ANALYSIS & PDE

Volume 13 No. 7 2020

msp

Analysis & PDE

msp.org/apde

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

The subscription price for 2020 is US $340/year for the electronic version, and $550/year (+$60, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscriber address should be sent to MSP.

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

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ANALYSIS AND PDE
Vol. 13, No. 7, 2020
https://doi.org/10.2140/apde.2020.13.1955

msp

REFINED MASS-CRITICAL STRICHARTZ ESTIMATES FOR SCHRÖDINGER OPERATORS

CASEY JAO

We develop refined Strichartz estimates at L^2 regularity for a class of time-dependent Schrödinger operators. Such refinements quantify near-optimizers of the Strichartz estimate and play a pivotal part in the global theory of mass-critical NLS. On one hand, the harmonic analysis is quite subtle in the L^2-critical setting due to an enormous group of symmetries, while on the other hand, the space-time Fourier analysis employed by the existing approaches to the constant-coefficient equation are not adapted to nontranslation-invariant situations, especially with potentials as large as those considered in this article.

Using phase-space techniques, we reduce to proving certain analogues of (adjoint) bilinear Fourier restriction estimates. Then we extend Tao's bilinear restriction estimate for paraboloids to more general Schrödinger operators. As a particular application, the resulting inverse Strichartz theorem and profile decompositions constitute a key harmonic analysis input for studying large-data solutions to the L^2-critical NLS with a harmonic oscillator potential in dimensions ≥ 2. This article builds on recent work of Killip, Visan, and the author in one space dimension.

1. Introduction

We prove sharpened forms of the Strichartz inequality for nontranslation-invariant linear Schrödinger equations with L^2 initial data. Recall that solutions to the linear constant-coefficient Schrödinger equation

$$i\partial_t u = -\tfrac{1}{2}\Delta u, \quad u(0,\cdot) = u_0 \in L^2(\mathbb{R}^d), \tag{1}$$

satisfy the Strichartz inequality [1977]

$$\|u\|_{L^{2(d+2)/d}_{t,x}(\mathbb{R}\times\mathbb{R}^d)} \leq C\|u(0,\cdot)\|_{L^2(\mathbb{R}^d)}. \tag{2}$$

On the other hand, it is also known if u a solution that comes close to saturating this inequality, then it must exhibit some "concentration"; see [Carles and Keraani 2007; Merle and Vega 1998; Moyua et al. 1999; Bégout and Vargas 2007]. Such inverse theorems may be equivalently formulated as a refined estimate

$$\|u\|_{L^{2(d+2)/d}_{t,x}} \lesssim \|u\|_X^{\theta}\,\|u(0,\cdot)\|_{L^2(\mathbb{R}^d)}^{1-\theta}, \tag{3}$$

where the norm X is weaker than the right side of (2) but measures the "microlocal concentration" of the solution. We pursue analogues of such refinements when the right side of (1) is replaced by a more general Schrödinger operator $-\frac{1}{2}\Delta + V(t,x)$.

MSC2010: primary 35Q41; secondary 42B37.
Keywords: inverse Strichartz estimates, bilinear restriction, Schrödinger operators.

Inverse theorems for the Strichartz inequality have provided a key input to the study of the L^2-critical NLS

$$i\partial_t u = -\tfrac{1}{2}\Delta u \pm |u|^{\frac{4}{d}}u, \quad u(0,\cdot) \in L^2(\mathbb{R}^d), \tag{4}$$

so termed because the rescaling $u \mapsto u_\lambda(t,x) := \lambda^{d/2}u(\lambda^2 t, \lambda x)$ preserves both (1) and the L^2-norm $M[u] := \|u(t)\|_{L^2(\mathbb{R}^d)} = \|u(0)\|_{L^2(\mathbb{R}^d)}$. Indeed, they are used to construct the profile decompositions underpinning the Bourgain–Kenig–Merle concentration compactness and rigidity method by identifying potential blowup scenarios for nonlinear solutions with large data. Using this method, the large-data global regularity problem for (4) was recently settled by Dodson [2012; 2015; 2016a; 2016b], building on earlier work of Killip, Visan, Tao, and Zhang [Killip et al. 2008; 2009; Tao et al. 2007]. For further discussion of this equation we refer the interested reader to the lecture notes [Killip and Visan 2013].

The large group of symmetries for the inequality (2) is a significant obstruction to characterizing its near-optimizers. Besides translation and scaling symmetry, both sides are also invariant under *Galilei* transformations

$$u \mapsto u_{\xi_0}(t,x) := e^{i[\langle x,\xi_0\rangle - \frac{1}{2}t|\xi_0|^2]}u(t, x - t\xi_0), \quad \xi_0 \in \mathbb{R}^d.$$

This last symmetry emerges only at L^2 regularity and creates an additional layer of complexity. In particular, while the Littlewood–Paley decomposition is extremely well-adapted to higher Sobolev regularity variants of (2), such as the $\dot{H}^1$-critical estimate

$$\|u\|_{L^{2(d+2)/(d-2)}_{t,x}} \lesssim \|\nabla u(0)\|_{L^2(\mathbb{R}^d)},$$

it is useless for inverting the L^2-critical estimate because one has no a priori knowledge of where the solution is concentrated in frequency. Instead, the mass-critical refinements cited above combine space-time Fourier-analytic arguments with restriction theory for the paraboloid.

In physical applications, one is naturally led to consider variants of the mass-critical equation (4) with external potentials, such as the harmonic oscillator

$$i\partial_t u = \Big(-\tfrac{1}{2}\Delta + \sum_j \omega_j^2 x_j^2\Big)u \pm |u|^{\frac{4}{d}}u, \quad u(0,\cdot) \in L^2(\mathbb{R}^d). \tag{5}$$

For instance, the cubic equation (with a $|u|^2 u$ nonlinearity) has been proposed as a model for Bose–Einstein condensates in a laboratory trap [Zhang 2000], where $\|u(t)\|_{L^2}$ represents the total number of particles, and in two space dimensions the critical Sobolev norm for this equation is precisely L^2.

While introducing the potential breaks scaling symmetry, one nonetheless expects solutions with highly concentrated initial data to be approximated, for short times, by solutions to the scale-invariant equation (4). Less obviously, the equation is invariant under "generalized" Galilei boosts, detailed in Lemma 1.1 below, where the spatial and frequency parameters act together on the solutions; in the constant-coefficient setting, this reduces to the usual independent space translation and Galilei boost symmetries.

This article develops refined Strichartz estimates for the linear equation

$$i\partial_t u = \big(-\tfrac{1}{2}\Delta + V\big)u, \quad u(0,\cdot) \in L^2(\mathbb{R}^d),$$

for a class $\mathcal{V}$ of real-valued potentials $V(t,x)$ that merely satisfy similar bounds as the harmonic oscillator and possibly also depend on time. Specifically, define

$$\mathcal{V} := \{V : \mathbb{R}\times\mathbb{R}^d \to \mathbb{R} : \|\partial_x^\alpha V\|_{L^\infty_{t,x}} \le M_{|\alpha|} \text{ for } 2 \le |\alpha| \le N = N(d)\} \tag{6}$$

for fixed constants $0 < M_1, M_2, \dots, M_N$. These estimates play a key role in the large-data theory for nontranslation-invariant L^2-critical Cauchy problems typified by (5). We briefly discuss the nonlinear problem in the last section of the introduction.

The case of one space dimension was treated in a previous joint work with Killip and Visan [Jao et al. 2019]. This paper extends the methods introduced there to higher dimensions.

1A. ***The setup.*** To clarify the structure of our arguments we begin with a slightly more general setup. Hence we consider time-dependent, real-valued symbols $a(t,x,\xi)$ which are measurable in t and satisfy

$$|\partial_x^\alpha \partial_\xi^\beta a| \le c_{\alpha\beta} \quad \text{for all } |\alpha| + |\beta| \ge 2. \tag{7}$$

Further, we assume the characteristic curvature condition

$$||\det a_{\xi\xi}| - 1| + |\|a_{\xi\xi}\| - 1| \le \varepsilon \tag{8}$$

for some small $0 < \varepsilon < 1$. For concreteness, all matrix norms in this article denote the Hilbert–Schmidt norm, but the exact choice of norm is inessential.

These hypotheses encompass several interesting situations:

- Schrödinger Hamiltonians with time-dependent scalar potentials $a = \frac{1}{2}|\xi|^2 + V(t,x)$, where $V \in \mathcal{V}$.
- Electromagnetic-type symbols $a = \frac{1}{2}|\xi|^2 + b(x,\xi) + V(t,x)$, where the first-order symbol $b(x,\xi)$ is real and satisfies $|\partial_x^\alpha \partial_\xi^\beta b| \le c_{\alpha\beta}$ for all $|\alpha| + |\beta| \ge 1$, and $V \in \mathcal{V}$ is a scalar potential as before.
- The frequency-1 portion of the Laplacian on a curved background.

For a symbol as defined above, write $a^w(t,x,D)$ for its Weyl quantization. Let $U(t,s)$ denote its unitary propagator on $L^2(\mathbb{R}^d)$, so that $u := U(t,s)u_s$ is the solution to the equation

$$(D_t + a^w(t,x,D))u = 0, \quad u(s,\cdot) = u_s \in L^2(\mathbb{R}^d). \tag{9}$$

Evolution equations of this type were studied in [Koch and Tataru 2005]. While translations and modulations do not preserve (9), they do preserve the class of equations defined by our assumptions. For an element (x_0,ξ_0) of classical phase space, define the "phase-space translation" operator $\pi(x_0,\xi_0)$ by

$$\pi(z_0)f(x) = e^{i\langle x - x_0, \xi_0\rangle} f(x - x_0).$$

Then a direct computation, as in the proof of [Koch and Tataru 2005, Proposition 4.3], yields:

Lemma 1.1. *If $U(t,s)$ is the propagator for the symbol a and $\sigma \mapsto z^\sigma = (x^\sigma, \xi^\sigma)$ is a bicharacteristic of a, then*

$$U(t,s)\pi(z_0^s) f = e^{i(\phi(t,z_0)-\phi(s,z_0))}\pi(z_0^t)U^{z_0}(t,s),$$

where U^{z_0} is the propagator for the equation

$$[D_t + (a^{z_0})^w(t,x,D)]u = 0,$$
$$a^{z_0}(t,z) = a(t, z_0^t + z) - \langle x, a_x(t,z_0^t)\rangle - \langle \xi, a_\xi(t,z_0^t)\rangle - a(z_0^t),$$

and the phase is defined by

$$\phi(t,z_0) = \int_0^t \langle a_\xi(\tau, z_0^\tau), \xi_0^\tau\rangle - a(\tau, z_0^\tau)\, d\tau.$$

Observe that the transformed symbol a^{z_0} satisfies the same estimates assumed of a. As a special case, symbols of the form $a = \frac{1}{2}|\xi|^2 + \langle A(t,x), \xi\rangle + \omega_{jk}(t)x^j x^k$ are themselves preserved by the mapping $a \mapsto a^{z_0}$ if $A = A_j\, dx^j$ is a 1-form whose components are linear functions of the space variables with time-dependent coefficients. In two and three space dimensions, such A are potentials for uniform magnetic fields.

The preceding hypotheses imply that (9) satisfies a local-in-time dispersive estimate:

Lemma 1.2. *If the symbol a satisfies the conditions (7) and (8), there exists $T_0 > 0$ such that the propagator $U(t,s)$ for the evolution equation (9) satisfies the estimate*

$$\|U(t,s)\|_{L_x^1 \to L_x^\infty} \lesssim |t-s|^{-\frac{d}{2}} \quad \textit{for all } |t-s| \le T_0. \tag{10}$$

Hence, the solutions to (9) satisfy local-in-time Strichartz estimates

$$\|u\|_{L_t^q L_x^r(I\times\mathbb{R}^d)} \lesssim_{|I|} \|u_s\|_{L^2(\mathbb{R}^d)}$$

for any compact time interval I, and for all Strichartz exponents *(q,r) satisfying $2 \le q, r \le \infty$, $\frac{2}{q} + \frac{d}{r} = \frac{d}{2}$, and $(q,r,d) \neq (2,\infty,2)$.*

Proof sketch. The dispersive estimate is shown in [Koch and Tataru 2005, Proposition 4.7] using wavepacket parametrices. Standard arguments (see [Ginibre and Velo 1995; Keel and Tao 1998]) then yield the Strichartz estimates. □

It suffices to choose the time increment T_0 so that

$$T_0 \le 1, \quad T_0\|a_{x\xi}\| + T_0^2\|a_{xx}\| \le \eta, \tag{11}$$

where $\eta = \eta(d)$ is a small parameter depending only on the dimension.

Remark. The concrete cases of scalar potentials and magnetic potentials were studied much earlier by Fujiwara [1979] and Yajima [1991], respectively, who proved the dispersive bound using Fourier integral parametrices.

We seek refinements of the Strichartz inequality analogous to those for the constant-coefficient equation. The earlier arguments for the constant-coefficient equation relied crucially on subtle bilinear estimates from Fourier restriction theory. We isolate and reformulate the technical lynchpin in the present context.

Hypothesis 1. *There exist $T_0 > 0$ and $1 < p < \frac{d+2}{d}$ such that the following holds: if $f, g \in L^2(\mathbb{R}^d)$ have frequency supports in sets of diameter $\lesssim N$ which are separated by distance $\sim N$, then*

$$\|U_\lambda^s(t) f U_\lambda^s(t) g\|_{L^p_{t,x}([-T_0,T_0]\times\mathbb{R}^d)} \lesssim N^{-\delta} \|f\|_{L^2(\mathbb{R}^d)} \|g\|_{L^2(\mathbb{R}^d)} \tag{12}$$

for all $s \in [-1, 1]$ and all $0 < \lambda \le 1$, where $U_\lambda^s(t) = U_\lambda^s(t, 0)$ are the propagators for the time-translated and rescaled symbols $a_\lambda^s := \lambda^2 a(s + \lambda^2 t, \lambda x, \lambda^{-1}\xi)$.

When $a = \frac{1}{2}|\xi|^2$, the scaling and translation parameters λ, s are extraneous, and inequalities of the form (12) are called (adjoint) bilinear Fourier restriction estimates. They were utilized in [Bégout and Vargas 2007] to obtain mass-critical Strichartz refinements in dimension 3 and higher (the results in dimensions 1 and 2, due to Carles and Keraani [2007], Merle and Vega [1998], and Moyua, Vargas and Vega [Moyua et al. 1999] utilized linear restriction estimates). For further discussion of such estimates, see for instance [Tao 2003].

In the first part of this paper, we connect (12) to Strichartz refinements. To measure concentration in the solution we test it against scaled, modulated, and translated wavepackets. Set

$$\psi(x) = c_d e^{-\frac{|x|^2}{2}}, \quad \psi_{x_0,\xi_0} = \pi(x_0, \xi_0)\psi, \quad c_d = 2^{-\frac{d}{2}}\pi^{-\frac{3d}{4}}, \tag{13}$$

where S_λ is the unitary rescaling $S_\lambda f(x) := \lambda^{-d/2} f(\lambda^{-1}x)$.

Theorem 1.3. *If Hypothesis 1 holds, then there exists $0 < \theta < 1$ such that for all initial data $u_0 \in L^2(\mathbb{R}^d)$ the solution u to (9) satisfies*

$$\|u\|_{L^{2(d+2)/d}([-1,1]\times\mathbb{R}^d)} \lesssim \Big(\sup_{0<\lambda\le 1,\, |t|\le 1,\, (x_0,\xi_0)\in T^*\mathbb{R}^d} |\langle S_\lambda \psi_{x_0,\xi_0}, u(t)\rangle_{L^2(\mathbb{R}^d)}| \Big)^\theta \|u_0\|_{L^2(\mathbb{R}^d)}^{1-\theta}. \tag{14}$$

The generality of our hypotheses requires us to formulate the estimates locally in time. Indeed, for most potentials the left side of the Strichartz estimate (14) is infinite if one integrates over $\mathbb{R}\times\mathbb{R}^d$; for instance, the harmonic oscillator potential $V = |x|^2$ admits periodic-in-time solutions. Nonetheless, our methods do yield (a new proof of) a global-in-time refined Strichartz estimate

$$\|u\|_{L^{2(d+2)/d}_{t,x}(\mathbb{R}\times\mathbb{R}^d)} \lesssim \Big(\sup_{\lambda>0,\ t\in\mathbb{R},\ (x_0,\xi_0)\in T^*\mathbb{R}^d} |\langle S_\lambda \psi_{x_0,\xi_0}, u(t)\rangle_{L^2(\mathbb{R}^d)}| \Big)^\theta \|u_0\|_{L^2(\mathbb{R}^d)}^{1-\theta}$$

for solutions to the constant-coefficient equation (1).

In applications to PDEs, such a refined estimate is nowadays interpreted in the framework of concentration compactness and yields profile decompositions via repeated application of the following:

Lemma 1.4. *Assume the estimate (14) holds. Let $u_n := U(t) f_n$ be a sequence of linear solutions with initial data $u_n(0) = f_n \in L^2(\mathbb{R}^d)$ such that $\|f_n\|_{L^2(\mathbb{R}^d)} \le A < \infty$ and $\|u_n\|_{L^{2(d+2)/d}_{t,x}} \ge \varepsilon > 0$. Then, after passing to a subsequence, there exist parameters*

$$\{(\lambda_n, t_n, x_n, \xi_n)\}_n \subset (0, 1] \times [-1, 1] \times \mathbb{R}^d_x \times \mathbb{R}^d_\xi$$

and a function $0 \neq \phi \in L^2(\mathbb{R}^d)$ *such that*

$$\pi(x_n, \xi_n)^{-1} S_{\lambda_n}^{-1} u_n \rightharpoonup \phi \quad \text{in } L^2,$$
$$\|\phi\|_{L^2} \gtrsim \varepsilon \left(\frac{\varepsilon}{A}\right)^{\frac{1-\theta}{\theta}}.$$

Further,

$$\|f_n\|_{L^2}^2 - \|f_n - U(t_n)^{-1} S_{\lambda_n} \pi(x_n, \xi_n) S_{\lambda_n} \phi\|_{L^2}^2 - \|U(t_n)^{-1} S_{\lambda_n} \pi(x_n, \xi_n) S_{\lambda_n} \phi\|_{L^2}^2 \to 0.$$

Proof. By the estimate (14), there exist $\lambda_n, t_n, x_n, \xi_n$ such that

$$|\langle S_{\lambda_n} \psi_{x_n,\xi_n}, U(t_n) f_n \rangle| = |\langle \psi, \pi(x_n, \xi_n)^{-1} S_{\lambda_n}^{-1} U(t_n) f_n \rangle| \gtrsim \varepsilon \left(\frac{\varepsilon}{A}\right)^{\frac{1-\theta}{\theta}}.$$

The sequence $\pi(x_n, \xi_n)^{-1} S_{\lambda_n}^{-1} U(t_n) f_n$ is bounded in L^2 and therefore converges weakly in L^2 to some ϕ after passing to a subsequence. The lower bound on $\|\phi\|_{L^2}$ is immediate, while

$$\begin{aligned}
\|f_n\|_{L^2}^2 - \|f_n - U(t_n)^{-1} S_{\lambda_n} \pi(x_n, \xi_n)\phi\|_{L^2}^2 - \|U(t_n)^{-1} S_{\lambda_n} \pi(x_n, \xi_n)\phi\|_{L^2}^2 \\
&= 2 \operatorname{Re}\langle f_n - U(t_n)^{-1} S_{\lambda_n} \pi(x_n, \xi_n)\phi, U(t_n)^{-1} S_{\lambda_n} \pi(x_n, \xi_n)\phi \rangle \\
&= 2 \operatorname{Re}\langle \pi(x_n, \xi_n)^{-1} S_{\lambda_n}^{-1} U(t_n) f_n - \phi, \phi \rangle \to 0. \qquad \square
\end{aligned}$$

Further discussion of profile decompositions and inverse Strichartz theorems may be found in the lecture notes [Killip and Visan 2013].

In the second part of this paper, we verify Hypothesis 1 for scalar potentials.

Theorem 1.5. *Consider a Schrödinger operator of the form* $H(t) = -\frac{1}{2}\Delta + V(t, x)$, *where* $V \in \mathcal{V}$. *Suppose* $S_1, S_2 \subset \mathbb{R}_\xi^d$ *are subsets of Fourier space with* $\operatorname{diam}(S_j) \le N$ *and* $c^{-1}N \ge \operatorname{dist}(S_1, S_2) \ge cN$ *for some* $0 < c < 1$. *There exists a constant* $\eta = \eta(c) > 0$ *such that if* $\tau_0 > 0$ *satisfies*

$$(\tau_0 + \tau_0^2)\|\partial_x^2 V\|_{L^\infty} < \eta,$$

then, for any $f, g \in L^2(\mathbb{R}^d)$ *with* $\operatorname{supp}(\hat{f}) \subset S_1$ *and* $\operatorname{supp}(\hat{g}) \subset S_2$, *the corresponding linear solutions* $u = U(t, 0) f$ *and* $v = U(t, 0)g$ *satisfy the estimate*

$$\|uv\|_{L^q([-T_0,T_0]\times\mathbb{R}^d)} \lesssim_\varepsilon N^{d-\frac{d+2}{q}+\varepsilon} \|f\|_{L^2} \|g\|_{L^2} \quad \text{for all } \frac{d+3}{d+1} \le q < \frac{d+2}{d}, \tag{15}$$

for any $\varepsilon > 0$, $N \ge 1$, *and* $V \in \mathcal{V}$.

For $V = 0$, the above estimate was conjectured by Klainerman and Machedon without the epsilon loss, and first proved in [Wolff 2001] for the wave equation and subsequently in [Tao 2003] for the Schrödinger equation (both with the epsilon loss). Strictly speaking, the time truncation is not present in the original formulations of those estimates, but may be easily removed by a rescaling and limiting argument.

Finally, while we make no attempt to address general magnetic potentials, a simple case with some physical relevance does essentially follow from the proof for scalar potentials. The necessary modifications for the following theorem are sketched in the last section.

Theorem 1.6. *The conclusion of the previous theorem holds for Schrödinger operators of the form* $H(t) = -\frac{1}{2}(\nabla - iA)^2 + V(t,x)$, *where* $A = A_j dx^j$ *is a* 1*-form whose components are linear in the space variables* (*i.e., the vector potential for a uniform magnetic field*), *and the condition on the time increment* τ_0 *is replaced by*

$$\tau_0 \|a_{x\xi}\| + (\tau_0 + \tau_0^2)\|a_{xx}\| < \eta.$$

We remark that the restriction estimate (12) does *not* hold for all symbols satisfying the conditions (7) and (8). For instance, it was observed by in [Vargas 2005] that when $U(t) = e^{it\partial_x\partial_y}$ is the "nonelliptic" Schrödinger propagator in two space dimensions (thus $a = \xi_x\xi_y$), the bilinear restriction estimate (7) can fail unless the frequency supports of the two inputs are not only disjoint but also separated in both Fourier coordinates. In fact, the refinement (14) as stated is *false* for the nonelliptic equation; for a correct formulation, one should enlarge the symmetry group on the right side to include the hyperbolic rescalings $u(x,y) \mapsto u(\mu x, \mu^{-1}y)$; see [Rogers and Vargas 2006].

While the classical bicharacteristics of elliptic and nonelliptic propagators seemingly have no qualitative difference — and indeed the dispersive estimates hold equally well for both — the quantum propagators have radically different behavior in terms of oscillations in time. If one compares the traveling wave solutions

$$e^{i\left[x\xi_x + y\xi_y - \frac{t}{2}(\xi_x^2+\xi_y^2)\right]}, \qquad e^{i\left[x\xi_x + y\xi_y - t\xi_x\xi_y\right]},$$

it is evident that unlike in the elliptic case two solutions to the nonelliptic equation which are well-separated in spatial frequency need not decouple in time.

The lesson of this counterexample is that while the dispersive and Strichartz estimates follow directly from properties of the classical Hamiltonian flow, an *inverse* Strichartz estimate depends more subtly on the temporal oscillations of the quantum evolution, which is connected to the bilinear decoupling estimates.

1B. ***The main ideas.*** Suppose one has initial data $u_0 \in L^2$ such that the corresponding solution u has nontrivial Strichartz norm. Then, we need to identify a bubble of concentration in u, characterized by several parameters that reflect the underlying symmetries in the problem. In the L^2-critical setting, the relevant features consist of a significant length scale λ_0 as well as the position x_0, frequency ξ_0, and time t_0 when concentration occurs.

The existing proofs of Strichartz refinements for the constant-coefficient equation first use space-time Fourier analysis (including restriction estimates) to identify a cube Q in Fourier space accounting for a significant portion of the space-time norm of u, which reveals the frequency center ξ_0 and scale λ_0 of the concentration. For example, [Bégout and Vargas 2007] first establishes an estimate of the form

$$\|e^{\frac{it\Delta}{2}} f\|_{L^{2(d+2)/d}} \lesssim \left(\sup_{Q \text{ dyadic cubes}} |Q|^{1-\frac{p}{2}} \int_Q |\hat{f}(\xi)|^p \, d\xi \right)^{\mu} \|f\|_{L^2(\mathbb{R}^d)}^{1-\mu p}.$$

Then, the time t_0 and position x_0 are recovered via a separate physical-space argument. These arguments ultimately rely on the fact that when $V = 0$, the equation is diagonalized by the Fourier transform.

For equations with variable coefficients, it is more natural to consider position x_0 and frequency ξ_0 together as a point in phase space, which propagates along the bicharacteristics for the equation. Following the approach in [Jao et al. 2019] for the one-dimensional equation, we work in the physical space and first isolate a significant time interval $[t_0 - \lambda_0^2, t_0 + \lambda_0^2]$, which also suggests a characteristic scale λ_0. Then x_0 and ξ_0 are recovered by phase-space techniques.

The first part of the argument in [Jao et al. 2019] carries over essentially unchanged; however, the ensuing phase-space analysis in higher dimensions is more involved and occupies the bulk of this article.

1C. ***An application to mass-critical NLS.*** This article was originally motivated by the problem of proving global well-posedness for the mass-critical quantum harmonic oscillator

$$i\partial_t u = \Big(-\tfrac{1}{2}\Delta + \sum_j \omega_j^2 x_j^2\Big)u \pm |u|^{\frac{4}{d}} u. \tag{16}$$

By spectral theory, the Cauchy problem for (16) is naturally posed in the "harmonic" Sobolev spaces

$$u_0 \in \mathcal{H}^s := \Big\{u_0 \in L^2 : \Big(-\Delta + \sum_j \omega_j^2 |x|^2\Big)^{\frac{s}{2}} u_0 \in L^2\Big\}.$$

Global existence for data in the "energy" space $\mathcal{H}^1$ was studied in [Zhang 2005]. More recently, Poiret, Robert, and Thomann [Poiret et al. 2014] established probabilistic well-posedness in two space dimensions for all subcritical cases $0 < s < 1$, as well as for other supercritical problems. Another recent contribution by Burq, Thomann, and Tzvetkov [Burq et al. 2013] constructs Gibbs measures and proves probabilistic global well-posedness for the critical case in one dimension.

It is well-known that the *isotropic* harmonic oscillator $\omega_j \equiv \frac{1}{2}$ may be "trivially" solved; to construct solutions on unit-length time intervals for arbitrary L^2 data, it suffices to observe that u is a solution of (4) on $\mathbb{R}_t \times \mathbb{R}_x^d$ if and only if its *Lens transform*

$$\mathcal{L}u(t,x) := \frac{1}{(\cos t)^{\frac{d}{2}}} u\Big(\tan t, \frac{x}{\cos t}\Big) e^{-\frac{i|x|^2 \tan t}{2}}$$

solves (16) on $\big(-\frac{\pi}{2} \times \frac{\pi}{2}\big)_t \times \mathbb{R}_x^d$ with the same initial data. However, this trick relies on algebraic cancellations that no longer hold for more general harmonic oscillators. For further discussion of the nonlinear harmonic oscillator as well as its connection with the Lens transform, consult [Carles 2011].

To solve (16) for large data in the critical space L^2, the concentration compactness and rigidity approach is much more promising. Experience has shown that constructing suitable profile decompositions is a core difficulty in implementing this strategy for dispersive equations with broken symmetries (e.g., loss of translation-invariance). For instance, see [Jao 2016] for the energy-critical variant of the quantum harmonic oscillator, as well as [Ionescu et al. 2012; Killip et al. 2016] for other energy-critical NLS on non-Euclidean domains. Thus this article supplies the main harmonic analysis input for the deterministic large-data theory of (16) at the critical regularity.

2. Preliminaries

2A. ***Notation.*** We use the Japanese bracket notation $\langle x\rangle := (1+|x|^2)^{1/2}$.

2B. ***Classical flow estimates.*** We collect some elementary properties of the classical Hamiltonian flow

$$\begin{cases} \dot{x} = a_\xi, & x(0) = y, \\ \dot{\xi} = -a_x, & \xi(0) = \eta. \end{cases} \tag{17}$$

Solutions to this system are *bicharacteristics*. For a point $z = (x, \xi)$ in phase space, let $\sigma \mapsto z^\sigma = (x^\sigma, \xi^\sigma)$ denote the bicharacteristic initialized at (x, ξ). Write $(y, \eta) \mapsto (x^t(y, \eta), \xi^t(y, \eta))$ for the flow map.

The linearization of (17) satisfies the following Gronwall estimates:

Lemma 2.1. *Suppose* $|t|\,\|\partial^2_{x,\xi} a\|_{L^\infty} \le 1$. *Then*

$$\begin{aligned} \frac{\partial x^t}{\partial \eta} &= \int_0^t a_{\xi\xi}(\tau, x^\tau, \xi^\tau)\,d\tau + O(t^2\|a_{x\xi}\|\,\|a_{\xi\xi}\|) + O(t^3\|a_{xx}\|\,\|a_{\xi\xi}\|^2), \\ \frac{\partial \xi^t}{\partial \eta} &= I + O(t\|a_{\xi x}\|) + O(t^2\|a_{xx}\|\,\|a_{\xi\xi}\|), \\ \frac{\partial x^t}{\partial y} &= I + O(t\|a_{x\xi}\|) + O(t^2\|a_{xx}\|\,\|a_{\xi\xi}\|), \\ \frac{\partial \xi^t}{\partial y} &= \int_0^t -a_{xx}(\tau, x^\tau, \xi^\tau)\,d\tau + O(t^2\|a_{xx}\|\,\|a_{x\xi}\|) + O(t^3\|a_{xx}\|^2\,\|a_{\xi\xi}\|). \end{aligned} \tag{18}$$

Proof. The linearized system takes the form

$$\begin{aligned} \dot{y} &= a_{\xi x} y + a_{\xi\xi}\eta, \\ \dot{\eta} &= -a_{xx} y - a_{x\xi}\eta. \end{aligned}$$

A preliminary application of Gronwall implies $|y(t)| + |\eta(t)| \lesssim |y(0)| + |\eta(0)|$.

Consider initial data $y(0) = I,\ \eta(0) = 0$. Then

$$|\eta(t)| \le \int_0^t |a_{xx} y|\,d\tau + \int_0^t |a_{x\xi}\eta(\tau)|\,d\tau,$$

so $|\eta(t)| \lesssim t\|a_{xx}\|$. Substituting this into the equation for y, we deduce

$$|y - I| \le \int_0^t |a_{\xi x} y|\,d\tau + \int_0^t |a_{\xi\xi}\eta|\,d\tau \lesssim t\|a_{\xi x}\| + t^2\|a_{\xi\xi}\|\,\|a_{xx}\|.$$

This in turn yields the refinement

$$\left| \eta(t) + \int_0^t a_{xx}\,d\tau \right| \lesssim t^2\|a_{xx}\|\,\|a_{\xi x}\| + t^3\|a_{xx}\|^2\,\|a_{\xi\xi}\|.$$

The case $y(0) = 0,\ \eta(0) = I$ is similar. We have

$$|y(t)| \le \int_0^t |a_{\xi\xi}\eta|\,d\tau + \int_0^t |a_{\xi x} y|\,d\tau \quad \Longrightarrow \quad |y(t)| \lesssim t\|a_{\xi\xi}\|,$$

which yields

$$|\eta(t)-I| \lesssim \int_0^t \|a_{xx}\|\,\|a_{\xi\xi}\|\tau\,d\tau + \int_0^t |a_{x\xi}\eta|\,d\tau \lesssim t\|a_{x\xi}\| + t^2\|a_{xx}\|\,\|a_{\xi\xi}\|,$$

$$\left| y(t) - \int_0^t a_{\xi\xi}\,d\tau \right| \lesssim t^2\|a_{\xi x}\|\,\|a_{\xi\xi}\| + t^3\|a_{xx}\|\,\|a_{\xi\xi}\|^2. \qquad \square$$

These imply, in view of the normalizations (8), the integrated estimates

$$\begin{aligned} x_1^t - x_2^t &= x_1^s - x_2^s + [I' + O(\varepsilon)](t-s)(\xi_1^s - \xi_2^s) \\ &\quad + O(|t-s|\,\|a_{x\xi}\|)(|x_1^s - x_2^s| + |t-s|\,|\xi_1^s - \xi_2^s|) \\ &\qquad + O(|t-s|^2\|a_{xx}\|)(|x_1^s - x_2^s| + |t-s|\,|\xi_1^s - \xi_2^s|), \\ \xi_1^t - \xi_2^t &= \xi_1^s - \xi_2^s + O(|t-s|\,\|a_{xx}\|)|x_1^s - x_2^s| \\ &\quad + O(|t-s|^2\,\|a_{xx}\|\,\|a_{x\xi}\|)|x_1^s - x_2^s| + O(|t-s|\,\|a_{x\xi}\|)|\xi_1^s - \xi_2^s| \\ &\qquad + O(|t-s|^3\|a_{xx}\|^2)|x_1^s - x_2^s| + O(|t-s|^2\,\|a_{xx}\|)|\xi_1^s - \xi_2^s|, \end{aligned} \tag{19}$$

where I' is an orthogonal matrix which equals the identity if $a_{\xi\xi}$ is positive-definite. In particular, we have:

Corollary 2.2. *If $|x_1^s - x_2^s| \le r$, then $|x_1^t - x_2^t| \ge Cr$ whenever $2Cr/|\xi_1^s - \xi_2^s| \le |t-s| \le T_0$.*

Physically, this means that two particles colliding with sufficiently large relative velocity will only interact once in the time window of interest.

Next, we record a technical lemma first proved in the 1-dimensional case [Jao et al. 2019, Lemma 2.2]. This is used in the proof of Lemma 4.3 below but the computations use the preceding estimates.

Lemma 2.3. *There exists a constant $C = C(\|\partial^2 a\|) > 0$ so that if $Q_\eta = (0,\eta) + [-1,1]^{2d} \subset T^*\mathbb{R}^d$ and $r \ge 1$, then*

$$\bigcup_{|t-t_0|\le \min(|\eta|^{-1},1)} \Phi(t)^{-1}(z_0^t + rQ_\eta) \subset \Phi(t_0)^{-1}(z_0^{t_0} + CrQ_\eta).$$

In other words, if the bicharacteristic z^t starting at $z \in T^*\mathbb{R}^d$ passes through the cube $z_0^t + rQ_\eta$ in phase space during some time window $|t-t_0| \le \min(|\eta|^{-1},1)$, then it must lie in the dilate $z_0^{t_0} + CrQ_\eta$ at time t_0.

Proof. If $z \in \Phi(t)^{-1}(z_0^t + rQ_\eta)$, by definition we have $|x^t - x_0^t| \le r$ and $|\xi^t - \xi_0^t - \eta| \le r$. Assuming that $|\eta| \ge 1$, the estimates (19) imply

$$\begin{aligned} |x^{t_0} - x_0^{t_0}| &\le r + |\eta|^{-1}(|\eta|+r) + O(|\eta|^{-1}\|\partial^2 a\|)(r + |\eta|^{-1}(|\eta|+r)) + O(|\eta|^{-2}\|\partial^2 a\|)(r + |\eta|^{-1}(|\eta|+r)) \\ &\le Cr, \\ |\xi^{t_0} - \xi_0^{t_0} - \eta| &\le r + O(|\eta|^{-1}\|a_{xx}\|)r + (|\eta|^{-2}\|a_{xx}\|\,\|a_{x\xi}\|)r + O(|\eta|^{-1}\|a_{x\xi}\|)(|\eta|+r) \\ &\qquad + (|\eta|^{-3}\|a_{xx}\|^2)r + O(|\eta|^{-2}\|a_{xx}|)(|\eta|+r) \\ &\le Cr. \end{aligned}$$

The case $|\eta| < 1$ is similar. $\square$

2C. *Wavepackets.* Let $R \geq 1$ be a scale and $z_0 = (x_0, \xi_0)$ be a point in phase space. A scale-R *wavepacket* at z_0 is a Schwartz function ϕ_{z_0} such that ϕ_{z_0} and its Fourier transform $\hat{\phi}_{z_0}$ concentrate in the regions $|x - x_0| \leq R^{1/2}$ and $|\xi - \xi_0| \leq R^{-1/2}$, respectively:

$$|(R^{\frac{1}{2}}\partial_x)^k \phi_{z_0}(x)| \lesssim_{k,N} \left\langle \frac{x - x_0}{R^{\frac{1}{2}}} \right\rangle^{-N}, \quad |(R^{-\frac{1}{2}}\partial_\xi)^k \hat{\phi}_{z_0}(\xi)| \lesssim_{k,N} \left\langle \frac{\xi - \xi_0}{R^{-\frac{1}{2}}} \right\rangle^{-N} \quad \text{for all } k, N \geq 0.$$

There are many ways to decompose L^2 functions into linear combinations of wavepackets. For the first part of this article, it is technically more convenient to use a continuous decomposition. Later on in Section 6C, we switch to a discrete version which is more common in the restriction theory literature.

In this section we recall a standard continuous wavepacket transform. To keep things simple we work at unit scale since that is all we shall need. For a function $f \in L^2(\mathbb{R}^d)$, its *Bargmann transform* or *FBI transform* is the function $Tf \in L^2(T^*\mathbb{R}^d)$ defined by

$$Tf(z) = \langle f, \psi_z \rangle_{L^2(\mathbb{R}^d)}, \qquad \psi_z = \pi(z)\psi \quad \text{as in (13)}.$$

The transform satisfies a Plancherel identity $\|Tf\|_{L^2(T^*\mathbb{R}^d)} = \|f\|_{L^2(\mathbb{R}^d)}$; dually, for any wavepacket coefficients $F \in L^2(T^*\mathbb{R}^d_z)$, one has

$$\|T^*F\|_{L^2_x} = \left\| \int_{T^*\mathbb{R}^d} F(z)\psi_z \, dz \right\|_{L^2_x} \leq \|F\|_{L^2_z}.$$

Indeed, TT^* is the orthogonal projection onto $TL^2(\mathbb{R}^d)$. Then as $T^*T = I$, any $f \in L^2(\mathbb{R}^d)$ can be resolved (nonuniquely) into a continuous superposition of wavepackets

$$f(x) = \int_{T^*\mathbb{R}^d} f_z \psi_z(x) \, dz.$$

Applying the propagator $U(t)$ to both sides and using linearity and the next lemma, one obtains a wavepacket decomposition

$$u(t,x) = \int u_z(t,x) \, dz, \quad u_z(t,x) = f_z[U(t)\psi_z](x),$$

of Schrödinger solutions. For brevity we sometimes omit the arguments and write $f = \int f_z \, dz$, $u = \int u_z \, dz$.

Lemma 2.4 (evolution of a packet). *If ψ_{z_0} is a scale-*1 *wavepacket, $U(t)$ is the propagator for* (9), *and $z_0 \mapsto z_0^t$ is the bicharacteristic starting at z_0, then $U(t)\psi_{z_0}$ is a scale-*1 *wavepacket concentrated at z_0^t for all $|t| = O(1)$.*

Proof sketch. Using Lemma 1.1 we reduce to the case $z_0 = 0$ and also ensure that the symbol $a(t, x, \xi)$ vanishes to second order at $(x, \xi) = (0, 0)$ in addition to satisfying the bounds (7). Then it suffices to show that propagator $U(t)$ for such symbols maps Schwartz functions to Schwartz functions on unit time scales. This is done using weighted Sobolev estimates as in [Koch and Tataru 2005, Section 4]. □

The term *wavepacket* shall also refer to space-time functions of the form $U(t)\psi_z$, not just the fixed time slices. Later it will be essential to exploit not just the space-time localization of wavepackets but also their phase as described in Lemma 1.1.

3. Choosing a length scale

We begin with the following lemma from [Jao et al. 2019, Proposition 3.1], obtained by a variant of the usual TT^* derivation of the Strichartz estimates. While that article concerned just Schrödinger operators with scalar potentials, the proof works equally well in the current more general setting.

Proposition 3.1. *Suppose $U(t,s)$ satisfies a local-in-time dispersive estimate as in Lemma 1.2. Let (q,r) be Strichartz exponents (i.e., satisfying the conditions in that lemma) with $2 < q < \infty$. Assume that $f \in L^2(\mathbb{R}^d)$ satisfies $\|f\|_{L^2(\mathbb{R}^d)} = 1$ and*

$$\|U(t)f\|_{L_t^q L_x^r([-1,1]\times\mathbb{R}^d)} \geq \varepsilon.$$

Then there is a time interval $J \subset [-1,1]$ such that

$$\|U(t,s)f\|_{L_t^{q-1} L_x^r(J\times\mathbb{R}^d)} \gtrsim |J|^{\frac{1}{q(q-1)}} \varepsilon^{\frac{q}{q-2}}.$$

Equivalently,

$$\|U(t,s)f\|_{L^q L^r} \lesssim \Big(\sup_{J\in[-1,1]} |J|^{-\frac{1}{q(q-1)}} \|U(t,s)f\|_{L_t^{q-1}L_x^r(J\times\mathbb{R}^d)}\Big)^{1-\frac{2}{q}} \|f\|_{L^2(\mathbb{R}^d)}^{\frac{2}{q}}.$$

Note that by pigeonholing we may always assume that $|J| \leq T_0$, where T_0 is the time increment selected in (11).

Now let (q,r) be the Strichartz exponents determined by the conditions $\frac{2}{q} + \frac{d}{r} = \frac{d}{2}$ and $q - 1 = r$. It is easy to see that $2 < r < \frac{2(d+2)}{d} < q < \infty$.

For each $J = [s-\mu, s+\mu] \subset [-1,1]$, we write

$$U(t,s)f = \left(\frac{T_0}{\mu}\right)^{\frac{d}{4}} \tilde{U}\left(\frac{T_0}{\mu}(t-s), 0\right) \tilde{f}\left(\sqrt{\frac{T_0}{\mu}}x\right), \quad \tilde{f} = \left(\frac{\mu}{T_0}\right)^{\frac{d}{4}} f\left(\sqrt{\frac{\mu}{T_0}}x\right),$$

where $\tilde{U}(t,s)$ is the propagator for the rescaled equation $(D_t + \tilde{a}^w)\tilde{u} = 0$, and

$$\tilde{a}(t,x,\xi) := \frac{\mu}{T_0} a\left(s + \frac{\mu}{T_0}t, \sqrt{\frac{\mu}{T_0}}x, \sqrt{\frac{T_0}{\mu}}\xi\right).$$

Changing variables, we obtain

$$|J|^{-\frac{1}{q(q-1)}} \|U(t,s)f\|_{L_t^{q-1}L^r(J\times\mathbb{R}^d)} = \|\tilde{U}(t)\tilde{f}\|_{L_t^{q-1}L_x^r([-T_0,T_0]\times\mathbb{R}^d)}.$$

By interpolating with $L^2_{t,x}([-T_0,T_0]\times\mathbb{R}^d)$, which is bounded by unitarity, we see that Theorem 1.3 would follow if we prove that for some $2 < q_0 < \frac{2(d+2)}{d}$ and $0 < \theta < 1$, the scale-1 refined estimate

$$\|U_\lambda^s(t)f\|_{L^{q_0}([-T_0,T_0]\times\mathbb{R}^d)} \lesssim \big(\sup_z |\langle \psi_z, f\rangle|\big)^{\theta} \|f\|_{L^2}^{1-\theta} \tag{20}$$

holds for all $s \in [-1,1]$, $0 < \lambda \leq 1$, where the notation $U_\lambda^s(t)$ is as in Hypothesis 1.

Over the next two sections we establish:

Proposition 3.2. *If Hypothesis 1 holds, then so does the estimate (20).*

4. A refined bilinear L^2 estimate

In previous work [Jao et al. 2019], we proved (20) when $d = 1$ with $q_0 = 4$ by viewing the inequality as a bilinear L^2 estimate and exploit orthogonality. Such a direct approach fails in $d \geq 2$ dimensions; since $2 < \frac{2(d+2)}{d} \leq 4$, the left side of (20) could well be infinite when $q_0 = 4$. To obtain a refined linear L^{q_0} estimate for $q_0 < \frac{2(d+2)}{d}$, we also begin by interpreting it as a refined bilinear $L^{q_0/2}$ estimate, but use dyadic decomposition and interpolate between two microlocalized estimates:

- A refined bilinear L^2 estimate ("refined" in the sense of exhibiting a sup over wavepacket coefficients) with some loss in the frequency separation of the inputs.
- A bilinear L^p estimate for some $p < \frac{d+2}{d}$ which yields gains in the frequency separation, essentially the content of Hypothesis 1.

This section discusses the former. In the next section we put together the two estimates, and the L^p estimate is established in the remainder of the paper.

Proposition 4.1. *Suppose $f = \int f_z \psi_z \, dz$ and $g = \int g_z \psi_z \, dz$ are $L^2(\mathbb{R}^d)$ initial data with corresponding Schrödinger evolutions $u = \int u_z \, dz$ and $v = \int v_z \, dz$, where $u_z(t,x) = f_z[U(t)\psi_z](x)$, $v_z(t,x) = g_z[U(t)\psi_z](x)$. Then*

$$\Big\| \int_{|\xi_1 - \xi_2| \sim N} u_{z_1} v_{z_2} \, dz_1 \, dz_2 \Big\|_{L^2([-T_0,T_0]\times\mathbb{R}^d)} \lesssim N^\alpha \big(\sup_z |f_z|^{\frac{1}{p'}} \|f_z\|_{L^2_z}^{\frac{1}{p}}\big)\big(\sup_z |g_z|^{\frac{1}{p'}} \|g_z\|_{L^2_z}^{\frac{1}{p}}\big) \tag{21}$$

for some $\alpha = \alpha(d)$ and $1 < p < 2$.

Proof. Square the left side and expand

$$\int f_{z_1} g_{z_2} \bar{f}_{z_3} \bar{g}_{z_4} K_N(z_1, z_2, z_3, z_4) \, dz_1 \, dz_2 \, dz_3 \, dz_4,$$

where $K_N := K \chi_{|\xi_1 - \xi_2| \sim N, \ |\xi_3 - \xi_4| \sim N}$, and

$$K(z_1, z_2, z_3, z_4) = \langle U(t)\psi_{z_1} U(t)\psi_{z_2}, \ U(t)\psi_{z_3} U(t)\psi_{z_4} \rangle_{L^2_{t,x}([-T_0,T_0]\times\mathbb{R}^d)}.$$

The estimate would follow if we could show that

$$N^{-\alpha} \langle z_1 - z_2 \rangle^\theta \langle z_3 - z_4 \rangle^\theta |K_N(\vec{z})| \text{ is a bounded operator on } L^2_{z_1,z_2} \text{ for some } \theta > 0, \tag{22}$$

as Young's inequality would then imply

$$\begin{aligned} \Big\| \int u_z \, dz \Big\|_{L^4}^2 &\lesssim \Big(\int |f_{z_1} g_{z_2}|^2 \langle z_1 - z_2 \rangle^{-2\theta} \, dz_1 \, dz_2 \Big)^{\frac{1}{2}} \Big(\int |f_{z_3} g_{z_4}|^2 \langle z_3 - z_4 \rangle^{-2\theta} \, dz_3 \, dz_4 \Big)^{\frac{1}{2}} \\ &\lesssim \sup_z |f_z|^{\frac{2}{p'}} \sup_z |g_z|^{\frac{2}{p'}} \|f\|_{L^2}^{\frac{2}{p}} \|g\|_{L^2}^{\frac{2}{p}} \quad \text{for some } 1 < p < 2. \end{aligned}$$

In view of the crude bound $|K(\vec{z})| \lesssim \min_{j,k} \langle z_j - z_k \rangle^{-1}$, which follows simply from the space-time supports of the wavepackets, (22) would follow from:

Lemma 4.2. *The localized kernel K_N satisfies*

$$\| |K_N|^{1-\delta} \|_{L^2_{z_1 z_2} \to L^2_{z_3 z_4}} \lesssim N^\alpha,$$

where α is a constant depending only on the dimension.

Proof of Lemma 4.2. In view of the unit scale spatial localization of the wavepackets and the propagation estimates (19), we may further truncate the kernel to the phase-space region

$$R = \{ |x_1 - x_2| \le 4|\xi_1 - \xi_2|, \ |x_3 - x_4| \le 4|\xi_3 - \xi_4| \}.$$

For instance, if $|x_1^s - x_2^s| \ge 4|\xi_1^s - \xi_2^s|$ and $|t-s| \le T_0$ with the parameter η in (11) chosen sufficiently small,

$$\begin{aligned} |x_1^t - x_2^t| &\ge (1 - |t-s|^2 \|\partial_x^2 V\|_{L^\infty} e^{|t-s|^2 \|\partial_x^2 V\|_{L^\infty}}) |x_1^s - x_2^s| \\ &\qquad - (|t-s| + |t-s|^3 \|\partial_x^2 V\|_{L^\infty} e^{|t-s|^2 \|\partial_x^2 V\|_{L^\infty}}) |\xi_1^s - \xi_2^s| \\ &\ge \tfrac{1}{2} |x_1^s - x_2^s| - \tfrac{3}{2} |t-s| \, |\xi_1^s - \xi_2^s| \\ &\ge \tfrac{1}{8} |x_1^s - x_2^s|. \end{aligned}$$

Therefore $|K_N(1-\chi_R)| \lesssim_M \langle x_1 - x_2 \rangle^{-M} \langle x_3 - x_4 \rangle^{-M} N^{-M}$ for any $M > 0$. Thus it suffices to prove that

$$\| K_N \chi_R \|_{L^2 \to L^2} \lesssim N^\alpha.$$

An estimate of this flavor was proved in the 1-dimensional case [Jao et al. 2019]. We shall argue similarly, but the proof is somewhat simpler since we aim for a cruder bound at this stage, completely ignoring temporal oscillations, and defer the more delicate analysis to the bilinear L^p estimate.

Partition the 4-particle phase space $(T^*\mathbb{R}^d)^4$ according to the degree of physical interaction between the particles. Let

$$\begin{aligned} E_0 &= \big\{ \vec{z} \in (T^*\mathbb{R}^d)^4 : \min_{|t| \le T_0} \max_{j,k} |x_j^t - x_k^t| \le 1 \big\}, \\ E_k &= \big\{ \vec{z} \in (T^*\mathbb{R}^d)^4 : 2^{k-1} < \min_{|t| \le T_0} \max_{j,k} |x_j^t - x_k^t| \le 2^k \big\}, \end{aligned}$$

and decompose the kernel into $K_N = \sum_{k \ge 0} K_N \chi_{E_k}$. Then we have the pointwise bound

$$|K(\vec{z})| \lesssim_M 2^{-kM} \frac{\langle \xi_1^{t(\vec{z})} + \xi_2^{t(\vec{z})} - \xi_3^{t(\vec{z})} - \xi_4^{t(\vec{z})} \rangle^{-M}}{\langle |\xi_1^{t(\vec{z})} - \xi_2^{t(\vec{z})}| + |\xi_3^{t(\vec{z})} - \xi_4^{t(\vec{z})}| \rangle}, \quad \vec{z} \in E_k, \tag{23}$$

where $t(\vec{z})$ is a time minimizing the "mutual distance" $\max_{i,j} |x_i^t - x_j^t|$. Further, the additional localization to R implies, by the estimates (19), that

$$\begin{aligned} |\xi_1^t - \xi_2^t - (\xi_1 - \xi_2)| &\lesssim \tfrac{1}{10} |\xi_1 - \xi_2|, \\ |\xi_3^t - \xi_4^t - (\xi_3 - \xi_4)| &\lesssim \tfrac{1}{10} |\xi_3 - \xi_4| \end{aligned}$$

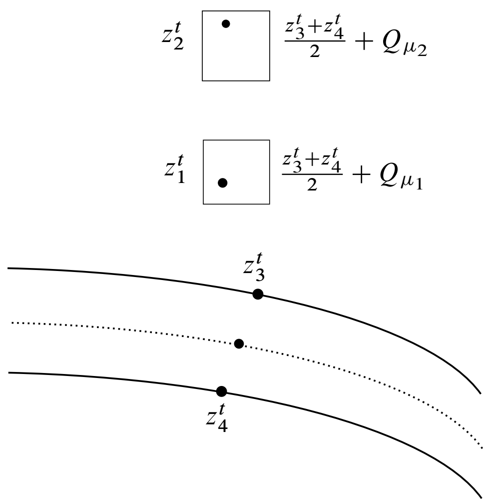

Figure 1. Z_{μ_1,μ_2} comprises all (z_1, z_2) such that z_1^t and z_2^t belong to the depicted phase-space boxes for t in the interval I.

for all $|t| \le T_0$. In particular $|\xi_1^{t(\vec{z})} - \xi_2^{t(\vec{z})}| \sim |\xi_3^{t(\vec{z})} - \xi_4^{t(\vec{z})}| \sim N$; thus, while the ξ_j^t may vary rapidly with time if x_j^t are extremely far from the origin, the *relative frequencies* retain the same order of magnitude.

Assuming the bound (23) for the moment, we apply Schur's test to complete the proof of Lemma 4.2. Fix (z_3, z_4) belonging to the projection $E_k \to T^*\mathbb{R}^d_{z_3} \times T^*\mathbb{R}^d_{z_4}$, define

$$E_k(z_3, z_4) = \{(z_1, z_2) : (z_1, z_2, z_3, z_4) \in E_k\},$$

and let t_1 be the time minimizing $|x_3^{t_1} - x_4^{t_1}| \le 2^k$. For any $(z_1, z_2) \in E_k(z_3, z_4)$, the mutual distance $\max_{j,k} |x_j^t - x_k^t|$ between x_1^t, x_2^t, x_3^t, x_4^t is minimized in the time window

$$I = \left\{ t : |t - t_1| \lesssim \min\left(1, \frac{2^k}{|\xi_3 - \xi_4|}\right) \right\},$$

as for all other times we have $|x_3^t - x_4^t| \gg 2^k$ (Corollary 2.2).

We estimate the size of the level sets of $|K|$. For a momentum $\xi \in \mathbb{R}^d$, denote by $Q_\xi = (0, \xi) + [-1, 1]^d \times [-1, 1]^d \subset T^*\mathbb{R}^d$ the unit phase-space box centered at $(0, \xi)$, and write $\Phi^t = \Phi(t, 0)$ for the propagator on classical phase space relative to time 0 for the Hamiltonian $h(x, \xi) = \frac{1}{2}|\xi|^2 + V(t, x)$. For $\mu_1, \mu_2 \in \mathbb{R}^d$, define

$$Z_{\mu_1,\mu_2} = \bigcup_{t \in I} (\Phi^t \otimes \Phi^t)^{-1} \left(\frac{z_3^t + z_4^t}{2} + 2^k Q_{\mu_1} \right) \times \left(\frac{z_3^t + z_4^t}{2} + 2^k Q_{\mu_2} \right).$$

This set is depicted schematically in Figure 1 when $k = 0$, and corresponds to the pairs of wave packets $(z_1, z_2) \in E_m(z_3, z_4)$ with momenta (μ_1, μ_2) relative to the wavepackets (z_3, z_4) at the "collision time" $t(\vec{z})$.

We note that $E_k(z_3, z_4) \subset \bigcup_{\mu_1,\mu_2\in\mathbb{Z}^d} Z_{\mu_1,\mu_2}$, and recall the following estimate from the 1-dimensional paper, whose proof we reproduce below for convenience:

Lemma 4.3.
$$|Z_{\mu_1,\mu_2}| \lesssim 2^{4dk} \max(1, |\mu_1, |, |\mu_2|)|I|. \tag{24}$$

Proof. Without loss assume $|\mu_1| \geq |\mu_2|$. Partition the interval I into subintervals of length $|\mu_1|^{-1}$ if $\mu_1 \neq 0$ and into subintervals of length 1 if $\mu_1 = 0$. For each t' in the partition, Lemma 2.3 implies that for some constant $C > 0$ we have

$$\bigcup_{|t-t'|\leq\min(1,|\mu_1|^{-1})} \Phi(t)^{-1}\Big(\frac{z_3^t + z_4^t}{2} + 2^k Q_{\mu_1}\Big) \subset \Phi(t')^{-1}\Big(\frac{z_3^{t'} + z_4^{t'}}{2} + C2^k Q_{\mu_1}\Big),$$
$$\bigcup_{|t-t'|\leq\min(1,|\mu_1|^{-1})} \Phi(t)^{-1}\Big(\frac{z_3^t + z_4^t}{2} + 2^k Q_{\mu_2}\Big) \subset \Phi(t')^{-1}\Big(\frac{z_3^{t'} + z_4^{t'}}{2} + C2^k Q_{\mu_2}\Big),$$

and so

$$\bigcup_{|t-t'|\leq\min(1,|\mu_1|^{-1})} (\Phi(t)\otimes\Phi(t))^{-1}\Big(\frac{z_3^t + z_4^t}{2} + 2^k Q_{\mu_1}\Big) \times \Big(\frac{z_3^t + z_4^t}{2} + 2^k Q_{\mu_2}\Big)$$
$$\subset (\Phi(t')\otimes\Phi(t'))^{-1}\Big(\frac{z_3^{t'} + z_4^{t'}}{2} + C2^k Q_{\mu_1}\Big) \times \Big(\frac{z_3^{t'} + z_4^{t'}}{2} + C2^k Q_{\mu_2}\Big).$$

By Liouville's theorem, the right side has measure $O(2^{4dk})$ in $(T^*\mathbb{R}^d)^2$. The claim follows by summing over the partition. □

For each $(z_1, z_2) \in E_k(z_3, z_4) \cap Z_{\mu_1,\mu_2}$, we have by definition

$$z_j^{t(\vec{z})} \in \frac{z_3^{t(\vec{z})} + z_4^{t(\vec{z})}}{2} + 2^k Q_{\mu_j}.$$

Thus

$$\xi_1^{t(\vec{z})} + \xi_2^{t(\vec{z})} - \xi_3^{t(\vec{z})} - \xi_4^{t(\vec{z})} = \mu_1 + \mu_2 + O(2^k),$$
$$\xi_1^{t(\vec{z})} - \xi_2^{t(\vec{z})} = \mu_1 - \mu_2 + O(2^k).$$

Hence when $(z_1, z_2) \in Z_{\mu_1,\mu_2}$, for any M we have

$$|K(\vec{z})| \lesssim_M 2^{-Mk} \frac{\langle\mu_1 + \mu_2\rangle^{-M}}{\langle|\mu_1 - \mu_2| + |\xi_3^{t(\vec{z})} - \xi_4^{t(\vec{z})}|\rangle}. \tag{25}$$

To apply Schur's test, we combine the estimates (24), (25), and evaluate

$$\int |K_N(z_1, z_2, z_3, z_4)|^{1-\delta} \chi_{E_k}(\vec{z})\, dz_1\, dz_2 \leq \sum_{\mu_1,\mu_2\in\mathbb{Z}^d} \int_{Z_{\mu_1,\mu_2}} |K_N^{1-\delta} \chi_{E_k}\, dz_2\, dz_2$$
$$\lesssim_M 2^{-Mk} \sum_{|\mu_1-\mu_2|\lesssim N+2^k} 2^{-Mk} \langle\mu_1 + \mu_2\rangle^{-M}$$
$$\lesssim N^d 2^{-(M-d)k}.$$

For fixed z_1, z_2, the integral over z_3 and z_4 is estimated the same way. This concludes the proof of Lemma 4.2, modulo some remarks on the crucial pointwise bound (23).

To obtain that estimate, we use Lemma 1.1 to write

$$K(\vec{z}) = \int e^{i\Phi} \prod_{j=1}^{4} U^{z_j}(t)\psi(x - x_j^t)\, dx\, dt,$$

$$\Phi(t, x; \vec{z}) = \sum_j \sigma_j [\langle x - x_j^t, \xi_j^t\rangle + \phi(t, x_0, \xi_0)],$$

where $\sigma = (+,+,-,-)$, and we set $\prod_j c_j := c_1 c_2 \bar{c}_3 \bar{c}_4$.

It is convenient to partition the integral further, writing

$$U^{\vec{z}_j}(t)\psi(x - x_j^t) = \sum_{\ell_j \geq 0} U^{\vec{z}_j}(t)\psi(x - x_j^t)\theta_{\ell_j}(x - x_j^t),$$

where $\sum_{\ell \geq 0} \theta_\ell$ is a partition of unity with θ_ℓ supported on the dyadic annulus of radius $\sim 2^\ell$. For $\vec{z} \in E_k$, only the terms

$$K_{\vec{\ell}}(\vec{z}) := \int e^{i\Phi} \prod_{j=1}^{4} U^{z_j}(t)\psi(x - x_j^t)\theta_{\ell_j}(x - x_j^t)\, dx\, dt,$$

with $\ell^* := \max_j \ell_j \gtrsim k$, will be nonzero.

By Lemma 2.1, the integral is supported on the space-time region

$$\left\{(t, x) : |t - t(\vec{z})| \lesssim \min\left(1, \frac{2^{\ell^*}}{\max_{i,j} |\xi_i^{t(\vec{z})} - \xi_k^{t(\vec{z})}|}\right) \text{ and } |x - x_j^t| \lesssim 2^{\ell_j}\right\},$$

and for all such t we have

$$|x_j^t - x_k^t| \lesssim 2^{\ell^*}, \quad |\xi_j^t - \xi_k^t - (\xi_j^{t(\vec{z})} - \xi_k^{t(\vec{z})})| \lesssim 2^{\ell^*}.$$

Integrating by parts in x, we may produce as many factors of $|\xi_1^t + \xi_2^t - \xi_3^t - \xi_4^t|^{-1}$ as desired and freeze $t = t(\vec{z})$ to obtain

$$|K_{\vec{\ell}}(\vec{z})| \lesssim_M 2^{-\ell^* M} \frac{\langle \xi_1^{t(\vec{z})} + \xi_2^{t(\vec{z})} - \xi_3^{t(\vec{z})} - \xi_4^{t(\vec{z})}\rangle^{-M}}{\langle |\xi_1^{t(\vec{z})} - \xi_2^{t(\vec{z})}| + |\xi_3^{t(\vec{z})} - \xi_4^{t(\vec{z})}|\rangle} \quad \text{for any } M \geq 0,$$

and the bound (23) follows upon summing over $\vec{\ell}$. □

This completes the proof of Proposition 4.1. □

5. Proof of Theorem 1.3

We prove Proposition 3.2 and hence Theorem 1.3. Begin with a Whitney decomposition of

$$(\mathbb{R}^d \times \mathbb{R}^d) \setminus \{(\xi, \xi) : \xi \in \mathbb{R}^d\} = \bigcup_{N \in 2^{\mathbb{Z}}} \bigcup_{Q \in \mathcal{Q}_N} Q,$$

where $\mathcal{Q}_N$ is the set of dyadic cubes in $\mathbb{R}^d \times \mathbb{R}^d$ with diameter $\sim N$ and distance $\sim N$ to the diagonal. For each $Q \in \mathcal{Q}_N$, its characteristic function factors into

$$\chi_N^Q(\xi_1, \xi_2) = \chi_N^{Q,1}(\xi_2)\chi_N^{Q,2}(\xi_2),$$

where $\chi_N^{Q,j}$ are characteristic functions of d-dimensional cubes of width N. Then we can take the decomposition

$$1(\xi_1, \xi_2) = \chi_0(\xi_1, \xi_2) + \sum_{N \geq 1} \sum_{Q \in \mathcal{Q}_N} \chi_N^{Q,1}(\xi_1)\chi_N^{Q,2}(\xi_2),$$

where $\chi_0(\xi_1, \xi_2)$ is supported on the set $|\xi_1 - \xi_2| \lesssim 1$.

Now suppose u and v are linear solutions with initial data $f = \int f_z \psi_z \, dz$ and $g = \int g_z \psi_z \, dz$, respectively, where $f_z = \langle f, \psi_z \rangle$ and $g_z = \langle g, \psi_z \rangle$. Writing $u_z = f_z U(t)\psi_z$, $v_z = g_z U(t)\psi_z$, we deduce as a consequence of Hypothesis 1 that

$$\Big\| \sum_{Q \in \mathcal{Q}_N} \int_Q u_{z_1} v_{z_2} \, dz_1 \, dz_2 \Big\|_{L^q([-T_0, T_0] \times \mathbb{R}^d)} \lesssim N^{-\delta} \| f_z \|_{L_z^2} \| g_z \|_{L_z^2} \tag{26}$$

for each $N \geq 1$. Indeed, for each cube Q the integral has a product structure

$$\begin{aligned} \int_Q u_{z_1} v_{z_2} \, dz_1 \, dz_2 &= \Big(\int u_{z_1} \chi_N^{Q,1}(\xi_1) \, dx_1 \, d\xi_1 \Big) \Big(\int v_{z_2} \chi_N^{Q,2}(\xi_2) \, dx_2 \, d\xi_2 \Big) \\ &= U(t) \Big[\int f_{z_1} \chi_N^{Q,1}(\xi_1) \psi_{z_1} \, dx_1 \, d\xi_1 \Big] U(t) \Big[\int g_{z_2} \chi_N^{Q,1}(\xi_2) \psi_{z_2} \, dx_2 \, d\xi_2 \Big]. \end{aligned}$$

By the rapid decay of the wavepackets, we may harmlessly insert frequency cutoffs $\tilde{\chi}_N^{Q,j}(D)$, where $\tilde{\chi}_N^{Q,j}$ are slightly fattened versions of $\chi_N^{Q,j}$ and still have supports separated by distance $\sim N$, and apply Hypothesis 1 to estimate

$$\begin{aligned} \Big\| \int_Q u_{z_1} v_{z_2} \, dz_1 \, dz_2 \Big\|_{L^q} &\lesssim N^{-\delta} \Big\| \int f_{z_1} \chi_N^{Q,1}(\xi_1) \, dx_1 \, d\xi_1 \Big\|_{L^2(\mathbb{R}^d)} \Big\| \int g_{z_2} \chi_N^{Q,2}(\xi_2) \, dx_2 \, d\xi_2 \Big\|_{L^2(\mathbb{R}^d)} \\ &\lesssim N^{-\delta} \| f_z \chi_N^{Q,1}(\xi) \|_{L_z^2} \| g_z \chi_N^{Q,2} \chi(\xi) \|_{L_z^2}. \end{aligned}$$

The left side of (26) is therefore bounded by

$$\begin{aligned} \sum_{Q \in \mathcal{Q}_N} N^{-\delta} \| f_z \chi_N^{Q,1}(\xi) \|_{L_z^2} \| g_z \chi_N^{Q,2} \chi(\xi) \|_{L_z^2} &\leq N^{-\delta} \Big(\sum_{Q \in \mathcal{Q}_N} \| f_z \chi_N^{Q,1}(\xi) \|_{L_z^2}^2 \Big)^{\frac{1}{2}} \Big(\sum_{Q \in \mathcal{Q}_N} \| g_z \chi_N^{Q,2}(\xi) \|_{L_z^2}^2 \Big)^{\frac{1}{2}} \\ &\lesssim N^{-\delta} \| f_z \|_{L_z^2} \| g_z \|_{L_z^2}, \end{aligned}$$

as claimed.

Now decompose the product uv into

$$uv = \int u_{z_1} v_{z_2} \chi_0(\xi_1, \xi_2) \, dz_1 \, dz_2 + \sum_{N \geq 1} \sum_{Q \in \mathcal{Q}_N} \int_Q u_{z_1} v_{z_2} \, dz_1 \, dz_2,$$

and estimate each group of terms in L^q for q between p and 2. For the sum over $\mathcal{Q}_N$ we interpolate between the L^p and L^2 bounds. Writing $\frac{1}{q} = \frac{1-\theta}{p} + \frac{\theta}{2}$, we have

$$\Big\| \sum_{Q \in \mathcal{Q}_N} \int_Q u_{z_1} v_{z_2}\, dz_1\, dz_2 \Big\|_{L^q} \le \Big\| \sum_{Q \in \mathcal{Q}_N} \int_Q u_{z_1} v_{z_2}\, dz_1\, dz_2 \Big\|_{L^p}^{1-\theta} \Big\| \sum_{Q \in \mathcal{Q}_N} \int_Q u_{z_1} v_{z_2}\, dz_1\, dz_2 \Big\|_{L^2}^{\theta}$$
$$\lesssim N^{-\delta(1-\theta)+\alpha\theta} \big[\big(\sup_z |\langle f, \psi_z| \big)^{\frac{1}{p'}} \big(\sup_z |\langle g, \psi_z| \big)^{\frac{1}{p'}} \big]^{\theta} (\|f\|_{L_x^2} \|g\|_{L_x^2})^{1-\theta+\frac{\theta}{p}}$$

and for q sufficiently close to p (hence θ sufficiently small) the exponent of N is negative.

For the "near-diagonal" sum, we interpolate between L^1 and L^2. For the L^1 bound we simply use Minkowski's inequality and the estimate $\|U(t)\psi_{z_1} U(t)\psi_{z_2}\|_{L^1} \lesssim_N \langle x_1 - x_2 \rangle^{-N}$ when $|\xi_1 - \xi_2| \le 1$ to obtain

$$\Big\| \int u_{z_1} v_{z_2} \chi_0(\xi_1, \xi_2)\, dx_1\, dx_2\, d\xi_1\, d\xi_2 \Big\|_{L_x^1} \lesssim \int |f_{z_1} g_{z_2}| \langle x_1 - x_2 \rangle^{-N} \chi_0(\xi_1, \xi_2)\, dz_1\, dz_2$$
$$\lesssim \|f_z\|_{L_z^2} \|g_z\|_{L_z^2},$$

which when combined with Proposition 4.1 yields

$$\Big\| \int u_{z_1} v_{z_2} \chi_0(\xi_1, \xi_2)\, dz_1\, dz_2 \Big\|_{L^q} \lesssim \Big\| \int u_{z_1} v_{z_2} \chi_0(\xi_1, \xi_2)\, dz_1\, dz_2 \Big\|_{L^1}^{1-\theta'} \Big\| \int u_{z_1} v_{z_2} \chi_0(\xi_1, \xi_2)\, dz_1\, dz_2 \Big\|_{L^2}^{\theta'}$$
$$\lesssim \big[(\sup_z |f_z|)^{\frac{1}{p'}} (\sup_z |g_z|)^{\frac{1}{p'}} \big]^{\theta'} (\|f_z\|_{L_z^2} \|g_z\|_{L_z^2})^{1-\theta'+\frac{\theta'}{p}}$$
$$\lesssim (\sup_z |\langle f, \psi_z \rangle| \sup_z |\langle g, \psi_z \rangle|)^{\frac{\theta}{p'}} (\|f\|_{L^2} \|g\|_{L^2})^{1-\theta+\frac{\theta}{p}}$$

for some $1 < p < 2$, where $\frac{1}{q} = 1 - \theta' + \frac{\theta'}{2}$.

Summing in N, we conclude that

$$\|uv\|_{L^q} \lesssim \big[(\sup_z |\langle f, \psi_z|)^{\frac{1}{p'}} (\sup_z |\langle g, \psi_z \rangle|)^{\frac{1}{p'}} \big]^{\theta} (\|f\|_{L_x^2} \|g\|_{L_x^2})^{1-\frac{\theta}{p'}}$$

for some $\theta = \theta(p) \in \big(1, \frac{d+2}{d}\big)$. Taking $u = v$ we obtain Proposition 3.2.

6. The restriction-type estimate

This purpose of this section is to prove Theorem 1.5.

We shall systematically use the following notation. For $N \ge 1$ and a potential V, we consider the rescaled potentials

$$V_N(t, x) := N^{-2} V(N^{-2} t, N^{-1} x).$$

Let $U(t, s)$ and $U_N(t, s)$ denote the propagators for the corresponding Schrödinger operators $H(t) := -\frac{1}{2}\Delta + V$ and $H_N(t) := -\frac{1}{2}\Delta + V_N$. We will often use the letter U to write the propagators for different potentials $V \in \mathcal{V}$; this ambiguity will not cause any serious issue, however, since all the estimates we shall need are valid uniformly over $\mathcal{V}$. Further, due to the time-translation invariance of our assumptions we shall usually just consider the propagator from time 0 and write $U(t) := U(t,0)$, $U_N(t) := U_N(t,0)$.

In the sequel, the letter C will denote a constant, depending only on the dimension d, which may change from line to line.

6A. ***Preliminary reductions.*** The hypotheses of Theorem 1.5 are invariant under various transformations of u and v:

- Galilei boosts $u(0) \mapsto \pi(z_0)u(0)$, $u \mapsto \pi(z_0^t)u^{z_0}$, where u^{z_0} satisfies $(D_t - \Delta + V^{z_0})u^{z_0} = 0$, $u^{z_0}(0) = u(0)$.
- Spatial rotations: for an orthogonal matrix g, $(g \cdot u)(t,x) := u(t, g^{-1} \cdot x)$ satisfies

$$[D_t(g \cdot u) - \Delta + (g \cdot V)](g \cdot u) = 0.$$

- Rescaling $u \mapsto u_\lambda = \lambda^{-d/2} u(\lambda^{-2}t, \lambda^{-1}x)$ for $\lambda > 1$. Then u_λ satisfies $(D_t - \Delta + V_\lambda)u_\lambda = 0$ with a smoother potential $V_\lambda(t,x) = \lambda^{-2}V(\lambda^{-2}t, \lambda^{-1}x)$.

We may and shall assume hereafter that V vanishes to second order at $x = 0$; that is, $V(t,0) = 0$ and $\partial_x V(t,0) = 0$ for all t. Indeed let $z_0^t = (x_0^t, \xi_0^t)$ be the bicharacteristic with $(x_0, \xi_0) = (0,0)$. Then by Lemma 1.1,

$$\begin{aligned} \|U(t)fU(t)g\|_{L^{(d+3)/(d+1)}} &= \|(\pi(z_0^t)U^{z_0}(t)f)(\pi(z_0^t)U^{z_0}(t)g)\|_{L^{(d+3)/(d+1)}} \\ &= \|U^{z_0}(t)fU^{z_0}(t)g\|_{L^{(d+3)/(d+1)}}, \end{aligned}$$

and the potential $V^{z_0}(t,x) = V(t, x_0^t + x) - V(t, x_0^t) - x\partial_x V(t, x_0^t)$ vanishes to second order at $x = 0$.

Theorem 1.5 is equivalent by rescaling to:

Theorem 6.1. *Given $S_1, S_2 \subset \mathbb{R}^d_\xi$ with* $\operatorname{diam}(S_j) \le 1$ *and* $c^{-1} \ge \operatorname{dist}(S_1, S_2) \ge c$ *for some* $0 < c < 1$, *there exists a constant $\eta = \eta(c) > 0$ such that if $V \in \mathcal{V}$ and $\tau_0 > 0$ satisfies*

$$(\tau_0 + \tau_0^2)\|\partial_x^2 V\|_{L^\infty_{t,x}} < \eta, \tag{27}$$

then, for any $f, g \in L^2(\mathbb{R}^d)$ with $\operatorname{supp}(\hat{f}) \subset S_1$ *and* $\operatorname{supp}(\hat{g}) \subset S_2$, *the corresponding Schrödinger solutions $u_N = U_N(t)f$ and $v_N = U_N(t)g$ satisfy the estimate*

$$\|u_N v_N\|_{L^q([-\tau_0 N^2, \tau_0 N^2] \times \mathbb{R}^d)} \lesssim_\varepsilon N^\varepsilon \|f\|_{L^2} \|g\|_{L^2} \quad \textit{for all } \frac{d+3}{d+1} \le q < \frac{d+2}{d}, \tag{28}$$

for any $\varepsilon > 0$ and $N \ge 1$.

In fact it suffices to take S_1 and S_2 of the form

$$S_1 = \left\{ \xi : \left| \xi - \frac{c}{2}e_1 \right| \le \frac{c}{100} \right\}, \quad S_2 = \left\{ \xi : \left| \xi + \frac{c}{2}e_1 \right| \le \frac{c}{100} \right\}. \tag{29}$$

General S_j can be reduced to this case by decomposing $\hat{f} = \sum_j \hat{f}_j$ and $\hat{g} = \sum_k \hat{g}_k$ into pieces supported in small balls and applying an appropriate Galilei boost and rotation for each pair (f_j, g_k) and possibly also a rescaling to bring the Fourier supports closer, which only reduces $\|\partial_x^2 V\|_{L^\infty}$. Henceforth we shall assume (29).

6B. ***General remarks.*** We use the induction-on-scales method pioneered in [Wolff 2001] for the cone and adapted in [Tao 2003] to the paraboloid. Our proof is modeled closely on Tao's treatment of the $V = 0$ case, and the reader may find it helpful to read the following exposition in parallel with [Tao 2003]. The main differences are as follows:

- The induction scheme (Section 6E) is complicated by the fact that frequency is not conserved, so one cannot directly apply an induction hypothesis which involves assumptions on the frequency supports at time 0 to a space-time ball at a later time.
- The low regularity of V in time makes the bilinear L^2 estimate (Section 6H) more delicate and we obtain weaker decay from temporal oscillations.
- In the final Kakeya-type estimate, the tubes in the key combinatorial lemma (Lemma 6.11, the analogue of Lemma 8.1 in Tao) are curved. Also, we need to be slightly more precise to compensate for the weaker decay in the L^2 bound.

6C. ***Discrete wavepacket decomposition.*** While the first part of this paper employed continuous wavepacket transforms, the following discrete decomposition, taken essentially from [Tao 2003], is more conventional in restriction theory and convenient for the combinatorial arguments involved. To each $z_0 = (x_0, \xi_0)$ in classical phase space with bicharacteristic $\gamma_{z_0}(t) = (x_0^t, \xi_0^t)$, we associate a space-time "tube"

$$T_{z_0} := \{(t, x) : |x - x_0^t| \le R^{\frac{1}{2}},\ |t| \le R\}.$$

For such a tube T, let $z(T) = (x(T), \xi(T))$ denote the corresponding initial point in phase space. A wavepacket ϕ associated to the bicharacteristic $z_0 \mapsto z_0^t$ is essentially supported in space-time on the tube T_{z_0}, and we shall often emphasize this fact by writing ϕ_T.

Lemma 6.2. *Let $u = U_N(t) f$ be a linear Schrödinger solution with* $\operatorname{supp}(\hat{f}) \subset S_1$. *For each $1 \le R \le N^2$, there exists a collection of tubes $\boldsymbol{T}$ and a decomposition*

$$u = \sum_{T \in \boldsymbol{T}} a_T \phi_T$$

into $R \times (R^{1/2})^d$ wave packets with the following properties:

- *Each $T \in \boldsymbol{T}$ satisfies $(x(T), \xi(T)) \in R^{1/2}\mathbb{Z}^d \times R^{-1/2}\mathbb{Z}^d$.*
- *Each wavepacket ϕ_T is a Schrödinger solution localized near the bicharacteristic $(x(T)^t, \xi(T)^t)$, i.e., it satisfies the pointwise bounds*

$$\begin{aligned} |(R^{\frac{1}{2}}\partial_x)^k \phi_T(t)| &\lesssim_{k,M} \left\langle \frac{x - x(T)^t}{R^{\frac{1}{2}}} \right\rangle^{-M} && \text{for all } k, M \ge 0, \\ |(R^{-\frac{1}{2}}\partial_\xi)^k \hat{\phi}_T(t)| &\lesssim_{k,M} \left\langle \frac{\xi - \xi(T)^t}{R^{-\frac{1}{2}}} \right\rangle^{-M} && \text{for all } k, M \ge 0. \end{aligned} \tag{30}$$

Moreover, $\hat{\phi}_T[0]$ is supported in an $R^{-1/2}$ neighborhood of $\xi(T) \in S_1$.

- *The complex coefficients a_T are square-summable*:

$$\sum_T |a_T|^2 \lesssim \|f\|_{L^2}^2.$$

Moreover, for any subcollection of tubes $\boldsymbol{T}' \subset \boldsymbol{T}$ and complex numbers a_T, one has

$$\Big\| \sum_{T \in \boldsymbol{T}'} a_T \phi_T \Big\|_{L^2}^2 \lesssim \sum_{T \in \boldsymbol{T}'} |a_T|^2.$$

A similar decomposition also holds for $v = U_N(t)g$.

Proof sketch. We outline the main steps as this construction is fairly standard; consult for instance [Tao 2003, Lemma 4.1]. Begin with partitions of unity $1 = \sum_{x_0 \in \mathbb{Z}^d} \eta(x - x_0)$ and $1 = \sum_{\xi_0 \in \mathbb{Z}^d} \chi(\xi - \xi_0)$ such that χ and $\hat{\eta}$ are compactly supported. By rescaling and quantizing, we obtain a pseudodifferential partition of unity used to decompose the initial data

$$f = \sum_{(x_0, \xi_0)} \eta\Big(\frac{x - x_0}{R^{\frac{1}{2}}} \Big) \chi(R^{\frac{1}{2}}(D - \xi_0)) f.$$

The propagation estimates then follow from the next lemma. □

Lemma 6.3. *If ϕ_{z_0} is a scale-R wavepacket concentrated at z_0, and $U_N(t)$ is the propagator for $H(t) = -\frac{1}{2}\Delta + V_N$, then $U_N(t)$ is a scale-R wavepacket concentrated at z_0^t for all $|t| \le R$.*

Proof. By rescaling we reduce to $R = 1$ and replace V by $V_{N/R^{1/2}}$ which also belongs to $\mathcal{V}$ since $N/R^{1/2} \ge 1$. Then the symbol $a = \frac{1}{2}|\xi|^2 + V_{N/R^{1/2}}(t, x)$ satisfies the estimates (7), and we can appeal to Lemma 2.4. □

6D. *Localization.* The proof of Theorem 6.1 begins with the observation that it suffices to establish the same estimate with the space-time norm restricted to a box of the form

$$\Omega_N = [-N^2, N^2] \times [-AN^2, AN^2]^d.$$

Theorem 6.4. *Assume the hypotheses and notation of Theorem 6.1 and replace c by $c/2$ and take* $\operatorname{diam}(S_j) \le \frac{11}{10}$. *Then there exists $A = A(c) > 0$ such that*

$$\|u_N v_N\|_{L^{(d+3)/(d+1)}(\Omega_N)} \lesssim_\varepsilon N^\varepsilon \|f\|_{L^2} \|g\|_{L^2} \tag{31}$$

for any $\varepsilon > 0$.

Remark. In the wavepacket decomposition of u_N and v_N, the Fourier supports of the wavepackets are contained in a slight dilate $S_j + B(0, CN^{-1})$ of S_j. Hence at various junctures we need to adjust various constants to accommodate this minor enlargement of Fourier supports.

The full theorem then follows from an approximate finite speed of propagation argument:

Lemma 6.5. *Theorem 6.4 implies Theorem 6.1.*

Proof of Lemma 6.5. Partition physical space $\mathbb{R}^d = \bigcup_{j \in \mathbb{Z}^d} Q_j$ into cubes of width $\sim N^2$, where Q_j denotes the cube with center $N^2 j \in N^2\mathbb{Z}^d$. Decompose $u := u_N$ and $v := v_N$ into $N^2 \times (N)^d$ wavepackets, and group the terms in the product according to their relative initial positions. Write

$$u = \sum_T a_T \phi_T = \sum_{j \in \mathbb{Z}^d} \sum_{T \in \boldsymbol{T}_j} u_T,$$
$$v = \sum_{T'} b_{T'} \phi_{T'} = \sum_{j' \in \mathbb{Z}^d} \sum_{T' \in \boldsymbol{T}'_{j'}} v_{T'},$$

where $\boldsymbol{T}_j = \{T \in \boldsymbol{T} : x(T) \in Q_j\}$ and similarly for $\boldsymbol{T}_{j'}$. Using the triangle inequality we estimate

$$\|uv\|_{L^{(d+3)/(d+1)}} \le \sum_{k \ge 0} \Big\| \sum_{|j-j'| \sim 2^k} \sum_{T \in \boldsymbol{T}_j,\, T' \in \boldsymbol{T}'_{j'}} u_T v_{T'} \Big\|_{L^{(d+3)/(d+1)}}. \tag{32}$$

For the k-th sum, note from (19) that if $(x_1, \xi_1) := (x(T), \xi(T))$ and $(x_2, \xi_2) := (x(T'), \xi(T'))$, we have

$$\begin{aligned} |x_1^t - x_2^t| &\ge (1 - Ct^2 \|\partial_x^2 V_N\|_{L^\infty})|x_1 - x_2| - (|t| + C|t|^3 \|\partial_x^2 V_N\|_{L^\infty})|\xi_1 - \xi_2| \\ &\ge (1 - C\tau_0^2 \|\partial_x^2 V\|_{L^\infty})|x_1 - x_2| - N^2(1 + C\tau_0^2 \|\partial_x^2 V\|_{L^\infty})|\xi_1 - \xi_2| \\ &\ge (1 - C\eta)|x_1 - x_2| - N^2(1 + C\eta)|\xi_1 - \xi_2|, \end{aligned}$$

where C hides the harmless Gronwall factor. As $|\xi_1 - \xi_2| \le c^{-1}$, there exists $k(c)$ such that if $|x_1 - x_2| \ge 2^k N^2$ and η is chosen small enough we obtain $|x_1^t - x_2^t| \gtrsim 2^k N^2$ for $k \ge k(c)$. Thus the tubes in $\boldsymbol{T}_j$ and $\boldsymbol{T}'_j$ are separated in space by distance $\gtrsim 2^k N^2$, and since each wavepacket ϕ_T decays rapidly away from its tube T in units of N, we have

$$\|\phi_T \phi_{T'}\|_{L^{(d+3)/(d+1)}} \lesssim 2^{-101dk} N^{-101d},$$

and estimate crudely as follows:

$$\begin{aligned} \Big\| \sum_{|j-j'| \sim 2^k} \sum_{T \in \boldsymbol{T}_j, T' \in \boldsymbol{T}'_j} u_T v_{T'} \Big\|_{L^{(d+3)/(d+1)}} &\lesssim 2^{-101dk} N^{-101d} \sum_{|j-j'| \sim 2^k} \sum_{T \in \boldsymbol{T}_j, T' \in \boldsymbol{T}'_{j'}} |a_T b_{T'}| \\ &\lesssim 2^{-101dk} N^{-100d} \sum_{|j-j'| \sim 2^k} \Big(\sum_{T \in \boldsymbol{T}_j} |a_T|^2 \Big)^{\frac12} \Big(\sum_{T' \in \boldsymbol{T}'_{j'}} |b_{T'}|^2 \Big)^{\frac12} \\ &\lesssim 2^{-100dk} N^{-100d} \Big(\sum_j \sum_{T \in \boldsymbol{T}_j} |a_T|^2 \Big)^{\frac12} \Big(\sum_j \sum_{T \in \boldsymbol{T}'_{j'}} |b_{T'}|^2 \Big)^{\frac12} \\ &\lesssim 2^{-100dk} N^{-100d} \|f\|_{L^2} \|g\|_{L^2}. \end{aligned}$$

For the "near diagonal" part of the sum (32), where $|j - j'| \le 2^{k(c)}$, we group the terms by their average initial positions:

$$\begin{aligned} &\Big\| \sum_{|j-j'| \lesssim 1} \sum_{T \in \boldsymbol{T}_j, T' \in \boldsymbol{T}'_{j'}} u_T v_{T'} \Big\|_{L^{(d+3)/(d+1)}} \\ &\qquad \le \sum_{m \in \mathbb{Z}^d + \mathbb{Z}^d} \sum_{|j-j'| \lesssim 1, j+j'=m} \Big\| \sum_{T \in \boldsymbol{T}_j,\, T' \in \boldsymbol{T}'_{j'}} u_T v_{T'} \Big\|_{L^{(d+3)/(d+1)}}. \end{aligned} \tag{33}$$

For each pair (j, j'), we translate the initial data by the midpoint

$$x_{jj'} := \frac{j+j'}{2} N^2$$

of Q_j and $Q_{j'}$, using Lemma 1.1 to write

$$u_T = \pi(z^t_{jj'}) a_T \tilde{\phi}_T =: \tilde{u}_T, \quad v_T = b_{T'} \pi(z^t_{jj'}) \tilde{\phi}_{T'} =: \tilde{v}_T,$$

where $z_{jj'} = (x_{jj'}, 0)$ and

$$\tilde{\phi}_T(t) = U^{(x_{jj'},0)}(t)\pi(-x_{jj'}, 0)\phi_T[0]$$

is a wavepacket solution for the modified potential $V^{(x_{jj'},0)}$. The norm on the right side above therefore can be written as

$$\Big\| \sum_{T \in \tilde{\boldsymbol{T}}_j, T' \in \tilde{\boldsymbol{T}}'_j} \tilde{u}_T \tilde{v}_{T'} \Big\|_{L^{(d+3)/(d+1)}},$$

where the initial positions $x(T)$ and $x(T')$ of the tubes now belong to the translated cubes $\widetilde{Q}_j := Q_j - x_{jj'}$, $\widetilde{Q}_{j'} - x_{jj'}$, which are now distance $\lesssim N^2$ from the origin (note however that the tubes in $\tilde{\boldsymbol{T}}_j$ are not simply translates of those in $\boldsymbol{T}_j$).

By simple bicharacteristic estimates and the wavepacket bounds (30), for large A the norm outside

$$\Omega_N := [-N^2, N^2] \times [-AN^2, AN^2]^d$$

is negligible:

$$\begin{aligned} \Big\| \sum_{T \in \tilde{\boldsymbol{T}}_j, T' \in \tilde{\boldsymbol{T}}'_j} \tilde{u}_T \tilde{v}_{T'} \Big\|_{L^{(d+3)/(d+1)}([-N^2,N^2]\times([-AN^2,AN^2]^c))} &\lesssim N^{-100d} \Big(\sum_{T \in \tilde{\boldsymbol{T}}_j} |a_T|^2 \Big)^{\frac{1}{2}} \Big(\sum_{T' \in \tilde{\boldsymbol{T}}'_j} |b_T|^2 \Big)^{\frac{1}{2}} \\ &\lesssim N^{-100d} \Big(\sum_{T \in \boldsymbol{T}_j} |a_T|^2 \Big)^{\frac{1}{2}} \Big(\sum_{T' \in \boldsymbol{T}'_j} |b_T|^2 \Big)^{\frac{1}{2}}. \end{aligned}$$

Inside Ω_N we invoke 6.4 using the fact that the $V^{(x_{jj'},0)}$ also satisfies the hypothesis (27) and that the wavepacket decompositions of u_N and v_N satisfy the relaxed Fourier support conditions in that proposition. Altogether, the right side of (33) is bounded by

$$\begin{aligned} &\sum_{m \in \mathbb{Z}^d + \mathbb{Z}^d} \ \sum_{|j-j'| \lesssim 1,\, j+j'=m} N^{\varepsilon} \Big(\sum_{T \in \boldsymbol{T}_j} |a_T|^2 \Big)^{\frac{1}{2}} \Big(\sum_{T' \in \boldsymbol{T}'_j} |b_T|^2 \Big)^{\frac{1}{2}} \\ &\qquad \lesssim N^{\varepsilon} \sum_m \Big(\sum_{|j-\frac{m}{2}| \lesssim 1} \sum_{T \in \boldsymbol{T}_j} |a_T|^2 \Big)^{\frac{1}{2}} \Big(\sum_{|j'-\frac{m}{2}| \lesssim 1} \sum_{T \in \boldsymbol{T}'_{j'}} |b_{T'}|^2 \Big)^{\frac{1}{2}} \\ &\qquad \lesssim N^{\varepsilon} \Big(\sum_T |a_T|^2 \Big)^{\frac{1}{2}} \Big(\sum_{T'} |b_{T'}|^2 \Big)^{\frac{1}{2}} \\ &\qquad \lesssim N^{\varepsilon} \|f\|_{L^2} \|g\|_{L^2}, \end{aligned}$$

thus recovering Theorem 6.1. □

6E. *Induction on scales.* Our induction scheme is set up slightly differently from Tao's to accommodate the nonconservation of frequency support of solutions.

In this section, we explicitly display the dependence of the propagator on the potential, and write $U_N^V(t) = U_N^V(t,0)$ for the propagator with potential V_N.

Let IH(α) denote the following statement:

There exists $C_\alpha > 0$ such that for each $N \geq 1$ and for all potentials $V \in \mathcal{V}_\eta$, the estimate

$$\|U_N^V(t) f U_N^V(t) g\|_{L^{(d+3)/(d+1)}(\Omega_N)} \leq C_\alpha N^{2\alpha} \|f\|_{L^2} \|g\|_{L^2} \tag{34}$$

holds for all $f, g \in L^2(\mathbb{R}^d)$ with $\hat{f}, \hat{g}$ supported in S_1 and S_2, respectively.

We prove:

<u>Inductive Step</u>: If IH(α) holds, then IH($\max((1-\delta)\alpha, C\delta) + \varepsilon$) holds for all $0 < \delta, \varepsilon \ll 1$.

By choosing δ and ε sufficiently small depending on α, we can always arrange that

$$\max((1-\delta)\alpha, C\delta) + C\varepsilon < \alpha - c\alpha^2$$

for some absolute constant c, and Theorem 6.4 follows.

The inductive hypothesis IH(α) shall be used to improve the estimate (34) over subregions $Q_R \subset \Omega_N$ at smaller scales $\operatorname{diam}(Q_R) \sim N^{2(1-\delta)} \ll N^2$.

Proposition 6.6. *Suppose* IH(α) *holds. Then for all* $1 \leq R \leq \frac{1}{16} N^2$ *and all space-time balls* $Q_R \subset 2\Omega_N$ *of diameter* R, *the estimate*

$$\|U_N^V(t) f U_N^V(t) g\|_{L^{(d+3)/(d+1)}(Q_R)} \leq C_\alpha R^\alpha \|f\|_{L^2} \|g\|_{L^2}$$

holds for all $f, g \in L^2(\mathbb{R}^d)$ *with* $\hat{f}, \hat{g}$ *supported in* $\widetilde{S}_1 := S_1 + B\big(0, \frac{c}{100}\big)$ *and* $\widetilde{S}_2 := S_2 + B\big(0, \frac{c}{100}\big)$, *respectively.*

Proof. We begin by estimating how much the Fourier supports can shift.

Lemma 6.7. *For* $1 \leq R \leq N^2$, *let* $Q_R \subset 2\Omega_N$ *be a space-time ball with center* (t_Q, x_Q) *and diameter* R. *Suppose the initial data* f, g *satisfy* $\operatorname{supp}(\hat{f}) \subset \widetilde{S}_1$ *and* $\operatorname{supp}(\hat{g}) \subset \widetilde{S}_2$. *There exist decompositions* $u(t_Q) = f_1 + f_2$ *and* $v(t_Q) = g_1 + g_2$, *with the following properties*:

- $\hat{f}_1$ *and* $\hat{g}_1$ *are supported in sets* S_1', S_2' *with* $\operatorname{diam}(S_j') \leq \frac{c}{10}$ *and* $\operatorname{dist}(S_1', S_2') \in \big[\frac{4c}{5}, \frac{5c}{4}\big]$.
- $\|f_2\|_{L^2} \lesssim N^{-100d} \|f\|_{L^2}$ *and* $\|g_2\|_{L^2} \lesssim N^{-100d} \|g\|_{L^2}$.

Proof. Begin by decomposing $u = U_N^V f$ and $v = U_N^V g$ into $N^2 \times (N)^d$ wavepackets:

$$u = \sum_{T \in \boldsymbol{T}_1} a_T \phi_T, \quad v = \sum_{T \in \boldsymbol{T}_2} b_T \phi_T. \tag{35}$$

By the spatial localization (30), we may ignore in u and v the packets whose tubes $T \in \boldsymbol{T}_j$ do not intersect $2Q_N := [-N^2, N^2] \times [-2AN^2, 2AN^2]$, as the portion of the sum involving those terms contributes at most $O(N^{-100d}) \|f\|_{L^2} \|g\|_{L^2}$. Thus there are $O(N^{2d})$ remaining terms.

Suppose ϕ_{T_1} and ϕ_{T_2} are wavepackets in the decomposition for u.

Let (x_1^t, ξ_1^t) and (x_2^t, ξ_2^t) be bicharacteristics with $|x_1|, |x_2| \le 2AN^2$. By (19), for $|t| \le \tau_0 N^2$ we have

$$|\xi_1^t - \xi_2^t - (\xi_1 - \xi_2)| \le C\tau_0 N^2 N^{-4} \|\partial_x^2 V\|_{L^\infty}(2AN^2 + \tau_0 N^2 |\xi_1 - \xi_2|) \\ \le C(\tau_0 A + \tau_0^2)\|\partial_x^2 V\|_{L^\infty} \le C\eta.$$

Therefore, recalling the definitions of $\widetilde{S}_j$, we see that we have $|\xi_1^{t_Q} - \xi_2^{t_Q}| \le \frac{c}{20} + C\eta$ if ξ_1, ξ_2 both belong to $\widetilde{S}_1$ or $\widetilde{S}_2$, while $|\xi_1^{t_Q} - \xi_2^{t_Q}| \in \left[\frac{9c}{10}, \frac{10c}{9}\right]$ if $\xi_1 \in \widetilde{S}_1$ and $\xi_2 \in \widetilde{S}_2$. Choose $\eta = \eta(c)$ sufficiently small.

Consequently, if

$$\widetilde{S}_j^t := \{\xi^t : \xi \in \widetilde{S}_j,\ |x| \le AN^2\} \tag{36}$$

denotes the set of frequencies of the wavepackets at time t, then $\operatorname{diam}(\widetilde{S}_j^t) \le \operatorname{diam}(S_j) + C\eta$ and $\operatorname{dist}(\widetilde{S}_1^t, \widetilde{S}_2^t) \ge \frac{9}{10}\operatorname{dist}(S_1, S_2)$. Now let S_j' denote $O(N^{-9/10})$ neighborhoods of $\widetilde{S}_j^t$, and take the decompositions

$$u(t_Q) = f_1 + f_2, \quad v(t_Q) = g_1 + g_2,$$

where $\hat{f}_1$ is supported on $\widetilde{S}_1$ and $\hat{f}_2$ on the complement, and similarly for g_1, g_2. For N large enough we have $\operatorname{dist}(S_1' S_2') \in \left[\frac{4c}{5}, \frac{5c}{4}\right]$. The estimates in the second bullet point now follow from the rapid decay of each wavepacket from its central frequency on the N^{-1} scale (the estimates (30) with $R = N^2$). $\square$

The proof of the proposition concludes with several applications of Lemma 1.1. Write

$$U(t, t_Q) f_1 = U(t, t_Q)\pi(x_Q, 0)\pi(-x_Q, 0) f_1 = \pi(z_Q^t) U^{z_Q}(t, t_Q)\tilde{f}_1 = \pi(z_Q^t)\tilde{u}(t + t_q),$$

where $z_Q^{t_Q} = (x_Q, 0)$. For $|t - t_Q| \le R$ and $|x_Q| \le AN^2$ we have $|x_Q^t - x_Q^{t_Q}| \le 2|t - t_Q| \le 2R$ provided that η is sufficiently small. Therefore, letting $\widetilde{Q}_R = 2(Q_R - (t_Q, x_Q))$,

$$\|uv\|_{L^{(d+3)/(d+1)}(Q_R)} \lesssim \|\tilde{u}\tilde{v}\|_{L^{(d+3)/(d+1)}(\widetilde{Q}_R)} + N^{-100d}\|f\|_{L^2}\|g\|_{L^2}.$$

It remains to consider the first term on the right side. The initial data $\tilde{f}_1$, $\tilde{g}_1$ for $\tilde{u}$ and $\tilde{v}$ have Fourier transforms supported in S_1', S_2'. We abuse notation and redefine

$$f := \tilde{f}_1, \quad g := \tilde{g}_1.$$

Cover $S_j' = \bigcup_k B_{j,k}$ by finitely overlapping balls of radius $\frac{c}{200}$. Using a subordinate partition of unity, we reduce to the case where $\operatorname{supp}\hat{f} \subset B_{1,k_1}$ and $\operatorname{supp}\hat{g} \subset B_{2,k_2}$. Again using Lemma 1.1, we may assume $B_{1,k_1} = -B_{2,k_2}$ and that their centers lie on the e_1-axis.

Since $2c \ge \operatorname{dist}(B_{1,k_1}, B_{2,k_2}) \ge \frac{c}{2}$, there exists some scaling factor $\lambda \in \left[\frac{1}{2}, 2\right]$ such that $\lambda^{-1} B_{j,k_j} \subset S_j$. Consider the rescalings

$$u_\lambda = U^V_{\frac{N}{\lambda}}(t) f_\lambda = U^{\widetilde{V}}_{(2R)^{1/2}}(t) f_\lambda, \quad v_\lambda = U^V_{\frac{N}{\lambda}}(t) g_\lambda = U^{\widetilde{V}}_{(2R)^{1/2}}(t) g_\lambda,$$

where

$$\widetilde{V}(t, x) = 2R\lambda^2 N^{-2} V(2R\lambda^2 N^{-2} t, (2R)^{\frac{1}{2}}\lambda N^{-1} x).$$

The potential $\widetilde{V}$ satisfies $\|\partial_x^2 \widetilde{V}\|_{L^\infty} \le \|\partial_x^2 V\|_{L^\infty}$ since $2R\lambda^2 N^{-2} \le 8RN^{-2} \le \frac{1}{2}$, and $\hat{u}_\lambda(0)$ and $\hat{v}_\lambda(0)$ are supported in S_1 and S_2. Hence we can apply IH(α) to conclude that

$$\|\tilde{u}\tilde{v}\|_{L^{(d+3)/(d+1)}(\widetilde{Q}_R)} \lesssim \|u_\lambda v_\lambda\|_{L^{(d+3)/(d+1)}(\widetilde{Q}_{2R})} \le C_\alpha R^\alpha \|f_\lambda\|_{L^2} \|g_\lambda\|_{L^2}. \qquad \square$$

From here on the argument hews closely to Tao's. We recall the following notation: write

$$A \lessapprox B$$

if $A \lesssim_\varepsilon N^\varepsilon B$ for all $N \gg 1$ and for all $\varepsilon > 0$.

To reiterate, we want to prove

$$\|U_N^V f U_N^V g\|_{L^{(d+3)/(d+1)}(\Omega_N)} \lessapprox N^{2\max((1-\delta)\alpha, C\delta)} \|f\|_{L^2} \|g\|_{L^2}, \tag{37}$$

assuming $\operatorname{supp}(\hat{f}) \subset S_1$ and $\operatorname{supp}(\hat{g}) \subset S_2$ with $\operatorname{diam}(S_j) \le 1$ and $\operatorname{dist}(S_1, S_2) \ge c$.

Normalize f and g in L^2, and take the decomposition

$$u := U_N^V f = \sum_T a_T \phi_T, \quad v := U_N^V = \sum_T b_T \phi_T.$$

As in the proof of Lemma 6.7, we discard all but the $O(N^{2d})$ wavepackets whose tubes intersect $2\Omega_N$. We also throw away the terms where $|a_T| = O(N^{-100d})$ or $|b_T| = O(N^{-100d})$, as that portion of the product can be bounded using the estimates (30) and Cauchy–Schwarz.

Consequently, in the decompositions of u and v we only consider the tubes T with $N^{-100d} \lesssim |a_T|, |b_T| \lesssim 1$. Partitioning the interval $[N^{-100d}, 1]$ into $\log N$ dyadic groups, we may further restrict to the tubes with $|a_T| \sim \gamma_1$ and $|b_T| \sim \gamma_2$ for dyadic numbers $N^{-100d} \lesssim \gamma_1, \gamma_2 \lesssim 1$. Let $\boldsymbol{T}_1$, $\boldsymbol{T}_2$ be the tubes for u and v, respectively, with this property. It therefore suffices to prove

$$\Big\| \sum_{T_1 \in \boldsymbol{T}_1} \phi_{T_1} \sum_{T_2 \in \boldsymbol{T}_2} \phi_{T_2} \Big\|_{L^{(d+3)/(d+1)}(\Omega_N)} \lessapprox (N^{2(1-\delta)\alpha} + N^{2C\delta}) \# \boldsymbol{T}_1^{\frac{1}{2}} \# \boldsymbol{T}_2^{\frac{1}{2}}$$

(we have absorbed the complex phases into the wavepackets).

We have in effect reduced to considering the region of the phase space $\{(x,\xi) : |x| \lesssim N^2,\ |\xi| \lesssim 1\}$, where the potential makes only a small perturbation to the Euclidean flow. For if $|x^s| \lesssim N^2$ and $|t-s| \le N^2$, one has

$$|x^t| \lesssim N^2,$$
$$|\xi^t - \xi^s| \le \int_s^t |\partial_x(V_N)(\tau, x^\tau)|\, d\tau \lesssim \int_s^t |x^\tau| \int_0^1 |\partial_x^2 V_N(\tau, s x^\tau)|\, ds\, d\tau \lesssim \tau_0 \|\partial_x^2 V\|_{L^\infty} \lesssim \eta.$$

Thus if $\xi \in S_j$, then ξ^t belongs to a small neighborhood of S_j provided that $\eta \ll c$ is a small multiple of c. For concreteness we choose η so that

$$|\xi^t - \xi^s| \le \frac{c}{100}. \tag{38}$$

6F. ***Coarse-scale decomposition.*** Following Tao, for small $\delta > 0$ we decompose $\Omega_N = \bigcup_{B \in \mathcal{B}'} B$ into $O(N^{2\delta d})$ smaller balls of radius $N^{2(1-\delta)}$ and estimate

$$\Big\| \sum_{T_1 \in \boldsymbol{T}_1} \sum_{T_2 \in \boldsymbol{T}_2} \phi_{T_1} \phi_{T_2} \Big\|_{L^{(d+3)/(d+1)}(\Omega_N)} \lesssim \sum_{B \in \mathcal{B}} \Big\| \sum_{T_1 \in \boldsymbol{T}_1} \sum_{T_2 \in \boldsymbol{T}_2} \phi_{T_1} \phi_{T_2} \Big\|_{L^{(d+3)/(d+1)}(B)}.$$

Let $\sim$ be a relation between tubes and balls to be specified later. Estimate the norm by the local part

$$\sum_{B \in \mathcal{B}} \Big\| \sum_{T_1 \sim B} \phi_{T_1} \sum_{T_2 \sim B} \phi_{T_2} \Big\|_{L^{(d+3)/(d+1)}(B)} \tag{39}$$

and the global part

$$\sum_{B \in \mathcal{B}} \Big\| \sum_{T_1 \nsim B \text{ or } T_2 \nsim B} \phi_{T_1} \phi_{T_2} \Big\|_{L^{(d+3)/(d+1)}(B)}. \tag{40}$$

We use Proposition 6.6 with $R = N^{2(1-\delta)} \le \frac{1}{16} N^2$ to estimate the local term by

$$\begin{aligned} (39) &\lessapprox \sum_{B \in \mathcal{B}} N^{2(1-\delta)\alpha} \Big(\sum_{T_1 \sim B} 1 \Big)^{\frac{1}{2}} \Big(\sum_{T_2 \sim B} 1 \Big)^{\frac{1}{2}} \\ &\lessapprox \Big(\sum_{T_1 \in \boldsymbol{T}_1} \#\{B : T_1 \sim B\} \Big)^{\frac{1}{2}} \Big(\sum_{T_2 \in \boldsymbol{T}_2} \#\{B : T_2 \sim B\} \Big)^{\frac{1}{2}} \lessapprox 1 \end{aligned}$$

if the relation $\sim$ is chosen so that each T is associated to $\lessapprox 1$ balls. Note that this step is why the Fourier supports are enlarged in that proposition, as $\operatorname{supp}(\hat{\phi}_{T_1}(0))$ is not quite contained in S_1.

Heuristically, a judicious choice of $\sim$ allows one to avoid the worst interactions that would otherwise occur in the bilinear L^2 estimate if one were to natively interpolate between L^1 and L^2. For example, if all the tubes were to intersect in a single ball B, it would be better to bound $L^{(d+3)/(d+1)}(B)$ directly using the inductive hypothesis rather than attempt to estimate $L^2(B)$.

The global piece (40) is controlled by interpolating between L^1 and L^2. By Cauchy–Schwarz and conservation of L^2 norm,

$$\begin{aligned} &\sum_{B} \Big\| \sum_{T_1 \nsim B \text{ or } T_2 \nsim B} \phi_{T_1} \phi_{T_2} \Big\|_{L^1(B)} \\ &\qquad \lesssim \sum_{B} \Big(\Big\| \sum_{T_1 \sim B} \phi_T \Big\|_{L^2(B)} + \Big\| \sum_{T_1 \nsim B} \phi_T \Big\|_{L^2(B)} \Big) \Big(\Big\| \sum_{T_2 \sim B} \phi_T \Big\|_{L^2(B)} + \Big\| \sum_{T_2 \nsim B} \phi_T \Big\|_{L^2(B)} \Big) \\ &\qquad \lesssim N^{2\delta} N^2 \# \boldsymbol{T}_1^{\frac{1}{2}} \# \boldsymbol{T}_2^{\frac{1}{2}}. \end{aligned} \tag{41}$$

The remaining sections prove the L^2 estimate

$$\Big\| \sum_{T_1 \nsim B \text{ or } T_2 \nsim B} \phi_{T_1} \phi_{T_2} \Big\|_{L^2(B)} \lessapprox N^{-\frac{d-1}{2}} N^{C\delta} \# \boldsymbol{T}_1^{\frac{1}{2}} \boldsymbol{T}_2^{\frac{1}{2}}. \tag{42}$$

6G. *Fine scale decomposition.* Cover $\Omega_N = \bigcup_{q\in\boldsymbol{q}} q$ by a finitely overlapping collection $\boldsymbol{q}$ of balls of radius N. It suffices to show

$$\sum_{q\in\boldsymbol{q}:q\subset 2B} \Big\| \sum_{T_1\nsim B \text{ or } T_2\nsim B} \phi_{T_1}\phi_{T_2} \Big\|_{L^2(q)}^2 \lessapprox N^{-(d-1)} N^{C\delta} \#\boldsymbol{T}_1\boldsymbol{T}_2.$$

We adopt the following notation from Tao. Fix $q \in \boldsymbol{q}$ and let μ_1, μ_2, λ_1 be dyadic numbers:

- $\boldsymbol{T}_j(q)$ is the set of tubes $T \in \boldsymbol{T}_j$ such that $T \cap N^\delta q$ is nonempty, where $N^\delta q$ denotes an N^δ neighborhood of q.
- $\boldsymbol{T}_j^{\nsim B}(q) = \{T \in \boldsymbol{T}_j(q) : T \nsim B\}$.
- $\boldsymbol{q}(\mu_1, \mu_2)$ is the set of balls q such that $\#\{T_j \in \boldsymbol{T}_j : T_j \cap N^\delta q \neq \phi\} \sim \mu_j$.
- $\lambda(T, \mu_1, \mu_2)$ is the number of (N^δ neighborhoods of) balls $q \in \boldsymbol{q}(\mu_1, \mu_2)$ that T intersects.
- $\boldsymbol{T}_j[\lambda_1, \mu_1, \mu_2]$ is the set of tubes $T \in \boldsymbol{T}_j$ such that $\lambda(T, \mu_1, \mu_2) \sim \lambda_1$.

Pigeonholing dyadically in μ_1, μ_2, and λ_1, it suffices to show

$$\sum_{q\in\boldsymbol{q}(\mu_1,\mu_2):q\subset 2B} \Big\| \sum_{T_1\in\boldsymbol{T}_1^{\nsim B}(q)\cap\boldsymbol{T}_1[\lambda_1,\mu_1,\mu_2]} \; \sum_{T_2\in\boldsymbol{T}_2(q)} \phi_{T_1}\phi_{T_2} \Big\|_{L^2(q)}^2 \lessapprox N^{C\delta} N^{-(d-1)} \#\boldsymbol{T}_1\#\boldsymbol{T}_2.$$

6H. *The L^2 bound.* Fix a ball $q = q(t_q, x_q) \in \boldsymbol{q}(\mu_1, \mu_2)$ centered at (t_q, x_q). Suppose want to estimate an expression of the form

$$\Big\| \sum_{T_1} \sum_{T_2} \phi_{T_1}\phi_{T_2} \Big\|_{L^2(q)}^2.$$

There are two main points to keep in mind:

- Only tubes that intersect $N^\delta q$ will make a nontrivial contribution; that is, tubes whose bicharacteristics (x^t, ξ^t) satisfy $|x^{t_q} - x_q| \le N^{1+\delta}$.
- To decouple the contributions of tubes that all overlap near $\boldsymbol{q}$, one needs to exploit oscillation in space and time. While Tao employs the space-time Fourier transform, we instead integrate by parts in space and time. Expanding out the L^2 norm

$$\sum_{T_1,T_2} \sum_{T_3,T_4} \langle \phi_{T_1}\phi_{T_2}, \phi_{T_3}\phi_{T_4} \rangle \tag{43}$$

and integrating by parts in both space and time, we shall obtain terms of the form

$$(N|\xi_1^t + \xi_2^t - \xi_3^t - \xi_4^t|)^{-1}, \quad (N||\xi_1^t - \xi_2^t|^2 - |\xi_3^t - \xi_4^t|^2|)^{-1},$$

where (x_j^t, ξ_j^t) are bicharacteristics with $|x_j^{t_q} - x_q| \le N^{1+\delta}$. Since, by (19), the relative frequencies $\xi_j^t - \xi_k^t$ vary by at most $O(N^{-2+2\delta})$ during the $O(N^{1+\delta})$ time window when the wavepackets intersect the ball $N^\delta q$, we can freeze $t = t_q$ above; see Lemma 6.10 below.

Hence, the integral (43) will be small unless $|x_j^{t_q} - x_q| \le N^{1+\delta}$ for all j *and* the frequencies ξ_j^t satisfy both resonance conditions

$$|\xi_1^{t_q} + \xi_2^{t_q} - \xi_3^{t_q} - \xi_4^{t_q}| = O(N^{-1}), \quad |\xi_1^{t_q} - \xi_2^{t_q}|^2 - |\xi_3^{t_q} - \xi_4^{t_q}|^2 = O(N^{-1}). \tag{44}$$

The preceding discussion motivates the following definition. Let

$$Z_{q,j} := \{(x,\xi) : |x| \le 2AN^2,\ \xi \in S_j,\ |x^{t_q} - x_q| \le N^{1+\delta}\}.$$

For frequencies ξ_1 and ξ_2', define the "space-time resonance" set

$$Z(\xi_1, \xi_2') = \big\{(x_1', \xi_1') \in Z_{q,1} : \text{ there exists } (x_2, \xi_2) \in Z_{q,2} \text{ such that } \xi_1 + \xi_2^{t_q} = (\xi_1')^{t_q} + \xi_2' \text{ and } |\xi_1 - \xi_2^{t_q}|^2 = |(\xi_1')^{t_q} - \xi_2'|^2\big\},$$

$$\pi(\xi_1, \xi_2') = \{(\xi_1')^{t_q} : (x_1', \xi_1') \in Z(\xi_1, \xi_2')\}.$$

This is a slight modification of Tao's definition which reflects the time dependence of frequency.

The following lemma follows from elementary geometry.

Lemma 6.8. *The set $\pi(\xi_1, \xi_2')$ is contained in the hyperplane passing through ξ_1 and orthogonal to $\xi_2' - \xi_1$ and is therefore transverse to $\zeta_2' - \zeta_1$ if ζ_1 and ζ_2' are small perturbations of ξ_1 and ξ_2', respectively.*

Due to the limited time regularity of the phase, we can actually integrate by parts just once in time. The resulting weaker decay still turns out to be just enough provided that we slightly refine the analogue of Tao's main combinatorial estimate for tubes (estimate (48) below). Hence we need to account more carefully for the contributions away from the "resonant set" π.

For ξ_1, ξ_2' and $k > 0$, define the "time nonresonance" sets

$$Z_0^t(\xi_1, \xi_2') = \big\{(x_1', \xi_1') \in Z_{q,1} : \text{ there exists } (x_2, \xi_2) \in Z_{q,2} \text{ such that } |\xi_1 + \xi_2^{t_q} - (\xi_1')^{t_q} - \xi_2'| \le N^{-1+C\delta} \text{ and } \big||\xi_1 - \xi_2^{t_q}|^2 = |(\xi_1')^{t_q} - \xi_2'|^2\big| \le N^{-1+C\delta}\big\},$$

$$Z_k^t(\xi_1, \xi_2') = \big\{(x_1', \xi_1') \in Z_{q,1} : \text{ for all } (x_2, \xi_2) \in Z_{q,2} \text{ with } |\xi_1 + \xi_2^{t_q} - (\xi_1')^{t_q} - \xi_2'| \le N^{-1+C\delta}, \big||\xi_1 - \xi_2^{t_q}|^2 - |(\xi_1')^{t_q} - \xi_2'|^2\big| \in (2^{k-1}N^{-1+C\delta}, 2^k N^{-1+C\delta}]\big\},$$

the "space nonresonance" set

$$Z^s(\xi_1, \xi_2') = \{(x_1', \xi_1') \in Z_{q,1} : |\xi_1 + \xi_2^{t_q} - (\xi_1')^{t_q} + \xi_2'| > N^{-1+C\delta} \text{ for all } (x_2, \xi_2) \in B_{q,2}\},$$

and the corresponding frequencies at time t_q

$$\pi_k^t(\xi_1, \xi_2') = \{(\xi_1')^{t_q} : (x_1', \xi_1') \in Z_k^t(\xi_1, \xi_2')\},$$
$$\pi^s(\xi_1, \xi_2') = \{(\xi_1')^{t_q} : (x_1', \xi_1') \in Z^s(\xi_1, \xi_2')\}.$$

An elementary computation shows that

$$\operatorname{dist}(\pi_k^t, \pi) \lesssim 2^k N^{-1+C\delta}. \tag{45}$$

Indeed, writing $\delta_1 := (\xi_1')^{t_q} - \xi_1$, $\delta_2 := \xi_2^{t_q} - \xi_2'$, and decomposing $\delta_j = \delta_j^{\parallel} + \delta_j^{\perp}$ into the components parallel and orthogonal to $\xi_1 - \xi_2'$, we have

$$\begin{aligned}|\xi_1 - \xi_2^{t_q}|^2 - |(\xi_1')^{t_q} - \xi_2'|^2 &= |\xi_1 - \xi_2' - \delta_2|^2 - |\delta_1 + \xi_1 - \xi_2'|^2 \\ &= -2\langle \xi_1 - \xi_2', \delta_1 + \delta_2\rangle + \delta_2^2 - \delta_1^2 \\ &= -2\langle \xi_1 - \xi_2', \delta_1^{\parallel} + \delta_2^{\parallel}\rangle + O(N^{-1+C\delta}) \quad (\text{since } |\delta_1 - \delta_2| \le N^{1+\delta}) \\ &= -4\langle \xi_1 - \xi_2', \delta_1^{\parallel}\rangle + O(N^{-1+C\delta}).\end{aligned}$$

Thus $|\langle(\xi_1')^{t_q} - \xi_1,\ \xi_1 - \xi_2'\rangle| \lesssim 2^k N^{-1+C\delta}$ and the claim follows from Lemma 6.8.

For $q \in \boldsymbol{q}(\mu_1, \mu_2)$ with $q \subset 2B$, define

$$\boldsymbol{T}_1^{\nsim B}(q, \lambda_1, \mu_1, \mu_2, \xi_1, \xi_2', k)$$

to be the collection of tubes $T \in \boldsymbol{T}_1^{\nsim B}(q) \cap \boldsymbol{T}_1[\lambda_1, \mu_1, \mu_2]$ such that $\xi(T)^{t_q} \in \pi_k^t(\xi_1, \xi_2')$. Set

$$\nu_k(q, \lambda_1, \mu_1, \mu_2) := \sup_{\xi_1 \in S_1,\ \xi_2' \in S_2} \#\boldsymbol{T}_1^{\sim B}(q, \lambda_1, \mu_1, \mu_2, \xi_1^{t_q}, (\xi_2')^{t_q}, k), \tag{46}$$

where $|x_1^{t_q} - x_q| + |(x_2')^{t_q} - x_q| \lesssim N^{1+\delta}$.

Then, the analogue of Tao's Lemma 7.1 is:

Lemma 6.9. *For each $q \in \boldsymbol{q}(\mu_1, \mu_2)$, we have*

$$\begin{aligned}&\Big\| \sum_{T_1 \in \boldsymbol{T}_1^{\nsim B}(q) \cap \boldsymbol{T}_1[\lambda_1, \mu_1, \mu_2]} \ \sum_{T_2 \in \boldsymbol{T}_2(q)} \phi_{T_1}\phi_{T_2} \Big\|_{L^2(q)}^2 \\ &\qquad \lessapprox N^{C\delta} N^{-(d-1)} \sup_k 2^{-k} \nu_k(q, \lambda_1, \mu_1, \mu_2) \#(\boldsymbol{T}_1^{\nsim B}(q) \cap \boldsymbol{T}_1[\lambda_1, \mu_1, \mu_2]) \#\boldsymbol{T}_2(q).\end{aligned}$$

Proof. For conciseness, set

$$\begin{aligned}\boldsymbol{T}_1' &:= \boldsymbol{T}_1^{\nsim B}(q) \cap \boldsymbol{T}_1[\lambda_1, \mu_1, \mu_2], \\ \boldsymbol{T}_2 &:= \boldsymbol{T}_2(q).\end{aligned}$$

Then the norm $L^2(q)$ is bounded by the norm $L^2(\eta_N\, dx\, dt)$, where $\eta_N(t)$ is a smooth weight equal to 1 on $|t - t_q| \le N^{1+\delta}$ and supported in $|t - t_q| \le 2N^{1+\delta}$:

$$\Big\| \sum_{T_1 \in \boldsymbol{T}_1'} \sum_{T_2 \in \boldsymbol{T}_2'} \phi_{T_1}\phi_{T_2} \Big\|_{L^2(\eta_N\, dx dt)}^2 = \sum_{T_1, T_1' \in \boldsymbol{T}_1'} \sum_{T_2, T_2' \in \boldsymbol{T}_2'} \langle \phi_{T_1}\phi_{T_2}, \phi_{T_1'}\phi_{T_2'}\rangle_{L^2(\chi_N\, dx dt)}.$$

By the bounds (30) and the transversality of the tubes in $\boldsymbol{T}_1'$ and $\boldsymbol{T}_2'$, the integrand has magnitude N^{-2d} and is essentially supported on a space-time ball of width N. Thus we have the crude bound

$$|\langle \phi_{T_1}\phi_{T_2}, \phi_{T_1'}\phi_{T_2'}\rangle| \lesssim N^{C\delta} N^{-2d} N^{d+1} = N^{C\delta} N^{-(d-1)}.$$

On the other hand, we may integrate by parts to obtain a more refined bound.

Lemma 6.10. *For each $k_1, k_2, \ell \geq 0$ and for all tubes $T_1, T_3 \in \boldsymbol{T}_1'$, $T_2, T_4 \in \boldsymbol{T}_2'$, we have*

$$|\langle \phi_{T_1}\phi_{T_2}, \phi_{T_3}\phi_{T_4}\rangle| \lesssim_{k_1,k_2} N^{C\delta} N^{-(d-1)} \min\big[N^{-\ell}|\xi_1^{t_q}+\xi_2^{t_q}-\xi_3^{t_q}-\xi_4^{t_q}|^{-\ell}, N^{-1}\big||\xi_1^{t_q}-\xi_2^{t_q}|^2-|\xi_3^{t_q}-\xi_4^{t_q}|^2\big|^{-1}\big].$$

Proof. The proof has a similar flavor to the earlier estimate (23) but takes advantage of oscillation in both space and time.

Let $z_j^t = (x_j^t, \xi_j^t)$ denote the bicharacteristic for ϕ_{T_j}, $j = 1, 2, 3, 4$. By Lemmas 1.1 and 6.2, we can write

$$\langle \phi_{T_1}\phi_{T_2}, \phi_{T_3}\phi_{T_4}\rangle = \int e^{i\Psi}\phi_1\phi_2\bar{\phi}_3\bar{\phi}_4\,\eta_N(t)\,dx\,dt, \tag{47}$$

where ϕ_j is a Schrödinger solution which satisfies

$$(N\partial_x)^k\phi_j(t,x) \lesssim_{k,M} N^{-\frac{d}{2}}\langle N^{-1}(x-x_j^t)\rangle^{-M},$$

and

$$\Psi = \sum_{j=1}^{4}\sigma_j\Big[\langle x-x_j^t, \xi_j^t\rangle - \int_0^\tau \tfrac{1}{2}|\xi_j^\tau|^2 - V(\tau, x_j^\tau)\,d\tau\Big], \quad \sigma = (+,+,-,-).$$

Using the rapid decay of each ϕ_j, we may harmlessly (with $O(N^{-100d})$ error) localize ϕ_j to an N^δ neighborhood of the tube T_j, so that $\phi_j(t)$ is supported in an $O(N^{1+\delta})$ neighborhood of the classical path x_j^t.

Then

$$\partial_x\Psi = \sum_j \sigma_j\xi_j^t, \quad -\partial_t\Psi = \frac{1}{2}\sum_j\sigma_j|\xi_j^t|^2 + \sum_j\sigma_j[V(t,x_j^t) + \langle x-x_j^t, \partial_x V(t,x_j^t)\rangle].$$

The first bound in the statement of the lemma results from integrating by parts in x, as in the proof of (23), to gain factors of $(N|\xi_1^t+\xi_2^t-\xi_3^t-\xi_4^t|)^{-1}$. Since

$$\xi_1^t+\xi_2^t-\xi_3^t-\xi_4^t = \xi_1^{t_q}+\xi_2^{t_q}-\xi_3^{t_q}-\xi_4^{t_q} + O(N^{-2+2\delta})$$

during the time window $|t-t_q| \leq O(N^{1+\delta})$ when $|x_j^t - x_q| \leq N^{1+\delta}$, we may replace t by t_q.

As in our work in one space dimension (more specifically, the proof of [Jao et al. 2019, Lemma 4.4]), instead of integrating by parts purely in time we use a vector field adapted to the average bicharacteristic for the four wavepackets ϕ_{T_j}. Defining

$$\bar{x}^t := \frac{1}{4}\sum_{j=1}^{4}x_j^t, \quad \bar{\xi}^t := \sum_{j=1}^{4}\xi_j^t, \quad L := \partial_t + \langle\bar{\xi}^t, \partial_x\rangle,$$

we compute as in that paper that

$$-L\Psi = \frac{1}{2}\sum\sigma_j|\bar{\xi}_j^t|^2 + \sum\sigma_j[V^{\bar{z}}(t,\bar{x}_j^t) + \langle x-x_j^t, \partial_x(V^{\bar{z}})(t,\bar{x}_j^t)\rangle],$$

where

$$\bar{x}_j^t := x_j^t - \bar{x}^t, \quad \bar{\xi}_j^t := \xi_j^t - \bar{\xi}^t$$

denote the coordinates of $\phi_{T_j}(t)$ in phase space relative to $(\bar{x}^t, \bar{\xi}^t)$; see Figure 2.

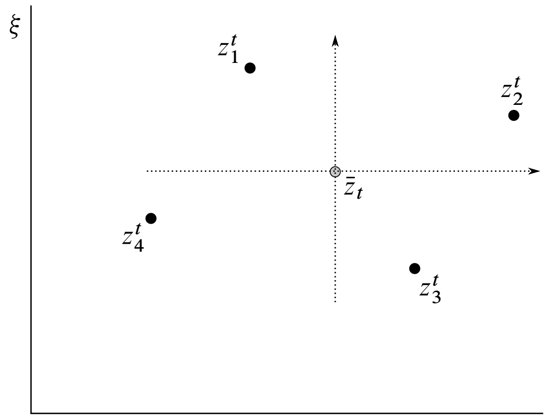

Figure 2. Phase space coordinates relative to the "center of mass".

We cannot yet integrate by parts since that would require two time derivatives of the phase Ψ, but the assumptions on V only allow Ψ to be differentiated once in time. However, we can decompose Ψ into $\Psi = \Psi_1 + \Psi_2$, where Ψ_2 has two time derivatives and accounts for the majority of the oscillation of $e^{i\Psi}$; indeed, we define Ψ_1 and Ψ_2 via the ODE

$$-L\Psi_2 = \frac{1}{2}\sum_j \sigma_j |\bar{\xi}_j^t|^2 = \frac{1}{4}(|\xi_1^{t_q} - \xi_2^{t_q}|^2 - |\xi_3^{t_q} - \xi_4^{t_q}|^2) + O(N^{-2+2\delta}),$$

$$-L\Psi_1 = \sum \sigma_j [V^{\bar{z}}(t, \bar{x}_j^t) + \langle x - x_j^t, \partial_x (V^{\bar{z}})(t, \bar{x}_j^t)\rangle] = O(N^{-2+2\delta}).$$

As before we have frozen $t = t_q$ in the main term with error at most $O(N^{-2+2\delta})$, and also used the estimates $|\bar{x}_j^t| \le \max_{j,k} |x_j^t - x_k^t| \lesssim N^{1+\delta}$, $|x - x_j^t| \lesssim N^{1+\delta}$ on the support of the integrand (47). Note also that the equation

$$\frac{d}{dt}\xi_j^t = -\partial_x V(t, x_j^t)$$

implies $L^2\Psi_2 = O(N^{-2})$. Now integrate by parts using the phase Ψ_2 to obtain

$$\text{RHS (47)} = \int e^{i\Psi_2} e^{i\Psi_1} \prod_j \phi_j \, \eta_N(t)\, dx\, dt = i \int e^{i\Psi_2} \Big\langle L, \frac{L\Psi_2}{|L\Psi_2|^2} \Big\rangle e^{i\Psi_1} \phi_1 \phi_2 \bar{\phi}_3 \bar{\phi}_4 \, \eta_N(t)\, dx\, dt$$

$$= i \int e^{i\Psi} \Big[-\frac{L^2\Psi_2}{|L\Psi_2|^2} + \Big\langle \frac{L\Psi_2}{|L\Psi_2|^2}, iL\Psi_1 + L \Big\rangle \Big] \phi_1 \phi_2 \bar{\phi}_3 \bar{\phi}_4 \, \eta_N(t)\, dx\, dt,$$

and the second bound in the lemma follows. □

Returning to the proof of Lemma 6.9, we decompose the sum into

$$\sum_{(T_1, T_2') \in \boldsymbol{T}_1' \times \boldsymbol{T}_2'} \Big[\sum_{T_1' \in \boldsymbol{T}_1^s} \sum_{T_2 \in \boldsymbol{T}_2'} + \sum_{0 \le k \lesssim \log N} \sum_{T_1' \in \boldsymbol{T}_{1,k}'} \sum_{T_2 \in \boldsymbol{T}_2'} \Big],$$

where $\boldsymbol{T}_1^s$ is the set of tubes in $\boldsymbol{T}_1'$ whose bicharacteristic $((x_1')^t, (\xi_1')^t)$ satisfies $(\xi_1')^{t_q} \in \pi^s(\xi_1^{t_q}, (\xi_2')^{t_q})$, and we abbreviate

$$\boldsymbol{T}_{1,k}' := \boldsymbol{T}_1^{\sim B}(q, \lambda_1, \mu_1, \mu_2, \xi_1^{t_q}, (\xi_2')^{t_q}, k).$$

The contribution from the "space nonresonance" terms $\boldsymbol{T}_1^s$ is $O(N^{-100d})$.

Now consider the k-th sum. Lemma 6.10 implies

$$|\langle \phi_{T_1}\phi_{T_2}, \phi_{T_1'}\phi_{T_2'}\rangle| \lesssim N^{C\delta} N^{-(d-1)} 2^{-k}.$$

For each $T_1' \in \boldsymbol{T}_1^{\sim B}(q, \lambda_1, \mu_1, \mu_2, \xi_1^{t_q}, (\xi_2')^{t_q}, k)$, the possible tubes T_2 correspond to the bicharacteristics (x_2^t, x_2^t) such that

$$|x_2^{t_q} - x_q| \le N^{1+\delta}, \quad \xi_1^{t_q} + \xi_2^{t_q} - (\xi_1')^{t_q} - (\xi_2')^{t_q} = O(N^{-1+C\delta}).$$

The preimage of this set under the time-t_q Hamiltonian flow map is an $(N^{1+C\delta})^d \times (N^{-1+C\delta})^{-d}$ box, so there are $O(N^{C\delta})$ choices of tubes T_2. Therefore, the k-th sum is at most

$$N^{C\delta} N^{-(d-1)} 2^{-k} \nu_k \# \boldsymbol{T}_1' \# \boldsymbol{T}_2',$$

whereupon the sum over k is replaced by the supremum at the cost of a $\log N$ factor. □

It remains to show that

$$\sum_{q \in \boldsymbol{q}(\mu_1,\mu_2): q \subset 2B} 2^{-k} \nu_k(q, \lambda_1, \mu_1, \mu_2) \#(\boldsymbol{T}_1^{\sim B}(q) \cap \boldsymbol{T}_1[\lambda_1, \mu_1, \mu_2]) \# \boldsymbol{T}_2(q) \lessapprox N^{C\delta} \# \boldsymbol{T}_1 \# \boldsymbol{T}_2. \tag{48}$$

6I. *Tube combinatorics.* This section begins exactly as in [Tao 2003, Section 8]. We define the relation $\sim$ between tubes and radius $N^{2(1-\delta)}$ balls. For a tube $T \in \boldsymbol{T}_1[\lambda_1, \mu_1, \mu_2]$, let $B(T, \lambda_1, \mu_1, \mu_2)$ be a ball $B \in \mathcal{B}$ that maximizes

$$\#\{q \in \boldsymbol{q}(\mu_1, \mu_2) : T \cap N^\delta q \ne \phi;\ q \cap B \ne \phi\}.$$

As T intersects roughly λ_1 (neighborhoods of) balls $q \in \boldsymbol{q}(\mu_1, \mu_2)$ in total and there are $O(N^{2\delta})$ many balls in $\mathcal{B}$, $B(T, \lambda_1, \mu_1, \mu_2)$ must intersect at least $N^{-2\delta}\lambda_1$ of those balls.

Declare $T \sim_{\lambda_1,\mu_1,\mu_2} B'$ if $T \in \boldsymbol{T}_1[\lambda_1, \mu_1, \mu_2]$ and $B' \subset 10B(T, \lambda_1, \mu_1, \mu_2)$. Finally, for $T \in \boldsymbol{T}_1$ set $T \sim B$ if $T \sim_{\lambda_1,\mu_1,\mu_2} B$ for some λ_1, μ_1, μ_2. Evidently $T \sim B$ for at most $(\log N)^3 \lessapprox 1$ many balls. The relation between tubes in $\boldsymbol{T}_2$ and balls in $\mathcal{B}$ is defined similarly.

Now we begin the proof of (48). On one hand,

$$\begin{aligned}\sum_{q \in \boldsymbol{q}(\mu_1,\mu_2)} \#(\boldsymbol{T}_1[\lambda_1, \mu_1, \mu_2] \cap \boldsymbol{T}_1(q)) &= \sum_{q \in \boldsymbol{q}(\mu_1,\mu_2)} \sum_{T_1 \in \boldsymbol{T}_1[\lambda_1,\mu_1,\mu_2] \cap \boldsymbol{T}_1(q)} 1_{T_1 \cap N^\delta q \ne 0} \\ &= \sum_{T \in \boldsymbol{T}_1[\lambda_1,\mu_1,\mu_2]} \sum_{q \in \boldsymbol{q}(\mu_1,\mu_2)} 1_{T_1 \cap N^\delta q \ne \phi} \lesssim \sum_{T \in \boldsymbol{T}_1} \lambda_1 = \lambda_1 \# \boldsymbol{T}_1.\end{aligned}$$

On the other hand, by definition $\#\boldsymbol{T}_2(q) \lesssim \mu_2$. The claim (48) would therefore follow if we could show

$$\nu_k(q_0, \lambda_1, \mu_1, \mu_2) \lessapprox 2^k N^{C\delta} \frac{\#\boldsymbol{T}_2}{\lambda_1 \mu_2} \tag{49}$$

for all $q_0 \in \boldsymbol{q}(\mu_1, \mu_2)$ such that $q_0 \subset 2B$.

Fix $\xi_1 \in S_1,\ \xi_2' \in S_2$, and a ball $q_0 = q_0(t_q, x_q)$. Recalling the definition (46) of ν_k, we need to show

$$\#\boldsymbol{T}_1^{\sim B}(q_0, \lambda_1, \mu_1, \mu_2, \xi_1^{t_q}, (\xi_2')^{t_q}, k) \lessapprox 2^k N^{C\delta} \frac{\#\boldsymbol{T}_2}{\lambda_1 \mu_2}.$$

For brevity write $\boldsymbol{T}_1' := \boldsymbol{T}_{1,k}^{\sim B}(q_0, \lambda_1, \mu_1, \mu_2, \xi_1^{t_q}, (\xi_2')^{t_q}, k)$.

Fix $T_1 \in \boldsymbol{T}_1'$. Since $T_1 \sim B$, the ball $2B(T_1, \lambda_1, \mu_1, \mu_2)$ has distance $\gtrsim N^{2(1-\delta)}$ from q_0. Thus

$$\#\{q \in \boldsymbol{q}(\mu_1, \mu_2) : T_1 \cap N^\delta q \neq \phi,\ \operatorname{dist}(q, q_0) \gtrapprox N^{2(1-\delta)}\} \gtrapprox N^{-2\delta} \lambda_1.$$

As each $q \in \boldsymbol{q}(\mu_1, \mu_2)$ intersects approximately μ_2 (N^δ-neighborhoods of) tubes in $\boldsymbol{T}_2$,

$$\#\{(q, T_2) \in \boldsymbol{q}(\mu_1, \mu_2) \times \boldsymbol{T}_2 : T_1 \cap N^\delta q \neq \phi,\ T_2 \cap N^\delta q \neq \phi,\ \operatorname{dist}(q, q_0) \gtrapprox N^{2(1-\delta)}\} \gtrapprox N^{-2\delta} \lambda_1 \mu_2.$$

Therefore

$$\#\big\{(q, T_1, T_2) \in \boldsymbol{q} \times \boldsymbol{T}_1' \times \boldsymbol{T}_2 : T_1 \cap N^\delta q \neq \phi,\ T_2 \cap N^\delta q \neq \phi,\ \operatorname{dist}(q, q_0) \gtrapprox N^{-2\delta} N^2\big\} \gtrapprox N^{-2\delta} \lambda_1 \mu_2 \#\boldsymbol{T}_1'.$$

On the other hand, the cardinality can be bounded above by the following analogue of Tao's Lemma 8.1:

Lemma 6.11. *For each* $T_2 \in \boldsymbol{T}_2$,

$$\#\big\{(q, T_1) \in \boldsymbol{q} \times \boldsymbol{T}_1' : T_1 \cap N^\delta q,\ T_2 \cap N^\delta q \neq \phi,\ \operatorname{dist}(q, q_0) \gtrapprox N^{-2\delta} N^2\big\} \lessapprox 2^k N^{C\delta}.$$

Proof. We estimate in two steps:

- For any tubes $T_1 \in \boldsymbol{T}_1'$ and $T_2 \in \boldsymbol{T}_2$, the intersection $N^\delta T_1 \cap N^\delta T_2$ is contained in a ball of radius $N^{C\delta}$.
- The number of tubes $T_1 \in \boldsymbol{T}_1'$ such that T_1 intersects $N^\delta T_2$ at distance $\gtrapprox N^{-2\delta} N^2$ from q_0 bounded above by $2^k N^{C\delta}$.

The first is evident from transversality. Hence we turn to the second claim.

In Tao's situation, the tubes in $\boldsymbol{T}_1'$ are all constrained to an $O(N^{-1+C\delta})$ neighborhood of a space-time hyperplane transverse to the tube T_2 (basically because of Lemma 6.8), and there are $O(N^{C\delta})$ many such tubes that intersect T_2 at distance $\gtrapprox N^{-2\delta} N^2$ from q_0. The extra 2^k factor results from the fact that we allow the tubes to deviate from that hyperplane by distance $2^k N^{-1+C\delta}$. Also, since our tubes are curved it is more convenient to work with their associated bicharacteristics instead of using Euclidean geometry in space-time.

Fix a tube $T_2 \in \boldsymbol{T}_2$ with ray $t \mapsto (x_2^t, \xi_2^t)$. Then, the tubes $T_1 \in \boldsymbol{T}_1'$ such that $N^\delta T_1 \cap N^\delta T_2$ are characterized by the property that

$$|x(T_1)^t - x_2^t| \lesssim N^{1+\delta} \quad \text{for some } |t - t_q| \gtrapprox N^{-2\delta} N^2.$$

We need to count the tubes in $\boldsymbol{T}_1'$ with this property. The bicharacteristics for such tubes emanate from the region

$$\begin{aligned}\Sigma := \{(x, \xi) : \operatorname{dist}(\xi, S_1) \le N^{-1+C\delta},\ \xi^{t_q} \in \pi_k^t,&\\ |x^{t_q} - x_q| \le N^{1+\delta},\ |x^t - x_2^t| \le N^{1+\delta} \text{ for some } |t - t_q| \gtrsim N^{-2\delta} N^2\};&\end{aligned}$$

hence it suffices to bound the cardinality of the intersection $(N\mathbb{Z}^d \times N^{-1}\mathbb{Z}^d) \cap \Sigma$.

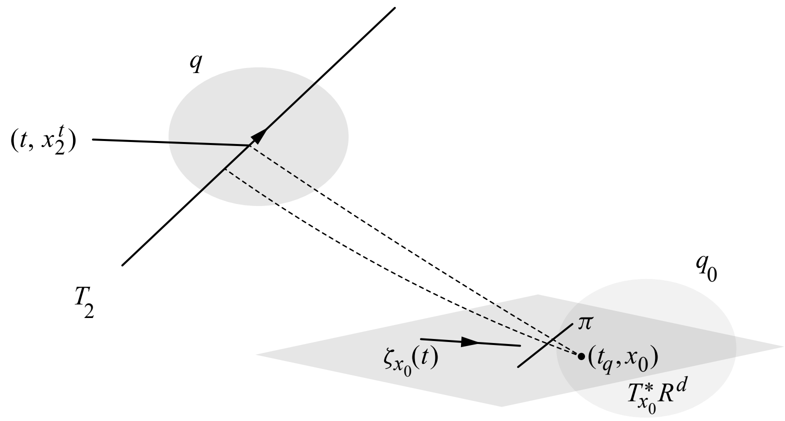

Figure 3. $\zeta_{x_0}(t) \in T^*_{x_0}\mathbb{R}^d$ is the set of tangent (covectors) for rays passing through (t_q, x_0) that intersect the ray (t, x_2^t) for the tube T_2 at times $|t - t_q| \gtrapprox N^{2-2\delta}$.

Denote by Σ^t the image of Σ under the time-t Hamiltonian flow map $(x, \xi) \mapsto (x^t, \xi^t)$. Recall from (36) that S_j^t denotes the image of the initial frequency set S_j for initial positions x with $|x| \lesssim N^2$; we saw earlier in (38) that S_j^t is a small perturbation of S_j.

Fix a basepoint x_0 with $|x_0 - x_q| \le N^{1+\delta}$. By Lemma 2.1 and the Hadamard global inverse function theorem, when $t \ne t_q$ we can parametrize the graph of the flow map $(x^{t_q}, \xi^{t_q}) \mapsto (x^t, \xi^t)$ by the variables

$$(x^{t_q}, x^t) \mapsto \big((x^{t_q}, \xi^{t_q}(x^t, x^{t_q})) \mapsto (x^t, \xi^t(x^{t_q}, x^t))\big).$$

Let $\xi(t, x) := \xi^{t_q}(x_0, x) \in T^*_{x_0}\mathbb{R}^d$ be the initial momentum $\xi(t, x) \in T^*_{x_0}\mathbb{R}^d$ such that the bicharacteristic with $x^{t_q} = x_0$ and $\xi^{t_q} = \xi(t, x)$ satisfies $x^t = x$.

Lemma 6.12. *Suppose at least one $T_1 \in \boldsymbol{T}_1'$ intersects $N^\delta T_2$. For $|t - t_q| \gtrapprox N^{-2\delta} N^2$, the curve $t \mapsto \zeta_{x_0}(t) := \xi(t, x_2^t) \in T^*_{x_0}\mathbb{R}^d$ is transverse to the hyperplane containing $\pi(\xi_1, \xi_2')$ for all $\xi_1 \in S_1^{t_q}$ and $\xi_2' \in S_2^{t_q}$ (see Figure 3). More precisely there exists $C(\eta) > 0$ such that*

$$\angle(\dot{\zeta}_{x_0}(t), \pi(\xi_1, \xi_2')) > C(\eta) \quad \textit{for all } \xi_1 \in S_1^{t_q},\ \xi_2' \in S_2^{t_q},$$

where the angle $\angle(v, W)$ between a vector v and a subspace W is defined in the usual manner. Moreover, for each t the image of an $N^{1+\delta}$ neighborhood of x_2^t under the map $x \mapsto \xi(t, x)$ belongs to an $N^{-1+C\delta}$ neighborhood of $\zeta_{x_0}(t)$.

Proof. By a slight abuse of notation we write $(x^t(y, \zeta), \xi^t(y, \zeta))$ for the bicharacteristic passing through (y, ζ) at time $t = t_q$ instead of $t = 0$. Both claims are consequences of Lemma 2.1, which yields

$$x_2^t = x^t(x_0, \zeta_{x_0}(t)), \quad \xi^{t_q}(x_0, \zeta_{x_0}(t)) = \zeta_{x_0}(t),$$
$$\xi_2^t = \frac{d}{dt} x_2^t = \xi^t(x_0, \zeta_{x_0}(t)) + \frac{\partial x^t}{\partial \zeta_{x_0}} \dot{\zeta}_{x_0}(t) = \xi^t(x_0, \zeta_{x_0}(t)) + (t - t_q)(I + O(\eta))\dot{\zeta}_{x_0}(t).$$

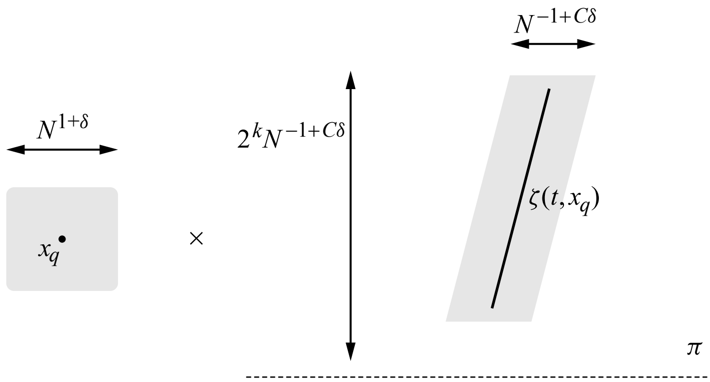

Figure 4. The phase-space region Σ^{t_q}.

Therefore

$$\dot{\zeta}_{x_0}(t) = (t-t_q)^{-1}(I + O(\eta))(\xi_2^t - \xi^t(x_0, \zeta_{x_0}(t))). \tag{50}$$

We claim that for any $C > 1$,

$$\operatorname{dist}(\zeta_{x_0}(t_q), S_1^{t_q}) \lesssim_C N^{-1+C\delta}. \tag{51}$$

Otherwise, as $|t-t_q| \gtrapprox N^{2(1-\delta)}$, for any ray (x_1^t, ξ_1^t) with $\xi_1 \in S_1$ and $|x_1^{t_q} - x_q| \le N^{1+\delta}$, the estimates (19) would imply

$$\begin{aligned}|x_1^t - x_2^t| &\gtrsim |t-t_q|\,|\xi_1^{t_q} - \zeta_{x_0}(t)| - |x_1^{t_q} - x_0| \\ &\gtrapprox N^{-1+C\delta} - N^{1+\delta} \gtrapprox N^{1+C\delta},\end{aligned}$$

so we get the contradiction that every $T_1 \in \boldsymbol{T}_1'$ misses T_2 by at least $N^{C\delta}$.

By the near-constancy (38) of the frequency variable and the definition (29) of S_j, the covector $\xi_2^t - \xi^t(x_0, \zeta_{x_0}(t))$ belongs a small perturbation (say, of magnitude at most $\frac{c}{50}$) of the difference set $S_2 - S_1 = -2ce_1 + B(0, \frac{c}{50})$, and hence by Lemma 6.8 is transverse to the hyperplane containing $\pi(\xi_1, \xi_2')$. The first claim now follows from (50).

