ANALYSIS & PDE

Volume 17 No. 3 2024

msp

Analysis & PDE
msp.org/apde

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

The subscription price for 2024 is US $440/year for the electronic version, and $690/year (+$65, 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.

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. 17 (2024), No. 3, pp. 757–830
DOI: 10.2140/apde.2024.17.757

SCHAUDER ESTIMATES FOR EQUATIONS WITH CONE METRICS, II

BIN GUO AND JIAN SONG

We continue our work on the linear theory for equations with conical singularities. We derive interior Schauder estimates for linear elliptic and parabolic equations with a background Kähler metric of conical singularities along a divisor of simple normal crossings. As an application, we prove the short-time existence of the conical Kähler–Ricci flow with conical singularities along a divisor with simple normal crossings.

1. Introduction 757
2. Preliminaries 763
3. Elliptic estimates 766
4. Parabolic estimates 807
5. Conical Kähler–Ricci flow 825
References 828

1. Introduction

This is a continuation of our paper [20]. Regularity of solutions of complex Monge–Ampère equations is a central problem in complex geometry. Complex Monge–Ampère equations with singular and degenerate data can be applied to study compactness and moduli problems of canonical Kähler metrics in Kähler geometry. In [43], Yau considered special cases of complex singular Monge–Ampère equations as generalizations of his solution to the Calabi conjecture. Conical singularities along complex hypersurfaces of a Kähler manifold are among the mildest singularities in Kähler geometry, and they have been extensively studied, especially in the case of Riemann surfaces [28; 41]. The study of such Kähler metrics with conical singularities has many geometric applications, for example, the Chern number inequality in various settings [38; 39]. Recently, Donaldson [14] initiated the program of studying analytic and geometric properties of Kähler metrics with conical singularities along a smooth complex hypersurface on a Kähler manifold. This is an essential step in the solution of the Yau–Tian–Donaldson conjecture relating existence of Kähler–Einstein metrics and algebraic K-stability on Fano manifolds [7; 8; 9; 40]. In [14], the Schauder estimate for linear Laplace equations with the conical background metric is established using classical potential theory. This is crucial for the openness of the continuity method to find a desirable

Research supported in part by the National Science Foundation grants DMS-17-11439 and DMS-17-10500. Both authors thank Duong H. Phong and Ved Datar for many insightful discussions.
MSC2010: 35J15, 35K10, 35K67.
Keywords: conic Kähler metrics, Schauder estimates.

(conical) Kähler–Einstein metric. Donaldson's Schauder estimate is generalized to the parabolic case [5] with a similar classical approach. There is also an alternative approach for the conical Schauder estimates using microlocal analysis [23]. Various global and local estimates and regularity are also derived in the conical setting [1; 6; 11; 12; 13; 15; 19; 24; 29; 32; 44; 45].

The Schauder estimates play an important role in linear PDE theory. Apart from the classical potential theory, various proofs have been established by different analytic techniques. In fact, the blow-up or perturbation techniques developed in [36; 42] (also see [2; 3; 33; 34]) are much more flexible and sharper than the classical method. The authors combined the perturbation method in [20] and geometric gradient estimates to establish sharp Schauder estimates for Laplace equations and heat equations on $\mathbb{C}^n$ with a background flat Kähler metric of conical singularities along the smooth hyperplane $\{z_1 = 0\}$ and derived explicit and optimal dependence on conical parameters.

In algebraic geometry, one often has to consider pairs (X, D) with X an algebraic variety of complex dimension n and the boundary divisor D a complex hypersurface of X. After possible log resolution, one can always assume the divisor D is a union of smooth hypersurfaces with simple normal crossings. The suitable category of Kähler metrics associated to (X, D) is the family of Kähler metrics on X with conical singularities along D. In order to study canonical Kähler metrics on pairs and related moduli problems, we are obliged to study regularity and asymptotics for complex Monge–Ampère equation with prescribed conical singularities of normal crossings. However, the linear theory is still missing and has been open for a while. The goal of this paper is to extend our result [20] and establish the sharp Schauder estimates for linear equations with background Kähler metric of conical singularities along divisors of simple normal crossings. We can apply and extend many techniques developed in [20]; however, new estimates and techniques have to be developed because, in the case of conical singularities along a single smooth divisor, the difficult estimate in the conical direction can sometimes be bypassed and reduced to estimates in the regular directions, while such treatment does not work in the case of simple normal crossings. One is forced to treat regions near high codimensional singularities directly with new and more delicate estimates beyond the scope of [20]. More crucially, the estimates in the mixed normal directions (see Section 3D) relies on those in Lemma 3.3, which is new compared to the case of a smooth divisor [20]. This enables us to compare the difference of mixed normal derivatives at two different points. Readers who are interested only in the case of smooth divisors are advised to omit Section 3D.

The standard local models for such conical Kähler metrics are described below.

Let $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p) \in (0, 1)^p$, $p \le n$, and let $\omega_{\boldsymbol{\beta}}$ (or $g_{\boldsymbol{\beta}}$) be the standard cone metric on $\mathbb{C}^p \times \mathbb{C}^{n-p}$ with cone singularity along $\mathcal{S} = \bigcup_{i=1}^p \mathcal{S}_i$, where $\mathcal{S}_i = \{z_i = 0\}$, that is,

$$\omega_{\boldsymbol{\beta}} = \sum_{j=1}^{p} \beta_j^2 \frac{\sqrt{-1}\, dz_j \wedge d\bar{z}_j}{|z_j|^{2(1-\beta_j)}} + \sum_{j=p+1}^{n} \sqrt{-1}\, dz_j \wedge d\bar{z}_j. \tag{1-1}$$

We shall use $s_{2p+1}, \dots, s_{2n}$ to denote the real coordinates of $\mathbb{C}^{n-p} = \mathbb{R}^{2n-2p}$ such that, for $j = p+1, \dots, n$,

$$z_j = s_{2j-1} + \sqrt{-1} s_{2j}.$$

In this paper we will study the conical Laplacian equation with the background metric $g_{\boldsymbol{\beta}}$ on $\mathbb{C}^n$

$$\Delta_{\boldsymbol{\beta}} u = f \quad \text{in } B_{\boldsymbol{\beta}}(0,1)\backslash\mathcal{S}, \tag{1-2}$$

where $B_{\boldsymbol{\beta}}(0,1)$ is the unit ball with respect to $g_{\boldsymbol{\beta}}$ centered at 0. The Laplacian $\Delta_{\boldsymbol{\beta}}$ is defined as

$$\Delta_{\boldsymbol{\beta}} u = \sum_{j,k} g_{\boldsymbol{\beta}}^{j\bar{k}} \frac{\partial^2 u}{\partial z_j \partial \bar{z}_k} = \sum_{j=1}^{p} \beta_j^{-1} |z_j|^{2(1-\beta_j)} \frac{\partial^2 u}{\partial z_j \partial \bar{z}_j} + \sum_{j=p+1}^{n} \frac{\partial^2 u}{\partial z_j \partial \bar{z}_j}.$$

We always assume

$$f \in C^0(B_{\boldsymbol{\beta}}(0,1)) \quad \text{and} \quad u \in C^0(\overline{B_{\boldsymbol{\beta}}(0,1)}) \cap C^2(B_{\boldsymbol{\beta}}(0,1)\backslash\mathcal{S}).$$

Throughout this paper, given a continuous function f, we write

$$\omega(r) := \omega_f(r) = \sup_{\substack{z,w \in B_{\boldsymbol{\beta}}(0,1) \\ d_{\boldsymbol{\beta}}(z,w) < r}} |f(z) - f(w)|$$

for the oscillation of f with respect to $g_{\boldsymbol{\beta}}$ in the ball $B_{\boldsymbol{\beta}}(0,1)$. It is clear that $\omega(2r) \le 2\omega(r)$ for any $r < \frac{1}{2}$. We say a continuous function f is *Dini continuous* if $\int_0^1 \omega(r)/r\,dr < \infty$.

Definition 1.1. We will write the (weighted) polar coordinates of z_j for $1 \le j \le p$ as

$$r_j = |z_j|^{\beta_j}, \quad \theta_j = \arg z_j.$$

We define D' to be one of the first-order operators $\{\partial/\partial s_{2p+1}, \dots, \partial/\partial s_{2n}\}$, and N_j to be one of the operators $\{\partial/\partial r_j, (\beta_j r_j)^{-1}(\partial/\partial\theta_j)\}$ which as vector fields are transversal to $\mathcal{S}_j$.

Our first main result is the Hölder estimates of the solution u to (1-2).

Theorem 1.2. *Suppose $\boldsymbol{\beta} \in \left(\frac{1}{2}, 1\right)^p$ and $f \in C^0(B_{\boldsymbol{\beta}}(0,1))$ is Dini continuous with respect to $g_{\boldsymbol{\beta}}$. Let $u \in C^0(\overline{B_{\boldsymbol{\beta}}(0,1)}) \cap C^2(B_{\boldsymbol{\beta}}(0,1)\backslash\mathcal{S})$ be the solution to* (1-2). *Then there exists $C = C(n, \boldsymbol{\beta}) > 0$ such that, for any two points $p, q \in B_{\boldsymbol{\beta}}\left(0, \frac{1}{2}\right)\backslash\mathcal{S}$,*

$$\begin{aligned} |(D')^2 u(p) - (D')^2 u(q)| + \sum_{j=1}^{p} \left| |z_j|^{2(1-\beta_j)} \frac{\partial^2 u}{\partial z_j \partial \bar{z}_j}(p) - |z_j|^{2(1-\beta_j)} \frac{\partial^2 u}{\partial z_j \partial \bar{z}_j}(q) \right| \\ \le C\left(d\|u\|_{L^\infty(B_{\boldsymbol{\beta}}(0,1))} + \int_0^d \frac{\omega(r)}{r}\,dr + d\int_d^1 \frac{\omega(r)}{r^2}\,dr \right), \end{aligned} \tag{1-3}$$

for any $1 \le j \le p$,

$$|N_j D' u(p) - N_j D' u(q)| \le C\left(d^{1/\beta_j - 1}\|u\|_{L^\infty(B_{\boldsymbol{\beta}}(0,1))} + \int_0^d \frac{\omega(r)}{r}\,dr + d^{1/\beta_j - 1}\int_d^1 \frac{\omega(r)}{r^{1/\beta_j}}\,dr \right), \tag{1-4}$$

and, for any $1 \le j, k \le p$ with $j \ne k$,

$$|N_j N_k u(p) - N_j N_k u(q)| \le C\left(d^{1/\beta_{\max} - 1}\|u\|_{L^\infty(B_{\boldsymbol{\beta}}(0,1))} + \int_0^d \frac{\omega(r)}{r}\,dr + d^{1/\beta_{\max} - 1}\int_d^1 \frac{\omega(r)}{r^{1/\beta_{\max}}}\,dr \right), \tag{1-5}$$

where $d = d_{\boldsymbol{\beta}}(p,q) > 0$ is the $g_{\boldsymbol{\beta}}$-distance of p and q and $\beta_{\max} = \max\{\beta_1, \dots, \beta_p\} \in \left(\frac{1}{2}, 1\right)$.

Remarks 1.3. (1) The number $\beta_{\max}$ on the right-hand side of (1-5) can be replaced by $\max\{\beta_j, \beta_k\}$.

(2) We assume $\boldsymbol{\beta} \in \left(\frac{1}{2}, 1\right)^p$ for the purposes of exposition and simplification of the statements of Theorems 1.2 and 1.7. When some of the angles β_j lie in $\left(0, \frac{1}{2}\right]$, the pointwise Hölder estimates in Theorem 1.2 are adjusted as follows: in (1-4), if $\beta_j \in \left(0, \frac{1}{2}\right]$, we replace the right-hand side by the right-hand side of (1-3); in (1-5), if both β_j and $\beta_k \in \left(0, \frac{1}{2}\right]$, we also replace the right-hand side by that of (1-3); if at least one of the β_j, β_k is bigger than $\frac{1}{2}$, (1-5) remains unchanged. The inequalities in Theorem 1.7 can be adjusted similarly. The proofs of these estimates are contained in the proof of the case when $\beta_j \in \left(\frac{1}{2}, 1\right)$ by using the corresponding estimates in (2-3).

An immediate corollary of Theorem 1.2 is a precise form of Schauder estimates for (1-2).

Corollary 1.4. *Given $\boldsymbol{\beta} \in (0,1)^p$ and $f \in C^{0,\alpha}_{\boldsymbol{\beta}}(\overline{B_{\boldsymbol{\beta}}(0,1)})$ for some $0 < \alpha < \min\{1, 1/\beta_{\max} - 1\}$, if $u \in C^0(B_{\boldsymbol{\beta}}(0,1)) \cap C^2(B_{\boldsymbol{\beta}}(0,1)\backslash\mathcal{S})$ solves* (1-2), *then $u \in C^{2,\alpha}_{\boldsymbol{\beta}}(B_{\boldsymbol{\beta}}(0,1))$. Moreover, for any compact subset $K \Subset B_{\boldsymbol{\beta}}(0,1)$, there exists a constant $C = C(n, \boldsymbol{\beta}, K) > 0$ such that the following estimate holds* (*see Definition 2.1 for the notations*):

$$\|u\|_{C^{2,\alpha}_{\boldsymbol{\beta}}(K)} \le C\left(\|u\|_{C^0(B_{\boldsymbol{\beta}}(0,1))} + \frac{\|f\|_{C^{0,\alpha}_{\boldsymbol{\beta}}(B_{\boldsymbol{\beta}}(0,1))}}{\alpha\left(\min\left\{\frac{1}{\beta_{\max}} - 1, 1\right\} - \alpha\right)}\right). \tag{1-6}$$

Remark 1.5. A scaling-invariant version of the Schauder estimate (1-6) is that, for any $0 < r < 1$, there exists a constant $C = C(n, \boldsymbol{\beta}, \alpha) > 0$ such that (see Definition 2.4 for the notations)

$$\|u\|^{*}_{C^{2,\alpha}_{\boldsymbol{\beta}}(B_{\boldsymbol{\beta}}(0,r))} \le C(\|u\|_{C^0(B_{\boldsymbol{\beta}}(0,r))} + \|f\|^{(2)}_{C^{0,\alpha}_{\boldsymbol{\beta}}(B_{\boldsymbol{\beta}}(0,r))}), \tag{1-7}$$

which follows from a standard rescaling argument by scaling r to 1.

Let g be a $C^{0,\alpha}_{\boldsymbol{\beta}}$-conical Kähler metric on $B_{\boldsymbol{\beta}}(0,1)$ (see Definition 3.31). By definition g is equivalent to $g_{\boldsymbol{\beta}}$. We consider the equation

$$\Delta_g u = f \quad \text{in } B_{\boldsymbol{\beta}}(0,1) \qquad \text{and} \qquad u = \varphi \quad \text{on } \partial B_{\boldsymbol{\beta}}(0,1) \tag{1-8}$$

for some $\varphi \in C^0(\partial B_{\boldsymbol{\beta}}(0,1))$. The following theorem is the generalization of Corollary 1.4 for nonflat background conical Kähler metrics, which is useful for applications of global geometric complex Monge–Ampère equations.

Theorem 1.6. *For any given $\boldsymbol{\beta} \in (0,1)^p$, $f \in C^{0,\alpha}_{\boldsymbol{\beta}}(\overline{B_{\boldsymbol{\beta}}(0,1)})$ and $\varphi \in C^0(\partial B_{\boldsymbol{\beta}}(0,1))$, there is a unique solution $u \in C^{2,\alpha}_{\boldsymbol{\beta}}(B_{\boldsymbol{\beta}}(0,1)) \cap C^0(\overline{B_{\boldsymbol{\beta}}(0,1)})$ to* (1-8). *Moreover, for any compact subset $K \Subset B_{\boldsymbol{\beta}}(0,1)$, there exists $C = C(n, \boldsymbol{\beta}, \alpha, g, K) > 0$ such that*

$$\|u\|_{C^{2,\alpha}_{\boldsymbol{\beta}}(K)} \le C(\|u\|_{C^0(B_{\boldsymbol{\beta}}(0,1))} + \|f\|_{C^{0,\alpha}_{\boldsymbol{\beta}}(B_{\boldsymbol{\beta}}(0,1))}).$$

Theorem 1.6 can immediately be applied to study complex Monge–Ampère equations with prescribed conical singularities along divisors of simple normal crossings, and most of the geometric and analytic results for canonical Kähler metrics with conical singularities along a smooth divisor can be generalized to the case of simple normal crossings.

We now turn to the parabolic Schauder estimates for the solution $u \in \mathscr{C}^0(\mathcal{Q}_{\boldsymbol\beta}) \cap \mathscr{C}^2(\mathcal{Q}^\#_{\boldsymbol\beta})$ to the equation

$$\frac{\partial u}{\partial t} = \Delta_{g_{\boldsymbol\beta}} u + f \tag{1-9}$$

for a Dini continuous function f in $\mathcal{Q}_{\boldsymbol\beta}$, where for convenience of notation we write

$$\mathcal{Q}_{\boldsymbol\beta} := B_{\boldsymbol\beta}(0,1) \times (0,1] \quad \text{and} \quad \mathcal{Q}^\#_{\boldsymbol\beta} := B_{\boldsymbol\beta}(0,1) \backslash \mathcal{S} \times (0,1].$$

Our second main theorem is the following pointwise estimate.

Theorem 1.7. *Suppose $\boldsymbol\beta \in \left(\frac{1}{2}, 1\right)^p$ and u is the solution to* (1-9). *Then there exists a computable constant $C = C(n, \boldsymbol\beta) > 0$ such that, for any $Q_p = (p, t_p)$, $Q_q = (q, t_q) \in B_{\boldsymbol\beta}\left(0, \frac{1}{2}\right) \backslash \mathcal{S} \times (\hat t, 1]$ (for some $\hat t \in (0,1)$),*

$$\begin{aligned} &|(D')^2 u(Q_p) - (D')^2 u(Q_q)| \\ &\qquad + \sum_{j=1}^p \left| |z_j|^{2(1-\beta_j)} \frac{\partial^2 u}{\partial z_j \partial \bar z_j}(Q_p) - |z_j|^{2(1-\beta_j)} \frac{\partial^2 u}{\partial z_j \partial \bar z_j}(Q_q) \right| + \left| \frac{\partial u}{\partial t}(Q_p) - \frac{\partial u}{\partial t}(Q_q) \right| \\ &\qquad\qquad \le C\left(\frac{d}{\hat t^{3/2}} \|u\|_{L^\infty(B_{\boldsymbol\beta}(0,1))} + \hat t^{-1} \int_0^d \frac{\omega(r)}{r}\, dr + \frac{d}{\hat t^{3/2}} \int_d^1 \frac{\omega(r)}{r^2}\, dr \right), \end{aligned}$$

and, for any $1 \le j \le p$,

$$|N_j D' u(Q_p) - N_j D' u(Q_q)| \le C\left(\frac{d^{1/\beta_j - 1}}{\hat t^{3/2}} \|u\|_{L^\infty(B_{\boldsymbol\beta}(0,1))} + \hat t^{-1} \int_0^d \frac{\omega(r)}{r}\, dr + \frac{d^{1/\beta_j - 1}}{\hat t^{3/2}} \int_d^1 \frac{\omega(r)}{r^{1/\beta_j}}\, dr \right),$$

and, for any $1 \le j, k \le p$ with $j \ne k$,

$$|N_j N_k u(Q_p) - N_j N_k u(Q_q)| \le C\left(\frac{d^{1/\beta_{\max} - 1}}{\hat t^{3/2}} \|u\|_{L^\infty(B_{\boldsymbol\beta}(0,1))} + \hat t^{-1} \int_0^d \frac{\omega(r)}{r}\, dr + \frac{d^{1/\beta_{\max} - 1}}{\hat t^{3/2}} \int_d^1 \frac{\omega(r)}{r^{1/\beta_{\max}}}\, dr \right),$$

where $d = d_{\mathcal{P},\boldsymbol\beta}(Q_p, Q_q) > 0$ is the parabolic $g_{\boldsymbol\beta}$-distance of Q_p and Q_q, $\beta_{\max} = \max\{\beta_1, \dots, \beta_p\}$, and $\omega(r)$ is the oscillation of f in $\mathcal{Q}_{\boldsymbol\beta}$ under the parabolic distance $d_{\mathcal{P},\boldsymbol\beta}$ (see Section 2A2).

If $f \in \mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(\mathcal{Q}_{\boldsymbol\beta})$ for some $\alpha \in (0, \min(1/\beta_{\max} - 1, 1))$, then we have the following precise estimates as the parabolic analogue of Corollary 1.4.

Corollary 1.8. *Suppose $\boldsymbol\beta \in (0,1)^p$ and $u \in \mathscr{C}^0(\mathcal{Q}_{\boldsymbol\beta}) \cap \mathscr{C}^2(\mathcal{Q}^\#_{\boldsymbol\beta})$ satisfies* (1-9). *Then there exists a constant $C = C(n, \boldsymbol\beta) > 0$ such that (see Definition 2.6 for the notations)*

$$\|u\|_{\mathscr{C}^{2+\alpha,(\alpha+2)/2}_{\boldsymbol\beta}(B_{\boldsymbol\beta}(0,1/2)\times(1/2,1])} \le C\left(\|u\|_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol\beta})} + \frac{\|f\|_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(\mathcal{Q}_{\boldsymbol\beta})}}{\alpha\left(\min\left\{\frac{1}{\beta_{\max}} - 1, 1\right\} - \alpha\right)} \right).$$

For general nonflat $\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}$-conical Kähler metrics g, we consider the linear parabolic equation

$$\frac{\partial u}{\partial t} = \Delta_g u + f \quad \text{in } \mathcal{Q}_{\boldsymbol\beta}, \qquad u = \varphi \quad \text{on } \partial_{\mathcal{P}} \mathcal{Q}_{\boldsymbol\beta}. \tag{1-10}$$

We then have the following parabolic Schauder estimates as an analogue of Theorem 1.6.

Theorem 1.9. *Given $\boldsymbol{\beta} \in (0,1)^p$, $f \in \mathscr{C}_{\boldsymbol{\beta}}^{\alpha,\alpha/2}(\overline{\mathcal{Q}_{\boldsymbol{\beta}}})$ and $\varphi \in \mathscr{C}^0(\partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol{\beta}})$, there exists a unique solution $u \in \mathscr{C}_{\boldsymbol{\beta}}^{2+\alpha,(\alpha+2)/2}(B_{\boldsymbol{\beta}}(0,1)\times(0,1]) \cap \mathscr{C}^0(\overline{\mathcal{Q}_{\boldsymbol{\beta}}})$ to the Dirichlet boundary value problem* (1-10). *For any compact subset $K \Subset B_{\boldsymbol{\beta}}(0,1)$ and $\varepsilon_0 > 0$, there exists $C = C(n,\boldsymbol{\beta},\alpha,K,\varepsilon_0,g) > 0$ such that the following interior Schauder estimate holds*:

$$\|u\|_{\mathscr{C}_{\boldsymbol{\beta}}^{2+\alpha,(2+\alpha)/2}(K\times[\varepsilon_0,1])} \leq C(\|u\|_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol{\beta}})} + \|f\|_{\mathscr{C}_{\boldsymbol{\beta}}^{\alpha,\alpha/2}(\mathcal{Q}_{\boldsymbol{\beta}})}).$$

Furthermore, if we assume $u|_{t=0} = u_0 \in C_{\boldsymbol{\beta}}^{2,\alpha}(B_{\boldsymbol{\beta}}(0,1))$, then $u \in \mathscr{C}_{\boldsymbol{\beta}}^{2+\alpha,(\alpha+2)/2}(B_{\boldsymbol{\beta}}(0,1)\times[0,1])$, and there exists a constant $C = C(n,\boldsymbol{\beta},\alpha,g,K) > 0$ such that

$$\|u\|_{\mathscr{C}_{\boldsymbol{\beta}}^{2+\alpha,(\alpha+2)/2}(K\times[0,1])} \leq C(\|u\|_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol{\beta}})} + \|f\|_{\mathscr{C}_{\boldsymbol{\beta}}^{\alpha,\alpha/2}(\mathcal{Q}_{\boldsymbol{\beta}})} + \|u_0\|_{C_{\boldsymbol{\beta}}^{2,\alpha}(B_{\boldsymbol{\beta}}(0,1))}).$$

As an application of Theorem 1.9, we derive the short-time existence of the conical Kähler–Ricci flow with background metric being conical along divisors with simple normal crossings.

Let (X, D) be a compact Kähler manifold, where $D = \sum_j D_j$ is a finite union of smooth divisors $\{D_j\}$ and D has only simple normal crossings. Let ω_0 be a $C_{\boldsymbol{\beta}}^{0,\alpha'}(X)$-conical Kähler metric with cone angle $2\pi\boldsymbol{\beta}$ along D (see Definition 2.8), let $\hat{\omega}_t$ be a family of conical metrics with bounded norm $\|\hat{\omega}\|_{\mathscr{C}_{\boldsymbol{\beta}}^{\alpha',\alpha'/2}}$, and let $\hat{\omega}_0 = \omega_0$. We consider the complex Monge–Ampère flow

$$\frac{\partial\varphi}{\partial t} = \log\left(\frac{(\hat{\omega}_t + \sqrt{-1}\partial\bar{\partial}\varphi)^n}{\omega_0^n}\right) + f \quad \text{and} \quad \varphi|_{t=0} = 0 \tag{1-11}$$

for some $f \in \mathscr{C}_{\boldsymbol{\beta}}^{\alpha',\alpha'/2}(X\times[0,1])$.

Theorem 1.10. *Given $\alpha \in (0,\alpha')$, there exists $T = T(n,\hat{\omega},f,\alpha',\alpha) > 0$ such that* (1-11) *admits a unique solution $\varphi \in \mathscr{C}_{\boldsymbol{\beta}}^{2+\alpha,(2+\alpha)/2}(X\times[0,T])$.*

An immediate corollary of Theorem 1.10 is the short-time existence for the conical Kähler–Ricci flow

$$\frac{\partial\omega}{\partial t} = -\operatorname{Ric}(\omega) + \sum_j (1-\beta_j)[D_j], \quad \omega|_{t=0} = \omega_0, \tag{1-12}$$

where $\operatorname{Ric}(\omega)$ is the unique extension of the Ricci curvature of ω from $X \setminus D$ to X, and $[D_j]$ denotes the current of integration over the component D_j. In addition we assume ω_0 is a $C_{\boldsymbol{\beta}}^{0,\alpha'}(X,D)$-conical Kähler metric such that

$$\omega_0^n = \frac{\Omega}{\prod_j (|s_j|_{h_j}^2)^{1-\beta_j}}, \tag{1-13}$$

where s_j and h_j are holomorphic sections and hermitian metrics, respectively, of the line bundle associated to D_j, and Ω is a smooth volume form.

Corollary 1.11. *For any given $\alpha \in (0,\alpha')$, there exists a constant $T = T(n,\omega_0,\alpha,\alpha') > 0$ such that the conical Kähler–Ricci flow* (1-12) *admits a unique solution $\omega = \omega_t$, where $\omega \in \mathscr{C}_{\boldsymbol{\beta}}^{\alpha,\alpha/2}(X\times[0,T])$ and, for each $t \in [0,T]$, ω_t is still a conical metric with cone angle $2\pi\boldsymbol{\beta}$ along D.*

Furthermore, ω is smooth in $X\backslash D \times (0,T]$ and the (*normalized*) *Ricci potentials of ω, by which we mean $\log(\omega^n/\omega_0^n)$, are still in $\mathscr{C}_{\boldsymbol{\beta}}^{2+\alpha,(2+\alpha)/2}(X\times[0,T])$.*

The short-time existence of the conical Kähler–Ricci flow with singularities along a smooth divisor is derived in [5] by adapting the elliptic potential techniques of Donaldson [14]. Corollary 1.11 treats the general case of conical singularities with simple normal crossings. There have been many results in the analytic aspects of the conical Ricci flow [5; 6; 15; 16; 24; 30; 43]. In [31], the conical Ricci flow on Riemann surfaces is completely classified with jumping conical structure in the limit. Such phenomena is also expected in higher dimension, but it requires much deeper and delicate technical advances both in analysis and geometry.

2. Preliminaries

We explain the notations and give some preliminary tools which will be used later in this section.

2A. ***Notations.*** To distinguish the elliptic from parabolic norms, we will use C to denote the norms in the elliptic case and $\mathscr{C}$ to denote the norms in the parabolic case.

We always assume the Hölder component α appearing in $C^{0,\alpha}_{\beta}$ or $\mathscr{C}^{\alpha,\alpha/2}_{\beta}$ (or other Hölder norms) is in $(0, \min\{\beta_{\max}^{-1} - 1, 1\})$.

2A1. *Elliptic case.* We will denote $d_{\beta}(x, y)$ to be the distance between two points $x, y \in \mathbb{C}^n$ under the metric g_{β}. $B_{\beta}(x, r)$ will be the metric ball under the metric induced by g_{β} with radius r and center x. It is well known that $(\mathbb{C}^n \backslash \mathcal{S}, g_{\beta})$ is geodesically convex, i.e., any two points $x, y \in \mathbb{C}^n \backslash \mathcal{S}$ can be joined by a g_{β}-minimal geodesic γ which is disjoint from $\mathcal{S}$.

Definition 2.1. We define the g_{β}-Hölder norm of functions $u \in C^0(B_{\beta}(0, r))$ for some $\alpha \in (0, 1)$ as

$$\|u\|_{C^{0,\alpha}_{\beta}(B_{\beta}(0,r))} := \|u\|_{C^0(B_{\beta}(0,r))} + [u]_{C^{0,\alpha}_{\beta}(B_{\beta}(0,r))},$$

where the seminorm is defined as

$$[u]_{C^{0,\alpha}_{\beta}(B_{\beta}(0,r))} := \sup_{x \neq y \in B_{\beta}(0,r)} \frac{|u(x) - u(y)|}{d_{\beta}(x, y)^{\alpha}}.$$

We denote by $C^{0,\alpha}_{\beta}(B_{\beta}(0, r))$ the subspace of all continuous functions u such that $\|u\|_{C^{0,\alpha}_{\beta}} < \infty$.

Definition 2.2. The $C^{2,\alpha}_{\beta}$-norm of a function u on $B_{\beta}(0, r) =: B_{\beta}$ is defined as

$$\begin{aligned}\|u\|_{C^{2,\alpha}_{\beta}(B_{\beta})} := \|u\|_{C^0(B_{\beta})} + \|\nabla_{g_{\beta}} u\|_{C^0(B_{\beta}, g_{\beta})} + \sum_{j=1}^{p} \|N_j D' u\|_{C^{0,\alpha}_{\beta}(B_{\beta})} \\ + \|(D')^2 u\|_{C^{0,\alpha}_{\beta}(B_{\beta})} + \sum_{1 \le j \neq k \le p} \|N_j N_k u\|_{C^{0,\alpha}_{\beta}(B_{\beta})} + \sum_{j=1}^{p} \left\| |z_j|^{2(1-\beta_j)} \frac{\partial^2 u}{\partial z_j \partial \bar{z}_j} \right\|_{C^{0,\alpha}_{\beta}(B_{\beta})}.\end{aligned}$$

We denote by $C^{2,\alpha}_{\beta}(B_{\beta}(0, r))$ the subspace of all continuous functions u such that $\|u\|_{C^{2,\alpha}_{\beta}} < \infty$.

Remark 2.3. These spaces are generalizations of those defined in [20] and are slight variations of those introduced in [14; 23].

Let us compare the Schauder estimates in [14; 23; 20] in the special case when $p = 1$, i.e., the conical singularities are supported on $\mathbb{C}^{n-1}$. The Hölder space of [14] is defined using a collection of differential operators as components of $\sqrt{-1}\partial\bar\partial$. The collection of differential operators in our definition for $C^{0,\alpha}_{\boldsymbol\beta}$ (see [20]) is given by

$$\left\{\frac{\partial}{\partial r}, r^{-1}\frac{\partial}{\partial\theta}, D', \frac{\partial^2}{\partial r^2}+r^{-1}\frac{\partial}{\partial r}+(\beta r)^{-2}\frac{\partial^2}{\partial\theta^2}, \frac{\partial}{\partial r}D', r^{-1}\frac{\partial}{\partial\theta}D'\right\},$$

while those defined in [23] for the Hölder space $\mathcal{D}^{0,\alpha}_{\omega}$ gives the collection

$$\left\{\frac{\partial}{\partial r}, r^{-1}\frac{\partial}{\partial\theta}, D', \frac{\partial^2}{\partial r^2}+r^{-1}\frac{\partial}{\partial r}+(\beta r)^{-2}\frac{\partial^2}{\partial\theta^2}, \frac{\partial}{\partial r}D', r^{-1}\frac{\partial}{\partial\theta}D', \frac{\partial^2}{\partial r\partial\theta}\right\}. \tag{2-1}$$

Here the operators D' are given in Definition 1.1. There seems to a typo in the original definition of (2-1) in [23, p. 104, (16)], where the factor r^{-1} is missing in the operator $r^{-1}\partial/\partial\theta D'$ (see [32, p. 57]). It was pointed out by the referee that this typo does not affect Proposition 3.3 in [23] since the correct operator was used in the proof. The space $\mathcal{D}^{0,\alpha}_{\omega}$ is introduced in [23] as an alternative definition for the Hölder space of [14] as a consequence of the Schauder estimates in [23, Proposition 3.3]. The Schauder estimates in [23] are stronger than those established in [14] by Donaldson and later in [20] by the authors because of the additional operator $\partial^2/\partial r\partial\theta$ in (2-1). This also implies that the two Hölder spaces from [14] and [20] must coincide. For interested readers, we refer to the survey paper [32] for more details on the characterization of the Hölder space of [14] in terms of the operators in (2-1).

For a given set $\Omega\subset B_{\boldsymbol\beta}(0,1)$, we define the following weighted (semi)norms.

Definition 2.4. Suppose $\sigma\in\mathbb{R}$ is a given real number and u is a $C^{2,\alpha}_{\boldsymbol\beta}$-function in Ω. We will write $d_x = d_{\boldsymbol\beta}(x,\partial\Omega)$ for any $x\in\Omega$. We define the weighted (semi)norms

$$[u]^{(\sigma)}_{C^{0,\alpha}_{\boldsymbol\beta}(\Omega)} = \sup_{x\neq y\in\Omega}\min(d_x,d_y)^{\sigma+\alpha}\frac{|u(x)-u(y)|}{d_{\boldsymbol\beta}(x,y)^{\alpha}},$$

$$\|u\|^{(\sigma)}_{C^0(\Omega)} = \sup_{x\in\Omega} d_x^{\sigma}|u(x)|,\quad [u]^{(\sigma)}_{C^1_{\boldsymbol\beta}(\Omega)} = \sup_{x\in\Omega\setminus\mathcal{S}} d_x^{\sigma+1}\Bigl(\sum_j |N_j u|(x)+|D'u|(x)\Bigr),$$

$$[u]^{(\sigma)}_{C^2_{\boldsymbol\beta}(\Omega)} = \sup_{x\in\Omega\setminus\mathcal{S}} d_x^{\sigma+2}|Tu(x)|,\quad [u]^{(\sigma)}_{C^{2,\alpha}_{\boldsymbol\beta}(\Omega)} = \sup_{x\neq y\in\Omega\setminus\mathcal{S}}\min(d_x,d_y)^{\sigma+2+\alpha}\frac{|Tu(x)-Tu(y)|}{d_{\boldsymbol\beta}(x,y)^{\alpha}},$$

$$\|u\|^{(\sigma)}_{C^{2,\alpha}_{\boldsymbol\beta}(\Omega)} = \|u\|^{(\sigma)}_{C^0(\Omega)}+[u]^{(\sigma)}_{C^1_{\boldsymbol\beta}(\Omega)}+[u]^{(\sigma)}_{C^2_{\boldsymbol\beta}(\Omega)}+[u]^{(\sigma)}_{C^{2,\alpha}_{\boldsymbol\beta}(\Omega)},$$

where T is the collection of operators of second-order

$$\left\{|z_j|^{2(1-\beta_j)}\frac{\partial^2}{\partial z_j\partial\bar z_j},\ N_jN_k\ (j\neq k),\ N_jD',\ (D')^2\right\}. \tag{2-2}$$

When $\sigma=0$, we write the norms above as $[\,\cdot\,]^*$ or $\|\cdot\|^*$ for simplicity of notation.

2A2. *Parabolic case.* We define $\mathcal{Q}_{\boldsymbol\beta}=\mathcal{Q}_{\boldsymbol\beta}(0,1)=B_{\boldsymbol\beta}(0,1)\times(0,1]$ to be a parabolic cylinder and

$$\partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol\beta} = (\overline{B_{\boldsymbol\beta}(0,1)}\times\{0\})\cup(\partial B_{\boldsymbol\beta}(0,1)\times(0,1])$$

to be the parabolic boundary of the cylinder $\mathcal{Q}_{\boldsymbol\beta}$. We write $\mathcal{S}_{\mathcal{P}} = \mathcal{S} \times [0, 1]$ for the singular set and $\mathcal{Q}^{\#}_{\boldsymbol\beta} = \mathcal{Q}_{\boldsymbol\beta} \backslash \mathcal{S}_{\mathcal{P}}$ for the complement of $\mathcal{S}_{\mathcal{P}}$. For any two space-time points $Q_i = (p_i, t_i)$, we define their parabolic distance $d_{\mathcal{P},\boldsymbol\beta}(Q_1, Q_2)$ as

$$d_{\mathcal{P},\boldsymbol\beta}(Q_1, Q_2) = \max\{\sqrt{|t_1 - t_2|},\ d_{\boldsymbol\beta}(p_1, p_2)\}.$$

Definition 2.5. We define the $g_{\boldsymbol\beta}$-Hölder norm of functions $u \in \mathscr{C}^0(\mathcal{Q}_{\boldsymbol\beta})$ for some $\alpha \in (0, 1)$ as

$$\|u\|_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(\mathcal{Q}_{\boldsymbol\beta})} := \|u\|_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol\beta})} + [u]_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(\mathcal{Q}_{\boldsymbol\beta})},$$

where the seminorm is

$$[u]_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(\mathcal{Q}_{\boldsymbol\beta})} := \sup_{Q_1 \neq Q_2 \in \mathcal{Q}_{\boldsymbol\beta}} \frac{|u(Q_1) - u(Q_2)|}{d_{\mathcal{P},\boldsymbol\beta}(Q_1, Q_2)^\alpha}.$$

We denote by $\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(\mathcal{Q}_{\boldsymbol\beta})$ the subspace of all continuous functions u such that $\|u\|_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(\mathcal{Q}_{\boldsymbol\beta})} < \infty$.

Definition 2.6. The $\mathscr{C}^{2+\alpha,(\alpha+2)/2}_{\boldsymbol\beta}$-norm of a function u on $\mathcal{Q}_{\boldsymbol\beta}$ is defined as

$$\|u\|_{\mathscr{C}^{2+\alpha,(\alpha+2)/2}_{\boldsymbol\beta}(\mathcal{Q}_{\boldsymbol\beta})} := \|u\|_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol\beta})} + \|\nabla_{g_{\boldsymbol\beta}} u\|_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol\beta}, g_{\boldsymbol\beta})} + \|\mathcal{T} u\|_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(\mathcal{Q}_{\boldsymbol\beta})},$$

where $\mathcal{T}$ is the collection of all the second-order operators in (2-2) with the first-order operator $\partial/\partial t$.

For a given set $\Omega \subset \mathcal{Q}_{\boldsymbol\beta}$ we define the following weighted (semi)norms.

Definition 2.7. Suppose $\sigma \in \mathbb{R}$ is a real number and u is a $\mathscr{C}^{2+\alpha,(\alpha+2)/2}_{\boldsymbol\beta}$-function in Ω. We will write $d_{\mathcal{P},Q} = d_{\mathcal{P},\boldsymbol\beta}(Q, \partial_{\mathcal{P}}\Omega)$ for any $Q \in \Omega$. We define the weighted (semi)norms

$$[u]^{(\sigma)}_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(\Omega)} = \sup_{Q_1 \neq Q_2 \in \Omega} \min(d_{\mathcal{P},Q_1}, d_{\mathcal{P},Q_2})^{\sigma+\alpha} \frac{|u(Q_1) - u(Q_2)|}{d_{\mathcal{P},\boldsymbol\beta}(Q_1, Q_2)^\alpha}, \quad \|u\|^{(\sigma)}_{\mathscr{C}^0(\Omega)} = \sup_{Q \in \Omega} d^{\sigma}_{\mathcal{P},Q} |u(Q)|,$$

$$[u]^{(\sigma)}_{\mathscr{C}^1_{\boldsymbol\beta}(\Omega)} = \sup_{Q \in \Omega \backslash \mathcal{S}_{\mathcal{P}}} d^{\sigma+1}_{\mathcal{P},Q} \Big(\sum_j |N_j u|(Q) + |D' u|(Q) \Big), \quad [u]^{(\sigma)}_{\mathscr{C}^{2,1}_{\boldsymbol\beta}(\Omega)} = \sup_{Q \in \Omega \backslash \mathcal{S}_{\mathcal{P}}} d^{\sigma+2}_{\mathcal{P},Q} |\mathcal{T} u(Q)|,$$

$$[u]^{(\sigma)}_{\mathscr{C}^{2+\alpha,(\alpha+2)/2}_{\boldsymbol\beta}(\Omega)} = \sup_{Q_1 \neq Q_2 \in \Omega \backslash \mathcal{S}_{\mathcal{P}}} \min(d_{\mathcal{P},Q_1}, d_{\mathcal{P},Q_2})^{\sigma+2+\alpha} \frac{|\mathcal{T} u(Q_1) - \mathcal{T} u(Q_2)|}{d_{\mathcal{P},\boldsymbol\beta}(Q_1, Q_2)^\alpha},$$

$$\|u\|^{(\sigma)}_{\mathscr{C}^{2+\alpha,(\alpha+2)/2}_{\boldsymbol\beta}(\Omega)} = \|u\|^{(\sigma)}_{\mathscr{C}^0(\Omega)} + [u]^{(\sigma)}_{\mathscr{C}^1_{\boldsymbol\beta}(\Omega)} + [u]^{(\sigma)}_{\mathscr{C}^{2,1}_{\boldsymbol\beta}(\Omega)} + [u]^{(\sigma)}_{\mathscr{C}^{2+\alpha,(\alpha+2)/2}_{\boldsymbol\beta}(\Omega)}.$$

When $\sigma = 0$, we write the norms above as $[\,\cdot\,]^*$ or $\|\cdot\|^*$ for simplicity of notation.

2A3. *Compact Kähler manifolds.* Let (X, D) be a compact Kähler manifold with a divisor $D = \sum_j D_j$ with simple normal crossings, i.e., on an open coordinate chart (U, z_j) of any $x \in D$, $D \cap U$ is given by $\{z_1 \cdots z_p = 0\}$ and $D_j \cap U = \{z_j = 0\}$ for any component D_j of D. We fix a finite cover $\{U_a, z_{a,j}\}$ of D.

Definition 2.8. A (singular) Kähler metric ω is called a conical metric with cone angle $2\pi\boldsymbol\beta$ along D if ω is equivalent to $\omega_{\boldsymbol\beta}$ locally on any coordinate chart U_a under the coordinates $\{z_{a,j}\}$, where $\omega_{\boldsymbol\beta}$ is the standard cone metric (1-1) with cone angle $2\pi\beta_j$ along $\{z_{a,j} = 0\}$, and ω is a smooth Kähler metric in the usual sense on $X \backslash \bigcup_a U_a$.

A conical metric ω is in $C^{0,\alpha}_{\boldsymbol\beta}(X, D)$ if ω is in $C^{0,\alpha}_{\boldsymbol\beta}(U_a)$ for each a and ω is smooth in the usual sense on $X \backslash \bigcup_a U_a$. Similarly we can define the $\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}$-conical Kähler metrics on $X \times [0, 1]$.

Definition 2.9. A continuous function $u \in C^0(X)$ is said to be in $C^{0,\alpha}_{\boldsymbol\beta}(X, D)$ if u is in $C^{0,\alpha}_{\boldsymbol\beta}(U_a)$ locally on each U_a and u is $C^{0,\alpha}$-continuous in the usual sense on $X\setminus\bigcup_a U_a$. We define the $C^{0,\alpha}_{\boldsymbol\beta}(X, D)$-norm of u

$$\|u\|_{C^{0,\alpha}_{\boldsymbol\beta}(X,D)} := \|u\|_{C^{0,\alpha}(X\setminus\bigcup_a U_a,\omega)} + \sum_a \|u\|_{C^{0,\alpha}_{\boldsymbol\beta}(U_a)}.$$

The $C^{0,\alpha}_{\boldsymbol\beta}(X, D)$-norm depends on the choice of finite covers, and another cover yields a different but equivalent norm. The space $C^{0,\alpha}_{\boldsymbol\beta}(X, D)$ is clearly independent of the choice of finite covers.

The other spaces and norms like $C^{2,\alpha}_{\boldsymbol\beta}(X, D)$, $\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(X\times[0, 1], D)$, etc., can be defined similarly.

2B. *A useful lemma.* We will frequently use the following elementary estimates from [20]. We write $B_{\mathbb{C}}(0, r)$ for the Euclidean ball in $\mathbb{C}$ with center 0 and radius $r > 0$.

Lemma 2.10 (Lemma 2.2 in [20]). *Given $r \in (0, 1]$, suppose $v \in C^0(B_{\mathbb{C}}(0, r)) \cap C^2(B_{\mathbb{C}}(0, r)\setminus\{0\})$ satisfies*

$$|z|^{2(1-\beta_1)}\frac{\partial^2 v}{\partial z\partial\bar z} = F \quad \text{in } B_{\mathbb{C}}(0, r)\setminus\{0\}$$

for some $F \in L^\infty(B_{\mathbb{C}}(0, r))$. Then we have the following pointwise estimate for any $z \in B_{\mathbb{C}}\big(0, \frac{9}{10}r\big)\setminus\{0\}$:

$$\left|\frac{\partial v}{\partial z}(z)\right| \le C\frac{\|v\|_{L^\infty}}{r} + C\|F\|_{L^\infty}\cdot\begin{cases} r^{2\beta_1-1} & \text{if } \beta_1\in\big(\tfrac12, 1\big),\\ |z|^{2\beta_1-1} & \text{if } \beta_1\in\big(0, \tfrac12\big),\\ \left|\log\Big(\dfrac{|z|}{2r}\Big)\right| & \text{if } \beta_1 = \tfrac12,\end{cases} \tag{2-3}$$

where the L^∞-norms are taken in $B_{\mathbb{C}}(0, r)$ and $C > 0$ is a uniform constant depending only on the angle β_1.

Finally we remark that the idea of the proof of the estimates in Theorems 1.2 and 1.7 is the same for general $2 \le p \le n$. To explain the argument more clearly, we prove the theorems assuming $p = 2$, i.e., the cone metric of $\omega_{\boldsymbol\beta}$ is singular along the two components $\mathcal{S}_1$ and $\mathcal{S}_2$.

3. Elliptic estimates

In this section, we will prove Theorems 1.2 and 1.6, the Schauder estimates for the Laplace equation (1-2). To begin with, we first observe the simple C^0-estimate based on the maximum principle.

Suppose $u \in C^2(B_{\boldsymbol\beta}(0, 1)\setminus\mathcal{S}) \cap C^0(\overline{B_{\boldsymbol\beta}(0, 1)})$ satisfies the equation

$$\begin{cases}\Delta_{\boldsymbol\beta} u = 0 & \text{in } B_{\boldsymbol\beta}(0, 1)\setminus\mathcal{S},\\ u = \varphi & \text{on } \partial B_{\boldsymbol\beta}(0, 1)\end{cases} \tag{3-1}$$

for some $\varphi \in C^0(\partial B_{\boldsymbol\beta}(0, 1))$, then we have the following lemma.

Lemma 3.1. *We have the maximum principle*

$$\inf_{\partial B_{\boldsymbol\beta}(0,1)}\varphi \le \inf_{B_{\boldsymbol\beta}(0,1)} u \le \sup_{B_{\boldsymbol\beta}(0,1)} u \le \sup_{\partial B_{\boldsymbol\beta}(0,1)}\varphi. \tag{3-2}$$

Proof. Consider the functions $\tilde u_\epsilon = u \pm \epsilon(\log|z_1|^2 + \log|z_2|^2)$ for any $\epsilon > 0$. By the proof of Lemma 2.1 in [20], (3-2) is established. □

The next step is to show (3-1) is solvable for suitable boundary values.

3A. *Conical harmonic functions.*

3A1. *Smooth approximating metrics.* Let $\epsilon \in (0, 1)$ be a given small positive number and define a smooth approximating Kähler metric g_ϵ on $B_{\boldsymbol\beta}(0, 1)$ as

$$g_\epsilon = \beta_1^2 \frac{\sqrt{-1}\, dz_1 \wedge d\bar z_1}{(|z_1|^2+\epsilon)^{1-\beta_1}} + \beta_2^2 \frac{\sqrt{-1}\, dz_2 \wedge d\bar z_2}{(|z_2|^2+\epsilon)^{1-\beta_2}} + \sum_{j=3}^{n} \sqrt{-1}\, dz_j \wedge d\bar z_j. \tag{3-3}$$

The g_ϵ are product metrics on $\mathbb{C} \times \mathbb{C} \times \mathbb{C}^{n-2}$. It is clear that their Ricci curvatures satisfy

$$\mathrm{Ric}(g_\epsilon) = \sqrt{-1}\partial\bar\partial \log((|z_1|^2+\epsilon)^{1-\beta_1}(|z_2|^2+\epsilon)^{1-\beta_2}) \geq 0.$$

Let $u_\epsilon \in C^2(B_{\boldsymbol\beta}(0, 1))$ be the solution to the equation

$$\Delta_{g_\epsilon} u_\epsilon = 0 \quad \text{in } B_{\boldsymbol\beta}(0, 1) \qquad \text{and} \qquad u_\epsilon = \varphi \quad \text{on } \partial B_{\boldsymbol\beta}(0, 1) \tag{3-4}$$

with a given $\varphi \in C^0(\partial B_{\boldsymbol\beta}(0, 1))$. Note that the metric balls $B_{\boldsymbol\beta}(0, 1)$ and $B_{g_\epsilon}(0, 1)$ are uniformly close when ϵ is sufficiently small, so for the following estimates we will work on $B_{\boldsymbol\beta}(0, 1)$.

Let u_ϵ be the harmonic function for $\Delta_\epsilon = \Delta_{g_\epsilon}$ as in (3-4), which we may assume without loss of generality is positive by replacing u_ϵ by $u_\epsilon - \inf u_\epsilon$ if necessary. We will study the Cheng–Yau-type gradient estimate of u_ϵ and the estimate of $\Delta_{1,\epsilon} u_\epsilon := (|z_1|^2+\epsilon)^{1-\beta_1}(\partial^2 u_\epsilon/\partial z_1 \partial\bar z_1)$. Let us recall Cheng–Yau's gradient estimate first.

In Sections 3A2–3A5, for convenience of notation, we will omit the subscript ϵ in g_ϵ and u_ϵ in the proofs of the lemmas.

3A2. *Cheng–Yau gradient estimate revisited.* We assume $u_\epsilon > 0$, as otherwise we could consider the function $u_\epsilon + \delta$ for some $\delta > 0$ and then let $\delta \to 0$. We fix a metric ball $B_{g_\epsilon}(p, R) \subset B_{\boldsymbol\beta}(0, 1)$ centered at some point $p \in B_{\boldsymbol\beta}(0, 1)$. Since $\mathrm{Ric}(g_\epsilon) \geq 0$, the Cheng–Yau gradient estimate holds for Δ_{g_ϵ}-harmonic functions.

Lemma 3.2 [10]. *Let $u_\epsilon \in C^2(B(p, R))$ be a positive Δ_{g_ϵ}-harmonic function. There exists a uniform constant $C = C(n) > 0$ such that* (*the metric balls are taken under the metric g_ϵ*)

$$\sup_{x \in B(p, 3R/4)} |\nabla u_\epsilon|_{g_\epsilon}(x) \leq C(n) \frac{\mathrm{osc}_{B(p,R)}\, u_\epsilon}{R}. \tag{3-5}$$

As we mentioned above, we will omit the ϵ in the subscript of u_ϵ and g_ϵ. The proof of the lemma is standard [10]. For completeness and to motivate the proofs of Lemmas 3.3 and 3.4, we sketch a proof. Defining $f = \log u$, it can be calculated that

$$\Delta f = \frac{\Delta u}{u} - \frac{|\nabla u|^2}{u^2} = -|\nabla f|^2. \tag{3-6}$$

Then by Bochner's formula we have

$$\begin{aligned} \Delta |\nabla f|^2 &= |\nabla\nabla f|^2 + |\nabla\bar\nabla f|^2 + 2\,\mathrm{Re}\langle \nabla f, \bar\nabla \Delta f\rangle + \mathrm{Ric}(\nabla f, \bar\nabla f) \\ &\geq |\nabla\bar\nabla f|^2 - 2\,\mathrm{Re}\langle \nabla f, \bar\nabla |\nabla f|^2\rangle. \end{aligned} \tag{3-7}$$

Let $\phi : [0, 1] \to [0, 1]$ be a standard cut-off function such that $\phi|_{[0,3/4]} = 1$, $\phi_{[5/6,1]} = 0$ and $0 < \phi < 1$ otherwise. Let $r(x) = d_{g_\epsilon}(p, x)$ be the distance function to p under the metric $g = g_\epsilon$. By abusing notation, we also write $\phi(x) = \phi(r(x)/R)$. It can be calculated by Laplacian comparison and the Bochner formula (3-7) that, at the (positive) maximum point $p_{\max}$ of $H := \phi^2|\nabla f|^2$,

$$\frac{2}{n}H^2 - \frac{4|\phi'|}{R}H^{3/2} - \frac{8(\phi')^2}{R^2}H + \frac{2H}{R^2}((2n-1)\phi\phi' + \phi\phi'' + (\phi')^2) \le 0.$$

Therefore, for any $x \in B\big(p, \frac{3}{4}R\big)$,

$$\frac{|\nabla u|^2}{u^2}(x) = |\nabla f(x)|^2 = H(x) \le H(p_{\max}) \le \frac{C(n)}{R^2}. \tag{3-8}$$

3A3. *Laplacian estimate in singular directions.* We will prove the estimates of

$$\Delta_{j,\epsilon} u_\epsilon := (|z_j|^2 + \epsilon)^{1-\beta_j} \frac{\partial u_\epsilon}{\partial z_j \partial \bar{z}_j}$$

for a Δ_{g_ϵ}-harmonic function u_ϵ.

Lemma 3.3. *Under the same assumptions as in Lemma 3.2, along the "bad" directions z_1 and z_2, we have that $\Delta_{1,\epsilon} u_\epsilon$ and $\Delta_{2,\epsilon} u_\epsilon$ satisfy the estimates*

$$\sup_{x \in B(p,R/2)} (|\Delta_{1,\epsilon} u_\epsilon|(x) + |\Delta_{2,\epsilon} u_\epsilon|(x)) \le C(n) \frac{\operatorname{osc}_{B(p,R)} u_\epsilon}{R^2}. \tag{3-9}$$

As in the proof of Cheng–Yau gradient estimates, we will work on the function $f = f_\epsilon = \log u$, and we only need to prove the estimate for $\Delta_{1,\epsilon} u_\epsilon$. We write $\Delta_{1,\epsilon} f := (|z_1|^2 + \epsilon)^{1-\beta_1}(\partial^2 f/\partial z_1 \partial \bar{z}_1)$.

As above, we will omit the subscript ϵ in $\Delta_{1,\epsilon} f$. We first observe that

$$\Delta_1 \Delta_{g_\epsilon} f = \Delta_{g_\epsilon} \Delta_1 f. \tag{3-10}$$

Equation (3-10) can be checked from the definitions using the property that g_ϵ is a product metric. Indeed

$$\begin{aligned}
\Delta_1 \Delta_{g_\epsilon} f &= (|z_1|^2 + \epsilon)^{1-\beta_1} \frac{\partial^2}{\partial z_1 \partial \bar{z}_1} \left((|z_1|^2 + \epsilon)^{1-\beta_1} \frac{\partial^2 f}{\partial z_1 \partial \bar{z}_1} + (|z_2|^2 + \epsilon)^{1-\beta_2} \frac{\partial^2 f}{\partial z_2 \partial \bar{z}_2} + \sum_j \frac{\partial^2 f}{\partial z_j \partial \bar{z}_j} \right) \\
&= (|z_1|^2 + \epsilon)^{1-\beta_1} \frac{\partial^2}{\partial z_1 \partial \bar{z}_1} \Delta_1 f + (|z_2|^2 + \epsilon)^{1-\beta_2} \frac{\partial^2}{\partial z_2 \partial \bar{z}_2} \Delta_1 f + \sum_{j=3}^n \frac{\partial^2}{\partial z_j \partial \bar{z}_j} \Delta_1 f = \Delta_{g_\epsilon} \Delta_1 f.
\end{aligned}$$

On the other hand, note that $\Delta_{g_\epsilon} f = \Delta_g f = -|\nabla f|^2$ by (3-6). Choosing a normal frame $\{e_1, \ldots, e_n\}$ at some point x such that $dg(x) = 0$ and $\Delta_1 f = f_{1\bar{1}}$, we calculate

$$\begin{aligned}
\Delta_1 |\nabla f|^2 = (f_j f_{\bar{j}})_{1\bar{1}} &= f_{j1} f_{\bar{j}\bar{1}} + f_{j\bar{1}} f_{\bar{j}1} + f_{j1\bar{1}} f_{\bar{j}} + f_j f_{\bar{j}1\bar{1}} \\
&= f_{j1} f_{\bar{j}\bar{1}} + f_{j\bar{1}} f_{\bar{j}1} + f_j f_{\bar{1}1\bar{j}} + f_{\bar{j}}(f_{1\bar{1}j} + f_m R_{1\bar{m}j\bar{1}}) \\
&= |\nabla_1 \nabla f|^2 + |\nabla_1 \overline{\nabla} f|^2 + 2\operatorname{Re}\langle \nabla f, \overline{\nabla} \Delta_1 f \rangle + f_m f_{\bar{j}} R_{1\bar{1}j\bar{m}} \\
&\ge (\Delta_1 f)^2 + 2\operatorname{Re}\langle \nabla f, \overline{\nabla} \Delta_1 f \rangle.
\end{aligned} \tag{3-11}$$

Then

$$\Delta(-\Delta_1 f) = -\Delta_1\Delta f = \Delta_1|\nabla f|^2 \ge (\Delta_1 f)^2 + 2\operatorname{Re}\langle\nabla f, \bar\nabla\Delta_1 f\rangle.$$

Let $\varphi : [0, 1] \to [0, 1]$ be a standard cut-off function such that $\varphi|_{[0,1/2]} = 1$ and $\varphi|_{[2/3,1]} = 0$. We also define $\varphi(x) = \varphi(r(x)/R)$. Then consider the function $G := \varphi^2 \cdot (-\Delta_1 f)$. We calculate

$$\begin{aligned}\Delta G &= \Delta(\varphi^2(-\Delta_1 f))\\ &= \varphi^2\Delta(-\Delta_1 f) + 2\operatorname{Re}\langle\nabla\varphi^2, \bar\nabla(-\Delta_1 f)\rangle + (-\Delta_1 f)\Delta\varphi^2\\ &\ge \varphi^2((\Delta_1 f)^2 + 2\operatorname{Re}\langle\nabla f, \bar\nabla\Delta_1 f\rangle) + 2\operatorname{Re}\langle\nabla\varphi^2, \bar\nabla(-\Delta_1 f)\rangle + (-\Delta_1 f)\Delta\varphi^2. \end{aligned}\tag{3-12}$$

We want to estimate the upper bound of G. If the maximum value of $G = \varphi^2(-\Delta_1 f)$ is negative, we are done. So we assume the maximum of G on $B(p, R)$ is *positive*, which is achieved at some point $p_{\max} \in B\big(p, \frac{2}{3}R\big)$. Hence, at $p_{\max}$ we have $(-\Delta_1 f) > 0$. By Laplacian comparison, $\Delta r \le (2n-1)/r$, and we get, at $p_{\max}$,

$$\Delta\varphi^2 \ge \frac{2}{R^2}((2n-1)\varphi\varphi' + \varphi\varphi'' + (\varphi')^2).\tag{3-13}$$

Thus, at $p_{\max}$, the last term on the right-hand side of (3-12) is greater than or equal to

$$(-\Delta_1 f)\frac{2}{R^2}((2n-1)\varphi\varphi' + \varphi\varphi'' + (\varphi')^2).$$

Substituting this into (3-12), it follows that, at $p_{\max}$, we have $\Delta G \le 0$ and $\nabla\Delta_1 f = -2\varphi^{-1}\Delta_1 f\nabla\varphi$, and hence

$$\begin{aligned}0 &\ge \Delta G\\ &\ge \varphi^2(\Delta_1 f)^2 + 2\varphi^2\operatorname{Re}\langle\nabla f, \bar\nabla\Delta_1 f\rangle + 4\varphi\operatorname{Re}\langle\nabla\varphi, \bar\nabla(-\Delta_1 f)\rangle + (-\Delta_1 f)\frac{2}{R^2}((2n-1)\varphi\varphi' + \varphi\varphi'' + (\varphi')^2)\\ &\ge \varphi^2(\Delta_1 f)^2 - 4\varphi|\Delta_1 f||\nabla f||\nabla\varphi| + 8\Delta_1 f|\nabla\varphi|^2 + (-\Delta_1 f)\frac{2}{R^2}((2n-1)\varphi\varphi' + \varphi\varphi'' + (\varphi')^2)\\ &= \frac{G^2}{\varphi^2} - 4\varphi^{-1}G|\nabla f||\nabla\varphi| - 8G\frac{|\nabla\varphi|^2}{\varphi^2} + \frac{2G}{R^2\varphi^2}((2n-1)\varphi\varphi' + \varphi\varphi'' + (\varphi')^2)\\ &\ge \frac{G^2}{\varphi^2} - 4\frac{|\varphi'||\nabla f|}{R\varphi}G - 8\frac{|\varphi'|^2}{R^2\varphi^2}G + \frac{2G}{R^2\varphi^2}((2n-1)\varphi\varphi' + \varphi\varphi'' + (\varphi')^2). \end{aligned}\tag{3-14}$$

Therefore, at $p_{\max} \in B\big(p, \frac{2}{3}R\big)$,

$$G^2 - 4\frac{\varphi|\varphi'\nabla f|}{R}G - 8\frac{|\varphi'|^2}{R^2}G + \frac{2G}{R^2}((2n-1)\varphi\varphi' + \varphi\varphi'' + (\varphi')^2) \le 0,$$

and combining (3-8) and the fact that φ, φ', φ'' are all uniformly bounded, we can get, at $p_{\max}$,

$$G^2 \le C(n)R^{-2}G \quad\Longrightarrow\quad G(p_{\max}) \le \frac{C(n)}{R^2}.$$

Then, for any $x \in B\big(p, \frac{1}{2}R\big)$, where $\varphi = 1$, we have

$$-\Delta_1 f(x) = G(x) \le G(p_{\max}) \le \frac{C(n)}{R^2}.$$

Moreover, recall that $f=\log u$ and $-\Delta_1 f=-\Delta_1 u/u+|\nabla_1 f|^2$, therefore it follows that

$$\sup_{x\in B(p,R/2)}\left(-\frac{\Delta_1 u}{u}(x)\right)\le \frac{C(n)}{R^2}. \tag{3-15}$$

This in particular implies that

$$\sup_{x\in B(p,R/2)}(-\Delta_1 u(x))\le C(n)\frac{\operatorname{osc}_{B(p,R/2)} u}{R^2}\le C(n)\frac{\operatorname{osc}_{B(p,R)} u}{R^2}. \tag{3-16}$$

On the other hand, consider the function $\hat u=\max_{B(p,R)} u-u$, which is still a positive g_ϵ-harmonic function with $\Delta_g \hat u=\Delta_{g_\epsilon}\hat u=0$. Applying (3-15) to the function $\hat u$, we get

$$\sup_{x\in B(p,R/2)}\left(\frac{\Delta_1 u(x)}{\max_{B(p,R)} u-u(x)}\right)=\sup_{x\in B(p,R/2)}\left(-\frac{\Delta_1\hat u}{\hat u}(x)\right)\le\frac{C(n)}{R^2}, \tag{3-17}$$

which yields

$$\sup_{x\in B(p,R/2)}\Delta_1 u(x)\le C(n)\frac{\operatorname{osc}_{B(p,R)} u}{R^2}. \tag{3-18}$$

Combining (3-18) and (3-16), we get

$$\sup_{x\in B(p,R/2)}|\Delta_1 u|(x)\le C(n)\frac{\operatorname{osc}_{B(p,R)} u}{R^2}. \tag{3-19}$$

3A4. *Mixed derivatives estimates.* In this subsection we will estimate the mixed derivatives

$$|\nabla_1\nabla_2 f|^2=\frac{\partial^2 f}{\partial z_1\partial z_2}\overline{\frac{\partial^2 f}{\partial z_1\partial z_2}}g^{1\bar 1}g^{2\bar 2}\quad\text{and}\quad |\nabla_1\nabla_{\bar 2} f|^2=\frac{\partial^2 f}{\partial z_1\partial \bar z_2}\overline{\frac{\partial^2 f}{\partial z_1\partial \bar z_2}}g^{1\bar 1}g^{2\bar 2},$$

where as before $f=\log u$ and u is a positive harmonic function of Δ_{g_ϵ}. Here for simplicity, we omit the subscript ϵ in u_ϵ, f_ϵ and g_ϵ. Observing that since $g_\epsilon=g$ is a product metric with the nonzero components $g_{k\bar k}$ depending only on z_k, it follows that the curvature tensor

$$R_{i\bar j k\bar l}=-\frac{\partial^2 g_{i\bar j}}{\partial z_k\partial \bar z_l}+g^{p\bar q}\frac{\partial g_{i\bar q}}{\partial z_k}\frac{\partial g_{p\bar j}}{\partial \bar z_l}$$

vanishes unless $i=j=k=l\in\{1,2\}$ and also $R_{i\bar i i\bar i}\ge 0$ for all $i=1,\dots,n$.

