mathbb{C}^2 \to \mathbb{C}^2$ such that:

(H1) $\Psi(\varphi(\zeta)) = \zeta$ for all $\zeta \in \Lambda'$.

(H2) $|\Psi(\zeta) - \zeta| < \mu\delta$ for all $\zeta \in \overline{\mathbb{B}}$, where μ is the number given by Lemma 2.3 for the set Λ' which appears in (3-8).

Consider the proper holomorphic embedding

$$\phi := \Psi \circ \phi_4 : \Omega' \hookrightarrow \mathbb{C}^2.$$

Properties (F4), (F7), (H2), and (3-8), together with the maximum principle, ensure that

$$|\phi(p) - \psi_{j-1}(p)| < (\mu + 1)\delta < \min\{\eta', \eta'', \epsilon_j\} \quad \text{for all } p \in D_{j-1}. \tag{3-12}$$

This inequality and (A1) guarantee that

$$\operatorname{dist}_\phi(p_0, bD_{j-1}) > j - 1 \tag{3-13}$$

and

$$\phi(bD_{j-1}) \subset r_j \mathbb{B} \setminus s'_{j-1}\overline{\mathbb{B}}. \tag{3-14}$$

Furthermore, since $\phi_4 : \Omega' \hookrightarrow \mathbb{C}^2$ is a proper holomorphic embedding, (F4), (F1), and the maximum principle ensure the existence of a smoothly bounded compact domain $D_j \subset \mathring{M}_j$ satisfying condition (1_j) and such that

$$\phi_4(\mathrm{b}D_j) \subset (r_{j+1} - \mu\delta)\mathbb{B} \setminus (s'_j + \mu\delta)\overline{\mathbb{B}} \subset \overline{\mathbb{B}}; \tag{3-15}$$

observe that $r_{j+1} - \mu\delta > s'_j + \mu\delta > r_j$ by (3-8) and (a). Thus, (H2) gives that

$$\phi(\mathrm{b}D_j) \subset r_{j+1}\mathbb{B} \setminus s'_j\overline{\mathbb{B}}. \tag{3-16}$$

Moreover, (3-15) and the maximum principle imply that $\phi_4(D_j) \subset (r_{j+1} - \mu\delta)\mathbb{B} \subset \overline{\mathbb{B}}$, and hence (H2) gives that

$$|\phi(p) - \phi_4(p)| < \mu\delta \quad \text{for all } p \in D_j. \tag{3-17}$$

Set

$$\psi_j := \phi|_{D_j} : D_j \to \mathbb{C}^2.$$

We claim that the pair $\Xi_j = \{D_j, \psi_j\}$ satisfies conditions (1_j)–(6_j). Indeed, (1_j) has been already checked, (2_j) is implied by (3-12), and $(4_j) = $ (3-16). On the other hand, since $\phi_4(D_j) \subset \mathbb{B}$, properties (F5) and (H2) ensure that $\varnothing = \psi_j(D_j \setminus \mathring{D}_{j-1}) \cap (t - \mu\delta)\overline{\mathbb{B}} \supset \psi_j(D_j \setminus \mathring{D}_{j-1}) \cap s'_{j-1}\overline{\mathbb{B}}$, where the latter inclusion follows from the fact that $s'_{j-1} < t - \mu\delta$; see (3-8). This proves (5_j). Further, (G2) and (H1) guarantee that $\Lambda_i \subset \psi_j(\mathring{D}_i)$ for all $i \in \{0, \ldots, j-1\}$, whereas (G3), (H1), and (3-16) ensure that $\Lambda_j \subset \psi_j(\mathring{D}_j)$; thereby proving (6_j). Finally, (3-14), (3-16), (3-17), (F6), and (A4) guarantee that $\mathrm{dist}_{\psi_j}(\mathrm{b}D_{j-1}, \mathrm{b}D_j) > 1$; take into account that $s'_j > r'_j$ in view of (a). This and (3-13) prove (3_j); recall that $p_0 \in \mathring{D}_{j-1} \Subset D_j$.

This concludes the construction of Ξ_j in Case 1.

Case 2: Assume that $\chi(M_j \setminus \mathring{M}_{j-1}) = -1$. In this case there is a smooth embedded compact path $\gamma \subset \mathring{M}_j \setminus \mathring{D}_{j-1}$ with its two endpoints lying in $\mathrm{b}D_{j-1}$, meeting transversely there, and otherwise disjoint from D_{j-1}, such that $D_{j-1} \cup \gamma$ is a strong deformation retract of M_j.

By (4_{j-1}) we may extend ψ_{j-1}, with the same name, to a smooth embedding $D_{j-1} \cup \gamma \hookrightarrow \mathbb{C}^2$ such that $\psi_{j-1}(\gamma) \subset r_j\mathbb{B} \setminus s'_{j-1}\overline{\mathbb{B}}$. Thus, in view of properties (1_{j-1})–(6_{j-1}), Mergelyan's theorem with interpolation applied to $\psi_{j-1} : D_{j-1} \cup \gamma \hookrightarrow \mathbb{C}^2$ guarantees the existence of a connected, smoothly bounded, compact domain $D'_{j-1} \subset \mathring{M}_j$ and an embedding $\psi'_{j-1} : D'_{j-1} \hookrightarrow \mathbb{C}^2$ of class $\mathscr{A}^1(D'_{j-1})$ with the following properties:

$(1'_{j-1})$ $D_{j-1} \Subset D'_{j-1} \Subset M_j$ and D'_{j-1} is a strong deformation retract of M_j.

$(2'_{j-1})$ $|\psi'_{j-1}(p) - \psi_{j-1}(p)| < \epsilon_j/2$ for all $p \in D_{j-1}$.

$(3'_{j-1})$ $\mathrm{dist}_{\psi'_{j-1}}(p_0, \mathrm{b}D'_{j-1}) > j - 1$.

$(4'_{j-1})$ $\psi'_{j-1}(\mathrm{b}D'_{j-1}) \subset r_j\mathbb{B} \setminus s'_{j-1}\overline{\mathbb{B}}$.

$(5'_{j-1})$ $\psi'_{j-1}(D'_{j-1} \setminus \mathring{D}_{j-1}) \cap s'_{j-1}\overline{\mathbb{B}} = \varnothing$.

$(6'_{j-1})$ $\Lambda_i \subset \psi'_{j-1}(\mathring{D}_i)$ for all $i \in \{0, \ldots, j-1\}$.

Since $\chi(M_j \setminus \mathring{D}'_{j-1}) = 0$, this reduces the proof to Case 1 and hence concludes the construction of the sequence $\Xi_j = \{D_j, \psi_j\}$ meeting properties (1_j)–(6_j), $j \in \mathbb{N}$.

This completes the proof of Theorem 3.1. □

To finish the paper we show how Theorem 3.1 implies Theorem 1.2.

Proof of Theorem 1.2. Let Λ and M be as in Theorem 1.2 and assume without loss of generality that $0 \in \Lambda$. By [Alarcón and López 2013b, Main Theorem, page 1795], there exists a complex structure on M such that the open Riemann surface M admits a proper holomorphic embedding $\phi : M \hookrightarrow \mathbb{C}^2$; up to composing with a translation, we may assume that $0 \in \phi(M)$. Take a number $0 < r < 1$ such that $\Lambda \cap r\overline{\mathbb{B}} = \{0\}$; since $\Lambda \subset \mathbb{B}$ is closed and discrete, every small enough $r > 0$ meets this requirement. Thus, there are a small number $0 < s < r$ and a smoothly bounded compact disk $K \subset M$ such that $\phi(\mathrm{b}K) \subset r\mathbb{B} \setminus s\overline{\mathbb{B}}$ and $0 \in \phi(\mathring{K})$.

Now, Theorem 3.1 applied to $\psi := \phi|_K$ furnishes a domain $D \subset M$ homeomorphic to M and a complete proper holomorphic embedding $\tilde{\psi} : D \hookrightarrow \mathbb{B}$ with $\Lambda \subset \tilde{\psi}(D)$. Since D is homeomorphic to the smooth surface M, there exists a complex structure on M (possibly different from the one used in the previous paragraph) such that the open Riemann surface M is biholomorphic to D. □

Acknowledgements

We thank Francisco J. López for helpful discussions about the paper.

Alarcón is supported by the Ramón y Cajal program of the Spanish Ministry of Economy and Competitiveness and is partially supported by the MINECO/FEDER grant no. MTM2014-52368-P, Spain.

References

[Alarcón and Forstnerič 2013] A. Alarcón and F. Forstnerič, "Every bordered Riemann surface is a complete proper curve in a ball", *Math. Ann.* **357**:3 (2013), 1049–1070. MR Zbl

[Alarcón and López 2013a] A. Alarcón and F. J. López, "Null curves in $\mathbb{C}^3$ and Calabi–Yau conjectures", *Math. Ann.* **355**:2 (2013), 429–455. MR Zbl

[Alarcón and López 2013b] A. Alarcón and F. J. López, "Proper holomorphic embeddings of Riemann surfaces with arbitrary topology into $\mathbb{C}^2$", *J. Geom. Anal.* **23**:4 (2013), 1794–1805. MR Zbl

[Alarcón and López 2016] A. Alarcón and F. J. López, "Complete bounded embedded complex curves in $\mathbb{C}^2$", *J. Eur. Math. Soc. (JEMS)* **18**:8 (2016), 1675–1705. MR Zbl

[Alarcón et al. 2015] A. Alarcón, B. Drinovec Drnovšek, F. Forstnerič, and F. J. López, "Every bordered Riemann surface is a complete conformal minimal surface bounded by Jordan curves", *Proc. Lond. Math. Soc.* (3) **111**:4 (2015), 851–886. MR Zbl

[Alarcón et al. 2016] A. Alarcón, J. Globevnik, and F. J. López, "A construction of complete complex hypersurfaces in the ball with control on the topology", *J. Reine Angew. Math.* (online publication November 2016).

[Bell and Narasimhan 1990] S. R. Bell and R. Narasimhan, "Proper holomorphic mappings of complex spaces", pp. 1–38 in *Several complex variables, VI*, edited by W. Barth and R. Narasimhan, Encyclopaedia Math. Sci. **69**, Springer, 1990. MR Zbl

[Bishop 1958] E. Bishop, "Subalgebras of functions on a Riemann surface", *Pacific J. Math.* **8** (1958), 29–50. MR Zbl

[Drinovec Drnovšek 2015] B. Drinovec Drnovšek, "Complete proper holomorphic embeddings of strictly pseudoconvex domains into balls", *J. Math. Anal. Appl.* **431**:2 (2015), 705–713. MR Zbl

[Forstnerič 2011] F. Forstnerič, *Stein manifolds and holomorphic mappings: the homotopy principle in complex analysis*, Ergebnisse der Mathematik (3) **56**, Springer, 2011. MR Zbl

[Forstnerič and Wold 2009] F. Forstnerič and E. F. Wold, "Bordered Riemann surfaces in $\mathbb{C}^2$", *J. Math. Pures Appl.* (9) **91**:1 (2009), 100–114. MR Zbl

[Forstnerič and Wold 2013] F. Forstnerič and E. F. Wold, "Embeddings of infinitely connected planar domains into $\mathbb{C}^2$", *Anal. PDE* **6**:2 (2013), 499–514. MR Zbl

[Forstnerič et al. 1996] F. Forstnerič, J. Globevnik, and B. Stensønes, "Embedding holomorphic discs through discrete sets", *Math. Ann.* **305**:3 (1996), 559–569. MR Zbl

[Globevnik 2015] J. Globevnik, "A complete complex hypersurface in the ball of $\mathbb{C}^N$", *Ann. of Math.* (2) **182**:3 (2015), 1067–1091. MR Zbl

[Globevnik 2016a] J. Globevnik, "Embedding complete holomorphic discs through discrete sets", *J. Math. Anal. Appl.* **444**:2 (2016), 827–838. MR Zbl

[Globevnik 2016b] J. Globevnik, "Holomorphic functions unbounded on curves of finite length", *Math. Ann.* **364**:3-4 (2016), 1343–1359. MR Zbl

[Jones 1979] P. W. Jones, "A complete bounded complex submanifold of $\mathbb{C}^3$", *Proc. Amer. Math. Soc.* **76**:2 (1979), 305–306. MR Zbl

[Martín et al. 2009] F. Martín, M. Umehara, and K. Yamada, "Complete bounded holomorphic curves immersed in $\mathbb{C}^2$ with arbitrary genus", *Proc. Amer. Math. Soc.* **137**:10 (2009), 3437–3450. MR Zbl

[Mergelyan 1951] S. N. Mergelyan, "On the representation of functions by series of polynomials on closed sets", *Doklady Akad. Nauk SSSR* (*N.S.*) **78** (1951), 405–408. In Russian. MR

[Runge 1885] C. Runge, "Zur Theorie der Analytischen Functionen", *Acta Math.* **6**:1 (1885), 245–248. MR Zbl

[Yang 1977a] P. Yang, "Curvature of complex submanifolds of C^n", pp. 135–137 in *Several complex variables* (*Proc. Sympos. Pure Math., Vol. XXX, Part 2*) (Williamstown, MA, 1975), edited by R. O. Wells, Jr., Amer. Math. Soc., Providence, RI, 1977. MR Zbl

[Yang 1977b] P. Yang, "Curvatures of complex submanifolds of $\mathbb{C}^n$", *J. Differential Geom.* **12**:4 (1977), 499–511. MR Zbl

Received 27 Jan 2017. Accepted 29 Jun 2017.

ANTONIO ALARCÓN: alarcon@ugr.es
Departamento de Geometría y Topología e Instituto de Matemáticas (IEMath-GR), Universidad de Granada, Granada, Spain

JOSIP GLOBEVNIK: josip.globevnik@fmf.uni-lj.si
Department of Mathematics and Institute of Mathematics, Physics and Mechanics, University of Ljubljana, Ljubljana, Slovenia

ANALYSIS AND PDE
Vol. 10, No. 8, 2017
dx.doi.org/10.2140/apde.2017.10.2001

FINITE-TIME DEGENERATION OF HYPERBOLICITY WITHOUT BLOWUP FOR QUASILINEAR WAVE EQUATIONS

JARED SPECK

In three spatial dimensions, we study the Cauchy problem for the wave equation $-\partial_t^2\Psi+(1+\Psi)^P\Delta\Psi=0$ for $P\in\{1,2\}$. We exhibit a form of stable Tricomi-type degeneracy formation that has not previously been studied in more than one spatial dimension. Specifically, using only energy methods and ODE techniques, we exhibit an open set of data such that Ψ is initially near 0, while $1+\Psi$ vanishes in finite time. In fact, generic data, when appropriately rescaled, lead to this phenomenon. The solution remains regular in the following sense: there is a high-order L^2-type energy, featuring degenerate weights only at the top-order, that remains bounded. When $P=1$, we show that any C^1 extension of Ψ to the future of a point where $1+\Psi=0$ must exit the regime of hyperbolicity. Moreover, the Kretschmann scalar of the Lorentzian metric corresponding to the wave equation blows up at those points. Thus, our results show that curvature blowup does not always coincide with singularity formation in the solution variable. Similar phenomena occur when $P=2$, where the vanishing of $1+\Psi$ corresponds to the failure of strict hyperbolicity, although the equation is hyperbolic at all values of Ψ.

The data are compactly supported and are allowed to be large or small as measured by an unweighted Sobolev norm. However, we assume that initially the spatial derivatives of Ψ are nonlinearly small relative to $|\partial_t\Psi|$, which allows us to treat the equation as a perturbation of the ODE $(d^2/dt^2)\Psi=0$. We show that for appropriate data, $\partial_t\Psi$ remains quantitatively negative, which simultaneously drives the degeneracy formation and yields a favorable spacetime integral in the energy estimates that is crucial for controlling some top-order error terms. Our result complements those of Alinhac and Lindblad, who showed that if the data are small as measured by a Sobolev norm with radial weights, then the solution is global.

1. Introduction

Many authors have studied model nonlinear wave equations as a means to gain insight into more challenging wave-like quasilinear equations, such as Einstein's equations of general relativity, the compressible Euler equations without vorticity, and the equations of elasticity. Motivated by the same considerations, in this paper, we study model quasilinear wave equations in three spatial dimensions. To simplify the presentation, we have chosen to restrict our attention to the 3-dimensional case only; with only modest additional effort, our results could be generalized to apply in any number of spatial dimensions. In a broad sense, we are interested in finding initial conditions without symmetry assumptions that lead to some kind of stable breakdown. In our main results, which we summarize just below (1A-2), we exhibit a type of stable degen-

Speck gratefully acknowledges support from NSF grant DMS-1162211, from NSF CAREER grant DMS-1454419, from a Sloan Research Fellowship provided by the Alfred P. Sloan foundation, and from a Solomon Buchsbaum grant administered by the Massachusetts Institute of Technology.
MSC2010: primary 35L80; secondary 35L05, 35L72.
Keywords: degenerate hyperbolic, strictly hyperbolic, Tricomi equation, weakly hyperbolic.

erate solution behavior, *distinct from blowup*, that, to the best of our knowledge, has not previously been studied in the context of quasilinear equations in more than one spatial dimension. Roughly, we show that there exists an open set of data such that certain principal coefficients in the equation vanish in finite time *without a singularity forming in the solution*. More precisely, the vanishing of the coefficients corresponds to the vanishing of the wave speed, which in turn is tied to other kinds of degeneracies described below. We note that Wong [2016] has obtained similar constructive results for axially symmetric timelike minimal submanifolds of Minkowski spacetime, a setting in which the equations of motion are a system of (effectively one-space-dimensional) quasilinear wave equations with principal coefficients that depend on the solution (but not its derivatives). Specifically, he showed that *all* axially symmetric solutions (without any smallness assumption) lead to a finite-time degeneracy caused by the vanishing of a principal coefficient in the evolution equations. We also note that in the case of one spatial dimension, results similar to ours are obtained in [Kato and Sugiyama 2013; Sugiyama 2013; 2016a; 2016b] using proofs by contradiction that rely on the method of Riemann invariants. However, since the method of Riemann invariants is not applicable in more than one spatial dimension and since we are interested in direct proofs, our approach here is quite different.

Through our study of model problems, we are aiming to develop approaches that might be useful for studying the kinds of degeneracies that might develop in solutions to more physically relevant quasilinear equations. One consideration behind this aim is that there are relatively few breakdown results for quasilinear equations compared to the semilinear case. A second consideration is that many of the techniques that have been used to study semilinear wave equations do not apply in the quasilinear case; see Section 1D for further discussion. A third consideration concerns fundamental limitations of semilinear model equations: they are simply incapable of exhibiting some of the most important degeneracies that can occur in solutions to quasilinear equations. In particular, the degeneracy exhibited by the solutions from our main results cannot occur in solutions to semilinear wave equations with principal part equal to the linear wave operator[1] $\square_m$. As a second example of breakdown that is unique to the quasilinear case, we note that the phenomenon of shock formation, described in more detail at the end of Section 1F, cannot occur in solutions to semilinear equations since, in the semilinear case, the evolution of characteristics is not influenced in any way[2] by the solution.

In view of the above discussion, it is significant that our analysis has robust features and could be extended to apply to a large class[3] of quasilinear equations. The robustness stems from the fact that our proofs are based only on energy estimates, ODE-type estimates, and the availability of an important monotonic spacetime integral (which we describe below) that arises in the energy estimates. However, rather than formulating a theorem about a general class of equations, we prefer to keep the paper short and to exhibit the main ideas by studying only the model equation (1A-1a) below in the cases $P = 1, 2$.

[1]Here, $\square_m := -\partial_t^2 + \Delta$ denotes the standard linear wave operator corresponding to the Minkowski metric $m := \mathrm{diag}(-1, 1, 1, 1)$ on $\mathbb{R}^{1+3}$.

[2]In the works on shock formation for quasilinear equations described in Section 1F, the intersection of the characteristics is tied to the blowup of some derivative of the solution.

[3]Of course, we can only hope to treat wave equations whose principal spatial coefficients vanish when evaluated at some finite values of the solution variable; equations such as $-\partial_t^2 \Psi + (1 + \Psi^2)\,\Delta\Psi = 0$ are manifestly immune to the kind of degeneracies under study here.

1A. ***Statement of the equations and summary of the main results.*** Specifically, in the cases $P = 1, 2$, we study the following model Cauchy problem on $\mathbb{R}^{1+3}$:

$$-\partial_t^2\Psi + (1+\Psi)^P \Delta\Psi = 0, \tag{1A-1a}$$

$$(\Psi|_{\Sigma_0}, \partial_t\Psi|_{\Sigma_0}) = (\mathring{\Psi}, \mathring{\Psi}_0), \tag{1A-1b}$$

where $(x^0 := t, x^1, x^2, x^3)$ is a fixed set of standard rectangular coordinates on $\mathbb{R}^{1+3}$, $\Delta := \sum_{a=1}^3 \partial_a^2$ is the standard Euclidean Laplacian on $\mathbb{R}^3$, and throughout,[4] $\Sigma_t := \{t\} \times \mathbb{R}^3 \simeq \mathbb{R}^3$. We sometimes denote the spatial coordinates by $\underline{x} := (x^1, x^2, x^3)$. Note that we can rewrite (1A-1a) as[5] $(g^{-1})^{\alpha\beta}(\Psi)\partial_\alpha\partial_\beta\Psi = 0$, where g is the Lorentzian (for $\Psi > -1$) metric

$$g := -dt^2 + (1+\Psi)^{-P}\sum_{a=1}^3 (dx^a)^2. \tag{1A-2}$$

This geometric perspective will be useful at various points in our discussion.

We now summarize our results; see Theorem 4.1 and Proposition 4.2 for precise statements.