The argument just given also implies the second statement: a ray with $x^{t_q} = x_0$ and $|x_2^t - x^t| \le N^{1+\delta}$ must satisfy $|\xi^{t_q} - \zeta_{x_0}(t)| \lesssim N^{-1+C\delta}$. □

By the second part of the lemma, the fiber of Σ^{t_q} in $T_{x_0}^*\mathbb{R}^d$ is contained in a "frequency tube"

$$\Theta(x_0) := \bigcup_{|t-t_q| \gtrapprox N^{2(1-\delta)}} B(\zeta_{x_0}(t), N^{-1+C\delta}).$$

As the basepoint x_0 varies in an $N^{1+\delta}$ neighborhood of x_q, the estimate (19) implies that the curve $\zeta_{x_0}(t)$ shifts by at most $O(N^{-1+3\delta})$. Hence the tubes $\Theta(x_0)$ are all contained in a dilate of $\Theta(x_q)$, which we denote by

$$\widetilde{\Theta}(x_q) := \bigcup_t B(\zeta_{x_q}(t), N^{-1+C\delta})$$

with a larger C.

Therefore, Σ^{t_q} is contained in the region

$$\widetilde{\Sigma}^{t_q} := \{(x,\xi) : |x - x_q| \le N^{1+\delta},\ \xi \in \pi_k^t \cap \widetilde{\Theta}(x_q) \subset \{\xi \in \widetilde{\Theta}(x_q) : \operatorname{dist}(\xi,\pi) \lesssim 2^k N^{-1+C\delta}\}\},$$

where for the last containment we recall the estimate (45). The region $\widetilde{\Sigma}^{t_q}$ is sketched in Figure 4. Using the previous lemma for the central curve ζ_{x_q}, the frequency projection $(x,\xi) \mapsto \xi)$ of $\widetilde{\Sigma}^{t_q}$ can be covered by approximately 2^k finitely overlapping cubes $\bigcup_{1 \le j \lesssim 2^k} Q_j$ of width $N^{-1+C\delta}$. By (19), the preimage of each box

$$B(x_q, N^{1+\delta}) \times Q_j$$

under the flow map $(x,\xi) \mapsto (x^{t_q}, \xi^{t_q})$ is contained in a $(CN^{1+C\delta})^d \times (CN^{-1+C\delta})^d$ box. The union of these preimages covers Σ and contains at most $O(2^k N^{C\delta})$ points in $N\mathbb{Z}^d \times N^{-1}\mathbb{Z}^d$. □

7. Remarks on magnetic potentials

We sketch the modifications needed to prove Theorem 1.6. The symbol for $H(t)$ is

$$a = \tfrac{1}{2}|\xi|^2 + \langle A, \xi\rangle + V(t,x),$$

where $A = A_j(t,x)dx^j$ and A_j are linear functions in the space variables with bounded time-dependent coefficients.

- Easy computation shows that the symbol map $a \mapsto a^{z_0}$ in Lemma 1.1 is

$$a^{z_0} = \tfrac{1}{2}|\xi|^2 + \langle A^{z_0}_{(1)}(t,x), \xi\rangle + \langle A^{z_0}_{(2)}(t,x), \xi_0^t\rangle + V^{z_0}_{(2)}(t,x),$$

where $A^{z_0}_{(1)}(t,x) = A(t, x_0^t + x) - A(t, x_0^t)$ and $A^{z_0}_{(2)}(t,x) = A(t, x_0^t + x) - \langle x, \partial_x A(t, x_0^t)\rangle - A(x_0^t)$, and similarly for V. Thus when A is linear, the first-order component of the symbol is exactly "Galilei-invariant", preserved by the transformation $a \mapsto a^{z_0}$ in Lemma 1.1.

- After rescaling, the inequality (15) takes the form

$$\|U_N f U_N g\|_{L^{(d+3)/(d+1)}([-\tau_0 N^2, \tau_0 N^2]\times\mathbb{R}^d)} \lesssim_\varepsilon N^\varepsilon \|f\|_{L^2}\|g\|_{L^2},$$

where $U_N(t)$ is the propagator for the rescaled symbol

$$a_N := N^{-2} a(N^{-2}t, N^{-1}x, N\xi) = \tfrac{1}{2}|\xi|^2 + N^{-2}\langle A(x), \xi\rangle + N^{-2} V(N^{-2}t, N^{-1}x).$$

- Exploiting Galilei-invariance, we may reduce to a spatially localized estimate as in Theorem 6.4. Note that in the region of phase space corresponding to that estimate $\{(x,\xi) : |x| \le N^2,\ |\xi| \lesssim 1\}$, and over an $O(N^2)$ time interval, both potential terms have strength $O(1)$ when integrated over the time interval $|t| \lesssim N^2$. However the magnetic term dominates near $x = 0$.

- Then, the rest of the previous proof can be mimicked with essentially no change except for Lemma 6.10. There, one argues essentially as before except the vector field L for integrating by parts should be replaced by

$$L := \partial_t + \langle \overline{a_\xi(z_j^t)}, \partial_x\rangle,$$

where $z_j^t = (x_j^t, \xi_j^t)$ and $\overline{a_\xi(z_j^t)} = \frac{1}{4}\sum_k a_\xi(z_k^t)$. Then one finds that

$$-L\Psi = \frac{1}{2}\sum_j \sigma_j |\bar{\xi}_j^t|^2 + \sum_j \sigma_j \langle A(\bar{x}_j^t), \xi_j^t\rangle + \sum_j \sigma_j [V^{\bar{z}}(t, \bar{x}_j^t) + \langle x - x_j^t, \partial_x(V^{\bar{z}})(t, \bar{x}_j^t)\rangle],$$

and decomposes as before $\Psi = \Psi_1 + \Psi_2$, where

$$-L\Psi_1 = \frac{1}{2}\sum_j \sigma_j |\bar{\xi}_j^t|^2 = |\xi_1^{t_q} - \xi_2^{t_q}|^2 - |\xi_3^{t_q} - \xi_4^{t_q}|^2 + O(N^{-1+\delta}),$$

$$-L\Psi_2 = \sum_j \sigma_j \langle A(\bar{x}_j^t), \xi_j^t\rangle + \sum_j \sigma_j [V^{\bar{z}}(t, \bar{x}_j^t) + \langle x - x_j^t, \partial_x(V^{\bar{z}})(t, \bar{x}_j^t)\rangle] = O(N^{-1+\delta}).$$

As in the proof of Lemma 6.10 the error terms are computed from the estimates (19), $|t - t_q| \lesssim N^{1+\delta}$, and $|\bar{x}_j^t| \lesssim N^{1+\delta}$. The errors are larger than before due to the magnetic term $a_{x\xi} = O(N^{-2})$ but are still acceptable.

Acknowledgements

The author is grateful to Michael Christ, Rowan Killip, Daniel Tataru, and Monica Visan for many helpful discussions and also wishes to thank the anonymous referee for numerous suggestions for improving the original manuscript. This research was partially supported by the National Science Foundation under award no. 1604623. Part of this work was completed during the 2017 Oberwolfach workshop "Nonlinear waves and dispersive equations".

References

[Bégout and Vargas 2007] P. Bégout and A. Vargas, "Mass concentration phenomena for the L^2-critical nonlinear Schrödinger equation", *Trans. Amer. Math. Soc.* **359**:11 (2007), 5257–5282. MR Zbl

[Burq et al. 2013] N. Burq, L. Thomann, and N. Tzvetkov, "Long time dynamics for the one dimensional non linear Schrödinger equation", *Ann. Inst. Fourier* (*Grenoble*) **63**:6 (2013), 2137–2198. MR Zbl

[Carles 2011] R. Carles, "Nonlinear Schrödinger equation with time dependent potential", *Commun. Math. Sci.* **9**:4 (2011), 937–964. MR Zbl

[Carles and Keraani 2007] R. Carles and S. Keraani, "On the role of quadratic oscillations in nonlinear Schrödinger equations, II: The L^2-critical case", *Trans. Amer. Math. Soc.* **359**:1 (2007), 33–62. MR Zbl

[Dodson 2012] B. Dodson, "Global well-posedness and scattering for the defocusing, L^2-critical nonlinear Schrödinger equation when $d \geq 3$", *J. Amer. Math. Soc.* **25**:2 (2012), 429–463. MR Zbl

[Dodson 2015] B. Dodson, "Global well-posedness and scattering for the mass critical nonlinear Schrödinger equation with mass below the mass of the ground state", *Adv. Math.* **285** (2015), 1589–1618. MR Zbl

[Dodson 2016a] B. Dodson, "Global well-posedness and scattering for the defocusing, L^2 critical, nonlinear Schrödinger equation when $d = 1$", *Amer. J. Math.* **138**:2 (2016), 531–569. MR Zbl

[Dodson 2016b] B. Dodson, "Global well-posedness and scattering for the defocusing, L^2-critical, nonlinear Schrödinger equation when $d = 2$", *Duke Math. J.* **165**:18 (2016), 3435–3516. MR Zbl

[Fujiwara 1979] D. Fujiwara, "A construction of the fundamental solution for the Schrödinger equation", *J. Analyse Math.* **35** (1979), 41–96. MR Zbl

[Ginibre and Velo 1995] J. Ginibre and G. Velo, "Generalized Strichartz inequalities for the wave equation", *J. Funct. Anal.* **133**:1 (1995), 50–68. MR Zbl

[Ionescu et al. 2012] A. D. Ionescu, B. Pausader, and G. Staffilani, "On the global well-posedness of energy-critical Schrödinger equations in curved spaces", *Anal. PDE* **5**:4 (2012), 705–746. MR Zbl

[Jao 2016] C. Jao, "The energy-critical quantum harmonic oscillator", *Comm. Partial Differential Equations* **41**:1 (2016), 79–133. MR Zbl

[Jao et al. 2019] C. Jao, R. Killip, and M. Vişan, "Mass-critical inverse Strichartz theorems for 1d Schrödinger operators", *Rev. Mat. Iberoam.* **35**:3 (2019), 703–730. MR Zbl

[Keel and Tao 1998] M. Keel and T. Tao, "Endpoint Strichartz estimates", *Amer. J. Math.* **120**:5 (1998), 955–980. MR Zbl

[Killip and Visan 2013] R. Killip and M. Vişan, "Nonlinear Schrödinger equations at critical regularity", pp. 325–437 in *Evolution equations*, edited by D. Ellwood et al., Clay Math. Proc. **17**, Amer. Math. Soc., Providence, RI, 2013. MR Zbl

[Killip et al. 2008] R. Killip, M. Visan, and X. Zhang, "The mass-critical nonlinear Schrödinger equation with radial data in dimensions three and higher", *Anal. PDE* **1**:2 (2008), 229–266. MR Zbl

[Killip et al. 2009] R. Killip, T. Tao, and M. Visan, "The cubic nonlinear Schrödinger equation in two dimensions with radial data", *J. Eur. Math. Soc.* (*JEMS*) **11**:6 (2009), 1203–1258. MR Zbl

[Killip et al. 2016] R. Killip, M. Visan, and X. Zhang, "Quintic NLS in the exterior of a strictly convex obstacle", *Amer. J. Math.* **138**:5 (2016), 1193–1346. MR Zbl

[Koch and Tataru 2005] H. Koch and D. Tataru, "Dispersive estimates for principally normal pseudodifferential operators", *Comm. Pure Appl. Math.* **58**:2 (2005), 217–284. 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* 8 (1998), 399–425. MR Zbl

[Moyua et al. 1999] A. Moyua, A. Vargas, and L. Vega, "Restriction theorems and maximal operators related to oscillatory integrals in $\mathbb{R}^3$", *Duke Math. J.* **96**:3 (1999), 547–574. MR Zbl

[Poiret et al. 2014] A. Poiret, D. Robert, and L. Thomann, "Probabilistic global well-posedness for the supercritical nonlinear harmonic oscillator", *Anal. PDE* **7**:4 (2014), 997–1026. MR Zbl

[Rogers and Vargas 2006] K. M. Rogers and A. Vargas, "A refinement of the Strichartz inequality on the saddle and applications", *J. Funct. Anal.* **241**:1 (2006), 212–231. MR Zbl

[Strichartz 1977] R. S. Strichartz, "Restrictions of Fourier transforms to quadratic surfaces and decay of solutions of wave equations", *Duke Math. J.* **44**:3 (1977), 705–714. MR Zbl

[Tao 2003] T. Tao, "A sharp bilinear restrictions estimate for paraboloids", *Geom. Funct. Anal.* **13**:6 (2003), 1359–1384. MR Zbl

[Tao et al. 2007] T. Tao, M. Visan, and X. Zhang, "Global well-posedness and scattering for the defocusing mass-critical nonlinear Schrödinger equation for radial data in high dimensions", *Duke Math. J.* **140**:1 (2007), 165–202. 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

[Yajima 1991] K. Yajima, "Schrödinger evolution equations with magnetic fields", *J. Analyse Math.* **56** (1991), 29–76. MR Zbl

[Zhang 2000] J. Zhang, "Stability of attractive Bose–Einstein condensates", *J. Statist. Phys.* **101**:3-4 (2000), 731–746. MR Zbl

[Zhang 2005] J. Zhang, "Sharp threshold for blowup and global existence in nonlinear Schrödinger equations under a harmonic potential", *Comm. Partial Differential Equations* **30**:10-12 (2005), 1429–1443. MR Zbl

Received 30 Oct 2017. Revised 29 Jul 2018. Accepted 6 Sep 2019.

CASEY JAO: cjao@math.toronto.edu
Department of Mathematics, University of Toronto, Toronto, ON, Canada

ANALYSIS AND PDE
Vol. 13, No. 7, 2020
https://doi.org/10.2140/apde.2020.13.1995

SCATTERING FOR DEFOCUSING ENERGY SUBCRITICAL NONLINEAR WAVE EQUATIONS

BENJAMIN DODSON, ANDREW LAWRIE, DANA MENDELSON AND JASON MURPHY

We consider the Cauchy problem for the defocusing power-type nonlinear wave equation in (1+3)-dimensions for energy subcritical powers p in the superconformal range $3 < p < 5$. We prove that any solution is global-in-time and scatters to free waves in both time directions as long as its critical Sobolev norm stays bounded on the maximal interval of existence.

1. Introduction 1995
2. Preliminaries 2001
3. Concentration compactness and the reduction of Theorem 1.1 2004
4. The soliton-like critical element 2016
5. The self-similar critical element 2031
6. Doubly concentrating critical element: the sword and shield 2042
7. The traveling-wave critical element 2050
Acknowledgments 2088
References 2088

1. Introduction

We study the Cauchy problem for the power-type nonlinear wave equation in $\mathbb{R}^{1+3}$,

$$\begin{cases} \Box u = \pm u|u|^{p-1}, \\ \vec{u}(0) = (u_0, u_1), \quad u = u(t,x), \quad (t,x) \in \mathbb{R}^{1+3}_{t,x}. \end{cases} \tag{1-1}$$

Here $\Box = -\partial_t^2 + \Delta$ so the "+" above yields the defocusing equation and the "−" yields the focusing equation. The equation has the following scaling symmetry: if $\vec{u}(t,x) = (u, \partial_t u)(t,x)$ is a solution, then so is

$$\vec{u}_\lambda(t,x) = \left(\lambda^{-\frac{2}{p-1}} u\left(\frac{t}{\lambda}, \frac{x}{\lambda}\right), \lambda^{-\frac{2}{p-1}-1} \partial_t u\left(\frac{t}{\lambda}, \frac{x}{\lambda}\right)\right). \tag{1-2}$$

The conserved energy, or Hamiltonian, is

$$E(\vec{u}(t)) = \int_{\{t\}\times\mathbb{R}^3} \frac{1}{2}(|u_t|^2 + |\nabla u|^2) \pm \frac{1}{p+1}|u|^{p+1}\,\mathrm{d}x = E(\vec{u}(0)),$$

which scales like

$$E(\vec{u}_\lambda) = \lambda^{3-2\frac{p+1}{p-1}} E(\vec{u}).$$

MSC2010: 35L71.
Keywords: nonlinear waves, scattering.

The energy is invariant under the scaling of the equation only when $p=5$, which is referred to as the energy-critical exponent. The range $p<5$ is called energy subcritical, since concentration of a solution by rescaling requires divergent energy; i.e., $\lambda \to 0$ implies $E(\vec{u}_\lambda) \to \infty$. Conversely, the range $p>5$ is called energy supercritical, and here $E(\vec{u}_\lambda) \to 0$ as $\lambda \to 0$; i.e., concentration by rescaling is energetically favorable.

Fixing p, the critical Sobolev exponent $s_p := \frac{3}{2} - \frac{2}{p-1}$ is defined to be the unique $s_p \in \mathbb{R}$ so that $\dot{H}^{s_p} \times \dot{H}^{s_p-1}(\mathbb{R}^3)$ is invariant under the scaling (1-2). We will often use the shorthand notation

$$\dot{\mathcal{H}}^s := \dot{H}^s \times \dot{H}^{s-1}(\mathbb{R}^3).$$

The power-type wave equation on $\mathbb{R}^{1+3}_{t,x}$ has been extensively studied. In the defocusing setting, the positivity of the conserved energy can be used to extend a local existence result to a global one for sufficiently regular initial data. Jörgens [1961] showed global existence for the defocusing equation for smooth compactly supported data. Strauss [1968] proved global existence for smooth solutions and moreover that these solutions decay in time and scatter to free waves — this remarkable paper was the first work that proved scattering for *any* nonlinear wave equation. There are many works extending the local well-posedness theorem of Lindblad and Sogge [1995] in $\dot{\mathcal{H}}^s$ for $s > s_p$ to an unconditional global well-posedness statement and we refer the reader to [Kenig et al. 2000; Gallagher and Planchon 2003; Bahouri and Chemin 2006; Roy 2009]. These works do not address global dynamics of the solution, in particular scattering. In the radial setting the first author has made significant advances in this direction, proving in [Dodson 2018b; 2019] an unconditional global well-posedness and scattering result for the defocusing cubic equation for data in a Besov space with the same scaling as $\dot{\mathcal{H}}^{1/2}$. In very recent work Dodson [2018a] has proved unconditional scattering for the defocusing equation for radial data in the critical Sobolev space in the entire range $3 \le p < 5$.

The goal of this paper is to address global dynamics for (1-1) in the nonradial setting. Our main result is the following theorem.

Theorem 1.1 (main theorem). *Consider (1-1) for energy subcritical exponents $3 < p < 5$ and with the* ***defocusing*** *sign. Let $\vec{u}(t) \in \dot{\mathcal{H}}^{s_p}(\mathbb{R}^3)$ be a solution to (1-1) on its maximal interval of existence $I_{\max}$. Suppose that*

$$\sup_{t \in I_{\max}} \|\vec{u}(t)\|_{\dot{\mathcal{H}}^{s_p}(\mathbb{R}^3)} < \infty. \tag{1-3}$$

Then, $\vec{u}(t)$ is defined globally in time, i.e., $I_{\max} = \mathbb{R}$. In addition, we have

$$\|u\|_{L^{2(p-1)}_{t,x}(\mathbb{R}^{1+3})} < \infty,$$

which implies that $\vec{u}(t)$ scatters to a free wave in both time directions; i.e., there exist solutions $\vec{v}^{\pm}_L(t) \in \dot{\mathcal{H}}^{s_p}(\mathbb{R}^3)$ to the free wave equation, $\Box v^{\pm}_L = 0$, so that

$$\|\vec{u}(t) - \vec{v}^{\pm}_L(t)\|_{\dot{\mathcal{H}}^{s_p}(\mathbb{R}^3)} \to 0 \quad \text{as } t \to \pm\infty.$$

A version of Theorem 1.1 restricted to radially symmetric data was established in [Shen 2013]; see also [Dodson and Lawrie 2015b] for the cubic power. This type of conditional scattering result first appeared in [Kenig and Merle 2010] in the setting of the 3-dimensional cubic NLS and has since attracted a great deal of research activity; see, e.g., [Kenig and Merle 2011a; 2011b; Killip and Visan 2011a; 2011b; Bulut 2012a; 2012b; Dodson and Lawrie 2015a; Rodriguez 2017; Duyckaerts et al. 2014; Duyckaerts and Roy 2017] for this type of result for the nonlinear wave equation.

In the energy-critical regime, the bound (1-3) is guaranteed by energy conservation, and the analogue of Theorem 1.1 was proved in the seminal works [Shatah and Struwe 1993; 1994; Bahouri and Shatah 1998; Bahouri and Gérard 1999]. In the energy-supercritical regime, the analogue of Theorem 1.1 was obtained in [Killip and Visan 2011a].

The regime treated in this work, namely energy-subcritical with nonradial data, necessitates several new technical developments, which may prove useful in contexts beyond the scope of Theorem 1.1.

Remark 1.2. It is conjectured that for the defocusing equation all solutions with data in $\dot{\mathcal{H}}^{s_p}$ scatter in both time directions as in the energy-critical case $p = 5$. Theorem 1.1 is a conditional result; specifically we do not determine a priori which data satisfy (1-3). It is perhaps useful to think of the theorem in its contrapositive formulation: if initial data in the critical space $\dot{\mathcal{H}}^{s_p}$ were to lead to an evolution that does not scatter in forward time, then the $\dot{\mathcal{H}}^{s_p}$ norm of the solution must diverge along at least one sequence of times tending to the maximal forward time of existence.

Remark 1.3. The dynamics are much different in the case of the energy subcritical *focusing* equation. In remarkable works, Merle and Zaag [2003; 2005] classified the blow up dynamics by showing that all blow-up solutions must develop the singularity *at the self-similar rate*. In the radial case, an infinite family of smooth self-similar solutions is constructed in [Bizoń et al. 2010]. Donninger and Schörkhuber [2012; 2017] address the stability of the self-similar blow up.

1A. *Comments about the proof.* The proof of Theorem 1.1 follows the fundamental concentration compactness/rigidity method which first appeared in [Kenig and Merle 2006; 2008]. The proof is by contradiction — if Theorem 1.1 were to fail, the profile decomposition of [Bahouri and Gérard 1999] would yield a minimal nontrivial solution to (1-1), referred to as a *critical element* and denoted by $\vec{u}_c$, that does not scatter. Here "minimal" refers to the size of the norm in (1-3). This standard construction is outlined in Section 3. The key feature of a critical element is that its trajectory is precompact modulo symmetries in the space $\dot{\mathcal{H}}^{s_p}$; see Proposition 3.3. The proof is completed by showing that this compactness property is too rigid for a nontrivial solution and thus the critical element cannot exist.

The major obstacle to rule out a critical element $\vec{u}_c(t)$ in this energy subcritical setting is the fact that $\vec{u}_c(t)$ is a priori at best an $\dot{\mathcal{H}}^{s_p}$ solution, while all known global monotonicity formulae, e.g., the conserved energy, virial and Morawetz-type inequalities require more regularity. In general, solutions to a semilinear wave equation are only as regular as their initial data because of the free propagator $S(t)$ in the Duhamel representation of a solution

$$\vec{u}_c(t_0) = S(t_0 - t)\vec{u}_c(t) + \int_t^{t_0} S(t_0 - \tau)(0, \pm|u|^{p-1}u(\tau))\,d\tau. \tag{1-4}$$

However, for a *critical element* the precompactness of its trajectory is at odds with the dispersion of the free part, $S(t_0 - t)\vec{u}(t)$, which means the first term on the right-hand-side above must vanish weakly as $t \to \sup I_{\max}$ or as $t \to \inf I_{\max}$, where $I_{\max}$ is as in Theorem 1.1. Thus, the Duhamel integral on the right-hand-side of (1-4) encodes the regularity of a critical element and additional regularity can be expected due to the nonlinearity. As in [Dodson and Lawrie 2015b] the gain in regularity at a fixed time t_0 is observed via the so-called "double Duhamel trick", which refers to the analysis of the pairing

$$\left\langle \int_{T_1}^{t_0} S(t_0-t)(0, \pm|u|^{p-1}u)\,\mathrm{d}t, \int_{t_0}^{T_2} S(t_0-\tau)(0, \pm|u|^{p-1}u)\,\mathrm{d}\tau \right\rangle, \tag{1-5}$$

where we take $T_1 < t_0$ and $T_2 > t_0$. The basic outline of this technique was introduced in [Tao 2007] and was used within the Kenig–Merle framework for nonlinear Schrödinger equations in [Killip and Visan 2010a; 2010b; 2013], and for nonlinear wave equations in, e.g., [Killip and Visan 2011a; Shen 2013]. This method is also closely related to the in/out decomposition used by Killip, Tao, and Visan [Killip et al. 2009, Section 6].

Here we employ several novel interpretations of the double Duhamel trick, substantially building on the simple implementation developed by the first two authors in the radial setting in [Dodson and Lawrie 2015a; 2015b] for $p = 3$, which exploited the sharp Huygens principle to overcome the difficulties arising from the both the slow $\langle t\rangle^{-1}$ decay of $S(t)$ in dimension 3 and the small power $p = 3$ that precluded this case from being treated by techniques introduced in earlier works. The general case (nonradial data) considered here requires several new ideas.

We briefly describe the set-up and several key components of the proof. A critical element has compact trajectory up to action by one-parameter families (indexed by $t \in I_{\max}(\vec{u}_c)$) of translations $x(t)$ that mark the *spatial center* of the bulk of $\vec{u}_c(t)$, and rescalings $N(t)$ that record the *frequency scale* at which $\vec{u}_c(t)$ is concentrated. In Section 3 we perform a reduction to four distinct behaviors of the parameters $x(t)$ and $N(t)$. First, following the language of [Killip and Visan 2011a] we distinguish between $x(t)$ that are *subluminal*, roughly that $|x(t)-x(\tau)| \le (1-\delta)|t-\tau|$ for some $\delta > 0$, and those that *fail to be subluminal*, i.e., if $x(t)$ forever moves at the speed of light, or more precisely, $|x(t)| \simeq |t|$ (in a certain sense) for all t. The latter case is quite delicate in this energy-subcritical setting and we introduce several new ideas to treat it; see Section 7. We elaborate further on these two cases.

Subluminal critical elements. When $x(t)$ is subluminal, we distinguish between what we call a *soliton-like critical element* where $N(t) = 1$, a *self-similar-like critical element* where $N(t) = t^{-1}$, $t > 0$, and a *global concentrating critical element* where $\limsup_{t\to\infty} N(t) = \infty$. These distinct cases are treated in Sections 4, 5, and 6 respectively.

In Section 4, we set out to show as in [Dodson and Lawrie 2015b] that soliton-like critical elements must be uniformly bounded in $\dot{\mathcal{H}}^{1+\epsilon} \cap \dot{\mathcal{H}}^{s_p}$ and hence the trajectory is precompact in $\dot{\mathcal{H}}^1$. Once this is accomplished we can access nonlinear monotonicity formulae to show that such critical elements cannot exist. In this latter step we employ a version of a standard argument based on virial identity, after shifting the spatial center of the solution to $x = 0$ by the Lorentz group, which is compactified by the bound in $\dot{\mathcal{H}}^1$. The heart of the argument in Section 4 is thus establishing the additional regularity of a soliton-like critical element. The goal, roughly, is to show that the pairing (1-5) can be estimated in $\dot{\mathcal{H}}^1$. In [Dodson and Lawrie 2015b] the proof relied crucially on radial Sobolev embedding. As this is no longer at our disposal in the current, nonradial setting, we have introduced a substantial reworking of the argument from [Dodson and Lawrie 2015b] that both simplifies it and removes the reliance on radial Sobolev embedding. Examining the pairing (1-5) at time $t_0 = 0$ we divide space-time into three types of regions; see Figure 1. The first region is a fixed time interval of the form $[t_0 - R, t_0 + R]$, where $R > 0$ is chosen so that the bulk of $\vec{u}_c(t)$ is captured by the light cone emanating from $(t_0, 0)$ in both time directions. In this region (1-5) is estimated using an argument based on Strichartz estimates, using crucially that $R > 0$ is finite and can be chosen independently of t_0 by compactness. The second region is the region of space-time exterior to this time interval and exterior to the cone. Here the $\dot{\mathcal{H}}^{s_p}$ norm of the solution is small on any fixed time slice and hence an argument based on the small-data theory can be used to absorb the time integrations in (1-5). Lastly, the heart of the double Duhamel trick is employed to note the interaction between the two regions in the interior of the light cone, one for times $t < -R$ and the other for times $t > R$ is identically $= 0$ by the sharp Huygens principle!

In Section 5 we show that a self-similar-like critical element cannot exist. Here we again use a double Duhamel argument centered at $t_0 \in (0, \infty)$, but with $T_1 = \inf I_{\max} = 0$ and $T_2 = \sup I_{\max} = \infty$ in (1-5). The argument-exploiting Huygens principle given in Section 4 no longer applies since the forward and backwards cones emanating from time, say, $t_0 = 1$ can never capture the bulk of the solution since $N(T) = T^{-1}$ is an expression of the fact that the solution is localized to the physical scale T at time T; see Remark 3.6. However, here we use a different argument based on a version of the long-time Strichartz estimates introduced in [Dodson 2012; 2016], which allow us to control Strichartz norms of the projection of $\vec{u}_c$ to high frequencies $k \gg 1$ on time intervals J which are long in the sense that $|J| \simeq 2^{\alpha k}$ for $\alpha \geq 1$.

In Section 6, $N(t)$ is no longer a given fixed function. We establish a dichotomy which we refer to colloquially as the sword or the shield: either additional regularity for the critical element can be established using essentially the same argument used in Section 4A, or a self-similar-like critical element can be extracted by passing to a suitable limit. To apply the argument from Section 4 the following must be true — fixing any time t_0, the amount of time (*but where now time is measured relative to the scale* $N(t)$) that one has to wait until the bulk of the solution is absorbed by the cone emanating from time t_0 must be uniform in t_0. We define functions $C_{\pm}(t_0)$ whose boundedness (or unboundedness) measures whether or not this criteria is satisfied; see the introduction to Section 6. The rest of the section is devoted to showing how to apply the arguments from Section 4 in the case where $C_{\pm}(t_0)$ are uniformly bounded, and how to extract a self-similar solution-like critical element in the case that one of $C_{\pm}(t_0)$ are not bounded.

Critical elements that are not subluminal. In Section 7 we show that critical elements with spatial center $x(t)$ traveling at the speed of light cannot exist. The technique in this section is novel and may be useful in other settings. First we note that such critical elements are easily ruled out for solutions with finite energy, as is shown in [Kenig and Merle 2008; Tao 2008a; 2008b; 2008c; 2009a; 2009b; Nakanishi and Schlag 2011] using an argument based on the conserved momentum, and even in the energy-supercritical setting; see [Killip and Visan 2011a] using the energy/flux identity. None of these techniques (which provide an a priori limit on the speed of $x(t)$) apply in our setting so we must rule out this critical element by other means, namely, *by first showing that such critical elements have additional regularity*.

In Section 3A we lay the necessary groundwork and show, using finite speed of propagation, that any such critical element must have a fixed scale; i.e., $N(t) = 1$ and $x(t)$ must choose a fixed preferred direction up to deviation in angle by $1/\sqrt{t}$. The model case one should consider is $x(t) = (t, 0, 0)$ for all $t \in \mathbb{R}$, which means that the bulk of $\vec{u}_c(t)$ travels along the x_1-axis at speed t. We are able to show that such critical elements have up to $1-\nu$ derivatives in the x_2- and x_3-directions for any $\nu > 0$. This is enough to show that such critical elements cannot exist via a Morawetz estimate adapted to the direction of $x(t)$ — this is the only place in the paper where the arguments are limited to the defocusing equation.

The technical heart of this section is the proof of extra regularity ($1-\nu$ derivatives) in the x_2- and x_3-directions. We again divide space-time into three regions. For a solution projected to a fixed frequency $N \gg 1$, we call region $\boldsymbol{A}$ the strip $[0, N^{1-\epsilon}] \times \mathbb{R}^3$ for $\epsilon > 0$ sufficiently small relative to ν. On this region we can control the solution by a version of the long-time Strichartz estimates proved in Section 7A. At time $t = N^{1-\epsilon}$ we then divide the remaining part of space-time for positive times into two regions. Region $\boldsymbol{B}$ is the set including all times $t \geq N^{1-\epsilon}$ exterior to the light cone of initial width $R(\eta_0)$ emanating from the point $(t, x) = (N^{1-\epsilon}, x(N^{1-\epsilon}))$ where $R(\eta_0)$ is chosen large enough so that $\vec{u}_v(N^{1-\epsilon})$ has $\dot{\mathcal{H}}^{s_p}$ norm less than η_0 exterior to the ball of radius $R(\eta_0)$ centered at $x(N^{1-\epsilon})$. The solution is then controlled on region $\boldsymbol{B}$ using small-data theory. Estimating the interaction of the two terms in the pair (1-5) on the remaining region $\boldsymbol{C}$ (the region $\{|x - x(N^{1-\epsilon})| \leq R(\eta_0) + t - N^{1-\epsilon}, t \geq N^{1-\epsilon}\}$) and the analogous region $\boldsymbol{C}'$ for negative times $\tau \leq -N^{1-\epsilon}$ provides the most delicate challenge. Any naive implementation of the double Duhamel trick based on Huygens principle is doomed to fail here since the left- and right-hand components of the pair (1-5) restricted to $\boldsymbol{C}$, $\boldsymbol{C}'$ interact *in the wave zone* $|x| \simeq |t|$. Furthermore, since we are in dimension $d = 3$, the $\langle t \rangle^{-1}$ decay from the wave propagator $S(t)$ in (1-5) is not sufficient for integration in time. For this reason we introduce an auxiliary frequency localization to frequencies $|(\xi_2, \xi_3)| \simeq M$ in the ξ_2- and ξ_3-directions after first localizing in all directions to frequencies $|\xi| \simeq N$. We call this angular frequency localization $\widehat{P}_{N,M}$. The key observation is that the intersection of the wave zone $\{|x| \simeq |t|\}$ with region $\boldsymbol{C}$ requires the spatial variable $x = (x_1, x_{2,3})$ to satisfy

$$\frac{|x_{2,3}|}{|x|} \ll \frac{M}{N}$$

for all $M \geq N^{s_p/(1-\nu)}$ as long as $\epsilon > 0$ is chosen small enough relative to ν, whereas application of $\widehat{P}_{N,M}$ restricts to frequencies $\xi = (\xi_1, \xi_{2,3})$ with

$$\frac{|\xi_{2,3}|}{|\xi|} \simeq \frac{M}{N}.$$

This yields *angular separation* in the kernel of $\widehat{P}_{N,M} S(t)$ and allows us to deduce arbitrary time decay for the worst interactions in (1-5); see Lemma 7.11. The remaining interactions in (1-5) are dealt with using an argument based on the sharp Huygens principle, which is complicated due to the blurring of supports caused by $\widehat{P}_{N,M}$.

Remark 1.4. The proof of Theorem 1.1 serves as the foundation for the more complicated case of the cubic equation, $p = 3$, as well as for the analogous result for the focusing equation; see for example [Dodson and Lawrie 2015b], where the focusing and defocusing equations are treated in the same framework in the radial setting.

Much of the argument given here carries over to the defocusing equation when $p = 3$. However, in this case we have $s_p = \frac{1}{2}$ and the critical space $\dot{\mathcal{H}}^{1/2}$ is the unique Sobolev space that is invariant under Lorentz transforms. This introduces several additional difficulties, described more in detail in Remark 3.14. Additionally, certain estimates in Section 7 fail at the $p = 3$ endpoint and would require modification.

Similarly the argument in Sections 4–6 applies equally well to the focusing equation. However the argument in Section 7 used to rule out the traveling-wave critical element is specific to the defocusing equation as it relies on a Morawetz-type estimate only valid in that setting.

2. Preliminaries

2A. ***Notation, definitions, inequalities.*** We write $A \lesssim B$ or $B \gtrsim A$ to denote $A \le CB$ for some $C > 0$. Dependence of implicit constants will be denoted with subscripts. If $A \lesssim B \lesssim A$, we write $A \simeq B$. We will use the notation $a\pm$ to denote the quantity $a \pm \epsilon$ for some sufficiently small $\epsilon > 0$.

We will denote by P_N the Littlewood–Paley projections onto frequencies of size $|\xi| \simeq N$ and by $P_{\le N}$ the projections onto frequencies of size $|\xi| \lesssim N$. Often we will consider the case when $N = 2^k$, $k \in \mathbb{Z}$, is a dyadic number and in this case we will employ the following notation: when write P_k with a *lowercase* subscript k this will mean projection onto frequencies $|\xi| \simeq 2^k$. We will often write u_N for $P_N u$, and similarly for $P_{\le N}$, $P_{>N}$, P_k, and so on.

These projections satisfy Bernstein's inequalities, which we state here.

Lemma 2.1 (Bernstein's inequalities [Tao 2006, Appendix A]). *Let* $1 \le p \le q \le \infty$ *and* $s \ge 0$. *Let* $f : \mathbb{R}^d \to \mathbb{R}$. *Then*

$$\|P_{\ge N} f\|_{L^p} \lesssim N^{-s} \||\nabla|^s P_{\ge N} f\|_{L^p},$$
$$\|P_{\le N} |\nabla|^s f\|_{L^p} \lesssim N^s \|P_{\le N} f\|_{L^p}, \quad \|P_N |\nabla|^{\pm s} f\|_{L^p} \simeq N^{\pm s} \|P_N f\|_{L^p},$$
$$\|P_{\le N} f\|_{L^q} \lesssim N^{\frac{d}{p}-\frac{d}{q}} \|P_{\le N} f\|_{L^p}, \quad \|P_N f\|_{L^q} \lesssim N^{\frac{d}{p}-\frac{d}{q}} \|P_N f\|_{L^p}.$$

We will write either

$$\|u\|_{L_t^q L_x^r(I\times\mathbb{R}^3)} \quad \text{or} \quad \|u\|_{L_t^q(I;L_x^r(\mathbb{R}^3))}$$

to denote the space-time norm

$$\left(\int_I \left(\int_{\mathbb{R}^3} |u(t,x)|^{q,r}\, dx\right)^{\frac{q}{r}} dt\right)^{\frac{1}{q}},$$

with the usual modifications if q or r equals infinity.

Given $s \in \mathbb{R}$ we define the space $\dot{\mathcal{H}}^s$ by

$$\dot{\mathcal{H}}^s = \dot{H}^s(\mathbb{R}^3) \times \dot{H}^{s-1}(\mathbb{R}^3).$$

For example, we work with initial data in $\dot{\mathcal{H}}^{s_p}$.

We also require the notion of a frequency envelope.

Definition 2.2 [Tao 2001, Definition 1]. A *frequency envelope* is a sequence $\beta = \{\beta_k\}$ of positive numbers with $\beta \in \ell^2$ satisfying the local constancy condition

$$2^{-\sigma|j-k|}\beta_k \lesssim \beta_j \lesssim 2^{\sigma|j-k|}\beta_k,$$

where $\sigma > 0$ is a small, fixed constant. If β is a frequency envelope and $(f, g) \in \dot{H}^s \times \dot{H}^{s-1}$ then we say that (f, g) *lies underneath* β if

$$\|(P_k f, P_k g)\|_{\dot{H}^s \times \dot{H}^{s-1}} \le \beta_k \quad \text{for all } k \in \mathbb{Z}.$$

Note that if (f, g) lies underneath β then we have

$$\|(f, g)\|_{\dot{H}^s \times \dot{H}^{s-1}} \lesssim \|\beta\|_{\ell^2(\mathbb{Z})}.$$

In practice, we will need to choose the parameter σ in the definition of frequency envelope sufficiently small depending on the power p of the nonlinearity.

We next record a commutator estimate.

Lemma 2.3. *Let χ_R be a smooth cutoff to $|x| \ge R$. For $0 \le s \le 1$ and $N \ge 1$,*

$$\|P_N \chi_R f - \chi_R P_N f\|_{L^2} \lesssim N^{-s}(NR)^{-(1-s)}\|f\|_{\dot{H}^s},$$
$$\|P_N \chi_R f - \chi_R P_N f\|_{L^2} \lesssim R^{-s}(NR)^{-(1-s)}\|f\|_{\dot{H}^{-s}}.$$

Proof. We write the commutator as an integral operator in the form

$$[P_N \chi_R f - \chi_R P_N f](x) = N^d \int K(N(x-y))[\chi_R(x) - \chi_R(y)] f(y)\, dy.$$

Thus, using the pointwise bound

$$|\chi_R(x) - \chi_R(y)| \lesssim N|x-y| \cdot N^{-1}R^{-1}$$

and Schur's test, we first find

$$\|P_N \chi_R f - \chi_R P_N f\|_{L^2} \lesssim N^{-1}R^{-1}\|f\|_{L^2}.$$

Next, a crude estimate via the triangle inequality, Bernstein's inequality, Hölder's inequality, and Sobolev embedding gives

$$\|P_N \chi_R f - \chi_R P_N f\|_{L^2} \lesssim N^{-1}\|\nabla(\chi_R f)\|_{L^2} + N^{-1}\|\nabla f\|_{L^2} \lesssim N^{-1}\|f\|_{\dot{H}^1}.$$

The first bound now follows from interpolation. For the second bound, we write

$$[P_N \chi_R f - \chi_R P_N f](x) = N^d \int K(N(x-y))[\chi_R(x) - \chi_R(y)]\nabla \cdot \nabla \Delta^{-1} f(y)\, dy$$

and integrate by parts. Estimating as above via Schur's test, we deduce

$$\|P_N \chi_R f - \chi_R P_N f\|_{L^2_x} \lesssim R^{-1}\||\nabla|^{-1} f\|_{L^2},$$

so that the second bound also follows from interpolation. ☐

2B. ***Strichartz estimates.*** The main ingredients for the small-data theory are Strichartz estimates for the linear wave equation in $\mathbb{R}^{1+3}$,

$$\begin{aligned} \Box v &= F, \\ \vec{v}(0) &= (v_0, v_1). \end{aligned} \tag{2-1}$$

A free wave means a solution to (2-1) with $F = 0$ and will be often denoted using the propagator notation $\vec{v}(t) = S(t)\vec{v}(0)$. We define a pair (r, q) to be wave-admissible in three dimensions if

$$q, r \geq 2, \quad \frac{1}{q} + \frac{1}{r} \leq \frac{1}{2}, \quad (q, r) \neq (2, \infty).$$

The Strichartz estimates stated below are standard and we refer to [Keel and Tao 1998; Lindblad and Sogge 1995; Sogge 2008].

Proposition 2.4 (Strichartz estimates [Keel and Tao 1998; Lindblad and Sogge 1995; Sogge 2008]). *Let $\vec{v}(t)$ solve (2-1) with data $\vec{v}(0) \in \dot{H}^s \times \dot{H}^{s-1}(\mathbb{R}^3)$, with $s > 0$. Let (q, r), and (a, b) be admissible pairs satisfying the gap condition*

$$\frac{1}{q} + \frac{3}{r} = \frac{1}{a'} + \frac{3}{b'} - 2 = \frac{3}{2} - s,$$

where (a', b') are the conjugate exponents of (a, b). Then, for any time interval $I \ni 0$ we have the bounds

$$\|v\|_{L_t^q(I; L_x^r)} \lesssim \|\vec{v}(0)\|_{\dot{H}^s \times \dot{H}^{s-1}} + \|F\|_{L_t^{a'}(I; L_x^{b'})}.$$

2C. ***Small data theory*: *global existence, scattering, perturbative theory.*** A standard argument based on Proposition 2.4 yields the scaling-critical small-data well-posedness and scattering theory. We define the following notation for a collection of function spaces that we will make extensive use of. In this subsection we fix $p \in [3, 5]$ (later we will fix $p \in (3, 5)$) and let $I \subset \mathbb{R}$ be a time interval. We define

$$S(I) := L_t^{2(p-1)}(I; L_x^{2(p-1)}(\mathbb{R}^3)).$$

For example, when $p = 3$, we have $S = L_{t,x}^4$, while for $p = 5$ we have $S = L_{t,x}^8$.

Remark 2.5. There are a few other function spaces related to

$$\dot{\mathcal{H}}^{s_p} := \dot{H}^{s_p} \times \dot{H}^{s_p - 1}(\mathbb{R}^3)$$

that will appear repeatedly in our analysis. First note the Sobolev embedding $\dot{H}^{s_p}(\mathbb{R}^3) \hookrightarrow L^{(3/2)(p-1)}(\mathbb{R}^3)$, which means

$$\|f\|_{L^{(3/2)(p-1)}(\mathbb{R}^3)} \lesssim \|f\|_{\dot{H}^{s_p}(\mathbb{R}^3)}.$$

Proposition 2.6 (small-data theory). *Let $3 \leq p < 5$ and suppose that $\vec{u}(0) = (u_0, u_1) \in \dot{H}^{s_p} \times \dot{H}^{s_p-1}(\mathbb{R}^3)$. Then there is a unique solution $\vec{u}(t) \in \dot{\mathcal{H}}^{s_p}$ with maximal interval of existence $I_{\max}(\vec{u}) = (T_-(\vec{u}), T_+(\vec{u}))$. Moreover, for any compact interval $J \subset I_{\max}$,*

$$\|u\|_{S(J)} < \infty.$$

Additionally, a globally defined solution $\vec{u}(t)$ on $t \in [0, \infty)$ scatters as $t \to \infty$ to a free wave if and only if $\|u\|_{S([0,\infty))} < \infty$. In particular, there exists a constant $\delta_0 > 0$ so that

$$\|\vec{u}(0)\|_{\dot{H}^{s_p} \times \dot{H}^{s_p - 1}} < \delta_0 \quad \Longrightarrow \quad \|u\|_{S(\mathbb{R})} \lesssim \|\vec{u}(0)\|_{\dot{H}^{s_p} \times \dot{H}^{s_p - 1}} \lesssim \delta_0$$

and thus $\vec{u}(t)$ scatters to free waves as $t \to \pm\infty$. Finally, we have the standard finite time blow-up criterion:

$$T_+(\vec{u}) < \infty \quad \Longrightarrow \quad \|u\|_{S([0, T_+(\vec{u}))} = +\infty.$$

An analogous statement holds if $-\infty < T_-(\vec{u})$.

The concentration compactness procedure in Section 3 requires the following nonlinear perturbation lemma for approximate solutions to (1-1).

Lemma 2.7 (perturbation lemma [Kenig and Merle 2006; 2008]). *There exist continuous functions $\epsilon_0, C_0 : (0, \infty) \to (0, \infty)$ so that the following holds true. Let $I \subset \mathbb{R}$ be an open interval (possibly unbounded) and $\vec{u}, \vec{v} \in C(I; \dot{H}^{s_p} \times \dot{H}^{s_p - 1})$ satisfy for some $A > 0$*

$$\|\vec{v}\|_{L^\infty(I; \dot{H}^{s_p} \times \dot{H}^{s_p - 1})} + \|v\|_{S(I)} \le A,$$

$$\||\nabla|^{s_p - \frac{1}{2}} \mathrm{eq}(u)\|_{L_t^{4/3}(I; L_x^{4/3})} + \||\nabla|^{s_p - \frac{1}{2}} \mathrm{eq}(v)\|_{L_t^{4/3}(I; L_x^{4/3})} + \|w_0\|_{S(I)} \le \epsilon \le \epsilon_0(A),$$

where $\mathrm{eq}(u) := \Box u \pm |u|^{p-1} u$ in the sense of distributions, and $\vec{w}_0(t) := S(t - t_0)(\vec{u} - \vec{v})(t_0)$ with $t_0 \in I$ fixed, but arbitrary. Then

$$\|\vec{u} - \vec{v} - \vec{w}_0\|_{L^\infty(I; \dot{H}^{s_p} \times \dot{H}^{s_p - 1})} + \|u - v\|_{S(I)} \le C_0(A)\epsilon.$$

In particular, $\|u\|_{S(I)} < \infty$.

3. Concentration compactness and the reduction of Theorem 1.1

We begin the proof of Theorem 1.1 using the concentration compactness and rigidity method of [Kenig and Merle 2006; 2008]. The concentration compactness aspect of the argument is by now standard and we follow the scheme from [Kenig and Merle 2010], which is a refinement of the scheme in [Kenig and Merle 2006; 2008]. The main conclusion of this section is the following: if Theorem 1.1 fails, there exists a minimal, nontrivial, nonscattering solution to (1-1), which we call a *critical element*.

We follow the notation from [Kenig and Merle 2010] for convenience. Given initial data $(u_0, u_1) \in \dot{H}^{s_p} \times \dot{H}^{s_p - 1}$, we let $\vec{u}(t) \in \dot{H}^{s_p} \times \dot{H}^{s_p - 1}$ be the unique solution to (1-1) with data $\vec{u}(0) = (u_0, u_1)$ and maximal interval of existence $I_{\max}(\vec{u}) := (T_-(\vec{u}), T_+(\vec{u}))$.

Given $A > 0$, set

$$\mathcal{B}(A) := \{(u_0, u_1) \in \dot{H}^{s_p} \times \dot{H}^{s_p - 1} : \|\vec{u}(t)\|_{L_t^\infty(I_{\max}(\vec{u}); \dot{H}^{s_p} \times \dot{H}^{s_p - 1})} \le A\}.$$

Definition 3.1. We say that $\mathcal{SC}(A; \vec{u}(0))$ holds if $\vec{u}(0) \in \mathcal{B}(A)$, $I_{\max}(\vec{u}) = \mathbb{R}$ and $\|u\|_{S(I)} < \infty$. In addition, we will say that $\mathcal{SC}(A)$ holds if, for every $(u_0, u_1) \in \mathcal{B}(A)$, one has $I_{\max}(\vec{u}) = \mathbb{R}$ and $\|u\|_{S(I)} < \infty$.

Remark 3.2. Recall from Proposition 2.6 that $\|u\|_{S(I)} < \infty$ if and only if $\vec{u}(t)$ scatters to a free waves as $t \to \pm\infty$. Thus, Theorem 1.1 is equivalent to the statement that $\mathcal{SC}(A)$ holds for all $A > 0$.

Now suppose that Theorem 1.1 *fails to be true*. By Proposition 2.6, there exists an $A_0 > 0$ small enough so that $\mathcal{SC}(A_0)$ holds. Since we are assuming that Theorem 1.1 fails, we can find a threshold value A_C so that for $A < A_C$, $\mathcal{SC}(A)$ holds, and for $A > A_C$, $\mathcal{SC}(A)$ fails. Note that we must have $0 < A_0 < A_C$. The Kenig–Merle concentration compactness argument is now used to produce a *critical element*, namely a minimal nonscattering solution $\vec{u}_c(t)$ to (1-1) so that $\mathcal{SC}(A_C, \vec{u}_c)$ fails, and which enjoys certain compactness properties.

We state a refined version of this result below, and we refer the reader to [Kenig and Merle 2010; Shen 2013; Tao et al. 2007; 2008] for the details. As usual, the deep foundations of the concentration compactness part of the Kenig–Merle framework are profile decompositions of [Bahouri and Gérard 1999] used in conjunction with the nonlinear perturbation theory in Lemma 2.7.

Proposition 3.3. *Suppose Theorem 1.1 fails to be true. Then, there exists a solution $\vec{u}(t)$ such that $\mathcal{SC}(A_C; \vec{u})$ fails, which we call a* ***critical element****. We can assume that $\vec{u}(t)$ does not scatter in either time direction, i.e.,*

$$\|u\|_{S((T_-(\vec{u}),0])} = \|u\|_{S([0,T_+(\vec{u}))} = \infty,$$

and moreover, there exist continuous functions

$$N : I_{\max}(\vec{u}) \to (0,\infty), \quad x : I_{\max}(\vec{u}) \to \mathbb{R}^3$$

so that the set

$$\left\{ \left(\frac{1}{N(t)^{\frac{2}{p-1}}} u\left(t, x(t) + \frac{\cdot}{N(t)}\right), \frac{1}{N(t)^{\frac{2}{p-1}+1}} u_t\left(t, x(t) + \frac{\cdot}{N(t)}\right)\right) : t \in I_{\max} \right\} \tag{3-1}$$

is precompact in $\dot{\mathcal{H}}^{s_p}$.

We make a few observations and reductions concerning the critical element found in Proposition 3.3. It will be convenient to proceed slightly more generally, starting by giving a name to the compactness property (3-1) satisfied by a critical element.

Definition 3.4. Let $I \ni 0$ be an interval and let $\vec{u}(t)$ be a nonzero solution to (1-1) on I. We will say $\vec{u}(t)$ has the *compactness property on I* if there are continuous functions $N : I \to (0,\infty)$ and $x : I \to \mathbb{R}^3$ so that the set

$$K_I := \left\{ \left(\frac{1}{N^{\frac{2}{p-1}}(t)} u\left(t, x(t) + \frac{\cdot}{N(t)}\right), \frac{1}{N^{\frac{2}{p-1}+1}(t)} u_t\left(t, x(t) + \frac{\cdot}{N(t)}\right)\right) : t \in I \right\}$$

is precompact in $\dot{\mathcal{H}}^{s_p}$.

We make the following standard remarks about solutions with the compactness property. We begin with a local constancy property for the modulation parameters.

Lemma 3.5 [Killip and Visan 2013, Lemma 5.18]. *Let $\vec{u}(t)$ have the compactness property on a time interval $I \subset \mathbb{R}$ with parameters $N(t)$ and $x(t)$. Then there exist constants $\epsilon_0 > 0$ and $C_0 > 0$ so that for every $t_0 \in I$ we have*

$$\left[t_0 - \frac{\epsilon_0}{N(t_0)}, t_0 + \frac{\epsilon_0}{N(t_0)}\right] \subset I,$$
$$\frac{1}{C_0} \le \frac{N(t)}{N(t_0)} \le C_0 \quad \textit{if } |t - t_0| \le \frac{\epsilon_0}{N(t_0)},$$
$$|x(t) - x(t_0)| \le \frac{C_0}{N(t_0)} \quad \textit{if } |t - t_0| \le \frac{\epsilon_0}{N(t_0)}.$$

Remark 3.6. For a solution with the *compactness property* on an interval I, we can, after modulation, control the $\dot{\mathcal{H}}^{s_p}$ tails uniformly in $t \in I$. Indeed, for any $\eta > 0$ there exists $R(\eta) < \infty$ such that

$$\int_{|x - x(t)| \ge \frac{R(\eta)}{N(t)}} ||\nabla|^{s_p} u(t,x)|^2 \,\mathrm{d}x + \int_{|\xi| \ge R(\eta)N(t)} |\xi|^{2s_p} |\hat{u}(t,\xi)|^2 \,\mathrm{d}\xi \le \eta,$$
$$\int_{|x - x(t)| \ge \frac{R(\eta)}{N(t)}} ||\nabla|^{s_p - 1} u_t(t,x)|^2 \,\mathrm{d}x + \int_{|\xi| \ge R(\eta)N(t)} |\xi|^{2(s_p-1)} |\hat{u}_t(t,\xi)|^2 \,\mathrm{d}\xi \le \eta$$

for all $t \in I$. We call $R(\cdot)$ the *compactness modulus*.

We also remark that any Strichartz norm of the linear part of the evolution of a solution with the compactness property on $I_{\max}$ vanishes as $t \to T_-$ and as $t \to T_+$. A concentration compactness argument then implies that the linear part of the evolution vanishes weakly in $\dot{\mathcal{H}}^{s_p}$, that is, for each $t_0 \in I_{\max}$,

$$S(t_0 - t)u(t) \rightharpoonup 0$$

weakly in $\dot{\mathcal{H}}^{s_p}$ as $t \nearrow \sup I$ and $t \searrow \inf I$; see [Tao et al. 2008, Section 6; Shen 2013, Proposition 3.6]. This implies the following lemma, which we use crucially in the proof of Theorem 1.1.

Lemma 3.7 [Tao et al. 2008, Section 6; Shen 2013, Proposition 3.6]. *Let $\vec{u}(t)$ be a solution to (1-1) with the compactness property on its maximal interval of existence $I = (T_-, T_+)$. Then for any $t_0 \in I$ we can write*

$$\int_{t_0}^{T} S(t_0 - s)(0, |u|^{p-1}u)\, ds \rightharpoonup \vec{u}(t_0) \quad \textit{as } T \nearrow T_+ \textit{weakly in } \dot{\mathcal{H}}^{s_p},$$
$$-\int_{T}^{t_0} S(t_0 - s)(0, |u|^{p-1}u)\, ds \rightharpoonup \vec{u}(t_0) \quad \textit{as } T \searrow T_- \textit{weakly in } \dot{\mathcal{H}}^{s_p}.$$

Remark 3.6 indicates that solutions $\vec{u}(t)$ with the compactness property have uniformly small tails in $\dot{\mathcal{H}}^{s_p}$, where "tails" are taken to be centered at $x(t)$, and relative to the frequency scale $N(t)$ at which the solutions are concentrating. We would like to use this fact to obtain lower bounds for norms of the solution $u(t)$. The immediate issue that arises is that the object that obeys compactness properties is the pair $\vec{u}(t,x) = (u(t,x), u_t(t,x))$ and, a priori, the solution could satisfy $u(t,x) = 0$ a fixed time t. Nonetheless, by averaging in time, such a lower bound still holds for the solution itself, $u(t)$. We can quantify this bound in several ways, starting with a result proved in [Killip and Visan 2011a].

Lemma 3.8 [Killip and Visan 2011a, Lemma 3.4]. *Let $\vec{u}(t)$ be a solution with the compactness property on $I_{\max} = \mathbb{R}$. Then, for any $A > 0$, there exists $\eta = \eta(A)$ such that*

$$\left|\left\{t \in \left[t_0, t_0 + \frac{A}{N(t_0)}\right] : \|u(t)\|_{L_x^{3/2(p-1)}} \geq \eta\right\}\right| \geq \frac{\eta}{N(t_0)}$$

for all $t_0 \in \mathbb{R}$.

Lemma 3.8 means that the $L_x^{(3/2)(p-1)}$ norm of $u(t)$ is nontrivial when averaged over intervals around t_0 of length comparable to $N(t_0)^{-1}$ uniformly in t_0. By combining this lemma with Remark 3.6 and Sobolev embedding we obtain the following as an immediate consequence.

Corollary 3.9 (averaged concentration around $x(t)$). *Fix any $\delta_0 > 0$. Let $\vec{u}(t)$ be a solution with the compactness property on $I_{\max} = \mathbb{R}$. There exists a constant $C > 0$ so that*

$$N(t_0) \int_{t_0}^{t_0 + \frac{\delta_0}{N(t_0)}} \int_{|x - x(t)| \leq \frac{C}{N(t)}} |u(t,x)|^{\frac{3}{2}(p-1)} \, dx \, dt \gtrsim 1$$

for all $t_0 \in \mathbb{R}$.

One can also deduce the following corollary, also proved in [Killip and Visan 2011a], which gives a lower bound on the localized S norm of $u(t)$.

Corollary 3.10 (S-norm concentration around $x(t)$). *Let $\vec{u}(t)$ be a solution with the compactness property on $I_{\max} = \mathbb{R}$. Then there exist constants $c, C > 0$ so that*

$$\int_{t_1}^{t_2} \int_{|x - x(t)| \leq \frac{C}{N(t)}} |u(t,x)|^{2(p-1)} \, dx \, dt \geq c \int_{t_1}^{t_2} N(t) \, dt$$

for any t_1, t_2 such that

$$t_2 - t_1 \geq \frac{1}{N(t_1)}.$$

Proof. The proof runs completely parallel to the argument in [Killip and Visan 2011a, proof of Corollary 3.5] given for the averaged potential energy. □

The fact that we have only averaged lower bounds on, e.g., the $L^{(3/2)(p-1)}$ norm of a critical element will not be too much trouble. We will often pair the above with the fact that the compactness parameters $N(t), x(t)$ are approximately locally constant; see Lemma 3.5.

Lastly, we also need the following estimate proved in [Dodson and Lawrie 2015b, Lemma 4.5].

Lemma 3.11 [Dodson and Lawrie 2015b, Lemma 4.5]. *Let $\vec{u}(t)$ have the compactness property on a time interval $I \subset \mathbb{R}$ with scaling parameter $N(t)$. Let $\eta > 0$. Then there exists $\delta > 0$ such that*

$$\|u\|_{L_{t,x}^{2(p-1)}([t_0 - \frac{\delta}{N(t_0)}, t_0 + \frac{\delta}{N(t_0)}] \times \mathbb{R}^3)} \leq \eta$$

uniformly in $t_0 \in I$.

3A. *Analysis of solutions with the compactness property.* In the next subsection, we will prove a classification result for solutions with the compactness property. Our goal is to gather together a list of possibilities for the compactness parameters $N(t)$ and $x(t)$ that is exhaustive in the sense that if we rule out the existence of all members of the list, then Theorem 1.1 is true. Before stating these cases, we need to distinguish between two scenarios based on how fast $x(t)$ is moving relative to the speed of light. To make this distinction precise, we have the following definition.

Definition 3.12. Let $\vec{u}(t)$ be a solution to (1-1) with the compactness property on $I = \mathbb{R}$ with parameters $x(t)$ and $N(t) \geq 1$. We will say that $x(t)$ is *subluminal* if there exists a constant $A > 1$ so that for all $t_0 \in \mathbb{R}$ there exists $t \in [t_0, t_0 + A/N(t_0)]$ such that

$$|x(t) - x(t_0)| \leq |t - t_0| - \frac{1}{AN(t_0)}.$$

Proposition 3.13. *Suppose $\vec{u}(t)$ is a solution to (1-1) with the compactness property on its maximal interval of existence $I_{\max}$ with compactness parameters $N(t)$ and $x(t)$. We can assume without loss of generality in the arguments that follow that $I_{\max}$, $N(t)$ and $x(t)$ fall into one of the following four scenarios*:

(I) <u>*Soliton-like critical element*</u>: $I_{\max} = \mathbb{R}$, $N(t) \equiv 1$ *for all* $t \in \mathbb{R}$ *and* $x(t)$ *is subluminal in the sense of Definition 3.12.*

(II) <u>*Two-sided concentrating critical element*</u>: $I_{\max} = \mathbb{R}$, $N(t) \geq 1$ *for all* $t \in \mathbb{R}$, $\limsup_{t\to\pm\infty} N(t) = \infty$, *and* $x(t)$ *is subluminal.*

(III) <u>*Self-similar-like critical element*</u>: $I_{\max} = (0, \infty)$, $N(t) = \frac{1}{t}$, *and* $x(t) \equiv 0$.

(IV) <u>*Traveling-wave critical element*</u>: $I_{\max} = \mathbb{R}$, $N(t) \equiv 1$ *for all* $t \in \mathbb{R}$ *and* $|x(t) - (t, 0, 0)| \lesssim \sqrt{|t|}$ *for all* $t \in \mathbb{R}$.

Remark 3.14. In the case $p = 3$, one must take into account the action of the Lorentz group, which will introduce additional cases to the list of critical elements in Proposition 3.13. For $p \neq 3$, the hypothesis (1-3) compactifies the action of the Lorentz group in the Bahouri–Gérard profile decomposition at regularity $\dot{\mathcal{H}}^{s_p}$, which is why only a translation $x(t)$ and scaling $N(t)$ appear in the descriptions of critical elements. However, because $\dot{\mathcal{H}}^{1/2}$ is invariant under action of the Lorentz group, one must confront critical elements with velocity $\ell(t)$ that approaches the speed of light. See [Ramos 2012; 2018] for Bahouri–Gérard-type profile decompositions in this setting.

Before proving Proposition 3.13, we note that ruling out cases (I)–(IV) in the statement of the proposition will prove our main result, Theorem 1.1. Hence we will now focus on establishing Proposition 3.13 and proving that such critical elements cannot exist.

We will prove this proposition in several steps. First, we will reduce the frequency parameter $N(t)$ to one of three possible cases. We state these reductions for $N(t)$, but we omit the proof as it follows readily from arguments similar to those in [Killip and Visan 2013, Theorem 5.25].

Proposition 3.15. *Let $\vec{u}(t)$ denote the critical element found in Proposition 3.3. Passing to subsequences, taking limits, using scaling considerations and time reversal, we can assume, without loss of generality, that $T_+(\vec{u}) = +\infty$, and that the frequency scale $N(t)$ and maximal interval of existence $I_{\max} = I_{\max}(\vec{u})$ satisfy one of the following three possibilities*:

- *Soliton-like scale*: $I_{\max} = \mathbb{R}$ *and*

$$N(t) \equiv 1 \quad \text{for all } t \in \mathbb{R}.$$

- *Doubly concentrating scale*: $I_{\max} = (-\infty, \infty)$ *and*

$$\limsup_{t \to T_-} N(t) = \infty, \quad \limsup_{t \to \infty} N(t) = \infty, \quad \text{and} \quad N(t) \geq 1 \quad \text{for all } t \in \mathbb{R}.$$

- *Self-similar scale*: $I_{\max} = (0, \infty)$ *and* $N(t) = t^{-1}$.

We will now make a few further reductions, mostly concerning the spatial center $x(t)$ of a critical element that is global in time.

We will show that in all cases where we have a solution with the compactness property with translation parameter $x(t)$ that fails to be subluminal, we may extract a traveling-wave solution. To prove this, we will need to analyze the properties of solutions with the compactness property and more specifically, properties of their spatial centers, $x(t)$. We turn to this analysis now. First, we note that in the case that $x(t)$ is subluminal (see Definition 3.12) we can derive the following consequence.

Lemma 3.16 [Killip and Visan 2011a, Proposition 4.3]. *Let $\vec{u}(t)$ be a solution to (1-1) with the compactness property on $I = \mathbb{R}$ with parameters $x(t)$ and $N(t) \geq 1$. Suppose $x(0) = 0$ and that $x(t)$ is subluminal in the sense of Definition 3.12. Then there exists a $\delta_0 > 0$ so that*

$$|x(t) - x(\tau)| \leq (1 - \delta_0)|t - \tau|$$

for all t, τ with

$$|t - \tau| \geq \frac{1}{\delta_0 N_{t,\tau}},$$

where $N_{t,\tau} := \inf_{s \in [t,\tau]} N(s)$.

Proof. See the proof of Proposition 4.3 in [Killip and Visan 2011a]. □

Using Lemma 3.5 together with Lemma 3.8 and a domain-of-dependence argument based on the finite speed of propagation, we obtain a preliminary bound on how fast $x(t)$ can grow. (See, e.g., [Killip and Visan 2011a, Proposition 4.1].)

Lemma 3.17. *Let $\vec{u}(t)$ have the compactness property on a time interval $I \subset \mathbb{R}$ with parameters $N(t)$ and $x(t)$. Then there exists a constant $C > 0$ so that for any $t_1, t_2 \in I$ we have*

$$|x(t_1) - x(t_2)| \leq |t_1 - t_2| + \frac{C}{N(t_1)} + \frac{C}{N(t_2)}. \tag{3-2}$$

In fact, if $\vec{u}(t)$ is global in time, we have $N(t)|t| \to \infty$ as $|t| \to \infty$ and we normalize so that $x(0) = 0$, from which the above yields

$$\lim_{t\to\pm\infty} \frac{|x(t)|}{|t|} \le 1. \tag{3-3}$$

Remark 3.18. We remark that by finite speed of propagation and compactness, we can assume that

$$\liminf_{t\to T_\pm(\vec{u})} |t|N(t) \in [1,\infty].$$

Note that according to the definition of the compactness property, the function $x(t)$ is not uniquely defined; indeed, one can always modify $x(t)$ up to a radius of $\mathcal{O}(N(t)^{-1})$, provided one also modifies the compactness modulus appropriately. Note, however, that the compactness property, together with monotone convergence, prevents $\vec{u}$ from concentrating on very narrow strips, as measured in units of $N(t)^{-1}$. See [Killip and Visan 2011a, Lemma 4.2].

Lemma 3.19. *Let $\vec{u}$ be a solution to (1-1) with the compactness property on an interval I. Then for any $\eta > 0$, there exists $c(\eta) > 0$ so that*

$$\sup_{\omega\in\mathbb{S}^2} \int_{|\omega\cdot[x-x(t)]|\le c(\eta)N(t)^{-1}} ||\nabla|^{s_p} u|^2 + ||\nabla|^{s_p-1} u_t|^2 \, dx < \eta.$$

To deal with ambiguity in the definition of $x(t)$, we use the notion of a "centered" spatial center as in [Killip and Visan 2011a], that is, a choice of $x(t)$ such that each plane through $x(t)$ partitions $\vec{u}(t)$ into two nontrivial pieces.

Definition 3.20. Let $\vec{u}$ be a solution to (1-1) with the compactness property on an interval I with spatial center $x(t)$. We call $x(t)$ *centered* if there exists $C(u) > 0$ such that, for all $\omega \in \mathbb{S}^2$ and $t \in I$,

$$\int_{\omega\cdot[x-x(t)]>0} ||\nabla|^{s_p} u(t,x)|^2 + ||\nabla|^{s_p-1} u_t(t,x)|^2 \, dx \ge C(u).$$

Proposition 3.21. *Let $\vec{u}$ be a global solution to (1-1) with the compactness property. Then there exists a centered spatial center for $\vec{u}$.*

Proof. The argument is similar to the proof of [Killip and Visan 2011a, Proposition 4.1]. Let $x(t)$ be any spatial center for $\vec{u}$. To shorten formulas, we introduce the notation

$$\varphi(t,x) = ||\nabla|^{s_p} u(t,x)|^2 + ||\nabla|^{s_p-1} u_t(t,x)|^2.$$

By compactness, there exists $C = C(u)$ large enough that

$$\inf_{t\in\mathbb{R}} \int_{B(t)} \varphi(t,x) \, dx \gtrsim_u 1, \quad \text{where } B(t) := \{x : |x - x(t)| \le CN(t)^{-1}\}.$$

Now set

$$\tilde{x}(t) = x(t) + \frac{\int_{B(t)} [x - x(t)]\varphi(t,x) \, dx}{\int_{B(t)} \varphi(t,x) \, dx}.$$

By definition, $|x(t)-\tilde{x}(t)| \le CN(t)^{-1}$, and hence $\tilde{x}(t)$ is a valid spatial center for $\vec{u}$ (one only needs to add C to the compactness modulus). We now claim that $\tilde{x}(t)$ is centered. To see this, first note that by construction one has

$$\int_{B(t)} \omega \cdot [x-\tilde{x}(t)]\varphi(t,x)\,\mathrm{d}x = 0.$$

On the other hand, combining nontriviality on $B(t)$ together with Lemma 3.19, we have

$$\int_{B(t)\cap|\omega\cdot[x-\tilde{x}(t)]|>cN(t)^{-1}} \varphi(t,x)\,\mathrm{d}x \gtrsim_u 1$$

for some $c = c(u) > 0$. Thus

$$\int_{B(t)} |\omega\cdot[x-\tilde{x}(t)]|\varphi(t,x)\,\mathrm{d}x \gtrsim_u N(t)^{-1},$$

and so

$$\int_{B(t)} \{\omega\cdot[x-\tilde{x}(t)]\}_+\varphi(t,x)\,\mathrm{d}x \gtrsim_u N(t)^{-1},$$

where "+" denotes the positive part. As $|x-\tilde{x}(t)| \le 2CN(t)^{-1}$ for $x \in B(t)$, we finally deduce

$$1 \lesssim_u \int_{B(t)} \frac{\{\omega\cdot[x-\tilde{x}(t)]\}_+}{2CN(t)^{-1}}\varphi(t,x)\,\mathrm{d}x \lesssim_u \int_{\omega\cdot[x-\tilde{x}(t)]>0} \varphi(t,x)\,\mathrm{d}x$$

for all $\omega \in \mathbb{S}^2$, as needed. □

Proposition 3.22. *Suppose that $\vec{u}(t)$ is a solution with the compactness property on $\mathbb{R}$ with parameters $N(t)$ and $x(t)$. Suppose in addition that $N(t) = 1$ for all $t \in \mathbb{R}$, and that $x(t)$ fails to be subluminal in the sense of Definition 3.12. Then there exists a (possibly different) solution $\vec{w}(s)$ to (1-1) with the compactness property on $\mathbb{R}$ with parameters $N(s)$ and $x(s)$ satisfying*

$$N(s) \equiv 1, \quad |x(s)-(s,0,0)| \lesssim \sqrt{|s|} \quad \textit{for all } s \in \mathbb{R}.$$

Proof. Let $\vec{u}(t)$ be a solution to (1-1) with the compactness property on $\mathbb{R}$ with parameters $N(t) \equiv 1$ and $x(t)$ failing to be subluminal. This means we can find a sequence t_m and intervals

$$I_m := [t_m, t_m+m]$$

such that

$$|x(t_m)-x(t)| \ge |t_m-t| - \frac{1}{m} \quad \text{for all } t \in I_m. \tag{3-4}$$

We construct a sequence as follows. Set

$$\vec{u}_m(0) := \vec{u}(t_m, \cdot - x(t_m)).$$

Using the precompactness of the trajectory of $\vec{u}$ modulo the translations by $x(t)$ we can (passing to a subsequence) extract a strong limit

$$\vec{u}_m(0) \to \vec{u}_\infty(0) \in \dot{\mathcal{H}}^{s_p} \quad \text{as } m \to \infty.$$

Let $\vec{u}_\infty(\tau)$ be the solution to (1-1) with initial data $\vec{u}_\infty(0)$. One can show that we must have $[0,\infty) \subset I_{\max}(\vec{u}_\infty)$ and that $\vec{u}_\infty$ satisfies the following compactness property on $[0,\infty)$: the set

$$K_\infty := \{\vec{u}_\infty(\tau, \cdot - x_\infty(\tau)) : \tau \in [0,\infty)\}$$

is precompact in $\dot{\mathcal{H}}^{s_p}(\mathbb{R}^3)$, where for each $\tau > 0$ the function $x_\infty(\tau)$ is defined by

$$x_\infty(\tau) := \lim_{m\to\infty} (x(t_m+\tau) - x(t_m)).$$

Note that for each $\epsilon > 0$ and for all $\tau \in [0,\infty)$ we can choose $M > 0$ large enough so that for all $m \geq M$ we have

$$|x(t_m+\tau) - x(t_m)| \geq |\tau| - \frac{1}{m},$$

where the last inequality follows from (3-4). Letting $m \to \infty$ above, we conclude that in fact

$$|x_\infty(\tau)| \geq \tau \quad \text{for all } \tau \in [0,\infty).$$

By finite speed of propagation (see (3-3)) we can conclude that in fact

$$\lim_{\tau\to\infty} \frac{|x_\infty(\tau)|}{\tau} = 1.$$

We now refine our solution again, this time constructing a suitable limit from $\vec{u}_\infty(\tau)$. First choose a sequence $0 < \sigma_m \to \infty$ such that, for $\tau \geq \sigma_m$, we have

$$\frac{|x_\infty(\tau)|}{\tau} \leq 1 + 2^{-m}$$

and set $\tau_m = \sigma_m + m$. Then by the previous two lines, it holds that

$$\tau \leq |x_\infty(\tau)| \leq \tau(1+2^{-m}) \quad \text{for all } \tau \in J_m := [\tau_m - m, \tau_m + m].$$

From (3-4) and the definition of x_∞ we have

$$|x_\infty(\tau) - x_\infty(t)| \geq |t-\tau| \quad \text{for all } t,\tau \in J_m. \tag{3-5}$$

As before we extract a limit from the sequence

$$\vec{u}_{\infty,m}(0) := \vec{u}_\infty(\tau_m, \cdot - x_\infty(\tau_m)) \to \vec{v}(0) \in \mathcal{H}^{s_p}$$

and we note that the solution $\vec{v}(s)$ to (1-1) with data $\vec{v}(0)$ has the compactness property on $\mathbb{R}$ with parameters $\widetilde{N}(s) \equiv 1$ and $\tilde{x}(s)$ defined by

$$\tilde{x}(s) := \lim_{m\to\infty} (x_\infty(\tau_m + s) - x_\infty(\tau_m)).$$

Using (3-5) along with (3-2) we see that for all $s_1, s_2 \in \mathbb{R}$ we have

$$|s_1 - s_2| \leq |\tilde{x}(s_1) - \tilde{x}(s_2)| \leq |s_1 - s_2| + \widetilde{C}$$

for some absolute constant $\widetilde{C} > 0$ and consequently

$$\lim_{s\to\pm\infty} \frac{|\tilde{x}(s)|}{|s|} = 1. \tag{3-6}$$

Now we express $\tilde{x}(s)$ in polar coordinates, finding $r(s) \geq 0$ and $\omega(s) \in \mathbb{S}^2$ so that

$$\tilde{x}(s) = r(s)\omega(s) \quad \text{for all } s \in [0,\infty).$$

Note that by (3-6) we have

$$\frac{r(s)}{s} \to 1 \quad \text{as } s \to \infty.$$

Since $\omega(s) \in \mathbb{S}^2$ we can find a sequence $s_m \to \infty$ and we can (up to passing to a subsequence) find a limit ω_0 so that

$$\omega(s_m) \to \omega_0 \quad \text{as } m \to \infty.$$

To prove the claim, it suffices to verify that

$$|\tilde{x}(s) - s\omega_0| \leq C\sqrt{s},$$

since then we obtain the desired result applying a fixed spatial rotation. Note that

$$|s_2\omega(s_2) - s_1\omega(s_1)|^2 = |s_1 - s_2|^2 + s_1 s_2 |\omega(s_2) - \omega(s_1)|^2.$$

By finite speed of propagation

$$|s_2\omega(s_2) - s_1\omega(s_1)|^2 \leq (|s_1 - s_2| + C)^2 = |s_1 - s_2|^2 + 2C|s_1 - s_2| + C^2,$$

and hence substituting this bound into the above equations we solve to obtain

$$|\omega(s_2) - \omega(s_1)| \leq \sqrt{\frac{2C|s_1 - s_2| + C^2}{s_1 s_2}}.$$

Then

$$\begin{aligned}|(s_n + s)\omega(s_n + s) - s_n\omega(s_n) - s\omega_0| &\leq |s_n + s|\,|\omega(s_n + s) - \omega(s_n)| + s|\omega(s_n) - \omega_0| \\ &\leq \sqrt{(2Cs + C^2)\Big(1 + \frac{s}{s_n}\Big)} + s|\omega(s_n) - \omega_0|,\end{aligned}$$

which implies

$$|\tilde{x}(s) - s\omega_0| \leq \sqrt{2Cs + C^2},$$

as required. □

In the case that $N(t) \geq 1$ and $x(t)$ is not subluminal, we will now show that we can also reduce to the case when $N(t) = 1$ for all $t \in \mathbb{R}$ and $x(t) = (t, 0, 0) + O(\sqrt{|t|})$. We will need the following lemma.

Lemma 3.23. *Let $\vec{u}(t)$ have the compactness property on $I \subset \mathbb{R}$ with parameters $N(t)$ and $x(t)$. Then there exists a constant $c \in (0,1)$ such that for any $t_1, t_2 \in I$ with $N(t_1) \leq N(t_2)$ it holds that*

$$|x(t_1) - x(t_2)| \geq |t_1 - t_2| - \frac{c}{N(t_1)} \quad \Longrightarrow \quad N(t_2) \leq \frac{1}{c^2} N(t_1).$$

Proof of Lemma 3.23. The argument adapts readily from [Killip and Visan 2011a, Lemma 4.4], using the arguments from Section 3A. Exploiting time-reversal symmetry, space-translation symmetry, and rotation symmetry, we may assume $t_1 < t_2$, $x(t_1) = 0$, and $x(t_2) = (x_1(t_2), 0, 0)$ with $x_1(t_2) \geq 0$. Further, we may choose $x(t)$ to be centered by Proposition 3.21.

Suppose for contradiction that for times t_1, t_2 as in the statement of the lemma,

$$|x(t_1) - x(t_2)| \geq |t_1 - t_2| - \frac{c}{N(t_1)}$$

but $cN(t_1)^{-1} \geq c^{-1} N(t_2)^{-1}$, where $c = c(u)$ will be chosen sufficiently small below.

Let $\psi : \mathbb{R} \to [0, \infty)$ be a cutoff so that $\psi = 1$ for $x \leq -1$ and $\psi = 0$ for $x \geq -\frac{1}{2}$. Set

$$\psi_2(x_1) = \psi\left(\frac{x_1 - x_1(t_2)}{cN(t_1)^{-1}}\right).$$

Then, given $\eta > 0$ and choosing $c = c(\eta)$ sufficiently small, we have

$$\|(\psi_2 u(t_2), \psi_2 u_t(t_2))\|_{\mathcal{H}^{s_p}} < \eta.$$

Choosing η small enough, the small-data theory and finite speed of propagation for (1-1) imply

$$\int_\Omega ||\nabla|^{s_p} u(t_1, x)|^2 + ||\nabla|^{s_p - 1} u_t(t_1, x)|^2 \, dx \lesssim \eta^2,$$

where

$$\Omega = \{x : x_1 \leq x_1(t_2) - (t_2 - t_1) - cN(t_1)^{-1}\}.$$

Using the assumption on $|x(t_2) - x(t_1)|$ and the normalizations above, one finds

$$\Omega \supset \{x : -e_1 \cdot [x - x(t_1)] \geq 2cN(t_1)^{-1}\},$$

so that

$$\int_{-e_1 \cdot [x - x(t_1)] \geq 2cN(t_1)^{-1}} ||\nabla|^{s_p} u(t_1, x)|^2 + ||\nabla|^{s_p - 1} u_t(t_1, x)|^2 \, dx \lesssim \eta^2.$$

On the other hand, choosing $c = c(\eta)$ sufficiently small, Lemma 3.19 implies

$$\int_{0 < -e_1 \cdot [x - x(t_1)] < 2cN(t_1)^{-1}} ||\nabla|^{s_p} u(t_1, x)|^2 + ||\nabla|^{s_p - 1} u_t(t_1, x)|^2 \, dx < \eta^2.$$

We now choose $\eta^2 \ll C(u)$, where $C(u)$ is as in Definition 3.20, to reach a contradiction to Proposition 3.21. □

We are now in a position to prove that we can extract a traveling-wave solution from any solution with compactness property with translation parameter $x(t)$ that fails to be subluminal.

Proposition 3.24. *Suppose that $\vec{u}(t)$ is a solution with the compactness property on $\mathbb{R}$ with parameters $N(t)$ and $x(t)$. Suppose that either $N(t)$ is soliton-like or doubly concentrating in the sense of Proposition 3.15 and that $x(t)$ fails to be subluminal in the sense of Definition 3.12. Then there exists a* *(possibly different) solution $\vec{w}(s)$ to (1-1) with the compactness property on $\mathbb{R}$ with parameters $N(s)$ and $x(s)$ satisfying*

$$N(s) \equiv 1, \quad |x(s) - (s, 0, 0)| \lesssim \sqrt{s} \quad \textit{for all } s \in \mathbb{R}.$$

Proof of Proposition 3.24. Note that by Proposition 3.22 it suffices to show that we can extract a solution with the compactness property on $\mathbb{R}$ with parameters $N(t) = 1$ and $x(t)$ failing to be subluminal. By our assumption that $x(t)$ fails to be subluminal, for each $m \in \mathbb{N}$ there exists $t_m \in \mathbb{R}$ so that

$$|x(t_m) - x(t)| \geq |t - t_m| - \frac{1}{mN(t_m)} \quad \text{for all } t \in I_m := \left[t_m, t_m + \frac{m}{N(t_m)}\right]. \tag{3-7}$$

We will show that $N(t) \simeq N(t_m)$ for all $t \in I_m$ with constants independent of m. First assume that

$$N(t_m) \leq N(t).$$

Then by Lemma 3.23 we can find a constant $c > 0$ so that

$$c^2 N(t) \leq N(t_m) \leq N(t) \quad \text{for all } t \in I_m.$$

Next assume that

$$N(t) \leq N(t_m).$$

This means that

$$-\frac{1}{N(t_m)} \geq -\frac{1}{N(t)}$$

and thus from (3-7) we see that

$$|x(t_m) - x(t)| \geq |t - t_m| - \frac{1}{mN(t_m)} \geq |t - t_m| - \frac{1}{MN(t)}.$$

Another application of Lemma 3.23 then gives

$$N(t) \leq N(t_m) \leq \frac{1}{c^2} N(t).$$

As we can assume in Lemma 3.23 that $c < 1$, we deduce that

$$c^2 N(t) \leq N(t_m) \leq \frac{1}{c^2} N(t) \quad \text{for all } t \in I_m. \tag{3-8}$$

We can then extract, in the usual manner a new solution $\vec{w}(s)$ with the compactness property on $[0, \infty)$ with

$$\tilde{N}(s) := \lim_{m \to \infty} \frac{N\left(t_m + \frac{s}{N(t_m)}\right)}{N(t_m)},$$
$$\tilde{x}(s) := \lim_{m \to \infty} N(t_m)\left(x\left(t_m + \frac{s}{N(t_m)}\right) - x(t_m)\right).$$

Note that by (3-8) we must have

$$c_1 \leq \tilde{N}(s) \leq C_1 \quad \text{for all } s \in [0, \infty).$$

Moreover, using (3-7), for each $\epsilon > 0$ we can find $M > 0$ large enough so that for each $m \geq M$ we have

$$|\tilde{x}(s)| + \epsilon \geq \left| N(t_m)\left(x\left(t_m + \frac{s}{N(t_m)} \right) - x(t_m) \right) \right|$$
$$\geq N(t_m) \left| \frac{s}{N(t_m)} - \frac{1}{mN(t_m))} \right| \geq s - \frac{1}{m}.$$

Letting $m \to \infty$ we obtain

$$|\tilde{x}(s)| \geq s \quad \text{for all } s \in [0, \infty).$$

Noting that $\tilde{x}(0) = 0$ and combining the above with (3-3), we conclude that

$$1 \leq \frac{|\tilde{x}(s)|}{s} \to 1 \quad \text{as } s \to \infty.$$

From here it is straightforward to obtain a new solution $\vec{w}(s)$ with the compactness property on all of $\mathbb{R}$ with parameters $N(s) \equiv 1$ and $x(s)$ failing to be subluminal in the sense of Definition 3.12, and we apply Proposition 3.22 to conclude. □

Finally, we now have the ingredients necessary to prove Proposition 3.13.

Proof of Proposition 3.13. Suppose $\vec{u}(t)$ is a solution to (1-1) with the compactness property on its maximal interval of existence $I_{\max}$ with compactness parameters $N(t)$ and $x(t)$. By Proposition 3.15, if the solution has the compactness property with $N(t) = t^{-1}$, then we may also assume without loss of generality that it has the compactness property with translation parameter $x(t) = 0$: by finite speed of propagation, $x(t)$ must remain bounded, and hence we may, up to passing to a subsequence, obtain a precompact solution with $x(t) = 0$ by applying a fixed translation. Thus, in the case that $N(t) = t^{-1}$ we obtain a self-similar solution; i.e., we have reduced to case (III).

In the remaining cases we must address different scenarios depending on whether or not $x(t)$ is subluminal in the sense of Definition 3.12. If $x(t)$ is subluminal, then we have reduced ourselves to cases (I) and (II). If $x(t)$ fails to be subluminal, then by Proposition 3.24 we can find a critical element as in the traveling-wave scenario, i.e., case (IV). □

4. The soliton-like critical element

In this section we show that the soliton-like critical element, that is, case (I) from Proposition 3.13, cannot exist. The main result is the following proposition:

Proposition 4.1. *There are no soliton-like critical elements for* (1-1)*, in the sense of case* (I) *of Proposition 3.13.*

We recall that soliton-like means that $\vec{u}(t)$ is a global solution to (1-1) with the compactness property on $\mathbb{R}$ as defined in Definition 3.4 with parameters $N(t) \equiv 1$, and $x(t)$ subluminal in the sense of Definition 3.12. We will show that any such solution with the compactness property is necessarily $\equiv 0$.

The proof will be accomplished in two main steps. We are ultimately aiming to employ a rigidity argument based on a virial identity, which will show that any such critical element must then be identically 0.

The key point here is that in order to access the virial identity, which is at $\dot{\mathcal{H}}^1$ regularity, and to use it to prove Proposition 4.1, we first must prove that our critical element actually lies in a precompact subset of $\dot{\mathcal{H}}^1$. Thus, we must first show that a soliton-like critical element must be more regular than expected. In fact, we will prove that the trajectory K of any soliton-like critical element (see Definition 3.4) must be precompact in $\dot{\mathcal{H}}^1 \cap \dot{\mathcal{H}}^{s_p}$.

Throughout this section, we assume towards a contradiction that $\vec{u}(t)$ is a critical element with $x(t)$ subluminal in the sense of Definition 3.12 and $N(t) \equiv 1$. In particular, by Lemma 3.16 there exists $\delta_0 > 0$ so that

$$|x(t) - x(\tau)| < (1 - \delta_0)|t - \tau| \quad \text{for all } |t - \tau| > \frac{1}{\delta_0}.$$

4A. *Additional regularity.* We first prove that if the soliton-like critical element $\vec{u}$ has some additional regularity to begin with, then we can achieve $\dot{\mathcal{H}}^1$ regularity. The key ingredient in our proof will be a double Duhamel argument, which will enable us to gain the requisite regularity for critical elements, while our main technical tool will be the use of a frequency envelope which controls the $\dot{\mathcal{H}}^1$ norm (see Definition 2.2). In order to exploit the sharp Huygens principle, we will use the following modified frequency projection operators: let $\psi \geq 0$ be a smooth function supported on $|x| \leq 2$ satisfying $\psi = 1$ on $|x| \leq 1$. For $k \geq 0$, let

$$Q_{<k} f(x) = \int_{\mathbb{R}^3} 2^{3k} \psi(2^k(x - y)) f(y)\, dy. \tag{4-1}$$

These satisfy the same estimates as the usual Littlewood–Paley projections (which instead use sharp cutoffs in frequency space), e.g., the Bernstein estimates in Lemma 2.1.

We summarize the main ingredient in Proposition 4.1, the aforementioned additional regularity result, in the following proposition.

Proposition 4.2. *Suppose $\vec{u}$ is a soliton-like critical element. Then*

$$\vec{u} \in L_t^\infty \dot{\mathcal{H}}^{s_p} \quad \Longrightarrow \quad \vec{u} \in L_t^\infty \dot{\mathcal{H}}^s$$

for some $s > 1$. In particular, the set

$$K := \{\vec{u}(t, \cdot - x(t)) : t \in \mathbb{R}\} \subset \dot{\mathcal{H}}^{s_p} \cap \dot{\mathcal{H}}^1$$

is precompact in $\dot{\mathcal{H}}^{s_p} \cap \dot{\mathcal{H}}^1$.

We will prove Proposition 4.2 in several steps. To make this precise, we define the parameter

$$s_0 = s_p + \frac{5 - p}{2p(p-1)} = \frac{3}{2} - \frac{5}{2p}. \tag{4-2}$$

This exponent is chosen so that $\dot{\mathcal{H}}^{s_0}$ has the same scaling as $L_t^p L_x^{2p}$, and we note that crucially $s_p < s_0 < 1$.

4B. *The jump from $\dot{\mathcal{H}}^{s_0}(\mathbb{R}^3)$ regularity to $\dot{\mathcal{H}}^1(\mathbb{R}^3)$ regularity.* We begin with the first, easier gain in regularity, namely passing from $\dot{\mathcal{H}}^{s_0}(\mathbb{R}^3)$ to $\dot{\mathcal{H}}^1(R^3)$.

Proposition 4.3. *Suppose $\vec{u}$ is a soliton-like critical element. Let $s_0 > s_p$ be defined as in (4-2). Then*

$$\vec{u} \in L_t^\infty \dot{\mathcal{H}}^{s_0} \quad \Longrightarrow \quad \vec{u} \in L_t^\infty \dot{\mathcal{H}}^1.$$

Proof. By time-translation symmetry, it suffices to estimate the $\dot{\mathcal{H}}^1$-norm at time $t = 0$. We complexify the solution, letting

$$w = u + \frac{i}{\sqrt{-\Delta}} u_t.$$

Then

$$\|w(t)\|_{\dot{H}^1} \simeq \|\vec{u}(t)\|_{\dot{H}^1 \times L^2},$$

and if $\vec{u}(t)$ solves (1-1), then $w(t)$ is a solution to

$$w_t = -i\sqrt{-\Delta} w \pm \frac{i}{\sqrt{-\Delta}} |u|^{p-1} u.$$

By Duhamel's principle, for any T, we have

$$w(0) = e^{iT\sqrt{-\Delta}} w(T) \pm \frac{i}{\sqrt{-\Delta}} \int_T^0 e^{i\tau\sqrt{-\Delta}} F(u)(\tau)\, \mathrm{d}\tau,$$

where $F(u) = |u|^{p-1}u$. By compactness (see Lemma 3.7),

$$\lim_{T\to\infty} Q_{<k} e^{-iT\sqrt{-\Delta}} w(T) = \lim_{T\to\infty} Q_{<k} e^{iT\sqrt{-\Delta}} w(-T) = 0 \tag{4-3}$$

as weak limits in $\dot{H}^1$ for any $k \geq 0$. We next write

$$\begin{aligned} Q_{<k} w(0) &= e^{-iT\sqrt{-\Delta}} Q_{<k} w(T) \mp \frac{1}{\sqrt{-\Delta}} \int_0^T e^{-it\sqrt{-\Delta}} Q_{<k} F(u(t))\, \mathrm{d}t \\ &= e^{iT\sqrt{-\Delta}} Q_{<k} w(-T) \mp \frac{1}{\sqrt{-\Delta}} \int_{-T}^0 e^{-it\sqrt{-\Delta}} Q_{<k} F(u(t))\, \mathrm{d}t. \end{aligned}$$

Using (4-3), and arguing as in [Dodson and Lawrie 2015b, Section 4] we can deduce

$$\begin{aligned} &\langle Q_{<k} w(0), Q_{<k} w(0)\rangle_{\dot{H}^1} \\ &\qquad = \lim_{T\to\infty} \Big\langle \int_0^T e^{-it\sqrt{-\Delta}} Q_{<k} F(u(t))\, \mathrm{d}t, \int_{-T}^0 e^{-i\tau\sqrt{-\Delta}} Q_{<k} F(u(s))\, \mathrm{d}t \Big\rangle_{L^2}. \end{aligned} \tag{4-4}$$

We fix a large parameter $R > 0$ to be determined below. Let δ_0 be as in the statement of Lemma 3.16 and take $T = 2R\delta_0^{-1}$. We define

$$\begin{aligned} &\text{region } \boldsymbol{A} := \{(t,x) : 0 \leq t \leq T\}, \\ &\text{region } \boldsymbol{B} := \{(t,x) : |x - x(T)| \geq R + |t - T|\}, \\ &\text{region } \boldsymbol{C} := \{(t,x) : |x - x(T)| < R + |t - T|\}. \end{aligned} \tag{4-5}$$

See Figure 1.

We will treat these regions separately. Our goal is to bound u on region $\boldsymbol{A}$ using that we are estimating the solution on a compact time interval, and on region $\boldsymbol{B}$ using the small-data theory and finite speed of propagation. We will then use the double Duhamel trick, together with the sharp Huygens principle on region $\boldsymbol{C}$, to conclude the proof.

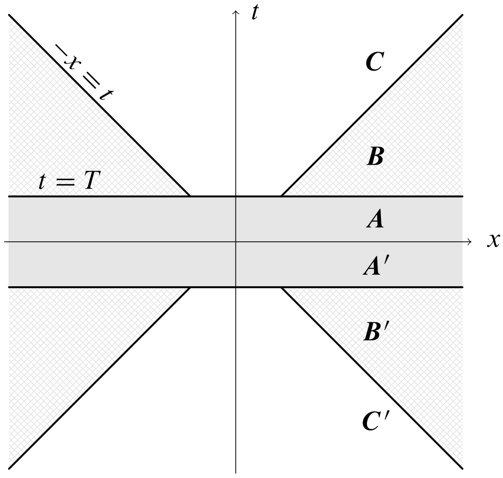

Figure 1. A depiction of the space-time regions $\boldsymbol{A}$, $\boldsymbol{A}'$, $\boldsymbol{B}$, $\boldsymbol{B}'$ and $\boldsymbol{C}$, $\boldsymbol{C}'$ in the case $x(t) \equiv 0$.

Let χ_R denote a smooth cutoff to the set

$$\{|x - x(T)| > R\} \subseteq \mathbb{R}^3.$$

Now fix a small parameter $\eta > 0$. By compactness of $\vec{u}$, if $R = R(\eta)$ is sufficiently large then we have

$$\|\chi_R \vec{u}(T)\|_{\mathcal{H}^{s_p}} \leq \eta. \tag{4-6}$$

We let $\vec{v} = (v, v_t)$ be the solution to (1-1) with initial data

$$\vec{v}(T) = \chi_R \vec{u}(T).$$

By finite speed of propagation, we have

$$u \equiv v \quad \text{for } |x - x(T)| \geq R + |t - T|.$$

We now rewrite (4-4), and abusing notation slightly, we define

$$\begin{aligned} \int_0^\infty e^{-it\sqrt{-\Delta}} Q_{<k} F(u(t))\, \mathrm{d}t &= A + B + C, \\ A &= \int_0^T e^{-it\sqrt{-\Delta}} Q_{<k} F(u(t))\, \mathrm{d}t, \\ B &= \int_T^\infty e^{-it\sqrt{-\Delta}} Q_{<k} F(v(t))\, \mathrm{d}t, \\ C &= \int_T^\infty e^{-it\sqrt{-\Delta}} Q_{<k} [F(u(t)) - F(v(t))]\, \mathrm{d}t. \end{aligned} \tag{4-7}$$

Note that the notation in (4-7) is such that each term relates to an estimate for the solution on the correspondingly named region from (4-5).

We can carry out a similar construction at time $-T$, yielding a small solution $\tilde{v}$ that agrees with u whenever $|x - x(-T)| \geq R + |t + T|$, and we obtain three terms in the negative time direction

$$\int_{-\infty}^{0} e^{-i\tau\sqrt{-\Delta}} Q_{<k} F(u(\tau))\,\mathrm{d}\tau = A' + B' + C',$$

$$\begin{aligned} A' &= \int_{-T}^{0} e^{-i\tau\sqrt{-\Delta}} Q_{<k} F(u(\tau))\,\mathrm{d}\tau, \\ B' &= \int_{-\infty}^{-T} e^{-i\tau\sqrt{-\Delta}} Q_{<k} F(\tilde{v}(\tau))\,\mathrm{d}\tau, \\ C' &= \int_{-\infty}^{-T} e^{-i\tau\sqrt{-\Delta}} Q_{<k} [F(u(\tau)) - F(\tilde{v}(\tau))]\,\mathrm{d}\tau. \end{aligned} \tag{4-8}$$

Using the elementary Hilbert space estimate

$$|\langle A + B + C, A' + B' + C' \rangle| \lesssim |A|^2 + |A'|^2 + |B|^2 + |B'|^2 + |\langle C, C' \rangle|$$

whenever $A + B + C = A' + B' + C'$, where the $|\cdot|^2$ denotes the square of the norm induced by the inner product, we may estimate

$$\langle Q_{<k} w(0), Q_{<k} w(0) \rangle_{\dot{H}^1}$$

by obtaining bounds for A, A' and B, B' and $\langle C, C' \rangle$.

Region A. To estimate the A and A' terms, first we establish the bound

$$\|u\|_{L^{2(p-1)}_{t,x}([-T,T]\times\mathbb{R}^3)} \lesssim \left(\frac{T}{\epsilon}\right)^{\frac{1}{2(p-1)}} \tag{4-9}$$

for some suitably small $\epsilon > 0$. To prove this, we rely on the fact that $\vec{u}$ is a soliton-like critical element. Fix $\eta > 0$. Since $N(t) = 1$, there exists $\epsilon > 0$ small enough that the $L^{2(p-1)}_{t,x}$-norm is bounded by η on any interval of length ϵ; see Lemma 3.11. Thus to obtain the desired bound, we divide $[-T, T]$ into $\sim \lceil T/\epsilon \rceil$ intervals J_k of length ϵ, and

$$\|u\|^{2(p-1)}_{L^{2(p-1)}_{t,x}([-T,T]\times\mathbb{R}^3)} \simeq \sum_{k=1}^{\lceil T/\epsilon \rceil} \|u\|^{2(p-1)}_{L^{2(p-1)}_{t,x}(J_k\times\mathbb{R}^3)} \lesssim \frac{T}{\epsilon}.$$

Using a similar argument together with Strichartz estimates and the hypothesis

$$\|u\|_{L^\infty_t \mathcal{H}^{s_0}} \lesssim 1,$$

we obtain

$$\|u\|_{L^p_t L^{2p}_x([-T,T]\times\mathbb{R}^3)} \lesssim \left(\frac{T}{\epsilon}\right)^{\frac{1}{p}} \|u\|_{L^\infty_t \mathcal{H}^{s_0}}. \tag{4-10}$$

Thus, using (4-9), (4-10) and Strichartz estimates, we can estimate

$$|A|^2 + |A'|^2 \lesssim \|u\|^p_{L^p_t L^{2p}_x([-T,T]\times\mathbb{R}^3)} \lesssim \left(\frac{T}{\epsilon}\right) \|u\|^p_{L^\infty_t \mathcal{H}^{s_0}}.$$

Region B. For the estimates of B and B', we use the small-data theory to bound the solutions v and $\tilde{v}$. We argue only for v as the estimates for $\tilde{v}$ are identical. By the small-data theory, for η chosen sufficiently small in (4-6), we have

$$\|v\|_{L^{2(p-1)}_{t,x}(\mathbb{R}^{1+3})} \lesssim \eta.$$

Using Strichartz estimates, we bound

$$\begin{aligned}\||\nabla|^{\frac{3(p-3)}{2p}} v\|_{L^p_t L^{2p/(p-2)}_x} &\lesssim \|v(T)\|_{\mathcal{H}^{s_0}} + \||\nabla|^{\frac{3(p-3)}{2p}} (|v|^{p-1}v)\|_{L^{2p/(p+2)}_t L^{p/(p-1)}_x} \\ &\lesssim \|u(T)\|_{\mathcal{H}^{s_0}} + \|v\|^{p-1}_{L^{2(p-1)}_{t,x}} \||\nabla|^{\frac{3(p-3)}{2p}} v\|_{L^p_t L^{2p/(p-2)}_x},\end{aligned}$$

with all space-time norms over $\mathbb{R}^{1+3}$. Note that $\left(p, \frac{2p}{p-2}\right)$ is wave-admissible. Thus, for η sufficiently small, we deduce

$$\||\nabla|^{\frac{3(p-3)}{2p}} v\|_{L^p_t L^{2p/(p-2)}_x} \lesssim \|u\|_{L^\infty_t \mathcal{H}^{s_0}(\mathbb{R}^{1+3})},$$

and hence it follows from Sobolev embedding that

$$\|v\|_{L^p_t L^{2p}_x(\mathbb{R}^{1+3})} \lesssim \|u\|_{L^\infty_t \mathcal{H}^{s_0}}.$$

Thus, we have shown that

$$|B|^2 + |B'|^2 \lesssim \|v\|^p_{L^p_t L^{2p}_x(\mathbb{R}^{1+3})} \lesssim \|u\|^p_{L^\infty_t \mathcal{H}^{s_0}}.$$

Region C. Finally, we claim that

$$\langle C, C'\rangle \equiv 0. \tag{4-11}$$

To see this, write

$$\langle C, C'\rangle = \int_T^\infty \int_{-\infty}^{-T} \left\langle e^{i(\tau-t)\sqrt{-\Delta}} Q_{<k}[F(u(t)) - F(v(t))], Q_{<k}[F(u(\tau)) - F(\tilde{v}(\tau))]\right\rangle d\tau\, dt,$$

and note that by subluminality and the fact that $x(0) = 0$, we have for $T = 2R\delta_0^{-1}$ the inclusion

$$\{|x - x(\pm T)| \le R\} \subset \{|x| \le (1 - 2^{-1}\delta_0)T\}.$$

We recall that the operator $Q_{<k}$ defined in (4-1) is given by convolution with the function $2^{3k}\psi(2^k x)$ for a fixed function $\psi \in C_0^\infty(\mathbb{R}^3)$. Hence, for $k \ge k_0$, a sufficiently large, fixed constant depending on the support of ψ, δ_0 and T, we can ensure

$$\operatorname{supp}\left(Q_{<k}[F(u(\tau)) - F(\tilde{v}(\tau))]\right) \subseteq \left\{|x| \le |\tau| - 4^{-1}\delta_0 T\right\}.$$

Similarly, using the properties of the $Q_{<k}$ and the sharp Huygens principle, we can ensure that for k sufficiently large,

$$\operatorname{supp}\left(e^{i(\tau-t)\sqrt{-\Delta}} Q_{<k}[F(u(t)) - F(v(t))]\right) \subseteq \{|x| > |t-\tau| - 4^{-1}\delta_0 T\}.$$

Since $t > 0$ and $\tau < 0$, we have $|t - \tau| > |\tau|$, this yields (4-11), as required.

Collecting these estimates, we obtain that

$$\|Q_{<k}w(0)\|_{\dot H^1}^2 = \langle Q_{<k}w(0), Q_{<k}w(0)\rangle_{\dot H^1} \lesssim 1$$

uniformly in $k \geq 0$. The desired result then follows. □

4C. ***The jump from $\dot{\mathcal{H}}^{s_p}(\mathbb{R}^3)$ regularity to $\dot{\mathcal{H}}^{s_0}(\mathbb{R}^3)$ regularity.*** Now we turn to the more difficult estimates. Here, we will need a finer analysis based on frequency envelope machinery. We prove the following.

Proposition 4.4. *Suppose $\vec u$ is a soliton-like critical element. Then*

$$\vec u \in L_t^\infty \dot{\mathcal{H}}^{s_p} \quad \implies \quad \vec u \in L_t^\infty \dot{\mathcal{H}}^{s}$$

for any $s_p \leq s < 1$.

Proof. Once again, we define

$$\begin{aligned} &\text{region } \boldsymbol{A} := \{(t,x) : |t| \leq T\},\\ &\text{region } \boldsymbol{B} := \{(t,x) : |x - x(T)| \geq R + |t-T|\},\\ &\text{region } \boldsymbol{C} := \{(t,x) : |x - x(T)| < R + |t-T|\}, \end{aligned}$$

with corresponding regions $\boldsymbol{A}'$, $\boldsymbol{B}'$, $\boldsymbol{C}'$ in the negative time direction. We further introduce

$$Q_k = Q_{<2k} - Q_{<k} \quad \text{for } k > 0, \qquad Q_0 = Q_{<0}.$$

By Schur's test, we can conclude that these frequency projections are a good partition of frequency space, in the sense that

$$\|f\|_{\dot H^s}^2 \sim \|Q_0 f\|_{\dot H^s}^2 + \sum_{k\geq 0} 2^{ks}\|Q_k f\|_{L^2}^2.$$

We will also need to introduce an exponent q satisfying

$$2 < q < \frac{2}{s_p}.$$

Region A. We begin by defining suitable frequency envelopes with a parameter $\sigma > 0$ to be determined shortly. We set

$$\begin{aligned} \gamma_k(t_0) &= \sum_j 2^{-\sigma|j-k|}\big[2^{s_p j}\|Q_j u(t_0)\|_{L^2} + 2^{j(s_p-1)}\|Q_j \partial_t u(t_0)\|_{L^2}\big],\\ \alpha_k(J) &= \sum_j 2^{-\sigma|j-k|}\big[2^{-j(\frac{2}{q}-s_p)}\|Q_j u\|_{L_t^q L_x^{2q/(q-2)}(J\times\mathbb{R}^3)}\\ &\qquad\qquad + 2^{j(\frac{2}{q}-1+s_p)}\|Q_j u\|_{L_t^{2q/(q-2)}L_x^q(J\times\mathbb{R}^3)}\big] \end{aligned} \tag{4-12}$$

for $k \geq 0$. Note $\big(q, \frac{2q}{q-2}\big)$ is sharp admissible and that each of the quantities appearing in the definition of β_k has the same scaling as $\dot{\mathcal{H}}^{s_p}$. We will choose

$$0 < \sigma < \frac{2}{q} - s_p.$$

Our goal is to prove that

$$\alpha_k([-T,T]) \lesssim \gamma_k(0) + C_0 2^{-k\sigma}, \tag{4-13}$$

where $C_0 = C_0(T)$.

We begin by recording the some space-time estimates for $\vec{u}$ that are consequences of the precompactness of the set K; see Definition 3.4. We fix $\eta > 0$. Since $N(t) = 1$, there exists $\epsilon > 0$ small enough that the $L^{2(p-1)}_{t,x}$ norm is $< \eta$ on any interval of length ϵ; see Lemma 3.11. Furthermore, we can find $k_0 = k_0(\eta)$ such that, for any $k > k_0$,

$$\|Q_{>k}u\|_{L^{2(p-1)}_{t,x}([-T,T]\times\mathbb{R}^3)} < \eta T^{\frac{1}{2(p-1)}}.$$

With these bounds in hand, we turn to the proof of (4-13). In the following, all space-time norms will be taken over $[-T,T]\times\mathbb{R}^3$. For any j, we decompose the nonlinearity as follows. Writing $u_{\le j} = Q_{\le j}u$ (and similarly for $u_{>j}$), we write

$$F(u) = F(u_{>k_0}) + F(u) - F(u_{>k_0}),$$

where $k_0(\eta)$ is as above. By Taylor's theorem, we have

$$F(u) = F(u_{>k_0}) + u_{\le k_0}\int_0^1 F'(\theta u_{\le k_0} + u_{>k_0})\,\mathrm{d}\theta,$$

and hence to estimate the nonlinearity, it suffices to estimate three types of terms

$$u^{p-1}_{>k_0}u_{>j}, \quad u^{p-1}_{>k_0}u_{\le j}, \quad u_{\le k_0}u^{p-1}.$$

Using the inhomogeneous Strichartz estimates, we obtain

$$\begin{aligned}
&2^{-j(\frac{2}{q}-s_p)}\left\|\int_0^t e^{i(t-s)\sqrt{-\Delta}}Q_jF(u(s))\,\mathrm{d}s\right\|_{L^q_tL^{2q/(q-2)}_x} \\
&\quad + 2^{j(\frac{2}{q}-1+s_p)}\left\|\int_0^t e^{i(t-s)\sqrt{-\Delta}}Q_jF(u(s))\,\mathrm{d}s\right\|_{L^{2q/(q-2)}_tL^q_x} \\
&\lesssim \min\{2^{j(\frac{2}{q}-1+s_p)}\|F(u)\|_{L^{q/(q-1)}_tL^{2q/(q+2)}_x}, 2^{-j(\frac{2}{q}-s_p)}\|F(u)\|_{L^{2q/(q+2)}_tL^{q/(q-1)}_x}\}.
\end{aligned}\tag{4-14}$$

Now let J be an interval with $|J| < \epsilon$, and let $t_0 = \inf J$. In the next estimates, all norms will be taken over $J\times\mathbb{R}^3$. Using Strichartz estimates, we estimate

$$\begin{aligned}
&2^{-j(\frac{2}{q}-s_p)}\|Q_ju\|_{L^q_tL^{2q/(q-2)}_x(J\times\mathbb{R}^3)} + 2^{j(\frac{2}{q}-1+s_p)}\|Q_ju\|_{L^{2q/(q-2)}_tL^q_x(J\times\mathbb{R}^3)} \\
&\quad\lesssim 2^{js_p}\|u_j(t_0)\|_{L^2_x} + 2^{j(s_p-1)}\|\partial_tu_j(t_0)\|_{L^2_x} + 2^{j(\frac{2}{q}-1+s_p)}\|u^{p-1}_{>k_0}u_{>j}\|_{L^{q/(q-1)}_tL^{2q/(q+2)}_x} \\
&\qquad + 2^{-j(\frac{2}{q}-s_p)}\|u^{p-1}_{>k_0}u_{\le j}\|_{L^{2q/(q+2)}_tL^{q/(q-1)}_x} + 2^{-j(\frac{2}{q}-s_p)}\|u_{\le k_0}u^{p-1}\|_{L^{2q/(q+2)}_tL^{q/(q-1)}_x} \\
&\quad\lesssim 2^{js_p}\|u_j(t_0)\|_{L^2_x} + 2^{j(s_p-1)}\|\partial_tu_j(t_0)\|_{L^2_x} + I + II + III.
\end{aligned}$$

We first estimate term I. We obtain

$$\begin{aligned}
&2^{j(\frac{2}{q}-1+s_p)}\|u_{>k_0}^{p-1}u_{>j}\|_{L_t^{q/(q-1)}L_x^{2q/(q+2)}}\\
&\qquad\lesssim 2^{j(\frac{2}{q}-1+s_p)}\|u_{>k_0}\|_{L^{2(p-1)}_{t,x}}^{p-1}\|u_{>j}\|_{L_t^{2q/(q-2)}L_x^q}\\
&\qquad\lesssim \eta^{(p-1)}T^{1/2}2^{j(\frac{2}{q}-1+s_p)}\sum_{\ell>j}2^{-\ell(\frac{2}{q}-1+s_p)}[2^{\ell(\frac{2}{q}-1+s_p)}\|Q_\ell u\|_{L_t^{2q/(q-2)}L_x^q}].
\end{aligned}$$

Similarly, for term II we obtain

$$\begin{aligned}
&2^{-j(\frac{2}{q}-s_p)}\|u_{>k_0}^{p-1}u_{\le j}\|_{L_t^{2q/(q+2)}L_x^{q/(q-1)}}\\
&\qquad\lesssim 2^{-j(\frac{2}{q}-s_p)}\|u_{>k_0}\|_{L^{2(p-1)}_{t,x}}^{p-1}\|u_{\le j}\|_{L_t^qL_x^{2q/(q-2)}}\\
&\qquad\lesssim 2^{-j(\frac{2}{q}-s_p)}\|u_{>k_0}\|_{L^{2(p-1)}_{t,x}}^{p-1}\sum_{\ell\le j}2^{\ell(\frac{2}{q}-s_p)}[2^{-\ell(\frac{2}{q}-s_p)}\|u_\ell\|_{L_t^qL_x^{2q/(q-2)}}].
\end{aligned}$$

Finally, for term III, using smallness of the interval and we obtain

$$\begin{aligned}
2^{-j(\frac{2}{q}-s_p)}\|u_{\le k_0}u^{p-1}\|_{L_t^{2q/(q+2)}L_x^{q/(q-1)}} &\lesssim 2^{-j(\frac{2}{q}-s_p)}\|u\|_{L^{2(p-1)}_{t,x}}^{p-1}\|u_{\le k_0}\|_{L_t^qL_x^{2q/(q-2)}}\\
&\lesssim 2^{-j(\frac{2}{q}-s_p)}2^{k_0(\frac{2}{q}-s_p)}\eta^{p-1}T^{\frac{1}{2}}.
\end{aligned}$$

Multiplying by $2^{-\sigma|j-k|}$ and summing in the above bounds, recalling that $\sigma<\frac{2}{q}-s_p$ in our definition of the frequency envelopes in (4-12), it follows that (for $t_0=\inf J$) we have

$$\gamma_k(t_1)+\alpha_k(J)\lesssim\gamma_k(t_0)+T^{\frac{1}{2}}\eta^{p-1}\alpha_k(J)+C_0(T)2^{-k\sigma}.$$

For $\eta\equiv\eta(T)$ small enough so that

$$C\eta^{p-1}T^{\frac{1}{2}}<\tfrac{1}{2}$$

with C the implicit constant in Strichartz estimates, this implies

$$\alpha_k(J)\lesssim\gamma_k(t_0)+C_02^{-k\sigma}.$$

Iterating this procedure $\lceil T/\epsilon\rceil$ times on $[-T,T]$, we may also conclude that

$$\gamma_k(t_0)\lesssim\gamma_k(0),$$

from which (4-13) follows by summing up these estimates.

Region B. To implement the double Duhamel argument, we will again consider the solution v to (1-1) with data $\vec{v}(T)=\chi_R\vec{u}(T)$. To control this solution, we define the frequency envelopes

$$\tilde{\gamma}_k(t_0)\quad\text{and}\quad\beta_k$$

analogously to (4-12), but with space-time norms over $\mathbb{R}^{1+3}$. We will prove

$$\beta_k \lesssim \gamma_k(0) + C_0 2^{-k\sigma}. \tag{4-15}$$

First observe that

$$\|v\|_{L^{2(p-1)}_{t,x}(\mathbb{R}^{1+3})} \lesssim \eta.$$

Thus

$$\begin{aligned}\||\nabla|^{-(\frac{2}{q}-s_p)} v\|_{L^q_t L^{2q/(q-2)}_x} + \||\nabla|^{\frac{2}{q}-1+s_p} v\|_{L^{2q/(q-2)}_t L^q_x} &\lesssim \|v(T)\|_{\mathcal{H}^{s_p}} + \||\nabla|^{\frac{2}{q}-1+s_p}(v|v|^{p-1})\|_{L^{q/(q-1)}_t L^{2q/(q+2)}_x} \\ &\lesssim \eta + \|v\|^{p-1}_{L^{2(p-1)}_{t,x}} \||\nabla|^{\frac{2}{q}-1+s_p} v\|_{L^{2q/(q-2)}_t L^q_x} \\ &\lesssim \eta + \eta^{p-1} \||\nabla|^{\frac{2}{q}-1+s_p} v\|_{L^{2q/(q-2)}_t L^q_x},\end{aligned}$$

which implies in particular that

$$\|v_{\le 1}\|_{L^q_t L^{2q/(q-2)}_x(\mathbb{R}^{1+3})} \lesssim \eta. \tag{4-16}$$

We now estimate β_k in essentially the same manner as α_k. The main difference is that we split at frequency 1 instead of at frequency k_0 as above. Estimating as above, but using (4-16), we deduce

$$\beta_k \lesssim \tilde{\gamma}_k(T) + \eta^{p-1}\beta_k + C_0 2^{-k\sigma},$$

which implies

$$\beta_k \lesssim \tilde{\gamma}_k(T) + C_0 2^{-k\sigma}. \tag{4-17}$$

In order to prove (4-15), we need to relate $\tilde{\gamma}_k(T)$ to $\gamma_k(0)$. Similar arguments as in (4-15) yield

$$\gamma_k(T) \lesssim \gamma_k(0) + \eta^{p-1}\beta_k + C_0 2^{-k\sigma} \lesssim \gamma_k(0) + C_0 2^{-k\sigma},$$

so it therefore suffices to relate $\tilde{\gamma}_k(T)$ to $\gamma_k(T)$. Using that $\vec{v}(T) = \chi_R \vec{u}(T)$, we apply the commutator estimate Lemma 2.3 to deduce

$$\begin{aligned}2^{ks_p}\|Q_k v(T)\|_{L^2} &\lesssim 2^{ks_p}\|Q_k u(T)\|_{L^2} + (2^k R)^{-(1-s_p)}\|u\|_{L^\infty_t \dot{H}^{s_p}}, \\ 2^{k(s_p-1)}\|Q_k \partial_t v(T)\|_{L^2} &\lesssim 2^{k(s_p-1)}\|Q_k \partial_t u(T)\|_{L^2} + 2^{-k}R^{-1}\|\partial_t u\|_{L^\infty_t \dot{H}^{s_p-1}}.\end{aligned}$$

In particular, since $\sigma < \frac{2}{q} - s_p < 1 - s_p$, we deduce that

$$\tilde{\gamma}_k(T) \lesssim \gamma_k(T) + C_0 2^{-k\sigma}.$$

Putting this together with (4-17) above, we conclude

$$\beta_k \lesssim \gamma_k(0) + C_0 2^{-k\sigma},$$

which completes the proof of (4-15).

We will now carry out the double Duhamel argument with the complexified solutions w. We write

$$\langle Q_j w(0), Q_j w(0)\rangle_{\dot H^1} = \lim_{s\to\infty}\Big\langle \int_0^s e^{-it\sqrt{-\Delta}} Q_j F(u(t))\,\mathrm{d}t, \int_{-s}^0 e^{-i\tau\sqrt{-\Delta}} Q_j F(u(\tau))\,\mathrm{d}\tau\Big\rangle_{\dot H^1} \tag{4-18}$$

and (as before) take the decomposition

$$\int_0^\infty e^{-it\sqrt{-\Delta}} Q_j F(u(t))\,\mathrm{d}t = A + B + C$$
$$= A' + B' + C' = \int_{-\infty}^0 e^{-i\tau\sqrt{-\Delta}} Q_j F(u(\tau))\,d\tau$$

for components as in (4-7) and (4-8). Once again, we rely on the algebraic inequality

$$\langle Q_j w(0), Q_j w(0)\rangle_{\dot H^1} \lesssim |A|^2 + |A'|^2 + |B|^2 + |B'|^2 + |\langle C, C'\rangle| \tag{4-19}$$

and we note that by construction and the argument above relying on the sharp Huygens principle, $\langle C, C'\rangle_{\dot{\mathcal{H}}^1} \equiv 0$.

To treat the other terms, we recall the definition of the frequency envelope α_k in (4-12), and we use (4-13) and (4-15). To this end, we multiply the left-hand side of (4-18) by $2^{-\sigma|j-k|}$ and we sum over $j \geq 0$ to obtain

$$\gamma_k(0) \lesssim \eta^{p-1}\gamma_k(0) + C_0 2^{-k\sigma},$$

which, choosing η sufficiently small depending only on the implicit constant, implies

$$\gamma_k(0) \lesssim C_0 2^{-k\sigma},$$

which yields $\vec u \in \dot{\mathcal{H}}^s$ for any $s_p \leq s < s_p + \sigma$. Since that we may choose any $\sigma < \frac{2}{q} - s_p$ and q arbitrarily close to 2, we deduce $\vec u \in L_t^\infty \dot{\mathcal{H}}^s$ for any $s_p \leq s < 1$. This completes the proof of Proposition 4.4. □

Propositions 4.3 and 4.4 immediately yield the following corollary.

Corollary 4.5. *Suppose $\vec u$ is a soliton-like critical element. Then*

$$\vec u \in L_t^\infty \dot{\mathcal{H}}^{s_p} \quad \Longrightarrow \quad \vec u \in L_t^\infty \dot{\mathcal{H}}^1.$$

4D. *The jump from $\dot{\mathcal{H}}^1(\mathbb{R}^3)$ regularity to $\dot{\mathcal{H}}^s(\mathbb{R}^3)$ regularity.* As mentioned above, in order to employ the rigidity argument based on a certain virial identity, we also need to prove that the trajectory of a critical element in fact lies in a precompact subset of $\dot{\mathcal{H}}^1$. We will achieve this by proving that in fact we can gain a bit more regularity; specifically we can place the solution in $\dot{\mathcal{H}}^s$ for some $s > 1$. The key idea here is that we actually have a bit of room in the previous estimates given the additional assumption of $\dot{\mathcal{H}}^1$ regularity, and this will provide some extra decay which we can use to establish the additional regularity.

Proposition 4.6. *Suppose $\vec u$ is a soliton-like critical element. Then $\vec u \in L_t^\infty \dot{\mathcal{H}}^s$ for some $s > 1$.*

Proof. Let v and $\tilde v$ be the solutions to the small-data Cauchy problems defined above. By small-data arguments $v(T) \in \dot{\mathcal{H}}^1(\mathbb{R}^3)$ and $\|v(T)\|_{\dot{\mathcal{H}}^{s_p}}$ small implies that

$$\|v\|_{L_t^{2q/(q-2)} L_x^q(\mathbb{R}\times\mathbb{R}^3)} + \||\nabla|^{1-s_p} v\|_{L_{t,x}^{2(p-1)}(\mathbb{R}\times\mathbb{R}^3)} \lesssim_T 1,$$
$$\|\tilde v\|_{L_t^{2q/(q-2)} L_x^q(\mathbb{R}\times\mathbb{R}^3)} + \||\nabla|^{1-s_p} \tilde v\|_{L_{t,x}^{2(p-1)}(\mathbb{R}\times\mathbb{R}^3)} \lesssim_T 1.$$

Furthermore, arguing as above and partitioning $[-T, T]$ into sufficiently small intervals, we obtain

$$\|u\|_{L_t^{2q/(q-2)}L_x^q([-T,T]\times\mathbb{R}^3)} + \||\nabla|^{1-s_p}u\|_{L_{t,x}^{2(p-1)}([-T,T]\times\mathbb{R}^3)} \lesssim_T 1.$$

These inequalities, together with the argument used to prove Proposition 4.4, as well as (4-19) and (4-14), establish that

$$\|Q_k u(0)\|_{\dot{\mathcal{H}}^1}^2 \lesssim_T 2^{-k\sigma}2^{-k(\frac{2}{q}-1+s_p)}.$$

Since we may choose any

$$\sigma < \frac{2}{q} - s_p,$$

and q arbitrarily close to 2, we have then shown that

$$\sum_k 2^{2\alpha k}\|Q_k u(0)\|_{\dot{\mathcal{H}}^1}^2 < \infty$$

for any $\alpha < \frac{1}{2}$, which concludes the proof. □

4E. *Rigidity for the soliton-like critical element.* Now we may prove that the soliton-like critical element is identically zero. We summarize this in the following proposition.

Proposition 4.7. *Let $\vec{u}(t) \in \dot{\mathcal{H}}^1$ be a global-in-time solution to (1-1) such that for subluminal $x(t)$ the set*

$$K = \{u(t, \cdot - x(t)), \partial_t u(t, \cdot - x(t)) : t \in \mathbb{R}\} \subset \dot{\mathcal{H}}^1 \cap \dot{\mathcal{H}}^{s_p}$$

is a precompact subset of $\dot{\mathcal{H}}^1 \cap \dot{\mathcal{H}}^{s_p}$. Then $\vec{u}(t) \equiv 0$.

As mentioned in the Introduction, we include a proof of rigidity for the soliton-like critical element in the *focusing setting* as well. The arguments that we use are similar to the ones given in [Côte et al. 2015, Section 3; Dodson and Lawrie 2015b; Rodriguez 2017] but with a modification. The key new ingredient here is that the subluminality of $x(t)$ compactifies the subset of the Lorentz group taking $(t, x(t))$ to $(t', 0)$; see also [Kenig and Merle 2006; Nakanishi and Schlag 2011] for a somewhat different approach that uses the Lorentz transform to show that critical elements must have zero momentum. The main ingredients in the proof are the following virial identities.

In what follows we let $r = |x|$ and set $\partial_r u = \nabla u \cdot (x/|x|)$.

Lemma 4.8 (virial identities). *Let $\chi \in C_0^\infty$ be a smooth radial function such that $\chi(r) = 1$ if $r \le 1$ and* supp $\chi \in \{r \le 2\}$. *For any $R > 0$ we define $\chi_R(r) = \chi(\frac{r}{R})$ and let $\vec{u}(t)$ be a solution to (1-1). Defining*

$$\Omega_{u(t)}(R) := \int_{|x|\ge R} |\nabla u|^2 + |\partial_t u|^2 + \frac{|u|^2}{|x|^2} + |u|^{p+1}\,\mathrm{d}x, \tag{4-20}$$

we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \partial_t u \mid \chi_R(r\partial_r u + u)\rangle = -E(\vec{u}) \pm \left(\frac{p-3}{p+1}\right)\|u\|_{L^{p+1}}^{p+1} + O(\Omega_{u(t)}(R)), \tag{4-21}$$

where the "+" above corresponds to the focusing equation and the "−" corresponds to the defocusing equation.

If $\vec{u}(t)$ solves the focusing equation, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big\langle \partial_t u : \chi_R\Big(r\partial_r u + \frac{1}{2}u\Big)\Big\rangle = -\frac{1}{2}\int |\partial_t u|^2 - 3\Big(\frac{1}{p+1} - \frac{1}{6}\Big)\int |u|^{p+1} + O(\Omega_{u(t)}(R)). \tag{4-22}$$

Proof of Proposition 4.7 for the focusing equation. We may assume that $x(0) = 0$. Since $x(t)$ is subluminal we can find $\delta > 0$ so that

$$|x(t) - x(\tau)| \le (1-\delta)|t-\tau|, \quad |x(t)| \le (1-\delta)|t| \tag{4-23}$$

for all $t, \tau \in \mathbb{R}$.

For convenience, we consider only the special case where

$$x(t) = (x_1(t), 0, 0) \quad \text{for all } t > 0,$$

as this contains the essential difficulties and the general argument is an easy modification of the one presented below. Recall that for each $\nu \in (-1, 1)$ we have a Lorentz transform L_ν defined by

$$L_\nu(t, x_1, x_1, x_3) = \left(\frac{t - \nu x_1}{\sqrt{1-\nu^2}}, \frac{x_1 - \nu t}{\sqrt{1-\nu^2}}, x_2, x_3\right) =: (t', x').$$

For any $T > 0$, set

$$\nu(T) := \frac{x_1(T)}{T}.$$

Then

$$-(1-\delta) \le \nu(T) \le 1-\delta \tag{4-24}$$

and the Lorentz transform $L_{\nu(T)}$ gives

$$L_{\nu(T)}(T, x_1(T), 0, 0) = (T', 0, 0, 0),$$

where

$$T' = \sqrt{T^2 - x_1(T)^2}. \tag{4-25}$$

Since $x(t)$ satisfies (4-23), we have the bounds

$$c_\delta T \le T' \le T$$

for $c_\delta := \sqrt{1-(1-\delta)^2} > 0$, which means that T' is comparable to T. For each $T > 0$ define

$$v_{\nu(T)}(t', x') := u \circ L_{\nu(T)}(t, x).$$

Then, since K above is precompact for $x(t)$ subluminal and since $\vec{v}_{\nu(T)}(t')$ as above is a fixed Lorentz transform of $\vec{u}(t, x)$, we can explicitly obtain a subluminal translation parameter $x'(t')$ with

$$x'(T') = 0,$$

by the choice of $\nu(T)$ above, such that the trajectory

$$K' := \{\vec{v}_{\nu(T)}(t', x - x'(t')) : t' \in \mathbb{R}\} \tag{4-26}$$

is precompact in $\dot{\mathcal{H}}^1 \cap \dot{\mathcal{H}}^{s_p}$; see for example [Duyckaerts et al. 2016, Section 6] or [Nakanishi and Schlag 2011, Chapter 2] where such claims are carefully justified. We will now establish the following.

Claim 4.9. *Consider a critical element for the focusing equation with $3 \le p < 5$. For each n there exists a time $T_n > 0$ such that for T_n' as in (4-25) we have*

$$\frac{1}{T_n'}\int_0^{T_n'}\int_{\mathbb{R}^3} |\partial_t v_{\nu(T_n)}(t,x)|^2 + |v_{\nu(T_n)}(t,x)|^{p+1}\,\mathrm{d}x\,\mathrm{d}t < \frac{1}{4n}.$$

Proof of Claim 4.9. Let $T>0$. Since $\vec{v}_{\nu(T)}$ solves the focusing equation we average (4-22) with $R = C_\delta T$ over the time interval $[0, T']$ for some constant C_δ to be specified below, yielding

$$\frac{1}{T'}\int_0^{T'}\int_{\mathbb{R}^3} |\partial_t v_{\nu(T)}(t,x)|^2 + |v_{\nu(T)}(t,x)|^{p+1}\,\mathrm{d}x\,\mathrm{d}t$$
$$\lesssim \frac{1}{T'}\Big|\langle \partial_t v_{\nu(T)}(t) \mid \chi_{2T}\, r\,\partial_r v_{\nu(T)}\rangle\big|_0^{T'}\Big| + \frac{1}{T'}\Big|\langle \partial_t v_{\nu(T)}(t) \mid \chi_{2T}\, v_{\nu(T)}\rangle\big|_0^{T'}\Big| + \frac{1}{T'}\int_0^{T'} \Omega_{v_{\nu(T)}(t)}(C_\delta T)\,\mathrm{d}t, \quad (4\text{-}27)$$

where $\Omega_{v_{\nu(T)}}(C_\delta T)$ is defined as in (4-20). Given $n > 0$, by (4-26), the subluminality of $x'(t')$, and the fact that

$$x'(0) = 0, \quad x'(T') = 0,$$

we can choose C_δ and $T = T_n$ large enough so that

$$\frac{1}{T_n'}\int_0^{T_n'} \Omega_{v_{\nu(T_n)}(t)}(C_\delta T)\,\mathrm{d}t \ll \frac{1}{n}.$$

Note that C_δ can be chosen independently of n. Next we estimate the first term on the right-hand side of (4-27). We treat only the case where the inner product is evaluated at $t = T'$, as the case when it is evaluated at $t = 0$ is similar. We have

$$\frac{1}{T'}|\langle \partial_t v_{\nu(T)}(T') \mid \chi_{2T}\cdot r\,\partial_r v_{\nu(T)}(T')\rangle| \lesssim \frac{T^{\frac12}}{T'}\|\partial_t v_{\nu(T)}(T')\|_{L^2}\|\nabla v_{\nu(T)}(T')\|_{L^2(|x|\le T^{1/2})}$$
$$+ \frac{C_\delta}{c_\delta}\|\partial_t v_{\nu(T)}(T')\|_{L^2}\|\nabla v_{\nu(T)}(T')\|_{L^2(T^{1/2}\le|x|\le C_\delta T)}.$$

Since $T' \simeq_\delta T$, the first term on the right-hand side above can be made as small as we like by choosing T_n large enough so that

$$\frac{T_n^{\frac12}}{T_n'} \ll \frac{1}{n}.$$

Similarly, for the second term on the right, we rely on the precompactness of K' in $\dot{\mathcal{H}}^1 \cap \dot{\mathcal{H}}^{s_p}$ and the fact that $x'(T_n') = 0$, which yields

$$\|\nabla v_{\nu(T_n)}(T_n')\|_{L^2(|x|\ge T_n^{1/2})} \ll \frac{1}{n}$$

for T_n large enough. The second term on the right-hand-side of (4-27) is estimated in a similar fashion. This completes the proof of Claim 4.9. ∎

Now, given this sequence of times T_n guaranteed by Claim 4.9 consider the sequence $\nu(T_n) := x_1(T_n)/T_n$. By (4-24) we can, passing to subsequence that we still denote by $\nu(T_n)$, find a fixed $\nu \in [-1-\delta, 1-\delta]$ with

$$\nu(T_n) \to \nu_0 \quad \text{as } n \to \infty. \tag{4-28}$$

Define

$$v_{\nu_0}(t', x') := u \circ L_{\nu_0}(t, x)$$

and note that this is a *fixed* Lorentz transform of u. It follows from Claim 4.9, (4-28), and a continuity argument that in fact

$$\frac{1}{T'_n} \int_0^{T'_n} \int_{\mathbb{R}^3} |\partial_t v_{\nu_0}(t, x)|^2 + |v_{\nu_0}(t, x)|^{p+1} \, \mathrm{d}x \, \mathrm{d}t < \frac{1}{2n}$$

after passing to a further subsequence. Using yet another continuity argument we can assume without loss of generality that $T'_n = M_n \in \mathbb{N}$; i.e.,

$$\frac{1}{M_n} \int_0^{M_n} \int_{\mathbb{R}^3} |\partial_t v_{\nu_0}(t, x)|^2 + |v_{\nu_0}(t, x)|^{p+1} \, \mathrm{d}x \, \mathrm{d}t < \frac{1}{n} \tag{4-29}$$

for some sequence $\{M_n\} \subset \mathbb{N}$ with $M_n \to \infty$. Now we claim that there exists a sequence of positive integers $m_n \to \infty$ such that

$$\int_{m_n}^{m_n+1} \int_{\mathbb{R}^3} |\partial_t v_{\nu_0}(t, x)|^2 + |v_{\nu_0}(t, x)|^{p+1} \, \mathrm{d}x \, \mathrm{d}t \to 0 \quad \text{as } n \to \infty. \tag{4-30}$$

If not, we could find $\epsilon_0 > 0$ such that for all $n \in \mathbb{Z}$ we have

$$\int_{m}^{m+1} \int_{\mathbb{R}^3} |\partial_t v_{\nu_0}(t, x)|^2 + |v_{\nu_0}(t, x)|^{p+1} \, \mathrm{d}x \, \mathrm{d}t \geq \epsilon_0.$$

However, summing up from 0 to $M_n - 1$ we would then have

$$\int_0^{M_n} \int_{\mathbb{R}^3} |\partial_t v_{\nu_0}(t, x)|^2 + |v_{\nu_0}(t, x)|^{p+1} \, \mathrm{d}x \, \mathrm{d}t \geq \epsilon_0 M_n,$$

which contradicts (4-29). Now, by (4-30) we have

$$\int_0^1 \int_{\mathbb{R}^3} |\partial_t v_{\nu_0}(m_n + t, x)|^2 + |v_{\nu_0}(m_n + t, x)|^{p+1} \, \mathrm{d}x \, \mathrm{d}t \to 0 \tag{4-31}$$

as $n \to \infty$. On the other hand, passing to a further subsequence, we can find $(V_0, V_1) \in \dot{\mathcal{H}}^1 \cap \dot{\mathcal{H}}^{s_p}$ such that

$$\vec{v}_{\nu_0}(m_n, \cdot - x'(m_n)) \to (V_0, V_1) \in \dot{\mathcal{H}}^1 \cap \dot{\mathcal{H}}^{s_p} \quad \text{as } n \to \infty.$$

Let $\vec{V}(t)$ be the solution to (1-1) with data (V_0, V_1). Then for some $t_0 > 0$ sufficiently small we have

$$\lim_{n \to \infty} \sup_{t \in [0, t_0]} \|\vec{v}_{\nu_0}(m_n + t, \cdot - x'(m_n)) - \vec{V}(t)\|_{\dot{\mathcal{H}}^1 \cap \dot{\mathcal{H}}^{s_p}} = 0. \tag{4-32}$$

However, from (4-31) we can then conclude that

$$\vec{V} \equiv 0,$$

from which we conclude from (4-32) and small-data arguments that

$$\vec{v}_{\nu_0} \equiv 0.$$

This means $\vec{u} \equiv 0$ as well, which finishes the proof. □

Proof of Proposition 4.7 for the defocusing equation. The argument is much easier if either $p = 3$ or if the equation is defocusing since (4-21) gives us coercive control over the energy. Indeed, arguing as in the proof of Claim 4.9, but using (4-21) instead of (4-22) we see that

$$E(\vec{v}_{\nu(T)}) = \frac{1}{T'} \int_0^{T'} E(\vec{v}_{\nu(T)})\, dt = o(1) \quad \text{as } T \to \infty$$

since, for each fixed T, the energy of $v_{\nu(T)}(t)$ is constant in time. However, since

$$v_{\nu(T)}(t', x') = u \circ L_{\nu(T)}(t, x),$$

we must have either $\limsup_{T\to\infty} |\nu(T)| = 1$, or $E(\vec{u}) = 0$. The former is impossible by (4-24). Hence $E(\vec{u}) = 0$. Therefore $\vec{u} \equiv 0$. □

Remark 4.10. Note the argument given above for the defocusing equation also works for the cubic focusing equation since (4-21) yields control of the full energy for $p = 3$. Arguing as above one can conclude that $E(u) = 0$. Since the only nonzero solutions with zero energy must blow up in both time directions [Killip et al. 2014] we conclude that the global-in-time solution satisfies $\vec{u} \equiv 0$; see [Dodson and Lawrie 2015b], where a version of this argument was carried out in detail.

5. The self-similar critical element

In this section, we assume towards a contradiction that $\vec{u}$ is a self-similar-like critical element as in Proposition 3.13, case (III). We will prove that any such $\vec{u}$ has finite energy, and in fact that $E(\vec{u}) = 0$. Since we are treating the defocusing equation, this implies $\vec{u} \equiv 0$. The arguments in this section can be readily adapted to the focusing setting as well.

More precisely, we will prove the following result.

Proposition 5.1. *There are no self-similar-like critical elements in the sense of case* (II) *of Proposition 3.13.*

As in Section 4, we will prove this proposition via two additional regularity arguments. We fix the following notation: let

$$s_0 = s_p + \frac{5-p}{2p(p-1)} = \frac{3}{2} - \frac{5}{2p}. \tag{5-1}$$

Proposition 5.2. *Let $\vec{u}$ be a self-similar-like critical element as in Proposition 3.13. Then,*

$$\|\vec{u}(T)\|_{\dot{\mathcal{H}}^{s_0}} \lesssim T^{-(s_0 - s_p)}$$

uniformly in $T > 0$.

Proposition 5.3. *Let $\vec{u}$ be a self-similar-like critical element as in Proposition 3.13. Let s_0 be as in (5-1) and suppose that*

$$\|\vec{u}(T)\|_{\dot{\mathcal{H}}^{s_0}} \lesssim T^{-(s_0-s_p)} \tag{5-2}$$

uniformly in $T > 0$. Then

$$\|\vec{u}(T)\|_{\dot{\mathcal{H}}^1} \lesssim T^{-p(s_0-s_p)}$$

uniformly in $T > 0$.

Proposition 5.3 will immediately imply Proposition 5.1.

Proof of Proposition 5.1 assuming Proposition 5.3. Note that the nonlinear component of the energy is controlled by the $\dot{H}^{3/2-3/(p+1)}(\mathbb{R}^3)$ norm by Sobolev embedding, and by interpolation we have

$$\dot{H}^{\frac{3}{2}-\frac{3}{p+1}}(\mathbb{R}^3) \subseteq \dot{H}^{s_p} \cap \dot{H}^1.$$

Thus the conserved energy $E(\vec{u})$ must be zero by sending $T \to \infty$ in Proposition 5.3. Then $E[\vec{u}] \equiv 0$, which implies that $\vec{u} \equiv 0$, which is impossible. □

Proposition 5.3 is the easier of the two additional regularity arguments, so we turn to this first.

5A. *The jump from $\dot{\mathcal{H}}^{s_0}(\mathbb{R}^3)$ to $\dot{\mathcal{H}}^1(\mathbb{R}^3)$ regularity.* We first prove that if $\vec{u}$ has some additional regularity, then we can achieve $\dot{\mathcal{H}}^1$ regularity, and hence reach the desired contradiction.

Proof of Proposition 5.3. Using $N(t) = t^{-1}$, we have

$$\|u\|_{L^{2(p-1)}_{t,x}([2^k,2^{k+1}]\times\mathbb{R}^3)} \lesssim 1$$

uniformly in k. Thus for any $0 < \eta \ll 1$, we can partition $[2^k, 2^{k+1}]$ into $C(\eta)$ intervals I_j so that

$$\|u\|_{L^{2(p-1)}_{t,x}(I_j\times\mathbb{R}^3)} < \eta.$$

On each such interval, we may argue using Strichartz estimates and a continuity argument together with (5-2) to deduce that

$$\|u\|_{L^p_t L^{2p}_x(I_j\times\mathbb{R}^3)} \lesssim 2^{-k(s_0-s_p)}$$

for each j. This implies

$$\|u\|_{L^p_t L^{2p}_x([2^k,2^{k+1}]\times\mathbb{R}^3)} \lesssim 2^{-k(s_0-s_p)}$$

uniformly in k. We once again complexify the solution. We let

$$w = u + \frac{i}{\sqrt{-\Delta}} u_t.$$

Once again, if $\vec{u}(t)$ solves (1-1), then $w(t)$ is a solution to

$$w_t = -i\sqrt{-\Delta} w \pm \frac{i}{\sqrt{-\Delta}} |u|^{p-1} u.$$

By compactness,

$$\lim_{T\to\infty} P_{\le k} e^{iT\sqrt{-\Delta}} w(-T) = 0$$

as weak limits in $\dot{H}^{s_0}$ for any $k \geq 0$. By Strichartz estimates, we have

$$\begin{aligned}\|P_{\leq k} w(T)\|_{\dot{H}^1} &\lesssim \|u|u|^{p-1}\|_{L_t^1 L_x^2([T,\infty)\times\mathbb{R}^3)} \\ &\lesssim \sum_{2^k \geq \frac{T}{2}} \|u|u|^{p-1}\|_{L_t^1 L_x^2([2^k,2^{k+1}]\times\mathbb{R}^3)} \lesssim \sum_{2^k \geq \frac{T}{2}} 2^{-kp(s_0-s_p)} \lesssim T^{-p(s_0-s_p)},\end{aligned}$$

which completes the proof. □

The jump from $\dot{\mathcal{H}}^{s_p}$ to $\dot{\mathcal{H}}^{s_0}$ regularity. It remains to prove Proposition 5.2. The main technical ingredient in the proof of Proposition 5.2 is a long-time Strichartz estimate.

Proposition 5.4 (long-time Strichartz estimate). *Let $\alpha \geq 1$ and*

$$2 < q < \frac{2}{s_p}.$$

Suppose $\vec{u}$ is a self-similar-like critical element as in Proposition 3.13 with compactness modulus function $R(\cdot)$. For any $\eta_0 > 0$, there exists $k_0 = k_0(R(\eta_0), \alpha)$ so that, for every $k > k_0$,

$$\||\nabla|^{\frac{3(p-3)}{2(p-1)}} u_{>k}\|_{L_t^{2(p-1)} L_x^{2(p-1)/(p-2)}([1,2^{\alpha(k-k_0)}]\times\mathbb{R}^3)} + \||\nabla|^{-(\frac{2}{q}-s_p)} u_{>k}\|_{L_t^q L_x^{2q/(q-2)}([1,2^{\alpha(k-k_0)}]\times\mathbb{R}^3)} < \eta_0.$$

Proof. We proceed by induction on $k > k_0$. Let $\eta_0 > 0$. Using compactness and the fact that $N(t) = t^{-1}$, we may find k_0 large enough that

$$\||\nabla|^{\frac{3(p-3)}{2(p-1)}} u_{>k_0}\|_{L_t^{2(p-1)} L_x^{2(p-1)/(p-2)}([1,2^{3\alpha}]\times\mathbb{R}^3)} + \||\nabla|^{-(\frac{2}{q}-s_p)} u_{>k_0}\|_{L_t^q L_x^{2q/(q-2)}([1,2^{3\alpha}]\times\mathbb{R}^3)} < \tfrac{1}{2}\eta_0.$$

This implies the result for $k_0 \leq k \leq 8k_0$.

To establish the induction step, by Taylor's theorem, we may take the decomposition

$$\begin{aligned}F(u) &= F(u_{>k-3}) + u_{\leq k-3} \int_0^1 F'(\theta u_{\leq k-3} + u_{>k-3}) \\ &= F(u_{>k-3}) + u_{\leq k-3} F'(u_{>k-3}) + u_{\leq k-3}^2 \int_0^1 \int_0^1 F''(\theta_1\theta_2 u_{<k-3} + u_{>k-3})\, d\theta_1\, d\theta_2.\end{aligned}$$

Hence, we can write the nonlinearity $F(u)$ as a sum of terms

$$P_{>k} F(u) = |P_{>k-3} u|^{p-1} P_{>k-3} u + P_{>k}(u_{\leq k-3} F'(u_{>k-3})) + P_{>k}(u_{\leq k-3}^2 P_{>k-3} F_2),$$

where

$$F_2 = \int_0^1 \int_0^1 F''(\theta_1\theta_2 u_{<k-3} + u_{>k-3})\, d\theta_1\, d\theta_2,$$

and we have used in the last term that

$$P_{>k}(u_{\leq k-3}^2 F_2) = P_{>k}(u_{\leq k-3}^2 P_{>k-3} F_2).$$

Note that $|F'(u_{>k-3})| \lesssim |u_{>k-3}|^{p-1}$ and $|F_2| \lesssim |u_{<k-3}|^{p-2} + |u_{>k-3}|^{p-2}$, and since the frequency projections are bounded on L^p, we will replace these terms with $|u|^{p-1}$ and $|u|^{p-2}$ respectively once we have a chosen a dual space in order to simplify the exposition of our estimates.

Fix exponents

$$\gamma = \frac{q}{2}, \quad \rho = \frac{6(-q+pq)}{12-12p-21q+13pq},$$

and note that $\gamma \in (1,2)$ for $q \in (2,4)$, while for $q = 2$, we have

$$\rho = \frac{6(p-1)}{7p-15} \in \left(\frac{6}{5}, 2\right).$$

In particular, by choosing q close to 2 we can guarantee that $\gamma, \rho \in (1,2)$. Furthermore,

$$\frac{1}{\gamma} + \frac{1}{\rho} - \frac{3}{2} = \frac{2(p-3)}{3(p-1)} \geq 0,$$

which guarantees that the conjugate exponent pair (γ', ρ') is wave-admissible.