We fix some notation: we will write $f_{12}=\nabla_1\nabla_2 f$ (in fact this is just the ordinary derivative of f with respect to g, since g is a product metric), $|f_{12}|_g^2=|\nabla_1\nabla_2 f|_g^2$, etc.

Let us first recall that (3-11) implies

$$\begin{aligned}
&\Delta(-\Delta_1 f-\Delta_2 f)\\
&=\sum_{k=1}^n(g^{1\bar 1}g^{k\bar k}f_{1k}f_{\bar 1\bar k}+g^{1\bar 1}g^{k\bar k}f_{1\bar k}f_{\bar 1k}+g^{2\bar 2}g^{k\bar k}f_{2k}f_{\bar 2\bar k}+g^{2\bar 2}g^{k\bar k}f_{2\bar k}f_{k\bar 2})\\
&\qquad\qquad -2\operatorname{Re}\langle\nabla f,\overline{\nabla}(-\Delta_1 f-\Delta_2 f)\rangle+f_1f_{\bar 1}g^{1\bar 1}g^{1\bar 1}R_{1\bar 1 1\bar 1}+f_2f_{\bar 2}g^{2\bar 2}g^{2\bar 2}R_{2\bar 2 2\bar 2}\\
&\ge\sum_{k=1}^n(|\nabla_1\nabla_k f|^2+|\nabla_1\nabla_{\bar k} f|^2+|\nabla_2\nabla_k f|^2+|\nabla_2\nabla_{\bar k} f|^2)-2\operatorname{Re}\langle\nabla f,\overline{\nabla}(-\Delta_1 f-\Delta_2 f)\rangle.
\end{aligned} \tag{3-20}$$

Next we calculate $\Delta|\nabla_1\nabla_2 f|^2$. For convenience of notation we will write $f^{12} = f_{\bar{1}\bar{2}} g^{1\bar{1}} g^{2\bar{2}}$, and hence $|\nabla_1\nabla_2 f|^2 = f_{12} f^{12}$. We calculate

$$\begin{aligned}\Delta|\nabla_1\nabla_2 f|^2 &= g^{k\bar{l}}(f_{12}f^{12})_{k\bar{l}} = g^{k\bar{k}}(f_{12}f^{12})_{k\bar{k}} \qquad \text{(since } g \text{ is a product metric)}\\ &= g^{k\bar{k}}(f_{12k\bar{k}}f^{12} + f_{12k}f^{12}{}_{,\bar{k}} + f_{12\bar{k}}f^{12}{}_{,k} + f_{12}f^{12}{}_{,k\bar{k}}).\end{aligned} \tag{3-21}$$

The first term on the right-hand side of (3-21) is (by Ricci identities and switching the indices)

$$\begin{aligned}&g^{k\bar{k}}f^{12}(f_{k1\bar{k}2} + g^{m\bar{m}}f_{m1}R_{k\bar{m}2\bar{k}} + g^{m\bar{m}}f_{km}R_{1\bar{m}2\bar{k}})\\ &\quad= g^{k\bar{k}}f^{12}(f_{k\bar{k}12} + g^{m\bar{m}}f_{m2}R_{k\bar{m}1\bar{k}} + g^{m\bar{m}}f_{m}R_{k\bar{m}1\bar{k}} + g^{m\bar{m}}f_{m1}R_{k\bar{m}2\bar{k}} + g^{m\bar{m}}f_{km}R_{1\bar{m}2\bar{k}})\\ &\quad= g^{k\bar{k}}f^{12}(f_{k\bar{k}12} + g^{m\bar{m}}f_{m2}R_{k\bar{m}1\bar{k}} + g^{m\bar{m}}f_{m1}R_{k\bar{m}2\bar{k}})\\ &\quad= g^{k\bar{k}}f^{12}f_{k\bar{k}12} + g^{1\bar{1}}g^{1\bar{1}}f^{12}f_{21}R_{1\bar{1}1\bar{1}} + g^{2\bar{2}}g^{2\bar{2}}f^{12}f_{12}R_{2\bar{2}2\bar{2}},\end{aligned} \tag{3-22}$$

and the last term on the right-hand side of (3-21) is the conjugate of the first term; hence we get

$$\begin{aligned}\Delta|\nabla_1\nabla_2 f|^2 = 2\,\mathrm{Re}(f^{12}(\Delta f)_{12}) + 2f^{12}f_{12}(g^{1\bar{1}}g^{1\bar{1}}R_{1\bar{1}1\bar{1}} + g^{2\bar{2}}g^{2\bar{2}}R_{2\bar{2}2\bar{2}})&\\ + g^{k\bar{k}}f_{12k}f^{12}{}_{,\bar{k}} + g^{k\bar{k}}f_{12\bar{k}}f^{12}{}_{,k}&.\end{aligned} \tag{3-23}$$

Recall from (3-6) that $\Delta f = -|\nabla f|^2$; hence the first term on the right-hand side of (3-23) is

$$\begin{aligned}2\,\mathrm{Re}(f^{12}(\Delta f)_{12}) &= 2\,\mathrm{Re}(f^{12}(-|\nabla f|^2)_{12})\\ &= -2\,\mathrm{Re}(f^{12}g^{k\bar{k}}(f_{k12}f_{\bar{k}} + f_{k1}f_{\bar{k}2} + f_{k2}f_{\bar{k}1} + f_k f_{\bar{k}12}))\\ &= -2\,\mathrm{Re}(f^{12}g^{k\bar{k}}(f_{12k}f_{\bar{k}} + f_{k1}f_{\bar{k}2} + f_{k2}f_{\bar{k}1} + f_k f_{12\bar{k}} - f_k f_m R_{1\bar{m}2k}g^{m\bar{m}}))\\ &= -4\,\mathrm{Re}\langle\nabla f, \bar{\nabla}|\nabla_1\nabla_2 f|^2\rangle - 2\,\mathrm{Re}(f^{12}g^{k\bar{k}}f_{k1}f_{2\bar{k}} + f^{12}g^{k\bar{k}}f_{k2}f_{\bar{k}1}).\end{aligned} \tag{3-24}$$

Combining (3-24) and (3-23), we get

$$\begin{aligned}\Delta|\nabla_1\nabla_2 f|^2 \geq -4\,\mathrm{Re}\langle\nabla f, \bar{\nabla}|\nabla_1\nabla_2 f|^2\rangle + \sum_k (f_{12k}f^{12k} + f_{12\bar{k}}f^{12\bar{k}})&\\ - 2\sum_k (|\nabla_1\nabla_2 f||\nabla_1\nabla_k f||\nabla_2\nabla_{\bar{k}} f| + |\nabla_1\nabla_2 f||\nabla_2\nabla_k f||\nabla_1\nabla_{\bar{k}} f|)&.\end{aligned} \tag{3-25}$$

On the other hand, by Kato's inequality we have

$$\begin{aligned}\Delta|\nabla_1\nabla_2 f|^2 &= 2|\nabla_1\nabla_2 f|\Delta|\nabla_1\nabla_2 f| + 2\big|\nabla|\nabla_1\nabla_2 f|\big|^2\\ &\leq 2|\nabla_1\nabla_2 f|\Delta|\nabla_1\nabla_2 f| + \sum_k |\nabla_k\nabla_1\nabla_2 f|^2 + |\nabla_{\bar{k}}\nabla_1\nabla_2 f|^2\\ &= 2|\nabla_1\nabla_2 f|\Delta|\nabla_1\nabla_2 f| + \sum_k f_{12k}f^{12k} + f_{12\bar{k}}f^{12\bar{k}}.\end{aligned} \tag{3-26}$$

Combining (3-25) and (3-26), it follows that

$$\Delta|\nabla_1\nabla_2 f| \geq -2\,\mathrm{Re}\langle\nabla f, \bar{\nabla}|\nabla_1\nabla_2 f|\rangle - \sum_k (|\nabla_1\nabla_k f||\nabla_2\nabla_{\bar{k}} f| + |\nabla_2\nabla_k f||\nabla_1\nabla_{\bar{k}} f|). \tag{3-27}$$

Combining (3-20) and (3-27) and applying the Cauchy–Schwarz inequality, we have

$$\Delta(|\nabla_1\nabla_2 f|+2(-\Delta_1 f-\Delta_2 f))\ge -2\,\mathrm{Re}\langle\nabla f,\overline{\nabla}(|\nabla_1\nabla_2 f|+2(-\Delta_1 f-\Delta_2 f))\rangle + \sum_{k=1}^{n}(|\nabla_1\nabla_k f|^2+|\nabla_1\nabla_{\bar k} f|^2+|\nabla_2\nabla_k f|^2+|\nabla_2\nabla_{\bar k} f|^2). \tag{3-28}$$

Note that the sum on the right-hand side of (3-27) is (recall under our notation $|\nabla_1\nabla_{\bar 1} f|^2=(\Delta_1 f)^2$) greater than or equal to

$$|\nabla_1\nabla_2 f|^2+|-\Delta_1 f|^2+|-\Delta_2 f|^2\ge \tfrac{1}{12}(|\nabla_1\nabla_2 f|+2(-\Delta_1 f-\Delta_2 f))^2,$$

so we get the equation

$$\Delta(|\nabla_1\nabla_2 f|+2(-\Delta_1 f-\Delta_2 f))\ge -2\,\mathrm{Re}\langle\nabla f,\overline{\nabla}(|\nabla_1\nabla_2 f|+2(-\Delta_1 f-\Delta_2 f))\rangle + \tfrac{1}{12}(|\nabla_1\nabla_2 f|+2(-\Delta_1 f-\Delta_2 f))^2. \tag{3-29}$$

Write

$$Q=\eta^2(|\nabla_1\nabla_2 f|+2(-\Delta_1 f-\Delta_2 f)) =: \eta^2 Q_1,$$

where $\eta(x)=\tilde\eta(r(x)/R)$ and $\tilde\eta$ is a cut-off function such that $\tilde\eta|_{[0,1/3]}=1$ and $\tilde\eta|_{[1/2,1]}=0$. The following arguments are similar to the previous two cases. We calculate

$$\begin{aligned}\Delta Q&=\eta^2\Delta Q_1+2\,\mathrm{Re}\langle\nabla\eta^2,\nabla Q_1\rangle+Q_1\Delta\eta^2\\ &\ge -2\eta^2\,\mathrm{Re}\langle\nabla f,\overline{\nabla}Q_1\rangle+2\,\mathrm{Re}\langle\nabla\eta^2,\nabla Q_1\rangle+\tfrac{1}{12}\eta^2Q_1^2+Q_1\Delta\eta^2.\end{aligned}\tag{3-30}$$

Apply the maximum principle to Q, and if $\max Q\le 0$, we are done. So we may assume that $\max Q>0$ and that it is attained at $p_{\max}$; thus at $p_{\max}$, we have $Q_1>0$, $\Delta Q\le 0$, $\nabla Q_1=-2\eta^{-1}Q_1\nabla\eta$ and

$$Q_1\Delta\eta^2\ge Q_1\frac{2}{R^2}((2n-1)\eta\eta'+\eta\eta''+(\eta')^2).$$

So, at $p_{\max}$,

$$\begin{aligned}0\ge\Delta Q&\ge 4\eta Q_1\,\mathrm{Re}\langle\nabla f,\overline{\nabla}\eta\rangle-8Q_1|\nabla\eta|^2+\eta^2\frac{Q_1^2}{12}+Q_1\frac{2}{R^2}((2n-1)\eta\eta'+\eta\eta''+(\eta')^2)\\ &=\frac{Q^2}{12\eta^2}+4Q\eta^{-1}\,\mathrm{Re}\langle\nabla f,\overline{\nabla}\eta\rangle-\frac{8Q}{\eta^2}\frac{(\eta')^2}{R^2}+\frac{2Q}{R^2\eta^2}((2n-1)\eta\eta'+\eta\eta''+(\eta')^2)\\ &\ge\frac{1}{\eta^2}\left(\frac{Q^2}{12}-\frac{40|\nabla f|}{R}Q-\frac{800}{R^2}Q-\frac{100n}{R^2}Q\right),\end{aligned}\tag{3-31}$$

where we choose η such that $|\eta'|,|\eta''|\le 10$, for example. Therefore, at $p_{\max}\in B\big(p,\tfrac12 R\big)$, we have

$$\frac{Q^2}{12}-Q\left(\frac{40|\nabla f|}{R}+\frac{800}{R^2}+\frac{100n}{R^2}\right)\le 0 \quad\Longrightarrow\quad Q(p_{\max})\le\frac{C(n)}{R^2},$$

since $\sup_{B(p,R/2)}|\nabla f|\le C(n)R^{-1}$ from the previous estimates. Then, for any $x\in B\big(p,\tfrac13 R\big)$, we have

$$Q_1(x)=\eta^2(x)Q_1(x)=Q(x)\le Q(p_{\max})\le\frac{C(n)}{R^2}.$$

Thus it follows that

$$|\nabla_1\nabla_2 f|(x) \le Q_1(x) + 2(\Delta_1 f(x) + \Delta_2 f(x)) \le \frac{C(n)}{R^2} + 2(\Delta_1 f(x) + \Delta_2 f(x)).$$

On the other hand, from $|\nabla_1\nabla_2 f| = |(\nabla_1\nabla_2 u/u) - (\nabla_1 u/u)(\nabla_2 u/u)|$, we get

$$\begin{aligned} |\nabla_1\nabla_2 u|(x) &\le |\nabla_1\nabla_2 f(x)|u(x) + u(x)\frac{|\nabla_1 u(x)|}{u}\frac{|\nabla_2 u(x)|}{u} \\ &\le C(n)\frac{u(x)}{R^2} + 2\Delta_1 u(x) + 2\Delta_2 u(x) + u(x)\frac{|\nabla_1 u(x)|}{u}\frac{|\nabla_2 u(x)|}{u} \\ &\le C(n)\frac{\operatorname{osc}_{B(p,R)} u}{R^2}. \end{aligned} \tag{3-32}$$

Therefore we obtain

$$\sup_{B(p,R/3)} |\nabla_1\nabla_2 u| \le C(n)\frac{\operatorname{osc}_{B(p,R)} u}{R^2}. \tag{3-33}$$

By exactly the same argument we get similar estimates for $|\nabla_1\nabla_{\bar 2} u|$ and $|\nabla_1\nabla_k u| + |\nabla_1\nabla_{\bar k} u|$ for $k \ne 1$.

Hence we have proved the following lemma.

Lemma 3.4. *There exists a constant $C(n) > 0$ such that, for the solution u_ϵ to* (3-4),

$$\sup_{B_{g_\epsilon}(0,R/2)} (|\nabla_i\nabla_j u_\epsilon|_{g_\epsilon} + |\nabla_i\nabla_{\bar j} u_\epsilon|_{g_\epsilon}) \le C(n)\frac{\operatorname{osc}_{B_{g_\epsilon}(0,R)} u_\epsilon}{R^2}$$

for all $i, j = 1, 2, \dots, n$.

3A5. *Convergence of u_ϵ.* In this subsection, we will show that the Dirichlet problem (3-1) admits a unique solution for any $\varphi \in C^0(\partial B_{\boldsymbol\beta}(0, 1))$. Here we will write $B_{\boldsymbol\beta} = B_{\boldsymbol\beta}(0, 1)$ for simplicity of notation.

Proposition 3.5. *For any $\varphi \in C^0(\partial B_{\boldsymbol\beta})$, the Dirichlet boundary value problem* (3-1) *admits a unique solution $u \in C^2(B_{\boldsymbol\beta}\setminus\mathcal{S}) \cap C^0(\overline{B_{\boldsymbol\beta}})$. Moreover, u satisfies the estimates in Lemmas 3.2–3.4 with u_ϵ replaced by u and the metric balls replaced by those under the metric $g_{\boldsymbol\beta}$, which we will refer to as "derivatives estimates" throughout this section.*

Proof. Given the estimates of u_ϵ as in Lemmas 3.2–3.4, we can derive the uniform local $C^{2,\alpha}$ estimates of u_ϵ on any compact subsets of $B_{\boldsymbol\beta}(0, 1)\setminus\mathcal{S}$.

The C^0 estimates of u_ϵ follow immediately from the maximum principle (see Lemma 3.1).

Take any compact subsets $K \Subset K' \Subset B_{\boldsymbol\beta}(0, 1)$. By Lemmas 3.2 and 3.3, we have

$$\sup_{K'}\left(|z_1|^{1-\beta_1}\left|\frac{\partial u_\epsilon}{\partial z_1}\right| + |z_2|^{1-\beta_2}\left|\frac{\partial u_\epsilon}{\partial z_2}\right| + \left|\frac{\partial u_\epsilon}{\partial s_j}\right|\right) \le C(n)\frac{\|u_\epsilon\|_\infty}{d(K', \partial B_{\boldsymbol\beta})}, \tag{3-34}$$

$$\sup_{K'}\left(|z_1|^{1-\beta_1}\left|\frac{\partial^2 u_\epsilon}{\partial s_k\partial z_1}\right| + |z_2|^{1-\beta_2}\left|\frac{\partial^2 u_\epsilon}{\partial s_k\partial z_2}\right| + \left|\frac{\partial^2 u_\epsilon}{\partial s_k\partial s_j}\right|\right) \le C(n)\frac{\|u_\epsilon\|_\infty}{d(K', \partial B_{\boldsymbol\beta})^2}, \tag{3-35}$$

and the third-order estimates

$$\sup_{K'}\left(|z_1|^{1-\beta_1}\left|\frac{\partial^3 u_\epsilon}{\partial z_1\partial s_k\partial s_l}\right| + |z_2|^{1-\beta_2}\left|\frac{\partial^3 u_\epsilon}{\partial z_2\partial s_k\partial s_l}\right| + \left|\frac{\partial^3 u_\epsilon}{\partial s_j\partial s_k\partial s_l}\right|\right) \le C(n)\frac{\|u_\epsilon\|_\infty}{d(K', \partial B_{\boldsymbol\beta})^3}. \tag{3-36}$$

Moreover, applying the gradient estimate to the Δ_{g_ϵ}-harmonic function $\Delta_{1,\epsilon}u_\epsilon$, we get

$$\sup_{K'}\left(|z_1|^{1-\beta_1}\left|\frac{\partial}{\partial z_1}\Delta_{1,\epsilon}u_\epsilon\right|+|z_2|^{1-\beta_2}\left|\frac{\partial}{\partial z_2}\Delta_{1,\epsilon}u_\epsilon\right|+\left|\frac{\partial}{\partial s_j}\Delta_{1,\epsilon}u_\epsilon\right|\right)\le C(n)\frac{\|u_\epsilon\|_\infty}{d(K',\partial B_{\boldsymbol\beta})^3}.$$

From (3-34)–(3-36), we see that the functions u_ϵ have uniform C^3-estimates in the "tangential directions" on any compact subset of $B_{\boldsymbol\beta}(0,1)$. Moreover, for any fixed small constant $\delta>0$, let $T_\delta(\mathcal{S})$ be the tubular neighborhood of $\mathcal{S}$. We consider the equation

$$\Delta_\epsilon u_\epsilon=(|z_1|^2+\epsilon)^{1-\beta_1}\frac{\partial^2 u_\epsilon}{\partial z_1\partial\bar z_1}+(|z_2|^2+\epsilon)^{1-\beta_2}\frac{\partial^2 u_\epsilon}{\partial z_2\partial\bar z_2}+\sum_{j=5}^{2n}\frac{\partial^2 u_\epsilon}{\partial s_j^2}=0\quad\text{on } K'\backslash T_{\delta/2}(\mathcal{S}),$$

which is strictly elliptic (with ellipticity depending only on $\delta>0$). Hence by standard elliptic Schauder theory, we also have $C^{2,\alpha}$-estimates of u_ϵ in the "transversal directions" (i.e., normal to $\mathcal{S}$) and the mixed directions on the compact subset $K\backslash T_\delta(\mathcal{S})$. By taking $\delta\to 0$ and $K\to B_{\boldsymbol\beta}$, and using a diagonal argument, up to a subsequence, the u_ϵ converge in $C^{2,\alpha}_{\mathrm{loc}}(B_{\boldsymbol\beta}\backslash\mathcal{S})$ to a function $u\in C^{2,\alpha}(B_{\boldsymbol\beta}\backslash\mathcal{S})$. Clearly, u satisfies the equation $\Delta_{\boldsymbol\beta}u=0$ on $B_{\boldsymbol\beta}\backslash\mathcal{S}$, and the estimates (3-34)–(3-36) hold for u outside $\mathcal{S}$, which implies that u can be continuously extended through $\mathcal{S}$ and defines a continuous function in $B_{\boldsymbol\beta}(0,1)$. It remains to check the boundary value of u.

Claim: $u=\varphi$ on $\partial B_{\boldsymbol\beta}(0,1)$. It remains to show the limit function u of u_ϵ satisfies the boundary condition $u=\varphi$ on $\partial B_{\boldsymbol\beta}(0,1)$, which will be proved by constructing suitable barriers as we did in [20].

The metric ball $B_{\boldsymbol\beta}(0,1)$ is given by

$$B_{\boldsymbol\beta}(0,1)=\left\{z\in\mathbb{C}^n\;\middle|\;d_{\boldsymbol\beta}(0,z)^2:=|z_1|^{2\beta_1}+|z_2|^{2\beta_2}+\sum_{j=5}^{2n}s_j^2<1\right\}.$$

$B_{\boldsymbol\beta}(0,1)\subset B_{\mathbb{C}^n}(0,1)$, and their boundaries only intersect at $\mathcal{S}_1\cap\mathcal{S}_2$, where $z_1=z_2=0$. Fix any point $q\in\partial B_{\boldsymbol\beta}(0,1)$ and consider the cases $q\in\mathcal{S}_1\cap\mathcal{S}_2$ and $q\notin\mathcal{S}_1\cap\mathcal{S}_2$.

Case 1: $q\in\mathcal{S}_1\cap\mathcal{S}_2$, i.e., $z_1(q)=z_2(q)=0$. Consider the point

$$q'=-q\in\partial B_{\boldsymbol\beta}(0,1)\cap\partial B_{\mathbb{C}^n}(0,1).$$

Since q is the *unique* farthest point from q' on $\partial B_{\boldsymbol\beta}(0,1)$ under the Euclidean distance, the function $\Psi_q(z):=d_{\mathbb{C}^n}(z,q')^2-4$ satisfies $\Psi_q(q)=0$ and $\Psi_q(z)<0$ for all $z\in\partial B_{\boldsymbol\beta}(0,1)\backslash\{q\}$. By the continuity of φ for any $\delta>0$, there is a small neighborhood V of q such that $\varphi(q)-\delta<\varphi(z)<\varphi(q)+\delta$ for all $z\in\partial B_{\boldsymbol\beta}(0,1)\cap V$, and, on $\partial B_{\boldsymbol\beta}(0,1)\backslash V$, we have that Ψ_q is bounded above by a negative constant. Hence we can define

$$\varphi_q(z):=\varphi(q)-\delta+A\Psi_q(z)<\varphi(z)$$

for all $z\in\partial B_{\boldsymbol\beta}(0,1)$ if A is chosen large enough. The function φ_q is Δ_{g_ϵ}-subharmonic; hence by the maximum principle we have $u_\epsilon(z)\ge\varphi_q(z)$ for all $z\in B_{\boldsymbol\beta}(0,1)$. Letting $\epsilon\to 0$ we get $u(z)\ge\varphi_q(z)$, taking $z\to q$ we have $\liminf_{z\to q}u(z)\ge\varphi(q)-\delta$, and since $\delta>0$ is arbitrary we have $\liminf_{z\to q}u(z)\ge\varphi(q)$.

By considering the barrier function $\varphi(q)+\delta-A\Psi_q(z)$ and using a similar argument it is not hard to see that $\limsup_{z\to q}u(z)\le\varphi(q)$; hence $\lim_{z\to q}u(z)=\varphi(q)$ and u is continuous up to $q\in\partial B_{\boldsymbol\beta}(0,1)$.

Case 2: $q \in \partial B_{\boldsymbol{\beta}}(0,1) \backslash \mathcal{S}_1 \cap \mathcal{S}_2$. We first consider the case when $z_1(q) \neq 0$ and $z_2(q) \neq 0$. The boundary $\partial B_{\boldsymbol{\beta}}(0,1)$ is smooth near q, and hence satisfies the exterior sphere condition. We choose an exterior Euclidean ball $B_{\mathbb{C}^n}(\tilde{q}, r_q)$ which is tangential to $\partial B_{\boldsymbol{\beta}}(0,1)$ (only) at q, i.e., under the Euclidean distance, q is the unique closest point to $\tilde{q}$ on $\partial B_{\boldsymbol{\beta}}(0,1)$. So the function

$$G(z) = \frac{1}{|z-\tilde{q}|^{2n-2}} - \frac{1}{r_q^{2n-2}} \tag{3-37}$$

satisfies $G(q) = 0$ and $G(z) < 0$ for all $z \in \partial B_{\boldsymbol{\beta}}(0,1) \backslash \{q\}$. We calculate

$$\begin{aligned}
\Delta_{g_\epsilon} G &= (|z_1|^2+\epsilon)^{-\beta_1+1} \frac{\partial^2 G}{\partial z_1 \partial \bar{z}_1} + (|z_2|^2+\epsilon)^{-\beta_2+1} \frac{\partial^2 G}{\partial z_2 \partial \bar{z}_2} + \sum_{k=3}^n \frac{\partial^2 G}{\partial z_k \partial \bar{z}_k} \\
&= ((|z_1|^2+\epsilon)^{-\beta_1+1}-1) \frac{\partial^2 G}{\partial z_1 \partial \bar{z}_1} + ((|z_2|^2+\epsilon)^{-\beta_2+1}-1) \frac{\partial^2 G}{\partial z_2 \partial \bar{z}_2} \\
&= \sum_{k=1}^2 (-n+1) \frac{(|z_k|^2+\epsilon)^{-\beta_k+1}-1}{|z-\tilde{q}|^{2n}} \left(-\frac{n|z_k-\tilde{q}_k|^2}{|z-\tilde{q}|^2} + 1 \right) \geq -C(q, r_q).
\end{aligned}$$

The function

$$\Psi_q(z) = A(d_{\boldsymbol{\beta}}(z,0)^2 - 1) + G(z)$$

is Δ_{g_ϵ}-subharmonic for $A \gg 1$, and $\Psi_q(q) = 0$ and $\Psi_q(z) < 0$ for all $z \in \partial B_{\boldsymbol{\beta}}(0,1) \backslash \{q\}$. We are in the same situation as Case 1, so by the same argument as above, we can show the continuity of u at such a boundary point q.

In the case when $z_1(q) \neq 0$ and $z_2(q) = 0$, the boundary $\partial B_{\boldsymbol{\beta}}(0,1)$ is not smooth at q and we cannot apply the exterior sphere condition to construct the barrier. Instead we use the geometry of the metric ball $B_{\boldsymbol{\beta}}(0,1)$. Consider the standard cone metric

$$g_{\beta_1} = \beta_1^2 \frac{dz_1 \otimes d\bar{z}_1}{|z_1|^{2(1-\beta_1)}} + \sum_{k=2}^n dz_k \otimes d\bar{z}_k$$

with cone singularity only along $\mathcal{S}_1 = \{z_1 = 0\}$. We observe that the metric ball $B_{\boldsymbol{\beta}}(0,1)$ is strictly contained in $B_{g_{\beta_1}}(0,1)$, and the boundaries of these balls are tangential at the points with vanishing z_2-coordinate. Thus $q \in \partial B_{\boldsymbol{\beta}}(0,1) \cap \partial B_{g_{\beta_1}}(0,1)$ and $\partial B_{g_{\beta_1}}(0,1)$ is smooth at q, so there exists an exterior sphere for $\partial B_{g_{\beta_1}}(0,1)$ at q. We define a similar function $G(z)$ as in (3-37), and, by the strict inclusion of the metric balls $B_{\boldsymbol{\beta}}(0,1) \subset B_{g_{\beta_1}}(0,1)$, it follows that $G(q) = 0$ and $G(z) < 0$ for all $z \in \partial B_{\boldsymbol{\beta}}(0,1) \backslash \{q\}$. The remaining argument is the same as before. □

Remark 3.6. For any constant $c \in \mathbb{R}$, the Dirichlet boundary value problem

$$\Delta_{g_{\boldsymbol{\beta}}} u = c \quad \text{in } B_{\boldsymbol{\beta}}(0,1) \backslash \mathcal{S} \qquad \text{and} \qquad u = \varphi \quad \text{on } \partial B_{\boldsymbol{\beta}}(0,1)$$

admits a solution $u \in C^2(B_{\boldsymbol{\beta}} \backslash \mathcal{S}) \cap C^0(\overline{B_{\boldsymbol{\beta}}})$ for any given $\varphi \in C^0(\partial B_{\boldsymbol{\beta}})$. This follows from the solution $\tilde{u}$ of (3-1) with boundary value $\tilde{\varphi} = \varphi - \frac{1}{2} c(n-2)^{-1} \sum_{j=5}^{2n} s_j^2$. Then the function $u = \tilde{u} + \frac{1}{2} c(n-2)^{-1} \sum_j s_j^2$ solves the equation above.

For later application, we prove the existence of solutions for a more general right-hand side of the Laplace equation with the standard background metric. This result is not needed to prove Theorem 1.2.

Proposition 3.7. *For any given $\varphi \in C^0(\partial B_{\boldsymbol\beta}(0,1))$ and $f \in C^{0,\alpha}_{\boldsymbol\beta}(\overline{B_{\boldsymbol\beta}(0,1)})$, the Dirichlet boundary value problem*

$$\begin{cases} \Delta_{g_{\boldsymbol\beta}} v = f & \text{in } B_{\boldsymbol\beta}(0,1)\backslash\mathcal{S}, \\ v = \varphi & \text{on } \partial B_{\boldsymbol\beta}(0,1) \end{cases} \tag{3-38}$$

admits a unique solution $v \in C^2(B_{\boldsymbol\beta}(0,1)\backslash\mathcal{S}) \cap C^0(\overline{B_{\boldsymbol\beta}(0,1)})$.

By Theorem 1.2, the solution v to (3-38) belongs to $C^{2,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta}(0,1)) \cap C^0(\overline{B_{\boldsymbol\beta}(0,1)})$.

Proof. The proof is similar to that of Proposition 3.5. As before, let g_ϵ be the approximating metrics (3-3) of $g_{\boldsymbol\beta}$ which are smooth metrics on $B_{\boldsymbol\beta}(0,1)$. By standard elliptic theory we can solve the equations

$$\begin{cases} \Delta_{g_\epsilon} v_\epsilon = f & \text{in } B_{\boldsymbol\beta}(0,1), \\ v_\epsilon = \varphi & \text{on } \partial B_{\boldsymbol\beta}(0,1). \end{cases} \tag{3-39}$$

For any compact subset $K \Subset B_{\boldsymbol\beta}(0,1)$ and small $\delta > 0$, we have a uniform $C^{2,\alpha'}$-bound of v_ϵ on $K\backslash T_\delta(\mathcal{S})$ for some $\alpha' < \alpha$. Thus v_ϵ converges in the $C^{2,\alpha'}$-norm to a function v on $K\backslash T_\delta(\mathcal{S})$ as $\epsilon \to 0$. By a standard diagonal argument, letting $K \to B_{\boldsymbol\beta}(0,1)$ and $\delta \to 0$, we can achieve

$$v_\epsilon \xrightarrow{C^{2,\alpha'}_{\mathrm{loc}}(B_{\boldsymbol\beta}(0,1)\backslash\mathcal{S})} v \in C^{2,\alpha'}_{\mathrm{loc}}(B_{\boldsymbol\beta}(0,1)\backslash\mathcal{S}) \quad \text{as } \epsilon \to 0.$$

Clearly v satisfies (3-38) in $B_{\boldsymbol\beta}(0,1)\backslash\mathcal{S}$. It only remains to show the boundary value of v coincides with φ and v is globally continuous in $B_{\boldsymbol\beta}(0,1)$.

Global continuity: $v \in C^0(B_{\boldsymbol\beta}(0,1))$. It suffices to show v is continuous at any $p \in \mathcal{S} \cap B_{\boldsymbol\beta}(0,1)$. Fix such a point p and take $R_0 > 0$ small enough that $B_{\mathbb{C}^n}(p, 10R_0) \cap \partial B_{\boldsymbol\beta}(0,1) = \varnothing$. We observe that $\frac{1}{2} g_{\mathbb{C}^n} \le g_\epsilon \le g_{\boldsymbol\beta}$, so for any $r \in \left(0, \frac{1}{2}\right)$,

$$B_{g_{\boldsymbol\beta}}(p,r) \subset B_{g_\epsilon}(p,r) \subset B_{\mathbb{C}^n}(p,2r). \tag{3-40}$$

In particular, the balls $B_{g_\epsilon}(p, 5R_0)$ are also disjoint with $\partial B_{\boldsymbol\beta}(0,1)$.

Since $\mathrm{Ric}(g_\epsilon) \ge 0$, we have the following Sobolev inequality [25]: there exists a constant $C = C(n) > 0$ such that, for any $h \in C^1_0(B_{g_\epsilon}(p,r))$,

$$\left(\int_{B_{g_\epsilon}(p,r)} h^{2n/(n-1)} \omega_\epsilon^n \right)^{(n-1)/n} \le C \left(\frac{r^{2n}}{\mathrm{Vol}_{g_\epsilon}(B_{g_\epsilon}(p,r))} \right)^{1/n} \int_{B_{g_\epsilon}(p,r)} |\nabla h|^2_{g_\epsilon} \omega_\epsilon^n. \tag{3-41}$$

It can be checked by straightforward calculations that $\mathrm{Vol}_{g_\epsilon}(B_{g_\epsilon}(p,1)) \ge c_0(n) > 0$ for some constant c_0 independent of ϵ. Then Bishop's volume comparison yields, for any $r \in (0,1)$,

$$C_1(n) r^{2n} \ge \mathrm{Vol}_{g_\epsilon}(B_{g_\epsilon}(p,r)) \ge c_1(n) r^{2n}.$$

Thus the Sobolev inequality (3-41) is reduced to

$$\left(\int_{B_{g_\epsilon}(p,r)} h^{2n/(n-1)} \omega_\epsilon^n \right)^{(n-1)/n} \le C \int_{B_{g_\epsilon}(p,r)} |\nabla h|^2_{g_\epsilon} \omega_\epsilon^n \quad \text{for all } h \in C^1_0(B_{g_\epsilon}(p,r)). \tag{3-42}$$

With (3-42) at hand, we can apply the same proof of the standard De Giorgi–Nash–Moser theory (see the proof of Corollary 4.18 in [22]) to derive the uniform Hölder continuity of v_ϵ at p, i.e., there exists a constant $C = C(n, \boldsymbol{\beta}, R_0) > 0$ such that

$$\operatorname{osc}_{B_{\boldsymbol{\beta}}(p,r)} v_\epsilon \le \operatorname{osc}_{B_{g_\epsilon}(p,r)} v_\epsilon \le C r^{\alpha''} \quad \text{for all } r \in (0, R_0)$$

for some $\alpha'' = \alpha''(n, \boldsymbol{\beta}, R_0) \in (0, 1)$, where in the first inequality we use the relation (3-40). Letting $\epsilon \to 0$ we see the continuity of v at p.

Boundary value: $v = \varphi$ on $\partial B_{\boldsymbol{\beta}}(0, 1)$. The proof is almost identical to that of Proposition 3.5. For example, the function $\varphi_q(z) = \varphi(q) - \delta + A\Psi_q(z)$ defined in Case 1 in the proof of Proposition 3.5 satisfies $\Delta_{g_\epsilon}\varphi_q(z) \ge \max_X f$ if $A > 0$ is taken large enough. Then from $\Delta_{g_\epsilon}(\varphi_q - v_\epsilon) \ge 0$ in $B_{\boldsymbol{\beta}}$ and $\varphi_q - \varphi \le 0$ on $\partial B_{\boldsymbol{\beta}}$, applying the maximum principle we get $\varphi_q \le v_\epsilon$ in $B_{\boldsymbol{\beta}}(0, 1)$. The remaining arguments are the same as in Proposition 3.5. Case 2 can be dealt with similarly. □

Remark 3.8. Let $H_0^1(B_{\boldsymbol{\beta}}(0, 1), g_{\boldsymbol{\beta}})$ be the completion of the space of $C_0^1(B_{\boldsymbol{\beta}}(0, 1))$-functions under the norm

$$\|\nabla u\|_{L^2(g_{\boldsymbol{\beta}})} = \left(\int_{B_{\boldsymbol{\beta}}(0,1)} |\nabla u|^2_{g_{\boldsymbol{\beta}}} \omega^n_{\boldsymbol{\beta}}\right)^{1/2}.$$

For any $h \in C_0^1(B_{\boldsymbol{\beta}}(0, 1))$, letting $\epsilon \to 0$ in (3-42), we get

$$\left(\int_{B_{\boldsymbol{\beta}}(p,r)} |h|^{2n/n-1} \omega^n_{\boldsymbol{\beta}}\right)^{(n-1)/n} \le C \int_{B_{\boldsymbol{\beta}}(p,r)} |\nabla h|^2_{g_{\boldsymbol{\beta}}} \omega^n_{\boldsymbol{\beta}} \tag{3-43}$$

for the same constant C in (3-42). That is, the Sobolev inequality also holds for the conical metric $\omega_{\boldsymbol{\beta}}$.

3B. *Tangential and Laplacian estimates.* In this section, we will prove the Hölder continuity of $\Delta_k u$ for $k = 1, 2$ and $(D')^2 u$ for the solution u to (1-2). The arguments of [20] can be adopted here. We recall that we assume $\beta_1, \beta_2 \in \left(\frac{1}{2}, 1\right)$. We fix some notations first.

For a given point $p \notin \mathcal{S}$, we define $r_p = d_{g_{\boldsymbol{\beta}}}(p, \mathcal{S})$, the $g_{\boldsymbol{\beta}}$-distance of p to the singular set $\mathcal{S}$. For simplicity of notation we will fix $\tau = \frac{1}{2}$ and an integer $k_p \in \mathbb{Z}_+$ to be the smallest integer such that $\tau^{k_p} < r_p$, and $k_{i,p} \in \mathbb{Z}_+$ the smallest integer $k_{i,p}$ such that $\tau^{k_{i,p}} < d_{\boldsymbol{\beta}}(p, \mathcal{S}_i)$ for $i = 1, 2$. So $k_p = \max\{k_{1,p}, k_{2,p}\}$. We write $p_1 \in \mathcal{S}_1$ and $p_2 \in \mathcal{S}_2$ for the projections of p to $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively.

For $j = 1, 2$, we will write

$$\Delta_j u := |z_j|^{2(1-\beta_j)} \frac{\partial^2 u}{\partial z_j \partial \bar{z}_j}.$$

We will consider a family of conical Laplace equations with different choices of $k \in \mathbb{Z}^+$.

(i) If $k \ge k_p$, the geodesic balls $B_{\boldsymbol{\beta}}(p, \tau^k)$ are disjoint from $\mathcal{S}$ and have smooth boundaries. We note that $g_{\boldsymbol{\beta}}$ is smooth on such balls. By standard theory we can solve the following Dirichlet problem for $u_k \in C^\infty(B_{\boldsymbol{\beta}}(p, \tau^k)) \cap C^0(\overline{B_{\boldsymbol{\beta}}(p, \tau^k)})$:

$$\begin{cases} \Delta_{\boldsymbol{\beta}} u_k = f(p) & \text{in } B_{\boldsymbol{\beta}}(p, \tau^k), \\ u_k = u & \text{on } \partial B_{\boldsymbol{\beta}}(p, \tau^k). \end{cases} \tag{3-44}$$

(ii) Without loss of generality, we assume $d_{\boldsymbol\beta}(p,\mathcal{S}_1)\le d_{\boldsymbol\beta}(p,\mathcal{S}_2)$, i.e., $k_{1,p}\ge k_{2,p}$. We now solve the following Dirichlet problem for $u_k\in C^2(B_{\boldsymbol\beta}(p_1,2\tau^k)\backslash\mathcal{S}_1)\cap C^0(\overline{B_{\boldsymbol\beta}(p_1,2\tau^k)})$ for $k_{2,p}+2\le k<k_{1,p}$:

$$\begin{cases}\Delta_{\boldsymbol\beta}u_k=f(p) & \text{in } B_{\boldsymbol\beta}(p_1,2\tau^k)\backslash\mathcal{S}_1,\\ u_k=u & \text{on } \partial B_{\boldsymbol\beta}(p_1,2\tau^k).\end{cases}\tag{3-45}$$

By similar arguments to those in the proof of Proposition 3.5 and Remark 3.6, such u_k exists.

(iii) For $2\le k\le k_{2,p}+1$, let $u_k\in C^2(B_{\boldsymbol\beta}(p_{1,2},2\tau^k)\backslash\mathcal{S})\cap C^0(\overline{B_{\boldsymbol\beta}(p_{1,2},2\tau^k)})$ solve the equation

$$\begin{cases}\Delta_{\boldsymbol\beta}u_k=f(p) & \text{in } B_{\boldsymbol\beta}(p_{1,2},2\tau^k),\\ u_k=u & \text{on } \partial B_{\boldsymbol\beta}(p_{1,2},2\tau^k),\end{cases}\tag{3-46}$$

whose existence follows from Remark 3.6. Here $p_{1,2}=(0;0;s(p))\in\mathcal{S}_1\cap\mathcal{S}_2$ is the projection of p_1 to $\mathcal{S}_2$.

We remark that we may take $f(p)=0$ by considering $\tilde u=u-\frac12 f(p)(n-2)^{-1}|s-s(p)|^2$. If the estimate holds for $\tilde u$, it also holds for u. So from now on we assume $f(p)=0$.

Lemma 3.9. *Let u_k be the solutions to* (3-44)–(3-46). *There exists a constant $C=C(n)>0$ such that, for all $k\in\mathbb{Z}_+$, we have the estimates*

$$\|u_k-u\|_{L^\infty(\hat B_k(p))}\le C(n)\tau^{2k}\omega(\tau^k),\tag{3-47}$$

where we define $\hat B_k(p)$ as

$$\hat B_k(p):=\begin{cases}B_{\boldsymbol\beta}(p,\tau^k) & \text{if } k\ge k_p,\\ B_{\boldsymbol\beta}(p_1,2\tau^k) & \text{if } k_{2,p}+2\le k<k_{1,p},\\ B_{\boldsymbol\beta}(p_{1,2},2\tau^k) & \text{if } 2\le k\le k_{2,p}+1\end{cases}\tag{3-48}$$

for different choices of $k\in\mathbb{Z}_+$.

We will also define $\lambda\hat B_k(p)$ to be the ball concentric with $\hat B_k(p)$ with radius scaled by $\lambda\in(0,1)$.

This lemma follows straightforwardly from Lemma 3.1 and the definition of $\omega(r)$, so we omit the proof. By the triangle inequality, we get the estimates

$$\|u_k-u_{k+1}\|_{L^\infty(\hat B_k/2)}\le C(n)\tau^{2k}\omega(\tau^k).\tag{3-49}$$

Since u_k-u_{k+1} are $g_{\boldsymbol\beta}$-harmonic functions on $\frac12\hat B_k$, applying the gradient and Laplacian estimates (3-5) and (3-9) for harmonic functions, we get the following lemma.

Lemma 3.10. *There exists a constant $C(n)>0$ such that, for all $k\in\mathbb{Z}_+$,*

$$\|D'u_k-D'u_{k+1}\|_{L^\infty(\hat B_k/3)}\le C(n)\tau^k\omega(\tau^k)\tag{3-50}$$

and

$$\sup_{(\hat B_k/3)\backslash\mathcal{S}}\left(\sum_{i=1}^2|\Delta_i(u_k-u_{k+1})|+|(D')^2u_k-(D')^2u_{k+1}|\right)\le C(n)\omega(\tau^k),\tag{3-51}$$

where we recall that D' denotes the first-order operators $\partial/\partial s_i$ for $i=5,\dots,2n$.

The following lemma can be proved by looking at the Taylor expansion of u_k at p for $k\gg1$ as in Lemma 2.8 of [20].

Lemma 3.11. *For $i = 1, 2$, we have the limits*

$$\lim_{k\to\infty} D'u_k(p) = D'u(p), \quad \lim_{k\to\infty} (D')^2u_k(p) = (D')^2u(p), \quad \lim_{k\to\infty} \Delta_i u_k(p) = \Delta_i u(p). \tag{3-52}$$

Combining Lemmas 3.10 and 3.11, we obtain estimates on the second-order (tangential) derivatives.

Proposition 3.12. *There exists a constant $C = C(n, \beta) > 0$ such that*

$$\sup_{B_\beta(0,1/2)\setminus\mathcal{S}} |(D')^2u| + |\Delta_i u| \le C\biggl(\|u\|_{L^\infty(B_\beta(0,1))} + \int_0^1 \frac{\omega(r)}{r}\,dr + |f(0)|\biggr). \tag{3-53}$$

Proof. From the triangle inequality we have, for any given $z \in B_{\boldsymbol\beta}\bigl(0, \frac{1}{2}\bigr)\setminus\mathcal{S}$,

$$\begin{aligned}|(D')^2u(z)| &\le \sum_{k=2}^{\infty} |(D')^2u_k(z) - (D')^2u_{k+1}(z)| + |(D')^2u_2(z)| \\ &\le C(n)\sum_{k=2}^{\infty} \omega(\tau^k) + C(n)\operatorname{osc}_{B_\beta(0,1)} u_0 \le C(n, \boldsymbol\beta)\biggl(\|u\|_{L^\infty} + \int_0^1 \frac{\omega(r)}{r}\,dr + |f(0)|\biggr).\end{aligned}$$

The estimates for $\Delta_i u$ can be proved similarly. □

For any other given point $q \in B_{\boldsymbol\beta}\bigl(0, \frac{1}{2}\bigr)\setminus\mathcal{S}$, we can solve a similar Dirichlet boundary problems as u_k with the metric balls centered at q, and we obtain a family of functions v_k such that

$$\Delta_{\boldsymbol\beta} v_k = f(q) \quad \text{in } \widetilde{B}_k(q), \qquad v_k = u \quad \text{on } \partial\widetilde{B}_k(q), \tag{3-54}$$

where $\widetilde{B}_k(q)$ are metrics balls centered at q given by

$$\widetilde{B}_k(q) = \widetilde{B}_k := \begin{cases} B_{\boldsymbol\beta}(q, \tau^k) & \text{if } k \ge k_q, \\ B_{\boldsymbol\beta}(q_i, 2\tau^k) & \text{if } k_{j,q} + 2 \le k < k_q \text{ (here } k_{i,q} = \max(k_{1,q}, k_{2,q}) \text{ and } j \ne i), \\ B_{\boldsymbol\beta}(q_{i,j}, 2\tau^k) & \text{if } k \le k_{j,q} + 1. \end{cases}$$

Similar estimates as in Lemmas 3.9–3.11 also hold for v_k within the balls $\widetilde{B}_k(q)$.

We are now ready to state the main result in this subsection on the continuity of second-order derivatives.

Proposition 3.13. *Let $d = d_{\boldsymbol\beta}(p, q) < \frac{1}{16}$. There exists a constant $C = C(n) > 0$ such that if u solves the conical Laplace equation* (1-2), *then the following holds for $i = 1, 2$:*

$$|\Delta_i u(p) - \Delta_i u(q)| + |(D')^2u(p) - (D')^2u(q)| \le C\biggl(d\|u\|_{L^\infty(B_\beta(0,1))} + \int_0^d \frac{\omega(r)}{r}\,dr + d\int_d^1 \frac{\omega(r)}{r^2}\,dr\biggr).$$

Proof. We only prove the estimate for $(D')^2u$; the estimates for $\Delta_i u$ can be dealt with in the same way.

We may assume $r_p = \min(r_p, r_q)$. We fix $\ell \in \mathbb{Z}$ such that τ^ℓ is comparable to d; more precisely, take

$$\tau^{\ell+4} \le d < \tau^{\ell+3} \quad \text{or} \quad \tau^{\ell+1} \le 8d \le \tau^\ell.$$

We calculate by the triangle inequality

$$\begin{aligned}|(D')^2u(p) - (D')^2u(q)| &\le |(D')^2u(p) - (D')^2u_\ell(p)| + |(D')^2u_\ell(p) - (D')^2u_\ell(q)| \\ &\qquad + |(D')^2u_\ell(q) - (D')^2v_\ell(q)| + |(D')^2v_\ell(q) - (D')^2u(q)| \\ &=: I_1 + I_2 + I_3 + I_4.\end{aligned}$$

We will estimate I_1–I_4 one by one.

I_1 and I_4: By (3-51) and (3-52), we have

$$I_1 = |(D')^2u(p) - (D')^2u_\ell(p)| \le C(n)\sum_{k=\ell}^{\infty}\omega(\tau^k),$$

and a similar estimate holds for I_4 as well:

$$I_4 = |(D')^2u(q) - (D')^2v_\ell(q)| \le C(n)\sum_{k=\ell}^{\infty}\omega(\tau^k).$$

I_3: By the choice of ℓ, it is not hard to see that $\frac{2}{3}\widetilde{B}_\ell(q) \subset \hat{B}_\ell(p)$. In particular, u_ℓ and v_ℓ are both defined on $\frac{2}{3}\widetilde{B}_\ell(q)$ and satisfy the equations

$$\Delta_{\boldsymbol{\beta}} u_\ell = f(p) \quad \text{and} \quad \Delta_{\boldsymbol{\beta}} v_\ell = f(q),$$

respectively, on this ball. From (3-47) for u_ℓ and from a similar estimate for v_ℓ, we get

$$\|u_\ell - v_\ell\|_{L^\infty(2\widetilde{B}_\ell(q)/3)} \le C\tau^{2\ell}\omega(\tau^\ell).$$

Consider the function

$$U := u_\ell - v_\ell - \frac{f(p)-f(q)}{2(n-2)}|s - s(\tilde{q})|^2, \tag{3-55}$$

where $\tilde{q}$ is the center of the ball $\widetilde{B}_\ell(q)$. U is $g_{\boldsymbol{\beta}}$-harmonic in $\frac{2}{3}\widetilde{B}_\ell(q)$ and satisfies the estimate

$$\|U\|_{L^\infty(2\widetilde{B}/3_\ell(q))} \le C\tau^{2\ell}\omega(\tau^\ell) + C\tau^{2\ell}\omega(d) \le C(n)\tau^{2\ell}\omega(\tau^\ell).$$

The derivatives estimates imply that

$$|(D')^2U(q)| \le C\tau^{-2\ell}\|U\|_{L^\infty(2\widetilde{B}/3_\ell(q))} \le C(n)\omega(\tau^\ell).$$

Hence

$$I_3 = |(D')^2u_\ell(q) - (D')^2v_\ell(q)| \le C(n)\omega(\tau^\ell).$$

I_2: This is a little more complicated than the previous estimates. We define $h_k = u_{k-1} - u_k$ for $k \le \ell$. We observe that h_k is $g_{\boldsymbol{\beta}}$-harmonic on $\hat{B}_k(p)$ and by (3-47) satisfies the L^∞-estimate $\|h_k\|_{\hat{B}_k(p)} \le C\tau^{2k}\omega(\tau^k)$ and the derivatives estimates $\|(D')^2h_k\|_{L^\infty(2\hat{B}_k(p)/3)}\omega(\tau^k)$. On the other hand, the function $(D')^2h_k$ is also $g_{\boldsymbol{\beta}}$-harmonic on $\frac{2}{3}\hat{B}_k(p)$, so the gradient estimate implies that

$$\|\nabla_{g_{\boldsymbol{\beta}}}(D')^2h_k\|_{L^\infty((\hat{B}_k(p)/2)\setminus\mathcal{S})} \le C\tau^{-k}\omega(\tau^k). \tag{3-56}$$

Integrating this along the minimal $g_{\boldsymbol{\beta}}$-geodesic γ connecting p and q and noting that γ avoids $\mathcal{S}$ since $(\mathbb{C}^n\setminus\mathcal{S}, g_{\boldsymbol{\beta}})$ is strictly geodesically convex, we get

$$|(D')^2h_k(p) - (D')^2h_k(q)| \le d\cdot\|\nabla_{g_{\boldsymbol{\beta}}}(D')^2h_k\|_{L^\infty((\hat{B}_k(p)/2)\setminus\mathcal{S})} \le dC\tau^{-k}\omega(\tau^k).$$

By the triangle inequality, for each $k \le \ell$,

$$I_2 = |(D')^2u_\ell(p) - (D')^2u_\ell(q)| \le |(D')^2u_2(p) - (D')^2u_2(q)| + dC\sum_{k=2}^{\ell}\tau^{-k}\omega(\tau^k). \tag{3-57}$$

Observe that $p, q \in \hat{B}_2(p)$ and the function $(D')^2 u_2$ is $g_{\boldsymbol{\beta}}$-harmonic on $\hat{B}_2(p)$. From (3-47) and derivatives estimates we have

$$\|(D')^2 u_2\|_{L^\infty(2\hat{B}_2(p)/3)} \le C\|u_2\|_{L^\infty(\hat{B}_2(p))} \le C(\|u\|_{L^\infty} + \omega(\tau^2)).$$

Again by the gradient estimate we have

$$\|\nabla_{g_{\boldsymbol{\beta}}}(D')^2 u_2\|_{L^\infty(\hat{B}_2(p)/2)} \le C(\|u\|_{L^\infty} + \omega(\tau^2)).$$

Integrating along the minimal geodesic γ we arrive at

$$|(D')^2 u_2(p) - (D')^2 u_2(q)| \le dC(\|u\|_{L^\infty} + \omega(\tau^2)).$$

Combining this with (3-57), we obtain

$$I_2 \le Cd\biggl(\|u\|_{L^\infty(B_{\boldsymbol{\beta}}(0,1))} + \sum_{k=2}^{\ell} \tau^{-k}\omega(\tau^k)\biggr).$$

Combing the estimates for I_1–I_4, we get

$$|(D')^2 u(p) - (D')^2 u(q)| \le C\biggl(d\biggl(\|u\|_{L^\infty(B_{\boldsymbol{\beta}}(0,1))} + \sum_{k=2}^{\ell} \tau^{-k}\omega(\tau^k)\biggr) + \sum_{k=\ell}^{\infty} \omega(\tau^k)\biggr).$$

Proposition 3.13 now follows from this and the fact that $\omega(r)$ is monotonically increasing. □

3C. ***Mixed normal-tangential estimates along the directions $\mathcal{S}$.*** Throughout this section, we fix two points $p, q \in B_{\boldsymbol{\beta}}\bigl(0, \frac{1}{2}\bigr)\backslash\mathcal{S}$ and assume $r_p \le r_q$. Recall that we defined the weighted "polar coordinates" (r_i, θ_i) for (z_1, z_2):

$$\rho_i = |z_i|, \quad r_i = \rho_i^{\beta_i}, \quad \theta_i = \arg z_i, \qquad i = 1, 2.$$

Under these coordinates,

$$\Delta_i u = |z_i|^{2(1-\beta_i)} \frac{\partial^2 u}{\partial z_i \partial \bar{z}_i} = \frac{\partial^2 u}{\partial r_i^2} + \frac{1}{r_i}\frac{\partial u}{\partial r_i} + \frac{1}{\beta_i^2 r_i^2}\frac{\partial^2 u}{\partial \theta_i^2}. \tag{3-58}$$

Let u_k and v_k be the solutions to (3-44)–(3-46) on $\hat{B}_k(p)$ and $\widetilde{B}_k(q)$, respectively. Recalling that $u_k - u_{k+1}$ satisfies (3-49) and applying gradient estimates to the $g_{\boldsymbol{\beta}}$-harmonic function $u_k - u_{k+1}$, we get the bound of $\|\nabla_{g_{\boldsymbol{\beta}}}(u_k - u_{k+1})\|_{L^\infty(\hat{B}_k(p)/3)}$, which in particular implies that, for $i = 1, 2$,

$$\biggl\| |z_i|^{1-\beta_i}\biggl(\frac{\partial u_k}{\partial z_i} - \frac{\partial u_{k+1}}{\partial z_i}\biggr)\biggr\|_{L^\infty(\hat{B}_k(p)/3)} \le C\tau^k\omega(\tau^k). \tag{3-59}$$

Similarly, $D'u_k - D'u_{k+1}$ is also $g_{\boldsymbol{\beta}}$-harmonic on $\frac{1}{2}\hat{B}_k(p)$, and applying gradient estimates to this function we get, for $i = 1, 2$,

$$\biggl\| |z_i|^{1-\beta_i}\biggl(\frac{\partial D'u_k}{\partial z_i} - \frac{\partial D'u_{k+1}}{\partial z_i}\biggr)\biggr\|_{L^\infty(\hat{B}_k(p)/3)} \le C\omega(\tau^k). \tag{3-60}$$

The next lemma can be proved in the same way as Lemma 2.10 of [20] since $p \notin \mathcal{S}$; we omit the proof.

Lemma 3.14. *For $i = 1, 2$, we have the limits*

$$\lim_{k\to\infty} \frac{\partial u_k}{\partial r_i}(p) = \frac{\partial u}{\partial r_i}(p), \quad \lim_{k\to\infty} \frac{\partial u_k}{r_i \partial \theta_i}(p) = \frac{\partial u}{r_i \partial \theta_i}(p)$$

and

$$\lim_{k\to\infty} \frac{\partial D' u_k}{\partial r_i}(p) = \frac{\partial D' u}{\partial r_i}(p), \quad \lim_{k\to\infty} \frac{\partial D' u_k}{r_i \partial \theta_i}(p) = \frac{\partial D' u}{r_i \partial \theta_i}(p). \tag{3-61}$$

Similar formulas also hold for v_k at the point q.