Summary of the main results: In the case $P = 1$, there exists an open subset of $H^6(\mathbb{R}^3) \times H^5(\mathbb{R}^3)$ comprising compactly supported initial data $(\mathring{\Psi}, \mathring{\Psi}_0)$ such that the solution Ψ, its *spatial* derivatives, and its *mixed space-time* derivatives initially satisfy a nonlinear smallness condition compared to[6] $\max_{\Sigma_0}[\mathring{\Psi}_0]_-$ and $1/\|\mathring{\Psi}_0\|_{L^\infty(\Sigma_0)}$, and such that the solution has the following property: the coefficient $1+\Psi$ in (1A-1a) vanishes at some time $T_\star \in (0, \infty)$. In fact, the finite-time vanishing of $1+\Psi$ *always* occurs if $\mathring{\Psi}_0$ is nontrivial and the data are appropriately rescaled; see Remark 2.4. Moreover,

$$\Psi \in C\big([0, T_\star), H^6(\mathbb{R}^3)\big) \cap L^2\big([0, T_\star], H^6(\mathbb{R}^3)\big) \cap C\big([0, T_\star], H^5(\mathbb{R}^3)\big),$$

while for any $N < 5$,

$$\partial_t\Psi \in C\big([0, T_\star), H^5(\mathbb{R}^3)\big) \cap L^\infty\big([0, T_\star], H^5(\mathbb{R}^3)\big) C\big([0, T_\star], H^N(\mathbb{R}^3)\big).$$

In addition, the Kretschmann scalar $\mathrm{Riem}(g)^{\alpha\beta\gamma\delta}\,\mathrm{Riem}(g)_{\alpha\beta\gamma\delta}$ blows up precisely at points where $1+\Psi$ vanishes, where $\mathrm{Riem}(g)$ denotes the Riemann curvature of g. Finally, the solution exits the regime of hyperbolicity at time $T_\star$ and thus it cannot be continued beyond $T_\star$ as a classical solution to a hyperbolic equation. In the case $P = 2$, similar results hold, the main differences being that Ψ is not necessarily an element of $L^2\big([0, T_\star], H^6(\mathbb{R}^3)\big)$ and that the strict hyperbolicity[7] breaks down when $1+\Psi$ vanishes but hyperbolicity[8] does not.[9] This leaves open, in the case $P = 2$, the possibility of classically extending the solution past time $T_\star$; see Section 1C2.

[4] Here we use the notation "$\simeq$" to mean "diffeomorphic to".

[5] Throughout we use Einstein's summation convention.

[6] Throughout, $[p]- := |\min\{p, 0\}|$.

[7] Equation (1A-1a) is said to be strictly hyperbolic in the direction ω if the symbol $p(\xi) := -\xi_0^2 + (1+\Psi)^P \sum_{a=1}^3 \xi_a^2$ has the following property: for any one-form $\xi \neq 0$, the polynomial $s \to p(\xi + s\omega)$ has two distinct real roots. It is straightforward to see that (1A-1a) is strictly hyperbolic in the direction $\omega := (1, 0, 0, 0)$ if $1+\Psi > 0$, and that it is *not* strictly hyperbolic in any direction if $1+\Psi = 0$.

[8] Here, by hyperbolic (in the direction ω), we mean that for all one-forms $\xi \neq 0$, the polynomial $s \to p(\xi + s\omega)$ from Footnote 7 has only real roots. Such polynomials are known as *hyperbolic polynomials*.

[9] In the literature, equations exhibiting this kind of degeneracy are often referred to as *weakly hyperbolic*.

1B. ***Paper outline.*** The remainder of the paper is organized as follows.

- In Section 1C, we provide some initial remarks expanding upon various aspects of our results.
- In Section 1D, we mention some techniques that have been used in studying the breakdown of solutions to semilinear equations. As motivation for the present work, we point out some limitations of the semilinear techniques for the study of quasilinear equations.
- In Section 1E we provide a brief overview of the proof of our main results.
- In Section 1F, we describe some connections between our results and prior work on degenerate hyperbolic PDEs.
- In Section 1G we summarize our notation.
- In Section 2, we state our assumptions on the initial data and introduce bootstrap assumptions.
- In Section 3, we use the bootstrap and data-size assumptions of Section 2 to derive a priori pointwise estimates and energy estimates. From the energy estimates, we deduce improvements of the bootstrap assumptions.
- In Section 4, we use the estimates of Section 3 to prove our main results.

1C. ***Initial remarks on the main results.*** As far as we know, there are no prior results in the spirit of our main results in more than one spatial dimension. There are, however, examples in which the Cauchy problem for a quasilinear wave equation has been solved (for suitable data without symmetry assumptions) and such that it was shown that some derivative of the solution blows up in finite time while the solution itself remains bounded. One class of such examples comprises shock formation results, which we describe in more detail at the end of Section 1F. A second example is [Luk 2013] on the formation of weak null singularities in a family of solutions to the Einstein-vacuum equations. Specifically, Luk exhibited a stable family of solutions such that the Christoffel symbols (which are, roughly speaking, the first derivatives of the solution) blow up along a null boundary, while the metric (that is, the solution itself) extends continuously past the null boundary. We stress that the degeneracy we have exhibited in our main results is much less severe than in the above results; there is no blowup in our solutions, except possibly at the top derivative level, due to the degeneracy of the weights in the energy (1E-1).

We also point out a connection between our work here and our joint works [Rodnianski and Speck 2014a; 2014b], in which we proved stable blowup results (without symmetry assumptions) for solutions to the linearized and nonlinear Einstein-scalar field and Einstein-stiff fluid systems. In the nonlinear problem, the wave speed became, relative to a geometrically defined coordinate system,[10] *infinite* at the singularity. Although the infinite wave speed is in the opposite direction of the degeneracy exhibited by our main results (in which the wave speed vanishes[11]), the analysis in [Rodnianski and Speck 2014a; 2014b] shares a key feature with that of the present work: the solution regime studied is such that the time derivatives dominate the evolution. That is, the spatial derivatives remain negligible, all the way up to the degeneracy; see

[10] Specifically, the Σ_t have constant mean curvature and the spatial coordinates are transported along the unit normal to Σ_t.
[11] Note that the effective wave speed for (1A-1a) is $(1+\Psi)^{P/2}$.

Section 1E for further discussion regarding this issue for the solutions under study here. Hence, those works and the present work all exhibit the stability of ODE-type behavior in some solutions to wave equations.

1C1. *Remarks on small data.* The methods of Alinhac [2003] and Lindblad [2008] yield that small-data solutions to (1A-1a) exist globally,[12] where the size of the data is measured by Sobolev norms with *radial weights*. Consequently, if $(\mathring{\Psi}, \mathring{\Psi}_0)$ are compactly supported data to which our main results apply, then for λ sufficiently large, the solution corresponding to the data $(\lambda^{-1}\mathring{\Psi}, \lambda^{-1}\mathring{\Psi}_0)$ is global. On the other hand, our main results apply to data that are allowed to be small in certain unweighted norms, as long as the spatial derivatives are "very small". How can we reconcile these two competing statements? The answer is that our data assumptions are nonlinear in nature and are *not* invariant under the rescaling $(\mathring{\Psi}, \mathring{\Psi}_0) \to (\lambda^{-1}\mathring{\Psi}, \lambda^{-1}\mathring{\Psi}_0)$ if λ is too large. We can sketch the situation as follows (see Section 2C for the precise nonlinear smallness assumptions that we use to close our proof): if ϵ is the size of $\nabla\Psi$ at time 0 (where ∇ denotes the spatial coordinate gradient) and δ is the size of $\partial_t\Psi$ at time 0, then, roughly speaking, some parts of our proof rely on[13] the assumption that $\epsilon\exp(C\delta^{-1}) \lesssim 1$. The point is that if λ is too large, then the assumption is not satisfied, the reason being that ϵ and δ both scale like λ^{-1}. One can contrast this against the discussion in Section 2D, where we note that a different scaling of the data always leads to our nonlinear smallness assumptions being satisfied.

1C2. *Remarks on extending the solution past the degeneracy.* It is of interest to know if and when the solutions provided by our main results can be extended, as solutions with some kind of Sobolev regularity,[14] past the time of first vanishing of $1+\Psi$. Although we do not address this question in this article, in this subsubsection, we describe what is known and some of the difficulties that one would encounter in attempting to answer it. The cases $P = 1$ and $P = 2$ in (1A-1a) correspond to different phenomena and hence we will discuss them separately, starting with the case $P = 1$.

Interesting results have recently been obtained in [Lerner et al. 2015] for equations related to (1A-1a). They *suggest* that in the case $P = 1$, it might not be possible to continue the solutions from our main results as Sobolev-class solutions in a spacetime neighborhood of a point at which $1+\Psi$ vanishes. Perhaps this is not surprising since, for the solutions under study, the case $P = 1$ corresponds to (1A-1a) changing from hyperbolic to elliptic type past the degeneracy (at least for C^1 solutions). Specifically, those authors proved a type of Hadamard ill-posedness for certain initial data for a class of quasilinear first-order systems in n spatial dimensions of the form

$$\partial_t u + \sum_{a=1}^{n} A^a(t,x,u)\,\partial_a u = F(t,x,u), \tag{1C-1}$$

[12] Since the equations do not satisfy the null condition, the asymptotics of the solution can be distorted compared to the case of solutions to the linear wave equation.

[13] For example, a careful analysis of the proof of inequality (3C-4) yields that the constant C in front of the $\mathring{\epsilon}^2$ term on the right-hand side depends on $\exp(\mathring{\delta}_*^{-1})$, where $\mathring{\delta}_*$ is defined in (2A-2). See Section 3A for our conventions regarding the dependence of constants on various parameters.

[14] The Cauchy–Kovalevskaya theorem could be used to prove an (admittedly unsatisfying) result showing that in the cases $P = 1, 2$, one can extend *analytic* solutions to (1A-1a) to exist in a spacetime neighborhood of a point at which $1+\Psi$ vanishes. Note that this shows that the blowup of the curvature of the metric of (1A-1a) that occurs when $1+\Psi = 0$ is not always an obstacle to continuing the solution classically.

where $(t,x)\in\mathbb{R}^{1+n}$, u is a map from $\mathbb{R}^{1+n}$ to $\mathbb{R}^N$ with n and N arbitrary, and the A^a are real $N\times N$ matrices. The authors proved several types of results in [Lerner et al. 2015], but here we describe only the ones that are most relevant for (1A-1a). Roughly, in Theorem 1.3 of that paper, for systems of type (1C-1) that satisfy some technical conditions, the authors studied perturbations of a background solution, denoted by $\phi=\phi(t,x)$, with the following property: the system (1C-1) is hyperbolic when evaluated at $(t,x,u)=(0,x,\mathring{\phi}(x))$, where $\mathring{\phi}(x):=\phi(0,x)$, but necessarily becomes elliptic at $(t,x,u)=(t,x,\phi(t,x))$ at any $t>0$ due to the branching[15] of the eigenvalues of the principal symbol. The assumptions of their Theorem 1.3 guarantee that the branching is stable under small perturbations. Roughly, for the solutions to (1A-1a) from our main results, a similar transition to ellipticity would occur in the case $P=1$ if one were able to classically extend the solution[16] past the time of first vanishing of $1+\Psi$. We now summarize the main aspects of [Lerner et al. 2015, Theorem 1.3]. We will use the notation $\mathring{u}$ to denote initial data for the system (1C-1) and u to denote the corresponding solution (if it exists). The theorem is, roughly, as follows: for any $m\in\mathbb{R}$ and $\alpha\in\left(\frac{1}{2},1\right]$, and any sufficiently small $T>0$, there is no H^m-neighborhood $\mathcal{U}$ of $\mathring{\phi}$ whose elements launch corresponding solutions obeying a bound roughly[17] of the type[18]

$$\sup_{\mathring{u}\in\mathcal{U}}\frac{\|u-\phi\|_{W_x^{1,\infty}L_t^\infty([0,T])}}{\|\mathring{u}-\mathring{\phi}\|_{H^m}^\alpha}<\infty.$$

Put differently, there exist data arbitrarily close to $\mathring{\phi}$ (as measured by a Sobolev norm of arbitrarily high order) such that either the solution does not exist or such that its deviation from ϕ becomes arbitrarily large in the low-order norm $\|\cdot\|_{W_x^{1,\infty}}$ in an arbitrarily short amount of time. It would be interesting to determine whether or not a similar result holds for initial data close to that of the data induced by the solutions to (1A-1a) from our main results at the time of first vanishing of $1+\Psi$.

We now discuss the case $P=2$. We are not aware of any results for Sobolev-class solutions to quasilinear equations that are relevant for extending solutions to (1A-1a) to exist in a spacetime neighborhood of a point at which $1+\Psi$ vanishes. As we will explain, the main technical difficulty that one encounters is that the solution might lose regularity past the degeneracy. In the case $P=2$, even though the strict hyperbolicity (see Footnote 7) of (1A-1a) breaks down when $1+\Psi$ vanishes (corresponding to a wave of zero speed), the hyperbolicity (see Footnote 8) of the equation nonetheless persists for *all* values of Ψ. The degeneracy is therefore less severe compared to the case $P=1$ and thus in principle, when $P=2$, the Sobolev-class solutions from our main results might be extendable, as a Sobolev-class solution, to a neighborhood of the points where $1+\Psi$ first vanishes. As we alluded to above, the lack of results in this direction might be tied to the following key difficulty: the best energy estimates available for degenerate[19] linear hyperbolic wave

[15] In [Lerner et al. 2015], the definition of hyperbolicity is that the polynomial (in λ) $p:=\det\left(\lambda I-\sum_{a=1}^n\xi_aA^a(t,x,u)\right)$ should have only real roots, which are eigenvalues of $\sum_{a=1}^n\xi_aA^a(t,x,u)$. Moreover, branching roughly means that the eigenvalues are real at $t=0$ but can have nonzero imaginary parts at arbitrarily small values of $t>0$.

[16] As we will explain, the solutions from our main results are such that Ψ is strictly decreasing in time at points where $1+\Psi$ vanishes.

[17] The precise results of [Lerner et al. 2015, Theorem 1.3] are localized in space, but here we omit those details for brevity.

[18] If $f=f(t,x)$, then $\|f\|_{W_x^{1,\infty}L_t^\infty([0,T])}:=\operatorname{ess\,sup}_{t\in[0,T]}\|f(t,\cdot)\|_{W^{1,\infty}}$.

[19] By degenerate, we mean that the wave equation is allowed to violate strict hyperbolicity at one or more points.

equations exhibit a loss of derivatives. By this, we roughly mean that the estimates for solutions Ψ to the linear equation are of the form $\|\Psi\|_{H^N(\Sigma_t)} \lesssim \|\mathring{\Psi}\|_{H^{N+d}(\Sigma_0)} + \|\mathring{\Psi}_0\|_{H^{N+d}(\Sigma_0)}$, where the loss of derivatives d (relative to the data) depends in a complicated way on the details of the degeneration of the coefficients in the equation; see Section 1F for further discussion. As is described in [Dreher 1999], in some cases, the loss of derivatives in the estimates is known to be saturated. Since proofs of well-posedness for nonlinear equations typically rely on estimates for linearized equations, any derivative loss would pose a serious obstacle to extending (in the case $P = 2$) the solution of (1A-1a) as a Sobolev-class solution in a spacetime neighborhood of points at which $1+\Psi$ vanishes. At the very least, one would need to rely on a method capable of handling a finite loss of derivatives in solutions to quasilinear equations. As is well-known [Hamilton 1982], in some cases, it is sometimes possible to handle a finite loss of derivatives using the Nash–Moser framework.

Despite the lack of results concerning extending the solution to (1A-1a) as a Sobolev-class solution past points at which $1+\Psi$ vanishes, there are constructive results in the class C^∞. Specifically, in one spatial dimension, Manfrin [1996] obtained well-posedness results that, for C^∞ initial data, allow one to locally continue the solution to (1A-1a) in the case $P = 2$ to a C^∞ solution that exists in a spacetime neighborhood of a point at which $1+\Psi$ vanishes; see Section 1F for further discussion. Manfrin [1999] also derived similar results in more than one spatial dimension, again treating the case of C^∞ data/solutions. We are also aware of a few results [Dreher 1999; Han et al. 2003] for quasilinear equations in more than one spatial dimension in which the authors proved local well-posedness in Sobolev spaces for equations featuring a degeneracy related to — but distinct from — the one under study here. However, the degeneracy in those works was created by a "prescribed semilinear factor" rather than a quasilinear-type solution-dependent factor. For this reason, it is not clear that the techniques used in those works are of relevance for trying to extend solutions to (1A-1a) beyond points where $1+\Psi$ vanishes; see the paragraph below (1F-5) for further discussion.

To close this subsubsection, we note that there are various well-posedness results [D'Ancona and Spagnolo 1992; Ebihara 1985; Ebihara et al. 1986] for degenerate wave equations of Kirchhoff type. An example of an equation of this type is

$$-\partial_t^2\Psi + F\left(\int_\Omega |\nabla\Psi|^2\,dx\right)\Delta\Psi = 0, \tag{1C-2}$$

where Ω is a bounded open set in $\mathbb{R}^n$ and $F = F(s) \geq 0$ satisfies various technical conditions (with $F = 0$ corresponding to the degeneracy). However, it remains open whether or not the techniques used in studying Kirchhoff-type equations are of relevance for proving local well-posedness for (1A-1a) (in the case $P = 2$) in regions where $1+\Psi$ is allowed to vanish.

1D. ***Remarks on methods used for studying blowup in solutions to semilinear wave equations.*** Although we are not aware of any other results in the spirit of the present work, there are many results exhibiting the most well-known type of degeneracy that can occur in solutions to wave equations in three spatial dimensions: the finite-time *blowup* of initially smooth solutions. Our main goal in this subsection is to recall some of the most important of these results but, at the same time, to describe some limitations of the proof techniques for the study of more general equations. We will focus only on constructive[20] results, by which we mean that the proofs provide a detailed description of the degeneracy formation and the mechanisms driving it, as in the present work. Constructive results, especially those proved via robust techniques, are clearly desirable if one aims to understand the mechanisms of breakdown in solutions to realistic physical and geometric systems. They are also important if one aims to continue the solution past the breakdown, as is sometimes possible if it is not too severe; see, for example, [Christodoulou and Lisibach 2016] for a recent result in spherical symmetry concerning weakly locally extending solutions to the relativistic Euler equations past the first shock singularity. Importantly, we will confine our discussion to prior results for semilinear equations since, as we mentioned earlier, aside from the shock formation results described at the end of Section 1F, most constructive breakdown results for wave equations in three spatial dimensions are blowup results for semilinear equations.

Specifically, most constructive breakdown results for wave equations in three spatial dimensions are blowup results for semilinear equations (or systems) of the form $\square_m \Psi = f(\Psi, \partial\Psi)$, where f is a smooth nonlinear term. Many important[21] approaches have been developed to prove constructive blowup for such equations, especially for scalar equations with $f = f(\Psi)$ given by a power law and for systems of wave-map type; see, for example, [Kenig and Merle 2008; Donninger 2010; Donninger and Schörkhuber 2012; 2014; Krieger and Schlag 2014; Krieger et al. 2008; 2009; Donninger et al. 2014; Rodnianski and Sterbenz 2010; Raphaël and Rodnianski 2012; Duyckaerts et al. 2012; Martel et al. 2014; Donninger and Krieger 2013]. There are also related results that are conditional in the sense that they do not guarantee that the solution will blow up. Instead they characterize the possible behaviors of the solution by providing information such as (i) how the singularity would form if the solution is not global and (ii) the structures of the data sets that lead to the various outcomes; see, for example, [Payne and Sattinger 1975; Struwe 2003; Nakanishi and Schlag 2011a; 2011b; 2012a; 2012b; Krieger et al. 2013a; 2013b; 2014; 2015; Killip et al. 2014].

Although the above results and others like them have yielded major advancements in our understanding of the blowup of solutions to semilinear equations, their proofs fundamentally rely on tools that are not typically applicable to quasilinear equations. Here are some important examples, where for brevity, we are not specific about exactly which semilinear equations have been treated with the stated technique:

- The existence of a conserved energy (which is not available for some important quasilinear equations, such as Einstein's equations[22]). This allows, among other things, for the application of techniques from Hamiltonian mechanics.

[20] Constructive proofs of blowup stand, of course, in contrast to proofs of breakdown by contradiction. There are many examples in the literature of proofs of blowup by contradiction for wave or wave-like equations. Two of the most important ones are Sideris' blowup result [1985] (proved by virial identity arguments) for the compressible Euler equations under a polytropic equation of state and John's proof [1981] of breakdown for several classes of semilinear and quasilinear wave equations in three spatial dimensions. See also the influential work [Levine 1974], in which he proved a nonconstructive blowup result for semilinear wave equations on an abstract Hilbert space.

[21] Arguably, the most sophisticated blowup results of this type have been proved for nonlinearities that correspond to energy critical equations.

[22] For asymptotically flat solutions to Einstein's equations, the ADM mass is conserved. However, in three spatial dimensions without symmetry assumptions, this quantity has thus far proven to be too weak to be of any use in controlling solutions.

• The invariance of the solutions under appropriate rescalings (which is not a feature of some important quasilinear equations, such as the compressible Euler equations[23]).

• The availability of well-posedness results in low regularity spaces such as the energy space (Lindblad [1998] showed that low regularity well-posedness fails for a large class of quasilinear equations in three spatial dimensions).

• The existence of a nontrivial ground state solution (corresponding to the existence of a soliton solution) and sharp classification results for the possible behaviors of the solution for initial data with energy less than the ground state: either there is finite-time blowup in both time directions or global existence, according to the sign of a functional (for quasilinear equations, there is no known analog of this kind of dichotomy). Moreover, in some cases, there are more complicated classification results available for solutions with energy just above the ground state.

• A characterization of a certain norm of the ground state as a size threshold separating global scattering solutions from ones that can blow up or exhibit other degenerate behavior (again, for quasilinear equations, there is no known analog of this kind of dichotomy).

• A characterization of the ground state as the universal blowup-profile under various assumptions.

• The availability of profile decompositions for sequences bounded in the natural energy space, which allows one to view the sequence as a superposition of linear solutions plus a small error (for quasilinear equations, there is no known analog of this).

• Channel-of-energy-type arguments showing that a portion of the solution propagates precisely at speed one (again, for quasilinear equations, there is no known[24] analog of this phenomenon).

• The possibility of sharply characterizing the spectrum, see for example [Costin et al. 2012], of linear operators tied to the dynamics (which, for quasilinear equations in many solution regimes, is exceedingly difficult).

Although the above methods are impressively powerful within their domain of applicability, since they do not seem to apply to quasilinear equations, we believe that it is important to develop new methods for studying the kinds of breakdown that can occur in the quasilinear case. It is for this reason that we have chosen to study the model wave equations (1A-1a).