By Strichartz estimates,

$$\begin{aligned}
&\||\nabla|^{\frac{3(p-3)}{2(p-1)}} u_{>k}\|_{L_t^{2(p-1)} L_x^{2(p-1)/(p-2)}} + \||\nabla|^{-(\frac{2}{q}-s_p)} u_{>k}\|_{L_t^q L_x^{2q/(q-2)}} \\
&\qquad \lesssim \|u_{>k}(1)\|_{\mathcal{H}^{s_p}} + \||\nabla|^{\frac{3(p-3)}{2(p-1)}} [u_{>k-3}]^p\|_{L_t^{2(p-1)/p} L_x^{2(p-1)/(2p-3)}} \\
&\qquad\quad + \||\nabla|^{-(\frac{2}{q}-s_p)} P_{>k}(u_{\leq k-3} u_{>k-3}^{p-1})\|_{L_t^{2q/(q+2)} L_x^{q/(q-1)}} \\
&\qquad\quad + \||\nabla|^{-\frac{4}{q}+3s_p} P_{>k}(u_{\leq k-3}^2 P_{>k}(u^{p-2}))\|_{L_t^\gamma L_x^\rho} \\
&\qquad := I + II + III,
\end{aligned}$$

where all space-time norms are over $[1, 2^{\alpha(k-k_0)}] \times \mathbb{R}^3$. We estimate term I as follows:

$$\begin{aligned}
&\||\nabla|^{\frac{3(p-3)}{2(p-1)}} [u_{>k-3}]^p\|_{L_t^{2(p-1)/p} L_x^{2(p-1)/(2p-3)}} \\
&\qquad \lesssim \|u_{>k-3}\|_{L_{t,x}^{2(p-1)}}^{p-1} \||\nabla|^{\frac{3(p-3)}{2(p-1)}} u_{>k-3}\|_{L_t^{2(p-1)} L_x^{2(p-1)/(p-2)}} \\
&\qquad \lesssim \||\nabla|^{\frac{3(p-3)}{2(p-1)}} u_{>k-3}\|_{L_t^{2(p-1)} L_x^{2(p-1)/(p-2)}}^p.
\end{aligned}$$

By induction, we have

$$\||\nabla|^{\frac{3(p-3)}{2(p-1)}} u_{>k-3}\|_{L_t^{2(p-1)} L_x^{2(p-1)/(p-2)}([1, 2^{\alpha(k-k_0)}/8] \times \mathbb{R}^3)} \leq \eta_0.$$

Thus, using $N(t) = t^{-1}$ and that

$$\int_{2^{\alpha(k-k_0)}/8}^{2^{\alpha(k-k_0)}} t^{-1}\, \mathrm{d}t = \log 2^{3\alpha} \sim 1,$$

and the fact that $N(t) \leq 1$ on $[2^{\alpha(k-k_0)}/8,\ 2^{\alpha(k-k_0)}]$ for $k > k_0 \gg 1$, we can deduce

$$\||\nabla|^{\frac{3(p-3)}{2(p-1)}} u_{>k-3}\|_{L_t^{2(p-1)} L_x^{2(p-1)/(p-2)}([2^{\alpha(k-k_0)}/8,\, 2^{\alpha(k-k_0)}] \times \mathbb{R}^3)} \leq \eta_0.$$

In particular, using (5-3), we obtain that

$$\||\nabla|^{\frac{3(p-3)}{2(p-1)}} [u_{>k-3}]^p\|_{L_t^{2(p-1)/p} L_x^{2(p-1)/(2p-3)}([1, 2^{\alpha(k-k_0)}] \times \mathbb{R}^3)} \lesssim \eta_0^p.$$

For term II, we estimate

$$2^{-k(\frac{2}{q}-s_p)}\|u_{\le k-3}\|_{L_t^q L_x^{2q/(q-2)}}\|u_{>k-3}\|^{p-1}_{L^{2(p-1)}_{t,x}} \lesssim \eta_0^{p-1}2^{-k(\frac{2}{q}-s_p)}\|u_{\le k-3}\|_{L_t^q L_x^{2q/(q-2)}}.$$

Fix $C_0 \ge 1$ to be determined below. We write

$$\begin{aligned}\|u_{\le k-3}\|_{L_t^q L_x^{2q/(q-2)}([1,2^{\alpha(k-k_0)}]\times\mathbb{R}^3)} &\lesssim \|u_{\le C_0}\|_{L_t^q L_x^{2q/(q-2)}([1,2^{\alpha(k-k_0)}]\times\mathbb{R}^3)}\\ &\quad+\|u_{C_0<\,\cdot\,\le k_0}\|_{L_t^q L_x^{2q/(q-2)}([1,2^{\alpha(k-k_0)}]\times\mathbb{R}^3)} \qquad (5\text{-}3)\\ &\quad+\sum_{k_0\le j\le k-3}\|u_j\|_{L_t^q L_x^{2q/(q-2)}([1,2^{\alpha(k-k_0)}]\times\mathbb{R}^3)}. \qquad (5\text{-}4)\end{aligned}$$

For (5-3), we have

$$\|u_{C_0<\,\cdot\,\le k_0}\|_{L_t^q L_x^{2q/(q-2)}([1,2^{\alpha(k-k_0)}]\times\mathbb{R}^3)} \lesssim C_0^{\frac{2}{q}-s_p}\log(2^{k-k_0}).$$

On the other hand, for $C_0 = C_0(\eta_0)$ large enough, we can estimate (5-4) by

$$\sum_{k_0\le j\le k-3}\|u_j\|_{L_t^q L_x^{2q/(q-2)}([1,2^{\alpha(k-k_0)}]\times\mathbb{R}^3)} \lesssim \eta_0 2^{k_0(\frac{2}{q}-s_p)}\log(2^{k-k_0}).$$

Finally, for $k_0 \le j \le k-3$ we first use the inductive hypothesis to write

$$\|P_j u\|_{L_t^q L_x^{2q/(q-2)}([1,2^{\alpha(j-k_0)}]\times\mathbb{R}^3)} \lesssim 2^{j(\frac{2}{q}-s_p)}\eta_0.$$

Arguing as we did for the high-frequency piece,

$$\|P_M u\|_{L_t^q L_x^{2q/(q-2)}([2^{\alpha(j-k_0)},2^{\alpha(k-k_0)}]\times\mathbb{R}^3)} \lesssim 2^{j(\frac{2}{q}-s_p)}\eta_0\log(2^{k-j}).$$

Thus

$$\sum_{k_0\le j\le k-3}\|u_j\|_{L_t^q L_x^{2q/(q-2)}([1,2^{\alpha(k-k_0)}]\times\mathbb{R}^3)} \lesssim \sum_{k_0\le j\le k-3}\eta_0 2^{j(\frac{2}{q}-s_p)}[1+\log(2^{k-j})] \lesssim \eta_0 2^{k(\frac{2}{q}-s_p)},$$

where we have used

$$\sum_{L>1} L^{-(\frac{2}{q}-s_p)}\log(L) \lesssim 1.$$

Collecting these estimates, we find

$$\|u_{\le k-3}\|_{L_t^q L_x^{2q/(q-2)}([1,2^{\alpha(k-k_0)}]\times\mathbb{R}^3)} \lesssim [C_0+\eta_0 2^{k_0(\frac{2}{q}-s_p)}]\log(2^{k-k_0})+\eta_0 2^{k(\frac{2}{q}-s_p)}, \qquad (5\text{-}5)$$

which yields

$$\begin{aligned}&\||\nabla|^{-(\frac{2}{q}-s_p)}P_{>k}(u_{\le k-3}u_{>k-3}^{p-1})\|_{L_t^{2q/(q+2)}L_x^{q/(q-1)}}\\ &\qquad\lesssim \eta_0^{p-1}2^{-k(\frac{2}{q}-s_p)}[C_0^{\frac{2}{q}-s_p}+\eta_0 2^{k_0(\frac{2}{q}-s_p)}]\log(2^{k-k_0})+\eta_0^p.\end{aligned}$$

Choosing k_0 possibly even larger, we deduce

$$\||\nabla|^{-(\frac{2}{q}-s_p)}P_{>k}(u_{\le k-3}u_{>k-3}^{p-1})\|_{L_t^{2q/(q+2)}L_x^{q/(q-1)}} \lesssim \eta_0^p.$$

Finally, we estimate term III. Since $-\frac{2}{q}+s_p<0$, we use the fractional chain rule and Bernstein estimates to obtain

$$\begin{aligned}\||\nabla|^{-\frac{4}{q}+3s_p}P_{>k}(u_{\le k-3}^2P_{>k}(u^{p-2}))\|_{L_t^\gamma L_x^\rho}&\\ &\lesssim 2^{-k\frac{4}{q}+2s_p}\|u_{\le k-3}\|^2_{L_t^qL_x^{2q/(q-2)}}\||\nabla|^{s_p}(u^{p-2})\|_{L_t^\infty L_x^{6(p-1)/(7p-15)}}\\ &\lesssim 2^{-2k(\frac{2}{q}-s_p)}\|u_{\le k-3}\|^2_{L_t^qL_x^{2q/(q-2)}}\|u\|^{p-3}_{L_t^\infty L_x^{3(p-1)/2}}\||\nabla|^{s_c}u\|_{L_t^\infty L_x^2}.\end{aligned}$$

Using (5-5) (and the conditions on k_0, C_0 given above), we conclude

$$\||\nabla|^{-\frac{4}{q}+3s_p}P_{>k}(u_{\le k-3}^2P_{>k}(u^{p-2}))\|_{L_t^\gamma L_x^\rho}\lesssim\eta_0^2.$$

Combining our estimates for terms I, II and III and choosing η_0 small, we conclude that

$$\||\nabla|^{\frac{3(p-3)}{2(p-1)}}u_{>k}\|_{L_t^{2(p-1)}L_x^{2(p-1)/(p-2)}}+\||\nabla|^{-(\frac{2}{q}-s_p)}u_{>k}\|_{L_t^qL_x^{2q/(q-2)}}\le\eta_0$$

on $[1,2^{\alpha(k-k_0)}]\times\mathbb{R}^3$, thus closing the induction and completing the proof. □

Finally, we arrive at the proof of the additional regularity Proposition 5.2.

Proof of Proposition 5.2. We compute

$$\|\vec{u}(1)\|^2_{\dot{\mathcal{H}}^{s_0}}\simeq\|w(1)\|^2_{\dot{H}^{s_0}}\lesssim\sum_{k\gg1}2^{2ks_0}\langle P_kw(1),P_kw(1)\rangle.$$

We use the double Duhamel argument based at $t=1$. For some $k\gg1$, we write

$$\langle P_kw(1),P_kw(1)\rangle=\int_0^1\int_1^\infty\langle e^{i(1-t)\sqrt{-\Delta}}P_kF(u(t)),e^{i(1-\tau)\sqrt{-\Delta}}P_kF(u(\tau))\rangle\,\mathrm{d}\tau\,\mathrm{d}t.$$

We fix $\alpha\ge1$, to be determined below, and split

$$\int_1^\infty e^{i(1-t)\sqrt{-\Delta}}P_kF(u(t))\,\mathrm{d}t=A_k+B_k,$$

where

$$A_k=\int_1^{2^{k\alpha}}e^{i(1-t)\sqrt{-\Delta}}P_kF(u(t))\,\mathrm{d}t,\qquad B_k=\int_{2^{k\alpha}}^\infty e^{i(1-t)\sqrt{-\Delta}}P_kF(u(t))\,\mathrm{d}t.$$

We also write

$$\int_0^1e^{i(1-\tau)\sqrt{-\Delta}}P_kF(u(s))\,\mathrm{d}s=Z_k.$$

We will use the estimate

$$|\langle A_k+B_k,Z_k\rangle|\le|A_k|^2+2|\langle B_k,Z_k\rangle|,$$

which follows from the fact that $A_k+B_k=Z_k$.

We first estimate the $\langle B_k,Z_k\rangle$ term. We expand

$$|\langle B_k,Z_k\rangle|\le\sum_{\ell\le0}\sum_{j\ge k\alpha}\int_{2^\ell}^{2^{\ell+1}}\int_{2^j}^{2^{j+1}}\big|\langle e^{-i(t-\tau)\sqrt{-\Delta}}P_kF(u(t)),P_kF(u(s))\rangle\big|\,d\tau\,\mathrm{d}t.$$

We claim that

$$\|P_k(u|u|^{p-1})\|_{L_t^2 L_x^1([2^\ell,2^{\ell+1}]\times\mathbb{R}^3)} \lesssim 2^{-ks_p}$$

uniformly in $\ell \geq 0$. Indeed, arguing as above, we can decompose the nonlinearity into two types of terms

$$u_{>k-1}u^{p-1} \quad \text{and} \quad uP_{>k-1}(u^{p-1}),$$

since if both u and $|u|^{p-1}$ are projected to low frequencies, the product vanishes when projected to high frequencies.

We thus have by Bernstein's inequality, Hölder's inequality, and the fractional chain rule that

$$\begin{aligned}\|P_k(u|u|^{p-1})\|_{L_t^2 L_x^1} &\lesssim \|u\|^{p-1}_{L^{2(p-1)}_{t,x}}\|u_{>k-1}\|_{L_t^\infty L_x^2} + \|u\|_{L^{2(p-1)}_{t,x}}\|P_{>k-1}(u^{p-1})\|_{L_t^{2(p-1)/(p-2)}L_x^{2(p-1)/(2p-3)}}\\ &\lesssim 2^{-ks_p}\|u\|^{p-1}_{L^{2(p-1)}_{t,x}}\||\nabla|^{s_p}u\|_{L_t^\infty L_x^2} \lesssim 2^{-ks_p},\end{aligned}$$

where all space-time norms are over $[2^\ell, 2^{\ell+1}]\times\mathbb{R}^3$.

Using dispersive estimates, we have, for any $j \geq k\alpha$ and $\ell \leq 0$,

$$\begin{aligned}&\int_{2^\ell}^{2^{\ell+1}}\int_{2^j}^{2^{j+1}} |\langle e^{-i(t-\tau)\sqrt{-\Delta}}P_kF(u(t)), P_kF(u(s))\rangle|\,\mathrm{d}\tau\,\mathrm{d}t\\ &\qquad\lesssim \int_{2^\ell}^{2^{\ell+1}}\int_{2^j}^{2^{j+1}} t^{-1}2^k\|P_k(u|u|^{p-1})(t)\|_{L_x^1}\|P_k(u|u|^{p-1})(\tau)\|_{L_x^1}\,\mathrm{d}\tau\,\mathrm{d}t\\ &\qquad\lesssim 2^{\frac{\ell}{2}}2^{-\frac{j}{2}}\|P_k(u|u|^{p-1})\|_{L_t^2L_x^1([2^\ell,2^{\ell+1}]\times\mathbb{R}^3)}\|P_k(u|u|^{p-1})\|_{L_t^2L_x^1([2^k,2^{k+1}]\times\mathbb{R}^3)}\\ &\qquad\lesssim 2^{\frac{\ell}{2}}2^{-\frac{j}{2}}2^{k(1-2s_p)}.\end{aligned}$$

Summing over $\ell \leq 0$ and $j \geq k\alpha$, we deduce that

$$|\langle B_k, Z_k\rangle_{\dot{H}_x^{s_p}}| \lesssim 2^{k(1-\frac{\alpha}{2})}. \tag{5-6}$$

We now turn to estimating the $|A_k|^2$ term. We will use a frequency envelope argument to establish the required bounds. Once again, we fix an exponent q satisfying

$$2 < q < \frac{s_p}{2}.$$

Let

$$\sigma < \min\Big\{s_p, \frac{2}{q} - s_p, \frac{4}{q} - 1 - s_p\Big\}. \tag{5-7}$$

and define

$$\gamma_k = \sum_j 2^{-\sigma|j-k|}\|w_j\|_{L_t^\infty \dot{H}^{s_p}([1,\infty)\times\mathbb{R}^3)}.$$

We will establish the following: Let $\eta_0 > 0$ and let $R(\cdot)$ denote the compactness modulus function of $\vec{u}$. Then there exists $k_0 \equiv k_0(\eta_0, R(\eta_0))$ sufficiently large that

$$\|A_k\|_{\dot{H}_x^{s_p}} \lesssim C(k_0)2^{-k(\frac{2}{q}-s_p)+} + \eta_0\sum_j 2^{-\sigma|j-k|}\|u_j\|_{L_t^\infty \dot{H}_x^{s_p}([1,\infty)\times\mathbb{R}^3)} \tag{5-8}$$

for all $k \geq k_0$. For $p > 3$, we write the nonlinearity as $F(u) = |u|^{p-3}u^3$ and then decompose u^3 by writing $u = u_{\leq k} + u_{>k}$. By further decomposing $u_{\leq k} = u_{\leq k_0} + u_{k_0 < \cdot \leq k}$, we are led to terms of the form

$$F(u) = |u|^{p-3}u_{>k}^3 \tag{5-9}$$
$$+ 3|u|^{p-3}u_{>k}^2 u_{\leq k_0} \tag{5-10}$$
$$+ 3|u|^{p-3}u_{>k}^2 u_{k_0 < \cdot \leq k} \tag{5-11}$$
$$+ |u|^{p-3}u_{\leq k_0} F_3 \tag{5-12}$$
$$+ |u|^{p-3}u_{k_0 < \cdot \leq k}^3 \tag{5-13}$$
$$+ 3|u|^{p-3}u_{>k} u_{\leq k} u_{\leq k_0} \tag{5-14}$$
$$+ 3|u|^{p-3}u_{>k} u_{\leq k} u_{k_0 < \cdot \leq k}, \tag{5-15}$$

where we have written

$$F_3 = u_{\leq k_0}^2 + 2u_{\leq k_0}u_{k_0 < \cdot \leq k} + u_{k_0 < \cdot \leq k}^2.$$

By Proposition 5.4, for any $\beta \geq 1$ there exists $k_0 \equiv k_0(R(\eta_0), \beta)$ so that for every $k > k_0$ we have

$$\| |\nabla|^{\frac{3(p-3)}{2(p-1)}} u_{>k} \|_{L_t^{2(p-1)} L_x^{2(p-1)/(p-2)}([1, 2^{\beta(k-k_0)}] \times \mathbb{R}^3)} + \| |\nabla|^{-(\frac{2}{q} - s_p)} u_{>k} \|_{L_t^q L_x^{2q/(q-2)}([1, 2^{\beta(k-k_0)}] \times \mathbb{R}^3)} < \eta_0.$$

Fix $\beta > \alpha$ and $k_1 = k_1(R(\eta_0), \alpha, \beta) \geq k_0$, which satisfies

$$2^{k_1(\beta - \alpha)} \geq 2^{k_0 \beta}.$$

Then $2^{\beta(k-k_0)} \geq 2^{k\alpha}$ for $k \geq k_1$, and hence, for every $k \geq k_1$, we have

$$\| |\nabla|^{\frac{3(p-3)}{2(p-1)}} u_{>k} \|_{L_t^{2(p-1)} L_x^{2(p-1)/(p-2)}([1, 2^{\alpha k}] \times \mathbb{R}^3)} + \| |\nabla|^{-(\frac{2}{q} - s_p)} u_{>k} \|_{L_t^q L_x^{2q/(q-2)}([1, 2^{\alpha k}] \times \mathbb{R}^3)} < \eta_0.$$

We will use this estimate repeatedly below. Furthermore, we may also establish identical long-time Strichartz estimates for

$$\| |\nabla|^{s_p - \frac{2}{r}} u_{>k} \|_{L_t^r L_x^{2r/(r-2)}},$$

where $2/s_p < r < 4$.

To estimate (5-9), we use the dual Strichartz pair

$$\left(\frac{r}{2}, \frac{6r(p-1)}{12 - 12p - 21r + 13pr} \right),$$

with $2/s_p < r < 4$. We note that this pair is dual admissible: writing the pair as (A, B), we have

$$\frac{1}{A} + \frac{1}{B} = \frac{13p - 21}{6p - 6} \geq \frac{3}{2}$$

for $p \geq 3$. Note that $A \in (1, 2)$ since $r \in (2, 4)$ and $B > 1$ for

$$r < \frac{12(p-1)}{7p - 15}.$$

This is compatible with $r > \frac{2}{s_p}$ when $p \in [3, 5)$. We can thus bound

$$2^{k(3s_p-\frac{4}{r})}\|u\|^{p-3}_{L_t^\infty L_x^{3(p-1)/2}}\|u_{>k}\|^2_{L_t^r L_x^{2r/(r-2)}}\sum_{k\le j}\|u_j\|_{L_t^\infty L_x^2} \lesssim \||\nabla|^{s_p-\frac{2}{r}}u_{>k}\|^2_{L_t^r L_x^{2r/(r-2)}}\sum_{k\le j}2^{(k-j)s_p}\|u_j\|_{L_t^\infty \dot H_x^{s_p}} \lesssim \eta_0\sum_{j>k}2^{(k-j)s_p}\|u_j\|_{L_t^\infty \dot H_x^{s_p}}.$$

For (5-10), we use the dual Strichartz pair

$$\left(\frac{2q(p-1)}{2p+q-2}, \frac{6q(p-1)}{6-15q+2p(5q-3)}\right). \tag{5-16}$$

We bound the contribution of this term by

$$2^{k(2s_p-\frac{2}{q})}\|u\|^{p-3}_{L_t^\infty L_x^{3(p-1)/2}}\|u_{>k}\|_{L_{t,x}^{2(p-1)}}\|u_{\le k_0}\|_{L_t^q L_x^{2q/(q-2)}}\|u_{>k}\|_{L_t^\infty L_x^2} \lesssim \eta_0 2^{-k(\frac{2}{q}-s_p)}2^{k_0(\frac{2}{q}-s_p)}\log 2^k\|u\|_{L_t^\infty \dot H_x^{s_p}} \lesssim \eta_0 2^{-k(\frac{2}{q}-s_p)}2^{k_0(\frac{2}{q}-s_p)}\log 2^k.$$

For (5-11), we use the same dual pair as in (5-16), and we obtain

$$2^{k(2s_p-\frac{2}{q})}\|u\|^{p-3}_{L_t^\infty L_x^{3(p-1)/2}}\|u_{>k}\|_{L_{t,x}^{2(p-1)}}\sum_{k_0\le j_1\le k\le j_2}\|u_{j_1}\|_{L_t^q L_x^{2q/(q-2)}}\|u_{j_2}\|_{L_t^\infty L_x^2} \lesssim \eta_0\sum_{k_0\le j_1\le k\le j_2}2^{j_1(\frac{2}{q}-s_p)}\log(2^{k-j_1})2^{-j_2 s_p}\|u_{j_2}\|_{L_t^\infty \dot H_x^{s_p}} \lesssim \eta_0\sum_{k\le j}2^{(k-j)s_p}\|u_j\|_{L_t^\infty \dot H_x^{s_p}}.$$

To estimate (5-12), we use the admissible dual pair $\left(\frac{q}{2}, \frac{6q}{7q-8}+\right)$. We choose ρ so that

$$\frac{3}{\rho} = \frac{2}{q} + \frac{4}{p-1} - \frac{3}{2}.$$

We bound the contribution of this term by

$$2^{-k(\frac{2}{q}-s_p)}\|u\|^{p-3}_{L_t^\infty L_x^{3(p-1)/2}}\|u_{\le k_0}\|_{L_t^q L_x^{2q/(q-2)}}\sum_{j_1\le j_2\le k}\|u_{j_1}\|_{L_t^q L_x^{2q/(q-2)}}\|u_{j_2}\|_{L_t^\infty L_x^{\rho+}} \lesssim 2^{-k(\frac{2}{q}-s_p)}2^{k_0(\frac{2}{q}-s_p)}\log 2^k\sum_{j_1\le j_2\le k}2^{j_1(\frac{2}{q}-s_p)}2^{j_2(s_p-\frac{2}{q}+)}\log(2^{k-j_1}) \lesssim 2^{-k(\frac{2}{q}-s_p)+}.$$

Now we estimate (5-13) as follows: we apply Strichartz estimates with the dual (sharp) admissible pair $\left(\frac{q}{2}, \frac{2q}{3q-4}\right)$. Then we obtain

$$2^{k\left(1-\frac{4}{q}+s_p\right)}\|u\|^{p-3}_{L_t^\infty L_x^{3(p-1)/2}} \sum \|u_{j_1}\|_{L_t^q L_x^{2q/(q-2)}} \|u_{j_2}\|_{L_t^q L_x^{2q/(q-2)}} \|u_{j_3}\|_{L_t^\infty L_x^{6(p-1)/(9-p)}},$$

where the sum is over $k_0 \le j_1 \le j_2 \le j_3 \le k$.

Now, for $k_0 \le j \le k$, we can estimate

$$\begin{aligned}\|u_j\|_{L_t^q L_x^{2q/(q-2)}([1,2^{\alpha k}]\times\mathbb{R}^3)} &\lesssim \|u_j\|_{L_t^q L_x^{2q/(q-2)}([1,2^{\alpha j}]\times\mathbb{R}^3)} + \|u_j\|_{L_t^q L_x^{2q/(q-2)}([2^{\alpha k},2^{\alpha k}]\times\mathbb{R}^3)} \\ &\lesssim \eta_0 \log(2^{k-j})2^{j\left(\frac{2}{q}-s_p\right)},\end{aligned}$$

using the long-time Strichartz estimate of Proposition 5.4 and we note the log comes from the second term. We also have

$$\|u_j\|_{L_t^\infty L_x^{6(p-1)/(9-p)}} \lesssim 2^{-j(1-s_p)}\|u_j\|_{L_t^\infty \dot{H}_x^{s_p}}.$$

This yields

$$\begin{aligned}\eta_0 2^{k\left(1-\frac{4}{q}+s_p\right)} \sum_{k_0\le j_1\le j_2\le j_3\le k} 2^{j_1\left(\frac{2}{q}-s_p\right)}\log(2^{k-j_1})2^{j_2\left(\frac{2}{q}-s_p\right)}\log(2^{k-j_2})2^{-j_3(1-s_p)}& \\ \lesssim \eta_0 \sum_{k_0\le j\le k} \log(2^{j-k})2^{(j-k)\left(\frac{4}{q}-1-s_p\right)}\|u_j\|_{L_t^\infty \dot{H}_x^{s_p}}&.\end{aligned}$$

Note that for this estimate, we need

$$q < \frac{4}{1+s_p} = \frac{8(p-1)}{5p-9},$$

which is compatible with $q > 2$ for $p \in [3,5)$.

For (5-14), we use the dual Strichartz pair

$$\left(\frac{q}{2}, \frac{6(pq-q)}{12-12p-21q+13pq}\right). \tag{5-17}$$

We bound the contribution of this term by

$$\begin{aligned}2^{k\left(3s_p-\frac{4}{q}\right)}\|u_{\le k_0}\|_{L_t^q L_x^{2q/(q-2)}} &\sum_{j_1\le k\le j_2} \|u_{j_1}\|_{L_t^q L_x^{2q/(q-2)}}\|u_{j_2}\|_{L_t^\infty L_x^2} \\ &\lesssim 2^{k\left(3s_p-\frac{4}{q}\right)}2^{k_0\left(\frac{2}{q}-s_p\right)} \sum_{j_1\le k\le j_2} 2^{j_1\left(\frac{2}{q}-s_p\right)}\log(2^{k-j_1})2^{-j_2 s_p}\|u_{j_2}\|_{L_t^\infty \dot{H}_x^{s_p}} \\ &\lesssim 2^{-k\left(\frac{2}{q}-s_p\right)}2^{k_0\left(\frac{2}{q}-s_p\right)}.\end{aligned}$$

Finally, for (5-15), using the same dual pair as in (5-17), and we estimate the contribution of this term by

$$\begin{aligned}
2^{k(3s_p-\frac{4}{q})}\|u\|^{p-3}_{L_t^\infty L_x^{3(p-1)/2}} &\sum_{j_1,k_0\le j_2\le k\le j_3} \|u_{j_1}\|_{L_t^q L_x^{2q/(q-2)}}\|u_{j_2}\|_{L_t^q L_x^{2q/(q-2)}}\|u_{j_3}\|_{L_t^\infty L_x^2}\\
&\lesssim \eta_0 2^{k(3s_p-\frac{4}{q})} \sum_{j_1\le j_2\le k\le j_3} 2^{k_1(\frac{2}{q}-s_p)}\log(2^{2k-j_1-j_2})2^{j_2(\frac{2}{q}-s_p)}2^{-j_3 s_p}\|u_{j_3}\|_{L_t^\infty \dot H_x^{s_p}}\\
&\lesssim \eta_0 \sum_{k\le j} 2^{(k-j)s_p}\|u_j\|_{L_t^\infty \dot H_x^{s_p}}.
\end{aligned}$$

Putting together all the estimates, we establish (5-8), which, together with (5-6) and the conditions on σ from (5-7), yields

$$\|w_k(1)\|_{\dot H_x^{s_p}} \lesssim 2^{k(\frac{1}{2}-\frac{\alpha}{4})} + 2^{-k(\frac{2}{q}-s_p)+} + \eta_0\sum_j 2^{-\sigma|j-k|}\|w_j\|_{L_t^\infty \dot H_x^{s_p}([1,\infty)\times\mathbb{R}^3)}$$

for all $k \gg 1$. For α large enough, we can guarantee that the second term dominates the first, and hence

$$\|w_k(1)\|_{\dot H_x^{s_p}} \lesssim 2^{-k(\frac{2}{q}-s_p)+} + \eta_0\sum_j 2^{-\sigma|j-k|}\|w_j\|_{L_t^\infty \dot H_x^{s_p}([1,\infty)\times\mathbb{R}^3)}$$

for all $k \gg 1$. We now rescale the solution u and use the fact that the rescaled solution $Tu(Tt, Tx)$ is also a self-similar solution for any $T > 1$ (with the same compactness modulus function as u). This yields

$$\|w_k\|_{L_t^\infty \dot H^{s_p}([1,\infty)\times\mathbb{R}^3)} \lesssim 2^{-k(\frac{2}{q}-s_p)+} + \eta_0\sum_j 2^{-\sigma|j-k|}\|w_j\|_{L_t^\infty \dot H_x^{s_p}([1,\infty)\times\mathbb{R}^3)}. \tag{5-18}$$

Let $0 < \eta < \sigma$. Then (5-18) implies that for $k \gg k_0$,

$$\gamma_k \lesssim 2^{-k\eta} + \eta_0\alpha_k,$$

and hence, we may conclude that

$$\|w(1)\|_{\dot H^{s_p+\eta}} \lesssim 1 \quad \text{for any } 0 < \eta < \sigma.$$

Using the same rescaling argument as above, and the relation between w and u, we ultimately deduce that

$$\|u(T)\|_{\mathcal{H}^{s_p+\delta}} \lesssim T^{-\eta},$$

which yields (5-2) provided we can choose

$$\eta = \frac{5-p}{2p(p-1)}.$$

Combining with the constraint $\eta < \frac{2}{q} - s_p$, this requires that we choose

$$2 < q < \frac{4p}{3p-5},$$

which is possible whenever $p \in [3, 5)$. For the other term appearing in the definition of σ, we find that we can choose

$$\eta = \frac{5-p}{2p(p-1)}$$

provided we take

$$q < \frac{8p}{5(p-1)},$$

which is similarly allowable by the requirement that $q > 2$ for $p \in [3, 5)$. This completes the proof of Proposition 5.2 and hence completes our treatment of the self-similar scenario. $\square$

6. Doubly concentrating critical element: the sword and shield

We now consider the case of the doubly concentrating critical element, that is, $N(t) \geq 1$ on $\mathbb{R} = I_{\max}$ and

$$\limsup_{t\to\pm\infty} N(t) = \infty.$$

By Proposition 3.13 we may assume in this case that $x(t)$ is subluminal in the sense of Definition 3.12. By Lemma 3.16 there exists $\delta_0 > 0$ so that

$$|x(t) - x(\tau)| \leq (1-\delta_0)|t-\tau| \tag{6-1}$$

for all t, τ with

$$|t-\tau| \geq \frac{1}{\delta_0 \inf_{s\in[t,\tau]} N(s)}.$$

The goal of this section is to prove the following proposition:

Proposition 6.1. *There are no doubly concentrating critical elements in the sense of case* (III) *of Proposition 3.13.*

To prove this proposition, we establish the following dichotomy: either additional regularity for the critical element can be established using essentially the same arguments used in Section 4, or a self-similar-like critical element can be extracted by passing to a suitable limit. To this end we define function $\tau : \mathbb{R} \to \mathbb{R}$ by

$$\tau(t) = \int_0^t N(s)\, \mathrm{d}s.$$

Since $N(t) > 0$ and $\lim_{t\to\pm\infty} \tau(t) = \pm\infty$, the function $\tau : [0, \infty) \to [0, \infty)$ is bijective. Hence for any $t_0 > 0$ and any $C_+ > 0$, there exists a unique $\kappa_+ = \kappa_+(t_0, C_+) > 0$ such that

$$t_0 + \frac{\kappa_+(t_0, C_+)}{N(t_0)} = \tau^{-1}(\tau(t_0) + C_+).$$

Similarly, for $t_0 < 0$ and any $C_- > 0$, we can define

$$t_0 - \frac{\kappa_-(t_0, C_-)}{N(t_0)} = \tau^{-1}(\tau(t_0) - C_-).$$

Fix $\eta > 0$ as in the small-data theory of Proposition 2.6, and let $R = R(\eta)$ be such that, for all $t \in \mathbb{R}$,

$$\int_{|x-x(t)|\geq \frac{R(\eta)}{N(t)}} ||\nabla|^{s_p} u(t,x)|^2 \, dx + \int_{|x-x(t)|\geq \frac{R(\eta)}{N(t)}} ||\nabla|^{s_p-1} u_t(t,x)|^2 \, dx \leq \eta; \tag{6-2}$$

see Remark 3.6. Now let $\chi(t) = \chi_{R,N}(t)$ be a smooth cutoff to the set

$$\left\{ |x-x(t)| \geq \frac{R(\eta)}{N(t)} \right\}.$$

By our choice of $R(\eta)$ we have

$$\|\chi(t)\vec{u}\|^2_{\dot{\mathcal{H}}^{s_p}} \lesssim \eta.$$

Since $N(t) \geq 1$ and by (6-1), for any t_0, there exists $C_+(t_0) \geq 1$ sufficiently large so that

$$\left| x\left(t_0 + \frac{\kappa_+(t_0, C_+(t_0))}{N(t_0)}\right) - x(t_0) \right| \leq \left| \frac{\kappa_+(t_0, C_+(t_0))}{N(t_0)} \right| - \frac{R(\eta)}{N(t_0 + \kappa_+(t_0, C_+(t_0))N(t_0)^{-1})},$$

and similarly for $C_-(t_0)$. By continuity we may assume that $C_\pm(t_0)$ are minimal with this property. Furthermore, for every t_0 there exists $C(t_0)$ such that, for some $t_1 \in \mathbb{R}$ satisfying

$$\tau(t_1) - \tau(t_0) \leq C(t_0),$$

there exist $t_- < t_1 < t_+$ with

$$\tau(t_1) - \tau(t_-) \leq 2C(t_0), \quad \tau(t_+) - \tau(t_1) \leq 2C(t_0),$$

which satisfies

$$|x(t_-) - x(t_1)| \leq |t_- - t_1| - \frac{R(\eta)}{N(t_-)} \quad \text{and} \quad |x(t_+) - x(t_1)| \leq |t_+ - t_1| - \frac{R(\eta)}{N(t_+)}.$$

We note that we define $C(t_0)$ instead of working directly with $C_\pm(t_0)$ so as to split the τ integral evenly forward and backward in time. Moreover, if one tries to work directly with t_0 instead of t_1, one runs into issues with Case 2 below.

It is clear from the definition that $C(t_0) \leq \sup(C_+(t_0), C_-(t_0))$, and thus is finite. However, $C_\pm(t_0)$ need not be uniformly bounded for $t_0 \in \mathbb{R}$, and hence neither does $C(t_0)$. We will now analyze several cases based on whether $C(t_0)$ are uniformly bounded for $t_0 \in \mathbb{R}$.

6A. ***Case 1*: *$C(t_0)$ are uniformly bounded.*** Here we work under the assumption that there exists a constant $C > 0$ such that $C(t_0) \leq C$ for all $t_0 \in \mathbb{R}$.

We show that essentially the same argument used in Section 4A can be used to show that such a critical element necessarily has the compactness property in $\dot{\mathcal{H}}^{s_p} \cap \dot{\mathcal{H}}^1$.

Proposition 6.2 (additional regularity). *Let $\vec{u}(t) \in \dot{\mathcal{H}}^{s_p}$ be a solution with the compactness property that is subluminal and doubly concentrating, as in case (III) of Proposition 3.13. Assume in addition that $C(t)$ is uniformly bounded as a function of $t \in \mathbb{R}$. Then $\vec{u}(t) \in \dot{\mathcal{H}}^1$ and satisfies the bound*

$$\|\vec{u}(t)\|_{\dot{\mathcal{H}}^1} \lesssim N(t)^{\frac{5-p}{2(p-1)}} \tag{6-3}$$

uniformly in $t \in \mathbb{R}$.

For the moment, we will assume Proposition 6.2, and we will use it to prove the following corollary.

Corollary 6.3. *Let $\vec{u}(t)$ satisfy the hypotheses of Proposition 6.2. Then $\vec{u}(t) \equiv 0$.*

Proof of Corollary 6.3 assuming Proposition 6.2. We begin by extracting from $\vec{u}(t)$ another solution with the compactness property on a half-infinite time interval $[0,\infty)$ but with new scaling parameter $\widetilde{N}(s) \to 0$ as $s \to \infty$. Let $\tilde{t}_m$ be any sequence of times with

$$\tilde{t}_m \to -\infty, \quad N(\tilde{t}_m) \to \infty \quad \text{as } m \to \infty.$$

Next choose another sequence $t_m \to -\infty$ by choosing t_m such that

$$N(t_m) := \max_{t \in [\tilde{t}_m, 0]} N(t).$$

Now define a sequence as follows: set

$$w_m(s,y) := \frac{1}{N(t_m)^{\frac{2}{p-1}}} u\Big(t_m + \frac{s}{N(t_m)}, x(t_m) + \frac{y}{N(t_m)}\Big),$$
$$\partial_t w_m(s,y) := \frac{1}{N(t_m)^{\frac{2}{p-1}+1}} \partial_t u\Big(t_m + \frac{s}{N(t_m)}, x(t_m) + \frac{y}{N(t_m)}\Big),$$

and set

$$\vec{w}_m := (w_m(0,y), \partial_t w_m(0,y)).$$

Then by the precompactness in $\dot{\mathcal{H}}^{s_p}$, there exists (after passing to a subsequence) $\vec{w}_\infty(y) \neq 0$ so that

$$\vec{w}_m \to \vec{w}_\infty \in \dot{\mathcal{H}}^{s_p}.$$

It is standard to show that $\vec{w}(s)$ (the evolution of $\vec{w}_\infty = \vec{w}(0)$) has the compactness property on $I = [0,\infty)$ with frequency parameter $\widetilde{N}(s)$ defined by

$$\widetilde{N}(s) := \lim_{m\to\infty} \frac{N\big(t_m + \frac{s}{N(t_m)}\big)}{N(t_m)},$$

and moreover that

$$\widetilde{N}(s) \le 1 \quad \text{for all } s \in [0,\infty),$$
$$\liminf_{s \to \pm\infty} \widetilde{N}(s) = 0.$$

By the uniform bounds of (6-3), we see that

$$\|\vec{w}(s)\|_{\dot{\mathcal{H}}^1} \lesssim \widetilde{N}(s)^{\frac{5-p}{2(p-1)}} \quad \text{for all } s \in [0,\infty),$$

and hence there exists a sequence of times $s_n \to \infty$ along which

$$\|\vec{w}(s_n)\|_{\dot{\mathcal{H}}^1} \lesssim \widetilde{N}(s_n)^{\frac{5-p}{2(p-1)}} \to 0 \quad \text{as } n \to \infty.$$

Using the above, Sobolev embedding, and interpolation, along the same sequence of times we have

$$\|w(s_n)\|_{L^{p+1}} \lesssim \|w(s_n)\|_{\dot{H}^{3(p-1)/(2(p+1))}} \to 0 \quad \text{as } n \to \infty.$$

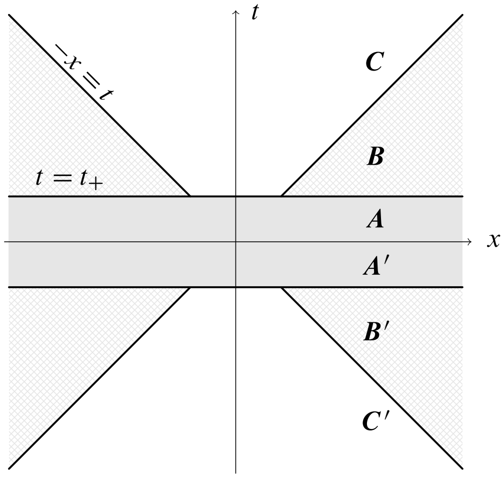

Figure 2. A depiction of the regions $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$.

But then, since the energy of $\vec{w}(s)$ is well-defined and conserved, we must have

$$E(\vec{w})=0.$$

For the defocusing equation we may immediately conclude that $\vec{w}(s)\equiv 0$. □

Remark 6.4. As in Section 4, these arguments readily adapt to the focusing setting.

Sketch of the proof of Proposition 6.2. The argument is nearly identical to the proof of Proposition 4.2 in Section 4; hence rather than repeat the entire proof, we instead summarize how the uniform boundedness of the numbers $C(t_0)$ allow us to proceed as in Section 4A. The main idea is that the boundedness of these constants means that for each $t_0\in\mathbb{R}$ we only have to wait a uniformly bounded amount of time, where time is measured relative to the scale $N(t)$, for the forward and backwards light cones based at $(t_0,x(t_0))$ to capture the bulk of the solution. Consequently, we can apply the same techniques that were developed in Section 4A directly and implement a double Duhamel argument. In order to estimate the norm at a time $t=t_0$, we recall the definitions of $t_1,t_\pm$ above and decompose space-time into three regions:

(A) Region $\boldsymbol{A}$: $[t_-,t_+]\times\mathbb{R}^3$.

(B) Region $\boldsymbol{B}$: the forward (resp. backward) light-cones from

$$\{t_+\}\times\{x:|x-x(t_1)|\geq|t_+-t_1|\},$$

and

$$\{t_-\}\times\{x:|x-x(t_1)|\geq|t_--t_1|\}.$$

(C) Region $\boldsymbol{C}$: $\mathbb{R}\times\mathbb{R}^4\setminus(\text{region } \boldsymbol{A}\cup\text{region } \boldsymbol{B})$.

On region $\boldsymbol{A}$, we control the solution by dividing the time interval $[t_-,t_+]$ into finitely many sufficiently small time strips on which we can use Lemma 3.11.

The main difficulty is that we need to ensure that we can uniformly control the number of small strips we will need to accomplish this (this type of uniform control was guaranteed in the Section 4 because we had $N(t) \equiv 1$ there). Here, the boundedness of the constants C is used to achieve this uniformity.

From Lemma 3.11 we know that for each $\eta > 0$ there exists $\delta > 0$ such that for all $t \in \mathbb{R}$

$$\|u\|_{L^{2(p-1)}_{t,x}([t-\frac{\delta}{N(t)},t+\frac{\delta}{N(t)}]\times\mathbb{R}^3)} \le \eta \quad \text{for all } t \in \mathbb{R}.$$

Fix this $\delta > 0$. Examining the proof of the estimates used to control the solution on region $\boldsymbol{A}$ in Section 4, see (4-5), we need to show that there exists a uniformly (in t_0) bounded number $M > 0$ of times t_m, $-M \le m \le M$ with $t_- \le t_m \le t_+$, and such that the corresponding intervals $I_{-M}, \dots, I_M$ with

$$I_M := \left[t_m - \frac{\delta}{N(t_m)}, t_m + \frac{\delta}{N(t_m)}\right]$$

satisfy

$$[t_-, t_+] \subset \bigcup_{m=-M}^{M} I_m.$$

In this case we obtain

$$\|u\|^{2(p-1)}_{L^{2(p-1)}_{t,x}([t_-,t_+]\times\mathbb{R}^3)} \lesssim \sum_{i=1}^{M} \|u\|^{2(p-1)}_{L^{2(p-1)}_{t,x}(I_i\times\mathbb{R}^3)} \lesssim \int_{t_-}^{t_+} N(t)\,\mathrm{d}t,$$

and, since

$$\int_{t_-}^{t_+} N(t)\,\mathrm{d}t = \tau(t_+) - \tau(t_-) \le 4C \tag{6-4}$$

by construction, this would yield the desired upper bound.

Hence, we now turn to the argument that intervals on which we can control the $L^{2(p-1)}_{t,x}$ will exhaust the time interval $[t_-, t_+]$ after finitely many steps. Since $|N'(t)| \lesssim N(t)^2$ on an interval of length $\delta/N(t)$, for any $t_1, t_2 \in [t_-, t_+]$, which satisfy $|t_1 - t_2| \le \delta/N(t_1)$, we have

$$N(t_1) - \delta N(t_1) \lesssim N(t_2) \lesssim N(t_1) + \delta N(t_1).$$

Consequently, for any $t_1 \in [t_-, t_+]$ we must have

$$\int_{t_1-\delta/N(t_1)}^{t_1+\delta/N(t_1)} N(t)\,\mathrm{d}t \ge (2\delta - 2\delta^2),$$

which for any $0 < \delta < \frac{1}{2}$ yields

$$\int_{t_1-\delta/N(t_1)}^{t_1+\delta/N(t_1)} N(t)\,\mathrm{d}t \ge \delta. \tag{6-5}$$

By (6-4) and (6-5),

$$4C \ge \int_{t_-}^{t_+} N(t)\,\mathrm{d}t = \int_{t_-}^{t_-+\delta/N(t_-)} N(t)\,\mathrm{d}t + \int_{t_-+\delta/N(t_-)}^{t_+} N(t)\,\mathrm{d}t = \int_{t_-+\delta/N(t_-)}^{t_+-\delta/N(t_+)} N(t)\,\mathrm{d}t + \delta;$$

hence the positivity of $N(t)$ implies that by iterating this procedure, we will be able to cover the whole interval $[t_-, t_+]$ in at most $4C/\delta$ many intervals of length $\delta/N(t)$, where we can control the $L_{t,x}^{2(p-1)}$ norm of the critical element.

On region $\boldsymbol{B}$, we use (6-2) to apply the small-data theory at times $t_\pm$, which, together with finite speed of propagation, yields a uniform bound on the solution. Finally, on region $\boldsymbol{C}$, we may use the sharp Huygens principle exactly as in Section 4A.

All together, using arguments from Section 4, this will yield that

$$\|u(t_1)\|_{\dot{\mathcal{H}}^1} \lesssim N(t_1)^{\frac{5-p}{2(p-1)}}.$$

For more details, we refer the reader to [Dodson and Lawrie 2015a]. By continuation of regularity and (6-5), this implies

$$\|u(t_0)\|_{\dot{\mathcal{H}}^1} \lesssim N(t_1)^{\frac{5-p}{2(p-1)}},$$

where the implicit constant again depends on C. Finally, since $|N'(t)| \lesssim N(t)^2$,

$$N(t_0) \sim_C N(t_1),$$

which completes the proof. □

6B. *Case 2*: *C(t) is not uniformly bounded.* In this case we will show how to extract a self-similar-like critical element by taking an appropriate limit. The arguments from Section 5, specifically Proposition 5.3, then allow us to conclude that any such solution must be $\equiv 0$, which is a contradiction.

By assumption, there exist sequences $\{t_n\}$ such that

$$C(t_n) \geq 2n.$$

Now define

$$I_n = [t_n - \kappa_-(t_n, n)N(t_n)^{-1}, t_n + \kappa_+(t_n, n)N(t_n)^{-1}].$$

Borrowing language from [Tao et al. 2007], since $C(t_n) \geq 2n$, we show that all sufficiently late times $t \in I_n$ are future-focusing, that is,

$$\text{for all } \tau \in I_n \text{ such that } \tau > t, \quad |x(\tau) - x(t)| \geq |\tau - t| - \frac{R(\eta)}{N(\tau)},$$

or all sufficiently early times $t \in I_n$ are past-focusing, that is,

$$\text{for all } \tau \in I_n \text{ such that } \tau < t, \quad |x(t) - x(\tau)| \geq |t - \tau| - \frac{R(\eta)}{N(\tau)}.$$

Indeed, suppose that there exist $t_-^n, t_+^n \in I_n$ such that $\tau(t_+^n) - \tau(t_-^n) \geq C_n$ for some $C_n \nearrow \infty$ as $n \nearrow \infty$, t_-^n is future-focusing, t_+^n is past-focusing, and $t_-^n < t_+^n$. In that case,

$$N(t) \sim N(\tau) \quad \text{for all } t, \tau \in [t_-^n, t_+^n],$$

with constant independent of n. For n sufficiently large this violates subluminality.

Therefore, suppose without loss of generality that for n sufficiently large, all sufficiently late times, say all

$$t \in \Big[t_n + \kappa_+\Big(t_n, \frac{n}{2}\Big)N(t_n)^{-1}, t_n + \kappa_+(t_n, n)N(t_n)^{-1}\Big] = I_n',$$

are future-focusing. First, we note that if $t \in I_n$ is future-focusing, then for any $\tau \in I_n$, $\tau > t$,

$$N(\tau) \le \frac{R(\eta)}{c} \inf_{t<s<\tau} N(s). \tag{6-6}$$

Indeed, for any $\tau \in I,\ \tau > t$,

$$|x(\tau) - x(t)| \ge |\tau - t| - \frac{R(\eta)}{N(\tau)}.$$

Then if $N(t) \le cN(\tau)/R(\eta)$,

$$|x(t) - x(\tau)| \ge |t - \tau| - \frac{c}{N(t)},$$

and therefore, we conclude that

$$N(\tau) \le \frac{1}{c^2} N(t) \le \frac{N(\tau)}{cR(\eta)},$$

which is a contradiction for $R(\eta)$sufficiently large. Note that in the case of past-focusing times, a similar argument yields a lower bound in place of (6-6).

Consequently, for any $t \in I_n'$,

$$N(t) \le \inf_{\tau<t:\tau\in I_n'} N(\tau).$$

In particular, modifying by a constant, $N(t)$ may be replaced by $\widetilde{N}(t)$ on I_n', where

$$\widetilde{N}(t) := \widetilde{N}_n(t) = \inf_{t_n+\kappa_+(t_n,\frac{n}{2})N(t_n)^{-1}<\tau<t} N(\tau).$$

Clearly, $\widetilde{N}(t)$ is monotone decreasing. Furthermore, extracting appropriate limits, we may assume that $\widetilde{N}(t)$ must converge to t^{-1} as $n \to \infty$. The main idea is that forward in time, on longer and longer time intervals, the precompact solution expands to fill the light cone. This observation will enable us to extract a solution which "looks self-similar" on $[1, \infty)$ and we can then rescale that solution to extract a true self-similar solution on $[0, \infty)$. We proceed with this argument now.

We begin by simplifying our notation, setting

$$t_-^n = t_n + \kappa_+\Big(t_n, \frac{n}{2}\Big)N(t_n)^{-1},$$
$$t_+^n = t_n + \kappa_+(t_n, n)N(t_n)^{-1}.$$

By definition of subluminality (see Definition 3.12), it holds that uniformly for all $t \in I_n'$,

$$\widetilde{N}(t)(t - t_-^n) \lesssim 1,$$

independent of n. We further have that

$$\widetilde{N}(t)(t - t_-^n) \gtrsim 1$$

is also uniformly bounded for all $t \in I_n'$ such that $t - t_-^n \ge \delta/N(t_-^n)$ by finite propagation speed.

Now set

$$K_n := \Big[\kappa_+(t_n, n) - \kappa_+\Big(t_n, \frac{n}{2}\Big)\Big] N(t_n)^{-1} \cdot N(t_-^n).$$

Since

$$\int_{t_-^n}^{t_+^n} \widetilde{N}(t)\,\mathrm{d}t \sim \frac{n}{2} \to \infty, \tag{6-7}$$

and $\widetilde{N}(t) \le \widetilde{N}(t_-^n)$ for all $t \in I_n'$, we see that if $K_n \le C$ for all $n \in \mathbb{N}$, then

$$\int_{t_-^n}^{t_+^n} \widetilde{N}(t) \lesssim 1,$$

which contradicts (6-7). Hence we may conclude that K_n is unbounded. We can then define a rescaled sequence as follows: set

$$u_n(0,x) = \frac{1}{\widetilde{N}(t_-^n)^{\frac{2}{p-1}}} u\Big(t_-^n, x(t_-^n) + \frac{x}{\widetilde{N}(t_-^n)}\Big),$$
$$\partial_t u_n(0,x) = \frac{1}{\widetilde{N}(t_-^n)^{\frac{2}{p-1}+1}} u\Big(t_-^n, x(t_-^n) + \frac{x}{\widetilde{N}(t_-^n)}\Big)$$

and let

$$\vec{w}_n(1) = (u_n(0,x), \partial_t u_n(0,x)).$$

By precompactness of the trajectory of $\vec{u}$ in $\dot{\mathcal{H}}^{s_p}$ (modulo symmetries), the rescaled initial data converges; that is, $\vec{w}_n(1) \to w_\infty$ in $\dot{\mathcal{H}}^{s_p}$. We let $\vec{w}(s)$ be the evolution of $\vec{w}_\infty =: \vec{w}(1)$; then $\vec{w}_\infty$ has the compactness property with a new scaling parameter $\widehat{N}(s)$, given by

$$\widehat{N}(s) = \lim_{n\to\infty} \frac{\widetilde{N}\Big(t_-^n + \frac{s}{\widetilde{N}(t_-^n)}\Big)}{\widetilde{N}(t_-^n)}.$$

Hence we have

$$cs \le \frac{1}{\widehat{N}(s)} \le s \quad \text{for all } s > 1.$$

We may also assume without loss of generality that $\vec{w}_\infty$ has the compactness property with translation parameter $\tilde{x}(s) = 0$: by finite speed of propagation, $\tilde{x}(s)$ must remain bounded, and hence we may, up to passing to a subsequence, obtain a precompact solution with $\tilde{x}(s) = 0$ by applying a fixed translation. Finally, we consider one last sequence of times $\{s_n\}$ with $s_n \to \infty$ and we define

$$w_n(1,x) = \frac{1}{(s_n)^{\frac{2}{p-1}}} w\Big(s_n, \frac{x}{s_n}\Big), \quad \partial_t u_n(1,x) = \frac{1}{(s_n)^{\frac{2}{p-1}+1}} w\Big(s_n, \frac{x}{s_n}\Big).$$

We set

$$\vec{v}_n(1) = (w_n(1,x), \partial_t w_n(1,x)),$$

which gives rise to a corresponding solution $\vec{v}_n(\tilde{s})$ with $\widehat{N}(\tilde{s}) = \tilde{s}^{-1}$ on $[1/s_n, \infty)$. We then can take the limit $n \to \infty$, which yields convergence $\vec{v}_n \to \vec{v}_\infty$ in $\dot{\mathcal{H}}^{s_p}$, and a solution $\vec{v}$ with initial data $\vec{v}_\infty$ which is self-similar on $[0, \infty)$.

7. The traveling-wave critical element

In this section we preclude the possibility of the existence of a "traveling-wave" critical element.

Recall the definition of a traveling-wave critical element.

Definition 7.1 (traveling wave). We say $\vec{u}(t) \neq 0$ is a *traveling-wave critical element* if $\vec{u}(t)$ is a global-in-time solution to (1-1) such that the set

$$K := \{(u(t, x(t) + \cdot), \partial_t u(t, x(t) + \cdot)) : t \in \mathbb{R}\}$$

is precompact in $\dot{H}^{s_p} \times \dot{H}^{s_p - 1}(\mathbb{R}^3)$, where the function $x : \mathbb{R} \to \mathbb{R}^3$ satisfies

$$x(0) = 0,$$

$$|t| - C_1 \le |x(t)| \le |t| + C_1, \tag{7-1}$$

$$|x(t) - (t, 0, 0)| \le C_1 |t|^{\frac{1}{2}} \tag{7-2}$$

for some uniform constant $C_1 > 0$.

The main result of this section is the following theorem.

Proposition 7.2. *There are no traveling-wave critical elements in the sense of case* (IV) *of Proposition 3.13.*

To prove Proposition 7.2, we will show that any traveling-wave critical element would enjoy additional regularity in the x_2- and x_3-directions. This will allow us to utilize a direction-specific Morawetz-type estimate to reach a contradiction. We will require an additional technical ingredient, namely, a long-time Strichartz estimate in the spirit of [Dodson 2012; 2016].

7.1. *Main ingredients in the proof.* The long-time Strichartz estimates take the following form:

Suppose $\vec{u}(t)$ is a traveling-wave critical element for (1-1). Let $\epsilon > 0$ and $0 < \theta < \frac{2}{3}\epsilon$. For any $\eta_0 > 0$, there exists $N_0 = N_0(\eta_0)$ large enough such that for all $N \ge N_0$ and for all $t_0 \in \mathbb{R}$, we have:

Proposition 7.3 (long-time Strichartz estimate). *Suppose $\vec{u}(t)$ is a traveling-wave critical element for (1-1). Let $\epsilon \in (0, 1)$ be arbitrary. Then,*

$$\|u_{>N}\|_{S([t_0, t_0 + N^{1-\epsilon}])} = o_N(1) \quad \text{as } N \to \infty,$$

where $S(I)$ denotes any admissible, non-endpoint Strichartz norm at Sobolev regularity $s = s_p$ on the time interval I.

With the help of Proposition 7.3, we will also prove the following additional regularity result.

Proposition 7.4 (additional regularity). *Suppose $\vec{u}(t)$ is a traveling-wave critical element for (1-1). For any $0 < \nu < \frac{1}{2}$,*

$$\||\partial_2|^{1-\nu} u\|_{L_t^\infty L_x^2(\mathbb{R} \times \mathbb{R}^3)} + \||\partial_3|^{1-\nu} u\|_{L_t^\infty L_x^2(\mathbb{R} \times \mathbb{R}^3)} < \infty.$$

Using Propositions 7.3 and 7.4, we can then prove the following Morawetz-type estimate. In the sequel, we use the notation

$$x = (x_1, x_{2,3}).$$

Proposition 7.5 (Morawetz-type estimate). *Suppose $\vec{u}(t)$ is a traveling-wave critical element for (1-1). Then there exists $\delta > 0$ and $\epsilon > 0$ such that*

$$\lim_{T\to\infty} \frac{1}{T^{1-\epsilon}} \int_0^{T^{1-\epsilon}} \int_{|x_{2,3}|\le T^\delta} |u_{\le T}(t,x)|^{p+1}\,\mathrm{d}x\,\mathrm{d}t = 0.$$

Combining Proposition 7.5 with the nontriviality of critical elements will yield a contradiction and complete the proof of Proposition 7.2.

We turn to the proofs of the three preceding propositions. In Section 7B we also give the proof of Proposition 7.2.

7A. *Long-time Strichartz estimates.* In this subsection we prove the long-time Strichartz estimate, Proposition 7.3, and then deduce a few technical corollaries.

Proof of Proposition 7.3. For technical reasons we fix a small parameter $0 < \theta \ll 1$ and introduce the following norm: given a time interval I,

$$\begin{aligned}\|u\|_{S_\theta(I)} = \|u\|_{L^{2(p-1)}_{t,x}} &+ \||\nabla|^{-\frac{2-3\theta}{2(p-1)}} u\|_{L_t^{p-1} L_x^{2(p-1)/\theta}} \\ &+ \||\nabla|^{-\frac{1-\theta}{p-1}} u\|_{L_t^{2(p-1)/(2-\theta)} L_x^{2(p-1)/\theta}} + \||\nabla|^{s_p-\theta} u\|_{L_t^{2/\theta} L_x^{2/(1-\theta)}} \\ &+ \||\nabla|^{\frac{2s_p}{3}-\frac{1}{3}} u\|_{L_t^{6/(1+s_p)} L_x^{6/(2-s_p)}} + \||\nabla|^{\frac{3}{4}-\frac{3}{2(p-1)}} u\|_{L_t^{2(p-1)} L_x^4},\end{aligned} \tag{7-3}$$

where all space-time norms are over $I \times \mathbb{R}^3$. Restrictions will be put on θ below. One can check that each of these norms correspond to wave-admissible exponent pairs at $\dot{H}^{s_p}$ regularity; this already requires $0 < \theta < p-3$. We will prove Proposition 7.3 for the space S_θ and note here that the same estimates then easily follow for the whole family of admissible Strichartz norms. We also note that a nearly identical (but simpler) argument works in the case $p = 3$, with the caveat that we need to perturb away from the inadmissible $(2,\infty)$ endpoint.

Let $\eta_0 > 0$ and $\epsilon > 0$. We will actually prove that there exists $N_0 \gg 1$ such that for $N \ge N_0$, we have

$$\|u_{>N}\|_{S_\theta([t_0, t_0 + (\frac{N}{N_0})^{1-\epsilon}])} < \eta_0$$

for any $t_0 \in \mathbb{R}$ and $\theta < \frac{2}{3}\epsilon$. This implies the estimate appearing in the statement of Proposition 7.3 upon enlarging ϵ and N_0; indeed,

$$N^{1-\epsilon'} \le \left(\frac{N}{N_0}\right)^{1-\epsilon}$$

provided $N \ge N_0^{(1-\epsilon)/(\epsilon'-\epsilon)}$.

By compactness and $N(t) \equiv 1$, there exists N_0 sufficiently large such that

$$\|u_{>N_0}\|_{S_\theta([t_0, t_0 + 9^{1-\epsilon}])} < \tfrac{1}{2}\eta_0$$

for any $t_0 \in \mathbb{R}$. This implies the desired estimate for $N_0 \le N \le 9N_0$. We will prove the result for larger N by induction.

Note that by choosing N_0 possibly even larger, we can guarantee

$$\|P_{>N}\vec{u}\|_{L_t^\infty \mathcal{H}^{s_p}(\mathbb{R}\times\mathbb{R}^3)} < \tfrac{1}{2}\eta_0 \tag{7-4}$$

for any $N \ge N_0$.

Before completing the inductive step, we make a few simplifications. First, by time-translation invariance, it suffices to consider $t_0 = 0$. Next, to keep formulas within the margins, we will assume all space-time norms are over $\big[0, \big(\frac{N}{N_0}\big)^{1-\epsilon}\big] \times \mathbb{R}^3$ unless otherwise stated.

By Taylor's theorem, we can write

$$\begin{aligned} F(u) &= F(u_{\le N}) + u_{>N}\int_0^1 F'(u_{<N} + \theta u_{>N}) \\ &= F(u_{\le N}) + u_{>N}F'(u_{<N}) + u_{>N}^2 \iint_0^1 F''(u_{<N} + \theta_1\theta_2 u_{>N}) \\ &= F(u_{\le N}) + u_{>N}F'(u_{<N}) + u_{>N}^2 F''(u_{<N}) + u_{>N}^3 \iiint_0^1 F'''(u_{<N} + \theta_1\theta_2\theta_3 u_{>N}) \end{aligned}$$

for any N. Thus (ignoring absolute values and constants) we need to estimate four types of terms

$$u_{>\frac{N}{8}} u_{\le\frac{N}{8}}^{p-1} + u_{>\frac{N}{8}}^2 u_{<\frac{N}{8}}^{p-2} + u_{\le\frac{N}{8}}^p + u_{>\frac{N}{8}}^3 F_2 =: I + II + III + IV,$$

where

$$F_2 = \iiint_0^1 F'''(u_{<\frac{N}{8}} + \theta_1\theta_2\theta_3 u_{>\frac{N}{8}}).$$

We will estimate the contribution of each term using Strichartz estimates.

Term I. We let $0 \le \theta < p-3$ as in (7-3) and further impose $\theta < \frac{2}{3}\epsilon$. We estimate

$$\begin{aligned} \||\nabla|^{s_p-1}P_{>N}(u_{\le\frac{N}{8}}^{p-1} u_{\ge\frac{N}{8}})\|_{L_t^1 L_x^2} &\lesssim N^{s_p-1}\|u_{\le\frac{N}{8}}\|_{L_t^{p-1}L_x^{2(p-1)/\theta}}^{p-1} \|u_{>\frac{N}{8}}\|_{L_t^\infty L_x^{2/(1-\theta)}} \\ &\lesssim N^{-1+\frac{3\theta}{2}}\|u_{\le\frac{N}{8}}\|_{L_t^{p-1}L_x^{2(p-1)/\theta}}^{p-1} \||\nabla|^{s_p-\frac{3\theta}{2}} u_{>\frac{N}{8}}\|_{L_t^\infty L_x^{2/(1-\theta)}} \\ &\lesssim \big[N^{-\frac{2-3\theta}{2(p-1)}}\|u_{\le\frac{N}{8}}\|_{L_t^{p-1}L_x^{2(p-1)/\theta}}\big]^{p-1} \||\nabla|^{s_p} u_{>\frac{N}{8}}\|_{L_t^\infty L_x^2}. \end{aligned}$$

Recalling (7-4), it remains to prove

$$N^{-\frac{2-3\theta}{2(p-1)}}\|u_{\le\frac{N}{8}}\|_{L_t^{p-1}L_x^{2(p-1)/\theta}} \lesssim \eta_0.$$

We let $C_0 \gg 1$, to be determined shortly, and begin by splitting

$$\begin{aligned} N^{-\frac{2-3\theta}{2(p-1)}}\|u_{\le\frac{N}{8}}\|_{L_t^{p-1}L_x^{2(p-1)/\theta}} &\lesssim N^{-\frac{2-3\theta}{2(p-1)}}\|u_{\le C_0}\|_{L_t^{p-1}L_x^{2(p-1)/\theta}} \\ &\quad + N^{-\frac{2-3\theta}{2(p-1)}}\|u_{C_0\le\,\cdot\,\le N_0}\|_{L_t^{p-1}L_x^{2(p-1)/\theta}} \\ &\quad + N^{-\frac{2-3\theta}{2(p-1)}}\sum_{N_0\le M\le\frac{N}{8}}\|u_M\|_{L_t^{p-1}L_x^{2(p-1)/\theta}}. \end{aligned}$$

By Bernstein's inequality and $N(t) \equiv 1$, we can estimate

$$N^{-\frac{2-3\theta}{2(p-1)}} \|u_{\leq C_0}\|_{L_t^{p-1} L_x^{2(p-1)/\theta}} \lesssim N^{-\frac{2-3\theta}{2(p-1)}} C_0^{\frac{2-3\theta}{2(p-1)}} \left(\frac{N}{N_0}\right)^{\frac{(1-\epsilon)}{p-1}}$$

on $\left[0, \left(\frac{N}{N_0}\right)^{1-\epsilon}\right] \times \mathbb{R}^3$. To guarantee that the overall power of N is negative, we need

$$\frac{3\theta}{2} < \epsilon.$$

Thus, for N_0 sufficiently large depending on C_0, we may guarantee that

$$N^{-\frac{2-3\theta}{2(p-1)}} \|u_{\leq C_0}\|_{L_t^{p-1} L_x^{2(p-1)/\theta}} \lesssim \eta_0.$$

Next, choosing $C_0 = C_0(\eta_0)$ large enough and using $N(t) \equiv 1$, we estimate

$$\begin{aligned} N^{-\frac{2-3\theta}{2(p-1)}} \|u_{C_0 \leq \cdot \leq N_0}\|_{L_t^{p-1} L_x^{2(p-1)/\theta}} &\lesssim N^{-\frac{2-3\theta}{2(p-1)}} N_0^{\frac{2-3\theta}{2(p-1)}} \||\nabla|^{-\frac{2-3\theta}{2(p-1)}} u_{>C_0}\|_{L_t^{p-1} L_x^{2(p-1)/\theta}} \\ &\lesssim \eta_0 \left(\frac{N}{N_0}\right)^{-\frac{2-3\theta}{2(p-1)} + \frac{(1-\epsilon)}{p-1}} \lesssim \eta_0. \end{aligned}$$

For the final term, we begin by estimating

$$N^{-\frac{2-3\theta}{2(p-1)}} \sum_{N_0 \leq M \leq \frac{N}{8}} \|u_M\|_{L_t^{p-1} L_x^{2(p-1)/\theta}} \lesssim \sum_{N_0 \leq M \leq \frac{N}{8}} \left(\frac{M}{N}\right)^{\frac{2-3\theta}{2(p-1)}} \||\nabla|^{-\frac{2-3\theta}{2(p-1)}} u_M\|_{L_t^{p-1} L_x^{2(p-1)/\theta}}.$$

We now apply the inductive hypothesis to the last term. To do so, we divide the interval $\left[0, \left(\frac{N}{N_0}\right)^{1-\epsilon}\right]$ into $\approx \left(\frac{N}{M}\right)^{1-\epsilon}$ intervals of length $\left(\frac{M}{N_0}\right)^{1-\epsilon}$. Continuing from above, this leads to

$$N^{-\frac{2-3\theta}{2(p-1)}} \sum_{N_0 \leq M \leq \frac{N}{8}} \|u_M\|_{L_t^{p-1} L_x^{2(p-1)/\theta}} \lesssim \sum_{N_0 \leq M \leq \frac{N}{8}} \left(\frac{M}{N}\right)^{\frac{2-3\theta}{2(p-1)} - \frac{(1-\epsilon)}{p-1}} \eta_0 \lesssim \eta_0,$$

where we have used that the exponent appearing is, in this case, positive. This completes the estimation of term I.

Term II. We estimate

$$\begin{aligned} \||\nabla|^{s_p - 1} P_{>N}(u_{\leq \frac{N}{8}}^{p-2} u_{\geq \frac{N}{8}}^2)\|_{L_t^1 L_x^2} &\lesssim N^{s_p - 1} \|u_{>\frac{N}{8}}\|_{L_t^{2(p-1)} L_x^4}^2 \|u_{\leq \frac{N}{8}}\|_{L_t^{p-1} L_x^\infty}^{p-2} \\ &\lesssim N^{s_p - 1 + 1 - \frac{1}{p-1}} \|u_{>\frac{N}{8}}\|_{L_t^{2(p-1)} L_x^4}^2 N^{-\frac{p-2}{(p-1)}} \|u_{\leq \frac{N}{8}}\|_{L_t^{p-1} L_x^\infty}^{p-2} \\ &\lesssim \left[N^{\frac{3}{4} - \frac{3}{2(p-1)}} \|u_{>\frac{N}{8}}\|_{L_t^{2(p-1)} L_x^4}\right]^2 N^{-\frac{p-2}{(p-1)}} \|u_{\leq \frac{N}{8}}\|_{L_t^{p-1} L_x^\infty}^{p-2}. \end{aligned}$$

We can argue as above (now with $\theta = 0$) for the low-frequency term, and we note that $(2(p-1), 4)$ is a wave-admissible pair at regularity

$$\frac{3}{2} - \frac{1}{2(p-1)} - \frac{3}{4} = s_p - \left(\frac{3}{4} - \frac{3}{2(p-1)}\right),$$

and we conclude using the inductive hypothesis on

$$\| |\nabla|^{\frac{3}{4}-\frac{3}{2(p-1)}} u_{>\frac{N}{8}} \|_{L_t^{2(p-1)} L_x^4}.$$

Term III. Next using the fractional chain rule we estimate

$$\begin{aligned}\| |\nabla|^{s_p-1} P_{>N}(u^p_{\leq \frac{N}{8}})\|_{L_t^1 L_x^2} &\lesssim N^{s_p-2} \|u_{\leq \frac{N}{8}}\|^{p-1}_{L_t^{2(p-1)/(2-\theta)} L_x^{2(p-1)/\theta}} \| |\nabla| u_{\leq \frac{N}{8}} \|_{L_t^{2/\theta} L_x^{2/(1-\theta)}} \\ &\lesssim N^{-1+\theta} \|u_{\leq \frac{N}{8}}\|^{p-1}_{L_t^{2(p-1)/(2-\theta)} L_x^{2(p-1)/\theta}} N^{-\theta+s_p-1} \| |\nabla| u_{\leq \frac{N}{8}} \|_{L_t^{2/\theta} L_x^{2/(1-\theta)}}.\end{aligned}$$

To complete the estimation of term III, we need to prove

$$N^{-\frac{1-\theta}{p-1}} \|u_{\leq \frac{N}{8}}\|_{L_t^{2(p-1)/(2-\theta)} L_x^{2(p-1)/\theta}} + N^{-\theta+s_p-1} \| |\nabla| u_{\leq N} \|_{L_t^{2/\theta} L_x^{2/(1-\theta)}} \lesssim \eta_0.$$

For this, we argue as in term I; that is, we split $u_{\leq \frac{N}{8}}$ into

$$u_{\leq \frac{N}{8}} = u_{\leq C_0} + u_{C_0 \leq \cdot \leq N_0} + \sum_{N_0 \leq M \leq \frac{N}{8}} u_M.$$

and estimate each term separately, relying on the inductive hypothesis (and a splitting of the time interval) for the final sum. Comparing with those estimates, we see that this requires

$$\frac{1-\theta}{p-1} - \frac{1-\epsilon}{p-1} > 0$$

to deal with the first term and

$$\theta + 1 - s_p - \frac{\theta(1-\epsilon)}{2} > 0$$

to deal with the second term. These conditions are satisfied provided $0 < \theta < \epsilon$.

Term IV. We estimate

$$\|u^3_{>\frac{N}{8}} F_2\|_{L_t^{2/(1+s_p)} L_x^{2/(2-s_p)}} \lesssim \|u^p_{>\frac{N}{8}}\|_{L_t^{2/(1+s_p)} L_x^{2/(2-s_p)}} + \|u^3_{>\frac{N}{8}} u^{p-3}_{\leq \frac{N}{8}}\|_{L_t^{2/(1+s_p)} L_x^{2/(2-s_p)}}.$$

For the first expression we estimate

$$\|u^p_{>\frac{N}{8}}\|_{L_t^{2/(1+s_p)} L_x^{2/(2-s_p)}} = \|u_{>\frac{N}{8}}\|^p_{L_t^{2p/(1+s_p)} L_x^{2p/(2-s_p)}} \lesssim \eta_0^p,$$

while for the second expression we have

$$\|u^3_{>\frac{N}{8}} u^{p-3}_{\leq \frac{N}{8}}\|_{L_t^{2/(1+s_p)} L_x^{2/(2-s_p)}} \lesssim \|u_{>\frac{N}{8}}\|^3_{L_t^{6/(1+s_p)} L_x^{6/(2-s_p)}} \|u_{\leq \frac{N}{8}}\|^{p-3}_{L^\infty_{t,x}}.$$

Now,

$$\|u_{\leq \frac{N}{8}}\|^{p-3}_{L^\infty_{t,x}} \lesssim N^{\frac{2(p-3)}{p-1}} \|u_{<\frac{N}{8}}\|^{p-3}_{L_t^\infty L^{3(p-1)/2}} \lesssim N^{\frac{2(p-3)}{p-1}}.$$

For the first term, we see that $\left(\frac{6}{1+s_p}, \frac{6}{2-s_p}\right)$ is an admissible Strichartz pair at regularity

$$s_p + \frac{1}{3} - \frac{2s_p}{3} < s_p,$$

and hence

$$\|u^3_{>\frac{N}{8}} u^{p-3}_{\leq \frac{N}{8}}\|_{L_t^{2/(1+s_p)} L_x^{2/(2-s_p)}} \lesssim N^{1-2s_p} \| |\nabla|^{\frac{2s_p}{3}-\frac{1}{3}} u_{>\frac{N}{8}}\|^3_{L_t^{6/(1+s_p)} L_x^{6/(2-s_p)}} N^{\frac{2(p-3)}{p-1}} \|u_{<\frac{N}{8}}\|^{p-3}_{L_t^\infty L^{3(p-1)/2}}.$$

Finally, note that

$$-2s_p+1+\frac{2(p-3)}{p-1}=-3+1+\frac{4}{p-1}+\frac{2p-6}{p-1}=0.$$

Hence, by the inductive hypothesis, putting all the pieces of the argument together, we obtain

$$\|u_{>N}\|_{S_\theta([0,(\frac{N}{N_0})^{1-\epsilon}])}\le \tfrac{1}{2}\eta_0+C\eta_0^3,$$

which suffices to complete the induction for η_0 sufficiently small. □

We will need the following corollary of Proposition 7.3, which provides some control over the low frequencies as well.

Corollary 7.6 (control of low frequencies). *Suppose $\vec{u}$ is a traveling-wave critical element for (1-1). Let $\epsilon>0$ and $0<\theta<\frac{2}{3}\epsilon$. For any η_0 there exists N sufficiently large such that*

$$\|u_{\le N}\|_{L_t^{p-1}L_x^{2(p-1)/\theta}([t_0,t_0+N^{1-\epsilon}])\times\mathbb{R}^3)}\lesssim \eta_0 N^{\frac{2-3\theta}{2(p-1)}}, \tag{7-5}$$

$$\|u_{\le N}\|_{L_t^{2(p-1)/(2-\theta)}L_x^{2(p-1)/\theta}([t_0,t_0+N^{1-\epsilon}]\times\mathbb{R}^3)}\lesssim \eta_0 N^{\frac{1-\theta}{p-1}},$$

$$\||\nabla|u_{\le N}\|_{L_t^{2/\theta}L_x^{2/(1-\theta)}([t_0,t_0+N^{1-\epsilon}]\times\mathbb{R}^3)}\lesssim \eta_0 N^{\theta-s_p+1}$$

uniformly over $t_0\in\mathbb{R}$.

Proof. We let η_0 and choose $N_0=N_0(\eta_0)\ge 1$ as in Proposition 7.3. By time-translation invariance, it suffices to consider $t_0=0$. We focus our attention on (7-5), as the other estimates follow similarly. For $N\ge N_0$, we estimate

$$\begin{aligned}\|u_{\le N}\|_{L_t^{p-1}L_x^{2(p-1)/\theta}([0,N^{1-\epsilon}]\times\mathbb{R}^3)}\lesssim{}& \|u_{\le N_0}\|_{L_t^{p-1}L_x^{2(p-1)/\theta}([0,N^{1-\epsilon}]\times\mathbb{R}^3)}\\ &+\sum_{N_0\le M\le N}\|u_M\|_{L_t^{p-1}L_x^{2(p-1)/\theta}([0,M^{1-\epsilon}]\times\mathbb{R}^3)}\\ &+\sum_{N_0\le M\le N}\|u_{\le N}\|_{L_t^{p-1}L_x^{2(p-1)/\theta}([M^{1-\epsilon},N^{1-\epsilon}]\times\mathbb{R}^3)}.\end{aligned}$$

For the first term, we use Bernstein's inequality and $N(t)\equiv 1$ to get

$$\|u_{\le N_0}\|_{L_t^{p-1}L_x^{2(p-1)/\theta}([0,N^{1-\epsilon}]\times\mathbb{R}^3)}\lesssim N_0^{\frac{2-3\theta}{2(p-1)}}N^{\frac{1-\epsilon}{p-1}}.$$

Recalling that

$$\frac{1-\epsilon}{p-1}<\frac{2-3\theta}{2(p-1)},$$

we see that this term is acceptable provided we choose N sufficiently large.

Next, we use Proposition 7.3 to estimate

$$\sum_{N_0\le M\le N}\|u_M\|_{L_t^{p-1}L_x^{2(p-1)/\theta}([0,M^{1-\epsilon}]\times\mathbb{R}^3)}\lesssim \eta_0\sum_{N_0\le M\le N}M^{\frac{2-3\theta}{2(p-1)}}\lesssim \eta_0 N^{\frac{2-3\theta}{2(p-1)}},$$

which is also acceptable.

For the remaining term, we split $[M^{1-\epsilon}, N^{1-\epsilon}]$ into $\approx \left(\frac{N}{M}\right)^{1-\epsilon}$ intervals of length $M^{1-\epsilon}$. Applying Proposition 7.3 once more, we have

$$\sum_{N_0 \le M \le N} \|u_M\|_{L_t^{p-1} L_x^{2(p-1)/\theta}([M^{1-\epsilon}, N^{1-\epsilon}] \times \mathbb{R}^3)} \lesssim \eta_0 \sum_{N_0 \le M \le N} M^{\frac{2-3\theta}{2(p-1)}} \left(\frac{N}{M}\right)^{\frac{1-\epsilon}{p-1}} \lesssim \eta_0 N^{\frac{2-3\theta}{2(p-1)}},$$

where we recall $0 < \theta < \frac{2}{3}\epsilon$ in order to sum. This term is also acceptable, and so we complete the proof of (7-5) and Corollary 7.6. ☐

Finally, we will need certain long-time Strichartz estimates with regularity in the x_2- and x_3-directions.

Corollary 7.7 (long-time Strichartz estimates for $\nabla_{x_2,x_3} u$). *Suppose that Proposition 7.4 holds with $\nu > 0$. Then, for any $\nu_0 > \nu$,*

$$\| |\nabla_{x_{2,3}}|^{1-\nu_0} u_N \|_{L_t^{2/(1-s_p)} L_x^{2/s_p}([t_0, t_0+N^{1-\epsilon}])} \lesssim N^{1-s_p}.$$

Proof. We only sketch this argument as it follows in the same manner as the standard long-time Strichartz estimate with some additional technical details. First we note that $\left(\frac{2}{1-s_p}, \frac{2}{s_p}\right)$ is an admissible Strichartz pair at regularity $1 - s_p$. By compactness, it suffices to argue with $t_0 = 0$. Let $S(I)$ denote any collection of Strichartz pairs at regularity $s = 0$. We will show that

$$\| |\nabla_{x_{2,3}}|^{1-\nu_0} u_N \|_{S([t_0, t_0+N^{1-\epsilon}])} \lesssim 1 \tag{7-6}$$

for $\nu_0 > \nu$, from which the result follows.

Let $\vec{u}$ be a solution with the compactness property on $\mathbb{R}$ with $N(t) = 1$. By the Gagliardo–Nirenberg inequality

$$\| |\nabla_{2,3}|^{1-\nu_0} u \|_{L^2_{x_2,x_3}} \le C \| |\nabla_{2,3}|^{1-\nu} u \|^{\alpha}_{L^2_{x_2,x_3}} \|u\|^{1-\alpha}_{L^{2/(1-s_p)}_{x_2,x_3}} \tag{7-7}$$

for

$$\alpha = \frac{1 - s_p - \nu_0}{1 - s_p - \nu}.$$

Next, we observe the Sobolev embedding

$$\dot{H}_x^{s_p} \hookrightarrow L^2_{x_1} L_{x_{2,3}}^{\frac{2}{1-s_p}}, \tag{7-8}$$

which follows from Sobolev embedding in $\mathbb{R}^2$ and Plancherel:

$$\int \left(\int |u(x_1, x_{2,3})|^{\frac{2}{1-s_p}} \, dx_{2,3} \right)^{1-s_p} \mathrm{d}x_1 \lesssim \int \left| |\nabla_{x_{2,3}}|^{s_p} u \right|^2 \mathrm{d}x \sim \int \left| |\xi_{2,3}|^{s_p} \hat{u}(\xi) \right|^2 \mathrm{d}\xi \lesssim \int \left| |\xi|^{s_p} \hat{u}(\xi) \right|^2 \mathrm{d}\xi.$$

Thus we may take the $L^2_{x_1}$ norm of both sides of (7-7) and use Hölder's inequality on the right to conclude that the trajectory $|\nabla_{2,3}|^{1-\nu_0} u$ has the compactness property in L^2_x, and hence there exists $N_0 = N_0(\eta_0)$ such that, for all $N > N_0$,

$$\| P_{>N} |\nabla_{2,3}|^{1-\nu_0} u \|_{S([0, 9^{1-\epsilon}])} \le \eta_0 \quad \text{for all } N \ge N_0(\eta_0),$$

which proves the base case, that is, (7-6) holds for $N_0 \le N \le 9N_0$.

We now proceed to the inductive step. Suppose that (7-6) holds up to frequency N_1 for $N_1 \geq 9N_0$. We will show that (7-6) holds for $N = 2N_1$. The argument we employ is similar to a persistence of regularity argument. Note that $|\nabla_{2,3}|^{1-\nu_0}u$ solves the equation

$$\partial_t |\nabla_{2,3}|^{1-\nu_0}u - \Delta|\nabla_{2,3}|^{1-\nu_0}u = |\nabla_{2,3}|^{1-\nu_0}F(u).$$

By the Strichartz estimates we have

$$\begin{aligned}
&\|P_{>N}|\nabla_{2,3}|^{1-\nu_0}u\|_{S([0,(N/N_0)^{1-\epsilon}])} \\
&\qquad \lesssim \|P_{>N}|\nabla_{2,3}|^{1-\nu_0}\vec{u}\|_{L_t^\infty \dot{\mathcal{H}}_x^0([0,(N/N_0)^{1-\epsilon}])} + \|P_{>N}|\nabla_{2,3}|^{1-\nu_0}F(u)\|_{\mathcal{N}([0,(N/N_0)^{1-\epsilon}])},
\end{aligned}$$

where $\mathcal{N}$ is the dual space to S. Let $\widehat{P}_M$ denote a Fourier projection in the ξ_2, ξ_3 variables. The first term can be bounded using compactness, so we focus on the second term. We again write

$$\begin{aligned}
F(u) &= F(u_{\leq N}) + u_{>N}\int_0^1 F'(u_{<N} + \theta u_{>N}) \\
&= F(u_{\leq N}) + u_{>N}F'(u_{<N}) + u_{>N}^2 \iint_0^1 F''(u_{<N} + \theta_1\theta_2 u_{>N}) \\
&= F(u_{\leq N}) + u_{>N}F'(u_{<N}) + u_{>N}^2 F''(u_{<N}) + u_{>N}^3 \iiint_0^1 F'''(u_{<N} + \theta_1\theta_2\theta_3 u_{>N}).
\end{aligned}$$

We will estimate the first term as an example, since the other terms will be similar generalizations of the proof of Proposition 7.3. We have

$$\begin{aligned}
&\||\nabla|^{-1}|\nabla_{2,3}|^{1-\nu_0}P_{>N}F(u_{\leq \frac{N}{8}})\|_{L_t^1 L_x^2} \\
&\quad \lesssim N^{-1}\||\nabla_{2,3}|^{1-\nu_0}P_{>N}F(u_{\leq \frac{N}{8}})\|_{L_t^1 L_x^2} \\
&\quad \lesssim N^{-2}\||\nabla_{2,3}|^{1-\nu_0}u_{\leq \frac{N}{8}}\|^{p-1}_{L_t^{2(p-1)/(2-\theta)}L_x^{2(p-1)/\theta}}\||\nabla|u_{\leq \frac{N}{8}}\|_{L_t^{2/\theta}L_x^{2/(1-\theta)}} \\
&\qquad\qquad + N^{-2}\|u_{\leq \frac{N}{8}}\|^{p-1}_{L_t^{2(p-1)/(2-\theta)}L_x^{2(p-1)/\theta}}\||\nabla||\nabla_{2,3}|^{1-\nu_0}u_{\leq \frac{N}{8}}\|_{L_t^{2/\theta}L_x^{2/(1-\theta)}} \\
&\quad \lesssim N^{-1+\theta-s_p}\||\nabla_{2,3}|^{1-\nu_0}u_{\leq \frac{N}{8}}\|^{p-1}_{L_t^{2(p-1)/(2-\theta)}L_x^{2(p-1)/\theta}} N^{-1-\theta+s_p}\||\nabla|u_{\leq \frac{N}{8}}\|_{L_t^{2/\theta}L_x^{2/(1-\theta)}} \\
&\qquad\qquad + N^{-1+\theta}\|u_{\leq \frac{N}{8}}\|^{p-1}_{L_t^{2(p-1)/(2-\theta)}L_x^{2(p-1)/\theta}} N^{-1-\theta}\||\nabla||\nabla_{2,3}|^{1-\nu_0}u_{\leq \frac{N}{8}}\|_{L_t^{2/\theta}L_x^{2/(1-\theta)}},
\end{aligned}$$

and all four terms can be treated analogously to the low-frequency component in term I in Proposition 7.3. □

7B. *Proof of Propositions 7.5 and 7.2, assuming Proposition 7.4.* As mentioned above, the long-time Strichartz estimate (Proposition 7.3) will be a key ingredient to proving additional regularity (Proposition 7.4). Before turning to the rather technical proof, let us use Proposition 7.4 (together with Proposition 7.3 and Corollary 7.6) to prove the Morawetz estimate, Proposition 7.5. With the Morawetz estimate in hand, we can then quickly rule out the possibility of traveling waves and hence complete the proof of the main result, Proposition 7.2.

We recall the notation $x = (x_1, x_{2,3})$ and similarly write $\xi = (\xi_1, \xi_{2,3})$.

Proof of Proposition 7.5. Let $\psi : [0,\infty) \to \mathbb{R}$ be a smooth cutoff satisfying

$$\begin{cases} \psi(\rho) = 1, & \rho \le 1, \\ \psi(\rho) = 0, & r > 2. \end{cases}$$

We fix $R > 0$ to be determined below and let $\psi_R(\rho) = \psi\left(\frac{\rho}{R}\right)$. Next, let

$$\chi_R(r) = \frac{1}{r}\int_0^r \psi_R(s)\,\mathrm{d}s.$$

We collect a few useful identities,

$$\partial_k[x^k \chi_R] = \chi_R + \psi_R, \quad r\partial_r \chi_R = -\chi_R + \psi_R, \tag{7-9}$$

and we recall the Sobolev embedding (7-8).

In the following, we consider χ_R as a function of $|x_{2,3}|$. For $T > 0$ and

$$I := P_{\le T},$$

we define the Morawetz quantity

$$M(t) = \int_{\mathbb{R}^3} \chi_R I u_t\, x^k \partial_k I u \,\mathrm{d}x + \frac{1}{2}\int_{\mathbb{R}^3} (\chi_R + \psi_R) I u_t I u \,\mathrm{d}x,$$

where repeated indices are summed over $k \in \{2,3\}$.

We first compute the derivative of $M(t)$:

$$M'(t) = \int \chi_R I u_t (x^k \partial_k I u_t) + \frac{1}{2}\int (\chi_R + \psi_R)(I u_t)^2 + \int \chi_R [x^k \partial_k I u] I u_{tt} + \frac{1}{2}\int (\chi_R + \psi_R) I u I u_{tt}.$$

By (7-9) and integration by parts, we have

$$\int [x^k \chi_R]\partial_k \frac{1}{2}(I u_t)^2 = -\frac{1}{2}\int (\chi_R + \psi_R)(I u_t)^2,$$

so we are left to estimate

$$\int \chi_R [x^k \partial_k I u] I u_{tt} + \frac{1}{2}\int (\chi_R + \psi_R) I u I u_{tt}. \tag{7-10}$$

Using the equation for u yields

$$I u_{tt} = \Delta I u - F(I u) + [F(I u) - I F(u)], \quad \text{where } F(z) = |z|^{p-1} z.$$

We first consider the contribution of $\Delta I u$ to (7-10). We claim

$$\int x^k \chi_R [\partial_k I u] \Delta I u + \frac{1}{2}(\chi_R + \psi_R) I u \Delta I u \,\mathrm{d}x \le \frac{1}{2}\int \Delta(\chi_R + \psi_R)(I u)^2 \,\mathrm{d}x. \tag{7-11}$$

In the proof of (7-11) we will simplify notation by suppressing the operator I, suppressing the dependence on R, and writing $u_k = \partial_k u$. We turn to the proof.

We begin by considering the first term on the left-hand side of (7-11). Integrating by parts yields

$$\int x^k \chi u_k u_{jj} = -\int \partial_j [x^k \chi] u_k u_j + \frac{1}{2} x^k \chi \partial_k (u_j^2),$$

where $k \in \{2,3\}$ and $j \in \{1,2,3\}$. Writing $r = |x_{2,3}|$ and using (7-9), we have

$$\begin{aligned}\int \partial_j[x^k\chi]u_k u_j &= \int \delta_{jk}\chi u_j u_k + \frac{x^k x^j}{r^2} r\chi' u_j u_k \\ &= \int \delta_{jk}\chi u_j u_k + \frac{x^k x^j}{r^2}\psi u_j u_k - \frac{x^k x^j}{r^2}\chi u_j u_k,\end{aligned}$$

where we may now restrict to $j \in \{2,3\}$. Using the other identity in (7-9), we also have

$$\frac{1}{2}\int x^k \chi \partial_k (u_j)^2 = -\frac{1}{2}\int (\chi+\psi)u_j^2.$$

As for the second term on the left-hand side of (7-11), we have

$$\frac{1}{2}\int (\chi+\psi)uu_{jj} = \frac{1}{2}\int \partial_{jj}(\chi+\psi)u^2 - \frac{1}{2}(\chi+\psi)u_j^2.$$

Collecting the computations above, we find

$$\begin{aligned}&\int \big(x^k\chi_R[\partial_k Iu]\Delta Iu + \tfrac{1}{2}(\chi_R+\psi_R)Iu\Delta Iu\big)\,\mathrm{d}x \\ &\qquad = \frac{1}{2}\int \Delta(\chi+\psi)u^2 - \int\Big[\delta_{jk} - \frac{x^j x^k}{r^2}\Big]\chi u_j u_k - \int \psi\Big[\frac{x_{2,3}}{r}\cdot\nabla_{x_{2,3}}u\Big]^2\,\mathrm{d}x,\end{aligned}$$

which yields (7-11).

We next consider the contribution of $-F(Iu)$ to (7-10). Using (7-9) and integration by parts,

$$\begin{aligned}-\int\Big[x^k\chi_R\partial_k Iu + \tfrac{1}{2}(\chi_R+\psi_R)Iu\Big]F(Iu)\,\mathrm{d}x &= -\int (x^k\chi_R)\frac{1}{p+1}\partial_k|Iu|^{p+1} + \tfrac{1}{2}(\chi_R+\psi_R)|Iu|^{p+1}\,\mathrm{d}x \\ &= \int\Big(\frac{1}{p+1} - \frac{1}{2}\Big)(\chi_R+\psi_R)|Iu|^{p+1}\,\mathrm{d}x. \qquad (7\text{-}12)\end{aligned}$$

Hence, by (7-10), (7-11), and (7-12) and the fundamental theorem of calculus, we deduce

$$\begin{aligned}\iint_{|x_{2,3}|\le R} |Iu|^{p+1}\,\mathrm{d}x\,\mathrm{d}t &\lesssim \sup_{t\in J}|M(t)| + \iint \Delta(\chi_R+\psi_R)|Iu|^2\,\mathrm{d}x\,\mathrm{d}t \\ &\quad + \Big|\iint [x^k\chi_R\partial_k Iu + (\chi_R+\psi_R)Iu][F(Iu) - IF(u)]\,\mathrm{d}x\,\mathrm{d}t\Big|\end{aligned}$$

for any interval J. In the following, we choose $J = [0, T^{1-\epsilon}]$, where $\epsilon > 0$ will be chosen below and T is large enough that Proposition 7.3 and Corollary 7.6 hold.

We need to estimate the terms on the right-hand side of this inequality. We first bound $|M(t)|$. By Bernstein's inequality, the Sobolev embedding (7-8), and Proposition 7.4, we have

$$\begin{aligned}\sup_t |M(t)| &\lesssim \|Iu_t\|_{L_t^\infty L_x^2}\big(R\|\nabla_{x_{2,3}}Iu\|_{L_t^\infty L_x^2} + \|Iu\|_{L_t^\infty L_{x_1}^2 L_{x_{2,3}}^{2/(1-s_p)}}\|\chi_R+\psi_R\|_{L_{x_{2,3}}^{2/(1+s_p)}}\big) \\ &\lesssim T^{1-s_p}(RT^\nu + R^{1+s_p}\|u\|_{L_t^\infty \dot H_x^{s_p}}) \\ &\lesssim T^{1-s_p}(RT^\nu + R^{1+s_p}), \qquad (7\text{-}13)\end{aligned}$$

for $\nu > 0$ to be chosen sufficiently small below.

For the next term, we have

$$\begin{aligned}\iint \Delta(\chi_R+\psi_R)|Iu|^2\,\mathrm{d}x\,\mathrm{d}t &\lesssim T^{1-\epsilon}\|Iu\|^2_{L_t^\infty L^2_{x_1}L^{2/(1-s_p)}_{x_{2,3}}}\|\Delta(\chi_R+\psi_R)\|_{L^{2/(1+s_p)}_{x_{2,3}}}\\ &\lesssim T^{1-\epsilon}\|u\|^2_{L_t^\infty \dot{H}_x^{s_p}}R^{-(2-2s_p)}\\ &\lesssim T^{1-\epsilon}R^{-(2-2s_p)}. \end{aligned}\tag{7-14}$$

Now we turn to the final term. Arguing as in the long-time Strichartz estimates, we need to estimate terms of the form

$$u^p_{\le T}+u_{>T}u^{p-1}_{\le T}+u^2_{>T}u^{p-2}_{\le T}+u^3_{>T}F_2,$$

where F_2 involves both high and low frequencies. Thus we estimate

$$\iint [x^k\chi_R\partial_k Iu+(\chi_R+\psi_R)Iu][F(Iu)-IF(u)]\,\mathrm{d}x\,\mathrm{d}t$$

$$\lesssim \iint x^k\chi_R\partial_k u_{\le T}P_{>T}[u^p_{\le T}]\,\mathrm{d}x\,\mathrm{d}t \tag{7-15}$$

$$+\iint x^k\chi_R\partial_k u_{\le T}P_{>T}[|u_{>T}||u_{\le T}|^{p-1}]\,\mathrm{d}x\mathrm{d}t \tag{7-16}$$

$$+\iint x^k\chi_R\partial_k u_{\le T}P_{>T}[|u_{>T}|^2|u_{\le T}|^{p-2}]\,\mathrm{d}x\mathrm{d}t \tag{7-17}$$

$$+\iint x^k\chi_R\partial_k u_{\le T}P_{>T}[|u_{>T}|^3F_2]\,\mathrm{d}x\mathrm{d}t \tag{7-18}$$

$$+\iint (\chi_R+\psi_R)u_{\le T}P_{>T}[u^p_{\le T}]\,\mathrm{d}x\,\mathrm{d}t \tag{7-19}$$

$$+\iint (\chi_R+\psi_R)u_{\le T}P_{>T}[|u_{>T}||u_{\le T}|^{p-1]}\,\mathrm{d}x\,\mathrm{d}t \tag{7-20}$$

$$+\iint (\chi_R+\psi_R)u_{\le T}P_{>T}[|u_{>T}|^2|u_{\le T}|^{p-2}]\,\mathrm{d}x\,\mathrm{d}t \tag{7-21}$$

$$+\iint (\chi_R+\psi_R)u_{\le T}P_{>T}[|u_{>T}|^3F_2]\,\mathrm{d}x\,\mathrm{d}t, \tag{7-22}$$

where all the integrals are taken over $[0,T^{1-\epsilon}]\times\mathbb{R}^3$. We treat each of these terms separately.

We first consider (7-15). Estimating as in the long-time Strichartz estimates and using Corollary 7.6, we obtain

$$\begin{aligned}\left|\iint x^k\chi_R\partial_k u_{\le T}P_{>T}[u^p_{\le T}]\,\mathrm{d}x\,\mathrm{d}t\right| &\lesssim RT^{1-s_p}\|\nabla_{x_{2,3}}u_{\le T}\|_{L_t^\infty L_x^2}T^{s_p-2}\||\nabla|P_{>T}(u_{\le T})^p\|_{L_t^1L_x^2}\\ &\lesssim RT^{1-s_p+\nu}T^{s_p-2}\|u_{\le T}\|^{p-1}_{L_t^{2(p-1)/(2-\theta)}L_x^{2(p-1)/\theta}}\|\nabla u_{\le T}\|_{L_t^{2/\theta}L_x^{2/(1-\theta)}}\\ &\lesssim RT^{1-s_p+\nu}.\end{aligned}$$

We next consider (7-16). We let $0 < \theta < \frac{2}{3}\epsilon$. By Bernstein's inequality, Proposition 7.4, and Corollary 7.6, we obtain

$$\begin{aligned}\left|\iint x^k \chi_R \partial_k u_{\le T}[|u_{>T}|\,|u_{\le T}|^{p-1}]\,dx\,dt\right| &\lesssim R\|\nabla_{x_{2,3}} u_{\le T}\|_{L_t^\infty L_x^2}\|u_{\le T}\|^{p-1}_{L_t^{p-1}L_x^{2(p-1)/\theta}}\|u_{>T}\|_{L_t^\infty L_x^{2/(1-\theta)}}\\ &\lesssim RT^{1-s_p}\|\nabla_{x_{2,3}} u_{\le T}\|_{L_t^\infty L_x^2} T^{s_p-1}\|u_{\le T}\|^{p-1}_{L_t^{p-1}L_x^{2(p-1)/\theta}}\|u_{>T}\|_{L_t^\infty L_x^{2/(1-\theta)}}\\ &\lesssim RT^{1-s_p+\nu}.\end{aligned}$$

For (7-17) we again argue as in the proof of the long-time Strichartz estimates, and using Corollary 7.6, we obtain

$$\begin{aligned}\left|\iint x^k \chi_R \partial_k u_{\le T}[|u_{>T}|^2|u_{\le T}|^{p-2}]\,dx\,dt\right| &\lesssim R\|\nabla_{x_{2,3}} u_{\le T}\|_{L_t^\infty L_x^2}\|u_{\le T}\|^{p-2}_{L_t^{p-1}L_x^\infty}\|u_{>T}\|_{L_t^{2(p-1)}L_x^4}\\ &\lesssim RT^{\nu+s_p-1}\||\nabla_{x_{2,3}}|^{1-\nu} u_{\le T}\|_{L_t^\infty L_x^2} N^{1-s_p}\|u_{\le T}\|^{p-2}_{L_t^{p-1}L_x^\infty}\|u_{>T}\|_{L_t^{2(p-1)}L_x^4}\\ &\lesssim RT^{1-s_p+\nu}.\end{aligned}$$

For (7-18), we once again use the bounds from the proof of the long-time Strichartz estimates as well as Corollary 7.7, and we obtain

$$\begin{aligned}\left|\iint x^k \chi_R \partial_k u_{\le T} P_{>T}[|u_{>T}|^3 F_2]\,dx\,dt\right| &\lesssim R\|\nabla_{x_{2,3}} u_{\le T}\|_{L_t^{2/(1-s_p)}L_x^{2/s_p}}\|P_{>T}[|u_{>T}|^3F_2]\|_{L_t^{2/(1+s_p)}L_x^{2/(2-s_p)}}\\ &\lesssim R\sum_{N\le T} N^{\nu_0}\left(\frac{T}{N}\right)^{\frac{(1-\epsilon)(1-s_p)}{2}}\||\nabla_{x_{2,3}}|^{1-\nu_0} u_N\|_{L_t^{2/(1-s_p)}L_x^{2/s_p}}\\ &\lesssim RT^{\nu_0}\sum_{N\le T}\left(\frac{T}{N}\right)^{\frac{(1-\epsilon)(1-s_p)}{2}} N^{1-s_p}\\ &\lesssim RT^{1-s_p+\nu_0}\sum_{N\le T}\left(\frac{N}{T}\right)^{1-s_p-\frac{(1-\epsilon)(1-s_p)}{2}} \lesssim RT^{1-s_p+\nu_0}\end{aligned}$$

for any $\nu_0 > \nu$, where $\nu > 0$ is as in Proposition 7.4.

Arguing analogously for the remaining terms, the estimates are almost identical, up to noting that by Hölder's inequality in the x_2- and x_3-variables we have

$$\|(\chi_R+\psi_R)u_{\le T}\|_{L_t^\infty L_x^2} \lesssim R^{s_p}\|u_{\le T}\|_{L_t^\infty L_{x_1}^2 L_{x_2,x_3}^{2/(1-s_p)}},$$

which is controlled by the $\dot H^{s_p}$ norm by the Sobolev embedding (7-8). Thus we obtain for (7-19)–(7-21) the estimates

$$\begin{aligned}(7\text{-}19) &\lesssim R^{s_p}T^{1-s_p}\|u_{\le T}\|_{L_t^\infty L_{x_1}^2 L_{x_2,x_3}^{2/(1-s_p)}} T^{s_p-2}\||\nabla|P_{>T}[u_{\le T}^p]\|_{L_t^1L_x^2} \lesssim R^{s_p},\\ (7\text{-}20) &\lesssim R^{s_p}T^{1-s_p}\|u_{\le T}\|_{L_t^\infty L_{x_1}^2 L_{x_2,x_3}^{2/(1-s_p)}} T^{s_p-1}\|P_{>T}[u_{>T}u_{\le T}^{p-1}]\|_{L_t^1L_x^2} \lesssim R^{s_p},\\ (7\text{-}21) &\lesssim R^{s_p}T^{1-s_p}\|u_{\le T}\|_{L_t^\infty L_{x_1}^2 L_{x_2,x_3}^{2/(1-s_p)}} T^{s_p-1}\|P_{>T}[u_{>T}^2u_{\le T}^{p-2}]\|_{L_t^1L_x^2} \lesssim R^{s_p}.\end{aligned}$$

For the last term (7-22), we note that

$$\frac{2}{s_p} \le \frac{2}{1-s_p},$$

and we use Hölder's inequality in the $x_{2,3}$-variables to estimate

$$\begin{aligned}
(7\text{-}22) &\lesssim \|(\chi_R+\psi_R)u_{\le T}\|_{L_t^{2/(1-s_p)}L_x^{2/s_p}} \|P_{>T}[u_{>T}^3 F_2]\|_{L_t^{2/(1+s_p)}L_x^{2/(2-s_p)}} \\
&\lesssim T^{\frac{(1-\epsilon)(1-s_p)}{2}} T^{\frac{1-s_p}{2}} R^{2s_p-1} \|(\chi_R+\psi_R)u_{\le T}\|_{L_t^\infty L_{x_1}^2 L_{x_{2,3}}^{2/(1-s_p)}} \\
&\lesssim T^{1-s_p} R^{2s_p-1} \|(\chi_R+\psi_R)u_{\le T}\|_{L_t^\infty L_{x_1}^2 L_{x_{2,3}}^{2/(1-s_p)}}.
\end{aligned}$$

Now, using (7-13), (7-14), and our estimates for (7-15)–(7-22), we have established that

$$\iint_{|x_{2,3}|\le R} |u_{\le T}|^{p+1}\,dx\,dt \lesssim RT^{1+\nu-s_p} + R^{s_p} + R^{2s_p-1}T^{1-s_p}.$$

We now choose $R = T^{(1/2)+}$ to obtain that the right-hand side is $o(T^{1-\epsilon})$. This can be achieved provided $\nu+\epsilon < s_p - \frac{1}{2}$, and hence we complete the proof. □

As mentioned above, with the Morawetz estimate Proposition 7.5 in hand, we can quickly rule out traveling waves. The final ingredient we will need is the nontriviality for compact solutions appearing in Corollary 3.9. Combining this corollary with Proposition 7.5, we can now prove Proposition 7.2.

Proof of Proposition 7.2. Suppose toward a contradiction that $\vec{u}$ is a traveling-wave critical element for (1-1). It suffices to prove that

$$\int_0^{T^{1-\epsilon}} \int_{|x_{2,3}|\le T^{(1/2)+}} |u_{\le T}(t,x)|^{p+1}\,dx\,dt \gtrsim T^{1-\epsilon} \tag{7-23}$$

for T sufficiently large, as this contradicts Proposition 7.5. By Corollary 3.9, the definition of the critical element, and the fact that $N(t)\equiv 1$, there exists $C \gg 1$ and $T \gg 1$ large enough that

$$\int_{t_0}^{t_0+1} \int_{|x-x(t)|\le C} |u_{\le T}(t,x)|^{\frac{3(p-1)}{2}}\,dx\,dt \gtrsim_u 1 \tag{7-24}$$

for all $t_0 \in \mathbb{R}$. Recalling $|x(t)-(t,0,0)| \lesssim \sqrt{t}$ we see that for $T > C^2$ we have

$$\{|x-x(t)|\le C\} \subset \{|x_{2,3}| \le T^{\frac{1}{2}+}\}$$

for all $t \in [0, T^{1-\epsilon}]$. Thus (7-24) implies (7-23), as desired. □

7C. ***Additional regularity*: *proof of Proposition 7.4*.** Our final task is to prove Proposition 7.4, namely, additional regularity for traveling waves. More precisely, we can establish additional regularity in the directions orthogonal to the direction of travel.

Recall the notation $x = (x_1, x_{2,3})$. We similarly use $\xi = (\xi_1, \xi_{2,3})$ for the frequency variable. We also introduce the following modified Littlewood–Paley operators:

For $N, M \in 2^{\mathbb{Z}}$, we let $\widehat{P}_{N,>M}$ be the Fourier multiplier operator that is equal to 1 where

$$|\xi| \simeq N \quad \text{and} \quad |\xi_{2,3}| \gtrsim M.$$

We let $\widehat{P}_{N,M} = \widehat{P}_{N,>2M} - \widehat{P}_{N,>M}$, and we let $P_N = \widehat{P}_{N,\le M} + \widehat{P}_{N,>M}$.

We will occasionally abuse notation slightly and apply these multipliers to a vector, where this should be taken to mean applying these multipliers componentwise. We note that this notation differs from that of the previous sections; however, we would like to make explicit that N corresponds to ξ-frequencies, while M corresponds to those of $\xi_{2,3}$.

We fix $\nu > 0$. We begin with the observation that

$$\|P_{\leq N_0} u\|_{L_t^\infty \dot{H}_x^1} \lesssim_{N_0} 1. \tag{7-25}$$

We will choose the precise value of $N_0 \gg 1$ in the course of the proof. On the other hand, we have

$$\begin{aligned}\sum_{N>N_0} \||\nabla_{x_{2,3}}|^{1-\nu} \widehat{P}_{N,\leq N^{s_p/(1-\nu)}} u(t)\|_{L_x^2}^2 &\lesssim \sum_{N>N_0} N^{2[\frac{s_p}{1-\nu}(1-\nu)-s_p]} \||\nabla|^{s_p} u_N(t)\|_{L_x^2}^2 \\ &\lesssim \||\nabla|^{s_p} u(t)\|_{L_x^2}^2. \end{aligned}\tag{7-26}$$

Therefore, we are left to show that

$$\sum_{N\geq N_0} \sum_{C_0 N^{s_p/(1-\nu)} \leq M \leq N} M^{2(1-\nu)} \|\widehat{P}_{N,\geq M} u(t)\|_{L_x^2}^2 \lesssim 1$$

for some fixed $C_0 > 0$ (uniformly in t).

We will use a double Duhamel argument together with a frequency envelope to estimate this expression. We will estimate

$$\|\widehat{P}_{N,\geq M} u(t_0)\|_{L^2(\mathbb{R}^3)}^2 \simeq N^{-2s_p} \|\widehat{P}_{N,\geq M} u(t_0)\|_{\dot{H}^{s_p}(\mathbb{R}^3)}^2 = N^{-2s_p} \langle \widehat{P}_{N,\geq M} u(t_0),\ \widehat{P}_{N,\geq M} u(t_0)\rangle_{\dot{H}^{s_p}(\mathbb{R}^3)}.$$

We will show that there exists a frequency envelope $\gamma_{M,N}$ such that

$$\|\widehat{P}_{N,\geq M} u(t_0)\|_{\dot{H}^{s_p}(\mathbb{R}^3)} \lesssim \gamma_{N,M}(t_0)$$

and such that

$$\sum_{N\geq N_0} \left(\sum_{C_0 N^{s_p/(1-\nu)} \leq M \leq N} M^{2(1-\nu)} N^{-2s_p} \gamma_{N,M}(t_0)^2 \right) \lesssim 1.$$

Consequently, this will show that

$$\begin{aligned}\sum_{N\geq N_0} \sum_{C_0 N^{s_p/(1-\nu)} \leq M \leq N} & M^{2(1-\nu)} \|\widehat{P}_{N,\geq M} u(t_0)\|_{L_x^2}^2 \\ &\lesssim \sum_{N\geq N_0} \sum_{C_0 N^{s_p/(1-\nu)} \leq M \leq N} M^{2(1-\nu)} N^{-2s_p} \|\widehat{P}_{N,\geq M} u(t_0)\|_{\dot{H}_x^{s_p}}^2 \lesssim 1.\end{aligned}$$

Together with (7-25) and (7-26) (and time-translation invariance), this will imply

$$\||\partial_2|^{1-\nu} u\|_{L_t^\infty L_x^2} + \||\partial_3|^{1-\nu} u\|_{L_t^\infty L_x^2} < \infty,$$

and hence prove Proposition 7.4. Thus, we let

$$\Gamma_{N,M}(t_0) = N^{s_p} \|\widehat{P}_{N,\geq M} u(t_0)\|_{L^2(\mathbb{R}^3)},$$

and we fix some $\sigma > 0$ to be specified later. We define the frequency envelope

$$\gamma_{M,N}(t_0) = \sum_{N'} \sum_{M' \le M} \min\left\{ \frac{N}{N'}, \frac{N'}{N} \right\}^{\sigma} \cdot \left(\frac{M'}{M} \right)^{\sigma} \Gamma_{N',M'}(t_0).$$

By time-translation symmetry, it suffices to consider the case $t_0 = 0$. Once again, we complexify the solution, letting

$$w = u + \frac{i}{\sqrt{-\Delta}} u_t.$$

Then

$$\|w(t)\|_{\dot{H}^1} \simeq \|\vec{u}(t)\|_{\dot{H}^1 \times L^2},$$

and if $\vec{u}(t)$ solves (1-1), then $w(t)$ is a solution to

$$w_t = -i\sqrt{-\Delta} w \pm \frac{i}{\sqrt{-\Delta}} |u|^{p-1} u.$$

By Duhamel's principle, for any T, we have

$$w(0) = e^{iT\sqrt{-\Delta}} w(T) \pm \frac{i}{\sqrt{-\Delta}} \int_T^0 e^{i\tau\sqrt{-\Delta}} F(u)(\tau)\, \mathrm{d}\tau,$$

where $F(u) = |u|^{p-1}u$. To estimate $\gamma_{N,M}$, we write

$$\begin{aligned} \widehat{P}_{N,\ge M} w(0) &= \widehat{P}_{N,\ge M} e^{-iT\sqrt{-\Delta}} w(T) - \frac{1}{\sqrt{-\Delta}} \int_0^T e^{-it\sqrt{-\Delta}} \widehat{P}_{N,\ge M} F(u)\, \mathrm{d}t \\ &= \widehat{P}_{N,\ge M} e^{-iT\sqrt{-\Delta}} w(-T) - \frac{1}{\sqrt{-\Delta}} \int_{-T}^0 e^{-i\tau\sqrt{-\Delta}} \widehat{P}_{N,\ge M} F(u)\, \mathrm{d}\tau. \end{aligned}$$

When we pair these expressions and take $T \to \infty$, we use the facts that

$$e^{-iT\sqrt{-\Delta}} \widehat{P}_{N,\ge M} w(T) \rightharpoonup 0 \quad \text{and} \quad e^{iT\sqrt{-\Delta}} \widehat{P}_{N,\ge M} w(-T) \rightharpoonup 0,$$

and ultimately we are left to estimate

$$\left\langle \int_0^\infty S(-t) \widehat{P}_{N,\ge M} F(u)\, \mathrm{d}t, \int_{-\infty}^0 S(-\tau) \widehat{P}_{N,\ge M} F(u)\, \mathrm{d}\tau \right\rangle_{\dot{H}_x^{s_p}}.$$

where we have introduced the notation

$$S(t) := \frac{1}{\sqrt{-\Delta}} e^{it\sqrt{-\Delta}}$$

above.

As we have done in previous sections, we will estimate this expression by dividing space-time into three regions: a compact time interval, an outer region, and a region inside the light-cone. We note, however, that the arguments on the compact time interval and the region inside the light cone will be considerably different than in previous sections.

Thus, we let $\eta_0 > 0$ and $\epsilon > 0$ be sufficiently small parameters and define the smooth cut-off

$$\chi_0(t,x) = 1_{\{|x-x(N^{1-\epsilon})|\geq R(\eta_0)+(t-N^{1-\epsilon}),\, t\geq N^{1-\epsilon}\}},$$

where $R(\eta_0)$ is such that

$$\|\chi_0(N^{1-\epsilon},x)u(N^{1-\epsilon},x)\|_{\dot{H}^{s_p}} + \|\chi_0(N^{1-\epsilon},x)\partial_t u(N^{1-\epsilon},x)\|_{\dot{H}^{s_p-1}} \leq \eta_0.$$

By the small-data theory, we may solve the Cauchy problem

$$\begin{cases} v_{tt} - \Delta v + F(v) = 0 \quad \text{on } \mathbb{R}\times\mathbb{R}^3, \\ (v,\partial_t v)|_{t=0} = \big(\chi_0(N^{1-\epsilon},x)u(N^{1-\epsilon},x),\, \chi_0(N^{1-\epsilon},x)u_t(N^{1-\epsilon},x)\big) \in \dot{\mathcal{H}}^{s_p}(\mathbb{R}^3). \end{cases}$$

Note that by finite propagation speed, $v = u$ on the set

$$\{(t,x) : |x - x(N^{1-\epsilon})| \geq R(\eta_0) + (t - N^{1-\epsilon}),\ t \geq N^{1-\epsilon}\}.$$

We now write

$$\int_0^\infty S(-t)\widehat{P}_{N,\geq M}F(u)\,\mathrm{d}t = A + B + C,$$

where

$$\begin{aligned} A &= \int_{N^{1-\epsilon}}^\infty S(-t)\widehat{P}_{N,\geq M}F(v)\,\mathrm{d}t, \\ B &= \int_0^{N^{1-\epsilon}} S(-t)\widehat{P}_{N,\geq M}F(u)\,\mathrm{d}t, \\ C &= \int_{N^{1-\epsilon}}^\infty S(-t)\widehat{P}_{N,\geq M}[F(u)-F(v)]\,\mathrm{d}t \end{aligned} \tag{7-27}$$

and perform a similar decomposition in the negative time direction, yielding quantities A', B', C'. We will use the estimate

$$|\langle A+B+C, A'+B'+C'\rangle| \lesssim \|A\|^2_{\dot{H}^{s_p}_x} + \|A'\|^2_{\dot{H}^{s_p}_x} + \|B\|^2_{\dot{H}^{s_p}_x} + \|B'\|^2_{\dot{H}^{s_p}_x} + |\langle C, C'\rangle_{\dot{H}^{s_p}_x}| \tag{7-28}$$

whenever $A + B + C = A' + B' + C'$.

Term A. We first estimate $\langle A, A\rangle_{\dot{H}^{s_p}_x}$ and $\langle A', A'\rangle_{\dot{H}^{s_p}_x}$, where

$$A = \int_{N^{1-\epsilon}}^\infty S(-t)\widehat{P}_{N,\geq M}F(v)\,\mathrm{d}t \quad \text{and} \quad A' = \int_{-\infty}^{-N^{1-\epsilon}} S(-\tau)\widehat{P}_{N,\geq M}F(v)\,\mathrm{d}\tau.$$

We introduce two parameters q and r satisfying

$$2 < q < \min\left\{p-1, \frac{2}{s_p}, \frac{5p-9}{3p-7}\right\} \quad \text{and} \quad \frac{2}{s_p} \leq r \leq \min\left\{\frac{2p(p-1)}{2p-3}, 4+\right\},$$

and let $I = [N^{1-\epsilon}, \infty)$. We fix $\sigma > 0$ to be specified later, and we define

$$a_{N',M} = [(N')^{-(\frac{2}{q}-s_p)}\|\widehat{P}_{N',\geq M}v\|_{L^q_t L^{2q/(q-2)}_x} + (N')^{\frac{2}{r}-1+s_p}\|\widehat{P}_{N',\geq M}v\|_{L^{2r/(r-2)}_t L^r_x}],$$

and let

$$\alpha_{N,M} = \sum_{N'} \sum_{M' \le M} \left(\frac{M'}{M}\right)^{\sigma} \min\left\{\frac{N}{N'}, \frac{N'}{N}\right\}^{\sigma} a_{N',M'}.$$

All space-time norms are taken over $I \times \mathbb{R}^3$. Our goal is to prove the following result.

Lemma 7.8. *Let A, A' and $\gamma_{N,M}$ be as above, then*

$$\sum_{N'} \sum_{M' \le M} \left(\frac{M'}{M}\right)^{\sigma} \min\left\{\frac{N}{N'}, \frac{N'}{N}\right\}^{\sigma} (\|A\|_{\dot{H}^{s_p}_x} + \|A'\|_{\dot{H}^{s_p}_x}) \lesssim \eta_0^{p-1}\, \alpha_{N,M},$$

and we also have

$$\alpha_{N,M} \lesssim \gamma_{N,M}(N^{1-\epsilon}) + \eta_0^{p-1} \alpha_{N,M}. \tag{7-29}$$

Proof. On this region, we will use the small-data theory, which implies, in particular, that

$$\|v\|_{L^{2(p-1)}_{t,x}(\mathbb{R}\times\mathbb{R}^3)} \lesssim \eta_0.$$

By Strichartz estimates, we may write

$$\begin{aligned} N^{-(\frac{2}{q}-s_p)} \|\widehat{P}_{N,\ge M} v\|_{L^q_t L^{2q/(q-2)}_x} &+ N^{\frac{2}{r}-1+s_p} \|\widehat{P}_{N,\ge M} v\|_{L^{2r/(r-2)}_t L^r_x} \\ &\lesssim \|\widehat{P}_{N,\ge M}(v, v_t)(N^{1-\epsilon})\|_{\dot{\mathcal{H}}^{s_p}} + \|P_{N,\ge M} F(v)\|_{N(\mathbb{R})}, \end{aligned} \tag{7-30}$$

and recall that, by definition, we have

$$\|\widehat{P}_{N,\ge M}(v, v_t)(N^{1-\epsilon})\|_{\dot{H}^{s_p}} \lesssim \Gamma_{N,M}(N^{1-\epsilon}). \tag{7-31}$$

Here, we let the norm $\|F\|_{N(\mathbb{R})}$ denote any finite combination $\sum_j \|F_j\|_{N_j(\mathbb{R})}$, with $F = \sum_j F_j$ and each $N_j(\mathbb{R})$ being a dual admissible Strichartz space with the appropriate scaling and number of derivatives.

It will be useful to introduce the quantities

$$v_{\text{lo}} = \sum_{N'} \widehat{P}_{N',\le M} v \quad \text{and} \quad v_{\text{hi}} = \sum_{N'} \widehat{P}_{N',> M} v,$$

where "lo" and "hi" are meant to refer to the $\xi_{2,3}$-frequency component. We decompose the nonlinearity via

$$F(v) = F(v_{\text{lo}}) + v_{\text{hi}} \int_0^1 F'(v_{\text{lo}} + \theta v_{\text{hi}})\, \mathrm{d}\theta,$$

which we write schematically as

$$F(v) = F(v_{\text{lo}}) + v_{\text{hi}} v^{p-1}.$$

For the high-frequency (in M) component, we write

$$v_{\text{hi}} v^{p-1} = (P_{\le N} v_{\text{hi}}) v^{p-1} + (P_{\ge N} v_{\text{hi}}) v^{p-1},$$

and to estimate these terms we may use the dual Strichartz spaces

$$L_t^{\frac{2q}{q+2}} \dot{H}_x^{-(\frac{2}{q}-s_p), \frac{q}{q-1}} \quad \text{and} \quad L_t^{\frac{r}{r-1}} \dot{H}_x^{\frac{2}{r}-1+s_p, \frac{2r}{r+2}}$$

respectively. This yields

$$\begin{aligned}\|P_{N',\geq M}v_{\mathrm{hi}}v^{p-1}\|_{N(\mathbb{R})} &\lesssim (N')^{-(\frac{2}{q}-s_p)}\sum_{N''\leq N'}\|\widehat{P}_{N'',\geq M}v\|_{L_t^qL_x^{2q/(q-2)}(I\times\mathbb{R}^3)}\|v\|^{p-1}_{L^{2(p-1)}_{t,x}}\\ &\quad+(N')^{\frac{2}{r}-1+s_p}\sum_{N''\geq N'}\|\widehat{P}_{N'',\geq M}v\|_{L_t^{2r/(r-2)}L_x^r(I\times\mathbb{R}^3)}\|v\|^{p-1}_{L^{2(p-1)}_{t,x}}\\ &:=\mathcal{N}_1+\mathcal{N}_2;\end{aligned}$$

hence we conclude that

$$\sum_{N'}\min\Bigl\{\frac{N}{N'},\frac{N'}{N}\Bigr\}^{\sigma}\|P_{N',\geq M}v_{\mathrm{hi}}v^{p-1}\|_{N(\mathbb{R})}\lesssim\sum_{N'}\min\Bigl\{\frac{N}{N'},\frac{N'}{N}\Bigr\}^{\sigma}(\mathcal{N}_1+\mathcal{N}_2).$$

Thus, we argue in order to bound the $\mathcal{N}_1$ and $\mathcal{N}_2$ terms. We only treat the first term as an example since the other term follows analogously. We obtain

$$\begin{aligned}&\sum_{N'\leq N}\sum_{N''\leq N'}\Bigl(\frac{N'}{N}\Bigr)^{\sigma}(N')^{-(\frac{2}{q}-s_p)}\|\widehat{P}_{N'',\geq M}v\|_{L_t^qL_x^{2q/(q-2)}}\\ &\qquad\lesssim\sum_{N'\leq N}\sum_{N''\leq N'}\Bigl(\frac{N'}{N}\Bigr)^{\sigma}\Bigl(\frac{N''}{N'}\Bigr)^{\frac{2}{q}-s_p}(N'')^{-(\frac{2}{q}-s_p)}\|\widehat{P}_{N'',\geq M}v\|_{L_t^qL_x^{2q/(q-2)}}\\ &\qquad\lesssim\sum_{N''\leq N}\Bigl(\frac{N''}{N}\Bigr)^{\sigma}a_{N'',M}\sum_{N'\geq N''}\Bigl(\frac{N'}{N''}\Bigr)^{-(\frac{2}{q}-s_p)+\sigma}.\end{aligned}$$

Hence this term can be bounded by $\alpha_{N,M}$ provided $\sigma<\frac{2}{q}-s_p$.

We also have

$$\begin{aligned}&\sum_{N'\geq N}\Bigl(\frac{N}{N'}\Bigr)^{\sigma}(N')^{-(\frac{2}{q}-s_p)}\sum_{N''\leq N'}\|\widehat{P}_{N'',\geq M}v\|_{L_t^qL_x^{2q/(q-2)}(I\times\mathbb{R}^3)}\\ &\quad\lesssim\sum_{N''\leq N}\Bigl(\frac{N''}{N}\Bigr)^{\sigma}a_{N'',M}\sum_{N''\leq N',N\leq N'}\Bigl(\frac{N^2}{N'N''}\Bigr)^{\sigma}\Bigl(\frac{N'}{N''}\Bigr)^{-(\frac{2}{q}-s_p)}\\ &\qquad\qquad+\sum_{N''\geq N}\Bigl(\frac{N}{N''}\Bigr)^{\sigma}a_{N'',M}\sum_{N''\leq N'}\Bigl(\frac{N''}{N'}\Bigr)^{\sigma}\Bigl(\frac{N'}{N''}\Bigr)^{-(\frac{2}{q}-s_p)}.\end{aligned}$$

Using that in the first term

$$\Bigl(\frac{N^2}{N'N''}\Bigr)^{\sigma}\leq\Bigl(\frac{(N')^2}{N'N''}\Bigr)^{\sigma}=\Bigl(\frac{N'}{N''}\Bigr)^{\sigma},$$

we can bound this expression by $\gamma_{N,M}$ provided

$$\sum_{N''\leq N'}\Bigl(\frac{N'}{N''}\Bigr)^{\sigma}\Bigl(\frac{N'}{N''}\Bigr)^{-(\frac{2}{q}-s_p)}\lesssim 1\quad\Longleftrightarrow\quad\sigma<\frac{2}{q}-s_p,$$

and so we obtain

$$\sum_{N'}\sum_{M'\leq M}\Bigl(\frac{M'}{M}\Bigr)^{\sigma}\min\Bigl\{\frac{N}{N'},\frac{N'}{N}\Bigr\}^{\sigma}\|P_{N',\geq M'}v_{\mathrm{hi}}v^{p-1}\|_{N(\mathbb{R})}\lesssim\eta_0^{p-1}\alpha_{N,M}.\tag{7-32}$$

Next we estimate the contribution of the low-frequency piece. We can write

$$\|P_{N,\geq M}F(v_{\mathrm{lo}})\|_{N(\mathbb{R})} \leq M^{-2}\|P_{N,\geq M}\Delta_{x_2,x_3}F(v_{\mathrm{lo}})\|_{N(\mathbb{R})}.$$

Applying the chain rule and taking the decomposition $v = P_{\geq N}v + P_{\leq N}v$, we obtain for $j = 2, 3$ that

$$\begin{aligned}\partial_{x_j}F(v_{\mathrm{lo}}) &= \partial_{x_j}v_{\mathrm{lo}}F'(v_{\mathrm{lo}})\\ &= \partial_{x_j}P_{\leq N}v_{\mathrm{lo}}F'(v_{\mathrm{lo}}) + \partial_{x_j}P_{\geq N}v_{\mathrm{lo}}F'(v_{\mathrm{lo}}),\end{aligned}$$

and hence

$$\begin{aligned}\|P_{N,\geq M}\Delta_{x_2,x_3}F(v_{\mathrm{lo}})\|_{N(\mathbb{R})} &\leq \sum_{j=1}^{2}\|\partial_{x_j}P_{N,\geq M}(\partial_{x_j}v_{\mathrm{lo}})F'(v_{\mathrm{lo}})\|_{N(\mathbb{R})}\\ &\leq \sum_{j=1}^{2} M\|P_{N,\geq M}(\partial_{x_j}v_{\mathrm{lo}})F'(v_{\mathrm{lo}})\|_{N(\mathbb{R})}.\end{aligned}$$

Estimating as above, using the dual Strichartz spaces

$$L_t^{\frac{2q}{q+2}}\dot{H}_x^{-(\frac{2}{q}-s_p),\frac{q}{q-1}} \quad \text{and} \quad L_t^{\frac{r}{r-1}}\dot{H}_x^{\frac{2}{r}-1+s_p,\frac{2r}{r+2}},$$

we conclude that

$$\begin{aligned}\|P_{N,\geq M}F(v_{\mathrm{lo}})\|_{N(\mathbb{R})} &\lesssim N^{-(\frac{2}{q}-s_p)}\sum_{N'\leq N}\sum_{M'\leq M} M^{-1}\|\nabla_{x_2,x_3}\widehat{P}_{N',M'}v_{\mathrm{lo}}\|_{L_t^qL_x^{2q/(q-2)}(I\times\mathbb{R}^3)}\|v\|^{p-1}_{L_{t,x}^{2(p-1)}}\\ &\quad + N^{\frac{2}{r}-1+s_p}\sum_{N'\geq N}\sum_{M'\leq M} M^{-1}\|\nabla_{x_2,x_3}\widehat{P}_{N',M'}v_{\mathrm{lo}}\|_{L_t^{2r/(r-2)}L_x^r(I\times\mathbb{R}^3)}\|v\|^{p-1}_{L_{t,x}^{2(p-1)}}\\ &\lesssim N^{-(\frac{2}{q}-s_p)}\sum_{N'\leq N}\sum_{M'\leq M}\left(\frac{M'}{M}\right)\|\widehat{P}_{N',M'}v_{\mathrm{lo}}\|_{L_t^qL_x^{2q/(q-2)}(I\times\mathbb{R}^3)}\|v\|^{p-1}_{L_{t,x}^{2(p-1)}}\\ &\quad + N^{\frac{2}{r}-1+s_p}\sum_{N'\geq N}\sum_{M'\leq M}\left(\frac{M'}{M}\right)\|\widehat{P}_{N',M'}v_{\mathrm{lo}}\|_{L_t^{2r/(r-2)}L_x^r(I\times\mathbb{R}^3)}\|v\|^{p-1}_{L_{t,x}^{2(p-1)}}.\end{aligned}$$

To establish a bound for this expression, it is useful to introduce the notation

$$\tilde{a}_{N,M'} = \sum_{N'}\min\left\{\frac{N}{N'},\frac{N'}{N}\right\}^{\sigma} a_{N',M'}.$$

Thus summing over N and $M' \leq M$, we can again argue exactly as above to bound this expression by

$$\begin{aligned}\eta_0^{p-1}\sum_{M'\leq M}\left(\frac{M'}{M}\right)^{\sigma}\sum_{M''\leq M'}\left(\frac{M''}{M'}\right)\tilde{a}_{N,M''} &\leq \eta_0^{p-1}\sum_{M''\leq M}\left(\frac{M''}{M}\right)^{\sigma}\sum_{M''\leq M'}\left(\frac{M''}{M'}\right)\left(\frac{M'}{M''}\right)^{\sigma}\tilde{a}_{N,M''}\\ &\leq \eta_0^{p-1}\sum_{M''\leq M}\left(\frac{M''}{M}\right)^{\sigma}\tilde{a}_{N,M''}\sum_{M''\leq M'}\left(\frac{M''}{M'}\right)\left(\frac{M'}{M''}\right)^{\sigma}\\ &\lesssim \eta_0^{p-1}\alpha_{N,M}\end{aligned}$$

provided $\sigma < 1$. Thus, we obtain

$$\sum_{N'}\sum_{M'\leq M}\left(\frac{M'}{M}\right)^{\sigma}\min\left\{\frac{N}{N'},\frac{N'}{N}\right\}^{\sigma}\|P_{N,\geq M}F(v_{\mathrm{lo}})\|_{N(\mathbb{R})} \lesssim \eta_0^{p-1}\alpha_{N,M}. \tag{7-33}$$

By Strichartz estimates, we have

$$\|A'\|_{\dot H_x^{s_p}} + \|A\|_{\dot H_x^{s_p}} \lesssim \|P_{N,\ge M} F(v)\|_{N(I)}.$$

Hence, putting these bounds together with (7-32) and (7-33) we obtain

$$\sum_{N'}\sum_{M'\le M}\left(\frac{M'}{M}\right)^{\sigma}\min\left\{\frac{N}{N'},\frac{N'}{N}\right\}^{\sigma}(\|A\|_{\dot H_x^{s_p}} + \|A'\|_{\dot H_x^{s_p}}) \lesssim \eta_0^{p-1}\alpha_{N,M}.$$

Together with (7-30) and (7-31), we also have

$$\alpha_{N,M} \lesssim \gamma_{N,M}(N^{1-\epsilon}) + \eta_0^{p-1}\alpha_{N,M},$$

as required. □

Term B. We next estimate the terms $\langle B, B\rangle$ and $\langle B', B'\rangle$ from (7-27). On this region, we use the long-time Strichartz estimates (Proposition 7.3) and another frequency envelope argument for this contribution. In the following we suppose, unless otherwise specified, that norms be taken over

$$I := [0, N^{1-\epsilon}]\times\mathbb{R}^3.$$

We define

$$\begin{aligned} b_{N',M} = \Big[(N')^{-(\frac{2}{q}-s_p)}\|\widehat{P}_{N',\ge M}u\|_{L_t^q L_x^{2q/(q-2)}} + (N')^{\frac{3(p-3)}{2(p-1)}}\|P_{N',\ge M}u\|_{L_t^{2(p-1)}L_x^{2(p-1)/(p-2)}} \\ + (N')^{-\frac{1}{p-1}+\frac{3\theta}{2(p-1)}}\|P_{N',\ge M}u\|_{L_t^{p-1}L_x^{2(p-1)/\theta}} + (N')^{s_p-\frac{3\theta}{2}}\|P_{N',\ge M}u\|_{L_t^\infty L_x^{2/(1-\theta)}} \\ + (N')^{-\theta+s_p}\|P_{N',\ge M}u\|_{L_t^{2/\theta}L_x^{2/(1-\theta)}} + (N')^{\frac{\ell}{2}}\|P_{N',\ge M}u\|_{L_t^{2p/(1+s_p-p\ell)}L_x^{2p/(2-s_p)}} \\ + (N')^{\frac{2s_p}{3}-\frac{1}{3}}\|P_{N',\ge M}\|_{L_t^{6/(1+s_p)}L_x^{6/(2-s_p)}}\Big], \end{aligned}$$

where $\ell > 0$ will be determined more precisely below and θ is as in Proposition 7.3. These are just a collection of admissible Strichartz pairs at regularity s_p. We then define the frequency envelope

$$\beta_{N,M} = \sum_{N'}\sum_{M'\le M}\left(\frac{M'}{M}\right)^{\sigma}\min\left\{\frac{N}{N'},\frac{N'}{N}\right\}^{\sigma} b_{N',M'}.$$

Our goal in this section is to prove the following result.

Lemma 7.9. *Let B, B' and $\beta_{N,M}$ be as above. Then*

$$\sum_{N'}\sum_{M'\le M}\left(\frac{M'}{M}\right)^{\sigma}\min\left\{\frac{N}{N'},\frac{N'}{N}\right\}^{\sigma}(\|B\|_{\dot H_x^{s_p}} + \|B'\|_{\dot H_x^{s_p}}) \lesssim \eta_0^{p-1}\beta_{N,M},$$

and we also have

$$\gamma_{N,M}(N^{1-\epsilon}) + \beta_{N,M} \lesssim \gamma_{N,M}(0) + \eta_0^{p-1}\beta_{N,M}. \tag{7-34}$$

Proof. Fix $t_0 = 0$. Throughout, we will assume that $N \gg N_0$ as in the statement of the long-time Strichartz estimates. By Strichartz estimates

$$\|P_{N',\ge M}(u,u_t)\|_{L_t^\infty \dot{\mathcal{H}}^{s_p}([0,N^{1-\epsilon}])} + b_{N',M} \lesssim \|P_{N',\ge M}(u,u_t)(0)\|_{\mathcal{H}^{s_p}} + \|P_{N',\ge M}F\|_{N(I)}.$$

Once again, it will be useful to introduce the quantities

$$u_{\mathrm{lo}} = \sum_{N'} \widehat{P}_{N',\le M} u \quad \text{and} \quad u_{\mathrm{hi}} = \sum_{N'} \widehat{P}_{N',>M} u,$$

where "lo" and "hi" are meant to refer to the $\xi_{2,3}$-frequency component. We decompose the nonlinearity via

$$F(u) = F(u_{\mathrm{lo}}) + u_{\mathrm{hi}} \int_0^1 F'(u_{\mathrm{lo}} + \theta u_{\mathrm{hi}})\,\mathrm{d}\theta,$$

which we write schematically as

$$F(u) = F(u_{\mathrm{lo}}) + u_{\mathrm{hi}} u^{p-1}.$$

These two expressions will be estimated almost identically, up to requiring additional exponential gains for the low frequency (in M) term, $F(u_{\mathrm{lo}})$. We will only estimate this term since the other is easier. Arguing as above via the chain rule with the Laplacian in the $x_{2,3}$-directions, we have

$$\|P_{N,\ge M} F(u_{\mathrm{lo}})\|_{N(I)} \le M^{-1} \|P_{N,\ge M} (\nabla_{2,3} u_{\mathrm{lo}}) F'(u_{\mathrm{lo}})\|_{N(I)}.$$

We write

$$(\nabla_{2,3} u_{\mathrm{lo}}) F'(u_{\mathrm{lo}}) = (\nabla_{2,3} P_{\ge N} u_{\mathrm{lo}}) F'(u_{\mathrm{lo}}) + (\nabla_{2,3} P_{\le N} u_{\mathrm{lo}}) F'(u_{\mathrm{lo}}) := 1 + 2 \tag{7-35}$$

and we begin with term 1. We set

$$P_{\ge N} u_{\mathrm{lo}} := u_{\mathrm{lo},\ge N}, \quad P_{\le N} u_{\mathrm{lo}} := u_{\mathrm{lo},\le N},$$

and take the decomposition

$$\begin{aligned}
(\nabla_{2,3} u_{\mathrm{lo},\ge N}) F'(v_{\mathrm{lo}}) &= (\nabla_{2,3} u_{\mathrm{lo},\ge N}) F'(u_{\mathrm{lo},\le N}) + (\nabla_{2,3} u_{\mathrm{lo},\ge N}) u_{\mathrm{lo},\ge N} \int_0^1 F''(u_{\mathrm{lo},\le N} + \theta u_{\mathrm{lo},\ge N}) \\
&= (\nabla_{2,3} u_{\mathrm{lo},\ge N}) F'(u_{\mathrm{lo},\le N}) + \nabla_{2,3} u_{\mathrm{lo},\ge N} u_{\mathrm{lo},\ge N} F''(u_{\mathrm{lo},\le N}) \\
&\qquad + (\nabla_{2,3} u_{\mathrm{lo},\ge N}) u_{\mathrm{lo},\ge N}^2 \iint_0^1 F'''(u_{\mathrm{lo},\le N} + \theta_1\theta_2 u_{\mathrm{lo},\ge N}) \\
&:= 1.I + 1.II + 1.III.
\end{aligned}$$

Term 1.I. We estimate using Corollary 7.6 to get

$$\begin{aligned}
&M^{-1} \| |\nabla|^{s_p - 1} P_N \nabla_{2,3} u_{\mathrm{lo},\ge N} F'(u_{\mathrm{lo},\le N}) \|_{L_t^1 L_x^2} \\
&\quad \le M^{-1} N^{s_p - 1} \|\nabla_{2,3} u_{\mathrm{lo},\ge N} F'(u_{\mathrm{lo},\le N})\|_{L_t^1 L_x^2} \\
&\quad \le M^{-1} N^{s_p - 1} \|u_{\mathrm{lo},\le N}\|_{L_t^{p-1} L_x^{2(p-1)/\theta}}^{p-1} \|\nabla_{2,3} u_{\mathrm{lo},\ge N}\|_{L_t^\infty L_x^{2/(1-\theta)}} \\
&\quad \le N^{s_p - 1} \|u_{\mathrm{lo},\le N}\|_{L_t^{p-1} L_x^2(p-1)/\theta}^{p-1} \sum_{M' \le M} \sum_{N' \ge N} \left(\frac{M'}{M}\right) \|P_{N',M'} u\|_{L_t^\infty L_x^{2/(1-\theta)}} \\
&\quad \le N^{s_p - 1} N^{-s_p + \frac{3\theta}{2}} \|u_{\mathrm{lo},\le N}\|_{L_t^{p-1} L_x^{2(p-1)/\theta}}^{p-1} \\
&\qquad\qquad \times \sum_{M' \le M} \sum_{N' \ge N} \left(\frac{M'}{M}\right) \left(\frac{N'}{N}\right)^{-s_p + \frac{3\theta}{2}} (N')^{s_p - \frac{3\theta}{2}} \|P_{N',M'} u\|_{L_t^\infty L_x^{2/(1-\theta)}} \\
&\quad \le \eta_0^{p-1} \sum_{M' \le M} \sum_{N' \ge N} \left(\frac{M'}{M}\right) \left(\frac{N'}{N}\right)^{-s_p + \frac{3\theta}{2}} b_{N',M'}.
\end{aligned}$$

Term 1.II. We estimate

$$
\begin{aligned}
M^{-1} \| |\nabla|^{s_p-1} P_N \nabla_{2,3} u_{\mathrm{lo},\geq N} u_{\mathrm{lo},\geq N} F''(u_{\mathrm{lo},\leq N}) \|_{L_t^1 L_x^2} \\
&\lesssim M^{-1} N^{s_p-1} \|\nabla_{2,3} u_{\mathrm{lo},\geq N}\|_{L_t^{2(p-1)} L_x^4} \|u_{\mathrm{lo},\geq N}\|_{L_t^{2(p-1)} L_x^4} \|u_{\mathrm{lo},\leq N}\|^{p-2}_{L_t^{p-1} L_x^\infty} \\
&\lesssim M^{-1} N^{s_p-1} \|\nabla_{2,3} u_{\mathrm{lo},\geq N}\|_{L_t^{2(p-1)} L_x^4} \|u_{\mathrm{lo},\geq N}\|_{L_t^{2(p-1)} L_x^4} \|u_{\mathrm{lo},\leq N}\|^{p-2}_{L_t^{p-1} L_x^\infty} \\
&\lesssim \eta_0^{p-1} \sum_{M' \leq M} \sum_{N' \geq N} \left(\frac{M'}{M}\right) \left(\frac{N}{N'}\right)^{\frac{3}{4} - \frac{3}{2(p-1)}} (N')^{\frac{3}{4} - \frac{3}{2(p-1)}} \|P_{N',M'} u\|_{L_t^{2(p-1)} L_x^4} \\
&\lesssim \eta_0^{p-1} \sum_{M' \leq M} \sum_{N' \geq N} \left(\frac{M'}{M}\right) \left(\frac{N}{N'}\right)^{\frac{3}{4} - \frac{3}{2(p-1)}} b_{N',M'}.
\end{aligned}
$$

Term 1.III. As in the proof of term IV in the long-time Strichartz estimates, there are two terms. For the first we estimate

$$
\begin{aligned}
M^{-1} \|\nabla_{2,3} u_{\mathrm{lo},\geq N} u^2_{\mathrm{lo},\geq N} u^{p-3}_{\mathrm{lo},\geq N}\|_{L_t^{2/(1+s_p)} L_x^{2/(2-s_p)}} \\
&\lesssim M^{-1} N^{\frac{\ell}{2}} \|\nabla_{2,3} u_{\mathrm{lo},\geq N} u^2_{\mathrm{lo},\geq N} u^{p-3}_{\mathrm{lo},\geq N}\|_{L_t^{2/(1+s_p-\ell)} L_x^{2/(2-s_p)}} \\
&\lesssim M^{-1} N^{\frac{\ell}{2}} \|u_{>\frac{N}{8}}\|^{p-1}_{L_t^{2p/(1+s_p)} L_x^{2p/(2-s_p)}} \|u_{\mathrm{lo},\geq N}\|_{L_t^{2p/(1+s_p-p\ell)} L_x^{2p/(2-s_p)}} \\
&\lesssim \eta_0^{p-1} \sum_{M' \leq M} \sum_{N' \geq N} \left(\frac{M'}{M}\right) \left(\frac{N}{N'}\right)^{\frac{\ell}{2}} (N')^{\frac{\ell}{2}} \|P_{N',M'} u\|_{L_t^{2p/(1+s_p-p\ell)} L_x^{2p/(2-s_p)}} \\
&\lesssim \eta_0^{p-1} \sum_{M' \leq M} \sum_{N' \geq N} \left(\frac{M'}{M}\right) \left(\frac{N}{N'}\right)^{\frac{\ell}{2}} b_{N',M'},
\end{aligned}
$$

where we have used that for $p > 3$ and $\ell > 0$, the pair

$$
\left(\frac{2p}{1 + s_p - p\ell}, \frac{2p}{2 - s_p}\right)
$$

is wave-admissible at regularity

$$
\frac{3}{2} - \frac{1 + s_p - p\ell}{2p} - \frac{6 - 3s_p}{2p} = s_p - \frac{\ell}{2}.
$$

For the second term, we have

$$
\begin{aligned}
M^{-1} \|\nabla_{2,3} u_{\mathrm{lo},\geq N} u^2_{\mathrm{lo},\geq N} u^{p-3}_{\mathrm{lo},\leq N}\|_{L_t^{2/(1+s_p)} L_x^{2/(2-s_p)}} \\
&\lesssim \eta_0^{p-1} \sum_{M' \leq M} \sum_{N' \geq N} \left(\frac{M'}{M}\right) \left(\frac{N}{N'}\right)^{\frac{2s_p}{3} - \frac{1}{3}} (N')^{\frac{2s_p}{3} - \frac{1}{3}} \|P_{N',M'} u\|_{L_t^{6/(1+s_p)} L_x^{6/(2-s_p)}} \\
&\lesssim \eta_0^{p-1} \sum_{M' \leq M} \sum_{N' \geq N} \left(\frac{M'}{M}\right) \left(\frac{N}{N'}\right)^{\frac{2s_p}{3} - \frac{1}{3}} b_{N',M'}.
\end{aligned}
$$

This completes the estimation of term 1 in (7-35).

Now we turn to term 2 in (7-35), namely

$$\nabla_{2,3} P_{\leq N} u_{\mathrm{lo}} F'(u_{\mathrm{lo}}).$$

We take the decomposition

$$\begin{aligned}\nabla_{2,3} u_{\mathrm{lo},\leq N} F'(v_{\mathrm{lo}}) &= \nabla_{2,3} u_{\mathrm{lo},\leq N} F'(u_{\mathrm{lo},\leq N}) + \nabla_{2,3} u_{\mathrm{lo},\leq N} u_{\mathrm{lo},\geq N} \int_0^1 F''(u_{\mathrm{lo},\leq N} + \theta u_{\mathrm{lo},\geq N}) \\ &= \nabla_{2,3} u_{\mathrm{lo},\leq N} F'(u_{\mathrm{lo},\leq N}) + \nabla_{2,3} u_{\mathrm{lo},\geq N} u_{\mathrm{lo},\geq N} F''(u_{\mathrm{lo},\leq N}) \\ &\qquad + \nabla_{2,3} u_{\mathrm{lo},\leq N} u^2_{\mathrm{lo},\geq N} \iint F'''(u_{\mathrm{lo},\leq N} + \theta_1\theta_2 u_{\mathrm{lo},\geq N}) \\ &:= 2.I + 2.II + 2.III.\end{aligned}$$

We omit the estimates for the first two terms since they follow as above, and we focus on

$$\nabla_{2,3} u_{\mathrm{lo},\leq N} u^2_{\mathrm{lo},\geq N} \iint_0^1 F'''(u_{\mathrm{lo},\leq N} + \theta_1\theta_2 u_{\mathrm{lo},\geq N}) =: \nabla_{2,3} u_{\mathrm{lo},\leq N} u^2_{\mathrm{lo},\geq N} F_3.$$

Here, we will need to introduce some new exponent pairs compared to the proof of the long-time Strichartz estimates. We divide this expression into two parts:

$$\begin{aligned}&\| |\nabla|^{-\frac{p-3}{2(p-1)}+s_p} P_N (\nabla_{2,3} u_{\mathrm{lo},\leq N} u^2_{\mathrm{lo},\geq N} F)\|_{L_t^{(p-1)/(p-2)} L_x^1} \\ &\quad \lesssim N^{-\frac{p-3}{2(p-1)}+s_p} \|\nabla_{2,3} u_{\mathrm{lo},\leq N} u^{p-1}_{\mathrm{lo},\geq N}\|_{L_t^{(p-1)/(p-2)} L_x^1} \\ &\qquad\qquad + N^{-\frac{p-3}{2(p-1)}+s_p} \|\nabla_{2,3} u_{\mathrm{lo},\leq N} u^2_{\mathrm{lo},\geq N} u^{p-3}_{\mathrm{lo},\leq N}\|_{L_t^{(p-1)/(p-2)} L_x^1}.\end{aligned}$$

Note that $\left(\frac{p-1}{p-2}, 1\right)$ is dual wave-admissible for $p \geq 3$.