We are going to estimate the quantities

$$J := \left| \frac{\partial D' u}{\partial r_i}(p) - \frac{\partial D' u}{\partial r_i}(q) \right| \quad \text{and} \quad K := \left| \frac{\partial D' u}{r_i \partial \theta_i}(p) - \frac{\partial D' u}{r_i \partial \theta_i}(q) \right|, \qquad i = 1, 2.$$

Note that J, K correspond to $|N_j D'u(p) - N_j D'u(q)|$ in Theorem 1.2. We will estimate the case for $i = 1$ and J, since the other cases are completely the same. By the triangle inequality we have

$$\begin{aligned} J \le & \left| \frac{\partial D' u}{\partial r_i}(p) - \frac{\partial D' u_\ell}{\partial r_i}(p) \right| + \left| \frac{\partial D' u_\ell}{\partial r_i}(p) - \frac{\partial D' u_\ell}{\partial r_i}(q) \right| \\ & \qquad + \left| \frac{\partial D' u_\ell}{\partial r_i}(q) - \frac{\partial D' v_\ell}{\partial r_i}(q) \right| + \left| \frac{\partial D' v_\ell}{\partial r_i}(q) - \frac{\partial D' u}{\partial r_i}(q) \right| \\ =: & \, J_1 + J_2 + J_3 + J_4. \end{aligned}$$

Lemma 3.15. *There exists a constant $C(n) > 0$ such that J_1, J_3 and J_4 satisfy*

$$J_1 + J_4 \le C \sum_{k=\ell}^{\infty} \omega(\tau^k), \quad J_3 \le C\omega(\tau^\ell).$$

Proof. The estimates for J_1 and J_4 can be proved similarly to those of I_1 and I_4 in Section 3B, using (3-60) and (3-61). J_3 can be estimated in a similar way to I_3 in Section 3B, using (3-60). We omit the details. □

To estimate J_2, as in Section 3B we define $h_k := u_{k-1} - u_k$ for $2 \le k \le \ell$ which is $g_{\boldsymbol{\beta}}$-harmonic on $\hat{B}_k(p)$ and satisfies the L^∞-estimate $\|h_k\|_{L^\infty(\hat{B}_k(p))} \le C\tau^{2k}\omega(\tau^k)$ by (3-60). We rewrite (3-56) as

$$\|(D')^3 h_k\|_{L^\infty((\hat{B}_k(p)/2)\setminus\mathcal{S})} + \sum_{i=1}^{2} \left\| |z_i|^{1-\beta_i} \frac{\partial}{\partial z_i}(D')^2 h_k \right\|_{L^\infty((\hat{B}_k(p)/2)\setminus\mathcal{S})} \le C\tau^{-k}\omega(\tau^k). \tag{3-62}$$

Lemma 3.16. *There exists a constant $C = C(n, \boldsymbol{\beta}) > 0$ such that, for any $z \in \frac{1}{4}\hat{B}_k(p)\setminus\mathcal{S}$, the following pointwise estimate holds for all $k \le \min(\ell, k_p)$:*

$$\left| \frac{\partial D' h_k}{\partial r_1}(z) \right| + \left| \frac{\partial D' h_k}{r_1 \partial \theta_1}(z) \right| \le C r_1(z)^{1/\beta_1 - 1} \tau^{-k(1/\beta_1 - 1)} \omega(\tau^k).$$

Proof. We define a function F as

$$|z_1|^{2(1-\beta_1)} \frac{\partial^2 D' h_k}{\partial z_1 \partial \bar{z}_1} = -|z_2|^{2(1-\beta_2)} \frac{\partial^2 D' h_k}{\partial z_2 \partial \bar{z}_2} - \sum_{j=5}^{2n} \frac{\partial^2 D' h_k}{\partial s_j^2} =: F. \tag{3-63}$$

The Laplacian estimates (3-9) and derivatives estimates applied to the $g_{\boldsymbol\beta}$-harmonic function $D'h_k$ imply that F satisfies

$$\|F\|_{L^\infty(\hat B_k(p)/2)} \le C(n)\tau^{-k}\omega(\tau^k). \tag{3-64}$$

For any $k \le \min(\ell, k_p)$ and $x \in \mathcal{S}_1 \cap \frac14 \hat B_k(p)$, we have that $B_{\boldsymbol\beta}(x, \tau^k) \subset \frac13 \hat B_k(p)$. The intersection of $B_{\boldsymbol\beta}(x, \tau^k)$ with the complex plane $\mathbb{C}$ passing through x and orthogonal to the hyperplane $\mathcal{S}_1$ lies in a metric ball of radius τ^k under the standard cone metric $\hat g_{\beta_1}$ on $\mathbb{C}$. We view (3-63) as defined on the ball $\hat B := B_{\mathbb{C}}(x, (\tau^k)^{1/\beta_1}) \subset \mathbb{C}$. The estimate (2-3) applied to the function $D'h_k$ gives rise to

$$\sup_{B_{\mathbb{C}}(x,(\tau^k)^{1/\beta_1}/2)\setminus\{x\}} \left|\frac{\partial D'h_k}{\partial z_1}\right| \le C\frac{\|D'h_k\|_{L^\infty(\hat B)}}{(\tau^k)^{1/\beta_1}} + C\|F\|_{L^\infty(\hat B)}(\tau^k)^{2-1/\beta_1}.$$

Therefore, on $B_{\mathbb{C}}\big(x, \frac12(\tau^k)^{1/\beta_1}\big)\setminus\{x\}$,

$$\left|\frac{\partial D'h_k}{\partial r_1}(z)\right| + \left|\frac{\partial D'h_k}{r_1\partial\theta_1}(z)\right| \le \frac{1}{\beta_1} r_1^{1/\beta_1-1}\left|\frac{\partial D'h_k}{\partial z_1}(z)\right| \le C r_1^{1/\beta_1-1}\tau^{k(1-1/\beta_1)}\omega(\tau^k). \tag{3-65}$$

On other hand, since $B_{\mathbb{C}}\big(x, \frac12(\tau^k)^{1/\beta_1}\big) = B_{\hat g_{\beta_1}}(x, 2^{-\beta_1}\tau^k)$,

$$\tfrac14 \hat B_k(p) \subset \bigcup_{x\in\mathcal{S}_1\cap\hat B_k/4} B_{\mathbb{C}}\big(x, \tfrac12(\tau^k)^{1/\beta_1}\big). \tag{3-66}$$

Equation (3-65) implies the desired estimate on the balls $\frac14 \hat B_k(p)$. □

Remark 3.17. By similar arguments we also get the following estimates for any $k \le \min(\ell, k_p)$ and $z \in \frac14 \hat B_k(p)\setminus\mathcal{S}_1$:

$$\left|\frac{\partial (D')^2 h_k}{\partial r_1}(z)\right| + \left|\frac{\partial (D')^2 h_k}{r_1\partial\theta_1}(z)\right| \le C r_1(z)^{1/\beta_1-1}\tau^{-k/\beta_1}\omega(\tau^k). \tag{3-67}$$

Lemma 3.18. *There exists a constant $C = C(n, \boldsymbol\beta) > 0$ such that, for all $k \le \min(k_p, \ell)$ and $z \in \frac14 \hat B_k(p)\setminus\mathcal{S}$, the following pointwise estimates hold*:

$$\left|\frac{\partial^2 D'h_k}{r_1^2\partial\theta_1^2}(z)\right| + \left|\frac{\partial^2 D'h_k}{r_1\partial r_1\partial\theta_1}(z)\right| \le C r_1(z)^{1/\beta_1-2}\tau^{-k(1/\beta_1-1)}\omega(\tau^k), \tag{3-68}$$

$$\left|\frac{\partial^2 D'h_k}{\partial r_1^2}(z)\right| \le C r_1(z)^{1/\beta_1-2}\tau^{-k(1/\beta_1-1)}\omega(\tau^k). \tag{3-69}$$

Proof. Applying the gradient estimate to the $g_{\boldsymbol\beta}$-harmonic function $D'h_k$, we get

$$\left\|\frac{\partial D'h_k}{r_1\partial\theta_1}\right\|_{L^\infty(\hat B_k(p)/2)} \le \|\nabla_{g_{\boldsymbol\beta}} D'h_k\|_{L^\infty(\hat B_k(p)/2)} \le C\omega(\tau^k).$$

The function $\partial_{\theta_1} D'h_k$ is also a continuous $g_{\boldsymbol\beta}$-harmonic function, so the derivatives estimates implies, on $\frac13 \hat B_k(p)\setminus\mathcal{S}$,

$$|F_1| \le \left||z_2|^{2(1-\beta_2)}\frac{\partial^2(\partial_{\theta_1} D'h_k)}{\partial z_2\partial\bar z_2}\right| + \left|\frac{\partial^2(\partial_{\theta_1} D'h_k)}{\partial s_j^2}\right| \le C\tau^{-k}\omega(\tau^k),$$

where F_1 is defined as

$$|z_1|^{2(1-\beta_1)}\frac{\partial^2(\partial_{\theta_1}D'h_k)}{\partial z_1\partial\bar z_1}=-|z_2|^{2(1-\beta_2)}\frac{\partial^2(\partial_{\theta_1}D'h_k)}{\partial z_2\partial\bar z_2}-\sum_{j=5}^{2n}\frac{\partial^2(\partial_{\theta_1}D'h_k)}{\partial s_j^2}=:F_1. \tag{3-70}$$

We apply similar arguments as in the proof of Lemma 3.16. For any $x\in\mathcal{S}_1\cap\frac{1}{4}\hat B_k(p)$, we view (3-70) as defined on the $\mathbb{C}$-ball $B_{\mathbb{C}}(x,(\tau^k)^{1/\beta_1})$, and by the estimate (2-3) we have, on $B_{\mathbb{C}}(x,(\tau^k)^{1/\beta_1}/2)\backslash\{x\}$,

$$\left|\frac{\partial(\partial_{\theta_1}D'h_k)}{\partial z_1}\right|\le C\frac{\|\partial_{\theta_1}D'h_k\|_{L^\infty(\hat B_{\mathbb{C}})}}{(\tau^k)^{1/\beta_1}}+C\|F_1\|_{L^\infty(\hat B_{\mathbb{C}})}(\tau^k)^{2-1/\beta_1}.$$

Equivalently, this means that, on $B_{\mathbb{C}}(x,(\tau^k)^{1/\beta_1}/2)\backslash\{x\}$,

$$\left|\frac{\partial^2D'h_k}{\partial r_1\partial\theta_1}\right|+\left|\frac{\partial^2D'h_k}{r_1\partial\theta_1^2}\right|\le r_1^{1/\beta_1-1}\left|\frac{\partial(\partial_{\theta_1}D'h_k)}{\partial z_1}\right|\le Cr_1^{1/\beta_1-1}\tau^{k(1-1/\beta_1)}\omega(\tau^k).$$

Again by the inclusion (3-66), we get (3-68). The estimate (3-69) follows from Lemma 3.16, (3-68), (3-64) and the equation (from (3-63))

$$\frac{\partial^2D'h_k}{\partial r_1^2}=\frac{1}{r_1}\frac{\partial D'h_k}{\partial r_1}-\frac{1}{\beta_1^2r_1^2}\frac{\partial^2D'h_k}{\partial\theta_1^2}+F. \qquad\square$$

Lemma 3.19. *There exists a constant $C(n,\boldsymbol\beta)>0$ such that, for $k\le\min(k_{2,p},\ell)$, the following pointwise estimates hold for any $z\in\frac{1}{4}\hat B_k(p)\backslash\mathcal{S}$:*

$$\left|\frac{\partial}{\partial r_2}\left(\frac{\partial D'h_k}{\partial r_1}\right)(z)\right|+\left|\frac{\partial}{r_2\partial\theta_2}\left(\frac{\partial D'h_k}{\partial r_1}\right)(z)\right|\le C(n,\boldsymbol\beta)r_1^{1/\beta_1-1}r_2^{1/\beta_2-1}\tau^{-k(-1+1/\beta_1+1/\beta_2)}\omega(\tau^k). \tag{3-71}$$

Proof. By the Laplacian estimate in (3-9) for the harmonic function $D'h_k$ on $\frac{1}{2}\hat B_k(p)$, we have

$$\sup_{\hat B_k(p)/2.2}(|\Delta_1D'h_k|+|\Delta_2D'h_k|)\le C(n)\tau^{-2k}\operatorname{osc}_{\hat B_k(p)/2}(D'h_k)\le C(n)\tau^{-k}\omega(\tau^k). \tag{3-72}$$

Since $\Delta_1(D'h_k)$ is also $g_{\boldsymbol\beta}$-harmonic, the Laplacian estimates (3-9) imply

$$\sup_{\hat B_k(p)/2.4}(|\Delta_1\Delta_1D'h_k|+|\Delta_2\Delta_1D'h_k|)\le C(n)\tau^{-2k}\operatorname{osc}_{\hat B_k(p)/2.2}\Delta_1D'h_k\le C\tau^{-3k}\omega(\tau^k). \tag{3-73}$$

Now from the equation $\Delta_{\boldsymbol\beta}(\Delta_1D'h_k)=0$, we get

$$|z_1|^{2(1-\beta_1)}\frac{\partial^2}{\partial z_1\partial\bar z_1}\Delta_1D'h_k=-\Delta_2\Delta_1D'h_k-\sum_j\frac{\partial^2}{\partial s_j^2}\Delta_1D'h_k=:F_2. \tag{3-74}$$

From (3-73) and the Laplacian estimates (3-9), we see that $\sup_{\hat B_k(p)/2.4}|F_2|\le C\tau^{-3k}\omega(\tau^k)$. Using similar arguments, by considering $x\in\frac{1}{3}\hat B_k(p)\cap\mathcal{S}_1$, we obtain from (3-74) that, on $\hat B:=B_{\mathbb{C}}(x,(\tau^k)^{1/\beta_1}/2)\backslash\{x\}$,

$$\left|\frac{\partial}{\partial z_1}\Delta_1D'h_k\right|\le C\frac{\|\Delta_1D'h_k\|_{L^\infty(\hat B)}}{(\tau^k)^{1/\beta_1}}+C\|F_2\|_{L^\infty(\hat B)}(\tau^k)^{2-1/\beta_1}\le C\tau^{-k(1+1/\beta_1)}\omega(\tau^k).$$

This implies that, for any $z \in \frac{1}{3}\hat{B}_k(p)\backslash\mathcal{S}$,

$$\left|\frac{\partial}{\partial r_1}\Delta_1 D'h_k(z)\right| + \left|\frac{\partial}{r_1\partial\theta_1}\Delta_1 D'h_k(z)\right| \le Cr_1^{1/\beta_1-1}\tau^{-k(1+1/\beta_1)}\omega(\tau^k). \tag{3-75}$$

Now taking $\partial/\partial r_1$ on both sides of $\Delta_{\boldsymbol\beta} D'h_k = 0$, we get

$$|z_2|^{2(1-\beta_2)}\frac{\partial^2}{\partial z_2\partial\bar{z}_2}\left(\frac{\partial D'h_k}{\partial r_1}\right) = -\frac{\partial}{\partial r_1}(\Delta_1 D'h_k) - \sum_j \frac{\partial^2}{\partial s_j^2}\left(\frac{\partial D'h_k}{\partial r_1}\right) =: F_3. \tag{3-76}$$

From (3-75), for any $z \in \frac{1}{3}\hat{B}_k\backslash\mathcal{S}$, we have $|F_3|(z) \le Cr_1^{1/\beta_1-1}\tau^{-k(1+1/\beta_1)}\omega(\tau^k)$. By a similar argument, for any $y \in \frac{1}{3.2}\hat{B}_k(p)\cap\mathcal{S}_2$, we apply estimate (2-3) to $\partial D'h_k/\partial r_1$ and get, on $A_1 := B_{\mathbb{C}}\big(y, \frac{1}{2}(\tau^k)^{1/\beta_2}\big)\backslash\{y\}$ — the punctured ball in the complex plane $\mathbb{C}$ of (Euclidean) radius $\frac{1}{2}(\tau^k)^{1/\beta_2}$ and orthogonal to $\mathcal{S}_2$ passing through y — that

$$\begin{aligned}\left|\frac{\partial}{\partial z_2}\left(\frac{\partial D'h_k}{\partial r_1}\right)(z)\right| &\le C\frac{\left\|\frac{\partial D'h_k}{\partial r_1}\right\|_{L^\infty(A_1)}}{(\tau^k)^{1/\beta_2}} + C\|F_3\|_{L^\infty(A_1)}(\tau^k)^{2-1/\beta_2} \\ &\le Cr_1^{1/\beta_1-1}\tau^{-k(1/\beta_1+1/\beta_2-1)}\omega(\tau^k).\end{aligned} \tag{3-77}$$

Varying $y \in \frac{1}{3.2}\hat{B}_k(p)\cap\mathcal{S}_2$ we get, for any $z \in \frac{1}{4}\hat{B}_k\backslash\mathcal{S}$, that the following pointwise estimate holds:

$$\left|\frac{\partial}{\partial r_2}\left(\frac{\partial D'h_k}{\partial r_1}\right)(z)\right| + \left|\frac{\partial}{r_2\partial\theta_2}\left(\frac{\partial D'h_k}{\partial r_1}\right)(z)\right| \le Cr_1^{1/\beta_1-1}r_2^{1/\beta_2-1}\tau^{-k(1/\beta_1+1/\beta_2-1)}\omega(\tau^k). \qquad \square$$

Lemma 3.20. *Let $d = d_{\boldsymbol\beta}(p,q)$. There exists a constant $C(n,\boldsymbol\beta)$ such that, for all $k \le \ell$,*

$$\left|\frac{\partial D'h_k}{\partial r_1}(p) - \frac{\partial D'h_k}{\partial r_1}(q)\right| \le Cd^{1/\beta_1-1}\tau^{-k(1/\beta_1-1)}\omega(\tau^k), \tag{3-78}$$

$$\left|\frac{\partial D'h_k}{r_1\partial\theta_1}(p) - \frac{\partial D'h_k}{r_1\partial\theta_1}(q)\right| \le Cd^{1/\beta_1-1}\tau^{-k(1/\beta_1-1)}\omega(\tau^k). \tag{3-79}$$

Proof. We will consider the different cases $r_p = \min(r_p, r_q) \le 2d$ and $r_p = \min(r_p, r_q) > 2d$.

Case 1: $r_p \le 2d$. In this case, it is clear by the choice of ℓ that $r_p \approx \tau^{k_p} \le 2d \le \tau^{\ell+2}$, so $k_p \ge \ell+2$.

From our assumption when solving (3-45), $r_p = d_{\boldsymbol\beta}(p,\mathcal{S}_1)$, i.e., $r_1(p) = r_p \le 2d$. By the triangle inequality we have $r_1(q) \le 3d$. We also remark that, for $k \le \ell$, we have $\tau^k \ge \tau^\ell > 8d$. In particular, the geodesics considered below all lie inside the balls $\frac{1}{4}\hat{B}_k(p)$, and the estimates in Lemmas 3.16–3.19 hold for points on these geodesics.

Let the coordinates of the points p and q be given by

$$p = (r_1(p), \theta_1(p); r_2(p), \theta_2(p); s(p)) \quad\text{and}\quad q = (r_1(q), \theta_1(q); r_2(q), \theta_2(q); s(q)).$$

Let $\gamma : [0,d] \to B_{\boldsymbol\beta}(0,q)\backslash\mathcal{S}$ be the unique $g_{\boldsymbol\beta}$-geodesic connecting p and q. We know the curve γ is disjoint from $\mathcal{S}$, and we write

$$\gamma(t) = (r_1(t), \theta_1(t); r_2(t), \theta_2(t); s(t))$$

for the coordinates of $\gamma(t)$ for $t \in [0,d]$. By definition we have, for all $t \in [0,d]$,

$$|\gamma'(t)|^2_{g_{\boldsymbol\beta}} = (r_1'(t))^2 + \beta_1^2 r_1(t)^2(\theta_1'(t))^2 + (r_2'(t))^2 + \beta_2^2 r_2(t)^2(\theta_2'(t))^2 + |s'(t)|^2 = 1.$$

So $|s(p)-s(q)|\le d$ and $|r_i(p)-r_i(q)|\le d$ for $i=1,2$. We define

$$q' := (r_1(q),\theta_1(q);r_2(p),\theta_2(p);s(p)) \quad \text{and} \quad p' := (r_1(p),\theta_1(q);r_2(p),\theta_2(p);s(p)), \tag{3-80}$$

the points with coordinates related to p and q. Let γ_1 be the $g_{\boldsymbol\beta}$-geodesic connecting q and q', γ_2 be the $g_{\boldsymbol\beta}$-geodesic joining q' to p', and γ_3 be the $g_{\boldsymbol\beta}$-geodesic joining p' to p.

By the triangle inequality, we have

$$\begin{aligned}
&\left|\frac{\partial D'h_k}{\partial r_1}(p)-\frac{\partial D'h_k}{\partial r_1}(q)\right|\\
&\qquad\le \left|\frac{\partial D'h_k}{\partial r_1}(p)-\frac{\partial D'h_k}{\partial r_1}(p')\right|+\left|\frac{\partial D'h_k}{\partial r_1}(p')-\frac{\partial D'h_k}{\partial r_1}(q')\right|+\left|\frac{\partial D'h_k}{\partial r_1}(q')-\frac{\partial D'h_k}{\partial r_1}(q)\right|\\
&\qquad=: J_1'+J_2'+J_3'.
\end{aligned}$$

Integrating along γ_3, on which the points have fixed r_1-coordinate $r_1(p)$, we get (by (3-68))

$$J_1'=\left|\int_{\gamma_3}\frac{\partial}{\partial\theta_1}\left(\frac{\partial D'h_k}{\partial r_1}\right)d\theta_1\right|\le C(n,\boldsymbol\beta)r_1(p)^{1/\beta_1-1}\tau^{-k(1/\beta_1-1)}\omega(\tau^k). \tag{3-81}$$

Integrating along γ_2, we get (by (3-69))

$$\begin{aligned}
J_2'&=\left|\int_{\gamma_2}\frac{\partial}{\partial r_1}\left(\frac{\partial D'h_k}{\partial r_1}\right)dr_1\right|\le C(n,\boldsymbol\beta)\tau^{-k(1/\beta_1-1)}\omega(\tau^k)\left|\int_{r_1(p)}^{r_1(q)}t^{1/\beta_1-2}\,dt\right|\\
&=C(n,\boldsymbol\beta)\tau^{-k(1/\beta_1-1)}\omega(\tau^k)|r_1(p)^{1/\beta_1-1}-r_1(q)^{1/\beta_1-1}|\\
&\le C(n,\boldsymbol\beta)\tau^{-k(1/\beta_1-1)}\omega(\tau^k)|r_1(p)-r_1(q)|^{1/\beta_1-1}\\
&\le C(n,\boldsymbol\beta)\tau^{-k(1/\beta_1-1)}\omega(\tau^k)d^{1/\beta_1-1}.
\end{aligned}\tag{3-82}$$

To deal with J_3', we need to consider different choices of $k\le\ell$.

<u>Case 1a</u>: $k_{2,p}+1\le k\le\ell$. In this case, the balls $\hat B_k(p)$ are centered at $p_1\in\mathcal S_1$ (recall p_1 is the projection of p to $\mathcal S_1$; hence p and p_1 have the same $(r_2,\theta_2;s)$-coordinates). We have $\tau^{-k}\le 8^{-1}d^{-1}$ by the choice of ℓ. The balls $\hat B_k(p)$ are disjoint from $\mathcal S_2$, so we can introduce the smooth coordinates $w_2=z_2^{\beta_2}$, and under the coordinates $(r_1,\theta_1;w_2,z_3,\dots,z_n)$, the metric $g_{\boldsymbol\beta}$ becomes the smooth cone metric with conical singularity *only* along $\mathcal S_1$ with angle $2\pi\beta_1$. Therefore we can derive the following estimate as in (3-62):

$$\sup_{(\hat B_k(p)/2)\setminus\mathcal S_1}\left|\frac{\partial(D')^2h_k}{\partial r_1}\right|+\left|\frac{\partial}{\partial r_1}\left(\frac{\partial D'h_k}{\partial w_2}\right)\right|\le C\tau^{-k}\omega(\tau^k). \tag{3-83}$$

Since q and q' have the same (r_1,θ_1)-coordinates and $g_{\boldsymbol\beta}$ is a product metric, γ_1 is in fact a straight line segment (under the coordinates $(w_2,z_3,\dots,z_n)$) in the hyperplane with fixed (r_1,θ_1)-coordinates. Integrating over γ_1, we get

$$\begin{aligned}
J_3'\le\int_{\gamma_1}\left|\frac{\partial}{\partial w_2}\left(\frac{\partial D'h_k}{\partial r_1}\right)\right|+\sum_j\left|\frac{\partial}{\partial s_j}\left(\frac{\partial D'h_k}{\partial r_1}\right)\right|&\le C\tau^{-k}\omega(\tau^k)d_{\boldsymbol\beta}(q,q')\le C\tau^{-k}\omega(\tau^k)d\\
&\le C(n,\boldsymbol\beta)\tau^{-k(1/\beta_1-1)}\omega(\tau^k)d^{1/\beta_1-1}.
\end{aligned}$$

Case 1b: $k \le k_{2,p}$. In this case, the balls $\hat{B}_k(p)$ are centered at $p_{1,2} \in \mathcal{S}_1 \cap \mathcal{S}_2$ and $\tau^k \ge r_2(p)$. By the triangle inequality, $r_2(q) \le d + r_2(p) \le \frac{9}{8}\tau^k$. We choose the points

$$\tilde{q} = (r_1(q), \theta_1(q); r_2(p), \theta_2(p); s(q)) \quad \text{and} \quad \hat{q} = (r_1(q), \theta_1(q); r_2(q), \theta_2(p); s(q)). \tag{3-84}$$

Let $\tilde{\gamma}_1$ be the $g_{\boldsymbol{\beta}}$-geodesic joining q' to $\tilde{q}$, $\tilde{\gamma}$ the $g_{\boldsymbol{\beta}}$-geodesic joining $\tilde{q}$ to $\hat{q}$, and $\hat{\gamma}$ the $g_{\boldsymbol{\beta}}$-geodesic joining $\hat{q}$ to q. The curves $\tilde{\gamma}_1$, $\tilde{\gamma}$ and $\hat{\gamma}$ all lie in the hyperplane with constant (r_1, θ_1)-coordinates $(r_1(q), \theta_1(q))$. Then by the triangle inequality we have

$$J_3' \le \left| \frac{\partial D'h_k}{\partial r_1}(q') - \frac{\partial D'h_k}{\partial r_1}(\tilde{q}) \right| + \left| \frac{\partial D'h_k}{\partial r_1}(\tilde{q}) - \frac{\partial D'h_k}{\partial r_1}(\hat{q}) \right| + \left| \frac{\partial D'h_k}{\partial r_1}(\hat{q}) - \frac{\partial D'h_k}{\partial r_1}(q) \right| =: J_1'' + J_2'' + J_3''.$$

We will use frequently the inequalities $r_1(q) \le 3d$ and $\max(r_2(q), r_2(p)) \le 2\tau^k$ in the estimates below. Integrating along $\hat{\gamma}$ we get (by (3-71))

$$\begin{aligned} J_3'' &\le \left| \int_{\hat{\gamma}} \frac{\partial}{\partial \theta_2}\left(\frac{\partial D'h_k}{\partial r_1} \right) d\theta_2 \right| \le C r_1(q)^{1/\beta_1 - 1} r_2(q)^{1/\beta_2} \tau^{-k(-1+1/\beta_1+1/\beta_2)} \omega(\tau^k) \\ &\le C d^{1/\beta_1 - 1} \tau^{-k(1/\beta_1 - 1)} \omega(\tau^k). \end{aligned}$$

Integrating along $\tilde{\gamma}$ we get (again by (3-71))

$$\begin{aligned} J_2'' &\le \left| \int_{\tilde{\gamma}} \frac{\partial}{\partial r_2}\left(\frac{\partial D'h_k}{\partial r_1} \right) dr_2 \right| \le C r_1(q)^{1/\beta_1 - 1} \tau^{-k(-1+1/\beta_1+1/\beta_2)} \omega(\tau^k) \left| \int_{r_2(q)}^{r_2(p)} t^{1/\beta_2 - 1}\, dt \right| \\ &\le C r_1(q)^{1/\beta_1 - 1} \tau^{-k(-1+1/\beta_1+1/\beta_2)} \omega(\tau^k) \max(r_2(q), r_2(p))^{1/\beta_2 - 1} d \\ &\le C d^{1/\beta_1 - 1} \tau^{-k(1/\beta_1 - 1)} \omega(\tau^k). \end{aligned}$$

Integrating along $\tilde{\gamma}_1$ we get (by (3-67))

$$J_1'' \le \left| \int_{\tilde{\gamma}_1} \frac{\partial}{\partial s_j}\left(\frac{\partial D'h_k}{\partial r_1} \right) dt \right| \le C r_1(q)^{1/\beta_1 - 1} \tau^{-k/\beta_1} \omega(\tau^k) d \le C d^{1/\beta_1 - 1} \tau^{-k(1/\beta_1 - 1)} \omega(\tau^k).$$

Combining the three inequalities above, we get, in the case $k \le k_{2,p}$,

$$J_3' \le C d^{1/\beta_1 - 1} \tau^{-k(1/\beta_1 - 1)} \omega(\tau^k).$$

Combining the estimates on J_1', J_2' and J_3', we finish the proof of (3-78) in the case $r_p \le 2d$.

Case 2a: $r_p > 2d$ and $\ell \le k_p$. In this case $\tau^{k_p} \approx r_p > 2d \ge \tau^{\ell+3}$. From the triangle inequality we get $d_{\boldsymbol{\beta}}(\gamma(t), \mathcal{S}) \ge d$. In particular, the r_1 and r_2 coordinates of $\gamma(t)$ are both bigger than d. In this case $k \le \ell \le k_p$, and Lemmas 3.16–3.19 hold for the points in γ. So $r_1(\gamma(t)) \le r_1(p) + d \le 2\tau^k$. We calculate the gradient of $\partial D'h_k / \partial r_1$ along γ:

$$\begin{aligned} \left| \nabla_{g_{\boldsymbol{\beta}}} \frac{\partial D'h_k}{\partial r_1} \right|^2 = \left| \frac{\partial}{\partial r_1}\left(\frac{\partial D'h_k}{\partial r_1} \right) \right|^2 + \left| \frac{\partial}{\beta_1 r_1 \partial \theta_1}\left(\frac{\partial D'h_k}{\partial r_1} \right) \right|^2 + \left| \frac{\partial}{\partial r_2}\left(\frac{\partial D'h_k}{\partial r_1} \right) \right|^2 \\ + \left| \frac{\partial}{\beta_2 r_2 \partial \theta_2}\left(\frac{\partial D'h_k}{\partial r_1} \right) \right|^2 + \sum_j \left| \frac{\partial}{\partial s_j}\left(\frac{\partial D'h_k}{\partial r_1} \right) \right|^2. \end{aligned}$$

(1) When $k_{2,p}+1 \le k \le \ell$ we have by (3-83) that

$$\sup_{(\hat{B}_k/2)\setminus \mathcal{S}_1} \left|\frac{\partial}{\partial r_2}\left(\frac{\partial D' h_k}{\partial r_1}\right)\right| + \left|\frac{\partial}{r_2 \partial \theta_2}\left(\frac{\partial D' h_k}{\partial r_1}\right)\right| \le C\tau^{-k}\omega(\tau^k). \tag{3-85}$$

Thus by Lemma 3.18, (3-67) and (3-85), along γ we have

$$\left|\nabla_{g_\beta} \frac{\partial D' h_k}{\partial r_1}\right| \le C\omega(\tau^k)(d^{1/\beta_1 - 2}\tau^{-k(1/\beta_1 - 1)} + \tau^{-k})$$

Integrating along γ we get

$$\begin{aligned}\left|\frac{\partial D' h_k}{\partial r_1}(p) - \frac{\partial D' h_k}{\partial r_1}(q)\right| \le \int_\gamma \left|\nabla_{g_\beta} \frac{\partial D' h_k}{\partial r_1}\right| &\le C\omega(\tau^k)(d^{1/\beta_1 - 1}\tau^{-k(1/\beta_1 - 1)} + d\tau^{-k}) \\ &\le C\omega(\tau^k) d^{1/\beta_1 - 1}\tau^{-k(1/\beta_1 - 1)}.\end{aligned}$$

(2) When $k \le k_{2,p}$, we have $r_2(\gamma(t)) \le r_2(p) + d \le \tau^k + d \le \frac{9}{8}\tau^k$ and similar estimates hold for $r_1(\gamma(t))$ too. Then by Lemma 3.18, Lemma 3.19 and (3-67) along γ the following estimate holds

$$\left|\nabla_{g_\beta} \frac{\partial D' h_k}{\partial r_1}\right|(\gamma(t)) \le C\omega(\tau^k)(d^{1/\beta_1 - 2}\tau^{-k(1/\beta_1 - 1)} + \tau^{-k})$$

Integrating along γ we get

$$\begin{aligned}\left|\frac{\partial D' h_k}{\partial r_1}(p) - \frac{\partial D' h_k}{\partial r_1}(q)\right| \le \int_\gamma \left|\nabla_{g_\beta} \frac{\partial D' h_k}{\partial r_1}\right| &\le C\omega(\tau^k)(d^{1/\beta_1 - 1}\tau^{-k(1/\beta_1 - 1)} + d\tau^{-k}) \\ &\le C\omega(\tau^k) d^{1/\beta_1 - 1}\tau^{-k(1/\beta_1 - 1)}.\end{aligned}$$

This finishes the proof of the lemma in this case.

Case 2b: $r_p > 2d$ but $\ell \ge k_p + 1$. When $k \le k_p$, the estimate (3-78) follows in the same way as the case above. Hence it suffices to consider the case when $k_p + 1 \le k \le \ell$. In this case the balls $\hat{B}_k(p) = B_{\boldsymbol{\beta}}(p, \tau^k)$ and it can be seen by triangle inequality that the geodesic $\gamma \subset \frac{1}{3}\hat{B}_k(p)\setminus \mathcal{S}$. Since the metric balls $\hat{B}_k(p)$ are disjoint with $\mathcal{S}$ we can use the smooth coordinates $w_1 = z_1^{\beta_1}$ and $w_2 = z_2^{\beta_2}$ as before, and everything becomes smooth under these coordinates in $\hat{B}_k(p)$.

The estimate (3-79) can be shown by the same argument, so we skip the details. □

Iteratively applying (3-78) for $k \le \ell$, we get

$$\begin{aligned} J_2 = \left|\frac{\partial D' u_\ell}{\partial r_1}(p) - \frac{\partial D' u_\ell}{\partial r_1}(q)\right| &\le \left|\frac{\partial D' u_2}{\partial r_1}(p) - \frac{\partial D' u_2}{\partial r_1}(q)\right| + C d^{1/\beta_1 - 1}\sum_{k=3}^{\ell} \tau^{-k(1/\beta_1 - 1)}\omega(\tau^k) \\ &\le C d^{1/\beta_1 - 1}\left(\|u\|_{C^0} + \sum_{k=2}^{\ell}\tau^{-k(1/\beta_1 - 1)}\omega(\tau^k)\right),\end{aligned}$$

where the inequality

$$\left|\frac{\partial D' u_2}{\partial r_1}(p) - \frac{\partial D' u_2}{\partial r_1}(q)\right| \le C d^{1/\beta_1 - 1}\|u\|_{C^0}$$

can be proved by the same argument as in proving (3-78).

Combining the estimates for $J_1,\ J_2,\ J_3,\ J_4$ we finish the proof of (1-4).

We remark that in solving (3-45) we assume $r_1(p) \le r_2(p)$, so we need also to deal with the following case, whose proof is more or less parallel to that of Lemma 3.20, so we just point out the differences and sketch the proof.

Lemma 3.21. *Let $d = d_{\boldsymbol\beta}(p, q) > 0$. There exists a constant $C(n, \boldsymbol\beta) > 0$ such that, for all $k \le \ell$,*

$$\left|\frac{\partial D'h_k}{\partial r_2}(p) - \frac{\partial D'h_k}{\partial r_2}(q)\right| \le C d^{1/\beta_2-1}\tau^{-k(1/\beta_2-1)}\omega(\tau^k), \tag{3-86}$$

$$\left|\frac{\partial D'h_k}{r_2\partial \theta_2}(p) - \frac{\partial D'h_k}{r_2\partial \theta_2}(q)\right| \le C d^{1/\beta_2-1}\tau^{-k(1/\beta_2-1)}\omega(\tau^k). \tag{3-87}$$

Proof. We consider the cases when $k \le k_{1,p}$ and $k_{1,p}+1 \le k \le \ell$.

<u>Case 1</u>: $k_{1,p}+1 \le k \le \ell$. The balls $\hat B_k(p)$ are disjoint with $\mathcal{S}_2$, so we can introduce the complex coordinate $w_2 = z_2^{\beta_2}$ on these balls as before. Let t_1 and t_2 be the real and imaginary parts of w_2, respectively. The derivatives estimates imply that

$$\|\partial_{w_2} D'h_k\|_{L^\infty(\hat B_k(p)/2)} \le C\omega(\tau^k) \quad \text{and} \quad \|\partial^2_{w_2} D'h_k\|_{L^\infty(\hat B_k(p)/2)} \le C\tau^{-k}\omega(\tau^k),$$

where $\partial^2_{w_2}$ denotes the full second-order derivatives in the $\{t_1, t_2\}$-directions. Also

$$\left\|\frac{\partial}{\partial r_1}\left(\frac{\partial D'h_k}{\partial w_2}\right)\right\|_{L^\infty(\hat B_k(p)/2)} + \left\|\frac{\partial}{r_1\partial \theta_1}\left(\frac{\partial D'h_k}{\partial w_2}\right)\right\|_{L^\infty(\hat B_k(p)/2)} \le C\tau^{-k}\omega(\tau^k).$$

Since

$$\frac{\partial}{\partial r_2} = \frac{w_2}{\beta_2 r_2}\frac{\partial}{\partial w_2} + \frac{\bar w_2}{\beta_2 r_2}\frac{\partial}{\partial \bar w_2}, \tag{3-88}$$

we have

$$\frac{\partial}{\partial w_2}\left(\frac{\partial D'h_k}{\partial r_2}\right) = \frac{1}{r_2}\frac{\partial D'h_k}{\partial w_2} - \frac{|w_2|^2}{2r_2^3}\frac{\partial D'h_k}{\partial w_2} - \frac{\bar w_2\cdot\bar w_2}{2r_2^3}\frac{\partial D'h_k}{\partial \bar w_2} + \frac{w_2}{r_2}\partial^2_{w_2} D'h_k,$$

and we have, on $\frac12 \hat B_k(p)$,

$$\left|\frac{\partial}{\partial w_2}\left(\frac{\partial D'h_k}{\partial r_2}\right)\right| \le \frac{C}{r_2}\omega(\tau^k) + C\tau^{-k}\omega(\tau^k)$$

and

$$\left\|\frac{\partial}{\partial r_1}\left(\frac{\partial D'h_k}{\partial r_2}\right)\right\|_{L^\infty(\hat B_k(p)/2)} + \left\|\frac{\partial}{r_1\partial \theta_1}\left(\frac{\partial D'h_k}{\partial r_2}\right)\right\|_{L^\infty(\hat B_k(p)/2)} \le C\tau^{-k}\omega(\tau^k).$$

Therefore,

$$\left|\nabla_{g_{\boldsymbol\beta}}\frac{\partial D'h_k}{\partial r_2}\right|^2 = \left|\frac{\partial^2 D'h_k}{\partial r_1\partial r_2}\right|^2 + \left|\frac{\partial^2}{r_1\partial\theta_1\partial r_2}\right|^2 + \left|\frac{\partial^2 D'h_k}{\partial w_2\partial r_2}\right|^2 + \sum_j\left|\frac{\partial^2 D'h_k}{\partial s_j\partial r_2}\right|^2 \le C(\tau^{-k}\omega(\tau^k))^2 + C\frac{1}{r_2^2}\omega(\tau^k)^2.$$

In this case we know that $r_1(p) \approx \tau^{k_p} \ge 2\tau^k \ge \tau^\ell > 8d$, so along γ

$$r_2(\gamma(t)) \ge r_2(p) - d \ge r_1(p) - d \ge \tfrac74\tau^k.$$

Integrating along γ we get

$$\left|\frac{\partial D'h_k}{\partial r_2}(p)-\frac{\partial D'h_k}{\partial r_2}(q)\right|\le\int_\gamma\left|\nabla_{g_\beta}\frac{\partial D'h_k}{\partial r_2}\right|\le C\tau^{-k}\omega(\tau^k)d\le C\tau^{-k(1/\beta_2-1)}\omega(\tau^k)d^{1/\beta_2-1}.$$

Case 2: $k\le k_{1,p}$. This case is the same as in the proof of (3-78), replacing r_1 by r_2 and β_1 by β_2. We omit the details.

We can prove (3-87) similarly. □

3D. ***Mixed normal directions.*** In this section, we deal with Hölder continuity of the four mixed derivatives

$$\frac{\partial^2 u}{\partial r_1\partial r_2},\quad \frac{\partial^2 u}{r_1\partial\theta_1\partial r_2},\quad \frac{\partial^2 u}{r_2\partial r_1\partial\theta_2},\quad \frac{\partial^2 u}{r_1r_2\partial\theta_1\partial\theta_2},\tag{3-89}$$

which by our previous notation correspond to N_1N_2u. Since the proof for each of them is more or less the same, we will only prove Hölder continuity for $\partial^2u/\partial r_1\partial r_2$. The following holds at p and q by the same reasoning of Lemma 3.11:

$$\lim_{k\to\infty}\frac{\partial^2 u_k}{\partial r_1\partial r_2}(p)=\frac{\partial^2 u}{\partial r_1\partial r_2}(p),\qquad \lim_{k\to\infty}\frac{\partial^2 v_k}{\partial r_1\partial r_2}(q)=\frac{\partial^2 u}{\partial r_1\partial r_2}(q).$$

By the triangle inequality,

$$\begin{aligned}\left|\frac{\partial^2 u}{\partial r_1\partial r_2}(p)-\frac{\partial^2 u}{\partial r_1\partial r_2}(q)\right|&\le\left|\frac{\partial^2 u}{\partial r_1\partial r_2}(p)-\frac{\partial^2 u_\ell}{\partial r_1\partial r_2}(p)\right|+\left|\frac{\partial^2 u_\ell}{\partial r_1\partial r_2}(p)-\frac{\partial^2 u_\ell}{\partial r_1\partial r_2}(q)\right|\\&\qquad+\left|\frac{\partial^2 u_\ell}{\partial r_1\partial r_2}(q)-\frac{\partial^2 v_\ell}{\partial r_1\partial r_2}(q)\right|+\left|\frac{\partial^2 v_\ell}{\partial r_1\partial r_2}(q)-\frac{\partial^2 u}{\partial r_1\partial r_2}(q)\right|\\&=:L_1+L_2+L_3+L_4.\end{aligned}$$

Lemma 3.22. *We have the estimate*

$$L_1+L_4\le\sum_{k=\ell}^\infty\omega(\tau^k).$$

Proof. We consider the cases when $k\ge k_p+1$ and $\ell\le k\le k_p$.

Case 1: $k\ge k_p+1$. In this case the balls $\hat B_k(p)$ are disjoint from $\mathcal{S}$ and we can introduce the smooth coordinates $w_1=z_1^{\beta_1}$ and $w_2=z_2^{\beta_2}$. Under the coordinates $\{w_1,w_2,z_3,\dots,z_n\}$, the cone metric $g_{\boldsymbol\beta}$ becomes the standard Euclidean metric $g_{\mathbb{C}^n}$ and the metric balls $\hat B_k(p)$ become the standard Euclidean balls with the same radius and center p. Since the $g_{\boldsymbol\beta}$-harmonic functions u_k-u_{k+1} satisfy (3-49), by standard gradient estimates for Euclidean harmonic functions, we get

$$\sup_{\hat B_k(p)/2.1}\left|D_{w_1}D_{w_2}(u_k-u_{k-1})\right|\le C\omega(\tau^k),$$

where we use D_{w_i} to denote either $\partial/\partial w_i$ or $\partial/\partial\bar w_i$ for simplicity. From (3-88) and a similar formula for $\partial/\partial r_1$, we get

$$\sup_{\hat B_k(p)/2.1}\left|\frac{\partial^2}{\partial r_1\partial r_2}(u_k-u_{k-1})\right|\le C\omega(\tau^k).\tag{3-90}$$

Case 2a: $\ell \geq k_{2,p}+1$ and $\ell \leq k_p = k_{1,p}$. For all $\ell \leq k$, the balls $\hat{B}_k(p)$ are disjoint from $\mathcal{S}_2$ and centered at p_1. We can still use $w_2 = z_2^{\beta_2}$ as the smooth coordinate. The cone metric $g_{\boldsymbol{\beta}}$ becomes smooth in the w_2-variable, and we can apply the standard gradient estimate to the $g_{\boldsymbol{\beta}}$-harmonic function $D_{w_2}(u_k - u_{k-1})$ to get

$$\sup_{\hat{B}_k(p)/2.2} \left|\frac{\partial}{\partial r_1} D_{w_2}(u_k - u_{k-1})\right| + \left|\frac{\partial}{r_1 \partial \theta_1} D_{w_2}(u_k - u_{k-1})\right| \leq C\omega(\tau^k).$$

Again by (3-88), we get

$$\sup_{\hat{B}_k(p)/2.2} \left|\frac{\partial^2}{\partial r_1 \partial r_2}(u_k - u_{k-1})\right| + \left|\frac{\partial^2}{r_1 \partial \theta_1 \partial r_2}(u_k - u_{k-1})\right| \leq C\omega(\tau^k). \tag{3-91}$$

Case 2b: $\ell \leq k_{2,p}$ and $k \geq k_{2,p}+1$. This case can be dealt with similarly as above.

Case 2c: $\ell \leq k \leq k_{2,p}$. In this case $r_2(p) \approx \tau^{k_{2,p}} \leq \tau^k \leq \tau^\ell \approx 8d$. Now the balls $\hat{B}_k(p)$ are centered at $p_{1,2} \in \mathcal{S}_1 \cap \mathcal{S}_2$. We can proceed as in the proof of Lemma 3.19, with the harmonic functions $u_k - u_{k-1}$ replacing the $D' h_k$ in that lemma to prove that, for any $z \in \frac{1}{3}\hat{B}_k(p) \setminus \mathcal{S}$,

$$\begin{aligned}\left|\frac{\partial^2}{\partial r_1 \partial r_2}(u_k - u_{k-1})\right|(z) &+ \left|\frac{\partial^2}{r_2 \partial \theta_2 \partial r_1}(u_k - u_{k-1})\right|(z) \\ &\leq C(n,\boldsymbol{\beta}) r_1(z)^{1/\beta_1 - 1} r_2(z)^{1/\beta_2 - 1} \tau^{-k(-2+1/\beta_1+1/\beta_2)} \omega(\tau^k).\end{aligned}$$

In particular, the estimate in each case holds at p, and from $r_1(p) \leq r_2(p) \leq \tau^k$ we obtain

$$\left|\frac{\partial^2}{\partial r_1 \partial r_2}(u_k - u_{k-1})\right|(p) + \left|\frac{\partial^2}{r_2 \partial \theta_2 \partial r_1}(u_k - u_{k-1})\right|(p) \leq C\omega(\tau^k). \tag{3-92}$$

Combining each case above, by (3-90)–(3-92), we get, for all $k \geq \ell$,

$$\left|\frac{\partial^2 u}{\partial r_1 \partial r_2}(u_k - u_{k-1})\right|(p) \leq C(n,\boldsymbol{\beta})\omega(\tau^k).$$

Therefore, by the triangle inequality,

$$L_1 \leq \sum_{k=\ell+1}^{\infty} \left|\frac{\partial^2 u}{\partial r_1 \partial r_2}(u_k - u_{k-1})\right|(p) \leq C(n,\boldsymbol{\beta}) \sum_{k=\ell+1}^{\infty} \omega(\tau^k).$$

The estimate for L_4 can be dealt with similarly by studying the derivatives of v_k at q. □

Lemma 3.23. $$L_3 \leq C(n,\boldsymbol{\beta})\omega(\tau^\ell).$$

Proof. As in the proof of Lemma 3.22, we consider the cases $\ell \geq k_{1,p}+1$, $k_{1,p} \geq \ell \geq k_{2,p}$ and $\ell \leq k_{2,p}-1$.

Case 1: $\ell \geq k_{1,p}+1$. Here the ball $\hat{B}_\ell(p)$ is equal to $B_{\boldsymbol{\beta}}(p,\tau^\ell)$, the function U defined in (3-55) is $g_{\boldsymbol{\beta}}$-harmonic in $\frac{1}{2}\hat{B}_\ell(p)$, and $\sup_{\hat{B}_\ell(p)/2}|U| \leq C\omega^{2\ell}\omega(\tau^\ell)$. Since the ball $\frac{1}{2}\hat{B}_\ell(p)$ is disjoint from $\mathcal{S}$, we have that w_1 and w_2 are well defined on $\frac{1}{2}\hat{B}_\ell(p)$, and thus we have the derivatives estimates

$$\sup_{\hat{B}_\ell(p)/3}\left|\frac{\partial^2 U}{\partial r_1 \partial r_2}\right| \leq \sup_{\hat{B}_\ell(p)/3}\left|D_{w_1} D_{w_2} U\right| \leq C(n,\boldsymbol{\beta})\omega(\tau^\ell).$$

In particular, at $q \in \frac{1}{3}\hat{B}_\ell(p)$,

$$L_3 = \left|\frac{\partial^2 u_\ell}{\partial r_1 \partial r_2}(q) - \frac{\partial^2 v_\ell}{\partial r_1 \partial r_2}(q)\right| = \left|\frac{\partial^2 U}{\partial r_1 \partial r_2}(q)\right| \le C(n, \boldsymbol{\beta})\omega(\tau^\ell).$$

<u>Case 2</u>: $k_{1,p} \ge \ell \ge k_{2,p}$. Here the ball $\hat{B}_\ell(p)$ is equal to $B_{\boldsymbol{\beta}}(p_1, 2\tau^\ell)$, the function U defined in (3-55) is $g_{\boldsymbol{\beta}}$-harmonic and well defined in a ball

$$B_q := B_{\boldsymbol{\beta}}\big(q, \tfrac{1}{10}\tau^\ell\big) \subset \tfrac{1}{2.2}\hat{B}_\ell(p),$$

and $\sup_{\hat{B}_\ell(p)/2} |U| \le C\omega^{2\ell}\omega(\tau^\ell)$. Since $\frac{1}{2}\hat{B}_\ell(p)$ is disjoint from $\mathcal{S}_2$, we have that w_2 is well defined on $\frac{1}{2.2}\hat{B}_\ell(p)$, and thus we have the derivatives estimates

$$\sup_{B_q/2}\left|\frac{\partial^2 U}{\partial r_1 \partial r_2}\right| \le \sup_{B_q/2}\left|\frac{\partial}{\partial r_1} D_{w_2} U\right| \le C(n, \boldsymbol{\beta})\omega(\tau^\ell).$$

In particular, at $q \in \frac{1}{2}B_q$, we have

$$L_3 = \left|\frac{\partial^2 u_\ell}{\partial r_1 \partial r_2}(q) - \frac{\partial^2 v_\ell}{\partial r_1 \partial r_2}(q)\right| = \left|\frac{\partial^2 U}{\partial r_1 \partial r_2}(q)\right| \le C(n, \boldsymbol{\beta})\omega(\tau^\ell).$$

<u>Case 3</u>: $\ell \le k_{2,p} - 1$. Here $r_2(p) \approx \tau^{k_{2,p}} \le \tau^{\ell+1} < 8d$, so

$$r_2(q) \le r_2(p) + d \le \tfrac{5}{8}\tau^\ell \quad \text{and} \quad r_1(q) \le d + r_1(p) \le d + r_2(p) \le \tfrac{5}{8}\tau^\ell.$$

Therefore the ball $\widetilde{B}_\ell(q)$ is centered at either q_1, q_2 or $q_{1,2} \in \mathcal{S}_1 \cap \mathcal{S}_2$, with radius $2\tau^\ell$. It follows that the function U defined in (3-55) is well defined on the ball $\frac{1}{1.8}\hat{B}_\ell(p)$.

By the same strategy as in the proof of Lemma 3.19, with the harmonic function $D'h_k$ in that lemma replaced by U on the metric ball $\frac{1}{1.8}\hat{B}_\ell(p)$, we can prove that, for any $z \in \frac{1}{2}\hat{B}_\ell(p)\setminus\mathcal{S}$,

$$\left|\frac{\partial^2 U}{\partial r_1 \partial r_2}(z)\right| \le C(n, \boldsymbol{\beta}) r_1^{1/\beta_1 - 1} r_2^{1/\beta_2 - 1} \tau^{-\ell(-2+1/\beta_1+1/\beta_2)}\omega(\tau^\ell).$$

Applying this inequality at q, we get

$$L_3 = \left|\frac{\partial^2 (u_\ell - v_\ell)}{\partial r_1 \partial r_2}(q)\right| \le C(n, \boldsymbol{\beta}) r_1(q)^{1/\beta_1 - 1} r_2(q)^{1/\beta_2 - 1} \tau^{-\ell(-2+1/\beta_1+1/\beta_2)}\omega(\tau^\ell) \le C(n, \boldsymbol{\beta})\omega(\tau^\ell).$$

In sum, in all cases $L_3 \le C(n, \boldsymbol{\beta})\omega(\tau^\ell)$. □

Lemma 3.24. *There exists a constant $C = C(n, \boldsymbol{\beta}) > 0$ such that, for all $k \le \ell$ and $z \in \frac{1}{3}\hat{B}_k(p)\setminus\mathcal{S}$,*

$$\left|\frac{\partial}{\partial \theta_1}\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right)(z)\right| + \left|\left(\frac{\partial^3 h_k}{r_1 \partial \theta_1^2 \partial r_2}\right)\right|$$
$$\le C \cdot \begin{cases} r_1^{1/\beta_1 - 1}\tau^{-k(-1+1/\beta_1)}\omega(\tau^k) & \text{if } k \in [k_{2,p}+1, \min(\ell, k_p)], \\ r_1^{1/\beta_1 - 1} r_2^{1/\beta_2 - 1}\tau^{-k(-2+1/\beta_1+1/\beta_2)}\omega(\tau^k) & \text{if } k \le k_{2,p}. \end{cases} \tag{3-93}$$

Proof. The proof is parallel to that of Lemma 3.19. The function $\partial h_k/\partial\theta_1$ is $g_{\boldsymbol\beta}$-harmonic on $\hat{B}_k(p)$, and by the Laplacian estimates (3-9), we have

$$\sup_{\hat{B}_k(p)/1.2}\left(\left|\Delta_1\frac{\partial h_k}{\partial\theta_1}\right|+\left|\Delta_2\frac{\partial h_k}{\partial\theta_1}\right|\right)\le C(n,\boldsymbol\beta)\omega(\tau^k).$$

The function $\Delta_2(\partial h_k/\partial\theta_1)$ is also $g_{\boldsymbol\beta}$-harmonic, so the Laplacian estimates (3-9) imply

$$\sup_{\hat{B}_k(p)/1.4}\left(\left|\Delta_1\Delta_2\frac{\partial h_k}{\partial\theta_1}\right|+\left|\Delta_2\Delta_2\frac{\partial h_k}{\partial\theta_1}\right|+\left|(D')^2\Delta_2\frac{\partial h_k}{\partial\theta_1}\right|\right)\le C\tau^{-2k}\left(\operatorname{osc}_{\hat{B}_k(p)/1.2}\Delta_2\frac{\partial h_k}{\partial\theta_1}\right)\le C\tau^{-2k}\omega(\tau^k).$$

We consider

$$|z_2|^{2(1-\beta_2)}\frac{\partial^2}{\partial z_2\partial\bar z_2}\left(\Delta_2\frac{\partial h_k}{\partial\theta_1}\right)=-\Delta_1\Delta_2\frac{\partial h_k}{\partial\theta_1}-\sum_j\frac{\partial^2}{\partial s_j^2}\Delta_2\frac{\partial h_k}{\partial\theta_1}=:F_5,\tag{3-94}$$

where the function F_5 satisfies $\sup_{\hat{B}_k(p)/1.4}|F_5|\le C\tau^{-2k}\omega(\tau^k)$.

<u>Case 1</u>: $k_{2,p}+1\le k\le\min(\ell,k_p)$. Here we introduce the smooth coordinate $w_2=z_2^{\beta_2}$ in the ball $\frac{1}{1.5}\hat{B}_k(p)$ as before. Since this ball is disjoint from $\mathcal{S}_2$, under the coordinates $(r_1,\theta_1;w_2,z_3,\dots,z_n)$ we can use the usual standard gradient estimate to the $g_{\boldsymbol\beta}$-harmonic function $\Delta_2(\partial h_k/\partial\theta_1)$ to obtain

$$\sup_{\hat{B}_k(p)/2}\left|\frac{\partial}{\partial r_2}\left(\Delta_2\frac{\partial h_k}{\partial\theta_1}\right)\right|+\left|\frac{\partial}{r_2\partial\theta_2}\left(\Delta_2\frac{\partial h_k}{\partial\theta_1}\right)\right|\le C\tau^{-k}\omega(\tau^k).\tag{3-95}$$

<u>Case 2</u>: $k\le k_{2,p}$. Here the ball $\hat{B}_k(p)$ is centered at $p_{1,2}$. We apply the usual estimate (2-3) to the function $\Delta_2(\partial h_k/\partial\theta_1)$, the solution to (3-94), on any $\mathbb{C}$-ball $A_2:=B_{\mathbb{C}}(y,(\tau^k)^{1/\beta_2})$ for any $y\in\mathcal{S}_2\cap\frac{1}{1.6}\hat{B}_k(p)$, where A_2 denotes the Euclidean ball in the complex plane orthogonal to $\mathcal{S}_2$ and passing through y. Then, for any $z\in B_{\mathbb{C}}(y,(\tau^k)^{1/\beta_2}/2)\backslash\{y\}$,

$$\left|\frac{\partial}{\partial z_2}\left(\Delta_2\frac{\partial h_k}{\partial\theta_1}\right)(z)\right|\le C\frac{\|\Delta_2\frac{\partial h_k}{\partial\theta_1}\|_{L^\infty(A_2)}}{\tau^{k/\beta_2}}+C\|F_5\|_{L^\infty(A_2)}(\tau^k)^{2-1/\beta_2}\le C\tau^{-k/\beta_2}\omega(\tau^k).$$

This implies that, on $\frac12\hat{B}_k(p)\backslash\mathcal{S}$,

$$\left|\frac{\partial}{\partial r_2}\left(\Delta_2\frac{\partial h_k}{\partial\theta_1}\right)\right|+\left|\frac{\partial}{r_2\partial\theta_2}\left(\Delta_2\frac{\partial h_k}{\partial\theta_1}\right)\right|\le Cr_2^{1/\beta_2-1}\tau^{-k/\beta_2}\omega(\tau^k).\tag{3-96}$$

Taking $\partial/\partial r_2$ on both sides of $\Delta_{\boldsymbol\beta}(\partial h_k/\partial\theta_1)=0$, we get

$$|z_1|^{2(1-\beta_1)}\frac{\partial^2}{\partial z_1\partial\bar z_1}\left(\frac{\partial^2 h_k}{\partial r_2\partial\theta_1}\right)=-\frac{\partial}{\partial r_2}\left(\Delta_2\frac{\partial h_k}{\partial\theta_1}\right)-\sum_j\frac{\partial^2}{\partial s_j^2}=:F_6.\tag{3-97}$$

It is not hard to see from (3-95), (3-96) and standard derivatives estimates that, on $\frac{1}{1.8}\hat{B}_k(p)\backslash\mathcal{S}$,

- in Case 1 when $k_{2,p}+1\le k\le\min(\ell,k_p)$, we have $|F_6|\le C\tau^{-k}\omega(\tau^k)$,
- in Case 2 when $k\le k_{2,p}$, we have $|F_6|\le Cr_2^{1/\beta_2-1}\tau^{-k/\beta_2}\omega(\tau^k)$.

Then by applying estimate (2-3) to the function $\partial^2 h_k/\partial r_2\partial\theta_1$ on any $\mathbb{C}$-ball $A_3 := B_{\mathbb{C}}(x, (\tau^k)^{1/\beta_1})$ for any $x \in \frac{1}{1.8}\hat{B}_k(p)\cap\mathcal{S}_1$, we get that, on $B_{\mathbb{C}}(x, (\tau^k)^{1/\beta_1}/2)\backslash\{x\}$,

$$\begin{aligned}&\left|\frac{\partial}{\partial r_1}\left(\frac{\partial^2 h_k}{\partial r_2\partial\theta_1}\right)\right| + \left|\frac{\partial}{r_1\partial\theta_1}\left(\frac{\partial^2 h_k}{\partial r_2\partial\theta_1}\right)\right| \\ &\qquad\le C r_1^{1/\beta_1-1}\frac{\left\|\frac{\partial^2 h_k}{\partial r_2\partial\theta_1}\right\|_{L^\infty(A_3)}}{\tau^{k/\beta_1}} + C r_1^{1/\beta_1-1}\|F_6\|_{L^\infty(A_3)}\tau^{k(2-1/\beta_1)} \\ &\qquad\le C\cdot\begin{cases} r_1^{1/\beta_1-1}\tau^{-k(-1+1/\beta_1)}\omega(\tau^k) & \text{if } k\in[k_{2,p}+1, \min(\ell, k_p)], \\ r_1^{1/\beta_1-1}r_2^{1/\beta_2-1}\tau^{-k(-2+1/\beta_1+1/\beta_2)}\omega(\tau^k) & \text{if } k\le k_{2,p}.\end{cases}\end{aligned}$$

Therefore this estimate holds on $\frac{1}{3}\hat{B}_k(p)\backslash\mathcal{S}$. □

Lemma 3.25. *For any $k\le\ell$ and any point $z\in\frac{1}{3}\hat{B}_k(p)\backslash\mathcal{S}$,*

$$\begin{aligned}\left|\frac{\partial^2 h_k}{\partial r_1\partial r_2}(z)\right| &\le C\cdot\begin{cases} r_1^{1/\beta_1-1}\tau^{-k(-1+1/\beta_1)}\omega(\tau^k) & \text{if } k\in[k_{2,p}+1, \min(\ell, k_p)], \\ r_1^{1/\beta_1-1}r_2^{1/\beta_2-1}\tau^{-k(-2+1/\beta_1+1/\beta_2)}\omega(\tau^k) & \text{if } k\le k_{2,p}.\end{cases} \\ \left|\frac{\partial^2 D'h_k}{\partial r_1\partial r_2}(z)\right| &\le C\cdot\begin{cases} r_1^{1/\beta_1-1}\tau^{-k/\beta_1}\omega(\tau^k) & \text{if } k\in[k_{2,p}+1, \min(\ell, k_p)], \\ r_1^{1/\beta_1-1}r_2^{1/\beta_2-1}\tau^{-k(-1+1/\beta_1+1/\beta_2)}\omega(\tau^k) & \text{if } k\le k_{2,p}.\end{cases}\end{aligned}\tag{3-98}$$

Proof. This follows from almost the same argument as in the proof of Lemma 3.24, by studying the harmonic functions h_k and $D'h_k$ instead of $\partial h_k/\partial\theta_1$. □

Lemma 3.26. *For any $k\le\ell$ and any $z\in\frac{1}{3}\hat{B}_k(p)\backslash\mathcal{S}$,*

$$\begin{aligned}&\left|\frac{\partial}{\partial r_1}\left(\frac{\partial^2 h_k}{\partial r_1\partial r_2}\right)\right|(z) \\ &\le C\omega(\tau^k)\cdot\begin{cases}\tau^{-k} + r_1(z)^{1/\beta_1-2}\tau^{-k(1/\beta_1-1)} & \text{if } k\in[k_{2,p}+1, \min(\ell, k_p)], \\ r_2(z)^{1/\beta_2-1}\tau^{-k/\beta_2} + r_1(z)^{1/\beta_1-2}r_2(z)^{1/\beta_2-1}\tau^{-k(-2+1/\beta_1+1/\beta_2)} & \text{if } k\le k_{2,p}.\end{cases}\end{aligned}$$

Proof. By the Laplacian estimates (3-9) we have

$$\sup_{(\hat{B}_k(p)/1.2)\backslash\mathcal{S}} |\Delta_1 h_k| + |\Delta_2 h_k| \le C(n, \boldsymbol{\beta})\omega(\tau^k).\tag{3-99}$$

Applying again the Laplacian estimates (3-9) to the $g_{\boldsymbol{\beta}}$-harmonic function $\Delta_1 h_k$, we have

$$\sup_{\hat{B}_k(p)/1.4}(|\Delta_1\Delta_1 h_k| + |\Delta_2\Delta_1 h_k| + |(D')^2\Delta_1 h_k|)\le C(n)\tau^{-2k}\omega(\tau^k).$$

We consider the equation

$$|z_2|^{2-2\beta_2}\frac{\partial^2}{\partial z_2\partial\bar{z}_2}\Delta_1 h_k = -\Delta_1\Delta_1 h_k - \sum_j\frac{\partial^2}{\partial s_j^2}\Delta_1 h_k =: F_7.\tag{3-100}$$

From the estimates above, $\|F_7\|_{L^\infty(\hat{B}_k(p)/1.8)}\le C\tau^{-2k}\omega(\tau^k)$.