1E. ***Brief overview of the analysis.*** As we mentioned earlier, the solutions that we study are such that $\mathring{\Psi}$, $\nabla\mathring{\Psi}_0$ (where ∇ denotes the spatial coordinate gradient), and sufficiently many of their spatial derivatives are "nonlinearly small" (in appropriate norms) compared to $[\mathring{\Psi}_0]_- := |\min\{\mathring{\Psi}_0, 0\}|$ and $1/\|\mathring{\Psi}_0\|_{L^\infty(\Sigma_0)}$. A key aspect of our work is that we are able to propagate the smallness, long enough for the coefficient $1+\Psi$ in (1A-1a) to vanish. Put differently, our main results show that under the smallness assumptions, the solution to (1A-1a) behaves in many ways like a solution to the second-order ODE $(d^2/dt^2)\Psi = 0$. The reason that Ψ vanishes for the first time is that $\partial_t\Psi$ is sufficiently negative at one or more spatial

[23]In particular, the fluid equation of state does not generally enjoy any useful scaling transformation properties.

[24]It is conceivable that channel-of-energy-type results might hold for certain quasilinear wave equations in various solution regimes, since channel-of-energy-type arguments seem to be somewhat stable under perturbations.

points, a condition that persists by the previous remarks. To control solutions, we use the (nonconserved) energy[25]

$$\mathcal{E}_{[2,5]}(t) := \sum_{k'=2}^{5} \int_{\Sigma_t} |\partial_t \nabla^{k'} \Psi|^2 + (1+\Psi)^P |\nabla \nabla^{k'} \Psi|^2 + |\nabla^{k'} \Psi|^2 \, d\underline{x}. \tag{1E-1}$$

We avoid using low-order energies corresponding to $k' = 0, 1$ in (1E-1) because for the solution regime under consideration such energies would contain terms that are allowed to be large, and we prefer to work only with small energies. Hence, to control the low-order derivatives of Ψ, we derive ODE-type estimates that rely in part on the energy estimates for its higher derivatives and Sobolev embedding. Analytically, the main challenge is that the vanishing of $1+\Psi$ leads to the degeneracy of the top-order spatial derivative terms in (1E-1), which makes it difficult to control some top-order error integrals in the energy estimates.

To close the energy estimates, we exploit the following monotonicity, which is available due to our assumptions on the data:

$\partial_t \Psi$ is *quantitatively strictly negative* in a neighborhood of points where $1+\Psi$ is close to 0.

This quantitative negativity yields, in our energy identities, the spacetime error integral

$$\int_{s=0}^{t} \int_{\Sigma_s} (\partial_t \Psi)(1+\Psi)^{P-1} |\nabla \nabla^{k'} \Psi|^2 \, d\underline{x} \, ds, \tag{1E-2}$$

which has a "friction-type" sign in regions where $1+\Psi$ is close to 0 but positive; see the spacetime integral on the left-hand side of (3C-3). It turns out that the availability of this spacetime integral compensates for the degeneracy of the energy (1E-1) and yields integrated control over the spatial derivatives up to top-order; *this is the key to closing the proof.*

1F. *Comparing with and contrasting against other results for degenerate hyperbolic equations.* For solutions such that $1+\Psi$ is near 0, (1A-1a) can be viewed as a "nearly degenerate" quasilinear hyperbolic PDE. For this reason, the proofs of our main results have ties to some prior results on degenerate hyperbolic PDEs, which we now discuss. In one spatial dimension, various aspects of degenerate hyperbolic PDEs have been explored in the literature, such as the branching of singularities [Alinhac 1978; Amano and Nakamura 1981; 1982; 1983; 1984], uniqueness of solutions for equations that are hyperbolic in one region but that can change type [Ruziev and Reissig 2016], and conditions that are *necessary* for well-posedness [Yagdzhyan 1989]. However, in the rest of this subsection, we will discuss only well-posedness results since they are most relevant in the context of our main results.

In one spatial dimension, there are many results on well-posedness, in various function spaces, for degenerate linear wave equations for the form

$$-\partial_t^2 \Psi + a(t,x)\, \partial_x^2 \Psi + b(t,x)\, \partial_x \Psi + c(t,x)\, \partial_t \Psi = f(t,x), \tag{1F-1}$$

where $a(t,x) \geq 0$ and $a(t,x) = 0$ corresponds to degeneracy via a breakdown of strict hyperbolicity. For example, if the coefficients $a(t,x)$, $b(t,x)$, and $c(t,x)$ are *analytic* and satisfy certain technical assumptions, then it is known [Nishitani 1984] that (1F-1) is well-posed for C^∞ data; see also [Nishitani

[25] Throughout, $d\underline{x} := dx^1 dx^2 dx^3$ denotes the standard Euclidean volume form on Σ_t.

1980] for similar results. There are also results on well-posedness for degenerate semilinear equations. For example, in [D'Ancona and Trebeschi 2001], the authors used a Nash–Moser argument to prove C^∞ local well-posedness for semilinear equations of the form

$$-\partial_t^2\Psi + a(t,x)\,\partial_x^2\Psi = f(t,x,\Psi,\partial_t\Psi,\partial_x\Psi), \tag{1F-2}$$

where $a(t,x) \geq 0$ is analytic and a and f satisfy appropriate technical assumptions. We clarify that in contrast to our work here, in the above works, the authors solved the equation in a spacetime neighborhood of points at which the degeneracy occurs.

A serious limitation of the above results is that techniques relying on analyticity assumptions are of little use for studying quasilinear Cauchy problems with Sobolev-class data, such as the problems we consider here. Fortunately, well-posedness results for degenerate linear equations that do not rely on analyticity assumptions are also known; see, for example, [Oleĭnik 1970; Taniguchi and Tozaki 1980; D'Ancona 1994; Han et al. 2006; Herrmann 2012; Herrmann et al. 2013; Han and Liu 2015]. We note in particular that the results of [Oleĭnik 1970; Taniguchi and Tozaki 1980; Han et al. 2006; Herrmann 2012; Herrmann et al. 2013; Han and Liu 2015] provide Sobolev estimates for the solution in terms of a Sobolev norm of the data, with a finite loss of derivatives. We also mention the related works [Ascanelli 2006; 2007], in which the author proves well-posedness results (in C^∞ and Gevrey spaces) for linear wave equations with two kinds of degeneracies: (i) the breakdown of strict hyperbolicity (corresponding to the vanishing of certain coefficients) and (ii) the blowup of the time derivatives of certain coefficients in the wave equation. We also mention the works [Ivriĭ 1975; Ishida and Yagdjian 2002], in which the authors obtain necessary and sufficient conditions for the Gevrey space well-posedness of degenerate linear hyperbolic equations.

Most relevant for our work here is Manfrin's aforementioned proof [1996] of C^∞ well-posedness for various degenerate *quasilinear* wave equations in one spatial dimension (see also [Manfrin 1999] for a similar result in more than one spatial dimension and the related work [Boiti and Manfrin 2000]), including those of the form

$$-\partial_t^2\Psi + \Psi^{2k}a(t,x,\Psi)\,\partial_x^2\Psi = f(t,x,\Psi), \tag{1F-3}$$

where $k \geq 1$ is an integer and $a(t,x,\Psi)$ is uniformly bounded from above and from below, strictly away from 0 (and $\Psi = 0$ corresponds to the degeneracy). More precisely, for C^∞ initial data, Manfrin used weighted energy estimates and Nash–Moser estimates to prove local well-posedness for solutions to (1F-3). The energy estimate weights are complicated to construct and are based on dividing spacetime into various regions with the help of "separating functions" adapted to the degeneracy. Note that Manfrin's results apply to our model equation[26] (1A-1a) in the case $P = 2$. However, it is an open problem whether or not his results can be extended to yield a local well-posedness result for (1F-3) with data in Sobolev spaces.

To further explain these results and their connection to our work here, we consider the simple Tricomi-type equation

$$-\partial_t^2\Psi + a(t)\,\Delta\Psi = 0, \tag{1F-4}$$

[26]More precisely, the role of "$1+\Psi$" in (1A-1a) is played by "Ψ" in (1F-3).

where $a(t) \geq 0$. It is known [Colombini and Spagnolo 1982] that in one spatial dimension, the linear (1F-4) can be *ill-posed*,[27] even if $a = a(t)$ is C^∞. Hence, it should not be taken for granted that we can (for suitable data) solve (1A-1a) in Sobolev spaces all the way up to the time of first vanishing of $1 + \Psi$. Roughly, what can go wrong in an attempt to solve the linear equation (1F-4) is that $a(t)$ can be highly oscillatory near a point t_0 with $a(t_0) = 0$. In fact, in the example from [Colombini and Spagnolo 1982], $a(t)$ oscillates *infinitely many* times near t_0. This generates, in the basic energy identity, an uncontrollable term involving the ratio $((d/dt)a(t))/a(t)$ and leads to ill-posedness in domains $[A, B) \times \mathbb{R}$ when $t_0 \in [A, B)$.

In all of the aforementioned well-posedness results, the technical conditions imposed on the coefficients rule out the infinite oscillatory behavior from [Colombini and Spagnolo 1982] that led to ill-posedness. To provide a more concrete example, we note that in one spatial dimension, Han [2010] derived degenerate energy estimates for linear wave equations of the form

$$-\partial_t^2 \Psi + a(t, x)\, \partial_x^2 \Psi + b_0(t, x)\, \partial_t \Psi + b(t, x)\, \partial_x \Psi + c(t, x)\Psi = f(t, x), \tag{1F-5}$$

where the coefficients satisfy certain technical conditions, including, roughly speaking, that $a(t, x) \geq 0$ behaves like $t^m + c_{m-1}(x)t^{m-1} + \cdots + c_1(x)t + c_0(x)$. In particular, even though a is allowed to vanish at some points, it does not exhibit highly oscillatory behavior in the t-direction. In [Han et al. 2006], similar results were derived in $n \geq 1$ spatial dimensions.

We now describe Dreher's Ph.D. thesis [1999], which involves the study of equations that share some common features with (1A-1a) near the degeneracy $1 + \Psi = 0$. Specifically, in his thesis, Dreher proved local well-posedness results in Sobolev spaces for several classes of quasilinear hyperbolic PDEs in any number of dimensions with various kinds of space and time degeneracies. However, a key difference between the equations studied by Dreher in his thesis and our work is that the degeneracies there were "prescribed" in the sense that they were caused only by degenerate semilinear factors that explicitly depend on the time and space variables. That is, if one deletes the degenerate semilinear factors, then one obtains a strictly hyperbolic PDE for which local well-posedness follows from standard techniques. Dreher made technical assumptions on the degenerate semilinear factors that were sufficient for proving well-posedness. In contrast, the degeneracy caused by $1 + \Psi = 0$ in (1A-1a) is "purely quasilinear" in nature. The following model equation in one spatial dimension gives a sense of the kinds of prescribed degeneracy treated by Dreher:

$$-\partial_t^2 \Psi + t^2 f(\Psi)\, \partial_x^2 \Psi = 0, \tag{1F-6}$$

where f is smooth and satisfies $f(\Psi) > 0$. We stress that the absence of strict hyperbolicity in a neighborhood of Σ_0 is *not* caused by the quasilinear factor $f(\Psi)$, but rather by the semilinear factor t^2. A related example, coming from geometry, is the aforementioned work [Han et al. 2003], in which the authors proved the existence of local C^k embeddings of surfaces of nonnegative Gaussian curvature into $\mathbb{R}^3$. The quasilinear system of PDEs studied there degenerated at points where the Gauss curvature of the surface vanishes. As in [Dreher 1999], the degeneracy was "prescribed" in the sense that it was caused

[27] In [Colombini and Spagnolo 1982], which addressed solutions in one spatial dimension, the authors exhibited a smooth function $a(t) \geq 0$ with $a(t_0) = 0$ for some $t_0 > 0$ and data such that there is no distributional solution to (1F-4) with the given data that extends past time t_0.

by the Gauss curvature (which is "known"). Hence, the authors were free to make technical assumptions on the Gauss curvature to ensure the local well-posedness of the PDE system.

We now give another example of prior work that is closely connected to our main results. In [Ruan et al. 2016], the authors proved local well-posedness in homogeneous Sobolev spaces on domains of the form $[0, T) \times \mathbb{R}^n$ for semilinear Tricomi equations of the form

$$-\partial_t^2 \Psi + t^P \Delta \Psi = f(\Psi), \tag{1F-7}$$

where $P \in \mathbb{N}$ and f is a nonlinearity such that f and f' obey certain $P, n-$dependent power-law growth bounds at ∞. See [Ruan et al. 2015a; 2015b] for related results. Note that the coefficient t^P in (1F-7) does not oscillate; once again, this is the key difference compared to the ill-posedness result for (1F-4) mentioned above. As we described in Section 1E, (1F-7) is a good model for the solutions to (1A-1a) provided by our main results in the sense that the degenerating coefficient $(1+\Psi)^P$ in (1A-1a) behaves in some ways, when $1+\Psi$ is small, like[28] the coefficient t^P (near $t=0$) in (1F-7).

In view of the above discussion, we believe that one should not expect to be able to solve (1A-1a) in Sobolev spaces all the way up to points with $1+\Psi=0$ unless one makes assumptions on the data that preclude highly oscillatory behavior in regions where $1+\Psi$ is small. In this article, we avoid the oscillatory behavior by exploiting the *relative* largeness of $[\partial_t \Psi]_-$ and the *relative* smallness of $\partial_t^2 \Psi$ in regions where $1+\Psi$ is small, which are present at time 0 and which we propagate; see the estimates (3B-2) and (3C-5c). As we have mentioned, the relative largeness of $[\partial_t \Psi]_-$ can be viewed as a kind of monotonicity in the problem. One might say that this monotonicity makes up for the lack of remarkable structure in (1A-1a), including that it is not an Euler–Lagrange equation, its solutions admit no known coercive conserved quantities, and the nonlinearities are not signed. As we described in Section 1E, this monotonicity yields an important signed spacetime integral that we use to close the energy estimates; see the spacetime integral on the left-hand side of (3C-4). The largeness of $[\partial_t \Psi]_-$ is connected to so-called *Levi-type* conditions that have appeared in the literature. Roughly, a Levi condition is a quantitative relationship between the sizes of various coefficients in the equation and their derivatives. As an example, we note that in their study [D'Ancona and Trebeschi 2001] of well-posedness for (1F-2) with analytic coefficients, the authors studied linearized equations of the form (1F-1) under the Levi condition $|b(t,x)| \lesssim |a(t,x)| + |\partial_t \sqrt{a}(t,x)|$; the Levi condition allowed them, for the linearized equation, to construct suitable weights for the energy estimates (even at points where a vanishes), which were sufficient for proving well-posedness. In the problems under study here, the largeness of $[\partial_t \Psi]_-$ at points with $1+\Psi=0$ can be viewed as a Levi-type condition for the coefficient $(1+\Psi)^P$ in (1A-1a), which allows us to control various error terms that arise when we derive energy estimates for the solution's higher derivatives.

The degenerate energy estimates featured in our proofs have some features in common with the foundational works [Alinhac 1999a; 1999b; 2001; 2002; Christodoulou 2007] on the formation of shock singularities in solutions to quasilinear wave equations in two or three spatial dimensions; see also the

[28] The key point is that since our solutions are such that $\partial_t \Psi < 0$ when $1+\Psi=0$, it follows that $1+\Psi$ behaves, to first order, linearly in t near points where it vanishes.

follow-up works [Christodoulou and Miao 2014; Speck et al. 2016; Speck 2016; Ding et al. 2015a; 2015b; 2017] and the survey article [Holzegel et al. 2016]. In those works, the authors constructed a dynamic geometric coordinate system that degenerated in a precise fashion[29] as the shock formed. Consequently, relative to the geometric coordinates, the solution remains rather smooth,[30] which was a key fact used to control error terms. A crucial feature of the proofs is that the energy estimates[31] contained weights that vanished at the shock, which is in analogy with the vanishing of the weight $(1+\Psi)^P$ in (1E-1) at the degeneracy. A second crucial feature of the proofs of shock formation is that they relied on the fact that the weight has a *quantitatively strictly negative time derivative* in a neighborhood of points where it vanishes. This yields a critically important monotonic spacetime integral that is in analogy with the one, (1E-2), that we use to control various error terms in the present work.

We close this subsection by noting that the degeneracy that we encounter in our study of (1A-1a) is related to — but distinct from — a particular kind of absence of strict hyperbolicity that has been studied in the context of the compressible Euler equations for initial data satisfying the physical vacuum condition; see, for example, [Coutand et al. 2010; Coutand and Shkoller 2011; 2012; Jang and Masmoudi 2009; 2011]. The key difference between those works and ours is that in those works, the degeneracies occurred along the fluid-vacuum boundary in spacelike directions rather than a timelike one. In particular, the degeneracy was already present at time 0. More precisely, at time 0, the fluid density vanished at a certain rate, meaning that the density's derivative in the (spacelike) normal direction to the vacuum boundary satisfied a quantitative signed condition. It turns out that this condition yields a signed integral in the energy identities that is essential for closing the energy estimates. The signed integral exploited in those works is analogous to the integral (1E-2), but the integrals in the above papers were available because the solution's (spacelike) normal derivative had a sign, which is in contrast to the sign of the timelike derivative $\partial_t\Psi$ exploited in the present work. With the help of the signed integral, the authors of the above papers were able to prove degenerate energy estimates with weights that vanished at a certain rate in the normal direction to the vacuum boundary. Ultimately, these degenerate estimates allowed them to prove local well-posedness in Sobolev spaces with weights that degenerate at the fluid-vacuum boundary.

1G. *Notation.* In this subsection, we summarize some notation that we use throughout.

- $\{x^\alpha\}_{\alpha=0,1,2,3}$ denotes the standard rectangular coordinates on $\mathbb{R}^{1+3}=\mathbb{R}\times\mathbb{R}^3$, and $\partial_\alpha := \partial/\partial x^\alpha$ denotes the corresponding coordinate partial derivative vector fields; $x^0\in\mathbb{R}$ is the time coordinate and $\underline{x} := (x^1,x^2,x^3)\in\mathbb{R}^3$ are the spatial coordinates.
- We often use the alternate notation $x^0=t$ and $\partial_0=\partial_t$.

[29] In essence, the authors straightened out the characteristics via a solution-dependent change of coordinates.

[30] The high-order geometric energies were allowed to blow up at the shock, which led to enormous technical complications in the proofs. Note that this possible high-order energy blowup is distinct from the formation of the shock singularity, which corresponds to the blowup of a low-order Cartesian coordinate partial derivative of the solution.

[31] There are many shock-formation results for solutions to quasilinear equations in one spatial dimension, with important contributions coming from Riemann [1860], Oleĭnik [1957], Lax [1957], Klainerman and Majda [1980], John [1974; 1981; 1984], and many others. However, those results are based exclusively on the method of characteristics and hence, unlike in the case of two or more spatial dimensions, the proofs do not rely on energy estimates.

- Greek "spacetime" indices such as α vary over 0, 1, 2, 3 and Latin "spatial" indices such as a vary over 1, 2, 3. We use primed indices, such as a', in the same way that we use their nonprimed counterparts. We use Einstein's summation convention in that repeated indices are summed over their respective ranges.
- We raise and lower indices with g^{-1} and g respectively (*not* with the Minkowski metric!).
- We sometimes omit the arguments of functions appearing in pointwise inequalities. For example, we sometimes write $|f| \leq C\mathring{\epsilon}$ instead of $|f(t, \underline{x})| \leq C\mathring{\epsilon}$.
- $\nabla^k \Psi$ denotes the array comprising all derivatives of order k of Ψ with respect to the rectangular spatial coordinate vector fields. We often use the alternate notation $\nabla\Psi$ in place of $\nabla^1\Psi$. For example, $\nabla\Psi := (\partial_1\Psi, \partial_2\Psi, \partial_3\Psi)$.
- $|\nabla^{\leq k}\Psi| := \sum_{k'=0}^{k} |\nabla^{k'}\Psi|$.
- $|\nabla^{[a,b]}\Psi| := \sum_{k'=a}^{b} |\nabla^{k'}\Psi|$.
- $H^N(\Sigma_t)$ denotes the standard Sobolev space of functions on Σ_t with corresponding norm

$$\|f\|_{H^N(\Sigma_t)} := \left\{ \sum_{a_1+a_2+a_3 \leq N} \int_{\underline{x}\in\mathbb{R}^3} |\partial_1^{a_1}\partial_2^{a_2}\partial_3^{a_3} f(t, \underline{x})|^2 \, d\underline{x} \right\}^{1/2}.$$

In the case $N = 0$, we use the notation "L^2" in place of "H^0".

- $L^\infty(\Sigma_t)$ denotes the standard Lebesgue space of functions on Σ_t with corresponding norm $\|f\|_{L^\infty(\Sigma_t)} := \operatorname{ess\,sup}_{\underline{x}\in\mathbb{R}^3} |f(t, \underline{x})|$.
- If A and B are two quantities, then we often write $A \lesssim B$ to indicate that "there exists a constant $C > 0$ such that $A \leq CB$".
- We sometimes write $\mathcal{O}(B)$ to denote a quantity A with the following property: there exists a constant $C > 0$ such that $|A| \leq C|B|$.

2. Assumptions on the initial data and bootstrap assumptions

In this short section, we state our assumptions on the data $(\Psi|_{\Sigma_0}, \partial_t\Psi|_{\Sigma_0}) := (\mathring{\Psi}, \mathring{\Psi}_0)$ for the model equation (1A-1a) and formulate bootstrap assumptions that are convenient for studying the solution. We also show that there exist data satisfying the assumptions.

2A. ***Assumptions on the data.*** We assume that the initial data are compactly supported and satisfy the size assumptions

$$\|\nabla^{\leq 4}\mathring{\Psi}\|_{L^\infty(\Sigma_0)} + \|\nabla^{[1,3]}\mathring{\Psi}_0\|_{L^\infty(\Sigma_0)} + \|\nabla^2\mathring{\Psi}\|_{H^4(\Sigma_0)} + \|\nabla^2\mathring{\Psi}_0\|_{H^3(\Sigma_0)} \leq \mathring{\epsilon}, \quad \|\mathring{\Psi}_0\|_{L^\infty(\Sigma_0)} \leq \mathring{\delta}, \quad \text{(2A-1)}$$

where $\mathring{\epsilon}$ and $\mathring{\delta}$ are two data-size parameters that we will discuss below (roughly, $\mathring{\epsilon}$ will have to be small for our proofs to close). Roughly speaking, in our analysis, we will approximately propagate the above size assumptions all the way up until the time of breakdown in hyperbolicity, except at the top derivative level. More precisely, we are not able to uniformly control the top-order spatial derivatives of Ψ in the norm $\|\cdot\|_{L^2(\Sigma_t)}$ up to the time of breakdown due to the presence of degenerate weights in our energy (see Definition 3.4).