For the first term, we have a bound of

$$\begin{aligned}&N^{-\frac{p-3}{2(p-1)}-s_p} \| |\nabla|^{s_p} u_{>N}\|^2_{L_t^\infty L_x^2} \|u_{\leq N}\|^{p-3}_{L_t^{p-1} L_x^\infty} \sum_{M' \leq M} M' \|P_{M'} u_{\leq N}\|_{L_t^{p-1} L_x^\infty} \\ &\qquad \lesssim \eta_0^{p-1} \sum_{M' \leq M} \sum_{N' \leq N} M' \left(\frac{N'}{N}\right)^{\frac{1}{p-1}} (N')^{-\frac{1}{p-1}} \|P_{N',M'} u\|_{L_t^{p-1} L_x^\infty}.\end{aligned}$$

For the second term we estimate

$$\begin{aligned}&N^{-\frac{p-3}{2(p-1)}+s_p} \|\nabla_{2,3} u_{\mathrm{lo},\leq N} u^{p-1}_{\mathrm{lo},\geq N}\|_{L_t^{(p-1)/(p-2)} L_x^1} \\ &\quad \lesssim N^{-\frac{p-3}{2(p-1)}+s_p} \|u_{\geq N}\|^{p-3}_{L_{t,x}^{2(p-1)}} \|u_{\geq N}\|^2_{L_t^4 L_x^{4(p-1)/(p+1)}} \\ &\qquad\qquad \times \sum_{M' \leq M} \sum_{N' \leq N} M' \left(\frac{N'}{N}\right)^{\frac{2}{p-1}} (N')^{-\frac{2}{p-1}} \|P_{N',M'} u\|_{L_t^\infty L_x^\infty}.\end{aligned}$$

Now we note that the pair

$$\left(4, \frac{4(p-1)}{p+1}\right)$$

is wave-admissible at regularity

$$\frac{3}{2} - \frac{1}{4} - \frac{3p+3}{4(p-1)} = s_p + \frac{2}{p-1} - \frac{1}{4} - \frac{3p+3}{4(p-1)}.$$

Noting that

$$\frac{3p+3}{4(p-1)} = \frac{3(p+1)}{4(p-1)} > \frac{3}{(p-1)},$$

we see that this is number strictly less than s_p. Thus we obtain a bound of

$$\eta_0^{p-1} \sum_{M' \le M} \sum_{N' \le N} M' \left(\frac{N'}{N}\right)^{\frac{2}{p-1}} (N')^{-\frac{2}{p-1}} \|P_{N',M'} u\|_{L_t^\infty L_x^\infty}.$$

Arguing as in the estimates for term A, we may determine the restrictions on σ. First, we need to assume that $\sigma < 1$ so that we can perform the summation in M, and we further require that σ be bounded above by the power appearing on the N'/N factor when $N' \le N$ and the N/N' factor when $N \le N'$. Examining the exponents in the definition of $\mathcal{S}_{N,M}$, this amounts to requiring σ smaller than the smallest (in absolute values) exponent in that expression, and hence we may assume the most restrictive of these will be taking $\sigma < \ell/2$ in term 1.III.

Provided this is the case, we obtain

$$\sum_{N'} \sum_{M' \le M} \left(\frac{M'}{M}\right)^{\sigma} \min\left\{\frac{N}{N'}, \frac{N'}{N}\right\}^{\sigma} \|P_{N', \ge M} F(u)\|_{N(I)} \lesssim \eta_0^{p-1} \beta_{N,M},$$

and since, by Strichartz estimates

$$\|B\|_{\dot{H}_x^{s_p}} + \|B'\|_{\dot{H}_x^{s_p}} \lesssim \|P_{N', \ge M} F\|_{N(I)}.$$

we have

$$\sum_{N'} \sum_{M' \le M} \left(\frac{M'}{M}\right)^{\sigma} \min\left\{\frac{N}{N'}, \frac{N'}{N}\right\}^{\sigma} (\|B\|_{\dot{H}_x^{s_p}} + \|B'\|_{\dot{H}_x^{s_p}}) \lesssim \eta_0^{p-1} \beta_{N,M},$$

as well as the estimate

$$\gamma_{N,M}(N^{1-\epsilon}) + \beta_{N,M} \lesssim \gamma_{N,M}(0) + \eta_0^{p-1} \beta_{N,M},$$

as required. □

Term C. We turn to the $\langle C, C' \rangle$ term (see (7-27) and (7-28)). In this section we prove the following lemma.

Lemma 7.10. *Let C, C' be defined as in (7-27), and let $M \ge C_0 N^{s_p/(1-\nu)}$. Then, for any $L \in \mathbb{N}$ we have*

$$|\langle C, C' \rangle_{\dot{H}_x^{s_p}}| \lesssim_L \frac{1}{M^L},$$

where the implicit constant above depends only on L.

Proof of Lemma 7.10. We introduce the notation

$$G(u, v)(t) = F(u(t)) - F(v(t)),$$

which we may abbreviate as $G(t)$ or even G. We are faced with estimating

$$\langle C, C' \rangle_{\dot{H}_x^{s_p}} \simeq N^{2s_p} \langle C, C' \rangle_{L_x^2},$$

where

$$\begin{aligned}\langle C, C'\rangle_{L^2_x} &= \int_{-\infty}^{-N^{1-\epsilon}} \int_{N^{1-\epsilon}}^{\infty} \langle S(-t)\widehat{P}_{N,M} G(u,v)(t), S(-\tau)\widehat{P}_{N,M} G(u,v)(\tau)\rangle_{L^2_x}\,\mathrm{d}t\,\mathrm{d}\tau \\ &= \int_{-\infty}^{-N^{1-\epsilon}} \int_{N^{1-\epsilon}}^{\infty} \langle G(u,v)(t),\ S(t-\tau)\widehat{P}^2_{N,M}\frac{1}{|\nabla|} G(u,v)(\tau)\rangle_{L^2_x}\,\mathrm{d}t\,\mathrm{d}\tau. \end{aligned} \tag{7-36}$$

Since $M \geq C_0 N^{s_p/(1-\nu)}$, it suffices to show that

$$|\langle C, C'\rangle_{L^2_x}| \lesssim_L \frac{1}{M^L},$$

and all inner products in this proof will be L^2_x inner products.

For each fixed t, τ as above we estimate the pairing,

$$\langle G(u,v)(t), S(t-\tau)\widehat{P}^2_{N,M} G(u,v)(\tau)\rangle_{L^2_x}.$$

Recall that by the definition of $\vec{v}(\tau)$, $G(u,v)(\tau)$ is supported in the region

$$\mathcal{G}_\pm(\tau) := \{x : |x - x(\pm N^{1-\epsilon})| \leq R(\eta_0) + |\tau| - N^{1-\epsilon} \text{ for all } \pm\tau \geq N^{1-\epsilon}\}. \tag{7-37}$$

This points to an immediate problem in any naive implementation of the double Duhamel trick by way of the Huygens principle as performed in previous sections. Namely, the support of the $S(t-\tau)$ evolution of $G(u,v)(\tau)$ intersects with the support of $G(u,v)(t)$ in the "wave zone", i.e., near the boundary of the light cone where the kernel of $S(t-\tau)$ only yields $\langle t-\tau\rangle^{-1}$ decay, which is not sufficient for integration in time. However, we are saved here by a gain in *angular separation* in the wave zone guaranteed by our directional frequency localization $\widehat{P}_{N,M}$. Indeed, application of $\widehat{P}_{N,M}$ restricts to frequencies $\xi = (\xi_1, \xi_{2,3})$ with

$$\frac{|\xi_{2,3}|}{|\xi|} \simeq \frac{M}{N},$$

whereas for any $x = (x_1, x_{2,3}) \in \mathcal{G}(t) \cap \{(t,x) : |x| \geq t - R(\eta_0)\}$ we claim that

$$\frac{|x_{2,3}|}{|x|} \ll \frac{M}{N}$$

for all $M \geq N^{s_p/(1-\nu)}$. We establish this fact in Lemma 7.11 below.

We introduce some additional notation. Let $R(\cdot)$ be the compactness modulus function. For given $t \in \mathbb{R}$ let

$$\begin{aligned}\mathcal{C}_{\text{ext}}(t) &:= \{x : |x| \geq |t| - R(\eta_0)\}, \\ \mathcal{C}_{\text{int}}(t) &:= \{x : |x| \leq |t| - R(\eta_0)\}.\end{aligned} \tag{7-38}$$

We decompose $\langle C, C'\rangle$ as follows. First, we write

$$G(u,v)(t) = G(u,v)(t)1_{\mathcal{C}_{\text{ext}}(t)} + G(u,v)(t)1_{\mathcal{C}_{\text{int}}(t)}.$$

Using this decomposition in (7-36) leads to four terms:

$$\iint \left\langle S(t-\tau)\frac{1}{|\nabla|}\widehat{P}_{N,M}\,1_{\mathcal{C}_{\mathrm{ext}}}(\tau)G(\tau),\, 1_{\mathcal{C}_{\mathrm{ext}}}(t)G(t)\right\rangle \mathrm{d}t\,\mathrm{d}\tau, \tag{7-39}$$

$$\iint \left\langle S(t-\tau)\frac{1}{|\nabla|}\widehat{P}_{N,M}\,1_{\mathcal{C}_{\mathrm{ext}}}(\tau)G(\tau),\, 1_{\mathcal{C}_{\mathrm{int}}}(t)G(t)\right\rangle \mathrm{d}t\,\mathrm{d}\tau, \tag{7-40}$$

$$\iint \left\langle S(t-\tau)\frac{1}{|\nabla|}\widehat{P}_{N,M}\,1_{\mathcal{C}_{\mathrm{int}}}(\tau)G(\tau),\, 1_{\mathcal{C}_{\mathrm{ext}}}(t)G(t)\right\rangle \mathrm{d}t\,\mathrm{d}\tau, \tag{7-41}$$

$$\iint \left\langle S(t-\tau)\frac{1}{|\nabla|}\widehat{P}_{N,M}\,1_{\mathcal{C}_{\mathrm{int}}}(\tau)G(\tau),\, 1_{\mathcal{C}_{\mathrm{int}}}(t)G(t)\right\rangle \mathrm{d}t\,\mathrm{d}\tau, \tag{7-42}$$

where the integrals are over $[-\infty,-N^{1-\epsilon}]\times[N^{1-\epsilon},\infty]$. We will refer to these terms as $C_{\mathrm{ext-ext}}$, $C_{\mathrm{ext-int}}$, $C_{\mathrm{int-ext}}$ and $C_{\mathrm{int-int}}$ respectively, and we will handle these terms separately below. Were it not for the frequency localization $\widehat{P}_{N,M}$ all but the first term above would vanish using the support properties of $G(u,v)$, together with the particular pairing of the cutoffs $1_{\mathcal{C}_{\mathrm{int}}}$ and $1_{\mathcal{C}_{\mathrm{ext}}}$, and the sharp Huygens principle. On the other hand, whereas in previous scenarios (e.g., the subluminal soliton) the first term would vanish, in the present setting there truly is an interaction between these two terms. This is the origin of the essential technical difficulty faced in the present scenario, and indeed we will find that the first term (7-39) requires the most careful analysis. The crucial observation is that in this setting we can rely on angular separation to exhibit decay.

The term $C_{\mathrm{ext-ext}}$. We will rely crucially on the following two lemmas, which together make precise the gain in decay from angular separation.

Lemma 7.11 (angular separation in the wave zone). *For any $c>0$ there exists $N_0=N_0(c)>0$ with the following property. Fix $\nu\in(0,1)$ and let $\epsilon>0$ be any number with*

$$\epsilon<\frac{2s_p}{1-\nu}-1.$$

Let (t,x) satisfy

$$|t|\geq N^{1-\epsilon},\quad x=(x_1,x_{2,3})\in\mathcal{G}(t)\cap\mathcal{C}_{\mathrm{ext}}(t),$$

where $\mathcal{G}(t)$ are defined in (7-37), (7-38). Then,

$$\frac{|x_{2,3}|}{|x|}\lesssim\frac{1}{N^{\frac{1}{2}-\frac{\epsilon}{2}}}\leq c\frac{M}{N} \tag{7-43}$$

for all $N\geq N_0$ and $M\geq N^{s_p/(1-\nu)}$.

See Figure 3 for a depiction of Lemma 7.11.

Next, we show that if we restrict to those $x\in\mathbb{R}^3$ satisfying (7-43) then we get strong pointwise decay for the kernel of the operator $S(t)\frac{1}{|\nabla|}\widehat{P}_{N,M}^2$.

To state the result, we define

$$\mathcal{S}_N:=\left\{x\in\mathbb{R}^3:\frac{|x_{2,3}|}{|x|}\lesssim\frac{1}{N^{\frac{1}{2}-\frac{\epsilon}{2}}}\right\}. \tag{7-44}$$

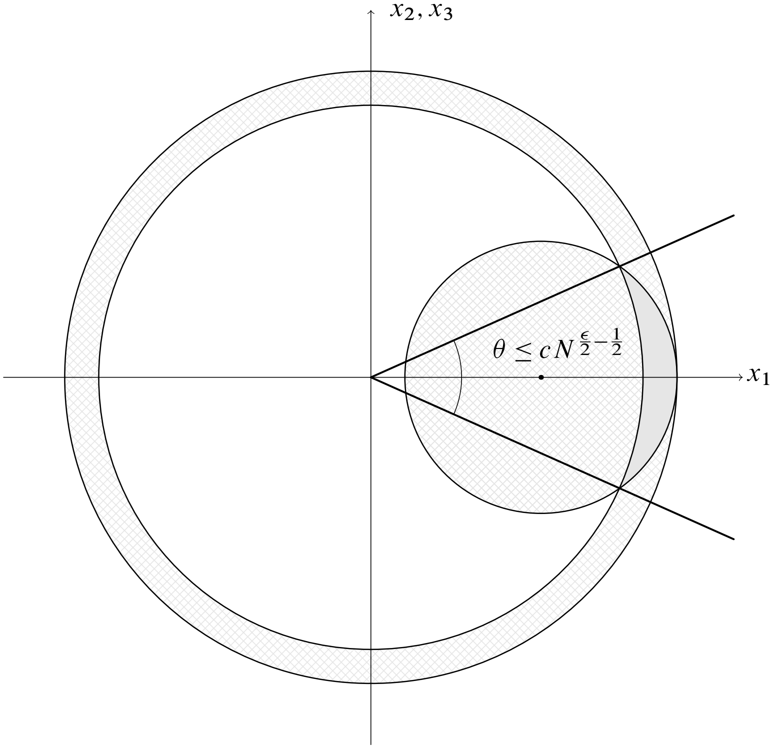

Figure 3. The dark gray region above represents the region $\mathcal{G}_+(t) \cap \mathcal{C}_{\text{ext}}$ in space at fixed time $t > N^{1-\epsilon}$.

Lemma 7.12 (kernel estimates via angular separation). *Let $K_{N,M}(t,x)$ denote the kernel of the operator $S(t)\frac{1}{|\nabla|}\widehat{P}^2_{N,M}$. Let $N \geq N_0$ where N_0 is as in the hypothesis of Lemma 7.11. Then, for any L,*

$$|1_{\mathcal{S}_N}(x)K_{N,M}(t,x)| \lesssim_L N\frac{N^L}{M^L}\frac{1}{\langle M|x|\rangle^L} \quad \text{for all } t \geq N^{1-\epsilon}, \tag{7-45}$$

where $\mathcal{S}_N$ is the set defined in (7-44) and where we have used the notation $\langle z\rangle := (1+|z|^2)^{1/2}$ above.

Proof of Lemma 7.11. We assume that $t \geq 0$. Since we are assuming

$$0 < \epsilon < \frac{2s_p}{1-\nu} - 1$$

and that $M \geq N^{s_p/(1-\nu)}$, it suffices to show the first inequality in (7-43), i.e., that

$$\frac{|x_{2,3}|}{|x|} \lesssim \frac{1}{N^{\frac{1}{2}-\frac{\epsilon}{2}}} \tag{7-46}$$

for all $x \in \mathcal{G}(t) \cap \mathcal{C}_{\text{ext}}(t)$ for some uniform constant.

First, we claim that (7-46) holds at time $t = N^{1-\epsilon}$. Suppose

$$x \in \mathcal{G}_+(N^{1-\epsilon}) \cap \mathcal{C}_{\text{ext}}(N^{1-\epsilon}).$$

By the definition of traveling wave (i.e., (7-1) and (7-2)) we have

$$|x| \simeq N^{1-\epsilon}, \quad |x_{2,3}| \lesssim N^{\frac{1}{2}-\frac{\epsilon}{2}}$$

and thus,

$$\frac{|x_{2,3}|}{|x|} \lesssim N^{\frac{\epsilon}{2}-\frac{1}{2}}$$

as desired.

Now suppose $t > N^{1-\epsilon}$. We introduce some notation. Let $\theta_{x(N^{1-\epsilon})}$ denote the angle between the unit vector $\vec{e}_1$ (the unit vector in the positive x_1-direction) and the vector $x(N^{1-\epsilon})$, where we recall that $x(t)$ denotes the spatial center of $\vec{u}$. Above, we have just shown that

$$|\sin(\theta_{x(N^{1-\epsilon})})| \simeq |\theta_{x(N^{1-\epsilon})}| \le A_1 N^{\frac{\epsilon}{2}-\frac{1}{2}}$$

for some uniform constant $A_1 > 0$. To finish the proof it will suffice to show that for any $x \in \mathcal{G}(t) \cap \mathcal{C}_{\text{ext}}(t)$, the angle $\theta_{(x,x(N^{1-\epsilon}))}$ formed between the vectors x and $x(N^{1-\epsilon})$ satisfies

$$|\theta_{(x,x(N^{1-\epsilon}))}| \le A_2 N^{\frac{\epsilon}{2}-\frac{1}{2}}$$

for some other uniform constant $A_2 > 0$, as then the sine of the total angle between x and the x_1-axis, i.e., $|x_{2,3}|/|x|$ would satisfy (7-46). To get a hold of $\theta_{(x,x(N^{1-\epsilon}))}$ we square both sides of the inequality defining the set $\mathcal{G}_+(t)$. For $x \in \mathcal{G}(t)$ we have

$$|x|^2 - 2x \cdot x(N^{1-\epsilon}) + |x(N)^{1-\epsilon}|^2 \le (R(\eta_0) + t - N^{1-\epsilon})^2.$$

Using that $x \cdot x(N^{1-\epsilon}) = |x||x(N^{1-\epsilon})| \cos \theta_{(x,x(N^{1-\epsilon}))}$ the above yields the inequality

$$-2|x||x(N^{1-\epsilon})| \cos \theta_{(x,x(N^{1-\epsilon}))} \le (R(\eta_0) + t - N^{1-\epsilon})^2 - |x|^2 - |x(N^{1-\epsilon})|^2.$$

Bootstrapping, we may assume that $\theta_{(x,x(N^{1-\epsilon}))}$ is small enough to use the estimate

$$\cos \theta_{(x,x(N^{1-\epsilon}))} \le 1 - \frac{\theta^2_{(x,x(N^{1-\epsilon}))}}{4}.$$

Plugging the above in we arrive at the inequality

$$\begin{aligned} \frac{\theta^2_{(x,x(N^{1-\epsilon}))}}{2} &\le 2 + \frac{1}{|x||x(N^{1-\epsilon})|}((R(\eta_0) + t - N^{1-\epsilon})^2 - |x|^2 - |x(N^{1-\epsilon})|^2) \\ &\le \frac{2|x||x(N^{1-\epsilon})| + (R(\eta_0) + t - N^{1-\epsilon})^2 - |x|^2 - |x(N)^{1-\epsilon}|^2}{|x||x(N^{1-\epsilon})|}. \end{aligned}$$

The requirement that $x \in C_{\text{ext}}(t)$, finite speed of propagation, and (7-1) imply that we have

$$t - R(\eta_0) \le |x| \le t + R(\eta_0) \quad \text{and} \quad N^{1-\epsilon} - R(\eta_0) \le |x(N^{1-\epsilon})| \le N^{1-\epsilon} + R(\eta_0).$$

Plugging the above into the previous line gives

$$\begin{aligned}\frac{\theta^2_{(x,x(N^{1-\epsilon}))}}{2} &\le \frac{2(t+R(\eta_0))(N^{1-\epsilon}+R(\eta_0))+(R(\eta_0)+t-N^{1-\epsilon})^2}{(t-R(\eta_0))(N^{1-\epsilon}-R(\eta_0))}-\frac{(t-R(\eta_0))^2-(N^{1-\epsilon}-R(\eta_0))^2}{(t-R(\eta_0))(N^{1-\epsilon}-R(\eta_0))}\\ &= \frac{6tR(\eta_0)+2N^{1-\epsilon}R(\eta_0)+R(\eta_0)^2}{(t-R(\eta_0))(N^{1-\epsilon}-R(\eta_0))}\\ &\lesssim \frac{1}{N^{1-\epsilon}}+\frac{1}{t}.\end{aligned}$$

Taking the square root and noting that $t \ge N^{1-\epsilon}$ we arrive at

$$|\theta_{(x,x(N^{1-\epsilon}))}| \lesssim \frac{1}{N^{\frac{1}{2}-\frac{\epsilon}{2}}}$$

as desired. □

Next, we prove Lemma 7.12.

Proof of Lemma 7.12. The kernel $K_{N,M}$ of the operator $S(t)\frac{1}{|\nabla|}\widehat{P}^2_{N,M}$ is given by

$$K_{N,M}(t,x) := \int e^{ix\cdot\xi}|\xi|^{-2}e^{it|\xi|}\phi^2\Big(\frac{|\xi|}{N}\Big)\phi^2\Big(\frac{|(\xi_2,\xi_3)|}{M}\Big)\,\mathrm{d}\xi,$$

where $\phi \in C_0^\infty(\mathbb{R})$ is satisfies $\phi(r)=1$ if $1 \le r \le 2$ and $\operatorname{supp}\phi \in \big(\frac{1}{4},4\big)$. Now, recall that we are restricting to only those $x \in \mathcal{S}_N$, as defined in (7-44). We express any such x in spherical coordinates

$$x = |x|(\cos\theta_x, \sin\theta_x\cos\omega, \sin\theta_x\sin\omega),$$

where θ_x denotes the angle formed by x and the unit vector in the e_1-direction. And recall that any $x \in \mathcal{S}_N$ satisfies

$$\frac{|x_{2,3}|}{|x|} = \sin\theta_x \simeq |\theta_x| \lesssim \frac{1}{N^{1-\epsilon}}. \tag{7-47}$$

Similarly, we change to the spherical variables

$$\xi = |\xi|(\cos\theta_\xi, \sin\theta_\xi\cos\alpha, \sin\theta_\xi\sin\alpha)$$

in the integral defining $K_{N,M}$ and note that because of the frequency localization $\widehat{P}_{N,M}$ we have

$$\frac{|\xi_{2,3}|}{|\xi|} = \sin\theta_\xi \simeq \frac{M}{N}. \tag{7-48}$$

This yields

$$K_{N,M}(t,x) = \int_0^{2\pi}\int_0^{\pi}\int_{\frac{N}{4}}^{4N} e^{i|x||\xi|f(\theta_x,\theta_\xi,\omega,\alpha)}|\xi|^{-2}e^{it|\xi|}\phi^2\Big(\frac{|\xi|}{N}\Big)\phi^2\Big(\frac{|\xi|\sin\theta_\xi}{M}\Big)|\xi|^2\sin\theta_\xi\,\mathrm{d}|\xi|\,\mathrm{d}\theta_\xi\,\mathrm{d}\alpha,$$

where the angular phase function $f(\theta_x,\theta_\xi,\omega,\alpha)$ is given by

$$f(\theta_x,\theta_\xi,\omega,\alpha) = \cos\theta_x\cos\theta_\xi + \sin\theta_x\sin\theta_\xi(\cos\omega\cos\alpha+\sin\omega\sin\alpha).$$

The idea is that the angular separation between x and ξ given by (7-47) and (7-48) allows us to integrate by parts in θ_ξ. Indeed, using (7-47) and (7-48) we have the lower bound

$$\begin{aligned}\left|\frac{d}{d\theta_\xi}[|x||\xi|f(\theta_x,\theta_\xi,\omega,\alpha)]\right| &= |x||\xi|\left|-\cos\theta_x\sin\theta_\xi+\sin\theta_x\cos\theta_\xi(\cos\omega\cos\alpha+\sin\omega\sin\alpha)\right|\\ &\geq |x||\xi|\left(\frac{M}{N}-O\left(\frac{1}{N^{\frac{1}{2}-\frac{\epsilon}{2}}}\right)\right)\\ &\gtrsim |x|M.\end{aligned}$$

Moreover, note that for any $L\in\mathbb{N}$ and $M\lesssim N$

$$\left|\frac{d^L}{d\theta_\xi^L}\left(\phi^2\left(\frac{|\xi|\sin\theta_\xi}{M}\right)\sin\theta_\xi\right)\right|\lesssim\frac{N^L}{M^L}.$$

Thus, integration by parts L-times in θ_ξ yields the estimate

$$|K_{N,M}(t,x)|\lesssim_L N^2\frac{N^L}{M^L}\frac{1}{\langle M|x|\rangle^L}\quad\text{for all } t\geq N^{1-\epsilon},\quad x\in\mathcal{G}(t)\cap\mathcal{C}_{\text{ext}}(t),$$

as desired. □

We can now estimate (7-39). Here will rely crucially on Lemmas 7.11 and 7.12. First we write,

$$\begin{aligned}&\left\langle S(t-\tau)\frac{1}{|\nabla|}\widehat{P}_{N,M}1_{\mathcal{C}_{\text{ext}}}(\tau)G(u,v)(\tau),\ 1_{\mathcal{C}_{\text{ext}}}(t)G(u,v)(t)\right\rangle\\ &\qquad\qquad= \langle K_{N,M}(t-\tau)*1_{\mathcal{C}_{\text{ext}}}(\tau)G(u,v)(\tau),\ 1_{\mathcal{C}_{\text{ext}}}(t)G(u,v)(t)\rangle.\end{aligned}$$

We claim that in fact the above can be expressed as

$$\begin{aligned}&\langle K_{N,M}(t-\tau)*1_{\mathcal{C}_{\text{ext}}}(\tau)G(\tau),1_{\mathcal{C}_{\text{ext}}}(t)G(t)\rangle\\ &\qquad=\left\langle(1_{\mathcal{S}_N}(\cdot)1_{\{|\cdot|\geq\frac{1}{2}|t-\tau|\}}(\cdot)K_{N,M}(t-\tau))*1_{\mathcal{C}_{\text{ext}}}(\tau)G(\tau),\ 1_{\mathcal{C}_{\text{ext}}}(t)G(t)\right\rangle,\end{aligned}\tag{7-49}$$

where the set $\mathcal{S}_N$ is defined in (7-44). Indeed, note that above we have

$$x\in\mathcal{G}_+(t)\cap\mathcal{C}_{\text{ext}}(t)\quad\text{and}\quad y\in\mathcal{G}_-(\tau)\cap\mathcal{C}_{\text{ext}}(\tau),\tag{7-50}$$

where $\mathcal{G}_\pm$ are as in (7-37) and C_{ext} is as in (7-38). Thus,

$$|x-y|\geq|t-\tau|-2R(\eta_0)\geq\tfrac{1}{2}|t-\tau|$$

as long as N is chosen large enough. Similarly by (7-50) we have $|x-y|\geq|x|$ and $|x-y|\geq|y|$ and thus,

$$\frac{|x_{2,3}-y_{2,3}|}{|x-y|}\leq\frac{|x_{2,3}|}{|x|}+\frac{|y_{2,3}|}{|y|}\lesssim\frac{1}{N^{\frac{1}{2}-\frac{\epsilon}{2}}},$$

where in the last inequality above we used Lemma 7.11. This proves the equality in (7-49).

Now, let q_p denote the Sobolev embedding exponent for $\dot{H}^{s_p}$, i.e., $q_p=\frac{3(p-1)}{2}$. Note that $q_p\geq p$ for $p\geq 3$ and $(q_p/p)'\geq 2$ for $p>0$ (where x' denotes the Hölder dual of x). By Hölder's and Young's inequalities we then have

$$\begin{aligned}&\left|\left\langle (1_{S_N}(\cdot)1_{\{|\cdot|\geq\frac12|t-\tau|\}}(\cdot)K_{N,M}(t-\tau)) * 1_{\mathcal{C}_{\text{ext}}}(\tau)G(\tau),\ 1_{\mathcal{C}_{\text{ext}}}(t)G(t)\right\rangle\right| \\ &\qquad \leq \|1_{S_N}(\cdot)1_{\{|\cdot|\geq\frac12|t-\tau|\}}(\cdot)K_{N,M}(t-\tau)\|_{L_x^{(q_p/p)'/2}} \|G(u,v)(t)\|_{L_x^{q_p/p}} \|G(u,v)(\tau)\|_{L_x^{q_p/p}}.\end{aligned}$$

Using (7-45) we see that

$$\begin{aligned}\|1_{S_N}(\cdot)1_{\{|\cdot|\geq\frac12|t-\tau|\}}(\cdot)K_{N,M}(t-\tau)\|_{L_x^{(q_p/p)'/2}} &\lesssim \frac{N^{L+1}}{M^{2L}}\left(\int_{|x|\geq\frac12|t-\tau|}\frac{1}{|x|^{2L(\frac{q_p}{p})'}}\,\mathrm{d}x\right)^{\frac{2}{(\frac{q_p}{p})'}} \\ &\lesssim \frac{N^{L+1}}{M^{2L}}\frac{1}{|t-\tau|^{L-1}}.\end{aligned} \tag{7-51}$$

Since $G(u,v) = F(u) - F(v)$ we have

$$\|G(u,v)(t)\|_{L_x^{q_p/p}} \lesssim \|u(t)\|_{L_x^{q_p}}^p + \|v(t)\|_{L_x^{q_p}}^p \lesssim \|u(t)\|_{\dot H_x^{s_p}}^p + \|v(t)\|_{\dot H^{s_p}}^p.$$

Putting this all together we arrive at the estimate

$$\begin{aligned}&\left|\int_{-\infty}^{-N^{1-\epsilon}}\int_{N^{1-\epsilon}}^{\infty}\left\langle S(t-\tau)\widehat{P}_{N,M}1_{\mathcal{C}_{\text{ext}}}(\tau)G(u,v)(\tau), 1_{\mathcal{C}_{\text{ext}}}(t)G(u,v)(t)\right\rangle \mathrm{d}t\,\mathrm{d}\tau\right| \\ &\quad\lesssim_L \int_{-\infty}^{-N^{1-\epsilon}}\int_{N^{1-\epsilon}}^{\infty}\frac{N^{L+1}}{M^{2L}}\frac{1}{|t-\tau|^{L-1}}(\|u(t)\|_{\dot H_x^{s_p}}^p + \|v(t)\|_{\dot H^{s_p}}^p)(\|u(\tau)\|_{\dot H_x^{s_p}}^p + \|v(\tau)\|_{\dot H^{s_p}}^p)\,\mathrm{d}t\,\mathrm{d}\tau \\ &\quad\lesssim_L N^{L+1+(1-\epsilon)(2-L)}M^{-2L}(\|u\|_{L_t^\infty \dot H^{s_p}}^{2p} + \|v\|_{L_t^\infty \dot H^{s_p}}^{2p}) \\ &\quad\lesssim_L M^{-L},\end{aligned}$$

where to obtain the last line we ensure that $\epsilon > 0$ is small enough so that when $M \geq N^{s_p/(1-\nu)}$ we also have $M^L \geq N^{4+\epsilon L}$. We have proved that

$$\text{(7-39)} \lesssim_L M^{-L},$$

as desired. This completes the treatment of the $C_{\text{ext}-\text{ext}}$ term.

The term $C_{\text{int}-\text{int}}$. Here we will use a combination of arguments based on sharp Huygens principle and the techniques developed to deal with the previous term $C_{\text{ext}-\text{ext}}$.

First we record an estimate for the kernel of the modified frequency projection.

Lemma 7.13. *Let $p_{N,M}^2$ denote the kernel of the operator $\widehat{P}_{N,M}^2$. Then,*

$$|p_{N,M}^2(x)| \lesssim_L \frac{N^3}{\langle N|x|\rangle^L} + \frac{N^3}{\langle M|x|\rangle^L}. \tag{7-52}$$

Next, consider the following decomposition of the forward cone centered at $(t,x) = (N^{1-\epsilon}, x(N^{1-\epsilon}))$ of width $R(\eta_0)$, i.e., the set

$$\mathcal{G}_+ := \bigcup_{t\geq N^{1-\epsilon}}\mathcal{G}_+(t),$$

where $\mathcal{G}(t)$ is defined as in (7-37). This decomposition is depicted in Figure 4.

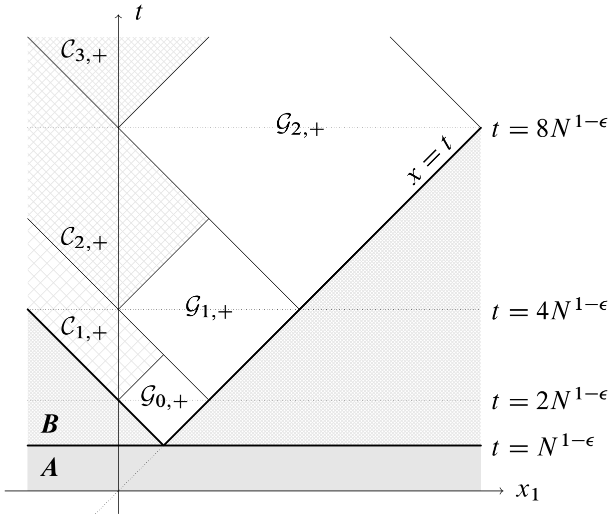

Figure 4. A depiction of the first few regions $\mathcal{C}_{j,+}$ and $\mathcal{G}_{+,j}$ within the region C.

We write

$$\mathcal{G}_+ = \bigcup_{j\ge 1} \mathcal{C}_{+,j} \cup \bigcup_{j\ge 0} \mathcal{G}_{+,j}.$$

We define $\mathcal{C}_{+,j}, G_{+,j}$ as follows. First, set

$$\widetilde{\mathcal{C}}_{+,1} := \{(t,x) : |x - x(2N^{1-\epsilon})| \ge R(\eta_0) + t - 2N^{1-\epsilon},\ t \ge 2N^{1-\epsilon}\} \cap \mathcal{G}_+$$

and for $j \ge 1$

$$\widetilde{\mathcal{C}}_{+,j} := \big\{\{(t,x) : |x - x(2^j N^{1-\epsilon})| \ge R(\eta_0) + t - 2^j N^{1-\epsilon},\ t \ge 2^j N^{1-\epsilon}\} \cap \mathcal{G}_+\big\} \setminus \mathcal{C}_{+,j-1}.$$

For $j \ge 0$, define sets $\widetilde{\mathcal{G}}_{+,j}$ to be the regions

$$\widetilde{\mathcal{G}}_{+,j} = \{(t,x) : |x - x(2^j N^{1-\epsilon})| \le R(\eta_0) + t - 2^j N^{1-\epsilon},\ 2^j N^{1-\epsilon} \le t \le 2^{j+1} N^{1-\epsilon}\} \cap \mathcal{G}_+.$$

Then we define

$$\mathcal{C}_{+,j} := \widetilde{\mathcal{C}}_{+,j} \cap \{(t,x) : |x| \le t - 2^j N^{1-\epsilon},\ t \ge 2^j N^{1-\epsilon}\},$$
$$\mathcal{G}_{+,j} := \widetilde{\mathcal{G}}_{+,j} \cup [\widetilde{\mathcal{C}}_{+,j+1} \setminus \mathcal{C}_{+,j+1}].$$

The regions $\mathcal{C}_{+,j}$ and $\mathcal{G}_{+,j}$ are depicted in Figure 4.

Now, split the integrand of (7-42) in the four pieces,

$$(7\text{-}42) = \int_{-\infty}^{-N^{1-\epsilon}} \int_{N^{1-\epsilon}}^{\infty} (I + II + III + IV)\, dt\, d\tau,$$

where

$$I = \sum_{j,k} \Big\langle \widehat{P}_{N,M}^2 \frac{1}{|\nabla|} S(t-\tau)[1_{\mathcal{C}_{-,j}} 1_{\mathcal{C}_{\rm int}} G](\tau), [1_{\mathcal{C}_{+,k}} 1_{\mathcal{C}_{\rm int}} G](t) \Big\rangle, \tag{7-53}$$

$$II = \sum_{j,k} \Big\langle \widehat{P}_{N,M}^2 \frac{1}{|\nabla|} S(t-\tau)[1_{\mathcal{C}_{-,j}} 1_{\mathcal{C}_{\rm int}} G](\tau), [1_{\mathcal{G}_{+,k}} 1_{\mathcal{C}_{\rm int}} G](t) \Big\rangle, \tag{7-54}$$

$$III = \sum_{j,k} \Big\langle \widehat{P}_{N,M}^2 \frac{1}{|\nabla|} S(t-\tau)[1_{\mathcal{G}_{-,j}} 1_{\mathcal{C}_{\rm int}} G](\tau), [1_{\mathcal{C}_{+,k}} 1_{\mathcal{C}_{\rm int}} G](t) \Big\rangle, \tag{7-55}$$

$$IV = \sum_{j,k} \Big\langle \widehat{P}_{N,M}^2 \frac{1}{|\nabla|} S(t-\tau)[1_{\mathcal{G}_{-,j}} 1_{\mathcal{C}_{\rm int}} G](\tau), [1_{\mathcal{G}_{+,k}} 1_{\mathcal{C}_{\rm int}} G](t) \Big\rangle. \tag{7-56}$$

First we estimate the term (7-53) above. The key points are the following. First, by the support properties of $1_{\mathcal{C}_{+,k}} 1_{\mathcal{C}_{\rm int}}(\tau, y)$, $1_{\mathcal{C}_{-,j}} 1_{\mathcal{C}_{\rm int}}(t, y)$ and the sharp Huygens principle, we must have

$$|x-y| \gtrsim (2^j+2^k)N^{1-\epsilon} \quad \text{for all } x \in \operatorname{supp}(1_{\mathcal{C}_{+,k}} 1_{\mathcal{C}_{\rm int}} G(u,v))(t),\ \ y \in \operatorname{supp}[S(t-\tau) 1_{\mathcal{C}_{-,j}} 1_{\mathcal{C}_{\rm int}} G(u,v)](\tau). \tag{7-57}$$

Second, by the definitions of the space-time cutoffs $1_{\mathcal{C}_{-,j}}$ and $1_{\mathcal{C}_{+,k}}$, the functions $1_{\mathcal{C}_{-,j}} u(\tau)$ and $1_{\mathcal{C}_{+,k}} u(t)$ are restricted to the exterior *small-data* regime and we thus have

$$\begin{aligned} \|1_{\mathcal{C}_{-,j}} 1_{\mathcal{C}_{\rm int}} G(u,v)\|_{\mathcal{N}((-\infty,-2^j N^{1-\epsilon}])} &\lesssim \|\vec{u}\|_{L_t^\infty \dot{\mathcal{H}}^{s_p}} \lesssim 1, \\ \|1_{\mathcal{C}_{+,k}} 1_{\mathcal{C}_{\rm int}} G(u,v)\|_{\mathcal{N}([2^j N^{1-\epsilon},\infty))} &\lesssim \|\vec{u}\|_{L_t^\infty \dot{\mathcal{H}}^{s_p}} \lesssim 1, \end{aligned} \tag{7-58}$$

where $\mathcal{N}$ denote suitable dual spaces.

We argue as follows. For any $q \ge 2$, and up to fattening the projection $\widehat{P}_{N,M}$, we have

$$\begin{aligned} &\Big|\Big\langle \widehat{P}_{N,M}^2 \frac{1}{|\nabla|} S(t-\tau)[1_{\mathcal{C}_{-,j}} 1_{\mathcal{C}_{\rm int}} G(u,v)](\tau), [1_{\mathcal{C}_{+,k}} 1_{\mathcal{C}_{\rm int}} G(u,v)](t) \Big\rangle\Big| \\ &\quad \lesssim \|1_{\{|\cdot| \gtrsim (2^j+2^k)N^{1-\epsilon}\}} p_{N,M}\|_{L^1} \|P_N |\nabla|^{-1-s_p+\frac{2}{q}} S(t-\tau)[1_{\mathcal{C}_{-,j}} 1_{\mathcal{C}_{\rm int}} G(u,v)](\tau)\|_{L_x^q} \\ &\qquad \times \| |\nabla|^{s_p-\frac{2}{q}} [1_{\mathcal{C}_{+,k}} 1_{\mathcal{C}_{\rm int}} G(u,v)](t)\|_{L^{q'}}. \end{aligned}$$

We estimate last line above as follows. Note that by (7-52) and (7-57) (and the lower bound on M), we have

$$\|1_{\{|\cdot| \gtrsim (2^j+2^k)N^{1-\epsilon}\}} p_{N,M}\|_{L^1} \lesssim_L \frac{N^3}{[(2^j+2^k)N^{1-\epsilon}]^{L-1}}.$$

By the dispersive estimate for the wave equation and noting that $|t-\tau| \ge 2N^{1-\epsilon}$ we have

$$\begin{aligned} &\|P_N |\nabla|^{-1-s_p+\frac{2}{q}} S(t-\tau)[1_{\mathcal{C}_{-,j}} 1_{\mathcal{C}_{\rm int}} G(u,v)](\tau)\|_{L_x^q} \\ &\qquad \lesssim \frac{1}{|t-\tau|^{1-\frac{2}{q}}} N^{1-\frac{4}{q}} \|P_N |\nabla|^{-1-s_p+\frac{2}{q}} [1_{\mathcal{C}_{-,j}} 1_{\mathcal{C}_{\rm int}} G(u,v)](\tau)\|_{L_x^{q'}} \\ &\qquad \lesssim \frac{1}{|t-\tau|^{1-\frac{2}{q}}} N^{-2s_p} \|P_N |\nabla|^{s_p-\frac{2}{q}} [1_{\mathcal{C}_{-,j}} 1_{\mathcal{C}_{\rm int}} G(u,v)](\tau)\|_{L_x^{q'}}. \end{aligned}$$

Thus, using the above, Bernstein's inequality, the Hardy–Littlewood–Sobolev inequality, and (7-58) in the last line below we have

$$\begin{aligned}
&\int_{N^{1-\epsilon}}^{\infty}\int_{-\infty}^{-N^{1-\epsilon}} \text{(7-53)}\,\mathrm{d}t\,\mathrm{d}\tau \\
&\lesssim_L \sum_{j,k\geq 1}\frac{N^3N^{-2s_p}}{[(2^j+2^k)N^{1-\epsilon}]^{L-1}}\int_{N^{1-\epsilon}}^{\infty}\int_{-\infty}^{-N^{1-\epsilon}}\biggl(\frac{1}{|t-\tau|^{1-\frac{2}{q}}}\||\nabla|^{s_p-\frac{2}{q}}[1_{\mathcal{C}_{-,j}}1_{\mathcal{C}_{\mathrm{int}}}G(u,v)](\tau)\|_{L_x^{q'}} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \times\||\nabla|^{s_p-\frac{2}{q}}[1_{\mathcal{C}_{-,k}}1_{\mathcal{C}_{\mathrm{int}}}G(u,v)](t)\|_{L_x^{q'}}\biggr)\,\mathrm{d}t\,\mathrm{d}\tau \\
&\lesssim_L \sum_{j,k\geq 1}\frac{N^3N^{-2s_p}}{[(2^j+2^k)N^{1-\epsilon}]^{L-1}}\||\nabla|^{s_p-\frac{2}{q}}[1_{\mathcal{C}_{+,k}}1_{\mathcal{C}_{\mathrm{int}}}G(u,v)]\|_{L_t^{2q/(q+2)}L_x^{q'}} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \times\||\nabla|^{s_p-\frac{2}{q}}[1_{\mathcal{C}_{-,j}}1_{\mathcal{C}_{\mathrm{int}}}G(u,v)]\|_{L_t^{2q/(q+2)}L_x^{q'}} \\
&\lesssim_L \sum_{j,k\geq 1}\frac{N^3N^{-2s_p}}{[(2^j+2^k)N^{1-\epsilon}]^{L-1}}\lesssim_L N^{-L/2},
\end{aligned}$$

where in the second-to-last line we have fixed $q>2$ above and note that the norms above are dual sharp admissible Strichartz pairs (e.g., one can take $q=4$).

Next, consider the term (7-56). Here we cannot rely exclusively on separation of supports because the $S(t-\tau)$ evolution of the term localized to $\mathcal{G}_{-,j}$ has some of its support within $2R(\eta_0)$ of the term localized to $\mathcal{G}_{+,k}$ for all j,k. The saving grace is that the pieces of the supports of $S(t-\tau)[1_{\mathcal{G}_{-,j}}1_{\mathcal{C}_{\mathrm{int}}}G(u,v)](\tau)$ and $[1_{\mathcal{G}_{+,k}}1_{\mathcal{C}_{\mathrm{int}}}G(u,v)](t)$ that are close to each other (say within $2^{\alpha j}+2^{\alpha k}$ for some small parameter $\alpha>0$) come along with *angular separation* in the sense of Lemma 7.12. To make this precise we must further subdivide $G_{\pm,k}$ as follows.

Let $\alpha>0$ be a small parameter to be fixed below. Let

$$\begin{aligned}
\mathcal{G}_{+,k,\mathrm{in}} &:= \mathcal{G}_{+,k}\cap\{(t,x):|x|\leq t-2^{\alpha k}N^{\alpha(1-\epsilon)}\},\\
\mathcal{G}_{+,k,\mathrm{out}} &:= \mathcal{G}_{+,k}\cap\{(t,x):|x|\geq t-2^{\alpha k}N^{\alpha(1-\epsilon)}\}.
\end{aligned}$$

We decompose (7-56) as follows, noting symmetry in j,k means it suffices to consider only the sum for $j\geq k$. We write (7-56) in the form

$$\sum\biggl\langle \widehat{P}_{N,M}^2\frac{1}{|\nabla|}S(t-\tau)[1_{\mathcal{G}_{-,j,\mathrm{in}}}1_{\mathcal{C}_{\mathrm{int}}}G],\ [1_{\mathcal{G}_{+,k,\mathrm{in}}}1_{\mathcal{C}_{\mathrm{int}}}G]\biggr\rangle \tag{7-59}$$

$$+\sum\biggl\langle \widehat{P}_{N,M}^2\frac{1}{|\nabla|}S(t-\tau)[1_{\mathcal{G}_{-,j,\mathrm{in}}}1_{\mathcal{C}_{\mathrm{int}}}G],\ [1_{\mathcal{G}_{+,k,\mathrm{out}}}1_{\mathcal{C}_{\mathrm{int}}}G]\biggr\rangle \tag{7-60}$$

$$+\sum\biggl\langle \widehat{P}_{N,M}^2\frac{1}{|\nabla|}S(t-\tau)[1_{\mathcal{G}_{-,j,\mathrm{out}}}1_{\mathcal{C}_{\mathrm{int}}}G],\ [1_{\mathcal{G}_{+,k,\mathrm{in}}}1_{\mathcal{C}_{\mathrm{int}}}G]\biggr\rangle \tag{7-61}$$

$$+\sum\biggl\langle \widehat{P}_{N,M}^2\frac{1}{|\nabla|}S(t-\tau)[1_{\mathcal{G}_{-,j,\mathrm{out}}}1_{\mathcal{C}_{\mathrm{int}}}G],\ [1_{\mathcal{G}_{+,k,\mathrm{out}}}1_{\mathcal{C}_{\mathrm{int}}}G]\biggr\rangle, \tag{7-62}$$

where the sums are over $j,k\geq 0$ with $j\geq k$, $G=G(u,v)$, and the pairings are evaluated at τ,t.

The key point will be that on the outer regions $\mathcal{G}_{-,j,\mathrm{out}}$, $\mathcal{G}_{+,k,\mathrm{out}}$ we can recover the same angular separation used to treat the term $C_{\mathrm{ext-ext}}$ and on the inner regions $\mathcal{G}_{-,j,\mathrm{in}}$ and $\mathcal{G}_{+,k,\mathrm{in}}$ we obtain sufficient separation in support between the two factors after evolution by $S(t-\tau)$ to get enough decay in j,k after the application of $\widehat{P}^2_{N,M}\frac{1}{|\nabla|}$.

Lemma 7.14 (angular separation in $\mathcal{G}_{\pm,j,\mathrm{out}}$). *Let $\alpha>0$ and let $S_{N,\alpha}$ be the set*

$$\mathcal{S}_{N,\alpha}:=\left\{x\in\mathbb{R}^3:\frac{|x_{2,3}|}{|x|}\lesssim\frac{1}{N^{\frac{(1-\alpha)(1-\epsilon)}{2}}}\right\}. \tag{7-63}$$

Then, there exists $\alpha>0$ small enough and $N_0>0$ large enough so that for all $x\in\mathcal{G}_{\pm,j,\mathrm{out}}$ we have

$$x\in S_{N,\alpha}\quad\textit{and}\quad\frac{1}{N^{\frac{(1-\alpha)(1-\epsilon)}{2}}}\ll\frac{M}{N}$$

for all $N\geq N_0$ and $M\geq N^{s_p/(1-\nu)}$ and for all $j\geq 0$.

Proof. It suffices to consider $x\in\mathcal{G}_{+,j,\mathrm{out}}$. The proof is nearly identical to the proof of Lemma 7.11, but here we have allowed the region $\mathcal{G}_{+,j,\mathrm{out}}$ to deviate farther from the boundary of the cone as j (and hence t) gets larger. As in Lemma 7.12 we have

$$|\sin(\theta_{x(2^jN^{1-\epsilon})})|\simeq|\theta_{x(2^jN^{1-\epsilon})}|\leq A_1N^{\frac{\epsilon}{2}-\frac{1}{2}}$$

independently of $j\geq 0$. To finish the proof it suffices to show that for any $x\in\mathcal{G}_{+,j,\mathrm{out}}$, the angle $\theta_{(x,x(2^jN^{1-\epsilon}))}$ formed between the vectors x and $x(2^jN^{1-\epsilon})$ satisfies

$$|\theta_{(x,x(2^jN^{1-\epsilon}))}|\leq A_2\frac{1}{N^{\frac{(1-\alpha)(1-\epsilon)}{2}}}$$

for some other uniform constant $A_2>0$, as then the sine of the total angle between x and the x_1-axis, i.e., $|x_{2,3}|/|x|$ would satisfy (7-63). Note that for any $(t,x)\in\mathcal{G}_{+,j,\mathrm{out}}$

$$2^jN^{1-\epsilon}-2^{\alpha j}N^{\alpha(1-\epsilon)}\leq|x|\leq 2^{j+1}N^{1-\epsilon}+2^{\alpha j}N^{\alpha(1-\epsilon)}.$$

Arguing as in the proof of Lemma 7.11 we see that for any $(t,x)\in\mathcal{G}_{+,j,\mathrm{out}}$

$$\theta^2_{(x,x(2^jN^{1-\epsilon}))}\lesssim\frac{2t2^{\alpha j}N^{\alpha(1-\epsilon)}}{(t-2^{\alpha j}N^{\alpha(1-\epsilon)})(2^jN^{1-\epsilon})}\lesssim\frac{1}{2^{(1-\alpha)j}N^{(1-\alpha)(1-\epsilon)}},$$

as desired. □

With Lemma 7.14 in hand, we can estimate the term (7-62) in an identical fashion as the term (7-39), noting that applications of Lemma 7.12 are still valid in this new setting because for $x\in G_{+,k,\mathrm{out}}$ and $y\in G_{-,j,\mathrm{out}}$ we have

$$\frac{|x_{2,3}-y_{2,3}|}{|x-y|}\lesssim\frac{|x_{2,3}|}{|x|}+\frac{|y_{2,3}|}{|y|}\lesssim\frac{1}{N^{(1-\alpha)(1-\epsilon)}}\ll\frac{M}{N},$$

i.e., sufficient angular separation since the Fourier variable ξ satisfies

$$\frac{|\xi_{2,3}|}{|\xi|}\simeq\frac{M}{N}.$$

Moreover we have

$$|x-y| \simeq (2^j+2^k)N^{1-\epsilon} \quad \text{if } x \in G_{+,k,\text{out}},\ y \in G_{-,j,\text{out}}.$$

This means that we are free to write,

$$\begin{aligned}&\Big\langle \widehat{P}^2_{N,M}\frac{1}{|\nabla|}S(t-\tau)[1_{\mathcal{G}_{-,j,\text{out}}}1_{\mathcal{C}_{\text{int}}}G(u,v)](\tau),\ [1_{\mathcal{G}_{+,k,\text{out}}}1_{\mathcal{C}_{\text{int}}}G(u,v)](t)\Big\rangle \\ &\qquad = \langle [1_{S_{N,\alpha}}1_{\{|\cdot|\simeq(2^j+2^k)N^{1-\epsilon}\}}K_{N,M}] * [1_{\mathcal{G}_{-,j,\text{out}}}1_{\mathcal{C}_{\text{int}}}G(u,v)](\tau), [1_{\mathcal{G}_{+,k,\text{out}}}1_{\mathcal{C}_{\text{int}}}G(u,v)](t)\rangle. \quad (7\text{-}64)\end{aligned}$$

Mimicking the estimates of (7-62) we see that as in (7-51) we have

$$\|1_{S_{N,\alpha}}(\cdot)1_{\{|\cdot|\simeq(2^j+2^k)N^{1-\epsilon}\}}(\cdot)K_{N,M}(t-\tau)\|_{L_x^{(qp/p)'/2}} \lesssim_L \frac{N^{L+1}}{M^{2L}}\frac{1}{[(2^j+2^k)N^{1-\epsilon}]^L}.$$

This allows us to sum in j,k, and we obtain

$$\int_{-\infty}^{-N^{1-\epsilon}}\int_{N^{1-\epsilon}}^{\infty} (7\text{-}62)\,dt\,d\tau \lesssim_L \frac{1}{M^L}.$$

To handle the term (7-59) we rely on the following observation: by the support properties of $1_{\mathcal{C}_{+,k}}1_{\mathcal{C}_{\text{int}}}(\tau,y)$, $1_{\mathcal{C}_{-,j}}1_{\mathcal{C}_{\text{int}}}(t,y)$ and the sharp Huygens principle, we must have

$$|x-y| \gtrsim (2^j+2^k)N^{1-\epsilon}$$

for all

$$x \in \operatorname{supp}(1_{\mathcal{G}_{+,k,\text{in}}}1_{\mathcal{C}_{\text{int}}}G(u,v))(t)$$

and

$$y \in \operatorname{supp} S(t-\tau)[1_{\mathcal{G}_{-,j,\text{in}}}1_{\mathcal{C}_{\text{int}}}G(u,v)](\tau).$$

Hence,

$$\begin{aligned}&\Big|\Big\langle \widehat{P}^2_{N,M}\frac{1}{|\nabla|}S(t-\tau)[1_{\mathcal{G}_{-,j,\text{in}}}1_{\mathcal{C}_{\text{int}}}G(u,v)](\tau),\ [1_{\mathcal{G}_{+,k,\text{in}}}1_{\mathcal{C}_{\text{int}}}G(u,v)](t)\Big\rangle\Big| \\ &\qquad \lesssim \|1_{\{|\cdot|\gtrsim(2^j+2^k)N^{1-\epsilon}\}}p_{N,M}\|_{L_x^{(qp/p)'/2}}N^{-1}\|P_N S(t-\tau)[1_{\mathcal{C}_{-,j}}1_{\mathcal{C}_{\text{int}}}G(u,v)](\tau)\|_{L_x^{qp/p}}\| \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \times[1_{\mathcal{C}_{+,k}}1_{\mathcal{C}_{\text{int}}}G(u,v)](t)\|_{L_x^{qp/p}} \\ &\qquad \lesssim_L \frac{1}{[(2^j+2^k)N^{1-\epsilon}]^L}(\|u\|^{2p}_{L_t^\infty \dot{H}^{s_p}} + \|v\|^{2p}_{L_t^\infty \dot{H}^{s_p}}).\end{aligned}$$

Hence,

$$\int_{-\infty}^{-N^{1-\epsilon}}\int_{N^{1-\epsilon}}^{\infty} (7\text{-}59)\,dt\,d\tau \lesssim_L \frac{1}{N^L}.$$

Next, for the term (7-60) we note that the same argument used to treat (7-59) applies. However, we note that here we only obtain spatial separation of $2^jN^{1-\epsilon}$. Nonetheless, since $j \geq k$ we have

$$2^jN^{1-\epsilon} \simeq (2^j+2^k)N^{1-\epsilon}$$

and hence we are able to sum in j, k, obtaining

$$\int_{-\infty}^{-N^{1-\epsilon}} \int_{N^{1-\epsilon}}^{\infty} \text{(7-60)}\, \mathrm{d}t\, \mathrm{d}\tau \lesssim_L \frac{1}{M^L}.$$

Lastly, consider the term (7-61). Here we use a mix of the arguments used to control (7-59) and (7-62). In particular we split the sum into two pieces noting that if $j \simeq k$ then the same argument used to estimate (7-59) applies since the spatial supports are separated by $\simeq 2^k N^{1-\epsilon} \simeq (2^j + 2^k) N^{1-\epsilon}$. If $j \gg k$, we obtain enough angular separation argument to use the same argument used to bound (7-62), since in this case we have

$$\frac{|x_{2,3} - y_{2,3}|}{|x - y|} \simeq \frac{|y_{2,3}|}{|y|} \lesssim \frac{1}{N^{(1-\alpha)(1-\epsilon)}} \ll \frac{M}{N}$$

for all $x \in \mathcal{G}_{+,k,\mathrm{in}}$ and $y \in \mathcal{G}_{-,j,\mathrm{out}}$ as long as $j \gg k$. We obtain

$$\int_{-\infty}^{-N^{1-\epsilon}} \int_{N^{1-\epsilon}}^{\infty} \text{(7-59)}\, \mathrm{d}t\, \mathrm{d}\tau \lesssim_L \frac{1}{N^L} + \frac{1}{M^L}.$$

This completes the estimation of (7-56).

At this point, the mixed terms (7-54) and (7-55) (i.e., the remaining contributions to the $C_{\mathrm{int-int}}$ term), as well as the $C_{\mathrm{int-ext}}$ and $C_{\mathrm{ext-int}}$ terms ((7-40) and (7-41)) can be handled with a combination of the techniques developed above. For example, after further subdividing $\mathcal{G}_-$ in the regions $\mathcal{C}_{-,j}$ and $\mathcal{G}_{-,j}$ consider the term of the form,

$$\sum_{j\geq 0} \int_{-2^{j+2}N^{1-\epsilon}}^{-2^j N^{1-\epsilon}} \int_{N^{1-\epsilon}}^{\infty} \Big\langle \widehat{P}^2_{N,M} \frac{1}{|\nabla|} S(t-\tau)[1_{\mathcal{G}_{-,j}} 1_{\mathcal{C}_{\mathrm{int}}} G(u,v)](\tau), 1_{\mathcal{C}_{\mathrm{ext}}} G(u,v)](t) \Big\rangle \,\mathrm{d}t\, \mathrm{d}\tau.$$

Fixing a large constant $K_1 > 0$, we can divide the above into two further pieces, namely

$$\sum_{j\geq 0} \int_{-2^{j+2}N^{1-\epsilon}}^{-2^j N^{1-\epsilon}} \int_{N^{1-\epsilon}}^{K_1 2^j N^{1-\epsilon}} \Big\langle \widehat{P}^2_{N,M} \frac{1}{|\nabla|} S(t-\tau)[1_{\mathcal{G}_{-,j}} 1_{\mathcal{C}_{\mathrm{int}}} G(u,v)](\tau),\ 1_{\mathcal{C}_{\mathrm{ext}}} G(u,v)](t) \Big\rangle \,\mathrm{d}t\, \mathrm{d}\tau$$
$$+ \sum_{j\geq 0} \int_{-2^{j+2}N^{1-\epsilon}}^{-2^j N^{1-\epsilon}} \int_{K_1 2^j N^{1-\epsilon}}^{\infty} \Big\langle \widehat{P}^2_{N,M} \frac{1}{|\nabla|} S(t-\tau)[1_{\mathcal{G}_{-,j}} 1_{\mathcal{C}_{\mathrm{int}}} G(u,v)](\tau),\ 1_{\mathcal{C}_{\mathrm{ext}}} G(u,v)](t) \Big\rangle \,\mathrm{d}t\, \mathrm{d}\tau.$$

For the first term on the right-hand-side above we can copy the argument used to estimate (7-59). Indeed by the sharp Huygens principle the spatial supports (before the application of $P_{N,M}$) are separated for each fixed t, τ by a distance of at least $\simeq 2^j N^{1-\epsilon} \simeq_{K_1} |t-\tau|$. For the second term above we can choose $K_1 \gg 1$ large enough to guarantee enough angular separation between the spatial and Fourier variables to mimic a combination of the arguments used to estimate (7-39) (where one integrates in t) and (7-62) (where one sums in j). The remaining interactions are handled similarly. We omit the details.

We have thus proved that

$$|\langle C, C' \rangle| \lesssim_L \frac{1}{N^L} + \frac{1}{M^L} \lesssim_L \frac{1}{M^L},$$

which finally completes the proof of Lemma 7.10. ☐

We are now prepared to conclude the frequency envelope argument and the proof of Proposition 7.4.

Proof of Proposition 7.4. Recall that we are trying to prove that

$$\sum_{N \geq N_0} \sum_{C_0 N^{s_p/(1-\nu)} \leq M \leq N} M^{2(1-\nu)} \| \widehat{P}_{N, \geq M} u(t) \|_{L_x^2}^2 \lesssim 1,$$

for some fixed $C_0 > 0$, for which it suffices to prove that

$$\sum_{N \geq N_0} \sum_{C_0 N^{s_p/(1-\nu)} \leq M \leq N} M^{2(1-\nu)} N^{-2s_p} \| \widehat{P}_{N, \geq M} u(t) \|_{\dot{H}_x^{s_p}}^2 \lesssim 1.$$

Once again, by time-translation invariance, we argue for $t = 0$. Recall that

$$|\langle \widehat{P}_{N,\geq M} u(0),\ \widehat{P}_{N,\geq M} u(0) \rangle_{\dot{H}_x^{s_p}}| \lesssim \|A\|_{\dot{H}^{s_p}}^2 + \|A'\|_{\dot{H}^{s_p}}^2 + \|B\|_{\dot{H}^{s_p}}^2 + \|B'\|_{\dot{H}^{s_p}}^2 + |\langle C, C' \rangle_{\dot{H}_x^{s_p}}|,$$

and hence by Lemmas 7.8, 7.9 and 7.10, we obtain

$$\begin{aligned} \gamma_{N,M}(0) &= \sum_{N', M' \geq M} \min\left\{ \frac{N}{N'}, \frac{N'}{N} \right\}^{\sigma} \left(\frac{M'}{M} \right)^{\sigma} \| \widehat{P}_{N, \geq M} u(0) \|_{\dot{H}_x^{s_p}}^2 \\ &\lesssim \eta_0^{p-1} \alpha_{N,M} + \eta_0^{p-1} \beta_{N,M} + M^{-L}. \end{aligned} \tag{7-65}$$

Furthermore by (7-29) and (7-34),

$$\alpha_{N,M} \lesssim \gamma_{N,M}(N^{1-\epsilon}) + \eta_0^{p-1} \alpha_{N,M},$$

and

$$\gamma_{N,M}(N^{1-\epsilon}) + \beta_{N,M} \lesssim \gamma_{N,M}(0) + \eta_0^{p-1} \beta_{N,M}.$$

Hence

$$\beta_{N,M} \lesssim \gamma_{N,M}(0), \quad \alpha_{N,M} \lesssim \gamma_{N,M}(N^{1-\epsilon}) \lesssim \gamma_{N,M}(0),$$

and we conclude from (7-65) that

$$\gamma_{N,M}(0) \lesssim \eta_0^{p-1} \gamma_{N,M}(0) + M^{-L},$$

which implies

$$\gamma_{N,M}(0) \lesssim M^{-L}$$

for any $L \gg 1$. Consequently, we have established that

$$\sum_{N \geq N_0} \sum_{M \geq C_0 N^{s_p/(1-\nu)}} M^{2(1-\nu)} N^{-2s_p} \gamma_{N,M}(0)^2 \lesssim 1,$$

which concludes the proof. □

Acknowledgments

Dodson gratefully acknowledges support from NSF grant DMS-1500424 and NSF grant DMS-1764358. Lawrie gratefully acknowledges support from NSF grant DMS-1700127. Mendelson gratefully acknowledges support from NSF grant DMS-1800697. Murphy gratefully acknowledges support from NSF grant DMS-1400706. The authors thank the MSRI program "New challenges in PDE: deterministic Dynamics and randomness in high and infinite dimensional systems", where this work began. Dodson and Lawrie also thank the IHES program "Trimester on nonlinear waves", where part of this work was completed. Dodson and Mendelson gratefully acknowledge support from the Institute for Advanced Study, where part of this work was completed.

References

[Bahouri and Chemin 2006] H. Bahouri and J.-Y. Chemin, "On global well-posedness for defocusing cubic wave equation", *Int. Math. Res. Not.* **2006** (2006), art. id. 54873, 12 pp. 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

[Bahouri and Shatah 1998] H. Bahouri and J. Shatah, "Decay estimates for the critical semilinear wave equation", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **15**:6 (1998), 783–789. MR Zbl

[Bizoń et al. 2010] P. Bizoń, P. Breitenlohner, D. Maison, and A. Wasserman, "Self-similar solutions of the cubic wave equation", *Nonlinearity* **23**:2 (2010), 225–236. MR Zbl

[Bulut 2012a] A. Bulut, "The defocusing cubic nonlinear wave equation in the energy-supercritical regime", pp. 1–11 in *Recent advances in harmonic analysis and partial differential equations*, edited by A. R. Nahmod et al., Contemp. Math. **581**, Amer. Math. Soc., Providence, RI, 2012. MR Zbl

[Bulut 2012b] A. Bulut, "Global well-posedness and scattering for the defocusing energy-supercritical cubic nonlinear wave equation", *J. Funct. Anal.* **263**:6 (2012), 1609–1660. MR Zbl

[Côte et al. 2015] R. Côte, C. E. Kenig, A. Lawrie, and W. Schlag, "Characterization of large energy solutions of the equivariant wave map problem, I", *Amer. J. Math.* **137**:1 (2015), 139–207. MR Zbl

[Dodson 2012] B. Dodson, "Global well-posedness and scattering for the defocusing, L^2-critical nonlinear Schrödinger equation when $d \geq 3$", *J. Amer. Math. Soc.* **25**:2 (2012), 429–463. MR Zbl

[Dodson 2016] B. Dodson, "Global well-posedness and scattering for the defocusing, L^2-critical, nonlinear Schrödinger equation when $d = 2$", *Duke Math. J.* **165**:18 (2016), 3435–3516. MR Zbl

[Dodson 2018a] B. Dodson, "Global well-posedness and scattering for the radial, defocusing, cubic nonlinear wave equation", preprint, 2018. arXiv

[Dodson 2018b] B. Dodson, "Global well-posedness and scattering for the radial, defocusing, cubic wave equation with almost sharp initial data", *Comm. Partial Differential Equations* **43**:10 (2018), 1413–1455. MR Zbl

[Dodson 2019] B. Dodson, "Global well-posedness and scattering for the radial, defocusing, cubic wave equation with initial data in a critical Besov space", *Anal. PDE* **12**:4 (2019), 1023–1048. MR Zbl

[Dodson and Lawrie 2015a] B. Dodson and A. Lawrie, "Scattering for radial, semi-linear, super-critical wave equations with bounded critical norm", *Arch. Ration. Mech. Anal.* **218**:3 (2015), 1459–1529. MR Zbl

[Dodson and Lawrie 2015b] B. Dodson and A. Lawrie, "Scattering for the radial 3D cubic wave equation", *Anal. PDE* **8**:2 (2015), 467–497. 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 2017] R. Donninger and B. Schörkhuber, "Stable blowup for wave equations in odd space dimensions", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **34**:5 (2017), 1181–1213. MR Zbl

[Duyckaerts and Roy 2017] T. Duyckaerts and T. Roy, "Blow-up of the critical Sobolev norm for nonscattering radial solutions of supercritical wave equations on $\mathbb{R}^3$", *Bull. Soc. Math. France* **145**:3 (2017), 503–573. MR Zbl

[Duyckaerts et al. 2014] T. Duyckaerts, C. Kenig, and F. Merle, "Scattering for radial, bounded solutions of focusing supercritical wave equations", *Int. Math. Res. Not.* **2014**:1 (2014), 224–258. MR Zbl

[Duyckaerts et al. 2016] T. Duyckaerts, C. Kenig, and F. Merle, "Solutions of the focusing nonradial critical wave equation with the compactness property", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **15** (2016), 731–808. MR Zbl

[Gallagher and Planchon 2003] I. Gallagher and F. Planchon, "On global solutions to a defocusing semi-linear wave equation", *Rev. Mat. Iberoamericana* **19**:1 (2003), 161–177. MR Zbl

[Jörgens 1961] K. Jörgens, "Das Anfangswertproblem im Grossen für eine Klasse nichtlinearer Wellengleichungen", *Math. Z.* **77** (1961), 295–308. MR Zbl

[Keel and Tao 1998] M. Keel and T. Tao, "Endpoint Strichartz estimates", *Amer. J. Math.* **120**:5 (1998), 955–980. MR Zbl

[Kenig and Merle 2006] C. E. Kenig and F. Merle, "Global well-posedness, scattering and blow-up for the energy-critical, focusing, non-linear Schrödinger equation in the radial case", *Invent. Math.* **166**:3 (2006), 645–675. MR Zbl

[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

[Kenig and Merle 2010] C. E. Kenig and F. Merle, "Scattering for $\dot{H}^{1/2}$ bounded solutions to the cubic, defocusing NLS in 3 dimensions", *Trans. Amer. Math. Soc.* **362**:4 (2010), 1937–1962. MR Zbl

[Kenig and Merle 2011a] C. E. Kenig and F. Merle, "Nondispersive radial solutions to energy supercritical non-linear wave equations, with applications", *Amer. J. Math.* **133**:4 (2011), 1029–1065. MR Zbl

[Kenig and Merle 2011b] C. E. Kenig and F. Merle, "Radial solutions to energy supercritical wave equations in odd dimensions", *Discrete Contin. Dyn. Syst.* **31**:4 (2011), 1365–1381. MR Zbl

[Kenig et al. 2000] C. E. Kenig, G. Ponce, and L. Vega, "Global well-posedness for semi-linear wave equations", *Comm. Partial Differential Equations* **25**:9-10 (2000), 1741–1752. MR Zbl

[Killip and Visan 2013] R. Killip and M. Vişan, "Nonlinear Schrödinger equations at critical regularity", pp. 325–437 in *Evolution equations*, edited by D. Ellwood et al., Clay Math. Proc. **17**, Amer. Math. Soc., Providence, RI, 2013. MR Zbl

[Killip and Visan 2010a] R. Killip and M. Visan, "Energy-supercritical NLS: critical $\dot{H}^s$-bounds imply scattering", *Comm. Partial Differential Equations* **35**:6 (2010), 945–987. MR Zbl

[Killip and Visan 2010b] 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

[Killip and Visan 2011a] R. Killip and M. Visan, "The defocusing energy-supercritical nonlinear wave equation in three space dimensions", *Trans. Amer. Math. Soc.* **363**:7 (2011), 3893–3934. MR Zbl

[Killip and Visan 2011b] R. Killip and M. Visan, "The radial defocusing energy-supercritical nonlinear wave equation in all space dimensions", *Proc. Amer. Math. Soc.* **139**:5 (2011), 1805–1817. MR Zbl

[Killip et al. 2009] R. Killip, T. Tao, and M. Visan, "The cubic nonlinear Schrödinger equation in two dimensions with radial data", *J. Eur. Math. Soc. (JEMS)* **11**:6 (2009), 1203–1258. 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

[Lindblad and Sogge 1995] H. Lindblad and C. D. Sogge, "On existence and scattering with minimal regularity for semilinear wave equations", *J. Funct. Anal.* **130**:2 (1995), 357–426. MR Zbl

[Merle and Zaag 2003] F. Merle and H. Zaag, "Determination of the blow-up rate for the semilinear wave equation", *Amer. J. Math.* **125**:5 (2003), 1147–1164. MR Zbl

[Merle and Zaag 2005] F. Merle and H. Zaag, "Determination of the blow-up rate for a critical semilinear wave equation", *Math. Ann.* **331**:2 (2005), 395–416. MR Zbl

[Nakanishi and Schlag 2011] K. Nakanishi and W. Schlag, *Invariant manifolds and dispersive Hamiltonian evolution equations*, European Mathematical Society (EMS), Zürich, 2011. MR Zbl

[Ramos 2012] J. Ramos, "A refinement of the Strichartz inequality for the wave equation with applications", *Adv. Math.* **230**:2 (2012), 649–698. MR Zbl

[Ramos 2018] J. Ramos, "Nonlinear profile decomposition for the $\dot{H}^{1/2} \times \dot{H}^{-1/2}(\mathbb{R}^d)$ energy subcritical wave equation", *J. Funct. Anal.* **275**:10 (2018), 2614–2646. MR Zbl

[Rodriguez 2017] C. Rodriguez, "Scattering for radial energy-subcritical wave equations in dimensions 4 and 5", *Comm. Partial Differential Equations* **42**:6 (2017), 852–894. MR Zbl

[Roy 2009] T. Roy, "Adapted linear-nonlinear decomposition and global well-posedness for solutions to the defocusing cubic wave equation on $\mathbb{R}^3$", *Discrete Contin. Dyn. Syst.* **24**:4 (2009), 1307–1323. MR Zbl

[Shatah and Struwe 1993] J. Shatah and M. Struwe, "Regularity results for nonlinear wave equations", *Ann. of Math.* (2) **138**:3 (1993), 503–518. MR Zbl

[Shatah and Struwe 1994] J. Shatah and M. Struwe, "Well-posedness in the energy space for semilinear wave equations with critical growth", *Internat. Math. Res. Notices* 7 (1994), 303–309. MR Zbl

[Shen 2013] R. Shen, "On the energy subcritical, nonlinear wave equation in $\mathbb{R}^3$ with radial data", *Anal. PDE* **6**:8 (2013), 1929–1987. MR Zbl

[Sogge 2008] C. D. Sogge, *Lectures on non-linear wave equations*, 2nd ed., International Press, Boston, 2008. MR Zbl

[Strauss 1968] W. A. Strauss, "Decay and asymptotics for $cmu = F(u)$", *J. Functional Analysis* **2** (1968), 409–457. MR Zbl

[Tao 2001] T. Tao, "Global regularity of wave maps, I: Small critical Sobolev norm in high dimension", *Internat. Math. Res. Notices* 6 (2001), 299–328. MR Zbl

[Tao 2006] T. Tao, *Nonlinear dispersive equations: local and global analysis*, CBMS Regional Conference Series in Mathematics **106**, Amer. Math. Soc., Providence, RI, 2006. MR Zbl

[Tao 2007] T. Tao, "A (concentration-)compact attractor for high-dimensional non-linear Schrödinger equations", *Dyn. Partial Differ. Equ.* **4**:1 (2007), 1–53. MR Zbl

[Tao 2008a] T. Tao, "Global regularity of wave maps, III: Large energy from $\mathbb{R}^{1+2}$ to hyperbolic spaces", preprint, 2008. arXiv

[Tao 2008b] T. Tao, "Global regularity of wave maps, IV: Absence of stationary or self-similar solutions in the energy class", preprint, 2008. arXiv

[Tao 2008c] T. Tao, "Global regularity of wave maps, V: Large data local wellposedness and perturbation theory in the energy class", preprint, 2008. arXiv

[Tao 2009a] T. Tao, "Global regularity of wave maps, VI: Abstract theory of minimal-energy blowup solutions", preprint, 2009. arXiv

[Tao 2009b] T. Tao, "Global regularity of wave maps, VII: Control of delocalised or dispersed solutions", preprint, 2009. arXiv

[Tao et al. 2007] T. Tao, M. Visan, and X. Zhang, "Global well-posedness and scattering for the defocusing mass-critical nonlinear Schrödinger equation for radial data in high dimensions", *Duke Math. J.* **140**:1 (2007), 165–202. MR Zbl

[Tao et al. 2008] T. Tao, M. Visan, and X. Zhang, "Minimal-mass blowup solutions of the mass-critical NLS", *Forum Math.* **20**:5 (2008), 881–919. MR Zbl

Received 16 Oct 2018. Accepted 6 Sep 2019.

BENJAMIN DODSON: dodson@math.jhu.edu
Department of Mathematics, Johns Hopkins University, Baltimore, MD, United States

ANDREW LAWRIE: alawrie@mit.edu
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, United States

DANA MENDELSON: dana@math.uchicago.edu
Department of Mathematics, University of Chicago, Chicago, IL, United States

JASON MURPHY: jason.murphy@mst.edu
Department of Mathematics and Statistics, Missouri University of Science and Technology, Rolla, MO, United States

ANALYSIS AND PDE
Vol. 13, No. 7, 2020
https://doi.org/10.2140/apde.2020.13.2091

NEW FORMULAS FOR THE LAPLACIAN OF DISTANCE FUNCTIONS AND APPLICATIONS

FABIO CAVALLETTI AND ANDREA MONDINO

The goal of the paper is to prove an exact representation formula for the Laplacian of the distance (and more generally for an arbitrary 1-Lipschitz function) in the framework of metric measure spaces satisfying Ricci curvature lower bounds in a synthetic sense (more precisely in essentially nonbranching $\mathsf{MCP}(K, N)$-spaces). Such a representation formula makes apparent the classical upper bounds together with lower bounds and a precise description of the singular part. The exact representation formula for the Laplacian of a general 1-Lipschitz function holds also (and seems new) in a general complete Riemannian manifold.

We apply these results to prove the equivalence of $\mathsf{CD}(K, N)$ and a dimensional Bochner inequality on signed distance functions. Moreover we obtain a measure-theoretic splitting theorem for infinitesimally Hilbertian, essentially nonbranching spaces satisfying $\mathsf{MCP}(0, N)$.

1. Introduction 2091
2. Prerequisites 2098

Part I. A representation formula for the Laplacian 2109
3. Transport set and disintegration 2109
4. Representation formula for the Laplacian 2115

Part II. Applications 2133
5. The singular part of the Laplacian 2133
6. $\mathsf{CD}(K, N)$ is equivalent to a (K, N)-Bochner-type inequality 2136
7. Splitting theorem under $\mathsf{MCP}(0, N)$ 2140
Acknowledgements 2145
References 2145

1. Introduction

The Laplacian comparison theorem for the distance function from a point in a manifold with Ricci curvature bounded from below is one of the most fundamental results in Riemannian geometry. The local version states that if (M, g) is a smooth Riemannian manifold of dimension $N \geq 2$ satisfying

MSC2010: 49J52, 53C23.
Keywords: Ricci curvature, optimal transport, Laplacian comparison, distance function.

$\mathrm{Ric}_g \geq (N-1)g$ then, calling $\mathsf{d}_p(\cdot) := \mathsf{d}(p, \cdot)$ the distance from a point $p \in M$, until the distance function is smooth the following upper bound holds:

$$\Delta \mathsf{d}_p \leq (N-1) \cot \mathsf{d}_p. \tag{1-1}$$

Of course here Δ denotes the Laplacian (also called Laplace–Beltrami operator) of the Riemannian manifold (M, g) and cot is the cotangent (for a general lower bound $\mathrm{Ric}_g \geq Kg$, an analogous upper bound holds by replacing the right-hand side of (1-1) with the suitable (hyperbolic-)trigonometric function). The result is very classical and can be proved either via the Bochner inequality (see for instance [Cheeger 2001, Section 2]) or by Jacobi fields computations (see for instance [Petersen 1998, Chapter 7]).

It was Calabi [1958] who first extended the upper bound (1-1) to the whole manifold in the weak sense of barriers. Cheeger and Gromoll [1971], in their celebrated proof of the splitting theorem, then proved that the upper bound (1-1) also holds globally on M in a distributional sense (see also [Cheeger 2001, Section 4]). Since those classical works, the Laplacian comparison theorem has become a fundamental technical tool in the investigation of Riemannian manifolds satisfying Ricci curvature lower bounds (see for instance [Cheeger 2001; Cheeger and Colding 1996; 1997; 2000a; 2000b; Colding 1996a; 1996b; 1997; Colding and Naber 2012; Li and Yau 1986; Petersen 1998]).

We finally mention that recently Mantegazza, Mascellani, and Uraltsev [Mantegazza et al. 2014] obtained an exact representation formula for the distributional Hessian (and Laplacian) of the distance function from a point and that Gigli [2015] extended to the nonsmooth setting the upper bound (1-1).

The goal of this paper is to sharpen the Laplacian comparison theorem in several ways. First of all we will give an *exact* representation formula for the Laplacian of a general distance function (and for a general 1-Lipschitz function on its transport set; see later for the details) which describes exactly also the singular part concentrated on the cut locus. Such a representation formula will hold on *every complete* Riemannian manifold, without any curvature assumption. When specialised to Riemannian manifolds with Ricci curvature bounded below, such an exact representation formula will make apparent not only the celebrated *global upper bound* (1-1) but also *a lower bound on the regular part of the Laplacian*. The results will be proved in the much higher generality of (not necessarily smooth) metric measure spaces satisfying Ricci curvature lower bounds in a synthetic sense (more precisely, essentially nonbranching $\mathsf{MCP}(K, N)$-spaces); see the final part of the introduction.

In order to fix the ideas, we start the introduction discussing the smooth setting of Riemannian manifolds.

Let us introduce some notation in order to state the results. Given a point $p \in M$, denote by $\mathcal{C}_p$ the cut locus of p. The negative gradient flow $g_t : M \to M$ of the distance function d_p induces a partition $\{X_\alpha\}_{\alpha \in Q}$ of $M \setminus (\{p\} \cup \mathcal{C}_p)$ into minimising geodesics; each X_α is called a (transport) ray and Q is a suitable set of indices. We will denote the initial and final points of the ray X_α as $a(X_\alpha)$ and $b(X_\alpha)$ respectively; it is not hard to see that $a(X_\alpha) \in \mathcal{C}_p$ and $b(X_\alpha) = p$ for every $\alpha \in Q$. Let us stress that in this case the endpoints $a(X_\alpha), b(X_\alpha)$ are not elements of the ray X_α (in general, endpoints may or may not be elements of the ray, depending on the specific case; see also Remark 3.1). Such a partition induces a disintegration (the nonexpert reader can think of a kind of "nonstraight Fubini theorem") of the

Riemannian volume measure $\mathfrak{m}$ into measures $\mathfrak{m}_\alpha = h_\alpha \mathcal{H}^1 \llcorner_{X_\alpha}$ concentrated on X_α:

$$\mathfrak{m} = \int_Q h_\alpha \mathcal{H}^1 \llcorner_{X_\alpha} \, \mathfrak{q}(d\alpha), \tag{1-2}$$

where $\mathfrak{q}$ is a suitable probability measure on the set of indices Q. We refer to Section 3A for all the details on the disintegration formula. Here we only mention that once the probability $\mathfrak{q}$ is fixed within a suitable family of probability measures, then the functions h_α are uniquely determined.

The fact that (M, g) satisfies $\mathrm{Ric}_g \geq (N-1)g$ is inherited by the disintegration as concavity properties of the densities h_α; for the details see Section 3. For simplicity of notation, we will define

$$(\log h_\alpha)'(x) := \frac{d}{dt}\Big|_{t=0} \log h_\alpha(g_t(x));$$

thanks to the disintegration (1-2) and the (semi-)concavity of h_α along X_α, the quantity $(\log h_\alpha)'$ is well-defined $\mathfrak{m}$-a.e.

The first main result of the paper is an exact representation formula for the Laplacian of the distance function in nonsmooth spaces satisfying synthetic lower bounds on the Ricci curvature (see later in the introduction). In order to fix the ideas, we state it here for smooth Riemannian manifolds. We denote by $C_c(M)$ the space of real-valued continuous functions with compact support in M endowed with the final topology and by $(C_c(M))'$ its dual space made of real-valued continuous linear functionals on $C_c(M)$.

Theorem 1.1. *Let (M, g) be a smooth complete N-dimensional Riemannian manifold, where $N \geq 2$. Fix $p \in M$, and consider $\mathsf{d}_p := \mathsf{d}(p, \cdot)$ and an associated disintegration $\mathfrak{m} = \int_Q h_\alpha \mathcal{H}^1 \llcorner_{X_\alpha} \mathfrak{q}(d\alpha)$.*

Then $\Delta \mathsf{d}_p$ is an element of $(C_c(M))'$ with the representation formula

$$\Delta \mathsf{d}_p = -(\log h_\alpha)' \mathfrak{m} - \int_Q h_\alpha \delta_{a(X_\alpha)} \, \mathfrak{q}(d\alpha). \tag{1-3}$$

It can be written as the sum of three Radon measures

$$\Delta \mathsf{d}_p = [\Delta \mathsf{d}_p]^+_{\mathrm{reg}} - [\Delta \mathsf{d}_p]^-_{\mathrm{reg}} + [\Delta \mathsf{d}_p]_{\mathrm{sing}},$$

with

$$[\Delta \mathsf{d}_p]^\pm_{\mathrm{reg}} = -[(\log h_\alpha)']^\pm \mathfrak{m}, \quad [\Delta \mathsf{d}_p]_{\mathrm{sing}} = -\int_Q h_\alpha \delta_{a(X_\alpha)} \, \mathfrak{q}(d\alpha) \leq 0,$$

where $\pm$ stands for the positive and negative parts. Here, $[\Delta \mathsf{d}_p]_{\mathrm{reg}} := [\Delta \mathsf{d}_p]^+_{\mathrm{reg}} - [\Delta \mathsf{d}_p]^-_{\mathrm{reg}}$ is the regular part of $\Delta \mathsf{d}_p$ (i.e., absolutely continuous with respect to $\mathfrak{m}$), and $[\Delta \mathsf{d}_p]_{\mathrm{sing}}$ is the singular part.

In particular, if (M, g) is compact $\Delta \mathsf{d}_p$ is a finite signed Borel (and in particular Radon) measure.

Moreover, if $\mathrm{Ric}_g \geq Kg$ for some $K \in \mathbb{R}$, the following comparison results hold true (for simplicity here we assume $K = N-1$ for the bounds corresponding to a general $K \in \mathbb{R}$; see (4-15)):

$$\Delta \mathsf{d}_p \leq (N-1) \cot \mathsf{d}_p \mathfrak{m}, \tag{1-4}$$

$$[\Delta \mathsf{d}_p]_{\mathrm{reg}} = -(\log h_\alpha)' \mathfrak{m} \geq -(N-1) \cot \mathsf{d}_{a(X_\alpha)} \mathfrak{m}. \tag{1-5}$$

Remark 1.2 (on the lower bound (1-5)). Denote by $\mathcal{C}_p := \{a(X_\alpha)\}_{\alpha \in Q}$ the cut locus of p and by g_t the negative gradient flow of d_p at time t. More precisely, g_t is defined ray by ray as the translation by t in the direction of the negative gradient of d_p for $t \in (0, |X_\alpha|)$, where $|X_\alpha|$ denotes the length of the transport ray X_α, i.e., $|X_\alpha| = \mathsf{d}(a(X_\alpha), b(X_\alpha)) = \mathsf{d}(a(X_\alpha), p)$. Then for every $\varepsilon > 0$ there exists $C_{K,N,\varepsilon} > 0$ so that

$$[\Delta \mathsf{d}_p]_{\text{reg}} \geq -C_{K,N,\varepsilon}\mathfrak{m} \quad \text{on } \{x = g_t(a_\alpha) : t \geq \varepsilon\} \supset \{x \in X : \mathsf{d}(x, \mathcal{C}_p) \geq \varepsilon\}.$$

Let us stress that such a lower bound depends just on the dimension N, on the lower bound $K \in \mathbb{R}$ over the Ricci tensor, and on the distance $\varepsilon > 0$ from the cut locus $\mathcal{C}_p$, but is independent of the specific manifold (M, g).

We will prove the next more general statement for any signed distance function. Let us first give some definitions: Given a continuous function $v : M \to \mathbb{R}$ such that $\{v = 0\} \neq \varnothing$, the function

$$d_v : M \to \mathbb{R}, \quad d_v(x) := \mathsf{d}(x, \{v = 0\}) \operatorname{sgn}(v), \tag{1-6}$$

is called the *signed distance function* (from the zero-level set of v). With a slight abuse of notation, we denote by d both the distance between points and the induced distance between sets; more precisely

$$\mathsf{d}(x, \{v = 0\}) := \inf\{\mathsf{d}(x, y) : y \in \{v = 0\}\}.$$

Analogously to d_p, a signed distance function d_v induces a partition of M (up to a set of measure zero) into rays $\{X_\alpha\}_{\alpha \in Q}$ and a corresponding disintegration of the Riemannian volume measure $\mathfrak{m}$. The orientation of the rays is analogous. More precisely, if X_α is a transport ray associated with d_v and $a(X_\alpha)$, $b(X_\alpha)$ are its initial and final points, then $d_v(b(X_\alpha)) \leq 0$, $d_v(a(X_\alpha)) \geq 0$, so that transport rays are oriented from $\{v \geq 0\}$ towards $\{v \leq 0\}$.

Theorem 1.3. *Let (M, g) be a smooth complete N-dimensional Riemannian manifold, where $N \geq 2$.*

Consider the signed distance function d_v for some continuous function $v : X \to \mathbb{R}$ and an associated disintegration

$$\mathfrak{m} = \int_Q h_\alpha \mathcal{H}^1 \llcorner_{X_\alpha} \mathfrak{q}(d\alpha).$$

Then Δd_v^2 is an element of $(C_c(M))'$ with the representation formula

$$\Delta d_v^2 = 2(1 - d_v (\log h_\alpha)')\mathfrak{m} - 2 \int_Q (h_\alpha d_v)[\delta_{a(X_\alpha)} - \delta_{b(X_\alpha)}]\, \mathfrak{q}(d\alpha). \tag{1-7}$$

It can be written as the sum of three Radon measures

$$\Delta d_v^2 = [\Delta d_v^2]^+_{\text{reg}} - [\Delta d_v^2]^-_{\text{reg}} + [\Delta d_v^2]_{\text{sing}},$$

with

$$[\Delta d_v^2]^\pm_{\text{reg}} := 2(1 - d_v(\log h_\alpha)')^\pm \mathfrak{m}, \quad [\Delta d_v^2]_{\text{sing}} := -2 \int_Q (h_\alpha d_v)[\delta_{a(X_\alpha)} - \delta_{b(X_\alpha)}]\, \mathfrak{q}(d\alpha) \leq 0,$$

where $\pm$ stands for the positive and negative parts; in particular if (M, g) is compact, Δd_v^2 is a finite signed Borel (and in particular Radon) measure.

Moreover, if $\mathrm{Ric}_g \ge Kg$ *for some* $K \in \mathbb{R}$, *the following comparison results hold true* (*for simplicity here we assume* $K = N-1$ *for the bounds corresponding to a general* $K \in \mathbb{R}$; *see* (4-23), (4-24)):

$$[\Delta d_v^2]^+_{\mathrm{reg}} \le 2\mathfrak{m} + 2(N-1)\mathsf{d}(\{v=0\}, x)(\cot \mathsf{d}_{b(X_\alpha)}\mathfrak{m}\llcorner_{\{v\ge 0\}} + \cot \mathsf{d}_{a(X_\alpha)}\mathfrak{m}\llcorner_{\{v<0\}}), \tag{1-8}$$

$$[\Delta d_v^2]^-_{\mathrm{reg}} \le 2\mathfrak{m} - 2(N-1)\mathsf{d}(\{v=0\}, \cdot)(\cot \mathsf{d}_{a(X_\alpha)}\mathfrak{m}\llcorner_{\{v\ge 0\}} + \cot \mathsf{d}_{b(X_\alpha)}\mathfrak{m}\llcorner_{\{v<0\}}). \tag{1-9}$$

We will also present a general statement (Corollary 4.10) valid for any 1-Lipschitz function $u : M \to \mathbb{R}$, provided the rays of the induced disintegration satisfy a suitable integrability condition (roughly, they should not be too short), obtaining the same representation formula together with the two-sided estimate we mentioned before.

An interesting feature of Corollary 4.10 is that it will hold *for every* 1-Lipschitz function $u : X \to \mathbb{R}$. Let us stress that the 1-Lipschitz assumption is clearly a *first-order* condition, with no information on second-order derivatives. Nevertheless, Corollary 4.10 will imply that in a general complete Riemannian manifold it is possible to deduce some information on the *second derivatives once restricted to a suitable subset*. More precisely, if one considers only the set of points “saturating the 1-Lipschitz assumption” then the Laplacian of u is a continuous linear functional on C_c. We stress that we will obtain an *exact representation formula* of Δu (restricted to such a set) which, in the case the Ricci curvature of the ambient N-manifold is bounded below by $K \in \mathbb{R}$, will give a *two-sided bound* on the regular part in terms of K, N. We refer to Corollary 4.10 for the details.

Up to now we focussed the introduction on the setting of complete Riemannian manifolds (satisfying Ricci curvature lower bounds). However, everything will be proved in the much higher generality of (possibly nonsmooth) essentially nonbranching, metric measure spaces $(X, \mathsf{d}, \mathfrak{m})$ satisfying the measure contraction property $\mathsf{MCP}(K, N)$ for some $K \in \mathbb{R}$, $N \in (1, \infty)$. We refer to Section 2A for the detailed definitions; here let us just recall that $\mathsf{MCP}(K, N)$, introduced independently in [Ohta 2007a] and [Sturm 2006b], is the weakest among the synthetic conditions of Ricci curvature bounded below by K and dimension bounded above by N for metric measure spaces. In particular it is strictly weaker than the celebrated curvature dimension condition $\mathsf{CD}(K, N)$ pioneered in [Lott and Villani 2009; Sturm 2006a; 2006b] and than the (weaker) reduced curvature dimension condition $\mathsf{CD}^*(K, N)$ [Bacher and Sturm 2010]. The essential nonbranching condition, introduced by T. Rajala and Sturm [2014], roughly amounts to asking that W_2-geodesics are concentrated on nonbranching geodesics.

Remark 1.4 (notable examples of spaces fitting in the framework of the paper). The class of essentially nonbranching $\mathsf{MCP}(K, N)$ spaces include many remarkable families of spaces, among them:

- Smooth Finsler manifolds where the norm on the tangent spaces is strongly convex, and which satisfy lower Ricci curvature bounds. More precisely we consider a C^∞-manifold M, endowed with a function $F : TM \to [0, \infty]$ such that $F|_{TM\setminus\{0\}}$ is C^∞ and for each $p \in M$ it holds that $F_p := T_pM \to [0, \infty]$ is a strongly convex norm; i.e.,

$$g^p_{ij}(v) := \frac{\partial^2(F_p^2)}{\partial v^i \partial v^j}(v) \quad \text{is a positive definite matrix at every } v \in T_pM \setminus \{0\}.$$

Under these conditions, it is known that one can write the geodesic equations and geodesics do not branch; in other words these spaces are nonbranching. We also assume (M, F) to be geodesically complete and endowed with a C^∞ measure $\mathfrak{m}$ in such a way that the associated metric measure space $(X, F, \mathfrak{m})$ satisfies the $\mathsf{MCP}(K, N)$ condition; see [Ohta 2007b; Ohta and Sturm 2014].

• Sub-Riemannian manifolds. The following are all examples of essentially nonbranching $\mathsf{MCP}(K, N)$-spaces: the $(2n+1)$-dimensional Heisenberg group [Juillet 2009], any corank-1 Carnot group [Rizzi 2016], any ideal Carnot group [Rifford 2013], any generalised H-type Carnot group of rank k and dimension n [Barilari and Rizzi 2018].

• Strong $\mathsf{CD}^*(K, N)$ spaces, and in particular $\mathsf{RCD}^*(K, N)$ spaces (see below). The class of $\mathsf{RCD}^*(K, N)$ spaces includes the following remarkable subclasses:

  – Measured Gromov Hausdorff limits of Riemannian N-dimensional manifolds satisfying Ricci $\geq K$; see [Ambrosio et al. 2014b; Gigli et al. 2015b].

  – Finite-dimensional Alexandrov spaces with curvature bounded from below; see [Petrunin 2011].

In the context of metric measure spaces satisfying Ricci curvature lower bounds in a synthetic form, the Laplacian comparison theorem in its classical form (1-1) was established in [Gigli 2015]. More precisely, that work developed a notion of a possibly multivalued Laplacian holding on a general metric measure space $(X, \mathsf{d}, \mathfrak{m})$; it also introduces a property of the space called *infinitesimal strict convexity*, which grants, among other things, uniqueness of the Laplacian. Finally, assuming infinitesimal strict convexity and $\mathsf{CD}^*(K, N)$, a sharp upper bound for the Laplacian of a general Kantorovich potential for the W_2 distance is obtained, in particular, for d_p^2. The comparison in [Gigli 2015] is stated for $\mathsf{CD}^*(K, N)$ but the same proof, in the case of d_p^2, works assuming the weaker $\mathsf{MCP}(K, N)$.

Our results therefore extend the ones in [Gigli 2015] removing the assumption of infinitesimal strict convexity (hence including the possibility of a multivalued Laplacian, see Definition 2.12); moreover we precisely describe the Laplacian of a general signed distance function or a 1-Lipschitz function with sufficiently long transport rays, obtaining also a lower bound on the regular part and a representation formula for the singular part. We stress the fundamental role of the exact representation formulas: it will be the key in our application to the Bochner inequality (signed distance functions) and for the splitting theorem (general 1-Lipschitz function); see the discussions below.

We conclude this part on the related results in the literature mentioning that the Laplacian comparison results [Gigli 2015, Theorem 5.14, Corollary 5.15] seem to claim the stronger conclusion that $\Delta \mathsf{d}_p^2$ is a Radon measure in the classical sense (see Definition 2.11 and comments shortly afterwards). This however seems to not follow from the proof, when (X, d) is not compact: $\Delta \mathsf{d}_p^2$ is proved to be an element of $(C_c(X))'$ so, by the Riesz theorem, it is a difference of positive Radon measures but it may fail to be a Borel measure (see [Gigli 2015, Proposition 4.13] and the application of the Riesz theorem in the last part of its proof). We will therefore adapt the definition of Laplacian (see Definition 2.12), weakening [Gigli 2015, Definition 4.4]. With this new definition also [Gigli 2015, Proposition 4.13] together with its applications seem to work.

The second part of the paper is devoted to applications.

In Section 6 we will use the representation formula for the Laplacian to show that, under essential nonbranching, the $\mathsf{CD}(K, N)$ condition is equivalent to a dimensional Bochner inequality on signed distance functions. The Bochner inequality corresponds to an *Eulerian* formulation of Ricci curvature lower bounds, while the $\mathsf{CD}(K, N)$ condition, based on convexity of entropies along W_2-geodesics of probability measures, corresponds to a *Lagrangian* approach.

It has been a long-standing open problem, see for instance the celebrated book [Villani 2009, Open Problem 17.38, Conclusions and Open Problems, p. 923], to show that the Eulerian and the Lagrangian formulations of Ricci curvature lower bounds are equivalent. Such an equivalence has already been proved to hold true under the additional assumption that the heat flow $\mathrm{H}_t : L^2(X, \mathfrak{m}) \to L^2(X, \mathfrak{m})$ is linear for every $t \geq 0$ (or, equivalently, the Cheeger energy $\mathrm{Ch}(f) := \int_X |\nabla f|_w^2 \, \mathfrak{m}$ satisfies the parallelogram identity). The class of $\mathsf{CD}(K, N)$ spaces satisfying such a linearity condition is called $\mathsf{RCD}(K, N)$. After its birth in [Ambrosio et al. 2014b] (see also [Ambrosio et al. 2015a]) for $N = \infty$ and further developments for $N < \infty$ (see [Ambrosio et al. 2019; Erbar et al. 2015; Gigli 2015] and the subsequent [Cavalletti and Milman 2016]), the theory of metric measure spaces satisfying $\mathsf{RCD}(K, N)$ (called $\mathsf{RCD}(K, N)$-spaces for short) has been flourishing in the last years (for a survey of results, see [Villani 2019; Ambrosio 2018]).

The equivalence between $\mathsf{RCD}(K, N)$ and the Bochner inequality (properly written in a weak form, called Bakry-Émery condition $\mathrm{BE}(K, N)$) was proved for $N = \infty$ by Ambrosio, Gigli, and Savaré [Ambrosio et al. 2014b; 2015b], and in the finite-dimensional case by Erbar, Kuwada, and Sturm [Erbar et al. 2015] and Ambrosio, Mondino, and Savaré [Ambrosio et al. 2019].

Let us stress that the linearity of the heat flow was a crucial assumption in all of the aforementioned works.

The equivalence between the Bochner inequality and $\mathsf{CD}(K, N)$ was proved also in *smooth* Finsler manifolds by Ohta and Sturm. In [Ohta and Sturm 2014] no linearity of the heat flow is assumed, on the other hand the smoothness of the Finsler structure is heavily used in the computations. In the present paper, in contrast to the aforementioned works, we assume *neither that the heat flow is linear nor that the space is smooth*, thus showing that the equivalence between *Lagrangian* and *Eulerian* approaches to Ricci curvature lower bounds holds in the higher generality of nonsmooth "possibly Finslerian" spaces.

The proof of the equivalence seems also to follow rather easily once the representation formula for the Laplacian of signed distance functions is at our disposal. Here we also crucially use [Cavalletti and Milman 2016], where it is shown that a control on the behaviour of signed distance functions is sufficient to control the geometry of the space (see the statement: $\mathsf{CD}^1(K, N)$ implies $\mathsf{CD}(K, N)$). This also motivates our interest in the Laplacian of this family of functions (Theorem 4.14).

A second application is a measure-theoretic splitting theorem stating, roughly, that an infinitesimally Hilbertian (i.e., the Cheeger energy satisfies the parallelogram identity), essentially nonbranching $\mathsf{MCP}(0, N)$ space containing a line is isomorphic as a *measure space* to a splitting (for the precise statement see Theorem 7.1).

For smooth Riemannian manifolds [Cheeger and Gromoll 1971], as well as for Ricci-limits [Cheeger and Colding 1996] and $\mathsf{RCD}(0, N)$ spaces [Gigli 2013], the splitting theorem has a stronger statement giving an *isometric splitting*. However under the assumptions of Theorem 7.1 it is not conceivable to expect also a splitting of the metric. Indeed the Heisenberg group $\mathbb{H}^n$ is an example of a nonbranching infinitesimally Hilbertian $\mathsf{MCP}(0, N)$ space [Juillet 2009] containing a line, which is homeomorphic and isomorphic as a measure space to a splitting (indeed it is homeomorphic to $\mathbb{R}^n$ and the measure is exactly the n-dimensional Lebesgue measure) but it is not isometric to a splitting.

2. Prerequisites

In this section we review the basic material needed throughout the paper. The standing assumptions are that (X, d) is a complete, proper and separable metric space endowed with a positive Radon measure $\mathfrak{m}$ satisfying $\mathrm{supp}(\mathfrak{m}) = X$. The triple $(X, \mathsf{d}, \mathfrak{m})$ is said to be a metric measure space, m.m.s. for short.

The properness assumption is motivated by the synthetic Ricci curvature lower bounds we will assume to hold.

2A. ***Essentially nonbranching,* $\mathsf{MCP}(K, N)$ *and* $\mathsf{CD}(K, N)$ *metric measure spaces.*** We denote by

$$\mathrm{Geo}(X) := \{\gamma \in C([0, 1], X) : \mathsf{d}(\gamma_s, \gamma_t) = |s - t| \mathsf{d}(\gamma_0, \gamma_1) \text{ for every } s, t \in [0, 1]\}$$

the space of constant-speed geodesics. The metric space (X, d) is a *geodesic space* if and only if for each $x, y \in X$ there exists $\gamma \in \mathrm{Geo}(X)$ so that $\gamma_0 = x,\ \gamma_1 = y$.

Recall that, for complete geodesic spaces, local compactness is equivalent to properness (a metric space is proper if every closed ball is compact).

We denote by $\mathcal{P}(X)$ the space of all Borel probability measures over X and by $\mathcal{P}_2(X)$ the space of probability measures with finite second moment. $\mathcal{P}_2(X)$ can be endowed with the L^2-Kantorovich–Wasserstein distance W_2 defined as follows: for $\mu_0, \mu_1 \in \mathcal{P}_2(X)$, set

$$W_2^2(\mu_0, \mu_1) := \inf_\pi \int_{X \times X} \mathsf{d}^2(x, y)\, \pi(dx\, dy), \tag{2-1}$$

where the infimum is taken over all $\pi \in \mathcal{P}(X \times X)$ with μ_0 and μ_1 as the first and the second marginals. The space (X, d) is geodesic if and only if the space $(\mathcal{P}_2(X), W_2)$ is geodesic.

For any $t \in [0, 1]$, let e_t denote the evaluation map

$$\mathrm{e}_t : \mathrm{Geo}(X) \to X, \quad \mathrm{e}_t(\gamma) := \gamma_t.$$

Any geodesic $(\mu_t)_{t\in[0,1]}$ in $(\mathcal{P}_2(X), W_2)$ can be lifted to a measure $\nu \in \mathcal{P}(\mathrm{Geo}(X))$, so that $(\mathrm{e}_t)_\sharp \nu = \mu_t$ for all $t \in [0, 1]$.

Given $\mu_0, \mu_1 \in \mathcal{P}_2(X)$, we denote by $\mathrm{OptGeo}(\mu_0, \mu_1)$ the space of all $\nu \in \mathcal{P}(\mathrm{Geo}(X))$ for which $(\mathrm{e}_0, \mathrm{e}_1)_\sharp \nu$ realises the minimum in (2-1). Such a ν will be called *dynamical optimal plan*. If (X, d) is geodesic, then the set $\mathrm{OptGeo}(\mu_0, \mu_1)$ is nonempty for any $\mu_0, \mu_1 \in \mathcal{P}_2(X)$.

We will also consider the subspace $\mathcal{P}_2(X, \mathsf{d}, \mathfrak{m}) \subset \mathcal{P}_2(X)$ formed by all those measures absolutely continuous with respect to $\mathfrak{m}$.

A set $G \subset \mathrm{Geo}(X)$ is a set of nonbranching geodesics if and only if for any $\gamma^1, \gamma^2 \in G$, it holds

$$\text{there exists } \bar{t} \in (0, 1) \text{ such that, for all } t \in [0, \bar{t}], \quad \gamma_t^1 = \gamma_t^2 \quad \Longrightarrow \quad \gamma_s^1 = \gamma_s^2 \quad \text{for all } s \in [0, 1].$$

In this paper we will only consider essentially nonbranching spaces; let us recall their definition (introduced in [Rajala and Sturm 2014]).

Definition 2.1. A metric measure space $(X, \mathsf{d}, \mathfrak{m})$ is *essentially nonbranching* (e.n.b. for short) if and only if for any $\mu_0, \mu_1 \in \mathcal{P}_2(X)$, with μ_0, μ_1 absolutely continuous with respect to $\mathfrak{m}$, any element of $\mathrm{OptGeo}(\mu_0, \mu_1)$ is concentrated on a set of nonbranching geodesics.

It is clear that if (X, d) is a smooth Riemannian manifold then any subset $G \subset \mathrm{Geo}(X)$ is a set of nonbranching geodesics, in particular any smooth Riemannian manifold is essentially nonbranching.

In order to formulate curvature properties for $(X, \mathsf{d}, \mathfrak{m})$ we recall the definition of the distortion coefficients: for $K \in \mathbb{R}$, $N \in [1, \infty)$, $\theta \in (0, \infty)$, $t \in [0, 1]$, set

$$\tau_{K,N}^{(t)}(\theta) := t^{1/N} \sigma_{K,N-1}^{(t)}(\theta)^{(N-1)/N}, \tag{2-2}$$

where the σ-coefficients are defined as follows: given two numbers $K, N \in \mathbb{R}$ with $N \geq 0$, we set, for $(t, \theta) \in [0, 1] \times \mathbb{R}_+$,

$$\sigma_{K,N}^{(t)}(\theta) := \begin{cases} \infty & \text{if } K\theta^2 \geq N\pi^2, \\ \dfrac{\sin(t\theta\sqrt{K/N})}{\sin(\theta\sqrt{K/N})} & \text{if } 0 < K\theta^2 < N\pi^2, \\ t & \text{if } K\theta^2 < 0 \text{ and } N = 0, \text{ or if } K\theta^2 = 0, \\ \dfrac{\sinh(t\theta\sqrt{-K/N})}{\sinh(\theta\sqrt{-K/N})} & \text{if } K\theta^2 \leq 0 \text{ and } N > 0. \end{cases} \tag{2-3}$$

Let us also recall the definition of the Rényi entropy functional $\mathcal{E}_N : \mathcal{P}(X) \to [0, \infty]$,

$$\mathcal{E}_N(\mu) := \int_X \rho^{1-1/N}(x)\, \mathfrak{m}(dx), \tag{2-4}$$

where $\mu = \rho\mathfrak{m} + \mu^s$ with $\mu^s \perp \mathfrak{m}$.

Next we recall the definition of $\mathsf{MCP}(K, N)$ given independently in [Ohta 2007a] and [Sturm 2006b]. On general metric measure spaces the two definitions slightly differ, but on essentially nonbranching spaces they coincide (see for instance Appendix A in [Cavalletti and Mondino 2017a] or Proposition 9.1 in [Cavalletti and Milman 2016]). We report the one given in [Ohta 2007a].

Definition 2.2 (MCP condition). Let $K \in \mathbb{R}$ and $N \in [1, \infty)$. A metric measure space $(X, \mathsf{d}, \mathfrak{m})$ satisfies $\mathsf{MCP}(K, N)$ if for any $\mu_0 \in \mathcal{P}_2(X)$ of the form

$$\mu_0 = \frac{1}{\mathfrak{m}(A)} \mathfrak{m}\llcorner_A$$

for some Borel set $A \subset X$ with $\mathfrak{m}(A) \in (0, \infty)$, and any $o \in X$ there exists $\nu \in \mathrm{OptGeo}(\mu_0, \delta_o)$ such that

$$\frac{1}{\mathfrak{m}(A)} \mathfrak{m} \geq (\mathrm{e}_t)_\sharp\big(\tau_{K,N}^{(1-t)}(\mathsf{d}(\gamma_0, \gamma_1))\nu(d\gamma)\big) \quad \text{for all } t \in [0, 1]. \tag{2-5}$$

From [Cavalletti and Milman 2016, Proposition 9.1], in the setting of essentially nonbranching spaces Definition 2.2 is equivalent to the following condition: for all $\mu_0, \mu_1 \in \mathcal{P}_2(X)$ with $\mu_0 \ll \mathfrak{m}$ and $\mathrm{supp}(\mu_1) \subset \mathrm{supp}(\mathfrak{m})$, there exists a unique $\nu \in \mathrm{OptGeo}(\mu_0, \mu_1)$, ν is induced by a map (i.e., $\nu = S_\sharp(\mu_0)$

for some map $S : X \to \mathrm{Geo}(X)$), $\mu_t := (e_t)_\# \nu \ll \mathfrak{m}$ for all $t \in [0, 1)$, and writing $\mu_t = \rho_t \mathfrak{m}$, we have for all $t \in [0, 1)$

$$\rho_t^{-1/N}(\gamma_t) \geq \tau_{K,N}^{(1-t)}(\mathsf{d}(\gamma_0, \gamma_1))\rho_0^{-1/N}(\gamma_0) \quad \text{for } \nu\text{-a.e. } \gamma \in \mathrm{Geo}(X). \tag{2-6}$$

The curvature-dimension condition was introduced independently in [Lott and Villani 2009] and [Sturm 2006a; 2006b]; let us recall its definition.

Definition 2.3 (CD condition). Let $K \in \mathbb{R}$ and $N \in [1, \infty)$. A metric measure space $(X, \mathsf{d}, \mathfrak{m})$ satisfies $\mathsf{CD}(K, N)$ if for any two $\mu_0, \mu_1 \in \mathcal{P}_2(X, \mathsf{d}, \mathfrak{m})$ with bounded support there exist $\nu \in \mathrm{OptGeo}(\mu_0, \mu_1)$ and $\pi \in \mathcal{P}(X \times X)$ a W_2-optimal plan such that $\mu_t := (e_t)_\sharp \nu \ll \mathfrak{m}$ and for any $N' \geq N$, $t \in [0, 1]$,

$$\mathcal{E}_{N'}(\mu_t) \geq \int \tau_{K,N'}^{(1-t)}(\mathsf{d}(x, y))\rho_0^{-1/N'} + \tau_{K,N'}^{(t)}(\mathsf{d}(x, y))\rho_1^{-1/N'} \, \pi(dx\, dy). \tag{2-7}$$

Throughout this paper, we will always assume the proper metric measure space $(X, \mathsf{d}, \mathfrak{m})$ satisfies $\mathsf{MCP}(K, N)$ for some $K, N \in \mathbb{R}$, and is essentially nonbranching. This will imply in particular that (X, d) is geodesic.

It is not difficult to see that if $(X, \mathsf{d}, \mathfrak{m})$ satisfies $\mathsf{CD}(K, N)$ then it also satisfies $\mathsf{MCP}(K, N)$, but the converse implication is false in general (for example the sub-Riemannian Heisenberg group satisfies $\mathsf{MCP}(K, N)$ for some suitable K, N, but does not satisfy $\mathsf{CD}(K', N')$ for any choice of K', N').

It is worth recalling that if (M, g) is a Riemannian manifold of dimension n and $h \in C^2(M)$ with $h > 0$, then the m.m.s. $(M, \mathsf{d}_g, h\,\mathrm{Vol}_g)$ (where d_g and Vol_g denote the Riemannian distance and volume induced by g) satisfies $\mathsf{CD}(K, N)$ with $N \geq n$ if and only if (see [Sturm 2006b, Theorem 1.7])

$$\mathrm{Ric}_{g,h,N} \geq Kg, \quad \mathrm{Ric}_{g,h,N} := \mathrm{Ric}_g - (N-n)\frac{\nabla_g^2 h^{1/(N-n)}}{h^{1/(N-n)}}.$$

In particular if $N = n$, the generalised Ricci tensor $\mathrm{Ric}_{g,h,N} = \mathrm{Ric}_g$ makes sense only if h is constant.

A variant of the CD condition, called reduced curvature dimension condition and denoted by $\mathsf{CD}^*(K, N)$ [Bacher and Sturm 2010], asks for the same inequality (2-7) as $\mathsf{CD}(K,N)$ but the coefficients $\tau_{K,N}^{(t)}(\mathsf{d}(\gamma_0, \gamma_1))$ and $\tau_{K,N}^{(1-t)}(\mathsf{d}(\gamma_0, \gamma_1))$ are replaced by $\sigma_{K,N}^{(t)}(\mathsf{d}(\gamma_0, \gamma_1))$ and $\sigma_{K,N}^{(1-t)}(\mathsf{d}(\gamma_0, \gamma_1))$, respectively. For both definitions there is a local version and it was recently proved in [Cavalletti and Milman 2016] that on an essentially nonbranching m.m.s. with $\mathfrak{m}(X) < \infty$, the $\mathsf{CD}^*_{\mathrm{loc}}(K, N)$, $\mathsf{CD}^*(K, N)$, $\mathsf{CD}_{\mathrm{loc}}(K, N)$, $\mathsf{CD}(K, N)$ conditions are all equivalent for all $K \in \mathbb{R}$, $N \in (1, \infty)$, via the $\mathsf{CD}^1(K, N)$ condition defined in terms of the L^1-optimal transport problem. For more details we refer to [Cavalletti and Milman 2016].

2B. ***Lipschitz functions and Laplacians in metric measure spaces.*** We recall some facts about calculus in metric measure spaces following the approach of [Ambrosio et al. 2014a; 2014b; Gigli 2015] with the slight difference that here we confine the presentation to the (easier) setting of Lipschitz functions (instead of Sobolev), as in the paper we will work in such a framework. For this subsection it is enough to assume the metric space (X, d) is complete and separable and $\mathfrak{m}$ is a nonnegative locally finite measure.

A function $f : X \to \mathbb{R}$ is Lipschitz (or more precisely L-Lipschitz) if there exists a constant $L \geq 0$ such that

$$|f(x) - f(y)| \leq L\,\mathsf{d}(x, y) \quad \text{for all } x, y \in X.$$

The minimal constant $L \geq 0$ satisfying the last inequality is called *global Lipschitz constant* of f and is denoted by $\mathrm{Lip}(f)$.

We denote by $\mathrm{LIP}(X)$ the space of real-valued Lipschitz functions on (X, d) and by $\mathrm{LIP}_c(\Omega) \subset \mathrm{LIP}(X)$ the subspace of Lipschitz functions of X with compact support contained in the open subset $\Omega \subset X$.

Given $f \in \mathrm{LIP}(X)$, the *local Lipschitz constant* $|Df|(x_0)$ of f at $x_0 \in X$ is defined as

$$|Df|(x_0) := \limsup_{x \to x_0} \frac{|f(x) - f(x_0)|}{\mathsf{d}(x, x_0)} \quad \text{if } x_0 \text{ is not isolated,} \qquad |Df|(x_0) = 0 \quad \text{otherwise.}$$

It is clear that $|Df| \leq \mathrm{Lip}(f)$ on all X.

Definition 2.4. Let $f, u \in \mathrm{LIP}(X)$. Define the functions $D^{\pm} f(\nabla u) : X \to \mathbb{R}$ by

$$D^{+} f(\nabla u) := \inf_{\varepsilon > 0} \frac{|D(u + \varepsilon f)|^2 - |Du|^2}{2\varepsilon},$$

while $D^{-} f(\nabla u)$ is obtained by replacing $\inf_{\varepsilon > 0}$ with $\sup_{\varepsilon < 0}$.

If $D^{+} f(\nabla u) = D^{-} f(\nabla u)$ $\mathfrak{m}$-a.e. for all $f, u \in \mathrm{LIP}(X)$, then $(X, \mathsf{d}, \mathfrak{m})$ is said to be (Lipschitz-) infinitesimally strictly convex and we set $Df(\nabla u) := D^{+} f(\nabla u)$; if moreover $Df(\nabla u) = Du(\nabla f)$ $\mathfrak{m}$-a.e. for all $f, u \in \mathrm{LIP}(X)$, then $(X, \mathsf{d}, \mathfrak{m})$ is said to be (Lipschitz)-infinitesimally Hilbertian.

Remark 2.5. Given $f, u \in \mathrm{LIP}(X)$, it is easily seen the map $\varepsilon \mapsto |D(u + \varepsilon f)|^2$ is convex and real-valued. Thus

$$\inf_{\varepsilon > 0} \frac{|D(u + \varepsilon f)|^2 - |Du|^2}{2\varepsilon} = \liminf_{\varepsilon \downarrow 0} \frac{|D(u + \varepsilon f)|^2 - |Du|^2}{2\varepsilon},$$
$$\sup_{\varepsilon < 0} \frac{|D(u + \varepsilon f)|^2 - |Du|^2}{2\varepsilon} = \limsup_{\varepsilon \uparrow 0} \frac{|D(u + \varepsilon f)|^2 - |Du|^2}{2\varepsilon}.$$

Remark 2.6. The local doubling and Poincaré conditions will be satisfied throughout the paper as we will work in essentially nonbranching $\mathsf{MCP}(K, N)$-spaces, with $K \in \mathbb{R}$, $N \in (1, \infty)$ thanks to [von Renesse 2008, Corollary p. 28]. The standing assumptions in that paper are $\mathsf{MCP}(K, N)$ and that the set

$$C_x := \{y \in X : \text{there exists } \gamma^1 \neq \gamma^2 \in \mathrm{Geo}(X) \text{ such that } x = \gamma_0^1 = \gamma_0^2,\ y = \gamma_1^1 = \gamma_1^2\}$$

has $\mathfrak{m}$-measure zero for $\mathfrak{m}$-a.e. $x \in X$.

In an essentially nonbranching $\mathsf{MCP}(K, N)$ space the previous property can be obtained as follows: for any $r > 0$ invoke [Cavalletti and Mondino 2017a, Theorem 5.2] with $\mu_0 := \mathfrak{m}\llcorner_{B_r(x)}/\mathfrak{m}(B_r(x))$ and $\mu_1 := \delta_x$; existence of a map pushing μ_0 to the unique element of $\mathrm{OptGeo}(\mu_0, \mu_1)$ yields that $\mathfrak{m}(C_x \cap B_r(x)) = 0$, actually for any $x \in X$.

Remark 2.7. The notions of infinitesimally strictly convex and infinitesimally Hilbertian have been introduced in [Ambrosio et al. 2014b; Gigli 2015] in the setting of Sobolev spaces, with the local Lipschitz constant replaced by the minimal weak upper gradient. The corresponding Lipschitz counterparts that we defined above have been already considered in [Mondino 2015] and coincide with the ones of [Gigli 2015] provided the space satisfies doubling and Poincaré locally, thanks to a deep result of [Cheeger 1999]. Thanks to Remark 2.6 we will avoid therefore the prefix "Lipschitz" in the corresponding notions, for simplicity of notation.

Definition 2.8 (test plans, [Ambrosio et al. 2014a]). Let $(X, \mathsf{d}, \mathfrak{m})$ be a metric measure space as above and $\pi \in \mathcal{P}(C([0,1], X))$. We say that π is a test plan provided it has bounded compression; i.e., there exists $C > 0$ such that

$$(\mathrm{e}_t)_\sharp \pi = \mu_t \leq C\mathfrak{m} \quad \text{for all } t \in [0,1],$$

and

$$\iint_0^1 |\dot\gamma_t|^2 \, dt \, \pi(d\gamma) < \infty.$$

Definition 2.9 (plans representing gradients). Let $(X, \mathsf{d}, \mathfrak{m})$ be an m.m.s., $g \in \mathrm{LIP}(X)$ and π a test plan. We say that π represents the gradient of g provided it is a test plan and we have

$$\liminf_{t\to 0} \int \frac{g(\gamma_t) - g(\gamma_0)}{t} \, \pi(d\gamma) \geq \frac{1}{2} \int |Dg|^2(\gamma_0) \, \pi(d\gamma) + \frac{1}{2} \limsup_{t \to 0} \frac{1}{t} \iint_0^t |\dot\gamma_s|^2 \, ds \, \pi(d\gamma)$$

Theorem 2.10 [Ambrosio et al. 2014b, Lemma 4.5; Gigli 2015, Theorem 3.10]. *Let $f, u \in \mathrm{LIP}(X)$ and π be any plan representing the gradient of u; then*

$$\begin{aligned}\int D^+ f(\nabla u)(\mathrm{e}_0)_\sharp \, \pi &\geq \limsup_{t\to 0} \int \frac{f(\gamma_t) - f(\gamma_0)}{t} \, \pi(d\gamma) \\ &\geq \liminf_{t\to 0} \int \frac{f(\gamma_t) - f(\gamma_0)}{t} \, \pi(d\gamma) \\ &\geq \int D^- f(\nabla u)(\mathrm{e}_0)_\sharp \, \pi.\end{aligned}$$

In particular, if $(X, \mathsf{d}, \mathfrak{m})$ is infinitesimally strictly convex then

$$\int_X Df(\nabla u)(\mathrm{e}_0)_\sharp \, \pi = \lim_{t\to 0} \int \frac{f(\gamma_t) - f(\gamma_0)}{t} \, \pi(d\gamma).$$

In order to define the Laplacian, let us recall the definition of Radon functional. For simplicity, from now on, we will assume (X, d) to be locally compact (this will be satisfied throughout the paper as we will work in the setting of $\mathsf{MCP}(K, N)$ spaces which are, even more strongly, locally doubling).

Definition 2.11. • A *Radon functional over an open set* $\Omega \subset X$ is a linear functional $T : \mathrm{LIP}_c(\Omega) \to \mathbb{R}$ such that for every compact subset $W \subset \Omega$ there exists a constant $C_W \geq 0$ so that

$$|T(f)| \leq C_W \max_W |f| \quad \text{for all } f \in \mathrm{LIP}_c(\Omega) \text{ with } \operatorname{supp}(f) \subset W.$$

• A *nonnegative Radon measure over an open set* $\Omega \subset X$ is a Borel, nonnegative measure $\mu : \mathcal{B}(\Omega) \to [0, +\infty]$ that is locally finite; i.e., for any $x \in \Omega$ there exists a neighbourhood U_x of finite μ-measure: $\mu(U_x) < +\infty$. A nonnegative Radon measure is said to be *finite* if $\mu(X) < \infty$.

• A *signed Radon measure over an open set* $\Omega \subset X$ is a Borel measure $\mu : \mathcal{B}(\Omega) \to \mathbb{R} \cup \{\pm\infty\}$ that can be written as $\mu = \mu^+ - \mu^-$ with μ^+, μ^- nonnegative Radon measures, where at least one of the two is finite.

A signed Radon measure is said to be *finite* if, denoting by $\|\mu\| := \mu^+ + \mu^-$ the total variation measure, it holds $\|\mu\|(X) < \infty$.

Note that, by the classical Riesz–Markov–Kakutani representation theorem, for every *nonnegative* Radon functional T over X there exists a nonnegative Radon measure μ_T representing T via integration, i.e.,

$$T(f) = \int_X f(x)\, \mu_T(dx) \quad \text{for all } f \in \mathrm{LIP}_c(X).$$

In particular, every Radon functional can be written as the sum of two Radon measures (i.e., the positive and negative parts, respectively).

Let us stress that the nonnegativity assumption is crucial. Indeed a general Radon functional may not be representable by a measure; for example consider $X = \mathbb{R}$, $\Omega = \mathbb{R} \setminus \{0\}$ and $T : \mathrm{LIP}_c(\Omega) \to \mathbb{R}$ defined by

$$T : \mathrm{LIP}_c(\Omega) \to \mathbb{R}, \qquad T(f) := \int_\Omega \frac{f(x)}{x}\, dx.$$

It is straightforward to see that T is a real-valued Radon functional over Ω but cannot be represented by a signed Radon measure over Ω, the point being that $(-\infty, 0)$ would have "measure" $-\infty$ and $(0, +\infty)$ would have "measure" $+\infty$, thus failing the additivity axiom. An expert reader may recognise that $T(f)$ is (up to a multiplicative constant) the Hilbert transform of f evaluated at 0.

Definition 2.12. Let $\Omega \subset X$ be an open subset and let $u \in \mathrm{LIP}(X)$. We say that u is in the domain of the Laplacian of Ω, and write $u \in D(\mathbf{\Delta}, \Omega)$, provided there exists a Radon functional T over Ω such that for any $f \in \mathrm{LIP}_c(\Omega)$ it holds

$$\int_X D^- f(\nabla u)\, \mathfrak{m} \le -T(f) \le \int_X D^+ f(\nabla u)\, \mathfrak{m}. \tag{2-8}$$

In this case we write $T \in \mathbf{\Delta} u_{\llcorner\Omega}$. In the case T can be represented by a signed measure μ over Ω, with a slight abuse of notation we will identify T with μ and write $\mu \in \mathbf{\Delta} u_{\llcorner\Omega}$.

Let us stress that in general there is not a unique operator T satisfying (2-8); in other words the Laplacian can be multivalued.

2C. *Synthetic Ricci lower bounds over the real line.* Given $K \in \mathbb{R}$ and $N \in (1, \infty)$, a nonnegative Borel function h defined on an interval $I \subset \mathbb{R}$ is called an $\mathsf{MCP}(K, N)$ density on I if for all $x_0, x_1 \in I$ and $t \in [0, 1]$

$$h(tx_1 + (1-t)x_0) \ge \sigma_{K,N-1}^{(1-t)}(|x_1 - x_0|)^{N-1} h(x_0). \tag{2-9}$$

Even though it is a folklore result, we will include a proof of the following fact:

Lemma 2.13. *A one-dimensional metric measure space, that for simplicity we directly identify with* $(I, |\cdot|, h\mathcal{L}^1)$, *satisfies* $\mathsf{MCP}(K, N)$ *if and only there exists* $\tilde{h}$, *an* $\mathsf{MCP}(K, N)$ *density, such that* $h = \tilde{h}$ $\mathcal{L}^1$*-a.e. on* I.

Proof. Assume h is an $\mathsf{MCP}(K,N)$ density on I. From [Cavalletti and Milman 2016, Proposition 9.1(iv)], it will be enough to prove (2-6) under the additional assumption that

$$\mu_0 = \frac{1}{\mathfrak{m}(A)} \chi_A \mathfrak{m}$$

for some $A \subset I$ such that $0 < \mathfrak{m}(A) < \infty$, with $\mathfrak{m} = h\mathcal{L}^1$.

Without any loss in generality we assume $o = 0 \in I$. Given then any $A \subset I$ as above, the unique W_2 geodesic $(\mu_t)_{t\in[0,1]}$ connecting μ_0 to δ_0 is

$$\mu_t = (f_t)_\sharp \mu_0, \qquad f_t(x) = (1-t)x.$$

Then using the change of variable formula,

$$\mu_t = \rho_t \mathfrak{m}, \qquad \rho_t(x) = \frac{h(x/(1-t))}{h(x)} \frac{\chi_A(x/(1-t))}{(1-t)\mathfrak{m}(A)},$$

implying that

$$\left(\frac{\rho_t(f_t(x))}{\rho_0(x)}\right)^{-1/N} = \left(\frac{(1-t)h((1-t)x)}{h(x)}\right)^{1/N} \geq (1-t)^{1/N} \sigma_{K,N-1}^{(1-t)}(|x|)^{(N-1)/N} = \tau_{K,N}^{(1-t)}(|x|),$$

proving (2-6). In order to prove the converse implication, we fix $x_1 = 0 = o$ and take

$$\mu_0 := \frac{1}{\mathcal{L}^1(A)} \mathcal{L}^1 \llcorner_A, \qquad A \subset I,\ 0 < \mathcal{L}^1(A) < \infty.$$

Then

$$\mu_t := \frac{1}{\mathcal{L}^1((1-t)A)} \mathcal{L}^1 \llcorner_{(1-t)A}$$

is the unique W_2-geodesic connecting μ_0 to δ_o. Hence (2-9) can be applied to

$$\mu_t = \rho_t \mathfrak{m}, \qquad \rho_t(x) = \frac{1}{(1-t)\mathcal{L}^1(A)} \frac{\chi_{(1-t)A}(x)}{h(x)}.$$

Then (2-9) along (μ_t) implies the claim. □

The estimate (2-9) implies several known properties that we collect in what follows. To write them in a unified way we define for $\kappa \in \mathbb{R}$ the function $s_\kappa : [0,+\infty) \to \mathbb{R}$ (on $[0, \pi/\sqrt{\kappa})$ if $\kappa > 0$),

$$s_\kappa(\theta) := \begin{cases} (1/\sqrt{\kappa}) \sin(\sqrt{\kappa}\theta) & \text{if } \kappa > 0, \\ \theta & \text{if } \kappa = 0, \\ (1/\sqrt{-\kappa}) \sinh(\sqrt{-\kappa}\theta) & \text{if } \kappa < 0. \end{cases} \tag{2-10}$$

For the moment we confine ourselves to the case $I = (a,b)$ with $a, b \in \mathbb{R}$; hence (2-9) implies

$$\left(\frac{s_{K/(N-1)}(b-x_1)}{s_{K/(N-1)}(b-x_0)}\right)^{N-1} \leq \frac{h(x_1)}{h(x_0)} \leq \left(\frac{s_{K/(N-1)}(x_1-a)}{s_{K/(N-1)}(x_0-a)}\right)^{N-1} \tag{2-11}$$

for $x_0 \le x_1$ (see the proof of Lemma 2.17 for the easier estimate in the case $K = 0$). Hence denoting by $D = b - a$ the length of I, for any $\varepsilon > 0$ it follows that

$$\sup\left\{\frac{h(x_1)}{h(x_0)} : x_0, x_1 \in [a+\varepsilon, b-\varepsilon]\right\} \le C_\varepsilon, \tag{2-12}$$

where C_ε only depends on K, N, provided $2\varepsilon \le D \le 1/\varepsilon$.

Moreover (2-11) implies that h is locally Lipschitz in the interior of I and an easy manipulation of it (see [Cavalletti and Milman 2016, Lemma A.9]) yields the following bound on the derivative of h:

$$-(N-1)\frac{s'_{K/(N-1)}(b-x)}{s_{K/(N-1)}(b-x)} \le (\log h)'(x) \le (N-1)\frac{s'_{K/(N-1)}(x-a)}{s_{K/(N-1)}(x-a)} \tag{2-13}$$

if $x \in (a, b)$ is a point of differentiability of h. Finally if $K > 0$, then $b - a \le \pi\sqrt{(N-1)/K}$.

Remark 2.14. The estimate (2-11) also implies that an $\mathsf{MCP}(K, N)$ density $h : (a, b) \to (0, \infty)$, $a, b \in \mathbb{R}$, can always be extended to a continuous function on the closed interval $[a, b]$. Notice indeed that the map

$$(a, b) \ni x \mapsto \frac{h(x)}{(s_{K/(N-1)}(b-x))^{N-1}}$$

is nondecreasing and strictly positive. Hence the following limit exists and is a real number:

$$\lim_{x\to a} \frac{h(x)}{(s_{K/(N-1)}(b-x))^{N-1}}.$$

Since $b-a > 0$, we obtain that also the limit $\lim_{x\to a} h(x)$ exists, for every $K \le 0$ and for $K > 0$ provided $b - a \ne \pi\sqrt{(N-1)/K}$. The case $K > 0$ and $b - a = \pi\sqrt{(N-1)/K}$ follows by rigidity: (2-11) implies

$$\frac{\sin(\pi - x_1\sqrt{K/(N-1)})}{\sin(\pi - x_0\sqrt{K/(N-1)})} \le \frac{h(x_1)}{h(x_0)} \le \frac{\sin(x_1\sqrt{K/(N-1)})}{\sin(x_0\sqrt{K/(N-1)})},$$

showing that $h(x)$, up to a renormalisation constant, coincides with $\sin(x\sqrt{K/(N-1)})$. To show that h can also be extended to a continuous function at b, one can argue as above starting from the nonincreasing property of the function

$$(a, b) \ni x \mapsto \frac{h(x)}{(s_{K/(N-1)}(x-a))^{N-1}},$$

following again from (2-11).

The next lemma was stated and proved in [Cavalletti and Milman 2016, Lemma A.8] under the CD condition; as the proof only uses $\mathsf{MCP}(K, N)$ we report it in this more general version.

Lemma 2.15. *Let h denote an $\mathsf{MCP}(K, N)$ density on a finite interval (a, b), $N \in (1, \infty)$, which integrates to* 1. *Then*

$$\sup_{x\in(a,b)} h(x) \le \frac{1}{b-a}\begin{cases} N, & K \ge 0, \\ \left(\int_0^1 (\sigma^{(t)}_{K,N-1}(b-a))^{N-1}\, dt\right)^{-1}, & K < 0. \end{cases} \tag{2-14}$$

In particular, for fixed K and N, h is uniformly bounded from above as long as $b - a$ is uniformly bounded away from 0 *(and from above if $K < 0$).*

From the previous auxiliary results we obtain the following lemma that will be used throughout the paper.

Lemma 2.16. *Let h denote an $\mathsf{MCP}(K,N)$ density on a finite interval (a,b), $N\in(1,\infty)$, which integrates to* 1. *Then*

$$\int_{(a,b)} |h'(x)|\,dx \le \frac{1}{b-a} C^{(K,N)}_{(b-a)} \tag{2-15}$$

for some $C^{(K,N)}_{(b-a)}>0$ with the property that, for fixed $K\in\mathbb{R}$ and $N\in(1,\infty)$, it holds

$$\sup_{r\in(0,R)} C^{(K,N)}_r<\infty \quad \text{for every } R>0, \qquad \lim_{r\uparrow\infty} C^{(K,N)}_r=\infty. \tag{2-16}$$

Proof. Case 1: $K\le 0$. The two inequalities in (2-13) give for each point $x\in(a,b)$ of differentiability of h

$$\begin{aligned} w_1 &:= h'(x)+(N-1)\frac{s'_{K/(N-1)}(b-x)}{s_{K/(N-1)}(b-x)}h(x)\ge 0,\\ w_2 &:= h'(x)-(N-1)\frac{s'_{K/(N-1)}(x-a)}{s_{K/(N-1)}(x-a)}h(x)\le 0. \end{aligned} \tag{2-17}$$

Thus, we can write

$$\begin{aligned} \int_{[a,b]} |h'|\,dx \le \int_{[a,a+b-a/(2)]} w_1\,dx &+ \int_{[a,a+(b-a)/2]} |w_1-h'|\,dx \\ -\int_{[a+(b-a)/2,b]} w_2\,dx &+ \int_{[a+(b-a)/2,b]} |w_2-h'|\,dx. \end{aligned} \tag{2-18}$$

First of all, observing that for $K\le 0$ one has

$$\frac{s'_{K/(N-1)}(t)}{s_{K/(N-1)}(t)}\ge 0$$

for all $t\ge 0$, we get

$$\begin{aligned} \int_{[a,a+(b-a)/2]} w_1\,dx &\le h\Big(a+\frac{b-a}{2}\Big)-h(a)+(N-1)\|h\|_{L^\infty(a,b)}\log\Big(\frac{s_{K/(N-1)}(b-a)}{s_{K/(N-1)}((b-a)/2)}\Big)\\ &\le C^{(K,N)}_{(b-a)}\|h\|_{L^\infty(a,b)}, \end{aligned}$$

$$\int_{[a,a+(b-a)/2]} |w_1-h'|\,dx = \int_{[a,a+(b-a)/2]} (N-1)\frac{s'_{K/(N-1)}(b-x)}{s_{K/(N-1)}(b-x)}\,h(x)\le C^{(K,N)}_{(b-a)}\|h\|_{L^\infty(a,b)},$$

where $r\mapsto C^{(K,N)}_r$ satisfies (2-16). The bounds for the second line of (2-18) are analogous. Thus we conclude

$$\int_{[a,b]} |h'|\,dx\le C^{(K,N)}_{(b-a)}\|h\|_{L^\infty(a,b)},$$

which, recalling (2-14), gives the claim (2-15).

Case 2: $K>0$. In order to simplify the notation, we assume $K=N-1>0$ (so that $b-a\le\pi$), $a=0$ and $b-a=D\le\pi$. The discussion for general $K>0$, $a<b\in[0,\pi]$ is analogous.

We first consider the case $D \le \pi/2$. Using (2-11), notice that

$$\frac{h(x)}{\sin(D-x)} \le \frac{h(D/2)}{\sin(D/2)} \quad \text{for all } x \in [0, D/2],$$
$$\frac{h(x)}{\sin(x)} \le \frac{h(D/2)}{\sin(D/2)} \quad \text{for all } x \in [D/2, D].$$

For $x \in [0, D/2]$, these yield (recall that $\cos(x) \ge 0$)

$$\omega_0'(x) := h'(x) + \frac{\cos(D-x)}{\sin(D/2)} h(D/2) \ge h'(x) + \frac{\cos(D-x)}{\sin(D-x)} h(x) \ge 0,$$

and for $x \in [D/2, D]$ (recall that $\cos(x) \ge 0$)

$$\omega_1'(x) := h'(x) - \frac{\cos(x)}{\sin(D/2)} h(D/2) \le h'(x) - \frac{\cos(x)}{\sin(x)} h(x) \le 0.$$

Then we can collect all the estimates together:

$$\begin{aligned} \int_{[0,D]} |h'(x)| &\le \int_{[0,D/2]} \omega_0'(x)\,dx + \int_{[0,D/2]} |\omega_0'(x) - h'(x)|\,dx \\ &\quad - \int_{[D/2,D]} \omega_1'(x)\,dx + \int_{[D/2,D]} |\omega_1'(x) - h'(x)|\,dx \\ &\le C\|h\|_{L^\infty(0,D)}. \end{aligned} \tag{2-19}$$

The claim (2-15) then follows applying Lemma 2.15.

If $D > \pi/2$, like in the case $K \le 0$, the two inequalities in (2-13) give for each point $x \in (0, D)$ of differentiability of h

$$h'(x) + (N-1)\frac{\cos(D-x)}{\sin(D-x)} h(x) \ge 0,$$
$$h'(x) - (N-1)\frac{\cos(x)}{\sin(x)} h(x) \le 0.$$

Hence for $x \in (0, D-\pi/2)$ we have $h'(x) \ge 0$ and for $x \in [\pi/2, D]$ we have $h'(x) \le 0$. Then Lemma 2.15 and the bound $D \le \pi$ imply that

$$\begin{aligned} \int_{[0,D-\pi/2]\cup[\pi/2,D]} |h'(x)|\,dx &= \int_{[0,D-\pi/2]} h'(x)\,dx - \int_{[\pi/2,D]} h'(x)\,dx \\ &\le 4 \sup_{[0,D]} |h| \le \frac{4N}{D}. \end{aligned}$$

In order to complete the proof it is then enough to bound $\int_{[D-\pi/2,\pi/2]} |h'(x)|\,dx$. Since (2-19) was obtained for any h MCP-density on $[0, D]$ with $D \le \pi/2$ without using the assumption of $\int h = 1$, it implies

$$\int_{[0,\pi/2]} |h'(x)|\,dx \le C\|h\|_{L^\infty[0,\pi/2]}$$

for any MCP-density on $[0, D]$ with $D \ge \pi/2$. Lemma 2.15 gives the claim. $\square$

In the proof of the splitting theorem for $\mathsf{MCP}(0,N)$ spaces we will use the next lemma.

Lemma 2.17. *Let h be a $\mathsf{MCP}(0,N)$ measure on the whole real line $\mathbb{R}$. Then h is identically equal to a real constant.*

Proof. We show that $h(x_0)=h(x_1)$ for all $x_0,x_1\in\mathbb{R}$. The $\mathsf{MCP}(0,N)$ condition reads as

$$h(tx_1+(1-t)x_0)\ge(1-t)^{N-1}h(x_0).$$

For $a<z<b$ apply the previous estimate for $z=x_0$ and $x_1=b$. It implies

$$\frac{h(tb+(1-t)z)}{h(z)}\ge(1-t)^{N-1};$$

if $w\in(z,b)$ and $w=tb+(1-t)z$ for some $t\in(0,1)$, then $1-t=(b-w)/(b-z)$. Plugging in the previous inequality the explicit expression of $(1-t)$ and repeating the argument taking now $x_0=a$ and $x_1=z$, we obtain the next two-sided estimate

$$\left(\frac{b-x_1}{b-x_0}\right)^{N-1}\le\frac{h(x_1)}{h(x_0)}\le\left(\frac{x_1-a}{x_0-a}\right)^{N-1}, \tag{2-20}$$

valid for all $a\le x_0\le x_1\le b$. Since

$$\lim_{b\to+\infty}\left(\frac{b-x_1}{b-x_0}\right)^{N-1}=1=\lim_{a\to-\infty}\left(\frac{x_1-a}{x_0-a}\right)^{N-1},$$

and since (2-20) holds for all $a\in(-\infty,x_0)$ and all $b\in(x_1,+\infty)$, the thesis follows. □

We now review a few facts about $\mathsf{CD}(K,N)$ densities of the real line (see [Cavalletti and Milman 2016, Appendix]). Given $K\in\mathbb{R}$ and $N\in(1,\infty)$, a nonnegative Borel function h defined on an interval $I\subset\mathbb{R}$ is called a $\mathsf{CD}(K,N)$ density on I if for all $x_0,x_1\in I$ and $t\in[0,1]$

$$h^{1/(N-1)}((1-t)x_0+tx_1)\ge h^{1/(N-1)}(x_0)\sigma^{(1-t)}_{K,N}(|x_1-x_0|)+h^{1/(N-1)}(x_1)\sigma^{(t)}_{K,N-1}(|x_1-x_0|). \tag{2-21}$$

A one-dimensional metric measure space, say $(I,|\cdot|,h\mathcal{L}^1)$, satisfies $\mathsf{CD}(K,N)$ if and only h has a continuous representative $\tilde h$ that is a $\mathsf{CD}(K,N)$ density.

We will make use of the fact that a $\mathsf{CD}(K,N)$ density $h:I\to[0,\infty)$ is locally semiconcave in the interior; i.e., for all x_0 in the interior of I, there exists $C_{x_0}\in\mathbb{R}$ so that $h(x)-C_{x_0}x^2$ is concave in a neighbourhood of x_0.

Recall moreover that if $f:I\to\mathbb{R}$ denotes a convex function on an open interval $I\subset\mathbb{R}$, it is well known that the left and right derivatives $f^{\prime,-}$ and $f^{\prime,+}$ exist at every point in I and that f is locally Lipschitz; in particular, f is differentiable at a given point if and only if the left and right derivatives coincide. Denoting by $D\subset I$ the differentiability points of f in I, it is also well known that $I\setminus D$ is at most countable. Clearly, all of these results extend to locally semiconvex and locally semiconcave functions as well. We finally recall the next regularisation property for $\mathsf{CD}(K,N)$ densities obtained in [Cavalletti and Milman 2016, Proposition A.10]

Proposition 2.18. *Let h be a $\mathsf{CD}(K, N)$ density on an interval (a, b). Let ψ_ε denote a nonnegative C^2 function supported on $[-\varepsilon, \varepsilon]$ with $\int \psi_\varepsilon = 1$. For any $\varepsilon \in (0, (b-a)/2)$, define the function h^ε on $(a+\varepsilon, b-\varepsilon)$ by*

$$\log h^\varepsilon := \log h * \psi_\varepsilon := \int \log h(y)\psi_\varepsilon(x-y)\,dy.$$

Then h^ε is a C^2-smooth $\mathsf{CD}(K, N)$ density on $(a+\varepsilon, b-\varepsilon)$.

Part I. A representation formula for the Laplacian

3. Transport set and disintegration

Throughout this section we assume $(X, \mathsf{d}, \mathfrak{m})$ to be a metric measure space with $\mathrm{supp}(\mathfrak{m}) = X$ and (X, d) geodesic and proper (and hence complete).

3A. ***Disintegration of σ-finite measures.*** To any 1-Lipschitz function $u : X \to \mathbb{R}$ there is a naturally associated d-cyclically monotone set

$$\Gamma_u := \{(x, y) \in X \times X : u(x) - u(y) = \mathsf{d}(x, y)\}. \tag{3-1}$$

Its transpose is given by $\Gamma_u^{-1} = \{(x, y) \in X \times X : (y, x) \in \Gamma_u\}$. We define the *transport relation* R_u and the *transport set* $\mathcal{T}_u$ as

$$R_u := \Gamma_u \cup \Gamma_u^{-1}, \quad \mathcal{T}_u := P_1(R_u \setminus \{x = y\}), \tag{3-2}$$

where $\{x = y\}$ denotes the diagonal $\{(x, y) \in X^2 : x = y\}$ and P_i is the projection onto the i-th component. Recall that $\Gamma_u(x) = \{y \in X : (x, y) \in \Gamma_u\}$ denotes the section of Γ_u through x in the first coordinate, and similarly for $R_u(x)$ (through either of the coordinates by symmetry). Since u is 1-Lipschitz, Γ_u, Γ_u^{-1} and R_u are closed sets, and so are $\Gamma_u(x)$ and $R_u(x)$.

Also recall the following definitions, introduced in [Cavalletti 2014]:

$$A_+ := \{x \in \mathcal{T}_u : \text{there exists } z, w \in \Gamma_u(x) \text{ such that } (z, w) \notin R_u\},$$
$$A_- := \{x \in \mathcal{T}_u : \text{there exists } z, w \in \Gamma_u^{-1}(x) \text{ such that } (z, w) \notin R_u\}.$$

$A_\pm$ are called the *sets of forward and backward branching points*, respectively. If $x \in A_+$ and $(y, x) \in \Gamma_u$ then necessarily also $y \in A_+$ (as $\Gamma_u(y) \supset \Gamma_u(x)$ by the triangle inequality); similarly, if $x \in A_-$ and $(x, y) \in \Gamma_u$ then necessarily $y \in A_-$.