Case 1: $k_{2,p}+1\le k\le \min(\ell,k_p)$. Here we directly apply the gradient estimate to $\Delta_1 h_k$ to get

$$\sup_{(\hat B_k(p)/1.5)\setminus\mathcal{S}}\left|\frac{\partial}{\partial r_2}\Delta_1 h_k\right|+\left|\frac{\partial}{r_2\partial\theta_2}\Delta_1 h_k\right|\le C\tau^{-k}\omega(\tau^k). \tag{3-101}$$

Case 2: $k\le k_{2,p}$. Here the balls $\hat B_k(p)$ are centered at $p_{1,2}$, and we can apply the usual $\mathbb{C}$-ball type estimate to get that, for any $z\in\frac12\hat B_k(p)\setminus\mathcal{S}$,

$$\begin{aligned}\left|\frac{\partial}{\partial r_2}\Delta_1 h_k\right|(z)+\left|\frac{\partial}{r_2\partial\theta_2}\Delta_1 h_k\right| &\le Cr_2(z)^{1/\beta_2-1}\frac{\|\Delta_1 h_k\|_{L^\infty}}{\tau^{k/\beta_2}}+Cr_2(z)^{1/\beta_2-1}\|F_7\|_{L^\infty}\tau^{k(2-1/\beta_2)}\\ &\le Cr_2(z)^{1/\beta_2-1}\tau^{-k/\beta_2}\omega(\tau^k).\end{aligned}$$

Recall that

$$\frac{\partial}{\partial r_1}\left(\frac{\partial^2 h_k}{\partial r_1\partial r_2}\right)=\frac{\partial}{\partial r_2}\Delta_1 h_k-\frac{1}{r_1}\frac{\partial^2 h_k}{\partial r_1\partial r_2}-\frac{1}{\beta_1^2r_1^2}\frac{\partial^3 h_k}{\partial\theta_1^2\partial r_2},$$

from which we derive that, for any $z\in\frac12\hat B_k(p)\setminus\mathcal{S}$,

$$\begin{aligned}&\left|\frac{\partial}{\partial r_1}\left(\frac{\partial^2 h_k}{\partial r_1\partial r_2}\right)\right|(z)\\ &\le C\omega(\tau^k)\begin{cases}\tau^{-k}+r_1(z)^{1/\beta_1-2}\tau^{-k(1/\beta_1-1)} & \text{if } k\in[k_{2,p}+1,\min(\ell,k_p)],\\ r_2(z)^{1/\beta_2-1}\tau^{-k/\beta_2}+r_1(z)^{1/\beta_1-2}r_2(z)^{1/\beta_2-1}\tau^{-k(-2+1/\beta_1+1/\beta_2)} & \text{if } k\le k_{2,p}.\end{cases}\end{aligned}$$

□

Lemma 3.27. *There exists a constant $C=C(n,\boldsymbol{\beta})>0$ such that, for all $k\le\ell$ and $z\in\frac13\hat B_k(p)\setminus\mathcal{S}$,*

$$\begin{aligned}&\left|\frac{\partial}{\partial\theta_2}\left(\frac{\partial^2 h_k}{\partial r_1\partial r_2}\right)(z)\right|+\left|\left(\frac{\partial^3 h_k}{r_2\partial\theta_2^2\partial r_1}\right)(z)\right|\\ &\qquad\le C\omega(\tau^k)\begin{cases}r_1^{1/\beta_1-1}\tau^{-k(-1+1/\beta_1)} & \textit{if } k\in[k_{2,p}+1,\min(\ell,k_p)],\\ r_1^{1/\beta_1-1}r_2^{1/\beta_2-1}\tau^{-k(-2+1/\beta_1+1/\beta_2)} & \textit{if } k\le k_{2,p}.\end{cases}\end{aligned}\tag{3-102}$$

Proof. It follows from the Laplacian estimates (3-9) that

$$\sup_{\hat B_k(p)/1.2}\left(\left|\Delta_1\frac{\partial h_k}{\partial\theta_2}\right|+\left|\Delta_2\frac{\partial h_k}{\partial\theta_2}\right|\right)\le C(n)\omega(\tau^k).$$

Again by (3-9), we have

$$\sup_{\hat B_k(p)/1.4}\left(\left|\Delta_1\Delta_1\frac{\partial h_k}{\partial\theta_2}\right|+\left|\Delta_2\Delta_1\frac{\partial h_k}{\partial\theta_2}\right|+\left|(D')^2\Delta_1\frac{\partial h_k}{\partial\theta_2}\right|\right)\le C\tau^{-2k}\omega(\omega^k).$$

We look at the equation

$$|z_1|^{2(1-\beta_1)}\frac{\partial^2}{\partial z_1\partial\bar z_1}\left(\Delta_1\frac{\partial h_k}{\partial\theta_2}\right)=-\Delta_2\Delta_1\frac{\partial h_k}{\partial\theta_2}-\sum_j\frac{\partial^2}{\partial s_j^2}\left(\Delta_1\frac{\partial h_k}{\partial\theta_2}\right)=:F_8$$

and note that

$$\sup_{\hat B_k(p)/1.4}|F_8|\le C\tau^{-2k}\omega(\tau^k).$$

As we did before, by estimate (2-3) it follows that, for any $z \in \frac{1}{2}\hat{B}_k(p)\backslash\mathcal{S}$ (remember here $k \le \min(\ell, k_p)$),

$$\left|\frac{\partial}{\partial r_1}\Delta_1 \frac{\partial h_k}{\partial \theta_2}\right|(z) + \left|\frac{\partial}{r_1\partial \theta_1}\Delta_1 \frac{\partial h_k}{\partial \theta_2}\right|(z) \le C r_1(z)^{1/\beta_1 - 1}\frac{\left\|\Delta_1 \frac{\partial h_k}{\partial \theta_2}\right\|_{L^\infty}}{\tau^{k/\beta_1}} + C r_1(z)^{1/\beta_1-1}\|F_8\|_{L^\infty}\tau^{k(2-1/\beta_1)}$$
$$\le C r_1(z)^{1/\beta_1-1}\tau^{-k/\beta_1}\omega(\tau^k).$$

Taking $\partial/\partial r_1$ on both sides of the equation $\Delta_{\boldsymbol{\beta}}(\partial h_k/\partial\theta_2) = 0$, we get

$$|z_2|^{2(1-\beta_2)}\frac{\partial^2}{\partial z_2 \partial \bar{z}_2}\left(\frac{\partial^2 h_k}{\partial r_1 \partial \theta_2}\right) = -\frac{\partial}{\partial r_1}\left(\Delta_1 \frac{\partial h_k}{\partial \theta_2}\right) - \sum_j \frac{\partial}{\partial r_1}\left(\frac{\partial^2}{\partial s_j^2}\frac{\partial h_k}{\partial \theta_2}\right) =: F_9. \tag{3-103}$$

Here $|F_9(z)| \le C r_1(z)^{1/\beta_1-1}\tau^{-k/\beta_1}\omega(\tau^k)$ for any $z \in \frac{1}{2}\hat{B}_k(p)\backslash\mathcal{S}$. Therefore, by the usual $\mathbb{C}$-ball argument,

- when $k \le k_{2,p}$, for any $z \in \frac{1}{3}\hat{B}_k(p)\backslash\mathcal{S}$, we have

$$\left|\frac{\partial}{\partial r_2}\left(\frac{\partial^2 h_k}{\partial r_1 \partial \theta_2}\right)(z)\right| + \left|\frac{\partial}{r_2\partial \theta_2}\left(\frac{\partial^2 h_k}{\partial r_1 \partial \theta_2}\right)(z)\right| \le C r_2(z)^{1/\beta_2-1} r_1(z)^{1/\beta_1-1}\tau^{k(2-1/\beta_1-1/\beta_2)}\omega(\tau^k),$$

- when $k_{2,p}+1 \le k \le \min(\ell, k_p)$, we have

$$\left|\frac{\partial}{\partial r_2}\left(\frac{\partial^2 h_k}{\partial r_1 \partial \theta_2}\right)(z)\right| + \left|\frac{\partial}{r_2\partial \theta_2}\left(\frac{\partial^2 h_k}{\partial r_1 \partial \theta_2}\right)(z)\right| \le C r_1(z)^{1/\beta_1-1}\tau^{k(1-1/\beta_1)}\omega(\tau^k). \qquad \square$$

Lemma 3.28. *For any $k \le \ell$ and any $z \in \frac{1}{3}\hat{B}_k(p)\backslash\mathcal{S}$,*

$$\left|\frac{\partial}{\partial r_2}\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right)(z)\right|$$
$$\le C\omega(\tau^k)\begin{cases} r_1(z)^{1/\beta_1-1}\tau^{-k/\beta_1} + r_1^{1/\beta_1-1} r_2^{-1}\tau^{-k(-1+1/\beta_1)} & \text{if } k \in [k_{2,p}+1, \min(\ell, k_p)], \\ r_1(z)^{1/\beta_1-1}\tau^{-k/\beta_1} + r_1^{1/\beta_1-1} r_2^{1/\beta_2-2}\tau^{-k(-2+1/\beta_1+1/\beta_2)} & \text{if } k \le k_{2,p}. \end{cases} \tag{3-104}$$

Proof. We first observe that

$$\frac{\partial}{\partial r_2}\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right) = \frac{\partial}{\partial r_1}\Delta_2 h_k - \frac{1}{r_2}\frac{\partial^2 h_k}{\partial r_1 \partial r_2} - \frac{1}{\beta_2^2 r_2^2}\frac{\partial^2}{\partial \theta_2^2}\left(\frac{\partial h_k}{\partial r_1}\right).$$

It can be shown by the $\mathbb{C}$-ball argument that, for any $z \in \frac{1}{2}\hat{B}_k(p)\backslash\mathcal{S}$,

$$\left|\frac{\partial}{\partial r_1}\Delta_2 h_k(z)\right| \le C r_1(z)^{1/\beta_1-1}\tau^{-k/\beta_1}\omega(\tau^k).$$

From Lemma 3.25, we have, for any $z \in \frac{1}{2}\hat{B}_k(p)\backslash\mathcal{S}$,

$$\left|\frac{1}{r_2}\frac{\partial^2 h_k}{\partial r_1 \partial r_2}(z)\right| \le C\cdot\begin{cases} r_1^{1/\beta_1-1} r_2^{-1}\tau^{-k(-1+1/\beta_1)}\omega(\tau^k) & \text{if } k \in [k_{2,p}+1, \min(\ell, k_p)], \\ r_1^{1/\beta_1-1} r_2^{1/\beta_2-2}\tau^{-k(-2+1/\beta_1+1/\beta_2)}\omega(\tau^k) & \text{if } k \le k_{2,p}. \end{cases}$$

From Lemma 3.27, we have, for any $z \in \frac{1}{2}\hat{B}_k(p)\backslash\mathcal{S}$,

$$\left|\frac{1}{r_2^2}\frac{\partial^3 h_k}{\partial r_1 \partial \theta_2^2}(z)\right| \le C\omega(\tau^k)\cdot\begin{cases} r_1^{1/\beta_1-1} r_2^{-1}\tau^{-k(-1+1/\beta_1)} & \text{if } k \in [k_{2,p}+1, \min(\ell, k_p)], \\ r_1^{1/\beta_1-1} r_2^{1/\beta_2-2}\tau^{-k(-2+1/\beta_1+1/\beta_2)} & \text{if } k \le k_{2,p}. \end{cases}$$

Therefore, for any $z \in \frac{1}{3}\hat{B}_k(p)\backslash\mathcal{S}$, we have

$$\left|\frac{\partial}{\partial r_2}\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right)(z)\right| \leq C\omega(\tau^k)\cdot\begin{cases} r_1(z)^{1/\beta_1-1}\tau^{-k/\beta_1} + r_1^{1/\beta_1-1} r_2^{-1}\tau^{-k(-1+1/\beta_1)} & \text{if } k \in [k_{2,p}+1, \min(\ell, k_p)], \\ r_1(z)^{1/\beta_1-1}\tau^{-k/\beta_1} + r_1^{1/\beta_1-1} r_2^{1/\beta_2-2}\tau^{-k(-2+1/\beta_1+1/\beta_2)} & \text{if } k \leq k_{2,p}. \end{cases} \qquad \square$$

It remains to estimate L_2. For simplicity, we write $h_k := -u_k + u_{k-1}$ as before, where we take $k \leq \ell$. We will define $\beta_{\max} = \max(\beta_1, \beta_2)$.

Lemma 3.29. *Let $d = d_{\boldsymbol{\beta}}(p, q)$. There exists a constant $C(n, \boldsymbol{\beta}) > 0$ such that, for all $k \leq \ell$,*

$$\begin{aligned}\left|\frac{\partial^2 h_k}{\partial r_1 \partial r_2}(p) - \frac{\partial^2 h_k}{\partial r_1 \partial r_2}(q)\right| &\leq C\omega(\tau^k)\tau^{-k(1/\beta_1-1)} d^{1/\beta_1-1} \\ &\leq C\omega(\tau^k)\tau^{-k(1/\beta_{\max}-1)} d^{1/\beta_{\max}-1}.\end{aligned}$$

Proof. Case 1: First we assume that $r_p \leq 2d$, so that $r_q \leq 3d$ and $\ell + 2 \leq k_p$. In particular, the balls $\hat{B}_k(p)$ are centered at either $p_1 \in \mathcal{S}_1$ or 0, depending on whether $k \geq k_{2,p}+1$ or $k \leq k_{2,p}$. As in the proof of Lemma 3.20, let $\gamma : [0, d] \to B_{\boldsymbol{\beta}}(0, 1)\backslash S$ be the $g_{\boldsymbol{\beta}}$-geodesic connecting p and q, let the two points q' and p' be defined as in (3-80), and let $\gamma_1, \gamma_2, \gamma_3$ be the $g_{\boldsymbol{\beta}}$-geodesics defined in that lemma. By the triangle inequality we calculate

$$\begin{aligned}&\left|\frac{\partial^2 h_k}{\partial r_1 \partial r_2}(p) - \frac{\partial^2 h_k}{\partial r_1 \partial r_2}(q)\right| \\ &\leq \left|\frac{\partial^2 h_k}{\partial r_1 \partial r_2}(p) - \frac{\partial^2 h_k}{\partial r_1 \partial r_2}(p')\right| + \left|\frac{\partial^2 h_k}{\partial r_1 \partial r_2}(p') - \frac{\partial^2 h_k}{\partial r_1 \partial r_2}(q')\right| + \left|\frac{\partial^2 h_k}{\partial r_1 \partial r_2}(q') - \frac{\partial^2 h_k}{\partial r_1 \partial r_2}(q)\right| =: L_1' + L_2' + L_3'.\end{aligned}$$

Integrating along γ_3, where the coordinates $(r_1; r_2, \theta_2; z_3, \ldots, z_n)$ are the same as p, we get (by (3-93))

$$L_1' = \left|\int_{\gamma_3} \frac{\partial}{\partial \theta_1}\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right) d\theta_1\right| \leq C\omega(\tau^k)\cdot\begin{cases} r_1(p)^{1/\beta_1-1}\tau^{-k(-1+1/\beta_1)} & \text{if } k \in [k_{2,p}+1, \ell], \\ r_1(p)^{1/\beta_1-1} r_2(p)^{1/\beta_2-1}\tau^{-k(-2+1/\beta_1+1/\beta_2)} & \text{if } k \leq k_{2,p}. \end{cases}$$

Integrating along γ_2, where the coordinates $(\theta_1; r_2, \theta_2; z_3, \ldots, z_n)$ are the same as p' or q', we get by the estimate in Lemma 3.26 that

$$\begin{aligned}L_2' &= \left|\int_{\gamma_2} \frac{\partial}{\partial r_1}\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right) dr_1\right| \\ &\leq C\omega(\tau^k)\cdot\begin{cases} \tau^{-k} d + \tau^{-k(1/\beta_1-1)}|r_1(p) - r_1(q)|^{1/\beta_1-1} & \text{if } k \in [k_{2,p}+1, \ell], \\ r_2(p)^{1/\beta_2-1}\tau^{-k/\beta_2} d \\ \qquad + r_2(p)^{1/\beta_2-1}\tau^{-k(-2+1/\beta_1+1/\beta_2)}|r_1(p) - r_1(q)|^{1/\beta_1-1} & \text{if } k \leq k_{2,p} \end{cases} \\ &\leq C\omega(\tau^k)\cdot\begin{cases} \tau^{-k} d + \tau^{-k(1/\beta_1-1)} d^{1/\beta_1-1} & \text{if } k \in [k_{2,p}+1, \ell], \\ r_2(p)^{1/\beta_2-1}\tau^{-k/\beta_2} d + r_2(p)^{1/\beta_2-1}\tau^{-k(-2+1/\beta_1+1/\beta_2)} d^{1/\beta_1-1} & \text{if } k \leq k_{2,p}. \end{cases}\end{aligned}$$

To deal with the term L_3', we consider two cases for k: $\ell \geq k \geq k_{2,p}+1$ and $k \leq k_{2,p}$.

Case 1a: $k_{2,p}+1 \le k \le \ell$. In this case the balls $\hat{B}_k(p)$ are centered at $p_1 \in \mathcal{S}_1$. Here $\tau^{-k} \le \tau^{-\ell} \le 8^{-1}d^{-1}$ and $\tau^k \le \tau^{k_{2,p}+1} \le \frac{1}{2}r_2(p)$, so $r_2(q) \ge -d + r_2(p) \ge \tau^k$. The balls $\hat{B}_k(p)$ are disjoint from $\mathcal{S}_2$, and we can use the smooth coordinate $w_2 = z_2^{\beta_2}$ as before. The functions $D_{w_2} D' h_k$ are $g_{\boldsymbol{\beta}}$-harmonic; hence by the gradient estimate we have

$$\sup_{(\hat{B}_k(p)/1.2)\setminus\mathcal{S}_1} |\nabla_{g_{\boldsymbol{\beta}}}(D_{w_2} D' h_k)| \le C(n) \frac{\|D_{w_2} D' h_k\|_{L^\infty(\hat{B}_k(p)/1.1)}}{\tau^k} \le C\tau^{-k}\omega(\tau^k).$$

From (3-88), we get

$$\sup_{(\hat{B}_k(p)/1.2)\setminus\mathcal{S}_1} \left| \frac{\partial^2}{\partial r_1 \partial r_2} D' h_k \right| \le C(n)\tau^{-k}\omega(\tau^k). \tag{3-105}$$

Recalling that $r_1(p) = r_p \le 2d \le \frac{1}{2}\tau^k$, the triangle inequality implies $r_1(q) \le 3d \le \frac{1}{2}\tau^k$. The points in γ_1 have fixed (r_1,θ_1)-coordinates $(r_1(q),\theta_1(q))$, so integrating along γ_1 we get (by (3-104) and (3-105))

$$\begin{aligned} L_3' &\le \int_{\gamma_1} \left| \frac{\partial}{\partial r_2}\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right)\right| + \left| \frac{\partial}{r_2 \partial \theta_2}\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right)\right| + \left| D'\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right)\right| \\ &\le Cd\omega(\tau^k)(r_1(q)^{1/\beta_1-1}\tau^{-k/\beta_1} + \tau_1(q)^{1/\beta_1-1}\min(r_2(p),r_2(q))^{-1}\tau^{-k(1/\beta_1-1)} + \tau^{-k}) \\ &\le C\tau^{-k}\omega(\tau^k)\cdot d \le C\tau^{-k(1/\beta_1-1)}\omega(\tau^k)d^{1/\beta_1-1}. \end{aligned}$$

Case 1b: $k \le k_{2,p}$. In this case $\tau^k \ge \tau^{k_{2,p}} \ge r_2(p)$ and $\tau^k \ge \tau^\ell \ge 8d$. Thus $r_2(q) \le r_2(p) + d \le \frac{3}{2}\tau^k$. We choose points $\tilde{q}$ and $\hat{q}$ as in (3-84), and let $\tilde{\gamma}_1$, $\tilde{\gamma}$ and $\hat{\gamma}$ be $g_{\boldsymbol{\beta}}$-geodesics defined as in the proof of Lemma 3.20. Then we have

$$L_3' \le \left| \frac{\partial^2 h_k}{\partial r_1 \partial r_2}(q') - \frac{\partial^2 h_k}{\partial r_1 \partial r_2}(\tilde{q})\right| + \left| \frac{\partial^2 h_k}{\partial r_1 \partial r_2}(\tilde{q}) - \frac{\partial^2 h_k}{\partial r_1 \partial r_2}(\hat{q})\right| + \left| \frac{\partial^2 h_k}{\partial r_1 \partial r_2}(\hat{q}) - \frac{\partial^2 h_k}{\partial r_1 \partial r_2}(q)\right| =: L_1'' + L_2'' + L_3''.$$

We will estimate L_1'', L_2'' and L_3'' term by term by integrating appropriate functions along the geodesics $\tilde{\gamma}_1$, $\tilde{\gamma}$ and $\hat{\gamma}$ as follows: The points in $\hat{\gamma}$ have fixed $(r_1,\theta_1;r_2;s)$-coordinates $(r_1(q),\theta_1(q);r_2(q);s(q))$, so (by (3-102))

$$\begin{aligned} L_3'' = \left| \int_{\hat{\gamma}} \frac{\partial}{\partial \theta_2}\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right) d\theta_2 \right| &\le C\omega(\tau^k) r_1(q)^{1/\beta_1-1} r_2(q)^{1/\beta_2-1}\tau^{-k(-2+1/\beta_1+1/\beta_2)} \\ &\le C\omega(\tau^k) r_1(q)^{1/\beta_1-1}\tau^{-k(-1+1/\beta_1)} \le C\tau^{-k(1/\beta_1-1)}\omega(\tau^k)d^{1/\beta_1-1}. \end{aligned}$$

Integrating along $\tilde{\gamma}$, where the points have constant r_1-coordinate $r_1(q)$, we get (by (3-104))

$$\begin{aligned} L_2'' &= \left| \int_{\tilde{\gamma}} \frac{\partial}{\partial r_2}\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right) dr_2 \right| \\ &\le C\omega(\tau^k)\big(r_1(q)^{1/\beta_1-1}\tau^{-k/\beta_1}|r_2(q)-r_2(p)| \\ &\qquad\qquad + r_1(q)^{1/\beta_1-1}\tau^{-k(-2+1/\beta_1+1/\beta_2)}|r_2(q)^{1/\beta_2-1} - r_2(p)^{1/\beta_2-1}|\big) \\ &\le C\omega(\tau^k)(r_1(q)^{1/\beta_1-1}\tau^{-k/\beta_1}d + r_1(q)^{1/\beta_1-1}\tau^{-k(-2+1/\beta_1+1/\beta_2)}d^{1/\beta_2-1}) \\ &\le C\omega(\tau^k)r_1(q)^{1/\beta_1-1}\tau^{-k(-1+1/\beta_1)} \\ &\le C\tau^{-k(1/\beta_1-1)}\omega(\tau^k)d^{1/\beta_1-1}. \end{aligned}$$

Integrating along $\tilde{\gamma}_1$, where the points have constant $(r_1, \theta_1; r_2, \theta_2)$-coordinates, we have (by (3-71))

$$\begin{aligned} L_1'' \le \int_{\tilde{\gamma}_1} \left| D'\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right)\right| &\le C r_1(q)^{1/\beta_1 - 1} r_2(p)^{1/\beta_2 - 1} \tau^{-k(-1+1/\beta_1+1/\beta_2)} d \\ &\le C d \tau^{-k}\omega(\tau^k) \le C d^{1/\beta_1 - 1}\tau^{-k(1/\beta_1 - 1)}\omega(\tau^k). \end{aligned}$$

Combining both cases, we conclude that

$$L_3' \le C\tau^{-k(1/\beta_1-1)}\omega(\tau^k) d^{1/\beta_1 - 1}.$$

Then by the estimates above for L_1' and L_2', we finally get, for all $k \le \ell$,

$$\left|\frac{\partial^2 h_k}{\partial r_1 \partial r_2}(p) - \frac{\partial^2 h_k}{\partial r_1 \partial r_2}(q)\right| \le C\omega(\tau^k)\tau^{-k(1/\beta_1 - 1)} d^{1/\beta_1 - 1} \le C\omega(\tau^k)\tau^{-k(1/\beta_{\max} - 1)} d^{1/\beta_{\max} - 1},$$

where in the last inequality we use the fact that $\tau^{-k} d \le \frac{1}{8} < 1$ when $k \le \ell$. Hence we finish the proof of Lemma 3.29 in the case $r_p \le 2d$.

Now we deal with the remaining cases.

Case 2: Here we assume $\min(r_p, r_q) = r_p \ge 2d$ and $\ell \le k_p$. In this case $\tau^{k_p} \approx r_p \ge 2d \ge \tau^{\ell+3}$, so $\ell + 3 \ge k_p$. It follows by the triangle inequality that $d_{\boldsymbol{\beta}}(\gamma(t), \mathcal{S}) \ge d$, where γ is the $g_{\boldsymbol{\beta}}$-geodesic joining p to q as before. In particular, this implies that $\min(r_1(\gamma(t)), r_2(\gamma(t))) \ge d$.

Since $\ell \le k_p$, Lemmas 3.24–3.28 hold for all $k \le \ell$ and $r_1(p) \approx \tau^{k_p} \le \tau^\ell$, so

$$r_1(\gamma(t)) \le d + r_1(p) \le \tfrac{9}{8}\tau^\ell \le \tfrac{9}{8}\tau^k.$$

We calculate the gradient of $\partial^2 h_k / \partial r_1 \partial r_2$ along the geodesic γ as

$$\begin{aligned} \left|\nabla_{g_{\boldsymbol{\beta}}} \frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right|^2 (\gamma(t)) = \left|\frac{\partial}{\partial r_1}\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right)\right|^2 + \left|\frac{1}{\beta_1 r_1 \partial \theta_1}\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right)\right|^2 + \left|\frac{\partial}{\partial r_2}\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right)\right|^2 \\ + \left|\frac{\partial}{\beta_2 r_2 \partial \theta_2}\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right)\right|^2 + \sum_j \left|\frac{\partial}{\partial s_j}\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right)\right|^2. \end{aligned}$$

Case 2a: $k_{2,p} + 1 \le k \le \ell$. Here along γ we have

$$r_2(\gamma(t)) \ge r_2(p) - d \ge \tau^k - d \ge \tfrac{7}{8}\tau^k.$$

Then by Lemmas 3.24–3.28, along γ we have

$$\left|\nabla_{g_{\boldsymbol{\beta}}} \frac{\partial^2 h_k}{\partial r_1 \partial r_2}(\gamma(t))\right| \le C\omega(\tau^k)(\tau^{-k} + d^{1/\beta_1 - 2}\tau^{-k(1/\beta_1 - 1)}).$$

Integrating this inequality along γ we get

$$\begin{aligned} \left|\frac{\partial^2 h_k}{\partial r_1 \partial r_2}(p) - \frac{\partial^2 h_k}{\partial r_1 \partial r_2}(q)\right| \le \int_\gamma \left|\nabla_{g_{\boldsymbol{\beta}}} \frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right| &\le C\omega(\tau^k)(d\tau^{-k} + d^{1/\beta_1 - 1}\tau^{-k(1/\beta_1 - 1)}) \\ &\le C d^{1/\beta_1 - 1}\tau^{-k(1/\beta_1 - 1)}\omega(\tau^k). \end{aligned}$$

Case 2b: $k \le k_{2,p}$. Here along γ we have

$$r_2(\gamma(t)) \le r_2(p) + d \le \tau^k + d \le \tfrac{9}{8}\tau^k.$$

Then by Lemmas 3.24–3.28, along γ we have

$$\left|\nabla_{g_\beta} \frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right| \le C\omega(\tau^k)(\tau^{-k} + d^{1/\beta_1 - 2}\tau^{-k(1/\beta_1 - 1)}).$$

Integrating this inequality along γ we again get

$$\left|\frac{\partial^2 h_k}{\partial r_1 \partial r_2}(p) - \frac{\partial^2 h_k}{\partial r_1 \partial r_2}(q)\right| \le \int_\gamma \left|\nabla_{g_\beta} \frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right| \le C d^{1/\beta_1 - 1}\tau^{-k(1/\beta_1 - 1)}\omega(\tau^k).$$

Case 3: Here we assume $\min(r_p, r_q) = r_p \ge 2d$ but $\ell \ge k_p + 1$. The case when $k \le k_p$ can be dealt with by the same argument as in Case 2, so we omit it and only consider the case when $k_p + 1 \le k \le \ell$. Here $r_2(p) \ge r_1(p) \ge \tau^k \ge \tau^\ell > 8d$, and hence

$$r_1(\gamma(t)) \ge \tfrac{7}{8}\tau^k \quad \text{and} \quad r_2(\gamma(t)) \ge \tfrac{7}{8}\tau^k$$

for any point $\gamma(t)$ in the geodesic γ. By the triangle inequality it follows that $\gamma \subset \frac{1}{3}\hat{B}_k(p) = B_{\boldsymbol\beta}\big(p, \frac{1}{3}\tau^k\big)$.

As before, we can introduce smooth coordinates $w_1 = z_1^{\beta_1}$ and $w_2 = z_2^{\beta_2}$, and $g_{\boldsymbol\beta}$ becomes the standard smooth Euclidean metric $g_{\mathbb{C}^n}$ under these coordinates. Moreover, h_k is the usual Euclidean harmonic function $\Delta_{g_{\mathbb{C}^n}} h_k = 0$ on $\hat{B}_k(p)$. By the standard derivatives estimates we have

$$\sup_{\hat{B}_k(p)/2} (|D^3_{w_1,w_2} h_k| + |D'(D^2_{w_1,w_2}) h_k|) \le C\tau^{-k}\omega(\tau^k).$$

From the equation

$$\frac{\partial^2 h_k}{\partial r_1 \partial r_2} = \frac{w_1 w_2}{r_1 r_2}\frac{\partial^2 h_k}{\partial w_1 \partial w_2} + \frac{\bar{w}_1 w_2}{r_1 r_2}\frac{\partial^2 h_k}{\partial \bar{w}_1 \partial w_2} + \frac{w_1 \bar{w}_2}{r_1 r_2}\frac{\partial^2 h_k}{\partial w_1 \partial \bar{w}_2} + \frac{\bar{w}_1 \bar{w}_2}{r_1 r_2}\frac{\partial^2 h_k}{\partial \bar{w}_1 \partial \bar{w}_2}$$

we see that, for $i = 1, 2$,

$$\sup_{\hat{B}_k(p)/2}\left|\frac{\partial}{\partial w_i}\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right)\right| \le \frac{C}{r_i}\omega(\tau^k) + C\tau^{-k}\omega(\tau^k) \quad \text{and} \quad \sup_{\hat{B}_k(p)/2}\left|D'\left(\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right)\right| \le C\tau^{-k}\omega(\tau^k).$$

From this we see that

$$\sup_\gamma \left|\nabla_{g_\beta}\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right| \le \sup_\gamma \left(C\tau^{-k}\omega(\tau^k) + \frac{C}{r_1}\omega(\tau^k) + \frac{C}{r_2}\omega(\tau^k)\right) \le C\tau^{-k}\omega(\tau^k).$$

Integrating along γ we get

$$\begin{aligned}\left|\frac{\partial^2 h_k}{\partial r_1 \partial r_2}(p) - \frac{\partial^2 h_k}{\partial r_1 \partial r_2}(q)\right| \le \int_\gamma \left|\nabla_{g_\beta}\frac{\partial^2 h_k}{\partial r_1 \partial r_2}\right| &\le C d\tau^{-k}\omega(\tau^k) \\ &\le C d^{1/\beta_1 - 1}\tau^{-k(1/\beta_1 - 1)}\omega(\tau^k).\end{aligned}$$

Combining the estimates in all three cases, we finish the proof of Lemma 3.29. □

By Lemma 3.29,

$$L_2 = \left|\frac{\partial^2 u_\ell}{\partial r_1 \partial r_2}(p) - \frac{\partial^2 u_\ell}{\partial r_1 \partial r_2}(q)\right| \leq \left|\frac{\partial^2 u_2}{\partial r_1 \partial r_2}(p) - \frac{\partial^2 u_2}{\partial r_1 \partial r_2}(q)\right| + C d^{1/\beta_{\max}-1} \sum_{k=3}^{\ell} \tau^{-k(1/\beta_{\max}-1)} \omega(\tau^k). \tag{3-106}$$

To finish the proof, it suffices to estimate the first term on the right-hand side of the above equation. Recall that we assume u_2 is a $g_{\boldsymbol{\beta}}$-harmonic function defined on the ball $\hat{B}_2(p)$, which is centered at $p_{1,2} \in \mathcal{S}_1 \cap \mathcal{S}_2$ and has radius $2\tau^2$. We also know u_2 satisfies the L^∞-estimate by the maximum principle: there exists some $C = C(n) > 0$ such that

$$\|u_2\|_{L^\infty(\hat{B}_k(p))} \leq C(\|u\|_{L^\infty(B_{\boldsymbol{\beta}}(0,1))} + \omega(\tau^2)). \tag{3-107}$$

Recall that the proofs of the estimates in Lemmas 3.24–3.28 in the case when $k \leq k_{2,p}$ work for any $g_{\boldsymbol{\beta}}$-harmonic functions defined on suitable balls, and we can repeat the arguments there replacing the L^∞-estimate of h_k, namely $\|h_k\|_{L^\infty} \leq C\tau^{2k}\omega(\tau^k)$, by the L^∞-estimate of u_2 given in (3-107) to get similar estimates as in those lemmas. We will omit the details. Given these estimates, we can repeat the proof of Lemma 3.29 to prove the estimates

$$\left|\frac{\partial^2 u_2}{\partial r_1 \partial r_2}(p) - \frac{\partial^2 u_2}{\partial r_1 \partial r_2}(q)\right| \leq C d^{1/\beta_{\max}-1}(\|u\|_{L^\infty(B_{\boldsymbol{\beta}}(0,1))} + \omega(\tau^2)).$$

This inequality, combined with (3-106), gives the final estimate of the L_2 term, that is

$$L_2 \leq C d^{1/\beta_{\max}-1} \|u\|_{L^\infty(B_{\boldsymbol{\beta}}(0,1))} + C d^{1/\beta_{\max}-1} \sum_{k=2}^{\ell} \tau^{-k(1/\beta_{\max}-1)} \omega(\tau^k). \tag{3-108}$$

By Lemmas 3.22 and 3.23 and the estimate (3-108) for L_2, we are ready to prove the following estimate; see (1-5).

Proposition 3.30. *For given $p, q \in B_{\boldsymbol{\beta}}\big(0, \frac{1}{2}\big) \backslash \mathcal{S}$, there is a constant $C = C(n, \boldsymbol{\beta}) > 0$ such that*

$$\left|\frac{\partial^2 u}{\partial r_1 \partial r_2}(p) - \frac{\partial^2 u}{\partial r_1 \partial r_2}(q)\right| \leq C\left(d^{1/\beta_{\max}-1} \|u\|_{L^\infty(B_{\boldsymbol{\beta}}(0,1))} + \int_0^d \frac{\omega(r)}{r}\, dr + d^{1/\beta_{\max}-1} \int_d^1 \frac{\omega(r)}{r^{1/\beta_{\max}}}\, dr\right).$$

Proof. From Lemmas 3.22 and 3.23 and the estimate (3-108) for L_2, we have

$$\begin{aligned}\left|\frac{\partial^2 u}{\partial r_1 \partial r_2}(p) - \frac{\partial^2 u}{\partial r_1 \partial r_2}(q)\right| &\leq C\left(d^{1/\beta_{\max}-1} \|u\|_{L^\infty(B_{\boldsymbol{\beta}}(0,1))} + d^{1/\beta_{\max}-1} \sum_{k=2}^{\ell} \tau^{-k(1/\beta_{\max}-1)} \omega(\tau^k) + \sum_{k=\ell}^{\infty} \omega(\tau^k)\right) \\ &\leq C\left(d^{1/\beta_{\max}-1} \|u\|_{L^\infty(B_{\boldsymbol{\beta}}(0,1))} + \int_0^d \frac{\omega(r)}{r}\, dr + d^{1/\beta_{\max}-1} \int_d^1 \frac{\omega(r)}{r^{1/\beta_{\max}}}\, dr\right),\end{aligned}$$

where the last inequality follows from the fact that $\omega(r)$ is monotonically increasing. □

Finally, we remark that the estimates for the other operators in (3-89) follow similarly; we omit the proofs and state that the estimates are the same as the estimates for $\partial^2 u/\partial r_1 \partial r_2$ in Proposition 3.30.

3E. ***Nonflat conical Kähler metrics.*** In this section, we will consider the Schauder estimates for general conical Kähler metrics on $B_{\boldsymbol\beta}(0,2)\subset\mathbb{C}^n$ with cone angle $2\pi\boldsymbol\beta$ along the simple normal crossing hypersurface $\mathcal{S}$. Let ω be such a metric. By definition, there exists a constant $C\ge 1$ such that

$$C^{-1}\omega_{\boldsymbol\beta}\le\omega\le C\omega_{\boldsymbol\beta}\quad\text{in } B_{\boldsymbol\beta}(0,2)\backslash\mathcal{S},\tag{3-109}$$

where $\omega_{\boldsymbol\beta}$ is the standard flat conical metric as before. Since ω is closed and $B_{\boldsymbol\beta}(0,2)$ is simply connected, we can write $\omega=\sqrt{-1}\partial\bar\partial\phi$ for some strictly plurisubharmonic function ϕ. By elliptic regularity, ϕ is Hölder continuous under the Euclidean metric on $B_{\boldsymbol\beta}(0,2)$.

We fix $\alpha\in(0,\min\{1/\beta_{\max}-1,1\})$.

Definition 3.31. We say $\omega=g$ is a $C^{0,\alpha}_{\boldsymbol\beta}$ Kähler metric on $B_{\boldsymbol\beta}(0,2)$ if it satisfies (3-109) and the Kähler potential ϕ of ω belongs to $C^{2,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta}(0,2))$.

We are interested in studying the Laplacian equation

$$\Delta_g u=f\quad\text{in } B_{\boldsymbol\beta}(0,1),\tag{3-110}$$

where $f\in C^{0,\alpha}_{\boldsymbol\beta}(\overline{B_{\boldsymbol\beta}(0,1)})$ and $u\in C^{2,\alpha}_{\boldsymbol\beta}$. We will prove the following scaling-invariant interior Schauder estimates. The proof closely follows that of Theorem 6.6 in [18], so we mainly focus on the differences.

Proposition 3.32. *There exists a constant* $C=C(n,\boldsymbol\beta,\|g\|^*_{C^{0,\alpha}_{\boldsymbol\beta}})>0$ *such that, if* $u\in C^{2,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta}(0,1))$ *satisfies* (3-110), *then*

$$\|u\|^*_{C^{2,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta}(0,1))}\le C(\|u\|_{C^0(B_{\boldsymbol\beta}(0,1))}+\|f\|^{(2)}_{C^{0,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta}(0,1))}).\tag{3-111}$$

Proof. Given any points $x_0\ne y_0\in B_{\boldsymbol\beta}(0,1)$, assume $d_{x_0}=\min(d_{x_0},d_{y_0})$; recall $d_x=d_{\boldsymbol\beta}(x,\partial B_{\boldsymbol\beta}(0,1))$. Let $\mu\in\left(0,\frac14\right)$ be a small number to be determined later. Write $d=\mu d_{x_0}$, and define $B:=B_{\boldsymbol\beta}(x_0,d)$ and $\frac12 B:=B_{\boldsymbol\beta}\left(x_0,\frac12 d\right)$.

Case 1: $d_{\boldsymbol\beta}(x_0,y_0)<\frac12 d$.

Case 1a: $B_{\boldsymbol\beta}(x_0,d)\cap\mathcal{S}=\varnothing$. We introduce smooth complex coordinates $\{w_1=z_1^{\beta_1},w_2=z_2^{\beta_2},z_3,\dots,z_n\}$ on $B_{\boldsymbol\beta}(x_0,d)$, under which $g_{\boldsymbol\beta}$ becomes the Euclidean metric and the components of g become C^α in the usual sense. Equation (3-110) has C^α leading coefficients, and we can apply Theorem 6.6 in [18] to conclude that (the following inequality is understood in the new coordinates)

$$[u]^*_{C^{2,\alpha}(B)}\le C(\|u\|_{C^0(B)}+\|f\|^{(2)}_{C^{0,\alpha}(B)}).\tag{3-112}$$

Recall that T denotes the second-order operators appearing in (2-2). Let D denote the ordinary first-order operators in $\{w_1,w_2,z_3,\dots,z_n\}$. We calculate

$$\begin{aligned}|Tu(x_0)-Tu(y_0)|&\le|D^2u(x_0)-D^2u(y_0)|+\frac{d_{\boldsymbol\beta}(x_0,y_0)}{d}(|D^2u(x_0)|+|D^2u(y_0)|)\\&\le\frac{4d_{\boldsymbol\beta}(x_0,y_0)^\alpha}{d^{2+\alpha}}[u]^*_{C^{2,\alpha}(B)}+\frac{4d_{\boldsymbol\beta}(x_0,y_0)}{d^3}[u]^*_{C^2(B)}\\&\le\frac{8d_{\boldsymbol\beta}(x_0,y_0)^\alpha}{d^{2+\alpha}}[u]^*_{C^{2,\alpha}(B)}+C\frac{d_{\boldsymbol\beta}(x_0,y_0)^\alpha}{d^{2+\alpha}}\|u\|_{C^0(B)}\ \text{(by the interpolation inequality).}\end{aligned}$$

Then we get

$$d_{x_0}^{2+\alpha}\frac{|Tu(x_0)-Tu(y_0)|}{d_{\beta}(x_0,y_0)^{\alpha}} \le \frac{C}{\mu^{2+\alpha}}\|f\|^{(2)}_{C^{0,\alpha}_{\beta}(B)} + \frac{C}{\mu^{2+\alpha}}\|u\|_{C^0(B)}. \tag{3-113}$$

Case 1b: $B_{\beta}(x_0,d)\cap\mathcal{S}\neq\varnothing$. Let $\hat{x}_0\in\mathcal{S}$ be the nearest possible point x_0 to $\mathcal{S}$. We consider the ball $\hat{B}:=B_{\beta}(\hat{x}_0,2d)$, which is contained in $B_{\beta}(0,1)$ by the triangle inequality. As in [14], we introduce a (nonholomorphic) basis of $T^*_{1,0}(\mathbb{C}^n\backslash\mathcal{S})$

$$\{\epsilon_j := dr_j+\sqrt{-1}\beta_j r_j\,d\theta_j,\, dz_k\}_{j=1,2;\,k=3,\dots,n},$$

and the dual basis of $T_{1,0}(\mathbb{C}^n\backslash\mathcal{S})$

$$\left\{\gamma_j := \frac{\partial}{\partial r_j}-\sqrt{-1}\frac{1}{\beta_j r_j}\frac{\partial}{\partial\theta_j},\, \frac{\partial}{\partial z_k}\right\}_{j=1,2;\,k=3,\dots,n}.$$

We can write the $(1,1)$-form ω in the basis $\{\epsilon_j\wedge\bar{\epsilon}_k,\, \epsilon_j\wedge d\bar{z}_k,\, dz_k\wedge\bar{\epsilon}_j,\, dz_j\wedge d\bar{z}_k\}$ as

$$\omega = g_{\epsilon_j\bar{\epsilon}_k}\epsilon_j\wedge\bar{\epsilon}_k + g_{\epsilon_j\bar{k}}\epsilon_j\wedge d\bar{z}_k + g_{k\bar{\epsilon}_j}\,dz_k\wedge\bar{\epsilon}_j + g_{j\bar{k}}\,dz_j\wedge d\bar{z}_k, \tag{3-114}$$

where

$$\begin{aligned} g_{\epsilon_j\bar{\epsilon}_k} &= \sqrt{-1}\partial\bar{\partial}\phi(\gamma_j,\bar{\gamma}_k), & g_{\epsilon_j\bar{k}} &= \sqrt{-1}\partial\bar{\partial}\phi\left(\gamma_j,\frac{\partial}{\partial\bar{z}_k}\right),\\ g_{k\bar{\epsilon}_j} &= \sqrt{-1}\partial\bar{\partial}\phi\left(\frac{\partial}{\partial z_k},\bar{\gamma}_j\right), & g_{k\bar{j}} &= \frac{\partial^2}{\partial z_k\partial\bar{z}_j}\phi. \end{aligned} \tag{3-115}$$

We remark that all the second-order derivatives of ϕ appearing in (3-115) are linear combinations of $|z_j|^{2-2\beta_j}(\partial^2/\partial z_j\partial z_{\bar{j}})N_jN_k$ $(j\neq k)$, N_jD' and $(D')^2$, which are studied in Theorem 1.2. The standard metric ω_{β} becomes the identity matrix under the basis above for $(1,1)$-forms. If ω is $C^{0,\alpha}_{\beta}$, all the coefficients in the expression for ω in (3-114) are $C^{0,\alpha}_{\beta}$-continuous, and the cross terms $g_{\epsilon_j\bar{\epsilon}_k}$ $(j\neq k)$ and $g_{\epsilon_j\bar{k}}$ tend to zero when approaching the corresponding singular sets $\mathcal{S}_j$ or $\mathcal{S}_k$. Moreover, the limit of $g_{j\bar{k}}\,dz_j\wedge d\bar{z}_k$ when approaching $\mathcal{S}_1\cap\cdots\cap\mathcal{S}_p$ defines a Kähler metric on it. Rescaling or rotating the coordinates if necessary we may assume at $\hat{x}_0\in\mathcal{S}$ that $g_{\epsilon_j\bar{\epsilon}_j}(\hat{x}_0)=1$, $g_{j\bar{k}}(\hat{x}_0)=\delta_{jk}$ and the cross terms vanish at $\hat{x}_0$. Let ω_{β} be the standard cone metric under these new coordinates near $\hat{x}_0$. We can rewrite (3-110) as

$$\Delta_g u(z) = \Delta_{g_{\beta}}u(z)+\eta(z)\cdot i\partial\bar{\partial}u(z) = f(z) \quad \text{for all } z\notin\mathcal{S}$$

for some hermitian matrix $\eta(z)=(\eta^{j\bar{k}})^n_{j,k=1}$, $\eta^{j\bar{k}}=g^{j\bar{k}}(z)-g^{j\bar{k}}_{\beta}$. It is not hard to see the term $\eta(z)\cdot i\partial\bar{\partial}u$ can be written as

$$\sum_{j,k=1}^{2}(g^{\epsilon_j\bar{\epsilon}_k}(z)-\delta_{jk})u_{\epsilon_j\bar{\epsilon}_j} + 2Re\sum_{\substack{1\le j\le 2\\ 3\le k\le n}} g^{\epsilon_j\bar{k}}u_{\epsilon_j\bar{k}} + \sum_{j,k=3}^{n}(g^{j\bar{k}}(z)-\delta_{jk})u_{j\bar{k}}, \tag{3-116}$$

where g with the upper indices denotes the inverse matrix of g. We consider the equivalent form of (3-110) on $\hat{B}$:

$$\Delta_{g_{\beta}}u = f-\eta\cdot\sqrt{-1}\partial\bar{\partial}u =: \hat{f}, \quad u\in C^0(\hat{B})\cap C^2(\hat{B}\backslash\mathcal{S}).$$

Observing that $x_0, y_0 \in B_{\boldsymbol\beta}\big(\hat x_0, \frac32 d\big)$, we can apply the scaled inequality (1-7) of Theorem 1.2 to conclude that

$$d^{2+\alpha}\frac{|Tu(x_0)-Tu(y_0)|}{d_{\boldsymbol\beta}(x_0,y_0)^\alpha}\le C(\|u\|_{C^0(\hat B)}+\|\hat f\|^{(2)}_{C^{0,\alpha}_{\boldsymbol\beta}(\hat B)});$$

thus

$$d^{2+\alpha}_{x_0}\frac{|Tu(x_0)-Tu(y_0)|}{d_{\boldsymbol\beta}(x_0,y_0)^\alpha}\le \frac{C}{\mu^{2+\alpha}}(\|u\|_{C^0(\hat B)}+\|\hat f\|^{(2)}_{C^{0,\alpha}_{\boldsymbol\beta}(\hat B)}). \tag{3-117}$$

<u>Case 2</u>: $d_{\boldsymbol\beta}(x_0,y_0)\ge\frac12 d$.

$$d^{2+\alpha}_{x_0}\frac{|Tu(x_0)-Tu(y_0)|}{d_{\boldsymbol\beta}(x_0,y_0)^\alpha}\le 4d^{2+\alpha}_{x_0}\frac{|Tu(x_0)|+|Tu(y_0)|}{d^\alpha}\le\frac{8}{\mu^\alpha}[u]^*_{C^2_{\boldsymbol\beta}(B_{\boldsymbol\beta}(0,1))}. \tag{3-118}$$

Combining (3-113), (3-117) and (3-118) we get

$$\begin{aligned}d^{2+\alpha}_{x_0}\frac{|Tu(x_0)-Tu(y_0)|}{d_{\boldsymbol\beta}(x_0,y_0)^\alpha}\le\frac{8}{\mu^\alpha}[u]^*_{C^2_{\boldsymbol\beta}(B_{\boldsymbol\beta}(0,1))}&+\frac{C}{\mu^{2+\alpha}}(\|u\|_{C^0(\hat B)}+\|\hat f\|^{(2)}_{C^{0,\alpha}_{\boldsymbol\beta}(\hat B)})\\&+\frac{C}{\mu^{2+\alpha}}\|f\|^{(2)}_{C^{0,\alpha}_{\boldsymbol\beta}(B)}+\frac{C}{\mu^{2+\alpha}}\|u\|_{C^0(B)}.\end{aligned} \tag{3-119}$$

By definition it is easy to see that (writing $B_{\boldsymbol\beta}=B_{\boldsymbol\beta}(0,1)$)

$$\|f\|^{(2)}_{C^{0,\alpha}_{\boldsymbol\beta}(B)}\le C\mu^2\|f\|^{(2)}_{C^0(B_{\boldsymbol\beta})}+C\mu^{2+\alpha}[f]^{(2)}_{C^{0,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta})}\le\mu^2\|f\|^{(2)}_{C^{0,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta})}.$$

We calculate

$$\begin{aligned}\|\hat f\|^{(2)}_{C^{0,\alpha}_{\boldsymbol\beta}(\hat B)}&\le\|\eta\|^{(0)}_{C^{0,\alpha}_{\boldsymbol\beta}(\hat B)}\|Tu\|^{(2)}_{C^{0,\alpha}_{\boldsymbol\beta}(\hat B)}+\|f\|^{(2)}_{C^{0,\alpha}_{\boldsymbol\beta}(\hat B)}\\&\le C_0[g]^*_{C^{0,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta})}\mu^\alpha(\mu^2[u]^*_{C^2_{\boldsymbol\beta}(B_{\boldsymbol\beta})}+\mu^{2+\alpha}[u]^*_{C^{2,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta})})+\mu^2\|f\|^{(2)}_{C^{0,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta})}\\&\le C_0[g]^*_{C^{0,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta})}\mu^\alpha(C(\mu)\|u\|_{C^0(B_{\boldsymbol\beta})}+2\mu^{2+\alpha}[u]^*_{C^{2,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta})})+\mu^2\|f\|^{(2)}_{C^{0,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta})},\end{aligned}$$

$$\frac{8}{\mu^\alpha}[u]^*_{C^2_{\boldsymbol\beta}(B_{\boldsymbol\beta})}\le\mu^\alpha[u]^*_{C^{2,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta})}+C(\mu)\|u\|_{C^0(B_{\boldsymbol\beta})}.$$

If we choose $\mu>0$ small enough that $\mu^\alpha(2C_0[g]^*_{C^{0,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta})}+1)\le\frac12$, then we get from (3-119) and the inequalities above that

$$d^{2+\alpha}_{x_0}\frac{|Tu(x_0)-Tu(y_0)|}{d_{\boldsymbol\beta}(x_0,y_0)^\alpha}\le\tfrac12[u]^*_{C^{2,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta})}+C(\mu)(\|u\|_{C^0(B_{\boldsymbol\beta})}+\|f\|^{(2)}_{C^{0,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta})}).$$

Taking the supremum over $x_0\ne y_0\in B_{\boldsymbol\beta}(0,1)$, we conclude from the inequality above that

$$[u]^*_{C^{2,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta})}\le C(\|u\|_{C^0(B_{\boldsymbol\beta})}+\|f\|^{(2)}_{C^{0,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta})}).$$

Proposition 3.32 then follows from interpolation inequalities. □

Remark 3.33. It follows easily from the proof of Proposition 3.32 that estimate (3-111) also holds for metric balls $B_{\boldsymbol\beta}(p,R)\subset B_{\boldsymbol\beta}(0,1)$ whose center p may not lie in $\mathcal S$.

Remark 3.34. The Schauder estimate was first established by Donaldson [14] for a background cone metric with singularity along a smooth divisor assuming $u \in C^{2,\alpha}_{\beta}$; this latter assumption was removed by Brendle [1] in the case $\beta \in \left(0, \frac{1}{2}\right)$ and by Jeffres, Mazzeo and Rubinstein [23], requiring only a weak solution. Jeffres, Mazzeo and Rubinstein [23] then extend the results to nonflat background metrics using a perturbation argument. This is the first time a Schauder estimate for the linear conic equation in the smooth divisor case appeared in the literature with a full proof.

An immediate corollary to Proposition 3.32 is the following interior Schauder estimate.

Corollary 3.35. *Suppose u satisfies* (3-110). *For any compact subset $K \Subset B_{\boldsymbol{\beta}}(0, 1)$, there exists a constant $C = C(n, \boldsymbol{\beta}, K, \|g\|_{C^{0,\alpha}_{\boldsymbol{\beta}}(B_{\boldsymbol{\beta}}(0,1))}) > 0$ such that*

$$\|u\|_{C^{2,\alpha}_{\boldsymbol{\beta}}(K)} \leq C(\|u\|_{C^0(B_{\boldsymbol{\beta}}(0,1))} + \|f\|_{C^{0,\alpha}_{\boldsymbol{\beta}}(B_{\boldsymbol{\beta}}(0,1))}).$$

Next we will show that (3-110) admits a unique $C^{2,\alpha}_{\boldsymbol{\beta}}$-solution for any $f \in C^{0,\alpha}_{\boldsymbol{\beta}}(\overline{B_{\boldsymbol{\beta}}(0, 1)})$ and boundary value $\varphi \in C^0(\partial B_{\boldsymbol{\beta}}(0, 1))$. We will follow the argument in Section 6.5 in [18]. In the following we write $B_{\boldsymbol{\beta}} = B_{\boldsymbol{\beta}}(0, 1)$ for simplicity.

Lemma 3.36. *Let $\sigma \in (0, 1)$ be a given number. Suppose $u \in C^{2,\alpha}_{\boldsymbol{\beta}}(B_{\boldsymbol{\beta}})$ solves* (3-110), *$\|u\|^{(-\sigma)}_{C^0(B_{\boldsymbol{\beta}})} < \infty$ and $\|f\|^{(2-\sigma)}_{C^{0,\alpha}_{\boldsymbol{\beta}}(B_{\boldsymbol{\beta}})} < \infty$. Then there exists a $C = C(n, \boldsymbol{\beta}, \alpha, g, \sigma) > 0$ such that*

$$\|u\|^{(-\sigma)}_{C^{2,\alpha}_{\boldsymbol{\beta}}(B_{\boldsymbol{\beta}})} \leq C(\|u\|^{(-\sigma)}_{C^0(B_{\boldsymbol{\beta}})} + \|f\|^{(2-\sigma)}_{C^{0,\alpha}_{\boldsymbol{\beta}}(B_{\boldsymbol{\beta}})}).$$

Proof. Given the estimates in Proposition 3.32, the proof is identical to that of Lemma 6.20 in [18]. We omit the details. □

Lemma 3.37. *Let $u \in C^2_{\boldsymbol{\beta}}(B_{\boldsymbol{\beta}}) \cap C^0(\overline{B_{\boldsymbol{\beta}}})$ solve the equation $\Delta_g u = f$ and $u \equiv 0$ on $\partial B_{\boldsymbol{\beta}}$. For any $\sigma \in (0, 1)$, there exists a constant $C = C(n, \boldsymbol{\beta}, \sigma, g) > 0$ such that*

$$\|u\|^{(-\sigma)}_{C^0(B_{\boldsymbol{\beta}})} = \sup_{x \in B_{\boldsymbol{\beta}}} d_x^{-\sigma} |u(x)| \leq C \sup_{x \in B_{\boldsymbol{\beta}}} d_x^{2-\sigma} |f(x)| = C\|f\|^{(2-\sigma)}_{C^0(B_{\boldsymbol{\beta}})},$$

where $d_x = d_{\boldsymbol{\beta}}(x, \partial B_{\boldsymbol{\beta}})$ as before.

Proof. Consider the function $w_1 = (1 - d^2_{\boldsymbol{\beta}})^\sigma$, where $d_{\boldsymbol{\beta}}(x) = d_{\boldsymbol{\beta}}(x, 0)$. We calculate

$$\begin{aligned}\Delta_g w_1 &= \sigma(1 - d^2_{\boldsymbol{\beta}})^{\sigma-2}(-(1 - d^2_{\boldsymbol{\beta}}) \operatorname{tr}_g g_{\boldsymbol{\beta}} - (1 - \sigma)|\nabla d^2_{\boldsymbol{\beta}}|^2_g) \\ &\leq \sigma(1 - d^2_{\boldsymbol{\beta}})^{\sigma-2}(-C^{-1}(1 - d^2_{\boldsymbol{\beta}}) - 4C^{-1} d^2_{\boldsymbol{\beta}}(1 - \sigma)) \leq -c_0 \sigma(1 - d^2_{\boldsymbol{\beta}})^{\sigma-2}.\end{aligned}$$

Take a large constant $A > 1$ such that, for $w = A w_1$,

$$\Delta_g w \leq -(1 - d_{\boldsymbol{\beta}})^{\sigma-2} \leq -\frac{|f|}{N} \quad \text{in } B_{\boldsymbol{\beta}},$$

where

$$N = \sup_{x \in B_{\boldsymbol{\beta}}} d_x^{2-\sigma} |f(x)| = \sup_{x \in B_{\boldsymbol{\beta}}} (1 - d_{\boldsymbol{\beta}}(x))^{2-\sigma} |f(x)|.$$

Hence $\Delta_g(Nw \pm u) \leq 0$, and from the definition of w we also have $w|_{\partial B_{\boldsymbol{\beta}}} \equiv 0$. By the maximum principle we obtain $|u(x)| \leq Nw \leq CN(1 - d_{\boldsymbol{\beta}}(x))^\sigma = CN d_x^\sigma$, and hence the lemma is proved. □

Proposition 3.38. *Given any function $f \in C^{0,\alpha}_{\beta}(\overline{B_\beta})$, the Dirichlet problem $\Delta_g u = f$ in B_β and $u \equiv 0$ on ∂B_β admits a unique solution $u \in C^{2,\alpha}_{\beta}(B_\beta) \cap C^0(\overline{B_\beta})$.*

Proof. The proof of this proposition is almost identical to that of Theorem 6.22 in [18]. For completeness, we provide the detailed argument. Fix $\sigma \in (0, 1)$ and define a family of operators $\Delta_t = t\Delta_g + (1-t)\Delta_{g_\beta}$. It is straightforward to see that Δ_t is associated to some cone metric which also satisfies (3-109). We study the Dirichlet problem

$$\Delta_t u_t = f \quad \text{in } B_\beta, \qquad u_t \equiv 0 \quad \text{on } \partial B_\beta. \tag{$*_t$}$$

Equation $(*_0)$ admits a unique solution $u_0 \in C^{2,\alpha}_{\beta}(B_\beta) \cap C^0(\overline{B_\beta})$ by Proposition 3.7. By Theorem 5.2 in [18], in order to apply the continuity method to solve $(*_1)$, it suffices to show Δ_t^{-1} defines a bounded linear operator between some Banach spaces. More precisely, define

$$\mathcal{B}_1 := \{u \in C^{2,\alpha}_{\beta}(B_\beta) \mid \|u\|^{(-\sigma)}_{C^{2,\alpha}_{\beta}(B_\beta)} < \infty\},$$
$$\mathcal{B}_2 := \{f \in C^{0,\alpha}_{\beta}(B_\beta) \mid \|f\|^{(2-\sigma)}_{C^{0,\alpha}_{\beta}(B_\beta)} < \infty\}.$$

By definition any $u \in \mathcal{B}_1$ is continuous on $\overline{B_\beta}$ and $u = 0$ on ∂B_β. By Lemmas 3.36 and 3.37, we have

$$\|u\|_{\mathcal{B}_1} = \|u\|^{(-\sigma)}_{C^{2,\alpha}_{\beta}(B_\beta)} \le C\|f\|^{(2-\sigma)}_{C^{0,\alpha}_{\beta}(B_\beta)} = C\|\Delta_t u\|_{\mathcal{B}_2},$$

for some constant C independent of $t \in [0, 1]$. Thus $(*_1)$ admits a solution $u \in \mathcal{B}_1$. □

Corollary 3.39. *For any given $\varphi \in C^0(\partial B_\beta)$ and $f \in C^{0,\alpha}_{\beta}(\overline{B_\beta})$, the Dirichlet problem*

$$\Delta_g u = f \quad \text{in } B_\beta \qquad \text{and} \qquad u = \varphi \quad \text{on } \partial B_\beta, \tag{3-120}$$

admits a unique solution $u \in C^{2,\alpha}_{\beta}(B_\beta) \cap C^0(\overline{B_\beta})$.