Before we can proceed, we must first introduce the crucial parameter $\mathring{\delta}_*$ that controls the time of first breakdown in hyperbolicity; our analysis shows that for $\mathring{\epsilon}$ sufficiently small, the time of first breakdown is $\{1+\mathcal{O}(\mathring{\epsilon})\}\mathring{\delta}_*^{-1}$; see also Remark 2.2.

Definition 2.1 (the parameter that controls the time of breakdown in hyperbolicity). We define the data-dependent parameter $\mathring{\delta}_*$ as

$$\mathring{\delta}_* := \max_{\Sigma_0}[\mathring{\Psi}_0]_-. \tag{2A-2}$$

Remark 2.2 (the relevance of $\mathring{\delta}_*$). The solutions that we study are such that[32] $\mathring{\Psi} \sim 0$ and $\partial_t^2\Psi \sim 0$ (throughout the evolution). Hence, by the fundamental theorem of calculus, we have $\partial_t\Psi(t,\underline{x}) \sim \mathring{\Psi}_0(\underline{x})$ and $1+\Psi(t,\underline{x}) \sim 1+t\mathring{\Psi}_0(\underline{x})$. From this last expression, we see that $1+\Psi$ is expected to vanish for the first time at approximately $t=\mathring{\delta}_*^{-1}$. See Lemma 3.1 for the precise statements.

2B. *Bootstrap assumptions.*

In proving our main results, we find it convenient to rely on a set of bootstrap assumptions, which we provide in this subsection.

The size of $T_{(\mathrm{Boot})}$: We assume that $T_{(\mathrm{Boot})}$ is a bootstrap time with

$$0 < T_{(\mathrm{Boot})} \le 2\mathring{\delta}_*^{-1}. \tag{2B-1}$$

The assumption (2B-1) gives us a sufficient margin of error to prove that finite-time degeneration of hyperbolicity occurs, as we explained in Remark 2.2.

Degeneracy has not yet occurred: We assume that for $(t,\underline{x}) \in [0,T_{(\mathrm{Boot})}) \times \mathbb{R}^3$ we have

$$0 < 1+\Psi(t,\underline{x}). \tag{2B-2}$$

L^∞ bootstrap assumptions: We assume that for $t \in [0,T_{(\mathrm{Boot})})$, we have

$$\|\Psi\|_{L^\infty(\Sigma_t)} \le 2\mathring{\delta}_*^{-1}\mathring{\delta} + \epsilon^{1/2}, \tag{2B-3a}$$

$$\|\partial_t\Psi\|_{L^\infty(\Sigma_t)} \le \mathring{\delta} + \epsilon^{1/2}, \tag{2B-3b}$$

$$\|\nabla^{[1,3]}\Psi\|_{L^\infty(\Sigma_t)} \le \epsilon, \quad \|\partial_t\nabla^{[1,3]}\Psi\|_{L^\infty(\Sigma_t)} \le \epsilon, \quad \|\partial_t^2\nabla^{\le 1}\Psi\|_{L^\infty(\Sigma_t)} \le \epsilon, \tag{2B-3c}$$

where $\epsilon > 0$ is a small bootstrap parameter; we describe our smallness assumptions in the next subsection.

Remark 2.3 (the solution remains compactly supported in space). From (2B-3a), we deduce that the wave speed $(1+\Psi)^{P/2}$ associated to (1A-1a) remains uniformly bounded from above for $(t,\underline{x}) \in [0,T_{(\mathrm{Boot})}) \times \mathbb{R}^3$. Hence, there exists a large, data-dependent ball $B \subset \mathbb{R}^3$ such that $\Psi(t,\underline{x})$ vanishes for $(t,\underline{x}) \in [0,T_{(\mathrm{Boot})}) \times B^c$, where B^c denotes the complement of B in $\mathbb{R}^3$.

2C. *Smallness assumptions.*

For the rest of the article, when we say that "A is small relative to B," we mean that $B > 0$ and that there exists a continuous increasing function $f:(0,\infty)\to(0,\infty)$ such that $A \le f(B)$. In principle, the functions f could always be chosen to be polynomials with positive

[32] Here "$A \sim B$" imprecisely indicates that A is well-approximated by B.

coefficients or exponential functions. However, to avoid lengthening the paper, we typically do not specify the form of f.

Throughout the rest of the paper, we make the following relative smallness assumptions. We continually adjust the required smallness in order to close our estimates.

- The bootstrap parameter ϵ from Section 2B is small relative to $\mathring{\delta}^{-1}$, where $\mathring{\delta}$ is the data-size parameter from (2A-1).
- ϵ is small relative to the data-size parameter $\mathring{\delta}_*$ from (2A-2).

The first assumption will allow us to control error terms that, roughly speaking, are of size $\epsilon\mathring{\delta}^k$ for some integer $k \geq 0$. The second assumption is relevant because the expected degeneracy-formation time is approximately $\mathring{\delta}_*^{-1}$ (see Remark 2.2); the assumption will allow us to show that various error products featuring a small factor ϵ remain small for $t \leq 2\mathring{\delta}_*^{-1}$, which is plenty of time for us to show that $1+\Psi$ vanishes.

Finally, we assume that

$$\epsilon^{3/2} \leq \mathring{\epsilon} \leq \epsilon, \tag{2C-1}$$

where $\mathring{\epsilon}$ is the data smallness parameter from (2A-1).

2D. *Existence of data.* It is easy to construct data such that the parameters $\mathring{\epsilon}$, $\mathring{\delta}$, and $\mathring{\delta}_*$ satisfy the relative size assumptions stated in Section 2C. For example, we can start with *any* smooth compactly supported data $(\mathring{\Psi}, \mathring{\Psi}_0)$ such that $\min_{\mathbb{R}^3} \mathring{\Psi}_0 < 0$. We then consider the one-parameter family

$$\big({}^{(\lambda)}\mathring{\Psi}(\underline{x}), {}^{(\lambda)}\mathring{\Psi}_0(\underline{x})\big) := \big(\lambda^{-1}\mathring{\Psi}(\underline{x}), \mathring{\Psi}_0(\lambda^{-1}\underline{x})\big).$$

One can check that for $\lambda > 0$ sufficiently large, all of the size assumptions of Section 2C are satisfied. The proof relies on the simple scaling identities

$$\nabla^k\, {}^{(\lambda)}\mathring{\Psi}(\underline{x}) = \lambda^{-1}(\nabla^k \mathring{\Psi})(\underline{x}), \tag{2D-1a}$$

$$\nabla^k\, {}^{(\lambda)}\mathring{\Psi}_0(\underline{x}) = \lambda^{-k}(\nabla^k \mathring{\Psi}_0)(\lambda^{-1}\underline{x}) \tag{2D-1b}$$

and

$$\|\nabla^k\, {}^{(\lambda)}\mathring{\Psi}\|_{L^2(\Sigma_0)} = \lambda^{-1}\|\mathring{\Psi}\|_{L^2(\Sigma_0)}, \tag{2D-2a}$$

$$\|\nabla^k\, {}^{(\lambda)}\mathring{\Psi}_0\|_{L^2(\Sigma_0)} = \lambda^{3/2-k}\|\mathring{\Psi}_0\|_{L^2(\Sigma_0)}. \tag{2D-2b}$$

Remark 2.4 (degeneracy occurs for solutions launched by any appropriately rescaled nontrivial data). The discussion in Section 2D can easily be extended to show that if $\mathring{\Psi}_0$ is nontrivial, then one *always* generates data to which our results apply by considering the rescaled data $({}^{(\lambda)}\mathring{\Psi}, {}^{(\lambda)}\mathring{\Psi}_0)$ with λ sufficiently large. More precisely, if $\min_{\mathbb{R}^3} \mathring{\Psi}_0 = 0$, then we must have $\max_{\mathbb{R}^3} \mathring{\Psi}_0 > 0$; in this case, the degeneracy in the solution generated by the rescaled data occurs in the past rather than the future.

3. A priori estimates

In this section, we use the data-size assumptions and the bootstrap assumptions of Section 2 to derive a priori estimates for the solution. This is the main step in the proof our results.

3A. ***Conventions for constants.*** In our estimates, the explicit constants $C > 0$ and $c > 0$ are free to vary from line to line. *These constants are allowed to depend on the data-size parameters $\mathring{\delta}$ and $\mathring{\delta}_*^{-1}$ from Section 2A*. However, the constants can be chosen to be independent of the parameters $\mathring{\epsilon}$ and ϵ whenever $\mathring{\epsilon}$ and ϵ are sufficiently small relative to $\mathring{\delta}^{-1}$ and $\mathring{\delta}_*$ in the sense described in Section 2C. For example, under our conventions, we have that $\mathring{\delta}_*^{-2}\epsilon = \mathcal{O}(\epsilon)$.

3B. ***Pointwise estimates.*** In this subsection, we derive pointwise estimates for Ψ and the inhomogeneous terms in the commuted wave equation.

We start with a simple lemma that provides sharp pointwise estimates for Ψ and $\partial_t \Psi$.

Lemma 3.1 (pointwise estimates for Ψ and $\partial_t\Psi$). *Under the data-size assumptions of Section 2A, the bootstrap assumptions of Section 2B, and the smallness assumptions of Section 2C, the following pointwise estimates hold for $(t, \underline{x}) \in [0, T_{(\mathrm{Boot})}) \times \mathbb{R}^3$:*

$$\Psi(t,\underline{x}) = t\mathring{\Psi}_0(\underline{x}) + \mathcal{O}(\epsilon), \tag{3B-1a}$$

$$\partial_t\Psi(t,\underline{x}) = \mathring{\Psi}_0(\underline{x}) + \mathcal{O}(\epsilon). \tag{3B-1b}$$

Proof. To derive (3B-1b), we first use the bootstrap assumptions to deduce $\|(1+\Psi)^P \Delta\Psi\|_{L^\infty(\Sigma_t)} \leq C\epsilon$. Hence, from (1A-1a), we deduce the pointwise bound $|\partial_t^2\Psi| \leq C\epsilon$. From this estimate and the fundamental theorem of calculus, we conclude the desired bound (3B-1b). The bound (3B-1a) then follows from the fundamental theorem of calculus, (3B-1b), and the small-data bound $\|\mathring{\Psi}\|_{L^\infty(\Sigma_0)} \leq \mathring{\epsilon} \leq \epsilon$. □

The next proposition captures the monotonicity that is present at points where $1+\Psi$ is small. It is of critical importance for the energy estimates.

Proposition 3.2 (monotonicity near the degeneracy). *Under the data-size assumptions of Section 2A, the bootstrap assumptions of Section 2B, and the smallness assumptions of Section 2C, the following statement holds for $(t, \underline{x}) \in [0, T_{(\mathrm{Boot})}) \times \mathbb{R}^3$:*

$$\Psi(t,\underline{x}) \leq -\tfrac{1}{2} \quad\Longrightarrow\quad \partial_t\Psi(t,\underline{x}) \leq -\tfrac{1}{8}\mathring{\delta}_*, \tag{3B-2}$$

where $\mathring{\delta}_*$ is the data-dependent parameter from Definition 2.1.

Proof. To prove (3B-2), we first use the estimates (3B-1a) and (3B-1b) to deduce that $\Psi(t,\underline{x}) = t\,\partial_t\Psi(t,\underline{x}) + \mathcal{O}(\epsilon)$. Hence, if $\Psi(t,\underline{x}) \leq -\frac{1}{2}$, then $t\,\partial_t\Psi(t,\underline{x}) \leq -\frac{1}{4}$. Recalling that $0 \leq t < 2\mathring{\delta}_*^{-1}$, see (2B-1), we conclude (3B-2). □

We now derive pointwise estimates for the inhomogeneous terms in the commuted wave equation.

Lemma 3.3 (pointwise estimates for the inhomogeneous terms). *Let Ψ be a solution to the wave equation (1A-1a). For $k = 2, 3, 4, 5$ and $P = 1, 2$, consider following wave equation,*[33] *obtained by commuting*

[33] We do not bother to state the precise form of $F^{(k)}$ here.

(1A-1a) *with* ∇^k:

$$-\partial_t^2 \nabla^k \Psi + (1+\Psi)^P \Delta \nabla^k \Psi = F^{(k)}. \tag{3B-3}$$

Under the data-size assumptions of Section 2A, the bootstrap assumptions of Section 2B, and the smallness assumptions of Section 2C, the following pointwise estimates hold for $(t, \underline{x}) \in [0, T_{(\mathrm{Boot})}) \times \mathbb{R}^3$:

$$|F^{(k)}| \le C\epsilon |\nabla^{[2,k+1]}\Psi| \qquad (P=1), \tag{3B-4}$$

$$|F^{(k)}| \le C\epsilon (1+\Psi)|\nabla^{k+1}\Psi| + \epsilon |\nabla^{[2,k]}\Psi| \quad (P=2). \tag{3B-5}$$

Proof. We first consider the case $P = 1$. Commuting (1A-1a) with ∇^k, we compute that

$$|F^{(k)}| \le C \sum_{\substack{a+b \le k+2 \\ 1 \le a,\, 2 \le b \le k+1}} |\nabla^a \Psi| |\nabla^b \Psi|.$$

The desired estimate (3B-4) then follows as a simple consequence of this bound and the bootstrap assumptions. The proof of (3B-5) is similar, the difference being that when $P = 2$, we have the bound

$$|F^{(k)}| \le C\epsilon (1+\Psi)|\nabla^{k+1}\Psi| + C \sum_{\substack{a+b \le k+2 \\ 1 \le a \le k,\, 2 \le b \le k}} |\nabla^a \Psi| |\nabla^b \Psi|. \qquad \square$$

3C. *Energy estimates.* We will use the following energy, which corresponds to between two and five commutations of the wave equation with ∇, in order to control solutions.

Definition 3.4 (the energy). We define

$$\mathcal{E}_{[2,5]}(t) := \sum_{k'=2}^{5} \int_{\Sigma_t} |\partial_t \nabla^{k'} \Psi|^2 + (1+\Psi)^P |\nabla \nabla^{k'} \Psi|^2 + |\nabla^{k'} \Psi|^2 \, d\underline{x}. \tag{3C-1}$$

We now provide the basic energy identity satisfied by solutions.

Lemma 3.5 (basic energy identity). *Let* Ψ *be a solution to the wave equation (1A-1a). Let* $\mathcal{E}_{[2,5]}$ *be the energy defined in (3C-1) and let* $F^{(k)}$ *be the inhomogeneous term from (3B-3). Then for* $P = 1, 2$, *we have the energy identity*

$$\begin{aligned}
\mathcal{E}_{[2,5]}(t) = \mathcal{E}_{[2,5]}(0) &+ P \sum_{k'=2}^{5} \int_{s=0}^{t} \int_{\Sigma_s} (\partial_t \Psi)(1+\Psi)^{P-1} |\nabla \nabla^{k'} \Psi|^2 \, d\underline{x}\, ds \\
&- 2P \sum_{k'=2}^{5} \int_{s=0}^{t} \int_{\Sigma_s} (1+\Psi)^{P-1} (\nabla \Psi) \cdot (\nabla \nabla^{k'} \Psi)(\partial_t \nabla^{k'} \Psi) \, d\underline{x}\, ds \\
&- 2 \sum_{k'=2}^{5} \int_{s=0}^{t} \int_{\Sigma_s} (\partial_t \nabla^{k'} \Psi) F^{(k')} \, d\underline{x}\, ds + 2 \sum_{k'=2}^{5} \int_{s=0}^{t} \int_{\Sigma_s} (\partial_t \nabla^{k'} \Psi)(\nabla^{k'} \Psi) \, d\underline{x}\, ds.
\end{aligned} \tag{3C-2}$$

Proof. The identity (3C-2) is standard and can verified by taking the time derivative of both sides of (3C-1), using (3B-3) for substitution, integrating by parts over Σ_t, and then integrating the resulting identity in time. $\square$

With the help of Lemma 3.5, we now derive an inequality satisfied by the energy $\mathcal{E}_{[2,5]}$.

Proposition 3.6 (integral inequality for the energy). *Let $\mathcal{E}_{[2,5]}$ be the energy defined in (3C-1). Let $\mathbf{1}_{\{-1<\Psi\leq -1/2\}}$ be the characteristic function of the spacetime subset $\{(t,\underline{x}) \mid -1<\Psi(t,\underline{x})\leq -\frac{1}{2}\}$ and define $\mathbf{1}_{\{-1/2<\Psi\}}$ analogously. Let $\mathring{\delta}_*$ be the data-size parameter from Definition 2.1. Under the data-size assumptions of Section 2A, the bootstrap assumptions of Section 2B, and the smallness assumptions of Section 2C, the following integral inequality holds for $P=1,2$ and $t\in[0,T_{(\text{Boot})})$:*

$$\begin{aligned}
\mathcal{E}_{[2,5]}(t)&+\tfrac{1}{8}P\mathring{\delta}_*\sum_{k'=2}^{5}\int_{s=0}^{t}\int_{\Sigma_s}\mathbf{1}_{\{-1<\Psi\leq -1/2\}}(1+\Psi)^{P-1}|\nabla\nabla^{k'}\Psi|^2\,d\underline{x}\,ds\\
&\leq \mathcal{E}_{[2,5]}(0)+C\sum_{k'=2}^{5}\int_{s=0}^{t}\int_{\Sigma_s}|\partial_t\nabla^{k'}\Psi|^2\,d\underline{x}\,ds+C\sum_{k'=2}^{5}\int_{s=0}^{t}\int_{\Sigma_s}|\nabla^{k'}\Psi|^2\,d\underline{x}\,ds\\
&\quad+C\sum_{k'=2}^{5}\int_{s=0}^{t}\int_{\Sigma_s}\mathbf{1}_{\{-1/2<\Psi\}}(1+\Psi)^{P-1}|\nabla\nabla^{k'}\Psi|^2\,d\underline{x}\,ds\\
&\quad+C\epsilon\sum_{k'=2}^{5}\int_{s=0}^{t}\int_{\Sigma_s}\mathbf{1}_{\{-1<\Psi\leq -1/2\}}(1+\Psi)^{2(P-1)}|\nabla\nabla^{k'}\Psi|^2\,d\underline{x}\,ds.
\end{aligned}\tag{3C-3}$$

Proof. We must bound the terms appearing in the energy identity (3C-2). We give the proof only for the case $P=1$ since the case $P=2$ can be handled using similar arguments. We start by bounding the first sum on the right-hand side of (3C-2); this is the only term that requires careful treatment. We split the integration domain $[0,t]\times\mathbb{R}^3$ into two pieces: a piece in which $-1<\Psi\leq-\frac{1}{2}$ and a piece in which $\Psi>-\frac{1}{2}$. To bound the first piece, we use the estimate (3B-2) to deduce that whenever $-1<\Psi\leq-\frac{1}{2}$, the integrand satisfies $(\partial_t\Psi)|\nabla\nabla^{k'}\Psi|^2\leq-\frac{1}{8}\mathring{\delta}_*|\nabla\nabla^{k'}\Psi|^2$. We can therefore bring all of the corresponding integrals over to the left-hand side of (3C-3) as *positive* integrals, as is indicated there. To bound the second piece, we use the estimate (3B-1b) to bound $\partial_t\Psi$ in L^∞ by $\leq C$, which allows us to bound the integrand by $C|\nabla\nabla^{k'}\Psi|^2$. It follows that since $\Psi>-\frac{1}{2}$ (by assumption), the integrals under consideration are bounded by the third sum on the right-hand side of (3C-3).

To bound the second sum on the right-hand side of (3C-2), we first use the bootstrap assumption (2B-3c) to bound the integrand factor $\nabla\Psi$ in L^∞ by $\leq\epsilon$. Thus, using Young's inequality, we bound the terms under consideration by the sum of the first, third, and fourth sums on the right-hand side of (3C-3).

To bound the third sum on the right-hand side of (3C-2), we use (3B-4) and Young's inequality to bound the integrand by $C\epsilon\sum_{k'=2}^{5}|\partial_t\nabla^{k'}\Psi|^2+C\epsilon\sum_{k'=2}^{6}|\nabla^{k'}\Psi|^2$. It is easy to see that the corresponding integrals are bounded by the right-hand side of (3C-3).

Finally, using Young's inequality, we bound last sum on the right-hand side of (3C-2) by the first two sums on the right-hand side of (3C-3). □

In the next proposition, we use Proposition 3.6 to derive our main a priori energy estimates. We also derive improvements of the bootstrap assumptions (2B-3).