Consider the *nonbranched transport set*

$$\mathcal{T}_u^{\mathrm{nb}} := \mathcal{T}_u \setminus (A_+ \cup A_-), \tag{3-3}$$

and define the *nonbranched transport relation*

$$R_u^{\mathrm{nb}} := R_u \cap (\mathcal{T}_u^{\mathrm{nb}} \times \mathcal{T}_u^{\mathrm{nb}}).$$

In was shown in [Cavalletti 2014] (see also [Bianchini and Cavalletti 2013]) that R_u^{nb} is an equivalence relation over $\mathcal{T}_u^{\mathrm{nb}}$ and that for any $x \in \mathcal{T}_u^{\mathrm{nb}}$, $R_u(x) \subset (X, \mathsf{d})$ is isometric to a closed interval in $(\mathbb{R}, |\cdot|)$, and $R_u^{\mathrm{nb}}(x) \subset (X, \mathsf{d})$ is isometric to either a closed, semiclosed or open interval in $(\mathbb{R}, |\cdot|)$.

Therefore, from the nonbranched transport relation R_u^{nb}, one obtains a partition of the nonbranched transport set $\mathcal{T}_u^{\mathrm{nb}}$ into a disjoint family (of equivalence classes) $\{X_\alpha\}_{\alpha\in Q}$, each of them isometric to a real interval (depending on the situation, the interval can be bounded or unbounded, closed, semiclosed or open). Here Q is any set of indices. Concerning the measurability, as the space (X, d) is proper, $\mathcal{T}_u$ and $A_\pm$ are σ-compact sets and, consequently, $\mathcal{T}_u^{\mathrm{nb}}$ and R_u^{nb} are Borel.

Remark 3.1 (initial and final points). It will be useful to isolate two families of distinguished points of the transport set, the *sets of initial and final points*, respectively:

$$a := \{x \in \mathcal{T}_u : \text{there does not exist } y \in \mathcal{T}_u,\ y \neq x, \text{such that } (y, x) \in R_u\},$$
$$b := \{x \in \mathcal{T}_u : \text{there does not exist } y \in \mathcal{T}_u,\ y \neq x, \text{such that } (x, y) \in R_u\}.$$

Notice that no inclusion of the form $a \subset A_+,\ b \subset A_-$ is valid. For instance consider

$$X = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geq 0\}$$

endowed with the Euclidean distance and

$$u(x) := \mathrm{dist}(x, \{x_1 = 0\});$$

then $a = \{x_1 = 0\}$ and $A_\pm = \varnothing$. In particular, sets a and b may or may not be subsets of $\mathcal{T}_u^{\mathrm{nb}}$. See also the discussion right above (1-2). Curvature assumptions will, however, imply that a and b have measure zero. We will also use the notation $a(X_\alpha)$, $b(X_\alpha)$ to denote the starting and final points, respectively, of the transport set X_α, whenever they exist.

Once a partition of the nonbranched transport set $\mathcal{T}_u^{\mathrm{nb}}$ is at our disposal, a decomposition of the reference measure $\mathfrak{m}\llcorner_{\mathcal{T}_u^{\mathrm{nb}}}$ can be obtained using the disintegration theorem. In the recent literature of optimal transportation, the disintegration formulas have always been obtained under the additional assumption of finiteness of the measure $\mathfrak{m}(X) < \infty$. We will therefore spend few words on how to use disintegration theorem to obtain a disintegration associated to the family of transport rays without assuming $\mathfrak{m}(X) < \infty$.

We first introduce the quotient map $\mathfrak{Q} : \mathcal{T}_u^{\mathrm{nb}} \to Q$ induced by the partition

$$\alpha = \mathfrak{Q}(x) \iff x \in X_\alpha. \tag{3-4}$$

The set of indices (or quotient set) Q can be endowed with the quotient σ-algebra $\mathscr{Q}$ (of the σ-algebra $\mathscr{X}$ over X of $\mathfrak{m}$-measurable subsets),

$$C \in \mathscr{Q} \iff \mathfrak{Q}^{-1}(C) \in \mathscr{X},$$

i.e., the finest σ-algebra on Q such that $\mathfrak{Q}$ is measurable.

The set of indices Q can be identified with any subset of $\bar{Q} \subset X$ satisfying the following two properties:

- For all $x \in \mathcal{T}_u^{\mathrm{nb}}$ there exists a unique $\bar{x} \in \bar{Q}$ such that $(x, \bar{x}) \in R_u^{\mathrm{nb}}$.
- If $x, y \in \mathcal{T}_u^{\mathrm{nb}}$ and $(x, y) \in R_u^{\mathrm{nb}}$, then $\bar{x} = \bar{y}$.

In particular $\bar{Q}$ has to contain a single element for each equivalence class X_α.

Another way to obtain a quotient set is to look instead first for an explicit quotient map: in particular, any map $\bar{\mathfrak{Q}} : \mathcal{T}_u^{\mathrm{nb}} \to \mathcal{T}_u^{\mathrm{nb}}$ satisfying the properties

- $(x, \bar{\mathfrak{Q}}(x)) \in R_u^{\mathrm{nb}}$,
- if $(x, y) \in R_u^{\mathrm{nb}}$, then $\bar{\mathfrak{Q}}(x) = \bar{\mathfrak{Q}}(y)$,

will be a quotient map for the equivalence relation R_u^{nb} over $\mathcal{T}_u^{\mathrm{nb}}$; then the quotient set associated to $\bar{\mathfrak{Q}}$ will be the set $\{x \in R_u^{\mathrm{nb}} : x = \bar{\mathfrak{Q}}(x)\}$.

Existence of $\bar{Q}$ or of $\bar{\mathfrak{Q}}$ can be always deduced by the axiom of choice. However, in order to apply the disintegration theorem, measurability properties are needed.

A rather explicit construction of the quotient map has been already obtained under the additional assumption of $\mathfrak{m}(X) < \infty$ (see [Cavalletti and Mondino 2017b; 2018, Lemma 3.8]); however, $\mathfrak{m}(X) < \infty$ did not play any role in the proof and we therefore simply report the next statement.

We will denote by $\mathcal{A}$ the σ-algebra generated by the analytic sets of X.

Lemma 3.2 (Q is locally contained in level sets of u). *There exists an $\mathcal{A}$-measurable quotient map $\mathfrak{Q} : \mathcal{T}_u^{\mathrm{nb}} \to Q$ such that the quotient set $Q \subset X$ is $\mathcal{A}$-measurable and can be written locally as a level set of u in the following sense*:

$$Q = \bigcup_{n \in \mathbb{N}} Q_n, \qquad Q_n \subset u^{-1}(l_n),$$

where $l_n \in \mathbb{Q}$ and $Q_i \cap Q_j = \varnothing$ for $i \neq j$.

Lemma 3.2 allows us to apply the disintegration theorem (see [Cavalletti and Milman 2016, Section 6.3]), provided the ambient measure $\mathfrak{m}$ is suitably modified into a finite measure. To this aim, the next elementary lemma will be useful.

Lemma 3.3. *Let $\mathfrak{m}$ be a σ-finite measure over the proper metric space (X, d) with $\operatorname{supp}(\mathfrak{m}) = X$. Then there exists a Borel function $f : X \to (0, \infty)$ satisfying*

$$\inf_{\mathcal{K}} f > 0 \quad \textit{for any compact subset } \mathcal{K} \subset X, \qquad \int_{\mathcal{T}_u^{\mathrm{nb}}} f \, \mathfrak{m} = 1. \tag{3-5}$$

Proof. Since by assumption (X, d) is proper, for every $x_0 \in X$ and $R > 0$ the closed metric ball $\bar{B}_R(x_0)$ is compact. Thus, using that $\mathfrak{m}$ is σ-finite and $\operatorname{supp}(\mathfrak{m}) = X$, we get

$$0 < \mathfrak{m}(B_n(x_0) \setminus B_{n-1}(x_0)) < \infty \quad \text{for all } n \in \mathbb{N}_{\geq 1}.$$

It is then readily checked that $f : X \to (0, \infty)$ defined by

$$f := \frac{1}{2^n \mathfrak{m}(B_n(x_0) \setminus B_{n-1}(x_0))}$$

on $B_{n+1}(x_0) \setminus B_n(x_0)$ for all $n \in \mathbb{N}_{\geq 1}$ satisfies (3-5). □

Under the assumption that $\mathfrak{m}$ is σ-finite, let $f : X \to (0, \infty)$ satisfy (3-5), set

$$\mu := f \mathfrak{m}\llcorner_{\mathcal{T}_u^{\mathrm{nb}}}, \tag{3-6}$$

and define the normalised quotient measure

$$\mathfrak{q} := \mathfrak{Q}_\sharp \mu. \tag{3-7}$$

Notice that $\mathfrak{q}$ is a Borel probability measure over X. It is straightforward to check that

$$\mathfrak{Q}_\sharp(\mathfrak{m}\llcorner_{\mathcal{T}_u^{\rm nb}}) \ll \mathfrak{q}.$$

Take indeed $E \subset Q$ with $\mathfrak{q}(E)=0$; then by definition $\int_{\mathfrak{Q}^{-1}(E)} f(x)\,\mathfrak{m}(dx)=0$, implying $\mathfrak{m}(\mathfrak{Q}^{-1}(E))=0$, since $f>0$.

From the disintegration theorem [Fremlin 2003, Section 452], we deduce the existence of a map

$$Q \ni \alpha \mapsto \mu_\alpha \in \mathcal{P}(X)$$

satisfying the following properties:

(1) For any μ-measurable set $B \subset X$, the map $\alpha \mapsto \mu_\alpha(B)$ is $\mathfrak{q}$-measurable.

(2) For $\mathfrak{q}$-a.e. $\alpha \in Q$, μ_α is concentrated on $\mathfrak{Q}^{-1}(\alpha)$.

(3) For any μ-measurable set $B \subset X$ and $\mathfrak{q}$-measurable set $C \subset Q$, the following disintegration formula holds:

$$\mu(B \cap \mathfrak{Q}^{-1}(C)) = \int_C \mu_\alpha(B)\,\mathfrak{q}(d\alpha).$$

Finally the disintegration is $\mathfrak{q}$-essentially unique; i.e., if any other map $Q \ni \alpha \mapsto \bar{\mu}_\alpha \in \mathcal{P}(X)$ satisfies the previous three points, then

$$\bar{\mu}_\alpha = \mu_\alpha \quad \mathfrak{q}\text{-a.e. } \alpha \in Q.$$

Hence once $\mathfrak{q}$ is given (recall that $\mathfrak{q}$ depends on f from Lemma 3.3), the disintegration is unique up to a set of $\mathfrak{q}$-measure zero. In the case $\mathfrak{m}(X)<\infty$, the natural choice, that we tacitly assume, is to take as f the characteristic function of $\mathcal{T}_u^{\rm nb}$ divided by $\mathfrak{m}(\mathcal{T}_u^{\rm nb})$ so that $\mathfrak{q} := \mathfrak{Q}_\sharp(\mathfrak{m}\llcorner_{\mathcal{T}_u^{\rm nb}}/\mathfrak{m}(\mathcal{T}_u^{\rm nb}))$.

All the previous properties will be summarised saying that $Q \ni \alpha \mapsto \mu_\alpha$ is a *disintegration of* μ *strongly consistent with respect to* $\mathfrak{Q}$.

It follows from [Fremlin 2003, Proposition 452F] that

$$\int_X g(x)\,\mu(dx) = \int_Q \int g(x)\,\mu_\alpha(dx)\,\mathfrak{q}(d\alpha)$$

for every $g : X \to \mathbb{R}\cup\{\pm\infty\}$ such that $\int g\,\mu$ is well-defined in $\mathbb{R}\cup\{\pm\infty\}$. Hence picking $g=1/f$ (where f is the one used to define μ), we get

$$\mathfrak{m}\llcorner_{\mathcal{T}_u^{\rm nb}} = \int_Q \frac{\mu_\alpha}{f}\,\mathfrak{q}(d\alpha); \tag{3-8}$$

the previous identity has to be understood with test functions as the previous formula.

Defining $\mathfrak{m}_\alpha := \mu_\alpha/f$, we obtain that $\mathfrak{m}_\alpha$ is a nonnegative Radon measure over X satisfying all the measurability properties (with respect to $\alpha \in Q$) of μ_α and giving a disintegration of $\mathfrak{m}\llcorner_{\mathcal{T}_u^{\rm nb}}$ strongly consistent with respect to $\mathfrak{Q}$. Moreover, for every compact subset $\mathcal{K} \subset X$, it holds

$$\frac{1}{\sup_{\mathcal{K}} f}\mu_\alpha(\mathcal{K}) \le \mathfrak{m}_\alpha(\mathcal{K}) = \frac{\mu_\alpha}{f}(\mathcal{K}) \le \frac{1}{\inf_{\mathcal{K}} f} \quad \text{for } \mathfrak{q}\text{-a.e. } \alpha \in Q. \tag{3-9}$$

In the next statement, we summarise what we have obtained so far concerning the disintegration of a σ-finite reference measure $\mathfrak{m}$ with respect to the nonbranched transport relation induced by any 1-Lipschitz function $u : X \to \mathbb{R}$.

We denote by $\mathcal{M}_+(X)$ the space of nonnegative Radon measures over X.

Theorem 3.4. *Let $(X, \mathsf{d}, \mathfrak{m})$ be any geodesic and proper (hence complete) m.m.s. with* $\operatorname{supp}(\mathfrak{m}) = X$ *and $\mathfrak{m}$ σ-finite. Then for any 1-Lipschitz function $u : X \to \mathbb{R}$, the measure $\mathfrak{m}$ restricted to the nonbranched transport set $\mathcal{T}_u^{\mathrm{nb}}$ admits the disintegration formula*

$$\mathfrak{m}\llcorner_{\mathcal{T}_u^{\mathrm{nb}}} = \int_Q \mathfrak{m}_\alpha \, \mathfrak{q}(d\alpha),$$

where $\mathfrak{q}$ is a Borel probability measure over $Q \subset X$ such that $\mathfrak{Q}_\sharp(\mathfrak{m}\llcorner_{\mathcal{T}_u^{\mathrm{nb}}}) \ll \mathfrak{q}$ and the map $Q \ni \alpha \mapsto \mathfrak{m}_\alpha \in \mathcal{M}_+(X)$ satisfies the following properties:

(1) *For any $\mathfrak{m}$-measurable set B, the map $\alpha \mapsto \mathfrak{m}_\alpha(B)$ is $\mathfrak{q}$-measurable.*

(2) *For $\mathfrak{q}$-a.e. $\alpha \in Q$, $\mathfrak{m}_\alpha$ is concentrated on $\mathfrak{Q}^{-1}(\alpha) = R_u^{\mathrm{nb}}(\alpha)$ (strong consistency).*

(3) *For any $\mathfrak{m}$-measurable set B and $\mathfrak{q}$-measurable set C, the following disintegration formula holds:*

$$\mathfrak{m}(B \cap \mathfrak{Q}^{-1}(C)) = \int_C \mathfrak{m}_\alpha(B) \, \mathfrak{q}(d\alpha).$$

(4) *For every compact subset $\mathcal{K} \subset X$ there exists a constant $C_{\mathcal{K}} \in (0, \infty)$ such that*

$$\mathfrak{m}_\alpha(\mathcal{K}) \le C_{\mathcal{K}} \quad \text{for } \mathfrak{q}\text{-a.e. } \alpha \in Q.$$

Moreover, for any $\mathfrak{q}$ as above such that $\mathfrak{Q}_\sharp(\mathfrak{m}\llcorner_{\mathcal{T}_u^{\mathrm{nb}}}) \ll \mathfrak{q}$, the disintegration is $\mathfrak{q}$-essentially unique (see above).

3B. *Localisation of Ricci bounds.* Under the additional assumption of a synthetic lower bound on the Ricci curvature, one can obtain regularity properties both on $\mathcal{T}_u^{\mathrm{nb}}$ and on the conditional measures $\mathfrak{m}_\alpha$. As some of these results were obtained assuming $\mathfrak{m}(X) < \infty$, in what follows we review how to obtain the same regularity with no finiteness assumption on $\mathfrak{m}$. First of all recall that, for any $K \in \mathbb{R}$ and $N \in (1, \infty)$, $\mathsf{CD}(K, N)$ implies $\mathsf{MCP}(K, N)$, which in turn implies that $\mathfrak{m}$ is σ-finite. Thus Theorem 3.4 can be applied.

Lemma 3.5. *Let $(X, \mathsf{d}, \mathfrak{m})$ be an essentially nonbranching m.m.s. with* $\operatorname{supp}(\mathfrak{m}) = X$ *and satisfying* $\mathsf{MCP}(K, N)$ *for some $K \in \mathbb{R}$, $N \in (1, \infty)$. Then for any 1-Lipschitz function $u : X \to \mathbb{R}$, it holds* $\mathfrak{m}(\mathcal{T}_u \setminus \mathcal{T}_u^{\mathrm{nb}}) = 0$.

Lemma 3.5 has been proved in [Cavalletti 2014] for metric measure spaces $(X, \mathsf{d}, \mathfrak{m})$ satisfying $\mathsf{RCD}(K, N)$ with $N < \infty$ and $\operatorname{supp}(\mathfrak{m}) = X$. The $\mathsf{RCD}(K, N)$ assumption was used in that proof only to have at our disposal the following property: given $\mu_0, \mu_1 \in \mathcal{P}(X)$ with $\mu_0 \ll \mathfrak{m}$, there exists a unique optimal dynamical plan for the W_2-distance and it is induced by a map. In [Cavalletti and Mondino 2017a, Theorem 1.1] this property is also satisfied by an e.n.b. metric measure space satisfying $\mathsf{MCP}(K, N)$

with $\mathrm{supp}(\mathfrak{m}) = X$, without any finiteness assumption on $\mathfrak{m}$. Hence Lemma 3.5 can be proved following verbatim [Cavalletti 2014].

Building on [Cavalletti and Mondino 2017a], in [Cavalletti and Milman 2016, Theorem 7.10] additional information on the transport rays was proved: for $\mathfrak{q}$-a.e. $\alpha \in Q$ it holds

$$R_u(\alpha) = \overline{R_u^{\mathrm{nb}}(\alpha)} \supset R_u^{\mathrm{nb}}(\alpha) \supset \mathring{R}_u(\alpha), \tag{3-10}$$

with the latter to be interpreted as the relative interior. The additional assumption of $\mathfrak{m}(X) < \infty$ was used in the proof only to obtain the existence of a disintegration of $\mathfrak{m}$ strongly consistent with the nonbranched equivalence relation. Hence from Theorem 3.4 also (3-10) is valid in the present framework.

To conclude, we assert that the localisation results for the synthetic Ricci curvature lower bounds $\mathsf{MCP}(K, N)$ and $\mathsf{CD}(K, N)$, with $K, N \in \mathbb{R}$ and $N > 1$, are valid also in our framework.

- *Localisation of* $\mathsf{MCP}(K, N)$. In [Bianchini and Cavalletti 2013, Theorem 9.5], assuming nonbranching and the $\mathsf{MCP}(K, N)$ condition, it is proved (adopting slightly different notation) that for $\mathfrak{q}$-a.e. $\alpha \in Q$ it holds

$$\mathfrak{m}_\alpha = h_\alpha \mathcal{H}^1 \llcorner_{X_\alpha},$$

where $\mathcal{H}^1$ denotes the one-dimensional Hausdorff measure. Moreover, the one-dimensional metric measure space $(\overline{X}_\alpha, \mathsf{d}, \mathfrak{m}_\alpha)$, isomorphic to $([0, D_\alpha], |\cdot|, h_\alpha \mathcal{L}^1)$, is proved to satisfy $\mathsf{MCP}(K, N)$; here $\overline{X}_\alpha$ stands for the closure of the transport ray X_α with respect to d. Note that $\overline{X}_\alpha$ might not be a subset of $\mathcal{T}_u^{\mathrm{nb}}$ because of its endpoints but this will not affect any argument as $\mathfrak{m}_\alpha(\overline{X}_\alpha \setminus X_\alpha) = 0$. No finiteness assumption was assumed in [Bianchini and Cavalletti 2013, Theorem 9.5] and, since here we restrict to the nonbranched transport set, the arguments can be carried over to give the same statement.

- *Localisation of* $\mathsf{CD}(K, N)$. The localisation of $\mathsf{CD}(K, N)$ was proved in [Cavalletti and Mondino 2017a, Theorem 5.1] under the assumption $\mathfrak{m}(X) = 1$. Nevertheless, in that work the $\mathsf{CD}(K, N)$ condition was assumed to be valid only locally; i.e., the space was assumed to satisfy $\mathsf{CD}_{\mathrm{loc}}(K, N)$. In particular the proof first shows that the one-dimensional metric measure space $(X_\alpha, \mathsf{d}, \mathfrak{m}_\alpha)$ satisfies $\mathsf{CD}_{\mathrm{loc}}(K, N)$ for $\mathfrak{q}$-a.e. $\alpha \in Q$ and then, thanks to the local-to-global property of one-dimensional $\mathsf{CD}(K, N)$ condition, concludes with the full claim. Hence, if $(X, \mathsf{d}, \mathfrak{m})$ is e.n.b. and satisfies $\mathsf{CD}(K, N)$, since by Theorem 3.4 a disintegration formula is at our disposal and the reference measure $\mathfrak{m}$ is locally finite, one can repeat the arguments in [Cavalletti and Mondino 2017a, Theorem 5.1] and obtain that the one-dimensional metric measure space $(\overline{X}_\alpha, \mathsf{d}, \mathfrak{m}_\alpha)$, isomorphic to $([0, D_\alpha], |\cdot|, h_\alpha \mathcal{L}^1)$, satisfies $\mathsf{CD}(K, N)$.

We summarise the above discussion in the next statement.

Theorem 3.6. *Let* $(X, \mathsf{d}, \mathfrak{m})$ *be an essentially nonbranching m.m.s. with* $\mathrm{supp}(\mathfrak{m}) = X$ *and satisfying* $\mathsf{MCP}(K, N)$ *for some* $K \in \mathbb{R}$, $N \in (1, \infty)$.

Then, for any 1*-Lipschitz function* $u : X \to \mathbb{R}$, *there exists a disintegration of* $\mathfrak{m}$ *strongly consistent with* R_u^{nb} *satisfying*

$$\mathfrak{m}\llcorner_{\mathcal{T}_u^{\mathrm{nb}}} = \int_Q \mathfrak{m}_\alpha \, \mathfrak{q}(d\alpha), \qquad \mathfrak{q}(Q) = 1.$$

Moreover, for $\mathfrak{q}$*-a.e.* α, $\mathfrak{m}_\alpha$ *is a Radon measure with* $\mathfrak{m}_\alpha = h_\alpha \mathcal{H}^1 \llcorner_{X_\alpha} \ll \mathcal{H}^1 \llcorner_{X_\alpha}$ *and* $(\bar{X}_\alpha, \mathsf{d}, \mathfrak{m}_\alpha)$ *satisfies* $\mathsf{MCP}(K, N)$.

If, additionally, $(X, \mathsf{d}, \mathfrak{m})$ *satisfies* $\mathsf{CD}_{\mathrm{loc}}(K, N)$, *then* h_α *is a* $\mathsf{CD}(K, N)$ *density on* X_α *for* $\mathfrak{q}$*-a.e.* α.

It is worth recalling that, once we know that $(\bar{X}_\alpha, \mathsf{d}, \mathfrak{m}_\alpha)$ satisfies $\mathsf{MCP}(K, N)$, it is straightforward to get that $\mathfrak{m}_\alpha = h_\alpha \mathcal{H}^1 \llcorner_{X_\alpha}$ for some density h_α. We refer to Section 2C for all the properties satisfied by one-dimensional metric measure spaces satisfying lower Ricci curvature bounds.

We conclude the section by specialising the results to the smooth framework of Riemannian manifolds (cf. [Klartag 2017]).

Corollary 3.7. *Let* (M, g) *be a complete* N*-dimensional Riemannian manifold, where* $N \geq 2$, *and let* $\mathfrak{m}$ *denote its Riemannian volume measure.*

Then, for any 1*-Lipschitz function* $u : M \to \mathbb{R}$, *there exists a disintegration of* $\mathfrak{m}$ *strongly consistent with* R_u^{nb} *satisfying*

$$\mathfrak{m}\llcorner_{\mathcal{T}_u^{\mathrm{nb}}} = \int_Q \mathfrak{m}_\alpha \, \mathfrak{q}(d\alpha), \qquad \mathfrak{q}(Q) = 1.$$

Moreover:

(1) *For* $\mathfrak{q}$*-a.e.* α, $\mathfrak{m}_\alpha$ *is a Radon measure with* $\mathfrak{m}_\alpha = h_\alpha \mathcal{H}^1 \llcorner_{X_\alpha} \ll \mathcal{H}^1 \llcorner_{X_\alpha}$.

(2) *For every* $x \in M$ *there exist a (compact, geodesically convex) neighbourhood* U *of* x *and* $\bar{K} \in \mathbb{R}$ *such that* $h_\alpha \llcorner_U$ *is a* $\mathsf{CD}(\bar{K}, N)$ *density on* $X_\alpha \cap U$ *for* $\mathfrak{q}$*-a.e.* α.

(3) *If, additionally,* $\mathrm{Ric}_g \geq K g$ *for some* $K \in \mathbb{R}$, *then* h_α *is a* $\mathsf{CD}(K, N)$ *density on* X_α *for* $\mathfrak{q}$*-a.e.* α.

Proof. The corollary follows directly from Theorems 3.4 and 3.6, reasoning as follows. A complete Riemannian manifold is geodesic and proper; hence Theorem 3.4 implies the first part of the claim,

$$\mathfrak{m}\llcorner_{\mathcal{T}_u^{\mathrm{nb}}} = \int_Q \mathfrak{m}_\alpha \, \mathfrak{q}(d\alpha), \qquad \mathfrak{q}(Q) = 1,$$

and for $\mathfrak{q}$-a.e. α, $\mathfrak{m}_\alpha$ is a Radon measure with $\mathfrak{m}_\alpha = h_\alpha \mathcal{H}^1 \llcorner_{X_\alpha} \ll \mathcal{H}^1 \llcorner_{X_\alpha}$.

Moreover every point $x \in M$ admits a geodesically convex compact neighbourhood U where, by compactness, the Ricci tensor is bounded below by some $\bar{K} \in \mathbb{R}$. In particular $(U, \mathsf{d}, \mathfrak{m}\llcorner_U)$ is an essentially nonbranching $\mathsf{CD}(\bar{K}, N)$ space and thus we can apply Theorem 3.6 to $(U, \mathsf{d}, \mathfrak{m}\llcorner_U)$. Since the partition associated to $u : U \to \mathbb{R}$ is given by the restriction of transport rays, the quotient measure of $\mathfrak{m}$ restricted to $U \cap \mathcal{T}_u^{\mathrm{nb}}$ will be absolutely continuous with respect to $\mathfrak{q}$; hence by $\mathfrak{q}$-essential uniqueness of disintegration we deduce that $h_\alpha \llcorner_U$ is a $\mathsf{CD}(\bar{K}, N)$ density on $X_\alpha \cap U$ for $\mathfrak{q}$-a.e. α. The third claim is already contained in Theorem 3.6. □

4. Representation formula for the Laplacian

From now on we will assume $(X, \mathsf{d}, \mathfrak{m})$ to be an e.n.b. metric measure space satisfying $\mathsf{MCP}(K, N)$ for some $K \in \mathbb{R}$ and $N \in (1, \infty)$. In particular $(X, \mathsf{d}, \mathfrak{m})$ is a locally doubling and Poincaré space (recall Remark 2.6).

We will obtain an explicit representation formula for the Laplacian for a general 1-Lipschitz function

$$u : X \to \mathbb{R}, \quad |u(x) - u(y)| \le \mathsf{d}(x, y),$$

assuming a mild regularity property on $\mathcal{T}_u$, the associated transport set defined in Section 3.

A distinguished role will be played by a particular family of 1-Lipschitz functions, namely the so-called signed distance functions. Such a class played a key role in the recent proof [Cavalletti and Milman 2016] of the local-to-global property of $\mathsf{CD}(K, N)$ under the e.n.b. assumption.

Definition 4.1 (signed distance function). Given a continuous function $v : (X, \mathsf{d}) \to \mathbb{R}$ so that $\{v = 0\} \neq \varnothing$, the function

$$d_v : X \to \mathbb{R}, \quad d_v(x) := \mathsf{d}(x, \{v = 0\}) \operatorname{sgn}(v), \tag{4-1}$$

is called the signed distance function (from the zero-level set of v).

With a slight abuse of notation, we denote by d both the distance between points and the induced distance between sets; more precisely

$$\mathsf{d}(x, \{v = 0\}) := \inf\{\mathsf{d}(x, y) : y \in \{v = 0\}\}.$$

Lemma 4.2. *The signed distance function d_v is* 1*-Lipschitz on $\{v \ge 0\}$ and $\{v \le 0\}$. If (X, d) is a length space, then d_v is* 1*-Lipschitz on the entire X.*

For the proof we refer to [Cavalletti and Milman 2016, Lemma 8.4].

We now fix once and for all a 1-Lipschitz function $u : X \to \mathbb{R}$. In order not to have empty statements, throughout the section we will assume that $\mathfrak{m}(\mathcal{T}_u) > 0$.

4A. ***Representing the gradient of $-u$.*** The translation along $\mathcal{T}_u^{\mathrm{nb}}$ is defined as

$$g : \mathbb{R} \times \mathcal{T}_u^{\mathrm{nb}} \to \mathcal{T}_u^{\mathrm{nb}} \subset X, \quad \operatorname{graph}(g) = \{(t, x, y) \in \mathbb{R} \times R_u^{\mathrm{nb}} : u(x) - u(y) = t\}.$$

Since R_u^{nb} is Borel, the same applies to g, while $\operatorname{Dom}(g) = P_{12}(\operatorname{graph}(g))$ is analytic. We will write g_t for $g(t, \cdot)$. Notice that

$$\operatorname{graph}(g_t) = \{(x, y) \in R_u^{\mathrm{nb}} : u(x) - u(y) = t\}$$

is Borel as well and thus for $t \in \mathbb{R}$

$$\operatorname{Dom}(g_t) = \mathcal{T}_u^{\mathrm{nb}}(t) := \{x \in \mathcal{T}_u^{\mathrm{nb}} : \text{there exists } y \in R_u^{\mathrm{nb}}(x) \text{ with } u(x) - u(y) = t\}$$

is an analytic set. The rough intuitive picture is of course that g_t plays the role of negative gradient flow of u, restricted to the points of maximal slope 1. In order to handle the case when $\mathfrak{m}(\mathcal{T}_u^{\mathrm{nb}}) = +\infty$, it is useful to introduce the following definition.

Definition 4.3. A measurable subset $E \subset X$ is said to be R_u^{nb}*-convex* if for any $x \in \mathcal{T}_u^{\mathrm{nb}}$ the set $E \cap R_u^{\mathrm{nb}}(x)$ is isometric to an interval.

For every bounded R_u^{nb}-convex subset $E \subset \mathcal{T}_u^{\mathrm{nb}}(2\varepsilon)$ with $\mathfrak{m}(E) > 0$, consider the function $\Lambda : E \to C([0,1]; X)$ defined by

$$[0,1] \ni \tau \mapsto \Lambda(x)_\tau := \begin{cases} g_\tau(x), & \tau \in [0,\varepsilon], \\ g_\varepsilon(x), & \tau \in [\varepsilon, 1], \end{cases}$$

and set

$$\pi_E := \frac{1}{\mathfrak{m}(E)} \Lambda_\sharp \mathfrak{m}\llcorner_E. \tag{4-2}$$

Note that

$$\mathfrak{m}(E)(\mathrm{e}_\tau)_\sharp \pi_E = (\mathrm{e}_\tau \circ \Lambda)_\sharp \mathfrak{m}\llcorner_E = \begin{cases} (g_\tau)_\sharp \mathfrak{m}\llcorner_E =: \mathfrak{m}_E^\tau, & \tau \in [0,\varepsilon], \\ (g_\varepsilon)_\sharp \mathfrak{m}\llcorner_E =: \mathfrak{m}_E^\varepsilon, & \tau \in [\varepsilon, 1]. \end{cases} \tag{4-3}$$

The rough intuitive idea is of course that $\mathfrak{m}_E^\tau$ is the push forward of $\mathfrak{m}\llcorner_E$ via the negative gradient flow of u at time τ.

Proposition 4.4. *Let $(X, \mathsf{d}, \mathfrak{m})$ be an e.n.b. metric measure space satisfying $\mathsf{MCP}(K, N)$ and u be as before. For every bounded R_u^{nb}-convex subset $E \subset \mathcal{T}_u^{\mathrm{nb}}(2\varepsilon)$ with $\mathfrak{m}(E) > 0$, the measure π_E defined in (4-2) is a test plan representing the gradient of $-u$ (see Definition 2.9).*

Proof. Fix $t \in [0, \varepsilon]$. First of all write

$$\mathfrak{m}(E)(\mathrm{e}_t)_\sharp \pi_E = \mathfrak{m}_E^t = \int_Q (g_t)_\sharp \mathfrak{m}_\alpha\llcorner_E \, \mathfrak{q}(d\alpha). \tag{4-4}$$

Since $\mathfrak{m}_\alpha\llcorner_E = h_\alpha \mathcal{H}^1\llcorner_{X_\alpha \cap E}$, we have

$$(g_t)_\sharp \mathfrak{m}_\alpha\llcorner_E = \frac{h_\alpha \circ g_{-t}}{h_\alpha} \mathfrak{m}_\alpha\llcorner_{g_t(E)}. \tag{4-5}$$

Identifying $X_\alpha \cap \big(\bigcup_{t\in[0,\varepsilon]} g_t(E)\big)$ with an interval $[a_\alpha, b_\alpha] \subset \mathbb{R}$ (for the sake of the argument we assume the interval to be closed, but all the other cases are completely analogous), from (2-11), for $x \in [a_\alpha + t, b_\alpha - 2\varepsilon + t]$ and $t \le \varepsilon$ it holds

$$\frac{h_\alpha(x-t)}{h_\alpha(x)} \le \left[\frac{s_{K/(N-1)}(b_\alpha - x + t)}{s_{K/(N-1)}(b_\alpha - x)}\right]^{N-1} \le C_\varepsilon \quad \text{for all } x \in [a_\alpha + t, b_\alpha - 2\varepsilon + t] \text{ and } t \le \varepsilon, \tag{4-6}$$

where the last inequality follows from the fact that $b_\alpha - x \ge 2\varepsilon - t \ge \varepsilon > 0$. We stress that $C_\varepsilon > 0$ is independent of $\alpha \in Q$. The combination of (4-4), (4-5) and (4-6) gives that

$$(\mathrm{e}_t)_\sharp \pi \le \frac{C_\varepsilon}{\mathfrak{m}(E)} \mathfrak{m}$$

for all $t \in [0, 1]$; i.e., π_E has bounded compression. Moreover since $\mathfrak{m}(E) < \infty$, and $|\dot\gamma| = 1$ for π-a.e. γ, it follows that π_E is a test plan (Definition 2.8).

We now prove that π_E represents the gradient of $-u$. Since by construction $u(x) - u(g_\tau(x)) = \tau$ for $\mathfrak{m}\llcorner E$-a.e. x, we have

$$\liminf_{\tau\to 0} \int \frac{u(\gamma_0) - u(\gamma_\tau)}{\tau} \pi_E(d\gamma) = \frac{1}{\mathfrak{m}(E)} \liminf_{\tau \to 0} \int_E \frac{u(x) - u(g_\tau(x))}{\tau} \mathfrak{m}(dx) = 1.$$

Hence the claim (recall Definition 2.9) follows by the fact that the 1-Lipschitz regularity of u implies $|Du| \le 1$ $\mathfrak{m}$-a.e. and thus

$$1 \ge \frac{1}{2\mathfrak{m}(E)} \int_{\mathcal{T}_u^{\mathrm{nb}}(2\varepsilon)} |Du|^2(x)\, \mathfrak{m}(dx) + \tfrac{1}{2}\pi_E(C([0,1]; X)). \qquad \square$$

In the next statement and in the rest of the paper, we will often consider the restriction of a Lipschitz function f to some transport ray $R_u^{\mathrm{nb}}(\alpha)$ giving a real variable Lipschitz function: $[a_\alpha, b_\alpha] \ni t \mapsto f(g(t, a_\alpha))$. It will make sense then to compute the t-derivative of the previous map: whenever it exists, we will use the notation

$$f'(x) := \lim_{t\to 0} \frac{f(g(t,x)) - f(x)}{t}. \tag{4-7}$$

Note that f' is roughly the directional derivative of f "in the direction of $-\nabla u$". Observe that, if $(X, \mathsf{d}, \mathfrak{m})$ is $\mathsf{MCP}(K, N)$ e.n.b., for every $f \in \mathrm{LIP}(X)$ the quantity f' is well-defined $\mathfrak{m}$-a.e. on $\mathcal{T}_u$.

Theorem 4.5. *Let $(X, \mathsf{d}, \mathfrak{m})$ be an e.n.b. metric measure space satisfying $\mathsf{MCP}(K, N)$ and u be as before. Then for any Lipschitz function $f : X \to \mathbb{R}$ it holds*

$$D^- f(-\nabla u) \le f' \le D^+ f(-\nabla u) \quad \mathfrak{m}\text{-a.e. on } \mathcal{T}_u. \tag{4-8}$$

Proof. Given $f \in \mathrm{LIP}(X)$, fix $\varepsilon > 0$ and let $E \subset \mathcal{T}_u^{\mathrm{nb}}(2\varepsilon)$ be any bounded R_u^{nb}-convex subset with $\mathfrak{m}(E) > 0$. Theorem 2.10 together with Proposition 4.4 and (4-3) implies

$$\begin{aligned}\int_E D^- f(-\nabla u)\, \mathfrak{m} &\le \liminf_{\tau\to 0} \int_E \frac{f(g_\tau(x)) - f(x)}{\tau}\, \mathfrak{m}(dx) \\ &\le \limsup_{\tau\to 0} \int_E \frac{f(g_\tau(x)) - f(x)}{\tau}\, \mathfrak{m}(dx) \le \int_E D^+ f(-\nabla u)\, \mathfrak{m}.\end{aligned}$$

To conclude it is enough to observe that

$$\int_E \frac{f(g_\tau(x)) - f(x)}{\tau}\, \mathfrak{m}(dx) = \int_Q \int_{E\cap X_\alpha} \frac{f(g_\tau(x)) - f(x)}{\tau}\, \mathfrak{m}_\alpha(dx)\, \mathfrak{q}(d\alpha),$$

and notice that for each $\alpha \in Q$ the incremental ratio $(f(g_\tau(x)) - f(x))/\tau$ converges to $f'(x)$ for $\mathfrak{m}_\alpha$-a.e. $x \in X_\alpha$ and is dominated by the Lipschitz constant of f. Therefore, by the dominated convergence theorem, for each E as above it holds

$$\int_E D^- f(-\nabla u)\, \mathfrak{m} \le \int_E f'\, \mathfrak{m} \le \int_E D^+ f(-\nabla u)\, \mathfrak{m}.$$

The claim follows by the arbitrariness of $\varepsilon > 0$ and $E \subset \mathcal{T}_u^{\mathrm{nb}}(2\varepsilon)$. $\square$

The chain rule [Gigli 2015, Proposition 3.15] combined with Theorem 4.5 allows us to obtain:

Corollary 4.6. *Let $(X, \mathsf{d}, \mathfrak{m})$ be an e.n.b. metric measure space satisfying $\mathsf{MCP}(K, N)$ and u be as before. Then for any Lipschitz function $f : X \to \mathbb{R}$*

$$D^- f(-\nabla u^2) \le 2uf' \le D^+ f(-\nabla u^2),$$

where the inequalities hold true $\mathfrak{m}$-a.e. over $\mathcal{T}_u$.

Proof. We show that $D^- f(-\nabla u^2) \le 2uf'$; the argument for proving $2uf' \le D^+ f(-\nabla u^2)$ is completely analogous.

By the chain rule [Gigli 2015, Proposition 3.15], we know that

$$D^- f(-\nabla u^2) = 2u D^{-\operatorname{sgn} u} f(-\nabla u).$$

Combining the last identity with Theorem 4.5 yields

$$D^- f(-\nabla u^2) = \begin{cases} 2uD^+ f(-\nabla u) \le 2uf' & \mathfrak{m}\text{-a.e. on } \{u \le 0\}, \\ 2uD^- f(-\nabla u) \le 2uf' & \mathfrak{m}\text{-a.e. on } \{u \ge 0\}, \end{cases}$$

giving the claim. □

4B. ***A formula for the Laplacian of a general* 1*-Lipschitz function.*** The next proposition, which is key to showing that Δu is a Radon functional, follows from Lemmas 2.15 and 2.16. We use the notation that $a(X_\alpha)$ and $b(X_\alpha)$ denote the initial and final points respectively of the transport ray X_α. Recall also that h_α is positive and differentiable a.e. on X_α; in particular $\log h_\alpha$ is well-defined and differentiable a.e. along X_α.

Proposition 4.7. *Let* $(X, \mathsf{d}, \mathfrak{m})$ *be an e.n.b. metric measure space satisfying* $\mathsf{MCP}(K, N)$ *for some* $K \in \mathbb{R}$, $N \in (1, \infty)$. *Let* $u : X \to \mathbb{R}$ *be a* 1*-Lipschitz function with associated disintegration* $\mathfrak{m}\llcorner_{\mathcal{T}_u^{\mathrm{nb}}} = \int_Q \mathfrak{m}_\alpha \, \mathfrak{q}(d\alpha)$, *with* $\mathfrak{q}(Q) = 1$, $\mathfrak{m}_\alpha = h_\alpha \mathcal{H}^1\llcorner_{X_\alpha}$, $h_\alpha \in L^1(\mathcal{H}^1\llcorner_{X_\alpha})$ *for* $\mathfrak{q}$*-a.e.* $\alpha \in Q$. *Assume that*

$$\int_Q \frac{1}{\mathsf{d}(a(X_\alpha), b(X_\alpha))} \, \mathfrak{q}(d\alpha) < \infty.$$

Then $T_u : \mathrm{LIP}_c(X) \to \mathbb{R}$

$$T_u(f) := \int_Q \int_{X_\alpha} (\log h_\alpha)' f \, \mathfrak{m}_\alpha \, \mathfrak{q}(d\alpha) + \int_Q (h_\alpha f)(a(X_\alpha)) - (h_\alpha f)(b(X_\alpha)) \, \mathfrak{q}(d\alpha) \tag{4-9}$$

is a Radon functional over X.

Proof. Fix any bounded open subset $W \subset X$ and observe that we can find a bounded R_u^{nb}-convex measurable subset $E \subset \mathcal{T}_u^{\mathrm{nb}}$ such that $W \cap \mathcal{T}_u^{\mathrm{nb}} \subset E$ (take for instance on each X_α the convex-hull of $W \cap X_\alpha$) and

$$\mathsf{d}(a(X_\alpha \cap E), b(X_\alpha \cap E)) \ge \min\{1, \mathsf{d}(a(X_\alpha), b(X_\alpha))\} \quad \text{for all } \alpha \in Q. \tag{4-10}$$

Note that E depends just on W and the ray relation R_u^{nb}. For any $f \in \mathrm{LIP}_c(X)$ with $\mathrm{supp}(f) \subset W$, it is clear that

$$\begin{aligned} &\int_Q \int_{X_\alpha} (\log h_\alpha)' f \, \mathfrak{m}_\alpha \, \mathfrak{q}(d\alpha) + \int_Q (h_\alpha f)(a(X_\alpha)) - (h_\alpha f)(b(X_\alpha)) \, \mathfrak{q}(d\alpha) \\ &\quad = \int_Q \int_{X_\alpha \cap E} (\log h_\alpha)' f \, \mathfrak{m}_\alpha \, \mathfrak{q}(d\alpha) + \int_Q (h_\alpha f)(a(X_\alpha \cap E)) - (h_\alpha f)(b(X_\alpha \cap E)) \, \mathfrak{q}(d\alpha). \end{aligned}$$

Since E is bounded, we have $\sup_{\alpha \in Q} \mathsf{d}(a(X_\alpha \cap E), b(X_\alpha \cap E)) \le C_W$ for some $C_W \in (0, \infty)$ depending only on $W \subset X$. Moreover, Theorem 3.4(4) implies $\sup_{\alpha \in Q} \int_{X_\alpha \cap E} h_\alpha \, d\mathcal{H}^1 \le C_W$. Therefore, applying

Lemmas 2.15 and 2.16 to the renormalised densities

$$\tilde{h}_\alpha := \frac{1}{\int_{X_\alpha \cap E} h_\alpha \, d\mathcal{H}^1} h_\alpha$$

and rescaling back to get h_α, recalling also (4-10) we infer

$$\sup_{X_\alpha \cap E} h_\alpha(x) + \int_{X_\alpha \cap E} |h'_\alpha| \, d\mathcal{H}^1 \le C_W \frac{1}{\mathsf{d}(a(X_\alpha), b(X_\alpha))} \quad \text{for } \mathfrak{q}\text{-a.e. } \alpha \in \mathfrak{Q}(E) \subset Q.$$

We can thus estimate

$$\left| \int_Q \int_{X_\alpha \cap E} (\log h_\alpha)' f \, \mathfrak{m}_\alpha \, \mathfrak{q}(d\alpha) + \int_Q (h_\alpha f)(a(X_\alpha \cap E)) - (h_\alpha f)(b(X_\alpha \cap E)) \, \mathfrak{q}(d\alpha) \right| \le \left(C_W \int_Q \frac{1}{\mathsf{d}(a(X_\alpha), b(X_\alpha))} \, \mathfrak{q}(d\alpha) \right) \max |f|.$$

We can therefore conclude that (4-9) defines a Radon functional. □

The first main result follows by combining Theorem 4.5 and Proposition 4.7.

Theorem 4.8. *Let $(X, \mathsf{d}, \mathfrak{m})$ be an e.n.b. metric measure space with $\operatorname{supp}(\mathfrak{m}) = X$ and satisfying $\mathsf{MCP}(K, N)$ for some $K \in \mathbb{R}$, $N \in (1, \infty)$. Let $u : X \to \mathbb{R}$ be a 1-Lipschitz function with associated disintegration $\mathfrak{m}\llcorner_{\mathcal{T}_u^{\mathrm{nb}}} = \int_Q \mathfrak{m}_\alpha \, \mathfrak{q}(d\alpha)$, with $\mathfrak{q}(Q) = 1$, $\mathfrak{m}_\alpha = h_\alpha \mathcal{H}^1 \llcorner_{X_\alpha}$, $h_\alpha \in L^1(\mathcal{H}^1 \llcorner_{X_\alpha})$ for $\mathfrak{q}$-a.e. $\alpha \in Q$. Assume that*

$$\int_Q \frac{1}{\mathsf{d}(a(X_\alpha), b(X_\alpha))} \, \mathfrak{q}(d\alpha) < \infty.$$

Then, for any open subset $U \subset X$ such that $\mathfrak{m}(U \setminus \mathcal{T}_u) = 0$, it holds $u \in D(\mathbf{\Delta}, U)$. More precisely, $T_U : \mathrm{LIP}_c(U) \to \mathbb{R}$, defined by

$$T_U(f) := -\int_Q f h'_\alpha \mathcal{H}^1 \llcorner_{X_\alpha \cap U} \, \mathfrak{q}(d\alpha) + \int_Q (f h_\alpha)(b(X_\alpha)) - (f h_\alpha)(a(X_\alpha)) \, \mathfrak{q}(d\alpha),$$

is a Radon functional with $T_U \in \mathbf{\Delta} u\llcorner_U$. Moreover, writing $T_U = T_U^{\mathrm{reg}} + T_U^{\mathrm{sing}}$, with

$$T_U^{\mathrm{reg}}(f) := -\int_Q f h'_\alpha \mathcal{H}^1 \llcorner_{X_\alpha \cap U} \, \mathfrak{q}(d\alpha), \quad T_U^{\mathrm{sing}}(f) := \int_Q (f h_\alpha)(b(X_\alpha)) - (f h_\alpha)(a(X_\alpha)) \, \mathfrak{q}(d\alpha),$$

it holds that T_U^{reg} can be represented by $T_U^{\mathrm{reg}} = -(\log h_\alpha)' \mathfrak{m}\llcorner_U$ and satisfies the bounds

$$-(N-1) \frac{s'_{K/(N-1)}(\mathsf{d}(b(X_\alpha), x))}{s_{K/(N-1)}(\mathsf{d}(b(X_\alpha), x))} \le (\log h_\alpha)'(x) \le (N-1) \frac{s'_{K/(N-1)}(\mathsf{d}(x, a(X_\alpha)))}{s_{K/(N-1)}(\mathsf{d}(x, a(X_\alpha)))}. \tag{4-11}$$

Remark 4.9 (interpretation in the case X_α is unbounded). Let us explicitly note that, in the case the ray X_α is isometric to $(-\infty, b)$ (respectively $(a, +\infty)$), then by definition $(f h_\alpha)(a(X_\alpha)) = 0$ (resp. $(f h_\alpha)(b(X_\alpha)) = 0$). Let us discuss the case $K = -(N-1)$, the other cases being analogous. In the case the ray X_α is isometric to $(-\infty, b)$ (respectively $(a, +\infty)$), then the upper bound (resp. the lower bound) in (4-11) should be interpreted as $(\log h_\alpha)' \le N - 1$ (resp. $(\log h_\alpha)' \ge -(N-1)$). In particular, if for

$\mathfrak{q}$-a.e. $\alpha \in Q$ the ray X_α is isometric to $(-\infty, +\infty)$, then for any open subset $U \subset X$ with $\mathfrak{m}(U \setminus \mathcal{T}_u) = 0$ the singular part T_U^{sing} vanishes and it holds $-(N-1)\mathfrak{m}\llcorner_U \leq T_U^{\text{reg}} \leq (N-1)\mathfrak{m}\llcorner_U$.

Proof of Theorem 4.8. Fix an arbitrary open subset $U \subset X$ such that $\mathfrak{m}(U \setminus \mathcal{T}_u) = 0$. Let $f : X \to \mathbb{R}$ be any Lipschitz function compactly supported in U and let f' be defined $\mathfrak{m}$-a.e. by (4-7). Recall that the closure of the transport ray $(\overline{X}_\alpha, \mathsf{d}, \mathfrak{m}_\alpha)$ is isomorphic to a (possibly unbounded, possibly not open) real interval $[a(X_\alpha), b(X_\alpha)]$ endowed with the weighted measure $h_\alpha \mathcal{L}^1$, so we can integrate by parts Lipschitz functions on X_α analogously as on a weighted real interval.

Via an integration by parts, we thus obtain

$$\int_{X_\alpha} h_\alpha(x) f'(x)\, \mathcal{H}^1(dx) = -\int_{X_\alpha} h'_\alpha(x) f(x)\, \mathcal{H}^1(dx) + (h_\alpha f)(b(X_\alpha)) - (h_\alpha f)(a(X_\alpha)) \quad \mathfrak{q}\text{-a.e. } \alpha,$$

which, together with Theorem 3.6, gives

$$\int_U f'(x)\, \mathfrak{m}(dx) = -\int_Q \int_{X_\alpha} h'_\alpha(x) f(x)\, \mathcal{H}^1(dx) + (h_\alpha f)(b(X_\alpha)) - (h_\alpha f)(a(X_\alpha))\, \mathfrak{q}(d\alpha). \tag{4-12}$$

Proposition 4.7 ensures that, under the assumptions of Theorem 4.8, the expression

$$T_{\Delta u}(f) := -\int_Q f h'_\alpha \mathcal{H}^1\llcorner_{X_\alpha}\, \mathfrak{q}(d\alpha) + \int_Q (h_\alpha f)(b(X_\alpha)) - (h_\alpha f)(a(X_\alpha))\, \mathfrak{q}(d\alpha)$$

defines a Radon functional on U.

The combination of (4-12) with Theorem 4.5 gives

$$\int_U D^- f(-\nabla u)\, \mathfrak{m} \leq T_{\Delta u}(f) \leq \int_U D^+ f(-\nabla u)\, \mathfrak{m}.$$

Noting that (see [Gigli 2015, Proposition 3.15])

$$D^- f(-\nabla u) = -D^+ f(\nabla u), \quad D^+ f(-\nabla u) = -D^- f(\nabla u) \quad \mathfrak{m}\text{-a.e.},$$

the previous inequalities imply

$$\int_U D^- f(\nabla u)\, \mathfrak{m} \leq -T_{\Delta u}(f) \leq -\int_U D^+ f(\nabla u)\, \mathfrak{m}.$$

Recalling (2-13), the proof of all the claims is complete. □

The next result, dealing with smooth Riemannian manifolds, can be proved using Corollary 3.7 in the proof of Theorem 4.8 and following verbatim the arguments. Let us just mention that the Laplacian here is single-valued, i.e., $\{T_U\} = \boldsymbol{\Delta} u\llcorner_U$, since on a smooth Riemannian manifold (M, g) it holds $D^+ f(\nabla u) = D^- f(\nabla u) = g(\nabla f, \nabla u)$.

Corollary 4.10. *Let (M, g) be an N-dimensional complete Riemannian manifold, where $N \geq 2$. Let $u : M \to \mathbb{R}$ be a 1-Lipschitz function with associated disintegration $\mathfrak{m}\llcorner_{\mathcal{T}_u} = \int_Q \mathfrak{m}_\alpha\, \mathfrak{q}(d\alpha)$, with $\mathfrak{q}(Q) = 1$, $\mathfrak{m}_\alpha = h_\alpha \mathcal{H}^1\llcorner_{X_\alpha}$, $h_\alpha \in L^1(\mathcal{H}^1\llcorner_{X_\alpha})$ for $\mathfrak{q}$-a.e. $\alpha \in Q$. Assume that*

$$\int_Q \frac{1}{\mathsf{d}(a(X_\alpha), b(X_\alpha))}\, \mathfrak{q}(d\alpha) < \infty.$$

Then, for any open subset $U \subset M$ such that $\mathfrak{m}(U \setminus \mathcal{T}_u) = 0$, it holds $u \in D(\mathbf{\Delta}, U)$. More precisely, $T_U : \mathrm{LIP}_c(U) \to \mathbb{R}$, defined by

$$T_U(f) := -\int_Q f h'_\alpha \mathcal{H}^1 \llcorner_{X_\alpha \cap U} \, \mathfrak{q}(d\alpha) + \int_Q (f h_\alpha)(b(X_\alpha)) - (f h_\alpha)(a(X_\alpha)) \, \mathfrak{q}(d\alpha),$$

is a Radon functional with $\{T_U\} = \mathbf{\Delta} u \llcorner_U$. Moreover, writing $T_U = T_U^{\mathrm{reg}} + T_U^{\mathrm{sing}}$, with

$$T_U^{\mathrm{reg}}(f) := -\int_Q f h'_\alpha \mathcal{H}^1 \llcorner_{X_\alpha \cap U} \, \mathfrak{q}(d\alpha), \quad T_U^{\mathrm{sing}}(f) := \int_Q (f h_\alpha)(b(X_\alpha)) - (f h_\alpha)(a(X_\alpha)) \, \mathfrak{q}(d\alpha),$$

it holds that T_U^{reg} can be represented by $T_U^{\mathrm{reg}} = -(\log h_\alpha)' \mathfrak{m} \llcorner_U$.

In addition, if $\mathrm{Ric}_g \geq K g$ for some $K \in \mathbb{R}$, then the following bounds hold:

$$-(N-1) \frac{s'_{K/(N-1)}(\mathsf{d}(b(X_\alpha), x))}{s_{K/(N-1)}(\mathsf{d}(b(X_\alpha), x))} \leq (\log h_\alpha)'(x) \leq (N-1) \frac{s'_{K/(N-1)}(\mathsf{d}(x, a(X_\alpha)))}{s_{K/(N-1)}(\mathsf{d}(x, a(X_\alpha)))}. \tag{4-13}$$

Specialising Corollary 4.10 to the distance function gives Theorem 1.1; we briefly discuss the details below.

Proof of Theorem 1.1. Fix $p \in M$.

Step 1: $u := \mathsf{d}_p := \mathsf{d}(p, \cdot)$ satisfies the assumptions of Corollary 4.10.

Since by hypothesis (M, g) is complete, any point $x \in M$ can be joined to p with a length-minimising geodesic. Thus $\mathcal{T}_{\mathsf{d}_p} = M$, $b(X_\alpha) = p$ and $a(X_\alpha) \in \mathcal{C}_p$ for every $\alpha \in Q$. Moreover, there exists $\varepsilon = \varepsilon(p) > 0$ such that all the minimising geodesics X_α emanating from p have length $\mathsf{d}(a(X_\alpha), b(X_\alpha)) = \mathsf{d}(a(X_\alpha), p) \geq \varepsilon$. Since by construction $\mathfrak{q}(Q) = 1$, we conclude that the assumptions of Corollary 4.10 are satisfied.

Step 2: The representation formula (1-3) holds. We are left to show that

$$\int_Q h_\alpha \delta_{b(X_\alpha)} \, \mathfrak{q}(d\alpha) = 0.$$

Clearly, it is enough to show that

$$h_\alpha(p) = 0 \quad \text{for } \mathfrak{q}\text{-a.e. } \alpha \in Q. \tag{4-14}$$

Suppose by contradiction that there exists $\bar{Q} \subset Q$, where $h_\alpha(p) \geq c > 0$, with $\mathfrak{q}(\bar{Q}) > 0$. For simplicity of notation, we identify the minimising geodesic X_α with the real interval $[a_\alpha, b_\alpha]$, where p corresponds to b_α. Then by Fatou's lemma it holds

$$\begin{aligned} \infty > \omega_N = \liminf_{r \downarrow 0} \frac{\mathfrak{m}(B_r(p))}{r^N} &\geq \liminf_{r \downarrow 0} \int_{\bar{Q}} \frac{1}{r^N} \int_{[b_\alpha - r, b_\alpha]} h_\alpha(t) \, dt \, \mathfrak{q}(d\alpha) \\ &\geq \int_{\bar{Q}} \liminf_{r \downarrow 0} \frac{1}{r} \int_{[b_\alpha - r, b_\alpha]} \frac{h_\alpha(t)}{r^{N-1}} \, dt \, \mathfrak{q}(d\alpha) = \infty, \end{aligned}$$

giving a contradiction and thus proving the claim (4-14).

Step 3: We define three nonnegative Radon measures $[\Delta \mathsf{d}_p]^{\pm}_{\mathrm{reg}} := -[(\log h_\alpha)']^{\pm}\mathfrak{m}$ and $[\Delta \mathsf{d}_p]_{\mathrm{sing}} := -\int_Q h_\alpha \delta_{a(X_\alpha)}\, \mathfrak{q}(d\alpha)$, and let $\Delta \mathsf{d}_p = [\Delta \mathsf{d}_p]^{+}_{\mathrm{reg}} - [\Delta \mathsf{d}_p]^{-}_{\mathrm{reg}} + [\Delta \mathsf{d}_p]_{\mathrm{sing}}$.

Combining Corollary 3.7(2) with (2-13), it follows that $[\Delta \mathsf{d}_p]_{\mathrm{reg}} := -(\log h_\alpha)'\mathfrak{m}$ defines a Radon functional; by the Riesz theorem, its positive and negative parts are thus Radon measures. Also $[\Delta \mathsf{d}_p]_{\mathrm{sing}} := -\int_Q h_\alpha \delta_{a(X_\alpha)}\, \mathfrak{q}(d\alpha) = \Delta \mathsf{d}_p - [\Delta \mathsf{d}_p]_{\mathrm{reg}}$ is a nonpositive Radon functional (as a difference of Radon functionals) and thus, by the Riesz theorem, it defines a Radon measure.

Step 4: Upper and lower bounds in the case $\mathrm{Ric}_g \geq Kg$ for some $K \in \mathbb{R}$.

If $\mathrm{Ric}_g \geq Kg$, by Corollary 3.7(3) we know that h_α is a $\mathsf{CD}(K, N)$ (and in particular $\mathsf{MCP}(K, N)$) density over X_α for $\mathfrak{q}$-a.e. α. Thus (2-13) gives the bounds

$$-(N-1)\frac{s'_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)})}{s_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)})}\mathfrak{m} \leq [\Delta \mathsf{d}_p]_{\mathrm{reg}} \leq (N-1)\frac{s'_{K/(N-1)}(\mathsf{d}_p)}{s_{K/(N-1)}(\mathsf{d}_p)}\mathfrak{m}, \tag{4-15}$$

completing the proof. □

Remark 4.11 (on the bounds under the assumption $\mathrm{Ric}_g \geq Kg$). A few comments are in order:

- *The upper bound* in (4-15) is the celebrated Laplacian comparison theorem. Note that a similar upper bound is proved above to hold more generally for the (regular part of the) Laplacian of a (rather) general 1-Lipschitz function (4-13) in the high generality of e.n.b. $\mathsf{MCP}(K, N)$-spaces (4-11).

- *The case of the round sphere*. Let $p, q \in \mathbb{S}^N$ be a couple of antipodal points; clearly the cut locus of p coincides with q. In this case, choosing $u = \mathsf{d}_p$ in the construction above gives the partition of $\mathbb{S}^N \setminus \{p, q\}$ into meridians, and each ray is a meridian without its endpoints p, q, oriented from q to p. Theorem 1.1 thus yields

$$-(N-1)\cot \mathsf{d}_q \leq \Delta \mathsf{d}_p \leq (N-1)\cot \mathsf{d}_p \quad \text{on } \mathbb{S}^N.$$

Note that (for the round sphere) the same conclusion could be achieved by applying the Laplacian comparison theorem to d_p and to d_q and using that $\mathsf{d}_p = \pi - \mathsf{d}_q$.

- *The lower bound for a smooth Riemannian manifold.* Arguing analogously to the spherical case, one can achieve the lower bound along a (minimising) geodesic $\gamma : [0, 1] \to M$ with (M, g) satisfying $\mathrm{Ric}_g \geq (N-1)g$ (see [Colding and Naber 2012, Lemma 3.2]). In this case, the function $x \mapsto \mathsf{d}_{\gamma_0}(x) + \mathsf{d}_{\gamma_1}(x)$ achieves its minimum $\mathsf{d}(\gamma_0, \gamma_1)$ along $\gamma([0, 1])$; thus $\Delta(\mathsf{d}_{\gamma_0} + \mathsf{d}_{\gamma_1}) \geq 0$ along $\gamma((0, 1))$ and, applying the upper bound (1-1) to d_{γ_0}, d_{γ_1} and exploiting the linearity of the Laplacian we get

$$-(N-1)\cot \mathsf{d}_{\gamma_1} \leq \Delta \mathsf{d}_{\gamma_0} \leq (N-1)\cot \mathsf{d}_{\gamma_0} \quad \text{along } \gamma((0, 1)). \tag{4-16}$$

"Gluing" together all the inequalities (4-16) corresponding to all the (minimising) geodesics emanating from p gives (4-15). Clearly this argument holds for smooth Riemannian manifolds, but in situations where the space is a priori not smooth and the Laplacian is a priori not linear (as for e.n.b. $\mathsf{MCP}(K, N)$-spaces), one has to argue differently. As the reader could already appreciate (see, e.g., the proof of Theorem 1.1), we attacked the problem by using techniques from L^1-optimal transport.

A crucial fact in order to apply Theorem 4.8 to the distance function in the smooth case was that the cut locus of a point p is at strictly positive distance from p. This fact is clearly not at our disposal in the general setting of an e.n.b. $\mathsf{MCP}(K, N)$ space (e.g., the boundary of a convex body in $\mathbb{R}^3$ whose cut locus is dense). In Section 4C we will thus argue differently, showing first the result for the distance squared, and then getting the claim for the distance via the chain rule.

4C. ***A formula for the Laplacian of a signed distance function.*** The goal of the subsection is to prove the existence of the Laplacian of d_v and d_v^2 as Radon measures and to show upper and lower bounds; let us stress that, contrary to the previous subsection, here there will be no integrability assumption on the reciprocal of the length of the transport rays.

Recall that given a continuous function $v : (X, \mathsf{d}) \to \mathbb{R}$ so that $\{v = 0\} \neq \varnothing$, the signed distance function

$$d_v : X \to \mathbb{R}, \quad d_v(x) := \mathsf{d}(x, \{v = 0\}) \operatorname{sgn}(v),$$

is 1-Lipschitz.

Notice also that since (X, d) is proper, $\mathcal{T}_{d_v} \supset X \setminus \{v = 0\}$. Indeed, given $x \in X \setminus \{v = 0\}$, consider the distance minimising $z \in \{v = 0\}$ (whose existence is guaranteed by the compactness of closed bounded sets). Then $(x, z) \in R_{d_v}$ and thus $x \in \mathcal{T}_{d_v}$, as $x \neq z$. The next remark follows.

Remark 4.12. Let X_α be any transport ray associated with d_v and let $a(X_\alpha), b(X_\alpha)$ be its starting and final points, respectively. Then

$$d_v(b(X_\alpha)) \leq 0, \quad d_v(a(X_\alpha)) \geq 0,$$

whenever $b(X_\alpha)$ and $a(X_\alpha)$ exist.

The next lemma will be key to showing the existence of the Laplacian of d_v^2 as a Radon measure.

Lemma 4.13. *Let $(X, \mathsf{d}, \mathfrak{m})$ be an e.n.b. metric measure space satisfying $\mathsf{MCP}(K, N)$ for some $K \in \mathbb{R}$, $N \in (1, \infty)$.*

The expression

$$\begin{aligned} \nu := {} & 2\left(1 + \mathsf{d}(\{v = 0\}, x)(N-1)\frac{s'_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}\right)\mathfrak{m}\llcorner_{\{v \geq 0\}} \\ & + 2\left(1 + \mathsf{d}(\{v = 0\}, x)(N-1)\frac{s'_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}\right)\mathfrak{m}\llcorner_{\{v < 0\}} \end{aligned} \tag{4-17}$$

defines a signed Radon measure over X. More precisely:

<u>Case 1</u>: $K > 0$. *In this case ν is a signed finite measure on X satisfying $\nu \leq C_{K,N}\mathfrak{m}$. Then*:

- *If*
$$\sup_{x \in \{v \geq 0\}} \mathsf{d}(x, b(X_\alpha)) < \pi\sqrt{(N-1)/K}, \qquad \sup_{x \in \{v < 0\}} \mathsf{d}(x, a(X_\alpha)) < \pi\sqrt{(N-1)/K}$$
then ν has density bounded in $L^\infty(X, \mathfrak{m})$.

- *If*

$$\sup_{x\in\{v\geq 0\}} \mathsf{d}(x, b(X_\alpha)) = \pi\sqrt{(N-1)/K} \quad \text{or} \quad \sup_{x\in\{v<0\}} \mathsf{d}(x, a(X_\alpha)) = \pi\sqrt{(N-1)/K}$$

then there exist exactly two points $\bar{a}, \bar{b} \in X$ *with* $\mathsf{d}(\bar{a}, \bar{b}) = \pi\sqrt{(N-1)/K}$ *such that for* $\mathfrak{q}$*-a.e.* α

$$a(X_\alpha) = \bar{a}, \quad b(X_\alpha) = \bar{b},$$

and ν *has density bounded in*

$$L^\infty_{\mathrm{loc}}(\{v \geq 0\} \setminus \{\bar{a}\}, \mathfrak{m}) \cap L^\infty_{\mathrm{loc}}(\{v \leq 0\} \setminus \{\bar{b}\}, \mathfrak{m}) \cap L^1(X, \mathfrak{m}).$$

Moreover, in this case $(X, \mathfrak{m})$ *is isomorphic to a spherical suspension as a measure space. If in addition* $(X, \mathsf{d}, \mathfrak{m})$ *is an* $\mathsf{RCD}(K, N)$ *space, then* $(X, \mathsf{d}, \mathfrak{m})$ *is isomorphic to a spherical suspension as a metric measure space.*

Case 2: $K = 0$. *In this case* $\nu = 2\mathfrak{m}$ *is a nonnegative Radon measure; if* $b(X_\alpha)$ *or* $a(X_\alpha)$ *does not exist, the two ratios in* (4-17) *are posed by definition equal to* 0*, respectively.*

Case 3: $K < 0$. *In this case* ν *is a nonnegative Radon measure. If* $b(X_\alpha)$ *or* $a(X_\alpha)$ *does not exist, the two ratios in* (4-17) *are posed by definition equal to* 1*, respectively.*

Proof. Case 2: For $K = 0$ the bounds are a straightforward consequence of the definition of the coefficients $s_{K/(N-1)}$ given in (2-10).

Case 3: For $K < 0$ observe that, since $(0, \infty) \ni t \mapsto \coth t \in (0, \infty)$ is decreasing and $\mathsf{d}(\{v = 0\}, x) \leq \mathsf{d}_{b(X_\alpha)}(x)$ for all $x \in \{v \leq 0\}$, it holds

$$\begin{aligned} 0 \leq \mathsf{d}(\{v=0\}, x)\frac{s'_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))} &\leq \mathsf{d}(\{v=0\}, x)\frac{s'_{K/(N-1)}(\mathsf{d}(\{v=0\}, x))}{s_{K/(N-1)}(\mathsf{d}(\{v=0\}, x))} \\ &= \mathsf{d}(\{v=0\}, x)\sqrt{\frac{-K}{N-1}}\coth\left(\sqrt{\frac{-K}{N-1}}\mathsf{d}(\{v=0\}, x)\right) \quad \text{for all } x \in \{v \leq 0\}. \end{aligned}$$

Since the function

$$[0, \infty) \ni t \mapsto t\coth\left(\sqrt{\frac{-K}{N-1}}t\right)$$

is locally bounded and the discussion for the second line of (4-17) is completely analogous, the claim follows.

Case 1: For $K > 0$, recall that an $\mathsf{MCP}(K, N)$-space has diameter at most $\pi\sqrt{(N-1)/K}$. Since $(0, \pi) \ni t \mapsto \cot t$ is decreasing and $\mathsf{d}(\{v = 0\}, x) \leq \mathsf{d}_{b(X_\alpha)}(x)$ for all $x \in \{v \leq 0\}$, it holds

$$\begin{aligned} \mathsf{d}(\{v=0\}, x)\frac{s'_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))} &\leq \mathsf{d}(\{v=0\}, x)\frac{s'_{K/(N-1)}(\mathsf{d}(\{v=0\}, x))}{s_{K/(N-1)}(\mathsf{d}(\{v=0\}, x))} \\ &= \mathsf{d}(\{v=0\}, x)\sqrt{\frac{K}{N-1}}\cot\left(\sqrt{\frac{K}{N-1}}\mathsf{d}(\{v=0\}, x)\right) \quad \text{for all } x \in \{v \leq 0\}. \end{aligned}$$

It is easily checked that

$$\sup_{t\in[0,\pi\sqrt{(N-1)/K}]}\left(1+t\sqrt{K(N-1)}\cot\left(t\sqrt{\frac{K}{N-1}}\right)\right)\le C'_{K,N};$$

thus the bound $\nu \le C_{K,N}\mathfrak{m}$ follows.

Since

$$\inf_{t\in[0,\pi\sqrt{(N-1)/K}-\varepsilon]}\cot\left(t\sqrt{\frac{K}{N-1}}\right)>-\infty \quad \text{for every } \varepsilon\in\left(0,\pi\sqrt{\frac{N-1}{K}}\right],$$

we have that ν is a measure with L^∞-bounded density unless the second bullet of case 1 holds.

To discuss the second bullet of case 1 we assume $K = N-1$ in order to simplify the notation, the case for general $K > 0$ being completely analogous.

Using the maximal diameter theorem (proved in [Ohta 2007b] in the nonbranched $\mathsf{MCP}(N-1, N)$-setting and easily extendable to the present e.n.b situation) one can show that all the rays X_α are of length π; for the reader's convenience we give a self-contained argument. Let $X_{\bar\alpha}$ be a ray of length π and X_α be any other ray; then

$$\begin{aligned}\mathsf{d}(a(X_\alpha),b(X_\alpha))+\pi &= \mathsf{d}(a(X_\alpha),b(X_\alpha))+\mathsf{d}(a(X_{\bar\alpha}),b(X_\alpha))+\mathsf{d}(b(X_\alpha),b(X_{\bar\alpha}))\\ &\ge \mathsf{d}(a(X_{\bar\alpha}),b(X_\alpha))+\mathsf{d}(a(X_\alpha),b(X_{\bar\alpha})),\end{aligned}\tag{4-18}$$

where the first equality follows from [Ohta 2007b, Lemma 5.2] (since $|X_{\bar\alpha}|=\pi$, for each $x\in X$, $\mathsf{d}(x,a(X_{\bar\alpha}))+\mathsf{d}(x,b(X_{\bar\alpha}))=\pi$). By d-cyclical monotonicity also the reverse inequality is valid giving

$$\mathsf{d}(a(X_\alpha),b(X_\alpha))+\pi=\mathsf{d}(a(X_{\bar\alpha}),b(X_\alpha))+\mathsf{d}(a(X_\alpha),b(X_{\bar\alpha})).\tag{4-19}$$

In particular, $a(X_{\bar\alpha})\neq b(X_\alpha)$; indeed otherwise (4-19) would give $\mathsf{d}(a(X_\alpha),b(X_\alpha))+\pi=\mathsf{d}(a(X_\alpha),b(X_{\bar\alpha}))$, which, by virtue of the Myers diameter bound, would imply $a(X_\alpha)=b(X_\alpha)$. Contradicting the fact that the rays have strictly positive length.

Adding $\mathsf{d}(b(X_\alpha),b(X_{\bar\alpha}))-\pi$ (resp. $\mathsf{d}(a(X_\alpha),a(X_{\bar\alpha}))-\pi$) to both sides of (4-19) and using again [Ohta 2007b, Lemma 5.2], we get

$$\begin{aligned}\mathsf{d}(a(X_\alpha),b(X_\alpha))+\mathsf{d}(b(X_\alpha),b(X_{\bar\alpha}))&=\mathsf{d}(a(X_\alpha),b(X_{\bar\alpha})),\\ \mathsf{d}(a(X_\alpha),b(X_\alpha))+\mathsf{d}(a(X_\alpha),a(X_{\bar\alpha}))&=\mathsf{d}(a(X_{\bar\alpha}),b(X_\alpha)).\end{aligned}$$

Summing up the last two identities, together with (4-19), yields

$$\mathsf{d}(a(X_\alpha),b(X_\alpha))+\mathsf{d}(a(X_\alpha),a(X_{\bar\alpha}))+\mathsf{d}(b(X_{\bar\alpha}),b(X_\alpha))=\pi.$$

Since $\mathsf{d}(a(X_{\bar\alpha}),b(X_{\bar\alpha}))=\pi$, the last identity forces the four points $a(X_{\bar\alpha}), a(X_\alpha), b(X_\alpha), b(X_{\bar\alpha})$ to lie on the same geodesic γ. If $a(X_\alpha)\neq a(X_{\bar\alpha})$ then $a(X_\alpha)$ would be an internal point of γ, contradicting that $a(X_\alpha)$ is the initial point of the nonextendible ray X_α, and if $b(X_\alpha)\neq b(X_{\bar\alpha})$ then $b(X_\alpha)$ would be an internal point of γ, contradicting that $b(X_\alpha)$ is the final point of the nonextendible ray X_α.

Moreover $(X, \mathfrak{m})$ is isomorphic as a measure space to a spherical suspension over any transport ray of length π [Ohta 2007b, p. 235].

We are left to show that the density of ν is in $L^1(X, \mathfrak{m})$. By symmetry it is enough to show that

$$\int_{\{v\geq 0\}} \big|1+(N-1)\mathsf{d}(\{v=0\},x)\cot(\mathsf{d}(\bar{b},x))\big|\,\mathfrak{m}(dx)<\infty. \tag{4-20}$$

Notice that, for every fixed $\varepsilon \in [0, \pi/2]$, the integrand is bounded for $\mathsf{d}(\bar{b}, x) \in [\varepsilon, \pi-\varepsilon]$.

Since $\bar{b} \in \{v \leq 0\}$, if $\bar{b}$ is an accumulation point for $\{v \geq 0\}$, then $v(\bar{b}) = 0$. As v is strictly decreasing on the rays, which cover a dense subset, it follows that $\{v = 0\} = \{\bar{b}\}$. Thus, in this case, the integrand becomes $1 + \mathsf{d}(\bar{b}, x)\cot(\mathsf{d}(\bar{b}, x))$, which is bounded for $\mathsf{d}(\bar{b}, x) \in [0, \varepsilon]$.

We now show that the integral is finite also on

$$\{x : \mathsf{d}(\bar{b}, x) \in [\pi-\varepsilon, \pi]\} \cap \{v \geq 0\}.$$

Since

$$\mathsf{d}(\{v=0\}, x) \leq \mathsf{d}(\bar{b}, x),$$

it is enough to show that

$$\int_{\{v\geq 0\}\cap \mathsf{d}(\bar{b},x)\in[\pi-\varepsilon,\pi]} |1+(N-1)\mathsf{d}(\bar{b},x)\cot(\mathsf{d}(\bar{b},x))|\,\mathfrak{m}(dx)<\infty. \tag{4-21}$$

Recalling that $(X, \mathfrak{m})$ is isomorphic as a measure space to a spherical suspension over any transport ray of length π, the integral in (4-21) is bounded by

$$\begin{aligned}\int_{[\pi-\varepsilon,\pi]} |1+(N-1)t\cot t|\sin^{N-1}(t)\,dt &= \int_{[0,\varepsilon]} [(N-1)(\pi-s)\cot s-1]\sin^{N-1}(s)\,ds \\ &= (N-1)\pi\int_{[0,\varepsilon]} s^{N-2}\,ds+O(\varepsilon)<\infty,\end{aligned}$$

since $N > 1$. This concludes the proof that the density of ν is in $L^1(X, \mathfrak{m})$. The stronger rigidity statement under the stronger $\mathsf{RCD}(K, N)$ assumption is a direct consequence of the maximal diameter theorem proved in [Ketterer 2015] in the $\mathsf{RCD}(K, N)$-setting. □

Corollary 4.6 and Lemma 4.13 have far-reaching consequences.

Theorem 4.14. *Let* $(X, \mathsf{d}, \mathfrak{m})$ *be an e.n.b. metric measure space satisfying* $\mathsf{MCP}(K, N)$ *for some* $K \in \mathbb{R}$, $N \in (1, \infty)$.

Consider the signed distance function d_v *for some continuous function* $v : X \to \mathbb{R}$ *and the associated disintegration*

$$\mathfrak{m}\llcorner_{X\setminus\{v=0\}} = \int_Q h_\alpha \mathcal{H}^1\llcorner_{X_\alpha}\,\mathfrak{q}(d\alpha).$$

Then $d_v^2 \in D(\mathbf{\Delta})$ *and one element of* $\mathbf{\Delta}(d_v^2)$*, which we denote by* Δd_v^2*, has the representation formula*

$$\Delta d_v^2 = 2(1-d_v(\log h_\alpha)')\mathfrak{m} - 2\int_Q (h_\alpha d_v)[\delta_{a(X_\alpha)}-\delta_{b(X_\alpha)}]\,\mathfrak{q}(d\alpha). \tag{4-22}$$

Moreover Δd_v^2 is a sum of two signed Radon measures and the following comparison results hold true:

$$\Delta d_v^2 \le \nu := 2\mathfrak{m} + 2(N-1)\mathsf{d}(\{v=0\},x)\frac{s'_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}\mathfrak{m}\llcorner_{\{v\ge 0\}} + 2(N-1)\mathsf{d}(\{v=0\},x)\frac{s'_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}\mathfrak{m}\llcorner_{\{v<0\}}, \tag{4-23}$$

$$\begin{aligned}[\Delta d_v^2]^{\mathrm{reg}} &:= 2(1-d_v(\log h_\alpha)')\mathfrak{m} \\ &\ge 2\mathfrak{m} - 2(N-1)\mathsf{d}(\{v=0\},x)\frac{s'_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}\mathfrak{m}\llcorner_{\{v\ge 0\}} \\ &\quad - 2(N-1)\mathsf{d}(\{v=0\},x)\frac{s'_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}\mathfrak{m}\llcorner_{\{v<0\}},\end{aligned} \tag{4-24}$$

where $[\Delta d_v^2]^{\mathrm{reg}}$ is the regular part of Δd_v^2 (i.e., absolutely continuous with respect to $\mathfrak{m}$).

Proof. Fix any compactly supported Lipschitz function $f : X \to \mathbb{R}$ and integrate by parts on each ray X_α to obtain

$$\begin{aligned}&\int_{X_\alpha} d_v(x) f'(x) h_\alpha(x)\,\mathcal{H}^1(dx) \\ &= -\int_{X_\alpha} f(x) d'_v(x) h_\alpha(x)\,\mathcal{H}^1(dx) - \int_{X_\alpha} f(x) d_v(x) h'_\alpha(x)\,\mathcal{H}^1(dx) + (f d_v h_\alpha)(b(X_\alpha)) - (f d_v h_\alpha)(a(X_\alpha)) \\ &= \int_{X_\alpha} f(x) h_\alpha(x)\,\mathcal{H}^1(dx) - \int_{X_\alpha} f(x) d_v(x) h'_\alpha(x)\,\mathcal{H}^1(dx) + (f d_v h_\alpha)(b(X_\alpha)) - (f d_v h_\alpha)(a(X_\alpha)) \\ &= \int_{X_\alpha} f(x)\big(1 - d_v(x)(\log h_\alpha)'(x)\big) h_\alpha(x)\,\mathcal{H}^1(dx) + (f d_v h_\alpha)(b(X_\alpha)) - (f d_v h_\alpha)(a(X_\alpha)).\end{aligned} \tag{4-25}$$

Then considering along each ray X_α the two regions $\{v \ge 0\}$ and $\{v < 0\}$, we notice that (2-13) gives

$$\begin{aligned}-d_v(x)(\log h_\alpha)'(x) &\le \mathsf{d}(\{v=0\},x)(N-1)\frac{s'_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}\chi_{\{v\ge 0\}}(x) \\ &\quad + \mathsf{d}(\{v=0\},x)(N-1)\frac{s'_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}\chi_{\{v<0\}}(x) =: V_\alpha(x).\end{aligned}$$

Hence we can collect the estimates, using Remark 4.12, and obtain

$$\int_{X_\alpha} d_v(x) f'(x) h_\alpha(x)\,\mathcal{H}^1(dx) \le \int_{X_\alpha} (1 + V_\alpha(x)) f(x) h_\alpha(x)\,\mathcal{H}^1(dx),$$

provided f is nonnegative. Thanks to Lemma 4.13,

$$\nu = 2\int_Q (1 + V_\alpha(x))\mathfrak{m}_\alpha\,\mathfrak{q}(d\alpha) = 2(1+V)\mathfrak{m}$$

is a well-defined Radon (possibly signed) measure.

Hence, continuing from (4-25), the expression

$$\Delta d_v^2 := 2\int_Q (h_\alpha - d_v h'_\alpha)\mathcal{H}^1\llcorner_{X_\alpha}\,\mathfrak{q}(d\alpha) + 2\int_Q (h_\alpha d_v)[\delta_{b(X_\alpha)} - \delta_{a(X_\alpha)}]\,\mathfrak{q}(d\alpha), \tag{4-26}$$

once restricted to bounded subsets, defines a Borel measure with values in $\mathbb{R}\cup\{-\infty\}$ which satisfies $\Delta d_v^2 \le \nu$. Now, combining Theorem 3.6 with (4-25) and (4-26), we get

$$\int_X f\,\Delta d_v^2(dx) = 2\int_Q\int_{X_\alpha} d_v(x)f'(x)h_\alpha(x)\,\mathcal{H}^1(dx)\,\mathfrak{q}(d\alpha) = 2\int_{\mathcal{T}_{d_v}} d_v(x)f'(x)\,\mathfrak{m}(dx)$$

for any compactly supported Lipschitz function $f : X \to \mathbb{R}$. Therefore, Corollary 4.6 yields

$$\int_{\mathcal{T}_{d_v}} D^- f(-\nabla d_v^2)\,\mathfrak{m} \le \int_X f\,\Delta d_v^2(dx) \le \int_{\mathcal{T}_{d_v}} D^+ f(-\nabla d_v^2)\,\mathfrak{m}$$

for any compactly supported Lipschitz function $f : X \to \mathbb{R}$. Since $X\setminus\mathcal{T}_{d_v} \subset \{v=0\} = \{d_v = 0\}$, from the locality properties of differentials (see [Gigli 2015, equation (3.7)]) we can turn the previous inequalities into

$$\int_X D^- f(-\nabla d_v^2)\,\mathfrak{m} \le \int_X f\,\Delta d_v^2(dx) \le \int_X D^+ f(-\nabla d_v^2)\,\mathfrak{m}, \tag{4-27}$$

valid for any compactly supported Lipschitz function $f : X \to \mathbb{R}$. In order to show that $d_v^2 \in D(\boldsymbol{\Delta})$ with $\Delta d_v^2 \in \boldsymbol{\Delta}(d_v^2)$, we are thus left to prove that Δd_v^2 is a signed Radon measure.

We now claim that Δd_v^2 is a sum of two Radon measures over X. Since $\Delta d_v^2 \le \nu$ with ν a signed Radon measure, thanks to the Riesz–Markov–Kakutani representation theorem it is enough to show that Δd_v^2 defines a Radon functional.

To this aim, fix a compact subset $W \subset X$ and fix a compactly supported Lipschitz cutoff function $\chi_W : X \to [0,1]$ satisfying $\chi_W \equiv 1$ on W. First observe that, using (4-27), for any Lipschitz function $f : X \to \mathbb{R}$ with $\mathrm{supp}(f) \subset W$ we have

$$\left|\int_X \chi_W\,\Delta d_v^2(dx)\right| \le 2\big(\max_{x\in\mathrm{supp}(\chi_W)} d_v(x)\big)\,\mathrm{Lip}(\chi_W)\mathfrak{m}(\mathrm{supp}(\chi_W)) \in (0,\infty),$$
$$\left|\int_X (f\chi_W)\,\Delta d_v^2(dx)\right| \le 2\big(\max_{x\in\mathrm{supp}(\chi_W)} d_v(x)\big)\,\mathrm{Lip}(f\chi_W)\mathfrak{m}(\mathrm{supp}(\chi_W)) \in (0,\infty).$$

Thus for any Lipschitz function $f : X \to \mathbb{R}$ with $\mathrm{supp}(f) \subset W$, using that $\Delta d_v^2 \le \nu \le \nu^+$, on one hand we have

$$\begin{aligned}\int_X f\,\Delta d_v^2 &= -\int_X (\max f - f)\chi_W\,\Delta d_v^2 + \int_X (\max f)\chi_W\,\Delta d_v^2\\ &\ge -\int_X (\max f - f)\chi_W\nu^+ - C_W(\max f),\end{aligned} \tag{4-28}$$

where $C_W := 2(\mathrm{Lip}\,\chi_W)\max_{x\in\mathrm{supp}(\chi_W)} \mathsf{d}_p(x)\mathfrak{m}(\mathrm{supp}(\chi_W)) \in (0,\infty)$ depends only on χ_W.

On the other hand,

$$\begin{aligned}\int_X f\,\Delta d_v^2 &= \int_X f^+\,\Delta d_v^2 - \int_X f^-\,\Delta d_v^2 \\ &\le \int_X f^+ \nu^+ + \int_X (\max f^- - f^-)\chi_W \nu^+ + C_W(\max f^-) \\ &\le \max|f|\big(\nu^+(W) + \nu^+(\mathrm{supp}(\chi_W)) + C_W\big). \end{aligned} \tag{4-29}$$

The combination of (4-28) and (4-29) gives that, for every compact subset $W \subset X$, there exists a constant $C'_W = 2(\nu^+(\mathrm{supp}(\chi_W)) + \mathrm{Lip}\,\chi_W \max_{x\in\mathrm{supp}(\chi_W)} d_v(x)\mathfrak{m}(\mathrm{supp}(\chi_W))) \in (0,\infty)$ such that

$$\left|\int_X f\,\Delta d_v^2\right| \le C'_W \max|f|$$

for every Lipschitz function $f : X \to \mathbb{R}$ with $\mathrm{supp}(f) \subset W$, showing that Δd_v^2 is a Radon functional and thus $d_v^2 \in D(\mathbf{\Delta})$ with $\Delta d_v^2 \in \mathbf{\Delta}(d_v^2)$.

In order to complete the proof we are left with showing (4-24): again from (2-13)

$$\begin{aligned} -d_v(x)(\log h_\alpha)'(x) \ge -(N-1)\mathsf{d}(\{v=0\},x)\frac{s'_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}\chi_{\{v\ge 0\}}(x) \\ -(N-1)\mathsf{d}(\{v=0\},x)\frac{s'_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}\chi_{\{v<0\}}(x), \end{aligned}$$

and the claim is proved. □

Remark 4.15. • In the case X is bounded, in the proof of Theorem 4.14 one can pick $W = X$ and $\chi_W \equiv 1$, giving that the total variation of Δd_v^2 is bounded by $\|\Delta d_v^2\| \le 2\nu^+(X)$.

• Theorem 1.3 can be proved using Corollary 3.7 in the proof of Theorem 4.14 and following verbatim the arguments. Uniqueness of the representation of the Laplacian follows then from infinitesimal Hilbertianity of smooth manifolds.

The representation formula for the Laplacian of the signed distance function on $X \setminus \{v=0\}$ follows from Theorem 4.14 by the chain rule [Gigli 2015, Proposition 4.11].

Corollary 4.16. *Let $(X, \mathsf{d}, \mathfrak{m})$ be an e.n.b. metric measure space satisfying $\mathsf{MCP}(K,N)$ for some $K \in \mathbb{R}$, $N \in (1,\infty)$.*

Consider the signed distance function d_v for some continuous function $v : X \to \mathbb{R}$ and the associated disintegration

$$\mathfrak{m}\llcorner_{X\setminus\{v=0\}} = \int_Q h_\alpha \mathcal{H}^1\llcorner_{X_\alpha}\,\mathfrak{q}(d\alpha).$$

Then

(1) *$|d_v| \in D(\mathbf{\Delta}, X\setminus\{v=0\})$ and one element of $\mathbf{\Delta}(|d_v|)\llcorner_{X\setminus\{v=0\}}$, which we denote by $\Delta|d_v|\llcorner_{X\setminus\{v=0\}}$, is the Radon functional on $X\setminus\{v=0\}$ with the representation formula*

$$\Delta|d_v|\llcorner_{X\setminus\{v=0\}} = -\,\mathrm{sgn}(v)(\log h_\alpha)'\mathfrak{m}\llcorner_{X\setminus\{v=0\}} - \int_Q (h_\alpha[\delta_{a(X_\alpha)\cap\{v>0\}} + \delta_{b(X_\alpha)\cap\{v<0\}}]\,\mathfrak{q}(d\alpha). \tag{4-30}$$

Moreover the following comparison results hold true:

$$\Delta|d_v|\llcorner_{X\setminus\{v=0\}} \le (N-1)\frac{s'_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}\mathfrak{m}\llcorner_{\{v>0\}} + (N-1)\frac{s'_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}\mathfrak{m}\llcorner_{\{v<0\}}, \tag{4-31}$$

$$\begin{aligned}[\Delta|d_v|\llcorner_{X\setminus\{v=0\}}]^{\mathrm{reg}} &:= -\operatorname{sgn}(v)(\log h_\alpha)'\mathfrak{m}\llcorner_{X\setminus\{v=0\}} \\ &\ge -(N-1)\frac{s'_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}\mathfrak{m}\llcorner_{\{v>0\}} \\ &\quad -(N-1)\frac{s'_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}\mathfrak{m}\llcorner_{\{v<0\}},\end{aligned} \tag{4-32}$$

where $[\Delta|d_v|\llcorner_{X\setminus\{v=0\}}]^{\mathrm{reg}}$ *is the regular part of* $\Delta|d_v|\llcorner_{X\setminus\{v=0\}}$ (*i.e., absolutely continuous with respect to* $\mathfrak{m}$).

(2) $d_v \in D(\mathbf{\Delta}, X\setminus\{v=0\})$ *and one element of* $\mathbf{\Delta}(d_v)\llcorner_{X\setminus\{v=0\}}$, *which we denote by* $\Delta d_v\llcorner_{X\setminus\{v=0\}}$, *is the Radon functional on* $X\setminus\{v=0\}$ *with the representation formula*

$$\Delta d_v\llcorner_{X\setminus\{v=0\}} = -(\log h_\alpha)'\mathfrak{m}\llcorner_{X\setminus\{v=0\}} - \int_Q (h_\alpha[\delta_{a(X_\alpha)\cap\{v>0\}} - \delta_{b(X_\alpha)\cap\{v<0\}}]\,\mathfrak{q}(d\alpha). \tag{4-33}$$

Moreover the following comparison results hold true:

$$\Delta d_v\llcorner_{X\setminus\{v=0\}} \le (N-1)\frac{s'_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{b(X_\alpha)}(x))}\mathfrak{m}\llcorner_{X\setminus\{v=0\}} + \int_Q h_\alpha\delta_{b(X_\alpha)\cap\{v<0\}}\,\mathfrak{q}(d\alpha), \tag{4-34}$$

$$\Delta d_v\llcorner_{X\setminus\{v=0\}} \ge -(N-1)\frac{s'_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}\mathfrak{m}\llcorner_{X\setminus\{v=0\}} - \int_Q (h_\alpha[\delta_{a(X_\alpha)\cap\{v>0\}}]\,\mathfrak{q}(d\alpha). \tag{4-35}$$

Proof. Writing $\operatorname{sgn}(v)d_v = \sqrt{d_v^2}$, a direct application of the chain rule [Gigli 2015, Proposition 4.11] combined with Theorem 4.14 gives that $|d_v| \in D(\mathbf{\Delta}, X\setminus\{v=0\})$ and that $\Delta|d_v|$ defined in (4-30) is an element of $\mathbf{\Delta}|d_v|\llcorner_{X\setminus\{v=0\}}$. The comparison results (4-31), (4-32) follow from the definition (4-30) together with (2-13).

Since $d_v = \operatorname{sgn}(v)\,|d_v|$, it is clear that $d_v \in D(\mathbf{\Delta}, X\setminus\{v=0\})$, with

$$\mathbf{\Delta}(d_v)\llcorner_{X\setminus\{v=0\}} = \operatorname{sgn}(v)\,\mathbf{\Delta}(|d_v|)\llcorner_{X\setminus\{v=0\}};$$

thus $\Delta d_v\llcorner_{X\setminus\{v=0\}}$defined in (4-33) is an element of $\mathbf{\Delta}(d_v)\llcorner_{X\setminus\{v=0\}}$ and the comparison results (4-34), (4-35) follow again from (2-13). ☐

We now specialise the above results to the distance function from a point $p \in X$; i.e., we pick $v = \mathsf{d}_p$ so that $\{v=0\} = p$ and $v \ge 0$ everywhere. Note that, in this case, $b(X_\alpha) = p$ for $\mathfrak{q}$-a.e. $\alpha \in Q$.

Corollary 4.17. *Let* $(X, \mathsf{d}, \mathfrak{m})$ *be an e.n.b. metric measure space satisfying* $\mathsf{MCP}(K, N)$ *for some* $K \in \mathbb{R}$, $N \in (1, \infty)$. *Fix* $p \in X$, *consider* $\mathsf{d}_p := \mathsf{d}(p, \cdot)$ *and the associated disintegration*

$$\mathfrak{m} = \int_Q h_\alpha \mathcal{H}^1 \llcorner X_\alpha \, \mathfrak{q}(d\alpha).$$

Then $\mathsf{d}_p^2 \in D(\mathbf{\Delta})$ *and one element of* $\mathbf{\Delta}(\mathsf{d}_p^2)$, *which we denote by* $\Delta \mathsf{d}_p^2$, *is a sum of two signed Radon measures and satisfies the representation formula*

$$\Delta \mathsf{d}_p^2 = 2(1 - \mathsf{d}_p (\log h_\alpha)')\mathfrak{m} - 2 \int_Q h_\alpha \mathsf{d}_p \delta_{a(X_\alpha)} \, \mathfrak{q}(d\alpha). \tag{4-36}$$

Moreover, the following comparison results hold true:

$$\Delta \mathsf{d}_p^2 \le \nu := 2\left(1 + (N-1)\mathsf{d}_p(x) \frac{s'_{K/(N-1)}(\mathsf{d}_p(x))}{s_{K/(N-1)}(\mathsf{d}_p(x))}\right)\mathfrak{m}, \tag{4-37}$$

$$[\Delta \mathsf{d}_p^2]^{\mathrm{reg}} := 2(1 - \mathsf{d}_p (\log h_\alpha)')\mathfrak{m} \ge 2\left(1 - (N-1)\mathsf{d}_p \frac{s'_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}\right)\mathfrak{m}, \tag{4-38}$$

where $[\Delta \mathsf{d}_p^2]^{\mathrm{reg}}$ *is the regular part of* $\Delta \mathsf{d}_p^2$ (*i.e., absolutely continuous with respect to* $\mathfrak{m}$).

Remark 4.18 (on the lower bound (4-38)). Denote by $\mathcal{C}_p := \{a(X_\alpha)\}_{\alpha \in Q}$ the cut locus of p. Then for every $\varepsilon > 0$ there exists $C_\varepsilon > 0$ so that for every bounded subset $W \subset X$ it holds

$$\begin{aligned}[\Delta \mathsf{d}_p^2]^{\mathrm{reg}}_{\llcorner W} &\ge 2\left(1 - (N-1)\mathsf{d}_p \frac{s'_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}\right)\mathfrak{m}_{\llcorner W} \\ &\ge -C_{\varepsilon, W}\mathfrak{m}_{\llcorner W} \quad \text{on } W \cap \{x = g_t(a_\alpha) : t \ge \varepsilon\} \supset W \cap \{x \in X : \mathsf{d}(x, \mathcal{C}_p) \ge \varepsilon\}.\end{aligned}$$

The representation formula for the Laplacian of the distance function follows from Corollary 4.17 by the chain rule [Gigli 2015, Proposition 4.11], writing $\mathrm{sgn}(v) d_v = \sqrt{d_v^2}$.

Corollary 4.19. *Let* $(X, \mathsf{d}, \mathfrak{m})$ *be an e.n.b. metric measure space satisfying* $\mathsf{MCP}(K, N)$ *for some* $K \in \mathbb{R}$, $N \in (1, \infty)$. *Fix* $p \in X$, *consider* $\mathsf{d}_p := \mathsf{d}(p, \cdot)$ *and the associated disintegration*

$$\mathfrak{m} = \int_Q h_\alpha \mathcal{H}^1 \llcorner_{X_\alpha} \, \mathfrak{q}(d\alpha).$$

Then $\mathsf{d}_p \in D(\mathbf{\Delta}, X \setminus \{p\})$ *and one element of* $\mathbf{\Delta} \mathsf{d}_{p \llcorner X\setminus\{p\}}$, *which we denote by* $\Delta \mathsf{d}_{p \llcorner X\setminus\{p\}}$, *is a Radon functional with the representation formula*

$$\Delta \mathsf{d}_{p \llcorner X\setminus\{p\}} = -(\log h_\alpha)'\mathfrak{m} - \int_Q h_\alpha \delta_{a(X_\alpha)} \, \mathfrak{q}(d\alpha). \tag{4-39}$$

Moreover, the following comparison results hold true:

$$\Delta \mathsf{d}_{p \llcorner X\setminus\{p\}} \le (N-1) \frac{s'_{K/(N-1)}(\mathsf{d}_p(x))}{s_{K/(N-1)}(\mathsf{d}_p(x))}\mathfrak{m}, \tag{4-40}$$

$$[\Delta \mathsf{d}_{p \llcorner X\setminus\{p\}}]^{\mathrm{reg}} := -(\log h_\alpha)'\mathfrak{m} \ge -(N-1) \frac{s'_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}{s_{K/(N-1)}(\mathsf{d}_{a(X_\alpha)}(x))}\mathfrak{m}, \tag{4-41}$$

where $[\Delta \mathsf{d}_{p \llcorner X\setminus\{p\}}]^{\mathrm{reg}}$ *is the regular part of* $\Delta \mathsf{d}_{p \llcorner X\setminus\{p\}}$ (*i.e., absolutely continuous with respect to* $\mathfrak{m}$).

Remark 4.20. Corollary 4.19 should be compared with [Gigli 2015, Corollary 5.15, Remark 5.16], where it was proved that $\mathsf{d}_p \in D(\mathbf{\Delta}, X \setminus \{p\})$ together with the upper bound (4-40) under the assumption that $(X, \mathsf{d}, \mathfrak{m})$ is an infinitesimally strictly convex $\mathsf{MCP}(K, N)$-space.

Let us stress that, by the very definition, the Laplacian in the infinitesimally strictly convex setting is single-valued, simplifying the treatment.

One novelty of Corollary 4.19 is that the infinitesimal strict convexity is replaced by the essentially nonbranching property which, a priori, does not exclude a multivalued Laplacian. In addition to that, the geometrically new content of Corollary 4.19 when compared with [Gigli 2015] is that it contains an *exact representation formula* (4-39) which also gives the new lower bound (4-41).

Part II. Applications

In Part II of the paper we collect all the main applications of the results obtained in Part I.

5. The singular part of the Laplacian

In order to state the next corollary recall that from essentially nonbranching and $\mathsf{MCP}(K, N)$ it follows that for every fixed $p \in X$ and $\mathfrak{m}$-a.e. $x \in X$ (precisely on $\mathcal{T}^{\mathrm{nb}}_{\mathsf{d}_p}$) there exists a unique geodesic γ^x starting from x and arriving at p, i.e., $\gamma^x_0 = x$ and $\gamma^x_1 = p$. For each $t \in [0, 1]$, define the map

$$T_t : \mathcal{T}^{\mathrm{nb}}_{\mathsf{d}_p} \to \mathcal{T}^{\mathrm{nb}}_{\mathsf{d}_p}, \quad T_t(x) := \gamma^x_t. \tag{5-1}$$

It is worth noting that T_t is also the W_2-optimal transport map from the (renormalised) ambient measure $\mathfrak{m}$ to δ_p, provided $\mathfrak{m}(X) < \infty$.

The goal of the next proposition is to get some refined information on the cut locus $\mathcal{C}_p$ of p; more precisely, we infer an upper bound on an optimal transport-type Minkowski content of $\mathcal{C}_p$.

Proposition 5.1. *Let $(X, \mathsf{d}, \mathfrak{m})$ be an e.n.b. metric measure space satisfying $\mathsf{MCP}(K, N)$ for some $K \in \mathbb{R}$, $N \in (1, \infty)$. Fix any point $p \in X$ and consider for each $t \in [0, 1]$ the map T_t defined by* (5-1).

Then, for every bounded open subset $W \subset X$ it holds

$$\limsup_{\varepsilon \downarrow 0} \frac{\mathfrak{m}((X \setminus T_\varepsilon(X)) \cap W)}{\varepsilon} \le \|[\Delta \mathsf{d}_p^2]_{\mathrm{sing}}\|(W) < \infty. \tag{5-2}$$

Remark 5.2 (geometric meaning of Proposition 5.1). Fix $p \in X$, and consider $\mathsf{d}_p := \mathsf{d}(p, \cdot)$ and the associated disintegration $\mathfrak{m} = \int_Q h_\alpha \mathcal{H}^1 \llcorner_{X_\alpha} \mathfrak{q}(d\alpha)$. Then the cut locus $\mathcal{C}_p$ of p coincides with the set of initial points $\{a(X_\alpha)\}_{\alpha \in Q}$ of the transport rays. The set $X \setminus T_\varepsilon(X)$ thus can be seen as an "optimal transport neighbourhood" of the cut locus $\mathcal{C}_p$ and therefore (5-2) gives an optimal transport-type estimate on a weak version of the codimension-1 Minkowski content of $\mathcal{C}_p$.

Since the cut locus of a point in an e.n.b. $\mathsf{MCP}(K, N)$ space can be dense (this can be the case already for the boundary of a convex body in $\mathbb{R}^3$), one cannot expect an upper bound on the classical codimension-1 Minkowski content of $\mathcal{C}_p$. The bound (5-2) looks interesting already in the classical setting of a smooth Riemannian manifold. Indeed it is well known that $\mathcal{C}_p$ is rectifiable with locally finite codimension-1

Hausdorff measure (see for instance [Mantegazza and Mennucci 2003]), but in the literature it seems not to be present any (local) bound on its codimension-1 Minkowski content.

Proof. If X is bounded, one can choose $W = X$ and the proof is easier (there is no need to introduce an intermediate set U in the arguments below); we thus discuss directly the case when X is not bounded.

Let $U \supset W$ be a bounded open subset such that W is compactly contained in U, in particular $\mathsf{d}(W, X \setminus U) > 0$.

With a slight abuse of notation, for ease of writing, in the next computations we identify the ray $(X_\alpha, \mathsf{d}, \mathfrak{m}_\alpha)$ with the real interval $((a_\alpha, b_\alpha), |\cdot|, h_\alpha \mathcal{L}^1)$ isomorphic to it as an m.m.s.

Recalling from Remark 2.14 that $h_\alpha : X_\alpha \simeq (a_\alpha, b_\alpha) \to \mathbb{R}^+$ is continuous up to the initial point a_α, it is clear that

$$h_\alpha(a(X_\alpha))\, \mathsf{d}_p(a(X_\alpha)) = \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \int_{[a_\alpha, a_\alpha + \varepsilon |X_\alpha|]} h_\alpha(s)\, ds,$$

where $|X_\alpha|$ denotes the length of the transport ray X_α, i.e., $|X_\alpha| = \mathsf{d}(a(X_\alpha), b(X_\alpha)) = \mathsf{d}(a(X_\alpha), p)$. Hence, for any bounded open subset $U \subset X$ it holds

$$\begin{aligned} \|[\Delta \mathsf{d}_p^2]^{\mathrm{sing}}\|(U) &= \int_{\{\alpha \in Q : a(X_\alpha) \in U\}} (h_\alpha \mathsf{d}_p)(a(X_\alpha))\, \mathfrak{q}(d\alpha) \\ &= \int_{\{\alpha \in Q : a(X_\alpha) \in U\}} \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{[a_\alpha, a_\alpha + \varepsilon |X_\alpha|]} h_\alpha(s)\, ds\, \mathfrak{q}(d\alpha), \end{aligned}$$

where $\|[\Delta \mathsf{d}_p^2]^{\mathrm{sing}}\|(U)$ denotes the total variation measure of U. Since by Corollary 4.17 we know that $\|[\Delta \mathsf{d}_p^2]^{\mathrm{sing}}\|(U) < \infty$, by Fatou's lemma we infer

$$\limsup_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \int_{\{\alpha \in Q : a(X_\alpha) \in U\}} \int_{[a_\alpha, a_\alpha + \varepsilon |X_\alpha|]} h_\alpha(s)\, ds\, \mathfrak{q}(d\alpha) \le \|[\Delta \mathsf{d}_p^2]^{\mathrm{sing}}\|(U) < \infty. \tag{5-3}$$

We then look for a more convenient expression of the left-hand side of the previous inequality. First, note that for ε sufficiently small such that

$$\varepsilon/(1-\varepsilon) < \frac{\mathsf{d}(W, X \setminus U)}{\mathsf{d}_p(W)}$$

it holds

$$\int_Q \int_{[a_\alpha, a_\alpha + \varepsilon |X_\alpha|] \cap W} h_\alpha(s)\, ds\, \mathfrak{q}(d\alpha) \le \int_{\{\alpha \in Q : a(X_\alpha) \in U\}} \int_{[a_\alpha, a_\alpha + \varepsilon |X_\alpha|]} h_\alpha(s)\, ds\, \mathfrak{q}(d\alpha). \tag{5-4}$$

Recalling the definition of the map T_t given in (5-1), we now claim that

$$\mathfrak{m}((X \setminus T_\varepsilon(X)) \cap W) = \int_Q \int_{[a_\alpha, a_\alpha + \varepsilon |X_\alpha|] \cap W} h_\alpha(s)\, ds\, \mathfrak{q}(d\alpha). \tag{5-5}$$

Indeed, on the one hand, by the disintegration theorem, Theorem 3.6, we know that

$$\mathfrak{m}((X \setminus T_\varepsilon(X)) \cap W) = \int_Q \int_{X_\alpha \cap (X \setminus T_\varepsilon(X)) \cap W} h_\alpha(s)\, ds\, \mathfrak{q}(d\alpha).$$

On the other hand, since trivially

$$X_\alpha \cap (X \setminus T_t(X)) \cap W = X_\alpha \setminus T_t(X) \cap W,$$

and since, as T_t is translating along $\mathcal{T}^{\mathrm{nb}}_{\mathsf{d}_p}$, one has $X_\alpha \setminus T_t(X) = X_\alpha \setminus T_t(X_\alpha)$, we obtain

$$X_\alpha \cap (X \setminus T_t(X)) \cap W = X_\alpha \setminus T_t(X_\alpha) \cap W.$$

The claim (5-5) follows. The combination of (5-3), (5-4) and (5-5) gives

$$\limsup_{\varepsilon \downarrow 0} \frac{\mathfrak{m}((X \setminus T_\varepsilon(X)) \cap W)}{\varepsilon} \le \|[\Delta \mathsf{d}_p^2]^{\mathrm{sing}}\|(U) < \infty$$

for every U bounded open subset compactly containing the open set W. Since from Theorem 4.14 we know that $\Delta \mathsf{d}_p^2$ is a Radon measure, the thesis (5-2) follows. □

We next give some sufficient condition implying that the densities h_α, given by the disintegration theorem, Theorem 3.6, are null at the final points.

Lemma 5.3. *Let* $(X, \mathsf{d}, \mathfrak{m})$ *be an e.n.b.* $\mathsf{MCP}(K, N)$ *space for some* $K \in \mathbb{R}$, $N \in (1, \infty)$.

Let $u = \mathsf{d}_p = \mathsf{d}(p, \cdot)$ *for some* $p \in X$ *and consider the disintegration associated to* d_p:

$$\mathfrak{m} = \int_Q h_\alpha \mathcal{H}^1 \llcorner_{X_\alpha} \mathfrak{q}(d\alpha).$$

Assume there exists $s > 1$ *such that*

$$\liminf_{r \downarrow 0} \frac{\mathfrak{m}(B_r(p))}{r^s} < \infty; \tag{5-6}$$

then $h_\alpha(p) = 0$ *for* $\mathfrak{q}$*-a.e.* $\alpha \in Q$.

More generally, for any 1*-Lipschitz function* u*, denoting by*

$$\mathfrak{m}\llcorner_{\mathcal{T}_u^{\mathrm{nb}}} = \int_Q h_\alpha \mathcal{H}^1 \llcorner_{X_\alpha} \mathfrak{q}(d\alpha)$$

the associated disintegration, it holds that

$$\left\| \int_Q h_\alpha(a(X_\alpha)) \delta_{a(X_\alpha)}\, \mathfrak{q}(d\alpha) \right\| \le \liminf_{r \downarrow 0} \frac{\mathfrak{m}\big(\bigcup_\alpha [a(X_\alpha), a(X_\alpha) + r]\big)}{r} = \beta \in [0, +\infty], \tag{5-7}$$

where the leftmost term is the total variation of the corresponding measure.

Proof. Suppose by contradiction the claim was false, i.e., there exists $\bar{Q} \subset Q$ where $h_\alpha(p) \ge c > 0$, with $\mathfrak{q}(\bar{Q}) > 0$. Observe that a.e. transport ray X_α ends in p, i.e., $b(X_\alpha) = p$ for $\mathfrak{q}$-a.e. $\alpha \in Q$. As usual, we identify the transport ray X_α with the real interval $[a_\alpha, b_\alpha]$ (the cases of semiclosed and open intervals are analogous). Then by Fatou's lemma it holds

$$\begin{aligned}\liminf_{r \downarrow 0} \frac{\mathfrak{m}(B_r(p))}{r^s} &\ge \liminf_{r \downarrow 0} \int_{\bar{Q}} \frac{1}{r^s} \int_{[b_\alpha - r, b_\alpha]} h_\alpha(t)\, dt\, \mathfrak{q}(d\alpha) \\ &\ge \int_{\bar{Q}} \liminf_{r \downarrow 0} \frac{1}{r} \int_{[b_\alpha - r, b_\alpha]} \frac{h_\alpha(t)}{r^{s-1}}\, dt\, \mathfrak{q}(d\alpha) = \infty,\end{aligned}$$

giving a contradiction and proving the claim.

The second part of the lemma follows along analogous arguments. □

Remark 5.4. If $(X, \mathsf{d}, \mathfrak{m})$ is an $\mathsf{RCD}(K, N)$ space not isometric to a circle or to a (possibly unbounded) real interval then (5-6) is satisfied for $\mathfrak{m}$-a.e. $p \in X$.

Indeed if $(X, \mathsf{d}, \mathfrak{m})$ is an $\mathsf{RCD}(K, N)$ space, using the rectifiability result [Mondino and Naber 2019, Theorem 1.1] (see also [Gigli et al. 2015a] and compare with [Cheeger and Colding 1997; 2000a; 2000b]) together with the absolute continuity of the reference measure $\mathfrak{m}$ with the respect to the Hausdorff measure of the bi-Lipschitz charts obtained independently in [Kell and Mondino 2018, Theorem 1.2] and [Gigli and Pasqualetto 2016, Theorem 3.5], it follows that for $\mathfrak{m}$-a.e. $p \in X$ there exists $n = n(p) \in \mathbb{N} \cap [1, \infty)$ such that

$$\liminf_{r \downarrow 0} \frac{\mathfrak{m}(B_r(p))}{r^n} < \infty.$$

If moreover we assume (X, d) not to be isometric to a circle or to a (possibly unbounded) real interval, then by [Kitabeppu and Lakzian 2016] it follows that $n(p) > 1$ for $\mathfrak{m}$-a.e. $p \in X$.

If $(X, \mathsf{d}, \mathfrak{m})$ is an $\mathsf{MCP}(K, N)$ space then the validity of (5-6) is not known.

6. $\mathsf{CD}(K, N)$ is equivalent to a (K, N)-Bochner-type inequality

The Bochner inequality is one of the most fundamental estimates in geometric analysis. For a smooth N-dimensional Riemannian manifold (M, g) with $\mathrm{Ricci}_g \geq Kg$, for some $K \in \mathbb{R}$, it states that for any smooth function $u \in C^3(M)$ it holds

$$\frac{1}{2}\Delta|\nabla u|^2 - \langle \nabla u, \nabla \Delta u\rangle \geq K|\nabla u|^2 + |\nabla^2 u|^2 \geq K|\nabla u|^2 + \frac{1}{N}(\Delta u)^2,$$

where $|\nabla^2 u|^2$ is the Hilbert–Schmidt norm of the Hessian matrix $\nabla^2 u$ and the rightmost inequality follows directly by the Cauchy–Schwarz inequality. Note in particular that if u is a distance function, then on an open dense set of full measure, $|\nabla u|^2 = 1$ and the Hessian is a block matrix with vanishing slot in the direction of the "gradient of the distance"; in particular, for a distance function the inequality can be improved to

$$-\langle \nabla u, \nabla \Delta u\rangle \geq K + \frac{1}{N-1}(\Delta u)^2 \quad \text{a.e.} \tag{6-1}$$

Finally, note that the term $-\langle \nabla u, \nabla \Delta u\rangle$ corresponds to "the derivative of Δu in the direction of $-\nabla u$"; thus, if we consider the transport set associated to u, such a term would correspond to what we denoted as $(\Delta u)'$. Since in a general m.m.s. it is not clear there is enough regularity to write $(\Delta u)'$, it is natural to consider the following version of (6-1) "integrated along transport rays":

$$\Delta u(g_t(x)) - \Delta u(x) \geq Kt + \frac{1}{N-1}\int_{(0,t)} (\Delta u)^2(g_s(x))\, ds. \quad \text{a.e. } x, t. \tag{6-2}$$

This is the (K, N)-Bochner inequality that will be proved to be equivalent to the $\mathsf{CD}(K, N)$ condition.

In order to state the results, it is useful to recall that given a 1-Lipschitz function u on an e.n.b. $\mathsf{CD}(K, N)$ space there is a natural disintegration of $\mathfrak{m}$ restricted to the transport set $\mathcal{T}_u^{\mathrm{nb}}$ (see Theorem 3.6):

$$\mathfrak{m}\llcorner_{\mathcal{T}_u^{\mathrm{nb}}} = \int_Q h_\alpha \mathcal{H}^1 \llcorner_{X_\alpha}\, \mathfrak{q}(d\alpha). \tag{6-3}$$

We will define $\mathrm{int}(\mathcal{T}_u^{\mathrm{nb}}) := \bigcup_{\alpha\in Q} \mathring{X}_\alpha$, where $\mathring{X}_\alpha$ stands for the relative interior of X_α; it can also be identified by isometry with the open interval (a_α, b_α).

The function h_α in (6-3) is a $\mathsf{CD}(K, N)$ density on (a_α, b_α), so in particular it is semiconcave; thus if D_α is the set of differentiability points of h_α, then $(a_\alpha, b_\alpha) \setminus D_\alpha$ is countable.

Our next result roughly states that the (K, N)-Bochner-type inequality (6-2) holds for those 1-Lipschitz functions for which we have found an explicit representation formula for the Laplacian, namely those 1-Lipschitz functions satisfying the hypothesis of Theorem 4.8 and for any distance function with sign d_v. Recall that, for any u belonging to these classes of functions, Δu outside of the initial and final points of transport rays forming $\mathcal{T}_u^{\mathrm{nb}}$ is absolutely continuous with respect to $\mathfrak{m}$.

Theorem 6.1 ($\mathsf{CD}_{\mathrm{loc}}(K, N)$+e.n.b.$\Longrightarrow(K, N)$-Bochner-type inequality). *Let $(X, \mathsf{d}, \mathfrak{m})$ be an e.n.b. metric measure space satisfying $\mathsf{CD}_{\mathrm{loc}}(K, N)$. Then the following hold*:

(1) *Let $u : X \to \mathbb{R}$ be any 1-Lipschitz function such that $\int_Q |X_\alpha|^{-1}\, \mathfrak{q}(d\alpha) < \infty$. Then for $\mathfrak{q}$-a.e. $\alpha \in Q$, for each $x \in X_\alpha$ it holds*

$$\Delta u(g_t(x)) - \Delta u(x) \geq Kt + \frac{1}{N-1} \int_{(0,t)} (\Delta u)^2 (g_s(x))\, ds \tag{6-4}$$

for all $t \in \mathbb{R}$ such that $g_t(x) \in \mathcal{T}_u$, up to a countable set depending only on α.

(2) *Let $u = d_v$ be a signed distance function. Then for $\mathfrak{q}$-a.e. $\alpha \in Q$, for each $x \in \mathring{X}_\alpha \setminus \{v = 0\}$ the (K, N)-Bochner-type inequality (6-4) holds for all $t \in \mathbb{R}$ such that $g_t(x) \in \mathring{X}_\alpha \setminus \{v = 0\}$ and $\mathrm{sgn}(d_v(x)) = \mathrm{sgn}(d_v(g_t(x)))$, provided the densities $\Delta d_v(x)$ and $\Delta d_v(g_t(x))$ exist.*

Proof. We prove just (1), the proof of (2) being completely analogous (using Corollary 4.16 in place of Theorem 4.8).

Fix $\alpha \in Q$ and $x \in \mathrm{int}(R_u^{\mathrm{nb}}(\alpha)) = (a_\alpha, b_\alpha)$ for which the representation of Δu given by Theorem 4.8 is valid:

$$\Delta u(x) = -(\log h_\alpha)'(x).$$

In particular h_α is differentiable at x. As observed above, for each α, $\Delta u(x)$ is defined on $D_\alpha \subset (a_\alpha, b_\alpha)$, with $(a_\alpha, b_\alpha) \setminus D_\alpha$ countable. Therefore the claim reduces to showing for $\mathfrak{q}$-a.e. $\alpha \in Q$ that

$$(\log h_\alpha)'(x) - (\log h_\alpha)'(g_t(x)) \geq Kt + \frac{1}{N-1} \int_{(0,t)} ((\log h_\alpha)'(g_s(x)))^2\, ds, \tag{6-5}$$

whenever $x, g_t(x) \in D_\alpha$. To prove (6-5), consider a nonnegative C^2 function ψ supported on $[-1, 1]$ with $\int \psi = 1$. Let $\psi_\varepsilon(x) := \psi(x/\varepsilon)$; of course ψ_ε is supported on $[-\varepsilon, \varepsilon]$ with $\int \psi_\varepsilon = 1$. Define the function h_α^ε on $(a_\alpha + \varepsilon, b_\alpha - \varepsilon)$ by

$$\log h_\alpha^\varepsilon := \log h_\alpha * \psi_\varepsilon. \tag{6-6}$$

Since by Theorem 3.6 we know that h_α is a $\mathsf{CD}(K, N)$ density, and h_α^ε is a C^2-smooth $\mathsf{CD}(K, N)$ density on $(a_\alpha + \varepsilon, b_\alpha - \varepsilon)$ by Proposition 2.18; in particular (6-5) is satisfied by h_α^ε. Taking the limit as $\varepsilon \to 0$, we obtain that $(\log h_\alpha^\varepsilon)' \to (\log h_\alpha)'$ pointwise on D_α and in $L^1((a_\alpha, b_\alpha))$. Thus we can pass into the limit as $\varepsilon \to 0$ in (6-5) and get that it is also satisfied by h_α. □

Also the converse implication holds, giving a complete equivalence between the (K, N)-Bochner-type inequality (6-4) on signed distance functions and the $\mathsf{CD}(K, N)$ condition.

Theorem 6.2 ($\mathsf{MCP}(K', N')$ + e.n.b. + (K, N)-Bochner-type inequality $\Longrightarrow \mathsf{CD}(K, N)$). *Let $(X, \mathsf{d}, \mathfrak{m})$ be an e.n.b. metric measure space satisfying $\mathsf{MCP}(K', N')$ for some $K' \in \mathbb{R}$, $N' \in (1, \infty)$, with $\mathfrak{m}(X) < \infty$. Assume that, for every signed distance function $d_v : X \to \mathbb{R}$, for $\mathfrak{q}$-a.e. $\alpha \in Q$, for each $x \in \mathring{X}_\alpha \setminus \{v = 0\}$ it holds*

$$\Delta d_v(g_t(x)) - \Delta d_v(x) \geq Kt + \frac{1}{N-1} \int_{(0,t)} (\Delta d_v)^2(g_s(x))\, ds \tag{6-7}$$

for all $t \in \mathbb{R}$ such that $g_t(x) \in \mathring{X}_\alpha \setminus \{v = 0\}$ and $\operatorname{sgn}(d_v(x)) = \operatorname{sgn}(d_v(g_t(x)))$, provided the densities $\Delta d_v(x)$ and $\Delta d_v(g_t(x))$ exist.

Then $(X, \mathsf{d}, \mathfrak{m})$ satisfies $\mathsf{CD}(K, N)$.

Remark 6.3. We briefly comment on the statement of Theorem 6.2. Using the assumption of e.n.b. and of $\mathsf{MCP}(K', N')$, we deduce from Corollary 4.16 that any $d_v \in D(\mathbf{\Delta}, X \setminus \{v = 0\})$. Therefore, in the assumption (6-7), we consider $\Delta d_v(g_t(x))$ only for those $g_t(x)$ belonging to $\{v > 0\}$ or to $\{v < 0\}$, provided $x \in \{v > 0\}$ or $x \in \{v < 0\}$ respectively.

Let us also comment on the assumptions $\mathsf{CD}_{\text{loc}}(K, N)$ vs. $\mathsf{CD}(K, N)$ and $\mathfrak{m}(X) < \infty$ in the last two results. It was proved in [Cavalletti and Milman 2016] that, under the assumption $\mathfrak{m}(X) < \infty$, an e.n.b. $\mathsf{CD}_{\text{loc}}(K, N)$ space satisfies $\mathsf{CD}(K, N)$ globally; on the other hand the implication is open without the assumption $\mathfrak{m}(X) < \infty$. We thus assumed $\mathsf{CD}_{\text{loc}}(K, N)$ in Theorem 6.1 as, a priori, it is more general and still gives that all the conditional densities h_α are $\mathsf{CD}(K, N)$ densities (see Theorem 3.6).

Proof. We show that, given any 1-Lipschitz function $\varphi : X \to \mathbb{R}$, the conditional probabilities associated to the transport set $\mathcal{T}_\varphi^{\text{nb}}$ of φ satisfy $\mathsf{CD}(K, N)$. From [Cavalletti and Milman 2016] it will then follow that $(X, \mathsf{d}, \mathfrak{m})$ satisfy $\mathsf{CD}(K, N)$.

<u>Step 1</u>: Let us fix $\varphi : X \to \mathbb{R}$ a 1-Lipschitz function and the associated nonbranched transport set $\mathcal{T}_\varphi^{\text{nb}}$. Fix also $c \in \mathbb{R}$, let $\varphi_c := \varphi - c$ and consider the associated signed distance function d_{φ_c} from the level set $\{\varphi = c\}$.

Note that the function d_{φ_c} coincides with φ_c along $(R_\varphi^{\text{nb}})^{-1}(\{\varphi = c\})$, i.e., along each transport ray of φ having nonempty intersection with $\{\varphi = c\}$.

Indeed, fix any $x \in \mathcal{T}_{\varphi_c}^{\text{nb}}$ with $\varphi(x) \geq c$ (the argument for $\varphi(x) \leq c$ is analogous) such that there exists $y \in R_\varphi^{\text{nb}}(x)$ with $\varphi(y) = c$ (i.e., $x \in (R_\varphi^{\text{nb}})^{-1}(\{\varphi = c\})$); then for any other $z \in \{\varphi = c\}$ it holds

$$\mathsf{d}(x, y) = \varphi(x) - \varphi(y) = \varphi(x) - \varphi(z) \leq \mathsf{d}(x, z),$$

showing that $\mathsf{d}(x, y) = d_{\varphi_c}(x)$ and that $d_{\varphi_c}(x) = \varphi(x) - \varphi(y) = \varphi(x) - c = \varphi_c(x)$. Hence if $x \in (R_\varphi^{\text{nb}})^{-1}(\{\varphi = c\})$, then

$$|d_{\varphi_c}(x) - d_c(y)| = \mathsf{d}(x, y)$$

for some $(x, y) \in (R_\varphi^{\text{nb}})$ implying $(x, y) \in R_{d_{\varphi_c}}$. Since a branching structure for d_{φ_c} inside $(R_\varphi^{\text{nb}})^{-1}(\{\varphi = c\})$ will imply a branching structure for φ_c, this implies that on $(R_\varphi^{\text{nb}})^{-1}(\{\varphi = c\})$ the equivalence relation R_φ^{nb} implies $R_{d_{\varphi_c}}^{\text{nb}}$. In particular it follows that $\mathcal{T}_\varphi^{\text{nb}} \cap (R_\varphi^{\text{nb}})^{-1}(\{\varphi = c\}) \subset \mathcal{T}_{d_{\varphi_c}}^{\text{nb}}$.

Step 2: Consider the disintegrations associated to $\mathcal{T}_\varphi^{\mathrm{nb}}$ and to $\mathcal{T}_{d_{\varphi c}}^{\mathrm{nb}}$ via Theorem 3.6:

$$\mathfrak{m}\llcorner_{\mathcal{T}_\varphi^{\mathrm{nb}}} = \int_{Q_\varphi} \mathfrak{m}_{\alpha,\varphi}\, \mathfrak{q}_\varphi(d\alpha), \qquad \mathfrak{m}\llcorner_{\mathcal{T}_{d_{\varphi c}}^{\mathrm{nb}}} = \int_{Q_{d_{\varphi c}}} \mathfrak{m}_{\alpha,d_{\varphi c}}\, \mathfrak{q}_{d_{\varphi c}}(d\alpha),$$

with $\mathfrak{m}_{\alpha,\varphi} = h_{\alpha,\varphi}\mathcal{H}^1\llcorner_{X_{\alpha,\varphi}}$ and $\mathfrak{m}_{\alpha,d_{\varphi c}} = h_{\alpha,d_{\varphi c}}\mathcal{H}^1\llcorner_{X_{\alpha,d_{\varphi c}}}$.

From Step 1 and the uniqueness of the disintegration, it follows that up to a constant factor

$$h_{\alpha,\varphi} = h_{\alpha,d_{\varphi c}} \quad \text{on } X_{\alpha,\varphi},$$

for all those α such that $X_{\alpha,\varphi} \cap \{\varphi = c\} \neq \varnothing$. Moreover from Corollary 4.16 we deduce that

$$\Delta d_{\varphi c}\llcorner_{\mathring{X}_{\alpha,d_{\varphi c}} \cap \{\varphi \neq c\}} = -(\log h_{\alpha,d_{\varphi c}})'.$$

The last two identities together with the assumption (6-7) applied to d_{φ_c} imply that for all those α such that $X_{\alpha,\varphi} \cap \{\varphi = c\} \neq \varnothing$, for each $x \in \mathring{X}_{\alpha,\varphi} \cap \{\varphi \neq c\}$ it holds

$$-[(\log h_{\alpha,\varphi})'(g_t(x)) - (\log h_{\alpha,\varphi})'(x)] \geq Kt + \frac{1}{N-1}\int_{(0,t)} [(\log h_{\alpha,\varphi})']^2(g_s(x))\, ds \tag{6-8}$$

for all those t such that $\varphi(g_t(x)) > c$ provided $\varphi(x) > c$ (and appropriate modifications if $\varphi(x) < c$). Identifying $\mathring{X}_\alpha$ with the isometric real interval (a_α, b_α) and denoting with c_α the unique point corresponding to $\mathring{X}_\alpha \cap \{\varphi = c\}$, (6-8) becomes

$$-[(\log h_{\alpha,\varphi})'(x+t) - (\log h_{\alpha,\varphi})'(x)] \geq Kt + \frac{1}{N-1}\int_{(0,t)} [(\log h_{\alpha,\varphi})']^2(x+s)\, ds \tag{6-9}$$

for each $x \in (a_\alpha, c_\alpha)$ and t such that $x + t \leq c_\alpha$. We again regularise by logarithmic convolution, i.e., as in (6-6). In order to simplify the notation, we will omit the subscript φ. We have

$$(\log h_\alpha^\varepsilon)'(x) = \int (\log h_\alpha)'(y)\psi_\varepsilon(x-y)\, dx,$$

$$(\log h_\alpha^\varepsilon)'(y) - (\log h_\alpha^\varepsilon)'(y+t) = \int [(\log h_\alpha)'(x) - (\log h_\alpha)'(x+t)]\psi_\varepsilon(x-y)\, dx.$$

Moreover

$$\begin{aligned}\iint_{(0,t)} ((\log h_\alpha)'(x+s))^2\psi_\varepsilon(x-y)\, ds\, dx &= \int_{(0,t)}\int ((\log h_\alpha)'(x+s))^2\psi_\varepsilon(x-y)\, dx\, ds \\ &\geq \int_{(0,t)} \left(\int (\log h_\alpha)'(x+s)\psi_\varepsilon(x-y)\, dx\right)^2 ds \\ &= \int_{(0,t)} \log(h_\alpha^\varepsilon)'(y+s)^2\, ds.\end{aligned}$$

Hence (6-9) is valid for $\log h_{\alpha,\varphi}^\varepsilon$ for each $\varepsilon > 0$ implying (just differentiate in t) that $h_{\alpha,\varphi}^\varepsilon$ is a $\mathsf{CD}(K, N)$ density on (a_α, c_α). Letting $\varepsilon \downarrow 0$ we obtain that $h_{\alpha,\varphi}$ is a $\mathsf{CD}(K, N)$ density on (a_α, c_α). From the arbitrariness of c, we conclude that $h_{\alpha,\varphi}$ is a $\mathsf{CD}(K, N)$ density. Hence $(X, \mathsf{d}, \mathfrak{m})$ satisfies $\mathsf{CD}^1_{\mathrm{Lip}}(K, N)$

(see [Cavalletti and Milman 2016] for the definition of $\mathsf{CD}^1_{\mathrm{Lip}}(K,N)$). Then we can conclude using [Cavalletti and Milman 2016] that $(X,\mathsf{d},\mathfrak{m})$ satisfies $\mathsf{CD}(K,N)$. □

7. Splitting theorem under MCP(0, N)

Before stating the main result of the section, let us introduce some notation.

Given a metric space (X,d), a curve $\bar\gamma:\mathbb{R}\to X$ is called *line* if it is an isometric immersion, i.e.,

$$\bar\gamma:\mathbb{R}\to X,\qquad \mathsf{d}(\bar\gamma_t,\bar\gamma_s)=|t-s|\quad\text{for all } s,t\in\mathbb{R}.$$

To a line $\bar\gamma:\mathbb{R}\to X$ we associate the Busemann functions

$$\mathsf{b}^+(x):=\lim_{t\to+\infty}\mathsf{d}(x,\bar\gamma_t)-t.$$

Straightforwardly from the triangle inequality, one can check that the Busemann functions are well-defined maps $\mathsf{b}^\pm:X\to\mathbb{R}$ and

$$|\mathsf{b}^\pm(x)-\mathsf{b}^\pm(y)|\le\mathsf{d}(x,y).$$

Since $\mathsf{b}^\pm$ are 1-Lipschitz functions, we can consider the associated nonbranching transport set $\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^\pm}$ defined in (3-3).

Theorem 7.1 (splitting theorem). *Let $(X,\mathsf{d},\mathfrak{m})$ be an e.n.b. infinitesimally Hilbertian $\mathsf{MCP}(0,N)$ space containing a line. Then $(X,\mathfrak{m})$ is isomorphic as a measure space to a splitting $Q\times\mathbb{R}$.*

More precisely the following holds. Denoting by $\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+}=\bigcup_{\alpha\in Q}X_\alpha$ the nonbranching transport set induced by b^+ with the associated (disjoint) decomposition in transport rays, it holds that $\mathfrak{m}(X\setminus\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+})=0$ and the map

$$\Phi:\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+}\to Q\times\mathbb{R},\qquad x\mapsto\Phi(x):=(\alpha(x),\mathsf{b}^+(x)),\tag{7-1}$$

is an isomorphism of measures spaces, i.e.,

- *Φ is a bijection,*
- *Φ induces an isomorphism between the σ-algebra of $\mathfrak{m}$-measurable subsets of $\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+}$ and the σ-algebra of $\mathfrak{q}\otimes\mathcal{L}^1$-measurable subsets of $Q\times\mathbb{R}$, where $\mathfrak{q}$ is quotient measure in the disintegration $\mathfrak{m}\llcorner_{\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+}}=\int_Q\mathfrak{m}_\alpha\,\mathfrak{q}(d\alpha)$ given by Theorem 3.6.*
- *$\Phi_\sharp\mathfrak{m}\llcorner_{\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+}}=\mathfrak{q}'\otimes\mathcal{L}^1$. Here $\mathfrak{q}'$ is a nonnegative measure over Q equivalent to $\mathfrak{q}$, i.e., $\mathfrak{q}'\ll\mathfrak{q}$ and $\mathfrak{q}\ll\mathfrak{q}'$.*

Moreover, for every $\alpha\in Q$, the map $\mathsf{b}^+:X_\alpha\to\mathbb{R}$ is an isometry.

If in addition (X,d) is nonbranching, then X is homeomorphic to a splitting $Q\times\mathbb{R}$. More precisely, $X=\mathcal{T}_{\mathsf{b}^+}=\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+}$ and the map $\Phi:X\to Q\times\mathbb{R}$ defined in (7-1) is an homeomorphism. Here the set of rays Q is induced with the compact-open topology as a subset of $C(\mathbb{R},X)$, where each ray is parametrised by $(\mathsf{b}^+)^{-1}$; i.e.,

given $\beta\in Q$, $\{\alpha_n\}_{n\in\mathbb{N}}\subset Q$, it holds $\beta=\lim_{n\to\infty}\alpha_n$ if and only if

$$0=\lim_{n\to\infty}\sup_{t\in I}\mathsf{d}\big(X_{\alpha_n}((\mathsf{b}^+)^{-1}(t)),X_\beta((\mathsf{b}^+)^{-1}(t))\big)\quad\textit{for every compact interval } I\subset\mathbb{R}.\tag{7-2}$$

Remark 7.2. For smooth Riemannian manifolds [Cheeger and Gromoll 1971], as well as for Ricci-limits [Cheeger and Colding 1996] and $\mathsf{RCD}(0, N)$ spaces [Gigli 2013], the splitting theorem has a stronger statement giving an *isometric splitting*. However under the assumptions of Theorem 7.1 it is not conceivable to expect also a splitting of the metric. Indeed the Heisenberg group $\mathbb{H}^n$ is an example of a nonbranching infinitesimally Hilbertian $\mathsf{MCP}(0, N)$ space [Juillet 2009] containing a line, which is homeomorphic and isomorphic as measure space to a splitting (indeed it is homeomorphic to $\mathbb{R}^n$ and the measure is exactly the n-dimensional Lebesgue measure) but it is not isometric to a splitting.

We start by establishing some preliminary lemmas on the properties of Busemann functions.

Lemma 7.3. *For any proper geodesic space, $\mathcal{T}_{\mathsf{b}^{\pm}} = X$.*

Proof. Fix any $x \in X$ and $s > 0$. For each $t \in \mathbb{R}$, consider a unit speed geodesic

$$\gamma^t : [0, \mathsf{d}(x, \bar{\gamma}_t)] \to X \quad \text{such that} \quad \gamma^t_0 = x \quad \text{and} \quad \gamma^t_{\mathsf{d}(x,\bar{\gamma}_t)} = \bar{\gamma}_t.$$

From the triangle inequality, $\lim_{t\to\pm\infty} \mathsf{d}(x, \bar{\gamma}_t) = \infty$. Hence any fixed $s > 0$, for $|t|$ sufficiently large, belongs to the domain of γ^t. Consider then the trivial identities

$$\mathsf{d}(x, \bar{\gamma}_t) - t - \mathsf{d}(\gamma^t_s, \bar{\gamma}_t) + t = \mathsf{d}(x, \gamma^t_s) = s > 0.$$

Taking the limit as $t \to +\infty$ and using uniform convergence gives

$$\mathsf{b}^+(x) - \mathsf{b}^+(y) = \mathsf{d}(x, y) = s > 0,$$

where y is any accumulation point of $\{\gamma^t_s\}_{t\geq 0}$. In particular this shows that each point $x \in X$ can be moved forwardly with respect to b^+ (into y) proving in particular that $x \in \mathcal{T}_{\mathsf{b}^+}$. The proof for b^- can be achieved along the same lines. □

The proof of Lemma 7.3 also proves the following corollary.

Corollary 7.4. *Let (X, d) be proper and geodesic. Then $\mathfrak{b}_{\mathsf{b}^{\pm}} = \emptyset$; i.e., the set of final points associated to b^+ and to b^- are both empty.*

Applying results from Part I we easily obtain the following result.

Proposition 7.5. *Let $(X, \mathsf{d}, \mathfrak{m})$ be an e.n.b. metric measure space satisfying $\mathsf{MCP}(0, N)$ containing a line. Then $\mathsf{b}^{\pm} \in D(\mathbf{\Delta}, X)$ and there exists a Radon measure $\Delta\mathsf{b}^{\pm} \in \mathbf{\Delta}\mathsf{b}^{\pm}$ satisfying*

$$\Delta\mathsf{b}^{\pm} \leq 0. \tag{7-3}$$

Proof. We only prove the claim for b^+, the proof for b^- being analogous. First of all from Theorem 3.6, we have the disintegration

$$\mathfrak{m} = \int_Q h_\alpha \mathcal{H}^1 \llcorner_{X_\alpha} \mathfrak{q}(d\alpha).$$

Thanks to Corollary 7.4 we deduce that each ray X_α is isomorphic to a right half-line (or to a full line), in particular it has infinite length.

The combination of Theorem 4.8 with Lemma 7.3 thus gives that $\mathsf{b}^+ \in D(\mathbf{\Delta}, X)$ and that

$$\Delta \mathsf{b}^+ := -\int_Q h'_\alpha \mathcal{H}^1 \llcorner_{X_\alpha} \mathfrak{q}(d\alpha) - \int_Q h_\alpha \delta_{a(X_\alpha)\cap U} \mathfrak{q}(d\alpha)$$

defines a Radon measure $\Delta \mathsf{b}^+ \in \mathbf{\Delta} \mathsf{b}^+$. We are left to show that $\Delta \mathsf{b}^+ \leq 0$.

As above, we identify X_α with the right half-line $[a_\alpha, \infty)$ endowed with the $\mathsf{MCP}(0, N)$ density h_α. Using (2-11), we deduce that for $a_\alpha < x_0 \leq x_1 < b < \infty$ it holds

$$\left(\frac{b - x_1}{b - x_0}\right)^{N-1} \leq \frac{h_\alpha(x_1)}{h_\alpha(x_0)}.$$

Letting $b \to \infty$, it follows that $h_\alpha(x_0) \leq h_\alpha(x_1)$, showing that $h'_\alpha \geq 0$ whenever h'_α exists.

Thus $\Delta \mathsf{b}^+ \leq 0$ and the proposition follows. □

Observe also that, by triangle inequality, one has

$$\mathsf{d}(x, \bar\gamma_t) - t + \mathsf{d}(x, \bar\gamma_{-s}) - s \geq 0.$$

Setting $\mathsf{b} := \mathsf{b}^+ + \mathsf{b}^-$ and letting $t, s \to \infty$, it gives

$$\mathsf{b} \geq 0 \quad \text{on } X \qquad \text{and} \qquad \mathsf{b} \equiv 0 \quad \text{on } \bar\gamma. \tag{7-4}$$

From now on we assume $(X, \mathsf{d}, \mathfrak{m})$ to be infinitesimally Hilbertian, which is equivalent to assuming that the Laplacian Δ is single-valued (on its domain) and linear. Proposition 7.5 then implies

$$\mathsf{b} := \mathsf{b}^+ + \mathsf{b}^- \in D(\mathbf{\Delta}, X), \qquad \Delta \mathsf{b} \leq 0. \tag{7-5}$$

It is worth noting that (7-5) will be the only implication of the paper where infinitesimal Hilbertianity plays a role. We now want to combine (7-5) and (7-4) with the strong maximum principle in order to infer that $\mathsf{b} \equiv 0$. The next statement was proved in [Björn and Björn 2011, Theorem 9.13] (actually we report a slightly weaker statement which will suffice for our purposes).

Theorem 7.6 (strong maximum principle). *Let $(X, \mathsf{d}, \mathfrak{m})$ be a metric measure space supporting a local weak $(1, 2)$-Poincaré inequality with $\mathfrak{m}$ locally doubling. Let $u \in \mathrm{LIP}(X)$ and $\Omega \subset X$ be a connected bounded open subset.*

If u attains its maximum in an interior point of Ω and

$$\int_\Omega |\nabla u|^2 \, \mathfrak{m} \leq \int_\Omega |\nabla(u + f)|^2 \, \mathfrak{m} \quad \textit{for all } f \in \mathrm{LIP}(X),\ \mathrm{supp}(f) \subset \Omega,\ f \leq 0, \tag{7-6}$$

then u is constant on Ω.

Let us discuss the validity of the strong maximum principle in our setting. Clearly, from Bishop–Gromov inequality it follows that an $\mathsf{MCP}(0, N)$ space is doubling. Moreover, essentially nonbranching $\mathsf{MCP}(0, N)$ spaces satisfy a local weak $(1, 1)$-Poincaré inequality [von Renesse 2008] (that work assumes negligibility of cut-locus from $\mathfrak{m}$-a.e. point that is satisfied whenever the space is essentially nonbranching, see Remark 2.6), which in turns implies that the space supports a local weak $(1, 2)$-Poincaré inequality. In conclusion if $(X, \mathsf{d}, \mathfrak{m})$ is an essentially nonbranching $\mathsf{MCP}(0, N)$ space, then the strong maximum principle holds. The simple link between (7-6) and the measure-valued Laplacian was established in [Gigli and Mondino 2013, Theorem 4.3]; for completeness, below we report the argument together with the desired conclusion $\mathsf{b} \equiv 0$.

Lemma 7.7. *Let* $(X, \mathsf{d}, \mathfrak{m})$ *be an infinitesimally Hilbertian, essentially nonbranching, metric measure space satisfying* $\mathsf{MCP}(0, N)$. *Assume* (X, d) *contains a line and let* $\mathsf{b} := \mathsf{b}^+ + \mathsf{b}^-$.

Then $\mathsf{b} \equiv 0$ *on* X.

Proof. It is enough to prove that (7-5) implies (7-6) for $u := -\mathsf{b}$; then the claim will follow by the combination of (7-4) with Theorem 7.6.

Let $\Omega \subset X$ be a connected bounded open subset and $f \in \mathrm{LIP}(X)$ be nonpositive with $\mathrm{supp}(f) \subset \Omega$. Since the map $\varepsilon \mapsto \int_\Omega |\nabla(-\mathsf{b} + \varepsilon f)|^2\, \mathfrak{m}$ is convex and $\Delta \mathsf{b} \leq 0$, we have

$$\begin{aligned}\int_\Omega |\nabla(-\mathsf{b} + f)|^2\, \mathfrak{m} - \int_\Omega |\nabla(-\mathsf{b})|^2\, \mathfrak{m} &\geq \lim_{\varepsilon \downarrow 0} \int_\Omega \frac{|\nabla(-\mathsf{b} + \varepsilon f)|^2 - |\nabla(-\mathsf{b})|^2}{\varepsilon}\, \mathfrak{m} \\ &= -2 \int_\Omega \langle \nabla \mathsf{b}, \nabla f \rangle\, \mathfrak{m} = 2 \int_\Omega f\, \Delta \mathsf{b} \geq 0,\end{aligned}$$

proving (7-6) for $u := -\mathsf{b}$. □

Lemma 7.8. *Let* $(X, \mathsf{d}, \mathfrak{m})$ *be an infinitesimally Hilbertian, essentially nonbranching, metric measure space satisfying* $\mathsf{MCP}(0, N)$. *Assume* (X, d) *contains a line. Let* $\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+} = \bigcup_{\alpha \in Q} X_\alpha$ *be the ray decomposition of the nonbranching transport set* $\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+}$ *associated to* b^+.

Then for each $\alpha \in Q$, *the ray* X_α *is isometric to* $\mathbb{R}$; *in other words* $a(X_\alpha) = \varnothing = b(X_\alpha)$.

Proof. From Lemma 7.7 we know that $\mathsf{b}^+ = -\mathsf{b}^-$ on all X. It follows that

$$\{(x, y) \in R^{\mathrm{nb}}_{\mathsf{b}^+}\} = \{(y, x) \in R^{\mathrm{nb}}_{\mathsf{b}^-}\}.$$

Thus $\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+} = \mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^-}$ with the same ray decomposition (from the support sense); clearly, on each ray, the orientation induced by b^+ is the opposite from the one induced by b^-. In particular, the set of initial points for b^+ coincides with the set of final points for b^-:

$$\mathfrak{a}_{\mathsf{b}^+} := \{x \in \mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+} : (y, x) \in R^{\mathrm{nb}}_{\mathsf{b}^+} \iff y = x\} = \{x \in \mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^-} : (x, y) \in R^{\mathrm{nb}}_{\mathsf{b}^-} \iff x = y\} =: \mathfrak{b}_{\mathsf{b}^-}.$$

Since from Corollary 7.4 the set of final points for b^- is empty, i.e., $\mathfrak{b}_{\mathsf{b}^-} = \varnothing$, it follows that both the sets of initial and final points for b^+ are empty; in other words, each ray X_α is isometric to $\mathbb{R}$. □

Proof of the splitting theorem, Theorem 7.1. By combining the lemmas above we can quickly get the first part of Theorem 7.1. Indeed, from Lemma 7.3 we already know that $X = \mathcal{T}_{\mathsf{b}^+}$ and, from Lemma 3.5 we know that $\mathfrak{m}(\mathcal{T}_{\mathsf{b}^+} \setminus \mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+}) = 0$; thus the claim $\mathfrak{m}(X \setminus \mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+}) = 0$ is proved.

Moreover, Theorem 3.6 ensures that there exists a disintegration of $\mathfrak{m}$ satisfying

$$\mathfrak{m}\llcorner_{\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+}} = \int_Q \mathfrak{m}_\alpha\, \mathfrak{q}(d\alpha), \qquad \mathfrak{q}(Q) = 1,$$

where, for $\mathfrak{q}$-a.e. α, $\mathfrak{m}_\alpha$ is a Radon measure $\mathfrak{m}_\alpha \ll \mathcal{H}^1\llcorner_{X_\alpha}$ and $(X_\alpha, \mathsf{d}, \mathfrak{m}_\alpha)$ satisfies $\mathsf{MCP}(0, N)$.

From Lemma 7.8 we know that (X_α, d) is isometric to the real line (note that the isometry is simply $\mathsf{b}^+ : X_\alpha \to \mathbb{R}$), and thus Lemma 2.17 implies that $\mathfrak{m}_\alpha = c_\alpha \mathcal{H}^1 \llcorner_{X_\alpha}$ for some constant $c_\alpha > 0$, for $\mathfrak{q}$-a.e. $\alpha \in Q$.