Proof. We may extend φ continuously to B_β and assume $\varphi \in C^0(\overline{B_\beta})$. Take a sequence of functions $\varphi_k \in C^{2,\alpha}_{\beta}(\overline{B_\beta}) \cap C^0(\overline{B_\beta})$ which converges uniformly to φ on $\overline{B_\beta}$. The Dirichlet problem

$$\Delta_g v_k = f - \Delta_g \varphi_k \quad \text{in } B_\beta \qquad \text{and} \qquad v_k = 0 \quad \text{on } \partial B_\beta$$

admits a unique solution $v_k \in C^{2,\alpha}_{\beta}(B_\beta) \cap C^0(\overline{B_\beta})$. Thus the function $u_k := v_k + \varphi_k \in C^{2,\alpha}_{\beta}$ satisfies $\Delta_g u_k = f$ in B_β and $u_k = \varphi_k$ on ∂B_β. By the maximum principle, u_k is uniformly bounded in $C^0(\overline{B_\beta})$. Corollary 3.35 gives uniformly $C^{2,\alpha}_{\beta}(K)$-bounds on any compact subset $K \Subset B_\beta$. Letting $k \to \infty$ and $K \to B_\beta$, by a diagonal argument and up to a subsequence, $u_k \to u \in C^{2,\alpha}_{\beta}(B_\beta)$. On the other hand, from $\Delta_g(u_k - u_l) = 0$, we see that $\{u_k\}$ is a Cauchy sequence in $C^0(\overline{B_\beta})$; thus u_k converges uniformly to u on $\overline{B_\beta}$. Hence $u \in C^0(\overline{B_\beta})$, and u satisfies (3-120). □

Corollary 3.40. *Given $f \in C^{0,\alpha}_{\beta}(B_\beta)$, suppose u is a weak solution to the equation $\Delta_g u = f$ in the sense that*

$$\int_{B_\beta} \langle \nabla u, \nabla \varphi \rangle \omega_g^n = -\int_{B_\beta} f\varphi\omega_g^n \quad \textit{for all } \varphi \in H^1_0(B_\beta),$$

then $u \in C^{2,\alpha}_{\beta}(B_\beta)$.

Proof. We first observe that the Sobolev inequality (3-43) also holds for the metric g, since g is equivalent to $g_{\boldsymbol\beta}$. The metric space $(B_{\boldsymbol\beta}, g)$ also has maximal volume growth/decay, so we can apply the same proof of De Giorgi–Nash–Moser theory [22] to conclude that u is continuous in $B_{\boldsymbol\beta}$. The standard elliptic theory implies that $u \in C^{2,\alpha}_{\mathrm{loc}}(B_{\boldsymbol\beta}\backslash\mathcal{S})$. For any $r \in (0, 1)$, by Corollary 3.39, the Dirichlet problem $\Delta_g \tilde{u} = f$ in $B_{\boldsymbol\beta}(0, r)$, $\tilde{u} = u$ on $\partial B_{\boldsymbol\beta}(0, r)$ admits a unique solution $\tilde{u} \in C^{2,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta}(0, r)) \cap C^0(\overline{B_{\boldsymbol\beta}(0, r)})$. Then $\Delta_g(u - \tilde{u}) = 0$ in $B_{\boldsymbol\beta}(0, r)$ and $u - \tilde{u} = 0$ on $\partial B_{\boldsymbol\beta}(0, r)$. By the maximum principle, we get $u = \tilde{u}$ in $B_{\boldsymbol\beta}(0, r)$, so we conclude $u \in C^{2,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta}(0, r))$. Since $r \in (0, 1)$ is arbitrary, we get $u \in C^{2,\alpha}_{\boldsymbol\beta}(B_{\boldsymbol\beta})$. □

Corollary 3.41. *Let X be a compact Kähler manifold and $D = \sum_j D_j$ be a divisor with simple normal crossings. Let g be a conical Kähler metric with cone angle $2\pi\boldsymbol\beta$ along D. Suppose $u \in H^1(g)$ is a weak solution to the equation $\Delta_g u = f$ in the sense that*

$$\int_X \langle \nabla u, \nabla\varphi\rangle \omega_g^n = -\int_X f\varphi\omega_g^n \quad \text{for all } \varphi \in C^1(X)$$

for some $f \in C^{0,\alpha}_{\boldsymbol\beta}(X)$. Then $u \in C^{2,\alpha}_{\boldsymbol\beta}(X) \cap C^0(X)$ and there exists a constant $C = C(n, \boldsymbol\beta, g, \alpha)$ such that

$$\|u\|_{C^{2,\alpha}_{\boldsymbol\beta}(X)} \le C(\|u\|_{C^0(X)} + \|f\|_{C^{0,\alpha}_{\boldsymbol\beta}(X)}).$$

Proof. We choose finite covers of D, $\{B_a\}$ and $\{B'_a\}$, with $B'_a \Subset B_a$ and centers in D. By assumption u is a weak solution to $\Delta_g u = f$ in each B_a, so by Corollary 3.40 we conclude that $u \in C^{2,\alpha}_{\boldsymbol\beta}(B_a)$ for each B_a. On $X\backslash\mathcal{S}$, the metric g is smooth so standard elliptic theory implies that $u \in C^{2,\alpha}_{\mathrm{loc}}(X\backslash\mathcal{S})$. Since $\{B_a\}$ covers D, we have $u \in C^{2,\alpha}_{\boldsymbol\beta}(X)$.

We can apply Corollary 3.35 to obtain that, for some constant $C > 0$,

$$\|u\|_{C^{2,\alpha}_{\boldsymbol\beta}(B'_a)} \le C(\|u\|_{C^0(B_a)} + \|f\|_{C^{\alpha}_{\boldsymbol\beta}(B_a)}).$$

On $X\backslash\bigcup_a\{B'_a\}$ the metric g is smooth, so the usual Schauder estimates apply. We finish the proof of the corollary using the definition of $C^{2,\alpha}_{\boldsymbol\beta}(X)$; see Definition 2.9. □

Remark 3.42. Let (X, D, g) be as in Corollary 3.41. It is easy to see by the variational method that weak solutions to $\Delta_g u = f$ always exist for any $f \in L^2(X, \omega_g^n)$ satisfying $\int_X f\omega_g^n = 0$.

4. Parabolic estimates

In this section, we will study the heat equation with background metric $\omega_{\boldsymbol\beta}$ and prove the Schauder estimates for solutions $u \in C^0(\mathcal{Q}_{\boldsymbol\beta}) \cap \mathscr{C}^{2,1}(\mathcal{Q}^\#_{\boldsymbol\beta})$ to the equation

$$\frac{\partial u}{\partial t} = \Delta_{g_{\boldsymbol\beta}} u + f \tag{4-1}$$

for a function $f \in \mathscr{C}^0(\mathcal{Q}_{\boldsymbol\beta})$ with some better regularity.

4A. ***Conical heat equations.*** In this section, we will show that, for any $\varphi \in \mathscr{C}^0(\partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol\beta})$, the Dirichlet problem (4-2) admits a unique $\mathscr{C}^{2,1}(\mathcal{Q}^\#_{\boldsymbol\beta}) \cap \mathscr{C}^0(\overline{\mathcal{Q}_{\boldsymbol\beta}})$-solution in $\mathcal{Q}_{\boldsymbol\beta}$. We first observe that a maximum principle argument yields the uniqueness of the solution.

Suppose $u \in \mathscr{C}^{2,1}(\mathcal{Q}_{\boldsymbol\beta}^{\#}) \cap \mathscr{C}^0(\overline{\mathcal{Q}_{\boldsymbol\beta}})$ solves the Dirichlet problem

$$\begin{cases} \dfrac{\partial u}{\partial t} = \Delta_{g_{\boldsymbol\beta}} u & \text{in } \mathcal{Q}_{\boldsymbol\beta}, \\ u = \varphi & \text{on } \partial_{\mathcal{P}} \mathcal{Q}_{\boldsymbol\beta} \end{cases} \tag{4-2}$$

for some given continuous function $\varphi \in \mathscr{C}^0(\partial_{\mathcal{P}} \mathcal{Q}_{\boldsymbol\beta})$. As in Lemma 3.1, it follows from the maximum principle that

$$\inf_{\partial_{\mathcal{P}} \mathcal{Q}_{\boldsymbol\beta}} u \le \inf_{\mathcal{Q}_{\boldsymbol\beta}} u \le \sup_{\mathcal{Q}_{\boldsymbol\beta}} u \le \sup_{\partial_{\mathcal{P}} \mathcal{Q}_{\boldsymbol\beta}} u. \tag{4-3}$$

So the $\mathscr{C}^{2,1}(\mathcal{Q}_{\boldsymbol\beta}^{\#}) \cap \mathscr{C}^0(\overline{\mathcal{Q}_{\boldsymbol\beta}})$-solution to (4-2) is unique, if it exists.

We prove the existence of solutions to (4-2). As before, we use an approximation argument. Let g_ϵ be the smooth approximation metrics in $B_{\boldsymbol\beta}$ from (3-3). Let u_ϵ be the $\mathscr{C}^{2,1}(\mathcal{Q}_{\boldsymbol\beta}) \cap \mathscr{C}^0(\overline{\mathcal{Q}_{\boldsymbol\beta}})$-solution to

$$\frac{\partial u_\epsilon}{\partial t} = \Delta_{g_\epsilon} u_\epsilon \quad \text{in } \mathcal{Q}_{\boldsymbol\beta} \qquad \text{and} \qquad u_\epsilon = \varphi \quad \text{on } \partial_{\mathcal{P}} \mathcal{Q}_{\boldsymbol\beta}. \tag{4-4}$$

4A1. *Estimates of u_ϵ.* We first recall the Li–Yau gradient estimates [26; 35] for positive solutions to the heat equations.

Lemma 4.1. *Let (M, g) be a complete manifold with $\mathrm{Ric}(g) \ge 0$ and $B(p, R)$ be the geodesic ball with center $p \in M$ and radius $R > 0$. Let u be a positive solution to the heat equation $\partial_t u - \Delta_g u = 0$ on $B(p, R)$. Then there exists $C = C(n) > 0$ such that, for all $t > 0$,*

$$\sup_{B(p,2R/3)} \left(\frac{|\nabla u|^2}{u^2} - \frac{2\dot u_t}{u} \right) \le \frac{C}{R^2} + \frac{2n}{t},$$

where $\dot u_t = \partial u / \partial t$.

By considering the functions $u_\epsilon - \inf u_\epsilon$ and $\sup u_\epsilon - u_\epsilon$, from Lemma 4.1, we see that there exists a constant $C = C(n) > 0$ such that, for any $R \in (0, 1)$ and $t \in (0, R^2)$,

$$\sup_{B_{g_\epsilon}(0,2R/3)} |\nabla u_\epsilon|^2_{g_\epsilon} \le C \left(\frac{1}{R^2} + \frac{1}{t} \right) (\mathrm{osc}_R\, u_\epsilon)^2, \tag{4-5}$$

$$\sup_{B_{g_\epsilon}(0,2R/3)} |\Delta_{g_\epsilon} u_\epsilon| = \sup_{B_{g_\epsilon}(0,2R/3)} \left| \frac{\partial u_\epsilon}{\partial t} \right| \le C \left(\frac{1}{R^2} + \frac{1}{t} \right) \mathrm{osc}_R\, u_\epsilon, \tag{4-6}$$

where $\mathrm{osc}_R\, u_\epsilon := \mathrm{osc}_{B_{g_\epsilon}(0,R) \times (0,R^2)}\, u_\epsilon$ is the oscillation of u_ϵ in the cylinder $B_{g_\epsilon}(0, R) \times (0, R^2)$. Replacing u_ϵ by $u_\epsilon - \inf u_\epsilon$, we may assume $u_\epsilon > 0$ and define $f_\epsilon = \log u_\epsilon$. Then we have

$$\frac{\partial f_\epsilon}{\partial t} = \Delta_{g_\epsilon} f_\epsilon + |\nabla f_\epsilon|^2.$$

Let $\varphi(x) = \varphi(r(x)/R)$, where φ is a cut-off function equal to 1 on $\left[0, \frac{3}{5}\right]$ and 0 on $\left[\frac{2}{3}, \infty\right)$ satisfying the inequalities $|\varphi''| \le 10$ and $(\varphi')^2 \le 10\varphi$. Let $r(x)$ be the distance function under g_ϵ to the center 0.

Lemma 4.2. *There exists a constant $C = C(n) > 0$ such that, for any small $\epsilon > 0$,*

$$\sup_{B_{g_\epsilon}(0,3R/5)} |\Delta_i u_\epsilon| \le C \left(\frac{1}{t} + \frac{1}{R^2} \right) \mathrm{osc}_R\, u_\epsilon \quad \textit{for all } t \in (0, R^2),$$

where we write $\Delta_i u_\epsilon := (|z_i|^2 + \epsilon)^{1-\beta_i} (\partial^2 u_\epsilon / \partial z_i \partial z_{\bar i})$ for $i = 1, \ldots, p$.

Proof. We only prove the case when $i = 1$. We define $F := t\varphi(-\Delta_1 f_\epsilon - 2\dot{f}_\epsilon)$, and we calculate

$$\begin{aligned}\Big(\frac{\partial}{\partial t} - \Delta_{g_\epsilon}\Big)(-\Delta_1 f_\epsilon - 2\dot{f}_\epsilon) &= -|\nabla_2\nabla f_\epsilon|^2 - |\nabla_1\overline{\nabla} f_\epsilon|^2 - 2\,\mathrm{Re}\langle\nabla f_\epsilon, \overline{\nabla}(-\Delta_1 f_\epsilon - 2\dot{f}_\epsilon)\rangle - R_{1\bar{1}j\bar{k}} f_{\epsilon,\bar{j}} f_{\epsilon,k}\\ &\le -(-\Delta_1 f_\epsilon)^2 - 2\,\mathrm{Re}\langle\nabla f_\epsilon, \overline{\nabla}(-\Delta_1 f_\epsilon - 2\dot{f}_\epsilon)\rangle.\end{aligned}$$

F achieves its maximum at a point (p_0, t_0), where we may assume $F(p_0, t_0) > 0$; otherwise we are already done. In particular, $p_0 \in B_{g_\epsilon}\big(0, \tfrac{2}{3}r\big)$ by the definition of φ and $t_0 > 0$. Then at (p_0, t_0), we have

$$\begin{aligned}0 &\le \Big(\frac{\partial}{\partial t} - \Delta_{g_\epsilon}\Big)F\\ &= \frac{F}{t_0} + t_0\varphi\Big(\frac{\partial}{\partial t} - \Delta_{g_\epsilon}\Big)(-\Delta_1 f_\epsilon - 2\dot{f}_\epsilon) - \frac{F}{\varphi}\Delta_{g_\epsilon}\varphi - 2t_0\,\mathrm{Re}\Big\langle\nabla\varphi, \overline{\nabla}\Big(\frac{F}{t_0\varphi}\Big)\Big\rangle\\ &\le \frac{F}{t_0} + t_0\varphi\Big(-(-\Delta_1 f_\epsilon)^2 - 2\frac{F}{t_0\varphi^2}\,\mathrm{Re}\langle\nabla f_\epsilon, \overline{\nabla}\varphi\rangle\Big) + C\frac{F}{R^2\varphi}(\varphi' + \varphi'') + 2\frac{F}{R^2\varphi^2}(\varphi')^2, \qquad (4\text{-}7)\end{aligned}$$

where we use the Laplacian comparison and the fact that $\nabla F = 0$ at (p_0, t_0). The second term on the right-hand side satisfies (we write $\widetilde{F} := -\Delta_1 f_\epsilon - 2\dot{f}_\epsilon$ for convenience of notation)

$$\begin{aligned}t_0\varphi\Big(-(-\Delta_1 f_\epsilon)^2 - 2\frac{F}{t_0\varphi^2}\,\mathrm{Re}\langle\nabla f_\epsilon, \overline{\nabla}\varphi\rangle\Big) &\le t_0\varphi\Big(-\widetilde{F}^2 - 4\widetilde{F}\dot{f}_\epsilon - 4(\dot{f}_\epsilon)^2 + \frac{2\widetilde{F}}{\varphi}\frac{|\nabla f_\epsilon||\varphi'|}{R}\Big)\\ &\le t_0\varphi\Big(-\widetilde{F}^2 - 4\widetilde{F}\dot{f}_\epsilon + 2\widetilde{F}|\nabla f_\epsilon|^2 + \frac{\widetilde{F}|\varphi'|^2}{2R^2\varphi^2}\Big)\\ &\le t_0\varphi\Big(-\widetilde{F}^2 + \frac{\widetilde{F}|\varphi'|^2}{2R^2\varphi^2} + C\frac{\widetilde{F}}{t_0} + C\frac{\widetilde{F}}{R^2}\Big) \qquad \text{(by Lemma 4.1)}\\ &= -\frac{F^2}{t_0\varphi} + C\frac{F}{2R^2\varphi} + C\frac{F}{t_0} + C\frac{F}{R^2}.\end{aligned}$$

Inserting this into (4-7), we get, for some constant $C = C(n) > 0$, at (p_0, t_0),

$$-F^2 + C\varphi F + \frac{t_0\varphi F}{R^2} + Ct_0\frac{F}{R^2} \ge 0,$$

from which we obtain $F(p_0, t_0) \le Ct_0/R^2 + C$. By the choice of (p_0, t_0), we can see that

$$\sup_{B_{g_\epsilon}(0,R/2)} (-\Delta_1 f_\epsilon - 2\dot{f}_\epsilon) \le C\Big(\frac{1}{R^2} + \frac{1}{t}\Big) \quad \text{for all } t \in (0, R^2),$$

which implies that, on $B_{g_\epsilon}\big(0, \tfrac{3}{5}R\big) \times (0, R^2)$,

$$-\Delta_1 u_\epsilon \le \dot{u}_\epsilon + C\Big(\frac{1}{t} + \frac{1}{R^2}\Big)u_\epsilon. \qquad (4\text{-}8)$$

Applying (4-8) to the function $\sup u_\epsilon - u_\epsilon$, we obtain, on $B_{g_\epsilon}\big(0, \tfrac{3}{5}R\big) \times (0, R^2)$,

$$|\Delta_1 u_\epsilon| \le |\dot{u}_\epsilon| + C\Big(\frac{1}{t} + \frac{1}{R^2}\Big)\,\mathrm{osc}_R\, u_\epsilon \le C\Big(\frac{1}{t} + \frac{1}{R^2}\Big)\,\mathrm{osc}_R\, u_\epsilon$$

by (4-6). Thus we finish the proof of the lemma. ☐

Lemma 4.3. *There exists a constant $C = C(n) > 0$ such that*

$$\sup_{i\neq j}\ \sup_{B_{g_\epsilon}(0,R/2)} (|\nabla_i\nabla_j u_\epsilon| + |\nabla_i\overline{\nabla}_j u_\epsilon|) \le C\left(\frac{1}{t}+\frac{1}{R^2}\right)\operatorname{osc}_R u_\epsilon$$

for all $t\in(0,R^2)$. Recall, here $|\nabla_i\nabla_j u_\epsilon|^2 = \nabla_i\nabla_j u_\epsilon \nabla_{\bar i}\nabla_{\bar j} u_\epsilon g_\epsilon^{i\bar i} g_\epsilon^{j\bar j}$ (no summation over i, j is taken).

Proof. We only prove the estimate for $|\nabla_1\nabla_2 u_\epsilon|$. The other estimates are similar, so we omit their proofs.

By calculations similar to those used to derive (3-27), we have

$$\left(\frac{\partial}{\partial t}-\Delta_{g_\epsilon}\right)|\nabla_1\nabla_2 f_\epsilon| \le 2\operatorname{Re}\langle\nabla f_\epsilon, \overline{\nabla}|\nabla_1\nabla_2 f_\epsilon|\rangle + \sum_k (|\nabla_1\nabla_k f_\epsilon||\nabla_2\nabla_{\bar k} f_\epsilon| + |\nabla_2\nabla_k f_\epsilon||\nabla_1\nabla_{\bar k} f_\epsilon|), \tag{4-9}$$

and similar to (3-20),

$$\begin{aligned}\left(\frac{\partial}{\partial t}-\Delta_{g_\epsilon}\right)(-\Delta_1 f_\epsilon - \Delta_2 f_\epsilon) &\le 2\operatorname{Re}\langle\nabla f_\epsilon, \overline{\nabla}(-\Delta_1 f_\epsilon - \Delta_2 f_\epsilon)\rangle \\ &\quad - \sum_k (|\nabla_1\nabla_k f_\epsilon|^2 + |\nabla_1\nabla_{\bar k} f_\epsilon|^2 + |\nabla_2\nabla_k f_\epsilon|^2 + |\nabla_2\nabla_{\bar k} f_\epsilon|^2). \end{aligned}\tag{4-10}$$

Combining (4-10), (4-9) and the Cauchy–Schwarz inequality, we get

$$\begin{aligned}&\left(\frac{\partial}{\partial t}-\Delta_{g_\epsilon}\right)(|\nabla_1\nabla_2 f_\epsilon| + 2(-\Delta_1 f_\epsilon - \Delta_2 f_\epsilon))\\ &\le 2\operatorname{Re}\langle\nabla f_\epsilon, \overline{\nabla}(|\nabla_1\nabla_2 f_\epsilon| + 2(-\Delta_1 f_\epsilon - \Delta_2 f_\epsilon))\rangle - \sum_k (|\nabla_1\nabla_k f_\epsilon|^2 + |\nabla_1\nabla_{\bar k} f_\epsilon|^2 + |\nabla_2\nabla_k f_\epsilon|^2 + |\nabla_2\nabla_{\bar k} f_\epsilon|^2)\\ &\le 2\operatorname{Re}\langle\nabla f_\epsilon, \overline{\nabla}(|\nabla_1\nabla_2 f_\epsilon| + 2(-\Delta_1 f_\epsilon - \Delta_2 f_\epsilon))\rangle - \tfrac{1}{10}(|\nabla_1\nabla_2 f_\epsilon| + 2(-\Delta_1 f_\epsilon - \Delta_2 f_\epsilon))^2.\end{aligned}$$

We define a cut-off function η similar to φ in the proof of Lemma 4.2 such that $\eta = 1$ on $B_{g_\epsilon}\big(0, \frac{1}{2}R\big)$ and η vanishes outside $B_{g_\epsilon}\big(0, \frac{3}{5}R\big)$. We write

$$G = t\eta(|\nabla_1\nabla_2 f_\epsilon| + 2(-\Delta_1 f_\epsilon - \Delta_2 f_\epsilon) - 2\dot f_\epsilon).$$

Like we did for F in the proof of Lemma 4.2, we argue similarly that at the maximum point (p_0, t_0) of G, for which we assume $G(p_0, t_0) > 0$,

$$\begin{aligned}0 \le \left(\frac{\partial}{\partial t}-\Delta_{g_\epsilon}\right)G &\le \frac{G}{t_0} - \frac{G^2}{t_0\eta} + C\frac{G}{R^2\eta} + C\frac{G}{t_0} + C\frac{G}{R^2} + C\frac{G}{R^2}\frac{\eta'+\eta''}{\eta} + \frac{2G}{R^2\eta^2}(\eta')^2\\ &\le \frac{1}{t_0\eta}\left(-G^2 + C\eta G + \frac{t_0\eta G}{R^2} + Ct_0\frac{G}{R^2}\right),\end{aligned}$$

so it follows that $G(p_0, t_0) \le C(1 + t_0/R^2)$. Therefore by the definition of G, on $B_{g_\epsilon}\big(0, \frac{1}{2}R\big)\times(0, R^2)$,

$$|\nabla_1\nabla_2 f_\epsilon| + 2(-\Delta_1 f_\epsilon - \Delta_2 f_\epsilon) - 2\dot f_\epsilon \le C\left(\frac{1}{R^2}+\frac{1}{t}\right),$$

and thus by Lemmas 4.1 and 4.2, we conclude that, on $B_{g_\epsilon}\big(0, \frac{1}{2}R\big)\times(0, R^2)$,

$$|\nabla_1\nabla_2 u_\epsilon| \le \dot u_\epsilon + 2|\Delta_1 u_\epsilon| + 2|\Delta_2 u_\epsilon| + \frac{|\nabla u_\epsilon|^2}{u_\epsilon} + Cu_\epsilon\left(\frac{1}{R^2}+\frac{1}{t}\right) \le C\left(\frac{1}{t}+\frac{1}{R^2}\right)\operatorname{osc}_R u_\epsilon,$$

as desired. □

4A2. *Existence of a solution u to* (4-2). We will show the limit function of u_ϵ as $\epsilon \to 0$ solves (4-2).

Proposition 4.4. *Given any $R \in (0,1)$ and any $\varphi \in \mathscr{C}^0(\partial_{\mathcal{P}} \mathcal{Q}_{\boldsymbol\beta}(0,R))$, there exists a unique function $u \in \mathscr{C}^{2,1}(\mathcal{Q}_{\boldsymbol\beta}(0,R)^{\#}) \cap C^0(\overline{\mathcal{Q}_{\boldsymbol\beta}(0,R)})$ solving* (4-2). *Moreover, there exists a constant $C = C(n,\boldsymbol\beta) > 0$ such that, for any $t \in (0,R^2)$ we have* (*defining $B_{\boldsymbol\beta}(r)^{\#} := B_{\boldsymbol\beta}(0,r)\backslash\mathcal{S}$*),

$$\sup_{B_{\boldsymbol\beta}(R/2)^{\#}} \left(\sum_{j=1}^{p} |z_j|^{2-2\beta_j} \left| \frac{\partial u}{\partial z_j} \right|^2 + |D'u|^2 \right) \le C\left(\frac{1}{t} + \frac{1}{R^2}\right)(\operatorname{osc}_R u)^2, \tag{4-11}$$

$$\sup_{B_{\boldsymbol\beta}(R/2)^{\#}} \left(\sum_{i\neq j} (|\nabla_i \nabla_j u|_{g_{\boldsymbol\beta}} + |\nabla_i \overline{\nabla}_j u|_{g_{\boldsymbol\beta}}) + \left| \frac{\partial u}{\partial t} \right| \right) \le C\left(\frac{1}{t} + \frac{1}{R^2}\right)\operatorname{osc}_R u, \tag{4-12}$$

$$\sup_{B_{\boldsymbol\beta}(R/2)^{\#}} \left(\sum_{j=1}^{p} |\nabla_{g_{\boldsymbol\beta}} \Delta_j u| + |\nabla_{g_{\boldsymbol\beta}} (D')^2 u| + \left| \nabla_{g_{\boldsymbol\beta}} \frac{\partial u}{\partial t} \right| \right) \le C\left(\frac{1}{t} + \frac{1}{R^2}\right)^{3/2}\operatorname{osc}_R u, \tag{4-13}$$

where by abusing notation we write $\operatorname{osc}_R u := \operatorname{osc}_{B_{\boldsymbol\beta}(0,R)\times(0,R^2)} u$.

Proof. Let u_ϵ be the $\mathscr{C}^{2,1}$-solution to (4-4). The $\mathscr{C}^0$-norm of u_ϵ follows from the maximum principle (4-3).

To prove the higher-order estimates, for any fixed compact subset $K \Subset B_{\boldsymbol\beta}(0,R)$ and $\delta > 0$, standard parabolic Schauder theory yields uniform $\mathscr{C}^{4+\alpha,(4+\alpha)/2}$-estimates of u_ϵ on $(K\backslash T_\delta \mathcal{S}) \times (\delta, R^2]$ for any $\alpha \in (0,1)$. As $\epsilon \to 0$, u_ϵ converges in $\mathscr{C}^{4+\alpha,(4+\alpha)/2}(K\backslash T_\delta\mathcal{S} \cap (\delta,R^2])$ to some function u which is also $\mathscr{C}^{4+\alpha,(4+\alpha)/2}$ in $(K\backslash T_\delta\mathcal{S}) \times (\delta,R^2]$. Letting $\delta \to 0$ and $K \to B_{\boldsymbol\beta}(0,R)$ and using a diagonal argument, we can assume that

$$u_\epsilon \xrightarrow{\mathscr{C}^{4+\alpha,(4+\alpha)/2}_{\mathrm{loc}}(B_{\boldsymbol\beta}(0,R)^{\#}\times(0,R^2])} u \quad \text{as } \epsilon \to 0.$$

Letting $\epsilon \to 0$, estimate (4-11) follows from (4-5); (4-12) is a consequence of Lemma 4.3; and (4-13) follows by applying the gradient estimate (4-5) to the Δ_{g_ϵ}-harmonic functions $\Delta_j u_\epsilon$, $(D')^2 u_\epsilon$ and $\partial u_\epsilon/\partial t$, and then letting $\epsilon \to 0$.

The gradient estimate (4-11) implies that, for any compact $K \Subset B_{\boldsymbol\beta}(0,R)$,

$$\sup_{K\backslash\mathcal{S}_j} \left| \frac{\partial u}{\partial z_j} \right| \le \frac{C(n,K,\boldsymbol\beta)(\operatorname{osc}_R u)^2}{t} |z_j|^{\beta_j - 1} \quad \text{for all } t \in (0,R^2).$$

From this, for any $t \in (0,R^2)$, we see that $u(\,\cdot\,,t)$ can be continuously extended to $\mathcal{S}$, and thus we have $u \in C^0(B_{\boldsymbol\beta}(0,R) \times (0,R^2))$.

It only remains to show $u = \varphi$ on $\partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol\beta}(0,R)$. Fix an arbitrary point $(q_0,t_0) \in \partial_{\mathcal{P}}(\mathcal{Q}_{\boldsymbol\beta}(0,R))$.

<u>Case 1</u>: $t_0 = 0$ and $q_0 \in \overline{B_{\boldsymbol\beta}(0,R)}$. We define a barrier function $\phi_1(z,t) = e^{-d_{\mathbb{C}^n}(z,q_0)^2 - \lambda t} - 1$, where $\lambda > 0$ is to be determined. If $\lambda \ge 4n$, we calculate

$$\begin{aligned}\left(\frac{\partial}{\partial t} - \Delta_{g_\epsilon}\right)\phi_1 &= -\lambda e^{-d_{\mathbb{C}^n}(z,q_0)^2 - \lambda t} - (-\Delta_{g_\epsilon} d^2_{\mathbb{C}^n} + |\nabla d^2_{\mathbb{C}^n}|^2_{g_\epsilon}) e^{-d_{\mathbb{C}^n}(z,q_0)^2 - \lambda t} \\ &\le \left(-\lambda + \sum_{j=1}^{p} (|z_j|^2 + \epsilon)^{1-\beta_j} + (n-p)\right) e^{-d_{\mathbb{C}^n}(z,q_0)^2 - \lambda t} < 0.\end{aligned}$$

On the other hand, $\phi_1(q_0, t_0) = 0$ and $\phi_1(z, t) < 0$ for any $(z, t) \neq (q_0, t_0)$. For any $\varepsilon > 0$, we can find a small neighborhood $V \cap \partial_{\mathcal{P}}(\mathcal{Q}_{\boldsymbol{\beta}}(0, R))$ of (q_0, t_0) such that, on V, we have $\varphi(q_0, t_0) + \varepsilon > \varphi(z, t) > \varphi(q_0, t_0) - \varepsilon$, since φ is continuous. On $\partial_{\mathcal{P}}(\mathcal{Q}_{\boldsymbol{\beta}}(0, R)) \backslash V$, the function ϕ_1 is bounded above by a negative constant. Therefore the function $\phi_1^- := \varphi(q_0, t_0) - \varepsilon + A\phi_1(z, t) \leq \varphi(z, t)$ for any $(z, t) \in \partial_{\mathcal{P}}(\mathcal{Q}_{\boldsymbol{\beta}}(0, R))$ if $A \gg 1$. Therefore, by the maximum principle, $\phi_1^-(z, t) \leq u_\epsilon(z, t)$ for any $(z, t) \in \mathcal{Q}_{\boldsymbol{\beta}}(0, R)$. Letting $\epsilon \to 0$, we have $\phi_1^-(z, t) \leq u(z, t)$. Letting $(z, t) \to (q_0, t_0)$ yields $\varphi(q_0, t_0) - \varepsilon \leq \liminf_{(z,t)\to(q_0,t_0)} u(z, t)$. Setting $\varepsilon \to 0$, we conclude that $\varphi(q_0, t_0) \leq \liminf_{(z,t)\to(q_0,t_0)} u(z, t)$. Considering $\phi_1^+(z, t) = \varphi(q_0, t_0) + \varepsilon - A\phi_1(z, t)$ and using an argument similar to that above, we can get $\varphi(q_0, t_0) \geq \limsup_{(z,t)\to(q_0,t_0)} u(z, t)$. Thus u coincides with φ at (q_0, t_0).

Case 2: $t_0 > 0$ and $q_0 \in \partial B_{\boldsymbol{\beta}}(0, R) \cap (\mathcal{S}_1 \cap \mathcal{S}_2)$. In this case $z_1(q_0) = z_2(q_0) = 0$. We define $q_0' = -q_0 \in \partial B_{\boldsymbol{\beta}}(0, R)$ to be the (Euclidean) opposite point to q_0. For some small $\delta > 0$, define

$$\phi_2(z, t) = d_{\mathbb{C}^n}(z, q_0')^2 - 4R^2 - \delta(t - t_0)^2.$$

Then $\phi_2(q_0, z_0) = 0$ and $\phi(z, t) < 0$ for any $(z, t) \neq (q_0, t_0)$. We calculate $\partial_t \phi_2 - \Delta_{g_\epsilon} \phi_2 \leq 0$. By an argument similar to Case 1, replacing ϕ_1 by ϕ_2 we get $\lim_{(z,t)\to(q_0,t_0)} u(z, t) = \varphi(q_0, t_0)$.

Case 3: $t_0 > 0$ and $q_0 \in \partial B_{\boldsymbol{\beta}}(0, R) \backslash (\mathcal{S}_1 \cap \mathcal{S}_2)$. As in Case 2 in the proof of Proposition 3.5, we define a similar function G. Define $\phi_3(z, t) = A(d_{\boldsymbol{\beta}}(z, 0)^2 - R^2) + G(z) - \delta(t - t_0)^2$ for $A \gg 1$ and small $\delta > 0$. Then we can calculate that $\partial_t \phi_3 \leq \Delta_{g_\epsilon} \phi_3$, $\phi_3(q_0, t_0) = 0$ and $\phi_3(z, t) < 0$ for any other $(z, t) \neq (q_0, t_0)$. Similar arguments to those in Case 1 proves that

$$\lim_{(z,t)\to(q_0,t_0)} u(z, t) = \varphi(q_0, t_0).$$

Combining the three cases above, we obtain that u coincides with φ on $\partial_{\mathcal{P}} \mathcal{Q}_{\boldsymbol{\beta}}(0, R)$. Thus the Dirichlet problem (4-2) admits a unique solution $u \in \mathscr{C}^0(\overline{\mathcal{Q}_{\boldsymbol{\beta}}(0, R)}) \cap \mathscr{C}^{2,1}(\mathcal{Q}_{\boldsymbol{\beta}}(0, R)^{\#})$. □

Corollary 4.5. *Given any functions $f \in \mathscr{C}_{\boldsymbol{\beta}}^{\alpha,\alpha/2}(\overline{\mathcal{Q}_{\boldsymbol{\beta}}})$ and $\varphi \in \mathscr{C}^0(\partial_{\mathcal{P}} \mathcal{Q}_{\boldsymbol{\beta}})$, there exists a unique solution $v \in \mathscr{C}^{2,1}(\mathcal{Q}_{\boldsymbol{\beta}}^{\#}) \cap \mathscr{C}^0(\overline{\mathcal{Q}_{\boldsymbol{\beta}}})$ to the Dirichlet problem*

$$\frac{\partial v}{\partial t} = \Delta_{g_{\boldsymbol{\beta}}} v + f \quad \text{in } \mathcal{Q}_{\boldsymbol{\beta}} \qquad \text{and} \qquad v = \varphi \quad \text{on } \partial_{\mathcal{P}} \mathcal{Q}_{\boldsymbol{\beta}}. \tag{4-14}$$

Proof. Let $v_\epsilon \in \mathscr{C}^{2+\alpha,(2+\alpha)/2}(\mathcal{Q}_{\boldsymbol{\beta}}) \cap \mathscr{C}^0(\mathcal{Q}_{\boldsymbol{\beta}})$ be the unique solution to the equations

$$\frac{\partial v_\epsilon}{\partial t} = \Delta_{g_\epsilon} v_\epsilon + f \quad \text{in } \mathcal{Q}_{\boldsymbol{\beta}} \qquad \text{and} \qquad v_\epsilon = \varphi \quad \text{on } \partial_{\mathcal{P}} \mathcal{Q}_{\boldsymbol{\beta}}.$$

For any compact subset $K \Subset B_{\boldsymbol{\beta}}(0, 1)$ and $\delta \in (0, 1)$, the standard Schauder estimates for parabolic equations provide uniform $\mathscr{C}^{2+\alpha,(2+\alpha)/2}$-estimates for v_ϵ on $K \backslash T_\delta \mathcal{S} \times (\delta^2, 1)$. Then $v_\epsilon \to v$ for some $v \in \mathscr{C}^{2+\alpha,(2+\alpha)/2}(K \backslash T_\delta \mathcal{S} \times (\delta^2, 1))$. Taking $\delta \to 0$ and $K \to B_{\boldsymbol{\beta}}(0, 1)$ and using a diagonal argument, we get that v_ϵ converges in $\mathscr{C}_{\mathrm{loc}}^{2+\alpha,(2+\alpha)/2}(B_{\boldsymbol{\beta}} \backslash \mathcal{S} \times (0, 1))$ to v and v satisfies the equation $\partial v / \partial t = \Delta_{g_{\boldsymbol{\beta}}} v + f$ on $B_{\boldsymbol{\beta}} \backslash \mathcal{S} \times (0, 1)$.

It only remains to show $v \in \mathscr{C}^0(\mathcal{Q}_{\boldsymbol{\beta}})$ and $v = \varphi$ on $\partial_{\mathcal{P}} \mathcal{Q}_{\boldsymbol{\beta}}$. The same proof as in Cases 1, 2 and 3 in Proposition 4.4 yields that v must coincide with φ on $\partial_{\mathcal{P}} \mathcal{Q}_{\boldsymbol{\beta}}$, since we can always choose $A > 1$ large enough that (for example in Case 1) $\partial \phi_1^- / \partial t - \Delta_{g_\epsilon} \phi_1^- \leq \inf_{\mathcal{Q}_{\boldsymbol{\beta}}} f \leq \partial v_\epsilon / \partial t - \Delta_{g_\epsilon} v_\epsilon$. To see the continuity of v in $\mathcal{Q}_{\boldsymbol\beta}$, because of the Sobolev inequality (3-42) for metric spaces $(B_{\boldsymbol\beta}, g_\epsilon)$ and by the proof of the standard De Giorgi–Nash–Moser theory for parabolic equations, we conclude that, for any $p \in \mathcal{S}$ and $t_0 \in (0,1)$, there exists a small number $R_0 = R_0(p, t_0)$ such that, on the cylinder $\widetilde{\mathcal{Q}}_{R_0} := B_{\boldsymbol\beta}(p, R_0) \times (t_0 - R_0^2, t_0)$, we have $\operatorname{osc}_{\widetilde{\mathcal{Q}}_r} v_\epsilon \le C r^{\alpha'}$ for any $r \in (0, R_0)$ and some $\alpha' \in (0,1)$. Therefore $\operatorname{osc}_{\widetilde{\mathcal{Q}}_r} v \le C r^{\alpha'}$ and v is continuous at (p, t_0), as desired.

The uniqueness of the solution to (4-14) follows from the maximum principle. □

Remark 4.6. Corollary 4.5 is not needed in the proof of Theorem 1.7. So by Theorem 1.7, the solution u to (4-14) is in $\mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(2+\alpha)/2}(\mathcal{Q}_{\boldsymbol\beta}) \cap \mathscr{C}^0(\overline{\mathcal{Q}_{\boldsymbol\beta}})$.

4B. ***Sketched proof of Theorem 1.7.*** With Proposition 4.4, we can prove the Schauder estimates for the solution $u \in \mathscr{C}^0(\overline{\mathcal{Q}_{\boldsymbol\beta}}) \cap \mathscr{C}^{2,1}(\mathcal{Q}_{\boldsymbol\beta}^\#)$ to (4-1) for a Dini-continuous function f by making use of almost the same arguments as in the proof of Theorem 1.2. We will not provide the full details and only point out the main differences. For any given points $Q_p = (p, t_p)$, $Q_q = (q, t_q) \in \big(B_{\boldsymbol\beta}\big(0, \tfrac{1}{2}\big)\backslash\mathcal{S}\big) \times (\hat{t}, 1)$, to define the approximating functions u_k as in (3-44), we define u_k in this case as the solution to the heat equation

$$\frac{\partial u_k}{\partial t} = \Delta_{g_{\boldsymbol\beta}} u_k + f(Q_p) \quad \text{in } \hat{B}_k(p) \times (t_p - \hat{t}\cdot\tau^{2k}, t_p], \qquad u_k = u \quad \text{on } \partial_{\mathcal{P}}(\hat{B}_k(p) \times (t_p - \hat{t}\cdot\tau^{2k}, t_p]),$$

where $\hat{B}_k(p)$ is defined in (3-48). We can now apply the estimates in Proposition 4.4 to the functions u_k or $u_k - u_{k-1}$, instead of those in Lemmas 3.3 and 3.4 as we did in Sections 3B, 3C and 3D, to prove the Schauder estimates for u. Thus we finish the proof of Theorem 1.7. □

4C. ***Interior Schauder estimate for nonflat conical Kähler metrics.*** Let $g = \sqrt{-1} g_{j\bar{k}}(z,t)\, dz_j \wedge dz_{\bar{k}}$ be a $\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}$ conical Kähler metric on $\mathcal{Q}_{\boldsymbol\beta}$ with conical singularity along $\mathcal{S}$; that is, $g(\,\cdot\,, t)$ is a $C_{\boldsymbol\beta}^{0,\alpha}$ conical Kähler metric (from Section 3E) for any $t \in [0,1]$, and the coefficients of g in the basis $\{\epsilon_j \wedge \bar{\epsilon}_k, \dots\}$ are $\tfrac{1}{2}\alpha$-Hölder continuous in $t \in [0,1]$. Suppose $u \in \mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(2+\alpha)/2}(\mathcal{Q}_{\boldsymbol\beta})$ satisfies the equation

$$\frac{\partial u}{\partial t} = \Delta_g u + f \quad \text{in } \mathcal{Q}_{\boldsymbol\beta} \tag{4-15}$$

for some $f \in \mathscr{C}^{\alpha,\alpha/2}(\overline{\mathcal{Q}_{\boldsymbol\beta}})$.

Proposition 4.7. *There exists a constant $C = C(n, \boldsymbol\beta, \alpha, g)$ such that*

$$\|u\|^*_{\mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(2+\alpha)/2}(\mathcal{Q}_{\boldsymbol\beta})} \le C\big(\|u\|_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol\beta})} + \|f\|^{(2)}_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\mathcal{Q}_{\boldsymbol\beta})}\big).$$

Proof. The proof is parallel to that of Proposition 3.32. Given any two points $P_x = (x, t_x)$, $P_y = (y, t_x) \in \mathcal{Q}_{\boldsymbol\beta}$, we may assume $d_{P_x} = \min\{d_{P_x}, d_{P_y}\} > 0$, where $d_{P_x} := d_{\mathcal{P},\boldsymbol\beta}(P_x, \partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol\beta})$ is the parabolic distance of P_x to the parabolic boundary $\partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol\beta}$. Let $\mu \in \big(0, \tfrac{1}{4}\big)$ be a positive number to be determined later. Define $d := \mu d_{P_x}$, $\mathcal{Q} := B_{\boldsymbol\beta}(x, d) \times (t_x - d^2, t_x]$ the "parabolic ball" centered at P_x, and $\tfrac{1}{2}\mathcal{Q} := B_{\boldsymbol\beta}\big(x, \tfrac{1}{2}d\big) \times \big(t_x - \tfrac{1}{4}d^2, t_x\big]$.

<u>Case 1</u>: $d_{\mathcal{P},\boldsymbol\beta}(P_x, P_y) < \tfrac{1}{2}d$. In this case we always have $P_y \in \tfrac{1}{2}\mathcal{Q}$.

<u>Case 1a</u>: $B_{\boldsymbol\beta}(x, d) \cap \mathcal{S} = \varnothing$. As in the proof of Proposition 3.32, we can introduce smooth complex coordinates $\{w_1, w_2, z_3, \dots, z_n\}$ on $B_{\boldsymbol\beta}(x, d)$ under which $g_{\boldsymbol\beta}$ becomes the standard Euclidean metric and the components of g are $\mathscr{C}^{\alpha,\alpha/2}$ in the usual sense on $\mathcal{Q}$. The leading coefficients and constant term f

in (4-15) are both $\mathscr{C}^{\alpha,\alpha/2}$ in the usual sense, so we apply the standard parabolic Schauder estimates (see Theorem 4.9 in [27]) to get that there exists some constant $C = C(n, \boldsymbol{\beta}, \alpha, g)$ independent of $\mathcal{Q}$ such that

$$[u]^*_{\mathscr{C}^{2+\alpha,(2+\alpha)/2}(\mathcal{Q})} \le C(\|u\|_{\mathscr{C}^0(\mathcal{Q})} + \|f\|^{(2)}_{\mathscr{C}^{\alpha,\alpha/2}(\mathcal{Q})}). \tag{4-16}$$

Let D denote the ordinary first-order operators in the coordinates $\{w_1, w_2, z_3, \dots, z_n\}$. We calculate

$$\begin{aligned}|Tu(P_x) - Tu(P_y)| &\le |D^2u(P_x) - D^2u(P_y)| + \frac{d_{\mathcal{P},\boldsymbol{\beta}}(P_x, P_y)}{d}(|D^2u(P_x)| + |D^2u(P_y)|)\\ &\le \frac{4d_{\mathcal{P},\boldsymbol{\beta}}(P_x, P_y)^\alpha}{d^{2+\alpha}}[u]^*_{\mathscr{C}^{2+\alpha,(2+\alpha)/2}(\mathcal{Q})} + \frac{4d_{\mathcal{P},\boldsymbol{\beta}}(P_x, P_y)}{d^3}[u]^*_{\mathscr{C}^{2,1}(\mathcal{Q})}\\ &\le \frac{8d_{\mathcal{P},\boldsymbol{\beta}}(P_x, P_y)^\alpha}{d^{2+\alpha}}[u]^*_{\mathscr{C}^{2+\alpha,(2+\alpha)/2}(\mathcal{Q})} + C\frac{d_{\mathcal{P},\boldsymbol{\beta}}(P_x, P_y)^\alpha}{d^{2+\alpha}}\|u\|_{\mathscr{C}^0(\mathcal{Q})}\end{aligned}$$

and

$$\left|\frac{\partial u}{\partial t}(P_x) - \frac{\partial u}{\partial t}(P_y)\right| \le \frac{4d_{\mathcal{P},\boldsymbol{\beta}}(P_x, P_y)^\alpha}{d^{2+\alpha}}[u]^*_{\mathscr{C}^{2+\alpha,(2+\alpha)/2}(\mathcal{Q})}.$$

Recall $\mathcal{T}$ denotes the operators in T and $\partial/\partial t$; then by (4-16) it follows that

$$d_{P_x}^{2+\alpha}\frac{|\mathcal{T}u(P_x) - \mathcal{T}u(P_y)|}{d_{\mathcal{P},\boldsymbol{\beta}}(P_x, P_y)^\alpha} \le \frac{C}{\mu^{2+\alpha}}\|f\|^{(2)}_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\mathcal{Q})} + \frac{C}{\mu^{2+\alpha}}\|u\|_{\mathscr{C}^0(\mathcal{Q})}. \tag{4-17}$$

Case 1b: $B_{\boldsymbol{\beta}}(x, d) \cap \mathcal{S} \ne \varnothing$. Let $\hat{x} \in \mathcal{S}$ be the projection of x onto $\mathcal{S}$ and $\hat{P}_x = (\hat{x}, t_x)$ be the corresponding space-time point. Define $\hat{\mathcal{Q}} := B_{\boldsymbol{\beta}}(\hat{x}, 2d) \times (t_x - 4d^2, t_x]$. As in Case 1b in the proof of Proposition 3.32, we may choose suitable enough complex coordinates that $g_{\epsilon_j\bar{\epsilon}_k}(\hat{P}_x) = \delta_{jk}$ and, for $j, k \ge p+1$, we have $g_{j\bar{k}}(\hat{P}_x) = \delta_{jk}$ and the cross terms in the expansion of g in (3-114) vanish at $\hat{P}_x$. Thus (4-15) can be rewritten as

$$\frac{\partial u}{\partial t} = \Delta_{g_{\boldsymbol{\beta}}}u + \eta \cdot \sqrt{-1}\partial\bar{\partial}u + f =: \Delta_{g_{\boldsymbol{\beta}}}u + \tilde{f}, \quad u \in \mathscr{C}^0(\hat{\mathcal{Q}}) \cap \mathscr{C}^{2,1}(\hat{\mathcal{Q}}^\#),$$

for some $(1, 1)$-form η as in the proof of Proposition 3.32. From the rescaled version of Theorem 1.7 we conclude that

$$d^{2+\alpha}\frac{|\mathcal{T}u(P_x) - \mathcal{T}u(P_y)|}{d_{\mathcal{P},\boldsymbol{\beta}}(P_x, P_y)^\alpha} \le C(\|u\|_{\mathscr{C}^0(\hat{\mathcal{Q}})} + \|\tilde{f}\|^{(2)}_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\hat{\mathcal{Q}})}).$$

Hence

$$d_{P_x}^{2+\alpha}\frac{|\mathcal{T}u(P_x) - \mathcal{T}u(P_y)|}{d_{\mathcal{P},\boldsymbol{\beta}}(P_x, P_y)^\alpha} \le \frac{C}{\mu^{2+\alpha}}(\|u\|_{\mathscr{C}^0(\hat{\mathcal{Q}})} + \|\tilde{f}\|^{(2)}_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\hat{\mathcal{Q}})}). \tag{4-18}$$

Case 2: $d_{\mathcal{P},\boldsymbol{\beta}}(P_x, P_y) \ge \frac{1}{2}d$. Here we calculate (recall $\mathcal{Q}_{\boldsymbol{\beta}} := B_{\boldsymbol{\beta}}(0, 1) \times (0, 1]$)

$$d_{P_x}^{2+\alpha}\frac{|\mathcal{T}u(P_x) - \mathcal{T}u(P_y)|}{d_{\mathcal{P},\boldsymbol{\beta}}(P_x, P_y)^\alpha} \le 4d_{P_x}^{2+\alpha}\frac{|\mathcal{T}u|(P_x) + |\mathcal{T}u|(P_y)}{d^\alpha} \le \frac{8}{\mu^\alpha}[u]^*_{\mathscr{C}^{2,1}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})}. \tag{4-19}$$

Combining (4-17)–(4-19), we obtain

$$\begin{aligned}d_{P_x}^{2+\alpha}\frac{|\mathcal{T}u(P_x) - \mathcal{T}u(P_y)|}{d_{\mathcal{P},\boldsymbol{\beta}}(P_x, P_y)^\alpha} \le \frac{8}{\mu^\alpha}[u]^*_{\mathscr{C}^{2,1}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})} + \frac{C}{\mu^{2+\alpha}}(\|u\|_{\mathscr{C}^0(\hat{\mathcal{Q}})} + \|\tilde{f}\|^{(2)}_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\hat{\mathcal{Q}})})&\\ + \frac{C}{\mu^{2+\alpha}}\|f\|^{(2)}_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\mathcal{Q})} + \frac{C}{\mu^{2+\alpha}}\|u\|_{\mathscr{C}^0(\mathcal{Q})}.&\end{aligned}$$

Observe that, for any $P \in \mathcal{Q}$ or $P \in \hat{\mathcal{Q}}$, we have $d_{\mathcal{P},\boldsymbol{\beta}}(P, \partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol{\beta}}) \geq (1-2\mu)d_{P_x}$. Then it follows from the definition that

$$\|f\|^{(2)}_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\mathcal{Q})} \leq C\mu^2\|f\|^{(2)}_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol{\beta}})} + C\mu^{2+\alpha}[f]^{(2)}_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})} \leq C\mu^2\|f\|^{(2)}_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})}.$$

We calculate

$$\begin{aligned}\|\tilde{f}\|^{(2)}_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\hat{\mathcal{Q}})} &\leq \|\eta\|^{(0)}_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\hat{\mathcal{Q}})}\|Tu\|^{(2)}_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\hat{\mathcal{Q}})} + \|f\|^{(2)}_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\hat{\mathcal{Q}})}\\ &\leq C_1[g]^*_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})}\mu^\alpha(\mu^2[u]^*_{\mathscr{C}^{2,1}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})} + \mu^{2+\alpha}[u]^*_{\mathscr{C}^{2+\alpha,(2+\alpha)/2}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})}) + \mu^2\|f\|^{(2)}_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})}\\ &\leq C_1[g]^*_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})}\mu^\alpha(C(\mu)[u]^*_{\mathscr{C}^{2,1}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})} + 2\mu^{2+\alpha}[u]^*_{\mathscr{C}^{2+\alpha,(2+\alpha)/2}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})}) + \mu^2\|f\|^{(2)}_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})},\end{aligned}$$

where in the last inequality we use the interpolation inequality, by which we also have

$$\frac{8}{\mu^\alpha}[u]^*_{\mathscr{C}^{2,1}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})} \leq \mu^\alpha[u]^*_{\mathscr{C}^{2+\alpha,(2+\alpha)/2}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})} + C(\mu)\|u\|_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol{\beta}})}.$$

If μ is chosen small enough that $\mu^\alpha(2C_1[g]^*_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})}+1) < \frac{1}{2}$, combining the above inequalities yields

$$d^{2+\alpha}_{P_x}\frac{|\mathcal{T}u(P_x)-\mathcal{T}u(P_y)|}{d_{\mathcal{P},\boldsymbol{\beta}}(P_x, P_y)^\alpha} \leq \tfrac{1}{2}[u]^*_{\mathscr{C}^{2+\alpha,(2+\alpha)/2}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})} + C(\mu)(\|u\|_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol{\beta}})} + \|f\|^{(2)}_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})}).$$

Taking the supremum over all $P_x \neq P_y \in \mathcal{Q}_{\boldsymbol{\beta}}$, we obtain

$$[u]^*_{\mathscr{C}^{2+\alpha,(2+\alpha)/2}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})} \leq C(\|u\|_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol{\beta}})} + \|f\|^{(2)}_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})}).$$

The proposition is proved by invoking the interpolation inequalities. □

Remark 4.8. It follows from the proof that the estimates in Proposition 4.7 also hold on $\mathcal{Q}_{\boldsymbol{\beta}}(p, R) := B_{\boldsymbol{\beta}}(p, R) \times (0, R^2) \subset \mathcal{Q}_{\boldsymbol{\beta}}$, i.e., the cylinder whose spatial center p may not lie in $\mathcal{S}$.

It is easy to derive the following local Schauder estimate for $\mathscr{C}^{2+\alpha,(2+\alpha)/2}_{\boldsymbol{\beta}}$-solutions to (4-15) from Proposition 4.7.

Corollary 4.9. *Let $K \Subset B_{\boldsymbol{\beta}}(0, 1)$ be a compact subset and $\varepsilon_0 \in (0, 1)$ be a given number. With the same assumptions as in Proposition 4.7, there exists a constant $C = C(n, \boldsymbol{\beta}, \alpha, g, K, \varepsilon_0) > 0$ such that*

$$\|u\|_{\mathscr{C}^{2+\alpha,(2+\alpha)/2}(K\times[\varepsilon_0,1])} \leq C(\|u\|_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol{\beta}})} + \|f\|_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})}).$$

With the interior Schauder estimates in Proposition 4.7, we can show the existence of $\mathscr{C}^{2+\alpha,(2+\alpha)/2}_{\boldsymbol{\beta}}(\mathcal{Q}_{\boldsymbol{\beta}})$-solutions to the Dirichlet problem

$$\frac{\partial u}{\partial t} = \Delta_g u + f \quad \text{in } \mathcal{Q}_{\boldsymbol{\beta}} \qquad \text{and} \qquad u = \varphi \quad \text{on } \partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol{\beta}} \tag{4-20}$$

for any given $f \in \mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}(\overline{\mathcal{Q}_{\boldsymbol{\beta}}})$ and $\varphi \in \mathscr{C}^0(\partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol{\beta}})$. We first show the existence of solutions to (4-20) in the case $\varphi \equiv 0$.

Lemma 4.10. *Let $\sigma \in (0,1)$ be given and $u \in \mathscr{C}_{\boldsymbol{\beta}}^{2+\alpha,(2+\alpha)/2}(\mathcal{Q}_{\boldsymbol{\beta}})$ solve* (4-20), *with $\|u\|^{(-\sigma)}_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol{\beta}})} < \infty$ and $\|f\|^{(2-\sigma)}_{\mathscr{C}_{\boldsymbol{\beta}}^{\alpha,\alpha/2}(\mathcal{Q}_{\boldsymbol{\beta}})} < \infty$. Then there is a constant $C = C(n, \alpha, \boldsymbol{\beta}, g, \sigma) > 0$ such that*

$$\|u\|^{(-\sigma)}_{\mathscr{C}_{\boldsymbol{\beta}}^{2+\alpha,(2+\alpha)/2}(\mathcal{Q}_{\boldsymbol{\beta}})} \le C\big(\|u\|^{(-\sigma)}_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol{\beta}})} + \|f\|^{(2-\sigma)}_{\mathscr{C}_{\boldsymbol{\beta}}^{\alpha,\alpha/2}(\mathcal{Q}_{\boldsymbol{\beta}})}\big).$$

Proof. The lemma follows from the definitions of the norms and the estimates in Proposition 4.7. □

Lemma 4.11. *Suppose $u \in \mathscr{C}_{\boldsymbol{\beta}}^{2,1}(\mathcal{Q}_{\boldsymbol{\beta}}) \cap \mathscr{C}^0(\overline{\mathcal{Q}_{\boldsymbol{\beta}}})$ satisfies $\partial u/\partial t = \Delta_g u + f$ and $u \equiv 0$ on $\partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol{\beta}}$. For any $\sigma \in (0,1)$, there exists a constant $C = C(n, \boldsymbol{\beta}, g, \sigma) > 0$ such that*

$$\|u\|^{(-\sigma)}_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol{\beta}})} = \sup_{P_x \in \mathcal{Q}_{\boldsymbol{\beta}}} d_{P_x}^{-\sigma}|u(P_x)| \le C \sup_{P_x \in \mathcal{Q}_{\boldsymbol{\beta}}} d_{P_x}^{2-\sigma}|f(P_x)| = C\|f\|^{(2-\sigma)}_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol{\beta}})},$$

where $d_{P_x} = d_{\mathcal{P},\boldsymbol{\beta}}(P_x, \partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol{\beta}})$ is the parabolic distance of P_x to the parabolic boundary $\partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol{\beta}}$.

Proof. We write $N := \|f\|^{(2-\sigma)}_{C^0(\mathcal{Q}_{\boldsymbol{\beta}})} < \infty$ and $P_x = (x, t_x)$. Define functions

$$w_1(P_x) = (1 - d_{\boldsymbol{\beta}}(x)^2)^\sigma \quad \text{and} \quad w_2(P_x) = t_x^{\sigma/2},$$

where $d_{\boldsymbol{\beta}}(x) = d_{\boldsymbol{\beta}}(x, 0)$ is the $g_{\boldsymbol{\beta}}$-distance between x and 0. Observe that $d_{P_x} = \min\{1 - d_{\boldsymbol{\beta}}(x), t_x^{1/2}\}$ by definition. By a straightforward calculation there is a constant $c_0 > 0$ such that

$$\left(\frac{\partial}{\partial t} - \Delta_g\right) w_1 \ge c_0 (1 - d_{\boldsymbol{\beta}}(x))^{\sigma-2} \quad \text{and} \quad \left(\frac{\partial}{\partial t} - \Delta_g\right) w_2 \ge c_0 (t_x^{1/2})^{\sigma-2}.$$

By the maximum principle we get

$$|u(P_x)| \le N c_0^{-1}(w_1(P_x) + w_2(P_x)) \quad \text{for all } P_x \in \mathcal{Q}_{\boldsymbol{\beta}}. \tag{4-21}$$

We take the decomposition of $\mathcal{Q}_{\boldsymbol{\beta}}$ into different regions, $\mathcal{Q}_{\boldsymbol{\beta}} = \Omega_1 \cup \Omega_2$, where

$$\Omega_1 := \{P_x \in \mathcal{Q}_{\boldsymbol{\beta}} \mid t_x^{1/2} > 1 - d_{\boldsymbol{\beta}}(x)\},$$
$$\Omega_2 := \{P_x \in \mathcal{Q}_{\boldsymbol{\beta}} \mid t_x^{1/2} \le 1 - d_{\boldsymbol{\beta}}(x)\}.$$

Inequality (4-21) implies that, on the parabolic boundaries $\partial_{\mathcal{P}}\Omega_1$ and $\partial_{\mathcal{P}}\Omega_2$, we have $|u(P_x)| \le 2Nc_0^{-1} d_{P_x}^\sigma$. On Ω_1 we have $(\partial/\partial t - \Delta_g)(2Nc_0^{-1}w_1 \pm u) \ge 0$ and $2Nc_0^{-1}w_1 \pm u \ge 0$ on $\partial_{\mathcal{P}}\Omega_1$, so the maximum principle implies that $2Nc_0^{-1}w_1 \pm u \ge 0$ in Ω_1, i.e., $|u(P_x)| \le 2Nc_0^{-1}d_{P_x}^\sigma$ in Ω_1. Similarly we also have $2Nc_0^{-1}w_2 \pm u \ge 0$ in Ω_2, and thus $|u(P_x)| \le 2Nc_0^{-1}d_{P_x}^\sigma$ in Ω_2. In conclusion, we get

$$|u(P_x)| \le 2c_0^{-1} N d_{P_x}^\sigma \quad \text{for all } P_x \in \mathcal{Q}_{\boldsymbol{\beta}}. \qquad \square$$

Proposition 4.12. *If $\varphi \equiv 0$, equation* (4-20) *admits a unique solution $u \in \mathscr{C}_{\boldsymbol{\beta}}^{2+\alpha,(2+\alpha)/2}(\mathcal{Q}_{\boldsymbol{\beta}}) \cap \mathscr{C}^0(\overline{\mathcal{Q}_{\boldsymbol{\beta}}})$ for any $f \in \mathscr{C}_{\boldsymbol{\beta}}^{\alpha,\alpha/2}(\overline{\mathcal{Q}_{\boldsymbol{\beta}}})$.*

Proof. Uniqueness follows from the maximum principle, so it suffices to show existence. We will use the continuity method. Define a continuous family of linear operators as follows: for $s \in [0,1]$, let $L_s := s(\partial/\partial t - \Delta_g) + (1-s)(\partial/\partial t - \Delta_{g_{\boldsymbol{\beta}}})$. It can been seen that $L_s = \partial/\partial t - \Delta_{g_s}$ for some conical

Kähler metric g_s which is uniformly equivalent to $g_{\boldsymbol\beta}$ and has uniform $\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}$-estimate. So the interior Schauder estimates holds also for L_s. Fix $\sigma \in (0, 1)$. Define

$$\mathcal{B}_1 := \{u \in \mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(2+\alpha)/2}(\mathcal{Q}_{\boldsymbol\beta}) \mid \|u\|^{(-\sigma)}_{\mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(2+\alpha)/2}(\mathcal{Q}_{\boldsymbol\beta})} < \infty\},$$
$$\mathcal{B}_2 := \{f \in \mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\mathcal{Q}_{\boldsymbol\beta}) \mid \|f\|^{(2-\sigma)}_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\mathcal{Q}_{\boldsymbol\beta})} < \infty\}.$$

Observe that any $u \in \mathcal{B}_1$ is continuous in $\overline{\mathcal{Q}_{\boldsymbol\beta}}$ and vanishes on $\partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol\beta}$. L_s defines a continuous family of linear operators from $\mathcal{B}_1$ to $\mathcal{B}_2$. By Lemmas 4.10 and 4.11 we have

$$\|u\|_{\mathcal{B}_1} \le C(\|u\|^{(-\sigma)}_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol\beta})} + \|L_s u\|_{\mathcal{B}_2}) \le C\|L_s u\|_{\mathcal{B}_2} \quad \text{for all } s \in [0, 1] \text{ and for all } u \in \mathcal{B}_1.$$

By Corollary 4.5 and Remark 4.6, L_0 is invertible, thus by Theorem 5.2 in [18], L_1 is also invertible. □

Corollary 4.13. *For any $\varphi \in \mathscr{C}^0(\partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol\beta})$ and $f \in \mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\overline{\mathcal{Q}_{\boldsymbol\beta}})$, equation* (4-20) *admits a unique solution $u \in \mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(2+\alpha)/2}(\mathcal{Q}_{\boldsymbol\beta}) \cap \mathscr{C}^0(\overline{\mathcal{Q}_{\boldsymbol\beta}})$.*

Proof. The proof is identical to that of Corollary 3.39 by an approximation argument. We may assume $\varphi \in \mathscr{C}^0(\overline{\mathcal{Q}_{\boldsymbol\beta}})$ and choose a sequence $\varphi_k \in \mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(2+\alpha)/2}(\overline{\mathcal{Q}_{\boldsymbol\beta}})$ which converges uniformly to φ on $\overline{\mathcal{Q}_{\boldsymbol\beta}}$. The equations

$$\frac{\partial v_k}{\partial t} = \Delta_g v_k + f - \Delta_g \varphi_k \qquad \text{and} \qquad v_k \equiv 0 \quad \text{on } \partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol\beta}$$

admit a unique $\mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(2+\alpha)/2}$-solution by Proposition 4.12. The interior Schauder estimates in Corollary 4.9 imply that $u_k := v_k + \varphi_k$ converges in $\mathscr{C}_{\boldsymbol\beta,\mathrm{loc}}^{2+\alpha,(\alpha+2)/2}$ to some function u in $\mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(2+\alpha)/2}(\mathcal{Q}_{\boldsymbol\beta})$ which solves (4-20). The $\mathscr{C}^0$-convergence $u_k \to u$ is uniform on $\overline{\mathcal{Q}_{\boldsymbol\beta}}$ by the maximum principle, so $u = \varphi$ on $\partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol\beta}$, as desired. □

We recall the definition of weak solutions and refer to Section 7.1 in [17] for the notations.