Proposition 3.7 (a priori energy estimates and improvement of the bootstrap assumptions). *Let $\mathring{\delta}_*$ be the data-size parameter from (2A-2) and let $\mathbf{1}_{\{-1<\Psi\leq-1/2\}}$ be the characteristic function of the spacetime* *subset* $\{(t, \underline{x}) \mid -1 < \Psi(t, \underline{x}) \leq -\frac{1}{2}\}$. *There exists a constant* $C > 0$ *such that under the data-size assumptions of Section 2A, the bootstrap assumptions of Section 2B, and the smallness assumptions of Section 2C, the following a priori energy estimate holds for* $P = 1, 2$ *and* $t \in [0, T_{(\text{Boot})})$:

$$\mathcal{E}_{[2,5]}(t) + \tfrac{1}{16} P \mathring{\delta}_* \sum_{k'=2}^{5} \int_{s=0}^{t} \int_{\Sigma_s} \mathbf{1}_{\{-1<\Psi\leq -1/2\}} (1+\Psi)^{P-1} |\nabla \nabla^{k'} \Psi|^2 \, d\underline{x} \, ds \leq C \mathring{\epsilon}^2. \tag{3C-4}$$

Moreover, we have the following estimates, which are a strict improvement of the bootstrap assumptions (2B-3) for $\mathring{\epsilon}$ *sufficiently small*:

$$\|\Psi\|_{L^\infty(\Sigma_t)} \leq 2\mathring{\delta}_*^{-1} \mathring{\delta} + C\mathring{\epsilon}, \tag{3C-5a}$$

$$\|\partial_t \Psi\|_{L^\infty(\Sigma_t)} \leq \mathring{\delta} + C\mathring{\epsilon}, \tag{3C-5b}$$

$$\|\nabla^{[1,3]} \Psi\|_{L^\infty(\Sigma_t)} \leq C\mathring{\epsilon}, \quad \|\partial_t \nabla^{[1,3]} \Psi\|_{L^\infty(\Sigma_t)} \leq C\mathring{\epsilon}, \quad \|\partial_t^2 \nabla^{\leq 1} \Psi\|_{L^\infty(\Sigma_t)} \leq C\mathring{\epsilon}. \tag{3C-5c}$$

Proof. We give the proof only for the case $P = 1$ since the case $P = 2$ can be handled using similar arguments. To obtain (3C-4), we first note that for ϵ sufficiently small relative to $\mathring{\delta}_*$, we can absorb the last sum on the right-hand side of (3C-3) into the second term on the left-hand side, which at most reduces its coefficient from $\frac{1}{8}\mathring{\delta}_*$ to $\frac{1}{16}\mathring{\delta}_*$, as is stated on the left-hand side of (3C-4). Moreover, since $\mathbf{1}_{\{-1/2<\Psi\}} \leq C\,\mathbf{1}_{\{-1/2<\Psi\}}(1+\Psi)$, we have the bound $\int_{\Sigma_s} \mathbf{1}_{\{-1/2<\Psi\}} |\nabla\nabla^{k'}\Psi|^2 \, d\underline{x} \leq C\mathcal{E}_{[2,5]}(s)$ for the terms in the next-to-last sum on the right-hand side of (3C-3). The remaining Σ_s integrals are easily seen to be bounded in magnitude by $\leq C\mathcal{E}_{[2,5]}(s)$. Also using the data bound $\mathcal{E}_{[2,5]}(0) \leq C\mathring{\epsilon}^2$, which follows from our data assumptions (2A-1), we obtain

$$\mathcal{E}_{[2,5]}(t) + \tfrac{1}{16} \mathring{\delta}_* \sum_{k'=2}^{5} \int_{s=0}^{t} \int_{\Sigma_s} \mathbf{1}_{\{-1<\Psi\leq -1/2\}} |\nabla \nabla^{k'} \Psi|^2 \, d\underline{x} \, ds \leq C \mathring{\epsilon}^2 + c \int_{s=0}^{t} \mathcal{E}_{[2,5]}(s) \, ds. \tag{3C-6}$$

From (3C-6), Grönwall's inequality, and (2B-1), we conclude

$$\mathcal{E}_{[2,5]}(t) + \tfrac{1}{16} \mathring{\delta}_* \sum_{k'=2}^{5} \int_{s=0}^{t} \int_{\Sigma_s} \mathbf{1}_{\{-1<\Psi\leq -1/2\}} |\nabla \nabla^{k'} \Psi|^2 \, d\underline{x} \, ds \leq C \exp(ct) \mathring{\epsilon}^2 \leq C \mathring{\epsilon}^2,$$

which is the desired bound (3C-4).

The estimates (3C-5c) for $\nabla^{[2,3]}\Psi$ and $\partial_t \nabla^{[2,3]}\Psi$ then follow from (3C-4) and the Sobolev embedding result $H^2(\mathbb{R}^3) \hookrightarrow L^\infty(\mathbb{R}^3)$. Next, we take up to one ∇ derivative of (1A-1a) and use the already obtained L^∞ estimates and the bootstrap assumptions to obtain the bound $\|\partial_t^2 \nabla^{\leq 1}\Psi\|_{L^\infty(\Sigma_t)} \leq C\mathring{\epsilon}$ stated in (3C-5c). Using this bound, the fundamental theorem of calculus, and the data assumptions $\|\nabla \mathring{\Psi}_0\|_{L^\infty(\Sigma_0)} \leq \mathring{\epsilon}$ and $\|\mathring{\Psi}_0\|_{L^\infty(\Sigma_0)} \leq \mathring{\delta}$, we obtain the bounds $\|\partial_t \nabla\Psi\|_{L^\infty(\Sigma_t)} \leq C\mathring{\epsilon}$ and $\|\partial_t \Psi\|_{L^\infty(\Sigma_t)} \leq \mathring{\delta} + C\mathring{\epsilon}$, which in particular yields (3C-5b). Using a similar argument based on the already obtained bound $\|\partial_t \nabla\Psi\|_{L^\infty(\Sigma_t)} \leq C\mathring{\epsilon}$, we deduce $\|\nabla\Psi\|_{L^\infty(\Sigma_t)} \leq C\mathring{\epsilon}$. Similarly, from the already obtained bound $\|\partial_t \Psi\|_{L^\infty(\Sigma_t)} \leq \mathring{\delta} + C\mathring{\epsilon}$, the fundamental theorem of calculus, the initial data bound $\|\mathring{\Psi}\|_{L^\infty(\Sigma_0)} \leq \mathring{\epsilon}$, and the fact that $0 \leq t < T_{(\text{Boot})} \leq 2\mathring{\delta}_*^{-1}$, we deduce $\|\Psi\|_{L^\infty(\Sigma_t)} \leq 2\mathring{\delta}_*^{-1}\mathring{\delta} + C\mathring{\epsilon}$, that is, (3C-5a). □

4. The main results

We now derive our main results, namely Theorem 4.1 and Proposition 4.2.

Theorem 4.1 (stable finite-time degeneration of hyperbolicity). *Let* $(\mathring{\Psi}, \mathring{\Psi}_0) \in H^6(\mathbb{R}^3) \times H^5(\mathbb{R}^3)$ *be compactly supported initial data* (1A-1b) *for the wave equation* (1A-1a) *with* $P \in \{1, 2\}$ *and let* Ψ *denote the corresponding solution. Let*

$$\mathcal{M}(t) := \min_{(s,\underline{x})\in[0,t]\times\mathbb{R}^3} \{1 + \Psi(s, \underline{x})\}. \tag{4-1}$$

Let $\mathring{\epsilon}$, $\mathring{\delta}$, *and* $\mathring{\delta}_*$ *be the data-size parameters from* (2A-1)–(2A-2) *and assume that* $\mathring{\delta} > 0$ *and* $\mathring{\delta}_* > 0$. *Note that if* $\mathring{\epsilon}$ *is sufficiently small, then* $\mathcal{M}(0) = 1 + \mathcal{O}(\mathring{\epsilon}) > 0$. *If* $\mathring{\epsilon}$ *is sufficiently small relative to* $\mathring{\delta}^{-1}$ *and* $\mathring{\delta}_*$ *in the sense described in* Section 2C, *then the following conclusions hold.*

Breakdown in hyperbolicity precisely at time $T_\star$: *There exists a* $T_\star > 0$ *satisfying*

$$T_\star = \{1 + \mathcal{O}(\mathring{\epsilon})\}\mathring{\delta}_*^{-1} \tag{4-2}$$

such that the solution exists classically on the slab $[0, T_\star) \times \mathbb{R}^3$ *and such that the following inequality holds for* $0 \le t < T_\star$:

$$\mathcal{M}(t) > 0. \tag{4-3}$$

Moreover,

$$\lim_{t\uparrow T_\star} \mathcal{M}(t) = 0. \tag{4-4}$$

Regularity properties on $[0, T_\star) \times \mathbb{R}^3$: *On the slab* $[0, T_\star) \times \mathbb{R}^3$, *the solution satisfies the energy bounds* (3C-4), *the* L^∞ *estimates* (3C-5) *and the pointwise estimates* (3B-1) (*with* $C\mathring{\epsilon}$ *on the right-hand side in place of* ϵ *in the latter two estimates*). *Moreover,*

$$\Psi \in C\big([0, T_\star), H^6(\mathbb{R}^3)\big) \cap L^\infty\big([0, T_\star), H^5(\mathbb{R}^3)\big), \tag{4-5a}$$

$$\partial_t\Psi \in C\big([0, T_\star), H^5(\mathbb{R}^3)\big) \cap L^\infty\big([0, T_\star), H^5(\mathbb{R}^3)\big). \tag{4-5b}$$

Regularity properties on $[0, T_\star] \times \mathbb{R}^3$: Ψ *extends to a classical solution on the closed slab* $[0, T_\star] \times \mathbb{R}^3$ *enjoying the following regularity properties*: *for any* $N < 5$, *we have*

$$\Psi \in C\big([0, T_\star], H^5(\mathbb{R}^3)\big), \tag{4-6a}$$

$$\partial_t\Psi \in C\big([0, T_\star], H^N(\mathbb{R}^3)\big) \cap L^\infty\big([0, T_\star], H^5(\mathbb{R}^3)\big). \tag{4-6b}$$

In particular, the L^∞ *estimates* (3C-5) *and the pointwise estimates* (3B-1) (*with* $C\mathring{\epsilon}$ *on the right-hand side in place of* ϵ *in these estimates*) *hold on* $[0, T_\star] \times \mathbb{R}^3$. *Moreover, in the case*[34] $P = 1$, *we have*

$$\Psi \in L^2\big([0, T_\star], H^6(\mathbb{R}^3)\big). \tag{4-7}$$

Description of the breakdown along $\Sigma_{T_\star}$: *The set*

$$\Sigma_{T_\star}^{\text{Degen}} := \{q \in \Sigma_{T_\star} \mid 1 + \Psi(q) = 0\} \tag{4-8}$$

[34] In the proof of the theorem, we clarify why our proof of (4-7) relies on the assumption $P = 1$.

is nonempty and we have the estimate

$$\sup_{\Sigma_{T_\star}^{\text{Degen}}} \partial_t \Psi \leq -\tfrac{1}{8}\mathring{\delta}_*. \tag{4-9}$$

*In particular, in the case $P = 1$, the hyperbolicity of the wave equation breaks down on $\Sigma_{T_\star}^{\text{Degen}}$ in the following sense: if $q \in \Sigma_{T_\star}^{\text{Degen}}$, then any C^1 extension of Ψ to any spacetime neighborhood Ω_q containing q necessarily contains points p such that (1A-1a) is elliptic at $\Psi(p)$. In contrast, in the case $P = 2$, only the **strict** hyperbolicity (in the sense of Footnote 7) of (1A-1a) breaks down for the first time at $T_\star$.*

Proof. Let $T_\star$ be the supremum of the set of times $T_{(\text{Boot})}$ subject to inequality (2B-1) and such that the solution exists classically on the slab $[0, T_{(\text{Boot})}) \times \mathbb{R}^3$, has the same Sobolev regularity as the initial data, and satisfies the bootstrap assumptions of Section 2B with $\epsilon := C_*\mathring{\epsilon}$, where C_* is described just below. By standard local well-posedness, see for example [Hörmander 1997], if $\mathring{\epsilon}$ is sufficiently small and $C_* > 1$ is sufficiently large, note that this is consistent with the assumed inequalities (2C-1), then $T_\star > 0$. Next, we state the following standard continuation result, which can be proved, for example, by making straightforward modifications to the proof of [Hörmander 1997, Theorem 6.4.11]: if $T_\star < 2\mathring{\delta}_*^{-1}$, then the solution can be classically continued to a slab of the form $[0, T_\star + \triangle] \times \mathbb{R}^3$ (for some $\triangle > 0$ with $T_\star + \Delta < 2\mathring{\delta}_*^{-1}$) on which the solution has the same Sobolev regularity as the initial data and on which the bootstrap assumptions hold, as long as the bootstrap inequalities (2B-3) are strictly satisfied for $t \in [0, T_\star)$ and $\inf_{t\in[0,T_\star)} \mathcal{M}(t) > 0$. It follows that either (i) $T_\star = 2\mathring{\delta}_*^{-1}$, (ii) that the bootstrap inequalities (2B-3) are saturated at some time $t \in [0, T_\star)$, or (iii) that $\inf_{t\in[0,T_\star)} \mathcal{M}(t) = 0$. If C_* is chosen to be sufficiently large and $\mathring{\epsilon}$ is chosen to be sufficiently small, then the a priori estimates (3C-5) ensure that the bootstrap inequalities (2B-3) are in fact strictly satisfied for $t \in [0, T_\star)$. Moreover, from (2A-2) and the estimate (3B-1a) (which is now known to hold with ϵ replaced by $C\mathring{\epsilon}$), we see that $\min_{\Sigma_t}(1+\Psi) = 1 - \mathring{\delta}_* t + \mathcal{O}(\mathring{\epsilon})$ and thus, in fact, case (iii) occurs with $T_\star = \mathring{\delta}_*^{-1} + \mathcal{O}(\mathring{\epsilon}) = \{1 + \mathcal{O}(\mathring{\epsilon})\}\mathring{\delta}_*^{-1} < 2\mathring{\delta}_*^{-1}$ and $\lim_{t\uparrow T_\star} \mathcal{M}(t) = 0$. From the above reasoning, we easily deduce that the energy bound (3C-4) holds for $t \in [0, T_\star)$ and, since the a priori estimates (3C-5) show that the bootstrap assumptions hold with ϵ replaced by $C\mathring{\epsilon}$, that the pointwise estimates (3B-1) hold for $(t, \underline{x}) \in [0, T_\star) \times \mathbb{R}^3$ with ϵ replaced by $C\mathring{\epsilon}$.

In the rest of this proof, we sometimes silently use the simple facts that $\Psi, \partial_t\Psi \in L^\infty\big([0, T_\star), H^1(\mathbb{R}^3)\big)$. These facts do not follow from the energy estimates (3C-4) (since the energy does not directly control Ψ, $\partial_t\Psi$, $\nabla\Psi$ or $\nabla\partial_t\Psi$), but instead follow from (3C-5) and the compactly supported (in space) nature of the solution. To proceed, we easily conclude from the definition of $\mathcal{E}_{[2,5]}(t)$ and the fact that the estimates (3C-4) and (3C-5b)–(3C-5c) hold on $[0, T_\star)$ that $\partial_t\Psi \in L^\infty\big([0, T_\star], H^5(\mathbb{R}^3)\big)$, as is stated in (4-6b). Also, this fact trivially implies the corresponding statement in (4-5b), where the closed time interval is replaced with $[0, T_\star)$. The facts that $\Psi \in C\big([0, T_\star), H^6(\mathbb{R}^3)\big)$ and $\partial_t\Psi \in C\big([0, T_\star), H^5(\mathbb{R}^3)\big)$, as is stated in (4-5a) and (4-5b), are standard results that can be proved using energy estimates and simple facts from functional analysis. We omit the details and instead refer the reader to [Speck 2008, Section 2.7.5]. We note that in proving these "soft" facts, it is important that $\mathcal{M}(t) > 0$ for $t \in [0, T_\star)$, which implies that standard techniques for strictly hyperbolic equations can be used. To obtain the conclusion

$\Psi \in L^2\big([0, T_\star], H^6(\mathbb{R}^3)\big)$ in the case $P = 1$, as is stated in (4-7), we simply use the fact that the energy bounds (3C-4) hold on $[0, T_\star)$ (including the bound for the spacetime integral term on the left-hand side). Note that the same argument does not apply in the case $P = 2$ since in this case, the spacetime integral on the left-hand side of (3C-4) features the degenerate weight $1 + \Psi$. The fact that $\Psi \in C\big([0, T_\star], H^5(\mathbb{R}^3)\big)$, as is stated in (4-6a), is a simple consequence of the fundamental theorem of calculus, the fact that $\mathring{\Psi} \in H^6(\Sigma_0)$, and the already proven fact that $\partial_t \Psi \in L^\infty\big([0, T_\star], H^5(\mathbb{R}^3)\big)$. To obtain that for $N < 5$ we have $\partial_t \Psi \in C\big([0, T_\star], H^N(\mathbb{R}^3)\big)$, as is stated in (4-6b), we first use (1A-1a), the fact that $\Psi \in C\big([0, T_\star], H^5(\mathbb{R}^3)\big)$, and the standard Sobolev calculus to obtain $\partial_t^2 \Psi \in C\big([0, T_\star], H^3(\mathbb{R}^3)\big)$. Hence, from the fundamental theorem of calculus and the fact that $\mathring{\Psi}_0 \in H^5(\Sigma_0)$, we obtain $\partial_t \Psi \in C\big([0, T_\star], H^3(\mathbb{R}^3)\big)$. From this fact and the fact $\partial_t \Psi \in L^\infty\big([0, T_\star], H^5(\mathbb{R}^3)\big)$, we obtain, by interpolating[35] between L^2 and H^5, the desired conclusion $\partial_t \Psi \in C\big([0, T_\star], H^N(\mathbb{R}^3)\big)$.

Next, we note that the arguments given in the first paragraph of this proof imply that $\mathcal{M}$ extends as a continuous decreasing function defined for $t \in [0, T_\star]$ such that $\mathcal{M}(t) > 0$ for $t \in [0, T_\star)$ and such that $\mathcal{M}(T_\star) = 0$. Also using that $\Psi \in C\big([0, T_\star], H^5(\mathbb{R}^3)\big) \subset C\big([0, T_\star], C^3(\mathbb{R}^3)\big)$, we deduce, in view of definitions (4-1) and (4-8), that $\Sigma_{T_\star}^{\text{Degen}}$ is nonempty. Moreover, from (3B-2) and the fact that $\partial_t \Psi \in C\big([0, T_\star], H^{4.9}(\mathbb{R}^3)\big) \subset C\big([0, T_\star], C^3(\mathbb{R}^3)\big)$, we find that the estimate (4-9) holds. In addition, in view of (4-9), we see that in the case $P = 1$, if $q \in \Sigma_{T_\star}^{\text{Degen}}$, then any C^1 extension of Ψ to a neighborhood of q contains points p such that $1 + \Psi(p) < 0$, which renders (1A-1a) elliptic. This is in contrast to the case $P = 2$ in the sense that (1A-1a) is hyperbolic for all values of Ψ. □

Theorem 4.1 yields that Ψ remains regular, all the way up to the time $T_\star$. However, as the next proposition shows, a type of invariant blowup does in fact occur at time $T_\star$ in both the cases $P = 1, 2$. The blowup is tied to the Riemann curvature of the metric g.

Proposition 4.2 (blowup of the Kretschmann scalar). *Let $g = g(\Psi)$ denote the spacetime metric defined in (1A-2) and let $\mathrm{Riem}(g)$ denote the Riemann curvature tensor[36] of g. Under the assumptions and conclusions of Theorem 4.1, we have the following estimate for the Kretschmann scalar $\mathrm{Riem}(g)^{\alpha\beta\gamma\delta}\,\mathrm{Riem}(g)_{\alpha\beta\gamma\delta}$ on $[0, T_\star] \times \mathbb{R}^3$:*

$$\mathrm{Riem}(g)^{\alpha\beta\gamma\delta}\,\mathrm{Riem}(g)_{\alpha\beta\gamma\delta} = \frac{15}{2}\frac{(\partial_t \Psi)^4}{(1+\Psi)^4} + \mathcal{O}\left(\frac{\mathring{\epsilon}}{(1+\Psi)^3}\right) \quad (P = 1), \tag{4-10a}$$

$$\mathrm{Riem}(g)^{\alpha\beta\gamma\delta}\,\mathrm{Riem}(g)_{\alpha\beta\gamma\delta} = 60\frac{(\partial_t \Psi)^4}{(1+\Psi)^4} + \mathcal{O}\left(\frac{\mathring{\epsilon}}{(1+\Psi)^3}\right) \quad (P = 2). \tag{4-10b}$$

In particular, $\mathrm{Riem}(g)^{\alpha\beta\gamma\delta}\,\mathrm{Riem}(g)_{\alpha\beta\gamma\delta}$ is bounded for $0 \le t < T_\star$, while by (3B-2) and (4-10) at time $T_\star$, $\mathrm{Riem}(g)^{\alpha\beta\gamma\delta}\,\mathrm{Riem}(g)_{\alpha\beta\gamma\delta}$ blows up precisely on the subset $\Sigma_{T_\star}^{\text{Degen}}$ defined in (4-8).

Proof. We prove only (4-10a) since the proof of (4-10b) is similar. The identities that we state in this proof rely on the form of the metric (1A-2). We first note the following simple decomposition formula,

[35] Here, we mean the following standard inequality: if $f \in H^5(\Sigma_t)$ and $0 \le N \le 5$, then there exists a constant $C_N > 0$ such that $\|f\|_{H^N(\Sigma_t)} \le C_N \|f\|_{L^2(\Sigma_t)}^{1-N/5} \|f\|_{H^5(\Sigma_t)}^{N/5}$.