Define the measure $\mathfrak{q}'$ on Q as

$$\mathfrak{q}'(B) = \int_B c_\alpha\, \mathfrak{q}(d\alpha) \quad \text{for any } \mathfrak{q}\text{-measurable subset } B \subset Q.$$

It is clear that $\mathfrak{q}' \ll \mathfrak{q}$ and that $\mathfrak{q} \ll \mathfrak{q}'$, i.e., they are equivalent measures, and that

$$\mathfrak{m}\llcorner_{\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+}} = \int_Q \mathcal{H}^1 \llcorner_{X_\alpha}\, \mathfrak{q}'(d\alpha).$$

The last disintegration formula is equivalent to claiming that the map

$$\Phi : \mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+} \to Q \times \mathbb{R}, \qquad x \mapsto \Phi(x) := (\alpha(x), \mathsf{b}^+(x)),$$

is an isomorphism of measures spaces, i.e., Φ induces an isomorphism between the σ-algebra of $\mathfrak{m}$-measurable subsets of $\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+}$ and the σ-algebra of $\mathfrak{q}\otimes\mathcal{L}^1$-measurable subsets of $Q\times\mathbb{R}$, and $\Phi_\sharp \mathfrak{m}\llcorner_{\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+}} = \mathfrak{q}'\otimes\mathcal{L}^1$. It is also clear that $\Phi : \mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+} \to Q \times \mathbb{R}$ is bijective, as $\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+} = \bigcup_{\alpha\in Q} X_\alpha$ is a partition, and $\mathsf{b}^+ : X_\alpha \to \mathbb{R}$ is an isometry for every $\alpha \in Q$.

We now prove the second part of Theorem 7.1. From the very definition (3-3) of the nonbranched transport set $\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+}$, if (X, d) is nonbranching then $\mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+} = \mathcal{T}_{\mathsf{b}^+}$. Thus, Lemma 7.3 gives $X = \mathcal{T}_{\mathsf{b}^+} = \mathcal{T}^{\mathrm{nb}}_{\mathsf{b}^+}$.

From the first part, we already know that $\Phi : X \to Q \times \mathbb{R}$ is bijective. Since convergence in Q (see (7-2)) is equivalent to the local uniform convergence of the rays, it is clear that Φ^{-1} is continuous.

It is then enough to show that Φ is continuous. We argue by contradiction. Assume that there exists a sequence $\{x_n\}_{n\in\mathbb{N}} \subset X$ with $x_n \to x$ in X such that $\{(\alpha(x_n), \mathsf{b}^+(x_n))\}_{n\in\mathbb{N}}$ does not converge to $(\alpha(x), \mathsf{b}^+(x))$. Since $\mathsf{b}^+ : X \to \mathbb{R}$ is continuous (actually it is even 1-Lipschitz), it is clear that $\mathsf{b}^+(x_n) \to \mathsf{b}^+(x)$ and thus it must be that $\{\alpha(x_n)\}_{n\in\mathbb{N}}$ does not converge to $\alpha(x)$. By the definition (7-2) of convergence in Q, it follows that, up to subsequences, it holds

$$0 < \varepsilon = \lim_{n\to\infty} \sup_{t\in I} \mathsf{d}\big(X_{\alpha(x_n)}((\mathsf{b}^+)^{-1}(t)), X_{\alpha(x)}((\mathsf{b}^+)^{-1}(t))\big) \quad \text{for some compact interval } I \subset \mathbb{R}. \tag{7-7}$$

As already observed, $\mathsf{b}^+ : X_\beta \to \mathbb{R}$ is an isometry for every $\beta \in Q$ and thus it can be used to parametrise each ray; in the formula above as well as in the following we fix such a parametrisation.

Since by assumption $x_n \to x$, for every closed interval $I \subset \mathbb{R}$ containing $\mathsf{b}^+(x)$, it is clear that the union of the images of the rays $X_{\alpha(x_n)}$ restricted to I are all contained in a compact subset of X. Thus, the by Arzelà–Ascoli theorem, such restrictions converge uniformly to a geodesic γ of X passing through x. By a standard diagonal argument, γ can be extended to a geodesic defined on the whole $\mathbb{R}$ and

$$X_{\alpha(x_n)} \to \gamma \quad \text{uniformly on compact intervals.} \tag{7-8}$$

Recalling that the relation R_{b^+} is closed (see (3-1) and (3-2)) we get that γ is a ray passing through x, i.e., $\gamma = X_\beta$ for some $\beta \in Q$. Since the rays are pairwise disjoint, it follows that $\beta = \alpha(x)$.

Therefore (7-8) contradicts (7-7). □

Acknowledgements

Mondino acknowledges the support of the EPSRC First Grant EP/R004730/1 and of the ERC Starting Grant 802689. The authors wish to thank the anonymous reviewers for the careful reading and for their comments, which led to an improvement in the exposition.

References

[Ambrosio 2018] L. Ambrosio, "Calculus, heat flow and curvature-dimension bounds in metric measure spaces", pp. 301–340 in *Proc. Int. Congress Math., I* (Rio de Janiero, 2018), World Sci., Singapore, 2018.

[Ambrosio et al. 2014a] L. Ambrosio, N. Gigli, and G. Savaré, "Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below", *Invent. Math.* **195**:2 (2014), 289–391. MR Zbl

[Ambrosio et al. 2014b] L. Ambrosio, N. Gigli, and G. Savaré, "Metric measure spaces with Riemannian Ricci curvature bounded from below", *Duke Math. J.* **163**:7 (2014), 1405–1490. MR Zbl

[Ambrosio et al. 2015a] L. Ambrosio, N. Gigli, A. Mondino, and T. Rajala, "Riemannian Ricci curvature lower bounds in metric measure spaces with σ-finite measure", *Trans. Amer. Math. Soc.* **367**:7 (2015), 4661–4701. MR Zbl

[Ambrosio et al. 2015b] L. Ambrosio, N. Gigli, and G. Savaré, "Bakry–Émery curvature-dimension condition and Riemannian Ricci curvature bounds", *Ann. Probab.* **43**:1 (2015), 339–404. MR Zbl

[Ambrosio et al. 2019] L. Ambrosio, A. Mondino, and G. Savaré, *Nonlinear diffusion equations and curvature conditions in metric measure spaces*, Mem. Amer. Math. Soc. **1270**, Amer. Math. Soc., Providence, RI, 2019. MR Zbl

[Bacher and Sturm 2010] K. Bacher and K.-T. Sturm, "Localization and tensorization properties of the curvature-dimension condition for metric measure spaces", *J. Funct. Anal.* **259**:1 (2010), 28–56. MR Zbl

[Barilari and Rizzi 2018] D. Barilari and L. Rizzi, "Sharp measure contraction property for generalized H-type Carnot groups", *Commun. Contemp. Math.* **20**:6 (2018), art. id. 1750081. MR Zbl

[Bianchini and Cavalletti 2013] S. Bianchini and F. Cavalletti, "The Monge problem for distance cost in geodesic spaces", *Comm. Math. Phys.* **318**:3 (2013), 615–673. MR Zbl

[Björn and Björn 2011] A. Björn and J. Björn, *Nonlinear potential theory on metric spaces*, EMS Tracts in Math. **17**, Eur. Math. Soc., Zürich, 2011. MR Zbl

[Calabi 1958] E. Calabi, "An extension of E. Hopf's maximum principle with an application to Riemannian geometry", *Duke Math. J.* **25**:1 (1958), 45–56. MR Zbl

[Cavalletti 2014] F. Cavalletti, "Monge problem in metric measure spaces with Riemannian curvature-dimension condition", *Nonlinear Anal.* **99** (2014), 136–151. MR Zbl

[Cavalletti and Milman 2016] F. Cavalletti and E. Milman, "The globalization theorem for the curvature dimension condition", preprint, 2016. arXiv

[Cavalletti and Mondino 2017a] F. Cavalletti and A. Mondino, "Optimal maps in essentially non-branching spaces", *Commun. Contemp. Math.* **19**:6 (2017), art. id. 1750007. MR Zbl

[Cavalletti and Mondino 2017b] F. Cavalletti and A. Mondino, "Sharp and rigid isoperimetric inequalities in metric-measure spaces with lower Ricci curvature bounds", *Invent. Math.* **208**:3 (2017), 803–849. MR Zbl

[Cavalletti and Mondino 2018] F. Cavalletti and A. Mondino, "Almost Euclidean isoperimetric inequalities in spaces satisfying local Ricci curvature lower bounds", *Int. Math. Res. Not.* (online publication April 2018).

[Cheeger 1999] J. Cheeger, "Differentiability of Lipschitz functions on metric measure spaces", *Geom. Funct. Anal.* **9**:3 (1999), 428–517. MR Zbl

[Cheeger 2001] J. Cheeger, *Degeneration of Riemannian metrics under Ricci curvature bounds*, Scuola Normale Superiore, Pisa, 2001. MR Zbl

[Cheeger and Colding 1996] J. Cheeger and T. H. Colding, "Lower bounds on Ricci curvature and the almost rigidity of warped products", *Ann. of Math.* (2) **144**:1 (1996), 189–237. MR Zbl

[Cheeger and Colding 1997] J. Cheeger and T. H. Colding, "On the structure of spaces with Ricci curvature bounded below, I", *J. Differential Geom.* **45**:3 (1997), 406–480. MR Zbl

[Cheeger and Colding 2000a] J. Cheeger and T. H. Colding, "On the structure of spaces with Ricci curvature bounded below, II", *J. Differential Geom.* **54**:1 (2000), 13–35. MR Zbl

[Cheeger and Colding 2000b] J. Cheeger and T. H. Colding, "On the structure of spaces with Ricci curvature bounded below, III", *J. Differential Geom.* **54**:1 (2000), 37–74. MR Zbl

[Cheeger and Gromoll 1971] J. Cheeger and D. Gromoll, "The splitting theorem for manifolds of nonnegative Ricci curvature", *J. Differential Geom.* **6** (1971), 119–128. MR Zbl

[Colding 1996a] T. H. Colding, "Large manifolds with positive Ricci curvature", *Invent. Math.* **124**:1-3 (1996), 193–214. MR Zbl

[Colding 1996b] T. H. Colding, "Shape of manifolds with positive Ricci curvature", *Invent. Math.* **124**:1-3 (1996), 175–191. MR Zbl

[Colding 1997] T. H. Colding, "Ricci curvature and volume convergence", *Ann. of Math.* (2) **145**:3 (1997), 477–501. MR Zbl

[Colding and Naber 2012] T. H. Colding and A. Naber, "Sharp Hölder continuity of tangent cones for spaces with a lower Ricci curvature bound and applications", *Ann. of Math.* (2) **176**:2 (2012), 1173–1229. MR Zbl

[Erbar et al. 2015] M. Erbar, K. Kuwada, and K.-T. Sturm, "On the equivalence of the entropic curvature-dimension condition and Bochner's inequality on metric measure spaces", *Invent. Math.* **201**:3 (2015), 993–1071. MR Zbl

[Fremlin 2003] D. H. Fremlin, *Measure theory, IV: Topological measure spaces, I*, Torres Fremlin, Colchester, UK, 2003.

[Gigli 2013] N. Gigli, "The splitting theorem in non-smooth context", preprint, 2013. arXiv

[Gigli 2015] N. Gigli, *On the differential structure of metric measure spaces and applications*, Mem. Amer. Math. Soc. **1113**, Amer. Math. Soc., Providence, RI, 2015. MR Zbl

[Gigli and Mondino 2013] N. Gigli and A. Mondino, "A PDE approach to nonlinear potential theory in metric measure spaces", *J. Math. Pures Appl.* (9) **100**:4 (2013), 505–534. MR Zbl

[Gigli and Pasqualetto 2016] N. Gigli and E. Pasqualetto, "Behaviour of the reference measure on RCD spaces under charts", preprint, 2016. arXiv

[Gigli et al. 2015a] N. Gigli, A. Mondino, and T. Rajala, "Euclidean spaces as weak tangents of infinitesimally Hilbertian metric measure spaces with Ricci curvature bounded below", *J. Reine Angew. Math.* **705** (2015), 233–244. MR Zbl

[Gigli et al. 2015b] N. Gigli, A. Mondino, and G. Savaré, "Convergence of pointed non-compact metric measure spaces and stability of Ricci curvature bounds and heat flows", *Proc. Lond. Math. Soc.* (3) **111**:5 (2015), 1071–1129. MR Zbl

[Juillet 2009] N. Juillet, "Geometric inequalities and generalized Ricci bounds in the Heisenberg group", *Int. Math. Res. Not.* **2009**:13 (2009), 2347–2373. MR Zbl

[Kell and Mondino 2018] M. Kell and A. Mondino, "On the volume measure of non-smooth spaces with Ricci curvature bounded below", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **18**:2 (2018), 593–610. MR Zbl

[Ketterer 2015] C. Ketterer, "Cones over metric measure spaces and the maximal diameter theorem", *J. Math. Pures Appl.* (9) **103**:5 (2015), 1228–1275. MR Zbl

[Kitabeppu and Lakzian 2016] Y. Kitabeppu and S. Lakzian, "Characterization of low dimensional $\mathrm{RCD}^*(K, N)$ spaces", *Anal. Geom. Metr. Spaces* **4**:1 (2016), 187–215. MR Zbl

[Klartag 2017] B. Klartag, *Needle decompositions in Riemannian geometry*, Mem. Amer. Math. Soc. **1180**, Amer. Math. Soc., Providence, RI, 2017. MR Zbl

[Li and Yau 1986] P. Li and S.-T. Yau, "On the parabolic kernel of the Schrödinger operator", *Acta Math.* **156**:3-4 (1986), 153–201. MR Zbl

[Lott and Villani 2009] J. Lott and C. Villani, "Ricci curvature for metric-measure spaces via optimal transport", *Ann. of Math.* (2) **169**:3 (2009), 903–991. MR Zbl

[Mantegazza and Mennucci 2003] C. Mantegazza and A. C. Mennucci, "Hamilton–Jacobi equations and distance functions on Riemannian manifolds", *Appl. Math. Optim.* **47**:1 (2003), 1–25. MR Zbl

[Mantegazza et al. 2014] C. Mantegazza, G. Mascellani, and G. Uraltsev, "On the distributional Hessian of the distance function", *Pacific J. Math.* **270**:1 (2014), 151–166. MR Zbl

[Mondino 2015] A. Mondino, "A new notion of angle between three points in a metric space", *J. Reine Angew. Math.* **706** (2015), 103–121. MR Zbl

[Mondino and Naber 2019] A. Mondino and A. Naber, "Structure theory of metric measure spaces with lower Ricci curvature bounds", *J. Eur. Math. Soc.* **21**:6 (2019), 1809–1854. MR Zbl

[Ohta 2007a] S.-i. Ohta, "On the measure contraction property of metric measure spaces", *Comment. Math. Helv.* **82**:4 (2007), 805–828. MR Zbl

[Ohta 2007b] S.-i. Ohta, "Products, cones, and suspensions of spaces with the measure contraction property", *J. Lond. Math. Soc.* (2) **76**:1 (2007), 225–236. MR Zbl

[Ohta and Sturm 2014] S.-i. Ohta and K.-T. Sturm, "Bochner–Weitzenböck formula and Li–Yau estimates on Finsler manifolds", *Adv. Math.* **252** (2014), 429–448. MR Zbl

[Petersen 1998] P. Petersen, *Riemannian geometry*, Graduate Texts in Math. **171**, Springer, 1998. MR Zbl

[Petrunin 2011] A. Petrunin, "Alexandrov meets Lott–Villani–Sturm", *Münster J. Math.* **4** (2011), 53–64. MR Zbl

[Rajala and Sturm 2014] T. Rajala and K.-T. Sturm, "Non-branching geodesics and optimal maps in strong $\mathrm{CD}(K,\infty)$-spaces", *Calc. Var. Partial Differential Equations* **50**:3-4 (2014), 831–846. MR Zbl

[von Renesse 2008] M.-K. von Renesse, "On local Poincaré via transportation", *Math. Z.* **259**:1 (2008), 21–31. MR Zbl

[Rifford 2013] L. Rifford, "Ricci curvatures in Carnot groups", *Math. Control Relat. Fields* **3**:4 (2013), 467–487. MR Zbl

[Rizzi 2016] L. Rizzi, "Measure contraction properties of Carnot groups", *Calc. Var. Partial Differential Equations* **55**:3 (2016), art. id. 60. MR Zbl

[Sturm 2006a] K.-T. Sturm, "On the geometry of metric measure spaces, I", *Acta Math.* **196**:1 (2006), 65–131. MR Zbl

[Sturm 2006b] K.-T. Sturm, "On the geometry of metric measure spaces, II", *Acta Math.* **196**:1 (2006), 133–177. MR Zbl

[Villani 2009] C. Villani, *Optimal transport: old and new*, Grundlehren der Math. Wissenschaften **338**, Springer, 2009. MR Zbl

[Villani 2019] C. Villani, "Inégalités isopérimétriques dans les espaces métriques mesurés (d'après F. Cavalletti & A. Mondino)", exposé 1127, pp. 213–265 in *Séminaire Bourbaki,* 2016/2017, Astérisque **407**, Soc. Math. France, Paris, 2019. MR

Received 20 Nov 2018. Revised 21 Jul 2019. Accepted 6 Sep 2019.

FABIO CAVALLETTI: cavallet@sissa.it
Mathematics Area, SISSA, Trieste, Italy

ANDREA MONDINO: andrea.mondino@maths.ox.ac.uk
Current address: *Mathematical Institute, University of Oxford, Oxford, United Kingdom*
Mathematics Institute, University of Warwick, Coventry, United Kingdom

ANALYSIS AND PDE
Vol. 13, No. 7, 2020
https://doi.org/10.2140/apde.2020.13.2149

msp

CONVEX SETS EVOLVING BY VOLUME-PRESERVING FRACTIONAL MEAN CURVATURE FLOWS

ELEONORA CINTI, CARLO SINESTRARI AND ENRICO VALDINOCI

We consider the volume-preserving geometric evolution of the boundary of a set under fractional mean curvature. We show that smooth convex solutions maintain their fractional curvatures bounded for all times, and the long-time asymptotics approach round spheres. The proofs are based on a priori estimates on the inner and outer radii of the solutions.

1. Introduction

Let $E_0 \subset \mathbb{R}^n$ be a smooth compact convex set, and let $\mathcal{M}_0 = \partial E_0$. For a fixed $s \in (0, 1)$, we consider the evolution of $\mathcal{M}_0$ by volume-preserving fractional mean curvature flow, that is, the family of immersions $F : \mathcal{M}_0 \times [0, T) \to \mathbb{R}^n$ which satisfies

$$\begin{cases} \partial_t F(p,t) = [-H_s(p,t) + h(t)]\nu(p,t), & p \in \mathcal{M}_0,\ t \geqslant 0, \\ F(p,0) = p & p \in \mathcal{M}_0. \end{cases} \tag{1}$$

Here $H_s(p, t)$ and $\nu(p, t)$ denote respectively the fractional mean curvature of order s and the normal vector of the hypersurface $\mathcal{M}_t := F(\mathcal{M}_0, t)$ at the point $F(p, t)$, while the function $h(t)$ is defined as

$$h(t) = \frac{1}{|\mathcal{M}_t|} \int_{\mathcal{M}_t} H_s(x)\, d\mu, \tag{2}$$

where $d\mu$ denotes the surface measure on $\mathcal{M}_t$. With this choice of $h(t)$, the set E_t enclosed by $\mathcal{M}_t$ has constant volume. An interesting feature of this flow is that the fractional s-perimeter of E_t is decreasing, and the monotonicity is strict unless E_t is a sphere.

Fractional (or nonlocal) mean curvature was first defined by Caffarelli, Roquejoffre and Savin [Caffarelli et al. 2010]. It arises naturally when performing the first variation of the fractional perimeter, a nonlocal notion of perimeter introduced in the same paper. We will recall the definitions of these quantities in Section 2. Minimizers of the fractional perimeter are usually called nonlocal minimal sets, and their boundaries nonlocal minimal surfaces. Fractional perimeter and mean curvature have also found application in other contexts, such as image reconstruction and nonlocal capillarity models; see, e.g., [Bosch and Stoll 2015; Maggi and Valdinoci 2017].

Nonlocal minimal surfaces have attracted the interest of many researchers in the last years. One of the main issues is the study of their regularity and the classification of nonlocal minimal cones: many

MSC2010: 53C44, 35R11, 35B40.
Keywords: geometric evolution equations, fractional partial differential equations, fractional perimeter, fractional mean curvature flow, asymptotic behavior of solutions.

results have been obtained, see [Caffarelli et al. 2010; Caffarelli and Valdinoci 2013; Barrios et al. 2014; Savin and Valdinoci 2013; Cinti et al. 2019; Cabré et al. 2020], which exhibit interesting analogies and differences with respect to the classical case. Among the important differences, we mention in particular the fact that fractional minimal surfaces can stick at the boundary of (even smooth and convex) domains, and occupy all the domain for small values of the fractional parameter; see [Dipierro et al. 2017]: these features are in sharp contrast with the classical case and they reveal the important role of the contributions coming from infinity in the geometric displacements of nonlocal minimal surfaces.

A related topic of investigation consists in the study of sets which are stationary for the fractional perimeter, i.e., sets having vanishing nonlocal mean curvature. This is a weaker notion than minimality, and some examples are helicoids and a nonlocal version of catenoids; see, e.g., [Dávila et al. 2018; Cinti et al. 2016]. Sets with constant nonlocal mean curvature, such as Delaunay-type surfaces, have been studied in [Cabré et al. 2018a; 2018b; 2018c; Dávila et al. 2016]. In addition, in [Cabré et al. 2018a; Ciraolo et al. 2018] an analogue of the Alexandrov theorem in the nonlocal setting was proved, which will be crucial for our purposes: any, regular enough, bounded set having constant fractional mean curvature is necessarily a ball.

Before introducing our results, let us recall some properties of the *classical* mean curvature flow, where the speed of the hypersurface is given by the usual mean curvature. This flow has been widely studied in the last decades, both for its geometric interest and for its relevance in physical models describing the dynamics of interfaces. The equation satisfied by the immersion is a parabolic PDE, and smooth solutions exist locally; however, they can become singular in finite time due to curvature blowup. For this reason, various notions of weak solutions have been introduced during the years which allow for the continuation of the evolution after the formation of singularities; see, e.g., [Chen et al. 1991; Evans and Spruck 1991].

An important feature of classical mean curvature flow is that, roughly speaking, it deforms general hypersurfaces into some canonical profiles, possibly after rescaling near the singularities. Such a behavior is related to the diffusive character of the flow and is of great interest for geometric applications. The first result on asymptotic convergence was obtained in [Huisken 1984] in the $h(t) \equiv 0$ case. He proved that *convex hypersurfaces remain smooth up to a finite maximal time at which they shrink to a point, and that they converge to a round sphere after rescaling*. Shortly afterwards, in [Huisken 1987], he obtained an analogous result for the *volume-preserving flow*: in this case, the solution exists for all times and converges to a sphere as $t \to +\infty$. In later years, many researchers have studied the convergence to a sphere for other kinds of geometric flows, with a speed driven by more general functions of the (classical) principal curvatures; see, e.g., [Andrews et al. 2013; Andrews and Wei 2017]. As a possible application of these results, we point out that the convergence to a sphere along a suitable flow can be used to obtain generalizations or alternative proofs of classical geometric inequalities, such as the isoperimetric inequality, or inequalities in convex analysis like the ones by Minkowski or Alexandrov and Fenchel; see, e.g., [McCoy 2005; Schulze 2008; Guan and Li 2009; Andrews et al. 2018].

By contrast, the study of fractional mean curvature flow has started only recently and very few results are known. The existence and uniqueness of weak solutions in the viscosity sense for the flow in the $h(t) \equiv 0$ case have been obtained by various authors with different approaches [Imbert 2009;

Caffarelli and Souganidis 2010; Chambolle et al. 2015]. In particular, Caffarelli and Souganidis [2010] proved convergence to motion by fractional mean curvature of a threshold dynamics scheme. After this, Chambolle, Novaga and Ruffini [Chambolle et al. 2017] extended the results in [Caffarelli and Souganidis 2010] to the anisotropic case and to the presence of an external driving force (that is $h(t) \neq 0$) and proved that the scheme preserves convexity, and, as a consequence, also the limit geometric evolution is convexity-preserving. On the other hand, the existence of smooth solutions has been established only recently in [Julin and La Manna 2020]. Their main result states the short-time existence of a unique classical solution for both the fractional mean curvature flow and the volume-preserving flow, starting from a $C^{1,1}$ initial datum.

Some qualitative properties of smooth solutions were analyzed in [Sáez and Valdinoci 2019], while the formation of neckpinch singularities was studied in [Cinti et al. 2018]. The occurrence of fattening for the fractional mean curvature flow and its generalizations were studied in [Cesaroni et al. 2019].

The aim of this paper is to study the *convergence to a sphere of the solutions of the nonlocal flow* (1) *with convex initial data*. This can be regarded as the first attempt to investigate the asymptotic behavior of solutions to fractional flows, in a similar spirit to the above-mentioned works in the classical case. Our main results are some a priori estimates on smooth solutions, which give a uniform control on the geometry of the evolving surfaces, and establish that the fractional curvature remains uniformly bounded along the flow. As a consequence, we can show that any smooth solution, satisfying suitable regularity assumptions, exists for all times and converges to a sphere. The method is inspired by the one of [Andrews 2001; Sinestrari 2015] in the classical setting and is based on the monotonicity along the flow of the fractional isoperimetric ratio, i.e., the ratio between suitable powers of the fractional perimeter and the enclosed volume. This monotonicity property is specific to the volume-preserving case, and so the approach used here does not apply when $h(t) \equiv 0$, although we expect that case to exhibit a similar behavior, at least if s is suitably close to 1. On the other hand, we include in this paper the treatment of more general flows in the volume-preserving setting, with a *nonlinear speed* of the form $\Phi(H_s)$, with $\Phi(\cdot)$ a positive increasing function satisfying suitable structural assumptions.

Let us describe our results in more detail. For this, let us denote by $\underline{\rho}_E$ and $\bar{\rho}_E$ the inner radius and the outer radius of a set $E \subset \mathbb{R}^n$, namely

$$\begin{aligned} \underline{\rho}_E &:= \sup\{r > 0 : \text{there exists } x_o \in \mathbb{R}^n \text{ such that } B_r(x_o) \subset E\}, \\ \bar{\rho}_E &:= \inf\{r > 0 : \text{there exists } x_o \in \mathbb{R}^n \text{ such that } B_r(x_o) \supset E\}. \end{aligned} \tag{3}$$

Then our main estimates can be stated as follows:

Theorem 1.1. *Let E_0 be a smooth compact convex set of $\mathbb{R}^n$ and let $\mathcal{M}_0 = \partial E_0$. Let $F : \mathcal{M}_0 \times [0, T) \to \mathbb{R}^n$, with $0 < T \leqslant +\infty$, be a solution of* (1) *of class $C^{2,\beta}$ for some $\beta > s$. Then there exist positive constants $0 < R_1 \leqslant R_2$, $0 < K_1 \leqslant K_2$, only depending on E_0, such that*

$$\begin{gathered} R_1 \leqslant \underline{\rho}_{E_t} \leqslant \bar{\rho}_{E_t} \leqslant R_2, \\ K_1 \leqslant H_s(p, t) \leqslant K_2, \quad p \in \mathcal{M}_0, \end{gathered}$$

for all $t \in [0, T)$.

As mentioned above, in [Chambolle et al. 2017] it is proven that the nonlocal mean curvature flow with forcing term ($h(t) \not\equiv 0$) preserves convexity. As a consequence, we know that solutions of problem (1) starting from a convex initial datum stay convex for all times.

The proof of Theorem 1.1 relies on a series of delicate estimates based on a nonlocal analysis of geometric flavor, which turns out to be significantly different with respect to the classical case.

Let us describe some intermediate steps in the proof of Theorem 1.1, which we believe to be of interest on their own. One of these results, Proposition 3.1, shows that a bound on the fractional isoperimetric ratio of a convex set implies a bound on the ratio between the outer and inner radii. A similar result was known in the classical case, but the proof in the nonlocal setting is quite different. Another crucial step of our argument is provided by Proposition 4.2, where we estimate the fractional mean curvature in terms of another nonlocal quantity, which has some formal analogy with the norm of the second fundamental form in the classical case. However, since there is no fractional analogue of the second fundamental form, as shown in [Abatangelo and Valdinoci 2014], there is no obvious relation as in the classical case. By suitable estimates of the surface integrals involved, we obtain an inequality which suffices for the purposes of this paper; on the other hand, it would be interesting to investigate further these topics and to derive sharper inequalities in the future.

Theorem 1.1 easily implies that a solution of (1) exists for all times and converges to a sphere as $t \to +\infty$, provided it satisfies suitable regularity and continuation properties. Roughly speaking, we need to know that the solution remains smooth and does not develop singularities as long as the fractional curvature is bounded. More precisely, we assume that there exists a smooth solution of (1) satisfying the following property for some $\beta > s$:

(R) If H_s is bounded on $\mathcal{M}_t$ for all $t \in [0, T_0)$ for some $T_0 \leqslant T$, where T is the maximal time of existence, then the $C^{2,\beta}$-norm of $\mathcal{M}_t$, up to translations, is also bounded for $t \in [0, T)$ by a constant only depending on the supremum of H_s. In addition, either $T_0 = T = +\infty$, or $T_0 < T$.

By "up to translations", we mean that $\mathcal{M}_t$ is not assumed to remain in a bounded set of $\mathbb{R}^n$, and that the $C^{2,\beta}$ bound applies after possibly composing the flow with a suitable, time-dependent, translation (e.g., the one fixing the barycenter). We give below more comments on the possibility of this behavior. For solutions satisfying (R), the following result holds.

Theorem 1.2. *Let E_0 be a smooth compact convex set of $\mathbb{R}^n$ and let $\mathcal{M}_0 = \partial E_0$. Let $F : \mathcal{M}_0 \times [0, T) \to \mathbb{R}^n$, with $0 < T \leqslant +\infty$, be a solution of (1) of class $C^{2,\beta}$ for some $\beta > s$ which satisfies property (R). Then $T = +\infty$, and $\mathcal{M}_t$ converges to a round sphere as $t \to +\infty$ in $C^{2,\beta}$ norm, possibly up to translations.*

Regarding assumption (R), we observe that it is a natural analogue of some properties which are well known in the classical case, see, e.g., [Huisken 1984, Sections 7–8], and are consequences of the standard parabolic theory. In the fractional setting, the validity of such an assumption is an open problem at the current stage. The only available results in this direction [Julin and La Manna 2020] imply, roughly speaking, that the last claim in (R) is true: if the $C^{1,\beta}$ norm of the solution remains bounded, for some $\beta > s$, then the smooth solution exists for all times. On the other hand, the boundedness of the fractional curvature gives directly $C^{1,\beta}$ bounds only for $\beta \leqslant s$. It can be hoped that solutions of the flow enjoy further regularity, in analogy with some regularity studies on elliptic and parabolic nonlocal problems; see, e.g., [Barrios et al. 2014; Chang-Lara and Dávila 2014a; 2014b; Dipierro et al. 2020]. In this respect, this paper should be regarded as a part of a broader program, which we plan to pursue further in future work.

As observed above, in Theorem 1.2 the convergence to a sphere is in principle only "up to translations", in the sense that the limit set, which is geometrically a sphere, could keep translating indefinitely. In the classical case, the possibility of the additional translation is ruled out either as a consequence of additional estimates on the convergence rate, see, e.g., [Bertini and Pipoli 2017], or by maximum-principle techniques based on reflection methods [Chow and Gulliver 1996; McCoy 2004; Andrews and Wei 2017]. We think that it would be interesting to understand whether these methods can be extended to the nonlocal setting.

The paper is organized as follows:

- In Section 2, we give some preliminaries and we recall the evolution laws of some geometric quantities associated to $\mathcal{M}_t$.
- Section 3 contains our a priori estimates on the inner and outer radii of convex solutions and a lower bound for H_s.
- Section 4 deals with some integral estimates which allow us to bound the fractional mean curvature with the nonlocal analogue of the norm of the nonlocal second fundamental form.
- In Section 5 we prove our key result, which gives an upper bound on the fractional mean curvature.
- In Section 6, we treat the more general case of a flow whose speed is of the form $\Phi(H_s)$, proving an upper bound on the fractional mean curvature.
- Finally, in Section 7, we prove convergence to a sphere in both the standard and the general cases.

2. Preliminaries

Consider a set $E \subset \mathbb{R}^n$, with boundary $\mathcal{M} := \partial E$, and let $s \in (0, 1)$. Given $x \in \mathcal{M}$, the *fractional mean curvature* of order s of E (equivalently, of $\mathcal{M}$) at x is defined by

$$H_s(x) = s(1-s) \lim_{\varepsilon \to 0^+} \int_{\mathbb{R}^n \setminus B_\varepsilon(x)} \frac{\tilde{\chi}_E(y)}{|x-y|^{n+s}}\, dy, \tag{4}$$

where

$$\tilde{\chi}_E(y) = \begin{cases} 1 & \text{if } y \in E^c, \\ -1 & \text{if } y \in E. \end{cases}$$

If $\mathcal{M}$ is smooth, then the fractional mean curvature is well-defined at each point and is a regular function. In fact, the following result is known; see [Figalli et al. 2015, Proposition 6.3; Cabré et al. 2018a, Proposition 2.1].

Theorem 2.1. *Suppose ∂E is of class $C^{1,\beta}$, with $\beta > s$. Then the right-hand side of (4) is well-defined and finite for all $x \in \partial E$ and defines a continuous function on ∂E. If in addition ∂E is of class $C^{2,\beta}$, with $\beta > s$, then $H_s \in C^1(\partial E)$ and its derivative in a tangential direction $v \in T_x\mathcal{M}$ is given by*

$$\frac{\partial H_s}{\partial v}(x) = s(1-s)(n+s) \lim_{\varepsilon \to 0^+} \int_{\mathbb{R}^n \setminus B_\varepsilon(x)} \tilde{\chi}_E(y) \frac{\langle y-x, v\rangle}{|x-y|^{n+s+2}}\, dy. \tag{5}$$

By using the divergence theorem and estimating the boundary terms on $\partial B_\varepsilon(x)$ with techniques similar to the proof of [Cabré et al. 2018a, Proposition 2.1], we can prove that, under the hypotheses of the previous theorem, H_s and its gradient can be written as boundary integrals on $\mathcal{M}$ as follows:

$$H_s(x) = 2(1-s) \lim_{\varepsilon\to 0^+} \int_{\mathcal{M}\setminus B_\varepsilon(x)} \frac{\langle y-x, \nu(y)\rangle}{|x-y|^{n+s}}\, d\mu(y), \tag{6}$$

$$\frac{\partial H_s}{\partial v}(x) = 2s(1-s) \lim_{\varepsilon\to 0^+} \int_{\mathcal{M}\setminus B_\varepsilon(x)} \frac{\nu(y)\cdot v}{|x-y|^{n+s}}\, d\mu(y). \tag{7}$$

We also recall that the *fractional perimeter* of E, as introduced in [Caffarelli et al. 2010], is defined as

$$\mathrm{Per}_s(E) = s(1-s) \int_E \int_{E^c} \frac{dx\, dy}{|x-y|^{n+s}}.$$

Then fractional mean curvature arises as the first variation of the fractional perimeter along a deformation of E; see (8) later.

We state a general criterion for the convergence of singular integrals on the boundary of a smooth compact set E. Suppose that ∂E is of class $C^{1,\beta}$, for some $\beta > s$, and that $f \in C^2(\partial E)$. Then, for any given $x \in \partial E$, the quantity

$$\lim_{\varepsilon\to 0^+} \int_{\mathcal{M}\setminus B_\varepsilon(x)} \frac{f(y)-f(x)}{|x-y|^{n+s}}\, d\mu(y)$$

exists and is finite.This can be proved by standard arguments. Roughly speaking, the contribution of the first-order approximation of $f(y)-f(x)$ around x cancels by symmetry reasons. The remaining terms are of order $O(|y-x|)^{1+\beta}$, by the smoothness of ∂E and of f, and this ensures convergence of the integral. In the following, for simplicity of notation, we will write singular integrals as the ones above as if they were ordinary integrals, with the implicit meaning that they are taken in the principal value sense.

We now recall some notation and general results about geometric evolutions of sets and hypersurfaces. Let us consider a time-dependent family of sets E_t evolving smoothly from a given initial set E_0. We can consider the corresponding evolution of the boundaries, and study the map $F : \mathcal{M}_0 \times [0, T) \to \mathbb{R}^n$, where $\mathcal{M}_0 = \partial E_0$ and $\mathcal{M}_t := \partial E_t$. Let us denote by $V(p, t) := \langle \partial_t F(p, t), \nu(p, t)\rangle$ the normal component of the speed of our flow.

We first recall the properties of the evolution of the classical geometric quantities associated to the hypersurfaces $\mathcal{M}_t$. As in [Huisken 1984], we denote by g_{ij} the components of the metric tensor in a given coordinate system, by g^{ij} its inverse, by h_{ij} the second fundamental form, by $H = h_{ij}g^{ij}$ the mean curvature and by $|A|^2 = h_{ij}g^{jl}h_{lk}g^{ki}$ the squared norm of the second fundamental form. If $\lambda_1 \leqslant \cdots \leqslant \lambda_{n-1}$ denote the principal curvatures at a given point, then $H = \lambda_1 + \cdots + \lambda_{n-1}$, while $|A|^2 = \lambda_1^2 + \cdots + \lambda_{n-1}^2$. We also denote by $\nabla^{\mathcal{M}_t}$, $\Delta^{\mathcal{M}_t}$ respectively the tangential gradient and the Laplace–Beltrami operator defined on $\mathcal{M}_t$.

We denote by $p, q, \ldots$ the points on $\mathcal{M}_0$ and by $x, y, \ldots$ the points on $\mathcal{M}_t$ for positive t, as well as the general points in $\mathbb{R}^n$. For simplicity of notation, when considering the speed V on $\mathcal{M}_t$ for a fixed t, we will usually write $V(x)$ with $x \in \mathcal{M}_t$ instead of $V(p, t)$, with $x = F(p, t)$. We will use similar conventions for all other quantities defined on the evolving hypersurfaces. We also denote by $d\mu$ the surface measure along $\mathcal{M}_t$. In this notation, we recall [Huisken and Polden 1999, Theorem 3.2 and Lemmata 7.4, 7.5 and 7.6] and we have:

Lemma 2.2. *The geometric quantities associated to $\mathcal{M}_t$ satisfy the following equations*:

(i) $\partial_t g_{ij} = 2V h_{ij}$ *and* $\partial_t g^{ij} = -2V h^{ij}$.

(ii) $\partial_t d\mu = V H \, d\mu$.

(iii) $\partial_t \nu = -\nabla^{\mathcal{M}_t} V$.

(iv) $\partial_t h_{ij} = -\nabla_i^{\mathcal{M}_t} \nabla_j^{\mathcal{M}_t} V + h_{ik} g^{km} h_{mj} V$.

(v) $\partial_t H = -\Delta^{\mathcal{M}_t} V - |A|^2 V$.

(vi) $\frac{d}{dt}|E_t| = \int_{\mathcal{M}_t} V(x)\, d\mu$ *and* $\frac{d}{dt}|M_t| = \int_{\mathcal{M}_t} V(x) H(x)\, d\mu$.

Next we recall the evolution of some nonlocal quantities; see [Caffarelli et al. 2010; Dávila et al. 2018, Appendix B, Proposition B.2; Sáez and Valdinoci 2019, Theorem 14].

Lemma 2.3. (i) *The fractional perimeter evolves according to*

$$\frac{d}{dt} \operatorname{Per}_s(E_t) = \int_{\mathcal{M}_t} H_s(x) V(x)\, d\mu. \tag{8}$$

(ii) *The fractional mean curvature satisfies the equation*

$$\frac{\partial_t H_s}{2s(1-s)} = -\int_{\mathcal{M}_t} \frac{V(y)-V(x)}{|y-x|^{n+s}}\, d\mu(y) - V(x) \int_{\mathcal{M}_t} \frac{1-\nu(y)\cdot\nu(x)}{|y-x|^{n+s}}\, d\mu(y). \tag{9}$$

We remark that there is a clear analogy between these equations and their classical counterparts. Indeed, as proved in [Dávila et al. 2018, Appendix A], we have, for a general smooth function f defined on a (fixed) hypersurface $\mathcal{M}$,

$$\lim_{s\to 1^-} 2s(1-s) \int_{\mathcal{M}} \frac{f(y)-f(x)}{|y-x|^{n+s}}\, d\mu(y) = \omega_n \Delta^{\mathcal{M}} f(x), \tag{10}$$

where ω_n is the volume of the unit ball of $\mathbb{R}^n$. In addition,

$$\lim_{s\to 1^-} 2s(1-s) \int_{\mathcal{M}} \frac{1-\nu(y)\cdot\nu(x)}{|y-x|^{n+s}}\, d\mu(y) = \omega_n |A|^2. \tag{11}$$

From now on, we assume that the map $F : \mathcal{M}_0 \times [0, T) \to \mathbb{R}^n$ satisfies (1). This corresponds to the normal speed

$$V(p,t) = -H_s(p,t) + h(t),$$

with $h(t)$ defined as in (2).

Then Lemma 2.2(vi) implies the enclosed volume E_t remains constant in time, while by Lemma 2.3(i) the fractional perimeter decreases according to

$$\partial_t \operatorname{Per}_s(E_t) = \int_{\mathcal{M}_t} [-H_s(x)+h(t)] H_s(x)\, d\mu = -\int_{\mathcal{M}_t} [H_s(x)-h(t)]^2\, d\mu \leqslant 0. \tag{12}$$

We conclude this section by recalling the analogue of the Alexandrov theorem in the nonlocal setting.

Theorem 2.4 [Cabré et al. 2018a, Theorem 1.1; Ciraolo et al. 2018, Theorem 1.1]. *Let E be a bounded open set of class $C^{1,s}$ with constant nonlocal mean curvature. Then, E is a ball.*

We point out that, by (12) and Theorem 2.4, the monotonicity of $\mathrm{Per}_s(E_t)$ is strict unless E_t is a sphere.

3. Bounds on inner and outer radii

Given a bounded set $E \subset \mathbb{R}^n$ with nonempty interior and $\omega \in \partial B_1$, we denote by $w_E(\omega)$ the width of the set E in direction ω; i.e.,

$$w_E(\omega) := \sup_{x,y\in E} (x-y)\cdot\omega. \tag{13}$$

Notice that w_E is the distance between the two hyperplanes orthogonal to ω touching E from outside. We also set

$$\underline{w}_E := \inf_{\omega\in\partial B_1} w_E(\omega) \quad \text{and} \quad \overline{w}_E := \sup_{\omega\in\partial B_1} w_E(\omega).$$

By construction, we have

$$\overline{w}_E = \mathrm{diam}(E). \tag{14}$$

Recalling the notation in (3), if E is convex, it is known that

$$\underline{\rho}_E \geqslant \frac{\underline{w}_E}{n+1} \quad \text{and} \quad \bar{\rho}_E \leqslant \frac{\overline{w}_E}{\sqrt{2}}; \tag{15}$$

see, e.g., [Andrews 1994, Lemma 5.4].

Using this notation, the following result holds true:

Proposition 3.1. *For any bounded, convex set $E \subset \mathbb{R}^n$ with nonempty interior, we have*

$$\underline{w}_E \geqslant c\left(\frac{|E|}{\mathrm{Per}_s(E)}\right)^{1/s}, \tag{16}$$

$$\underline{\rho}_E \geqslant c\left(\frac{|E|}{\mathrm{Per}_s(E)}\right)^{1/s}, \tag{17}$$

$$\overline{w}_E \leqslant C(\mathrm{Per}_s(E))^{(n-1)/s}|E|^{(1+s-n)/s}, \tag{18}$$

$$\bar{\rho}_E \leqslant C(\mathrm{Per}_s(E))^{(n-1)/s}|E|^{(1+s-n)/s}, \tag{19}$$

$$\frac{\bar{\rho}_E}{\underline{\rho}_E} \leqslant C(\mathrm{Per}_s(E))^{n/s}|E|^{(s-n)/s} \tag{20}$$

for suitable constants $C > c > 0$ only depending on n, s.

Proof. First of all, we observe that

$$\text{it is enough to prove (16),} \tag{21}$$

since, after that, the claims in (17), (18), (19) and (20) would follow. Indeed, if (16) holds true, then (17) follows directly from (15).

Now we prove (18) assuming that (16) (and so (17)) holds true. To this aim, we observe that we can suppose that

$$\overline{w}_E \geqslant 4\underline{\rho}_E. \tag{22}$$

Indeed, suppose instead that the opposite inequality holds. Then, we use the nonlocal isoperimetric inequality (see [Frank et al. 2008]) to see that

$$|E|^{1/n} = |E|^{(n-s)(n-1)/(ns)}|E|^{(1+s-n)/s} \leqslant C_1(\mathrm{Per}_s(E))^{(n-1)/s}|E|^{(1+s-n)/s}$$

for some $C_1 > 0$. Accordingly, since $|E|^{1/n} \geqslant |B_{\underline{\rho}_E}|^{1/n} = C_2\underline{\rho}_E$, for some $C_2 > 0$, we obtain

$$C_2\underline{\rho}_E \leqslant C_1(\mathrm{Per}_s(E))^{(n-1)/s}|E|^{(1+s-n)/s},$$

and so, if the opposite inequality holds in (22),

$$\frac{C_2\overline{w}_E}{4} \leqslant C_1(\mathrm{Per}_s(E))^{(n-1)/s}|E|^{(1+s-n)/s},$$

which says that (18) is satisfied.

Consequently, we may assume that (22) holds true. Thus, after a translation we may suppose that $B_{\underline{\rho}_E} \subseteq E$ and there exists $p \in \overline{E}$ with $|p| \geqslant \overline{w}_E/2 - \underline{\rho}_E$. We stress that, in view of (22),

$$|p| \geqslant \frac{\overline{w}_E}{4} =: \ell.$$

Since E is convex, the convex hull of p with $B_{\underline{\rho}_E}$ lies in $\overline{E}$ and therefore $|E| \geqslant \tilde{c}\underline{\rho}_E{}^{n-1}\ell$ for some $\tilde{c} > 0$. This and (17) imply

$$\overline{w}_E = 4\ell \leqslant \frac{4|E|}{\tilde{c}\underline{\rho}_E{}^{n-1}} \leqslant \frac{4|E|(\mathrm{Per}_s(E))^{(n-1)/s}}{\tilde{c}c^{n-1}|E|^{(n-1)/s}},$$

which gives (18), as desired.

Then, from (18) and (15), one obtains (19). Finally, (20) clearly follows from (17) and (19). This completes the proof of (21).

In view of (21), from now on we focus on the proof of (16). To this aim, after a rigid motion, we may suppose that $\underline{w}_E$ is realized in the vertical direction, and, more precisely, that

$$E \subseteq \{x_n \in [-\underline{w}_E, 0]\}. \tag{23}$$

We denote by π the projection onto $\mathbb{R}^{n-1} \times \{0\}$ and $E' := \pi(E)$. We consider a nonoverlapping tiling of $\mathbb{R}^{n-1} \times \{0\}$ by cubes $\{Q_i\}_{i\in\mathbb{N}}$ which have side length equal to $\underline{w}_E/\sqrt{n-1}$ (hence, their diagonal is equal to $\underline{w}_E$). We denote by $\mathbb{N}_\star$ the set of indices $i \in \mathbb{N}$ for which Q_i intersects E'. Let also

$$Q := \bigcup_{i\in\mathbb{N}_\star} Q_i \quad \text{and} \quad F := Q \times (0, \underline{w}_E].$$

Due to (23), we know that F lies outside E and therefore

$$\begin{aligned}\mathrm{Per}_s(E) &\geqslant \iint_{E\times F} \frac{dx\,dy}{|x-y|^{n+s}} \\ &= \int_{-\underline{w}_E}^{0} dx_n \int_Q dx' \int_0^{\underline{w}_E} dy_n \int_Q dy' \frac{\chi_E(x',x_n)}{|x-y|^{n+s}} \\ &\geqslant \sum_{i\in\mathbb{N}_\star} \int_{-\underline{w}_E}^{0} dx_n \int_{Q_i} dx' \int_0^{\underline{w}_E} dy_n \int_{Q_i} dy' \frac{\chi_E(x',x_n)}{|x-y|^{n+s}}.\end{aligned}$$

Now we remark that if $x', y' \in Q$, $x_n \in [-\underline{w}_E, 0]$ and $y_n \in (0, \underline{w}_E]$, we have

$$|x-y|^2 = |x'-y'|^2 + |x_n-y_n|^2 \leqslant \underline{w}_E{}^2 + (2\underline{w}_E)^2 = 5\underline{w}_E{}^2.$$

As a consequence,

$$\begin{aligned}\mathrm{Per}_s(E) &\geqslant \frac{1}{5^{(n+s)/2}\underline{w}_E{}^{n+s}} \sum_{i\in\mathbb{N}_\star} \int_{-\underline{w}_E}^{0} dx_n \int_{Q_i} dx' \int_0^{\underline{w}_E} dy_n \int_{Q_i} dy' \chi_E(x',x_n) \\ &= \frac{1}{5^{(n+s)/2}\underline{w}_E{}^{n+s}} \left(\frac{\underline{w}_E}{\sqrt{n-1}}\right)^{n-1} \underline{w}_E \sum_{i\in\mathbb{N}_\star} \int_{-\underline{w}_E}^{0} dx_n \int_{Q_i} dx' \chi_E(x',x_n) \\ &= \frac{1}{5^{(n+s)/2}\underline{w}_E{}^{n+s}} \left(\frac{\underline{w}_E}{\sqrt{n-1}}\right)^{n-1} \underline{w}_E |E|,\end{aligned}$$

where we used (23) once again in the last identity. This estimate plainly implies (16), as desired. □

For completeness, we point out an interesting geometric consequence of the estimate in (20) in terms of the nonlocal isoperimetric ratio

$$\mathcal{I}_s(E) := \frac{(\mathrm{Per}_s(E))^n}{|E|^{n-s}}.$$

Indeed, formula (20) states that if the nonlocal isoperimetric ratio of E is bounded, then so is the ratio between the inner and outer radius of E and, more precisely,

$$\frac{\bar{\rho}_E}{\underline{\rho}_E} \leqslant C(\mathcal{I}_s(E))^{1/s}.$$

In the local case when $s = 1$, this formula was already known; see, e.g., [Andrews 2001, Proposition 5.1; Sinestrari 2015, Proposition 2.1].

As an immediate consequence of the results of this section, we obtain:

Corollary 3.2. *Let E_0 be a convex subset of $\mathbb{R}^n$ and $\mathcal{M}_0 = \partial E_0$. Let $F : \mathcal{M}_0 \times [0,T) \to \mathbb{R}^n$, with $0 < T \leqslant +\infty$, be a solution of (1). Then there exist positive constants $0 < R_1 \leqslant R_2$, only depending on E_0, such that*

$$R_1 \leqslant \underline{\rho}_{E_t} \leqslant \bar{\rho}_{E_t} \leqslant R_2 \quad \textit{for all } t \in [0,T).$$

In addition, there exists $K_1 > 0$ such that $H_s(p,t) \geqslant K_1$ for all $(p,t) \in \mathcal{M}_0 \times [0,T)$.

Proof. As already mentioned in the Introduction, we know that the evolution given by (1) preserves convexity, as established in [Chambolle et al. 2017]; hence we have that E_t is convex for all $0 < t < T$.

By definition, we have

$$\omega_n \underline{\rho}_{E_t}^n \leqslant |E_t| \leqslant \omega_n \bar{\rho}_{E_t}^n.$$

Since $|E_t|$ is constant, this gives an upper bound on $\underline{\rho}_{E_t}$ and a lower bound on $\bar{\rho}_{E_t}$ in terms of $|E_0|$. On the other hand, since $\mathrm{Per}_s(E_t)$ is decreasing in time, inequality (20) gives a uniform bound on the ratio $\bar{\rho}_{E_t}/\underline{\rho}_{E_t}$. These properties together yield the first assertion.

To prove the lower bound on H_s, let us consider an arbitrary point $x \in \mathcal{M}_t$. Since E_t is convex, it is contained in the half-space $\{y \in \mathbb{R}^n : (y-x)\cdot\nu(x) \leqslant 0\}$. Moreover, by definition, the diameter of E_t is not greater than $2\bar{\rho}_{E_t}$, which is less than $2R_2$. Therefore, if we introduce the half-balls

$$B_+ = \{y \in B_{2R_2}(x) : (y-x)\cdot\nu(x) \geqslant 0\}, \qquad B_- = \{y \in B_{2R_2}(x) : (y-x)\cdot\nu(x) \leqslant 0\},$$

we have that $E_t \subset B_-$. It follows that

$$\begin{aligned}\frac{1}{s(1-s)} H_s(x) &= \int_{E_t^c} \frac{dy}{|x-y|^{n+s}} - \int_{E_t} \frac{dy}{|x-y|^{n+s}} \\ &\geqslant \int_{\mathbb{R}^n \setminus B_{2R_2}(x)} \frac{dy}{|x-y|^{n+s}} + \int_{B_+} \frac{dy}{|x-y|^{n+s}} - \int_{B_-} \frac{dy}{|x-y|^{n+s}} \\ &= \int_{\mathbb{R}^n \setminus B_{2R_2}(x)} \frac{dy}{|x-y|^{n+s}} = \int_{|z|\geqslant 2R_2} \frac{dz}{|z|^{n+s}},\end{aligned}$$

where the last integral is independent of x, t. □

The previous result contains the first part of the statement of Theorem 1.1 (the bounds on inner and outer radii and the lower bound for H_s). To conclude the proof of Theorem 1.1 it remains to establish the upper bound for the fractional mean curvature, which will be done in Section 5.

We conclude this section with the following observation. Corollary 3.2 ensures that, at any given time, there exists a ball of radius R_1 contained in E_t. However, the center of the ball may be different at different times. We want to show that, by choosing a smaller radius, we can find a ball with fixed center which remains inside E_t for a time interval with fixed length.

Lemma 3.3. *For any $t_0 \geqslant 0$, we can find $x_0 \in \mathbb{R}^n$ such that*

$$B_{R_1/2}(x_0) \subset E_t \quad \textit{for all } t \in [t_0, t_0 + t^*],$$

where $t^* > 0$ only depends on n, s, R_1.

Proof. As in [Andrews 2001; McCoy 2004], we use a comparison argument. Volume-preserving curvature flows in general do not satisfy an avoidance principle. However, if E_t evolves by (1) and F_t evolves by the standard fractional mean curvature flow (corresponding to $h(t) \equiv 0$) then an easy maximum principle argument shows that if $F_{t_0} \subset E_{t_0}$ at a certain time t_0, then we also have $F_t \subset E_t$ for all $t \geqslant t_0$.

In our case, we can use comparison with a shrinking ball. From the previous corollary, there exists x_0 such that $B_{R_1}(x_0) \subset E_t$. We set $F_{t_0} = B_{R_1}(x_0)$ and we denote by F_t the evolution of F_{t_0} for $t \geqslant t_0$ by standard fractional mean curvature flow, which is a shrinking sphere. We let t^* be the time such that $F_{t_0+t^*} = B_{R_1/2}(x_0)$, whose value only depends on n, s, R_1. Then the comparison argument yields the conclusion. □

4. Integral surface estimates for convex sets

We collect in this section some estimates on weighted integrals along the boundary of a convex set. We start with a uniform estimate of the weighted surface of a convex set only dependent on its inner and outer radii.

Lemma 4.1. *Let $\beta > 1$ and let $E \subset \mathbb{R}^n$ be a bounded, convex set with nonempty interior. Then, there exists a constant $C > 0$, depending on n, such that, for any $\in \partial E$, we have*

$$\int_{\partial E} \frac{d\mu(y)}{|x-y|^{n-\beta}} \leqslant C \frac{\bar\rho_E}{\underline{\rho}_E}\left[\frac{1}{\beta-1} + \left(\frac{\bar\rho_E}{\underline{\rho}_E}\right)^{n-2}\right](\operatorname{diam}(E))^{\beta-1}.$$

Proof. We can suppose that x is the origin. By definition, there exists $p \in E$ such that $B_{\underline{\rho}_E}(p) \subseteq E$. By convexity, the convex envelope of 0 and $B_{\underline{\rho}_E}(p)$ lies in $\bar E$. Up to a rotation, we can assume that $p = (0, \dots, |p|)$. This easily implies, again by convexity, that $B_{\underline{\rho}_E/2}(0) \cap \partial E$ is the graph of a Lipschitz function f, with Lipschitz constant bounded by $2|p|/\underline{\rho}_E \leqslant 4\bar\rho_E/\underline{\rho}_E$.

Let us set $\delta := \underline{\rho}_E/2$ and $M := \bar\rho_E/\underline{\rho}_E$. In addition, let us denote by $C', C'', \dots$ constants depending only on n. We can estimate, using the fact that $\beta > 1$,

$$\int_{\partial E\cap B_\delta} \frac{d\mu(y)}{|y|^{n-\beta}} \leqslant \int_{\substack{y'\in\mathbb{R}^{n-1}\\ |y'|\leqslant\delta}} \frac{\sqrt{1+|\nabla f(y')|^2}}{|y'|^{n-\beta}}\, dy' \leqslant C'M\int_0^\delta \frac{\tau^{n-2}}{\tau^{n-\beta}}\, d\tau = \frac{C'M}{\beta-1}\delta^{\beta-1}. \tag{24}$$

The remaining part of the integral satisfies

$$\int_{\partial E\setminus B_\delta} \frac{d\mu(y)}{|y|^{n-\beta}} \leqslant \frac{1}{\delta^{n-\beta}}\int_{\partial E\setminus B_\delta} d\mu(y) \leqslant \frac{\mu(\partial E)}{\delta^{n-\beta}}. \tag{25}$$

Now we observe that

$$\mu(\partial E) \leqslant \mu(B_{\bar\rho_E}). \tag{26}$$

Indeed, we know that there exists $q \in E$ such that $B_{\bar\rho_E}(q) \supseteq E$. Let us denote by $\Pi_E : \mathbb{R}^n \to E$ the projection on the convex set E. Then Π_E maps $\partial B_{\bar\rho_E}(q)$ onto ∂E and is nonexpansive, from which (26) follows.

As a consequence of (25) and (26), we obtain that

$$\int_{\partial E\setminus B_\delta} \frac{d\mu(y)}{|y|^{n-\beta}} \leqslant \frac{C''{\bar\rho_E}^{n-1}}{\delta^{n-\beta}} = C'''M^{n-1}\delta^{\beta-1}.$$

This and (24) imply the desired result (recall also (14) and (15)). □

Now we obtain a bound on the fractional mean curvature in terms of the integral quantity which appears in the last term of (9). In view of (11), one can consider this estimate as the fractional counterpart of the elementary property that the classical mean curvature is bounded by the norm of the second fundamental form. An estimate of this kind is more delicate to obtain in the nonlocal case, since the fractional mean curvature cannot be realized by the average of finitely many directional curvatures, and so methods involving linear algebra cannot be applied; see [Abatangelo and Valdinoci 2014]. We give here a proof in the case of convex sets, but it is natural to expect that a similar property should hold in a more general setting.

Proposition 4.2. *Let $E \subset \mathbb{R}^n$ be a convex set with $C^{1,\alpha}$ boundary, with $\alpha \in (s, 1)$. Then, there exists $C > 0$, depending on n and on the ratio $\bar{\rho}_E/\underline{\rho}_E$, such that, for every $x \in \partial E$, we have*

$$H_s(x) \leqslant C(\operatorname{diam}(E))^{(1-s)/2}\left((1-s)\int_{\partial E} \frac{1-\nu(y)\cdot\nu(x)}{|x-y|^{n+s}}\,d\mu(y)\right)^{1/2}.$$

Proof. Given $x \in \partial E$, with exterior normal $\nu(x)$, from the convexity of E we have that $\{p \in \mathbb{R}^n : (p-x)\cdot\nu(x) > 0\}$ touches E from outside at p. As a consequence, if $y \in \partial E$, we have that $(y-x)\cdot\nu(x) \leqslant 0$ and therefore, recalling (6), we have

$$\begin{aligned}
\frac{1}{2(1-s)}H_s(x) &= \int_{\partial E} \frac{(y-x)\cdot\nu(y)}{|x-y|^{n+s}}\,d\mu(y)\\
&= \int_{\partial E} \frac{(y-x)\cdot\nu(x)}{|x-y|^{n+s}}\,d\mu(y) + \int_{\partial E} \frac{(y-x)\cdot(\nu(y)-\nu(x))}{|x-y|^{n+s}}\,d\mu(y)\\
&\leqslant \int_{\partial E} \frac{(y-x)\cdot(\nu(y)-\nu(x))}{|x-y|^{n+s}}\,d\mu(y)\\
&\leqslant \int_{\partial E} \frac{|\nu(y)-\nu(x)|}{|x-y|^{n+s-1}}\,d\mu(y)\\
&= \int_{\partial E} \frac{|\nu(y)-\nu(x)|}{|x-y|^{(n+s)/2}}\,\frac{d\mu(y)}{|x-y|^{(n+s-2)/2}}.
\end{aligned}$$

Hence, exploiting Hölder's inequality,

$$\frac{1}{2(1-s)}H_s(x) \leqslant \sqrt{\int_{\partial E} \frac{|\nu(y)-\nu(x)|^2}{|x-y|^{n+s}}\,d\mu(y)}\ \sqrt{\int_{\partial E} \frac{d\mu(y)}{|x-y|^{n+s-2}}}.$$

Since we have $|\nu(y)-\nu(x)|^2 = 2(1-\nu(y)\cdot\nu(x))$, the desired result follows easily from Lemma 4.1 with $\beta := 2-s > 1$. □

5. Upper bound on the fractional curvature

In this section, we show that the bounds on the inner and outer radii imply that the fractional mean curvature of our solution is bounded from above. This, together with Corollary 3.2, will conclude the proof of Theorem 1.1.

To this purpose, we adapt to the nonlocal setting a technique originally introduced in [Tso 1985]. We consider the support function on the evolving hypersurface

$$u(p,t) = \langle F(p,t), \nu(p,t)\rangle.$$

By Lemma 2.2(iii) and the representation (7) of the gradient of H_s, we find that u evolves according to

$$\begin{aligned}
\partial_t u &= \langle \partial_t F, \nu\rangle + \langle F, \partial_t \nu\rangle\\
&= -H_s + h + \langle F, \nabla^{\mathcal{M}} H_s\rangle = -H_s + h + 2s(1-s)\int_{\mathcal{M}_t} \frac{x^T\cdot\nu(y)}{|y-x|^{n+s}}\,d\mu(y).
\end{aligned} \tag{27}$$

From Lemma 3.3, we know that for any t_0 there exists $x_0 \in \mathbb{R}^n$ such that $B_{R_1/2}(x_0) \subset E_t$ for any $t \in [t_0, t_0 + t^*]$. For simplicity, we perform our computations in the case $x_0 = 0$. By the convexity of E_t, we deduce that $u \geqslant R_1/2$ on $\mathcal{M}_t$ for all $t \in [t_0, t_0 + t^*]$. We then set $\alpha = R_1/4$ and we consider the function

$$W = \frac{H_s}{u - \alpha}.$$

Since

$$\alpha \leqslant u - \alpha \leqslant \operatorname{diam}(\mathcal{M}_t) - \alpha \leqslant 2\bar{\rho}_{E_t} - \alpha,$$

we deduce from Corollary 3.2 that

$$\frac{1}{C} \leqslant \frac{W}{H_s} \leqslant C \tag{28}$$

for some C only depending on n, s and the initial data.

Let us now analyze the evolution equation satisfied by W. By Lemma 2.3(ii), the fractional mean curvature satisfies the equation

$$\frac{\partial_t H_s}{2s(1-s)} = \int_{\mathcal{M}_t} \frac{H_s(y) - H_s(x)}{|y-x|^{n+s}}\, d\mu(y) + (H_s(x) - h(t)) \int_{\mathcal{M}_t} \frac{1 - \nu(y)\cdot\nu(x)}{|y-x|^{n+s}}\, d\mu(y).$$

Recalling (27) and neglecting the positive terms containing $h(t)$, we find

$$\begin{aligned}
\frac{\partial_t W(x,t)}{2s(1-s)} &= \frac{1}{u(x)-\alpha} \int_{\partial E_t} \frac{H_s(y) - H_s(x)}{|y-x|^{n+s}}\, d\mu(y) + \frac{H_s(x) - h(t)}{u(x)-\alpha} \int_{\partial E_t} \frac{1-\nu(y)\cdot\nu(x)}{|y-x|^{n+s}}\, d\mu(y) \\
&\qquad - \frac{H_s(x)}{(u(x)-\alpha)^2}\left(\frac{-H_s(x) + h(t)}{2s(1-s)} + \int_{\partial E_t} \frac{x^T \cdot \nu(y)}{|y-x|^{n+s}}\, d\mu(y)\right) \\
&< \frac{1}{u(x)-\alpha} \int_{\partial E_t} \frac{H_s(y) - H_s(x)}{|y-x|^{n+s}}\, d\mu(y) + \frac{H_s(x)}{u(x)-\alpha} \int_{\partial E_t} \frac{1-\nu(y)\cdot\nu(x)}{|y-x|^{n+s}}\, d\mu(y) \\
&\qquad - \frac{H_s(x)}{(u(x)-\alpha)^2}\left(\frac{-H_s(x)}{2s(1-s)} + \int_{\partial E_t} \frac{x^T \cdot \nu(y)}{|y-x|^{n+s}}\, d\mu(y)\right).
\end{aligned} \tag{29}$$

We can write

$$\int_{\partial E_t} \frac{H_s(y)-H_s(x)}{|y-x|^{n+s}}\, d\mu(y) = \int_{\partial E_t} (u(y)-\alpha)\frac{W(y)-W(x)}{|y-x|^{n+s}}\, d\mu(y) + W(x)\int_{\partial E_t} \frac{u(y)-u(x)}{|y-x|^{n+s}}\, d\mu(y). \tag{30}$$

Observe also

$$\begin{aligned}
\int_{\partial E_t} \frac{u(y)-u(x)}{|y-x|^{n+s}}\, d\mu(y) &= \int_{\partial E_t} \frac{y\cdot\nu(y) - x\cdot\nu(x)}{|y-x|^{n+s}}\, d\mu(y) \\
&= \int_{\partial E_t} \frac{(y-x)\cdot\nu(y)}{|y-x|^{n+s}}\, d\mu(y) + \int_{\partial E_t} \frac{(x^T + u(x)\nu(x))\cdot(\nu(y)-\nu(x))}{|y-x|^{n+s}}\, d\mu(y) \\
&= \frac{1}{2(1-s)} H_s(x) + \int_{\partial E_t} \frac{x^T\cdot\nu(y)}{|y-x|^{n+s}}\, d\mu(y) - u(x)\int_{\partial E_t} \frac{1-\nu(x)\cdot\nu(y)}{|y-x|^{n+s}}\, d\mu(y).
\end{aligned} \tag{31}$$

From (30) and (31) we deduce that

$$\frac{1}{u(x)-\alpha}\int_{\partial E_t}\frac{H_s(y)-H_s(x)}{|y-x|^{n+s}}\,d\mu(y)-\frac{H_s(x)}{(u(x)-\alpha)^2}\int_{\partial E_t}\frac{x^T\cdot\nu(y)}{|y-x|^{n+s}}\,d\mu(y)$$
$$=\frac{1}{u(x)-\alpha}\int_{\partial E_t}\frac{(W(y)-W(x))(u(y)-\alpha)}{|y-x|^{n+s}}\,d\mu(y)+\frac{1}{2(1-s)}W^2-u(x)\frac{W(x)}{u(x)-\alpha}\int_{\partial E_t}\frac{1-\nu(x)\cdot\nu(y)}{|y-x|^{n+s}}\,d\mu(y).$$

We then conclude from (29)

$$\frac{\partial_t W}{2s(1-s)}<\frac{1}{u(x)-\alpha}\int_{\partial E_t}\frac{(W(y)-W(x))(u(y)-\alpha)}{|y-x|^{n+s}}\,d\mu(y)+\frac{1+s}{2s(1-s)}W^2-\alpha\frac{W(x)}{u(x)-\alpha}\int_{\partial E_t}\frac{1-\nu(x)\cdot\nu(y)}{|y-x|^{n+s}}\,d\mu(y). \tag{32}$$

Recalling the estimate of Proposition 4.2 and (28), we immediately obtain:

Corollary 5.1. *At any point where the spatial maximum for $W(\cdot,t)$ is attained, we have*

$$\partial_t W<C_1W^2-C_2W^3 \tag{33}$$

for constants C_1, C_2 only depending on n, s and the initial data.

We are now ready to prove the upper bound on the fractional mean curvature.

Theorem 5.2. *Let E_0 be a convex subset of $\mathbb{R}^n$ and $\mathcal{M}_0=\partial E_0$. Let $F:\mathcal{M}_0\times[0,T)\to\mathbb{R}^n$, with $0<T\leqslant+\infty$, be a solution of (1) of class $C^{2,\beta}$ for some $\beta>s$. Then there exists $K_2>0$, only depending on n, s, E_0, such that*

$$H_s(p,t)\leqslant K_2,\qquad p\in\mathcal{M}_0,$$

for all $t\in[0,T)$.

Proof. Let us take an arbitrary $t_0\in[0,T)$. We know from Lemma 3.3 that there exists $x_0\in\mathbb{R}^n$ such that $B_{R_1/2}(x_0)\subset E_t$ for any $t\in[t_0,t_0+t^*]$. In addition, setting $W=H_s(\langle x-x_0,\nu\rangle-R_1/4)^{-1}$, we know that the maximum of W satisfies inequality (33) in this time interval. We need a little care because the point x_0 depends on t_0 and therefore the function W is defined differently in different intervals.

Let us set for simplicity $F(w)=C_1w^2-C_2w^3$ to denote the right-hand side of (33). We observe that $F(w)<0$ for $w>C_1/C_2$. Let us denote by $\tilde w(t)$ the solution of the equation $\tilde w'(t)=F(\tilde w(t))$ defined for $t>0$ and satisfying $\tilde w(t)\to+\infty$ as $t\to 0^+$. It is easily seen that such a function exists and is implicitly defined by the formula

$$\int_{\tilde w(t)}^{+\infty}\frac{dw}{C_2w^3-C_1w^2}=t.$$

In addition, $\tilde w(t)$ is defined for all $t\in(0,+\infty)$ and decreases monotonically from $+\infty$ to C_1/C_2.

We now treat differently the cases $t_0=0$ and $t_0>0$. If $t_0=0$, using the sign properties of the right-hand side of (33), we obtain

$$W(p,t)\leqslant\max\left\{\max_{\mathcal{M}_0}W,\,\frac{C_1}{C_2}\right\},\qquad p\in\mathcal{M},\ t\in[0,t^*].$$

Taking into account (28), this implies

$$H_s(p,t) \leqslant C', \qquad p \in \mathcal{M},\ t \in [0, t^*], \tag{34}$$

for a suitable constant C'. If $t_0 > 0$, we observe instead that, again by (33),

$$W(p, t_0 + \tau) \leqslant \tilde{w}(\tau), \qquad \tau \in [0, t^*].$$

In particular, since $\tilde{w}$ is monotone,

$$W(p, t_0 + \tau) \leqslant \tilde{w}(t^*/2), \qquad \tau \in [t^*/2, t^*].$$

Using (28), it follows that

$$H_s(p,t) \leqslant C'', \qquad p \in \mathcal{M},\ t \in [t_0 + t^*/2, t_0 + t^*]. \tag{35}$$

By the arbitrariness of t_0, we conclude from (34)–(35) that $H_s(p,t) \leqslant K_2 := \max\{C', C''\}$ for all p, t. □

6. The case of a nonlinear speed

In this section we study a generalization of problem (1) in which the velocity is given by a general function of the fractional mean curvature. More precisely, we consider

$$\begin{cases} \partial_t F(p,t) = [-\Phi(H_s(p,t)) + \varphi(t)]\nu(p,t), & p \in \mathcal{M}_0,\ t \geqslant 0, \\ F(p,0) = p, & p \in \mathcal{M}_0, \end{cases} \tag{36}$$

where

$$\varphi(t) = \frac{1}{|\mathcal{M}_t|} \int_{\mathcal{M}_t} \Phi(H_s(x))\, d\mu.$$

We assume that $\Phi : [0, +\infty) \to [0, +\infty)$ is a C^2 function, satisfying the following properties:

(i) $\lim_{a \to +\infty} \Phi(a) = +\infty$.

(ii) $\Phi'(a) > 0$ for every $a > 0$.

(iii) $\lim_{a \to +\infty} \Phi'(a) a^2 / \Phi(a) = +\infty$.

Typical examples are functions of the form $\Phi(a) = a^p$ with $p > 0$, but hold in many other cases, e.g., $\Phi(a) = e^a$ or $\Phi(a) = \ln(a+1)$. Assumption (ii) ensures that $\Phi(H_s)$ satisfies the monotonicity assumption (A) in [Chambolle et al. 2015, Section 2] (monotonicity with respect to set inclusion). Hence, by [Chambolle et al. 2015, Theorem 2.21], problem (36) is well-posed and admits a viscosity solution, at least in the case $\varphi \equiv 0$ considered in that paper. In the case of a general $\Phi(H_s)$, the local existence result of smooth solutions is not yet known; there is also no result on the invariance of convexity, since the result in [Chambolle et al. 2017] does not apply. In the classical case, convexity is preserved under some additional structural hypotheses on Φ, see [Bertini and Sinestrari 2018; Andrews and Wei 2017], and it is likely that similar results hold in the fractional case. We will not address these issues here and we will assume instead a priori the existence of a convex smooth solution.

The aim of this section is to prove that Theorem 1.1 holds also for the more general problem (36). We first have the following lemma. As before, we denote by E_t the set enclosed by $\mathcal{M}_t$.

Lemma 6.1. *Flow (36) keeps the volume of E_t constant and decreases its fractional perimeter* $\mathrm{Per}_s(E_t)$.

Proof. The first part of the statement is an easy consequence of the choice of $\varphi(t)$. The second part follows exactly as in the proof of [Bertini and Sinestrari 2018, Lemma 3.1] in the local case. □

The uniform bounds on inner and outer radii and the lower bound for H_s are obtained exactly as for the $\Phi(H_s) = H_s$ case (see Section 3), since they just rely on convexity and on the fact that the flow preserves volume and decreases the s-perimeter. Hence, we immediately have the following:

Proposition 6.2. *Let $F : \mathcal{M}_0 \times [0, T) \to \mathbb{R}^n$, with $0 < T \leqslant +\infty$, be a smooth convex solution of (36). Then there exist positive constants $0 < R_1 \leqslant R_2$, only depending on E_0, such that*

$$R_1 \leqslant \underline{\rho}_{E_t} \leqslant \bar{\rho}_{E_t} \leqslant R_2 \quad \textit{for all } t \in [0, T).$$

In addition, there exists $K_1 > 0$ such that $H_s(p, t) \geqslant K_1$ for all $(p, t) \in \mathcal{M}_0 \times [0, T)$.

From the previous proposition, we deduce again that, by choosing a smaller radius, we can find a ball with fixed center which remains inside E_t for a time interval with fixed length; that is, Lemma 3.3 holds also for solutions of the nonlinear flow (36). The proof of this fact is again by a comparison argument and we refer to [Bertini and Sinestrari 2018, Lemma 3.6] for the details.

In order to prove the analogue of Theorem 1.1 for the flow (36), it remains to establish the upper bound on H_s.

Proposition 6.3. *Let $F : \mathcal{M}_0 \times [0, T) \to \mathbb{R}^n$, with $0 < T \leqslant +\infty$, be a smooth convex solution of (36). We have that, at any time $t \in [0, T)$,*

$$\Phi(H_s) \leqslant K_3,$$

where K_3 is a positive constant depending only on n, s, and E_0.

Proof. The proof is similar to the one given in Section 5. We show in detail how the argument is adapted to the case of a general speed, for the sake of clarity. We consider again the support function

$$u(p, t) = \langle F(p, t), \nu(p, t)\rangle.$$

If now F evolves according to (36), recalling Lemma 2.2(iii) and the expression for $\nabla^{\mathcal{M}} H_s$, we have

$$\begin{aligned} \partial_t u &= \langle \partial_t F, \nu\rangle + \langle F, \partial_t \nu\rangle \\ &= -\Phi(H_s) + \varphi(t) + \Phi'(H_s)\langle F, \nabla^{\mathcal{M}} H_s\rangle \\ &= -\Phi(H_s) + \varphi(t) + 2s(1-s)\Phi'(H_s) \int_{\mathcal{M}_t} \frac{x^T \cdot \nu(y)}{|x-y|^{n+s}} \, d\mu(y). \end{aligned} \tag{37}$$

Moreover, using Lemma 2.2(ii), we have that the fractional mean curvature satisfies

$$\frac{\partial_t H_s(x)}{2s(1-s)} = \int_{\mathcal{M}_t} \frac{\Phi(H_s(y)) - \Phi(H_s(x))}{|x-y|^{n+s}} \, d\mu(y) + (\Phi(H_s(x) - \varphi(t)) \int_{\mathcal{M}_t} \frac{1 - \nu(x)\cdot\nu(y)}{|x-y|^{n+s}} \, d\mu(y). \tag{38}$$

We define, much as before, but with the new velocity $\Phi(H_s)$,

$$W = \frac{\Phi(H_s)}{u(x) - \alpha},$$

where α is chosen in the same way as in Section 5. We have that

$$
\begin{aligned}
\frac{\partial_t W}{2s(1-s)} &= \frac{1}{2s(1-s)}\left[\frac{\Phi'(H_s)\partial_t H_s}{u(x)-\alpha} - \frac{\Phi(H_s)\partial_t u}{(u(x)-\alpha)^2}\right] \\
&= \frac{\Phi'(H_s(x))}{u(x)-\alpha}\left[\int_{\mathcal{M}_t} \frac{\Phi(H_s(y))-\Phi(H_s(x))}{|x-y|^{n+s}}\,d\mu(y) + (\Phi(H_s(x))-\varphi(t))\int_{\mathcal{M}_t}\frac{1-\nu(x)\cdot\nu(y)}{|x-y|^{n+s}}\,d\mu(y)\right] \\
&\qquad - \frac{\Phi(H_s(x))}{(u(x)-\alpha)^2}\left[\frac{-\Phi(H_s(x))+\varphi(t)}{2s(1-s)} + \Phi'(H_s(x))\int_{\mathcal{M}_t}\frac{x^T\cdot\nu(y)}{|x-y|^{n+s}}\,d\mu(y)\right] \\
&< \frac{\Phi'(H_s(x))}{u(x)-\alpha}\left[\int_{\mathcal{M}_t} \frac{\Phi(H_s(y))-\Phi(H_s(x))}{|x-y|^{n+s}}\,d\mu(y) + \Phi(H_s(x))\int_{\mathcal{M}_t}\frac{1-\nu(x)\cdot\nu(y)}{|x-y|^{n+s}}\,d\mu(y)\right] \\
&\qquad - \frac{\Phi(H_s(x))}{(u(x)-\alpha)^2}\left[\frac{-\Phi(H_s(x))}{2s(1-s)} + \Phi'(H_s(x))\int_{\mathcal{M}_t}\frac{x^T\cdot\nu(y)}{|x-y|^{n+s}}\,d\mu(y)\right]. \qquad (39)
\end{aligned}
$$

By the definition of W we have that

$$
\begin{aligned}
&\int_{\mathcal{M}_t} \frac{\Phi(H_s(y))-\Phi(H_s(x))}{|x-y|^{n+s}}\,d\mu(y) \\
&\qquad = \int_{\mathcal{M}_t} (u(y)-\alpha)\frac{(W(y)-W(x))}{|x-y|^{n+s}}\,d\mu(y) + W(x)\int_{\mathcal{M}_t}\frac{u(y)-u(x)}{|x-y|^{n+s}}\,d\mu(y). \qquad (40)
\end{aligned}
$$

Moreover, formula (31) holds unchanged, since it is independent of the velocity:

$$
\int_{\mathcal{M}_t}\frac{u(y)-u(x)}{|x-y|^{n+s}}\,d\mu(y) = \frac{1}{2(1-s)}H_s(x) + \int_{\mathcal{M}_t}\frac{x^T\cdot\nu(y)}{|x-y|^{n+s}}\,d\mu(y) - u(x)\int_{\mathcal{M}_t}\frac{1-\nu(x)\cdot\nu(y)}{|x-y|^{n+s}}\,d\mu(y). \qquad (41)
$$

We combine now (40) and (41) to get

$$
\begin{aligned}
&\frac{1}{u(x)-\alpha}\int_{\mathcal{M}_t}\frac{\Phi(H_s(y))-\Phi(H_s(x))}{|x-y|^{n+s}}\,d\mu(y) \\
&= \frac{1}{u(x)-\alpha}\int_{\mathcal{M}_t}(u(y)-\alpha)\frac{W(y)-W(x)}{|x-y|^{n+s}}\,d\mu(y) \\
&\qquad + \frac{W(x)}{u(x)-\alpha}\left[\frac{1}{2(1-s)}H_s(x) + \int_{\mathcal{M}_t}\frac{x^T\cdot\nu(y)}{|x-y|^{n+s}}\,d\mu(y) - u(x)\int_{\mathcal{M}_t}\frac{1-\nu(x)\cdot\nu(y)}{|x-y|^{n+s}}\,d\mu(y)\right] \\
&= \frac{1}{u(x)-\alpha}\int_{\mathcal{M}_t}(u(y)-\alpha)\frac{W(y)-W(x)}{|x-y|^{n+s}}\,d\mu(y) + \frac{H_s(x)\Phi(H_s(x))}{2(1-s)(u(x)-\alpha)^2} \\
&\qquad + \frac{W(x)}{u(x)-\alpha}\int_{\mathcal{M}_t}\frac{x^T\cdot\nu(y)}{|x-y|^{n+s}}\,d\mu(y) - \frac{W(x)}{u(x)-\alpha}u(x)\int_{\mathcal{M}_t}\frac{1-\nu(x)\cdot\nu(y)}{|x-y|^{n+s}}\,d\mu(y). \qquad (42)
\end{aligned}
$$

Finally, plugging (42) into (39), we obtain

$$\frac{\partial_t W}{2s(1-s)} < \Phi'(H_s)\left[\frac{1}{u(x)-\alpha}\int_{\mathcal{M}_t}\frac{(W(y)-W(x))(u(y)-\alpha)}{|x-y|^{n+s}}\,d\mu(y)\right] + \frac{W^2}{2s(1-s)} + \Phi'(H_s)W\left[\frac{H_s(x)}{2(1-s)(u(x)-\alpha)} - \frac{\alpha}{u(x)-\alpha}\int_{\mathcal{M}_t}\frac{1-\nu(x)\cdot\nu(y)}{|x-y|^{n+s}}\,d\mu(y)\right]. \quad (43)$$

This inequality is the analogue of estimate (32) in the presence of a nonlinear speed Φ. Again, we use Proposition 4.2 to bound the last term and we get

$$\frac{\partial_t W}{2s(1-s)} < \Phi'(H_s)\left[\frac{1}{u(x)-\alpha}\int_{\mathcal{M}_t}\frac{(W(y)-W(x))(u(y)-\alpha)}{|x-y|^{n+s}}\,d\mu(y)\right] + C_1W^2 + \frac{W\Phi'(H_s)H_s}{(u(x)-\alpha)}[C_2-C_3H_s]. \quad (44)$$

Setting $\widetilde{W}(t) = \sup_{\mathcal{M}_t} W(x,t)$, we have

$$\partial_t\widetilde{W}(t) \leqslant C_1\widetilde{W}^2 + \frac{\widetilde{W}\Phi'(H_s)H_s}{u-\alpha}[C_2 - C_3H_s],$$

where $H_s = H_s(\tilde{x}, t)$ for a suitable $\tilde{x}$ such that $W(\tilde{x}, t) = \widetilde{W}(t)$.

We choose now $K > 3C_2/C_3$, so that $H_s \geqslant K$ implies $C_2 - C_3H_s \leqslant -2C_3H_s/3$. Suppose now that there exists t^* such that $\widetilde{W}(t^*) \geqslant \Phi(K)/\alpha$. Recalling that $u-\alpha \geqslant \alpha$ and using the monotonicity of Φ, we deduce that $H_s(x^*, t^*) \geqslant K$ for any x^* such that $W(x^*, t^*) = \widetilde{W}(t^*)$. Hence, at $t = t^*$, we have

$$\partial_t\widetilde{W} \leqslant C_1\widetilde{W}^2 - \frac{2C_3\widetilde{W}\Phi'(H_s)H_s^2}{3(u-\alpha)} \leqslant \widetilde{W}^2\left[C_1 - \frac{2C_3}{3}\frac{\Phi'(H_s)H_s^2}{\Phi(H_s)}\right].$$

By property (iii) of Φ, we can choose K large enough so that if $H_s \geqslant K$ we have

$$C_1 - \frac{2C_3}{3}\frac{\Phi'(H_s)H_s^2}{\Phi(H_s)} < -1,$$

which gives

$$\partial_t\widetilde{W} \leqslant -\widetilde{W}^2.$$

From this last estimate, the conclusion follows by a comparison argument, exactly as in the proof of [Bertini and Sinestrari 2018, Proposition 3.7]. □

As a consequence of the boundedness of the speed $\Phi(H_s)$ and of property (i) satisfied by Φ, and recalling Proposition 6.2, we deduce the following:

Corollary 6.4. *We have that H_s is uniformly bounded in $(0, T)$.*

7. Convergence to a sphere

In this section we prove our convergence result (Theorem 1.2 for the case $\Phi(H_s) = H_s$), which for the general problem (36) reads as follows:

Theorem 7.1. *Let $F : \mathcal{M}_0 \times [0, T) \to \mathbb{R}^n$, with $0 < T \leqslant +\infty$, be a smooth convex solution of (36) of class $C^{2,\beta}$ for some $\beta > s$ which satisfies property (R). Then $T = +\infty$, and $\mathcal{M}_t$ converges to a round sphere as $t \to +\infty$ in $C^{2,\beta}$ norm, possibly up to translations.*

We first observe that, by the lower and upper bounds on H_s, we have that $\Phi'(H_s)$ is bounded from above and below by positive constants for every $t \in [0, +\infty)$.

The crucial step in the proof of Theorem 7.1 is the following result.

Proposition 7.2. *Under our assumption, we have that*

$$\lim_{t\to+\infty} \max_{\mathcal{M}_t} |\Phi(H_s)(x) - \varphi(t)| = 0.$$

Proof. The proof follows the one in [Bertini and Sinestrari 2018, Proposition 4.4]. For any t, let $\overline{H}_s(t)$ be such that $\Phi(\overline{H}_s(t)) = \varphi(t)$. Then, recalling (8), we have

$$\begin{aligned}\frac{d}{dt}\operatorname{Per}_s(E_t) &= \int_{\mathcal{M}_t} H_s\varphi\, d\mu - \int_{\mathcal{M}_t} H_s\Phi(H_s)\, d\mu \\ &= \int_{\mathcal{M}_t} (H_s - \overline{H}_s)(\Phi(\overline{H}_s) - \Phi(H_s))\, d\mu \\ &= -\int_{\mathcal{M}_t} |H_s - \overline{H}_s||\Phi(\overline{H}_s) - \Phi(H_s)|\, d\mu.\end{aligned}$$

Hence, using the boundedness of Φ', we deduce that

$$\frac{d}{dt}\operatorname{Per}_s(E_t) \leqslant -\frac{1}{\sup \Phi'}\int_{\mathcal{M}_t} |\Phi(\overline{H}_s) - \Phi(H_s)|^2\, d\mu = -\frac{1}{\sup \Phi'}\int_{\mathcal{M}_t} |\Phi(\overline{H}_s) - \varphi|^2\, d\mu.$$

Suppose now, by contradiction, that there exists $\varepsilon > 0$ such that $|\Phi(H_s) - \varphi| = \varepsilon$ at some point $(\bar{p}, \bar{t})$. By our regularity assumption and using Theorem 2.1, we have that H_s is uniformly Lipschitz; therefore there exists a uniform radius $r(\varepsilon) > 0$ for which

$$|\Phi(H_s) - \varphi| > \frac{\varepsilon}{2} \quad \text{in } B((\bar{p}, \bar{t}), r(\varepsilon)),$$

which implies

$$\frac{d}{dt}\operatorname{Per}_s(E_t) \leqslant -\eta(\varepsilon) \quad \text{for any } t \in [\bar{t} - r(\varepsilon), \bar{t} + r(\varepsilon)],$$

for some $\eta > 0$. The fact that $\operatorname{Per}_s(E_t) > 0$ and is decreasing in time implies that the above property cannot hold for $\bar{t}$ arbitrarily large. This shows that $|\Phi(H_s) - \varphi|$ tends to zero uniformly. □

We are now ready to give the proof of our convergence result.

Proof of Theorem 7.1. Using our regularity assumption (R) and the uniform bounds for H_s of Corollary 6.4 and Theorem 1.1, we deduce that the flow exists for all $t \in [0, \infty)$ and that the hypersurfaces $\mathcal{M}_t$, possibly up to translations, are bounded in the $C^{2,\beta}$ norm uniformly in t. Hence, the $\mathcal{M}_t$ are precompact in $C^{2,\beta'}$ for $\beta' < \beta$. By Proposition 7.2 and the stability results of [Cozzi 2015], we have that any possible subsequential limit as $t \to +\infty$ has constant fractional curvature. Then Theorem 2.4 ensures that the limit is a ball, with radius uniquely determined by the volume constraint. The uniqueness of the subsequential limit easily implies that the whole family $\mathcal{M}_t$ converges to a sphere as $t \to +\infty$. □

Acknowledgements

Cinti was supported by MINECO grant MTM2014-52402-C3-1-P and is part of the Catalan research group 2014 SGR 1083. Cinti and Sinestrari were supported by the group GNAMPA of INdAM (Istituto Nazionale di Alta Matematica). Cinti and Valdinoci were supported by the European Research Council Starting Grant "EPSILON" (Elliptic PDEs and symmetry of interfaces and layers for odd nonlinearities) no. 277749. Valdinoci was supported by the Australian Research Council Discovery Project "NEW" (Nonlocal equations at work) no. DP170104880. The authors are members of INdAM/GNAMPA.

References

[Abatangelo and Valdinoci 2014] N. Abatangelo and E. Valdinoci, "A notion of nonlocal curvature", *Numer. Funct. Anal. Optim.* **35**:7-9 (2014), 793–815. MR Zbl

[Andrews 1994] B. Andrews, "Contraction of convex hypersurfaces in Euclidean space", *Calc. Var. Partial Differential Equations* **2**:2 (1994), 151–171. MR Zbl

[Andrews 2001] B. Andrews, "Volume-preserving anisotropic mean curvature flow", *Indiana Univ. Math. J.* **50**:2 (2001), 783–827. MR Zbl

[Andrews and Wei 2017] B. Andrews and Y. Wei, "Volume-preserving flows by powers of the k-th mean curvature", preprint, 2017. arXiv

[Andrews et al. 2013] B. Andrews, J. McCoy, and Y. Zheng, "Contracting convex hypersurfaces by curvature", *Calc. Var. Partial Differential Equations* **47**:3-4 (2013), 611–665. MR Zbl

[Andrews et al. 2018] B. Andrews, X. Chen, and Y. Wei, "Volume preserving flow and Alexandrov–Fenchel type inequalities in hyperbolic space", preprint, 2018. arXiv

[Barrios et al. 2014] B. Barrios, A. Figalli, and E. Valdinoci, "Bootstrap regularity for integro-differential operators and its application to nonlocal minimal surfaces", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **13**:3 (2014), 609–639. MR Zbl

[Bertini and Pipoli 2017] M. C. Bertini and G. Pipoli, "Volume preserving non-homogeneous mean curvature flow in hyperbolic space", *Differential Geom. Appl.* **54**:B (2017), 448–463. MR Zbl

[Bertini and Sinestrari 2018] M. C. Bertini and C. Sinestrari, "Volume-preserving nonhomogeneous mean curvature flow of convex hypersurfaces", *Ann. Mat. Pura Appl.* (4) **197**:4 (2018), 1295–1309. MR Zbl

[Bosch and Stoll 2015] J. Bosch and M. Stoll, "A fractional inpainting model based on the vector-valued Cahn–Hilliard equation", *SIAM J. Imaging Sci.* **8**:4 (2015), 2352–2382. MR Zbl

[Cabré et al. 2018a] X. Cabré, M. M. Fall, J. Solà-Morales, and T. Weth, "Curves and surfaces with constant nonlocal mean curvature: meeting Alexandrov and Delaunay", *J. Reine Angew. Math.* **745** (2018), 253–280. MR Zbl

[Cabré et al. 2018b] X. Cabré, M. M. Fall, and T. Weth, "Delaunay hypersurfaces with constant nonlocal mean curvature", *J. Math. Pures Appl.* (9) **110** (2018), 32–70. MR Zbl

[Cabré et al. 2018c] X. Cabré, M. M. Fall, and T. Weth, "Near-sphere lattices with constant nonlocal mean curvature", *Math. Ann.* **370**:3-4 (2018), 1513–1569. MR Zbl

[Cabré et al. 2020] X. Cabré, E. Cinti, and J. Serra, "Stable s-minimal cones in $\mathbb{R}^3$ are flat for $s \sim 1$", *J. Reine Angew. Math.* **764** (2020), 157–180. MR Zbl

[Caffarelli and Souganidis 2010] L. A. Caffarelli and P. E. Souganidis, "Convergence of nonlocal threshold dynamics approximations to front propagation", *Arch. Ration. Mech. Anal.* **195**:1 (2010), 1–23. MR Zbl

[Caffarelli and Valdinoci 2013] L. Caffarelli and E. Valdinoci, "Regularity properties of nonlocal minimal surfaces via limiting arguments", *Adv. Math.* **248** (2013), 843–871. MR Zbl

[Caffarelli et al. 2010] L. Caffarelli, J.-M. Roquejoffre, and O. Savin, "Nonlocal minimal surfaces", *Comm. Pure Appl. Math.* **63**:9 (2010), 1111–1144. MR Zbl

[Cesaroni et al. 2019] A. Cesaroni, S. Dipierro, M. Novaga, and E. Valdinoci, "Fattening and nonfattening phenomena for planar nonlocal curvature flows", *Math. Ann.* **375**:1-2 (2019), 687–736. MR Zbl

[Chambolle et al. 2015] A. Chambolle, M. Morini, and M. Ponsiglione, "Nonlocal curvature flows", *Arch. Ration. Mech. Anal.* **218**:3 (2015), 1263–1329. MR Zbl

[Chambolle et al. 2017] A. Chambolle, M. Novaga, and B. Ruffini, "Some results on anisotropic fractional mean curvature flows", *Interfaces Free Bound.* **19**:3 (2017), 393–415. MR Zbl

[Chang-Lara and Dávila 2014a] H. Chang-Lara and G. Dávila, "Regularity for solutions of non local parabolic equations", *Calc. Var. Partial Differential Equations* **49**:1-2 (2014), 139–172. MR Zbl

[Chang-Lara and Dávila 2014b] H. Chang-Lara and G. Dávila, "Regularity for solutions of nonlocal parabolic equations, II", *J. Differential Equations* **256**:1 (2014), 130–156. MR Zbl

[Chen et al. 1991] Y. G. Chen, Y. Giga, and S. Goto, "Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations", *J. Differential Geom.* **33**:3 (1991), 749–786. MR Zbl

[Chow and Gulliver 1996] B. Chow and R. Gulliver, "Aleksandrov reflection and nonlinear evolution equations, I: The n-sphere and n-ball", *Calc. Var. Partial Differential Equations* **4**:3 (1996), 249–264. MR Zbl

[Cinti et al. 2016] E. Cinti, J. Davila, and M. Del Pino, "Solutions of the fractional Allen–Cahn equation which are invariant under screw motion", *J. Lond. Math. Soc.* (2) **94**:1 (2016), 295–313. MR Zbl

[Cinti et al. 2018] E. Cinti, C. Sinestrari, and E. Valdinoci, "Neckpinch singularities in fractional mean curvature flows", *Proc. Amer. Math. Soc.* **146**:6 (2018), 2637–2646. MR Zbl

[Cinti et al. 2019] E. Cinti, J. Serra, and E. Valdinoci, "Quantitative flatness results and BV-estimates for stable nonlocal minimal surfaces", *J. Differential Geom.* **112**:3 (2019), 447–504. MR Zbl

[Ciraolo et al. 2018] G. Ciraolo, A. Figalli, F. Maggi, and M. Novaga, "Rigidity and sharp stability estimates for hypersurfaces with constant and almost-constant nonlocal mean curvature", *J. Reine Angew. Math.* **741** (2018), 275–294. MR Zbl

[Cozzi 2015] M. Cozzi, "On the variation of the fractional mean curvature under the effect of $C^{1,\alpha}$ perturbations", *Discrete Contin. Dyn. Syst.* **35**:12 (2015), 5769–5786. MR Zbl

[Dávila et al. 2016] J. Dávila, M. del Pino, S. Dipierro, and E. Valdinoci, "Nonlocal Delaunay surfaces", *Nonlinear Anal.* **137** (2016), 357–380. MR Zbl

[Dávila et al. 2018] J. Dávila, M. del Pino, and J. Wei, "Nonlocal s-minimal surfaces and Lawson cones", *J. Differential Geom.* **109**:1 (2018), 111–175. MR Zbl

[Dipierro et al. 2017] S. Dipierro, O. Savin, and E. Valdinoci, "Boundary behavior of nonlocal minimal surfaces", *J. Funct. Anal.* **272**:5 (2017), 1791–1851. MR Zbl

[Dipierro et al. 2020] S. Dipierro, O. Savin, and E. Valdinoci, "Nonlocal minimal graphs in the plane are generically sticky", *Comm. Math. Phys.* **376**:3 (2020), 2005–2063. MR Zbl

[Evans and Spruck 1991] L. C. Evans and J. Spruck, "Motion of level sets by mean curvature, I", *J. Differential Geom.* **33**:3 (1991), 635–681. MR Zbl

[Figalli et al. 2015] A. Figalli, N. Fusco, F. Maggi, V. Millot, and M. Morini, "Isoperimetry and stability properties of balls with respect to nonlocal energies", *Comm. Math. Phys.* **336**:1 (2015), 441–507. MR Zbl

[Frank et al. 2008] R. L. Frank, E. H. Lieb, and R. Seiringer, "Hardy–Lieb–Thirring inequalities for fractional Schrödinger operators", *J. Amer. Math. Soc.* **21**:4 (2008), 925–950. MR Zbl

[Guan and Li 2009] P. Guan and J. Li, "The quermassintegral inequalities for k-convex starshaped domains", *Adv. Math.* **221**:5 (2009), 1725–1732. MR Zbl

[Huisken 1984] G. Huisken, "Flow by mean curvature of convex surfaces into spheres", *J. Differential Geom.* **20**:1 (1984), 237–266. MR Zbl

[Huisken 1987] G. Huisken, "The volume preserving mean curvature flow", *J. Reine Angew. Math.* **382** (1987), 35–48. MR Zbl

[Huisken and Polden 1999] G. Huisken and A. Polden, "Geometric evolution equations for hypersurfaces", pp. 45–84 in *Calculus of variations and geometric evolution problems* (Cetraro, 1996), edited by S. Hildebrandt and M. Struwe, Lecture Notes in Math. **1713**, Springer, 1999. MR Zbl

[Imbert 2009] C. Imbert, "Level set approach for fractional mean curvature flows", *Interfaces Free Bound.* **11**:1 (2009), 153–176. MR Zbl

[Julin and La Manna 2020] V. Julin and D. A. La Manna, "Short time existence of the classical solution to the fractional mean curvature flow", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **37**:4 (2020), 983–1016. MR Zbl

[Maggi and Valdinoci 2017] F. Maggi and E. Valdinoci, "Capillarity problems with nonlocal surface tension energies", *Comm. Partial Differential Equations* **42**:9 (2017), 1403–1446. MR Zbl

[McCoy 2004] J. A. McCoy, "The mixed volume preserving mean curvature flow", *Math. Z.* **246**:1-2 (2004), 155–166. MR Zbl

[McCoy 2005] J. A. McCoy, "Mixed volume preserving curvature flows", *Calc. Var. Partial Differential Equations* **24**:2 (2005), 131–154. MR Zbl

[Sáez and Valdinoci 2019] M. Sáez and E. Valdinoci, "On the evolution by fractional mean curvature", *Comm. Anal. Geom.* **27**:1 (2019), 211–249. MR Zbl

[Savin and Valdinoci 2013] O. Savin and E. Valdinoci, "Regularity of nonlocal minimal cones in dimension 2", *Calc. Var. Partial Differential Equations* **48**:1-2 (2013), 33–39. MR Zbl

[Schulze 2008] F. Schulze, "Nonlinear evolution by mean curvature and isoperimetric inequalities", *J. Differential Geom.* **79**:2 (2008), 197–241. MR Zbl

[Sinestrari 2015] C. Sinestrari, "Convex hypersurfaces evolving by volume preserving curvature flows", *Calc. Var. Partial Differential Equations* **54**:2 (2015), 1985–1993. MR Zbl

[Tso 1985] K. Tso, "Deforming a hypersurface by its Gauss–Kronecker curvature", *Comm. Pure Appl. Math.* **38**:6 (1985), 867–882. MR Zbl

Received 21 Nov 2018. Revised 19 Jul 2019. Accepted 6 Sep 2019.

ELEONORA CINTI: eleonora.cinti5@unibo.it
Dipartimento di Matematica, Università degli Studi di Bologna, Bologna, Italy

CARLO SINESTRARI: sinestra@mat.uniroma2.it
Dipartimento di Ingegneria Civile e Ingegneria Informatica, Università di Roma "Tor Vergata", Rome, Italy

ENRICO VALDINOCI: enrico.valdinoci@uwa.edu.au
Department of Mathematics and Statistics, University of Western Australia, Crawley, WA, Australia

ANALYSIS AND PDE
Vol. 13, No. 7, 2020
https://doi.org/10.2140/apde.2020.13.2173

msp

C^*-ALGEBRAS ISOMORPHICALLY REPRESENTABLE ON l^p

MARCH T. BOEDIHARDJO

Let $p \in (1, \infty)\backslash\{2\}$. We show that every homomorphism from a C^*-algebra $\mathcal{A}$ into $B(l^p(J))$ satisfies a compactness property where J is any set. As a consequence, we show that a C^*-algebra $\mathcal{A}$ is isomorphic to a subalgebra of $B(l^p(J))$, for some set J, if and only if $\mathcal{A}$ is residually finite-dimensional.

1. Introduction

For $1 \le p < \infty$ and a set J, let $l^p(J)$ be the space

$$\left\{ f : J \to \mathbb{C} : \sum_{j \in J} |f(j)|^p < \infty \right\}$$

with norm

$$\|f\| = \left(\sum_{j \in J} |f(j)|^p \right)^{\frac{1}{p}}.$$

Two Banach algebras $\mathcal{A}_1$ and $\mathcal{A}_2$ are *isomorphic* if there exist a bijective homomorphism $\phi : \mathcal{A}_1 \to \mathcal{A}_2$ and $C > 0$ such that

$$\frac{1}{C}\|a\| \le \|\phi(a)\| \le C\|a\|$$

for all $a \in \mathcal{A}_1$. The algebras $\mathcal{A}_1$ and $\mathcal{A}_2$ are *isometrically isomorphic* if, moreover, ϕ can be chosen so that $\|\phi(a)\| = \|a\|$ for all $a \in \mathcal{A}_1$.

Gardella and Thiel [2020] showed that for $p \in [1, \infty)\backslash\{2\}$, a C^*-algebra $\mathcal{A}$ is isometrically isomorphic to a subalgebra of $B(l^p(J))$, for some set J, if and only if $\mathcal{A}$ is commutative. So it is natural to consider the question of whether this result holds if we relax the condition of isometrically isomorphic to isomorphic. In this paper, we show that for $p \in (1, \infty)\backslash\{2\}$, a C^*-algebra $\mathcal{A}$ is isomorphic to a subalgebra of $B(l^p(J))$, for some set J, if and only if $\mathcal{A}$ is residually finite-dimensional (Corollary 2.2). We prove this by showing that every homomorphism from a C^*-algebra $\mathcal{A}$ into $B(l^p(J))$ satisfies a compactness property (Theorem 2.1).

The proofs of the main results Theorem 2.1 and Corollary 2.2 in this paper are quite different from the proof of Gardella and Thiel's result. Lamperti's characterization [1958] of isometries on L^p, for $p \ne 2$, plays a crucial role in the proof of Gardella and Thiel's result, while uniform convexity of l^p, for $1 < p < \infty$, and an argument in probability that imitates the proof of Khintchine's inequality [Lindenstrauss and Tzafriri 1977, Theorem 2.b.3], for $p = 1$, are used in the proof of Theorem 2.1.

MSC2010: 46H20.
Keywords: l^p space, C^*-algebra.

2. Main results and proofs

Throughout this paper, the scalar field is $\mathbb{C}$. For algebras $\mathcal{A}_1$ and $\mathcal{A}_2$, a *homomorphism* $\phi : \mathcal{A}_1 \to \mathcal{A}_2$ is a bounded linear map such that $\phi(a_1a_2) = \phi(a_1)\phi(a_2)$ for all $a_1, a_2 \in \mathcal{A}$. For an element a of a C^*-algebra, $|a| = \sqrt{a^*a}$. The algebra of bounded linear operators on a Banach space $\mathcal{X}$ is denoted by $B(\mathcal{X})$ and the dual of $\mathcal{X}$ is denoted by $\mathcal{X}^*$. For $1 \le p \le \infty$, the l^p direct sum of Banach spaces $\mathcal{X}_\alpha$, for $\alpha \in \Lambda$, is denoted by $\left(\bigoplus_{\alpha\in\Lambda} \mathcal{X}_\alpha\right)_{l^p}$. Two Banach spaces $\mathcal{X}_1$ and $\mathcal{X}_2$ are *isomorphic* if there is an invertible operator $S : \mathcal{X}_1 \to \mathcal{X}_2$. A C^*-algebra $\mathcal{A}$ is *residually finite-dimensional* if for every $a \in \mathcal{A}$, there is a $*$-representation ϕ of $\mathcal{A}$ on a finite-dimensional space such that $\phi(a) \neq 0$.

Theorem 2.1. *Let $p \in (1, \infty)\backslash\{2\}$. Let J be a set. Let $\mathcal{A}$ be a C^*-algebra. Let $\phi : \mathcal{A} \to B(l^p(J))$ be a homomorphism. Then*

(i) *the norm closure of $\{\phi(a)x : a \in \mathcal{A},\ \|a\| \le 1\}$ in $l^p(J)$ is norm compact for every $x \in l^p(J)$, and*

(ii) *$\mathcal{A}/\ker\phi$ is a residually finite-dimensional C^*-algebra.*

Corollary 2.2. *Let $p \in (1, \infty)\backslash\{2\}$. A C^*-algebra $\mathcal{A}$ is isomorphic to a subalgebra of $B(l^p(J))$, for some set J, if and only if $\mathcal{A}$ is residually finite-dimensional.*

Theorem 2.1 and Corollary 2.2 will be proved at the end of this section after a series of lemmas are proved. Theorem 2.1 has an easier proof when ϕ is contractive. Indeed, if $\phi : \mathcal{A} \to B(l^p(J))$ is a contractive homomorphism, then the range of ϕ is in the algebra of diagonal operators on $l^p(J)$ by [Blecher and Phillips 2019, Proposition 2.12] (or by [Gardella and Thiel 2020, Lemma 5.2] when J is countable). Thus, $\{\phi(a)x : a \in \mathcal{A},\ \|a\| \le 1\}$ is norm relatively compact, for every $x \in l^p(J)$, and $\mathcal{A}/\ker\phi$ is commutative.

It is not known if Theorem 2.1 and Corollary 2.2 hold for $p = 1$. However, throughout their proofs, we use, in an essential way, the assumption that p is in the reflexive range. For example, in the proof of Theorem 2.1(i), we use the fact that every bounded sequence in $l^p(J)$ has a weakly convergent subsequence. In the proof of Corollary 2.2, we use a classical result of Pełczyński that the l^p direct sum of finite-dimensional Hilbert spaces is isomorphic to $l^p(J)$ for some set J. This result of Pełczyński holds only when p is in the reflexive range.

The structure of the proof of Theorem 2.1(i) goes as follows: If the closure of $\{\phi(a)x_0 : a \in \mathcal{A},\ \|a\| \le 1\}$ is not compact for some $x_0 \in l^p(J)$, then we can find a bounded sequence in $(b_k)_{k\in\mathbb{N}}$ in $\mathcal{A}$ such that $\phi(b_k)x_0 \to 0$ weakly, as $k \to \infty$, and $\inf_{k\in\mathbb{N}} \|\phi(b_k)x_0\| > 0$. Assume that $p > 2$. In Lemma 2.5, we show that $\phi(b_k) \to 0$ weakly implies that $\omega(b_k^*b_k) \to 0$ for all positive linear functionals $\omega : \mathcal{A} \to \mathbb{C}$ of the form $\omega(a) = y_0^*(\phi(a)x_0)$. This is proved by considering $\sum_{k=1}^n \delta_k b_k$ for random $\delta_1, \ldots, \delta_n$ in $\{-1, 1\}$ and by exploiting $p > 2$. Lemma 2.9 says that when $y_0^* \in (l^p(J))^*$ is suitably chosen, $\omega(b_k^*b_k) \to 0$ implies that $\|\phi(b_k)x_0\| \to 0$, which contradicts $\inf_{k\in\mathbb{N}} \|\phi(b_k)x\| > 0$. This is proved by using the uniform convexity of $l^p(J)$.

Theorem 2.1(ii) follows from Theorem 2.1(i) by using a GNS-type construction and a classical result about compact unitary representations of groups on Hilbert spaces.

The following two lemmas are needed for the proof of Lemma 2.5.

Lemma 2.3. *Let $\mathcal{A}$ be a unital C^*-algebra. Let $a \in \mathcal{A}$. Then there exists a sequence $(c_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ such that $\|c_n\| \le 1$ for all $n \in \mathbb{N}$ and $|a| = \lim_{n\to\infty} c_n a$.*

Proof. Without loss of generality, we may assume that $\|a\| \le 1$. For $n \in \mathbb{N}$, define $g_n \in C[0, 1]$ by

$$g_n(x) = \begin{cases} \frac{1}{\sqrt{x}}, & \frac{1}{n} \le x \le 1, \\ n\sqrt{n}x, & 0 \le x \le \frac{1}{n}. \end{cases}$$

Take $c_n = g_n(a^*a)a^*$. Then $c_n c_n^* = g_n(a^*a)a^*ag_n(a^*a)$. Note that

$$xg_n(x)^2 = \begin{cases} 1, & \frac{1}{n} \le x \le 1, \\ n^3x^3, & 0 \le x \le \frac{1}{n}. \end{cases}$$

Thus, $0 \le xg_n(x)^2 \le 1$ for all $x \in [0, 1]$ and so $0 \le c_n c_n^* \le 1$. Hence $\|c_n\| \le 1$.

We have

$$xg_n(x) = \begin{cases} \sqrt{x}, & \frac{1}{n} \le x \le 1, \\ n\sqrt{n}x^2, & 0 \le x \le \frac{1}{n}, \end{cases}$$

and so

$$|xg_n(x) - \sqrt{x}| \le \frac{1}{\sqrt{n}} \quad \text{for all } x \in [0, 1].$$

Since $c_n a = g_n(a^*a)a^*a$, it follows that

$$\|c_n a - \sqrt{a^*a}\| \le \frac{1}{\sqrt{n}}.$$

Thus, the result follows. □

Lemma 2.4. *Let $\mathcal{A}$ be a unital C^*-algebra. Let ω be a positive linear functional on $\mathcal{A}$. Let $a \in \mathcal{A}$. If $a \ge 0$ then*

$$\omega(a^2) \le \omega(a)^{\frac{2}{3}}\omega(a^4)^{\frac{1}{3}}.$$

Proof. There exists a measure μ on $[0, \|a\|]$ such that

$$\omega(f(a)) = \int f(x)\, d\mu(x),$$

for all $f \in C[0, \|a\|]$. So

$$\omega(a^2) = \int x^2\, d\mu(x) \le \left(\int x\, d\mu(x)\right)^{\frac{2}{3}} \left(\int x^4\, d\mu(x)\right)^{\frac{1}{3}} = \omega(a)^{\frac{2}{3}}\omega(a^4)^{\frac{1}{3}}.$$ □

Lemma 2.5. *Let $2 < p < \infty$. Let J be a set. Let $\mathcal{A}$ be a unital C^*-algebra. Let $\phi : \mathcal{A} \to B(l^p(J))$ be a unital homomorphism. Let $x_0 \in l^p(J)$. Let y_0^* be a bounded linear functional on $l^p(J)$. Define $\omega : \mathcal{A} \to \mathbb{C}$ by*

$$\omega(a) = y_0^*(\phi(a)x_0),$$

for $a \in \mathcal{A}$. Assume that ω is a positive linear functional. Let $(b_k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{A}$ such that $\|b_k\| \le 1$ for all $k \in \mathbb{N}$ and $\phi(b_k)x_0 \to 0$ weakly as $k \to \infty$. Then $\omega(b_k^* b_k) \to 0$ as $k \to \infty$.

Proof. By contradiction, suppose that $\omega(b_k^*b_k)$ does not converge to 0. Passing to a subsequence, we have that there exists $\gamma > 0$ such that $\omega(b_k^*b_k) \geq \gamma$ for all $k \in \mathbb{N}$.

Since $\|\phi(b_k)x_0\| \leq \|\phi\|\,\|x_0\|$ and $\phi(b_k)x_0 \to 0$ weakly, passing to a further subsequence, we may assume that there are $z_1, z_2, \ldots$ in $l^p(J)$ with disjoint supports such that $\|z_k\| \leq \|\phi\|\|x_0\|$ and $\|\phi(b_k)x_0 - z_k\| \leq 1/2^k$ for all $k \in \mathbb{N}$.

Let $n \in \mathbb{N}$. For each $\delta = (\delta_1, \ldots, \delta_n) \in \{-1, 1\}^n$, let

$$a_\delta = \left|\sum_{k=1}^n \delta_k b_k\right| \in \mathcal{A}.$$

By Lemma 2.4,

$$\omega(a_\delta^2) \leq \omega(a_\delta)^{\frac{2}{3}}\omega(a_\delta^4)^{\frac{1}{3}}.$$

Thus,

$$\mathbb{E}\omega(a_\delta^2) \leq [\mathbb{E}\omega(a_\delta)]^{\frac{2}{3}}[\mathbb{E}\omega(a_\delta^4)]^{\frac{1}{3}},$$

where $\mathbb{E}$ denotes expectation over $\delta = (\delta_1, \ldots, \delta_n)$ uniformly distributed on $\{-1, 1\}^n$.

Note that

$$\begin{aligned}\mathbb{E}\omega(a_\delta^2) &= \mathbb{E}\omega\left(\left(\sum_{k=1}^n \delta_k b_k\right)^*\left(\sum_{k=1}^n \delta_k b_k\right)\right)\\ &= \mathbb{E}\omega\left(\sum_{1\leq j,k\leq n} \delta_j\delta_k b_j^* b_k\right) = \sum_{1\leq j,k\leq n}\mathbb{E}(\delta_j\delta_k)\omega(b_j^*b_k) = \sum_{k=1}^n \omega(b_k^*b_k) \geq n\gamma.\end{aligned}$$

Therefore,

$$n\gamma \leq [\mathbb{E}\omega(a_\delta)]^{\frac{2}{3}}[\mathbb{E}\omega(a_\delta^4)]^{\frac{1}{3}}. \tag{2-1}$$

We have

$$a_\delta^4 = \left[\left(\sum_{k=1}^n \delta_k b_k\right)^*\left(\sum_{k=1}^n \delta_k b_k\right)\right]^2 = \sum_{1\leq k_1,\ldots,k_4\leq n} \delta_{k_1}\delta_{k_2}\delta_{k_3}\delta_{k_4} b_{k_1}^* b_{k_2} b_{k_3}^* b_{k_4}.$$

Since $\|b_k\| \leq 1$, it follows that

$$\mathbb{E}\omega(a_\delta^4) = \sum_{1\leq k_1,\ldots,k_4\leq n}\mathbb{E}(\delta_{k_1}\delta_{k_2}\delta_{k_3}\delta_{k_4})\omega(b_{k_1}^*b_{k_2}b_{k_3}^*b_{k_4}) \leq \sum_{1\leq k_1,\ldots,k_4\leq n}\mathbb{E}(\delta_{k_1}\delta_{k_2}\delta_{k_3}\delta_{k_4}).$$

Note that $\mathbb{E}(\delta_{k_1}\delta_{k_2}\delta_{k_3}\delta_{k_4}) = 0$ unless the following occurs:

$$(k_1 = k_2 \text{ and } k_3 = k_4) \quad \text{or} \quad (k_1 = k_3 \text{ and } k_2 = k_4) \quad \text{or} \quad (k_1 = k_4 \text{ and } k_2 = k_3).$$

Thus, $\mathbb{E}\omega(a_\delta^4) \leq 3n^2$. So by (2-1), we have $n\gamma \leq 3^{\frac{1}{3}}n^{\frac{2}{3}}[\mathbb{E}\omega(a_\delta)]^{\frac{2}{3}}$. Hence,

$$\mathbb{E}\omega(a_\delta) \geq \frac{\gamma^{\frac{3}{2}}}{3^{\frac{1}{2}}}n^{\frac{1}{2}}. \tag{2-2}$$

Fix $\delta \in \{-1, 1\}^n$. By Lemma 2.3,

$$\omega(a_\delta) = \omega\left(\left|\sum_{k=1}^n \delta_k b_k\right|\right) \leq \sup_{c\in\mathcal{A},\,\|c\|\leq 1}\left|\omega\left(c\sum_{k=1}^n \delta_k b_k\right)\right|.$$

For $c \in \mathcal{A}$ with $\|c\| \leq 1$,

$$\begin{aligned}\left|\omega\left(c\sum_{k=1}^{n}\delta_k b_k\right)\right| &= \left|y_0^*\left(\phi(c)\left(\sum_{k=1}^{n}\delta_k\phi(b_k)x_0\right)\right)\right| \\ &\leq \|y_0^*\|\,\|\phi\|\left\|\sum_{k=1}^{n}\delta_k\phi(b_k)x_0\right\| \\ &\leq \|y_0^*\|\,\|\phi\|\left(\left\|\sum_{k=1}^{n}\delta_k z_k\right\| + \sum_{k=1}^{n}\frac{1}{2^k}\right) \leq \|y_0^*\|\,\|\phi\|(\|\phi\|\,\|x_0\|n^{\frac{1}{p}}+1),\end{aligned}$$

where the last two inequalities follow from the fact that $z_1, z_2, \ldots$ have disjoint supports, $\|z_k\| \leq \|\phi\|\,\|x_0\|$ and $\|\phi(b_k)x_0 - z_k\| \leq 1/2^k$. Thus,

$$\omega(a_\delta) \leq \|y_0^*\|\,\|\phi\|(\|\phi\|\,\|x_0\|n^{\frac{1}{p}}+1) \quad \text{for all } \delta \in \{-1, 1\}^n.$$

So by (2-2),

$$\frac{\gamma^{\frac{3}{2}}}{3^{\frac{1}{2}}}n^{\frac{1}{2}} \leq \|y_0^*\|\,\|\phi\|(\|\phi\|\,\|x_0\|n^{\frac{1}{p}}+1).$$

Since n can be chosen to be arbitrarily large and $p > 2$, an absurdity follows. □

For $1 < p < 2$, we have the following result, where the order of b_k^* and b_k is switched, by using the dual l^p space in Lemma 2.5.

Lemma 2.6. *Let $1 < p < 2$. Let J be a set. Let $\mathcal{A}$ be a unital C^*-algebra. Let $\phi : \mathcal{A} \to B(l^p(J))$ be a unital homomorphism. Let $x_0 \in l^p(J)$. Let y_0^* be a bounded linear functional on l^p. Define $\omega : \mathcal{A} \to \mathbb{C}$ by*

$$\omega(a) = y_0^*(\phi(a)x_0),$$

for $a \in \mathcal{A}$. Let $(b_k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{A}$ such that $\|b_k\| \leq 1$ for all $k \in \mathbb{N}$ and such that the sequence $y_0^* \circ \phi(b_k)$ of bounded linear functionals on $l^p(J)$ converges to 0 weakly as $k \to \infty$. Assume that ω is a positive linear functional. Then $\omega(b_k b_k^*) \to 0$ as $k \to \infty$.

Proof. Let $\mathcal{A}_1$ be the unital C^*-algebra consisting of the same elements as $\mathcal{A}$ but with reverse order multiplication

$$a \cdot b = ba.$$

Define a unital homomorphism $\phi_1 : \mathcal{A}_1 \to B((l^p(J))^*)$ by

$$\phi_1(a)y^* = y^* \circ \phi(a),$$

for all $a \in \mathcal{A}_1,\ y^* \in (l^p(J))^*$. Define $\omega_1 : \mathcal{A}_1 \to \mathbb{C}$ by

$$\omega_1(a) = \omega(a) = x_0^{**}(\phi(a)y_0^*),$$

for all $a \in \mathcal{A}_1$, where x_0^{**} is the image of x_0 in the bidual $(l^p)^{**}$. By Lemma 2.5, the result follows. □

The following two lemmas are needed for the proof of Lemma 2.9.

Lemma 2.7 [Clarkson 1936]. *Let $1 < p < \infty$. Let J be a set. For every $\epsilon > 0$, there exists $\gamma > 0$ such that, for all $x, y \in l^p(J)$ satisfying $\|x\|, \|y\| \leq 1$ and $\|x+y\| > 2-\gamma$, we have $\|x-y\| < \epsilon$.*

Lemma 2.8 [Russo and Dye 1966]. *Let $\mathcal{A}$ be a unital C^*-algebra. Then the closed unital ball of $\mathcal{A}$ is the closed convex hull of the set of all unitary elements of $\mathcal{A}$.*

Lemma 2.9. *Let $1 < p < \infty$. Let J be a set. Let $\mathcal{A}$ be a unital C^*-algebra. Let $\phi : \mathcal{A} \to B(l^p(J))$ be a unital homomorphism. Let $x_0 \in l^p(J)$. Then there exists $y_0^* \in (l^p(J))^*$ such that $\omega : \mathcal{A} \to \mathbb{C}$,*

$$\omega(a) = y_0^*(\phi(a)x_0), \quad a \in \mathcal{A},$$

defines a positive linear functional and, for every $\epsilon > 0$, there exists $\gamma > 0$ such that whenever $a \in \mathcal{A}$ satisfies $\|a\| \leq 1$ and $\omega(a^*a) < \gamma$, we have $\|\phi(a)x_0\| < \epsilon$.

Proof. Let $\mathcal{U}(\mathcal{A})$ be the set of all unitary elements of $\mathcal{A}$. Let $(v_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{U}(\mathcal{A})$ such that

$$\lim_{n\to\infty} \|\phi(v_n)x_0\| = \sup_{u\in\mathcal{U}(\mathcal{A})} \|\phi(u)x_0\|.$$

For each $n \in \mathbb{N}$, let x_n^* be a bounded linear functional on $l^p(J)$ such that $\|x_n^*\| = 1$ and $x_n^*(\phi(v_n)x_0) = \|\phi(v_n)x_0\|$. Then $x_n^* \circ \phi(v_n)$ is a bounded sequence in $(l^p(J))^*$. Passing to a subsequence, we may assume that $x_n^* \circ \phi(v_n)$ converges weakly to a bounded linear functional $y_0^* \in (l^p(J))^*$ as $n \to \infty$. Thus, $\omega : \mathcal{A} \to \mathbb{C}$,

$$\omega(a) = y_0^*(\phi(a)x_0) = \lim_{n\to\infty} x_n^*(\phi(v_n a)x_0),$$

for $a \in \mathcal{A}$, defines a bounded linear functional on $\mathcal{A}$. Note that

$$\omega(1) = \lim_{n\to\infty} x_n^*(\phi(v_n)x_0) = \lim_{n\to\infty} \|\phi(v_n)x_0\| = \sup_{u\in\mathcal{U}(\mathcal{A})} \|\phi(u)x_0\|,$$

and, for every $u_0 \in \mathcal{U}(\mathcal{A})$,

$$|\omega(u_0)| = \lim_{n\to\infty} |x_n^*(\phi(v_n u_0)x_0)| \leq \sup_{u\in\mathcal{U}(\mathcal{A})} \|\phi(u)x_0\|.$$

So by Lemma 2.8, we have $\|\omega\| \leq \sup_{u\in\mathcal{U}(\mathcal{A})} \|\phi(u)x_0\|$. Thus, $\omega(1) = \|\omega\|$ and hence ω is a positive linear functional.

By contradiction, suppose that there are $\epsilon > 0$ and a sequence $(a_k)_{k\in\mathbb{N}}$ in $\mathcal{A}$ such that $\|a_k\| \leq 1$ and $\|\phi(a_k)x_0\| \geq \epsilon$ for all $k \in \mathbb{N}$ and $\omega(a_k^*a_k) \to 0$ as $k \to \infty$. We have

$$\|a_k\| \geq \frac{\|\phi(a_k)x_0\|}{\|\phi\|\,\|x_0\|} \geq \frac{\epsilon}{\|\phi\|\,\|x_0\|},$$

for all $k \in \mathbb{N}$. For $k \in \mathbb{N}$, let $b_k = a_k/\|a_k\|$. We have $\|b_k\| = 1$ and $\|\phi(b_k)x_0\| \geq \epsilon$ for all $k \in \mathbb{N}$ and $\omega(b_k^*b_k) \to 0$ as $k \to \infty$.

Since $\|x_n^*\| = 1$,

$$\begin{aligned}\liminf_{n\to\infty} \|\phi(v_n)\phi(1-|b_k|)x_0 + \phi(v_n)x_0\| &\geq \liminf_{n\to\infty} [x_n^*(\phi(v_n)\phi(1-|b_k|)x_0) + x_n^*(\phi(v_n)x_0)] \\ &= \omega(1-|b_k|) + \omega(1) = 2\omega(1) - \omega(|b_k|).\end{aligned}$$

Thus,

$$\liminf_{n\to\infty} \|\phi(v_n)\phi(1-|b_k|)x_0 + \phi(v_n)x_0\| \geq 2\omega(1) - \omega(|b_k|).$$

But

$$\|\phi(v_n)\phi(1-|b_k|)x_0\| \leq \sup_{b\in\mathcal{A},\, \|b\|\leq 1} \|\phi(b)x_0\|\,\|1-|b_k|\| \leq \sup_{u\in\mathcal{U}(\mathcal{A})} \|\phi(u)x_0\| = \omega(1)$$

and $\|\phi(v_n)x_0\| \leq \omega(1)$ for all $n \in \mathbb{N}$. Take

$$x = \frac{1}{\omega(1)}\phi(v_n)\phi(1-|b_k|)x_0 \quad \text{and} \quad y = \frac{1}{\omega(1)}\phi(v_n)x_0$$

in Lemma 2.7 and note that $\omega(|b_k|) \leq \omega(b_k^* b_k)^{\frac{1}{2}}\omega(1)^{\frac{1}{2}} \to 0$ as $k \to \infty$. We have

$$\lim_{k\to\infty} \limsup_{n\to\infty} \|\phi(v_n)\phi(1-|b_k|)x_0 - \phi(v_n)x_0\| = 0.$$

Thus,

$$\lim_{k\to\infty} \limsup_{n\to\infty} \|\phi(v_n)\phi(|b_k|)x_0\| = 0.$$

So $\|\phi(|b_k|)x_0\| \to 0$ as $k \to \infty$. Since

$$b_k = b_k\Big(|b_k| + \frac{1}{k}\Big)^{-1}\Big(|b_k| + \frac{1}{k}\Big) \quad \text{and} \quad \Big\| b_k\Big(|b_k| + \frac{1}{k}\Big)^{-1}\Big\| \leq 1,$$

it follows that $\|\phi(b_k)x_0\| \to 0$ as $k \to \infty$ which contradicts $\|\phi(b_k)x_0\| \geq \epsilon$. □

Proof of Theorem 2.1(i). Without loss generality, we may assume that $\mathcal{A}$ is unital by extending ϕ to a homomorphism from the unitization of $\mathcal{A}$ into $B(l^p(J))$. We may also assume that ϕ is unital since $\phi(1)$ is an idempotent on $l^p(J)$ and the range of every idempotent on $l^p(J)$ is isomorphic to $l^p(J_0)$ for some set J_0 [Pełczyński 1960; Johnson 2012].

Let $x_0 \in l^p$. Let $(a_k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{A}$ such that $\|a_k\| \leq \frac{1}{2}$ for all $k \in \mathbb{N}$. We need to show that $(\phi(a_k)x_0)_{k\in\mathbb{N}}$ has a norm-convergent subsequence.

Case 1: $p > 2$. Passing to a subsequence, we may assume that $(\phi(a_k)x_0)_{k\in\mathbb{N}}$ converges weakly to an element of $l^p(J)$. Thus, $\phi(a_{k_1} - a_{k_2})x_0 \to 0$ weakly as $k_1, k_2 \to \infty$.

By Lemma 2.5, we have

$$\lim_{k_1,k_2\to\infty} \omega((a_{k_1} - a_{k_2})^*(a_{k_1} - a_{k_2})) = 0$$

for every positive linear functional $\omega : \mathcal{A} \to \mathbb{C}$ of the form $\omega(a) = y_0^*(\phi(a)x_0)$ for $a \in \mathcal{A}$. By Lemma 2.9, we have $\lim_{k_1,k_2\to\infty} \|\phi(a_{k_1} - a_{k_2})x_0\| = 0$. So $(\phi(a_k)x_0)_{k\in\mathbb{N}}$ is norm-convergent.

Case 2: $p < 2$. Passing to a subsequence, we may assume that $(y_0^* \circ \phi(a_k^*))_{k\in\mathbb{N}}$ converges weakly to an element of $(l^p(J))^*$. Thus, $y^* \circ \phi(a_{k_1}^* - a_{k_2}^*) \to 0$ weakly as $k_1, k_2 \to \infty$ for every $y^* \in (l^p(J))^*$.

By Lemma 2.6, we have

$$\lim_{k\to\infty} \omega((a_{k_1}^* - a_{k_2}^*)(a_{k_1}^* - a_{k_2}^*)^*) = 0$$

for every positive linear functional $\omega : \mathcal{A} \to \mathbb{C}$ of the form $\omega(a) = y_0^*(\phi(a)x_0)$ for $a \in \mathcal{A}$. By Lemma 2.9, we have $\lim_{k_1,k_2\to\infty} \|\phi(a_{k_1} - a_{k_2})x_0\| = 0$. So $(\phi(a_k)x_0)_{k\in\mathbb{N}}$ is norm-convergent. □

Lemma 2.10 [Kerr and Li 2016, Theorem 2.24]. *Let G be a group. Let $\mathcal{H}$ be a Hilbert space. Let $\psi : G \to B(\mathcal{H})$ be a unital homomorphism such that $\psi(g)$ is unitary for all $g \in G$. If $\{\psi(g)x : g \in G\}$ is norm precompact in $\mathcal{H}$ for all $x \in \mathcal{H}$, then $\mathcal{H}$ is the direct sum of some finite-dimensional subspaces $\mathcal{H}_\alpha$, for $\alpha \in \Lambda$, such that $\mathcal{H}_\alpha$ is invariant under $\psi(g)$ for all $\alpha \in \Lambda$ and $g \in G$.*

Proof of Theorem 2.1(ii). As in the proof Theorem 2.1(i), we may assume that $\mathcal{A}$ is unital and ϕ is unital. We may also assume that $\ker\phi = \{0\}$. Let $a_0 \neq 0$. There exists $x_0 \in l^p(J)$ such that $\phi(a_0)x_0 \neq 0$. By Lemma 2.9, there exists $y_0^* \in (l^p(J))^*$ such that $\omega : \mathcal{A} \to \mathbb{C}$,

$$\omega(a) = y_0^*(\phi(a)x_0),$$

for $a \in \mathcal{A}$, defines a positive linear functional and $\omega(a_0^*a_0) \neq 0$.

Equip $\mathcal{A}$ with the positive semidefinite sesquilinear form

$$\langle a, b\rangle = \omega(b^*a),$$

for $a, b \in \mathcal{A}$. Consider the ideal $\mathcal{A}_0 = \{a \in \mathcal{A} : \langle a, a\rangle = 0\}$ of $\mathcal{A}$. Let $\mathcal{H}$ be the completion of the quotient space $\mathcal{A}/\mathcal{A}_0$. Then $\mathcal{H}$ is a Hilbert space. For each $a \in \mathcal{A}$, we can define a bounded linear operator on $\mathcal{H}$ by sending $b + \mathcal{A}_0$ to $ab + \mathcal{A}_0$ for $b \in \mathcal{A}$. So $\eta : \mathcal{A} \to B(\mathcal{H})$,

$$\eta(a)(b + \mathcal{A}_0) = ab + \mathcal{A}_0,$$

for $a, b \in \mathcal{A}$, defines a unital $*$-homomorphism. We have

$$\begin{aligned}\|\eta(a_1)(b + \mathcal{A}_0) - \eta(a_2)(b + \mathcal{A}_0)\| &= \omega(b^*(a_1 - a_2)^*(a_1 - a_2)b)\\ &= y_0^*(\phi(b^*(a_1 - a_2)^*(a_1 - a_2)b)x_0)\\ &\leq \|y_0^*\|\,\|\phi\|\,\|b^*\|\,\|a_1 - a_2\|\,\|\phi(a_1 - a_2)\phi(b)x_0\|,\end{aligned}$$

for all $a_1, a_2, b \in \mathcal{A}$. By Theorem 2.1(i), we have that $\{\phi(a)x_0 : a \in \mathcal{A},\ \|a\| \leq 1\}$ is norm precompact so $\{\eta(a)(b + \mathcal{A}_0) : a \in \mathcal{A},\ \|a\| \leq 1\}$ is norm precompact for all $b \in \mathcal{A}$. Let $\mathcal{U}(\mathcal{A})$ be the set of all unitary elements of $\mathcal{A}$. By Lemma 2.10, we have that $\mathcal{H}$ is the direct sum of some finite-dimensional subspaces $\mathcal{H}_\alpha$, for $\alpha \in \Lambda$, such that $\mathcal{H}_\alpha$ is invariant under $\eta(u)$ for all $\alpha \in \Lambda$ and $u \in \mathcal{U}(\mathcal{A})$. Note that $\mathcal{H}_\alpha$ is thus invariant under $\eta(a)$ for all $a \in \mathcal{A}$.

Since $\omega(a_0^*a_0) \neq 0$, we have $\eta(a_0) \neq 0$. So $\eta(a_0) \neq 0$ on $\mathcal{H}_{\alpha_0}$ for some $\alpha_0 \in \Lambda$. Thus, $\mathcal{A}$ is residually finite-dimensional. □

Proof of Corollary 2.2. One direction follows from Theorem 2.1. For the other direction, suppose that $\mathcal{A}$ is a residually finite-dimensional C^*-algebra. Then there is a collection $(\phi_\alpha)_{\alpha\in\Lambda}$ of $*$-representations of $\mathcal{A}$ on finite-dimensional Hilbert spaces $\mathcal{H}_\alpha$ such that $\|a\| = \sup_{\alpha\in\Lambda} \|\phi_\alpha(a)\|$ for all $a \in \mathcal{A}$. Define $\phi : \mathcal{A} \to B\big((\bigoplus_{\alpha\in\Lambda} \mathcal{H}_\alpha)_{l^p}\big)$ by $\phi = \bigoplus_{\alpha\in\Lambda} \phi_\alpha$. Thus ϕ is a norm-preserving homomorphism. However, it is a classical result of [Pełczyński 1960] that for $1 < p < \infty$, the l^p direct sum of finite-dimensional Hilbert spaces is isomorphic to $l^p(J)$ for some set J. Therefore, $\mathcal{A}$ is isomorphic to a subalgebra of $B(l^p(J))$, via the map $a \mapsto S\phi(a)S^{-1}$, where $S : \big(\bigoplus_{\alpha\in\Lambda} \mathcal{H}_\alpha\big)_{l^p} \to l^p(J)$ is any invertible operator. □

Acknowledgements

The author is grateful to the referee for some suggestions that improved the exposition. The author is supported by NSF DMS-1856221.

References

[Blecher and Phillips 2019] D. P. Blecher and N. C. Phillips, "L^p-operator algebras with approximate identities, I", *Pacific J. Math.* **303**:2 (2019), 401–457. MR Zbl

[Clarkson 1936] J. A. Clarkson, "Uniformly convex spaces", *Trans. Amer. Math. Soc.* **40**:3 (1936), 396–414. MR Zbl

[Gardella and Thiel 2020] E. Gardella and H. Thiel, "Extending representations of Banach algebras to their biduals", *Math. Z.* **294**:3-4 (2020), 1341–1354. MR Zbl

[Johnson 2012] B. Johnson, "Complemented subspaces of $\ell_p(I)$ for uncountable I", answer on MathOverflow, 2012, available at https://mathoverflow.net/questions/111882/complemented-subspaces-of-ell-pi-for-uncountable-i.

[Kerr and Li 2016] D. Kerr and H. Li, *Ergodic theory: independence and dichotomies*, Springer, 2016. MR Zbl

[Lamperti 1958] J. Lamperti, "On the isometries of certain function-spaces", *Pacific J. Math.* **8** (1958), 459–466. MR Zbl

[Lindenstrauss and Tzafriri 1977] J. Lindenstrauss and L. Tzafriri, *Classical Banach spaces*, vol. I: Sequence spaces, Ergebnisse der Mathematik und ihrer Grenzgebiete **92**, Springer, 1977. MR Zbl

[Pełczyński 1960] A. Pełczyński, "Projections in certain Banach spaces", *Studia Math.* **19** (1960), 209–228. MR Zbl

[Russo and Dye 1966] B. Russo and H. A. Dye, "A note on unitary operators in C^*-algebras", *Duke Math. J.* **33** (1966), 413–416. MR Zbl

Received 29 Dec 2018. Revised 19 Jun 2019. Accepted 6 Sep 2019.

MARCH T. BOEDIHARDJO: march@math.ucla.edu
Department of Mathematics, University of California, Los Angeles, CA, United States

ANALYSIS AND PDE
Vol. 13, No. 7, 2020
https://doi.org/10.2140/apde.2020.13.2183

msp

EXPONENTIAL CONVERGENCE OF PARABOLIC OPTIMAL TRANSPORT ON BOUNDED DOMAINS

FARHAN ABEDIN AND JUN KITAGAWA

We study the asymptotic behavior of solutions to the second boundary value problem for a parabolic PDE of Monge–Ampère type arising from optimal mass transport. Our main result is an exponential rate of convergence for solutions of this evolution equation to the stationary solution of the optimal transport problem. We derive a differential Harnack inequality for a special class of functions that solve the linearized problem. Using this Harnack inequality and certain techniques specific to mass transport, we control the oscillation in time of solutions to the parabolic equation, and obtain exponential convergence. Additionally, in the course of the proof, we present a connection with the pseudo-Riemannian framework introduced by Kim and McCann in the context of optimal transport, which is interesting in its own right.

1. Introduction

Given two smooth domains Ω, $\Omega^* \subset \mathbb{R}^n$, two probability measures μ, η defined respectively on Ω and Ω^*, and a Borel measurable *cost function* $c : \bar{\Omega} \times \bar{\Omega}^* \to \mathbb{R}$, the optimal transport problem is to find a μ-measurable map $T : \Omega \to \Omega^*$ satisfying $T_\# \mu = \eta$ (where $T_\# \mu(E) := \mu(T^{-1}(E))$ for all measurable $E \subset \Omega^*$) such that

$$\int_\Omega c(x, T(x))\, d\mu(x) = \max_{S_\# \mu = \eta} \int_\Omega c(x, S(x))\, d\mu(x). \tag{1}$$

Under mild assumptions on the cost function and the measures, it can be shown that the solution T to (1) exists; see, for example, [Brenier 1991; Gangbo and McCann 1996]. If the measures μ and η are absolutely continuous with respect to Lebesgue measure, and c satisfies the bitwist condition (6) below, the map T is μ-a.e. single-valued and can be determined by the implicit relation

$$\nabla_x c(x, T(x)) = \nabla u(x),$$

where the scalar-valued potential u is a *c-convex function* (see Definition 2.1) satisfying the Monge–Ampère-type equation

$$\begin{cases} \det[D^2 u(x) - A(x, \nabla u(x))] = B(x, \nabla u(x)), & x \in \Omega, \\ T(\Omega) = \Omega^*, \end{cases} \tag{2}$$

where A is a matrix-valued function and B is scalar-valued, defined in terms of the cost function c and the densities of the measures μ, η. The issue of existence and regularity of solutions to the PDE (2) has

Kitagawa's research was supported in part by National Science Foundation grant DMS-1700094.
MSC2010: 35K96, 58J35.
Keywords: parabolic optimal transport, Monge–Kantorovich, exponential convergence, Kim–McCann metric, Li–Yau Harnack inequality.

been an active area of research for many years. For higher-order regularity results, we refer the reader to [Ma et al. 2005; Trudinger and Wang 2009; Urbas 1997].

One possible approach to finding a solution to the PDE above is to solve the parabolic PDE

$$\begin{cases} \partial_t u(x,t) = \log\det[D^2u(x,t) - A(x,\nabla u(x,t))] - \log B(x,\nabla u(x,t)), & x \in \Omega,\ t > 0, \\ G(x,\nabla u(x,t)) = 0, & x \in \partial\Omega, \\ u(x,0) = u_0(x), & x \in \Omega, \end{cases} \tag{3}$$

for appropriate initial and boundary conditions u_0 and G (see Section 2), and view a stationary solution as $t \to \infty$ as a solution to (2). The study of existence, regularity, and asymptotic behavior of solutions to the parabolic problem (3) was initiated only recently through the works [Kitagawa 2012; Kim et al. 2012].

The main result of this paper is the following theorem on an exponential convergence rate of solutions to the parabolic equation (3). The notation $C_x^{k_1}C_t^{k_2}$ will denote functions on a space-time domain which are C^{k_1} in the space variable and C^{k_2} in the time variable, with corresponding norms finite. Our main result is as follows:

Theorem 1.1. *Suppose $u \in C_x^4C_t^3(\overline{\Omega}\times[0,\infty))$ is a solution on $\overline{\Omega}\times[0,\infty)$ to the parabolic equation (3) converging uniformly on $\overline{\Omega}$ to a stationary solution u^∞ as $t \to \infty$, and $\mathcal{K}$ is a constant such that*

$$\|u\|_{C_x^4C_t^2(\overline{\Omega}\times[0,\infty))} + \|c\|_{C^4(\overline{\Omega}\times\overline{\Omega}^*)} \le \mathcal{K}. \tag{4}$$

If the cost function c satisfies the bitwist condition (6), and Ω and Ω^* satisfy the c-convexity conditions (8) and (9), then

$$\|u(\cdot,t) - u^\infty\|_{L^\infty(\Omega)} \le C_1e^{-C_2t} \quad \text{for all } t \ge 0,$$

for some constants $C_1, C_2 > 0$ depending only on $\mathcal{K}$ and the dimension n.

Previous work in [Kitagawa 2012] establishes the existence of a function $u \in C_x^2C_t^1(\overline{\Omega}\times[0,\infty))$ that solves (3) for all times $t \ge 0$ and converges in $C^2(\overline{\Omega})$ to a function $u^\infty(\cdot)$ as $t \to \infty$, where $u^\infty(\cdot)$ satisfies the elliptic optimal transport equation (2). Using this result and a bootstrapping argument, we obtain the following corollary.

Corollary 1.2. *Suppose the cost function c satisfies the bitwist condition (6) and the Ma–Trudinger–Wang condition (10), and suppose Ω and Ω^* satisfy the c-convexity conditions (8) and (9) with $\delta, \delta^* > 0$. Suppose the source and target measures μ and η are absolutely continuous with smooth densities that are bounded away from zero and infinity on $\overline{\Omega}$ and $\overline{\Omega}^*$ respectively. Finally, suppose the initial condition $u_0 \in C^{4,\alpha}(\Omega)$ for some $\alpha \in (0,1]$ is locally, uniformly c-convex (as in Definition 2.1) and satisfies the boundary compatibility conditions (12). Then u satisfies the hypotheses of Theorem 1.1 above.*

Proof. Under the Ma–Trudinger–Wang condition (10) and the uniform c- and c^*-convexity of the domains (i.e., (8) and (9) with $\delta, \delta^* > 0$), global $C_x^{2,\alpha}C_t^{1,\alpha}$ estimates of the solution $u(x,t)$ to (3) were obtained in [Kitagawa 2012, Theorems 10.1 and 11.2, and Section 12]. Thus, by applying boundary Schauder estimates for linear uniformly parabolic equations in nondivergence form with uniformly oblique boundary conditions (see [Lieberman 1996, Theorems 4.23 and 4.31]) to the linearized equation (18), we obtain the desired higher regularity of u. □

Remark 1.3. One particular motivation for this exponential convergence result comes from numerics for optimal transport. Since the stationary state of (3) gives rise to the solution of the optimal transport problem between the measures μ and η, one could attempt to implement an algorithm that is initiated with some c-convex potential function and flows toward the desired solution via (3). Establishing quantitative rates of convergence for such an algorithm is consequently of paramount importance. One difficulty that should be noted here is that in the case with nonempty boundary, due to compatibility requirements with the boundary condition, there are some restrictions on what can be taken as an initial condition (compare to the case of no boundary, where one can simply take a constant function), and it is not always clear how to generate initial data that will still provide global existence. We plan to explore this issue of finding appropriate initial conditions in future work.

1A. ***Prior results and the contributions of this paper.*** The parabolic flow (3) on Riemannian manifolds with no boundary was considered by Kim, Streets, and Warren [Kim et al. 2012], under a strong form of the Ma–Trudinger–Wang condition (10); their methods strongly use that the boundary is empty. There, the authors prove exponential convergence of the solution u of (3) to the solution u^{∞} of the elliptic equation (2); see [Kim et al. 2012, Theorem 1.1]. Their proof relies on establishing a Li–Yau-type Harnack inequality for solutions to the linearization of (3), coupled with the observation that this linearization is actually a heat equation where the elliptic part is a conformal factor times the Laplace–Beltrami operator of a conformal change of a metric defined from the solution of the parabolic evolution itself; see [Kim et al. 2012, Proposition 5.1] and the discussion preceding Proposition 2.7 below.

However, presence of a boundary turns out to be a major obstruction to applying the methods of [Kim et al. 2012]. First, their method of introducing a conformal change of metric cannot be used in two dimensions: when there is no boundary, it is possible to convert the two-dimensional problem to a three-dimensional one, but such a technique simply does not work when the boundary is nonempty and is required to satisfy certain convexity properties. Second, the linearization of (3) is a Neumann boundary-value problem with respect to a time-varying Riemannian metric, for which there is no general known Harnack inequality. Existing results require that the metric itself satisfy some specific evolution, such as Ricci flow [Bailesteanu et al. 2010] or Gauss curvature flow [Chow 1991]. Thus while there is a sizable body of work on differential Harnack inequalities, none of them are directly applicable to the linearization of (3). We also mention the result [Schnürer and Smoczyk 2003], which treats a nonlinear evolution equation arising from Gauss curvature flow that resembles (3) in the case where the cost function is $c(x, y) = \langle x, y\rangle$, with nonempty boundary. The authors of [Schnürer and Smoczyk 2003] also obtain an exponential convergence result, but assume certain structural assumptions on the function B in (3) that are not satisfied in the optimal transport case, and impose additional constraints on the initial data u_0.

The contributions of this paper are as follows. First we show it is possible to obtain a Harnack inequality for a certain subclass of solutions to the linearized equation. In the interior, this can be shown by a series of estimates similar to that of [Kim et al. 2012] with no boundary, but as mentioned above, a different method must be employed to settle the two-dimensional case. In dealing with the boundary, we must carefully exploit the curvature conditions imposed on the boundaries of both the source and target domains in order to choose the correct class of solutions for which we can obtain the Harnack inequality. Once we have a Harnack inequality for such special solutions, we use the fact that solutions of the parabolic flow come from the optimal transport problem, and hence satisfy a mass-preservation condition (see Lemma 2.3 below), to finish the proof of exponential convergence. We heavily stress here that our approach diverges from the traditional proof of exponential convergence via the Harnack inequality, and crucially uses the fact that there is an underlying optimal transport problem. Additionally, we show this analysis of the boundary behavior can also be done by exploiting the pseudo-Riemannian structure introduced in [Kim and McCann 2010] for optimal transport. More specifically, we prove a relation between the second fundamental form with respect to the time-varying Riemannian metric on the source domain, with the Euclidean second fundamental forms of the source and target domains under c-exponential coordinates, which has not previously been explored.