Definition 4.14. We say a function u on $\mathcal{Q}_{\boldsymbol\beta}$ is a weak solution to the equation $\partial u/\partial t = \Delta_g u + f$ if:

(1) $u \in L^2(0, 1; H^1(B_{\boldsymbol\beta}))$ and $\partial u/\partial t \in L^2(0, 1; H^{-1}(B_{\boldsymbol\beta}))$.

(2) For any $v \in H_0^1(B_{\boldsymbol\beta})$ and $t \in (0, 1)$,

$$\int_{B_\beta} \frac{\partial u(x,t)}{\partial t} v(x)\omega_g^n = -\int_{B_\beta} \langle \nabla u(x,t), \nabla v(x)\rangle_g \omega_g^n + \int_{B_\beta} f(x,t)v(x)\omega_g^n.$$

On can use the classical *Galerkin approximations* to construct a weak solution to $\partial u/\partial t = \Delta_g u + f$ (see Section 7.1.2 in [17]). If f has better regularity, so does the weak solution u.

Lemma 4.15. *If $f \in \mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\mathcal{Q}_{\boldsymbol\beta})$, then any weak solution to*

$$\frac{\partial u}{\partial t} = \Delta_g u + f$$

belongs to $\mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(\alpha+2)/2}(\mathcal{Q}_{\boldsymbol\beta})$.

Proof. The Sobolev inequality holds for the metric g, so by the proof of the standard De Giorgi–Nash–Moser theory for parabolic equations we have that u is in fact continuous on $\mathcal{Q}_{\boldsymbol\beta}$. Since the metric g is smooth on $\mathcal{Q}^{\#}_{\boldsymbol\beta}$, the weak solution u is also a weak solution in $\mathcal{Q}^{\#}_{\boldsymbol\beta}$ with the smooth background metric, so we have that $u \in \mathscr{C}^{2+\alpha,(2+\alpha)/2}_{\mathrm{loc}}(\mathcal{Q}^{\#}_{\boldsymbol\beta})$ in the usual sense by the classical Schauder estimates. Thus it suffices to consider points at $\mathcal{S}$. We choose the worst such point $0 \in \mathcal{S}$ only, since the case when centers are in other components of $\mathcal{S}$ is even simpler. We fix the point $P_0 = (0, t_0) \in \mathcal{Q}_{\boldsymbol\beta}$ with $t_0 > 0$. Fix $r \in (0, \sqrt{t_0})$. By Corollary 4.13,

$$\frac{\partial v}{\partial t} = \Delta v + f \quad \text{in } \mathcal{Q}_{\boldsymbol\beta}(P_0, r) := B_{\boldsymbol\beta}(0, r) \times (t_0 - r^2, t_0]$$

with boundary value $v = u$ on $\partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol\beta}(P_0, r)$ admits a unique solution $v \in \mathscr{C}^{2+\alpha,(\alpha+2)/2}{}_{\boldsymbol\beta}(\mathcal{Q}_{\boldsymbol\beta}(P_0,r))$. Then by the maximum principle $u = v$ in $\mathcal{Q}_{\boldsymbol\beta}(P_0, r)$. Thus $u \in \mathscr{C}^{2+\alpha,(\alpha+2)/2}_{\boldsymbol\beta}(\mathcal{Q}_{\boldsymbol\beta}(P_0, r))$ too. Since the argument also works at other space-time points in $\mathcal{S}_{\mathcal{P}}$, we see that $u \in \mathscr{C}^{2+\alpha,(2+\alpha)/2}_{\boldsymbol\beta}(\mathcal{Q}_{\boldsymbol\beta})$, as desired. □

Corollary 4.16. *Let (X, g, D) be as in Corollary 3.41, and let $u_0 \in C^0(X)$ and $f \in \mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(X \times (0, 1])$ be given functions. The weak solution u to the equation*

$$\frac{\partial u}{\partial t} = \Delta_g u + f \quad \textit{in } X \times (0, 1], \qquad u|_{t=0} = u_0$$

always exists. Moreover, $u \in \mathscr{C}^{2+\alpha,(2+\alpha)/2}_{\boldsymbol\beta}(X \times (0, 1])$, and there exists a constant $C = C(n, g, \boldsymbol\beta, \alpha) > 0$ such that

$$\|u\|_{\mathscr{C}^{2+\alpha,(2+\alpha)/2}_{\boldsymbol\beta}(X\times(1/2,1])} \le C(\|u_0\|_{C^0(X)} + \|f\|_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(X\times(0,1])}).$$

Proof. The weak equation can be constructed using the *Galerkin approximations* [17]. The uniqueness is an easy consequence of the maximum principle. The regularity of u follows from the local results in Lemma 4.15. The estimate follows from the maximum principle, a covering argument as in Corollary 3.41, and the local estimates in Corollary 4.9. □

The *interior* estimate in Corollary 4.16 is not good enough to show the existence of solutions to nonlinear partial differential equations since the estimate becomes worse as t approaches 0. We need some global estimates in the whole time interval $t \in [0, 1]$ if the initial value u_0 has better regularity.

4D. ***Schauder estimate near $t = 0$.*** In this subsection, we will prove a Schauder estimate in the whole time interval for the solutions to the heat equation when the initial value is 0 or has better regularity. We consider the model case with the background metric $g_{\boldsymbol\beta}$ first, then we generalize the estimate to general nonflat conical Kähler metrics.

4D1. *The model case.* In this subsection, we will assume the background metric is $g_{\boldsymbol\beta}$. Let u be the solution to the equation

$$\frac{\partial u}{\partial t} = \Delta_{g_{\boldsymbol\beta}} u + f \quad \text{in } \mathcal{Q}_{\boldsymbol\beta}, \qquad u|_{t=0} \equiv 0, \tag{4-22}$$

and $u = \varphi \in \mathscr{C}^0$ on $\partial B_{\boldsymbol\beta} \times (0, 1]$, where $f \in \mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(\overline{\mathcal{Q}_{\boldsymbol\beta}})$. In the calculations below, we should have used the smooth approximating solutions u_ϵ, where $\partial_t u_\epsilon = \Delta_{g_\epsilon} u_\epsilon + f$ and $u_\epsilon = u$ on $\partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol\beta}$. But by letting $\epsilon \to 0$, the corresponding estimates also hold for u. So for simplicity, we will work directly on u.

We fix $0 < \rho < R \le 1$ and define $B_R := B_\beta(0, R)$ and $Q_R := B_R \times [0, R^2]$ in this section. Let u be the solution to (4-22). We first have the following Caccioppoli inequalities.

Lemma 4.17. *There exists a constant $C = C(n) > 0$ such that*

$$\sup_{t\in[0,\rho^2]} \int_{B_\rho} u^2\omega_\beta^n + \iint_{Q_\rho} |\nabla u|^2_{g_\beta}\omega_\beta^n\, dt \le C\biggl(\frac{1}{(R-\rho)^2}\iint_{Q_R} u^2\omega_\beta^n\, dt + (R-\rho)^2\iint_{Q_R} f^2\omega_\beta^n\, dt\biggr) \quad (4\text{-}23)$$

and

$$\sup_{t\in[0,\rho^2]} \int_{B_\rho} |\nabla u|^2_{g_\beta}\omega_\beta^n + \iint_{Q_\rho} (|\nabla\nabla u|^2_{g_\beta} + |\nabla\overline{\nabla} u|^2_{g_\beta})\omega_\beta^n\, dt \le C\biggl(\frac{1}{(R-\rho)^2}\iint_{Q_R} |\nabla u|^2_{g_\beta}\omega_\beta^n\, dt + \iint_{Q_R} (f - f_R)^2\omega_\beta^n\, dt\biggr), \quad (4\text{-}24)$$

where $f_R := |Q_R|^{-1}_{\omega_\beta} \iint_{Q_R} f\omega_\beta^n\, dt$ is the average of f over the cylinder Q_R.

Proof. We fix a cut-off function η such that $\operatorname{supp}\eta \subset B_R$, $\eta = 1$ on B_ρ, and $|\nabla\eta|_{g_\beta} \le 2/(R-\rho)$. Multiplying both sides of (4-22) by $\eta^2 u$ and integrating by parts, we get

$$\begin{aligned}\frac{d}{dt}\int_{B_R} \eta^2u^2 &= \int_{B_R} 2\eta^2 u\Delta_{g_\beta}u + 2\eta^2uf = \int_{B_R} -2\eta^2|\nabla u|^2_{g_\beta} - 4u\eta\langle\nabla u, \nabla\eta\rangle_{g_\beta} + 2\eta^2uf \\ &\le \int_{B_R} -\eta^2|\nabla u|^2_{g_\beta} + 4u^2|\nabla\eta|^2_{g_\beta} + \eta^2\frac{u}{(R-\rho)^2} + \eta^2(R-\rho)^2f^2.\end{aligned}$$

Equation (4-23) follows by integrating this inequality over $t \in [0, s^2]$ for all $s \le \rho$. To see (4-24), observe that the Bochner formula yields

$$\frac{\partial}{\partial t}|\nabla u|^2 \le \Delta_{g_\beta}|\nabla u|^2 - |\nabla\nabla u|^2_{g_\beta} - |\nabla\overline{\nabla}u|^2_{g_\beta} - 2\langle\nabla u, \nabla f\rangle_{g_\beta}.$$

Multiplying both sides of this inequality by η^2 and integrating by parts, we get

$$\begin{aligned}\frac{d}{dt}\int_{B_R} \eta^2|\nabla u|^2 &\le \int_{B_R} -2\eta\langle\nabla\eta, \nabla|\nabla u|^2\rangle_{g_\beta} - \eta^2|\nabla\nabla u|^2 - \eta^2|\nabla\overline{\nabla}u|^2 - 2\eta^2\langle\nabla u, \nabla f\rangle \\ &\le \int_{B_R} 4\eta|\nabla u||\nabla\eta|\bigl|\nabla|\nabla u|\bigr| - \eta^2|\nabla\nabla u|^2 - \eta^2|\nabla\overline{\nabla}u|^2 \\ &\qquad\qquad 4\eta|f - f_R||\nabla\eta||\nabla u| + 2\eta^2|f - f_R||\Delta_{g_\beta}u| \\ &\le \int_{B_R} -\tfrac{1}{2}\eta^2(|\nabla\nabla u|^2 + |\nabla\overline{\nabla}u|^2) + 10\eta^2|\nabla u|^2|\nabla\eta|^2 + 20\eta^2(f - f_R)^2.\end{aligned}$$

Then (4-24) follows by integrating this inequality over $t \in [0, s^2]$ for any $s \in [0, \rho]$. □

Combining (4-23) and (4-24) we conclude that

$$\sup_{t\in[0,R^2/4]} \int_{B_{R/2}} |\nabla u|^2 + \iint_{Q_{R/2}} |\Delta_{g_\beta}u|^2 \le \frac{C}{R^4}\iint_{Q_R} u^2 + CR^{2n+2}\|f\|^2_{\mathscr{C}^0(Q_R)} + CR^{2n+2+2\alpha}([f]_{\mathscr{C}_\beta^{\alpha,\alpha/2}(Q_R)})^2. \quad (4\text{-}25)$$

By a standard Moser iteration argument we get the following sub-mean-value inequality.

Lemma 4.18. *If in addition $f \equiv 0$, then there exists a constant $C = C(n, \boldsymbol{\beta}) > 0$ such that*

$$\sup_{Q_\rho} |u| \le C\biggl(\frac{1}{(R-\rho)^{2n+2}} \iint_{Q_R} u^2 \omega_{\boldsymbol{\beta}}^n \, dt\biggr)^{1/2}.$$

Proof. For any $p \ge 1$, multiplying both sides of the equation by $\eta^2 u_+^p$, where $u_+ = \max\{u, 0\}$, and integrating by parts, we get

$$\frac{d}{dt} \int_{B_R} \frac{\eta^2}{p+1} u_+^{p+1} = \int_{B_R} -p\eta^2 u_+^{p-1} |\nabla u_+|^2 - 2\eta u_+^p \langle \nabla u_+, \nabla \eta \rangle.$$

By the Cauchy–Schwarz inequality and integrating over $t \in [0, R^2]$, we conclude that

$$\sup_{s \in [0, R^2]} \int_{B_R} \eta^2 u_+^{p+1} \bigg|_{t=s} + \iint_{Q_R} |\nabla(\eta u_+^{(p+1)/2})|^2 \le \frac{C}{(R-\rho)^2} \iint_{Q_R} u_+^{p+1} \omega_{\boldsymbol{\beta}}^n \, dt =: A.$$

By the Sobolev inequality we get

$$\begin{aligned} \int_0^{R^2} \int_{B_R} (\eta^2 u_+^{p+1})^{1+1/n} &\le \int_0^{R^2} \biggl(\int_{B_R} \eta^2 u_+^{p+1}\biggr)^{1/n} \biggl(\int_{B_R} (\eta u_+^{(p+1)/2})^{2n/(n-1)}\biggr)^{(n-1)/n} \\ &\le A^{1/n} C \int_0^{R^2} \int_{B_R} |\nabla(\eta u_+^{(p+1)/2})|^2 \le C A^{(n+1)/n}. \end{aligned}$$

If we write

$$H(p, \rho) = \biggl(\int_0^{\rho^2} \int_{B_\rho} u_+^p\biggr)^{1/p},$$

the inequality above implies

$$H((p+1)\xi, \rho) \le \frac{C^{1/(p+1)}}{(R-\rho)^{2/(p+1)}} H(p+1, R),$$

where $\xi = (n+1)/n > 1$. Writing $p_k + 1 = 2\xi^k$ and $\rho_k = \rho + (R-\rho)2^{-k}$, we then have $H(p_{k+1}+1, \rho_{k+1}) \le H(p_k + 1, \rho_k)$. Iterating this inequality we get

$$H(\infty, \rho) = \sup_{Q_\rho} u_+ \le \frac{C}{(R-\rho)^{n+1}} \biggl(\iint_{Q_R} u_+^2\biggr)^{1/2}.$$

Similarly we get the same inequality for $u_- = \max\{-u, 0\}$. □

Corollary 4.19. *If in addition $f \equiv 0$, then there is a constant $C = C(n, \boldsymbol{\beta}) > 0$ such that*

$$\iint_{Q_\rho} u^2 \omega_{\boldsymbol{\beta}}^n \, dt \le C\Bigl(\frac{\rho}{R}\Bigr)^{2+2n} \iint_{Q_R} u^2 \omega_{\boldsymbol{\beta}}^n \, dt. \tag{4-26}$$

Proof. When $\rho \in \bigl[\frac{1}{2}R, R\bigr]$, the inequality is trivial; when $\rho \in \bigl[0, \frac{1}{2}R\bigr)$, it follows from Lemma 4.18. □

Lemma 4.20. *If in addition $f \equiv 0$, then there is a constant $C = C(n, \boldsymbol{\beta}) > 0$ such that, for any $\rho \in (0, R)$,*

$$\iint_{Q_\rho} u^2 \omega_{\boldsymbol{\beta}}^n \, dt \le C\Bigl(\frac{\rho}{R}\Bigr)^{2n+4} \iint_{Q_R} u^2 \omega_{\boldsymbol{\beta}}^n \, dt.$$

Proof. The inequality is trivial in the case $\rho \in \left[\frac{1}{2}R, R\right]$, so we assume $\rho < \frac{1}{2}R$. First we observe that $\Delta_{\boldsymbol\beta} u$ also satisfies the equations $\partial_t(\Delta_{\boldsymbol\beta} u) = \Delta_{\boldsymbol\beta}(\Delta_{\boldsymbol\beta} u)$ and $(\Delta_{\boldsymbol\beta} u)|_{t=0} \equiv 0$, so (4-26) holds with u^2 replaced by $(\Delta_{\boldsymbol\beta} u)^2$, i.e.,

$$\iint_{Q_\rho} (\Delta_{\boldsymbol\beta} u)^2 \omega_{\boldsymbol\beta}^n \, dt \le C\Big(\frac{\rho}{R}\Big)^{2+2n} \iint_{Q_R} (\Delta_{\boldsymbol\beta} u)^2 \omega_{\boldsymbol\beta}^n \, dt.$$

Since $u|_{t=0} = 0$, we have $u(x,t) = \int_0^t \partial_s u(x,s)\, ds$, and we calculate

$$\begin{aligned}\iint_{Q_\rho} u^2 \le \rho^4 \iint_{Q_\rho} \left|\frac{\partial u}{\partial t}\right|^2 = \rho^4 \iint_{Q_\rho} (\Delta_{\boldsymbol\beta} u)^2 &\le C\rho^4 \Big(\frac{\rho}{R}\Big)^{2n+2} \iint_{Q_{R/2}} (\Delta_{\boldsymbol\beta} u)^2 \\ &\le C\Big(\frac{\rho}{R}\Big)^{2n+6} \iint_{Q_R} u^2 \omega_{\boldsymbol\beta}^n \, dt \qquad \text{(by (4-25)).} \qquad \square\end{aligned}$$

Lemma 4.21. *Let u be a solution to* (4-22). *There exists a constant $C = C(n, \boldsymbol\beta, \alpha) > 0$ such that*

$$\frac{1}{\rho^{2n+2+2\alpha}} \iint_{Q_\rho} (\Delta_{\boldsymbol\beta} u)^2 \le \frac{C}{R^{2n+2+2\alpha}} \iint_{Q_R} (\Delta_{\boldsymbol\beta} u)^2 \omega_{\boldsymbol\beta}^n \, dt + C([f]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(Q_R)})^2.$$

Proof. Let $u = u_1 + u_2$, where

$$\frac{\partial u_1}{\partial t} = \Delta_{\boldsymbol\beta} u_1 + f_R \quad \text{in } Q_R, \qquad u_1 = u \quad \text{on } \partial_{\mathcal P} Q_R,$$

and

$$\frac{\partial u_2}{\partial t} = \Delta_{\boldsymbol\beta} u_2 + f - f_R \quad \text{in } Q_R, \qquad u_2 = 0 \quad \text{on } \partial_{\mathcal P} Q_R.$$

The function $\Delta_{\boldsymbol\beta} u_1$ satisfies the assumptions of Lemma 4.20. Thus

$$\iint_{Q_\rho} (\Delta_{\boldsymbol\beta} u_1)^2 \omega_{\boldsymbol\beta}^n \, dt \le C\Big(\frac{\rho}{R}\Big)^{2n+4} \iint_{Q_R} (\Delta_{\boldsymbol\beta} u_1)^2 \omega_{\boldsymbol\beta}^n \, dt.$$

Multiplying both sides of the equation for u_2 by $\dot u_2 = \partial u_2/\partial t$ and noting that $\dot u_2 = 0$ on $\partial B_R \times (0, R^2)$, we get

$$\begin{aligned}\int_{B_R} (\dot u_2)^2 = \int_{B_R} \dot u_2 \Delta_{\boldsymbol\beta} u_2 + \dot u_2 (f - f_R) &= \int_{B_R} -2\langle \nabla \dot u_2, \nabla u_2\rangle + \dot u (f - f_R) \\ &\le \int_{B_R} -\frac{\partial}{\partial t}|\nabla u_2|^2 + \tfrac{1}{2}(\dot u_2)^2 + 2(f - f_R)^2.\end{aligned}$$

Integrating over $t \in [0, R^2]$, we obtain

$$\iint_{Q_R} (\dot u_2)^2 \le -2 \int_{B_R} |\nabla u_2|^2 \bigg|_{t=R^2} + 4 \iint_{Q_R} (f - f_R)^2,$$

therefore

$$\iint_{Q_R} (\Delta_{\boldsymbol\beta} u_2)^2 \le 2 \iint_{Q_R} (\dot u_2)^2 + 2 \iint_{Q_R} (f - f_R)^2 \le C R^{2n+2+2\alpha} ([f]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(Q_R)})^2.$$

Then for $\rho < R$ we have

$$\iint_{Q_\rho} (\Delta_{\boldsymbol\beta} u)^2 \le 2\iint_{Q_\rho} (\Delta_{\boldsymbol\beta} u_1)^2 + 2\iint_{Q_\rho} (\Delta_{\boldsymbol\beta} u_2)^2$$
$$\le C\Big(\frac{\rho}{R}\Big)^{2n+4} \iint_{Q_R} (\Delta_{\boldsymbol\beta} u_1)^2 \omega_{\boldsymbol\beta}^n\, dt + CR^{2n+2+2\alpha}([f]_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(Q_R)})^2.$$

The estimate is proved by an iteration lemma (see Lemma 3.4 in [22]). □

Lemma 4.22. *Suppose u satisfies* (4-22). *There exists a constant $C = C(n, \boldsymbol\beta, \alpha) > 0$ such that, for any $\rho \in \big(0, \frac12 R\big)$,*

$$\iint_{Q_\rho} (\Delta_{\boldsymbol\beta} u - (\Delta_{\boldsymbol\beta} u)_\rho)^2 \omega_{\boldsymbol\beta}^n\, dt \le C M_R \rho^{2n+2+2\alpha},$$

where

$$M_R := \frac{1}{R^{4+2\alpha}}\|u\|^2_{\mathscr{C}^0(Q_R)} + \frac{1}{R^{2\alpha}}\|f\|^2_{\mathscr{C}^0(Q_R)} + ([f]_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(Q_R)})^2.$$

Proof. From Lemma 4.21, we get

$$\iint_{Q_\rho} (\Delta_{\boldsymbol\beta} u)^2 \le C\rho^{2+2n+2\alpha}\bigg(\frac{1}{R^{2n+2+2\alpha}}\iint_{Q_{2R/3}} (\Delta_{\boldsymbol\beta} u)^2 + ([f]_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(Q_{2R/3})})^2\bigg)$$
$$\le C\rho^{2+2n+2\alpha}\bigg(\frac{1}{R^{2n+6+2\alpha}}\iint_{Q_R} u^2 + \frac{1}{R^{2\alpha}}\|f\|^2_{C^0(Q_R)} + ([f]_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(Q_R)})^2\bigg) \quad \text{(by (4-25))}$$
$$\le C\rho^{2+2n+2\alpha} M_R.$$

On the other hand, by the Hölder inequality,

$$(\Delta_{\boldsymbol\beta} u)_\rho^2 = \frac{1}{|Q_\rho|^2_{g_{\boldsymbol\beta}}}\bigg(\iint_{Q_\rho} (\Delta_{\boldsymbol\beta} u)\omega_{\boldsymbol\beta}^n\, dt\bigg)^2 \le \frac{C}{\rho^{2+2n}}\iint_{Q_\rho} (\Delta_{\boldsymbol\beta} u)^2 \le C M_R \rho^{2\alpha}.$$

The lemma is proved by combining the two inequalities above. □

By Campanato's lemma (see Theorem 3.1 in Chapter 3 of [22]), we get the following.

Corollary 4.23. *There is a constant $C = C(n, \boldsymbol\beta, \alpha) > 0$ such that, for any $x \in B_{\boldsymbol\beta}\big(0, \frac34\big)$ and $R < \frac{1}{10}$,*

$$[\Delta_{\boldsymbol\beta} u]_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(B_{\boldsymbol\beta}(x,R/2)\times[0,R^2/4])}$$
$$\le C\bigg(\frac{1}{R^{2+\alpha}}\|u\|_{\mathscr{C}^0(B_{\boldsymbol\beta}(x,R)\times[0,R^2])} + \frac{1}{R^\alpha}\|f\|_{\mathscr{C}^0(B_{\boldsymbol\beta}(x,R)\times[0,R^2])} + [f]_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(B_{\boldsymbol\beta}(x,R)\times[0,R^2])}\bigg). \tag{4-27}$$

Lemma 4.24. *There exists a constant $C = C(n, \boldsymbol\beta, \alpha) > 0$ such that, for any $x \in B_{\boldsymbol\beta}\big(0, \frac34\big)$ and $R < \frac{1}{10}$,*

$$[Tu]_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(B_{\boldsymbol\beta}(x,R/2)\times[0,R^2/4])} + \bigg[\frac{\partial u}{\partial t}\bigg]_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(B_{\boldsymbol\beta}(x,R/2)\times[0,R^2/4])}$$
$$\le C\bigg(\frac{1}{R^{2+\alpha}}\|u\|_{\mathscr{C}^0(B_{\boldsymbol\beta}(x,R)\times[0,R^2])} + \frac{1}{R^\alpha}\|f\|_{\mathscr{C}^0(B_{\boldsymbol\beta}(x,R)\times[0,R^2])} + [f]_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol\beta}(B_{\boldsymbol\beta}(x,R)\times[0,R^2])}\bigg). \tag{4-28}$$

Proof. It follows from (4-27) and the elliptic Schauder estimates in Theorem 1.2 by adjusting R slightly that, for any $t \in \left[0, \frac{1}{4}R^2\right]$,

$$[Tu(\cdot\,,t)]_{C^{0,\alpha}_{\beta}(B_{\beta}(x,R/2))} \le C\left(\frac{1}{R^{2+\alpha}}\|u\|_{\mathscr{C}^0(B_{\beta}(x,R)\times[0,R^2])} + \frac{1}{R^{\alpha}}\|f\|_{\mathscr{C}^0(B_{\beta}(x,R)\times[0,R^2])} + [f]_{\mathscr{C}^{\alpha,\alpha/2}_{\beta}(B_{\beta}(x,R)\times[0,R^2])}\right),$$

that is, in the spatial variables the estimate (4-28) holds. It only remains to show the Hölder continuity of Tu in the time-variable. For this, we fix any two times $0 \le t_1 < t_2 \le \frac{1}{4}R^2$ and denote $r := \frac{1}{2}\sqrt{t_2 - t_1}$. For any $x_0 \in B_{\beta}\left(x, \frac{1}{4}R\right)$, $B_{\beta}(x_0, r) \subset B_{\beta}\left(x, \frac{1}{2}R\right)$. By (4-27) and the equation for u, it is not hard to see that the inequality (4-27) holds when $\Delta_{\beta} u$ on the left-hand side is replaced by $\dot{u} = \partial u/\partial t$. In particular,

$$\frac{|\dot{u}(y,t) - \dot{u}(y,t_1)|}{|t - t_1|^{\alpha/2}} \le A_R \quad \text{for all } y \in B_{\beta}\left(x, \tfrac{1}{2}R\right),$$

where A_R is defined to be the constant on the right-hand side of (4-27). Integrating over $t \in [t_1, t_2]$ we get

$$|u(y,t_2) - u(y,t_1) - \dot{u}(y,t_1)(t_2 - t_1)| \le CA_R(t_2 - t_1)^{1+\alpha/2}.$$

Thus, for any $y \in B_{\beta}(x_0, r)$,

$$\begin{aligned}|u(y,t_2) - u(y,t_1) - \dot{u}(x_0,t_1)(t_2 - t_1)| &\le |u(y,t_2) - u(y,t_1) - \dot{u}(y,t_1)(t_2 - t_1)| + |\dot{u}(x_0,t_1) - \dot{u}(y,t_1)|(t_2 - t_1)\\ &\le CA_R(t_2 - t_1)^{1+\alpha/2} + A_R r^{\alpha}(t_2 - t_1).\end{aligned}$$

Write

$$\tilde{u}(y) := u(y,t_2) - u(y,t_1) - \dot{u}(x_0,t_1)(t_2 - t_1),$$

which is a function on $B_{\beta}(x_0, r)$. We have that the function $\tilde{f} := \Delta_{\beta}\tilde{u} = \Delta_{\beta}u(\cdot\,,t_2) - \Delta_{\beta}u(\cdot\,,t_1)$ satisfies the inequalities $\|\tilde{f}\|_{C^0(B_{\beta}(x_0,r))} \le A_R(t_2 - t_1)^{\alpha}$ and $[\tilde{f}]_{C^{0,\alpha}_{\beta}(B_{\beta}(x_0,r))} \le A_R$ by (4-27). It follows from the rescaled version of Proposition 3.32 that

$$|T\tilde{u}|_{C^0(B_{\beta}(x_0,r/2))} \le C(n,\boldsymbol{\beta},\alpha)\left(\frac{\|\tilde{u}\|_{C^0(B_{\beta}(x_0,r))}}{r^2} + \|\tilde{f}\|_{C^0(B_{\beta}(x_0,r))} + r^{\alpha}[\tilde{f}]_{C^{0,\alpha}(B_{\beta}(x_0,r))}\right) \le C(t_2 - t_1)^{\alpha/2}A_R.$$

Therefore, for any $x_0 \in B_{\beta}\left(x, \frac{1}{4}R\right)$,

$$\frac{|Tu(x_0,t_2) - Tu(x_0,t_1)|}{|t_2 - t_1|^{\alpha/2}} \le CA_R.$$

It is then easy to see by the triangle inequality that (adjusting R slightly if necessary)

$$[Tu]_{\mathscr{C}^{\alpha,\alpha/2}_{\beta}(B_{\beta}(x,R/2)\times[0,R^2/4])} \le CA_R,$$

as desired. The estimate for $\dot{u}$ follows from the equation $\dot{u} = \Delta_g u + f$. □

Remark 4.25. By a simple parabolic rescaling of the metric and time, we see from (4-28) that, for any $0 < r < R < \frac{1}{10}$,

$$[Tu]_{\mathscr{C}^{\alpha,\alpha/2}_{\beta}(Q_r)} \le C\left(\frac{\|u\|_{\mathscr{C}^0(Q_R)}}{(R-r)^{2+\alpha}} + \frac{\|f\|_{\mathscr{C}^0(Q_R)}}{(R-r)^{\alpha}} + [f]_{\mathscr{C}^{\alpha,\alpha/2}_{\beta}(Q_R)}\right). \tag{4-29}$$

4D2. *The nonflat metric case.* In this subsection, we will consider the case when the background metrics are general nonflat $\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}$-conical Kähler metrics $g = g(z,t)$. Suppose $u \in \mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(2+\alpha)/2}(\mathcal{Q}_{\boldsymbol\beta})$ satisfies the equation

$$\frac{\partial u}{\partial t} = \Delta_g u + f \quad \text{in } \mathcal{Q}_{\boldsymbol\beta}, \qquad u|_{t=0} = 0, \tag{4-30}$$

and $u \in \mathscr{C}^0(\partial_{\mathcal{P}}\mathcal{Q}_{\boldsymbol\beta})$.

Proposition 4.26. *There exists a constant $C = C(n, \boldsymbol\beta, \alpha, g) > 0$ such that*

$$\|u\|_{\mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(2+\alpha)/2}(B_{\boldsymbol\beta}(0,1/2)\times[0,1/4])} \le C(\|u\|_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol\beta})} + \|f\|_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\mathcal{Q}_{\boldsymbol\beta})}).$$

Proof. Choosing suitable complex coordinates at the origin $x = 0$, we may assume the components of g in the basis $\{\epsilon_j \wedge \bar\epsilon_k, \dots\}$ satisfy $g_{\epsilon_j\bar\epsilon_k}(\,\cdot\,,0) = \delta_{jk}$ and $g_{j\bar k}(\,\cdot\,,0) = \delta_{jk}$ at the origin. As in the proof of Proposition 4.7, we can write (4-30) as

$$\frac{\partial u}{\partial t} = \Delta_{\boldsymbol\beta} u + \eta\cdot\sqrt{-1}\partial\bar\partial u + f =: \Delta_{\boldsymbol\beta} u + \hat f,$$

where η is given in the proof of Proposition 3.32. By (4-29) we get

$$[Tu]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\widetilde Q_r)} \le C\biggl(\frac{\|u\|_{\mathscr{C}^0(\widetilde Q_R)}}{(R-r)^{2+\alpha}} + \frac{1}{(R-r)^\alpha}\|\hat f\|_{\mathscr{C}^0(\widetilde Q_R)} + [\hat f]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\widetilde Q_R)}\biggr),$$

where $\widetilde Q_R := B_{\boldsymbol\beta}(0,R)\times[0,R^2]$. Observe that

$$\begin{aligned}\frac{1}{(R-r)^\alpha}\|\hat f\|_{\mathscr{C}^0(\widetilde Q_R)} &\le \frac{1}{(R-r)^\alpha}\|f\|_{\mathscr{C}^0(\widetilde Q_R)} + \frac{1}{(R-r)^\alpha}\|\eta\|_{\mathscr{C}^0(\widetilde Q_R)}\|Tu\|_{C^0(\widetilde Q_R)}\\ &\le \frac{1}{(R-r)^\alpha}\|f\|_{\mathscr{C}^0(\widetilde Q_R)} + \frac{[\eta]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\widetilde Q_R)}R^\alpha}{(R-r)^\alpha}(\varepsilon[Tu]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\widetilde Q_R)} + C(\varepsilon)\|u\|_{\mathscr{C}^0(\widetilde Q_R)})\end{aligned}$$

and

$$\begin{aligned}[\hat f]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\widetilde Q_R)} &\le [f]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\widetilde Q_R)} + \|\eta\|_{\mathscr{C}^0(\widetilde Q_R)}[Tu]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\widetilde Q_R)} + \|Tu\|_{\mathscr{C}^0(\widetilde Q_R)}[\eta]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\widetilde Q_R)}\\ &\le [f]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\widetilde Q_R)} + [\eta]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\widetilde Q_R)}R^\alpha[Tu]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\widetilde Q_R)}\\ &\qquad + [\eta]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\widetilde Q_R)}(\varepsilon[Tu]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\widetilde Q_R)} + C(\varepsilon)\|u\|_{\mathscr{C}^0(\widetilde Q_R)}).\end{aligned}$$

By choosing $R_0 = R_0(n,\boldsymbol\beta,\alpha,g) > 0$ small enough and suitable $\varepsilon > 0$, for any $0 < r < R < R_0 < \frac{1}{10}$, the combination of the above inequalities yields

$$[Tu]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\widetilde Q_r)} \le \tfrac12[Tu]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\widetilde Q_R)} + C\biggl(\frac{\|u\|_{\mathscr{C}^0(\widetilde Q_R)}}{(R-r)^{2+\alpha}} + \frac{1}{(R-r)^\alpha}\|f\|_{\mathscr{C}^0(\widetilde Q_R)} + [f]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\widetilde Q_R)}\biggr).$$

By Lemma 4.27 below (setting $\phi(r) = [Tu]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\widetilde Q_r)}$), we conclude that

$$[Tu]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(B_{\boldsymbol\beta}(0,R_0/2)\times[0,R_0^2/4])} \le C(\|u\|_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol\beta})} + \|f\|_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\mathcal{Q}_{\boldsymbol\beta})}).$$

This is the desired estimate when the center of the ball is the worst possible. For the other balls $B_{\boldsymbol\beta}(x,r)$ with center $x \in B_{\boldsymbol\beta}\bigl(0,\frac12\bigr)$, we can repeat the above procedures and use the smooth coordinates $w_j = z_j^{\beta_j}$

in case the ball is disjoint from $\mathcal{S}_j$. Finitely many such balls cover $B_{\boldsymbol\beta}\big(0, \frac{1}{2}\big)$, so we get

$$[Tu]_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(B_{\boldsymbol\beta}(0,1/2)\times[0,1/100])} \leq C(\|u\|_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol\beta})} + \|f\|_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\mathcal{Q}_{\boldsymbol\beta})}).$$

The proposition is proved by combining this inequality, the equation for u, interpolation inequalities, and the interior Schauder estimates in Corollary 4.9. □

Lemma 4.27 [22, Lemma 4.3]. *Let $\phi(t) \geq 0$ be bounded in $[0, T]$. Suppose, for any $0 < t < s \leq T$, we have*

$$\phi(t) \leq \tfrac{1}{2}\phi(s) + \frac{A}{(s-t)^a} + B$$

for some $a > 0$ and $A, B > 0$. Then it holds that, for any $0 < t < s \leq T$,

$$\phi(t) \leq c(a)\bigg(\frac{A}{(s-t)^a} + B\bigg).$$

Corollary 4.28. *Suppose u satisfies the equations*

$$\frac{\partial u}{\partial t} = \Delta_g u + f \quad \text{in } \mathcal{Q}_{\boldsymbol\beta} \qquad u|_{t=0} = u_0 \in C_{\boldsymbol\beta}^{2,\alpha}(B_{\boldsymbol\beta}(0, 1)).$$

Then

$$\|u\|_{\mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(\alpha+2)/2}(B_{\boldsymbol\beta}(0,1/2)\times[0,1])} \leq C(\|u\|_{\mathscr{C}^0(\mathcal{Q}_{\boldsymbol\beta})} + \|f\|_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(\mathcal{Q}_{\boldsymbol\beta})} + \|u_0\|_{C_{\boldsymbol\beta}^{2,\alpha}(B_{\boldsymbol\beta}(0,1))})$$

for some constant $C = C(n, \boldsymbol\beta, \alpha, g) > 0$.

Proof. We set $\hat u = u - u_0$ and $\hat f = f - \Delta_g u_0$. $\hat u$ satisfies the conditions in Proposition 4.26, so the corollary follows from Proposition 4.26 applied to $\hat u$ and triangle inequalities. □

Corollary 4.29. *In addition to the assumptions in Corollary 4.16, we also assume that $u_0 \in C_{\boldsymbol\beta}^{2,\alpha}(X)$. Then the weak solution to $\partial u/\partial t = \Delta_g u + f$ with $u|_{t=0} = u_0$ exists and is in $\mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(2+\alpha)/2}(X, \times[0, 1])$. Moreover, there is a $C = C(n, \boldsymbol\beta, \alpha, g) > 0$ such that*

$$\|u\|_{\mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(2+\alpha)/2}(X\times[0,1])} \leq C(\|f\|_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(X\times[0,1])} + \|u_0\|_{C_{\boldsymbol\beta}^{2,\alpha}(X)}). \tag{4-31}$$

Proof. Observe that by the maximum principle we have

$$\|u\|_{\mathscr{C}^0(X\times[0,1])} \leq \|f\|_{\mathscr{C}^0(X\times[0,1])} + \|u_0\|_{C^0(X)}.$$

Then (4-31) follows from Corollary 4.28 and a covering argument as in the proof of Corollary 3.41. □

5. Conical Kähler–Ricci flow

Let X be a compact Kähler manifold and $D = \sum_j D_j$ be a divisor with simple normal crossings. Let ω_0 be a fixed $C_{\boldsymbol\beta}^{0,\alpha'}(X)$ conical Kähler metric with cone angle $2\pi\boldsymbol\beta$ along D and $\hat\omega_t$ be a family of $\mathscr{C}_{\boldsymbol\beta}^{\alpha',\alpha'/2}$ conical metrics which are uniformly equivalent to ω_0, with $\hat\omega_0 = \omega_0$ and $\|\hat\omega\|_{\mathscr{C}_{\boldsymbol\beta}^{\alpha',\alpha'/2}(X\times[0,1])} \leq C_0$. We consider the complex Monge–Ampère equation

$$\begin{cases} \dfrac{\partial\varphi}{\partial t} = \log\bigg(\dfrac{(\hat\omega_t + \sqrt{-1}\partial\bar\partial\varphi)^n}{\omega_0^n}\bigg) + f, \\ \varphi|_{t=0} = 0, \end{cases} \tag{5-1}$$

where $f \in \mathscr{C}^{\alpha',\alpha'/2}{}_{\beta}(X\times[0,1])$ is a given function. We will use an inverse function theorem argument from [4] which was outlined in [21] to show the short-time existence of the flow (5-1).

Theorem 5.1. *There exists a small* $T = T(n, \boldsymbol{\beta}, \omega_0, f, \alpha, \alpha') > 0$ *such that* (5-1) *admits a unique solution* $\varphi \in \mathscr{C}_{\boldsymbol{\beta}}^{2+\alpha,(2+\alpha)/2}(X \times [0, T])$ *for any* $\alpha < \alpha'$.

Proof. The uniqueness of the solution follows from the maximum principle. We will break the proof of short-time existence into three steps.

Step 1. Let $u \in \mathscr{C}_{\boldsymbol{\beta}}^{2+\alpha',(2+\alpha')/2}(X \times [0, 1])$ be the solution to

$$\begin{cases} \dfrac{\partial u}{\partial t} = \Delta_{g_0} u + f & \text{in } X \times [0, 1], \\ u|_{t=0} = 0. \end{cases}$$

Thanks to Corollary 4.29, such a u exists and satisfies the estimate (4-31). We fix $\varepsilon > 0$, so that, as long as $\|\phi\|_{C^{2,\alpha}_{\boldsymbol{\beta}}(X)} \le \varepsilon$, we have that $\hat{\omega}_{t,\phi} := \hat{\omega}_t + \sqrt{-1}\partial\bar{\partial}\phi$ is equivalent to ω_0, i.e.,

$$C_0^{-1}\omega_0 \le \omega_{0,\phi} \le C_0\omega_0 \quad \text{and} \quad \|\hat{\omega}_{t,\phi}\|_{\mathscr{C}^{\alpha,\alpha/2}_{\boldsymbol{\beta}}} \le C_0.$$

We *claim* that, for $T_1 > 0$ small enough, $\|u\|_{\mathscr{C}^{2+\alpha,(2+\alpha)/2}_{\boldsymbol{\beta}}(X\times[0,T_1])} \le \varepsilon$. We first observe by (4-31) that

$$N := \|u\|_{\mathscr{C}^{2+\alpha',(\alpha'+2)/2}_{\boldsymbol{\beta}}(X\times[0,1])} \le C\|f\|_{\mathscr{C}^{\alpha',\alpha'/2}_{\boldsymbol{\beta}}(X\times[0,1])}.$$

It suffices to show that $[u]_{\mathscr{C}^{2+\alpha',(\alpha'+2)/2}_{\boldsymbol{\beta}}(X\times[0,T_1])}$ is small, since the lower-order derivatives are small because $u|_{t=0} = 0$. We calculate, for any $t_1, t_2 \in [0, T_1]$,

$$\frac{|Tu(x,t_1) - Tu(x,t_2)|}{|t_1 - t_2|^{\alpha/2}} + \frac{|\dot{u}(x,t_1) - \dot{u}(x,t_2)|}{|t_1 - t_2|^{\alpha/2}} \le N|t_1 - t_2|^{(\alpha'-\alpha)/2} \le \tfrac{1}{4}\varepsilon$$

if $NT_1^{(\alpha'-\alpha)/2} < \frac{1}{4}\varepsilon$. For any $x, y \in X$ and $t \in [0, T_1]$,

$$\frac{|Tu(x,t) - Tu(y,t)|}{d_{g_0}(x,y)^{\alpha}} \le N \min\left\{\frac{2T_1^{\alpha'/2}}{d_{g_0}(x,y)^{\alpha}}, d_{g_0}(x,y)^{\alpha'-\alpha}\right\} \le \tfrac{1}{2}\varepsilon.$$

The *claim* then follows from the triangle inequality.

We define a function

$$w(x,t) := \frac{\partial u}{\partial t}(x,t) - \log\left(\frac{(\hat{\omega}_t + \sqrt{-1}\partial\bar{\partial}u)^n}{\omega_0^n}\right)(x,t) - f(x,t) \quad \text{for all } (x,t) \in X \times [0, T_1].$$

It is clear that $w(x, 0) \equiv 0$.

Step 2: We consider the small ball

$$\mathcal{B} = \{\phi \in \mathscr{C}_{\boldsymbol{\beta}}^{2+\alpha,(2+\alpha)/2}(X \times [0, T_1]) \mid \|\phi\|_{\mathscr{C}^{2+\alpha,(\alpha+2)/2}_{\boldsymbol{\beta}}} \le \varepsilon,\ \phi(\,\cdot\,, 0) = 0\}$$

in the space $\mathscr{C}_{\boldsymbol{\beta}}^{2+\alpha,(2+\alpha)/2}(X \times [0, T_1])$. Then $u|_{t\in[0,T_1]} \in \mathcal{B}$ by the discussion in Step 1.

Define the differential map $\Psi : \mathcal{B} \to \mathscr{C}_{\boldsymbol{\beta}}^{\alpha,\alpha/2}(X \times [0, T_1])$ by

$$\Psi(\phi) = \frac{\partial\phi}{\partial t} - \log\left(\frac{(\hat{\omega}_t + \sqrt{-1}\partial\bar{\partial}\phi)^n}{\omega_0^n}\right) - f.$$

The map Ψ is well defined and C^1, where the differential $D\Psi_\phi$ at any $\phi \in \mathcal{B}$ is given by

$$D\Psi_\phi(v) = \frac{\partial v}{\partial t} - (\hat{g}_\phi)^{i\bar{j}} v_{i\bar{j}} = \frac{\partial v}{\partial t} - \Delta_{\hat{\omega}_{t,\phi}} v$$

for any

$$v \in T_\phi \mathcal{B} = \{v \in \mathscr{C}_{\boldsymbol{\beta}}^{2+\alpha,(2+\alpha)/2}(X \times [0, T_1]) \mid v(\,\cdot\,, 0) = 0\},$$

where $(\hat{g}_\phi)^{i\bar{j}}$ denotes the inverse of the metric $\hat{\omega}_t + \sqrt{-1}\partial\bar{\partial}\phi$. As a linear map,

$$D\Psi_\phi : T_\phi \mathcal{B} \to \mathscr{C}_{\boldsymbol{\beta}}^{\alpha,\alpha/2}(X \times [0, T_1])$$

is injective by the maximum principle and surjective by Corollary 4.29. Thus $D\Psi_\phi$ is invertible at any $\phi \in \mathcal{B}$. In particular, $D\Psi_u$ is invertible, and by the inverse function theorem, $\Psi : \mathcal{B} \to \mathscr{C}_{\boldsymbol{\beta}}^{\alpha,\alpha/2}(X \times [0, T_1])$ defines a local diffeomorphism from a small neighborhood of $u \in \mathcal{B}$ to an open neighborhood of $w = \Psi(u)$ in $\mathscr{C}_{\boldsymbol{\beta}}^{\alpha,\alpha/2}(X \times [0, T_1])$. This implies that, for any $\tilde{w} \in \mathscr{C}_{\boldsymbol{\beta}}^{\alpha,\alpha/2}(X \times [0, T_1])$ with $\|w - \tilde{w}\|_{\mathscr{C}_{\boldsymbol{\beta}}^{\alpha,\alpha/2}(X \times [0,T_1])} < \delta$ for some small $\delta > 0$, there exists a unique $\varphi \in \mathcal{B}$ such that $\Psi(\varphi) = \tilde{w}$.

Step 3. For a small $T_2 < T_1$ to be determined, we define a function

$$\tilde{w}(x, t) = \begin{cases} 0, & t \in [0, T_2], \\ w(x, t - T_2), & t \in [T_2, T_1]. \end{cases}$$

Since $u \in \mathscr{C}_{\boldsymbol{\beta}}^{2+\alpha',(2+\alpha')/2}$, we see that $w \in \mathscr{C}_{\boldsymbol{\beta}}^{\alpha',\alpha'/2}(X \times [0, T_1])$ with $M := \|w\|_{\mathscr{C}_{\boldsymbol{\beta}}^{\alpha',\alpha'/2}(X \times [0,T_1])} < \infty$. We *claim* that if T_2 is small enough, then $\|w - \tilde{w}\|_{\mathscr{C}_{\boldsymbol{\beta}}^{\alpha,\alpha/2}(X \times [0,T_1])} < \delta$. We write $\eta = w - \tilde{w}$. It is clear from the fact that $w(\,\cdot\,, 0) = 0$ that $\|\eta\|_{\mathscr{C}^0} \le \frac{1}{2}\delta$ if T_2 is small enough.

Spatial directions: If $t < T_2$ then

$$\frac{|\eta(x, t) - \eta(y, t)|}{d_{g_0}(x, y)^\alpha} = \frac{|w(x, t) - w(y, t)|}{d_{g_0}(x, y)^\alpha} \le M \min\left\{\frac{2T_2^{\alpha'/2}}{d_{g_0}(x, y)^\alpha}, d_{g_0}(x, y)^{\alpha'-\alpha}\right\} \le 2MT_2^{(\alpha'-\alpha)/2}.$$

If $t \in [T_2, T_1]$ then

$$\begin{aligned} \frac{|\eta(x, t) - \eta(y, t)|}{d_{g_0}(x, y)^\alpha} &= \frac{|w(x, t) - w(y, t) - w(x, t - T_2) + w(y, t - T_2)|}{d_{g_0}(x, y)^\alpha} \\ &\le 2M \min\left\{\frac{T_2^{\alpha'/2}}{d_{g_0}(x, y)^\alpha}, d_{g_0}(x, y)^{\alpha'-\alpha}\right\} \le 2MT_2^{(\alpha'-\alpha)/2}. \end{aligned}$$

Time direction: If $t, t' < T_2$ then

$$\frac{|\eta(x, t) - \eta(x, t')|}{|t - t'|^{\alpha/2}} = \frac{|w(x, t) - w(x, t')|}{|t - t'|^{\alpha/2}} \le M|t - t'|^{(\alpha'-\alpha)/2} \le MT_2^{(\alpha'-\alpha)/2}.$$

If $t, t' \in [T_2, T_1]$ then

$$\frac{|\eta(x, t) - \eta(x, t')|}{|t - t'|^{\alpha/2}} = \frac{|w(x, t) - w(x, t') - w(x, t - T_2) + w(x, t' - T_2)|}{|t - t'|^{\alpha/2}} \le 2MT_2^{(\alpha'-\alpha)/2}.$$

If $t < T_2 \le t' \le T_1$ then

$$\frac{|\eta(x,t)-\eta(x,t')|}{|t-t'|^{\alpha/2}} = \frac{|w(x,t)-w(x,t')+w(x,t'-T_2)|}{|t-t'|^{\alpha/2}} \le 2MT_2^{(\alpha'-\alpha)/2}.$$

Therefore, if we choose $T_2 > 0$ small enough that $2MT_2^{(\alpha'-\alpha)/2} < \frac{1}{4}\delta$, then we have

$$\frac{|\eta(x,t)-\eta(x,t')|}{|t-t'|^{\alpha/2}} + \frac{|\eta(x,t)-\eta(y,t)|}{d_{g_0}(x,y)^\alpha} \le \tfrac{1}{2}\delta \quad \text{for all } x \in X, \quad t,t' \in [0,T_1].$$

It then follows from the triangle inequality that

$$\begin{aligned}|\eta(x,t)-\eta(y,t')| &\le |\eta(x,t)-\eta(y,t)| + |\eta(y,t)-\eta(y,t')|\\ &\le \tfrac{1}{2}\delta(d_{g_0}(x,y)^\alpha + |t-t'|^{\alpha/2}) \le \tfrac{1}{2}\delta d_{\mathcal{P},g_0}((x,t),(y,t'))^\alpha.\end{aligned}$$

In conclusion, $\|\tilde{w}-w\|_{\mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(X\times[0,T_1])} < \delta$, so by Step 2 we conclude that there exists a $\varphi \in \mathcal{B}$ such that $\Psi(\varphi) = \tilde{w}$. Since $\tilde{w}|_{t\in[0,T_2]} \equiv 0$ by definition, $\varphi|_{t\in[0,T_2]}$ satisfies (5-1) for $t \in [0,T]$, where $T := T_2$. This shows the short-time existence of the flow (5-1). □

Proof of Corollary 1.11. Recall that in (1-13) we wrote $\omega_0^n = \Omega/\prod_j(|s_j|^2_{h_j})^{1-\beta_j}$, where Ω is a smooth volume form, s_j and h_j are holomorphic sections and hermitian metrics, respectively, of the line bundle associated to the component D_j. Choose a smooth reference form

$$\chi = \sqrt{-1}\partial\bar\partial \log\Omega - \sum_j (1-\beta_j)\sqrt{-1}\partial\bar\partial \log h_j.$$

Define the reference metrics $\hat\omega_t = \omega_0 + t\chi$ which are $\mathscr{C}_{\boldsymbol\beta}^{\alpha',\alpha'/2}$-conical and Kähler for small $t > 0$. Let φ be the $\mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(2+\alpha)/2}$-solution to (1-11) with $f \equiv 0$. Then it is straightforward to check that $\omega_t = \hat\omega_t + \sqrt{-1}\partial\bar\partial\varphi$ satisfies the conical Kähler–Ricci flow equation (1-12) and $\omega \in \mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(X\times[0,T])$ for some small $T > 0$.

The smoothness of ω in $X\backslash D \times (0,T]$ follows from the general smoothing properties of parabolic equations; see [37]. Taking $\partial/\partial t$ on both sides of (1-11) we get

$$\frac{\partial\dot\varphi}{\partial t} = \Delta_{\omega_t}\dot\varphi + \operatorname{tr}_{\omega_t}\chi \quad \text{and} \quad \dot\varphi|_{t=0} = 0.$$

By Corollary 4.29, $\dot\varphi \in \mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(2+\alpha)/2}(X\times[0,T])$ since $\operatorname{tr}_{\omega_t}\chi \in \mathscr{C}_{\boldsymbol\beta}^{\alpha,\alpha/2}(X\times[0,T])$. Therefore the normalized Ricci potential $\log(\omega_t^n/\omega_0^n)$ exists in $\mathscr{C}_{\boldsymbol\beta}^{2+\alpha,(2+\alpha)/2}(X\times[0,T])$. □

References

[1] S. Brendle, "Ricci flat Kähler metrics with edge singularities", *Int. Math. Res. Not.* **2013**:24 (2013), 5727–5766. MR Zbl

[2] L. A. Caffarelli, "Interior a priori estimates for solutions of fully nonlinear equations", *Ann. of Math.* (2) **130**:1 (1989), 189–213. MR Zbl

[3] L. A. Caffarelli, "Interior $W^{2,p}$ estimates for solutions of the Monge–Ampère equation", *Ann. of Math.* (2) **131**:1 (1990), 135–150. MR Zbl

[4] A. Chau and L.-F. Tam, "On a modified parabolic complex Monge–Ampère equation with applications", *Math. Z.* **269**:3-4 (2011), 777–800. MR Zbl

[5] X. Chen and Y. Wang, "Bessel functions, heat kernel and the conical Kähler–Ricci flow", *J. Funct. Anal.* **269**:2 (2015), 551–632. MR Zbl

[6] X. Chen and Y. Wang, "On the long time behaviour of the conical Kähler–Ricci flows", *J. Reine Angew. Math.* **744** (2018), 165–199. MR Zbl

[7] X. Chen, S. Donaldson, and S. Sun, "Kähler–Einstein metrics on Fano manifolds, I: Approximation of metrics with cone singularities", *J. Amer. Math. Soc.* **28**:1 (2015), 183–197. MR Zbl

[8] X. Chen, S. Donaldson, and S. Sun, "Kähler–Einstein metrics on Fano manifolds, II: Limits with cone angle less than 2π", *J. Amer. Math. Soc.* **28**:1 (2015), 199–234. MR Zbl

[9] X. Chen, S. Donaldson, and S. Sun, "Kähler–Einstein metrics on Fano manifolds, III: Limits as cone angle approaches 2π and completion of the main proof", *J. Amer. Math. Soc.* **28**:1 (2015), 235–278. MR Zbl

[10] S. Y. Cheng and S. T. Yau, "Differential equations on Riemannian manifolds and their geometric applications", *Comm. Pure Appl. Math.* **28**:3 (1975), 333–354. MR Zbl

[11] V. V. Datar, "On convexity of the regular set of conical Kähler–Einstein metrics", *Math. Res. Lett.* **23**:1 (2016), 105–126. MR Zbl

[12] V. V. Datar and J. Song, "A remark on Kähler metrics with conical singularities along a simple normal crossing divisor", *Bull. Lond. Math. Soc.* **47**:6 (2015), 1010–1013. MR Zbl

[13] V. Datar, B. Guo, J. Song, and X. Wang, "Connecting toric manifolds by conical Kähler–Einstein metrics", *Adv. Math.* **323** (2018), 38–83. MR Zbl

[14] S. K. Donaldson, "Kähler metrics with cone singularities along a divisor", pp. 49–79 in *Essays in mathematics and its applications*, edited by P. M. Pardalos and T. M. Rassias, Springer, 2012. MR Zbl

[15] G. Edwards, "A scalar curvature bound along the conical Kähler–Ricci flow", *J. Geom. Anal.* **28**:1 (2018), 225–252. MR Zbl

[16] G. Edwards, "Metric contraction of the cone divisor by the conical Kähler–Ricci flow", *Math. Ann.* **374**:3-4 (2019), 1525–1557. MR Zbl

[17] L. C. Evans, *Partial differential equations*, 2nd ed., Grad. Stud. Math. **19**, Amer. Math. Soc., Providence, RI, 2010. MR Zbl

[18] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Springer, 2001. MR Zbl

[19] H. Guenancia and M. Păun, "Conic singularities metrics with prescribed Ricci curvature: general cone angles along normal crossing divisors", *J. Differential Geom.* **103**:1 (2016), 15–57. MR Zbl

[20] B. Guo and J. Song, "Schauder estimates for equations with cone metrics, I", *Indiana Univ. Math. J.* **70**:5 (2021), 1639–1676. MR Zbl

[21] R. S. Hamilton, "Three-manifolds with positive Ricci curvature", *J. Differential Geom.* **17**:2 (1982), 255–306. MR Zbl

[22] Q. Han and F. Lin, *Elliptic partial differential equations*, Courant Lect. Notes in Math. **1**, Courant Inst. Math. Sci., New York, 1997. MR Zbl

[23] T. Jeffres, R. Mazzeo, and Y. A. Rubinstein, "Kähler–Einstein metrics with edge singularities", *Ann. of Math.* (2) **183**:1 (2016), 95–176. MR Zbl

[24] X. Jin, J. Liu, and X. Zhang, "Twisted and conical Kähler–Ricci solitons on Fano manifolds", *J. Funct. Anal.* **271**:9 (2016), 2396–2421. MR Zbl

[25] P. Li, *Geometric analysis*, Cambridge Stud. Adv. Math. **134**, Cambridge Univ. Press, 2012. MR Zbl

[26] P. Li and S.-T. Yau, "On the parabolic kernel of the Schrödinger operator", *Acta Math.* **156**:3-4 (1986), 153–201. MR Zbl

[27] G. M. Lieberman, *Second order parabolic differential equations*, World Sci., River Edge, NJ, 1996. MR Zbl

[28] F. Luo and G. Tian, "Liouville equation and spherical convex polytopes", *Proc. Amer. Math. Soc.* **116**:4 (1992), 1119–1129. MR Zbl

[29] R. Mazzeo, "Kähler–Einstein metrics singular along a smooth divisor", exposé 6 in *Journées équations aux dérivées partielles* (Saint-Jean-de-Monts, France, 1999), Univ. Nantes, 1999. MR Zbl

[30] R. Mazzeo, Y. A. Rubinstein, and N. Sesum, "Ricci flow on surfaces with conic singularities", *Anal. PDE* **8**:4 (2015), 839–882. MR Zbl

[31] D. H. Phong, J. Song, J. Sturm, and X. Wang, "The Ricci flow on the sphere with marked points", *J. Differential Geom.* **114**:1 (2020), 117–170. MR Zbl

[32] Y. A. Rubinstein, "Smooth and singular Kähler–Einstein metrics", pp. 45–138 in *Geometric and spectral analysis* (Montreal, 2012), edited by P. Albin et al., Contemp. Math. **630**, Amer. Math. Soc., Providence, RI, 2014. MR Zbl

[33] M. V. Safonov, "The classical solution of the elliptic Bellman equation", *Dokl. Akad. Nauk SSSR* **278**:4 (1984), 810–813. In Russian; translated in *Soviet Math. Dokl.* **30**:2 (1984), 482–485. MR Zbl

[34] M. V. Safonov, "Classical solution of second-order nonlinear elliptic equations", *Izv. Akad. Nauk SSSR Ser. Mat.* **52**:6 (1988), 1272–1287. In Russian; translated in *Math. USSR-Izv.* **33**:3 (1989), 597–612. MR Zbl

[35] R. Schoen and S.-T. Yau, *Lectures on differential geometry*, Conf. Proc. Lect. Notes Geom. Topol. **1**, International Press, Cambridge, MA, 1994. MR Zbl

[36] L. Simon, "Schauder estimates by scaling", *Calc. Var. Partial Differential Equations* **5**:5 (1997), 391–407. MR Zbl

[37] J. Song and G. Tian, "The Kähler–Ricci flow through singularities", *Invent. Math.* **207**:2 (2017), 519–595. MR Zbl

[38] J. Song and X. Wang, "The greatest Ricci lower bound, conical Einstein metrics and Chern number inequality", *Geom. Topol.* **20**:1 (2016), 49–102. MR Zbl

[39] G. Tian, "Kähler–Einstein metrics on algebraic manifolds", pp. 143–185 in *Transcendental methods in algebraic geometry* (Cetraro, Italy, 1994), edited by F. Catanese and C. Ciliberto, Lecture Notes in Math. **1646**, Springer, 1996. MR Zbl

[40] G. Tian, "K-stability and Kähler–Einstein metrics", *Comm. Pure Appl. Math.* **68**:7 (2015), 1085–1156. MR Zbl

[41] M. Troyanov, "Prescribing curvature on compact surfaces with conical singularities", *Trans. Amer. Math. Soc.* **324**:2 (1991), 793–821. MR Zbl

[42] X.-J. Wang, "Schauder estimates for elliptic and parabolic equations", *Chinese Ann. Math. Ser. B* **27**:6 (2006), 637–642. MR Zbl

[43] S. T. Yau, "On the Ricci curvature of a compact Kähler manifold and the complex Monge–Ampère equation, I", *Comm. Pure Appl. Math.* **31**:3 (1978), 339–411. MR Zbl

[44] H. Yin, "Analysis aspects of Ricci flow on conical surfaces", *Acta Math. Sin.* (*Engl. Ser.*) **38**:5 (2022), 807–857. MR Zbl

[45] H. Yin and K. Zheng, "Expansion formula for complex Monge–Ampère equation along cone singularities", *Calc. Var. Partial Differential Equations* **58**:2 (2019), art. id. 50. MR Zbl

Received 23 Oct 2018. Revised 6 Jul 2022. Accepted 11 Aug 2022.