[36] Our sign convention for curvature is $D_\alpha D_\beta X_\mu - D_\beta D_\alpha X_\mu = \mathrm{Riem}(g)_{\alpha\beta\mu\nu} X^\nu$, where D denotes the Levi-Civita connection of g and X is an arbitrary smooth vector field.

which relies on the standard symmetry and antisymmetry properties of the Riemann curvature tensor:

$$\mathrm{Riem}(g)^{\alpha\beta\gamma\delta}\,\mathrm{Riem}(g)_{\alpha\beta\gamma\delta} = \mathrm{Riem}(g)_{ab}^{\ \ cd}\,\mathrm{Riem}(g)_{cd}^{\ \ ab} + 4\,\mathrm{Riem}(g)_{a0}^{\ \ c0}\,\mathrm{Riem}(g)_{c0}^{\ \ a0} - 4g_{cc'}g_{dd'}g^{bb'}\,\mathrm{Riem}(g)_{0b}^{\ \ cd}\,\mathrm{Riem}(g)_{0b'}^{\ \ c'd'}. \tag{4-11}$$

Next, we let $\underline{g}$ denote the first fundamental form of Σ_t relative to g; that is, $\underline{g}_{ij} = g_{ij} = (1+\Psi)^{-P}\delta_{ij}$ for $i, j = 1, 2, 3$, where δ_{ij} denotes the standard Kronecker delta. We also let (recalling that $P = 1$ in the present context)

$$k^i_{\ j} := -(\underline{g}^{-1})^{ia}\big(\tfrac{1}{2}\mathcal{L}_{\partial_t}\underline{g}_{aj}\big) = \tfrac{1}{2}\{\partial_t \ln(1+\Psi)\}\delta^i_{\ j} \tag{4-12}$$

denote the $\big($type $\binom{1}{1}\big)$ second fundamental form of Σ_t relative to g, where $\mathcal{L}_{\partial_t}$ denotes Lie differentiation with respect to the vector field ∂_t and $\delta^i_{\ j}$ denotes the standard Kronecker delta. Standard calculations based on the Gauss and Codazzi equations for the Lorentzian manifold $(\mathbb{R}^{1+3}, g)$ yield, see for example [Rodnianski and Speck 2014b, Appendix A], the identities

$$\mathrm{Riem}(g)_{ab}^{\ \ cd} = k^c_{\ a}k^d_{\ b} - k^d_{\ a}k^c_{\ b} + \triangle_{ab}^{\ \ cd}, \tag{4-13}$$

$$\mathrm{Riem}(g)_{a0}^{\ \ c0} = (\partial_t \ln(1+\Psi))k^c_{\ a} + k^c_{\ e}k^e_{\ a} + \triangle_{a0}^{\ \ c0}, \tag{4-14}$$

$$\mathrm{Riem}(g)_{0b}^{\ \ cd} = \triangle_{0b}^{\ \ cd}, \tag{4-15}$$

where the error terms are

$$\triangle_{ab}^{\ \ cd} := \mathrm{Riem}(\underline{g})_{ab}^{\ \ cd}, \tag{4-16}$$

$$\triangle_{a0}^{\ \ c0} := -\frac{1}{1+\Psi}\partial_t((1+\Psi)k^c_{\ a}), \tag{4-17}$$

$$\triangle_{0b}^{\ \ cd} := (\underline{g}^{-1})^{ce}\,\partial_e(k^d_{\ b}) - (\underline{g}^{-1})^{de}\,\partial_e(k^c_{\ b}) + (\underline{g}^{-1})^{ce}\,\Gamma_{e\ f}^{\ d}k^f_{\ b} - (\underline{g}^{-1})^{ce}\Gamma_{e\ b}^{\ f}k^d_{\ f} - (\underline{g}^{-1})^{de}\Gamma_{e\ f}^{\ c}k^f_{\ b} + (\underline{g}^{-1})^{de}\Gamma_{e\ b}^{\ f}k^c_{\ f}, \tag{4-18}$$

and

$$\Gamma_{j\ k}^{\ i} := \tfrac{1}{2}(\underline{g}^{-1})^{ai}\{\partial_j\underline{g}_{ak} + \partial_k\underline{g}_{ja} - \partial_a\underline{g}_{jk}\} \tag{4-19}$$

are the Christoffel symbols[37] of $\underline{g}$. In (4-16), $\mathrm{Riem}(\underline{g})$ denotes the Riemann curvature tensor of $\underline{g}$. We note that in deriving (4-14) and (4-17), we have used the simple identity

$$-\partial_t(k^c_{\ a}) = (\partial_t \ln(1+\Psi))k^c_{\ a} - \frac{1}{1+\Psi}\partial_t((1+\Psi)k^c_{\ a}).$$

We will use the estimates of Theorem 4.1 to show that

$$\triangle_{ab}^{\ \ cd} := \mathcal{O}(\mathring{\epsilon})\frac{1}{1+\Psi}, \tag{4-20}$$

$$\triangle_{a0}^{\ \ c0} := \mathcal{O}(\mathring{\epsilon}), \tag{4-21}$$

$$\triangle_{0b}^{\ \ cd} := \mathcal{O}(\mathring{\epsilon})\frac{1}{1+\Psi}. \tag{4-22}$$

[37] Our index conventions for the Christoffel symbols are different than the ones used in many works on differential geometry.

The desired bound (4-10a) then follows from (4-11), (4-12), (4-13)–(4-15), (4-20)–(4-22), the simple estimates $g^{ij} = \mathcal{O}(1)(1+\Psi)$ and $g_{ij} = \mathcal{O}(1)(1+\Psi)^{-1}$, and straightforward calculations.

It remains for us to prove (4-20)–(4-22). To prove (4-20), we first use (4-19) and (3C-5) to deduce

$$\Gamma_{j\ k}^{\ i} = \mathcal{O}(\mathring{\epsilon})\frac{1}{1+\Psi}, \quad \partial_l \Gamma_{j\ k}^{\ i} = \mathcal{O}(\mathring{\epsilon})\frac{1}{(1+\Psi)^2}. \tag{4-23}$$

Since $\operatorname{Riem}(\underline{g})_{ab}^{\ \ cd}$ has the schematic structure $\operatorname{Riem}(\underline{g})_{ab}^{\ \ cd} = \underline{g}^{-1}\underline{\partial}\Gamma + \underline{g}^{-1}\Gamma \cdot \Gamma$ (where $\underline{\partial}$ denotes the gradient with respect to the spatial coordinates), we deduce from (4-23) and the simple estimate $(\underline{g}^{-1})^{ij} = \mathcal{O}(1)(1+\Psi)$ that

$$\operatorname{Riem}(\underline{g})_{ab}^{\ \ cd} = \mathcal{O}(\mathring{\epsilon})\frac{1}{1+\Psi},$$

which yields (4-20). To prove (4-21), we first use (4-12), (1A-1a), and the estimates (3C-5a) and (3C-5c) to deduce $\partial_t((1+\Psi)k^c_{\ a}) = \frac{1}{2}\partial_t^2\Psi\delta^c_{\ a} = \frac{1}{2}(1+\Psi)\,\Delta\Psi\delta^c_{\ a} = \mathcal{O}(\mathring{\epsilon})(1+\Psi)$. From this bound and (4-17), we conclude (4-21). To prove (4-22), we first use (4-12) and the estimates (3C-5) to deduce

$$k^i_{\ j} = \mathcal{O}(1)\frac{1}{1+\Psi}, \quad \partial_l k^i_{\ j} = \mathcal{O}(\mathring{\epsilon})\frac{1}{(1+\Psi)^2}. \tag{4-24}$$

From (4-23), (4-24), and the simple estimate $(\underline{g}^{-1})^{ij} = \mathcal{O}(1)(1+\Psi)$, we conclude (4-22). □

Acknowledgments

Speck would like to thank Yuusuke Sugiyama for bringing the references [Kato and Sugiyama 2013; Sugiyama 2013; 2016a; 2016b] to his attention, to Michael Dreher for pointing out the references [Manfrin 1996; 1999], to Willie Wong for pointing out the relevance of his work [Wong 2016], and to an anonymous referee for pointing out the work [Lerner et al. 2015].

References

[Alinhac 1978] S. Alinhac, "Branching of singularities for a class of hyperbolic operators", *Indiana Univ. Math. J.* **27**:6 (1978), 1027–1037. MR Zbl

[Alinhac 1999a] S. Alinhac, "Blowup of small data solutions for a quasilinear wave equation in two space dimensions", *Ann. of Math.* (2) **149**:1 (1999), 97–127. MR Zbl

[Alinhac 1999b] S. Alinhac, "Blowup of small data solutions for a class of quasilinear wave equations in two space dimensions, II", *Acta Math.* **182**:1 (1999), 1–23. MR Zbl

[Alinhac 2001] S. Alinhac, "The null condition for quasilinear wave equations in two space dimensions, II", *Amer. J. Math.* **123**:6 (2001), 1071–1101. MR Zbl

[Alinhac 2002] S. Alinhac, "A minicourse on global existence and blowup of classical solutions to multidimensional quasilinear wave equations", exposé I, in *Journées "Équations aux Dérivées Partielles"* (Forges-les-Eaux, 2002), Univ. Nantes, 2002. MR

[Alinhac 2003] S. Alinhac, "An example of blowup at infinity for a quasilinear wave equation", pp. 1–91 in *Autour de l'analyse microlocale*, edited by G. Lebeau, Astérisque **284**, Société Mathématique de France, Paris, 2003. MR Zbl

[Amano and Nakamura 1981] K. Amano and G. Nakamura, "Branching of singularities for degenerate hyperbolic operators and Stokes phenomena, II", *Proc. Japan Acad. Ser. A Math. Sci.* **57**:3 (1981), 164–167. MR Zbl

[Amano and Nakamura 1982] K. Amano and G. Nakamura, "Branching of singularities for degenerate hyperbolic operator and Stokes phenomena, III", *Proc. Japan Acad. Ser. A Math. Sci.* **58**:10 (1982), 432–435. MR Zbl

[Amano and Nakamura 1983] K. Amano and G. Nakamura, "Branching of singularities for degenerate hyperbolic operator and Stokes phenomena, IV", *Proc. Japan Acad. Ser. A Math. Sci.* **59**:2 (1983), 47–50. MR Zbl

[Amano and Nakamura 1984] K. Amano and G. Nakamura, "Branching of singularities for degenerate hyperbolic operators", *Publ. Res. Inst. Math. Sci.* **20**:2 (1984), 225–275. MR Zbl

[Ascanelli 2006] A. Ascanelli, "A degenerate hyperbolic equation under Levi conditions", *Ann. Univ. Ferrara Sez. VII Sci. Mat.* **52**:2 (2006), 199–209. MR Zbl

[Ascanelli 2007] A. Ascanelli, "Well posedness under Levi conditions for a degenerate second order Cauchy problem", *Rend. Semin. Mat. Univ. Padova* **117** (2007), 113–126. MR Zbl

[Boiti and Manfrin 2000] C. Boiti and R. Manfrin, "Some results of blow-up and formation of singularities for non-strictly hyperbolic equations", *Funkcial. Ekvac.* **43**:1 (2000), 87–119. MR Zbl

[Christodoulou 2007] D. Christodoulou, *The formation of shocks in 3-dimensional fluids*, Eur. Math. Soc., Zürich, 2007. MR Zbl

[Christodoulou and Lisibach 2016] D. Christodoulou and A. Lisibach, "Shock development in spherical symmetry", *Ann. PDE* **2**:1 (2016), art. id. 3. MR

[Christodoulou and Miao 2014] D. Christodoulou and S. Miao, *Compressible flow and Euler's equations*, Surveys of Modern Mathematics **9**, International Press, Somerville, MA, 2014. MR Zbl

[Colombini and Spagnolo 1982] F. Colombini and S. Spagnolo, "An example of a weakly hyperbolic Cauchy problem not well posed in C^∞", *Acta Math.* **148** (1982), 243–253. MR Zbl

[Costin et al. 2012] O. Costin, M. Huang, and W. Schlag, "On the spectral properties of $L_\pm$ in three dimensions", *Nonlinearity* **25**:1 (2012), 125–164. MR Zbl

[Coutand and Shkoller 2011] D. Coutand and S. Shkoller, "Well-posedness in smooth function spaces for moving-boundary 1-D compressible Euler equations in physical vacuum", *Comm. Pure Appl. Math.* **64**:3 (2011), 328–366. MR Zbl

[Coutand and Shkoller 2012] D. Coutand and S. Shkoller, "Well-posedness in smooth function spaces for the moving-boundary three-dimensional compressible Euler equations in physical vacuum", *Arch. Ration. Mech. Anal.* **206**:2 (2012), 515–616. MR Zbl

[Coutand et al. 2010] D. Coutand, H. Lindblad, and S. Shkoller, "A priori estimates for the free-boundary 3D compressible Euler equations in physical vacuum", *Comm. Math. Phys.* **296**:2 (2010), 559–587. MR Zbl

[D'Ancona 1994] P. D'Ancona, "Well posedness in C^∞ for a weakly hyperbolic second order equation", *Rend. Sem. Mat. Univ. Padova* **91** (1994), 65–83. MR Zbl

[D'Ancona and Spagnolo 1992] P. D'Ancona and S. Spagnolo, "Global solvability for the degenerate Kirchhoff equation with real analytic data", *Invent. Math.* **108**:2 (1992), 247–262. MR Zbl

[D'Ancona and Trebeschi 2001] P. D'Ancona and P. Trebeschi, "On the local solvability for a nonlinear weakly hyperbolic equation with analytic coefficients", *Comm. Partial Differential Equations* **26**:5-6 (2001), 779–811. MR Zbl

[Ding et al. 2015a] B. Ding, I. Witt, and H. Yin, "Blowup of classical solutions for a class of 3-D quasilinear wave equations with small initial data", *Differential Integral Equations* **28**:9-10 (2015), 941–970. MR Zbl

[Ding et al. 2015b] B. Ding, I. Witt, and H. Yin, "On the lifespan and the blowup mechanism of smooth solutions to a class of 2-D nonlinear wave equations with small initial data", *Quart. Appl. Math.* **73**:4 (2015), 773–796. MR Zbl

[Ding et al. 2017] B. Ding, I. Witt, and H. Yin, "Blowup time and blowup mechanism of small data solutions to general 2-D quasilinear wave equations", *Commun. Pure Appl. Anal.* **16**:3 (2017), 719–744. MR Zbl

[Donninger 2010] R. Donninger, "Nonlinear stability of self-similar solutions for semilinear wave equations", *Comm. Partial Differential Equations* **35**:4 (2010), 669–684. MR Zbl

[Donninger and Krieger 2013] R. Donninger and J. Krieger, "Nonscattering solutions and blowup at infinity for the critical wave equation", *Math. Ann.* **357**:1 (2013), 89–163. MR Zbl

[Donninger and Schörkhuber 2012] R. Donninger and B. Schörkhuber, "Stable self-similar blow up for energy subcritical wave equations", *Dyn. Partial Differ. Equ.* **9**:1 (2012), 63–87. MR Zbl

[Donninger and Schörkhuber 2014] R. Donninger and B. Schörkhuber, "Stable blow up dynamics for energy supercritical wave equations", *Trans. Amer. Math. Soc.* **366**:4 (2014), 2167–2189. MR Zbl

[Donninger et al. 2014] R. Donninger, M. Huang, J. Krieger, and W. Schlag, "Exotic blowup solutions for the u^5 focusing wave equation in $\mathbb{R}^3$", *Michigan Math. J.* **63**:3 (2014), 451–501. MR Zbl

[Dreher 1999] M. Dreher, *Local solutions to quasilinear weakly hyperbolic differential equations*, Ph.D. thesis, Technischen Universität Bergakademie Freiberg, 1999, http://www.math.uni-konstanz.de/∼dreher/papers/D-03.pdf.

[Duyckaerts et al. 2012] T. Duyckaerts, C. Kenig, and F. Merle, "Universality of the blow-up profile for small type II blow-up solutions of the energy-critical wave equation: the nonradial case", *J. Eur. Math. Soc.* (*JEMS*) **14**:5 (2012), 1389–1454. MR Zbl

[Ebihara 1985] Y. Ebihara, "On the existence of local smooth solutions for some degenerate quasilinear hyperbolic equations", *An. Acad. Brasil. Ciênc.* **57**:2 (1985), 145–152. MR Zbl

[Ebihara et al. 1986] Y. Ebihara, L. A. Medeiros, and M. M. Miranda, "Local solutions for a nonlinear degenerate hyperbolic equation", *Nonlinear Anal.* **10**:1 (1986), 27–40. MR Zbl

[Hamilton 1982] R. S. Hamilton, "The inverse function theorem of Nash and Moser", *Bull. Amer. Math. Soc.* (*N.S.*) **7**:1 (1982), 65–222. MR Zbl

[Han 2010] Q. Han, "Energy estimates for a class of degenerate hyperbolic equations", *Math. Ann.* **347**:2 (2010), 339–364. MR Zbl

[Han and Liu 2015] Q. Han and Y. Liu, "Degenerate hyperbolic equations with lower degree degeneracy", *Proc. Amer. Math. Soc.* **143**:2 (2015), 567–580. MR Zbl

[Han et al. 2003] Q. Han, J.-X. Hong, and C.-S. Lin, "Local isometric embedding of surfaces with nonpositive Gaussian curvature", *J. Differential Geom.* **63**:3 (2003), 475–520. MR Zbl

[Han et al. 2006] Q. Han, J.-X. Hong, and C.-S. Lin, "On the Cauchy problem of degenerate hyperbolic equations", *Trans. Amer. Math. Soc.* **358**:9 (2006), 4021–4044. MR Zbl

[Herrmann 2012] T. Herrmann, *H^∞ well-posedness for degenerate p-evolution operators*, Ph.D. thesis, Technischen Universität Bergakademie Freiberg, 2012, http://www.qucosa.de/fileadmin/data/qucosa/documents/9981/diss-herrmann-corp.pdf.

[Herrmann et al. 2013] T. Herrmann, M. Reissig, and K. Yagdjian, "H^∞ well-posedness for degenerate p-evolution models of higher order with time-dependent coefficients", pp. 125–151 in *Progress in partial differential equations*, edited by M. Reissig and M. Ruzhansky, Springer Proc. Math. Stat. **44**, Springer, 2013. MR Zbl

[Holzegel et al. 2016] G. Holzegel, S. Klainerman, J. Speck, and W. W.-Y. Wong, "Small-data shock formation in solutions to 3D quasilinear wave equations: an overview", *J. Hyperbolic Differ. Equ.* **13**:1 (2016), 1–105. MR Zbl

[Hörmander 1997] L. Hörmander, *Lectures on nonlinear hyperbolic differential equations*, Mathématiques & Applications **26**, Springer, 1997. MR Zbl

[Ishida and Yagdjian 2002] H. Ishida and K. Yagdjian, "On a sharp Levi condition in Gevrey classes for some infinitely degenerate hyperbolic equations and its necessity", *Publ. Res. Inst. Math. Sci.* **38**:2 (2002), 265–287. MR Zbl

[Ivriĭ 1975] V. J. Ivriĭ, "Conditions for correctness in Gevrey classes of the Cauchy problem for hyperbolic operators with characteristics of variable multiplicity", *Dokl. Akad. Nauk SSSR* **221**:6 (1975), 1253–1255. In Russian. MR

[Jang and Masmoudi 2009] J. Jang and N. Masmoudi, "Well-posedness for compressible Euler equations with physical vacuum singularity", *Comm. Pure Appl. Math.* **62**:10 (2009), 1327–1385. MR Zbl

[Jang and Masmoudi 2011] J. Jang and N. Masmoudi, "Vacuum in gas and fluid dynamics", pp. 315–329 in *Nonlinear conservation laws and applications*, edited by A. Bressan et al., IMA Vol. Math. Appl. **153**, Springer, 2011. MR Zbl

[John 1974] F. John, "Formation of singularities in one-dimensional nonlinear wave propagation", *Comm. Pure Appl. Math.* **27** (1974), 377–405. MR Zbl

[John 1981] F. John, "Blow-up for quasilinear wave equations in three space dimensions", *Comm. Pure Appl. Math.* **34**:1 (1981), 29–51. MR Zbl

[John 1984] F. John, "Formation of singularities in elastic waves", pp. 194–210 in *Trends and applications of pure mathematics to mechanics* (Palaiseau, 1983), Lecture Notes in Phys. **195**, Springer, 1984. MR Zbl

[Kato and Sugiyama 2013] K. Kato and Y. Sugiyama, "Blow up of solutions to the second sound equation in one space dimension", *Kyushu J. Math.* **67**:1 (2013), 129–142. MR Zbl

[Kenig and Merle 2008] C. E. Kenig and F. Merle, "Global well-posedness, scattering and blow-up for the energy-critical focusing non-linear wave equation", *Acta Math.* **201**:2 (2008), 147–212. MR Zbl

[Killip et al. 2014] R. Killip, B. Stovall, and M. Visan, "Blowup behaviour for the nonlinear Klein–Gordon equation", *Math. Ann.* **358**:1-2 (2014), 289–350. MR Zbl

[Klainerman and Majda 1980] S. Klainerman and A. Majda, "Formation of singularities for wave equations including the nonlinear vibrating string", *Comm. Pure Appl. Math.* **33**:3 (1980), 241–263. MR Zbl

[Krieger and Schlag 2014] J. Krieger and W. Schlag, "Full range of blow up exponents for the quintic wave equation in three dimensions", *J. Math. Pures Appl.* (9) **101**:6 (2014), 873–900. MR Zbl

[Krieger et al. 2008] J. Krieger, W. Schlag, and D. Tataru, "Renormalization and blow up for charge one equivariant critical wave maps", *Invent. Math.* **171**:3 (2008), 543–615. MR Zbl

[Krieger et al. 2009] J. Krieger, W. Schlag, and D. Tataru, "Slow blow-up solutions for the $H^1(\mathbb{R}^3)$ critical focusing semilinear wave equation", *Duke Math. J.* **147**:1 (2009), 1–53. MR Zbl

[Krieger et al. 2013a] J. Krieger, K. Nakanishi, and W. Schlag, "Global dynamics away from the ground state for the energy-critical nonlinear wave equation", *Amer. J. Math.* **135**:4 (2013), 935–965. MR Zbl

[Krieger et al. 2013b] J. Krieger, K. Nakanishi, and W. Schlag, "Global dynamics of the nonradial energy-critical wave equation above the ground state energy", *Discrete Contin. Dyn. Syst.* **33**:6 (2013), 2423–2450. MR Zbl

[Krieger et al. 2014] J. Krieger, K. Nakanishi, and W. Schlag, "Threshold phenomenon for the quintic wave equation in three dimensions", *Comm. Math. Phys.* **327**:1 (2014), 309–332. MR Zbl

[Krieger et al. 2015] J. Krieger, K. Nakanishi, and W. Schlag, "Center-stable manifold of the ground state in the energy space for the critical wave equation", *Math. Ann.* **361**:1-2 (2015), 1–50. MR Zbl

[Lax 1957] P. D. Lax, "Hyperbolic systems of conservation laws, II", *Comm. Pure Appl. Math.* **10** (1957), 537–566. MR Zbl

[Lerner et al. 2015] N. Lerner, T. T. Nguyen, and B. Texier, "The onset of instability in first-order systems", preprint, 2015. To appear in *J. Eur. Math. Soc.* (*JEMS*). arXiv

[Levine 1974] H. A. Levine, "Instability and nonexistence of global solutions to nonlinear wave equations of the form $Pu_{tt} = -Au + \mathcal{F}(u)$", *Trans. Amer. Math. Soc.* **192** (1974), 1–21. MR Zbl

[Lindblad 1998] H. Lindblad, "Counterexamples to local existence for quasilinear wave equations", *Math. Res. Lett.* **5**:5 (1998), 605–622. MR Zbl

[Lindblad 2008] H. Lindblad, "Global solutions of quasilinear wave equations", *Amer. J. Math.* **130**:1 (2008), 115–157. MR Zbl

[Luk 2013] J. Luk, "Weak null singularities in general relativity", preprint, 2013. arXiv

[Manfrin 1996] R. Manfrin, "Local solvability in C^∞ for quasi-linear weakly hyperbolic equations of second order", *Comm. Partial Differential Equations* **21**:9-10 (1996), 1487–1519. MR Zbl

[Manfrin 1999] R. Manfrin, "Well posedness in the C^∞ class for $u_{tt} = a(u)\Delta u$", *Nonlinear Anal.* **36**:2, Ser. A: Theory Methods (1999), 177–212. MR Zbl

[Martel et al. 2014] Y. Martel, F. Merle, P. Raphaël, and J. Szeftel, "Near soliton dynamics and singularity formation for L^2 critical problems", *Uspekhi Mat. Nauk* **69**:416 (2014), 77–106. In Russian; translated in *Russian Math. Surveys* **69**:2 (2014), 261–290.