1B. ***Outline and strategy of proof.*** The outline of the remainder of the paper and the strategy behind our proof are as follows. In Section 2 we give the necessary background for the optimal transport problem. We also recall the method of [Kim et al. 2012] for the proof of exponential convergence on manifolds with no boundary, and prove here the important parabolic estimate Proposition 2.7, although with a slightly different proof from that of Kim, Streets, and Warren. In Section 3 we obtain expressions for and estimates on the boundary condition acting on the relevant auxiliary function. For the benefit of the reader, we divide the proof of these estimates into the inner product case and the general cost function case. In Section 4 we finally obtain the exponential convergence result from the estimates derived in the previous sections; the proof we present relies on the underlying optimal transport structure of the problem. The final Section 5 provides the aforementioned alternative, geometric approach to the boundary estimates from Section 3.

2. Preliminaries

2A. ***Basic notions from optimal transport.*** We denote by D^2, ∇, and D_β the Hessian matrix, the gradient vector, and the directional derivative in the direction β of a given function with respect to the space variable x. Spatial partial derivatives will be denoted by subscript indices, with the actual variable specified when necessary, while D_x and D_p will be used for the derivative matrix of a mapping with respect to the variable in the subscript. We will also follow the convention of summing over repeated indices. Time derivatives will be denoted by ∂_t.

When considering a Riemannian manifold (M, g), we will denote the inner product and norm with respect to the metric g by $\langle \cdot , \cdot \rangle_g$ and $|\cdot|_g$ respectively. The notation ∇^g, Hess_g, Δ_g, and Ric_g will be used for the gradient, Hessian, Laplacian, and Ricci tensor with respect to g.

Regarding the cost function $c(x, y)$, derivatives in the x-variable will be denoted by subscripts preceding a comma, while derivatives in the y-variable will be denoted by subscripts following a comma. The notation $c^{i,j}$ denotes the entries of the inverse of the matrix $c_{i,j}$.

We will assume from here onward that Ω, Ω^* are open, smooth, bounded domains in $\mathbb{R}^n$. The outward-pointing unit normals to $\partial\Omega$ and $\partial\Omega^*$ will be denoted by ν and ν^* respectively. The function h^* will be a normalized defining function for Ω^*; i.e., $h^* = 0$ on $\partial\Omega$, $h^* < 0$ on Ω, and $\nabla h^* = \nu^*$ on $\partial\Omega^*$. The measures μ, η are assumed to be absolutely continuous with respect to n-dimensional Lebesgue measure, with densities ρ, ρ^* respectively satisfying the bounds $0 < \lambda \le \rho, \rho^* \le \Lambda < \infty$ and the mass balance condition

$$\int_\Omega \rho = \int_{\Omega^*} \rho^*. \tag{5}$$

We will also assume $c \in C^{4,\alpha}(\bar{\Omega} \times \bar{\Omega}^*)$ for some $\alpha \in (0, 1]$, and

$$\begin{aligned} &y \mapsto \nabla_x c(x, y) \text{ is a diffeomorphism for all } x \in \bar{\Omega},\\ &x \mapsto \nabla_y c(x, y) \text{ is a diffeomorphism for all } y \in \bar{\Omega}^*. \end{aligned} \tag{6}$$

For any $p \in \nabla_x c(x, \Omega^*)$ and $x \in \Omega$ (resp. $q \in \nabla_y c(\Omega, y)$ and $y \in \Omega^*$), we denote by $Y(x, p)$ (resp. $X(q, y)$) the unique element of Ω^* (resp. Ω) such that

$$(\nabla_x c)(x, Y(x, p)) = p \quad (\text{resp. } (\nabla_y c)(X(q, y), y) = q). \tag{7}$$

We say Ω *is c-convex with respect to* Ω^* if the set $\nabla_y c(\Omega, y)$ is a convex set for each $y \in \Omega^*$. Similarly, Ω^* *is* c^**-convex with respect to* Ω if the set $\nabla_x c(x, \Omega^*)$ is a convex set for each $x \in \Omega$. Analytically, these conditions are satisfied if we have

$$[\nu_i^j(x) - c^{\ell,k} c_{ij,\ell}(x, y)\nu^k(x)]\tau^i\tau^j \ge \delta|\tau|^2 \qquad \text{for all } x \in \partial\Omega,\ y \in \bar{\Omega}^*,\ \tau \in T_x(\partial\Omega), \tag{8}$$

$$[(\nu^*)_i^j(y) - c^{k,\ell} c_{\ell,ij}(x, y)(\nu^*)^k(x)](\tau^*)^i(\tau^*)^j \ge \delta^*|\tau^*|^2 \quad \text{for all } y \in \partial\Omega^*,\ x \in \bar{\Omega},\ \tau^* \in T_y(\partial\Omega^*) \tag{9}$$

for some constants $\delta, \delta^* \ge 0$ respectively, where we will always sum over repeated indices. If δ (resp. δ^*) is strictly positive, we say that Ω *is uniformly c-convex with respect to* Ω^* (resp. Ω^* *is uniformly* c^**-convex with respect to* Ω).

Define the matrix-valued function A by $A(x, p) := (D_x^2 c)(x, Y(x, p))$. Since $Y(x, p)$ satisfies the equation $(\nabla_x c)(x, Y(x, p)) = p$, we can differentiate implicitly in p to get

$$(D_{x,y}^2 c)(x, Y(x, p)) D_p Y(x, p) = \mathbb{I}_n.$$

Similarly, differentiating the equation $(\nabla_x c)(x, Y(x, p)) = p$ in x gives

$$(D_x^2 c)(x, Y(x, p)) + (D_{x,y}^2 c)(x, Y(x, p)) D_x Y(x, p) = 0.$$

We have chosen the convention $(DY)_{\ell m} = Y_m^\ell$ for differentiation either in the x- or p-variables. It follows that

$$A(x, p) = (D_x^2 c)(x, Y(x, p)) = -(D_p Y)^{-1}(x, p) D_x Y(x, p).$$

Definition 2.1. A function $\varphi : \Omega \to \mathbb{R}$ is said to be *c-convex* if for any point $x_0 \in \Omega$, there exists a $y_0 \in \Omega^*$ and $\lambda_0 \in \mathbb{R}$ such that

$$\begin{aligned} \varphi(x_0) &= c(x_0, y_0) + \lambda_0,\\ \varphi(x) &\ge c(x, y_0) + \lambda_0 \quad \text{for all } x \in \Omega. \end{aligned}$$

A function $\varphi \in C^2(\Omega)$ is said to be *locally, uniformly c-convex* if $D^2\varphi(x) - A(x, \nabla\varphi(x)) > 0$ as a matrix for every $x \in \overline{\Omega}$.

Although we will not use it explicitly in this paper, we also mention the, by now well-known, Ma–Trudinger–Wang condition. This condition (or rather a stronger version of it) was first used to obtain interior $C^{2,\alpha}$ regularity of solutions to the elliptic optimal transport equation (2) in [Ma et al. 2005]. It was proven to be a necessary condition for regularity theory in [Loeper 2009], and it was shown that classical solutions for the parabolic equation (3) exist under the same condition in [Kitagawa 2012].

Definition 2.2. The cost function $c(x, y)$ satisfies the Ma–Trudinger–Wang (MTW) condition if

$$D_{p_i p_j} A_{k\ell}(x, p)\xi^i \xi^j \eta^k \eta^\ell \geq 0 \quad \text{for all } x \in \overline{\Omega},\ p \in \nabla_x c(x, \Omega^*),\ \xi \perp \eta. \tag{10}$$

2B. *The parabolic optimal transport problem.* For a function $u \in C^4_x C^2_t(\overline{\Omega} \times [0, \infty))$ (which, in the sequel, will be the solution to the parabolic optimal transportation problem), we will employ the following notation:

$$\begin{aligned}
T(x, t) &= Y(x, \nabla u(x, t)),\\
B(x, p) &= |\det (D^2_{x,y}c)(x, Y(x, p))| \cdot \frac{\rho(x)}{\rho^*(Y(x, p))},\\
G(x, p) &= h^*(Y(x, p)),\\
\beta(x, t) &= \nabla_p G(x, p)|_{p=\nabla u},\\
W(x, t) &= D^2 u(x, t) - A(x, \nabla u(x, t)).
\end{aligned}$$

The components of the matrix $W(x, t)$ will be denoted by w_{ij}, while the components of the inverse matrix will be denoted by w^{ij}.

Using the above notation, we can now precisely state the parabolic optimal transportation problem. We seek to find a function $u \in C^4_x C^2_t(\overline{\Omega} \times [0, \infty))$ satisfying the evolution equation

$$\begin{cases}
\partial_t u(x, t) = \log\det[D^2 u(x, t) - A(x, \nabla u(x, t))] - \log B(x, \nabla u(x, t)), & x \in \Omega,\ t > 0,\\
G(x, \nabla u(x, t)) = 0, & x \in \partial\Omega,\ t > 0,\\
u(x, 0) = u_0(x), & x \in \Omega.
\end{cases} \tag{11}$$

We require the function $u_0 \in C^{4,\alpha}(\Omega)$ for some $\alpha \in (0, 1]$ to be locally, uniformly c-convex as in Definition 2.1 and satisfy

$$\begin{cases}
h^*(Y(x, \nabla u_0(x))) = 0 \quad \text{on } \partial\Omega,\\
T_0(\Omega) = \Omega^*,
\end{cases} \tag{12}$$

where $T_0(x) := Y(x, \nabla u_0(x))$.

Let us establish some basic facts which will be needed throughout.

Lemma 2.3. *The function $\theta(x, t) := \partial_t u(x, t)$ satisfies*

$$\int_\Omega e^{\theta(x,t)} \rho(x)\, dx = \int_{\Omega^*} \rho^*(y)\, dy \quad \textit{for all } t \geq 0. \tag{13}$$

Proof. Differentiating the identity $T(x, t) = Y(x, \nabla u(x, t))$, we obtain

$$T^k_{x_\ell}(x, t) = Y^k_{x_\ell}(x, \nabla u(x, t)) + Y^k_{p_j}(x, \nabla u(x, t)) u_{x_j x_\ell}, \quad k, \ell = 1, \dots, n.$$

In matrix notation,

$$\begin{aligned} D_x T(x,t) &= D_x Y(x, \nabla u(x,t)) + D_p Y(x, \nabla u(x,t)) D^2 u(x,t) \\ &= D_p Y(x, \nabla u(x,t))(D^2 u(x,t) - A(x, \nabla u(x,t)) \\ &= D_p Y(x, \nabla u(x,t)) W(x,t) \\ &= (D^2_{x,y} c)^{-1}(x, Y(x, \nabla u(x,t))) W(x,t). \end{aligned} \tag{14}$$

Consequently,

$$|\det D_x T(x,t)| = \frac{\det W(x,t)}{|\det(D^2_{x,y} c)(x, T(x,t))|}. \tag{15}$$

From (11), it follows that

$$e^{\partial_t u(x,t)} \rho(x) = |\det D_x T(x,t)| \rho^*(x, T(x,t)). \tag{16}$$

Integrating over Ω and using the change of variables formula yields the desired identity. □

Observe that, by (13) and the mass balance condition (5), θ must satisfy

$$\sup_\Omega \theta(\,\cdot\,, t) \geq 0 \quad \text{and} \quad \inf_\Omega \theta(\,\cdot\,, t) \leq 0 \quad \text{for all } t \geq 0. \tag{17}$$

Lemma 2.4. *Let ν denote the outward-pointing unit normal to Ω, and let W and β be defined as above. Then*

$$\nu(x) = \frac{W(x,t)\beta(x,t)}{|W(x,t)\beta(x,t)|} \quad \textit{for all } (x,t) \in \partial\Omega \times [0, \infty).$$

Proof. Fix $t \geq 0$. The boundary condition $G(x, \nabla u(x,t)) = 0$ on $\partial\Omega$ is equivalent to saying $h^*(T(x,t)) = 0$ on $\partial\Omega$. Therefore, by differentiating in any direction τ tangential to $\partial\Omega$, we get

$$h^*_k(T(x,t)) T^k_{x_i}(x,t) \tau^i = 0.$$

In matrix notation,

$$\langle W(x,t)(D_p Y)^T(x, \nabla u(x,t)) \nabla h^*(T(x,t)), \tau \rangle = 0.$$

By definition,

$$\beta(x,t) = (D_p Y)^T(x, \nabla u(x,t)) \nabla h^*(Y(x, \nabla u(x,t))).$$

Therefore,

$$\langle W(x,t)\beta(x,t), \tau \rangle = 0.$$

It follows that $W\beta$ is parallel to the unit outward-pointing normal vector field ν on $\partial\Omega$. Since $h^* < 0$ on Ω, we can write $W\beta = \chi\nu$, where $\chi \geq 0$. Notice that by (15) and (16), W is positive definite. By bitwist (6), and the fact that $\nabla h^* = \nu^*$, we also know β is nonzero. Consequently, $\chi = |W\beta|$ is nonzero. □

2C. *The linearized equation.* Differentiating (11) in t gives the following linear equation for θ:

$$\begin{cases} \mathcal{L}\theta := w^{ij}(\theta_{ij} - D_{p_k} A_{ij} \theta_k) + D_{p_k}(\log B)\theta_k - \partial_t \theta = 0 & \text{on } \mathcal{C}_T := \Omega \times [0, T], \\ D_\beta \theta = 0 & \text{on } \partial\Omega \times [0, T), \end{cases} \tag{18}$$

where $D_\beta\theta := \langle\beta, \nabla\theta\rangle$, and where, in the coefficients, $p = \nabla u(x,t)$. By the global C^2 estimates established in [Kitagawa 2012], the operator $\mathcal{L}$ is uniformly parabolic and, by Theorems 7.1 and 9.2 of that paper, the boundary condition $D_\beta\theta = 0$ is uniformly oblique for all time. Hence, there exist positive constants $c_1, c_2 > 0$ depending only on Ω, Ω^*, B, c and u_0, but independent of t, such that $w^{ij}\xi_i\xi_j \geq c_1|\xi|^2$ for all $(x,t) \in \Omega$ and $\xi \in \mathbb{R}^n$, and $\langle\beta, \nu\rangle \geq c_2 > 0$ for all $x \in \partial\Omega$, $t > 0$.

Solutions to the linearized equation (18) satisfy the following maximum principle; see also [Kitagawa 2012, Theorem 8.1].

Proposition 2.5. *Suppose v is a solution to the linearized equation (18). Then*

$$\max_{(x,t)\in\mathcal{C}_T} v(x,t) = \max_{x\in\Omega} v(x,0), \qquad \min_{(x,t)\in\mathcal{C}_T} v(x,t) = \min_{x\in\Omega} v(x,0).$$

Proof. By the parabolic maximum principle, the maximum of v occurs on the parabolic boundary $\partial_P\mathcal{C}_T := (\overline{\Omega}\times\{0\}) \cup (\partial\Omega\times(0,T))$. Suppose there exists $(x_0,t_0) \in \partial\Omega\times(0,T)$ such that $v(x_0,t_0) = \max_{(x,t)\in\mathcal{C}_T} v(x,t)$. It then follows from Hopf's lemma, see [Lieberman 1996, Lemma 2.8 and following paragraph], that $D_\beta v(x_0,t_0) > 0$. However, this violates the boundary condition $D_\beta v = 0$, and so the maximum cannot occur on $\partial\Omega\times(0,T)$. The argument for the minimum follows in similar fashion. □

2D. ***Exponential convergence on manifolds with no boundary.*** In this section we recall the proof for exponential convergence in the case of no boundary as done in [Kim et al. 2012]. At the end of the section, we reprove the parabolic estimate Proposition 2.7 for the linearized operator, but we note our method differs slightly from that of [Kim et al. 2012].

The authors of [Kim et al. 2012] consider the parabolic flow (11) on a Riemannian manifold with no boundary and show exponential convergence of the solution u of (11) to the solution u^∞ of the elliptic equation (2). A key ingredient in their proof of exponential convergence is a Li–Yau-type Harnack inequality for positive solutions v of the linearized equation $\mathcal{L}v = 0$; see [Kim et al. 2012, Theorem 5.2]. This strategy is motivated by the observation that the operator $\mathcal{L}$ is a heat-type equation with respect to the time-varying Riemannian metric g with components $g_{ij} = w_{ij}$ (see the discussion preceding Proposition 2.7 below).

Suppose v is a positive solution to the linearized equation $\mathcal{L}v = 0$ on $\mathcal{C}_T$, where $T > 0$ is chosen to be sufficiently large. Let $f = \log v$ and consider the quantity

$$F = t(|\nabla^g f|_g^2 - \alpha\,\partial_t f) = t(w^{ij} f_i f_j - \alpha\,\partial_t f), \tag{19}$$

where $\alpha > 0$ is a constant to be determined and ∇^g denotes the gradient of a function with respect to the metric g. It is shown in [Kim et al. 2012, Theorem 5.2] that F is sublinear in t everywhere in $\mathcal{C}_T$; that is, there exist constants $C_1, C_2 > 0$ (independent of T) such that $F(x,t) \leq C_1 + C_2 t$ for all $x \in \Omega$, $t \in [0,T]$. The sublinearity in t of F implies the differential Harnack inequality

$$w^{ij} f_i f_j - \alpha\,\partial_t f \leq \frac{C_1}{t} + C_2 \tag{20}$$

for some possibly different constants C_1 and $C_2 > 0$. A standard argument applying the fundamental theorem of calculus to f along an appropriate space-time curve, and then using (20) to estimate the term involving $\partial_t f$ (see, for instance, [Kim et al. 2012, p. 4345, Proof of Theorem 5.2]), yields the parabolic Harnack inequality

$$\sup_{\Omega} v(\,\cdot\,, t) \le C \inf_{\Omega} v(\,\cdot\,, t+1) \quad \text{for all } t \ge 1, \tag{21}$$

where $C > 0$ is a constant independent of t. One then applies (21) to the nonnegative solutions

$$v_k^+(x, t) := \sup_{\Omega} v(\,\cdot\,, k) - v(x, k+t) \quad \text{and} \quad v_k^-(x, t) := v(x, k+t) - \inf_{\Omega} v(\,\cdot\,, k), \qquad k = 0, 1, 2, \dots,$$

to obtain decay of oscillation of θ in time; see [Kim et al. 2012, Section 7.1]. This shows that θ converges exponentially fast to a constant function on Ω as $t \to \infty$. Invoking (17), we conclude that $\lim_{t\to\infty} \theta \equiv 0$, and so $u(\,\cdot\,, t)$ converges exponentially fast as $t \to \infty$ to a function $u^\infty(\,\cdot\,)$ solving (2).

Below we show the sublinearity of F, which is a standard argument provided here for completeness. The proof relies on an important parabolic inequality satisfied by F, (24), which we will prove in Proposition 2.7 below.

Proposition 2.6. *If F does not attain a positive maximum on $\partial\Omega \times (0, T)$, then there exist constants C_1' and $C_2' > 0$ independent of T such that*

$$F(x, t) \le C_1' + C_2' t \quad \textit{for all } (x, t) \in \mathcal{C}_T. \tag{22}$$

Proof. First note that $F(\,\cdot\,, 0) \equiv 0$ because $\inf_\Omega v(\,\cdot\,, 0) > 0$, and so the bound holds at $t = 0$. Suppose there exists a first time $\tau \in (0, T)$ such that $F(y, \tau) \ge C_1' + C_2'\tau$ for some $y \in \Omega$. By going further in time if necessary, we may assume there exists a point $(x_0, t_0) \in \overline{\Omega} \times (0, T]$ such that $F(x_0, t_0) > C_1' + C_2' t_0$ and F attains a local maximum at (x_0, t_0). If (x_0, t_0) is an interior point of $\mathcal{C}_T$, it follows from (24) that

$$C_1 F(x_0, t_0)^2 - F(x_0, t_0) - C_2 t_0^2 \le 0,$$

from which we conclude

$$F(x_0, t_0) \le \frac{1 + \sqrt{1 + 4C_1C_2t_0^2}}{2C_1} \le \widetilde{C}_1 + \widetilde{C}_2 t_0 \tag{23}$$

for a different set of constants $\widetilde{C}_1, \widetilde{C}_2 > 0$ and for $t_0 > 0$ sufficiently large. If C_1', C_2' were chosen at the beginning to satisfy $C_1' > \widetilde{C}_1$ and $C_2' > \widetilde{C}_2$, then we reach a contradiction based on (23). □

Thus it is clear that on a manifold with no boundary, Proposition 2.6 combined with the discussion above yields exponential convergence, as is shown in [Kim et al. 2012].

We finish this section by establishing the parabolic inequality (24) satisfied by F. It is shown in [Kim et al. 2012, Proposition 5.1] that if $n \ge 3$ and

$$\psi(x, t) := \left(\frac{\rho^*(T(x, t))^2 \,\det D_x T(x, t)}{|\det D^2_{x,y} c(x, T(x, t))|} \right)^{1/(n-2)},$$

then

$$\mathcal{L}v = \psi \Delta_{\psi g} v - \partial_t v,$$

where $\Delta_{\psi g}$ is the Laplace–Beltrami operator with respect to the time-varying metric ψg with $g_{ij} := w_{ij}$. By adapting the proof of the differential Harnack inequality for the heat equation established in [Li and Yau 1986], the authors of [Kim et al. 2012] establish a parabolic inequality for F similar to (24) in the case of manifolds with no boundary of dimension $n \geq 3$. The case $n = 2$ is treated in [Kim et al. 2012] through the introduction of a third dummy dimension in a manner giving the solution u of (11) a product structure; see [Kim et al. 2012, Section 7.1.2] for details. In the presence of a boundary, such an argument for dealing with the two-dimensional case is almost certain to fail due to the requirement of uniform c- and c^*-convexity of the domains involved.

We elect to take a different approach which considers the weighted Laplacian $\Delta_\phi := \Delta_g - \langle \nabla^g \phi, \nabla^g \cdot \rangle_g$ for the manifold with density $(\Omega, g, e^{-\phi} d\,\mathrm{Vol}_g)$, where

$$\phi(x,t) := \log\left(\frac{|\det D^2_{x,y} c(x, T(x,t))|}{\rho^*(T(x,t))^2 \det D_x T(x,t)} \right)^{1/2}.$$

It was first noted in [Warren 2014, Section 3] that for such a choice of weighted manifold, $\mathcal{L} = \Delta_\phi - \partial_t$. The advantage of using this representation of $\mathcal{L}$ is that the case of dimension $n = 2$ does not need to be treated separately. As mentioned above, in the case of nonempty boundary, the conversion of the two-dimensional problem to a three-dimensional one as in [Kim et al. 2012] cannot be carried out. To summarize, the following proof follows the spirit of [Kim et al. 2012, Section 6] (which in turn is based on [Li and Yau 1986, Theorem 1.2]), but the details differ as we use the representation of the linearized operator as a weighted Laplacian from [Warren 2014], in contrast with the conformal factor approach used in [Kim et al. 2012].

Proposition 2.7. *Under the same hypotheses as Theorem 1.1, there exist constants C_1, C_2, and $C_3 > 0$, depending only on the constant $\mathcal{K}$ defined in (4) and the dimension n, such that whenever v satisfies $\mathcal{L}v = 0$,*

$$\mathcal{L}F + 2\langle \nabla^g f, \nabla^g F \rangle_g \geq \frac{1}{t}(C_1 F^2 - F - C_2 t^2 + C_3 t |\nabla^g f|_g^2 F). \tag{24}$$

Proof. We recall the well-known weighted Bochner formula

$$\Delta_\phi(|\nabla^g f|_g^2) = 2\|\mathrm{Hess}_g f\|^2 + 2\langle \nabla^g f, \nabla^g(\Delta_\phi f)\rangle_g + 2\,\mathrm{Ric}_\phi(\nabla^g f, \nabla^g f), \tag{25}$$

where $\mathrm{Ric}_\phi := \mathrm{Ric}_g + \mathrm{Hess}_g \phi$. Clearly, $\mathrm{Ric}_\phi \geq -\mathcal{K}$, where $\mathcal{K}$ is defined in (4). Since $\mathcal{L}v = 0$, the function $f := \log v$ solves the equation

$$\partial_t f = \Delta_\phi f + |\nabla^g f|_g^2. \tag{26}$$

Consider the auxiliary function

$$F := t(|\nabla^g f|_g^2 - \alpha\, \partial_t f), \quad \alpha > 0.$$

By using (25), we obtain

$$\begin{aligned} \Delta_\phi F &= t(\Delta_\phi(|\nabla^g f|_g^2) - \alpha \Delta_\phi(\partial_t f)) \\ &= t\big(2\|\mathrm{Hess}_g f\|^2 + 2\langle \nabla^g f, \nabla^g(\Delta_\phi f)\rangle_g + 2\,\mathrm{Ric}_\phi(\nabla^g f, \nabla^g f) - \alpha \Delta_\phi(\partial_t f)\big). \end{aligned}$$

Direct computation shows that

$$\Delta_\phi(\partial_t f) \le \partial_t(\Delta_\phi f) + C(\|\mathrm{Hess}_g f\| + |\nabla^g f|_g),$$

where $C = C(\partial_t g, \partial_t \nabla g, \partial_t \nabla\phi) \ge 0$ depends only on $\mathcal{K}$. Therefore,

$$\begin{aligned}\Delta_\phi F &\ge t\big(2\|\mathrm{Hess}_g f\|^2 + 2\langle\nabla^g f, \nabla^g(\Delta_\phi f)\rangle_g + 2\,\mathrm{Ric}_\phi(\nabla^g f, \nabla^g f) - \alpha\,\partial_t(\Delta_\phi f) - \alpha C(\|\mathrm{Hess}_g f\| + |\nabla^g f|_g)\big)\\ &\ge t\big(\|\mathrm{Hess}_g f\|^2 + 2\langle\nabla^g f, \nabla^g(\Delta_\phi f)\rangle_g - \alpha\,\partial_t(\Delta_\phi f) - C_1|\nabla^g f|_g^2 - C_2\big),\end{aligned}$$

where we have used Cauchy's inequality and the lower bound for Ric_ϕ. From (26) and the definition of F, it follows that

$$\Delta_\phi f = -\bigg(\frac{F}{t} + (\alpha-1)\,\partial_t f\bigg).$$

Therefore,

$$\begin{aligned}2\langle\nabla^g f, \nabla^g(\Delta_\phi f)\rangle_g &= -2\bigg\langle\nabla^g f, \nabla^g\bigg(\frac{F}{t} + (\alpha-1)\,\partial_t f\bigg)\bigg\rangle_g\\ &= -\frac{2}{t}\langle\nabla^g f, \nabla^g F\rangle_g - 2(\alpha-1)\langle\nabla^g f, \nabla^g(\partial_t f)\rangle_g.\end{aligned}$$

Furthermore,

$$\partial_t F = \frac{F}{t} + t(\partial_t|\nabla^g f|_g^2 - \alpha\,\partial_t^2 f).$$

Therefore,

$$\begin{aligned}-\alpha\,\partial_t(\Delta_\phi f) &= \alpha\,\partial_t\bigg(\frac{F}{t} + (\alpha-1)\,\partial_t f\bigg)\\ &= \alpha\bigg(\frac{\partial_t F}{t} - \frac{F}{t^2} + (\alpha-1)\,\partial_t^2 f\bigg)\\ &= \alpha\bigg(\frac{\partial_t F}{t} - \frac{F}{t^2}\bigg) + (\alpha-1)\alpha\,\partial_t^2 f\\ &= \alpha\bigg(\frac{\partial_t F}{t} - \frac{F}{t^2}\bigg) + (\alpha-1)\bigg(\frac{F}{t^2} - \frac{\partial_t F}{t} + \partial_t|\nabla^g f|_g^2\bigg)\\ &= \frac{\partial_t F}{t} - \frac{F}{t^2} + (\alpha-1)\,\partial_t|\nabla^g f|_g^2.\end{aligned}$$

It follows that

$$\begin{aligned}&2\langle\nabla^g f, \nabla^g(\Delta_\phi f)\rangle_g - \alpha\,\partial_t(\Delta_\phi f)\\ &\qquad= \frac{1}{t}\bigg(\partial_t F - 2\langle\nabla^g f, \nabla^g F\rangle_g - \frac{F}{t}\bigg) + (\alpha-1)(\partial_t|\nabla^g f|_g^2 - 2\langle\nabla^g f, \nabla^g(\partial_t f)\rangle_g)\\ &\qquad\ge \frac{1}{t}\bigg(\partial_t F - 2\langle\nabla^g f, \nabla^g F\rangle_g - \frac{F}{t}\bigg) - C_3|\nabla^g f|_g^2,\end{aligned}$$

where we have used the fact

$$\partial_t|\nabla^g f|_g^2 \le 2\langle\nabla^g f, \nabla^g(\partial_t f)\rangle_g + \gamma|\nabla^g f|_g^2$$

for some constant $\gamma = \gamma(\partial_t g) \geq 0$ depending only on $\mathcal{K}$. Inserting the above inequality into the lower bound for $\Delta_\phi F$ yields

$$\Delta_\phi F \geq t\left(\|\mathrm{Hess}_g f\|^2 + \frac{1}{t}\left(\partial_t F - 2\langle \nabla^g f, \nabla^g F\rangle_g - \frac{F}{t}\right) - C_4|\nabla^g f|_g^2 - C_2\right).$$

Now since $\Delta_\phi f = \Delta_g f - \langle \nabla^g \phi, \nabla^g f\rangle_g$, we have

$$\begin{aligned}(\Delta_g f)^2 = (\Delta_\phi f + \langle \nabla^g\phi, \nabla^g f\rangle_g)^2 &= (\Delta_\phi f)^2 + \langle \nabla^g\phi, \nabla^g f\rangle_g^2 + 2(\Delta_\phi f)\langle \nabla^g\phi, \nabla^g f\rangle_g \\ &\geq (\Delta_\phi f)^2 + \langle \nabla^g\phi, \nabla^g f\rangle_g^2 - \frac{(\Delta_\phi f)^2}{2} - 2\langle \nabla^g\phi, \nabla^g f\rangle_g^2 \\ &= \frac{(\Delta_\phi f)^2}{2} - \langle \nabla^g\phi, \nabla^g f\rangle_g^2 \\ &\geq \frac{(\Delta_\phi f)^2}{2} - |\nabla^g\phi|_g^2|\nabla^g f|_g^2.\end{aligned}$$

Therefore, by the arithmetic-geometric mean inequality, we have

$$\|\mathrm{Hess}_g f\|^2 \geq \frac{1}{n}(\Delta_g f)^2 \geq \frac{(\Delta_\phi f)^2}{2n} - \frac{1}{n}|\nabla^g\phi|_g^2|\nabla^g f|_g^2.$$

Since $|\nabla^g\phi|_g \leq \mathcal{K}$, we obtain

$$\Delta_\phi F \geq \frac{t}{2n}(\Delta_\phi f)^2 + \partial_t F - 2\langle \nabla^g f, \nabla^g F\rangle_g - \frac{F}{t} - C_5 t|\nabla^g f|_g^2 - C_2 t.$$

Finally, by (26), and after relabeling constants, we conclude that

$$\Delta_\phi F + 2\langle \nabla^g f, \nabla^g F\rangle_g - \partial_t F \geq \frac{1}{t}[C_1 t^2(|\nabla^g f|_g^2 - \partial_t f)^2 - F - C_2 t^2|\nabla^g f|_g^2 - C_3 t^2].$$

Here the constants $C_1, C_2, C_3 > 0$ depend only on up to fourth-order derivatives of the cost function (through $\mathrm{Hess}_g \phi$) and the $C_x^4 C_t^1$ norm of the solution u to (11) (through the time derivative of g and bounds on the Ricci curvature of g), and hence only on $\mathcal{K}$ and on the dimension n.

Let $y = |\nabla^g f|_g^2$ and $z = \partial_t f$. Then for any $\alpha, \epsilon, \delta > 0$, we have the identity

$$(y-z)^2 = \left(\frac{1}{\alpha} - \frac{\epsilon}{2}\right)(y-\alpha z)^2 + \left(1 - \frac{\epsilon}{2} - \delta - \frac{1}{\alpha}\right)y^2 + \left(1 - \alpha + \frac{\epsilon}{2}\alpha^2\right)z^2 + \epsilon y(y-\alpha z) + \delta y^2.$$

We now choose $\alpha, \epsilon > 0$ such that

$$1 - \frac{\epsilon}{2} - \frac{1}{\alpha} > 0, \quad 1 - \alpha + \frac{\epsilon}{2}\alpha^2 \geq 0, \quad \frac{1}{\alpha} - \frac{\epsilon}{2} > 0.$$

Note that these conditions impose the restriction $\alpha > 1$. A direct verification shows that $\alpha = 2$ and $\epsilon = \frac{1}{2}$ satisfy the above inequalities. We then choose $\delta = \frac{1}{8} \in \left(0, 1 - \frac{\epsilon}{2} - \frac{1}{\alpha}\right) = \left(0, \frac{1}{4}\right)$. With these choices of α, ϵ, δ, we obtain (discarding the second and third terms in the expansion, and using that $F = t(y - \alpha z)$)

$$\Delta_\phi F + 2\langle \nabla^g f, \nabla^g F\rangle_g - \partial_t F \geq \frac{1}{t}\left[C_1 t^2\left\{\frac{F^2}{4t^2} + y\frac{F}{2t} + \frac{y^2}{8}\right\} - F - C_2 t^2 y - C_3 t^2\right].$$

Using Cauchy's inequality, we may eliminate the $-C_2t^2y$ and $C_1t^2y^2/8$ terms to get

$$\Delta_\phi F + 2\langle \nabla^g f, \nabla^g F\rangle_g - \partial_t F \geq \frac{1}{t}\left[C_1 t^2\left\{\frac{F^2}{4t^2} + y\frac{F}{2t}\right\} - F - C_4 t^2\right].$$

Relabeling constants, we have thus established an inequality of the form (24). □

3. Sublinearity of F on domains with boundary

On a domain with boundary, one must deal with the possibility that F attains a maximum at a point $(x_0, t_0) \in \partial_P \mathcal{C}_T = (\bar{\Omega} \times \{0\}) \cup (\partial\Omega \times (0, T))$, the parabolic boundary of the cylinder $\mathcal{C}_T$. Since $F \equiv 0$ on $\bar{\Omega} \times \{0\}$, it suffices to assume $(x_0, t_0) \in \partial\Omega \times (0, T)$. The original argument of [Li and Yau 1986, proof of Theorem 1.1] in the case of the heat equation eliminates the possibility of F attaining a nonnegative maximum on $\partial\Omega \times (0, T)$ by means of a contradiction to Hopf's lemma. For this, they require two additional hypotheses: namely, the solution to the heat equation also satisfies a Neumann boundary condition, and the boundary is mean-convex.

We will obtain a similar contradiction to Hopf's lemma only for the particular nonnegative solution $\Theta(x, t) := \sup_\Omega \theta(\cdot, 0) - \theta(x, t)$ of the linearized equation (18) (as well as for translations of Θ in time) by exploiting the boundary condition $D_\beta\Theta = 0$ on $\partial\Omega \times [0, T]$, and using the assumption that the domains Ω, Ω^* are respectively c-convex and c^*-convex. This gives the desired sublinearity at the boundary of the corresponding function F defined in (19) and establishes the Harnack inequality (21) for Θ, which turns out to be sufficient to prove the exponential convergence of $u(\cdot, t)$ to the steady state solution $u^\infty(\cdot)$ as $t \to \infty$ (see Section 4). As mentioned in the Introduction, it is unclear if such a sublinearity estimate at the boundary holds for an arbitrary nonnegative solution v of the linearized equation (18).

Let us carry on with the proof of the sublinearity of F outlined in Proposition 2.6, now assuming there exists $(x_0, t_0) \in \partial\Omega \times (0, T)$ such that $F(x_0, t_0) > C_1' + C_2' t_0$ and F attains a local maximum at (x_0, t_0). It follows from (24) that, in a spherical cap near (x_0, t_0), we have

$$\mathcal{L}F + 2\langle \nabla^g f, \nabla^g F\rangle_g \geq 0.$$

By the uniform obliqueness of β and Hopf's lemma, it follows that $D_\beta F(x_0, t_0) > 0$. Anticipating a contradiction, we proceed to explicitly compute $D_\beta F(x_0, t_0)$. We first make a rotation centered at x_0 so the directions $e_1, \ldots, e_{n-1}$ form an orthonormal basis for the tangent space to $\partial\Omega$ at x_0, and the direction e_n is the outward-pointing unit normal direction to $\partial\Omega$ at x_0. Differentiating F in these coordinates, we find that

$$\begin{aligned} D_\beta F(x_0, t_0) &= D_\beta|_{(x_0,t_0)} t(w^{ij} f_i f_j - \alpha\, \partial_t f) \\ &= t_0[(D_\beta w^{ij}) f_i f_j + 2w^{ij}(D_\beta f_i) f_j - \alpha D_\beta(\partial_t f)]\big|_{(x_0,t_0)} \\ &= t_0[-w^{i\ell} w^{jk}(D_\beta w_{\ell k}) f_i f_j + 2w^{ij}((D_\beta f)_i - \beta_i^k f_k) f_j - \alpha(\partial_t(D_\beta f) - (\partial_t \beta^k) f_k)]\big|_{(x_0,t_0)}. \end{aligned}$$

Now since $D_\beta f = D_\beta v/v = 0$ on $\partial\Omega$, we have $\partial_t(D_\beta f) = 0$ and $(D_\beta f)_i = 0$ for $i = 1, \ldots, n-1$. Therefore,

$$D_\beta F(x_0, t_0) = t_0[-w^{i\ell} w^{jk}(D_\beta w_{\ell k}) f_i f_j - 2w^{ij}\beta_i^k f_k f_j + 2w^{nj} f_j (D_\beta f)_n + \alpha(\partial_t \beta^k) f_k]\big|_{(x_0,t_0)}.$$

We claim $w^{nj} f_j = 0$ at (x_0, t_0). By Lemma 2.4, $W\beta$ is parallel to the outward-pointing unit normal vector ν on $\partial\Omega$, so $\nu = (1/\chi)W\beta$, where $\chi := |W\beta|$. Again since $D_\beta f = 0$ on $\partial\Omega$,

$$0 = \langle \beta, \nabla f\rangle = \langle W^{-1}W\beta, \nabla f\rangle = \langle W\beta, W^{-1}\nabla f\rangle.$$

Hence,

$$\tau := W^{-1}\nabla f \tag{27}$$

is tangent to $\partial\Omega$. In the coordinate system defined above, we have $\nu(x_0, t_0) = e_n$, and so $\tau^n(x_0, t_0) = 0$. Since $\tau^n = w^{nj} f_j$, the claim is proved. It follows that

$$D_\beta F(x_0, t_0) = t_0[-(D_\beta w_{k\ell})\tau^k\tau^\ell - 2\beta_i^k f_k \tau^i + \alpha(\partial_t \beta^k) f_k]\big|_{(x_0,t_0)}. \tag{28}$$

Note that since $\tau_n = 0$ at (x_0, t_0), it suffices to sum the indices in the first term over $k, \ell = 1, \dots, n-1$.

3A. *Inner product cost.* We first show how to explicitly compute $D_\beta F(x_0, t_0)$ in the case when the cost function is given by the Euclidean inner product on $\mathbb{R}^n$ (which is known to be equivalent to taking the cost function to be the Euclidean distance squared). There are a number of simplifications in this case, as $Y(x, p) = p$, $W(x, t) = D^2u(x, t)$, and c- and c^*-convexity of sets and functions reduce to the usual notions of convexity of the domains Ω and Ω^*.

Proposition 3.1. *If* $c(x, y) = \langle x, y\rangle$,

$$D_\beta F(x_0, t_0) = t_0[-\chi\langle (D\nu)\tau, \tau\rangle - \langle D^2h^*(\nabla u)\nabla f, \nabla f\rangle + \alpha\langle D^2h^*(\nabla u)\nabla\theta, \nabla f\rangle]\big|_{(x_0,t_0)}. \tag{29}$$

Proof. We have

$$W = D^2u, \quad \beta = \nabla h^*(\nabla u).$$

Consequently, $\nu = (1/\chi)(D^2u)\beta$. Differentiating ν^k in the e_ℓ-direction for $k, \ell = 1, \dots, n-1$, we find

$$\begin{aligned} \nu_\ell^k = \left(\frac{1}{\chi}u_{kr}\beta^r\right)_\ell &= \frac{1}{\chi}(u_{\ell kr}\beta^r + u_{kr}\beta_\ell^r) - \frac{\chi_\ell}{\chi^2}(u_{kr}\beta^r) \\ &= \frac{1}{\chi}(D_\beta u_{\ell k} + u_{kr}\beta_\ell^r) - (\log\chi)_\ell \nu^k. \end{aligned}$$

Solving for $D_\beta u_{\ell k}$, we obtain

$$D_\beta u_{\ell k} = \chi\nu_\ell^k - u_{kr}\beta_\ell^r + \chi(\log\chi)_\ell\nu^k.$$

Therefore at (x_0, t_0), we have (recall (27))

$$\begin{aligned} -(D_\beta u_{\ell k})\tau^\ell\tau^k &= -(\chi\nu_\ell^k - u_{kr}\beta_\ell^r + \chi(\log\chi)_\ell\nu^k)\tau^\ell\tau^k \\ &= -\chi\nu_\ell^k\tau^\ell\tau^k + u_{kr}\tau^k\beta_\ell^r\tau^\ell \\ &= -\chi\nu_\ell^k\tau^\ell\tau^k + f_r\beta_\ell^r\tau^\ell, \end{aligned}$$

where we sum the indices k, ℓ from 1 to $n-1$. Substituting this into (28) gives

$$D_\beta F(x_0, t_0) = t_0[-\chi\nu_\ell^k\tau^\ell\tau^k - \beta_i^k f_k\tau^i + \alpha(\partial_t\beta^k) f_k]\big|_{(x_0,t_0)}.$$

Since $\beta(x,t) = \nabla h^*(\nabla u(x,t))$, we find that

$$\beta_i^k f_k \tau^i = h^*_{k\ell}(\nabla u) u_{\ell i} f_k \tau^i = h^*_{k\ell}(\nabla u) f_k u_{\ell i} \tau^i = h^*_{k\ell}(\nabla u) f_k f_\ell,$$

and

$$(\partial_t \beta^k) f_k = h^*_{k\ell}(\nabla u)(\partial_t u_\ell) f_k = h^*_{k\ell}(\nabla u)\theta_\ell f_k;$$

hence (29) follows. □

3B. *General cost.* We now show how to explicitly compute $D_\beta F(x_0,t_0)$ in the case of a general cost.

Proposition 3.2. *We have*

$$D_\beta F(x_0,t_0) = t_0[-\chi(\nu_i^j - c^{r,\ell} c_{ij,r} \nu^\ell)\tau^i\tau^j - G_{p_k p_s}(x,\nabla u) f_k f_s + \alpha\, G_{p_k p_s}(x,\nabla u) f_k \theta_s]\big|_{(x_0,t_0)}. \tag{30}$$

Proof. We have

$$w_{jk}(x,t) = u_{jk}(x,t) - c_{jk}(x,T(x,t)), \qquad \beta^k(x,t) = h^*_\ell(Y(x,\nabla u(x,t)))Y^\ell_{p_k}(x,\nabla u(x,t)).$$

Recall that $\nu = (1/\chi)W\beta$. As in the case of the inner product cost, we differentiate ν^j in the e_i-direction for $i, j = 1, \dots, n-1$ to get

$$\begin{aligned}\nu_i^j &= \left(\frac{1}{\chi} w_{jk}\beta^k\right)_i = \frac{1}{\chi}((w_{jk})_i \beta^k + w_{jk}\beta_i^k) - \frac{\chi_i}{\chi^2}(w_{jk}\beta^k)\\ &= \frac{1}{\chi}((w_{jk})_i\beta^k + w_{jk}\beta_i^k) - (\log\chi)_i \nu^j.\end{aligned}$$

Differentiating w_{jk} gives

$$\begin{aligned}(w_{jk})_i &= u_{jki} - c_{jki} - c_{jk,r} T_i^r\\ &= (w_{ij})_k + c_{ij,r} T_k^r - c_{jk,r} T_i^r\\ &= (w_{ij})_k + c_{ij,r} c^{r,\ell} w_{\ell k} - c_{jk,r} c^{r,\ell} w_{\ell i},\end{aligned}$$

where we have used (14) in the final line. Therefore,

$$\begin{aligned}\nu_i^j &= \frac{1}{\chi}((w_{jk})_i\beta^k + w_{jk}\beta_i^k) - (\log\chi)_i\nu^j\\ &= \frac{1}{\chi}\big([(w_{ij})_k + c_{ij,r}c^{r,\ell}w_{\ell k} - c_{jk,r}c^{r,\ell}w_{\ell i}]\beta^k + w_{jk}\beta_i^k\big) - (\log\chi)_i\nu^j\\ &= \frac{1}{\chi}\big(D_\beta w_{ij} + [c_{ij,r}c^{r,\ell}w_{\ell k} - c_{jk,r}c^{r,\ell}w_{\ell i}]\beta^k + w_{jk}\beta_i^k\big) - (\log\chi)_i\nu^j.\end{aligned}$$

Solving for $D_\beta w_{ij}$, we obtain

$$D_\beta w_{ij} = \chi\nu_i^j - [c_{ij,r}c^{r,\ell}w_{\ell k} - c_{jk,r}c^{r,\ell}w_{\ell i}]\beta^k - w_{jk}\beta_i^k + \chi(\log\chi)_i\nu^j.$$

Therefore, at (x_0,t_0), we have (again using (27))

$$\begin{aligned}-(D_\beta w_{ij})\tau^i\tau^j &= -\big(\chi\nu_i^j - [c_{ij,r}c^{r,\ell}w_{\ell k} - c_{jk,r}c^{r,\ell}w_{\ell i}]\beta^k - w_{jk}\beta_i^k + \chi(\log\chi)_i\nu^j\big)\tau^i\tau^j\\ &= -(\chi\nu_i^j - c_{ij,r}c^{r,\ell}w_{\ell k}\beta^k + c_{jk,r}c^{r,\ell}w_{\ell i}\beta^k)\tau^i\tau^j + w_{jk}\beta_i^k\tau^i\tau^j\\ &= -\chi(\nu_i^j - c^{r,\ell}c_{ij,r}\nu^\ell)\tau^i\tau^j - c^{r,\ell}c_{jk,r}f_\ell\beta^k\tau^j + f_k\beta_i^k\tau^i\\ &= -\chi(\nu_i^j - c^{r,\ell}c_{ij,r}\nu^\ell)\tau^i\tau^j - c_{jk,r}h^*_s Y^s_{p_k} Y^r_{p_\ell} f_\ell\tau^j + f_k\beta_i^k\tau^i\end{aligned}$$

where we sum the indices i, j from 1 to $n-1$. It follows from (28) that

$$D_\beta F(x_0,t_0)=t_0[-\chi(\nu_i^j-c^{r,\ell}c_{ij,r}\nu^\ell)\tau^i\tau^j-c_{jk,r}h_s^*Y^s_{p_k}Y^r_{p_\ell}f_\ell\tau^j-f_k\beta_i^k\tau^i+\alpha\,\partial_t\beta^k f_k]\big|_{(x_0,t_0)}. \tag{31}$$

We compute

$$\beta_i^k=h^*_{\ell r}(Y^r_{x_i}+Y^r_{p_s}u_{si})Y^\ell_{p_k}+h^*_\ell(Y^\ell_{p_kx_i}+Y^\ell_{p_kp_s}u_{si}). \tag{32}$$

To simplify the first term, recall the identity (see (14))

$$Y^r_{x_i}+Y^r_{p_s}u_{si}=Y^r_{p_s}w_{si}.$$

For the second term in (32), we differentiate the equation $c_{i,\ell}Y^\ell_{p_k}=\delta_{ik}$ with respect to p_s and x_i to obtain

$$Y^\ell_{p_kp_s}=-c^{\ell,j}c_{j,rq}Y^r_{p_k}Y^q_{p_s}$$

and

$$Y^\ell_{p_kx_i}=-c^{\ell,j}c_{ij,r}Y^r_{p_k}+c^{\ell,j}c_{j,rq}Y^r_{p_k}Y^q_{p_s}c_{si}=-c_{ij,r}Y^\ell_{p_j}Y^r_{p_k}-Y^\ell_{p_kp_s}c_{si}.$$

Therefore,

$$Y^\ell_{p_kx_i}+Y^\ell_{p_kp_s}u_{si}=-c_{ij,r}Y^\ell_{p_j}Y^r_{p_k}+w_{si}Y^\ell_{p_kp_s}.$$

Substituting these into the expression (32) gives

$$\beta_i^k=h^*_{\ell r}Y^\ell_{p_k}Y^r_{p_s}w_{si}+h^*_\ell(-c_{ij,r}Y^\ell_{p_j}Y^r_{p_k}+w_{si}Y^\ell_{p_kp_s}).$$

Therefore,

$$\begin{aligned}f_k\beta_i^k\tau^i&=h^*_{\ell r}Y^\ell_{p_k}Y^r_{p_s}f_kw_{si}\tau^i-c_{ij,r}h^*_\ell Y^\ell_{p_j}Y^r_{p_k}f_k\tau^i+h^*_\ell Y^\ell_{p_kp_s}f_kw_{si}\tau^i\\&=h^*_{\ell r}Y^\ell_{p_k}Y^r_{p_s}f_kf_s-c_{ij,r}h^*_\ell Y^\ell_{p_j}Y^r_{p_k}f_k\tau^i+h^*_\ell Y^\ell_{p_kp_s}f_kf_s.\end{aligned}$$

Substituting into (31) and observing that the second term in the above expression cancels the term $-c_{jk,r}h_s^*Y^r_{p_\ell}Y^s_{p_k}f_\ell\tau^j$ in (31), we obtain

$$D_\beta F(x_0,t_0)=t_0[-\chi(\nu_i^j-c^{r,\ell}c_{ij,r}\nu^\ell)\tau^i\tau^j-(h^*_{\ell r}Y^\ell_{p_k}Y^r_{p_s}+h^*_\ell Y^\ell_{p_kp_s})f_kf_s+\alpha\beta_t^kf_k]\big|_{(x_0,t_0)}.$$

Next, we compute

$$\partial_t\beta^k=(h^*_{\ell r}Y^\ell_{p_k}Y^r_{p_s}+h^*_\ell Y^\ell_{p_kp_s})\theta_s.$$

Finally, noticing that $h^*_{\ell r}Y^\ell_{p_k}Y^r_{p_s}+h^*_\ell Y^\ell_{p_kp_s}=G_{p_kp_s}$, we obtain the claimed expression (30). □

4. Proof of exponential convergence

With Propositions 3.1 and 3.2 in hand, we may now prove our main result. We note that the proof presented here is different from the standard proof of exponential convergence via parabolic Harnack inequality outlined in Section 2D, and explicitly uses special properties of the underlying optimal transport problem. In particular, we must take into consideration the c- and c^*-convexity of both domains Ω and Ω^*, and judiciously choose special solutions of the linearized problem that will allow us to utilize the expressions for $D_\beta F$ obtained in Propositions 3.1 and 3.2.

Proof of Theorem 1.1. Consider the function

$$\Theta(x,t) = \sup_{\Omega} \theta(\,\cdot\,,0) - \theta(x,t),$$

which satisfies (18), and is nonnegative by Proposition 2.5. We claim $D_\beta F(x_0,t_0) \le 0$ when $v = \Theta$, which will contradict Hopf's lemma, thus proving F cannot attain a positive maximum on $\partial\Omega \times (0,T)$.

Let us first deal with the case of the inner product cost. Since the domain Ω is convex, we have $\langle (D\nu)\tau, \tau\rangle \ge 0$. Therefore, since $\chi \ge 0$, we obtain using Proposition 3.1

$$D_\beta F(x_0,t_0) \le t_0[-\langle D^2h^*(\nabla u)\nabla f, \nabla f\rangle + \alpha\langle D^2h^*(\nabla u)\nabla\theta, \nabla f\rangle]\big|_{(x_0,t_0)}. \tag{33}$$

Next, the convexity of Ω^* implies D^2h^* is nonnegative, so by substituting for $f = \log\Theta$ in (33), we find

$$D_\beta F(x_0,t_0) \le t_0\bigg[-\frac{1}{\Theta^2}\langle D^2h^*(\nabla u)\nabla\theta, \nabla\theta\rangle - \frac{\alpha}{\Theta}\langle D^2h^*(\nabla u)\nabla\theta, \nabla\theta\rangle\bigg]\bigg|_{(x_0,t_0)} \le 0.$$

This is the desired contradiction to Hopf's lemma. For general costs, we use Proposition 3.2, noticing that c-convexity of Ω with respect to Ω^*, given in (8), implies $(\nu_i^j - c^{r,\ell}c_{ij,r}\nu^\ell)\tau^i\tau^j \ge 0$, while the c^*-convexity of Ω^* with respect to Ω, given in (9), implies $G_{p_k p_s}$ is a nonnegative matrix.

It follows from Proposition 2.6 that with the choice $v = \Theta$, the corresponding function F defined in (19) is sublinear in time, and consequently the Harnack inequality (21) holds for Θ. Using this Harnack inequality, we now prove exponential convergence of $\theta(\,\cdot\,,t)$. The argument is similar to [Kim et al. 2012, Section 7], but differs in an essential manner. For each integer $k \ge 1$, consider the function

$$\Theta_k(x,t) := \sup_{\Omega} \theta(\,\cdot\,,k-1) - \theta(x,(k-1)+t).$$

The functions Θ_k are nonnegative by Proposition 2.5 and solve (18). Arguing as above, the corresponding functions F for $v = \Theta_k$ are also sublinear in t (with constants independent of k) and thus the Harnack inequality (21) holds for Θ_k. Applying (21) to Θ_k at $t = 1$ yields

$$\sup_{\Omega} \theta(\,\cdot\,,k-1) - \inf_{\Omega} \theta(\,\cdot\,,k) \le C\big(\sup_{\Omega} \theta(\,\cdot\,,k-1) - \sup_{\Omega} \theta(\,\cdot\,,k+1)\big). \tag{34}$$

Now by (17), we know $\inf_\Omega \theta(\,\cdot\,,k) \le 0$ for each k. Therefore, defining $\epsilon := (C-1)/C < 1$, we find

$$\sup_{\Omega} \theta(\,\cdot\,,k+1) \le \epsilon \sup_{\Omega} \theta(\,\cdot\,,k-1).$$

Iterating this inequality gives the exponential decay of the supremum

$$\sup_{\Omega} \theta(\,\cdot\,,t) \le \sup_{\Omega} \theta(\,\cdot\,,0)e^{-\sigma t}, \quad \text{where } e^{-\sigma} = \epsilon. \tag{35}$$

On the other hand, (34) implies

$$\inf_{\Omega} \theta(\,\cdot\,,k) \ge -(C-1)\sup_{\Omega} \theta(\,\cdot\,,k-1) + C\sup_{\Omega} \theta(\,\cdot\,,k+1) \ge -(C-1)\sup_{\Omega} \theta(\,\cdot\,,k-1),$$

where we have used (17) again to throw away the term $\sup_\Omega \theta(\,\cdot\,,k+1)$. Therefore, by (35), we obtain

$$\inf_{\Omega} \theta(\,\cdot\,,k) \ge -(C-1)\sup_{\Omega} \theta(\,\cdot\,,0)e^{-\sigma(k-1)}. \tag{36}$$

This implies the exponential convergence of $\inf_\Omega \theta(\,\cdot\,, t)$, which combined with (35) gives the desired exponential convergence of $\theta(\,\cdot\,, t)$ to zero. □

5. A geometric approach to sublinearity at the boundary

In this section, we present an alternative approach to the computation of $D_\beta F(x_0, t_0)$ arising in the boundary sublinearity above. We will accomplish this using geometric language, exploiting the pseudo-Riemannian framework for optimal transport developed in [Kim and McCann 2010]. All material in this section is new, to the best knowledge of the authors, and constitutes the first treatment of the boundary geometry of domains in the context of the Kim–McCann metric.

In order to stay in line with established conventions, in this section we will mostly follow the notation used in [Kim and McCann 2010]. Thus in this section only, we will refer to the source and target domains as Ω and $\bar{\Omega}$ respectively (in particular, $\bar{\Omega}$ does not denote the closure of a set), which we assume are subsets of some fixed Riemannian manifolds. Points with a bar above will belong to $\bar{\Omega}$, while those without will belong to Ω. We also adopt the Einstein summation convention with the caveat that any indices given by Greek letters will run from 1 to $2n$, while lower case Roman indices run between 1 and n with the convention that an index with a bar above will be that value with n added to it: in other words, $1 \le \gamma \le 2n,\ 1 \le i \le n$ and $\bar{i} := i + n$.

Additionally, we will switch sign conventions at this point to stay in line with the definitions of [Kim and McCann 2010]. This means that c will be replaced by $-c$ everywhere, and the optimal transport problem (1) that is considered will be a minimization instead of a maximization problem.

We also split the tangent and cotangent spaces of $\Omega \times \bar{\Omega}$ in the canonical way according to the product structure, which gives the splitting $dc = Dc \oplus \bar{D}c$ of the one form dc on $\Omega \times \bar{\Omega}$, and given any local coordinate system on $\Omega \times \bar{\Omega}$ we will use the notation X to denote the full $2n$-dimensional coordinate variable: thus given a point $X = (x, \bar{x}) \in \Omega \times \bar{\Omega}$, X^i will indicate the i-th coordinate of x with $1 \le i \le n$, and $X^{\bar{i}}$ will indicate the i-th coordinate of $\bar{x}$. We will also suppress the time variable in this section, as everything considered will be for a fixed time t (in fact, the time dependency of the potential u will be completely irrelevant in the results of this section). Finally, we use the notation

$$[\Omega]_{\bar{x}} := -\bar{D}c(\Omega, \bar{x}) \subset T^*_{\bar{x}}\bar{\Omega}, \quad [\bar{\Omega}]_x := -Dc(x, \bar{\Omega}) \subset T^*_x\Omega \quad \text{for any } (x, \bar{x}) \in \Omega \times \bar{\Omega}.$$

Equip Ω with the pullback metric $w := (\mathrm{Id} \times T)^* h$, where

$$h := \frac{1}{2}\begin{pmatrix} 0 & -\bar{D}Dc \\ -D\bar{D}c & 0 \end{pmatrix}$$

is the Kim–McCann (pseudo-Riemannian) metric on $\Omega \times \bar{\Omega}$ defined as in [Kim and McCann 2010, (2.1)]. By [Kim et al. 2010, Section 3.2], in Euclidean coordinates the coefficients of w at x are exactly $w_{ij}(x) = u_{ij}(x) + c_{ij}(x, T(x))$, and w is a Riemannian metric. We will write ∇^w and ∇^h for the Levi-Civita connections of w and h respectively, Γ for the Christoffel symbols of h, and $|\cdot|_w$ for the length of a vector in w. We will also metrically identify various cotangent spaces naturally with $\mathbb{R}^n$ through the underlying Riemannian metrics on Ω or $\bar{\Omega}$. The inner products and norms in these underlying metrics will be denoted by $\langle\,\cdot\,,\,\cdot\,\rangle$ and $|\cdot|$ respectively. Our main result of the section is the following.

Theorem 5.1. *Let II^w be the second fundamental form of $\partial\Omega$ defined with respect to the metric w, and fix a point $x_0 \in \partial\Omega$. If $\mathrm{II}^{\partial[\Omega]_{T(x_0)}}$, $\mathrm{II}^{\partial[\bar{\Omega}]_{x_0}}$ are the (Euclidean) second fundamental forms of $\partial[\Omega]_{T(x_0)}$ and $\partial[\bar{\Omega}]_{x_0}$ respectively, then for any $\tau_1, \tau_2 \in T_{x_0}\partial\Omega$ we have*

$$2|\beta(x_0)|_w \mathrm{II}^w_{x_0}(\tau_1,\tau_2) = |DT(x_0)\beta(x_0)|\mathrm{II}^{\partial[\Omega]_{T(x_0)}}_{-\bar{D}c(x_0,T(x_0))}(\hat{\tau}_1,\hat{\tau}_2) + |\beta(x_0)|\mathrm{II}^{\partial[\bar{\Omega}]_{x_0}}_{-Dc(x_0,T(x_0))}(\hat{\bar{\tau}}_1,\hat{\bar{\tau}}_2), \tag{37}$$

where

$$\hat{\tau}_i := -D\bar{D}c(x_0,T(x_0))\tau_i \in T^*_{T(x_0)}\bar{\Omega},$$
$$\hat{\bar{\tau}}_i := -\bar{D}Dc(x_0,T(x_0))DT(x_0)\tau_i \in T^*_{x_0}\Omega.$$

Proof. Fix any point $x_0 \in \partial\Omega$. Note by Lemma 2.4 that $\beta(x_0)$ is an (outward) normal to $\partial\Omega$ at x_0 with respect to the metric w. Then since $\mathrm{Id}\times T$ is an embedding of Ω into $\Omega\times\bar{\Omega}$, if ∇^h is the Levi-Civita connection of h, we have (using that τ_2 is tangent to $\partial\Omega$ in the second line)

$$\begin{aligned}
\mathrm{II}^w_{x_0}(\tau_1,\tau_2) &= w\left(\nabla^w_{\tau_1}\frac{\beta}{|\beta|_w},\tau_2\right) = |\beta|_w^{-1} w(\nabla^w_{\tau_1}\beta,\tau_2) + D_{\tau_1}(|\beta|_w^{-1})w(\beta,\tau_2)\\
&= |\beta|_w^{-1} w(\nabla^w_{\tau_1}\beta,\tau_2) = -|\beta|_w^{-1} w(\beta,\nabla^w_{\tau_1}\tau_2)\\
&= -|\beta|_w^{-1} h\big((\beta\oplus DT(x_0)\beta),\nabla^h_{(\tau_1\oplus DT(x_0)\tau_1)}(\tau_2\oplus DT(x_0)\tau_2)\big)\\
&= -|\beta|_w^{-1}(\beta\oplus DT(x_0)\beta)^\flat[\nabla^h_{(\tau_1\oplus DT(x_0)\tau_1)}(\tau_2\oplus DT(x_0)\tau_2)],
\end{aligned} \tag{38}$$

where $\flat$ is the operation of lowering the indices of a tangent vector to $\Omega\times\bar{\Omega}$ by the metric h. Next consider the mapping $\Phi(x,\bar{x}) := -Dc(x_0,\bar{x})\oplus(-\bar{D}c(x,T(x_0)))$. By the bitwist condition (6), Φ is a diffeomorphism on $\Omega\times\bar{\Omega}$; hence Φ^{-1} gives a global coordinate chart on the set. We will use hats to denote quantities related to h written in the coordinates given by Φ^{-1}, while quantities without hats will be in Euclidean coordinates. A quick calculation yields that

$$\frac{\partial\Phi^\delta}{\partial X^\gamma}(x_0,T(x_0)) = 2h_{\delta\gamma}(x_0,T(x_0)). \tag{39}$$

We will now calculate the Christoffel symbols $\widehat{\Gamma}^\delta_{\gamma\lambda}$ in the coordinates given by Φ^{-1}. By [Kim and McCann 2010, Lemma 4.1] the Christoffel symbols of h in Euclidean coordinates are identically zero unless all three of the indices are simultaneously between 1 and n, or between $n+1$ and $2n$. Thus the standard transformation law shows that in the coordinates given by Φ^{-1}, the only Christoffel symbols that can be nonzero are those where either the upper index is not barred and both lower indices are, or the upper index is barred and both lower indices are not. Since Ω is c-convex with respect to $\bar{\Omega}$, there is an n-dimensional cone $K(x_0)$ of directions that point inward to $[\Omega]_{T(x_0)}$ from the boundary point $-\bar{D}c(x_0,T(x_0))$. By [Kim and McCann 2010, Lemma 4.4], for any such direction v in this cone $K(x_0)$, any segment of the form $s\mapsto\Phi^{-1}(sv\oplus -Dc(x_0,T(x_0)))$ is a geodesic for h for small $s>0$. Thus plugging such a segment into the geodesic equations in Φ^{-1} coordinates yields for any fixed $\bar{i}$, at $(x_0,T(x_0))$,

$$0 = \widehat{\Gamma}^{\bar{i}}_{jk}v^j v^k.$$

Suppose $\{v_l\}_{l=1}^n$ is a linearly independent collection of vectors in $K(x_0)$; then for any $1 \le l_1 \ne l_2 \le n$ we have

$$0 = \widehat{\Gamma}^{\bar{i}}_{jk}(v^j_{l_1} + v^j_{l_2})(v^k_{l_1} + v^k_{l_2}) = \widehat{\Gamma}^{\bar{i}}_{jk} v^j_{l_1} v^k_{l_1} + \widehat{\Gamma}^{\bar{i}}_{jk} v^j_{l_2} v^k_{l_2} + \widehat{\Gamma}^{\bar{i}}_{jk} v^j_{l_1} v^k_{l_2} + \widehat{\Gamma}^{\bar{i}}_{jk} v^j_{l_2} v^k_{l_1} = 2\widehat{\Gamma}^{\bar{i}}_{jk} v^j_{l_1} v^k_{l_2},$$

which implies all Christoffel symbols of the form $\widehat{\Gamma}^{\bar{i}}_{jk}$ are also zero. A similar argument reversing the roles of Ω and $\overline{\Omega}$ yields that *all* Christoffel symbols of h are zero in the Φ^{-1} coordinates at the point $(x_0, T(x_0))$.

Now using (39), we see that the coordinates of the 1-form $(\beta \oplus DT(x_0)\beta)^\flat$ in Φ^{-1} are equal to the Euclidean coordinates of the tangent vector $\frac{1}{2}(\beta \oplus DT(x_0)\beta)$. Also we can calculate, for $i = 1$ or 2,

$$\begin{aligned}
(\widehat{\tau_i \oplus DT(x_0)\tau_i})^j &= \frac{\partial \Phi^j}{\partial X^{\bar{k}}}(x_0, T(x_0))(\tau_i \oplus DT(x_0)\tau_i)^{\bar{k}} = -c_{j\bar{k}}(DT(x_0)\tau_i)^k = \hat{\hat{\tau}}^j_i,\\
(\widehat{\tau_i \oplus DT(x_0)\tau_i})^{\bar{j}} &= \frac{\partial \Phi^{\bar{j}}}{\partial X^k}(x_0, T(x_0))(\tau_i \oplus DT(x_0)\tau_i)^k = -c_{k\bar{j}}\tau^k_i = \hat{\tau}^j_i,
\end{aligned}$$

where we have identified $T^*_{x_0}\Omega$ and $T^*_{T(x_0)}\overline{\Omega}$ with $\mathbb{R}^n$ to write the vectors $\hat{\tau}_i$ and $\hat{\hat{\tau}}_i$ defined in the statement of the theorem in Euclidean coordinates. Combining this fact with (39), we can write (38) in the coordinates given by Φ^{-1} as

$$-\frac{1}{2}|\beta|_w^{-1}\left(\hat{\hat{\tau}}^j_1 \sum_{i=1}^n \beta^i(\partial_{\hat{x}^j}\hat{\hat{\tau}}^i_2) + \hat{\tau}^l_1 \sum_{k=1}^n (DT(x_0)\beta)^k(\partial_{\hat{x}^l}\hat{\tau}^k_2)\right). \tag{40}$$

Now we can see that the function $h^*(Y(x_0, \cdot))$ is a defining function for the set $[\overline{\Omega}]_{x_0}$; hence identifying $T^*_{x_0}\Omega$ with $\mathbb{R}^n$ and differentiating yields that $\nabla_p h^*(Y(x_0, p))$ is in the outward normal direction for $p \in \partial[\overline{\Omega}]_{x_0}$. In particular, the unit outward normal vector to $\partial[\overline{\Omega}]_{x_0}$ at $-Dc(x_0, T(x_0))$ has coordinates given by $\beta^i/|\beta|$. A similar calculation involving $h(X(T(x_0), \cdot))$ yields that the coordinates of the unit outward normal vector to $\partial[\Omega]_{T(x_0)}$ at $-\bar{D}c(x_0, T(x_0))$ are given by $(DT(x_0)\beta)^k/|DT(x_0)\beta|$. Additionally, since each τ_i is tangent to $\partial\Omega$, we see that $\hat{\tau}_i$ and $\hat{\hat{\tau}}_i$ are respectively tangent to $\partial[\overline{\Omega}]_{x_0}$ and $\partial[\Omega]_{T(x_0)}$. Thus we calculate

$$\begin{aligned}
\mathrm{II}^{\partial[\overline{\Omega}]_{x_0}}_{-Dc(x_0,T(x_0))}(\hat{\hat{\tau}}_1, \hat{\hat{\tau}}_2) &= \left\langle \nabla_{\hat{\hat{\tau}}_1}\frac{\beta}{|\beta|}, \hat{\hat{\tau}}_2\right\rangle = |\beta|^{-1}\langle \nabla_{\hat{\hat{\tau}}_1}\beta, \hat{\hat{\tau}}_2\rangle + D_{\hat{\hat{\tau}}_1}\left(\frac{1}{|\beta|}\right)\langle \beta, \hat{\hat{\tau}}_2\rangle\\
&= |\beta|^{-1}\langle \nabla_{\hat{\hat{\tau}}_1}\beta, \hat{\hat{\tau}}_2\rangle = |\beta|^{-1}(D_{\hat{\hat{\tau}}_1}\langle \beta, \hat{\hat{\tau}}_2\rangle - \langle \beta, \nabla_{\hat{\hat{\tau}}_1}\hat{\hat{\tau}}_2\rangle) = -|\beta|^{-1}\langle \beta, \nabla_{\hat{\hat{\tau}}_1}\hat{\hat{\tau}}_2\rangle\\
&= -|\beta|^{-1}\hat{\hat{\tau}}^j_1 \sum_{i=1}^n \beta^i(\partial_{\hat{x}^j}\hat{\hat{\tau}}^i_2)
\end{aligned}$$

and likewise

$$\begin{aligned}
\mathrm{II}^{\partial[\Omega]_{T(x_0)}}_{-\bar{D}c(x_0,T(x_0))}(\hat{\tau}_1, \hat{\tau}_2) &= \left\langle \nabla_{\hat{\tau}_1}\frac{DT(x_0)\beta}{|DT(x_0)\beta|}, \hat{\tau}_2\right\rangle = -|DT(x_0)\beta|^{-1}\langle DT(x_0)\beta, \nabla_{\hat{\tau}_1}\hat{\tau}_2\rangle\\
&= -|DT(x_0)\beta|^{-1}\hat{\tau}^l_1 \sum_{k=1}^n (DT(x_0)\beta)^k(\partial_{\hat{x}^l}\hat{\tau}^k_2).
\end{aligned}$$

Comparing this with (40) completes the proof of the theorem. ∎

The relevance of the above theorem to our current exponential convergence result is as follows. In terms of the metric w, we see that the β-directional derivative of the first term in the function F defined by (19) is given by (at x_0)

$$\begin{aligned} D_\beta(w(\nabla^w f, \nabla^w f)) &= 2w(\nabla^w_\beta \nabla^w f, \nabla^w f) = \operatorname{Hess} f(\beta, \nabla^w f) \\ &= \operatorname{Hess} f(\nabla^w f, \beta) = 2w(\nabla^w_{\nabla^w f} \nabla^w f, \beta) = -2w(\nabla^w_{\nabla^w f}\beta, \nabla^w f) \\ &= -2|\beta|_w \mathrm{II}^w(\nabla^w f, \nabla^w f). \end{aligned}$$

Here we repeatedly used that $\nabla^w f$ is tangent to $\partial\Omega$ (due to the boundary condition $D_\beta v = 0$ and since $f = \log v$), while β is normal in the metric w, and we have used (38) in the last line. Under the c- and c^*-convexity conditions (8) and (9), the two terms on the right-hand side of (37) are nonnegative; hence, by Theorem 5.1, $D_\beta w(\nabla^w f, \nabla^w f)$ is nonpositive. Thus in order to obtain a contradiction with the Hopf lemma as in Section 4, all that remains is to evaluate the last term $-\alpha D_\beta(\partial_t f)$. Obtaining a sign on this term depends on the specific choice of the function v, as in Section 4.

Acknowledgments

The authors would like to thank Micah Warren for bringing our attention to the special case $n = 2$, and for pointing out the reference [Warren 2014].

References

[Bailesteanu et al. 2010] M. Bailesteanu, X. Cao, and A. Pulemotov, "Gradient estimates for the heat equation under the Ricci flow", *J. Funct. Anal.* **258**:10 (2010), 3517–3542. MR Zbl

[Brenier 1991] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions", *Comm. Pure Appl. Math.* **44**:4 (1991), 375–417. MR Zbl

[Chow 1991] B. Chow, "On Harnack's inequality and entropy for the Gaussian curvature flow", *Comm. Pure Appl. Math.* **44**:4 (1991), 469–483. MR Zbl

[Gangbo and McCann 1996] W. Gangbo and R. J. McCann, "The geometry of optimal transportation", *Acta Math.* **177**:2 (1996), 113–161. MR Zbl

[Kim and McCann 2010] Y.-H. Kim and R. J. McCann, "Continuity, curvature, and the general covariance of optimal transportation", *J. Eur. Math. Soc.* (*JEMS*) **12**:4 (2010), 1009–1040. MR Zbl

[Kim et al. 2010] Y.-H. Kim, R. J. McCann, and M. Warren, "Pseudo-Riemannian geometry calibrates optimal transportation", *Math. Res. Lett.* **17**:6 (2010), 1183–1197. MR Zbl

[Kim et al. 2012] Y.-H. Kim, J. Streets, and M. Warren, "Parabolic optimal transport equations on manifolds", *Int. Math. Res. Not.* **2012**:19 (2012), 4325–4350. MR Zbl

[Kitagawa 2012] J. Kitagawa, "A parabolic flow toward solutions of the optimal transportation problem on domains with boundary", *J. Reine Angew. Math.* **672** (2012), 127–160. MR Zbl

[Li and Yau 1986] P. Li and S.-T. Yau, "On the parabolic kernel of the Schrödinger operator", *Acta Math.* **156**:3-4 (1986), 153–201. MR Zbl

[Lieberman 1996] G. M. Lieberman, *Second order parabolic differential equations*, World Scientific, River Edge, NJ, 1996. MR Zbl

[Loeper 2009] G. Loeper, "On the regularity of solutions of optimal transportation problems", *Acta Math.* **202**:2 (2009), 241–283. MR Zbl

[Ma et al. 2005] X.-N. Ma, N. S. Trudinger, and X.-J. Wang, "Regularity of potential functions of the optimal transportation problem", *Arch. Ration. Mech. Anal.* **177**:2 (2005), 151–183. MR Zbl

[Schnürer and Smoczyk 2003] O. C. Schnürer and K. Smoczyk, "Neumann and second boundary value problems for Hessian and Gauss curvature flows", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **20**:6 (2003), 1043–1073. MR Zbl

[Trudinger and Wang 2009] N. S. Trudinger and X.-J. Wang, "On the second boundary value problem for Monge–Ampère type equations and optimal transportation", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **8**:1 (2009), 143–174. MR Zbl

[Urbas 1997] J. Urbas, "On the second boundary value problem for equations of Monge–Ampère type", *J. Reine Angew. Math.* **487** (1997), 115–124. MR Zbl

[Warren 2014] M. Warren, "On solutions to Cournot–Nash equilibria equations on the sphere", *Pacific J. Math.* **272**:2 (2014), 423–437. MR Zbl

Received 2 Jan 2019. Revised 21 Jun 2019. Accepted 6 Sep 2019.

FARHAN ABEDIN: abedinf1@msu.edu
Department of Mathematics, Michigan State University, East Lansing, MI, United States

JUN KITAGAWA: kitagawa@math.msu.edu
Department of Mathematics, Michigan State University, East Lansing, MI, United States

ANALYSIS AND PDE
Vol. 13, No. 7, 2020
https://doi.org/10.2140/apde.2020.13.2205

NUCLEAR DIMENSION OF SIMPLE STABLY PROJECTIONLESS C*-ALGEBRAS

JORGE CASTILLEJOS AND SAMUEL EVINGTON

We prove that $\mathcal{Z}$-stable, simple, separable, nuclear, nonunital C*-algebras have nuclear dimension at most 1. This completes the equivalence between finite nuclear dimension and $\mathcal{Z}$-stability for simple, separable, nuclear, nonelementary C*-algebras.

Introduction

The Elliott classification programme, a 40-year endeavour involving generations of researchers, asks the following question: when are K-theory and traces a complete invariant for simple, separable, nuclear C*-algebras?

Fundamentally, there are two cases to consider: the *unital* case and the *stably projectionless* case. (This dichotomy follows from Brown's theorem [1977] and is discussed further below.) Recall that a C*-algebra A is said to be stably projectionless if there are no nonzero projections in the matrix amplification $M_n(A)$ for any $n \in \mathbb{N}$. Stably projectionless, simple, separable, nuclear C*-algebras arise naturally as crossed products [Kishimoto and Kumjian 1996] and can also be constructed using inductive limits with a wide variety of K-theoretic and tracial invariants occurring [Blackadar 1980; Razak 2002; Tsang 2005; Jacelon 2013; Gong and Lin 2016; 2017; Elliott et al. 2017; 2020].

In the unital case, a definitive answer for when K-theory and traces form a complete invariant is now known [Kirchberg 1995; Phillips 2000; Gong et al. 2015; Elliott et al. 2015; Tikuisis et al. 2017; Winter 2014]. Firstly, Rosenberg and Schochet's universal coefficient theorem [1987] must hold for the C*-algebras concerned. Secondly, the C*-algebras must have *finite nuclear dimension* [Winter and Zacharias 2010]. This second condition has a geometric flavour and generalises the notation of finite covering dimension for topological spaces. Recent results [Gong and Lin 2016; 2017; Elliott et al. 2017; 2020] are now converging on a similar classification result in the stably projectionless case; the most recent developments will be discussed below.

A major programme of research now focuses on providing methods for verifying finite nuclear dimension in practice. In the unital setting, a recent result of the authors together with Tikuisis, White and Winter [Castillejos et al. 2019] shows that finite nuclear dimension can be accessed through the tensorial absorption condition known as $\mathcal{Z}$*-stability*, where $\mathcal{Z}$ is the *Jiang–Su algebra* (discussed in more detail below).

Evington is supported by NCN (2014/14/E/ST1/00525) and EPSRC (EP/R025061/1); Castillejos is partially supported by European Research Council Consolidator Grant 614195 RIGIDITY, and by a long term structural funding — a Methusalem grant of the Flemish Government.
MSC2010: 46L05, 46L35.
Keywords: C*-algebras, classification of C*-algebras.

In concrete examples, it can be very hard to prove directly that a C*-algebra has finite nuclear dimension. The strategy of verifying $\mathcal{Z}$-stability instead has recently been used to prove that certain unital, simple, separable, nuclear C*-algebras coming from dynamical systems are classifiable [Conley et al. 2018; Kerr and Szabó 2020]. However, since this strategy relies on [Castillejos et al. 2019], it has until now only been available in the unital setting.

In this paper, we consider and overcome the conceptual and technical challenges unique to the nonunital setting, allowing us prove the following:

Theorem A. *Let A be a simple, separable, nuclear, $\mathcal{Z}$-stable* C*-*algebra. Then A has nuclear dimension at most* 1.

For the following reasons, the nonunital case is harder than the unital case and needs new methods. Obviously, we cannot just unitise our C*-algebras because this breaks both simplicity and $\mathcal{Z}$-stability. A more fundamental issue is that nonunital, simple C*-algebras need not actually be *algebraically simple*. There can be nontrivial (nonclosed) ideals. Examples of such ideals are the domains of *unbounded traces*, which may now exist and must therefore be taken into account. Furthermore, [Castillejos et al. 2019] builds on the foundations of [Bosa et al. 2019a], which has a global assumption of unitality and makes explicit use of the unit at a number of critical points in the argument (an example is the 2×2 matrix trick of [Bosa et al. 2019a, Section 2], which is inspired by ideas of Connes).

To understand how we circumvent the issues associated to unbounded traces, it will be helpful to first discuss the folklore result, alluded to above, that Brown's theorem [1977] implies a dichotomy for simple C*-algebras between the unital and the stably projectionless cases.

Writing $\mathbb{K}$ for the C*-algebra of compact operators (on a separable, infinite-dimensional Hilbert space), recall that C*-algebras A, B are *stably isomorphic* if $A \otimes \mathbb{K} \cong B \otimes \mathbb{K}$. Suppose now that A is a simple, separable C*-algebra that is not stably projectionless. Then there exists a nonzero projection $p \in A \otimes \mathbb{K}$, and so the hereditary subalgebra $p(A \otimes \mathbb{K})p$ is unital. By [Brown 1977, Theorem 2.8], $p(A \otimes \mathbb{K})p$ is stably isomorphic to $A \otimes \mathbb{K}$, and hence stably isomorphic to A (see Section 2 for more details).

Crucial to proving Theorem A in general is the observation that the hypotheses and the conclusion depend only on the stable isomorphism class of A.[1] Hence, by [Castillejos et al. 2019, Theorem B] and the folklore result above based on Brown's theorem, it suffices to prove Theorem A in the stably projectionless case.

However, this folklore reduction is not enough for us. We go a step further and pass to a hereditary subalgebra $A_0 \subseteq A \otimes \mathbb{K}$ on which all tracial functionals are bounded and the set of tracial states $T(A_0)$ is weak* compact. The existence of such hereditary subalgebra follows from the Cuntz semigroup computation of [Elliott et al. 2011] for $\mathcal{Z}$-stable C*-algebras, and Brown's theorem assures us that A_0 is stably isomorphic to A. This second reduction puts us in a position where a similar proof strategy to that of [Bosa et al. 2019a] can be implemented, and where the key new ingredient from [Castillejos et al. 2019], *complemented partitions of unity* (CPoU), is also available.

[1] That is, A is a simple, separable, nuclear and $\mathcal{Z}$-stable C*-algebra if and only if $A \otimes \mathbb{K}$ is likewise, and $\dim_{\text{nuc}} A \leq 1$ if and only if $\dim_{\text{nuc}}(A \otimes \mathbb{K}) \leq 1$; see Proposition 2.3 for details and references.

Of course, we still have to deal with the global assumption of unitality in [Bosa et al. 2019a]. A key tool in this endeavour is our unitisation lemma for order-zero maps into ultrapowers (Lemma 4.2), which allows us to assume the domains of certain maps are unital in places where simplicity and $\mathcal{Z}$-stability are only really needed on the codomain side in [Bosa et al. 2019a].

We now turn to the broader context of Theorem A and its applications. As alluded to above, *nuclear dimension* for C*-algebras is a noncommutative dimension theory that reduces to the covering dimension of the spectrum in the commutative case. *Finite nuclear dimension* has proven to be a technically useful strengthening of nuclearity, that is both necessary for classification [Villadsen 1999; Rørdam 2003; Toms 2008; Giol and Kerr 2010] and a vital ingredient of the most recent classification theorems [Kirchberg 1995; Phillips 2000; Gong et al. 2015; Elliott et al. 2015; Tikuisis et al. 2017; Winter 2014].

The *Jiang–Su algebra* $\mathcal{Z}$ [1999] is a simple C*-algebra which plays a fundamental role in the classification of simple C*-algebras since A and $A \otimes \mathcal{Z}$ have the same K-theory and traces under mild hypotheses. A C*-algebra is said to be *$\mathcal{Z}$-stable* if $A \cong A \otimes \mathcal{Z}$. Moreover, any C*-algebra can be $\mathcal{Z}$-stabilised by tensoring with the Jiang–Su algebra because $\mathcal{Z} \cong \mathcal{Z} \otimes \mathcal{Z}$. In many ways, the Jiang–Su algebra is the C*-algebraic analogue of the hyperfinite II_1 factor $\mathcal{R}$ [Murray and von Neumann 1943], with $\mathcal{Z}$-stability analogous to the McDuff property [1970].

Combining Theorem A with the main results of [Winter 2012; Tikuisis 2014], we arrive at the following relationship between finite nuclear dimension and $\mathcal{Z}$-stability, which was conjectured in [Toms and Winter 2009].

Theorem B. *Let A be a nonelementary, simple, separable, nuclear* C*-*algebra. The following are equivalent*:

(i) *A has finite nuclear dimension.*

(ii) *A is $\mathcal{Z}$-stable.*

One striking consequence of Theorems A and B is that nuclear dimension can only attain three different values in the simple setting.

Corollary C. *The nuclear dimension of a simple* C*-*algebra is* 0, 1 *or* ∞.

This is in stark contrast to the commutative case, where all nonnegative integers can occur. Moreover, we remark that the C*-algebras of nuclear dimension zero are known to be precisely the approximately finite-dimensional C*-algebras [Winter and Zacharias 2010, Remark 2.2.(iii)].[2]

Whilst Corollary C is interesting in its own right, we believe the main applications of our results will be in classification of simple, stably projectionless C*-algebras. Theorem A opens up a new pathway to proving that concrete examples of stably projectionless, simple, separable, nuclear C*-algebras, such as C*-algebras coming from flows on C*-algebras or from actions of more general locally compact groups, have finite nuclear dimension: it now suffices to verify $\mathcal{Z}$-stability.

[2]In the nonseparable case, there are different and inequivalent definitions of *approximately finite-dimensional* [Farah and Katsura 2010]. The one required here is that any finite set is approximately contained in a finite-dimensional subalgebra; see for example [Castillejos 2017, Definition 2.2].

We end this introduction with a discussion of recent developments in the classification of simple, stably projectionless C*-algebras. As mentioned above, there has been impressive progress in recent years [Gong and Lin 2016; 2017; Elliott et al. 2020]. As in the unital case, the classification is via a functor constructed from the K-theory and the tracial data of the C*-algebra; this functor is called the *Elliott invariant* and is typically denoted $\mathrm{Ell}(\,\cdot\,)$; see [Gong and Lin 2017, Definition 2.9] for a precise definition.

By combining Theorem A with [Gong and Lin 2017, Theorem 1.2], one obtains a classification of simple, separable, nuclear C*-algebras in the UCT class that tensorially absorb the C*-algebra $\mathcal{Z}_0$ — a stably projectionless analogue of the Jiang–Su algebra introduced in [Gong and Lin 2017, Definition 8.1].

Corollary D. *Let A and B be simple, separable, nuclear* C*-*algebras which satisfy the UCT. Then*

$$A \otimes \mathcal{Z}_0 \cong B \otimes \mathcal{Z}_0 \quad \textit{if and only if} \quad \mathrm{Ell}(A \otimes \mathcal{Z}_0) \cong \mathrm{Ell}(B \otimes \mathcal{Z}_0).$$

Corollary D reduces to the celebrated Kirchberg–Phillips classification [Kirchberg 1995; Phillips 2000] in the traceless case and is otherwise a result about stably projectionless C*-algebras. For these C*-algebras, the difference between $\mathcal{Z}_0$-stability and $\mathcal{Z}$-stability, roughly speaking, comes down to how complex the interaction between the K-theory and traces is allowed to be; see [Gong and Lin 2017] for more details.

Structure of paper. Section 1 reviews the necessary preliminary material as appropriate to the nonunital setting. Section 2 is concerned with the invariance of C*-algebraic properties under stable isomorphism and the reduction argument outlined above. The next three sections generalise the necessary technical machinery from [Bosa et al. 2019a; Castillejos et al. 2019]. Section 3 concerns the existence of an order-zero embedding $\Phi : A \to A_\omega$ with appropriate finite-dimensional approximations. Section 4 contains the aforementioned unitisation lemma for order-zero maps into ultrapowers. Section 5 is devoted to a uniqueness theorem for maps into ultrapowers, which we shall use to compare (unitisations of) Φ and the canonical embedding $A \to A_\omega$. Theorem A and its corollaries are proved in Section 6, with analogous results for decomposition rank (a forerunner of nuclear dimension) proved in Section 7. Since some preliminary lemmas from [Bosa et al. 2019a] are stated only in the unital case, we include an Appendix with their nonunital versions.

1. Preliminaries

In this section, we recall the most important definitions and results that will be used in the sequel, and we introduce the notation used in this paper.

We write $\mathbb{K}$ to denote the C*-algebra of compact operators (on a separable, infinite-dimensional Hilbert space). Given a C*-algebra A, we write A_+ for the positive elements of A and $A_{+,1}$ for the positive contractions; we write $\mathrm{Ped}(A)$ for the Pedersen ideal of A, which is the minimal dense ideal of A (see [Pedersen 1979, Section 5.6]), and we write $A^\sim$ for the unitisation of A. Our convention is that, if A is already unital, then we adjoin a new unit, so $A^\sim \cong A \oplus \mathbb{C}$ as C*-algebras. For $S \subseteq A$ self-adjoint, we set $S^\perp := \{a \in A : ab = ba = 0 \text{ for all } b \in S\}$. For $\epsilon > 0$ and $a, b \in A$, the notation $a \approx_\epsilon b$ means $\|a - b\| < \epsilon$. For $a, b \in A$ with b self-adjoint, we write $a \lhd b$ to mean that $ab = ba = a$.

We use the common abbreviation c.p.c. for completely positive and contractive maps between C*-algebras. A c.p.c. map $\phi : A \to B$ is *order-zero* if it preserves orthogonality in the sense that, for $a, b \in A_+$, $\phi(a)\phi(b) = 0$ whenever $ab = 0$.

Following [Winter and Zacharias 2010, Definition 2.1], a C*-algebra A has *nuclear dimension at most* n if there is a net $(F_i, \psi_i, \phi_i)_{i \in I}$, where F_i is a finite-dimensional C*-algebra, $\psi_i : A \to F_i$ is a c.p.c. map, and $\phi_i : F_i \to A$ is a c.p. map, such that $\phi_i \circ \psi_i(a) \to a$ for all $a \in A$ and, moreover, each F_i decomposes into $n+1$ ideals $F_i = F_i^{(0)} \oplus \cdots \oplus F_i^{(n)}$ for which the restrictions $\phi_i|_{F_i^{(k)}}$ are c.p.c. order-zero. The *nuclear dimension* of A, denoted by $\dim_{\mathrm{nuc}} A$, is defined to be the smallest such n (and to be ∞, if no such n exists). The *decomposition rank*, a forerunner of nuclear dimension, is obtained if one additionally requires ϕ_i to be a c.p.c. map [Kirchberg and Winter 2004, Definition 3.1]. We shall denote the decomposition rank of a C*-algebra A by $\mathrm{dr}(A)$.

By a *trace* on a C*-algebra we will typically mean a tracial state, i.e., a positive linear functional $\tau : A \to \mathbb{C}$ of operator norm 1 such that $\tau(ab) = \tau(ba)$ for all $a, b \in A$. We write $T(A)$ for the set of tracial states on A endowed with the weak*-topology. More general notions of traces are discussed in Section 1A below. By a *cone* we will mean a convex subset C of a locally convex space that satisfies $C + C \subseteq C$, $\lambda C \subset C$ for $\lambda > 0$, and $C \cap (-C) = \{0\}$. A *base* for a cone C is a closed, convex, and bounded subset X such that for any nonzero $c \in C$ there exist unique $\lambda > 0$ and $x \in X$ such that $c = \lambda x$. By [Alfsen 1971, Theorem II.2.6], a cone is locally compact if and only if it has a compact base. A map $f : C \to D$ between cones is *linear* if $f(\lambda x + \mu y) = \lambda f(x) + \mu f(y)$ for $\lambda, \mu \geq 0$ and $x, y \in C$. If X is a compact base for the cone C, then any continuous affine map $X \to D$ extends uniquely to a continuous linear map $C \to D$.

1A. *Generalised traces.* In this preliminary section, we briefly discuss the generalisations of traces that arise in the general theory of C*-algebras.

Definition 1.1 (cf. [Blanchard and Kirchberg 2004, Definition 2.22]). A *quasitrace*[3] on a C*-algebra A is a function $\tau : A_+ \to [0, \infty]$ with $\tau(0) = 0$ such that

(i) $\tau(a^*a) = \tau(aa^*)$ for all $a \in A$,

(ii) $\tau(a + b) = \tau(a) + \tau(b)$ for all commuting elements $a, b \in A_+$,

(iii) τ extends to a function $\tau_2 : M_2(A)_+ \to [0, \infty]$ for which (i) and (ii) hold.

The quasitrace τ is *additive* if (ii) holds for all $a, b \in A_+$.[4] Setting $\mathrm{Dom}_{1/2}(\tau) := \{a \in A : \tau(a^*a) < \infty\}$, we say that τ is *densely defined* if $\mathrm{Dom}_{1/2}(\tau)$ is dense in A, and that τ is *bounded* if $\mathrm{Dom}_{1/2}(\tau) = A$.

We write $Q\widetilde{T}(A)$ for the cone of densely defined, lower-semicontinuous quasitraces; $\widetilde{T}(A)$ for the cone of densely defined, lower-semicontinuous, additive quasitraces; and $\widetilde{T}_b(A)$ for the cone of bounded, additive quasitraces. The topology on these cones is given by pointwise convergence on $\mathrm{Ped}(A)$.

[3] Strictly speaking, a *2-quasitrace*; however, we shall not need this terminology.

[4] We use the terminology *additive quasitrace* because we are reserving the word *trace* for tracial states. For additive quasitraces, condition (iii) is automatic with τ_2 given by the usual formula.

Since the traces on a C*-algebra will play a crucial role in the arguments of this paper, the following existence theorem of Blackadar–Cuntz is fundamental.

Theorem 1.2 [Blackadar and Cuntz 1982, Theorem 1.2]. *Let A be a simple* C*-*algebra such that $A \otimes \mathbb{K}$ contains no infinite projections. Then $Q\widetilde{T}(A) \neq 0$.*

It is an open question whether $Q\widetilde{T}(A) = \widetilde{T}(A)$ in general. However, when A is exact, this is a famous result of Haagerup; for the unital case see [Haagerup 2014] and for deducing the general case from that work see [Blanchard and Kirchberg 2004, Remark 2.29(i)].

Every $\tau \in Q\widetilde{T}(A)$ has a unique extension to a densely defined, lower-semicontinuous quasitrace on $A \otimes \mathbb{K}$ which is additive whenever τ is additive [Blanchard and Kirchberg 2004, Remark 2.27(viii)]. Therefore, we have canonical isomorphisms $Q\widetilde{T}(A) \cong Q\widetilde{T}(A \otimes \mathbb{K})$ and $\widetilde{T}(A) \cong \widetilde{T}(A \otimes \mathbb{K})$, which we treat as identifications. Furthermore, every $\tau \in \widetilde{T}_b(A)$ has a unique extension to a positive linear functional on A, which we also denote τ, satisfying the trace condition $\tau(ab) = \tau(ba)$ for all $a, b \in A$.

Let $a, b \in A_+$. If there exists a sequence $(x_n)_{n\in\mathbb{N}}$ in A such that $b = \sum_{n=1}^{\infty} x_n^* x_n$ and $\sum_{n=1}^{\infty} x_n x_n^* \leq a$, then b is said to be *Cuntz–Pedersen subequivalent* to a [1979]. Our notation for this subequivalence will be $b \preccurlyeq a$. The following proposition is proven by the same method as [Cuntz and Pedersen 1979, Proposition 4.7]. For the benefit of the reader, we give full details.

Proposition 1.3. *Let A be a simple, separable* C*-*algebra and $B \subseteq A$ a nonzero hereditary subalgebra. The restriction map $\rho : \widetilde{T}(A) \to \widetilde{T}(B)$ is a linear homeomorphism of cones.*

Proof. Since $\mathrm{Ped}(B) \subseteq \mathrm{Ped}(A)$, the restriction of a densely defined quasitrace on A is a densely defined quasitrace on B. Restriction also preserves additivity and lower-semicontinuity. Hence, ρ is well-defined. Continuity of ρ follows immediately from the fact that $\mathrm{Ped}(B) \subseteq \mathrm{Ped}(A)$, and it is clear that ρ is linear.

We now turn to proving that ρ is surjective. Let $\sigma \in \widetilde{T}(B)$. Define $\tau : A_+ \to [0, \infty]$ by $\tau(a) := \sup\{\sigma(b) : b \in B_+,\ b \preccurlyeq a\}$. The following properties of τ are easy to verify:

$$\tau(0) = 0, \tag{1-1}$$

$$\tau(a^*a) = \tau(aa^*), \qquad a \in A, \tag{1-2}$$

$$\tau(\lambda a) = \lambda\tau(a), \qquad \lambda \geq 0,\ a \in A_+, \tag{1-3}$$

$$\tau(a_1 + a_2) \geq \tau(a_1) + \tau(a_2), \qquad a_1, a_2 \in A_+. \tag{1-4}$$

Let $a_1, a_2 \in A_+$. Suppose $b \in B_+$ and $b \preccurlyeq a_1 + a_2$. By [Pedersen 1969, Corollary 1.2], there exist $b_1, b_2 \in A_+$ with $b = b_1 + b_2$ such that $b_1 \preccurlyeq a_1$ and $b_2 \preccurlyeq a_2$. Since B is a hereditary subalgebra, $b_1, b_2 \in B_+$. Hence,

$$\sigma(b) = \sigma(b_1) + \sigma(b_2) \leq \tau(a_1) + \tau(a_2). \tag{1-5}$$

Taking the supremum, we get $\tau(a_1 + a_2) \leq \tau(a_1) + \tau(a_2)$. Therefore, we have $\tau(a_1 + a_2) = \tau(a_1) + \tau(a_2)$. This completes the proof that τ is an additive quasitrace.

Since B is a hereditary subalgebra of A, the restriction of the Cuntz–Pedersen subequivalence relation on A to B is the same as the Cuntz–Pedersen subequivalence relation on B. It follows that $\tau|_{B_+}$ is σ. As σ is densely defined on B and A is simple, τ is densely defined.

Let $\tilde{\tau}(a) := \sup_{\epsilon>0} \tau((a-\epsilon)_+)$ be the *lower-semicontinuous regularisation* of τ; see [Blanchard and Kirchberg 2004, Remark 2.27(iv)] and [Elliott et al. 2011, Lemma 3.1]. Then $\tilde{\tau}$ is a densely defined, lower-semicontinuous, additive quasitrace on A, and we still have $\tilde{\tau}|_{B_+} = \sigma$ because σ is lower-semicontinuous. Therefore, ρ is surjective.

We now prove that ρ is injective. Let σ, τ, and $\tilde{\tau}$ be as above. Suppose $\psi \in \widetilde{T}(A)$ also satisfies $\psi|_B = \sigma$. Since $\psi(b) \leq \psi(a)$ whenever $b \preccurlyeq a$, we must have $\tau \leq \psi$. Since taking lower-semicontinuous regularisations is order-preserving, we have $\tilde{\tau} \leq \psi$. By [Elliott et al. 2011, Proposition 3.2], there exists $\phi \in \widetilde{T}(A)$ such that $\psi = \tilde{\tau} + \phi$. However, $\psi|_{B_+} = \tilde{\tau}|_{B_+} = \sigma$ and so ϕ vanishes on B_+. Since A is simple, it follows that $\phi = 0$ and so $\psi = \tilde{\tau}$. Therefore, ρ is injective.

Finally, we prove ρ that is a homeomorphism. Fix $b \in \mathrm{Ped}(B) \setminus \{0\}$. Note that b is also in $\mathrm{Ped}(A)$ and is full in both A and B by simplicity. Set $X_A := \{\tau \in \widetilde{T}(A) : \tau(b) = 1\}$ and $X_B := \{\tau \in \widetilde{T}(B) : \tau(b) = 1\}$. By [Tikuisis and Toms 2015, Proposition 3.4], X_A is a compact base for the cone $\widetilde{T}(A)$ and X_B is a compact base for the cone $\widetilde{T}(B)$. Since $b \in B$, we have that $\rho(X_A) = X_B$. Hence, ρ defines a continuous, affine bijection from X_A to X_B. Since X_A and X_B are compact Hausdorff spaces, ρ in fact defines an affine homeomorphism between compact bases for the cones $\widetilde{T}(A)$ and $\widetilde{T}(B)$. Therefore, ρ is a linear homeomorphism of the cones $\widetilde{T}(A)$ and $\widetilde{T}(B)$. □

1B. *Strict comparison.* We first recall the definition of *Cuntz subequivalence*. Let A be a C*-algebra and $a, b \in A_+$. Then $a \precsim b$ if and only if there exists a sequence $(x_n)_{n\in\mathbb{N}}$ in A such that

$$\lim_{n\to\infty} \|x_n^* b x_n - a\| = 0. \tag{1-6}$$

If $a \precsim b$ and $b \precsim a$, then a is said to be *Cuntz equivalent* to b. We shall write $[a]$ for the Cuntz equivalence class of a.

The *Cuntz semigroup* $\mathrm{Cu}(A)$ is the ordered abelian semigroup obtained by considering the Cuntz equivalence classes of positive elements in $A \otimes \mathbb{K}$ under orthogonal addition and the order induced by Cuntz subequivalence; see [Coward et al. 2008]. If one only considers the Cuntz equivalence classes of positive elements in $\bigcup_{k=1}^{\infty} M_k(A)$, then one obtains the *classical Cuntz semigroup* $W(A)$; see [Ara et al. 2011].

Informally, a C*-algebra A has *strict comparison* if traces determine the Cuntz comparison theory. In order to formalise this notion, we need to recall the *rank function* associated to a lower-semicontinuous quasitrace. Suppose $\tau : A_+ \to [0, \infty]$ is a lower-semicontinuous quasitrace. Then the rank function $d_\tau : (A \otimes \mathbb{K})_+ \to [0, \infty]$ is given by

$$d_\tau(a) = \lim_{n\to\infty} \tau(a^{1/n}), \tag{1-7}$$

where we have made use of the unique extension of τ to $A \otimes \mathbb{K}$. We have $d_\tau(a) \leq d_\tau(b)$ whenever $a, b \in (A \otimes \mathbb{K})_+$ satisfy $a \precsim b$ by [Blackadar and Handelman 1982, Theorem II.2.2]. Strict comparison can be viewed as a partial converse.

Since we will be adapting the methods of [Bosa et al. 2019a], we shall be working with the same definition of strict comparison that is used there.

Definition 1.4 [Bosa et al. 2019a, Definition 1.5]. A C*-algebra A has *strict comparison* (*of positive elements*, *with respect to bounded traces*) if

$$[\text{for all } \tau \in T(A),\quad d_\tau(a) < d_\tau(b)] \quad\Longrightarrow\quad a \precsim b \tag{1-8}$$

for $k \in \mathbb{N}$ and $a, b \in M_k(A)_+$.

We alert the reader to two facts about this definition. Firstly, it only concerns positive elements in $\bigcup_{k=1}^{\infty} M_k(A)$, so it is a property of the classical Cuntz semigroup $W(A)$. Secondly, we only require the condition $d_\tau(a) < d_\tau(b)$ to be shown when τ is a tracial state.

In light of the potential confusion that could arise from the variety of definitions of strict comparison that appear in the literature, we include a proof that $\mathcal{Z}$-stability implies strict comparison in the sense of Definition 1.4 for the benefit of the reader. The key ingredient in the proof is that $W(A)$ is almost unperforated whenever A is $\mathcal{Z}$-stable, which is due to [Rørdam 2004].

Proposition 1.5. *Let A be a simple, separable, $\mathcal{Z}$-stable* C**-algebra with $Q\widetilde{T}(A) = \widetilde{T}_b(A) \neq 0$. Then A has strict comparison of positive elements with respect to bounded traces.*

Proof. As A is $\mathcal{Z}$-stable, so is $A \otimes \mathbb{K}$. Hence, by [Rørdam 2004, Theorem 4.5], $\mathrm{Cu}(A) = W(A \otimes \mathbb{K})$ is almost unperforated. Applying [Elliott et al. 2011, Propositions 6.2 and 4.4], we find that A has strict comparison in the following sense: for all $a, b \in (A \otimes \mathbb{K})_+$ with $[a] \leq \infty[b]$ in $\mathrm{Cu}(A)$ if $d_\tau(a) < d_\tau(b)$ for all lower-semicontinuous quasitraces with $d_\tau(b) = 1$ then $[a] \leq [b]$ in $\mathrm{Cu}(A)$.

We show that under our hypothesis on A this implies that A has strict comparison in the sense of Definition 1.4. Consider $a, b \in M_k(A)_+$ and let $\epsilon > 0$ and $f_\epsilon : [0, 1] \to [0, 1]$ be the function that is 0 on $[0, \epsilon]$, affine on $[\epsilon, 2\epsilon]$ and 1 on $[2\epsilon, 1]$. Since $M_k(A)$ is simple, there exists $n \in \mathbb{N}$ such that $[f_\epsilon(a)] \leq n[b] \leq \infty[b]$ in $\mathrm{Cu}(A)$ by [Blackadar 2006, Corollary II.5.2.12]. As ϵ is arbitrary, we have $[a] \leq \infty[b]$.

Since $Q\widetilde{T}(A) = \widetilde{T}_b(A)$, if $d_\tau(a) < d_\tau(b)$ for all $\tau \in T(A)$, then $a \precsim b$ in $A \otimes \mathbb{K}$. As $M_k(A)$ is a hereditary subalgebra of $A \otimes \mathbb{K}$, we have $a \precsim b$ in $M_k(A)$ by [Kirchberg and Rørdam 2000, Lemma 2.2(iii)]. □

Remark 1.6. By replacing $M_k(A)$ with $A \otimes \mathbb{K}$ in the proof of Proposition 1.5, we see that (1-8) holds for all $a, b \in (A \otimes \mathbb{K})_+$. Therefore, A also has strict comparison by traces in the sense of [Ng and Robert 2016, Definition 3.1] under the hypotheses of Proposition 1.5.

1C. *Ultraproducts and Kirchberg's epsilon test.* Let ω be a free ultrafilter on $\mathbb{N}$, which we regard as fixed for the entirety of the paper. The *ultraproduct* $\prod_{n\to\omega} A_n$ of a sequence of C*-algebras $(A_n)_{n\in\mathbb{N}}$ is defined by

$$\prod_{n\to\omega} A_n := \frac{\prod_{n\in\mathbb{N}} A_n}{\bigl\{(a_n)_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} A_n : \lim_{n\to\omega} \|a_n\| = 0\bigr\}}. \tag{1-9}$$

The *ultrapower* A_ω of a C*-algebra A is the ultraproduct of the constant sequence $(A_n)_{n\in\mathbb{N}}$ with $A_n = A$ for all $n \in \mathbb{N}$. We identify A with the subalgebra of A_ω given by constant sequences $(a)_{n\in\mathbb{N}}$.

Every sequence $(\tau_n)_{n\in\mathbb{N}}$, where $\tau_n \in T(A_n)$, defines a tracial state on the ultrapower $\prod_{n\to\omega} A_n$ via $(a_n) \mapsto \lim_{n\to\omega} \tau_n(a_n)$. Tracial states of this form are known as *limit traces*. The set of all limit traces will be denoted by $T_\omega\big(\prod_{n\to\omega} A_n\big)$.

Not all traces on an ultraproduct are limit traces but we have the following density result due to [Ng and Robert 2016, Theorem 1.2] (generalising an earlier result of [Ozawa 2013, Theorem 8]).

Theorem 1.7 [Ng and Robert 2016; Ozawa 2013]. *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of simple, separable, $\mathcal{Z}$-stable* C*-*algebras with $Q\widetilde{T}(A_n) = \widetilde{T}_b(A_n)$ for all $n \in \mathbb{N}$. Then $T_\omega\big(\prod_{n\to\omega} A_n\big)$ is weak*-dense in $T\big(\prod_{n\to\omega} A_n\big)$.*

Proof. By Proposition 1.5 and Remark 1.6, each A_n has strict comparison by traces in the sense of [Ng and Robert 2016, Definition 3.1]. The result now follows by Theorem 1.2 of the same work. □

We shall also need uniform tracial ultraproducts. Recall that any trace $\tau \in T(A)$ defines a 2-seminorm $\|a\|_{2,\tau} := \tau(a^*a)^{1/2}$. The uniform 2-seminorm is then defined by

$$\|a\|_{2,T(A)} := \sup_{\tau\in T(A)} \|a\|_{2,\tau} = \sup_{\tau\in T(A)} \tau(a^*a)^{1/2}. \tag{1-10}$$

We can then define the *uniform tracial ultraproduct* of a sequence of C*-algebras $(A_n)_{n\in\mathbb{N}}$ by

$$\prod^{n\to\omega} A_n := \frac{\prod_{n\in\mathbb{N}} A_n}{\big\{(a_n)_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} A_n : \lim_{n\to\omega} \|a_n\|_{2,T(A_n)} = 0\big\}}. \tag{1-11}$$

The *uniform tracial ultrapower* A^ω of a C*-algebra A, which can be defined as the uniform tracial ultraproduct of the constant sequence $(A_n)_{n\in\mathbb{N}}$ with $A_n = A$ for all $n \in \mathbb{N}$, was introduced in [Castillejos et al. 2019]. We identify A with the subalgebra of A^ω given by constant sequences $(a)_{n\in\mathbb{N}}$.

Since $\|a\|_{2,T(A)} \le \|a\|$ for all $a \in A$, there exists a canonical surjection from the ultraproduct to the uniform tracial ultraproduct. The kernel of this *-homomorphism is the *trace kernel ideal* given by

$$\begin{aligned} J_{(A_n)} &:= \Big\{(a_n)_{n\in\mathbb{N}} \in \prod_{n\to\omega} A_n : \lim_{n\to\omega} \|a_n\|_{2,T(A_n)} = 0\Big\} \\ &= \Big\{x \in \prod_{n\to\omega} A_n : \|x\|_{2,\tau} = 0,\ \tau \in T_\omega\Big(\prod_{n\to\omega} A_n\Big)\Big\}. \end{aligned} \tag{1-12}$$

It follows that limit traces also induce traces on the uniform tracial ultraproduct. In the ultrapower case, we therefore use a unified notation $T_\omega(A)$ for the limit traces on A_ω or the induced traces on A^ω.

An important tool for working with ultrapowers are reindexing arguments, which allow one to find elements of the ultrapower exactly satisfying some given condition provided one can find elements of the ultrapower which approximately satisfy the condition for any given tolerance. A precise and very general formulation of such reindexing arguments is *Kirchberg's epsilon test*, which we state below.

Lemma 1.8 (Kirchberg's epsilon test [2006, Lemma A.1]). *Let $X_1, X_2, \dots$ be a sequence of nonempty sets, and for each $k, n \in \mathbb{N}$, let $f_n^{(k)} : X_n \to [0,\infty)$ be a function. Define $f_\omega^{(k)} : \prod_{n=1}^\infty X_n \to [0,\infty]$ by $f_\omega^{(k)}((s_n)_{n=1}^\infty) := \lim_{n\to\omega} f_n^{(k)}(s_n)$ for $(s_n) \in \prod_{n=1}^\infty X_n$. Suppose that for all $m \in \mathbb{N}$ and $\epsilon > 0$, there exists $(s_n)_{n=1}^\infty \in \prod_{n=1}^\infty X_n$ with $f_\omega^{(k)}((s_n)) < \epsilon$ for $k = 1, \dots, m$. Then there exists $(t_n)_{n=1}^\infty \in \prod_{n=1}^\infty X_n$ such that $f_\omega^{(k)}((t_n)) = 0$ for all $k \in \mathbb{N}$.*

1D. ***Stable rank one.*** A unital C*-algebra A is said to have *stable rank one* if the invertible elements form a dense subset. In this paper, we shall make use of the following nonunital generalisation.

Definition 1.9. Let A be a C*-algebra. We say that A has *stable rank one in* $A^{\sim}$ if every element of A is a limit of invertible elements in $A^{\sim}$.

In the unital case, $A^{\sim} \cong A \oplus \mathbb{C}$, so A has stable rank one in $A^{\sim}$ if and only if A has stable rank one. In the nonunital case, A having stable rank one in $A^{\sim}$ is weaker than requiring that $A^{\sim}$ itself have stable rank one; see [Robert 2016, Example 3.4].

A related notion is Robert's *almost stable rank one* [2016, Definition 3.1], which requires that, for all hereditary subalgebras $B \subseteq A$, B has stable rank one in $B^{\sim}$. Robert proved the following.

Theorem 1.10 [Robert 2016, Corollary 3.2]. *Let A be a $\mathcal{Z}$-stable, projectionless* C*-*algebra. Then A has almost stable rank one. In particular, A has stable rank one in $A^{\sim}$.*

We now prove that having stable rank one in the unitisation passes to ultraproducts. We employ the notation $[(a_n)]$ for the element of the ultraproduct defined by the bounded sequence (a_n). First, let us record that taking unitisations commutes with taking the ultraproduct. The proof of this lemma is straightforward and we omit it.

Lemma 1.11. *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of* C*-*algebras. The canonical inclusion $\prod_{n\to\omega} A_n \to \prod_{n\to\omega} A_n^{\sim}$ extends to an isomorphism*

$$\Big(\prod_{n\to\omega} A_n\Big)^{\sim} \cong \prod_{n\to\omega} A_n^{\sim}. \tag{1-13}$$

We now proceed to show that having stable rank one in the unitisation passes to ultraproducts.

Proposition 1.12. *Let (A_n) be a sequence of* C*-*algebras. Suppose for each $n \in \mathbb{N}$, A_n has stable rank one in $A_n^{\sim}$. Then $A_\omega := \prod_{n\to\omega} A_n$ has stable rank one in $A_\omega^{\sim}$.*

Proof. Let $x \in A_\omega$ and say $x = [(a_n)]$. By Theorem 1.10 and [Bosa et al. 2019a, Lemma 1.20], for each $n \in \mathbb{N}$ there is a unitary $u_n \in A_n^{\sim}$ such that $a_n \approx_{1/n} u_n|a_n|$. We then have $x = [(u_n)]|x| \in \prod_{n\to\omega} A_n^{\sim}$. By [Bosa et al. 2019a, Lemma 1.20] once more, x is a norm limit of invertible elements in $\prod_{n\to\omega} A_n^{\sim}$. By Lemma 1.11, $\prod_{n\to\omega} A_n^{\sim}$ is just $A_\omega^{\sim}$. □

1E. ***Complemented partitions of unity.*** The key technical tool in [Castillejos et al. 2019] was the complemented partitions of unity technique which enabled Theorem A to be proven in the unital case. This property is best formulated in terms of the tracial ultrapower A^ω of a separable C*-algebra with $T(A)$ nonempty and compact. These assumptions imply that A^ω is unital, with any sequential approximate identity representing the unit [loc. cit., Proposition 1.11]. We refer to [loc. cit., Definition G] of the same work for a detailed explanation of the ideas behind this definition.

Definition 1.13. Let A be a separable C*-algebra with $Q\widetilde{T}(A) = \widetilde{T}_b(A) \neq 0$ and $T(A)$ compact. We say that A has *complemented partitions of unity* (CPoU) if for every $\|\cdot\|_{2,T_\omega(A)}$-separable subset S of A^ω, every family $a_1, \dots, a_k \in (A^\omega)_+$, and any scalar

$$\delta > \sup_{\tau \in T_\omega(A)} \min_{i=1,\dots,k} \tau(a_i), \tag{1-14}$$

there exist orthogonal projections $p_1, \dots, p_k \in A^\omega \cap S'$ such that

$$p_1 + \cdots + p_k = 1_{A^\omega} \quad \text{and} \quad \tau(a_i p_i) \le \delta \tau(p_i), \quad \tau \in T_\omega(A),\ i = 1, \dots, k. \tag{1-15}$$

The following theorem gives sufficient conditions for a C*-algebra to have complemented partitions of unity. Although not necessary for our purposes, the hypothesis of $\mathcal{Z}$-stability can be weakened to *uniform property* Γ; see [Castillejos et al. 2019, Section 2] for more details.

Theorem 1.14 [Castillejos et al. 2019, Theorem I]. *Let A be a separable, nuclear, $\mathcal{Z}$-stable* C*-*algebra with $Q\widetilde{T}(A) = \widetilde{T}_b(A) \neq 0$ and $T(A)$ compact. Then A has complemented partitions of unity.*

2. Reductions

In this section, we show how Brown's theorem [1977, Theorem 2.8] can be used to reduce the task of proving Theorem A in general to proving it for unital C*-algebras and for stably projectionless C*-algebras with a compact trace space. We begin with the general statement of Brown's theorem.

Theorem 2.1 [Brown 1977, Theorem 2.8]. *Let B be a full hereditary subalgebra of a* C*-*algebra A. Suppose both A and B are σ-unital. Then B is stably isomorphic to A.*

In our applications, we shall be working with C*-algebras that are simple and separable. Hence, the fullness and σ-unitality conditions will be satisfied. We shall therefore use the following form of Brown's theorem.

Corollary 2.2. *Let B be a nonzero hereditary subalgebra of a simple, separable* C*-*algebra A. Then B is stably isomorphic to A.*

The utility of Brown's theorem for this paper derives from the fact that the hypotheses and conclusion of Theorem A are invariant under stable isomorphism. We state this formally below.

Proposition 2.3 [Kirchberg and Winter 2004; Winter and Zacharias 2010; Toms and Winter 2007]. *Let A be a* C*-*algebra. Then*:

(i) *A is simple if and only if $A \otimes \mathbb{K}$ is simple.*

(ii) *A is separable if and only if $A \otimes \mathbb{K}$ is separable.*

(iii) *A is nuclear if and only if $A \otimes \mathbb{K}$ is nuclear.*

(iv) $\mathrm{dr}(A) = \mathrm{dr}(A \otimes \mathbb{K})$.

(v) $\dim_{\mathrm{nuc}}(A) = \dim_{\mathrm{nuc}}(A \otimes \mathbb{K})$.

(vi) *A is separable and $\mathcal{Z}$-stable if and only if $A \otimes \mathbb{K}$ is separable and $\mathcal{Z}$-stable.*

Proof. Properties (i)–(iii) are well known; see for example [Blackadar 2006, Chapter IV.3]. Part (iv) is [Kirchberg and Winter 2004, Corollary 3.9]. Part (v) is [Winter and Zacharias 2010, Corollary 2.8]. Part (vi) is [Toms and Winter 2007, Corollary 3.2]. □

Next, we recall that a C*-algebra A is *stably projectionless* if there are no nonzero projections in $A \otimes \mathbb{K}$. By definition, this property is preserved under stable isomorphism. Stably projectionless C*-algebras can be viewed as highly nonunital C*-algebras. Indeed, the following folklore result establishes a dichotomy for simple, separable C*-algebras.

Proposition 2.4. *Let A be a nonzero, simple, separable* C*-*algebra. Then exactly one of the following holds*:

(a) *A is stably isomorphic to a unital* C*-*algebra.*

(b) *A is stably projectionless.*

Proof. Let A be a simple, separable C*-algebra A. Then $A \otimes \mathbb{K}$ is simple and separable by Proposition 2.3. Suppose that A is not stably projectionless. Then there exists a nonzero projection $p \in A \otimes K$. Set $B := p(A \otimes \mathbb{K})p$. Then B is a unital C*-algebra with unit $1_B = p$. Moreover, B is a nonzero hereditary subalgebra of $A \otimes \mathbb{K}$. Therefore, B is stably isomorphic to $A \otimes \mathbb{K}$ by Corollary 2.2, and hence is stably isomorphic to A.

Now suppose that A is stably isomorphic to a unital C*-algebra B. Then there exists an isomorphism $\phi : B \otimes \mathbb{K} \to A \otimes \mathbb{K}$. Writing 1_B for the unit of B and e_{ij} for the matrix units of $\mathbb{K}$, we have that $\phi(1_B \otimes e_{ii})$ is a nonzero projection in $A \otimes \mathbb{K}$. Hence, A cannot be stably projectionless. □

This dichotomy justifies the terminology *stably unital* for the nonstably projectionless, simple, separable C*-algebras. The stably unital case of Theorem A follows immediately from [Castillejos et al. 2019, Theorem B] together with Propositions 2.4 and 2.3. The stably projectionless case on the other hand requires a further reduction and a technical modification of the methods of [Bosa et al. 2019a]. The purpose of the additional reduction is to pass to the case where the trace space is compact, and it is based on the following folklore result.

Lemma 2.5. *Let A be a simple, separable* C*-*algebra with $\widetilde{T}(A) \neq 0$. Let $A_0 := \overline{a(A \otimes \mathbb{K})a}$ be the hereditary subalgebra generated by a nonzero positive contraction $a \in (A \otimes \mathbb{K})_{+,1}$ for which the function $\tau \mapsto d_\tau(a)$ is continuous and finite-valued. Then $\widetilde{T}(A_0) = \widetilde{T}_b(A_0) \neq 0$ and $T(A_0)$ is compact.*

Proof. By Proposition 1.3, the restriction map $\rho : \widetilde{T}(A \otimes \mathbb{K}) \to \widetilde{T}(A_0)$ is a linear homeomorphism. Let $\sigma \in \widetilde{T}(A_0)$. Then σ has an extension $\tau := \rho^{-1}(\sigma) \in \widetilde{T}(A \otimes \mathbb{K})$. By [Tikuisis 2014, Proposition 2.4], we have $\|\sigma\|_{A_0^*} = d_\tau(a) < \infty$. Therefore, $\widetilde{T}(A_0) = \widetilde{T}_b(A_0) \neq 0$. Since $\sigma \mapsto d_{\rho^{-1}(\sigma)}(a)$ is continuous, $T(A_0)$ is a weak*-closed subspace of the unit ball of A_0^*. Therefore, $T(A_0)$ is compact. □

We now explain how the results of [Elliott et al. 2011] can be used to prove the existence of positive contractions with continuous-rank functions under suitable hypotheses.

Proposition 2.6. *Let A be a simple, separable, $\mathcal{Z}$-stable* C*-*algebra with $Q\widetilde{T}(A) = \widetilde{T}(A) \neq 0$. Let $f : \widetilde{T}(A) \to [0, \infty)$ be a strictly positive, continuous, linear function. Then there exists a nonzero positive contraction $a \in (A \otimes \mathbb{K})_{+,1}$ with $d_\tau(a) = f(\tau)$ for all $\tau \in \widetilde{T}(A)$.*

Proof. Following [Elliott et al. 2011, Section 4.1], we write $F(\mathrm{Cu}(A))$ for the space of functionals on the Cuntz semigroup of A. In [loc. cit., Theorem 4.4], it is shown that all functionals on the Cuntz semigroup are of the form d_τ for some lower-semicontinuous quasitrace τ on A. However, we alert the reader to the fact that quasitraces are not assumed to be densely defined in [loc. cit.]. Since A is simple, this means that either $\tau \in Q\widetilde{T}(A)$ or τ is the *trivial quasitrace*, which satisfies $\tau(0) = 0$ and is infinite otherwise.

We now consider the topology on $F(\mathrm{Cu}(A))$, defined in general in [loc. cit., Section 4.1], and its relation to the topology on $\widetilde{T}(A)$, which is given by pointwise convergence on $\mathrm{Ped}(A)$. By [loc. cit., Theorem 4.4] and our assumption that all quasitraces are additive, the topology on $F(\mathrm{Cu}(A))$ agrees with the topology on set $\widehat{T}(A)$ of (not necessarily densely defined) lower-semicontinuous, additive quasitraces defined in [loc. cit., Section 3.2].[5] This topology is shown to be compact and Hausdorff in [loc. cit., Theorem 3.7]. By [loc. cit., Theorem 3.10], the restriction of this topology to $\widetilde{T}(A)$ is pointwise convergence on $\mathrm{Ped}(A)$. Since $\widehat{T}(A) \setminus \widetilde{T}(A)$ is just one point, it follows that the topology on $\widehat{T}(A)$ is simply the one point compactification of the topology on $\widetilde{T}(A)$.

By [Tikuisis and Toms 2015, Proposition 3.4], the cone $\widetilde{T}(A)$ has a compact base K. Since f is strictly positive and continuous, $\inf_{\tau \in K} f(\tau) > 0$. Hence, we may extend f to the one-point compactification of $\widetilde{T}(A)$ by setting $f(\infty) = \infty$ and the resulting map is still continuous. It follows that f defines an element of the dual cone $L(F(\mathrm{Cu}(A)))$; see [Elliott et al. 2011, Section 5.1]. Therefore, as A is $\mathcal{Z}$-stable, there exists $a \in (A \otimes \mathbb{K})_{+,1}$ such that $f(\tau) = d_\tau(a)$ for all $\tau \in \widetilde{T}(A)$ by [loc. cit., Theorem 6.6]. □

We end this section with the following summary of all the reductions.

Theorem 2.7. *Let A be a nonzero, simple, separable, exact, $\mathcal{Z}$-stable* C*-*algebra. Then one of the following holds*:

(a) *A is stably isomorphic to a unital* C*-*algebra.*

(b) *A is stably isomorphic to a stably projectionless* C*-*algebra A_0 with $Q\widetilde{T}(A_0) = \widetilde{T}_b(A_0) \neq 0$ and $T(A_0)$ is compact.*

Proof. Suppose (a) does not hold. Then A is stably projectionless by Proposition 2.4. By Theorem 1.2, $Q\widetilde{T}(A) \neq 0$. Since A is exact, the nonunital version of Haagerup's theorem gives $Q\widetilde{T}(A) = \widetilde{T}(A) \neq 0$; see [Haagerup 2014; Blanchard and Kirchberg 2004, Remark 2.29(i)]. Since $\widetilde{T}(A)$ is a cone with a compact base, there exists a strictly positive, continuous, linear function $f : \widetilde{T}(A) \to [0, \infty)$. By Proposition 2.6, there is a positive contraction $a \in (A \otimes \mathbb{K})_{+,1}$ such that $f(\tau) = d_\tau(a)$ for all $\tau \in \widetilde{T}(A)$. Set $A_0 := \overline{a(A \otimes \mathbb{K})a}$. By Lemma 2.5, $Q\widetilde{T}(A_0) = \widetilde{T}_b(A_0) \neq 0$ and $T(A_0)$ compact. By Corollary 2.2, A is stably isomorphic to A_0. Hence, A_0 is stably projectionless. □

3. Existence

Let A be a separable, nuclear C*-algebra with complemented partitions of unity (CPoU). In this section, we will construct a sequence of maps $A \xrightarrow{\theta_n} F_n \xrightarrow{\eta_n} A$, where F_n are finite-dimensional C*-algebras,

[5] In [Elliott et al. 2011], the notation $T(A)$ is used instead of $\widehat{T}(A)$, but this clashes with the notation for the tracial states used in this paper.

θ_n are c.p.c. maps and η_n are c.p.c. order-zero maps, which induces a *-homomorphism $A \to A^\omega$ that agrees with the diagonal inclusion $A \to A^\omega$.

We will do this in two steps. First, we will fix a trace τ and produce maps $A \to F \to A$ that approximate the identity map on A in $\|\cdot\|_{2,\tau}$-norm. We shall then construct the required sequence of maps using complemented partitions of unity (CPoU).

The following lemma will be deduced from [Castillejos et al. 2019, Lemma 5.1], but it can also be proved by directly applying the methods of [Brown et al. 2017, Lemma 2.5].

Lemma 3.1. *Let A be a separable, nuclear C^*-algebra and let $\tau \in T(A)$. For any finite subset $\mathcal{F} \subseteq A$ and $\epsilon > 0$ there exist a finite-dimensional C^*-algebra F, a c.p.c. map $\theta : A \to F$, and a c.p.c. order-zero map $\eta : F \to A$ such that*

$$\|\theta(a)\theta(b)\| < \epsilon \quad \text{for } a, b \in \mathcal{F} \text{ such that } ab = 0, \tag{3-1}$$

$$\|\eta \circ \theta(a) - a\|_{2,\tau} < \epsilon \quad \text{for } a \in \mathcal{F}. \tag{3-2}$$

If all traces are quasidiagonal,[6] *one can additionally arrange that*

$$\|\theta(ab) - \theta(a)\theta(b)\| < \epsilon, \quad a, b \in \mathcal{F}. \tag{3-3}$$

Proof. The trace τ extends to a trace on $A^\sim$. By [Castillejos et al. 2019, Lemma 5.1] applied to $A^\sim$, there exist a finite-dimensional F, a c.p.c. map $\tilde{\theta} : A^\sim \to F$, and a c.p.c. order-zero map $\tilde{\eta} : F \to A^\sim$ such that

$$\|\tilde{\theta}(a)\tilde{\theta}(b)\| < \tfrac{1}{2}\epsilon \quad \text{for } a, b \in \mathcal{F} \text{ satisfying } ab = 0, \tag{3-4}$$

$$\|\tilde{\eta} \circ \tilde{\theta}(a) - a\|_{2,\tau} < \tfrac{1}{2}\epsilon \quad \text{for } a \in \mathcal{F}. \tag{3-5}$$

Let $(e_n)_{n\in\mathbb{N}}$ be an approximate identity of A. Then $e_n \nearrow 1_{A^\sim}$ in $\|\cdot\|_{2,\tau}$. Hence, the c.p.c. maps $\hat{\eta}_n : F \to A$ given by $\hat{\eta}_n(x) = e_n\tilde{\eta}(x)e_n$ converge to $\tilde{\eta}$ in the point-$\|\cdot\|_{2,\tau}$ topology. The sequence $\hat{\eta}_n$ is asymptotically order-zero in $\|\cdot\|_{2,\tau}$. Since F is finite-dimensional, we can make use of order-zero lifting to obtain a sequence of c.p.c. order-zero maps $\eta_n : F \to A$ converging to $\tilde{\eta}$ in the point-$\|\cdot\|_{2,\tau}$ topology.[7]

Set $\theta := \tilde{\theta}|_A$. Choose $n \in \mathbb{N}$ such that $\|\eta_n(\theta(a)) - \tilde{\eta}(\theta(a))\|_{2,\tau} < \frac{1}{2}\epsilon$ for all $a \in \mathcal{F}$. We then have

$$\|\eta \circ \theta(a) - a\|_{2,\tau} < \|\tilde{\eta} \circ \theta(a) - a\|_{2,\tau} + \tfrac{1}{2}\epsilon < \epsilon, \quad a \in \mathcal{F}. \tag{3-6}$$

If all traces are quasidiagonal, the map $\tilde{\theta}$ given by [Castillejos et al. 2019, Lemma 5.1] is approximately a *-homomorphism. Hence, so is θ. □

With the previous lemma in hand, we can now utilise complemented partitions of unity (CPoU) to prove the following.

Lemma 3.2. *Let A be a separable, nuclear C^*-algebra with $Q\widetilde{T}(A) = \widetilde{T}_b(A) \neq 0$ and $T(A)$ compact. Suppose A has CPoU. Then there exists a sequence of c.p.c. maps $\phi_n : A \to A$ which factor through*

[6] See Definition 7.1.

[7] Indeed, let $J_\tau := \{(a_n)_{n\in\mathbb{N}} \in \ell^\infty(A) : \lim_{n\to\infty} \tau(a_n^* a_n) = 0\}$ and consider the diagonal map $(\hat{\eta}_n) : F \to \ell^\infty(A)/J_\tau$. This map is c.p.c. order-zero and so has a c.p.c. order-zero lift $(\eta_n) : F \to \ell^\infty(A)$ by [Winter 2009, Proposition 1.2.4].

finite-dimensional algebras F_n as

$$\begin{array}{ccc} A & \xrightarrow{\phi_n} & A \\ & {\scriptstyle\theta_n}\searrow \quad \nearrow{\scriptstyle\eta_n} & \\ & F_n & \end{array} \tag{3-7}$$

with θ_n c.p.c. and η_n c.p.c. order-zero, in such a way that the induced map $(\theta_n)_{n=1}^{\infty} : A \to \prod_{\omega} F_n$ is order-zero, and the induced map $\overline{\Phi} = (\phi_n)_{n=1}^{\infty} : A \to A^{\omega}$ agrees with the diagonal inclusion $A \to A^{\omega}$.

If all traces on A are quasidiagonal, then we may arrange that the induced map $(\theta_n)_{n=1}^{\infty} : A \to \prod_{\omega} F_n$ is a *-homomorphism.

Proof. As in [Castillejos et al. 2019, Lemma 5.2], by a standard application of Kirchberg's epsilon test, it suffices to show that for a finite set $\mathcal{F} \subseteq A$ and a tolerance $\epsilon > 0$, there is a sequence (F_n, θ_n, η_n) such that $\theta_n : A \to F_n$ is approximately order-zero (or approximately multiplicative if all traces are quasidiagonal), $\eta_n : F_n \to A$ is an order-zero map, and the induced map $\overline{\Phi}_\epsilon = (\eta_n \circ \theta_n)_{n=1}^{\infty} : A \to A^{\omega}$ satisfies $\|a - \overline{\Phi}_\epsilon(a)\|_{2,T_\omega(A)} < \epsilon$ for all $a \in \mathcal{F}$. In fact, we will arrange for all the F_n to be the same finite-dimensional algebra F, and all the θ_n to be the same map θ.

Let $\mathcal{F} \subseteq A$ be a finite subset and $\epsilon > 0$. By Lemma 3.1, for any $\tau \in T(A)$ there exist a finite-dimensional C*-algebra F_τ, a c.p.c. map $\theta_\tau : A \to F_\tau$, and an order-zero map $\eta_\tau : F_\tau \to A$ such that

$$\|\theta(a)\theta(b)\| < \epsilon, \qquad a, b \in \mathcal{F} \text{ such that } ab = 0, \tag{3-8}$$

$$\|\eta_\tau \circ \theta_\tau(x) - x\|_{2,\tau}^2 < \frac{\epsilon^2}{|\mathcal{F}|}, \quad x \in \mathcal{F}. \tag{3-9}$$

Set $a_\tau := \sum_{x \in \mathcal{F}} |x - \eta_\tau \tau \circ \theta_\tau(a)|^2$. By the compactness of $T(A)$, there exist $\tau_1, \dots, \tau_k \in T(A)$ such that for all $\tau \in T(A)$ there is some τ_i such that $\tau(a_{\tau_i}) < \epsilon^2$.

By CPoU, there exist pairwise orthogonal projections $p_1, \dots, p_k \in A^{\omega} \cap A'$ adding up to $1_{A^{\omega}}$ such that $\tau(a_i p_i) \le \epsilon^2 \tau(p_i)$ for all $\tau \in T_\omega(A)$. Set $F := \bigoplus_{i=1}^{k} F_{\tau_i}$, $\theta : A \to F$ and $\eta : F \to A^{\omega}$ by

$$\theta(a) := (\theta_{\tau_1}(a), \dots, \theta_{\tau_k}(a)), \quad \eta(x_1, \dots, x_k) := \sum_{i=1}^{k} \eta_{\tau_i}(x_i) p_i, \tag{3-10}$$

where $a \in A$ and $x_i \in F_{\tau_i}$. By construction (see [Castillejos et al. 2019, Lemma 5.2, equation (5.16)]), we obtain

$$\|a - \eta \circ \theta(a)\|_{2,T_\omega(A)} < \epsilon, \quad a \in \mathcal{F}. \tag{3-11}$$

By [Winter 2009, Proposition 1.2.4], $\eta : F \to A^{\omega}$ can be lifted to a sequence of order-zero maps $\eta_n : F \to A$. Thus the sequence (F, θ, η_n) is the required sequence.

Finally, if all traces are quasidiagonal, the map θ is approximately multiplicative by Lemma 3.1. Combining the previous argument with Kirchberg's epsilon test yields that the induced map $(\theta_n) : A \to \prod_{\omega} F_n$ is a *-homomorphism. □

4. Unitisation

In this section, we prove that a c.p.c. order-zero map $\phi : A \to B_\omega$ from a separable C*-algebra into a C*-ultrapower extends to a c.p.c. order-zero map $\phi^\sim : A^\sim \to B_\omega$. Moreover, under appropriate conditions, Dini's theorem can be used to construct an extension for which the tracial behaviour of $\phi^\sim(1_{A^\sim})$ is determined by ϕ. These results were inspired by the structure theory for order-zero maps developed in [Winter and Zacharias 2009] and the existence of *supporting order-zero maps* proved in [Bosa et al. 2019a, Lemma 1.14]. We begin with a technical lemma.

Lemma 4.1. *Let $\phi : A \to B$ be a c.p.c. order-zero map between C*-algebras. Suppose that $h \in B$ is a positive contraction such that*

$$\phi(a)\phi(b) = h\phi(ab), \quad a, b \in A_+. \tag{4-1}$$

Then the map $\phi^\sim : A^\sim \to B$ defined by $\phi^\sim(a + \lambda 1_{A^\sim}) := \phi(a) + \lambda h$ is c.p.c. order-zero.

Proof. By [Winter and Zacharias 2009, Corollary 4.1], there exists a *-homomorphism $\pi : C_0(0,1] \otimes A \to B_\omega$ such that $\phi(a) = \pi(t \otimes a)$ for all $a \in A$, where t denotes the canonical generator of the cone. In terms of π, equation (4-1) gives $h\pi(t \otimes ab) = \pi(t^2 \otimes ab)$ for $a, b \in A_+$, from which we deduce that

$$h\pi(t \otimes a) = \pi(t^2 \otimes a), \quad a \in A, \tag{4-2}$$

since $(A_+)^2 = A_+$ and A_+ spans A. It then follows that $h^n\pi(t^m \otimes a) = \pi(t^{n+m} \otimes a)$ for $a \in A$ and for all $n, m \in \mathbb{N}_{\geq 1}$, from which we obtain

$$g(h)\pi(f \otimes a) = \pi(gf \otimes a), \quad a \in A,\ f, g \in C_0(0,1], \tag{4-3}$$

since $\operatorname{span}\{t^n : n \in \mathbb{N}_{\geq 1}\}$ is dense in $C_0(0,1]$. Taking adjoints, we also obtain $\pi(f \otimes a)g(h) = \pi(fg \otimes a)$ for all $a \in A$, $f, g \in C_0(0,1]$.

We now define a map $\pi^\sim : C_0(0,1] \odot A^\sim \to B$ from the algebraic tensor product by setting $\pi^\sim(f \otimes (a + \lambda 1_{A^\sim})) := \pi(f \otimes a) + \lambda f(h)$ on elementary tensors. A straightforward computation using (4-3) and its adjoint shows that $\pi^\sim$ is a *-homomorphism. Hence, $\pi^\sim$ extends to a map $C_0(0,1] \otimes A^\sim \to B$. Finally, define $\phi^\sim : A^\sim \to B$ by $\phi^\sim(x) := \pi^\sim(t \otimes x)$. Then $\phi^\sim$ is a c.p.c. order-zero map and $\phi^\sim(a + \lambda 1_{A^\sim}) = \phi(a) + \lambda h$ as required. □

We now prove the unitisation lemma for order-zero maps.

Lemma 4.2. *Let A, B be C*-algebras with A separable and let $\phi : A \to B_\omega$ be a c.p.c. order-zero map*:

(a) *There exists a c.p.c. order-zero map $\phi^\sim : A^\sim \to B_\omega$ which extends ϕ.*

(b) *Suppose now that $T(B)$ is compact and nonempty. Let $(e_n)_{n \in \mathbb{N}}$ be an approximate unit for A and suppose the function*

$$\theta : \overline{T_\omega(B)}^{w*} \to [0,1], \quad \tau \mapsto \lim_{n \to \infty} \tau(\phi(e_n)),$$

is continuous. Then there exists a c.p.c. order-zero map $\phi^\sim : A^\sim \to B_\omega$ which extends ϕ and satisfies $\tau(\phi^\sim(1_{A^\sim})) = \theta(\tau)$ for all $\tau \in \overline{T_\omega(B)}^{w*}$.

Proof. (a) Let $(e_n)_{n\in\mathbb{N}}$ be an approximate unit for A. By [Winter and Zacharias 2009, Corollary 4.1], there exists a *-homomorphism $\pi : C_0(0,1]\otimes A \to B_\omega$ such that $\phi(a)=\pi(t\otimes a)$ for all $a\in A$, where t denotes the canonical generator of the cone. For any $a,b\in A_+$, we have

$$\begin{aligned}\lim_{n\to\infty}\phi(e_n)\phi(ab) &= \lim_{n\to\infty}\pi(t^2\otimes e_n ab)\\ &= \pi(t^2\otimes ab)=\phi(a)\phi(b). \end{aligned} \tag{4-4}$$

We shall now prove the existence of a positive contraction $h\in B_\omega$ such that (4-1) holds for all $a,b\in A_+$ by an application of Kirchberg's epsilon test (Lemma 1.8).

Let $X_n := B_{+,1}$ for all $n\in\mathbb{N}$. Let $\phi_n : A_+\to B_+$ be a sequence of functions such that $(\phi_n(a))_{n\in\mathbb{N}}$ is a representative for $\phi(a)$ for all $a\in A_+$. Fix a dense sequence $(a_n)_{n\in\mathbb{N}}$ in A_+. Define $f_n^{(r,s)} : X_n\to[0,\infty]$ for $r,s\in\mathbb{N}$ by

$$f_n^{(r,s)}(x) := \|x\phi_n(a_ra_s)-\phi_n(a_r)\phi_n(a_s)\|. \tag{4-5}$$

Then define $f_\omega^{(r,s)} : \prod_{n=1}^\infty X_n\to[0,\infty]$ by $(x_n)_{n\in\mathbb{N}}\mapsto \lim_{n\to\omega} f_n^{(r,s)}(x_n)$.

Let $m\in\mathbb{N}$ and $\epsilon>0$. By (4-4), there is $k\in\mathbb{N}$ such that

$$\|\phi(e_k)\phi(a_ra_s)-\phi(a_r)\phi(a_s)\|<\epsilon, \quad 1\le r,s\le m. \tag{4-6}$$

Let $x=(x_n)_{n\in\mathbb{N}}$ be a sequence of positive contractions in B representing $\phi(e_k)$. Then $f_\omega^{(r,s)}(x)<\epsilon$ whenever $1\le r,s\le m$. By Kirchberg's epsilon test, there exists a sequence of positive contractions $y=(y_n)_{n\in\mathbb{N}}$ in B such that $f_\omega^{(r,s)}(y)=0$ for all $r,s\in\mathbb{N}$. Let h be the positive contraction in B_ω represented by $(y_n)_{n\in\mathbb{N}}$. Then h satisfies (4-1) for all $a,b\in\{a_n : n\in\mathbb{N}\}$. By density, h satisfies (4-1) for all $a,b\in A_+$. The result now follows by Lemma 4.1.

(b) By Dini's theorem, $\tau(\phi(e_n))\nearrow\theta(\tau)$ uniformly for $\tau\in\overline{T_\omega(B)}^{w*}$.[8] For each $l\in\mathbb{N}$, set

$$\gamma_l := \max_{\tau\in\overline{T_\omega(B)}^{w*}}(\theta(\tau)-\tau(\phi(e_l))). \tag{4-7}$$

Then $\gamma_l\ge 0$ as $\tau(\phi(e_n))$ increases with n, and $\lim_{l\to\infty}\gamma_l=0$ as the convergence is uniform.

We shall now prove the existence of a positive contraction $h\in B_\omega$ such that (4-1) holds for all $a,b\in A_+$ and that

$$\tau(h)=\lim_{n\to\infty}\tau(\phi(e_n)), \quad \tau\in\overline{T_\omega(B)}^{w*}. \tag{4-8}$$

Once again, we use Kirchberg's epsilon test (Lemma 1.8).

Let X_n, ϕ_n, $f_n^{(r,s)}$, and $f_\omega^{(r,s)}$ be as in (a). Define $g_n^{(l,+)}, g_n^{(l,-)} : X_n\to[0,\infty]$ for $l\in\mathbb{N}$ by

$$g_n^{(l,+)}(x) := \max\Big(\sup_{\tau\in T(B)}(\tau(x)-\tau(\phi_n(e_l)))-\gamma_l, 0\Big), \tag{4-9}$$

$$g_n^{(l,-)}(x) := \max\Big(\sup_{\tau\in T(B)}(\tau(\phi_n(e_l))-\tau(x)), 0\Big). \tag{4-10}$$

Then define $g_\omega^{(l,+)}, g_\omega^{(l,-)} : \prod_{n=1}^\infty X_n\to[0,\infty]$ by $(x_n)_{n\in\mathbb{N}}\mapsto\lim_{n\to\omega} g_n^{(l,+)}(x_n)$ and $(x_n)_{n\in\mathbb{N}}\mapsto \lim_{n\to\omega} g_n^{(l,-)}(x_n)$ respectively.

[8] Our convention is that approximate units for C*-algebras are by default assumed to be increasing.

The key observation is that a sequence $x = (x_n)_{n\in\mathbb{N}}$ representing a positive contraction $b \in B_\omega$ satisfies $g_\omega^{(l,+)}(x) = g_\omega^{(l,-)}(x) = 0$ if and only if

$$\tau(\phi(e_l)) \leq \tau(b) \leq \tau(\phi(e_l)) + \gamma_l, \quad \tau \in \overline{T_\omega(B)}^{w*}. \tag{4-11}$$

Let $m \in \mathbb{N}$ and $\epsilon > 0$. By (4-4), there is $k > m$ such that

$$\|\phi(e_k)\phi(a_r a_s) - \phi(a_r)\phi(a_s)\| < \epsilon, \quad 1 \leq r, s \leq m. \tag{4-12}$$

Let $x = (x_n)_{n\in\mathbb{N}}$ be a sequence of positive contractions in B representing $\phi(e_k)$. Then $f_\omega^{(r,s)}(x) < \epsilon$ whenever $1 \leq r, s \leq m$. Furthermore, as $k > m$, we have by (4-7) that for any $l \leq m$

$$\tau(\phi(e_l)) \leq \tau(\phi(e_k)) \leq \theta(\tau) \leq \tau(\phi(e_l)) + \gamma_l, \quad \tau \in \overline{T_\omega(B)}^{w*}. \tag{4-13}$$

Therefore, $g_\omega^{(l,+)}(x) = g_\omega^{(l,-)}(x) = 0$ for $l \leq m$.

By Kirchberg's epsilon test, there exists a sequence of positive contractions $y = (y_n)_{n\in\mathbb{N}}$ in B such that $f_\omega^{(r,s)}(y) = g_\omega^{(l,+)}(y) = g_\omega^{(l,-)}(y) = 0$ for all $r, s, l \in \mathbb{N}$. Let h be the positive contraction in B_ω represented by $(y_n)_{n\in\mathbb{N}}$. Then h satisfies (4-1) for all $a, b \in A_+$ as in (a) and

$$\tau(\phi(e_l)) \leq \tau(h) \leq \tau(\phi(e_l)) + \gamma_l, \quad \tau \in \overline{T_\omega(B)}^{w*},\ l \in \mathbb{N}. \tag{4-14}$$

Letting $l \to \infty$, we obtain (4-8) because $\lim_{l\to\infty} \gamma_l = 0$. The result now follows by Lemma 4.1. □

5. The uniqueness theorem

In this section, we establish the uniqueness theorem for maps from a C*-algebra into a C*-ultrapower, which will be used to bound the nuclear dimension of $\mathcal{Z}$-stable C*-algebras. This theorem is a nonunital version of [Castillejos et al. 2019, Lemma 4.8] which in turn builds on [Bosa et al. 2019a, Theorem 5.5]. For notational convenience, we work with ultrapowers throughout rather than general ultraproducts.

Theorem 5.1 (cf. [Bosa et al. 2019a, Theorem 5.5]). *Let B be a simple, separable, $\mathcal{Z}$-stable C*-algebra with CPoU, stable rank one in $B^\sim$, $Q\widetilde{T}(B) = \widetilde{T}_b(B) \neq 0$, and $T(B)$ compact. Let A be a unital, separable, nuclear C*-algebra, let $\phi_1 : A \to B_\omega$ be a c.p.c. order-zero map such that $\phi_1(a)$ is full for all nonzero $a \in A$ and the induced map $\bar{\phi}_1 : A \to B^\omega$ is a *-homomorphism, and let $\phi_2 : A \to B_\omega$ be a c.p.c. order-zero map such that*

$$\tau \circ \phi_1 = \tau \circ \phi_2^m, \quad \tau \in T(B_\omega),\ m \in \mathbb{N}, \tag{5-1}$$

where order-zero functional calculus is used to interpret ϕ_2^m.[9] *Let $k \in \mathcal{Z}_+$ be a positive contraction with spectrum* [0, 1], *and define c.p.c. order-zero maps $\psi_i : A \to (B \otimes \mathcal{Z})_\omega$ by $\psi_i(a) := \phi_i(a) \otimes k$. Then ψ_1 and ψ_2 are unitarily equivalent in $(B \otimes \mathcal{Z})_\omega^\sim$.*

The proof of Theorem 5.1 follows by a careful adaptation of the arguments from [Bosa et al. 2019a; Castillejos et al. 2019] to handle the potential nonunitality of B. In the subsections that follow, we shall first review the key ingredients of the proof of [Castillejos et al. 2019, Lemma 4.8] and [Bosa et al. 2019a,

[9] Suppose $\phi_2(x) = \pi_2(t \otimes x)$ where $\pi_2 : C_0(0, 1] \otimes A \to B_\omega$ is a *-homomorphism and t is the canonical generator of $C_0(0, 1]$. Then $\phi_2^m(x) = \pi_2(t^m \otimes x)$; see [Winter and Zacharias 2009, Corollary 4.2].

Theorem 5.5] and explain clearly the modifications needed in the nonunital setting. We shall then return to the proof of Theorem 5.1.