BIN GUO: bguo@rutgers.edu
Department of Mathematics, Columbia University, New York, NY, United States
Current address: *Department of Mathematics & Computer Science, Rutgers University, Newark, NJ, United States*

JIAN SONG: jiansong@math.rutgers.edu
Department of Mathematics, Rutgers University, Piscataway, NJ, United States

ANALYSIS AND PDE
Vol. 17 (2024), No. 3, pp. 831–941
DOI: 10.2140/apde.2024.17.831

THE STABILITY OF SIMPLE PLANE-SYMMETRIC SHOCK FORMATION FOR THREE-DIMENSIONAL COMPRESSIBLE EULER FLOW WITH VORTICITY AND ENTROPY

JONATHAN LUK AND JARED SPECK

Consider a one-dimensional simple small-amplitude solution $(\varrho_{(\mathrm{bkg})}, v^1_{(\mathrm{bkg})})$ to the isentropic compressible Euler equations which has smooth initial data, coincides with a constant state outside a compact set, and forms a shock in finite time. Viewing $(\varrho_{(\mathrm{bkg})}, v^1_{(\mathrm{bkg})})$ as a plane-symmetric solution to the full compressible Euler equations in three dimensions, we prove that the shock-formation mechanism for the solution $(\varrho_{(\mathrm{bkg})}, v^1_{(\mathrm{bkg})})$ is stable against all sufficiently small and compactly supported perturbations. In particular, these perturbations are allowed to break the symmetry and have nontrivial vorticity and variable entropy.

Our approach reveals the full structure of the set of blowup-points at the first singular time: within the constant-time hypersurface of first blowup, the solution's first-order Cartesian coordinate partial derivatives blow up precisely on the zero level set of a function that measures the inverse foliation density of a family of characteristic hypersurfaces. Moreover, relative to a set of geometric coordinates constructed out of an acoustic eikonal function, the fluid solution and the inverse foliation density function remain smooth up to the shock; the blowup of the solution's Cartesian coordinate partial derivatives is caused by a degeneracy between the geometric and Cartesian coordinates, signified by the vanishing of the inverse foliation density (i.e., the intersection of the characteristics).

1.	Introduction	832
2.	Geometric setup	847
3.	Volume forms and energies	859
4.	Assumptions on the data and statement of the main theorems	862
5.	Reformulation of the equations and the remarkable null structure	867
6.	The bootstrap assumptions and statement of the main a priori estimates	869
7.	A localization lemma via finite speed of propagation	871
8.	Estimates for the geometric quantities associated to the acoustical metric	871
9.	Transport estimates for the specific vorticity and the entropy gradient	880
10.	Lower-order transport estimates for the modified fluid variables	884
11.	Top-order transport and elliptic estimates for the specific vorticity and the entropy gradient	889
12.	Wave estimates for the fluid variables	897
13.	Proving the L^∞ estimates	906
14.	Putting everything together	909
	Appendix: Proof of the wave estimates	916
	References	939

Luk is supported by a Terman fellowship and the NSF grants DMS-1709458 and DMS-2005435. Speck gratefully acknowledges support from NSF grant DMS-2054184 and NSF CAREER grant DMS-1914537.
MSC2020: primary 35L67; secondary 35L05, 35Q31, 76N10.
Keywords: compressible Euler equations, shock formation, stable singularity formation, wave breaking, vectorfield method, characteristics, eikonal function, null condition, null hypersurface, null structure.

1. Introduction

It is classically known — going back to the work of Riemann — that the compressible Euler equations admit solutions for which singularities develop from smooth initial data. Indeed, such examples can already be found in the plane symmetric isentropic case. In this case, the compressible Euler equations reduce to a 2×2 hyperbolic system in $1+1$-dimensions, which can be analyzed using Riemann invariants. In particular, it is easy to show that simple plane-symmetric solutions — solutions with one vanishing Riemann invariant — obey a Burgers-type equation, and that a shock can form in finite time. By a shock, we mean that the solution remains bounded but its first-order partial derivative with respect to the standard spatial coordinate blows up, and that the blowup is tied to the intersection of the characteristics.

In this article, we prove that a class of simple plane-symmetric isentropic small-amplitude shock-forming solutions to the compressible Euler equations are *stable* under small perturbations which break the symmetry and admit variable vorticity and entropy. In particular, the perturbed solutions develop a shock singularity in finite time. This provides the details of the argument sketched in [37; 52] and completes the program that we have initiated (partly joint also with Gustav Holzegel and Willie Wai-Yeung Wong) in [36; 37; 50; 52].

We will consider the spatial domain[1] $\Sigma \doteq \mathbb{R} \times \mathbb{T}^2 = \mathbb{R} \times (\mathbb{R}/\mathbb{Z})^2$ and a time interval I. Our unknowns are the density $\varrho : I \times \Sigma \to \mathbb{R}_{>0}$, the velocity $v : I \times \Sigma \to \mathbb{R}^3$, and the entropy $s : I \times \Sigma \to \mathbb{R}$. Relative to the standard Cartesian coordinates (t, x^1, x^2, x^3) on $I \times \mathbb{R} \times \mathbb{T}^2$, the compressible Euler equations can be expressed as

$$(\partial_t + v^a \partial_a)\varrho = -\varrho \operatorname{div} v, \tag{1-1}$$

$$(\partial_t + v^a \partial_a)v^j = -\frac{1}{\varrho}\delta^{ja}\partial_a p, \quad j = 1, 2, 3, \tag{1-2}$$

$$(\partial_t + v^a \partial_a)s = 0, \tag{1-3}$$

where (from now on) δ^{ij} denotes the Kronecker delta, $\operatorname{div} v \doteq \partial_a v^a$ is the Euclidean divergence of v, repeated lowercase Latin indices are summed over $i, j = 1, 2, 3$, and the pressure p relates to ϱ and s by a prescribed smooth *equation of state* $p = p(\varrho, s)$. In other words, the right-hand side of (1-2) can be expressed as

$$-\frac{1}{\varrho}\delta^{ja}\,\partial_a p = -\frac{1}{\varrho}p_{;\varrho}\delta^{ja}\,\partial_a \varrho - \frac{1}{\varrho}p_{;s}\delta^{ja}\,\partial_a s,$$

where $p_{;\varrho}$ denotes[2] the partial derivative of the equation of state with respect to the density at fixed s, and analogously for $p_{;s}$.

For the remainder of the paper:

(1) We fix a constant $\bar{\varrho} > 0$ and a constant solution $(\varrho, v^i, s) = (\bar{\varrho}, 0, 0)$ to (1-1)–(1-3).

(2) We fix an equation of state $p = p(\varrho, s)$ such that[3] $(\partial p/\partial \varrho)(\bar{\varrho}, 0) = 1$.

[1] It is only for technical convenience that we chose the spatial topology $\mathbb{R} \times \mathbb{T}^2$. Similar results also hold, for instance, on $\mathbb{R}^3$.

[2] Later in the paper, we will take the partial derivative of various quantities with respect to the logarithmic density ρ. If f is a function of the fluid unknowns, then $f_{;\rho}$ will denote the partial derivative of f with respect to ρ when the other fluid variables are held fixed. Similarly, $f_{;s}$ denotes the partial derivative of f with respect to s when the other fluid variables are held fixed.

[3] This normalization can always be achieved by a change of variables as long as $(\partial p/\partial \varrho)(\bar{\varrho}, 0) > 0$; see [36, footnote 19].

For notational convenience, we define the logarithmic density $\rho \doteq \log(\varrho/\bar{\varrho})$ and the speed of sound $c(\rho, s) \doteq \sqrt{\partial p/\partial \varrho}(\varrho, s)$. We will from now on think of c as a function of (ρ, s).

We will study perturbations of a shock-forming background solution $(\varrho_{(\mathrm{bkg})}, v^i_{(\mathrm{bkg})}, s_{(\mathrm{bkg})})$ arising from smooth initial data such that the following hold:

(1) The background solution is plane-symmetric and isentropic, i.e., $v^2_{(\mathrm{bkg})} = v^3_{(\mathrm{bkg})} = s_{(\mathrm{bkg})} = 0$, and $(\varrho_{(\mathrm{bkg})}, v^1_{(\mathrm{bkg})})$ are functions only of t and x^1.

(2) The background solution is simple, i.e., the Riemann invariant $\mathcal{R}^{(\mathrm{bkg})}_{(-)}$, satisfies

$$\mathcal{R}^{(\mathrm{bkg})}_{(-)} \doteq v^1_{(\mathrm{bkg})} - \int_0^{\rho_{(\mathrm{bkg})}} c(\rho', 0)\, d\rho' = 0.$$

(3) The background solution is initially compactly supported in an x^1-interval of length $\leq 2\mathring{\sigma}$, i.e., outside this interval, $(\varrho_{(\mathrm{bkg})}, v^i_{(\mathrm{bkg})}, s_{(\mathrm{bkg})}) \restriction_{t=0} = (\bar{\varrho}, 0, 0)$.

(4) At time 0 (and hence throughout the evolution), the Riemann invariant $\mathcal{R}^{(\mathrm{bkg})}_{(+)} \doteq v^1_{(\mathrm{bkg})} + \int_0^{\rho_{(\mathrm{bkg})}} c(\rho', 0)\, d\rho'$ has small $\leq \mathring{\alpha}$ amplitude.

(5) At time 0, the Cartesian spatial derivatives of $\mathcal{R}^{(\mathrm{bkg})}_{(+)}$ up to the third order[4] are bounded above pointwise by $\leq \mathring{\delta}^{(\mathrm{bkg})}$ (where $\mathring{\delta}^{(\mathrm{bkg})}$ is not necessarily small).

(6) The quantity[5] $\mathring{\delta}^{(\mathrm{bkg})}_*$ (where $\mathring{\delta}^{(\mathrm{bkg})}_*$ is not necessarily small) that controls the blowup-time satisfies[6]

$$\mathring{\delta}^{(\mathrm{bkg})}_* \doteq \frac{1}{2} \sup_{\{t=0\}} \left[\frac{1}{c} \left\{ \frac{1}{c} \frac{\partial c}{\partial \rho}(\rho_{(\mathrm{bkg})}, 0) + 1 \right\} (\partial_1 \mathcal{R}^{(\mathrm{bkg})}_{(+)}) \right]_+ > 0,$$

and the solution forms a shock at time[7] $T^{(\mathrm{bkg})}_{(\mathrm{sing})} = (\mathring{\delta}^{(\mathrm{bkg})}_*)^{-1}$.

The analysis for plane-symmetric solutions can be carried out easily using Riemann invariants. It is then straightforward to check that there exists a large class of plane-symmetric solutions satisfying (1)–(6) above.

We now provide a rough version of our main theorem; see Section 4B for a more precise statement.

Theorem 1.1 (main theorem, rough version). *Consider a plane-symmetric, shock-forming background solution* $(\varrho_{(\mathrm{bkg})}, v^i_{(\mathrm{bkg})}, s_{(\mathrm{bkg})})$ *satisfying* (1)–(6) *above, where the parameter* $\mathring{\alpha}$ *from point* (4) *is small. Consider a small perturbation of the initial data of this background solution satisfying the following assumptions* (*see Section 4A for the precise assumptions*):

[4]In the one-dimensional case, one only needs information about the data's first derivative to close a proof of blowup for a simple plane wave. However, when studying perturbations in three dimensions, we need estimates on these derivatives up to third order in order to close the proof. For example, the proof of the bound (8-23c) relies on having control of up to these third-order derivatives (as is provided by (8-20b)–(8-20c)), and we use the bound (8-23c) in the proof of Lemma 14.2 as well as in the proof of the energy estimates in the Appendix.

[5]One can check that this rules out the Chaplygin gas, whose speed of sound (after normalization) is given by $c(\rho, s) = \exp(-\rho)$. One can also check that for any other equation of state, it is possible to choose $\bar{\varrho}$ appropriately so that $\mathring{\delta}^{(\mathrm{bkg})}_* > 0$.

[6]Here, $[\,\cdot\,]_+$ denotes the positive part.

[7]In the plane-symmetric, isentropic, simple case, $\mathcal{R}^{(\mathrm{bkg})}_{(+)}$ satisfies the transport equation $\partial_t \mathcal{R}^{(\mathrm{bkg})}_{(+)} + (v^1_{(\mathrm{bkg})} + c(\rho_{(\mathrm{bkg})})) \cdot \partial_1 \mathcal{R}^{(\mathrm{bkg})}_{(+)} = 0$, and the blowup-time of $\partial_1 \mathcal{R}^{(\mathrm{bkg})}_{(+)}$ can easily be computed explicitly by commuting this transport equation with ∂_1 to obtain a Riccati-type ODE in $\partial_1 R^{(\mathrm{bkg})}_{(+)}$ along the integral curves of $\partial_t + (v^1_{(\mathrm{bkg})} + c(\rho_{(\mathrm{bkg})}))\partial_1$.

- *The perturbation is compactly supported in a region of x^1-length $\leq 2\mathring{\sigma}$.*
- *The perturbation belongs to a high-order Sobolev space, where the required Sobolev regularity is independent of the background solution and equation of state.*
- *The perturbation is small, where the smallness is captured by the small parameter $0 < \mathring{\varepsilon} \ll 1$, and the required smallness depends on the order of the Sobolev space, the equation of state, and the parameters of the background solution.*

Then the corresponding unique perturbed solution satisfies the following:

(1) *The solution is initially smooth, but it becomes singular at a time $T_{(\mathrm{sing})}$, which is a small perturbation of the background blowup-time $(\mathring{\delta}_*^{(\mathrm{bkg})})^{-1}$.*

(2) *Defining*[8] $\mathcal{R}_{(+)} \doteq v^1 + \int_0^{\rho} c(\rho', s)\, d\rho'$, *we have the singular behavior*

$$\limsup_{t \to T_{(\mathrm{sing})}^-} \sup_{\{t\}\times\Sigma} |\partial_1 \mathcal{R}_{(+)}| = +\infty. \tag{1-4}$$

(3) *Relative to a geometric coordinate system (t, u, x^2, x^3), where u is an eikonal function, the solution remains smooth, all the way up to time $T_{(\mathrm{sing})}$. In particular, the partial derivatives of the solution with respect to the geometric coordinates do not blow up.*

(4) *The blowup at time $T_{(\mathrm{sing})}$ is characterized by the vanishing of the inverse foliation density μ (see Definition 2.15) of a family of acoustically null hypersurfaces defined to be the level sets of u.*

(5) *In particular, the set of blowup-points at time $T_{(\mathrm{sing})}$ is characterized by*

$$\Big\{(u, x^2, x^3) \in \mathbb{R}\times\mathbb{T}^2 : \limsup_{(\tilde{t},\tilde{u},\tilde{x}^2,\tilde{x}^3)\to(T_{(\mathrm{sing})}^-,u,x^2,x^3)} |\partial_1\mathcal{R}_{(+)}|(\tilde{t}, \tilde{u}, \tilde{x}^2, \tilde{x}^3) = \infty\Big\} \\ = \{(u, x^2, x^3) \in \mathbb{R}\times\mathbb{T}^2 : \mu(T_{(\mathrm{sing})}, u, x^2, x^3) = 0\},$$

where $|\partial_1\mathcal{R}_{(+)}|(\tilde{t}, \tilde{u}, \tilde{x}^2, \tilde{x}^3)$ denotes the absolute value of the Cartesian partial derivative $\partial_1\mathcal{R}_{(+)}$ evaluated at the point with geometric coordinates $(\tilde{t}, \tilde{u}, \tilde{x}^2, \tilde{x}^3)$.

(6) *At the same time, as $T_{(\mathrm{sing})}$ is approached from below, the fluid variables ϱ, v^i, s all remain bounded, as do the specific vorticity $\Omega^i \doteq (\mathrm{curl}\, v)^i/(\varrho/\bar{\varrho})$ and the entropy gradient $S \doteq \nabla s$.*

The proof of Theorem 1.1 relies on two main ingredients: (i) Christodoulou's geometric theory of shock formation for irrotational and isentropic solutions, in which case the dynamics reduces to the study of quasilinear wave equations and (ii) a (re-)formulation of the compressible Euler equations as a quasilinear system of wave-transport equations, which was derived in [50], following the earlier works [36; 37] in the barotropic[9] case. This formulation exhibits remarkable null structures and regularity properties, which in total allow us to perturbatively control the vorticity and entropy gradient all the way up to the singular

[8] In higher dimensions or in the presence of dynamic entropy, $\mathcal{R}_{(+)}$ is *not* a Riemann invariant because its dynamics is not determined purely by a transport equation. Nonetheless, for comparison purposes, we continue to use the symbol $\mathcal{R}_{+}$ to denote this quantity.

[9] A barotropic fluid is such that the equation of state for the pressure is a function of the density alone, as opposed to being a function of the pressure and entropy.

time — even though generically, their first-order Cartesian partial derivatives blow up at the singularity. See Section 1A for further discussion of the proof.

Some remarks are in order.

Remark 1.2. Note that even though the rough Theorem 1.1 is formulated in terms of plane-symmetric background solutions, we do not actually "subtract off a background" in the proof. See Theorem 4.3 for the precise formulation.

Remark 1.3 (results building up towards Theorem 1.1).

- Concerning stability of simple plane-symmetric shock-forming solutions to the compressible Euler equations, the first result was our joint work with G. Holzegel and W. Wong [52], which proved the analog[10] of Theorem 1.1 in the case[11] where the perturbation is irrotational and isentropic (i.e., $\Omega \equiv 0$, $S \equiv 0$).

- In [36], we proved the first stable shock formation result without symmetry assumptions for the compressible Euler equations for open sets of initial data that can have nontrivial specific vorticity Ω. Specifically, in [36], we treated the two-dimensional barotropic compressible Euler equations (see footnote 9). One of the key points in [36] was our reformulation of equations into a system of quasilinear wave-transport equations which has favorable nonlinear null structures. This allowed us to use the full power of the geometric vectorfield method on the wave part of the system while treating the vorticity perturbatively.

- In [37], we considered three-dimensional barotropic compressible Euler flow and derived a similar reformulation of the equations that allowed for nonzero vorticity. In contrast to the two-dimensional case, the transport equation satisfied by the specific vorticity Ω featured vorticity-stretching source terms (of the schematic form $\Omega \cdot \partial v$). In order to handle the vorticity-stretching source terms in the framework of [36], we also showed in [37] that Ω satisfies a div-curl-transport system with source terms that are favorable from the point of view of regularity and from the point of view of null structure. We refer to Section 1A6 for further discussion of this point.

- To incorporate thermodynamic effects into compressible fluid flow, one must look beyond the family of barotropic equations of state, e.g., consider equations of state in which the pressure depends on the density and entropy.[12] Fortunately, in [50], it was shown that a similar good reformulation of the compressible Euler equations holds under an arbitrary equation of state (in which the pressure is a function of the density and the entropy) in the presence of vorticity and variable entropy. In the present paper, we use this reformulation to prove our main results; we recall it below as Theorem 5.1. The analysis in [50] is substantially more complicated compared to the barotropic case, and the basic setup requires the

[10] We remark that while [52] only explicitly stated a theorem in two spatial dimensions, the analogous result in three (or indeed higher) dimensions can be proved using similar arguments; see [52, Remarks 1.4,1.11].

[11] The main theorem in [52] is stated for general quasilinear wave equations. Particular applications to the relativistic compressible Euler equations in the irrotational and isentropic regime can be found in [52, Appendix B]. It applies equally well to the nonrelativistic case.

[12] Incorporating entropy into the analysis is expected to be especially important for studying weak solutions after the shock (see Section 1B4 for further discussion), since formal calculations [16] suggest that the entropy (even if initially zero) should jump across the shock hypersurface, which in turn should induce a jump in vorticity.

observation of some new structures tied to elliptic estimates for Ω and S, such as good regularity and null structures tied to the modified fluid variables from Definition 2.7.

This paper completes the program described above by giving the analytic details already sketched in [37; 50]. Chief among the analytic novelties in the present paper are the elliptic estimates for Ω and S at the top-order; see [37, Sections 1.3, 4.2.7], [50, Section 4.3] and Section 1A6. We also point out that there are other related works, which we discuss in Section 1B.

Remark 1.4 (blowup and boundedness of quantities involving higher derivatives). For generic perturbations, derivatives of fluid variables other than $\mathcal{R}_+$ (whose blowup was highlighted in (1-4)) can also blow up. In particular, while the ∂_2 and ∂_3 derivatives of the fluid variables are identically 0 for the plane-symmetric background solutions, for the perturbed solution, $\partial_2 v^i$, say, is generically unbounded at the singularity. This is because the perturbation changes the geometry of the solution, and the regular directions no longer align with the Cartesian directions.

On the other hand, there are indeed higher derivatives of the fluid variables that remain bounded up to the singular time. These include the specific vorticity and the entropy gradient that we already mentioned explicitly in Theorem 1.1. Moreover, any null-hypersurface-tangential geometric derivatives (see further discussions in Section 1A) of the fluid variables are also bounded up to the singular time. This is not just a curiosity, but rather is a fundamental aspect of the proof.

Remarkably, there are additionally quantities, denoted by $\mathcal{C}$ and $\mathcal{D}$ (these variables were identified in [50], see (2-5a)–(2-5b)), which are special combinations of up-to-second-order Cartesian coordinate derivatives of the fluid variables, which remain uniformly bounded up to the singularity (as do their derivatives in directions tangent to a family of null hypersurfaces); $\mathcal{C}$ and $\mathcal{D}$ are precisely the modified fluid variables mentioned in Remark 1.3. The existence of such regular higher-order quantities is not only an interesting fact, but is also quite helpful in controlling the solution up to the first singularity; see Section 1A.

Finally, as a comparison with our two-dimensional work [36], note that in the two-dimensional case, we proved that the specific vorticity remains Lipschitz (in Cartesian coordinates) up to the first singular time. This is no longer the case in three dimensions. Indeed, in the language of this paper, the improved regularity for the specific vorticity in [36] stems from the fact that in two dimensions, the Cartesian coordinate derivatives of the specific vorticity Ω coincide with $\mathcal{C}$.

Remark 1.5 (additional information on subclasses of solutions). Within the solution regime we study, we are able to derive additional information about the solution by making further assumptions on the data. For instance, there are open subsets of data such that the vorticity/entropy gradient are nonvanishing at the first singularity, and also open subsets of data such that the fluid variables remain Hölder[13] $C^{1/3}$ up to the singularity. See Section 4B for details.

Remark 1.6 (the maximal smooth development). The approach we take here allows us to analyze the solution up to the first singular time, and our main results yield a complete description of the set of blowup-points at that time (see, for example, conclusions (4)–(5) of Theorem 1.1). However, since

[13] The Hölder estimates hold only for an open subset of data satisfying certain nondegeneracy assumptions. They were not announced in [37; 50]. We were instead inspired by [9; 11] to include such estimates.

the compressible Euler equations are a hyperbolic system, it is desirable to go beyond our results by deriving a full description of the maximal smooth development of the initial data, in analogy with [15]. Understanding the maximal smooth development is particularly important for the shock development problem; see Section 1B4 below.

Our methods, at least on their own, are not enough to construct the maximal smooth development.[14] This is in part because our approach here relies on spatially global elliptic estimates on constant-t hypersurfaces; the point is that a full description of the smooth maximal development would require spatially localized estimates. On the other hand, the recent preprint [1] discovered an integral identity that allows the elliptic estimates to be localized, and thus gives hope that Theorem 1.1 can be extended to derive the structure of the full maximal smooth development.

Remark 1.7 (no universal blowup-profile). One of the main advantages of our geometric framework is that it works for many kinds of singular solutions, not just those exhibiting a specific blowup-profile. In particular, the solutions featured in Theorem 1.1 do not exhibit a universal blowup-profile. Although we do not rigorously study the full class of blowup-profiles exhibited by the solutions from Theorem 1.1, the full class is likely quite complicated to describe. This can already be seen in model case of Burgers' equation, where there are a continuum of possible blowup-profiles and corresponding blowup-rates [27] (recall that we work in the near plane-symmetric regime and our work includes, as special cases, plane-symmetric solutions, which are analogs of Burgers' equation solutions). A related issue is that at the time of first singularity formation, the set of blowup-points can be complicated and/or of infinite cardinality (as one can already see in the special case of plane-symmetric solutions, viewed as solutions in three dimensions with symmetry).

Remark 1.8 (the relativistic case). While our present work treats only the nonrelativistic case, it is likely that the relativistic case can also be treated in the same way. This is because the relativistic compressible Euler equations also admit a similar reformulation as we consider here, and likewise the variables in the reformulation also exhibit a very similar null structure [25].

In the remainder of the Introduction, we will first discuss the proof in Section 1A and then discuss some related works in Section 1B. We will end the introduction with an outline of the remainder of the paper.

1A. *Ideas of the proof.*

1A1. *The Christodoulou theory.* The starting point of our proof is the work of Christodoulou [15] on shock formation for quasilinear wave equations.[15] Consider the following model quasilinear covariant wave equation for the scalar function Ψ: $\square_{g(\Psi)}\Psi = 0$, where the Cartesian component functions $g_{\alpha\beta}$ are given (nonlinear in general) functions of Ψ, i.e., $g_{\alpha\beta} = g_{\alpha\beta}(\Psi)$. Our study of compressible Euler flow in

[14] Notice that in our earlier result [36] for the isentropic Euler equations in two spatial dimensions, we also only solved the equations up to the first singular time. However, there is an important difference. In the two-dimensional case, there does not seem to be a philosophical obstruction in extending [36] to provide a complete description of the maximal smooth development. In contrast, in the three-dimensional case it seems that ideas in [1] would be needed in a fundamental way.

[15] Strictly speaking, [15] is only concerned with the irrotational isentropic relativistic Euler equations. However, its methods apply to much more general quasilinear wave equations; see further discussions in [30; 48].

this paper essentially amounts to studying a system of similar equations with source terms and showing that the source terms do not radically distort the dynamics. This is possible only because the source terms have remarkable null structure, described below.

A key insight for studying the formation of shocks, going back to [15], is that it is advantageous to study the shock formation via a system of geometric coordinates. The point is that when appropriately constructed, such coordinates regularize the problem, which allows one to treat the problem of shock formation as if it were a standard local existence problem. More precisely, one constructs geometric coordinates, adapted to the flow, such that the solution remains regular relative to them.[16] However, the geometric coordinates *degenerate* relative to the Cartesian ones, and the blowup of the solution's first-order Cartesian coordinate partial derivatives can be derived as a consequence of this degeneracy.

To carry out this strategy, one must use the Lorentzian geometry associated to the acoustical metric g (see Definition 2.9). The following geometric objects are of central importance in implementing this program:

- A foliation by constant-u characteristic hypersurfaces $\mathcal{F}_u$ (where $g^{-1}(du, du) = 0$; see (2-13)). The function u is known as an "acoustic eikonal function".

- The inverse foliation density μ ($\doteq -1/g^{-1}(dt, du)$), where μ^{-1} measures the density of $\mathcal{F}_u$ with respect to the constant-t hypersurfaces.

- A frame of vectorfields $\{L, X, Y, Z\}$, where $\{L, Y, Z\}$ are tangent to $\mathcal{F}_u$ (with L being its null generator) and X is transversal to $\mathcal{F}_u$; see Figure 1, where we have suppressed the Z-direction.

- $\{L, X, Y, Z\}$ is a frame that is "comparable" to the Cartesian frame $\{\partial_t, \partial_1, \partial_2, \partial_3\}$, by which we mean the coefficients relating the frames to each other are size $\mathcal{O}(1)$.

- However, in the analysis, uniform boundedness estimates are generally available for the derivatives of quantities with respect to only the *rescaled* frame elements $\{L, \breve{X} \doteq \mu X, Y, Z\}$.

The analysis simultaneously yields control of the derivatives of Ψ with respect to the rescaled frame and gives also quantitative estimates on the geometry. In this geometric picture, the blowup is completely captured by $\mu \to 0$. The connection between the vanishing of μ and the blowup of some Cartesian coordinate partial derivative of Ψ can be understood as follows: one proves an estimate of the form $|\breve{X}\Psi| \approx 1$ (which is consistent with the uniform boundedness estimates mentioned above). In view of the relation $\breve{X} = \mu X$, this estimate implies that $|X\Psi|$ blows up like $1/\mu$ as $\mu \to 0$.

We now give a more detailed description of the behavior of the solution, with a focus on how it behaves at different derivative levels.

- As our discussion above suggested, at the lower derivative levels, derivatives of quantities with respect to the *rescaled* frame are *regular*, e.g., $L\Psi$, $\breve{X}\Psi$, $Y\Psi$, $Z\Psi$, $\ldots$, $L^3\breve{X}Y\Psi$, etc. are uniformly bounded.

[16]It should be emphasized that it is only at the low derivative levels that the solution is regular. The high-order geometric energies can still blow up, even though the low-order energies remain bounded. The possible growth of the high-order energies is one of the central technical difficulties in the problem, and we will discuss it below in more detail.

• As we highlighted above, the formation of the shock corresponds to $\mu \to 0$ in finite time, and moreover, the *nonrescaled* first-order derivative $X\Psi$ *blows up* in finite time, exactly at points where μ vanishes.

• The main difficulty in the proof is that the only known approach to the solution's regularity theory *with respect to the rescaled frame derivatives* that is able to avoid a loss of derivatives allows for the following possible scenario: the energy estimates are such that the high-order geometric energies might blow up when the shock forms. This leads to severe difficulties in the proof, especially considering that one needs to show that the low-order derivatives of the solution remain bounded in order to derive the singular high-order energy estimates.[17]

In [15], Christodoulou showed that the maximum possible blowup-rate of the high-order energies is of the form $\mu_\star^{-2P}(t)$, where P is a universal positive constant and $\mu_\star(t) \doteq \min\{1, \min_{\Sigma_t} \mu\}$. To reconcile this possible high-order energy blowup with the regular behavior at the lower derivative levels, one is forced to derive a hierarchy of energy estimates of the form, where M_* is a universal[18] positive integer:

$$\mathbb{E}_{N_{\mathrm{top}}}(t) \lesssim \mathring{\epsilon}^2 \mu_\star^{-2M_*+1.8}(t), \quad \mathbb{E}_{N_{\mathrm{top}}-1}(t) \lesssim \mathring{\epsilon}^2 \mu_\star^{-2M_*+3.8}(t), \quad \mathbb{E}_{N_{\mathrm{top}}-2}(t) \lesssim \mathring{\epsilon}^2 \mu_\star^{-2M_*+5.8}(t), \quad \ldots, \tag{1-5}$$

where $\mathbb{E}_N$ denotes the energy after N commutations and all energies are by assumption initially of small size $\mathring{\epsilon}^2$. In other words, the energy estimates become less singular by two powers of $\mu_\star$ for each descent below the top derivative level. Importantly, despite the possible blowup at higher orders, all the sufficiently low-order energies are bounded, which, by Sobolev embedding, is what allows one to show the uniform pointwise boundedness of the solution's lower-order derivatives:[19]

$$\sum_{N=1}^{N_{\mathrm{top}}-M_*} \mathbb{E}_N(t) \lesssim \mathring{\epsilon}^2. \tag{1-6}$$

1A2. *The nearly simple plane-symmetric regime.* Christodoulou's work [15] concerned compactly supported[20] initial data in $\mathbb{R}^3$, a regime in which dispersive effects dominate for a long time before the singularity formation processes eventually take over. In a joint work with Holzegel and Wong [52], we adapted the Christodoulou theory to the almost simple plane symmetric regime. The important point is that the commutators $\{L, Y, Z\}$, in addition to being regular derivatives near the singularity, also simultaneously capture the fact that the solution is "almost simple plane symmetric." Moreover, the following analytical considerations were fundamental to the philosophy of the proof in [52]:

[17] The possible high-order energy blowup has its origins in the presence of some difficult factors of $1/\mu$ in the top-order energy identities, where one must work hard to avoid a loss of derivatives. To close the energy estimates, one commutes the wave equation many times with the $\mathcal{F}_u$-tangent subset $\{L, Y, Z\}$ of the rescaled frame. The most difficult terms in the commuted wave equation are top-order terms in which all the derivatives fall onto the components of $\{L, Y, Z\}$. It turns out that due to the way the rescaled frame is constructed, the corresponding difficult error terms depend on the top-order derivatives of the eikonal function u. In Proposition A.4, we identify these difficult commutator terms. To avoid the loss of derivatives, one must work with modified quantities and use elliptic estimates. It is in this process that one creates difficult factors of $1/\mu$.

[18] Our proof of the universality of M_* in the presence of vorticity and entropy requires some new observations, described below (1-11).

[19] The lowest-order energy $\mathbb{E}_N(t)$ is excluded from this estimate because it is not of small size $\mathring{\epsilon}^2$, owing to the largeness of $\breve{X}\mathcal{R}_{(+)}$.

[20] More precisely, his work addressed compactly supported irrotational perturbations of constant, nonvacuum fluid solutions.

• All energy estimates can be closed by commuting only with tangential derivatives $\{L, Y, Z\}$ (and *without* $\breve{X}$). This is a slightly different strategy than we used in our paper [36] in the two-dimensional case, in which we closed the energy estimates by commuting the equations with strings of tangential derivatives $\{L, Y, Z\}$, as well as strings that contain up to one factor of $\breve{X}$. In [36], we also could have closed the energy estimates by commuting only with tangential derivatives $\{L, Y, Z\}$, but we would have had to work with the modified fluid variable $\mathcal{C}$ (which, though fundamental in three dimensions, was not needed in [36] due to the absence of the vorticity-stretching term) or to treat the Cartesian gradients $\partial_\alpha \Omega^i$ as independent unknowns.

• After being commuted with (at least one of) L, Y, Z, the wave equation solutions are *small*. In particular, we can capture the smallness from "nearly simple plane-symmetric" data without explicitly subtracting the simple plane-symmetric background solution; see also Remark 1.2.

1A3. *The reformulation of the equations.* In order to extend Christodoulou's theory so that it can be applied to the compressible Euler equations, a crucial first step is to reformulate the compressible Euler equations as a system of quasilinear wave equations and transport equations. Here, the transport part of the system refers to the vorticity and the entropy, and the intention is to handle them perturbatively.

As we mentioned earlier, the reformulation has been carried out in [36; 37; 50]. Here we highlight the main features and philosophy of the reformulation, and explain how we derived it.

(1) To the extent possible, formulate compressible Euler flow as a perturbation of a system of quasilinear wave equations.

(a) We compute $\square_g v^i$, $\square_g \rho$, and $\square_g s$, where $\square_g$ is the covariant wave operator associated to the acoustical metric (see (2-7)). Then using the compressible Euler equations (1-1)–(1-3), we eliminate and re-express many terms.

(b) We find that v^i, ρ, and s do not exactly satisfy wave equations; instead, the right-hand sides contain second derivatives of the fluid variables, which we will show to be *perturbative*, despite their appearance of being principal order in terms of the number of derivatives.

(2) The "perturbative" terms mentioned above are equal to *good transport variables* that we identify, specifically $(\Omega, S, \mathcal{C}, \mathcal{D})$. These variables behave better than what one might naïvely expect, from the points of view of their regularity and their singularity strength.

(a) While both $\Omega^i \doteq (\operatorname{curl} v)^i/(\varrho/\bar{\varrho})$ and $S \doteq \nabla s$ are derivatives of the fluid variables, they play a distinguished role since they satisfy independent transport equations, and obey better bounds than generic first derivatives of the fluid variables.

(b) We have introduced the *modified fluid variables* $\mathcal{C}^i$ and $\mathcal{D}$ (see Definition 2.7), which, up to lower-order correction terms, are equal to $(\operatorname{curl} \Omega)^i$ and $\Delta s = \operatorname{div} S$ respectively. These quantities satisfy better estimates than generic first derivatives of Ω and S, which is crucial for our proof.

1A4. *The remarkable null structure of the reformulation.* In the reformulation of compressible Euler flow, we consider the unknowns to be all of $(v^i, \rho, s, \Omega^i, S^i, \mathcal{C}^i, \mathcal{D})$. Note that these include not only the fluid variables, but also higher-order variables which can be derived from the fluid variables.

The equations satisfied by these variables take the following schematic form (see Theorem 5.1 for the precise equations):[21]

$$\square_g(v, \rho, s) = \overbrace{\partial(v, \rho)\cdot\partial(v, \rho)}^{\doteq I} + (\Omega, S)\cdot\partial(v, \rho) + (\mathcal{C}, \mathcal{D}), \tag{1-7}$$

$$B(\Omega, S) = (\Omega, S)\cdot\partial(v, \rho), \tag{1-8}$$

$$B(\mathcal{C}, \mathcal{D}) = \underbrace{\partial(v, \rho)\cdot\partial(\Omega, S)}_{\doteq II} + \underbrace{(\Omega, S)\cdot\partial(v, \rho)\cdot\partial(v, \rho)}_{\doteq III} + S\cdot S\cdot\partial(v, \rho). \tag{1-9}$$

Here, $\square_g$ is the covariant wave operator associated to the acoustical metric (see (2-7)) and $B \doteq \partial_t + v^a\partial_a$ is the transport operator associated with the material derivative (cf. (1-1)–(1-3)).

Although it is not apparent from the way we have written it, the system of equations (1-7)–(1-9) has a remarkable null structure! Importantly, the terms I, II and III are g-null forms: when decomposed in the $\{L, X, Y, Z\}$ frame, we do not have $X(v^i, \rho)\cdot X(v^i, \rho)$ in I and III, nor do we have $X(v, \rho)\cdot X(\Omega, S)$ in II.

Because $X(v^i, \rho)$ is the only derivative that blows up (while $\breve{X}(v^i, \rho)$ is bounded), it follows that given a g-null form $\mathcal{Q}$ in the fluid variables (see Definition 8.1 concerning g-null forms), such as $\mathcal{Q}(\partial v^i, \partial v^j)$, the quantity $\mu\mathcal{Q}(\partial v^i, \partial v^j)$ remains bounded up to the singularity, while a generic quadratic nonlinearity $\mathcal{Q}_{\text{bad}}$ would be such that $\mu\mathcal{Q}_{\text{bad}}(\partial v^i, \partial v^j)$ blows up when μ vanishes.

As is already observed in [48], a null form I on the right-hand side of the wave equation allows all the wave estimates in Section 1A1 to be proved. As we will discuss below, the null forms II and III in (1-9) will also be important for estimating the full system.

1A5. *Estimates for the transported variables.* To control solutions to the system (1-7)–(1-9), we in particular need to estimate the transport variables $(\Omega, S, \mathcal{C}, \mathcal{D})$ and understand how they interact with the wave variables (v, ρ, s) on the left-hand side of (1-7). Here, we will discuss the estimates at the low derivative levels. We will discuss the difficult technical issues of a potential loss of derivatives and the blowup of the higher-order energies in Sections 1A6 and 1A7 respectively.

We begin with two basic — but crucial — properties regarding the transport operator for the compressible Euler system, which were already observed in [36]:

- *The transport vectorfield B is* ***transversal*** *to the null hypersurfaces $\mathcal{F}_u$*; see Figure 1, where some integral curves of B are depicted. As a result, one *gains* a power of μ by integrating along B; i.e., for solutions ϕ to $B\phi = \mathfrak{F}$, we have $\|\phi\|_{L^\infty} \lesssim \|\mu\mathfrak{F}\|_{L^\infty}$.
- *μB is a* ***regular*** *vectorfield in the (t, u, x^2, x^3) differential structure.* Thus, if $B\phi = \mathfrak{F}$ and $\mu\mathfrak{F}$ has bounded $\{L, Y, Z\}$ derivatives, then ϕ also has bounded $\{L, Y, Z\}$ derivatives.

We now apply these observations to (1-8) and (1-9):

- Even though $\partial(v, \rho)$ blows up as the shock forms, $\mu\partial(v, \rho)$ remains regular. This is because $\mu\partial$ can be written as a linear combination of the rescaled frame vectorfields $\{\mu X, L, Y, Z\}$ (see Section 1A1) with

[21] Here, our notation above the brackets is such that $\partial(v, \rho)\cdot\partial(v, \rho)$ may contain all of $\partial v^i\partial v^j$, $\partial v^i\partial\rho$ and $\partial\rho\partial\rho$. A similar convention applies for other terms.

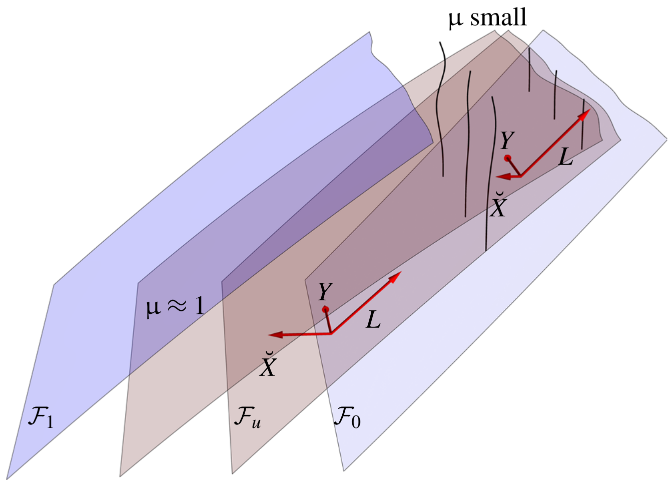

Figure 1. The dynamic vectorfield frame at two distinct points on $\mathcal{F}_u$ with the Z-direction suppressed, and the integral curves of the transport operator B for the specific vorticity and entropy.

coefficients that are $\mathcal{O}(1)$ or $\mathcal{O}(\upmu)$. Hence, the above observations imply that $(\varOmega, S)$ and their $\{L, Y, Z\}$ derivatives are bounded.

- The null structure and the bounds for the wave variables and $(\varOmega, S)$ together imply that the right-hand side of (1-9) is $\mathcal{O}(\upmu^{-1})$. Thus, $\mathcal{C}$, $\mathcal{D}$ and their $\{L, Y, Z\}$ derivatives are also bounded.

1A6. *Elliptic estimates for the vorticity and the entropy gradient.* Despite the favorable structure of (1-7)–(1-9), there is apparently a potential loss of derivatives. To see this, consider the following simple derivative count. Suppose we bound (v, ρ, s) with $N_{\text{top}} + 1$ derivatives. Equation (1-7) dictates[22] that we should control $(\mathcal{C}, \mathcal{D})$ with N_{top} derivatives. If we rely only on (1-8), then we can only bound N_{top} derivatives of $(\varOmega, S)$. However, this is insufficient: plugging this into (1-9) and using only transport estimates, we are only able to control $N_{\text{top}} - 1$ derivatives of $(\mathcal{C}, \mathcal{D})$, which is not enough.

The key to handling this difficulty is the observation that in fact, $\mathcal{C}$ and $\mathcal{D}$ can be used in conjunction with elliptic estimates to control one derivative of $\varOmega$ and S. This is because up to lower-order terms, $\mathcal{C} \approx \operatorname{curl} \varOmega$ and $\mathcal{D} \approx \operatorname{div} S$, while at the same time, by the definitions of $\varOmega$ and S — precisely that $\varOmega$ is almost a curl of a vectorfield and $S = \nabla s$ is an exact gradient — $\operatorname{div} \varOmega$ and $\operatorname{curl} S$ are of lower order in terms of the number of derivatives. It follows that we can control *all* first-order spatial derivatives of $\varOmega$ and S, including $\mathcal{C}$ and $\mathcal{D}$, using elliptic estimates.

1A7. *L^2 estimates for the transport variables and the high-order blowup-rate.* We end this section with a few comments on the L^2 energy estimates for the transport variables $(\varOmega, S)$ (and $(\mathcal{C}, \mathcal{D})$), with a focus on how to handle the degeneracies tied to the vanishing of $\upmu$.

First, due to the eventual vanishing of $\upmu$ and the corresponding blowup of the wave variables, we need to incorporate $\upmu$ weights into our analysis of the transport variables $(\varOmega, S)$ (and $(\mathcal{C}, \mathcal{D})$). In particular, we

[22] We use here the fact that inverting the wave operator gains one derivative.

need to incorporate $\upmu$ weights into the transport equations and energies so that the wave terms appearing as inhomogeneous terms in the energy estimates for the transport variables are regular. Importantly, despite the need to rely on $\upmu$ weights in some parts of the analysis, the "transport energy" that we construct controls a nondegenerate energy flux (i.e., an energy flux without $\upmu$ weights) on constant-u hypersurfaces $\mathcal{F}_u$. That this energy flux is bounded can be thought of as another manifestation of the transversality of the transport operator and $\mathcal{F}_u$. More precisely, with Σ_t denoting constant-t hypersurfaces, we have, roughly, L^2 estimates of the following form, where $\mathcal{P}^N$ is an order-N differential operator corresponding to repeated differentiation with respect to the $\mathcal{F}_u$-tangent vectorfields $\{L, Y, Z\}$:

$$\begin{aligned}\sup_{t'\in[0,t)} \|\sqrt{\upmu}\mathcal{P}^N(\varOmega, S)\|^2_{L^2(\Sigma_{t'})} + \sup_{u'\in[0,u)} \|\mathcal{P}^N(\varOmega, S)\|^2_{L^2(\mathcal{F}_{u'})} \\ \lesssim \text{data terms } + \text{regular wave terms } + \int_{u'=0}^{u'=u} \|\mathcal{P}^N(\varOmega, S)\|^2_{L^2(\mathcal{F}_{u'})}\, du'. \quad (1\text{-}10)\end{aligned}$$

Here, the nondegenerate energy flux (i.e., the energy along $\mathcal{F}_{u'}$ on the left-hand side of (1-10), which does not have a $\upmu$-weight) allows one to absorb the last term on the right-hand side of (1-10) using Grönwall's inequality[23] in u (as opposed to Grönwall's inequality in t which has a loss in $\upmu$). For the lower-order energies, the "regular wave terms" are indeed bounded (see (1-6)), which in total allows us to prove that the transport energies on the left-hand side of (1-10) are also bounded at the lower derivative levels.

Second, since the higher-order energies of the wave variables (v, ρ, s) can blow up as $\upmu_\star(t) \to 0$ (even in the absence of inhomogeneous terms; see (1-5)), (1-10) allows for the possibility that the higher-order energies of the transport variables $(\varOmega, S)$ (and $(\mathcal{C}, \mathcal{D})$) might also blow up. Hence, one needs to verify that there is consistency between the blowup-rates (with respect to powers of $\upmu_\star^{-1}$) associated to the different kinds of solution variables. That is, using (1-10) and the wave energy blowup-rates from (1-5), one needs to compute the expected blowup-rate of the transport variables and then plug these back into the energy estimates for the wave variables to confirm that the transport terms have an expected singularity strength that is consistent with wave energy blowup-rates. See, for example, the proof of Proposition 12.7.

Third, due to issues mentioned in Section 1A6, the transport estimates at the top-order are necessarily coupled with elliptic estimates. By their nature, the elliptic estimates treat derivatives in all spatial directions on the same footing. This clashes with the philosophy of bounding the solution with respect to the rescaled frame (which would mean that derivatives in the Y and Z frame directions should be more regular than those in the X-direction), and it leads to estimates that are singular in $\upmu_\star^{-1}$. To illustrate the difficulties and our approach to overcoming them, we first note that, suppressing many error terms, we can derive a top-order inequality of the following form, with $\underline{\partial}$ denoting Cartesian spatial derivatives and A denoting a constant depending on the equation of state:

$$\begin{aligned}&\|\sqrt{\upmu}\underline{\partial}\mathcal{P}^{N_{\text{top}}}(\varOmega, S)\|_{L^2(\Sigma_t)} \\ &\quad\le C\mathring{\varepsilon}^{3/2}\upmu^{-2M_*+2.8}(t) + A\int_{t'=0}^{t'=t} \upmu_\star^{-1}(t')\|\sqrt{\upmu}\underline{\partial}\mathcal{P}^{N_{\text{top}}}(\varOmega, S)\|_{L^2(\Sigma_{t'})}\, dt' + \cdots. \quad (1\text{-}11)\end{aligned}$$

[23] Our analysis takes place in regions of bounded u width, so that factors of e^{Cu} which arise in our Grönwall estimates can be bounded by a constant.

To apply Grönwall's inequality to (1-11), one must quantitatively control the behavior of the crucial "Grönwall factor" $\int_{t'=0}^{t'=t} A/\mu_\star(t')\,dt'$. A fundamental aspect of our analysis is that $\mu_\star(t)$ tends to 0 linearly[24] in t towards the blowup-time. It follows that one can at best prove an estimate of the form $\int_{t'=0}^{t'=t} \mu_\star^{-1}(t')\,dt' \lesssim \log(\mu_\star^{-1})(t)$ (recall that $\mu_\star(t) = \min\{1, \min_{\Sigma_t} \mu\}$, and see Proposition 8.11 for related estimates). Using only this estimate and applying Grönwall's inequality to (1-11), we find (ignoring the error terms "$\cdots$") that $\|\underline{\partial}\mathcal{P}^{N_{\mathrm{top}}}(\varOmega, S)\|_{L^2(\Sigma_t)} \lesssim \mathring{\epsilon}^{3/2}\mu_\star^{-\max\{\mathcal{O}(A),\, 2M_*-2.8\}}(t)$. Notice that unless A is small, the dominant blowup-rate in the problem would be the one corresponding to these elliptic estimates for $(\varOmega, S)$, which could in principle be much larger than the blowup-rates corresponding to the irrotational and isentropic case.[25]

However, we can prove a better result: we can show that the blowup-rates are not dominated by the top-order elliptic estimates for the transport variables, but rather by the blowup-rates for the wave variables.[26] The key to showing this is to replace the estimate (1-11) with a related L^2 estimate that features *weights in the eikonal function u*; see Proposition 11.4. Thanks to the u weights, the corresponding constant A in this analog of (1-11) can be chosen to be arbitrarily *small*, and thus the main contribution to the blowup-rate comes from the wave variables error terms, which are present in the "$\cdots$" on the right-hand side of (1-11). That this can be done is related to the fact that we have good flux estimates for top derivatives of $\mathcal{C}$ and $\mathcal{D}$ on $\mathcal{F}_u$. We refer to Propositions 11.2, 11.10, and 11.11 for the details.

1B. *Related works.*

1B1. *Shock formation in one spatial dimension.* One-dimensional shock formation has a long tradition starting from [45]. See the works of Lax [34], John [31], Liu [35], and Christodoulou and Raoul Perez [20], as well as the surveys [12; 24] for details.

1B2. *Multidimensional shock formation for quasilinear wave equations.* Multidimensional shock formation for quasilinear wave equations was first proven in Alinhac's groundbreaking papers [3; 4; 5]. Alinhac's methods allowed him to prove the formation of nondegenerate shock singularities which, roughly speaking, are shock singularities that are isolated within the constant-time hypersurface of first blowup. The problem was revisited in Christodoulou's monumental book [15], which concerned the quasilinear wave equations of irrotational and isentropic relativistic fluid mechanics. In this book, Christodoulou introduced methods that apply to a more general class of shock singularities than the nondegenerate ones treated by Alinhac and, for a large open subset of these solutions, are able to yield a complete description of the maximal smooth development, up to the boundary. This was the starting point of his follow-up breakthrough monograph [16] on the restricted shock development problem.

[24] The linear vanishing rate is crucial for the proof of Proposition 8.11 and for the Grönwall-type estimates for the energies that we carry out in Proposition 12.7 and in the Appendix. See (14-1) for a precise description of how $\mu_\star$ goes to 0.

[25] In principle, the largeness of A would not be an obstruction to closing the estimates. It would just mean that the number of derivatives needed to close the problem would increase in the presence of vorticity and entropy. We refer readers to the technical estimates in Section A9 for clarification on the role that the sizes of various constants play in determining the blowup-rates in the problem, as well as the number of derivatives needed to close the proof.

[26] In other words, our approach yields the same maximum possible high-order energy blowup-rates for the wave variables in the general case as it does for irrotational and isentropic solutions.

For quasilinear wave equations, there are many extensions, variations, and simplifications of [15], some of which adapted Christodoulou's geometric framework to other solution regimes. See, for instance, [14; 18; 19; 30; 41; 42; 48; 52].

1B3. *Multidimensional shock formation for the compressible Euler equations.* Multidimensional singularity formation for the compressible Euler equations without symmetry assumptions was first discovered by Sideris [47] via an indirect argument. A constructive proof of stable shock formation in a symmetry-reduced regime for which multidimensional phenomena (such as dispersion and vorticity) are present was given by Alinhac in [2]. See also [10; 11].

All the works in Section 1B2 on quasilinear wave equations can be used to obtain an analogous result for the compressible Euler equations in the irrotational and isentropic regime, where the dynamics reduces to a single, scalar quasilinear wave equation for a potential function. The regime of small, compact, irrotational perturbations of nonvacuum constant fluid states was treated in Christodoulou's aforementioned breakthrough work [15] in the relativistic case, and later in [19] in the nonrelativistic case.

Shock formation beyond the irrotational and isentropic regime was first proven in [36; 37; 50]. These are already discussed above; see Remark 1.3.

In very interesting recent works [10; 11], Buckmaster, Shkoller and Vicol provided a philosophically new proof of stable singularity formation without symmetry assumptions in three dimensions under adiabatic equations of state in a solution regime with vorticity and/or dynamic entropy for initial data such that precisely one singular point forms at the first singular time; these are analogs of the nondegenerate singularities that Alinhac studied [3; 4; 5] in the case of quasilinear wave equations. Moreover, in their regime (compare with Remark 1.7), they proved that the singularity is a perturbation of a self-similar Burgers shock. See also the two-dimensional precursor work [9] in symmetry, and the recent work [7], which, in two dimensions in azimuthal symmetry, constructed a set of shock-forming solutions whose cusp-like spatial behavior at the singularity is unstable (nongeneric).

1B4. *Shock development problem.* In the one-dimensional case, the theory of global solutions of small bounded variation (BV) norms [6; 28] allows one to study solutions that form shocks, as well as the subsequent interactions of the shocks in the corresponding weak solutions. In higher dimensions, the compressible Euler equations are ill-posed in BV spaces [44]. Nonetheless, in two or three dimensions, one still hopes to develop a theory that allows one to uniquely extend the solution as a piecewise smooth weak solution beyond the first shock singularity and to prove that the resulting solution has a propagating shock hypersurface. This is known as *the shock development problem.*

Even though the shock development problem for the compressible Euler equations in its full generality is open in higher dimensions, it has been solved under spherical symmetry in three dimensions, or in azimuthal symmetry in two dimensions. See [18; 55] and, most recently, [8].

In the irrotational and isentropic regime, the restricted shock development problem was solved in the recent monumental work [16] of Christodoulou *without any symmetry assumptions*. Here, the word "restricted" means that the approach of [16] does not exactly construct a weak solution to the compressible Euler equations, but instead yields a weak solution to a closely related hyperbolic PDE system such that the solution was "forced" to remain irrotational and isentropic. Nonetheless, this gives hope that under an arbitrary equation of state for the compressible Euler equations in three dimensions, one could construct a unique weak solution with a propagating shock hypersurface, starting from the first singular time exhibited in Theorem 1.1. To solve this problem would in particular require extending the ideas in [16] beyond the irrotational and isentropic regime. This is an outstanding open problem.

1B5. *Other singularities for the compressible Euler equations.* It has been known since [29; 46] that the compressible Euler equations admit self-similar solutions. Recently, this has been revisited by Merle, Raphaël, Rodnianski and Szeftel [39] to show that singularities more severe than shocks can arise in three dimensions starting from smooth initial data. See also [40; 38] for some spectacular applications.

1B6. *Singularity formation in related models.* For shock formation results concerning some other multi-speed hyperbolic problems, see [49; 51] by the second author.

Interestingly, there are also *nonhyperbolic* models with stable self-similar blowup-profiles modeled on a self-similar Burgers shock. Examples include the Burgers equation with transverse viscosity [23], the Burgers–Hilbert equations [54], and the fractal Burgers equation [13], as well as general dispersive or dissipative perturbations of the Burgers equation [43]. See also [21; 22].

1B7. *Other works.* The framework we introduced in [36; 37; 50] is useful in other low-regularity settings. See for example results on improved regularity for vorticity/entropy in [25], and results on local existence with rough data in [26; 53; 56].

1C. ***Structure of the paper.*** The remainder of the paper is structured as follows.

Sections 2–4 are introductory sections. We introduce the basic setup in Section 2, and we define the norms and energies in Section 3. The setup is similar to the setups in [36; 52]. Then in Section 4, we state our precise assumptions on the initial data and give a precise statement of our main results, which we split into several theorems and corollaries.

In Section 5, we recall the results of [50] on the reformulation of the equations, which is important for the remainder of the paper.

The bulk of paper is devoted to proving the main a priori estimates, which we state in Section 6 as Theorem 6.3. The proof of Theorem 6.3, which we provide in Section 14, relies on a set of bootstrap assumptions that we also state in Section 6. Next, after an easy (but crucial) finite-speed-of-propagation argument in Section 7, in Section 8, we cite various straightforward pointwise and L^∞ estimates for geometric quantities found in [52], and we complement these results with a few related ones that allow us to handle the transport variables.

We then turn to the main estimates in this paper. In Section 9, we carry out the transport estimates, specifically L^∞ estimates and energy estimates, for Ω, S and their derivatives. In Section 10, we prove analogous transport estimates for $\mathcal{C}$, $\mathcal{D}$, and their derivatives, except we delay the proof of the top-order estimates until the next section. In Section 11, we derive the top-order estimates for $\mathcal{C}$ and $\mathcal{D}$, which, as we described in Section 1A7, requires elliptic estimates in addition to transport estimates. In total, these estimates for the transport variables can be viewed as the main new contribution of the paper.

Next, in Section 12, we derive energy estimates for the fluid wave variables. For convenience, we have organized the wave equation estimates so that they rely on an auxiliary proposition, namely

Proposition 12.1, that provides estimates for solutions to the fluid wave equations in terms of various norms of their inhomogeneous terms, which for purposes of the proposition, we simply denote by $\mathfrak{G}$. To prove the final a priori energy estimates for the wave equations, which are located in Proposition 12.7, we must use the bounds for $\mathfrak{G}$ that we obtained in the previous sections, including the bounds for the transport variables. Since the auxiliary result Proposition 12.1 does not rely on the precise structure of $\mathfrak{G}$, it can be proved using essentially same arguments that have been used in previous works on shock formation for wave equations. For this reason, and to aid the flow of the paper, we delay the proof of Proposition 12.1 until the Appendix.

Next, in Section 13, we use the energy estimates to derive L^∞ estimates for the wave variables. In particular, these estimates yield improvements of the L^∞ bootstrap assumptions that we made in Section 6.

In Section 14, we combine the results of the previous sections to provide the proof of the main a priori estimates as well as the main theorems and their corollaries.

Finally, in the Appendix, we provide the details behind the proof of the auxiliary result Proposition 12.1. The proof relies on small modifications to the proofs of [36; 52] that account for the third spatial dimension (note that three dimensions wave equations were also handled in [15; 48]), as well as the presence of the inhomogeneous terms $\mathfrak{G}$ in the wave equations.

2. Geometric setup

In this section, we construct most of the geometric objects that we use to study shock formation and exhibit their basic properties.

2A. ***Notational conventions and remarks on constants.*** The precise definitions of some of the concepts referred to here are provided later in the article.

- Lowercase Greek spacetime indices α, β, etc. correspond to the Cartesian spacetime coordinates (see Section 2C) and vary over 0, 1, 2, 3. Lowercase Latin spatial indices a,b, etc. correspond to the Cartesian spatial coordinates and vary over 1, 2, 3. Uppercase Latin spatial indices A,B, etc. correspond to the coordinates on $\ell_{t,u}$ and vary over 2, 3. All lowercase Greek indices are lowered and raised with the acoustical metric g and its inverse g^{-1}, and *not with the Minkowski metric*. We use Einstein's summation convention in that repeated indices are summed.

- By "$\cdot$" we denote the natural contraction between two tensors. For example, if ξ is a spacetime one-form and V is a spacetime vectorfield, then $\xi \cdot V \doteq \xi_\alpha V^\alpha$.

- If ξ is an $\ell_{t,u}$-tangent one-form (as defined in Section 2J), then $\xi^\#$ denotes its $g\!\!\!/$-dual vectorfield, where $g\!\!\!/$ is the Riemannian metric induced on $\ell_{t,u}$ by g. Similarly, if ξ is a symmetric type-$\binom{0}{2}$ $\ell_{t,u}$-tangent tensor, then $\xi^\#$ denotes the type-$\binom{1}{1}$ $\ell_{t,u}$-tangent tensor formed by raising one index with $g\!\!\!/^{-1}$ and $\xi^{\#\#}$ denotes the type-$\binom{2}{0}$ $\ell_{t,u}$-tangent tensor formed by raising both indices with $g\!\!\!/^{-1}$.