[Nakanishi and Schlag 2011a] K. Nakanishi and W. Schlag, "Global dynamics above the ground state energy for the focusing nonlinear Klein–Gordon equation", *J. Differential Equations* **250**:5 (2011), 2299–2333. MR Zbl

[Nakanishi and Schlag 2011b] K. Nakanishi and W. Schlag, *Invariant manifolds and dispersive Hamiltonian evolution equations*, Eur. Math. Soc., Zürich, 2011. MR Zbl

[Nakanishi and Schlag 2012a] K. Nakanishi and W. Schlag, "Global dynamics above the ground state for the nonlinear Klein–Gordon equation without a radial assumption", *Arch. Ration. Mech. Anal.* **203**:3 (2012), 809–851. MR Zbl

[Nakanishi and Schlag 2012b] K. Nakanishi and W. Schlag, "Invariant manifolds around soliton manifolds for the nonlinear Klein–Gordon equation", *SIAM J. Math. Anal.* **44**:2 (2012), 1175–1210. MR Zbl

[Nishitani 1980] T. Nishitani, "The Cauchy problem for weakly hyperbolic equations of second order", *Comm. Partial Differential Equations* **5**:12 (1980), 1273–1296. MR Zbl

[Nishitani 1984] T. Nishitani, "A necessary and sufficient condition for the hyperbolicity of second order equations with two independent variables", *J. Math. Kyoto Univ.* **24**:1 (1984), 91–104. MR Zbl

[Oleĭnik 1957] O. A. Oleĭnik, "Discontinuous solutions of non-linear differential equations", *Uspehi Mat. Nauk* (*N.S.*) **12**:3(75) (1957), 3–73. In Russian; translated in *Amer. Math. Soc. Transl.* (*2*) **26** (1963), 95–172. MR Zbl

[Oleĭnik 1970] O. A. Oleĭnik, "On the Cauchy problem for weakly hyperbolic equations", *Comm. Pure Appl. Math.* **23** (1970), 569–586. MR

[Payne and Sattinger 1975] L. E. Payne and D. H. Sattinger, "Saddle points and instability of nonlinear hyperbolic equations", *Israel J. Math.* **22**:3-4 (1975), 273–303. MR

[Raphaël and Rodnianski 2012] P. Raphaël and I. Rodnianski, "Stable blow up dynamics for the critical co-rotational wave maps and equivariant Yang–Mills problems", *Publ. Math. Inst. Hautes Études Sci.* **115** (2012), 1–122. MR Zbl

[Riemann 1860] B. Riemann, "Über die Fortpflanzung ebener Luftwellen von endlicher Schwingungsweite", *Abh. König. Ges. Wiss. Göttengen* **8** (1860), 43–66.

[Rodnianski and Speck 2014a] I. Rodnianski and J. Speck, "A regime of linear stability for the Einstein-scalar field system with applications to nonlinear big bang formation", preprint, 2014. To appear in *Ann. of Math. (2)*. arXiv

[Rodnianski and Speck 2014b] I. Rodnianski and J. Speck, "Stable big bang formation in near-FLRW solutions to the Einstein-scalar field and Einstein-stiff fluid systems", preprint, 2014. arXiv

[Rodnianski and Sterbenz 2010] I. Rodnianski and J. Sterbenz, "On the formation of singularities in the critical O(3) σ-model", *Ann. of Math.* (2) **172**:1 (2010), 187–242. MR Zbl

[Ruan et al. 2015a] Z. Ruan, I. Witt, and H. Yin, "On the existence and cusp singularity of solutions to semilinear generalized Tricomi equations with discontinuous initial data", *Commun. Contemp. Math.* **17**:3 (2015), art. id. 1450028. MR Zbl

[Ruan et al. 2015b] Z. Ruan, I. Witt, and H. Yin, "On the existence of low regularity solutions to semilinear generalized Tricomi equations in mixed type domains", *J. Differential Equations* **259**:12 (2015), 7406–7462. MR Zbl

[Ruan et al. 2016] Z. Ruan, I. Witt, and H. Yin, "Minimal regularity solutions of semilinear generalized Tricomi equations", preprint, 2016. arXiv

[Ruziev and Reissig 2016] M. Ruziev and M. Reissig, "Tricomi type equations with terms of lower order", *Int. J. Dyn. Syst. Differ. Equ.* **6**:1 (2016), 1–15. MR

[Sideris 1985] T. C. Sideris, "Formation of singularities in three-dimensional compressible fluids", *Comm. Math. Phys.* **101**:4 (1985), 475–485. MR Zbl

[Speck 2008] J. R. Speck, *On the questions of local and global well-posedness for the hyperbolic PDEs occurring in some relativistic theories of gravity and electromagnetism*, Ph.D. thesis, Rutgers University, 2008, https://search.proquest.com/docview/304511925. MR

[Speck 2016] J. Speck, *Shock formation in small-data solutions to 3D quasilinear wave equations*, Mathematical Surveys and Monographs **214**, Amer. Math. Soc., Providence, RI, 2016. MR Zbl

[Speck et al. 2016] J. Speck, G. Holzegel, J. Luk, and W. Wong, "Stable shock formation for nearly simple outgoing plane symmetric waves", *Ann. PDE* **2**:2 (2016), art. id. 10. MR

[Struwe 2003] M. Struwe, "Equivariant wave maps in two space dimensions", *Comm. Pure Appl. Math.* **56**:7 (2003), 815–823. MR Zbl

[Sugiyama 2013] Y. Sugiyama, "Global existence of solutions to some quasilinear wave equation in one space dimension", *Differential Integral Equations* **26**:5-6 (2013), 487–504. MR Zbl

[Sugiyama 2016a] Y. Sugiyama, "Degeneracy in finite time of 1D quasilinear wave equations", *SIAM J. Math. Anal.* **48**:2 (2016), 847–860. MR Zbl

[Sugiyama 2016b] Y. Sugiyama, "Degeneracy in finite time of 1D quasilinear wave equations, II", preprint, 2016. arXiv

[Taniguchi and Tozaki 1980] K. Taniguchi and Y. Tozaki, "A hyperbolic equation with double characteristics which has a solution with branching singularities", *Math. Japon.* **25**:3 (1980), 279–300. MR Zbl

[Wong 2016] W. W.-Y. Wong, "Singularities of axially symmetric time-like minimal submanifolds in Minkowski space", preprint, 2016. arXiv

[Yagdzhyan 1989] K. A. Yagdzhyan, "The parametrix of the Cauchy problem for hyperbolic operators that are degenerate with respect to the space variables", *Izv. Akad. Nauk Armyan. SSR Ser. Mat.* **24**:5 (1989), 421–432. In Russian; translated in *Soviet J. Contemporary Math. Anal.* **24**: 5 (1990), 1–13. MR Zbl

Received 5 Feb 2017. Accepted 17 Jun 2017.

JARED SPECK: jspeck@math.mit.edu
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, United States

ANALYSIS AND PDE
Vol. 10, No. 8, 2017
dx.doi.org/10.2140/apde.2017.10.2031

DIMENSION OF THE MINIMUM SET FOR THE REAL AND COMPLEX MONGE–AMPÈRE EQUATIONS IN CRITICAL SOBOLEV SPACES

TRISTAN C. COLLINS AND CONNOR MOONEY

We prove that the zero set of a nonnegative plurisubharmonic function that solves $\det(\partial\bar{\partial}u) \geq 1$ in $\mathbb{C}^n$ and is in $W^{2,n(n-k)/k}$ contains no analytic subvariety of dimension k or larger. Along the way we prove an analogous result for the real Monge–Ampère equation, which is also new. These results are sharp in view of well-known examples of Pogorelov and Błocki. As an application, in the real case we extend interior regularity results to the case that u lies in a critical Sobolev space (or more generally, certain Sobolev–Orlicz spaces).

1. Introduction

In this paper we investigate the dimension of the singular set for the real and complex Monge–Ampère equations, assuming critical Sobolev regularity.

We first discuss the real case. It is well known that convex (e.g., viscosity) solutions to $\det D^2u = 1$ are not always classical solutions. Pogorelov constructed examples in dimension $n \geq 3$ of the form

$$u(x', x_n) = |x'|^{2-\frac{2}{n}} f(x_n)$$

that solve $\det D^2u = 1$ in $|x_n| < \rho$ for some $\rho > 0$ and some smooth, positive f. This example is $C^{1,\alpha}$ for $\alpha \leq 1 - \frac{2}{n}$, and $W^{2,p}$ for $p < \frac{n(n-1)}{2}$. Furthermore, this solution is not strictly convex, and it vanishes on $\{x' = 0\}$.

On the other hand, it is known that strictly convex solutions are smooth. The proof of this fact is closely related to the solution of the Dirichlet problem, which has a long history, beginning with work of Pogorelov [1971a; 1971b; 1973; 1978], Cheng and Yau [1976; 1977] and Calabi [1958]. Cheng and Yau [1976] solved the Minkowski problem on the sphere, and in [Cheng and Yau 1977] proved the existence of solutions to the Dirichlet problem which are smooth in the interior and Lipschitz up to the boundary. P. L. Lions [1983; 1985] gave an independent proof of this result. Caffarelli, Nirenberg and Spruck [Caffarelli et al. 1984] and Krylov [1983] established the existence of solutions smooth up to the boundary, provided the boundary data are $C^{3,1}$. Trudinger and Wang [2008] proved optimal boundary regularity results, where the optimality comes from earlier examples of Wang [1996].

Remark 1.1. In the case $n = 2$ it is a classical result of Alexandroff [1942] that solutions to $\det D^2u \geq 1$ are strictly convex.

Collins was supported in part by National Science Foundation grant DMS-1506652, the European Research Council and the Knut and Alice Wallenberg Foundation. Mooney was supported in part by National Science Foundation fellowship DMS-1501152 and by the ERC grant "Regularity and Stability in Partial Differential Equations (RSPDE)".
MSC2010: primary 32W20, 35J96; secondary 35B33, 35B65.
Keywords: Monge–Ampère, regularity, viscosity solution, Sobolev.

In view of the above discussion, to show interior regularity for $\det D^2u = 1$ it is enough to show strict convexity. (We remark that interior *estimates* generally depend on the modulus of strict convexity). Urbas [1988] showed strict convexity when u is in $C^{1,\alpha}$ for $\alpha > 1-\frac{2}{n}$, or in $W^{2,p}$ for $p > \frac{n(n-1)}{2}$. (Note that for these values of p, $W^{2,p}$ embeds into $C^{1,\alpha}$ for $\alpha > 1-\frac{2}{n-1}$, so neither result implies the other). Caffarelli [1990b] showed that if $1 \le \det D^2u < \Lambda$ and u is not strictly convex, then the graph of u contains an affine set with no interior extremal points, and if $\det D^2u \ge 1$, then the dimension of any affine set in the graph of u is strictly smaller than $\frac{n}{2}$ [Caffarelli 1993]; see also [Mooney 2015]. These results led to interior $C^{2,\alpha}$ and $W^{2,q}$ estimates for solutions with linear boundary data, when $\det D^2u$ is strictly positive and C^α, resp. C^0 [Caffarelli 1990a]. Finally, in [Mooney 2015] the second author showed that if $\det D^2u \ge 1$, then u is strictly convex away from a set of Hausdorff $(n-1)$-dimensional measure zero, and that this is optimal by example (even when $\det D^2u = 1$).

In view of the Pogorelov example, the $C^{1,\alpha}$ hypothesis in [Urbas 1988] is sharp, and the $W^{2,p}$ hypothesis is nearly sharp. In this paper we show interior regularity for the borderline case $p = \frac{n(n-1)}{2}$. Our result in the real case is:

Theorem 1.1. *Assume that u is a convex solution to* $\det D^2u \ge 1$ *in $B_1 \subset \mathbb{R}^n$, and let $0 < k < \frac{n}{2}$. If $u \in W^{2,p}(B_1)$ for some $p \ge \frac{n}{2k}(n-k)$, then the dimension of the set where u agrees with a tangent plane is at most $k-1$.*

Remark 1.2. In particular, if $u \in W^{2,\frac{n(n-1)}{2}}$, then it is strictly convex. We in fact show that u is strictly convex if Δu lies in Orlicz spaces that are slightly weaker than $L^{\frac{n(n-1)}{2}}$ (see Section 3), strengthening the result from [Urbas 1988]. Our result is sharp in view of the Pogorelov example.

As a consequence, we can extend interior estimates to the borderline case $p = \frac{n(n-1)}{2}$ (see Section 5). Interior estimates of this kind are often important in geometric applications, where one does not control the boundary data.

Remark 1.3. There are analogues of the Pogorelov example that vanish on sets of dimension k for any $k < \frac{n}{2}$ and are not in $W^{2,\frac{n}{2k}(n-k)}$ [Caffarelli 1993]. These show that Theorem 1.1 is also sharp in the case $k > 1$.

We now discuss the complex case. Like in the real case, there exist singular Pogorelov-type examples of the form

$$u(z', z_n) = |z'|^{2-\frac{2}{n}} f(z_n)$$

for $n \ge 2$, such that $\det(\partial\bar{\partial}u)$ is a strictly positive polynomial [Błocki 1999].

Remark 1.4. In fact, there are analogues of this example that vanish on sets of complex dimension k for any $k < n$. Furthermore, these singular examples are global.

Less is known about interior regularity for the complex Monge–Ampère equation $\det(\partial\bar{\partial}u) = 1$. Błocki and Dinew [2011] showed that if $u \in W^{2,p}$ for some $p > n(n-1)$, then u is smooth. This result relies on an important estimate of Kołodziej [1996]. The same result is true provided Δu is bounded; see, e.g., [Wang 2012]. In this case the point is that the operator becomes uniformly elliptic, and by its concavity an important $C^{2,\alpha}$ estimate of Evans and Krylov applies; see, e.g., [Caffarelli and Cabré 1995]. Thus far, there does not seem to be a geometric condition analogous to strict convexity that guarantees interior regularity.

However, if u is nonnegative then something can be said about analytic structures in the minimum set. A classical theorem of Harvey and Wells [1973] says that the minimum set of a smooth, strictly plurisubharmonic function is contained in a C^1, totally real submanifold. Dinew and Dinew [2016] recently showed that if $\det(\partial\bar{\partial}u)$ has a positive lower bound and $u \in C^{1,\alpha}$ for $\alpha > 1 - \frac{2k}{n}$, or C^{β} for $\beta > 2 - \frac{2k}{n}$ in the case $k > \frac{n}{2}$, then the minimum set of u contains no analytic subvarieties of dimension k or larger. We investigate the same situation assuming Sobolev regularity. In the complex case, our main result is:

Theorem 1.2. *Assume that u is a nonnegative plurisubharmonic function satisfying* $\det(\partial\bar{\partial}u) \geq 1$ *in the viscosity sense in $B_1 \subset \mathbb{C}^n$, and let $0 < k < n$. If $\Delta u \in L^p$ for some $p \geq \frac{n}{k}(n-k)$, then the zero set $\{u = 0\}$ contains no analytic subvarieties of dimension k or larger.*

Remark 1.5. We recall that $\det(\partial\bar{\partial}u) \geq 1$ in the viscosity sense if u is continuous in B_1, and whenever a quadratic polynomial P of the form $A_{\bar{j}i}z_i\bar{z}_j + \mathrm{Re}(Q(z_1, \ldots, z_n))$ such that A is a positive semidefinite Hermitian matrix touches u from above in B_1, then $\det(\partial\bar{\partial}P) \geq 1$.

Remark 1.6. Since $W^{2,\frac{n}{k}(n-k)}$ embeds into $C^{1,1-\frac{2k}{n-k}}$, this result is different from that in [Dinew and Dinew 2016]. It is sharp in view of Pogorelov-type examples.

Remark 1.7. It is not known whether all singularities of solutions to $\det(\partial\bar{\partial}u) = 1$ arise as analytic subvarieties, or that they occur on a complex analogue of the agreement set with a tangent plane. Thus, Theorem 1.2 does not immediately imply smoothness of solutions to $\det(\partial\bar{\partial}u) = 1$ when $u \in W^{2,n(n-1)}$ (unlike in the real case).

The critical Sobolev spaces arise naturally in geometric applications. For example, in complex dimension 2 the L^2 norm of the Laplacian is a scale invariant, monotone quantity whose concentration controls, at least qualitatively, the regularity of functions with Monge–Ampère mass bounded below. In this sense, Theorem 1.2 can be seen as a step toward understanding the regularity and compactness properties of sequences of (quasi)-PSH functions with lower bounds for the Monge–Ampère mass, which arise frequently in Kähler geometry.

The proof of Theorem 1.1 relies on two key observations. The first is that u grows at least like $\mathrm{dist}^{2-\frac{2k}{n}}$ away from a zero set of dimension k. The second is that the $W^{2,\frac{n}{2k}(n-k)}$ norm is invariant under the rescalings that fix the k-dimensional zero set, and preserve functions with this growth. By combining these observations with some convex analysis, we show that the mass of $(\Delta u)^{\frac{n}{2k}(n-k)}$ is at least some fixed positive constant in each dyadic annulus around the zero set.

In the complex case the strategy is similar, but an important difficulty is that we don't have convexity. We overcome this in two ways. First, using subharmonicity along complex lines, we can say that u grows at a certain rate from its zero set at many points. Second, we use a dichotomy argument: either the mass of $|D^2u|^{\frac{n(n-k)}{k}}$ is at least a small constant in an annulus around the zero set, or it is very large and concentrates close to the zero set. Using that the $W^{2,\frac{n(n-k)}{k}}$ norm is bounded, we can rule out the second case and proceed as before.

The paper is organized as follows. In Section 2 we prove some estimates from convex analysis that are useful in the real case. We then prove an analogue in the general setting that is useful in the complex case. In Section 3 we prove Theorem 1.1. In Section 4 we prove Theorem 1.2. Finally, in Section 5 we give some applications of Theorem 1.1 to interior estimates for the real Monge–Ampère equation.

2. Preliminaries

Here we prove some useful functional inequalities. The first inequality is from convex analysis. This will be used to prove Theorem 1.1. We then prove a certain analogue in the general setting. This will be used to prove Theorem 1.2.

Estimate from convex analysis.

Lemma 2.1. *Let $n \geq 2$ and let w be a nonnegative convex function on $B_2 \subset \mathbb{R}^n$, with $w(0) = 0$ and $\sup_{\partial B_1} w \geq 1$. Then there is some positive constant $c(n)$ such that*

$$\int_{B_2 \backslash B_1} \Delta w \, dx > c(n).$$

Proof. By integration by parts, we have

$$\int_{B_2 \backslash B_1} \Delta w \, dx = \int_{\partial B_2} \partial_r w \, ds - \int_{\partial B_1} \partial_r w \, ds,$$

where ∂_r denotes radial derivative. By convexity, $\partial_r w$ is increasing on radial lines. We conclude that

$$\int_{B_2 \backslash B_1} \Delta w \, dx \geq \frac{1}{2} \int_{\partial B_2} \partial_r w \, ds.$$

Assume that the maximum of w on ∂B_1 is achieved at e_n. By convexity, $w \geq 1$ in $B_2 \cap \{x_n \geq 1\}$; hence $\partial_r w > \frac{1}{2}$ on $\partial B_2 \cap \{x_n \geq 1\}$. Since $\partial_r w \geq 0$, the conclusion follows. □

As a consequence, the Sobolev regularity of a convex function whose maximum on ∂B_r grows like r^q is no better than that of r^q:

Lemma 2.2. *Assume that w is a nonnegative convex function on $B_1 \subset \mathbb{R}^n$ $(n \geq 2)$ such that $w(0) = 0$ and $\sup_{\partial B_r} w \geq r^q$ for some $q \in [1, 2)$ and all $r < 1$. Then*

$$\int_{B_1 \backslash B_r} (\Delta w)^{\frac{n}{2-q}} \, dx \geq c(n,q)\,|\log r|$$

for some $c(n,q) > 0$ and all $r \in \left(0, \frac{1}{2}\right)$.

Remark 2.3. We take $q \geq 1$ since convex functions are locally Lipschitz.

Proof. Fix $\rho < \frac{1}{2}$ and let $w_\rho(x) = \rho^{-q} w(\rho x)$. Note that the $L^{\frac{n}{2-q}}$ norm of Δw is invariant under such rescalings. We conclude from this observation and Lemma 2.1 that

$$\int_{B_{2\rho} \backslash B_\rho} (\Delta w)^{\frac{n}{2-q}} \, dx = \int_{B_2 \backslash B_1} (\Delta w_\rho)^{\frac{n}{2-q}} \, dx \geq c(n,q).$$

The estimate follows by summing this inequality over dyadic annuli. □

Remark 2.4. One can refine this estimate to Orlicz norms. Let $F : [0,\infty) \to [0,\infty)$ be a convex function with $F(0) = 0$. By Lemma 2.1 we have

$$\frac{1}{|B_{2r}\backslash B_r|}\int_{B_{2r}\backslash B_r} \Delta u \, dx > c(n)r^{q-2}.$$

Using Jensen's inequality and summing over dyadic annuli, we obtain

$$\int_{B_1} F(\Delta u/\lambda)\, dx \geq \sum_{k=1}^{\infty} |B_{2^{-k}}\backslash B_{2^{-k-1}}|\, F\big(c(n)2^{-k(q-2)}/\lambda\big).$$

In particular, the Orlicz norm $\|\Delta u\|_{L^F(B_1)}$ is equal to ∞ if

$$\int_1^{\infty} t^{-\frac{n}{2-q}} F(t)\, \frac{dt}{t} = \infty.$$

Examples $F(t)$ that agree with $t^{\frac{n}{2-q}}|\log t|^{-p}$ for t large and $0 \leq p \leq 1$ satisfy this condition, and give weaker norms than $L^{\frac{n}{2-q}}$. For a reference on Orlicz spaces, see, e.g., [Simon 2011].

Estimate without convex analysis. The following estimate is a certain analogue of Lemma 2.1, in the general setting.