5A. *The 2 × 2 matrix trick.* We begin by reviewing the 2×2 matrix trick, which converts the problem of unitary equivalence of maps into the problem of unitary equivalence of positive elements. The version stated below is very similar to [Bosa et al. 2019a, Lemma 2.3]; however, for our applications, we must weaken the stable rank one assumption and we have no need for the Kirchberg algebra case.

Proposition 5.2 (cf. [Bosa et al. 2019a, Lemma 2.3]). *Let A be a separable, unital* C**-algebra and B be a separable* C**-algebra. Let $\phi_1, \phi_2 : A \to B_\omega$ be c.p.c. order-zero maps and $\hat{\phi}_1, \hat{\phi}_2 : A \to B_\omega$ be supporting order-zero maps (as in (A-1)). Suppose that B_ω has stable rank one in $B_\omega^\sim$. Let $\pi : A \to M_2(B_\omega)$ be given by*

$$\pi(a) := \begin{pmatrix} \hat{\phi}_1(a) & 0 \\ 0 & \hat{\phi}_2(a) \end{pmatrix}, \quad a \in A, \tag{5-2}$$

and set $C := M_2(B_\omega) \cap \pi(A)' \cap \{1_{M_2(B_\omega^\sim)} - \pi(1_A)\}^\perp$. If

$$\begin{pmatrix} \phi_1(1_A) & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 0 \\ 0 & \phi_2(1_A) \end{pmatrix} \tag{5-3}$$

are unitarily equivalent in $C^\sim$, then ϕ_1 and ϕ_2 are unitarily equivalent in $B_\omega^\sim$.

Proof. Let

$$u = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} \in C^\sim \tag{5-4}$$

be a unitary implementing the unitary equivalence of the positive elements. Since B_ω has stable rank one in $B_\omega^\sim$, we have that $u_{21}^*\phi_2(1_A)$ is the limit of invertibles in $B_\omega^\sim$. Hence, by [Bosa et al. 2019a, Lemma 1.20] and Kirchberg's epsilon test, there is a unitary $w \in B_\omega^\sim$ with $u_{21}^*\phi_2(1_A) = w|u_{21}^*\phi_2(1_A)|$. Arguing exactly as in the proof of [Bosa et al. 2019a, Lemma 2.3], we obtain that $\phi_1(a) = w\phi_2(a)w^*$ for all $a \in A$. □

5B. *Property* (SI). Our goal in this section is to show that c.p.c. order-zero maps from separable, unital C*-algebras into ultrapowers of C*-algebras with compact trace space satisfy property (SI).

The following definition is a variant of [Bosa et al. 2019a, Definition 4.2], which in turn goes back to [Matui and Sato 2012], that allows us to handle cases when the codomain is not unital.

Definition 5.3. Let B be a simple, separable, C*-algebra with $Q\widetilde{T}(B) = \widetilde{T}_b(B) \neq 0$. Write J_{B_ω} for the trace kernel ideal. Let A be a separable, unital C*-algebra, let $\pi : A \to B_\omega$ be a c.p.c. order-zero map, and define

$$C := B_\omega \cap \pi(A)' \cap \{1_{B_\omega^\sim} - \pi(1_A)\}^\perp. \tag{5-5}$$

The map π has *property* (*SI*) if the following holds. For all $e, f \in C_+$ such that $e \in J_{B_\omega}$, $\|f\| = 1$ and f has the property that, for every nonzero $a \in A_+$, there exists $\gamma_a > 0$ such that

$$\tau(\pi(a)f^n) > \gamma_a, \quad \tau \in T_\omega(B),\ n \in \mathbb{N}, \tag{5-6}$$

there exists $s \in C$ such that

$$s^*s = e \quad \text{and} \quad fs = s. \tag{5-7}$$

The main result of this subsection is that under certain hypotheses, c.p.c. maps $A \to B_\omega$ have property (SI). This result is a nonunital version of [Bosa et al. 2019a, Lemma 4.4] and its proof is almost identical to the original proof. Since this result is one of the most delicate parts of this work, we include its proof.

Proposition 5.4 (cf. [Bosa et al. 2019a, Lemma 4.4]). *Let B be a simple, separable, $\mathcal{Z}$-stable* C*-*algebra with $Q\widetilde{T}(B) = \widetilde{T}_b(B) \neq 0$. Let A be a separable, unital, nuclear* C*-*algebra. Then every c.p.c. order-zero map $\pi : A \to B_\omega$ has property* (*SI*).

Proof. Let $\pi : A \to B_\omega$ be a c.p.c. order-zero map with A and B as in the statement. Let C be as in (5-5) and set $\bar{C} := C/(C \cap J_{B_\omega})$. Let $e, f \in C_+$ and γ_a be as in the definition of property (SI). As in the proof of [Bosa et al. 2019a, Lemma 4.4], it is enough to exhibit an element $s \in B_\omega$ approximately satisfying

$$s^*\pi(a)s = \pi(a)e \quad \text{for all } a \in A \text{ and } fs = s. \tag{5-8}$$

Let $\mathcal{F} \subseteq A$ be a finite subset of contractions and $\epsilon > 0$. Since B is $\mathcal{Z}$-stable, using Lemma A.6(ii) we can find a c.p.c. order-zero map $\alpha : \mathcal{Z} \to B_\omega \cap \pi(A)' \cap \{e, f\}'$ such that $\alpha(1_{\mathcal{Z}})$ acts as unit on $\pi(A)$. Therefore, we may define a new c.p.c. map $\tilde{\pi} : A \otimes \mathcal{Z} \to B_\omega$ by setting $\tilde{\pi}(a \otimes z) := \pi(a)\alpha(z)$. It follows by construction that $\pi(a) = \tilde{\pi}(a \otimes 1_{\mathcal{Z}})$ for $a \in A$. By [Winter and Zacharias 2009, Corollary 4.3], $\tilde{\pi}$ is a c.p.c. order-zero map and note that e and f are elements of the relative commutant $B_\omega \cap \tilde{\pi}(A \otimes \mathcal{Z})' \cap \{1_{B_\omega^\sim} - \tilde{\pi}(1_{A \otimes \mathcal{Z}})\}^\perp$.

Arguing as in the proof of [Bosa et al. 2019a, Lemma 4.4], for any $b \in (A \otimes \mathcal{Z})_+$, there exists a positive constant $\tilde{\gamma}_b$ such that

$$\tau(\tilde{\pi}(b) f^n) > \tilde{\gamma}_b, \quad \tau \in T_\omega(B_\omega),\ n \in \mathbb{N}. \tag{5-9}$$

Next, we will apply [loc. cit., Lemmas 4.7 and 4.8] to the unital, separable, nuclear C*-algebra A. Set $\mathcal{G} := \{x \otimes 1_{\mathcal{Z}} : x \in \mathcal{F}\} \subseteq A \otimes \mathcal{Z}$. Since no irreducible representation of $A \otimes \mathcal{Z}$ contains any compact operator, by [loc. cit., Lemma 4.8] there exist $L, N \in \mathbb{N}$, pairwise inequivalent pure states $\lambda_1, \dots, \lambda_L$ on $A \otimes \mathcal{Z}$ and elements $c_i, d_{i,l} \in A \otimes \mathcal{Z}$ for $i = 1, \dots, N,\ l = 1, \dots, L$ such that

$$x \approx_\epsilon \sum_{l=1}^{L} \sum_{i,j=1}^{N} \lambda_l(d_{i,l}^* x d_{j,l}) c_i^* c_j, \quad x \in \mathcal{G}. \tag{5-10}$$

By [loc. cit., Lemma 4.7], applied to the set $\{d_{i,l}^* x d_{j,l'} : x \in \mathcal{G},\ i, j = 1, \dots, N,\ l, l' = 1, \dots, L\}$, there exist positive contractions $a_1, \dots, a_L \in (A \otimes \mathcal{Z})_+$ such that for $l = 1, \dots, L,\ \lambda_l(a_l) = 1$ and

$$a_l d_{i,l}^* x d_{j,l} a_l \approx_\delta \lambda_l(d_{i,l}^* x d_{j,l}) a_l^2, \quad x \in \mathcal{G},\ i, j = 1, \dots, N, \tag{5-11}$$

and for $l \neq l'$,

$$a_l d_{i,l}^* x d_{j,l'} a_{l'} \approx_\delta 0, \quad x \in \mathcal{G},\ i, j = 1, \dots, N, \tag{5-12}$$

with $\delta := \epsilon/(N^2 L \max_k \|c_k\|^2)$. Note, the condition $\lambda_l(a_l) = 1$ ensures that the a_l have norm 1.

By hypothesis, B is simple, separable, $\mathcal{Z}$-stable and $Q\widetilde{T}(B) = \widetilde{T}_b(B) \neq 0$. Hence, by Proposition 1.5, B has strict comparison of positive elements by bounded traces. Thus, for $l = 1, \dots, L$, we may apply Lemma A.11 with a_l in place of a. Let $S_l \subseteq (A \otimes \mathcal{Z})_+ \setminus \{0\}$ denote the countable subset such that the conclusion of Lemma A.11 is satisfied with a_l in place of a.

Let $\hat{\pi} : A \otimes \mathcal{Z} \to B_\omega \cap \{f\}'$ be a supporting c.p.c. order-zero map for $\tilde{\pi}$. As in [loc. cit., Lemma 4.4], using (5-9) and Lemma A.4 twice (taking $x := 0$ and with $S_0 := \tilde{\pi}(S_1 \cup \dots \cup S_L)$), we find $t, h \in B_\omega \cap \hat{\pi}(A \otimes \mathcal{Z})' \cap \tilde{\pi}(A \otimes \mathcal{Z})'$ satisfying $h \lhd t \lhd f$ and, for every $b \in S_1 \cup \dots \cup S_L$,

$$\tau(\tilde{\pi}(b)h^n) \geq \tilde{\gamma}_b, \quad \tau \in T_\omega(B), \ n \in \mathbb{N}. \tag{5-13}$$

By Lemma A.11 (with $\tilde{\pi}$ in place of π), there is a contraction $r_l \in B_\omega$ such that $\tilde{\pi}(a_l)r_l = tr_l = r_l$ and $r_l^* r_l = e$. Using $t \lhd f \lhd \tilde{\pi}(1_A)$, we obtain $\tilde{\pi}(1_A)r_l = r_l$ for each l, and hence

$$r_l^* \tilde{\pi}(a_l^2) r_l = \tilde{\pi}(1_A)^{1/2} e \tilde{\pi}(1_A)^{1/2}. \tag{5-14}$$

Set

$$s := \sum_{l=1}^{L} \sum_{i=1}^{N} \hat{\pi}(d_{i,l} a_l) r_l \hat{\pi}(c_i) \in B_\omega. \tag{5-15}$$

Using $r_l = tr_l$, $t \lhd f$ and that t commutes with the image of $\hat{\pi}$, we can obtain $fs = s$. For $x \in \mathcal{F}$, the calculations of [loc. cit., Lemma 4.4, equation 4.46] show

$$s^* \pi(x) s = \pi(x) e, \tag{5-16}$$

as required. Then Kirchberg's epsilon test produces an element $s \in B_\omega$ that exactly satisfies (5-8). As in the proof of [loc. cit., Lemma 4.10], $s \in C$. ☐

5C. *Structural results for relative commutants.* Combining property (SI) with complemented partitions of unity (CPoU), one can now prove important structural properties for the relative commutant algebras $C := B_\omega \cap \pi(A)' \cap \{1_{\widetilde{B}_\omega} - \pi(1_A)\}^\perp$ arising from the 2×2 matrix trick.

Theorem 5.5 (cf. [Castillejos et al. 2019, Lemma 4.7]). *Let B be a simple, separable, $\mathcal{Z}$-stable* C*-*algebra with $Q\widetilde{T}(B) = \widetilde{T}_b(B) \neq 0$ and $T(B)$ compact. Suppose additionally that B has CPoU. Let A be a separable unital nuclear* C*-*algebra and $\pi : A \to B_\omega$ a c.p.c. order-zero map which induces a *-homomorphism $\bar{\pi} : A \to B^\omega$. Let*

$$C := B_\omega \cap \pi(A)' \cap \{1_{\widetilde{B}_\omega} - \pi(1_A)\}^\perp, \quad \overline{C} := C/(C \cap J_{B_\omega}). \tag{5-17}$$

Then:

(i) *All traces on C factor through $\overline{C}$.*

(ii) *C has strict comparison of positive elements by bounded traces.*

(iii) *The traces on C are the closed convex hull of traces of the form $\tau(\pi(a)\,\cdot\,)$ for $\tau \in T(B_\omega)$ and $a \in A_+$ with $\tau(\pi(a)) = 1$.*

First, we discuss two preliminary results, which originate from [Bosa et al. 2019a, Lemmas 3.20 and 3.22] and were generalised in [Castillejos et al. 2019, Lemmas 4.3 and 4.6], where the newly discovered CPoU was used in place of the earlier methods that required further assumptions on $T(B)$. Both results are proven by checking that these lemmas approximately hold for $\pi_\tau(B^\omega)''$ for any trace in $\tau \in \overline{T_\omega(B)}^{w*}$, which in turn follows from the fact that $\pi_\tau(B^\omega)''$ is a finite von Neumann algebra, and then using CPoU to patch local solutions together. In [loc. cit.], these results are stated for B unital, but the proofs do not make use of the unit. They only require that $T(B)$ is compact, as this guarantees that B^ω is unital [loc. cit., Proposition 1.11].

Lemma 5.6 (cf. [Castillejos et al. 2019, Lemma 4.3]). *Let B be a separable* C*-*algebra with $Q\widetilde{T}(B) = \widetilde{T}_b(B) \neq 0$ and $T(B)$ compact. Suppose B has CPoU. Let $S \subseteq B^\omega$ be a $\|\cdot\|_{2,T_\omega(B)}$-separable and self-adjoint subset, and let p be a projection in the centre of $B^\omega \cap S'$. Then $p(B^\omega \cap S')$ has strict comparison of positive elements by bounded traces.*

Proposition 5.7 (cf. [Castillejos et al. 2019, Lemma 4.6]). *Let B be a separable* C*-*algebra with $Q\widetilde{T}(B) = \widetilde{T}_b(B) \neq 0$ and $T(B)$ compact. Suppose B has CPoU. Let A be a separable, unital, nuclear* C*-*algebra and $\phi : A \to B^\omega$ a *-homomorphism. Set $C := B^\omega \cap \phi(A)' \cap \{1_{B^\omega} - \phi(1_A)\}^\perp$. Define T_0 to be the set of all traces on C of the form $\tau(\phi(a)\,\cdot\,)$, where $\tau \in T(B^\omega)$ and $a \in A_+$ satisfies $\tau(\phi(a)) = 1$.*

Suppose $z \in C$ is a contraction and $\delta > 0$ satisfies $|\rho(z)| \leq \delta$ for all $\rho \in T_0$. Write $K := 12 \cdot 12 \cdot (1+\delta)$. Then there exist contractions $w, x_1, \ldots, x_{10}, y_1, \ldots, y_{10} \in C$ such that

$$z = \delta w + K \sum_{i=1}^{10} [x_i, y_i]. \tag{5-18}$$

In particular, $T(C)$ is the closed convex hull of T_0.

With these preparatory results now established, we explain how to adapt the original proof of [Bosa et al. 2019a, Theorem 4.1] to prove Theorem 5.5.

Proof of Theorem 5.5. For (i), the proof of [loc. cit., Theorem 4.1(i)] still works in our situation with the following minor modifications. We use Lemma A.4 instead of [loc. cit., Lemma 1.18], Proposition 5.4 in place of [loc. cit., Lemma 4.4] and Lemma A.5 in place of [loc. cit., Lemma 1.19].

Similarly, for (ii) we use the proof from [loc. cit., Lemma 3.20] with the following modifications. Since B is $\mathcal{Z}$-stable, any matrix algebra embeds into $B^\omega \cap \bar{\pi}(A)' \cap \{\bar{c}\}'$ [Castillejos et al. 2019, Proposition 2.3]. We use Lemma 5.6 to see that $\bar{C}$ has strict comparison of positive elements by traces in place of [Bosa et al. 2019a, Lemma 3.20], and [Castillejos et al. 2019, Lemma 1.8] in place of [Bosa et al. 2019a, Lemma 3.10].

In the same vein, (iii) follows from (i), [Castillejos et al. 2019, Lemma 1.5], and Proposition 5.7. □

5D. ***Unitary equivalence of totally full positive elements.*** The main theorem of this section is a nonunital version of the classification of totally full positive elements up to unitary equivalence in relative commutant sequence algebras obtained in [Bosa et al. 2019a, Lemma 5.1].[10]

Let us begin by stating the following lemma which can be proved in exactly the same way as [Bosa et al. 2019a, Lemma 5.3] since the Robert–Santiago argument [2010] at the core of the proof has no

[10]Recall that a nonzero $h \in C_+$ is *totally full* if $f(h)$ is full in C for every nonzero $f \in C_0((0, \|h\|])_+$ [Bosa et al. 2019a, Definition 1.1].

unitality hypothesis. All that is required is to formally replace all occurrences of 1_{B_ω} with $1_{B_\omega^\sim}$, and replace [Bosa et al. 2019a, Lemma 1.17] with Lemma A.3, [loc. cit., Lemma 2.2] with Lemma A.8, and [loc. cit., Lemma 5.4] with Lemma A.9.

Lemma 5.8 (cf. [Bosa et al. 2019a, Lemma 5.3]). *Let B be a separable, $\mathcal{Z}$-stable* C*-*algebra and let A be a separable, unital* C*-*algebra. Let $\pi : A \to B_\omega$ be a c.p.c. order-zero map such that*

$$C := B_\omega \cap \pi(A)' \cap \{1_{B_\omega^\sim} - \pi(1_A)\}^\perp \tag{5-19}$$

is full in B_ω.

Assume that every full hereditary subalgebra D of C satisfies the following: if $x \in D$ is such that there exist totally full elements $e_l, e_r \in D_+$ such that $e_l x = x e_r = 0$, then there exists a full element $s \in D$ such that $sx = xs = 0$.

Let $a, b \in C_+$ be totally full positive contractions. Then a and b are unitarily equivalent in $C^\sim$ if and only if for every $f \in C_0(0, 1]_+$, $f(a)$ is Cuntz equivalent to $f(b)$ in C.

With this lemma in hand, we can now prove the main theorem of this section.

Theorem 5.9 (cf. [Bosa et al. 2019a, Theorem 5.1]). *Let B be a separable, $\mathcal{Z}$-stable* C*-*algebra with $Q\widetilde{T}(B) = \widetilde{T}_b(B) \neq 0$. Let A be a separable, unital* C*-*algebra and let $\pi : A \to B_\omega$ be a c.p.c. order-zero map such that*

$$C := B_\omega \cap \pi(A)' \cap \{1_{B_\omega^\sim} - \pi(1_A)\}^\perp \tag{5-20}$$

is full in B_ω and has strict comparison of positive elements with respect to bounded traces.

Let $a, b \in C_+$ be totally full positive elements. Then a and b are unitarily equivalent in $C^\sim$ if and only if $\tau(a^k) = \tau(b^k)$ for every $\tau \in T(C)$ and $k \in \mathbb{N}$.

Proof. Let $a, b \in C_+$ be totally full positive elements satisfying $\tau(a^k) = \tau(b^k)$ for every $\tau \in T(C)$ and $k \in \mathbb{N}$. Without loss of generality, assume that a and b are contractions.

After replacing [Bosa et al. 2019a, Lemma 1.22(iv)] with Lemma A.6(iv), part (i) of the proof of [loc. cit., Theorem 5.1] shows that the technical hypothesis of Lemma 5.8 is satisfied for every full hereditary subalgebra $D \subseteq C$. The argument of part (ii) of the proof of [loc. cit., Theorem 5.1] then shows that $f(a)$ is Cuntz equivalent to $f(b)$ for all $f \in C_0(0, 1]_+$. (This part of the proof of [loc. cit., Theorem 5.1] does not make any use of the unit; only strict comparison is needed.) By Lemma 5.8, a and b are unitarily equivalent by unitaries in $C^\sim$. The converse is straightforward. □

5E. *Proof of the uniqueness theorem.* We now have all the ingredients we need for the proof of Theorem 5.1.

Proof of Theorem 5.1. By hypothesis, $\bar{\phi}_1(1_A) \in B^\omega$ is a projection. Hence $d_\tau(\phi_1(1_A)) = \tau(\phi_1(1_A))$ and we immediately can conclude that the map $\tau \mapsto d_\tau(\phi_1(1_A))$ is continuous. Similarly, by (5-1), the map $\tau \mapsto d_\tau(\phi_2(1_A))$ is continuous. Hence, by Lemma A.1, there exist supporting order-zero maps $\hat{\phi}_1, \hat{\phi}_2 : A \to B_\omega$ such that

$$\tau(\hat{\phi}_i(a)) = \lim_{m\to\infty} \tau(\phi_i^{1/m}(a)), \quad a \in A,\ \tau \in T_\omega(B),\ i = 1, 2, \tag{5-21}$$

and the maps $\bar{\hat{\phi}}_i : A \to B^\omega$ are *-homomorphisms. In particular,

$$\tau(\hat{\phi}_2(a)) \overset{(5\text{-}1)}{=} \tau(\phi_1(a)), \quad a \in A,\ \tau \in T_\omega(B). \tag{5-22}$$

By Proposition 1.12, B_ω has stable rank one in $B_\omega^\sim$. Thus, we may use the 2×2 matrix trick (Proposition 5.2). Recall $\psi_i(a) := \phi_i(a) \otimes k$ and define $\hat{\psi}_1, \hat{\psi}_2 : A \to (B \otimes \mathcal{Z})_\omega$ by $\hat{\psi}_i(a) := \hat{\phi}_i(a) \otimes 1_{\mathcal{Z}}$, with $i = 1, 2$. It is immediate that $\hat{\psi}_i$ is a supporting order-zero map for ψ_i. Then define $\pi : A \to M_2(B_\omega) \subseteq M_2((B \otimes \mathcal{Z})_\omega)$ by

$$\pi(a) := \begin{pmatrix} \hat{\psi}_1(a) & 0 \\ 0 & \hat{\psi}_2(a) \end{pmatrix}, \quad a \in A, \tag{5-23}$$

and set $C := M_2((B \otimes \mathcal{Z})_\omega) \cap \pi(A)' \cap \{1_{M_2((B\otimes\mathcal{Z})_\omega^\sim)} - \pi(1_A)\}^\perp$. We will show that

$$h_1 := \begin{pmatrix} \psi_1(1_A) & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad h_2 := \begin{pmatrix} 0 & 0 \\ 0 & \psi_2(1_A) \end{pmatrix}$$

are unitarily equivalent in $C^\sim$. For nonzero $a \in A$, observe that

$$0 \le \begin{pmatrix} \psi_1(a) & 0 \\ 0 & 0 \end{pmatrix} \le \begin{pmatrix} \hat{\psi}_1(a) & 0 \\ 0 & \hat{\psi}_2(a) \end{pmatrix} = \pi(a), \tag{5-24}$$

and using that $\psi_1(a)$ is full in $(B \otimes \mathcal{Z})_\omega$ since $\phi_1(a)$ is full, we conclude that $\pi(a)$ is full in $M_2((B \otimes \mathcal{Z})_\omega)$. By construction, the induced map $\bar{\pi} : A \to M_2(B^\omega)$ is a *-homomorphism. Thus, by Theorem 5.5, C has strict comparison.

Notice that $h_1 \in C$ is full in $M_2(B_\omega)$, and hence C is also full in $M_2(B_\omega)$. Let ρ be a trace on C of the form $\tau(\pi(x)\,\cdot)$, where $\tau \in T(M_2((B \otimes \mathcal{Z})_\omega))$, $x \in A_+$ and $\tau(\pi(x)) = 1$. Set a trace $\tilde{\tau}$ on B_ω by $\tilde{\tau}(b) := \tau(1_{M_2} \otimes b \otimes 1_{\mathcal{Z}})$. Thus, as in [Bosa et al. 2019a, Theorem 5.5, equation (5.41)],

$$\rho(h_1^m) = \tfrac{1}{2}\tau_{\mathcal{Z}}(k^m) = \rho(h_2^m), \quad m \in \mathbb{N}. \tag{5-25}$$

By Theorem 5.5, equation (5-25) holds for any trace on C.

A standard strict comparison argument shows that $f(h_1)$ and $f(h_2)$ are full in C for any $f \in C_0(0,1]_+$, so h_1 and h_2 are totally full. By Theorem 5.9, h_1 is unitarily equivalent to h_2 in $C^\sim$. By the 2×2 matrix trick (Proposition 5.2), ψ_1 and ψ_2 are equivalent in $(B \otimes \mathcal{Z})_\omega^\sim$. □

6. Nuclear dimension and $\mathcal{Z}$-stability

In this section, we prove Theorems A and B, and deduce Corollaries C and D.

Theorem 6.1. *Let A be a simple, separable, nuclear and $\mathcal{Z}$-stable* C*-*algebra. Then* $\dim_{\text{nuc}} A \le 1$.

Proof. By Theorem 2.7, either A is stably isomorphic to a unital C*-algebra B, or A is stably isomorphic to a stably projectionless C*-algebra A_0 with $Q\widetilde{T}(A_0) = \widetilde{T}_b(A_0) \neq 0$ and $T(A_0)$ compact.

The stably unital case follows immediately from [Castillejos et al. 2019, Theorem B] together with Proposition 2.3. Indeed, if A is stably isomorphic to a unital C*-algebra B, then B is also simple, separable, nuclear and $\mathcal{Z}$-stable by Proposition 2.3. Hence, $\dim_{\mathrm{nuc}} B \le 1$ by [loc. cit., Theorem B]. Therefore, $\dim_{\mathrm{nuc}} A \le 1$ by a second application of Proposition 2.3.

We now consider the case when A is stably isomorphic to a stably projectionless C*-algebra A_0 with $Q\widetilde{T}(A_0) = \widetilde{T}_b(A_0) \neq 0$ and $T(A_0)$ compact. By Proposition 2.3, A_0 is simple, separable, nuclear and $\mathcal{Z}$-stable. Since A_0 is stably projectionless and $\mathcal{Z}$-stable, A_0 has stable rank one in $A_0^{\sim}$ by Theorem 1.10. Furthermore, A_0 has CPoU by Theorem 1.14.

In light of Proposition 2.3, it suffices to prove that $\dim_{\mathrm{nuc}} A_0 \le 1$. We now show this using the same fundamental strategy of [Bosa et al. 2019a] (taking into account the modification introduced in [Castillejos et al. 2019]). We shall estimate the nuclear dimension of the first factor embedding $j : A_0 \to A_0 \otimes \mathcal{Z}$, $j(x) = x \otimes 1_{\mathcal{Z}}$, in the sense of [Tikuisis and Winter 2014, Definition 2.2]. Since A_0 is $\mathcal{Z}$-stable and $\mathcal{Z}$ is strongly self-absorbing, we have $\dim_{\mathrm{nuc}}(A_0) = \dim_{\mathrm{nuc}}(j)$; see [loc. cit., Proposition 2.6].

Let $\iota : A_0 \to (A_0)_\omega$ be the canonical embedding. Let h be a strictly positive contraction in A_0, and let $(e_n)_{n\in\mathbb{N}}$ be the approximate identity given by $e_n := h^{1/n}$. Then $\lim_{n\to\infty} \tau(e_n) = 1$ for all $\tau \in T(A_0)$. Since $T(A_0)$ is compact, $\tau \circ \iota \in T(A_0)$ for all $\tau \in T_\omega(A_0)$ and so for all $\tau \in \overline{T_\omega(A_0)}^{w*}$. It follows that

$$\lim_{n\to\infty} \tau(\iota(e_n)) = 1, \qquad \tau \in \overline{T_\omega(A_0)}^{w*}. \tag{6-1}$$

Thus applying Lemma 4.1, we obtain a c.p.c. order-zero extension $\iota^{\sim} : A_0^{\sim} \to (A_0)_\omega$ with $\tau(\iota^{\sim}(1_{A_0^{\sim}})) = 1$ for all $\tau \in \overline{T_\omega(A_0)}^{w*}$. Writing $\overline{\iota^{\sim}} : A_0^{\sim} \to A_0^{\omega}$ for the induced map into the uniform tracial ultrapower, we observe that $1_{A_0^{\omega}} - \overline{\iota^{\sim}}(1_{A_0^{\sim}})$ is a positive element in A_0^{ω} that vanishes on all limit traces and so must be zero. Hence, $\overline{\iota^{\sim}}$ is a unital c.p.c. order zero map and so must be a unital *-homomorphism.

Let $(\phi_n : A_0 \to A_0)_{n=1}^{\infty}$ be the sequence of c.p.c. maps constructed in Lemma 3.2, which factorize as $\eta_n \circ \theta_n$ through finite-dimensional algebras F_n as in (3-7). By construction, the induced map $\Phi : A_0 \to (A_0)_\omega$ is c.p.c. order-zero and the induced map $\overline{\Phi} : A_0 \to A_0^{\omega}$ agrees with the diagonal inclusion $\bar{\iota} : A_0 \to A_0^{\omega}$. It follows that $\tau \circ \Phi = \tau \circ \iota$ for all $\tau \in \overline{T_\omega(A)}^{w*}$. Hence,

$$\lim_{n\to\infty} \tau(\Phi(e_n)) = 1, \qquad \tau \in \overline{T_\omega(A_0)}^{w*}. \tag{6-2}$$

Therefore, applying Lemma 4.1 again, we obtain a c.p.c. order-zero extension $\Phi^{\sim} : A_0^{\sim} \to (A_0)_\omega$ with $\tau(\Phi^{\sim}(1_{A_0^{\sim}})) = 1$ for all $\tau \in \overline{T_\omega(A_0)}^{w*}$. Arguing as before, $\overline{\Phi^{\sim}} : A_0^{\sim} \to A_0^{\omega}$ is a unital *-homomorphism. In fact, we have $\overline{\Phi^{\sim}} = \overline{\iota^{\sim}}$ since both maps agree on A_0 by construction and are unital.

We are almost ready to apply Theorem 5.1 to the c.p.c. order-zero maps $\iota^{\sim}$ and $\Phi^{\sim}$. We observe that A_0 is a simple, separable, $\mathcal{Z}$-stable with CPoU, stable rank one in $A_0^{\sim}$, $Q\widetilde{T}(A_0) = \widetilde{T}_b(A_0) \neq 0$, and $T(A_0)$ compact; that $A_0^{\sim}$ is unital, separable and nuclear; and that both maps induce a unital *-homomorphism $\overline{\iota^{\sim}} = \overline{\Phi^{\sim}} : A_0^{\sim} \to A_0^{\omega}$. Since $\overline{\iota^{\sim}} = \overline{\Phi^{\sim}}$ and both maps are *-homomorphisms, we have

$$\tau \circ \iota = \tau \circ \Phi^{m}, \qquad \tau \in \overline{T_\omega(A_0)}^{w*},\ m \in \mathbb{N}. \tag{6-3}$$

The tracial condition (5-1) follows because $T_\omega(A_0)$ is dense in $T((A_0)_\omega)$ by Theorem 1.7.

Before we may apply Theorem 5.1, we must show that $\iota^{\sim}(x)$ is full for all nonzero $x \in A_0^{\sim}$. By Proposition 1.5, A_0 has strict comparison by bounded traces because A_0 is simple, separable, $\mathcal{Z}$-stable and $Q\widetilde{T}(A_0) = \widetilde{T}_b(A_0) \neq 0$. Hence, $(A_0)_\omega$ has strict comparison in the sense of Lemma A.10.

Using that A_0 is simple and $T(A_0)$ is compact, the minimum $\gamma_a := \min_{\tau\in T(A)} \tau(a)$ exists and is strictly positive for any nonzero $a \in (A_0)_{+,1}$. Since $\tau \circ \iota \in T(A_0)$ for any $\tau \in \overline{T_\omega(A_0)}^{w*}$, we have $d_\tau(\iota(a)) \geq \gamma_a$ for any $\tau \in \overline{T_\omega(A_0)}^{w*}$. Hence, $\iota(a)$ is full in $(A_0)_\omega$ using Lemma A.10.

For any nonzero $x \in A_0^\sim$, the ideal I_x of $A_0^\sim$ generated by x contains a nonzero positive contraction $a \in A_{+,1}$. A simple computation using supporting order-zero maps shows that the ideal of $(A_0)_\omega$ generated by $\iota^\sim(x)$ contains $\iota^\sim(I_x)$, which is full since it contains the full element $\iota^\sim(a)$. Hence, $\iota^\sim(x)$ is full in $(A_0)_\omega$.

Fix a positive contraction $k \in \mathcal{Z}_+$ of full spectrum. Applying Theorem 5.1 to the maps $\iota^\sim$ and $\Phi^\sim$, we obtain unitaries $w^{(0)}, w^{(1)} \in (A_0 \otimes \mathcal{Z})^\sim_\omega$ such that

$$x \otimes k = w^{(0)}(\Phi(x) \otimes k)w^{(0)*}, \tag{6-4}$$

$$x \otimes (1_{\mathcal{Z}} - k) = w^{(1)}(\Phi(x) \otimes (1_{\mathcal{Z}} - k))w^{(1)*}, \quad x \in A. \tag{6-5}$$

Choose representing sequences $(w_n^{(0)})_{n=1}^\infty$ and $(w_n^{(1)})_{n=1}^\infty$ of unitaries in $(A_0 \otimes \mathcal{Z})^\sim$ for $w^{(0)}$ and $w^{(1)}$, respectively. We have c.p.c. maps $\theta_n \oplus \theta_n : A_0 \to F_n \oplus F_n$ and $\tilde{\eta}_n : F_n \oplus F_n \to A_0 \otimes \mathcal{Z}$, where

$$\tilde{\eta}_n(y_0, y_1) := w_n^{(0)}(\eta_n(y_0) \otimes k)w_n^{(0)*} + w_n^{(1)}(\eta_n(y_1) \otimes (1_{\mathcal{Z}} - k))w_n^{(1)*}. \tag{6-6}$$

Hence, $j(x)$ is the limit, as $n \to \omega$, of $(\tilde{\eta}_n \circ (\theta_n \oplus \theta_n)(x))_{n=1}^\infty$ and, since $\tilde{\eta}_n$ is the sum of two c.p.c. order-zero maps, $\dim_{\text{nuc}}(j) \leq 1$. □

Theorem 6.2. *Let A be a nonelementary, simple, separable, nuclear* C*-*algebra. Then A has finite nuclear dimension if and only if it is $\mathcal{Z}$-stable.*

Proof. Let A be a nonelementary, simple, separable, nuclear C*-algebra. If A is $\mathcal{Z}$-stable, then $\dim_{\text{nuc}}(A) \leq 1 < \infty$ by Theorem 6.1. Conversely, if $\dim_{\text{nuc}}(A) < \infty$, then A is $\mathcal{Z}$-stable by [Tikuisis 2014, Theorem 8.5] □

Corollary 6.3. *The nuclear dimension of a simple* C*-*algebra is* 0, 1 *or* ∞.

Proof. Let A be a simple, separable C*-algebra with finite nuclear dimension. Then, in particular, A is nuclear. If A is elementary, then $\dim_{\text{nuc}}(A) = 0$; otherwise, A is $\mathcal{Z}$-stable by Theorem 6.2. Hence, $\dim_{\text{nuc}}(A) \leq 1$ by Theorem 6.1. The nonseparable case follows from the separable one as in the proof of [Castillejos et al. 2019, Corollary C]. □

In [Elliott 1996, Theorem 5.2.2], a stably projectionless, simple, separable, nuclear C*-algebra with a unique trace, $K_0 = \mathbb{Z}$ and $K_1 = 0$ is constructed as a limit of 1-dimensional noncommutative CW complexes. By [Gong and Lin 2017, Theorem 1.4], there is a unique C*-algebra with these properties that has finite nuclear dimension and satisfies the UCT. This C*-algebra is denoted by $\mathcal{Z}_0$ [Gong and Lin 2017, Definition 8.1], reflecting its role as a stably projectionless analogue of the Jiang–Su algebra $\mathcal{Z}$. An important further property of $\mathcal{Z}_0$, which follows from its construction, is that $\mathcal{Z}_0$ is $\mathcal{Z}$-stable [Gong and Lin 2017, Remark 7.3, Definition 8.1].

It has recently been shown that simple, separable C*-algebras which satisfy the UCT and have finite nuclear dimension are classified up to stabilisation with $\mathcal{Z}_0$ by the Elliott invariant [Gong and Lin

2017, Theorem 1.2]. The appropriate form of the Elliott invariant in this setting is detailed in [loc. cit., Definition 2.9]. In light of the main result of this paper, we can weaken the hypothesis of finite nuclear dimension in [loc. cit., Theorem 1.2] to that of nuclearity.

Corollary 6.4 (cf. [Gong and Lin 2017, Theorem 1.2]). *Let A and B be simple, separable, nuclear* C^**-algebras which satisfy the UCT. Then*

$$A\otimes\mathcal{Z}_0\cong B\otimes\mathcal{Z}_0 \quad \textit{if and only if} \quad \mathrm{Ell}(A\otimes\mathcal{Z}_0)\cong\mathrm{Ell}(B\otimes\mathcal{Z}_0).$$

Proof. Since A and $\mathcal{Z}_0$ are simple, separable and nuclear, so is $A\otimes\mathcal{Z}_0$. Using that $\mathcal{Z}_0$ is $\mathcal{Z}$-stable, it follows that $A\otimes\mathcal{Z}_0$ is $\mathcal{Z}$-stable. Therefore, $\dim_{\mathrm{nuc}} A\otimes\mathcal{Z}_0\leq 1$ by Theorem A. Similarly, $\dim_{\mathrm{nuc}} B\otimes\mathcal{Z}_0\leq 1$. The result now follows from [Gong and Lin 2017, Theorem 1.2]. □

7. Decomposition rank and $\mathcal{Z}$-stability

Using the machinery developed to prove Theorem A, we can also prove similar results for the decomposition rank of simple $\mathcal{Z}$-stable C^*-algebras under suitable finiteness and quasidiagonality assumptions. To this end, we recall the definition of quasidiagonality for tracial states.

Definition 7.1 (cf. [Brown 2006, Definition 3.3.1]). Let A be a C^*-algebra. A tracial state $\tau\in T(A)$ is *quasidiagonal* if there exists a net[11] of c.p.c. maps $\phi_n : A\to M_{k_n}(\mathbb{C})$ with $\|\phi_n(ab)-\phi_n(a)\phi_n(b)\|\to 0$ and $\mathrm{tr}_{k_n}(\phi_n(a))\to\tau(a)$.

In the unital case, the c.p.c. maps in Definition 7.1 can be taken to be unital (see the proof of [Brown and Ozawa 2008, Lemma 7.1.4]). Moreover, a trace $\tau\in T(A)$ is quasidiagonal if and only if its extension to $A^\sim$ is quasidiagonal [Brown 2006, Proposition 3.5.10]. We write $T_{QD}(A)$ for the set of all quasidiagonal tracial states on A.

We can now state a decomposition-rank version of Theorem A.

Theorem 7.2. *Let A be a simple, separable, nuclear and $\mathcal{Z}$-stable* C^**-algebra. Suppose further that A is stably finite and that $T(B)=T_{QD}(B)$ for all nonzero hereditary subalgebras $B\subseteq A\otimes\mathbb{K}$. Then* $\mathrm{dr}(A)\leq 1$.

Proof. By Theorem 2.7, either A is stably isomorphic to a unital C^*-algebra B, or A is stably isomorphic to a stably projectionless C^*-algebra A_0 with $Q\widetilde{T}(A_0)=\widetilde{T}_b(A_0)\neq 0$ and $T(A_0)$ compact.

In the first case, B is simple, separable, nuclear and $\mathcal{Z}$-stable by Proposition 2.3. Moreover, B is finite and $T(B)=T_{QD}(B)$ by our additional hypotheses on A. Hence $\mathrm{dr}(B)\leq 1$ by [Castillejos et al. 2019, Theorem B]. By Proposition 2.3 once more, $\mathrm{dr}(A)\leq 1$.

In the second case, we have that $T(A_0)=T_{QD}(A_0)$ by our additional hypotheses on A, so in the proof of Theorem 6.1 the maps θ_n from Lemma 3.2 can be taken to be approximately multiplicative. Therefore, $\mathrm{dr}(A_0)\leq 1$ by [Bosa et al. 2019b, Lemma 1.9]. Hence, $\mathrm{dr}(A)\leq 1$ by Proposition 2.3. □

Remark 7.3. If A is a simple, separable, nuclear C^*-algebra in the UCT class, then $T(B)=T(B)_{QD}$ for all hereditary subalgebras $B\subseteq A\otimes\mathbb{K}$ by [Tikuisis et al. 2017, Theorem A] since the UCT class is closed under stable isomorphism.

[11] When A is separable, one can work with sequences instead of general nets.

As with nuclear dimension, we obtain a trichotomy result for decomposition rank as a corollary of Theorem 7.2.

Corollary 7.4. *The decomposition rank of a simple* C^**-algebra is* 0, 1 *or* ∞.

Proof. Elementary C^*-algebras have decomposition rank zero and so are covered by this result.

Let A be a nonelementary, simple, separable C^*-algebra with finite decomposition rank. Then A has finite nuclear dimension, and so is $\mathcal{Z}$-stable by [Tikuisis 2014, Corollary 8.6]. Since $\mathrm{dr}(A) < \infty$, A is stably finite and $T(A) = T_{QD}(A)$.[12] Moreover, by Corollary 2.2 and Proposition 2.3, $\mathrm{dr}(B) = \mathrm{dr}(A) < \infty$ for any nonzero hereditary subalgebra $B \subseteq A \otimes \mathbb{K}$. Therefore, we have $T(B) = T_{QD}(B)$. Now, $\mathrm{dr}(A) \leq 1$ by Theorem 7.2.

The nonseparable case follows from the separable case as in the proof of [Castillejos et al. 2019, Corollary C] since the proof of [Winter and Zacharias 2010, Proposition 2.6] works equally well for decomposition rank. □

Appendix: Nonunital lemmas

The purpose of this appendix is to state appropriate nonunital versions of the technical lemmas from [Bosa et al. 2019a]. In cases where substantial modifications to the proof are required, we give full details. In cases where the modifications are trivial, we refer the reader to the proof of the corresponding result from [loc. cit.] and explain the modifications in a remark.

We begin with the existence of supporting order-zero maps.

Lemma A.1 (cf. [Bosa et al. 2019a, Lemma 1.14]). *Let* A, B_n *be* C^**-algebras with* A *separable and unital, set* $B_\omega := \prod_\omega B_n$*, and suppose that* $S \subseteq B_\omega$ *is separable and self-adjoint. Let* $\phi : A \to B_\omega \cap S'$ *be a c.p.c. order-zero map. Then there exists a c.p.c. order-zero map* $\hat{\phi} : A \to B_\omega \cap S'$ *such that*

$$\phi(ab) = \hat{\phi}(a)\phi(b) = \phi(a)\hat{\phi}(b), \quad a, b \in A. \tag{A-1}$$

Suppose now that $T(B_n)$ *is nonempty for all* $n \in \mathbb{N}$. *If the map* $\tau \mapsto d_\tau(\phi(1_A))$ *from* $\overline{T_\omega(B_\omega)}^{w*}$ *to* $[0, 1] \subseteq \mathbb{R}$ *is continuous* (*with respect to the weak***-topology*) *then we can, in addition, arrange that*

$$\tau(\hat{\phi}(a)) = \lim_{m\to\infty} \tau(\phi^{1/m}(a)), \quad a \in A_+,\ \tau \in T_\omega(B_\omega), \tag{A-2}$$

where order-zero map functional calculus is used to interpret $\phi^{1/m}$*. In this case, the induced map* $\bar{\hat{\phi}} : A \to B^\omega$ *is a* **-homomorphism.*

Remarks. The proof of [Bosa et al. 2019a, Lemma 1.14] only actually requires continuity of $\tau \mapsto d_\tau(\phi(1_A))$ on $\overline{T_\omega(B_\omega)}^{w*}$ (as opposed to $T(B_\omega)$) and $\overline{T_\omega(B_\omega)}^{w*}$ is compact in the nonunital case too. There is no further use of the unitality of the B_n in the proof of [loc. cit., Lemma 1.14]. □

[12] One can reduce to the unital case because $\mathrm{dr}(A) = \mathrm{dr}(A^\sim)$ [Kirchberg and Winter 2004, Proposition 3.4]. Then $T(A) = T_{QD}(A)$ by [Bosa et al. 2019a, Proposition 8.5]. Stably finiteness of A follows from [Kirchberg and Winter 2004, Proposition 5.1] and [Brown and Ozawa 2008, Theorem 7.1.15] for example.

We now record some more straightforward applications of the Kirchberg's epsilon test. These results are almost identical to those proven in [Bosa et al. 2019a, Section 1]. However, we shall need slightly more general statements because we wish to apply them to the algebras of the form $B_\omega \cap S' \cap \{1_{B_\omega^\sim} - d\}^\perp$.

Lemma A.2 (cf. [Bosa et al. 2019a, Lemma 1.16]). *Let $(B_n)_{n=1}^\infty$ be a sequence of* C*-*algebras and set $B_\omega := \prod_\omega B_n$. Let S_1, S_2 be separable self-adjoint subsets of $B_\omega^\sim$, and let T be a separable subset of $B_\omega \cap S_1' \cap S_2^\perp$. Then there exists a contraction $e \in (B_\omega \cap S_1' \cap S_2^\perp)_+$ that acts as a unit on T, i.e., such that $et = te = t$ for every $t \in T$.*

Remarks. The only change to the statement is that S_1, S_2 are subsets of $B_\omega^\sim$ (as opposed to B_ω). The proof is not affected. □

Lemma A.3 (cf. [Bosa et al. 2019a, Lemma 1.17]). *Let $(B_n)_{n=1}^\infty$ be a sequence of* C*-*algebras and set $B_\omega := \prod_\omega B_n$. Let S_1, S_2 be separable self-adjoint subsets of $B_\omega^\sim$, and set $C := B_\omega \cap S_1' \cap S_2^\perp$.*

(i) *Let $h_1, h_2 \in C_+$. Then h_1 and h_2 are unitarily equivalent via a unitary from $C^\sim$ if and only if they are approximately unitarily equivalent*; *i.e., for any $\epsilon > 0$ there exists a unitary $u \in C^\sim$ with $uh_1u^* \approx_\epsilon h_2$.*

(ii) *Let $a \in C$. Then there exists a unitary $u \in C^\sim$ with $a = u|a|$ if and only if for each $\epsilon > 0$ there exists a unitary $u \in C^\sim$ with $a \approx_\epsilon u|a|$.*

(iii) *Let $h_1, h_2 \in C_+$. Then h_1 and h_2 are Murray–von Neumann equivalent if and only if they are approximately Murray–von Neumann equivalent*; *i.e., for any $\epsilon > 0$ there exists $x \in C$ with $xx^* \approx_\epsilon h_1$ and $x^*x \approx_\epsilon h_2$.*

Remarks. The statement of [Bosa et al. 2019a, Lemma 1.17] uses the convention that $C^\sim := C$ when C is already unital. In this paper, we use the convention that a new unit is still adjoined, so $C^\sim \cong C \oplus \mathbb{C}$ when C is unital. The choice of convention does not affect the validity of the lemma.[13] Apart from this, the only change to the statement is that S_1, S_2 are subsets of $B_\omega^\sim$ (as opposed to B_ω), which does not affect the proof. □

Lemma A.4 (cf. [Bosa et al. 2019a, Lemma 1.18]). *Let $(B_n)_{n=1}^\infty$ be a sequence of* C*-*algebras with $T(B_n)$ nonempty for each $n \in \mathbb{N}$. Write $B_\omega := \prod_\omega B_n$. Let S_0 be a countable self-adjoint subset of $(B_\omega)_+$ and let T be a separable self-adjoint subset of B_ω. If $x, f \in (B_\omega \cap S_0' \cap T')_+$ are contractions with $x \lhd f$ and with the property that for all $a \in S_0$ there exists $\gamma_a \geq 0$ such that $\tau(af^m) \geq \gamma_a$ for all $m \in \mathbb{N}$, $\tau \in T_\omega(B_\omega)$, then there exists a contraction $f' \in (B_\omega \cap S_0' \cap T')_+$ such that $x \lhd f' \lhd f$ and $\tau(a(f')^m) \geq \gamma_a$ for all $m \in \mathbb{N}$, $\tau \in T_\omega(B_\omega)$, and $a \in S_0$.*

If each B_n is simple, separable, $\mathcal{Z}$-stable and $Q\widetilde{T}(B_n) = \widetilde{T}_b(B_n) \neq 0$ for all $n \in \mathbb{N}$, then the above statement holds with $T(B_\omega)$ in place of $T_\omega(B_\omega)$.

Remarks. The only change to the proof of [Bosa et al. 2019a, Lemma 1.18] is to replace $\min_{\tau \in T(B_n)}$ with $\inf_{\tau \in T(B_n)}$ in [loc. cit., equation (1.34)], as the minimum need not exist in the nonunital case. The final sentence follows since $T_\omega(B_\omega)$ is weak*-dense in $T_\omega(B_\omega)$ under the additional hypotheses by Theorem 1.7. □

[13] For example, if C is unital, h_1, h_2 are unitary equivalent in C if and only if they are unitary equivalent in $C \oplus \mathbb{C}$.

Lemma A.5 (cf. [Bosa et al. 2019a, Lemma 1.19]). *Let $(B_n)_{n=1}^{\infty}$ be a sequence of separable* C*-*algebras with $T(B_n) \neq \varnothing$ for each $n \in \mathbb{N}$ and set $B_\omega := \prod_\omega B_n$. Let A be a separable, unital* C*-*algebra and let $\pi : A \to B_\omega$ be a c.p.c. order-zero map such that $\pi(1_A)$ is full and the induced map $\bar{\pi} : A \to B^\omega$ is a *-homomorphism. Define $C := B_\omega \cap \pi(A)' \cap \{1_{B_\omega^\sim} - \pi(1_A)\}^\perp$. Let $S \subseteq C$ be a countable self-adjoint subset and let $\overline{S}$ denote the image of S in B^ω*:

(i) *Then the image of $C \cap S'$ in B^ω is precisely*
$$\bar{\pi}(1_A)(B^\omega \cap \bar{\pi}(A)' \cap \bar{S}') = B^\omega \cap \bar{\pi}(A)' \cap \bar{S}' \cap \{1_{(B^\omega)^\sim} - \bar{\pi}(1_A)\}^\perp, \tag{A-3}$$
a C*-*subalgebra of B^ω with unit $\bar{\pi}(1_A)$.*

(ii) *Let $\tau \in T_\omega(B_\omega)$ be a limit trace and $a \in A_+$ and form the tracial functional $\rho := \tau(\pi(a)\cdot)$ on C. Then $\|\rho\| = \tau(\pi(a))$. If each B_n is additionally simple, $\mathcal{Z}$-stable and $Q\widetilde{T}(B_n) = \widetilde{T}_b(B_n) \neq 0$ for all $n \in \mathbb{N}$, then this holds for all traces $\tau \in T(B_\omega)$.*

Remarks. The proof of [Bosa et al. 2019a, Lemma 1.19] does not need the B_n to be unital. Note that the notation $\{1_{B_\omega^\sim} - \pi(1_A)\}^\perp$ is just an alternative notation for subalgebra on which $\pi(1_A)$ acts as a unit, and similarly for $\{1_{(B^\omega)^\sim} - \bar{\pi}(1_A)\}^\perp$. The final sentence of (ii) follows since $T_\omega(B_\omega)$ is weak*-dense in $T_\omega(B_\omega)$ under the additional hypotheses by Theorem 1.7. □

Next, we consider some properties of ultraproducts of separable, $\mathcal{Z}$-stable C*-algebras.

Lemma A.6 (cf. [Bosa et al. 2019a, Lemma 1.22]). *Let $(B_n)_{n\in\mathbb{N}}$ be a sequence of separable, $\mathcal{Z}$-stable* C*-*algebras and set $B_\omega := \prod_\omega B_n$. Then*:

(i) *If $S \subseteq B_\omega$ is separable, then there exist isomorphisms $\phi_n : B_n \to B_n \otimes \mathcal{Z}$ such that the induced isomorphism $\Phi : B_\omega \to \prod_\omega (B_n \otimes \mathcal{Z})$ maps $x \in S$ to $x \otimes 1_{\mathcal{Z}} \in \left(\prod_\omega B_n\right) \otimes \mathcal{Z} \subseteq \prod_\omega (B_n \otimes \mathcal{Z})$.*

(ii) *Let $S_1, S_2 \subseteq B_\omega^\sim$ be separable and self-adjoint. For any separable subset $T \subseteq B_\omega \cap S_1' \cap S_2^\perp$, there exists a c.p.c. order-zero map $\psi : \mathcal{Z} \to B_\omega \cap S_1' \cap S_2^\perp \cap T'$ such that $\psi(1_{\mathcal{Z}})$ acts as a unit on T.*

(ii′) *Let $S_1, S_2 \subseteq B_\omega^\sim$ be separable and self-adjoint. For any separable subalgebra $C \subseteq B_\omega \cap S_1' \cap S_2^\perp$, there exists a *-homomorphism $\Psi : C \otimes \mathcal{Z} \to B_\omega \cap S_1' \cap S_2^\perp$ such that $\Psi(x \otimes 1_{\mathcal{Z}}) = x$ for all $x \in C$.*

(iii) *If each B_n is projectionless, then B_ω has stable rank one in $B_\omega^\sim$.*

(iv) *If $S \subseteq B_\omega$ is separable and self-adjoint, and $b \in (B_\omega \cap S')_+$, then for any $n \in \mathbb{N}$ there exists $c \in (B_\omega \cap S')_+$ with $c \leq b$ such that $n[c] \leq [b] \leq (n+1)[c]$ in $W(B_\omega \cap S')$.*

Proof. Observe that (i) is the same as in [Bosa et al. 2019a, Lemma 1.22(i)] and follows as $\mathcal{Z}$ is strongly self-absorbing.

For (ii), by Lemma A.2, there exists a positive contraction $h \in B_\omega \cap S_1' \cap S_2^\perp$ that acts as a unit on T. Set $S := S_1 \cup S_2 \cup T \cup \{h\}$. Let $\Phi : B_\omega \to \prod_\omega (B_n \otimes \mathcal{Z})$ be the isomorphism from (i) with $\Phi(x) = x \otimes 1_{\mathcal{Z}}$ for all $x \in S$. Define a c.p.c. order-zero map $\psi' : \mathcal{Z} \to \prod_\omega (B_n \otimes \mathcal{Z})$ by $\psi'(z) := h \otimes z$. By the choice of h, $\psi(1_{\mathcal{Z}})$ acts as a unit on $T \otimes 1_{\mathcal{Z}}$ and the image of ψ lies in $(B_\omega \otimes 1_{\mathcal{Z}}) \cap (S_1 \otimes 1_{\mathcal{Z}})' \cap (S_2 \otimes 1_{\mathcal{Z}})^\perp \cap (T \otimes 1_{\mathcal{Z}})'$. Now set $\psi := \Phi^{-1} \circ \psi'$.

For (ii′), by part (ii), there exists a c.p.c. order-zero map $\psi : \mathcal{Z} \to B_\omega \cap S_1' \cap S_2^\perp \cap T'$ such that $\psi(1_{\mathcal{Z}})$ acts as a unit on C. Since $\mathcal{Z}$ is nuclear, we can define a c.p.c. order-zero map $\Psi : C \otimes \mathcal{Z} \to B_\omega \cap S_1' \cap S_2^\perp$ by $x \otimes z \mapsto x\psi(z)$. Let $x_1, x_2 \in C$ and $z_1, z_2 \in \mathcal{Z}$. Then

$$\begin{aligned}\Psi(x_1 \otimes z_1)\Psi(x_1 \otimes z_1) &= x_1\psi(z_1)x_2\psi(z_2)\\ &= x_1x_2\psi(1_{\mathcal{Z}})\psi(z_1z_2)\\ &= x_1x_2\psi(z_1z_2)\\ &= \Psi(x_1x_2 \otimes z_1z_2), \qquad \text{(A-4)}\end{aligned}$$

where we have used the order-zero identity in the second line. Hence, Ψ is in fact a *-homomorphism. Moreover, we have $\Psi(x \otimes 1_{\mathcal{Z}}) = x\psi(1_{\mathcal{Z}}) = x$ for all $x \in C$.

Part (iii) follows by combining Theorem 1.10 with Proposition 1.12.

For (iv), let C be the C*-algebra generated by b. By (ii′), there is a *-homomorphism $\Psi : C \otimes \mathcal{Z} \to B_\omega \cap S'$. By [Rørdam 2004, Lemma 4.2], there exists $e_n \in \mathcal{Z}_{+,1}$ with $n[e_n] \le [1_{\mathcal{Z}}] \le (n+1)[e_n]$. Set $c := \Psi(b \otimes e_n)$. □

The following lemmas are crucial to the results of [Bosa et al. 2019a, Section 5]. The proof of the first needs to be adapted slightly to the nonunital setting.

Lemma A.7 (cf. [Bosa et al. 2019a, Lemma 2.1]). *Let $(B_n)_{n=1}^\infty$ be a sequence of $\mathcal{Z}$-stable C*-algebras and set $B_\omega := \prod_\omega B_n$. Let $S \subseteq B_\omega$ be separable and self-adjoint, and let $d \in (B_\omega \cap S')_+$ be a contraction. Suppose that $x, f \in C := B_\omega \cap S' \cap \{1_{B_\omega^\sim} - d\}^\perp$ are such that $xf = fx = 0$, $f \ge 0$ and f is full in C. Then x is approximated by invertibles in $C^\sim$.*

Proof. By Lemma A.2 (with $T := \{x^*x, xx^*\}$, $S_1 := S$, $S_2 := \{1_{B_\omega^\sim} - d, f\}$), we obtain a contraction $e \in C_+$ such that $xx^*, x^*x \lhd e$ and $ef = 0$. Polar decomposition yields $ex = xe = x$. As in the proof of [Bosa et al. 2019a, Lemma 2.1], we may find a separable subalgebra C_0 of C containing x, e, and f such that f is full in C_0. By Lemma A.6(ii′), there is a *-homomorphism $\Psi : C_0 \otimes \mathcal{Z} \to C$ such that $\Psi(x \otimes 1_{\mathcal{Z}}) = x$ for all $x \in C_0$. By [Robert 2016, Lemma 2.1], $x \otimes 1_{\mathcal{Z}}$ is a product of two nilpotent elements $n_1, n_2 \in C_0 \otimes \mathcal{Z}$. It follows that $x = \Psi(x \otimes 1_{\mathcal{Z}}) = \Psi(n_1)\Psi(n_2)$ is the product of two nilpotent elements in C. If $y \in C$ is nilpotent and $\epsilon > 0$, the operator $y + \epsilon 1_{C^\sim}$ is invertible in $C^\sim$ (with inverse $-\sum_{k=1}^N (-\epsilon)^{-k} y^{k-1}$, where $N \in \mathbb{N}$ satisfies $y^N = 0$). Therefore, x can be approximated by invertible elements in $C^\sim$. □

Lemma A.8 (cf. [Bosa et al. 2019a, Lemma 2.2]). *Let $(B_n)_{n=1}^\infty$ be a sequence of $\mathcal{Z}$-stable C*-algebras and set $B_\omega := \prod_\omega B_n$. Let $S \subseteq B_\omega$ be separable and self-adjoint, and let $d \in (B_\omega \cap S')_+$ be a contraction. Suppose that $x, s \in C := B_\omega \cap S' \cap \{1_{B_\omega^\sim} - d\}^\perp$ are such that $xs = sx = 0$ and s is full in C. Then x is approximated by invertibles in $C^\sim$.*

Proof. Let C_0 be the C*-subalgebra of C generated by x and s. By Lemma A.6(ii′), there exists a *-homomorphism $\Psi : C_0 \otimes \mathcal{Z} \to C$ with $\Psi(y \otimes 1_{\mathcal{Z}}) = y$ for all $y \in C_0$. Let $z_1, z_2 \in \mathcal{Z}_+$ be nonzero orthogonal elements. Set $s' := \Psi(s \otimes z_1)$ and $f := \Psi(|s| \otimes z_2)$. As in the proof of [Bosa et al. 2019a, Lemma 2.2], it follows by Lemma A.7 (with s' in place of x) that s' is approximated by invertibles in $C^\sim$. We finish the proof exactly as in the proof of [loc. cit., Lemma 2.2], where we replace [loc. cit., Lemma 1.17] with Lemma A.3, and [loc. cit., Lemma 2.1] with Lemma A.7. □

Lemma A.9 (cf. [Bosa et al. 2019a, Lemma 5.4] and [Robert and Santiago 2010, Lemma 2]). *Let B be a separable, $\mathcal{Z}$-stable* C*-*algebra and let A be a separable, unital* C*-*algebra. Let $\pi : A \to B_\omega$ be a c.p.c. order-zero map such that*

$$C := B_\omega \cap \pi(A)' \cap \{1_{B_\omega^\sim} - \pi(1_A)\}^\perp \tag{A-5}$$

is full in B_ω. Assume that every full hereditary subalgebra D of C satisfies the following: if $x \in D$ is such that there exist totally full elements $e_l, e_r \in D_+$ such that $e_l x = x e_r = 0$, then there exists a full element $s \in D$ such that $sx = xs = 0$. Let $e, f, f', \alpha, \beta \in C_+$ be such that

$$\alpha \lhd e, \quad \alpha \sim \beta \lhd f, \quad \textit{and} \quad f \sim f' \lhd e. \tag{A-6}$$

Suppose also that there exist $d_e, d_f \in C_+$ that are totally full such that

$$\begin{aligned} d_e \lhd e, &\quad d_e \alpha = 0, \\ d_f \lhd f, &\quad d_f \beta = 0. \end{aligned} \tag{A-7}$$

Then there exists $e' \in C_+$ such that

$$\alpha \lhd e' \lhd e \quad \textit{and} \quad \alpha + e' \sim \beta + f. \tag{A-8}$$

Remarks. The proof of [Robert and Santiago 2010, Lemma 2] does not assume unitality. The proof from [Bosa et al. 2019a] is still valid after replacing [loc. cit., Lemma 1.17] with Lemma A.3, [loc. cit., Lemma 2.2] with Lemma A.8, and using $1_{B_\omega^\sim}$ in place of 1_{B_ω}. □

The following lemma concerns the interplay between strict comparison and ultraproducts in the nonunital setting.

Lemma A.10 (cf. [Bosa et al. 2019a, Lemma 1.23]). *Let $(B_n)_{n=1}^\infty$ be a sequence of* C*-*algebras with $T(B_n)$ nonempty and set $B_\omega := \prod_\omega B_n$. Suppose each B_n has strict comparison of positive elements with respect to bounded traces. Then B_ω has strict comparison of positive elements with respect to limit traces, in the following sense: if $a, b \in M_k(B_\omega)_+$ for some $k \in \mathbb{N}$ satisfy $d_\tau(a) < d_\tau(b)$ for all τ in the weak*-closure of $T_\omega(B_\omega)$, then $a \preceq b$.*

Remarks. The proof is identical to that of [Bosa et al. 2019a, Lemma 1.23]. However, it is important to note that, in the nonunital case, we do not necessarily have that $\overline{T_\omega(B_\omega)}^{w*} \subseteq T(B_\omega)$, as the later need not be closed. Indeed, we may have $0 \in \overline{T_\omega(B_\omega)}^{w*}$ in which case $d_\tau(a) < d_\tau(b)$ cannot hold for all $\tau \in \overline{T_\omega(B_\omega)}^{w*}$. □

Finally, we record a technical lemma needed for the proof of the main theorem of the property (SI) section.

Lemma A.11 (cf. [Bosa et al. 2019a, Lemma 4.9]). *Let B be a simple, separable,* C*-*algebra with $Q\widetilde{T}(B) = \widetilde{T}_b(B) \neq 0$. Suppose B has strict comparison of positive elements by bounded traces. Let A be a separable, unital* C*-*algebra and let $\pi : A \to B_\omega$ be a c.p.c. order-zero map. Let $a \in A_+$ be a positive* *contraction of norm* 1. *Then there exists a countable set* $S \subseteq A_+ \setminus \{0\}$ *such that the following holds*: *if* $e, t, h \in (B_\omega \cap \pi(A)' \cap \{1_{B_\omega^\sim} - \pi(1_A)\}^\perp)_+$ *are contractions such that*

$$e \in J_{B_\omega} \quad \textit{and} \quad h \lhd t, \tag{A-9}$$

and if for all $b \in S$, *there exists* $\gamma_b > 0$ *such that*

$$\tau(\pi(b)h) > \gamma_b, \quad \tau \in T_\omega(B_\omega), \tag{A-10}$$

then there exists a contraction $r \in B_\omega$ *such that*

$$\pi(a)r = tr = r \quad \textit{and} \quad r^*r = e. \tag{A-11}$$

Remarks. The proof of [Bosa et al. 2019a, Lemma 4.9] works in our situation using Lemma A.10 in place of [loc. cit., Lemma 1.23]. □

Acknowledgements

Part of this work was undertaken during a visit of Castillejos to IMPAN. Castillejos thanks Evington and IMPAN for their hospitality. Evington would like to thank George Elliott for his helpful comments on this research during Evington's secondment at the Fields Institute, which was supported by the EU RISE Network *Quantum Dynamics*. The authors would also like to thank Jamie Gabe, Gábor Szabó, Stefaan Vaes and Stuart White for useful comments on an earlier version of this paper.

References

[Alfsen 1971] E. M. Alfsen, *Compact convex sets and boundary integrals*, Ergebnisse der Mathematik (3) **57**, Springer, 1971. MR Zbl

[Ara et al. 2011] P. Ara, F. Perera, and A. S. Toms, "K-theory for operator algebras: classification of C^*-algebras", pp. 1–71 in *Aspects of operator algebras and applications*, edited by P. Ara et al., Contemp. Math. **534**, Amer. Math. Soc., Providence, RI, 2011. MR Zbl

[Blackadar 1980] B. E. Blackadar, "A simple C^*-algebra with no nontrivial projections", *Proc. Amer. Math. Soc.* **78**:4 (1980), 504–508. MR Zbl

[Blackadar 2006] B. Blackadar, *Operator algebras: theory of C^*-algebras and von Neumann algebras, operator algebras and non-commutative geometry, III*, Encycl. Math. Sci. **122**, Springer, 2006. MR Zbl

[Blackadar and Cuntz 1982] B. E. Blackadar and J. Cuntz, "The structure of stable algebraically simple C^*-algebras", *Amer. J. Math.* **104**:4 (1982), 813–822. MR Zbl

[Blackadar and Handelman 1982] B. Blackadar and D. Handelman, "Dimension functions and traces on C^*-algebras", *J. Funct. Anal.* **45**:3 (1982), 297–340. MR Zbl

[Blanchard and Kirchberg 2004] E. Blanchard and E. Kirchberg, "Non-simple purely infinite C^*-algebras: the Hausdorff case", *J. Funct. Anal.* **207**:2 (2004), 461–513. MR Zbl

[Bosa et al. 2019a] J. Bosa, N. P. Brown, Y. Sato, A. Tikuisis, S. White, and W. Winter, *Covering dimension of C^*-algebras and 2-coloured classification*, Mem. Amer. Math. Soc. **1233**, Amer. Math. Soc., Providence, RI, 2019. MR Zbl

[Bosa et al. 2019b] J. Bosa, J. Gabe, A. Sims, and S. White, "The nuclear dimension of $\mathcal{O}_\infty$-stable C^*-algebras", preprint, 2019. arXiv

[Brown 1977] L. G. Brown, "Stable isomorphism of hereditary subalgebras of C^*-algebras", *Pacific J. Math.* **71**:2 (1977), 335–348. MR Zbl

[Brown 2006] N. P. Brown, *Invariant means and finite representation theory of C^*-algebras*, Mem. Amer. Math. Soc. **865**, 2006. MR Zbl

[Brown and Ozawa 2008] N. P. Brown and N. Ozawa, *C^*-algebras and finite-dimensional approximations*, Grad. Studies in Math. **88**, Amer. Math. Soc., Providence, RI, 2008. MR Zbl

[Brown et al. 2017] N. P. Brown, J. R. Carrión, and S. White, "Decomposable approximations revisited", pp. 45–65 in *Operator algebras and applications*, edited by T. M. Carlsen et al., Abel Symp. **12**, Springer, 2017. MR Zbl

[Castillejos 2017] J. Castillejos, "Decomposable approximations and approximately finite dimensional C^*-algebras", *Math. Proc. Cambridge Philos. Soc.* **162**:1 (2017), 1–12. MR Zbl

[Castillejos et al. 2019] J. Castillejos, S. Evington, A. Tikuisis, S. White, and W. Winter, "Nuclear dimension of simple C^*-algebras", preprint, 2019. To appear in *Invent. Math.* arXiv

[Conley et al. 2018] C. T. Conley, S. C. Jackson, D. Kerr, A. S. Marks, B. Seward, and R. D. Tucker-Drob, "Følner tilings for actions of amenable groups", *Math. Ann.* **371**:1-2 (2018), 663–683. MR Zbl

[Coward et al. 2008] K. T. Coward, G. A. Elliott, and C. Ivanescu, "The Cuntz semigroup as an invariant for C^*-algebras", *J. Reine Angew. Math.* **623** (2008), 161–193. MR Zbl

[Cuntz and Pedersen 1979] J. Cuntz and G. K. Pedersen, "Equivalence and traces on C^*-algebras", *J. Funct. Anal.* **33**:2 (1979), 135–164. MR Zbl

[Elliott 1996] G. A. Elliott, "An invariant for simple C^*-algebras", pp. 61–90 in *Canadian Mathematical Society: 1945–1995, III*, edited by J. B. Carrell and R. Murty, Canadian Math. Soc., Ottawa, 1996. MR Zbl

[Elliott et al. 2011] G. A. Elliott, L. Robert, and L. Santiago, "The cone of lower semicontinuous traces on a C^*-algebra", *Amer. J. Math.* **133**:4 (2011), 969–1005. MR Zbl

[Elliott et al. 2015] G. A. Elliott, G. Gong, H. Lin, and Z. Niu, "On the classification of simple amenable C^*-algebras with finite decomposition rank, II", preprint, 2015. arXiv

[Elliott et al. 2017] G. A. Elliott, G. Gong, H. Lin, and Z. Niu, "The classification of simple separable KK-contractible C^*-algebras with finite nuclear dimension", preprint, 2017. arXiv

[Elliott et al. 2020] G. A. Elliott, G. Gong, H. Lin, and Z. Niu, "Simple stably projectionless C*-algebras with generalized tracial rank one", *J. Noncommut. Geom.* **14**:1 (2020), 251–347. MR

[Farah and Katsura 2010] I. Farah and T. Katsura, "Nonseparable UHF algebras, I: Dixmier's problem", *Adv. Math.* **225**:3 (2010), 1399–1430. MR Zbl

[Giol and Kerr 2010] J. Giol and D. Kerr, "Subshifts and perforation", *J. Reine Angew. Math.* **639** (2010), 107–119. MR Zbl

[Gong and Lin 2016] G. Gong and H. Lin, "On classification of non-unital simple amenable C^*-algebras, I", preprint, 2016. arXiv

[Gong and Lin 2017] G. Gong and H. Lin, "On classification of simple non-unital amenable C^*-algebras, II", preprint, 2017. arXiv

[Gong et al. 2015] G. Gong, H. Lin, and Z. Niu, "Classification of finite simple amenable $\mathfrak{Z}$-stable C^*-algebras", preprint, 2015. arXiv

[Haagerup 2014] U. Haagerup, "Quasitraces on exact C^*-algebras are traces", *C. R. Math. Acad. Sci. Soc. R. Can.* **36**:2-3 (2014), 67–92. MR Zbl

[Jacelon 2013] B. Jacelon, "A simple, monotracial, stably projectionless C^*-algebra", *J. Lond. Math. Soc.* (2) **87**:2 (2013), 365–383. MR Zbl

[Jiang and Su 1999] X. Jiang and H. Su, "On a simple unital projectionless C^*-algebra", *Amer. J. Math.* **121**:2 (1999), 359–413. MR Zbl

[Kerr and Szabó 2020] D. Kerr and G. Szabó, "Almost finiteness and the small boundary property", *Comm. Math. Phys.* **374**:1 (2020), 1–31. MR

[Kirchberg 1995] E. Kirchberg, "Exact C^*-algebras, tensor products, and the classification of purely infinite algebras", pp. 943–954 in *Proceedings of the International Congress of Mathematicians, II* (Zürich, 1994), edited by S. D. Chatterji, Birkhäuser, Basel, 1995. MR Zbl

[Kirchberg 2006] E. Kirchberg, "Central sequences in C^*-algebras and strongly purely infinite algebras", pp. 175–231 in *Operator algebras: the Abel symposium* (Oslo, 2004), edited by S. Neshveyev and C. Skau, Abel Symp. **1**, Springer, 2006. MR Zbl

[Kirchberg and Rørdam 2000] E. Kirchberg and M. Rørdam, "Non-simple purely infinite C^*-algebras", *Amer. J. Math.* **122**:3 (2000), 637–666. MR Zbl

[Kirchberg and Winter 2004] E. Kirchberg and W. Winter, "Covering dimension and quasidiagonality", *Int. J. Math.* **15**:1 (2004), 63–85. MR Zbl

[Kishimoto and Kumjian 1996] A. Kishimoto and A. Kumjian, "Simple stably projectionless C^*-algebras arising as crossed products", *Canad. J. Math.* **48**:5 (1996), 980–996. MR Zbl

[Matui and Sato 2012] H. Matui and Y. Sato, "$\mathcal{Z}$-stability of crossed products by strongly outer actions", *Comm. Math. Phys.* **314**:1 (2012), 193–228. MR Zbl

[McDuff 1970] D. McDuff, "Central sequences and the hyperfinite factor", *Proc. Lond. Math. Soc.* (3) **21** (1970), 443–461. MR Zbl

[Murray and von Neumann 1943] F. J. Murray and J. von Neumann, "On rings of operators, IV", *Ann. of Math.* (2) **44** (1943), 716–808. MR Zbl

[Ng and Robert 2016] P. W. Ng and L. Robert, "Sums of commutators in pure C^*-algebras", *Münster J. Math.* **9**:1 (2016), 121–154. MR Zbl

[Ozawa 2013] N. Ozawa, "Dixmier approximation and symmetric amenability for C^*-algebras", *J. Math. Sci. Univ. Tokyo* **20**:3 (2013), 349–374. MR Zbl

[Pedersen 1969] G. K. Pedersen, "Measure theory for C^* algebras, III", *Math. Scand.* **25** (1969), 71–93. Zbl

[Pedersen 1979] G. K. Pedersen, *C^*-algebras and their automorphism groups*, Lond. Math. Soc. Monogr. **14**, Academic Press, London, 1979. MR Zbl

[Phillips 2000] N. C. Phillips, "A classification theorem for nuclear purely infinite simple C^*-algebras", *Doc. Math.* **5** (2000), 49–114. MR Zbl

[Razak 2002] S. Razak, "On the classification of simple stably projectionless C^*-algebras", *Canad. J. Math.* **54**:1 (2002), 138–224. MR Zbl

[Robert 2016] L. Robert, "Remarks on $\mathcal{Z}$-stable projectionless C^*-algebras", *Glasg. Math. J.* **58**:2 (2016), 273–277. MR Zbl

[Robert and Santiago 2010] L. Robert and L. Santiago, "Classification of C^*-homomorphisms from $C_0(0, 1]$ to a C^*-algebra", *J. Funct. Anal.* **258**:3 (2010), 869–892. MR Zbl

[Rørdam 2003] M. Rørdam, "A simple C^*-algebra with a finite and an infinite projection", *Acta Math.* **191**:1 (2003), 109–142. MR Zbl

[Rørdam 2004] M. Rørdam, "The stable and the real rank of $\mathcal{Z}$-absorbing C^*-algebras", *Int. J. Math.* **15**:10 (2004), 1065–1084. MR Zbl

[Rosenberg and Schochet 1987] J. Rosenberg and C. Schochet, "The Künneth theorem and the universal coefficient theorem for Kasparov's generalized K-functor", *Duke Math. J.* **55**:2 (1987), 431–474. MR Zbl

[Tikuisis 2014] A. Tikuisis, "Nuclear dimension, $\mathcal{Z}$-stability, and algebraic simplicity for stably projectionless C^*-algebras", *Math. Ann.* **358**:3-4 (2014), 729–778. MR Zbl

[Tikuisis and Toms 2015] A. P. Tikuisis and A. Toms, "On the structure of Cuntz semigroups in (possibly) nonunital C^*-algebras", *Canad. Math. Bull.* **58**:2 (2015), 402–414. MR Zbl

[Tikuisis and Winter 2014] A. Tikuisis and W. Winter, "Decomposition rank of $\mathcal{Z}$-stable C^*-algebras", *Anal. PDE* **7**:3 (2014), 673–700. MR Zbl

[Tikuisis et al. 2017] A. Tikuisis, S. White, and W. Winter, "Quasidiagonality of nuclear C^*-algebras", *Ann. of Math.* (2) **185**:1 (2017), 229–284. MR Zbl

[Toms 2008] A. S. Toms, "On the classification problem for nuclear C^*-algebras", *Ann. of Math.* (2) **167**:3 (2008), 1029–1044. MR Zbl

[Toms and Winter 2007] A. S. Toms and W. Winter, "Strongly self-absorbing C^*-algebras", *Trans. Amer. Math. Soc.* **359**:8 (2007), 3999–4029. MR Zbl

[Toms and Winter 2009] A. S. Toms and W. Winter, "The Elliott conjecture for Villadsen algebras of the first type", *J. Funct. Anal.* **256**:5 (2009), 1311–1340. MR Zbl

[Tsang 2005] K.-W. Tsang, "On the positive tracial cones of simple stably projectionless C^*-algebras", *J. Funct. Anal.* **227**:1 (2005), 188–199. MR Zbl

[Villadsen 1999] J. Villadsen, "On the stable rank of simple C^*-algebras", *J. Amer. Math. Soc.* **12**:4 (1999), 1091–1102. MR Zbl

[Winter 2009] W. Winter, "Covering dimension for nuclear C^*-algebras, II", *Trans. Amer. Math. Soc.* **361**:8 (2009), 4143–4167. MR Zbl

[Winter 2012] W. Winter, "Nuclear dimension and $\mathcal{Z}$-stability of pure C^*-algebras", *Invent. Math.* **187**:2 (2012), 259–342. MR Zbl

[Winter 2014] W. Winter, "Localizing the Elliott conjecture at strongly self-absorbing C^*-algebras", *J. Reine Angew. Math.* **692** (2014), 193–231. MR Zbl

[Winter and Zacharias 2009] W. Winter and J. Zacharias, "Completely positive maps of order zero", *Münster J. Math.* **2** (2009), 311–324. MR Zbl

[Winter and Zacharias 2010] W. Winter and J. Zacharias, "The nuclear dimension of C^*-algebras", *Adv. Math.* **224**:2 (2010), 461–498. MR Zbl

Received 4 Mar 2019. Revised 9 Jun 2019. Accepted 13 Aug 2019.

JORGE CASTILLEJOS: jorge.castillejoslopez@kuleuven.be
Department of Mathematics, KU Leuven, Leuven, Belgium

SAMUEL EVINGTON: samuel.evington@glasgow.ac.uk
School of Mathematics and Statistics, University of Glasgow, Glasgow, United Kingdom

and

Institute of Mathematics, Polish Academy of Sciences, Warsaw, Poland

ANALYSIS AND PDE
Vol. 13, No. 7, 2020
https://doi.org/10.2140/apde.2020.13.2241

msp

ON THE REGULARITY OF MINIMIZERS FOR SCALAR INTEGRAL FUNCTIONALS WITH (p, q)-GROWTH

PETER BELLA AND MATHIAS SCHÄFFNER

We revisit the question of regularity for minimizers of scalar autonomous integral functionals with so-called (p, q)-growth. In particular, we establish Lipschitz regularity under the condition $\frac{q}{p} < 1 + \frac{2}{n-1}$ for $n \geq 3$, improving a classical result due to Marcellini (*J. Differential Equations* **90**:1 (1991), 1–30).

1. Introduction and main results

In this note, we consider the problem of regularity for local minimizers of

$$\mathcal{F}[u] := \int_\Omega f(\nabla u)\,dx, \tag{1}$$

where $\Omega \subset \mathbb{R}^n$, $n \geq 2$, is a bounded domain and $f : \mathbb{R}^n \to \mathbb{R}$ is a sufficiently smooth integrand satisfying (p, q)-growth of the form:

Assumption 1. There exist $0 < m \leq M < \infty$ such that $f \in C^2(\mathbb{R}^n)$ satisfies for all $z, \lambda \in \mathbb{R}^n$

$$\begin{cases} m|z|^p \leq f(z) \leq M(1 + |z|^q), \\ m(1 + |z|^2)^{\frac{p-2}{2}}|\lambda|^2 \leq \langle D^2 f(z)\lambda, \lambda\rangle \leq M(1 + |z|^2)^{\frac{q-2}{2}}|\lambda|^2. \end{cases} \tag{2}$$

Regularity properties of local minimizers of (1) in the case $p = q$ are classical; see, e.g., [Giusti 2003]. A systematic regularity theory in the case $p < q$ was initiated in [Marcellini 1989; 1991]. In particular, Marcellini [1991] proved:

(A) If $2 \leq p < q$ and $\frac{q}{p} < 1 + \frac{2}{n-2}$ if $n \geq 3$, then every local minimizer $u \in W^{1,q}_{\mathrm{loc}}(\Omega)$ of (1) satisfies $u \in W^{1,\infty}_{\mathrm{loc}}(\Omega)$.

(B) If $2 \leq p < q$ and $\frac{q}{p} < 1 + \frac{2}{n}$, then every local minimizer $u \in W^{1,p}_{\mathrm{loc}}(\Omega)$ of (1) satisfies $u \in W^{1,\infty}_{\mathrm{loc}}(\Omega)$.

We emphasize that establishing Lipschitz-regularity is the crucial point in the regularity theory for functionals with (p, q)-growth in the form (2). Indeed, local boundedness of the gradient implies that the nonstandard growth of f and $D^2 f$ becomes irrelevant and higher regularity (depending on the smoothness of f) follows by standard arguments; see, e.g., [Marcellini 1989, Chapter 7] and Corollary 7 below.

By now there is a large and quickly growing literature on regularity results for minimizers of functionals with (p, q)-growth and more general nonstandard growth; we refer to [Mingione 2006] for an overview.

MSC2010: 35B65.
Keywords: nonuniformly elliptic equations, local Lipschitz continuity, (p, q)-growth, nonstandard growth conditions.

Under additional structural assumptions on the growth of f, for example anisotropic growth of the form

$$m\sum_{i=1}^{n}|z_i|^{p_i} \le f(z) \le M\sum_{i=1}^{n}(1+|z_i|^q),$$

more precise and sharp assumptions on the involved exponents that ensure higher regularity are available in the literature; see, e.g., [Cupini et al. 2015; Fusco and Sbordone 1993]. Regularity results under general structural assumptions beyond polynomial growth can be found, e.g., in [Lieberman 1991; Marcellini 1993]; see also the recent result [Eleuteri et al. 2020], where convexity is only imposed "at infinity". Moreover, rather sharp conditions are known for certain nonautonomous functionals; see, e.g., [Baroni et al. 2018; Colombo and Mingione 2015; De Filippis and Mingione 2020; Esposito et al. 2004], where also Hölder-continuity of the integrand f in the space variable has to be balanced with p, q, and n. In [Carozza et al. 2014; Esposito et al. 1999] higher integrability results for autonomous integral functionals can be found that are also valid in the case of systems.

To the best of our knowledge, there is no improvement of the results (A) and (B) with respect to the relation between the exponents p, q and the dimension n available in the literature (without any additional structure assumption or further a priori assumptions on the minimizer, e.g., boundedness as in [Bousquet and Brasco 2020; Carozza et al. 2011]). In the present paper, we give such an improvement in the case $n \ge 3$. Before we state the results, we recall a standard notion of local minimizer in the context of functionals with (p, q)-growth.

Definition 2. We call $u \in W^{1,1}_{\mathrm{loc}}(\Omega)$ a local minimizer of $\mathcal{F}$ given in (1) if and only if

$$f(\nabla u) \in L^1_{\mathrm{loc}}(\Omega)$$

and

$$\int_{\operatorname{supp}\varphi} f(\nabla u)\,dx \le \int_{\operatorname{supp}\varphi} f(\nabla u+\nabla\varphi)\,dx$$

for any $\varphi \in W^{1,1}(\Omega)$ satisfying $\operatorname{supp}\varphi \Subset \Omega$.

The main results of the present paper can be summarized as:

Theorem 3. *Let $\Omega \subset \mathbb{R}^n$, $n \ge 2$, and suppose Assumption 1 is satisfied with $2 \le p \le q < \infty$ such that*

$$\frac{q}{p} < 1+\frac{2}{n-3} \quad \text{if } n \ge 4. \tag{3}$$

Let $u \in W^{1,q}_{\mathrm{loc}}(\Omega)$ be a local minimizer of the functional $\mathcal{F}$ given in (1). Then, $u \in W^{1,\infty}_{\mathrm{loc}}(\Omega)$.

Theorem 4. *Let $\Omega \subset \mathbb{R}^n$, $n \ge 2$ and suppose Assumption 1 is satisfied with $2 \le p \le q < \infty$ such that*

$$\frac{q}{p} < 1+\min\left\{1, \frac{2}{n-1}\right\}. \tag{4}$$

Let $u \in W^{1,1}_{\mathrm{loc}}(\Omega)$ be a local minimizer of the functional $\mathcal{F}$ given in (1). Then, $u \in W^{1,\infty}_{\mathrm{loc}}(\Omega)$.

Remark 5. Notice that Theorems 3 and 4 improve the results (A) and (B) with respect to the assumptions on $\frac{q}{p}$ in dimensions $n \ge 3$. The results in [Marcellini 1991] apply to more general situations in the sense that (smooth) spatial dependence of f is allowed, a bounded right-hand side is included and nonlinear elliptic equations that not need to be Euler–Lagrange equations of integral functionals of the type (1) are considered. In order to present the new ingredients in the simplest setting we focus on the case of autonomous integral functionals with no right-hand side (as in [Marcellini 1989]). Very recently [Beck and Mingione 2020] sharp criteria for Lipschitz-regularity of minimizers of variational integrals with respect to the right-hand side were obtained under the assumption $\frac{q}{p} < 1 + \frac{2}{n}$. It would be of interest to see whether such results can be extended to the case $\frac{q}{p} < 1 + \frac{2}{n-1}$ if $n \geq 3$.

Remark 6. We do not know whether assumptions (3) and (4) are respectively optimal in Theorems 3 and 4. It is known that Lipschitz-regularity and even boundedness of minimizers fail if $\frac{q}{p}$ is too large depending on the dimension n. In particular it is known that in order to ensure boundedness it is necessary that $\frac{q}{p} \to 1$ if $n \to \infty$; see [Giaquinta 1987; Hong 1992; Marcellini 1989; 1991] for related counterexamples. In particular, it is shown in [Hong 1992] that the functional

$$\int_\Omega |\nabla u|^2 + |u_{x_n}|^4 \, dx,$$

which satisfies (2) with $p = 2$ and $q = 4$, admits an unbounded minimizer if $n \geq 6$. Clearly, this does not match condition (4) in Theorem 4 and even not condition (3).

As already mentioned, once boundedness of the gradient is established, higher regularity follows by standard arguments; see, e.g., [Marcellini 1989, Proof of Theorem D]. Let us state (without proof) a rather direct consequence of Theorem 4.

Corollary 7. *Let $\Omega \subset \mathbb{R}^n$, $n \geq 2$, and suppose Assumption 1 is satisfied with $2 \leq p \leq q < \infty$ such that (4) holds. Moreover, suppose that $z \mapsto f(z)$ is of class $C^{k,\alpha}_{\mathrm{loc}}$ for some integer $k \geq 2$ and $\alpha \in (0,1)$. Let $u \in W^{1,1}_{\mathrm{loc}}(\Omega)$ be a local minimizer of the functional $\mathcal{F}$ given in (1). Then, $u \in C^{k+2,\alpha}_{\mathrm{loc}}(\Omega)$.*

The proofs of Theorems 3 and 4 are in several aspects similar to the approach of [Marcellini 1989; 1991]. Following [Marcellini 1991], we prove Theorem 3 appealing to the difference quotient method in order to differentiate the Euler–Lagrange equation and use a variant of Moser's iteration argument [1960] to prove boundedness of the gradient. The improvement compared to the previous results lies in a recent refinement of Moser's iteration argument in the context of the linear nonuniformly elliptic equation, obtained by us in [Bella and Schäffner 2019] (see [Bella and Schäffner 2020] for an application to finite difference equations and stochastic analysis). In order to illustrate the relation between Theorem 3 and local boundedness results for nonuniformly elliptic equation, we suppose for the moment that f satisfies (2) with $2 = p < q$. A local minimizer $u \in W^{1,q}_{\mathrm{loc}}(\Omega)$ of (1) satisfies the Euler–Lagrange equation

$$\nabla \cdot Df(\nabla u) = 0$$

and thus, by differentiating,

$$\nabla \cdot D^2 f(\nabla u) \nabla(\partial_j u) = 0 \quad \text{for } j = 1, \dots, n. \tag{5}$$

The coefficient $D^2 f(\nabla u)$ is nonuniformly elliptic and we have by (2) and the assumption $u \in W^{1,q}_{\mathrm{loc}}(\Omega)$

$$m|\lambda|^2 \leq \langle D^2 f(\nabla u)\lambda, \lambda\rangle \leq \mu|\lambda|^2, \quad \text{where } \mu := M(1 + |\nabla u|^2)^{\frac{q-2}{2}} \in L^{\frac{q}{q-2}}_{\mathrm{loc}}(\Omega)$$

(recall $p = 2$). Classic regularity results for linear nonuniformly elliptic equations, due to [Murthy and Stampacchia 1968; Trudinger 1971], yield local boundedness of $\partial_j u$ if

$$\frac{q-2}{q} < \frac{2}{n} \quad \Longrightarrow \quad \frac{q}{2} < \frac{n}{n-2} = 1 + \frac{2}{n-2},$$

which is precisely Marcellini's condition (A) (in the case $p = 2$). Very recently, we improved in [Bella and Schäffner 2019] the assumptions of [Murthy and Stampacchia 1968; Trudinger 1971] and established local boundedness and validity of the Harnack inequality for linear elliptic equations under essentially optimal assumptions on the integrability of the coefficients; see [Franchi et al. 1998] for related counterexamples. Applied to (5), the results of [Bella and Schäffner 2019] yield local boundedness of $\partial_j u$ if

$$\frac{q-2}{q} < \frac{2}{n-1} \quad \Longrightarrow \quad \frac{q}{2} < \frac{n-1}{n-3} = 1 + \frac{2}{n-3},$$

which is precisely condition (3). For $p > 2$ the results of [Bella and Schäffner 2019] applied to (5) do not give the claimed condition (3) and thus we need to combine the reasoning of [Marcellini 1991] with arguments of [Bella and Schäffner 2019] and provide an essentially self-contained proof of Theorem 3. Theorem 4 follows from Theorem 3 by a combination of an interpolation argument (similar to [Marcellini 1991, Theorem 3.1]) and a suitable approximation procedure (inspired by [Esposito et al. 1999]).

The paper is organized as follows: In Section 2, we recall some results from [Marcellini 1991] and present a technical lemma which is used to derive an improved version of the Caccioppoli inequality, which plays a prominent role in the proof of Theorem 3. In Section 3, we prove Theorem 3 and provide a useful a priori estimate via interpolation; see Corollary 12. Finally, in Section 4, we establish Theorem 4 as a consequence of Corollary 12 and an approximation argument.

2. Preliminary lemmas

For $\alpha \geq 2$ and $k > 0$, let $g_{\alpha,k} : \mathbb{R} \to \mathbb{R}$ be the unique $C^1(\mathbb{R})$-function satisfying

$$g_{\alpha,k}(t) = t(1+t^2)^{\frac{\alpha-2}{2}} \quad \text{for } |t| \leq k, \tag{6}$$

and which is affine on $\mathbb{R} \setminus \{|t| \leq k\}$. Moreover, we set

$$G_{\alpha,k}(t) := \frac{g_{\alpha,k}^2(t)}{g'_{\alpha,k}(t)}. \tag{7}$$

The following bounds on $G_{\alpha,k}$ are derived in [Marcellini 1991]

Lemma 8 [Marcellini 1991, Lemma 2.6]. *For every $\alpha \in [2, \infty)$ and $k > 0$ there exists $c = c(\alpha, k) \in [1, \infty)$ such that for all $t \in \mathbb{R}$*

$$G_{\alpha,k}(t) \leq c_{\alpha,k}(1+t^2), \tag{8}$$

$$G_{\alpha,k}(t) \leq 2\left(\frac{1+k^2}{k^2}\right)^{\frac{\alpha-2}{2}} (1+t^2)^{\frac{\alpha}{2}}. \tag{9}$$

Appealing to the difference quotient method, it was proven in [Marcellini 1991] that local minimizers of (1) satisfying $W^{1,q}_{\mathrm{loc}}(\Omega)$ integrability enjoy higher differentiability:

Lemma 9. *Let $\Omega \subset \mathbb{R}^n$, $n \geq 2$, and suppose Assumption 1 is satisfied with $2 \leq p \leq q < \infty$. Let $u \in W^{1,q}_{\mathrm{loc}}(\Omega)$ be a local minimizer of the functional $\mathcal{F}$ given in (1). Then, $u \in W^{2,2}_{\mathrm{loc}}(\Omega)$. Moreover, for every $\eta \in C^1_c(\Omega)$, any $s \in \{1, \dots, n\}$ and any $\alpha \geq 2$,*

$$\int_\Omega \eta^2 g'_{\alpha,k}(u_{x_s})(1+|\nabla u|^2)^{\frac{p-2}{2}}|\nabla u_{x_s}|^2\,dx \leq \frac{4M}{m}\int_\Omega |\nabla\eta|^2 G_{\alpha,k}(u_{x_s})(1+|\nabla u|^2)^{\frac{q-2}{2}}\,dx. \tag{10}$$

Lemma 9 is essentially proven in [Marcellini 1991]. However, estimate (10), which is the starting point for our analysis, is not explicitly stated in that work (as mentioned above, that work deals with more general equations, and additional terms appear on the right-hand side to which our methods do not directly apply) and thus we sketch the proof of Lemma 9 following the reasoning of [Marcellini 1991].

Proof of Lemma 9. First, we note that since $u \in W^{1,q}_{\mathrm{loc}}(\Omega)$ and $|Df(z)| \leq c(1+|z|)^{q-1}$ for some $c = c(M,n,q) \in [1,\infty)$ (by (2)), we obtain that u solves the Euler–Lagrange equation

$$\int_\Omega \langle Df(\nabla u), \nabla\varphi\rangle\,dx = 0 \quad \text{for all } \varphi \in W^{1,q}(\Omega) \text{ with } \operatorname{supp}\varphi \Subset \Omega. \tag{11}$$

For $s \in \{1, \dots, n\}$, we consider the difference quotient operator

$$\tau_{s,h} v := \frac{1}{h}(v(\cdot + he_s) - v), \quad \text{where } v \in L^1_{\mathrm{loc}}(\mathbb{R}^n).$$

Fix $\eta \in C^1_c(\Omega)$. Testing (11) with $\varphi := \tau_{s,-h}(\eta^2 g_{\alpha,k}(\tau_{s,h}u))$, we obtain

$$\begin{aligned}(I) &:= \int_\Omega \eta^2 g'_{\alpha,k}(\tau_{s,h}u)\langle \tau_{s,h}Df(\nabla u), \tau_{s,h}\nabla u\rangle\,dx \\ &= -2\int_\Omega \eta g_{\alpha,k}(\tau_{s,h}u)\langle \tau_{s,h}Df(\nabla u), \nabla\eta\rangle\,dx =: (II).\end{aligned}$$

Writing $\tau_{s,h}Df(\nabla u) = \frac{1}{h}Df(\nabla u + th\tau_{s,h}\nabla u)\big|_{t=0}^{t=1}$, the fundamental theorem of calculus yields

$$\begin{aligned}\int_\Omega\int_0^1 \eta^2 g'_{\alpha,k}(\tau_{s,h}u)\langle D^2 f(\nabla u + th\tau_{s,h}\nabla u))\tau_{s,h}\nabla u, \tau_{s,h}\nabla u\rangle\,dt\,dx \\ = (I) = (II) \\ = -2\int_\Omega\int_0^1 \eta g_{\alpha,k}(\tau_{s,h}u)\langle D^2 f(\nabla u + th\tau_{s,h}\nabla u)\tau_{s,h}\nabla u, \nabla\eta\rangle\,dt\,dx.\end{aligned} \tag{12}$$

Young's inequality and the definition of $G_{\alpha,k}$, see (7), then yield

$$|(II)| \leq \tfrac{1}{2}(I) + 2(III), \tag{13}$$

where

$$(III) := \int_\Omega\int_0^1 G_{\alpha,k}(\tau_{s,h}u)\langle D^2 f(\nabla u + th\tau_{s,h}\nabla u)\nabla\eta, \nabla\eta\rangle\,dt\,dx.$$

Combining (12), (13) with the assumptions on $D^2 f$, see (2), we obtain for all $\alpha \geq 2$

$$m \int_\Omega \int_0^1 \eta^2 g'_{\alpha,k}(\tau_{s,h}u)(1+|\nabla u + th\tau_{s,h}\nabla u|^2)^{\frac{p-2}{2}} |\tau_{s,h}\nabla u|^2 \, dx$$

$$\leq (I) \leq 4(III)$$

$$\leq 4M \int_\Omega \int_0^1 G_{\alpha,k}(\tau_{s,h}u)(1+|\nabla u + th\tau_{s,h}\nabla u|^2)^{\frac{q-2}{2}} |\nabla \eta|^2 \, dx. \quad (14)$$

Estimate (14) with $\alpha = 2$ (and thus $g_{2,k} = t$, $g'_{2,k} = 1$ and $G_{2,k}(t) = t^2$; see (6), (7)), the assumption $u \in W^{1,q}_{\mathrm{loc}}(\Omega)$ and the arbitrariness of $\eta \in C^1_c(\Omega)$ and $s \in \{1, \dots, n\}$ yield $u \in W^{2,2}_{\mathrm{loc}}(\Omega)$. Finally, by sending h to zero in (14) we obtain the desired estimate (10) (for this we use that $G_{\alpha,k}$ is quadratic for every $k > 0$, see (8), and thus $G_{\alpha,k}(\tau_{s,h}u) \to G_{\alpha,k}(u_{x_s})$ in $L^{\frac{q}{2}}(\Omega')$ for any $\Omega' \Subset \Omega$). □

To this point, we essentially recalled notation and statements from [Marcellini 1991]. Following that work, we will combine (10) with a Moser-iteration-type argument to establish the desired Lipschitz-estimate. In contrast to [Marcellini 1991], we optimize estimate (10) with respect to η, which will enable us to use Sobolev inequality on spheres instead of balls. The following lemma captures the needed improvement due to a suitable choice of the cut-off function η:

Lemma 10. *Fix $n \geq 2$. For given $0 < \rho < \sigma < \infty$ and $v \in L^1(B_\sigma)$ consider*

$$J(\rho, \sigma, v) := \inf\left\{ \int_{B_\sigma} |v||\nabla \eta|^2 \, dx \;\middle|\; \eta \in C^1_0(B_\sigma),\ \eta \geq 0,\ \eta = 1 \text{ in } B_\rho \right\}.$$

Then for every $\delta \in (0, 1]$

$$J(\rho, \sigma, v) \leq (\sigma - \rho)^{-(1+\frac{1}{\delta})} \left(\int_\rho^\sigma \left(\int_{S_r} |v| \right)^\delta dr \right)^{\frac{1}{\delta}}. \quad (15)$$

Proof of Lemma 10. Estimate (15) follows directly by minimizing among radial symmetric cut-off functions. Indeed, we obviously have for every $\varepsilon \geq 0$

$$J(\rho, \sigma, v) \leq \inf\left\{ \int_\rho^\sigma \eta'(r)^2 \left(\int_{S_r} |v| + \varepsilon \right) dr \;\middle|\; \eta \in C^1(\rho, \sigma),\ \eta(\rho) = 1,\ \eta(\sigma) = 0 \right\} =: J_{1\mathrm{d},\varepsilon}.$$

For $\varepsilon > 0$, the one-dimensional minimization problem $J_{1\mathrm{d},\varepsilon}$ can be solved explicitly and we obtain

$$J_{1\mathrm{d},\varepsilon} = \left(\int_\rho^\sigma \left(\int_{S_r} |v| + \varepsilon \right)^{-1} dr \right)^{-1}. \quad (16)$$

Let us give an argument for (16). First we observe that using the assumption $v \in L^1(B_\sigma)$ and a simple approximation argument we can replace $\eta \in C^1(\rho, \sigma)$ with $\eta \in W^{1,\infty}(\rho, \sigma)$ in the definition of $J_{1\mathrm{d},\varepsilon}$. Let $\tilde{\eta} : [\rho, \sigma] \to [0, \infty)$ be given by

$$\tilde{\eta}(r) := 1 - \left(\int_\rho^\sigma b(r)^{-1} \, dr \right)^{-1} \int_\rho^r b(r)^{-1} \, dr, \quad \text{where } b(r) := \textstyle\int_{S_r} |v| + \varepsilon.$$

Clearly, $\tilde\eta \in W^{1,\infty}(\rho,\sigma)$ (since $b \geq \varepsilon > 0$), $\tilde\eta(\rho)=1$, $\tilde\eta(\sigma)=0$, and thus

$$J_{1\mathrm{d},\varepsilon} \leq \int_\rho^\sigma \tilde\eta'(r)^2 b(r)\,dr = \left(\int_\rho^\sigma b(r)^{-1}\,dr\right)^{-1}.$$

The reverse inequality follows by Hölder's inequality: For every $\eta \in W^{1,\infty}(\rho,\sigma)$ satisfying $\eta(\rho)=1$ and $\eta(\sigma)=0$, we have

$$1 = \left(\int_\rho^\sigma \eta'(r)\,dr\right)^2 \leq \int_\rho^\sigma \eta'(r)^2 b(r)\,dr \int_\rho^\sigma b(r)^{-1}\,dr.$$

Clearly, the last two displayed formulas imply (16).

Next, we deduce (15) from (16). For every $s>1$, we obtain by Hölder inequality

$$\sigma - \rho = \int_\rho^\sigma \left(\frac{b}{b}\right)^{\frac{s-1}{s}} \leq \left(\int_\rho^\sigma b^{s-1}\right)^{\frac{1}{s}} \left(\int_\rho^\sigma \frac{1}{b}\right)^{\frac{s-1}{s}},$$

with b as above, and by (16) that

$$J_{1\mathrm{d},\varepsilon} \leq (\sigma-\rho)^{-\frac{s}{s-1}} \left(\int_\rho^\sigma \left(\int_{S_r} |v| + \varepsilon\right)^{s-1} dr\right)^{\frac{1}{s-1}}.$$

Sending ε to zero, we obtain (15) with $\delta = s-1>0$. □

3. Proof of Theorem 3

The main result of this section is the following

Theorem 11. *Let $\Omega \subset \mathbb{R}^n$, $n \geq 3$, and suppose Assumption 1 is satisfied with $2 \leq p < q < \infty$ such that (3) holds. Fix*

$$\theta = \frac{2q}{(n-1)p-(n-3)q} \quad \text{if } n \geq 4 \qquad \text{and} \qquad \theta > \frac{q}{p} \quad \text{if } n = 3. \tag{17}$$

Let $u \in W^{1,q}_{\mathrm{loc}}(\Omega)$ be a local minimizer of the functional $\mathcal{F}$ given in (1). Then, there exists $c = c(n,m,M,p,q,\theta) \in [1,\infty)$ such that for every $B_R(x_0) \Subset \Omega$ and any $\rho \in (0,1)$

$$\|(1+|\nabla u|^2)^{\frac{1}{2}}\|_{L^\infty(B_{\rho R}(x_0))} \leq c((1-\rho)R)^{-n\frac{\theta}{q}} \|(1+|\nabla u|^2)^{\frac{1}{2}}\|^{\theta}_{L^q(B_R(x_0))}. \tag{18}$$

Proof of Theorem 3. Theorem 11 contains the claim of Theorem 3 in the case $n \geq 3$ and $2 \leq p < q$. The remaining case $n=2$ is contained in [Marcellini 1991, Theorem 2.1] and the statement is classic for $p=q$. □

Proof of Theorem 11. Throughout the proof we write $\lesssim$ if $\leq$ holds up to a positive constant which depends only on n, m, M, p and q.

Step 1: One step improvement. Suppose that $B_2 \Subset \Omega$. We claim that for every

$$\gamma \in (0,1] \quad \text{satisfying} \quad \frac{n-3}{n-1} \leq \gamma \tag{19}$$

there exists $c = c(\gamma, n, m, M, p, q) \in [1, \infty)$ such that for every $\frac{1}{2} \le \rho < \sigma \le 1$ and any $\alpha \ge 2$

$$\|\phi_{\alpha+p-2}\|^2_{W^{1,2}(B_\rho)} \le c\alpha^2(\sigma-\rho)^{-(1+\frac{1}{\gamma})}\|\phi_{(\alpha+q-2)\gamma}\|^{\frac{2}{\gamma}}_{W^{1,2}(B_\sigma)}, \tag{20}$$

where we use the shorthand

$$\phi_\beta := \sum_{j=1}^n (1+u_{x_j}^2)^{\frac{\beta}{4}} \quad \text{for } \beta > 0. \tag{21}$$

Moreover, there exists $c = c(n, m, M, p, q) \in [1, \infty)$ such that for every $0 < \rho < \sigma \le 2$ and any $\alpha \ge 2$

$$\|\nabla\phi_{\alpha+p-2}\|^2_{L^2(B_\rho)} \le c\alpha^2(\sigma-\rho)^{-2}\|\phi_{\alpha+q-2}\|^2_{L^2(B_\sigma)}. \tag{22}$$

Substep 1.1: We claim that there exists $c = c(\gamma, n, q) \in [1, \infty)$ such that for every $k > 0$, $\alpha \ge 2$, $s \in \{1, \dots, n\}$, and $\frac{1}{2} \le \rho < \sigma \le 1$

$$\begin{aligned} I_{\alpha,k,s}(\rho,\sigma) &:= \inf\left\{\int_{B_\sigma} |\nabla\eta|^2 G_{\alpha,k}(u_{x_s})(1+|\nabla u|^2)^{\frac{q-2}{2}} \,\middle|\, \eta \in C_0^1(B_\sigma),\ \eta = 1 \text{ in } B_\rho\right\} \\ &\le c(\sigma-\rho)^{-(1+\frac{1}{\gamma})}\left(\frac{1+k^2}{k^2}\right)^{\frac{\alpha-2}{2}} \|\phi_{(q-2+\alpha)\gamma}\|^{\frac{2}{\gamma}}_{W^{1,2}(B_\sigma)}. \end{aligned} \tag{23}$$

Assumption $u \in W^{1,q}(B_1)$ and estimate (8) imply that $v := G_{\alpha,k}(u_{x_s})(1+|\nabla u|^2)^{\frac{q-2}{2}} \in L^1(B_1)$. Hence, Lemma 10 and (9) yield for every $\delta \in (0, 1]$

$$I_{\alpha,k,s}(\rho,\sigma) \le 2(\sigma-\rho)^{-(1+\frac{1}{\delta})}\left(\frac{1+k^2}{k^2}\right)^{\frac{\alpha-2}{2}}\left(\int_\rho^\sigma\left(\int_{S_r}(1+u_{x_s}^2)^{\frac{\alpha}{2}}(1+|\nabla u|^2)^{\frac{q-2}{2}}\right)^\delta dr\right)^{\frac{1}{\delta}}.$$

Appealing to Young's inequality, we find $c = c(n) \in [1, \infty)$ such that

$$\sum_{j=1}^n (1+u_{x_j}^2)^{\frac{\alpha}{2}} \sum_{j=1}^n (1+u_{x_j}^2)^{\frac{q-2}{2}} \le c\sum_{j=1}^n (1+u_{x_j}^2)^{\frac{\alpha+q-2}{2}} \tag{24}$$

(in fact (24) is valid with $c = 1 + \frac{1}{2}n(n-1)$; see [Marcellini 1991, Lemma 2.9]) and thus

$$\begin{aligned} (1+u_{x_s}^2)^{\frac{\alpha}{2}}(1+|\nabla u|^2)^{\frac{q-2}{2}} &\le n^{\max\{\frac{q-2}{2}-1,0\}} \sum_{j=1}^n (1+u_{x_j}^2)^{\frac{\alpha}{2}} \sum_{j=1}^n (1+u_{x_j}^2)^{\frac{q-2}{2}} \\ &\overset{(24)}{\le} cn^{\max\{\frac{q-2}{2}-1,0\}} \sum_{j=1}^n (1+u_{x_j}^2)^{\frac{\alpha+q-2}{2}} \\ &\le cn^{\max\{\frac{q-2}{2}-1,0\}} \phi^{\frac{2}{\gamma}}_{(\alpha+q-2)\gamma}, \end{aligned}$$

where in the first inequality we use Jensen's inequality in the case $\frac{q-2}{2} \ge 1$ and the discrete ℓ_s-ℓ_1 estimate, with $s \ge 1$, for $\frac{q-2}{2} \in (0, 1)$, and the third inequality is again the discrete ℓ_s-ℓ_1 estimate, with $s \ge 1$. Hence, we find $c = c(n, q) \in [1, \infty)$ such that

$$I_{\alpha,k,s}(\rho,\sigma) \le c(\sigma-\rho)^{-(1+\frac{1}{\delta})}\left(\frac{1+k^2}{k^2}\right)^{\frac{\alpha-2}{2}}\left(\int_\rho^\sigma\left(\int_{S_r}\phi^{\frac{2}{\gamma}}_{(\alpha+q-2)\gamma}\right)^\delta dr\right)^{\frac{1}{\delta}} \quad \text{for all } \delta \in (0, 1]. \tag{25}$$

To estimate the right-hand side in (25) we use the Sobolev inequality on spheres; i.e., for all $\gamma \in (0,1]$ there exists $c = c(n,\gamma) \in [1,\infty)$ such that for every $r > 0$

$$\Big(\int_{S_r} |\varphi|^{\frac{2}{\gamma}}\Big)^{\frac{\gamma}{2}} \le c\Big(\Big(\int_{S_r} |\nabla\varphi|^{(\frac{2}{\gamma})_*}\Big)^{\frac{1}{(2/\gamma)*}} + \frac{1}{r}\Big(\int_{S_r} |\varphi|^{(\frac{2}{\gamma})_*}\Big)^{\frac{1}{(2/\gamma)*}}\Big), \quad \text{where } \frac{1}{(\frac{2}{\gamma})_*} = \frac{\gamma}{2} + \frac{1}{n-1}. \tag{26}$$

Estimate (26) and assumption (19) in the form

$$\frac{1}{(\frac{2}{\gamma})_*} = \frac{\gamma}{2} + \frac{1}{n-1} \overset{(19)}{\ge} \frac{n-3}{2(n-1)} + \frac{1}{n-1} \ge \frac{1}{2}$$

yield

$$\begin{aligned}
&\Big(\int_\rho^\sigma \Big(\int_{S_r} \phi_{(\alpha+q-2)\gamma}^{\frac{2}{\gamma}}\Big)^\delta dr\Big)^{\frac{1}{\delta}} \\
&\quad \le c\Big(\int_\rho^\sigma \Big[\Big(\int_{S_r} |\nabla\phi_{(\alpha+q-2)\gamma}|^{(\frac{2}{\gamma})_*}\Big)^{\frac{1}{(2/\gamma)*}} + \frac{1}{r}\Big(\int_{S_r} \phi_{(\alpha+q-2)\gamma}^{(\frac{2}{\gamma})_*}\Big)^{\frac{1}{(2/\gamma)*}}\Big]^{\delta\frac{2}{\gamma}} dr\Big)^{\frac{1}{\delta}} \\
&\quad \le c\Big(\int_\rho^\sigma |S_r|^{(\frac{1}{(2/\gamma)*}-\frac{1}{2})\frac{2\delta}{\gamma}} \Big[\Big(\int_{S_r} |\nabla\phi_{(\alpha+q-2)\gamma}|^2\Big)^{\frac{1}{2}} + \frac{1}{r}\Big(\int_{S_r} \phi_{(\alpha+q-2)\gamma}^2\Big)^{\frac{1}{2}}\Big]^{\delta\frac{2}{\gamma}} dr\Big)^{\frac{1}{\delta}},
\end{aligned} \tag{27}$$

where $c = c(\gamma,n) \in [1,\infty)$. Combining (25) and (27) with the choice $\delta = \gamma$, we obtain the claimed estimate (23) (we can ignore the factors $|S_r|$ and $\frac{1}{r}$ in (27) by assumption $\frac{1}{2} \le \rho < \sigma \le 1$).

Substep 1.2: Proof of (20). Lemma 9 and estimate (23) yield for every $s \in \{1,\dots,n\}$

$$\int_{B_\rho} g'_{\alpha,k}(u_{x_s})(1+|\nabla u|^2)^{\frac{p-2}{2}}|\nabla u_{x_s}|^2\,dx \le c(\sigma-\rho)^{-(1+\frac{1}{\gamma})}\Big(\frac{1+k^2}{k^2}\Big)^{\frac{\alpha-2}{2}} \|\phi_{(q-2+\alpha)\gamma}\|_{W^{1,2}(B_\sigma)}^{\frac{2}{\gamma}},$$

where $c = c(\gamma,n,m,M,p,q) \in [1,\infty)$. Sending k to infinity and summing over s from 1 to n, we obtain (using $\lim_{k\to\infty} g'_{\alpha,k}(t) \ge (1+t^2)^{\frac{\alpha-2}{2}}$)

$$\int_{B_\rho} \sum_{j=1}^n (1+u_{x_j}^2)^{\frac{\alpha+p-4}{2}}|\nabla u_{x_j}|^2\,dx \le c(\sigma-\rho)^{-(1+\frac{1}{\gamma})}\|\phi_{(q-2+\alpha)\gamma}\|_{W^{1,2}(B_\sigma)}^{\frac{2}{\gamma}}.$$

Combining the above estimate with the pointwise inequality

$$|\nabla\phi_{\alpha+p-2}| \le \frac{\alpha+p-2}{2}\sum_{j=1}^n (1+u_{x_j}^2)^{\frac{\alpha+p-4}{4}}|\nabla u_{x_j}|, \tag{28}$$

we obtain that there exists $c = c(\gamma,n,m,M,p,q) \in [1,\infty)$ such that for all $\frac{1}{2} \le \rho < \sigma \le 1$ and $\alpha \ge 2$

$$\|\nabla\phi_{\alpha+p-2}\|_{L^2(B_\rho)}^2 \le c\alpha^2(\sigma-\rho)^{-(1+\frac{1}{\gamma})}\|\phi_{(q-2+\alpha)\gamma}\|_{W^{1,2}(B_\sigma)}^{\frac{2}{\gamma}}. \tag{29}$$

It remains to estimate $\|\phi_{\alpha+p-2}\|_{L^2(B_\rho)}$. For this, we use a version of the Poincaré inequality: for every $\varepsilon > 0$ there exists $c = c(\varepsilon,n) \in [1,\infty)$ such that for all $r > 0$ and $v \in H^1(B_r)$

$$\Big(\fint_{B_r} |v|^2\Big)^{\frac{1}{2}} \le c\Big(r\Big(\fint_{B_r} |\nabla v|^2\Big)^{\frac{1}{2}} + \Big(\fint_{B_r} |v|^\varepsilon\Big)^{\frac{1}{\varepsilon}}\Big). \tag{30}$$

We recall a proof of (30) at the end of this step. Inequality (30) with $v = \phi_{\alpha+p-2}$ and

$$\varepsilon = 2\gamma \frac{\alpha+q-2}{\alpha+p-2}$$

together with the inequality

$$1 \le \phi_{\alpha+p-2} \le n^{\max\{0,1-\frac{\alpha+p-2}{(\alpha+q-2)\gamma}\}} \phi_{(\alpha+q-2)\gamma}^{\frac{1}{\gamma}\frac{\alpha+p-2}{\alpha+q-2}}$$

(the second inequality follows by Jensens inequality if $(\alpha+q-2)\gamma/(\alpha+p-2) \ge 1$ and the discrete ℓ_s-ℓ_1 inequality, with $s \ge 1$, otherwise) yield

$$\|\phi_{\alpha+p-2}\|^2_{L^2(B_\rho)} \le c(\|\nabla\phi_{\alpha+p-2}\|^2_{L^2(B_\rho)} + \|\phi_{(\alpha+q-2)\gamma}\|^{\frac{2}{\gamma}\frac{\alpha+p-2}{\alpha+q-2}}_{L^2(B_\rho)}), \tag{31}$$

where $c = c(n,\gamma,p,q) \in [1,\infty)$ (note that $\rho \in [\frac{1}{2},1]$ and $\varepsilon \in [2\gamma, \frac{q}{p}2\gamma]$). The first term on the right-hand side in (31) can be estimated by (29) and the second term (using $p \le q$ and $\phi_\beta \ge 1$ for all $\beta > 0$) by

$$\|\phi_{(\alpha+q-2)\gamma}\|^{\frac{2}{\gamma}\frac{\alpha+p-2}{\alpha+q-2}}_{L^2(B_\rho)} \le c\|\phi_{(\alpha+q-2)\gamma}\|^{\frac{2}{\gamma}}_{L^2(B_\rho)}. \tag{32}$$

A combination of (29), (31) and (32) yield (20).

Finally, we recall an argument for (30): Clearly it suffices to proof the statement for $r = 1$. Given $\varepsilon > 0$, set

$$U_\varepsilon := \{x \in B_1 \mid |v(x)| \le \lambda_\varepsilon\}, \quad \text{where } \lambda_\varepsilon := \left(2 \fint_{B_1} |v|^\varepsilon\right)^{\frac{1}{\varepsilon}}.$$

The choice of λ_ε and the Markov inequality yield

$$|B_1 \setminus U_\varepsilon| \le \lambda^{-\varepsilon} \int_{B_1} |v|^\varepsilon \le \tfrac{1}{2}|B_1|$$

and thus $|U_\varepsilon| \ge \frac{1}{2}|B_1|$. Hence, by a suitable version of the Poincaré inequality, see, e.g., [Gilbarg and Trudinger 1998, (7.45), p. 164], there exists $c = c(n) \in [1,\infty)$ such that

$$\int_{B_1} \left| v - \fint_{U_\varepsilon} v \right|^2 \le c \int_{B_1} |\nabla v|^2.$$

The above inequality, the triangle inequality and

$$\fint_{U_\varepsilon} |v| \le 2\lambda_\varepsilon^{1-\varepsilon} \fint_{B_1} |v|^\varepsilon \le 2^{\frac{1}{\varepsilon}} \left(\fint_{B_1} |v|^\varepsilon\right)^{\frac{1}{\varepsilon}}$$

imply (30).

Substep 1.3: Proof of (22). This estimate is an intermediate step in the proof of [Marcellini 1991, Lemma 2.10], but for completeness we recall the argument. Lemma 9 with η being the affine cutoff function for B_ρ in B_σ yields for every $s \in \{1,\dots,n\}$

$$\int_{B_\rho} g'_{\alpha,k}(u_{x_s})(1+|\nabla u|^2)^{\frac{p-2}{2}} |\nabla u_{x_s}|^2\,dx \lesssim (\sigma-\rho)^{-2} \int_{B_\sigma} G_{\alpha_k}(u_{x_s})(1+|\nabla u|^2)^{\frac{q-2}{2}}\,dx$$

and by summing s from 1 to n and sending $k \to \infty$, we obtain

$$\int_{B_\rho} \sum_{j=1}^{n} (1+u_{x_j}^2)^{\frac{\alpha+p-4}{2}} |\nabla u_{x_j}|^2 \, dx \lesssim (\sigma-\rho)^{-2} \int_{B_\sigma} \phi_{\alpha+q-2}^2. \tag{33}$$

Estimate (22) is a consequence of (28) and (33).

Step 2: Iteration. Fix θ as in (17). We claim that there exists $c = c(n,m,M,p,q,\theta) \in [1,\infty)$ such that

$$\|(1+|\nabla u|^2)^{\frac{1}{2}}\|_{L^\infty(B_{\frac{1}{2}})} \leq c\|(1+|\nabla u|^2)^{\frac{1}{2}}\|_{L^q(B_2)}^{\theta}. \tag{34}$$

Set

$$\gamma = \frac{n-3}{n-1} \quad \text{if } n \geq 4 \quad \text{and} \quad \gamma = \frac{\frac{p}{q}\theta - 1}{\theta - 1} \quad \text{if } n = 3. \tag{35}$$

Note that the assumptions $p < q$ and $\theta > \frac{q}{p}$ yield

$$0 < \gamma < \frac{p}{q} \quad \text{if } n = 3. \tag{36}$$

We define a sequence $\{a_k\}_{k\in\mathbb{N}_0}$ by

$$\alpha_0 := 2, \qquad \alpha_k := \frac{1}{\gamma}(\alpha_{k-1}+p-2) - (q-2) \quad \text{for all } k \in \mathbb{N}.$$

By induction one sees that

$$\alpha_k = 2 + \left(\frac{p}{\gamma} - q\right) \sum_{i=0}^{k-1} \gamma^{-i} = 2 + \left(\frac{p}{\gamma} - q\right) \frac{\gamma^{-k}-1}{\gamma^{-1}-1} = 2 + p\frac{\gamma^{-k}-1}{1-\gamma}\left(1 - \gamma\frac{q}{p}\right) \quad \text{for all } k \in \mathbb{N}.$$

The choice of γ in (35), assumption (3), and (36) together with $p < q$ imply $1 - \gamma\frac{q}{p} > 0$ and $\gamma^{-1} > 1$; hence

$$\alpha_k \to \infty \quad \text{as } k \to \infty.$$

For $k \in \mathbb{N}$, set

$$\rho_k = \frac{1}{2} + \frac{1}{2^{k+1}}, \quad \sigma_k := \rho_k + \frac{1}{2^{k+1}} = \rho_{k-1}$$

(where $\rho_0 := 1$), and

$$A_k := \|\phi_{\alpha_k+p-2}\|_{W^{1,2}(B_{\rho_k})}^{\frac{2}{\alpha_k+p-2}} \quad \text{for all } k \in \mathbb{N}_0,$$

where ϕ_β, $\beta \geq 0$ is defined in (21). Since $\alpha_{k-1}+p-2 = (\alpha_k+q-2)\gamma$, estimate (20) for $\alpha = \alpha_k$ implies

$$A_k \leq (c2^{(k+1)(1+\frac{1}{\gamma})}\alpha_k^2)^{\frac{1}{\alpha_k+p-2}} A_{k-1}^{\frac{1}{\gamma}\frac{\alpha_{k-1}+p-2}{\alpha_k+p-2}} \quad \text{for every } k \in \mathbb{N},$$

where $c = c(\gamma,n,m,M,p,q) \in [1,\infty)$ as in (20) and thus by iteration

$$A_k \leq A_0^{\gamma^{-k}\prod_{i=1}^{k}\frac{\alpha_{i-1}+p-2}{\alpha_i+p-2}} \prod_{i=1}^{k} (c2^{(i+1)(1+\frac{1}{\gamma})}\alpha_i^2)^{\frac{1}{\alpha_i+p-2}}. \tag{37}$$

Note that for every $k \in \mathbb{N}$

$$\prod_{i=1}^{k} (c2^{(i+1)(1+\frac{1}{\gamma})}\alpha_i^2)^{\frac{1}{\alpha_i+p-2}} \leq \exp\left(\sum_{i=1}^{\infty} \frac{\log(c2^{(i+1)(1+\frac{1}{\gamma})}\alpha_i^2)}{\alpha_i+p-2}\right) = c(\gamma,n,m,M,p,q) < \infty$$

and

$$\gamma^{-k}\prod_{i=1}^{k}\frac{\alpha_{i-1}+p-2}{\alpha_i+p-2}=\gamma^{-k}\frac{\alpha_0+p-2}{\alpha_k+p-2}$$
$$=\gamma^{-k}\frac{p}{p\frac{\gamma^{-k}-1}{1-\gamma}\big(1-\gamma\frac{q}{p}\big)+p}=\Big(\frac{\gamma^{-k}}{\gamma^{-k}-1}\Big)\Big(\frac{1-\gamma}{1-\gamma\frac{q}{p}+\frac{1-\gamma}{\gamma^{-k}-1}}\Big).$$

Hence, sending $k\to\infty$ in (37), we obtain that there exists $c=c(n,m,M,p,q,\theta)\in[1,\infty)$ (note $\gamma=\gamma(n,p,q,\theta)<1$) such that

$$\|(1+|\nabla u|^2)^{\frac12}\|_{L^\infty(B_{\frac12})}\le cA_0^{\frac{1-\gamma}{1-\gamma(q/p)}}=c\|\phi_p\|_{W^{1,2}(B_1)}^{\frac{2(1-\gamma)}{p-\gamma q}}. \tag{38}$$

Estimate (22) and $2\le p\le q$ together with $\phi_\beta\ge1$ for all $\beta\ge0$ yield

$$\|\phi_p\|_{W^{1,2}(B_1)}^{\frac{2(1-\gamma)}{p-\gamma q}}\lesssim\|\phi_q\|_{L^2(B_2)}^{\frac{2(1-\gamma)}{p-\gamma q}}\lesssim\|(1+|\nabla u|^2)^{\frac12}\|_{L^q(B_2)}^{\frac{q}{p}\frac{1-\gamma}{1-\gamma\frac{q}{p}}}. \tag{39}$$

Estimates (38), (39) and the choice of γ in (35) imply (34).

<u>Step 3</u>: Conclusion. Fix $\rho\in(0,1)$ and $B_R(x_0)\Subset\Omega$. By scaling and translation, we deduce from Step 2 that

$$\|(1+|\nabla u|^2)^{\frac12}\|_{L^\infty(B_{\frac{R}{4}}(x_0))}\le cR^{-n\frac{\theta}{q}}\|(1+|\nabla u|^2)^{\frac12}\|_{L^q(B_R(x_0))}^{\theta}, \tag{40}$$

where $c=c(n,m,M,p,q,\theta)\in[1,\infty)$ is the same as in (34). Applying for every $y\in B_{\rho R}(x_0)$ estimate (40) with $B_R(x_0)$ replaced by $B_{(1-\rho)R}(y)\subset\Omega$, we obtain

$$\|(1+|\nabla u|^2)^{\frac12}\|_{L^\infty(B_{\frac{1-\rho}{4}R}(y))}\le c((1-\rho)R)^{-n\frac{\theta}{q}}\|(1+|\nabla u|^2)^{\frac12}\|_{L^q(B_R(x_0))}^{\theta}$$

and thus the claimed estimate (18) follows. □

By the same interpolation argument as in [Marcellini 1991, Theorem 3.1], we deduce from Theorem 11:

Corollary 12. *Let $\Omega\subset\mathbb{R}^n$, $n\ge3$, and suppose Assumption 1 is satisfied with $2\le p<q<\infty$ such that (4) holds. Let θ be given as in (17) with the additional constraint $\theta<\frac{q}{q-p}$ for $n=3$ and set*

$$\alpha:=\frac{\theta\frac{p}{q}}{1-\theta\big(1-\frac{p}{q}\big)}. \tag{41}$$

Let $u\in W^{1,q}_{\mathrm{loc}}(\Omega)$ be a local minimizer of the functional $\mathcal{F}$ given in (1). Then, there exists $c=c(n,m,M,p,q,\theta)\in[1,\infty)$ such that for every $B_{2R}(x_0)\Subset\Omega$

$$\|(1+|\nabla u|^2)^{\frac12}\|_{L^\infty(B_{\frac{R}{2}}(x_0))}\le cR^{-n\frac{\alpha}{p}}\|(1+|\nabla u|^2)^{\frac12}\|_{L^p(B_R(x_0))}^{\alpha}. \tag{42}$$

Remark 13. A direct calculation yields

$$\alpha=\frac{2p}{(n+1)p-(n-1)q}\quad\text{if } n\ge4.$$

For $n=3$, the assumption on θ in Corollary 12 reads $\theta \in \big(\frac{q}{p}, \frac{q}{q-p}\big)$. Since $2 \le p < q$, we have

$$\frac{q}{p} < \frac{q}{q-p} \quad \iff \quad \frac{q}{p} < 2,$$

where the second inequality is ensured by (4) (for $n=3$).

Proof of Corollary 12. We prove the statement for $x_0 = 0$ and $R = 1$; the general claim follows by scaling and translation. Throughout the proof we write $\lesssim$ if $\le$ holds up to a positive constant which depends only on n, m, M, p, q and θ.

For $\nu \in \mathbb{N} \cup \{0\}$, we set

$$\rho_\nu = 1 - \frac{1}{2^{1+\nu}}.$$

Combining the elementary interpolation inequality

$$\|(1+|\nabla u|^2)^{\frac12}\|_{L^q(B_{\rho_\nu})} \le \|(1+|\nabla u|^2)^{\frac12}\|_{L^p(B_{\rho_\nu})}^{\frac{p}{q}} \|(1+|\nabla u|^2)^{\frac12}\|_{L^\infty(B_{\rho_\nu})}^{1-\frac{p}{q}} \tag{43}$$

with estimate (18), we obtain for every $\nu \in \mathbb{N}$

$$\begin{aligned} \|(1+|\nabla u|^2)^{\frac12}\|_{L^\infty(B_{\rho_{\nu-1}})} &\overset{(18)}{\lesssim} \left(1 - \frac{\rho_{\nu-1}}{\rho_\nu}\right)^{-n\frac{\theta}{q}} \|(1+|\nabla u|^2)^{\frac12}\|_{L^q(B(\rho_\nu))}^{\theta} \\ &\overset{(43)}{\le} \left(1 - \frac{\rho_{\nu-1}}{\rho_\nu}\right)^{-n\frac{\theta}{q}} \|(1+|\nabla u|^2)^{\frac12}\|_{L^p(B(\rho_\nu))}^{\frac{p}{q}\theta} \|(1+|\nabla u|^2)^{\frac12}\|_{L^\infty(B(\rho_\nu))}^{(1-\frac{p}{q})\theta} \\ &\le c 2^{(1+\nu)n\frac{\theta}{q}} \|(1+|\nabla u|^2)^{\frac12}\|_{L^p(B_1)}^{\frac{p}{q}\theta} \|(1+|\nabla u|^2)^{\frac12}\|_{L^\infty(B_{\rho_\nu})}^{(1-\frac{p}{q})\theta}, \end{aligned} \tag{44}$$

where $c = c(n, n, m, M, p, q, \theta) \in [1, \infty)$. Iterating (44) from $\nu = 1$ to $\hat\nu$, we obtain

$$\begin{aligned} &\|(1+|\nabla u|^2)^{\frac12}\|_{L^\infty(B_{\frac12})} \\ &= \|(1+|\nabla u|^2)^{\frac12}\|_{L^\infty(B_{\rho_0})} \\ &\overset{(44)}{\le} 2^{n\frac{\theta}{q}\sum_{\nu=0}^{\hat\nu-1}(\nu+1)((1-\gamma)\theta)^\nu} \big(c\|(1+|\nabla u|^2)^{\frac12}\|_{L^p(B_1)}^{\frac{p}{q}\theta}\big)^{\sum_{\nu=0}^{\hat\nu-1}((1-\frac{p}{q})\theta)^\nu} \|(1+|\nabla u|^2)^{\frac12}\|_{L^\infty(B_1)}^{((1-\frac{p}{q})\theta)^{\hat\nu}}. \end{aligned} \tag{45}$$

The choice of θ and assumption (4) imply

$$0 < \left(1 - \frac{p}{q}\right)\theta < 1. \tag{46}$$

Indeed, (46) is ensured for $n=3$ by the assumption $\theta < \frac{q}{q-p}$ and for $n \ge 4$ by

$$0 < \left(1 - \frac{p}{q}\right)\theta \overset{(17)}{=} \frac{2(q-p)}{(n-1)p-(n-3)q} = 1 - \frac{(n+1)p-(n-1)q}{(n-1)p-(n-3)q} \overset{(4)}{<} 1.$$

Hence,

$$\sum_{\nu=0}^\infty (\nu+1)\left(\left(1-\frac{p}{q}\right)\theta\right)^\nu \lesssim 1 \quad \text{and} \quad \sum_{\nu=0}^\infty \left(\left(1-\frac{p}{q}\right)\theta\right)^\nu = \frac{1}{1-\theta\big(1-\frac{p}{q}\big)}.$$

Thus, estimates (18) and (45) yield for every $\hat{\nu} \in \mathbb{N}$

$$\|(1+|\nabla u|^2)^{\frac{1}{2}}\|_{L^\infty(B_{\frac{1}{2}})} \overset{(45)}{\lesssim} \|(1+|\nabla u|^2)^{\frac{1}{2}}\|_{L^p(B_1)}^{\frac{\theta(p/q)}{1-\theta(1-p/q)}} \|(1+|\nabla u|^2)^{\frac{1}{2}}\|_{L^\infty(B_1)}^{((1-\frac{p}{q})\theta)^{\hat{\nu}}}$$
$$\overset{(18)}{\lesssim} \|(1+|\nabla u|^2)^{\frac{1}{2}}\|_{L^p(B_1)}^{\alpha} \|(1+|\nabla u|^2)^{\frac{1}{2}}\|_{L^q(B_2)}^{\theta((1-\frac{p}{q})\theta)^{\hat{\nu}}}.$$

Assumptions $u \in W^{1,q}_{\mathrm{loc}}(\Omega)$ and $B_2 \Subset \Omega$ imply $\|(1+|\nabla u|^2)^{\frac{1}{2}}\|_{L^q(B_2)} < \infty$ and thus we find $\hat{\nu} \in \mathbb{N}$ such that

$$\|(1+|\nabla u|^2)^{\frac{1}{2}}\|_{L^q(B_2)}^{\theta((1-\frac{p}{q})\theta)^{\hat{\nu}}} \leq 2,$$

which finishes the proof. □

4. Proof of Theorem 4

The main result of this section is:

Theorem 14. *Let $\Omega \subset \mathbb{R}^n$, $n \geq 3$, and suppose Assumption 1 is satisfied with $2 \leq p < q < \infty$ such that (4) holds. Let θ be given as in (17) with the additional constraint $\theta < \frac{q}{q-p}$ for $n = 3$. Let $u \in W^{1,1}_{\mathrm{loc}}(\Omega)$ be a local minimizer of the functional $\mathcal{F}$ given in (1). Then there exists $c = c(n, m, M, p, q, \theta) \in [1, \infty)$ such that for every $B_{2R}(x_0) \Subset \Omega$*

$$\|\nabla u\|_{L^\infty(B_{\frac{R}{2}}(x_0))} \leq c\left(\fint_{B_R(x_0)} f(\nabla u)\,dx + 1\right)^{\frac{\alpha}{p}},$$

where α is given in (41).

Proof of Theorem 4. Theorem 14 contains the claim of Theorem 4 in the case $n \geq 3$ and $2 \leq p < q$. The remaining case $n = 2$ follows from a combination of [Marcellini 1991, Theorem 2.1] and [Esposito et al. 1999, Theorem 2.1], and the result is classic for $p = q$. □

Appealing to the a priori estimate of Corollary 12, the statement of Theorem 14 follows from by now well-established approximation arguments. Below, we present a proof of Theorem 14 that closely follows [Esposito et al. 1999, proof of Theorem 2.1, Step 3].

Proof of Theorem 14. Throughout the proof we write $\lesssim$ if $\leq$ holds up to a positive constant which depends only on n, m, M, p, q and θ.

We assume $B_2 \Subset \Omega$ and show

$$\|\nabla u\|_{L^\infty(B_{\frac{1}{8}})} \lesssim \left(\fint_{B_1} f(\nabla u)\,dx + 1\right)^{\frac{\alpha}{p}}. \tag{47}$$

Clearly the general claim follows by standard scaling, translation and covering arguments.

Following [Esposito et al. 1999], we introduce two small parameters $\sigma, \varepsilon \in (0, 1)$. Parameter $\sigma > 0$ is related to a perturbation f_σ of the integrand f

$$f_\sigma(\xi) := f(\xi) + \sigma|\xi|^q \quad \text{for every } \xi \in \mathbb{R}^n. \tag{48}$$

Since f satisfies (2) and $\sigma \in (0,1)$, the function f_σ satisfies (2) with M replaced by M' depending on M and q. The second parameter $\varepsilon > 0$ corresponds to a regularization u_ε of u, where $u_\varepsilon := u * \varphi_\varepsilon$ with $\varphi_\varepsilon := \varepsilon^{-n}\varphi\big(\frac{\cdot}{\varepsilon}\big)$ and φ being a nonnegative, radially symmetric mollifier; i.e., it satisfies

$$\varphi \geq 0, \quad \operatorname{supp}\varphi \subset B_1, \quad \int_{\mathbb{R}^n} \varphi(x)\,dx = 1, \quad \varphi(\cdot) = \tilde{\varphi}(|\cdot|) \quad \text{for some } \tilde{\varphi} \in C^\infty(\mathbb{R}).$$

Given $\varepsilon, \sigma \in (0,1)$, we denote by $v_{\varepsilon,\sigma} \in u_\varepsilon + W_0^{1,q}(B_1)$ the unique function satisfying

$$\int_{B_1} f_\sigma(\nabla v_{\varepsilon,\sigma})\,dx \leq \int_{B_1} f_\sigma(\nabla v)\,dx \quad \text{for all } v \in u_\varepsilon + W_0^{1,q}(B_1). \tag{49}$$

In view of Corollary 12, we have

$$\begin{aligned}
\|\nabla v_{\varepsilon,\sigma}\|_{L^\infty(B_{\frac{1}{8}})} &\overset{(42)}{\lesssim} \left(\int_{B_{\frac{1}{4}}} |\nabla v_{\varepsilon,\sigma}|^p\,dx + 1\right)^{\frac{\alpha}{p}} \\
&\overset{(2)}{\lesssim} \left(\int_{B_1} f_\sigma(\nabla v_{\varepsilon,\sigma})\,dx + 1\right)^{\frac{\alpha}{p}} \\
&\overset{(48),(49)}{\leq} \left(\int_{B_1} f(\nabla u_\varepsilon) + \sigma|\nabla u_\varepsilon|^q\,dx + 1\right)^{\frac{\alpha}{p}} \\
&\leq \left(\int_{B_{1+\varepsilon}} f(\nabla u)\,dx + \sigma \int_{B_1} |\nabla u_\varepsilon|^q\,dx + 1\right)^{\frac{\alpha}{p}},
\end{aligned} \tag{50}$$

where we used Jensen's inequality and the convexity of f in the last step. Similarly,

$$\begin{aligned}
m\int_{B_1} |\nabla v_{\varepsilon,\sigma}|^p\,dx &\overset{(2)}{\leq} \int_{B_1} f(\nabla v_{\varepsilon,\sigma})\,dx \overset{(48),(49)}{\leq} \int_{B_1} f(\nabla u_\varepsilon) + \sigma|\nabla u_\varepsilon|^q\,dx \\
&\leq \int_{B_{1+\varepsilon}} f(\nabla u)\,dx + \sigma\int_{B_1} |\nabla u_\varepsilon|^q\,dx.
\end{aligned} \tag{51}$$

Fix $\varepsilon \in (0,1)$. In view of (50) and (51), we find $w_\varepsilon \in u_\varepsilon + W_0^{1,p}(B_1)$ such that as $\sigma \to 0$, up to subsequence,

$$\begin{aligned}
v_{\varepsilon,\sigma} &\rightharpoonup w_\varepsilon && \text{weakly in } W^{1,p}(B_1), \\
\nabla v_{\varepsilon,\sigma} &\overset{*}{\rightharpoonup} \nabla w_\varepsilon && \text{weakly* in } L^\infty(B_{\frac{1}{8}}).
\end{aligned}$$

Hence, a combination of (50), (51) with the weak/weak* lower-semicontinuity of convex functionals yields

$$\|\nabla w_\varepsilon\|_{L^\infty(B_{\frac{R}{8}})} \leq \liminf_{\sigma \to 0} \|\nabla v_{\varepsilon,\sigma}\|_{L^\infty(B_{\frac{1}{8}})} \lesssim \left(\int_{B_{1+\varepsilon}} f(\nabla u)\,dx + 1\right)^{\frac{\alpha}{p}}, \tag{52}$$

$$m\int_{B_1} |\nabla w_\varepsilon|^p\,dx \leq \int_{B_1} f(\nabla w_\varepsilon)\,dx \leq \int_{B_{1+\varepsilon}} f(\nabla u)\,dx. \tag{53}$$

Since $w_\varepsilon \in u_\varepsilon + W_0^{1,q}(B_1)$ and $u_\varepsilon \to u$ in $W^{1,p}(B_1)$, we find by (53) a function $w \in u + W_0^{1,p}(B_1)$ such that, up to subsequence,

$$\nabla w_\varepsilon \rightharpoonup \nabla w \quad \text{weakly in } L^p(B_1).$$

Appealing to the bounds (52), (53) and lower semicontinuity, we obtain

$$\|\nabla w\|_{L^\infty(B_{\frac{1}{8}})} \lesssim \left(\int_{B_{1+\varepsilon}} f(\nabla u)\,dx + 1\right)^{\frac{\alpha}{p}}, \tag{54}$$

$$\int_{B_1} f(\nabla w)\,dx \leq \int_{B_1} f(\nabla u)\,dx. \tag{55}$$

Inequality (55), the strong convexity of f and the fact $w \in u + W_0^{1,p}(B_1)$ imply $w = u$ and thus the claimed estimate (47) is a consequence of (54). □

Acknowledgment

The authors were supported by the German Science Foundation DFG in context of the Emmy Noether Junior Research Group BE 5922/1-1.

References

[Baroni et al. 2018] P. Baroni, M. Colombo, and G. Mingione, "Regularity for general functionals with double phase", *Calc. Var. Partial Differential Equations* **57**:2 (2018), art. id. 62, 48 pp. MR Zbl

[Beck and Mingione 2020] L. Beck and G. Mingione, "Lipschitz bounds and nonuniform ellipticity", *Comm. Pure Appl. Math.* **73**:5 (2020), 944–1034. MR Zbl

[Bella and Schäffner 2019] P. Bella and M. Schäffner, "Local boundedness and Harnack inequality for solutions of linear non-uniformly elliptic equations", *Comm. Pure Appl. Math.* (online publication November 2019).

[Bella and Schäffner 2020] P. Bella and M. Schäffner, "Quenched invariance principle for random walks among random degenerate conductances", *Ann. Probab.* **48**:1 (2020), 296–316. MR

[Bousquet and Brasco 2020] P. Bousquet and L. Brasco, "Lipschitz regularity for orthotropic functionals with nonstandard growth conditions", *Rev. Mat. Iberoam.* **36**:7 (2020), 1989–2032. MR

[Carozza et al. 2011] M. Carozza, J. Kristensen, and A. Passarelli di Napoli, "Higher differentiability of minimizers of convex variational integrals", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **28**:3 (2011), 395–411. MR Zbl

[Carozza et al. 2014] M. Carozza, J. Kristensen, and A. Passarelli di Napoli, "Regularity of minimizers of autonomous convex variational integrals", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **13**:4 (2014), 1065–1089. MR Zbl

[Colombo and Mingione 2015] M. Colombo and G. Mingione, "Regularity for double phase variational problems", *Arch. Ration. Mech. Anal.* **215**:2 (2015), 443–496. MR Zbl

[Cupini et al. 2015] G. Cupini, P. Marcellini, and E. Mascolo, "Local boundedness of minimizers with limit growth conditions", *J. Optim. Theory Appl.* **166**:1 (2015), 1–22. MR Zbl

[De Filippis and Mingione 2020] C. De Filippis and G. Mingione, "On the regularity of minima of non-autonomous functionals", *J. Geom. Anal.* **30**:2 (2020), 1584–1626. MR Zbl

[Eleuteri et al. 2020] M. Eleuteri, P. Marcellini, and E. Mascolo, "Regularity for scalar integrals without structure conditions", *Adv. Calc. Var.* **13**:3 (2020), 279–300. MR

[Esposito et al. 1999] L. Esposito, F. Leonetti, and G. Mingione, "Higher integrability for minimizers of integral functionals with (p, q) growth", *J. Differential Equations* **157**:2 (1999), 414–438. MR Zbl

[Esposito et al. 2004] L. Esposito, F. Leonetti, and G. Mingione, "Sharp regularity for functionals with (p,q) growth", *J. Differential Equations* **204**:1 (2004), 5–55. MR Zbl

[Franchi et al. 1998] B. Franchi, R. Serapioni, and F. Serra Cassano, "Irregular solutions of linear degenerate elliptic equations", *Potential Anal.* **9**:3 (1998), 201–216. MR Zbl

[Fusco and Sbordone 1993] N. Fusco and C. Sbordone, "Some remarks on the regularity of minima of anisotropic integrals", *Comm. Partial Differential Equations* **18**:1-2 (1993), 153–167. MR Zbl

[Giaquinta 1987] M. Giaquinta, "Growth conditions and regularity, a counterexample", *Manuscripta Math.* **59**:2 (1987), 245–248. MR Zbl

[Gilbarg and Trudinger 1998] D. Gilbarg and N. Trudinger, *Elliptic partial differential equations of second order*, 2nd ed., Springer, 1998.

[Giusti 2003] E. Giusti, *Direct methods in the calculus of variations*, World Scientific, River Edge, NJ, 2003. MR Zbl

[Hong 1992] M. C. Hong, "Some remarks on the minimizers of variational integrals with nonstandard growth conditions", *Boll. Un. Mat. Ital. A* (7) **6**:1 (1992), 91–101. MR

[Lieberman 1991] G. M. Lieberman, "The natural generalization of the natural conditions of Ladyzhenskaya and Ural'tseva for elliptic equations", *Comm. Partial Differential Equations* **16**:2-3 (1991), 311–361. MR Zbl

[Marcellini 1989] P. Marcellini, "Regularity of minimizers of integrals of the calculus of variations with nonstandard growth conditions", *Arch. Rational Mech. Anal.* **105**:3 (1989), 267–284. MR Zbl

[Marcellini 1991] P. Marcellini, "Regularity and existence of solutions of elliptic equations with p,q-growth conditions", *J. Differential Equations* **90**:1 (1991), 1–30. MR Zbl

[Marcellini 1993] P. Marcellini, "Regularity for elliptic equations with general growth conditions", *J. Differential Equations* **105**:2 (1993), 296–333. MR Zbl

[Mingione 2006] G. Mingione, "Regularity of minima: an invitation to the dark side of the calculus of variations", *Appl. Math.* **51**:4 (2006), 355–426. MR Zbl

[Moser 1960] J. Moser, "A new proof of De Giorgi's theorem concerning the regularity problem for elliptic differential equations", *Comm. Pure Appl. Math.* **13** (1960), 457–468. MR Zbl

[Murthy and Stampacchia 1968] M. K. V. Murthy and G. Stampacchia, "Boundary value problems for some degenerate-elliptic operators", *Ann. Mat. Pura Appl.* (4) **80** (1968), 1–122. MR Zbl

[Trudinger 1971] N. S. Trudinger, "On the regularity of generalized solutions of linear, non-uniformly elliptic equations", *Arch. Rational Mech. Anal.* **42** (1971), 50–62. MR Zbl

Received 18 May 2019. Revised 10 Jul 2019. Accepted 6 Sep 2019.

PETER BELLA: peter.bella@math.tu-dortmund.de
Fakultät für Mathematik, Technische Universität Dortmund, Dortmund, Germany

MATHIAS SCHÄFFNER: math.schaeffner@gmail.com
Mathematisches Institut, Universität Leipzig, Leipzig, Germany

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 13 No. 7 2020

Refined mass-critical Strichartz estimates for Schrödinger operators 1955
CASEY JAO

Scattering for defocusing energy subcritical nonlinear wave equations 1995
BENJAMIN DODSON, ANDREW LAWRIE, DANA MENDELSON and JASON MURPHY

New formulas for the Laplacian of distance functions and applications 2091
FABIO CAVALLETTI and ANDREA MONDINO

Convex sets evolving by volume-preserving fractional mean curvature flows 2149
ELEONORA CINTI, CARLO SINESTRARI and ENRICO VALDINOCI

C^*-algebras isomorphically representable on l^p 2173
MARCH T. BOEDIHARDJO

Exponential convergence of parabolic optimal transport on bounded domains 2183
FARHAN ABEDIN and JUN KITAGAWA

Nuclear dimension of simple stably projectionless C*-algebras 2205
JORGE CASTILLEJOS and SAMUEL EVINGTON

On the regularity of minimizers for scalar integral functionals with (p, q)-growth 2241
PETER BELLA and MATHIAS SCHÄFFNER