Lemma 2.5. *Let w be a nonnegative function on $B_2 \subset \mathbb{R}^n$ with $w(0) = 0$, and let $p > \frac{n}{2}$. Then there exists $c_0 > 0$ depending on n, p such that for all $\epsilon \in (0,1)$, there exists some $\delta(\epsilon, n, p)$ such that either*

$$\left(\int_{B_2\backslash B_\epsilon} |D^2 w|^p \, dx\right)^{\frac{1}{p}} \geq \delta \sup_{\partial B_1} w,$$

or

$$\left(\int_{B_{2\epsilon}} |D^2 w|^p \, dx\right)^{\frac{1}{p}} \geq c_0 \epsilon^{\frac{n}{p}-2} \sup_{\partial B_1} w.$$

Proof. After multiplying by a constant we may assume that $\sup_{\partial B_1} w = 1$. Assume that the first case is not satisfied. Then by the Sobolev–Poincaré and Morrey inequalities we have

$$\|w - l\|_{L^\infty(B_2\backslash B_\epsilon)} < C(n, p, \epsilon)\,\delta$$

for some linear function l. Take δ so small that the right side is less than $\frac{1}{8}$.

By the hypotheses on w, we have $l(0) > \frac{1}{2}$. Indeed, after a rotation we have $l(e_n) > \frac{7}{8}$ and $l(-2e_n) \geq -\frac{1}{8}$.

Let $\tilde{w}(x) = (w - l)(\epsilon x)$. Then $|\tilde{w}| < \frac{1}{8}$ in $B_2\backslash B_1$, and furthermore, $\tilde{w}(0) < -\frac{1}{2}$. It follows again from standard embeddings that

$$\left(\int_{B_2} |D^2 \tilde{w}|^p \, dx\right)^{\frac{1}{p}} > c_0(n, p).$$

Scaling back, we obtain the desired inequality. □

3. Proof of Theorem 1.1

We recall some estimates on the geometry of solutions to $\det D^2u \geq 1$. The first says that the volume of sublevel sets grows at most as fast as for the paraboloids with Hessian determinant 1:

Lemma 3.1. *Assume* $\det D^2u \geq 1$ *in a convex subset of* $\mathbb{R}^n$ *containing* 0, *with* $u \geq 0$ *and* $u(0) = 0$. *Then*

$$|\{u < h\}| < C(n)h^{\frac{n}{2}}$$

for all $h > 0$.

The proof follows from the affine invariance of the Monge–Ampère equation and a quadratic barrier; see, e.g., [Mooney 2015], Lemma 2.2.

Using Lemma 3.1 we can quantify how quickly u grows from a singularity. Below we fix $n \geq 3$ and $0 < k < \frac{n}{2}$, and we write $(x, y) \in \mathbb{R}^n$ with $x \in \mathbb{R}^{n-k}$ and $y \in \mathbb{R}^k$.

Lemma 3.2. *Assume that* $\det D^2u \geq 1$ *in* $\{|x| < 1\} \cap \{|y| < 1\} \subset \mathbb{R}^n$, *with* $u \geq 0$ *and* $u = 0$ *on* $\{x = 0\}$. *Then for all* $r < 1$ *we have*

$$\inf_{y} \sup_{|x|=r} u(x, y) > c(n)r^{2-\frac{2k}{n}}.$$

Proof. Take $c = c(n)$ small and assume by way of contradiction that for some r_0 the conclusion is false. Set $h = cr_0^{2-\frac{2k}{n}}$. Then for some y_0 we have

$$\{y = y_0\} \cap \left\{|x| < (c^{-1}h)^{\frac{n}{2}\frac{1}{n-k}}\right\} \subset \{u < h\}.$$

Since $\{u < h\}$ is convex, it contains the convex hull of the set on the left and $\pm e_n$. We conclude that

$$|\{u < h\}| \geq \tilde{c}(n)(c^{-1}h)^{\frac{n}{2}},$$

which contradicts Lemma 3.1 for c small. $\square$

The main theorem follows from the growth established in Lemma 3.2 and the convex analysis estimate Lemma 2.2.

Proof of Theorem 1.1. Assume that u agrees with a tangent plane on a set of dimension k. After subtracting the tangent plane, translating and rescaling, we may assume that $u \geq 0$ on $\{|x| < 1\} \cap \{|y| < 1\}$, and that $u = 0$ on $\{x = 0\}$. By Lemma 3.2, we also have that

$$\inf_{\{|y|<1\}} \sup_{|x|=r} u(x, y) \geq c(n)r^{2-\frac{2k}{n}}.$$

Apply Lemma 2.2 on the slices $\{y = \text{const.}\}$, taking $q = 2 - \frac{2k}{n}$ and replacing n by $n - k$, and integrate in y to conclude that

$$\int_{\{|y|<1\} \cap \{r<|x|<1\}} (\Delta u)^{\frac{n}{2k}(n-k)}\, dx \geq c(n, k)|\log r|.$$

Taking $r \to 0$ completes the proof. $\square$

Remark 3.3. By Remark 2.4, one obtains the same result if Δu is in the (weaker) Orlicz space L^F for any convex $F : [0, \infty) \to [0, \infty)$ satisfying $F(0) = 0$ and

$$\int_1^\infty t^{-\frac{n(n-k)}{2k}} F(t) \frac{dt}{t} = \infty. \tag{1}$$

4. Proof of Theorem 1.2

We first prove an analogue of Lemma 3.2. We fix $n \geq 2$ and $0 < k < n$, and we use coordinates $(z, w) \in \mathbb{C}^n$ with $z \in \mathbb{C}^{n-k}$ and $w \in \mathbb{C}^k$.

Lemma 4.1. *Assume that* $\det(\partial\bar{\partial}u) \geq 1$ *in* $\{|z| < 1\} \cap \{|w| < 1\} \subset \mathbb{C}^n$ *in the viscosity sense, with* $u \geq 0$ *and* $u = 0$ *on* $\{z = 0\}$. *Then for all* $r < 1$ *we have*

$$\sup_{|w|<\frac{1}{4}} \sup_{|z|=r} u(z, w) \geq c(n) r^{2-\frac{2k}{n}}.$$

Proof. Take $c = c(n)$ small and assume by way of contradiction that for some r_0 the conclusion is false. Let $h = c r_0^{2-\frac{2k}{n}}$. Since u is subharmonic on the slices $\{w = \text{const.}\}$, by the maximum principle we have

$$\left\{|w| < \tfrac{1}{4}\right\} \cap \left\{|z| < (c^{-1}h)^{\frac{n}{2(n-k)}}\right\} \subset \{u < h\}.$$

(Note that the volume of the set on the left is much larger than h^n for c small.) The proof then proceeds as in the real case. For c small, the convex quadratics $Q_t = 2h(16|w|^2 + (c^{-1}h)^{-\frac{n}{n-k}}|z|^2) + t$ are supersolutions that lie strictly above u on $\partial\left(\{|w| < \frac{1}{4}\} \cap \{|z| < r_0\}\right)$ for $t \geq 0$. For some $t \geq 0$, Q_t touches u from above somewhere inside this set, contradicting that u is a viscosity subsolution. $\square$

Proof of Theorem 1.2. Assume that the minimum set of u contains an analytic subvariety of dimension k. After a biholomorphic transformation and a rescaling, we may assume that $u \geq 0$ on $\{|z| < 1\} \cap \{|w| < 1\}$ and $u = 0$ on $\{z = 0\}$ (see, e.g., [Dinew and Dinew 2016, Theorem 32] for details) and that

$$\|u\|_{W^{2,\frac{n(n-k)}{k}}(\{|w|<1\}\cap\{|z|<1\})} = C_0 < \infty.$$

(Here we used elliptic theory: Δu controls $D^2 u$ in L^p for $1 < p < \infty$.)

For any $r < \frac{1}{2}$ we define

$$u_r(z, w) = \frac{1}{r^{2-\frac{2k}{n}}} u(rz, w).$$

We claim that there exist $\epsilon, \delta > 0$ small depending on n, k, C_0 (but not r) such that

$$\int_{\{|w|<1\}\cap\{\epsilon<|z|<2\}} |D_z^2 u_r|^{\frac{n(n-k)}{k}} |dz|\,|dw| > \delta. \tag{2}$$

Here D_z^2 denotes the Hessian in the z-variable. We first indicate how to complete the proof given the claim. The invariance of this norm under the rescalings used to obtain u_r gives that

$$\int_{\{|w|<1\}\cap\{(\frac{\epsilon}{2})r<|z|<r\}} |D^2 u|^{\frac{n(n-k)}{k}} |dz|\,|dw| > \delta$$

for all $r < 1$. By summing this over the annuli $\{|w| < 1\} \cap \{(\frac{\epsilon}{2})^{k+1} < |z| < (\frac{\epsilon}{2})^k\}$ we eventually contradict the upper bound on the $W^{2,\frac{n(n-k)}{k}}$ norm of u.

We now prove the claim. By Lemma 4.1, there exists some $(z_0, w_0) \in \{|z| = 1\} \cap \{|w| < \frac{1}{4}\}$ with $u_r(z_0, w_0) \geq c(n) > 0$. Let

$$M(w) = u_r(z_0, w).$$

Since $M(w)$ is positive and subharmonic, we have by the mean value inequality that

$$\int_{\{|w|<1\}} M(w)\,|dw| > c(n) > 0. \tag{3}$$

By Lemma 2.5, for all ϵ small, there exists $\delta(n, k, \epsilon)$ such that either

$$\left(\int_{\{\epsilon<|z|<2\}} |D_z^2 u_r(z, w)|^{\frac{n(n-k)}{k}}\,|dz|\right)^{\frac{k}{n(n-k)}} \geq \delta M(w)$$

or

$$\left(\int_{\{|z|<2\}} |D_z^2 u_r(z, w)|^{\frac{n(n-k)}{k}}\,|dz|\right)^{\frac{k}{n(n-k)}} \geq c(n,k)\epsilon^{-\frac{2(n-k)}{n}} M(w).$$

Let A_ϵ be the set of w such that the first case holds. We conclude from the scale-invariance of the norm we consider that

$$C_0 \geq c(n,k)\int_{A_\epsilon^c}\left(\int_{\{|z|<2\}} |D_z^2 u_r|^{\frac{n(n-k)}{k}}\,|dz|\right)^{\frac{k}{n(n-k)}} |dw| \geq c(n,k)\epsilon^{-\frac{2(n-k)}{n}} \int_{A_\epsilon^c} M(w)\,|dw|.$$

By taking $\epsilon(n, k, C_0)$ small, we conclude that the mass of $M(w)$ in A_ϵ^c is less than its mass in A_ϵ. We conclude from the estimate (3) that

$$\left(\int_{\{|w|<1\}\cap\{\epsilon<|z|<2\}} |D_z^2 u_r|^{\frac{n(n-k)}{k}}\,|dz|\,|dw|\right)^{\frac{k}{n(n-k)}} \geq c(n,k)\,\delta\int_{A_\epsilon} M(w)\,|dw| \geq c(n,k)\,\delta(n,k,C_0). \quad \square$$

Remark 4.2. To justify the above computations, on the slices $\{w = \text{const.}\}$ we use that, for almost every w in $\{|w| < 1\}$, the restrictions $f_w(z) := u(z, w)$ are in $W^{2,\frac{n(n-k)}{k}}(\{|z| < 1\})$ with $D_z^2 f_w = D_z^2 u(\cdot, w)$. To see this, let $\{u^j\}$ be a sequence of smooth functions approximating u in $W^{2,\frac{n(n-k)}{k}}(B_1)$, and apply the Fubini theorem to $|u^j - u| + |\nabla u^j - \nabla u| + |D^2 u^j - D^2 u|$.

Remark 4.3. Theorem 1.2 actually implies a slightly more general result. Namely, if u is plurisubharmonic on B_1 and satisfies $\det \partial\bar{\partial} u \geq 1$, and $\Delta u \in L^p$ for some $p \geq \frac{n}{k}(n - k)$, then u cannot be pluriharmonic when restricted to any analytic set of dimension greater than or equal to k. This follows from Theorem 1.2 and the proof of Theorem 35 in [Dinew and Dinew 2016].

5. Applications

As a consequence of Theorem 1.1 we obtain interior estimates for the real Monge–Ampère equation depending on the $W^{2,p}$ norm of the solution for any $p \geq \frac{n(n-1)}{2}$. This extends a result of Urbas [1988] to the equality case $p = \frac{n(n-1)}{2}$.

Remark 5.1. In fact, we obtain interior estimates depending on certain Orlicz norms that are slightly weaker than $L^{\frac{n(n-1)}{2}}$.

We recall the definition of sections of a convex function. Let u be a convex function on $B_1 \subset \mathbb{R}^n$. If l is a supporting linear function to u at $x \in B_1$, we set

$$S_h^l(x) = \{u < l + h\} \cap B_1.$$

Lemma 5.2. *Assume that* $\det D^2u \geq 1$ *in* $B_1 \subset \mathbb{R}^n$, *and that* $\|u\|_{W^{2,p}(B_1)} < C_0$ *for some* $p \geq \frac{n(n-1)}{2}$. *Then there exists* $h_0 > 0$ *depending only on* n, p *and* C_0 *such that*

$$S_{h_0}^l(x) \Subset B_1$$

for all $x \in B_{\frac{1}{2}}$ *and supporting linear functions* l *at* x.

Proof. The result follows from a standard compactness argument using the closedness of the condition $\det D^2u \geq 1$ under uniform convergence, the lower semicontinuity of the $W^{2,p}$ norm under weak convergence and Theorem 1.1. □

Remark 5.3. The conclusion is the same if $\|\Delta u\|_{L^F(B_1)} < C_0$ for some F satisfying condition (1) for $k = 1$, and in addition, e.g., $\|u\|_{W^{2,2}(B_1)} < C_0$. The argument is by compactness again, but one has to work harder to extract a limit whose Hessian has bounded Orlicz norm. Rather than using weak $W^{2,2}$ convergence of a subsequence $\{u_k\}$, invoke the Banach–Saks theorem and use the strong convergence in $W^{2,2}$ of Cesàro means $\frac{1}{N}\sum_{k=1}^{N} u_k$. The convexity of F then implies that the Hessian of the limit has bounded Orlicz norm.

(In order to use Banach–Saks we need control of Δu in L^p for some $p > 1$, which does not follow from bounded Orlicz norm. This is the reason for the second condition).

Interior, e.g., $C^{2,\alpha}$ estimates and $W^{2,q}$ estimates in terms of $\|u\|_{W^{2,n(n-1)/2}(B_1)}$ follow, where the estimates also depend on n, α and the C^α norm (resp. n, q and the modulus of continuity) of $\det D^2u$.

Indeed, by Lemma 5.2 we have $S_{h_0}^l(x) \Subset B_1$ for some universal h_0 and all $x \in B_{\frac{1}{2}}$. Since $\det D^2u$ also has an upper bound (depending on the C^α norm or modulus of continuity of $\det D^2u$), we have the lower volume bound $|S_{h_0}^l(x)| > ch_0^{\frac{n}{2}}$ for compactly contained sections [Caffarelli 1990b]. Combining this with the diameter estimate $\operatorname{diam}(S_{h_0}^l(x)) < 2$, we see that the eigenvalues of the affine transformations normalizing these sections (taking B_1 to their John ellipsoids) are bounded above and below by positive universal constants. The estimates follow by applying Caffarelli's results [1990a] in the normalized sections, scaling back, and doing a covering argument.

Acknowledgements

Mooney would like to thank A. Figalli for comments and X.-J. Wang for encouragement. Collins would like to thank the math department at Chalmers University, where the majority of this work was completed, for a perfect working environment.

References

[Alexandroff 1942] A. Alexandroff, "Smoothness of the convex surface of bounded Gaussian curvature", *C. R. (Doklady) Acad. Sci. URSS (N.S.)* **36** (1942), 195–199. In Russian. MR Zbl

[Błocki 1999] Z. Błocki, "On the regularity of the complex Monge–Ampère operator", pp. 181–189 in *Complex geometric analysis in Pohang* (Pohang, 1997), Contemp. Math. **222**, Amer. Math. Soc., Providence, RI, 1999. MR Zbl

[Błocki and Dinew 2011] Z. Błocki and S. Dinew, "A local regularity of the complex Monge–Ampère equation", *Math. Ann.* **351**:2 (2011), 411–416. MR Zbl

[Caffarelli 1990a] L. A. Caffarelli, "Interior $W^{2,p}$ estimates for solutions of the Monge–Ampère equation", *Ann. of Math.* (2) **131**:1 (1990), 135–150. MR Zbl

[Caffarelli 1990b] L. A. Caffarelli, "A localization property of viscosity solutions to the Monge–Ampère equation and their strict convexity", *Ann. of Math.* (2) **131**:1 (1990), 129–134. MR Zbl

[Caffarelli 1993] L. A. Caffarelli, "A note on the degeneracy of convex solutions to Monge Ampère equation", *Comm. Partial Differential Equations* **18**:7-8 (1993), 1213–1217. MR Zbl

[Caffarelli and Cabré 1995] L. A. Caffarelli and X. Cabré, *Fully nonlinear elliptic equations*, American Mathematical Society Colloquium Publications **43**, Amer. Math. Soc., Providence, RI, 1995. MR Zbl

[Caffarelli et al. 1984] L. Caffarelli, L. Nirenberg, and J. Spruck, "The Dirichlet problem for nonlinear second-order elliptic equations, I: Monge–Ampère equation", *Comm. Pure Appl. Math.* **37**:3 (1984), 369–402. MR Zbl

[Calabi 1958] E. Calabi, "Improper affine hyperspheres of convex type and a generalization of a theorem by K. Jörgens", *Michigan Math. J.* **5** (1958), 105–126. MR Zbl

[Cheng and Yau 1976] S. Y. Cheng and S. T. Yau, "On the regularity of the solution of the n-dimensional Minkowski problem", *Comm. Pure Appl. Math.* **29**:5 (1976), 495–516. MR Zbl

[Cheng and Yau 1977] S. Y. Cheng and S. T. Yau, "On the regularity of the Monge–Ampère equation $\det(\partial^2 u/\partial x_i \partial s x_j) = F(x, u)$", *Comm. Pure Appl. Math.* **30**:1 (1977), 41–68. MR Zbl

[Dinew and Dinew 2016] S. Dinew and Ż. Dinew, "The minimum sets and free boundaries of strictly plurisubharmonic functions", *Calc. Var. Partial Differential Equations* **55**:6 (2016), art. id. 148. MR Zbl

[Harvey and Wells 1973] F. R. Harvey and R. O. Wells, Jr., "Zero sets of non-negative strictly plurisubharmonic functions", *Math. Ann.* **201** (1973), 165–170. MR Zbl

[Kołodziej 1996] S. Kołodziej, "Some sufficient conditions for solvability of the Dirichlet problem for the complex Monge–Ampère operator", *Ann. Polon. Math.* **65**:1 (1996), 11–21. MR Zbl

[Krylov 1983] N. V. Krylov, "Boundedly inhomogeneous elliptic and parabolic equations in a domain", *Izv. Akad. Nauk SSSR Ser. Mat.* **47**:1 (1983), 75–108. In Russian; translated in *Math. USSR-Izv.* **22**:1 (1984), 67–97. MR Zbl

[Lions 1983] P.-L. Lions, "Sur les équations de Monge–Ampère, I", *Manuscripta Math.* **41**:1-3 (1983), 1–43. MR Zbl

[Lions 1985] P.-L. Lions, "Sur les équations de Monge–Ampère", *Arch. Rational Mech. Anal.* **89**:2 (1985), 93–122. MR Zbl

[Mooney 2015] C. Mooney, "Partial regularity for singular solutions to the Monge–Ampère equation", *Comm. Pure Appl. Math.* **68**:6 (2015), 1066–1084. MR Zbl

[Pogorelov 1971a] A. V. Pogorelov, "The Dirichlet problem for the multidimensional analogue of the Monge–Ampère equation", *Dokl. Akad. Nauk SSSR* **201** (1971), 790–793. In Russian; translated in *Sov. Math., Dokl.* **12** (1971), 1727–1731. MR Zbl

[Pogorelov 1971b] A. V. Pogorelov, "The regularity of the generalized solutions of the equation $\det(\partial^2 u/\partial x^i\, \partial x^j) = \varphi(x^1, x^2, \dots, x^n) > 0$", *Dokl. Akad. Nauk SSSR* **200** (1971), 534–537. In Russian; translated in *Sov. Math., Dokl.* **12** (1971), 1436–1440. MR Zbl

[Pogorelov 1973] A. V. Pogorelov, *Extrinsic geometry of convex surfaces*, Translations of Mathematical Monographs **35**, Amer. Math. Soc., Providence, RI, 1973. MR Zbl

[Pogorelov 1978] A. V. Pogorelov, *The Minkowski multidimensional problem*, Scripta Series in Mathematics **5**, V. H. Winston & Sons, Washington, D.C., 1978. MR Zbl

[Simon 2011] B. Simon, *Convexity: an analytic viewpoint*, Cambridge Tracts in Mathematics **187**, Cambridge University Press, 2011. MR Zbl

[Trudinger and Wang 2008] N. S. Trudinger and X.-J. Wang, "Boundary regularity for the Monge–Ampère and affine maximal surface equations", *Ann. of Math.* (2) **167**:3 (2008), 993–1028. MR Zbl

[Urbas 1988] J. I. E. Urbas, "Regularity of generalized solutions of Monge–Ampère equations", *Math. Z.* **197**:3 (1988), 365–393. MR Zbl

[Wang 1996] X.-J. Wang, "Regularity for Monge–Ampère equation near the boundary", *Analysis* **16**:1 (1996), 101–107. MR Zbl

[Wang 2012] Y. Wang, "On the $C^{2,\alpha}$-regularity of the complex Monge–Ampère equation", *Math. Res. Lett.* **19**:4 (2012), 939–946. MR Zbl

Received 23 Mar 2017. Revised 18 Jun 2017. Accepted 17 Jul 2017.

TRISTAN C. COLLINS: tcollins@math.harvard.edu
Department of Mathematics, Harvard University, Cambridge, MA, United States

CONNOR MOONEY: connor.mooney@math.ethz.ch
Department of Mathematics, ETH Zurich, Zurich, Switzerland

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at msp.org/apde.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in APDE are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

ANALYSIS & PDE

Volume 10 No. 8 2017