- If V is an $\ell_{t,u}$-tangent vectorfield, then $V_\flat$ denotes its $g\!\!\!/$-dual one-form.

- If V and W are vectorfields, then $V_W \doteq V^\alpha W_\alpha = g_{\alpha\beta} V^\alpha W^\beta$.

• If ξ is a one-form and V is a vectorfield, then $\xi_V \doteq \xi_\alpha V^\alpha$. We use similar notation when contracting higher-order tensorfields against vectorfields. For example, if ξ is a type-$\binom{0}{2}$ tensorfield and V and W are vectorfields, then $\xi_{VW} \doteq \xi_{\alpha\beta} V^\alpha W^\beta$.

• Unless otherwise indicated, all quantities in our estimates that are not explicitly under an integral are viewed as functions of the geometric coordinates (t, u, x^2, x^3). Unless otherwise indicated, integrands have the functional dependence established below in Definition 3.1.

• $[Q_1, Q_2] = Q_1 Q_2 - Q_2 Q_1$ denotes the commutator of the operators Q_1 and Q_2.

• $A \lesssim B$ means that there exists $C > 0$ such that $A \leq CB$. $A \approx B$ means that $A \lesssim B$ and $B \lesssim A$. $A = \mathcal{O}(B)$ means that $|A| \lesssim |B|$.

• The constants C are free to vary from line to line. *These constants, and implicit constants as well, are allowed to depend on the equation of state, the background* $\bar{\varrho}$, *the maximum number of times* N_{top} *that we commute the equations, and the parameters* $\mathring{\sigma}$, $\mathring{\delta}$ *and* $\mathring{\delta}_*^{-1}$ from Section 4A.

• Constants $C_\blacklozenge$ are also allowed to vary from line to line, but unlike C, the $C_\blacklozenge$ are *only allowed to depend on the equation of state and the background* $\bar{\varrho}$.

• In the Appendix, there appear *absolute constants* M_{abs}, which can be chosen to be independent of the equation of state and all other parameters in the problem.

• For our proof to close, the high-order energy blowup-rate parameter M_* needs to be chosen to be large in a manner that depends only on M_{abs}; hence, M_* can also be chosen to be an absolute constant.

• The integer N_{top} denotes the maximum number of times we need to commute the equations to close the estimates. For our proof to close, N_{top} needs to be chosen to be large in a manner that depends only on M_*. N_{top} could be chosen to be an absolute constant, but we choose to think of it as a parameter that we are free to adjust so that we can study solutions with arbitrary sufficiently large regularity.

• For our proof to close, the data-size parameters $\mathring{\alpha}$ and $\mathring{\epsilon}$ must be chosen to be sufficiently small, where the required smallness is clarified in Theorem 6.3. We always assume that $\mathring{\epsilon}^{1/2} \leq \mathring{\alpha}$.

• $A \lesssim_\blacklozenge B$ means that $A \leq C_\blacklozenge B$, with $C_\blacklozenge$ as above. Similarly, $A = \mathcal{O}_\blacklozenge(B)$ means that $|A| \leq C_\blacklozenge |B|$.

• For example, $\mathring{\delta}_*^{-2} = \mathcal{O}(1)$, $2 + \mathring{\alpha} + \mathring{\alpha}^2 = \mathcal{O}_\blacklozenge(1)$, $\mathring{\alpha}\mathring{\epsilon} = \mathcal{O}(\mathring{\epsilon})$, $C_\blacklozenge \mathring{\alpha}^2 = \mathcal{O}_\blacklozenge(\mathring{\alpha})$, $N!\,\mathring{\epsilon} = \mathcal{O}(\mathring{\epsilon})$, and $C\mathring{\alpha} = \mathcal{O}(1)$. Some of these examples are nonoptimal; e.g., we actually have $\mathring{\alpha}\mathring{\epsilon} = \mathcal{O}_\blacklozenge(\mathring{\epsilon})$.

• $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ respectively denote the standard floor and ceiling functions.

2B. *Caveats on citations.* Before we introduce our geometric setup, we should say that our setup is essentially the same as that in [36; 52], except for some small differences. We will therefore cite whenever possible the computations in [36; 52], except we will need to take into account the following differences:

• The work [52] allows for very general metrics, while in the present paper, we are only concerned with the acoustical metric for the compressible Euler equations. In citing [52], we sometimes adjust formulas to take into account the explicit form of the Cartesian metric components $g_{\alpha\beta}$ stated in Definition 2.9.

- The papers [36; 52] concern two spatial dimensions (with ambient manifold $\Sigma = \mathbb{R} \times \mathbb{T}$), while in the present paper, we are concerned with three spatial dimensions (with $\Sigma = \mathbb{R} \times \mathbb{T}^2$).
- In [52], the metric components $g_{\alpha\beta}$ were functions of a scalar function Ψ, as opposed to the array $\vec{\Psi}$ (defined in (2-3)). For this reason, we must make minor adjustments to many of the formulas from [52] to account for the fact that in the present article, $\vec{\Psi}$ is an array.

In all cases, our minor adjustments can easily be verified by examining the proof in [52].

2C. ***Basic setup and ambient manifold.*** We recall again the setup from the Introduction. We will work on the spacetime manifold $I \times \Sigma$ (with $I \subseteq \mathbb{R}$ a time interval and $\Sigma \doteq \mathbb{R} \times \mathbb{T}^2$ the spatial domain). We fix a standard Cartesian coordinate system $\{x^\alpha\}_{\alpha=0,1,2,3}$ on $I \times \Sigma$, where $t \doteq x^0 \in I$ is the time coordinate and $x \doteq (x^1, x^2, x^3) \in \mathbb{R} \times \mathbb{T}^2$ are the spatial coordinates.[27] We use the notation $\{\partial_\alpha\}_{\alpha=0,1,2,3}$ (or $\partial_t \doteq \partial_0$) to denote the Cartesian coordinate partial derivative vectorfields.

In this coordinate system, the *plane-symmetric* solutions are exactly those whose fluid variables are independent of (x^2, x^3).

2D. ***Fluid variables and new variables useful for the reformulation.*** As we already discussed in Section 1A3, at the heart of our approach is a reformulation of the compressible Euler equations in terms of new variables. We introduce these new variables in this subsection; see Definitions 2.3 and 2.7.

The basic fluid variables are (ϱ, v^i, s) (see the Introduction). We fix an equation of state $p = p(\varrho, s)$ and a constant $\bar{\varrho} > 0$ such that $p_{;\varrho}(\bar{\varrho}, 0) = 1$.

Definition 2.1. Define the *logarithmic density* ρ and the *speed of sound* $c(\rho, s)$ by

$$\rho = \log\left(\frac{\varrho}{\bar{\varrho}}\right), \quad c(\rho, s) = \sqrt{\frac{\partial p}{\partial \varrho}(\varrho, s)}.$$

Remark 2.2. As is suggested by our notation, we will consider $c(\rho, s)$ as a function of (ρ, s). The normalization of $p_{;\varrho}$ that we stated above is equivalent to

$$c(0, 0) = 1. \tag{2-1}$$

Definition 2.3 (the fluid variables arrays).

(1) Define the *almost Riemann invariants*[28] $\mathcal{R}_{(\pm)}$ as follows (recall Definition 2.1):

$$\mathcal{R}_{(\pm)} \doteq v^1 \pm F(\rho, s), \quad F(\rho, s) \doteq \int_0^\rho c(\rho', s)\, d\rho'. \tag{2-2}$$

[27] While the coordinates x^2, x^3 on $\mathbb{T}^2$ are only locally defined, the corresponding partial derivative vectorfields ∂_2, ∂_3 can be extended so as to form a global smooth frame on $\mathbb{T}^2$. Similar remarks apply to the one-forms dx^2, dx^3 These simple observations are relevant for this paper because when we derive estimates, the coordinate functions x^2, x^3 themselves are never directly relevant; what matters are estimates for the components of various tensorfields with respect to the frame $\{\partial_t, \partial_1, \partial_2, \partial_3\}$ and the basis dual coframe $\{dt, dx^1, dx^2, dx^3\}$, which are everywhere smooth.

[28] $\mathcal{R}_{(\pm)}$ coincide with the well-known *Riemann invariants* in the plane-symmetric isentropic case. Even though they are no longer "invariant" in our case, they are useful in capturing smallness.

(2) Define the *array of wave variables*:[29]

$$\vec{\Psi} \doteq (\Psi_1, \Psi_2, \Psi_3, \Psi_4, \Psi_5) \doteq (\mathcal{R}_{(+)}, \mathcal{R}_{(-)}, v^2, v^3, s). \tag{2-3}$$

Remark 2.4. We sometimes use the simpler notation Ψ in place of $\vec{\Psi}$ when there is no danger of confusion. At other times, we use the notation Ψ to denote a generic element of $\vec{\Psi}$. The precise meaning of the symbol Ψ will be clear from context.

Remark 2.5 (clarification on our approach to estimating ρ and v^1). Recall that we have introduced $\mathcal{R}_{(\pm)}$ to allow us to capture the fact that our solutions are perturbations of simple plane waves (for which only $\mathcal{R}_{(+)}$ is nonvanishing). In the one-dimensional isentropic case, $\{\mathcal{R}_{(+)}, \mathcal{R}_{(-)}\}$ can be taken to be the unknowns in place of $\{\rho, v^1\}$. A similar remark holds in the present three-dimensional case as well, provided we take into account the entropy. Specifically, from (2-1) and Definition 2.3, it follows that $v^1 = \frac{1}{2}(\mathcal{R}_{(+)} + \mathcal{R}_{(-)})$, and that when ρ, v^1, and s are sufficiently small (as is captured by the smallness parameters $\mathring{\alpha}$ and $\mathring{\epsilon}$ described at the beginning of Section 4A), we have (via the implicit function theorem) $\rho = (\mathcal{R}_{(+)} - \mathcal{R}_{(-)}) \cdot \widetilde{F}(\mathcal{R}_{(+)} - \mathcal{R}_{(-)}, s)$, where $\widetilde{F}$ is a smooth function. This allows us to control ρ and v^1 in terms of $\mathcal{R}_{(+)}$, $\mathcal{R}_{(-)}$, and s. Throughout the article, we use this observation without explicitly pointing it out. In particular, even though many of the equations we cite explicitly involve ρ and v^1, it should be understood that we always estimate these quantities in terms of the wave variables $\mathcal{R}_{(+)}$, $\mathcal{R}_{(-)}$, and s, which are featured in the array (2-3).

Definition 2.6 (Euclidean divergence and curl). Denote by[30] div and curl the Euclidean spatial divergence and curl operator. That is, given a Σ_t-tangent vectorfield $V = V^a \partial_a$, define

$$\operatorname{div} V \doteq \partial_a V^a, \qquad (\operatorname{curl} V)^i \doteq \epsilon_{iab} \partial_a V^b, \tag{2-4}$$

where ϵ_{iab} is the fully antisymmetric symbol normalized by $\epsilon_{123} = 1$.

Definition 2.7 (the higher-order variables).

(1) Define the *specific vorticity* to be the Σ_t-tangent vectorfield with the Cartesian spatial components

$$\varOmega^i \doteq \frac{(\operatorname{curl} v)^i}{\varrho/\bar{\varrho}} = \frac{(\operatorname{curl} v)^i}{\exp(\rho)}.$$

(2) Define the *entropy gradient* to be the Σ_t-tangent vectorfield with the Cartesian spatial components

$$S^i \doteq \partial_i s.$$

(3) Define the *modified fluid variables* by

$$\mathcal{C}^i \doteq \exp(-\rho)(\operatorname{curl} \varOmega)^i + \exp(-3\rho) c^{-2} \frac{p_{;s}}{\bar{\varrho}} S^a \partial_a v^i - \exp(-3\rho) c^{-2} \frac{p_{;s}}{\bar{\varrho}} (\partial_a v^a) S^i, \tag{2-5a}$$

$$\mathcal{D} \doteq \exp(-2\rho) \operatorname{div} S - \exp(-2\rho) S^a \partial_a \rho. \tag{2-5b}$$

We think of $\mathcal{C}$ as a Σ_t-tangent vectorfield with Cartesian spatial components given by (2-5a).

[29] Throughout, we consider $\vec{\Psi}$ as an array of scalar functions; we will not attribute any tensorial structure to the labeling index ι of Ψ_ι besides simple contractions, denoted by $\diamond$, corresponding to the chain rule; see Definition 2.13.

[30] This is in contrast to $\operatorname{d\!\!/iv}$; see Definition 2.33.

2E. ***The acoustical metric and related objects in Cartesian coordinates.*** Hidden within compressible Euler flow lies a geometric structure captured by the acoustical metric, which governs the dynamics of the sound waves. We introduce in this subsection the acoustical metric g in Cartesian coordinates.

Definition 2.8 (material derivative vectorfield). We define the *material derivative vectorfield* as follows relative to the Cartesian coordinates:

$$B \doteq \partial_t + v^a \partial_a. \tag{2-6}$$

Definition 2.9 (the acoustical metric). Define the *acoustical metric* g (in Cartesian coordinates) by

$$g \doteq -dt \otimes dt + c^{-2} \sum_{a=1}^{3} (dx^a - v^a\, dt) \otimes (dx^a - v^a\, dt). \tag{2-7}$$

The following lemma follows from straightforward computations.

Lemma 2.10 (the inverse acoustical metric). *The inverse of the acoustical metric g from* (2-7) *can be expressed as*

$$g^{-1} = -B \otimes B + c^2 \sum_{a=1}^{3} \partial_a \otimes \partial_a. \tag{2-8}$$

Remark 2.11 (closeness to the Minkowski metric). In our analysis, v and $c - 1$ will be small, where the smallness is captured by the parameters $\mathring{\alpha}$ and $\mathring{\varepsilon}$ described at the beginning of Section 4A. Recalling (2-7), we see that g will be L^∞-close to the Minkowski metric. It is therefore convenient to introduce the decomposition

$$g_{\alpha\beta}(\vec{\Psi}) = m_{\alpha\beta} + g^{(\text{small})}_{\alpha\beta}(\vec{\Psi}), \quad m_{\alpha\beta} \doteq \operatorname{diag}(-1, 1, 1, 1), \tag{2-9}$$

where m is the Minkowski metric and $g^{(\text{small})}_{\alpha\beta}(\vec{\Psi})$ is a smooth function of $\vec{\Psi}$ such that

$$g^{(\text{small})}_{\alpha\beta}(\vec{\Psi} = 0) = 0. \tag{2-10}$$

Definition 2.12 ($\vec{\Psi}$-derivatives of $g_{\alpha\beta}$). For $\alpha, \beta = 0, \dots, 3$ and $\imath = 1, \dots, 5$, we define

$$G^{\imath}_{\alpha\beta}(\vec{\Psi}) \doteq \frac{\partial}{\partial \Psi_\imath} g_{\alpha\beta}(\vec{\Psi}), \quad \vec{G}_{\alpha\beta} = \vec{G}_{\alpha\beta}(\vec{\Psi}) \doteq \big(G^1_{\alpha\beta}(\vec{\Psi}), G^2_{\alpha\beta}(\vec{\Psi}), G^3_{\alpha\beta}(\vec{\Psi}), G^4_{\alpha\beta}(\vec{\Psi}), G^5_{\alpha\beta}(\vec{\Psi})\big). \tag{2-11}$$

For each fixed $\imath \in \{1, \dots, 5\}$, we think of $\{G^{\imath}_{\alpha\beta}\}_{\alpha,\beta=0,\dots,3}$, as the Cartesian components of a spacetime tensorfield. Similarly, we think of $\{\vec{G}_{\alpha\beta}\}_{\alpha,\beta=0,\dots,3}$ as the Cartesian components of an array-valued spacetime tensorfield.

Definition 2.13 (operators involving $\vec{\Psi}$). Let U_1, U_2, V be vectorfields. We define

$$V\vec{\Psi} \doteq (V\Psi_1, V\Psi_2, V\Psi_3, V\Psi_4, V\Psi_5), \quad \vec{G}_{U_1 U_2} \diamond V\vec{\Psi} \doteq \sum_{\imath=1}^{5} G^{\imath}_{\alpha\beta} U_1^\alpha U_2^\beta V\Psi_\imath. \tag{2-12}$$

We use similar notation with other differential operators in place of vectorfield differentiation. For example, $\vec{G}_{U_1U_2} \diamond \not{\Delta}\vec{\Psi} \doteq \sum_{\imath=1}^{5} G^{\imath}_{\alpha\beta} U_1^\alpha U_2^\beta \not{\Delta}\Psi_\imath$ (where $\not{\Delta}$ is defined in Definition 2.33).

2F. ***The acoustic eikonal function and related constructions.*** To control the solution up to the shock, we will crucially rely on an eikonal function for the acoustical metric.

Definition 2.14 (acoustic eikonal function). The acoustic eikonal function (eikonal function for short) u solves the eikonal equation initial value problem

$$(g^{-1})^{\alpha\beta}\partial_\alpha u \partial_\beta u = 0, \quad \partial_t u > 0, \quad u \restriction_{t=0} = \mathring{\sigma} - x^1, \tag{2-13}$$

where $\mathring{\sigma} > 0$ is the constant controlling the initial support (recall Theorem 1.1).

Definition 2.15 (inverse foliation density). Define the inverse foliation density μ by

$$\mu \doteq \frac{-1}{(g^{-1})^{\alpha\beta}(\vec{\Psi})\partial_\alpha t \partial_\beta u} > 0. \tag{2-14}$$

Note that $1/\mu$ measures the density of the level sets of u relative to the constant-time hypersurfaces Σ_t. For the data that we will consider, we have $\mu \restriction_{\Sigma_0} \approx 1$. When μ vanishes, the level sets of u intersect and, as it turns out, $\max_{\alpha=0,1,2,3} |\partial_\alpha u|$ and $\max_{\alpha=0,1,2,3} |\partial_\alpha \mathcal{R}_{(+)}|$, blow up.

The following quantities, tied to μ, play an important role in our description of the singular behavior of our high-order energies.

Definition 2.16. Define $\mu_\star(t,u)$ and $\mu_\star(t)$ by[31]

$$\mu_\star(t,u) \doteq \min\bigl\{1, \min_{u' \le u} \mu(t,u')\bigr\}, \quad \mu_\star(t) \doteq \min\bigl\{1, \min_{\Sigma_t} \mu\bigr\}.$$

2G. ***Subsets of spacetimes.***

Definition 2.17 (subsets of spacetime). For $0 \le t'$ and $0 \le u'$, define

$$\Sigma_{t'} \doteq \{(t,x) \in \mathbb{R} \times (\mathbb{R} \times \mathbb{T}^2) \mid t = t'\}, \tag{2-15a}$$

$$\Sigma_{t'}^{u'} \doteq \{(t,x) \in \mathbb{R} \times (\mathbb{R} \times \mathbb{T}^2) \mid t = t',\ 0 \le u(t,x) \le u'\}, \tag{2-15b}$$

$$\mathcal{F}_{u'} \doteq \{(t,x) \in \mathbb{R} \times (\mathbb{R} \times \mathbb{T}^2) \mid u(t,x) = u'\}, \tag{2-15c}$$

$$\mathcal{F}_{u'}^{t'} \doteq \{(t,x) \in \mathbb{R} \times (\mathbb{R} \times \mathbb{T}^2) \mid 0 \le t \le t',\ u(t,x) = u'\}, \tag{2-15d}$$

$$\ell_{t',u'} \doteq \mathcal{F}_{u'}^{t'} \cap \Sigma_{t'}^{u'} = \{(t,x) \in \mathbb{R} \times (\mathbb{R} \times \mathbb{T}^2) \mid t = t',\ u(t,x) = u'\}, \tag{2-15e}$$

$$\mathcal{M}_{t',u'} \doteq \bigcup_{u \in [0,u']} \mathcal{F}_u^{t'} \cap \{(t,x) \in \mathbb{R} \times (\mathbb{R} \times \mathbb{T}^2) \mid 0 \le t < t'\}. \tag{2-15f}$$

We refer to the Σ_t and Σ_t^u as "constant time slices," the $\mathcal{F}_u$ and $\mathcal{F}_u^t$ as "null hyperplanes," "null hypersurfaces," "characteristics," or "acoustic characteristics," and the $\ell_{t,u}$ as "tori." Note that $\mathcal{M}_{t,u}$ is "open-at-the-top" by construction.

[31] By definition, $\mu_\star(t,u) \ge \mu_\star(t)$ for all $u \in \mathbb{R}$. Note that by the localization lemma (Lemma 7.1) we prove below, we have $\mu_\star(t) = \mu_\star(t, U_0)$. In most of the proof, it suffices to consider the function $\mu_\star(t)$ without considering $\mu_\star(t,u)$. The more refined definition for $\mu_\star(t,u)$ will only be referred to in the Appendix, so that the formulas take the same forms as their counterparts in [36; 52].

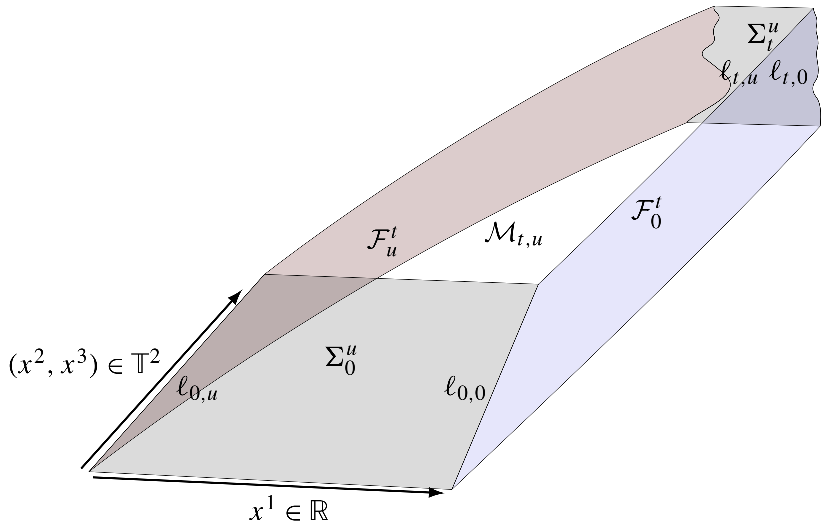

Figure 2. The spacetime region and various subsets. The (unlabeled and uncolored) flat front and back surfaces should be identified.

2H. *Important vectorfields, the rescaled frame, and the nonrescaled frame.*

Definition 2.18 (important vectorfields). (1) Define the *geodesic null vectorfield* by

$$L^{\nu}_{(\mathrm{Geo})} \doteq -(g^{-1})^{\nu\alpha}\, \partial_\alpha u. \tag{2-16}$$

(2) Define the *rescale null vectorfield* (recall the definition of μ in (2-14)) by

$$L \doteq \mu L_{(\mathrm{Geo})}. \tag{2-17}$$

(3) Define X to be the unique vectorfield that is Σ_t-tangent, g-orthogonal to the $\ell_{t,u}$, and normalized by

$$g(L, X) = -1. \tag{2-18}$$

Define the "rescaled" vectorfield $\breve{X}$ by

$$\breve{X} \doteq \mu X. \tag{2-19}$$

(4) Define Y and Z respectively to be the g-orthogonal projection[32] of the Cartesian partial derivative vectorfields ∂_2 and ∂_3 to the tangent space of $\ell_{t,u}$, i.e.,

$$Y \doteq \partial_2 - g(\partial_2, X)X, \qquad Z \doteq \partial_3 - g(\partial_3, X)X. \tag{2-20}$$

(5) We will use vectorfields in $\mathscr{P} \doteq \{L, Y, Z\}$ for commutation, and we therefore refer to them as *commutation vectorfields*. An element of $\mathscr{P}$ will often be denoted schematically by $\mathcal{P}$ (see also Definition 3.4).

We collect some basic properties of these vectorfields; see [52, (2.12), (2.13) and Lemma 2.1] for proofs.

[32]To see that Y and Z are tangent to $\ell_{t,u}$, one can use (2-18), (2-23), the fact that B is g-orthogonal to Σ_t, and the fact that ∂_i is tangent to Σ_t. Alternatively, see (2-30b).

Lemma 2.19 (basic properties of the vectorfields).

(1) $L_{(\mathrm{Geo})}$ *is geodesic and null, i.e.,*

$$g(L_{(\mathrm{Geo})}, L_{(\mathrm{Geo})}) = 0, \quad \mathscr{D}_{L_{(\mathrm{Geo})}} L_{(\mathrm{Geo})} = 0,$$

where $\mathscr{D}$ is the Levi-Civita connection associated to g.

(2) *The following identities hold*:

$$Lu = 0, \quad Lt = L^0 = 1, \qquad \breve{X}u = 1, \quad \breve{X}t = \breve{X}^0 = 0, \tag{2-21}$$

$$g(X, X) = 1, \quad g(\breve{X}, \breve{X}) = \mu^2, \quad g(L, X) = -1, \quad g(L, \breve{X}) = -\mu. \tag{2-22}$$

(3) *The vectorfield B (see (2-6)) is future-directed, g-orthogonal to Σ_t, and is normalized by $g(B,B)=-1$. Moreover,*

$$B = \partial_t + v^a \partial_a = L + X, \tag{2-23}$$

$$B_\alpha = -\delta_\alpha^0, \tag{2-24}$$

where δ_α^β is the Kronecker delta.

2I. *Transformations.* Having introduced various vectorfields in Section 2H, we now derive some related transformation formulas that we will use later on.

Definition 2.20 (coordinate vectorfields in geometric (t, u, x^2, x^3)-coordinates). Define $(\partial\!\!\!/_t, \partial\!\!\!/_u, \partial\!\!\!/_2, \partial\!\!\!/_3)$ to be the coordinate partial derivative vectorfields in the geometric (t, u, x^2, x^3)-coordinate system.

Definition 2.21 (Cartesian components of geometric vectorfields).

(1) Define L^i and X^i to be the Cartesian i-th components of L and X respectively. (Note $L^i + X^i - v^i = 0$; see (2-23).)

(2) Define[33] $L_{(\mathrm{small})}$ and $X_{(\mathrm{small})}$ by

$$L^1_{(\mathrm{small})} \doteq L^1 - 1, \quad L^2_{(\mathrm{small})} \doteq L^2, \quad L^3_{(\mathrm{small})} \doteq L^3, \tag{2-25a}$$

$$X^1_{(\mathrm{small})} \doteq X^1 + 1, \quad X^2_{(\mathrm{small})} \doteq X^2, \quad X^3_{(\mathrm{small})} \doteq X^3. \tag{2-25b}$$

Lemma 2.22 (relations between $\{\partial_\alpha\}_{\alpha=0,1,2,3}$ and $\{L, X, Y, Z\}$). *The following identities hold*:

$$\partial_t \doteq \partial_0 = L + X - v^a \partial_a, \tag{2-26a}$$

$$\partial_1 = c^{-2} X^1 X - \frac{X^2}{X^1} Y - \frac{X^3}{X^1} Z, \tag{2-26b}$$

$$\partial_2 = Y + (c^{-2} X^2) X, \quad \partial_3 = Z + (c^{-2} X^3) X. \tag{2-26c}$$

Proof. Equation (2-26a) is simply a restatement of (2-23), and (2-26c) follows from (2-20) and $g(\partial_A, X) = c^{-2} X^A$ for $A = 2, 3$ (see (2-7)). Finally, to obtain (2-26b), we write $X = X^a \partial_a$ and use (2-26c) to obtain

$$\partial_1 = \frac{1}{X^1}[1 - c^{-2}((X^2)^2 + (X^3)^2)]X - \frac{X^2}{X^1} Y - \frac{X^3}{X^1} Z.$$

This then implies (2-26b) since $\sum_{a=1}^3 (X^a)^2 = c^2$ by $g(X, X) = 1$ (see (2-22)) and (2-7). □

[33] The notation is suggestive of the fact that these quantities are of size $\mathcal{O}(\mathring{\alpha})$ (and hence small).

Lemma 2.23 (relation between $\{\partial\!\!\!/_\alpha\}$ and $\{L, X, Y, Z\}$). *The following identities hold, where repeated capital Latin indices are summed over* $A = 2, 3$:

$$L = \partial\!\!\!/_t + L^A \partial\!\!\!/_A, \qquad \breve{X} = \partial\!\!\!/_u + \mu X^A \partial\!\!\!/_A, \tag{2-27a}$$

$$Y = (1 - c^{-2}(X^2)^2)\partial\!\!\!/_2 - c^{-2} X^2 X^3 \partial\!\!\!/_3, \qquad Z = (1 - c^{-2}(X^3)^2)\partial\!\!\!/_3 - c^{-2} X^2 X^3 \partial\!\!\!/_2. \tag{2-27b}$$

Proof. Equation (2-27a) is an immediate consequence of (2-21) (and (2-19)).

To derive the first equation in (2-27b), simply note that $Yx^2 = 1 - c^{-2}(X^2)^2$ and $Yx^3 = -cX^2X^3$ by (2-26c), and that $Yt = Yu = 0$ since Y is $\ell_{t,u}$-tangent. The second equation in (2-27b) follows from similar reasoning. □

Lemma 2.24 (relation[34] between $\{\partial_a\}_{a=1,2,3}$, $\{\partial\!\!\!/_u, \partial\!\!\!/_2, \partial\!\!\!/_3\}$, and $\{\breve{X}, Y, Z\}$). *The following identities hold*:

$$\partial\!\!\!/_u = \frac{\mu c^2}{X^1}\partial_1 = \breve{X} - \mu c^2 \frac{X^2}{(X^1)^2} Y - \mu c^2 \frac{X^3}{(X^1)^2} Z, \tag{2-28a}$$

$$\partial\!\!\!/_2 = \partial_2 - \frac{X^2}{X^1}\partial_1 = \left\{1 + \left(\frac{X^2}{X^1}\right)^2\right\} Y + \frac{X^2 X^3}{(X^1)^2} Z, \tag{2-28b}$$

$$\partial\!\!\!/_3 = \partial_3 - \frac{X^3}{X^1}\partial_1 = \frac{X^2 X^3}{(X^1)^2} Y + \left\{1 + \left(\frac{X^3}{X^1}\right)^2\right\} Z. \tag{2-28c}$$

Proof. It suffices to derive the identities

$$\partial\!\!\!/_u x^1 = \frac{\mu c^2}{X^1}, \quad \partial\!\!\!/_2 x^1 = -\frac{X^2}{X^1}, \quad \partial\!\!\!/_3 x^1 = -\frac{X^3}{X^1}; \tag{2-29}$$

it is straightforward to see that the first identities in each of (2-28a)–(2-28c) follow from (2-29); the second identities in (2-28a)–(2-28c) then follow from the first ones and Lemma 2.22. To prove (2-29), we invert (2-27b) to obtain (with the help of the identity $\sum_{a=1}^{3}(X^a)^2 = c^2$, which follows from (2-22) and (2-7)):

$$\partial\!\!\!/_2 = \left\{\frac{c^2}{(X^1)^2} - \left(\frac{X^3}{X^1}\right)^2\right\} Y + \frac{X^2 X^3}{(X^1)^2} Z, \quad \partial\!\!\!/_3 = \frac{X^2 X^3}{(X^1)^2} Y + \left\{\frac{c^2}{(X^1)^2} - \left(\frac{X^2}{X^1}\right)^2\right\} Z.$$

On the other hand, by (2-26c), $Yx^1 = -c^{-2}X^2X^1$ and $Zx^1 = -c^{-2}X^3X^1$. Hence,

$$\partial\!\!\!/_2 x^1 = -\frac{X^2}{X^1}, \quad \partial\!\!\!/_3 x^1 = -\frac{X^3}{X^1}.$$

Plugging back into the second identity in (2-27a), we obtain

$$\partial\!\!\!/_u x^1 = \mu X^1 - \sum_{A=2}^{3} \mu X^A \partial\!\!\!/_A x^1 = \mu X^1 + \sum_{A=2}^{3} \mu \frac{(X^A)^2}{X^1} = \frac{\mu c^2}{X^1},$$

where we again used $\sum_{i=1}^{3}(X^i)^2 = c^2$. □

[34] We could also obtain $\partial\!\!\!/_t = \partial_t + (L^1 + (X^2L^2 + X^3L^3)/X_1)\partial_1$. Since this will not be explicitly needed, we will not prove it.

2J. *Projection tensorfields, $\vec{G}_{(\mathrm{frame})}$, and projected Lie derivatives.*

Definition 2.25 (projection tensorfields). We define the Σ_t projection tensorfield[35] $\underline{\Pi}$ and the $\ell_{t,u}$ projection tensorfield $\not{\Pi}$ relative to Cartesian coordinates as

$$\underline{\Pi}_\nu^{\ \mu} \doteq \delta_\nu^{\ \mu} + B_\nu B^\mu = \delta_\nu^{\ \mu} - \delta_\nu^{\ 0} L^\mu - \delta_\nu^{\ 0} X^\mu, \tag{2-30a}$$

$$\not{\Pi}_\nu^{\ \mu} \doteq \delta_\nu^{\ \mu} + X_\nu L^\mu + L_\nu (L^\mu + X^\mu) = \delta_\nu^{\ \mu} - \delta_\nu^{\ 0} L^\mu + L_\nu X^\mu. \tag{2-30b}$$

In (2-30a)–(2-30b), $\delta_\nu^{\ \mu}$ is the standard Kronecker delta. The last equalities in (2-30a) and (2-30b) follow from (2-23)–(2-24).

Definition 2.26 (projections of tensorfields). Given any type-$\binom{m}{n}$ spacetime tensorfield ξ, we define its Σ_t projection $\underline{\Pi}\xi$ and its $\ell_{t,u}$ projection $\not{\Pi}\xi$ as

$$(\underline{\Pi}\xi)^{\mu_1\cdots\mu_m}_{\nu_1\cdots\nu_n} \doteq \underline{\Pi}_{\widetilde{\mu}_1}^{\ \mu_1} \cdots \underline{\Pi}_{\widetilde{\mu}_m}^{\ \mu_m} \underline{\Pi}_{\nu_1}^{\ \widetilde{\nu}_1} \cdots \underline{\Pi}_{\nu_n}^{\ \widetilde{\nu}_n} \xi^{\widetilde{\mu}_1\cdots\widetilde{\mu}_m}_{\widetilde{\nu}_1\cdots\widetilde{\nu}_n}, \tag{2-31a}$$

$$(\not{\Pi}\xi)^{\mu_1\cdots\mu_m}_{\nu_1\cdots\nu_n} \doteq \not{\Pi}_{\widetilde{\mu}_1}^{\ \mu_1} \cdots \not{\Pi}_{\widetilde{\mu}_m}^{\ \mu_m} \not{\Pi}_{\nu_1}^{\ \widetilde{\nu}_1} \cdots \not{\Pi}_{\nu_n}^{\ \widetilde{\nu}_n} \xi^{\widetilde{\mu}_1\cdots\widetilde{\mu}_m}_{\widetilde{\nu}_1\cdots\widetilde{\nu}_n}. \tag{2-31b}$$

We say that a spacetime tensorfield ξ is Σ_t-tangent (respectively $\ell_{t,u}$-tangent) if $\underline{\Pi}\xi = \xi$ (respectively if $\not{\Pi}\xi = \xi$). Alternatively, we say that ξ is a Σ_t tensor (respectively $\ell_{t,u}$ tensor).

Definition 2.27 ($\ell_{t,u}$ projection notation). If ξ is a spacetime tensor, then $\not{\xi} \doteq \not{\Pi}\xi$.

If ξ is a symmetric type-$\binom{0}{2}$ spacetime tensor and V is a spacetime vectorfield, then $\not{\xi}_V \doteq \not{\Pi}(\xi_V)$, where ξ_V is the spacetime one-form with Cartesian components $\xi_{\alpha\nu} V^\alpha$, $(\nu = 0, 1, 2, 3)$.

Remark 2.28 (clarification of the symbols $(\not{\partial}_t, \not{\partial}_u, \not{\partial}_2, \not{\partial}_3)$). We caution that the coordinate partial derivative vectorfields $(\not{\partial}_t, \not{\partial}_u, \not{\partial}_2, \not{\partial}_3)$ from Definition 2.20 are not $\ell_{t,u}$ projections of other vectorfields; i.e., for $(\not{\partial}_t, \not{\partial}_u, \not{\partial}_2, \not{\partial}_3)$, we are not using the "slash conventions" of Definition 2.27.

Throughout, $\mathcal{L}_V \xi$ denotes the Lie derivative of the tensorfield ξ with respect to the vectorfield V. We often use the Lie bracket notation $[V, W] \doteq \mathcal{L}_V W$ when V and W are vectorfields.

Definition 2.29 (Σ_t- and $\ell_{t,u}$-projected Lie derivatives). If ξ is a tensorfield and V is a vectorfield, we define the Σ_t-projected Lie derivative $\underline{\mathcal{L}}_V \xi$ and the $\ell_{t,u}$-projected Lie derivative $\not{\mathcal{L}}_V \xi$ as

$$\underline{\mathcal{L}}_V \xi \doteq \underline{\Pi}\mathcal{L}_V \xi, \qquad \not{\mathcal{L}}_V \xi \doteq \not{\Pi}\mathcal{L}_V \xi. \tag{2-32}$$

Definition 2.30 (components of $\vec{G}$ relative to the nonrescaled frame). We define

$$\vec{G}_{(\mathrm{frame})} \doteq \{\vec{G}_{LL}, \vec{G}_{LX}, \vec{G}_{XX}, \not{\vec{G}}_L, \not{\vec{G}}_X, \not{\vec{G}}\}, \tag{2-33}$$

where $\vec{G}_{\alpha\beta}$ is defined in (2-11).

Our convention is that derivatives of $\vec{G}_{(\mathrm{frame})}$ form a new array consisting of the differentiated components. For example,

$$\not{\mathcal{L}}_L \vec{G}_{(\mathrm{frame})} \doteq \{L(\vec{G}_{LL}), L(\vec{G}_{LX}), \ldots, \not{\mathcal{L}}_L \not{\vec{G}}\},$$

[35] In (2-30a), we have corrected a sign error that occurred in [52, Definition 2.8].

where

$$L(\vec{G}_{LL}) \doteq \{L(G^1_{LL}), L(G^2_{LL}), \dots, L(G^5_{LL})\},$$
$$\not{\mathcal{L}}_L(\vec{\not{G}}_X) \doteq \{\not{\mathcal{L}}_L(\not{G}^1_X), \not{\mathcal{L}}_L(\not{G}^2_X), \dots, \not{\mathcal{L}}_L(\not{G}^5_X)\},$$

etc.

2K. *First and second fundamental forms and covariant differential operators.*

Definition 2.31 (first fundamental forms). Let $\underline{\Pi}$ and $\not{\Pi}$ be as in Definition 2.27. We define the first fundamental form $\underline{g}$ of Σ_t and the first fundamental form $\not{g}$ of $\ell_{t,u}$ as

$$\underline{g} \doteq \underline{\Pi} g, \quad \not{g} \doteq \not{\Pi} g. \tag{2-34}$$

We define the inverse first fundamental forms by raising the indices with g^{-1}:

$$(\underline{g}^{-1})^{\mu\nu} \doteq (g^{-1})^{\mu\alpha}(g^{-1})^{\nu\beta}\underline{g}_{\alpha\beta}, \quad (\not{g}^{-1})^{\mu\nu} \doteq (g^{-1})^{\mu\alpha}(g^{-1})^{\nu\beta}\not{g}_{\alpha\beta}, \tag{2-35}$$

where $\underline{g}$ is the Riemannian metric on Σ_t induced by g, while $\not{g}$ is the Riemannian metric on $\ell_{t,u}$ induced by g. Simple calculations imply that $(\underline{g}^{-1})^{\mu\alpha}\underline{g}_{\alpha\nu} = \underline{\Pi}_\nu^{\ \mu}$ and $(\not{g}^{-1})^{\mu\alpha}\not{g}_{\alpha\nu} = \not{\Pi}_\nu^{\ \mu}$.

Lemma 2.32 (identities for induced metrics). *In the (t, u, x^2, x^3)-coordinate system, we have*

$$\underline{g} = \frac{\mu^2 c^2}{(X^1)^2} du \otimes du - \mu \sum_{A=2}^{3} \frac{X^A}{(X^1)^2}(dx^A \otimes du + du \otimes dx^A) + \not{g}, \quad \not{g} = \sum_{A,B=2}^{3} c^{-2}\Big(\delta_{AB} + \frac{X^A X^B}{(X^1)^2}\Big) dx^A \otimes dx^B.$$

Moreover,

$$\not{g}^{-1} = \sum_{A,B=2}^{3} (c^2\delta^{AB} - X^A X^B)\not{\partial}_A \otimes \not{\partial}_B.$$

Proof. The identities for $\underline{g}$ and $\not{g}$ follow easily from Lemma 2.24 and the fact that $\underline{g}_{ij} = c^{-2}\delta_{ij}$ in Cartesian coordinates (see (2-7)). The identity for $\not{g}^{-1}$ follows from inverting the matrix $(\not{g}_{AB})_{A,B=2,3}$ and using the identity $\sum_{i=1}^{3}(X^i)^2 = c^2$, which follows from the first identity in (2-22) and (2-7). □

Definition 2.33 (differential operators associated to the metrics).

- $\mathscr{D}$ denotes the Levi-Civita connection of the acoustical metric g.
- $\not{\nabla}$ denotes the Levi-Civita connection of $\not{g}$.
- If f is a scalar function on $\ell_{t,u}$, then $\not{d} f \doteq \not{\nabla} f = \not{\Pi}\mathscr{D} f$, where $\mathscr{D} f$ is the gradient one-form associated to f.
- If ξ is an $\ell_{t,u}$-tangent one-form, then $\not{\mathrm{div}}\,\xi$ is the scalar function $\not{\mathrm{div}}\,\xi \doteq \not{g}^{-1} \cdot \not{\nabla}\xi$.
- Similarly, if V is an $\ell_{t,u}$-tangent vectorfield, then $\not{\mathrm{div}}\,V \doteq \not{g}^{-1} \cdot \not{\nabla} V_\flat$, where $V_\flat$ is the one-form $\not{g}$-dual to V.
- If ξ is a symmetric type-$\binom{0}{2}$ $\ell_{t,u}$-tangent tensorfield, then $\not{\mathrm{div}}\,\xi$ is the $\ell_{t,u}$-tangent one-form $\not{\mathrm{div}}\,\xi \doteq \not{g}^{-1} \cdot \not{\nabla}\xi$, where the two contraction indices in $\not{\nabla}\xi$ correspond to the operator $\not{\nabla}$ and the first index of ξ.
- $\not{\Delta} \doteq \not{g}^{-1} \cdot \not{\nabla}^2$ denotes the covariant Laplacian corresponding to $\not{g}$.

2L. ***Ricci coefficients.***

Definition 2.34 (Ricci coefficients).

(1) Define the *second fundamental form* k of Σ_t and the *null second fundamental form* χ of $\ell_{t,u}$ as

$$k \doteq \tfrac{1}{2}\underline{\mathcal{L}}_B \underline{g}, \quad \chi \doteq \tfrac{1}{2}\not{\mathcal{L}}_L \not{g}. \tag{2-36}$$

(2) Define ζ to be the $\ell_{t,u}$-tangent one-form whose components are given by

$$\zeta(\not{\partial}_A) \doteq g(\mathscr{D}_{\not{\partial}_A} L, X) = \mu^{-1} g(\mathscr{D}_{\not{\partial}_A} L, \breve{X}), \quad A = 2, 3. \tag{2-37}$$

(3) Given any symmetric type-$\binom{0}{2}$ $\ell_{t,u}$-tangent tensorfield ξ, define its *trace* by

$$\mathrm{tr}_{\not{g}}\, \xi \doteq (\not{g}^{-1})^{AB} \xi_{AB}.$$

Lemma 2.35 (useful identities for the Ricci coefficients). *The following identities hold*:[36]

$$\chi = g_{ab}(\not{d}L^a) \otimes (\not{d}x^b) + \tfrac{1}{2}\vec{\not{G}} \diamond L\vec{\Psi} + \tfrac{1}{2}\not{d}\vec{\Psi} \overset{\diamond}{\otimes} \vec{\not{G}}_L - \tfrac{1}{2}\vec{\not{G}}_L \overset{\diamond}{\otimes} \not{d}\vec{\Psi}, \tag{2-38a}$$

$$\mathrm{tr}_{\not{g}}\, \chi = g_{ab}\not{g}^{-1} \cdot \{(\not{d}L^a) \otimes (\not{d}x)^b\} + \tfrac{1}{2}\not{g}^{-1} \cdot \vec{\not{G}} \diamond L\vec{\Psi}, \tag{2-38b}$$

$$\not{k} = \tfrac{1}{2}\mu^{-1}\vec{\not{G}} \diamond \breve{X}\vec{\Psi} + \tfrac{1}{2}\vec{\not{G}} \diamond L\vec{\Psi} - \tfrac{1}{2}\vec{\not{G}}_L \overset{\diamond}{\otimes} \not{d}\vec{\Psi} - \tfrac{1}{2}\not{d}\vec{\Psi} \overset{\diamond}{\otimes} \vec{\not{G}}_L - \tfrac{1}{2}\vec{\not{G}}_X \overset{\diamond}{\otimes} \not{d}\vec{\Psi} - \tfrac{1}{2}\not{d}\vec{\Psi} \overset{\diamond}{\otimes} \vec{\not{G}}_X, \tag{2-38c}$$

$$\zeta = -\tfrac{1}{2}\mu^{-1}\vec{\not{G}}_L \diamond \breve{X}\vec{\Psi} + \tfrac{1}{2}\vec{\not{G}}_X \diamond L\vec{\Psi} - \tfrac{1}{2}\vec{G}_{LX} \diamond \not{d}\vec{\Psi} - \tfrac{1}{2}\vec{G}_{XX} \diamond \not{d}\vec{\Psi}. \tag{2-38d}$$

Proof. This is the same as [52, Lemmas 2.13, 2.15] except for small modifications incorporating the third dimension. □

2M. ***Pointwise norms.*** We always measure the magnitude of $\ell_{t,u}$ tensors[37] using $\not{g}$.

Definition 2.36 (pointwise norms). For any type-$\binom{m}{n}$ $\ell_{t,u}$ tensor $\xi^{\mu_1\cdots\mu_m}_{\nu_1\cdots\nu_n}$, we define

$$|\xi| \doteq \sqrt{\not{g}_{\mu_1\widetilde{\mu}_1} \cdots \not{g}_{\mu_m\widetilde{\mu}_m} (\not{g}^{-1})^{\nu_1\widetilde{\nu}_1} \cdots (\not{g}^{-1})^{\nu_n\widetilde{\nu}_n} \xi^{\mu_1\cdots\mu_m}_{\nu_1\cdots\nu_n} \xi^{\widetilde{\mu}_1\cdots\widetilde{\mu}_m}_{\widetilde{\nu}_1\cdots\widetilde{\nu}_n}}. \tag{2-39}$$

2N. ***Transport equations for the eikonal function quantities.*** The next lemma provides the transport equations that, in conjunction with (2-38b), we use to estimate the eikonal function quantities μ, $L^i_{(\mathrm{small})}$, and $\mathrm{tr}_{\not{g}}\, \chi$ below top order.

Lemma 2.37 ([52, Lemma 2.12] the transport equations satisfied by μ and $L^i_{(\mathrm{small})}$). *The following transport equations hold*:

$$L\mu = \tfrac{1}{2}\vec{G}_{LL} \diamond \breve{X}\vec{\Psi} - \tfrac{1}{2}\mu\vec{G}_{LL} \diamond L\vec{\Psi} - \mu\vec{G}_{LX} \diamond L\vec{\Psi}, \tag{2-40}$$

$$LL^i = \tfrac{1}{2}\vec{G}_{LL} \diamond (L\vec{\Psi})X^i - \vec{\not{G}}_L^{\#} \diamond (L\vec{\Psi}) \cdot \not{d}x^i + \tfrac{1}{2}\vec{G}_{LL} \diamond (\not{d}^{\#}\vec{\Psi}) \cdot \not{d}x^i. \tag{2-41}$$

[36] Here, $\vec{\not{G}}_L \overset{\diamond}{\otimes} \not{d}\vec{\Psi} \doteq \sum_{\iota=1}^{5} \not{G}_L^{\iota} \otimes \not{d}\Psi_\iota$, and similarly for the other terms involving $\overset{\diamond}{\otimes}$.

[37] Note that in contrast, for Σ_t tensors, we measure their magnitude using the Euclidean metric or an equivalent norm; see, for example, Definition 11.1.

2O. ***Calculations connected to the failure of the null condition.*** Many important estimates are tied to the coefficients $\vec{G}_{LL}$. In the next two lemmas, we derive expressions for $\vec{G}_{LL}$ and $\frac{1}{2}\vec{G}_{LL} \diamond \breve{X}\vec{\Psi}$. This presence of the latter term on the right-hand side (2-40) is tied to the failure of Klainerman's null condition [32] and thus one expects that the product must be nonzero for shocks to form; this is explained in more detail in the survey article [30] in a slightly different context.

Lemma 2.38 (formula for $\frac{1}{2}\vec{G}_{LL} \diamond \breve{X}\vec{\Psi}$). *Let F be the smooth function of* (ρ, s) *from* (2-2), *and let $F_{;s}$ denote its partial derivative with respect to s at fixed ρ. For solutions to* (1-1)–(1-3), *we have*

$$\frac{1}{2}\vec{G}_{LL} \diamond \breve{X}\vec{\Psi} = -\frac{1}{2}c^{-1}(c^{-1}c_{;\rho}+1)\{\breve{X}\mathcal{R}_{(+)} - \breve{X}\mathcal{R}_{(-)}\} - \frac{1}{2}\mu c^{-2}X^1\{L\mathcal{R}_{(+)} + L\mathcal{R}_{(-)}\} \\ - \mu c^{-2}(X^2 Lv^2 + X^3 Lv^3) - \mu c^{-1}c_{;s}X^a S^a + \mu c^{-1}(c^{-1}c_{;\rho}+1)F_{;s}X^a S^a. \quad (2\text{-}42)$$

Proof. This is the same as [36, Lemmas 2.45, 2.46], except for minor modifications incorporating the third dimension and the entropy (via the $c_{;s}$-dependent and $F_{;s}$-dependent products). □

3. Volume forms and energies

In this section, we first define geometric integration forms and corresponding integrals. We then define the energies and null fluxes which we will use in the remainder of the paper to derive a priori L^2-type estimates.

3A. ***Geometric forms and related integrals.*** We define our geometric integrals in terms of area and volume forms that remain nondegenerate relative to the geometric coordinates throughout the evolution (i.e., all the way up to the shock).

Definition 3.1 (geometric forms and related integrals). Define the area form $d\lambda_{g\!\!/}$ on $\ell_{t,u}$, the area form $d\underline{\varpi}$ on Σ_t^u, the area form $d\overline{\varpi}$ on $\mathcal{F}_u^t$, and the volume form $d\varpi$ on $\mathcal{M}_{t,u}$ as follows (relative to the (t, u, x^2, x^3)-coordinates):

$$d\lambda_{g\!\!/} = d\lambda_{g\!\!/}(t,u,x^2,x^3) \doteq \frac{dx^2\,dx^3}{c|X^1|}, \qquad d\underline{\varpi} = d\underline{\varpi}(t,u',x^2,x^3) \doteq d\lambda_{g\!\!/}(t,u',x^2,x^3)\,du',$$
$$d\overline{\varpi} = d\overline{\varpi}(t',u,x^2,x^3) \doteq d\lambda_{g\!\!/}(t',u,x^2,x^3)\,dt', \quad d\varpi = d\varpi(t',u',x^2,x^3) \doteq d\lambda_{g\!\!/}(t',u',x^2,x^3)\,du'\,dt'.$$

It is understood that unless we explicitly indicate otherwise, all integrals are defined with respect to the forms of Definition 3.1. Moreover, in our notation, we often suppress the variables with respect to which we integrate; i.e., we write $\int_{\ell_{t,u}} f\,d\lambda_{g\!\!/} \doteq \int_{(x^2,x^3)\in\mathbb{T}^2} f(t,u,x^2,x^3)\,d\lambda_{g\!\!/}(t,u,x^2,x^3)$, etc.

The following lemma clarifies the geometric and analytic significance of the forms from Definition 3.1.

Lemma 3.2 (identities concerning the forms).

(1) $d\lambda_{g\!\!/}$ *is the volume measure induced by* $g\!\!/$ *on* $\ell_{t,u}$.

(2) $\mu\,d\underline{\varpi}$ *is the volume measure induced by* $\underline{g}$ *on* Σ_t^u.

(3) *Let dx be the standard Euclidean volume measure on Σ_t^u, i.e., $dx = dx^1\,dx^2\,dx^3$ relative to the Cartesian spatial coordinates. Then*

$$dx = \mu c^3\,d\underline{\varpi}. \quad (3\text{-}1)$$

Proof. A computation based on Lemma 2.32 and the identity $\sum_{a=1}^{3}(X^a)^2 = c^2$ (which follows from (2-22) and (2-7)) yields that $\det \not{g} = 1/(c^2(X^1)^2)$. Since $d\lambda_{\not{g}} = \sqrt{\det \not{g}}\, dx^2\, dx^3$, we thus obtain (1).

Next, we again use Lemma 2.32 and the identity $\sum_{a=1}^{3}(X^a)^2 = c^2$ to compute that relative to the (u, x^2, x^3)-coordinates, we have $\det \underline{g} = \mu^2/(c^2(X^1)^2)$. Taking the square root, we see that the volume measure induced by $\underline{g}$ on Σ_t^u is given in the (u, x^2, x^3)-coordinates by $\mu/(c|X^1|)\, du\, dx^2\, dx^3$, which gives (2).

Finally, we obtain (3) from (2) via (2-7), which implies that relative to the Cartesian spatial coordinates, the canonical volume form induced by $\underline{g}$ on Σ_t is $c^{-3} dx^1\, dx^2\, dx^3$. □

3B. *The definitions of the energies and null fluxes.*

3B1. *Forms and conventions.*

Definition 3.3 (volume forms for L^p norms). For $p \in \{1, 2\}$, we define L^p norms with respect to the volume forms introduced in Definition 3.1. That is, for scalar functions or $\ell_{t,u}$-tangent tensorfields ξ, we define

$$\|\xi\|_{L^p(\ell_{t,u})} \doteq \left(\int_{\ell_{t,u}} |\xi|^p\, d\lambda_{\not{g}}\right)^{1/p}, \quad \|\xi\|_{L^p(\mathcal{F}_u^t)} \doteq \left(\int_{\mathcal{F}_u^t} |\xi|^p\, d\overline{\varpi}\right)^{1/p},$$
$$\|\xi\|_{L^p(\Sigma_t^u)} \doteq \left(\int_{\Sigma_t^u} |\xi|^p\, d\underline{\varpi}\right)^{1/p}, \quad \|\xi\|_{L^p(\Sigma_t)} \doteq \left(\int_{\Sigma_t} |\xi|^p\, d\underline{\varpi}\right)^{1/p},$$
$$\|\xi\|_{L^p(\mathcal{M}_{t,u})} \doteq \left(\int_{\mathcal{M}_{t,u}} |\xi|^p\, d\varpi\right)^{1/p}.$$

Definition 3.4 (conventions with variable arrays and differentiated quantities).

(1) Given the fluid variable array $\vec{\Psi}$ in Definition 2.3, define

$$|\vec{\Psi}| = |\Psi| \doteq \max_{\iota=1,\dots,5} |\Psi_\iota|.$$

We also set

$$|\varOmega| \doteq \max_{a=1,2,3} |\varOmega^a|,$$

and similarly for the other Σ_t-tangent tensorfields such as S and $\mathcal{C}$ that correspond to the transport variables. For $p = 2$ or $p = \infty$, define also

$$\|\Psi\|_{L^p(\ell_{t,u})} \doteq \max_{\iota=1,\dots,5} \|\Psi_\iota\|_{L^p(\ell_{t,u})},$$

and similarly for $L^p(\Sigma_t^u)$, $L^p(\Sigma_t)$, $L^p(\mathcal{F}_u^t)$, and $L^p(\mathcal{M}_{t,u})$. Similarly, we set

$$\|\varOmega\|_{L^p(\ell_{t,u})} \doteq \max_{a=1,2,3} \|\varOmega^a\|_{L^p(\ell_{t,u})},$$

and we analogously define L^p norms of other Σ_t-tangent tensorfields that correspond to the transport variables, such as S and $\mathcal{C}$.

(2) When estimating multiple solution variables simultaneously, we use the following convention (for $p = 2$ or $p = \infty$):

$$\|(\varOmega, S)\|_{L^p(\ell_{t,u})} \doteq \max\{\|\varOmega\|_{L^p(\ell_{t,u})},\ \|S\|_{L^p(\ell_{t,u})}\},$$

and similarly for $L^p(\Sigma_t^u)$, $L^p(\Sigma_t)$, $L^p(\mathcal{F}_u^t)$, and $L^p(\mathcal{M}_{t,u})$.

(3) Let $\mathscr{P} \doteq \{L, Y, Z\}$ be the set of commutation vectorfields and

$$\mathscr{P}^{(N)} \doteq \{\mathcal{P}_1\mathcal{P}_2 \cdots \mathcal{P}_N \mid \mathcal{P}_i \in \mathscr{P} \text{ for } 1 \le i \le N\}.$$

For any smooth scalar function ϕ, define

$$|\mathcal{P}^N\phi| \doteq \max_{\mathcal{P}_1,\dots,\mathcal{P}_N \in \mathscr{P}^{(N)}} |\mathcal{P}_1 \cdots \mathcal{P}_N \phi|.$$

For $p = 2$ or $p = \infty$, the L^p norms are defined similarly, with

$$\|\mathcal{P}^N\phi\|_{L^p(\ell_{t,u})} \doteq \||\mathcal{P}^N\phi|\|_{L^p(\ell_{t,u})}, \quad \text{etc.}$$

Moreover, we let $\mathcal{P}^N\Omega$ denote the Σ_t-tangent vectorfield with Cartesian spatial components $\mathcal{P}^N\Omega^i$, and we define

$$|\mathcal{P}^N\Omega| \doteq \max_{\mathcal{P}_1\cdots\mathcal{P}_N \in \mathscr{P}^{(N)}} \max_{a=1,2,3} |\mathcal{P}_1 \cdots \mathcal{P}_N \Omega^a|,$$
$$\|\mathcal{P}^N\Omega\|_{L^p(\ell_{t,u})} \doteq \||\mathcal{P}^N\Omega|\|_{L^p(\ell_{t,u})}, \quad \text{etc.,}$$

and similarly for other Σ_t-tangent tensorfields that correspond to the transport variables, such as S and $\mathcal{C}$.

(4) Similarly, we let $\not{\mathscr{P}} \doteq \{Y, Z\}$ be the set of $\ell_{t,u}$-tangent commutation vectorfields and define

$$\not{\mathscr{P}}^{(N)} \doteq \{\not{\mathcal{P}}_1\not{\mathcal{P}}_2 \cdots \not{\mathcal{P}}_N \mid \not{\mathcal{P}}_i \in \not{\mathscr{P}} \text{ for } 1 \le i \le N\},$$
$$|\not{\mathcal{P}}^N\phi| \doteq \max_{\not{\mathcal{P}}_1,\dots,\not{\mathcal{P}}_N \in \not{\mathscr{P}}^{(N)}} |\not{\mathcal{P}}_1 \cdots \not{\mathcal{P}}_N \phi|, \quad \text{etc.}$$

The importance of distinguishing the subset of $\ell_{t,u}$-tangent commutation vectorfields from the full set $\mathscr{P}$ will be made clear in the Appendix.[38]

(5) We use the following conventions for sums:

$$|\mathcal{P}^{[N_1,N_2]}\phi| \doteq \sum_{N'=N_1}^{N_2} |\mathcal{P}^{N'}\phi|, \quad |\mathcal{P}^{\le N}\phi| \doteq |\mathcal{P}^{[0,N]}\phi|,$$
$$|\not{\mathcal{P}}^{[N_1,N_2]}\phi| \doteq \sum_{N'=N_1}^{N_2} |\not{\mathcal{P}}^{N'}\phi|, \quad |\not{\mathcal{P}}^{\le N}\phi| \doteq |\not{\mathcal{P}}^{[0,N]}\phi|.$$

(6) We will combine the above conventions. For instance,

$$|\mathcal{P}^N(\Omega, S)| \doteq \max_{\mathcal{P}_1\cdots\mathcal{P}_N \in \mathscr{P}^{(N)}} \max\{|\mathcal{P}_1 \cdots \mathcal{P}_N \Omega|,\ |\mathcal{P}_1 \cdots \mathcal{P}_N S|\},$$
$$|\mathcal{P}^N\Psi| \doteq \max_{\iota=1,\dots,5} \max_{\mathcal{P}_1\cdots\mathcal{P}_N \in \mathscr{P}^{(N)}} |\mathcal{P}_1 \cdots \mathcal{P}_N \Psi_\iota|.$$

3B2. *Definitions of the energies.* We are now ready to introduce the main energies we use to control the solution.

[38]As in the two-dimensional case, the most difficult error terms in the wave equation energy estimates are commutator terms involving the pure $\ell_{t,u}$-tangent derivatives of the null mean curvature of the $\ell_{t,u}$.

Wave energies:

$$\mathbb{E}_N(t,u) \doteq \sup_{t'\in[0,t]} (\|\breve{X}\mathcal{P}^N\Psi\|^2_{L^2(\Sigma^u_{t'})} + \|\sqrt{\upmu}\mathcal{P}^{N+1}\Psi\|^2_{L^2(\Sigma^u_{t'})}), \tag{3-2a}$$

$$\mathbb{F}_N(t,u) \doteq \sup_{u'\in[0,u]} (\|L\mathcal{P}^N\Psi\|^2_{L^2(\mathcal{F}^t_{u'})} + \|\sqrt{\upmu}\not{d}\mathcal{P}^N\Psi\|^2_{L^2(\mathcal{F}^t_{u'})}), \tag{3-2b}$$

$$\mathbb{K}_N(t,u) \doteq \|\not{d}\mathcal{P}^N\Psi\|^2_{L^2(\mathcal{M}_{t,u}\cap\{\upmu\le 1/4\})}, \tag{3-2c}$$

$$\mathbb{Q}_N(t,u) \doteq \mathbb{E}_N(t,u) + \mathbb{F}_N(t,u), \tag{3-2d}$$

$$\mathbb{W}_N(t,u) \doteq \mathbb{E}_N(t,u) + \mathbb{F}_N(t,u) + \mathbb{K}_N(t,u). \tag{3-2e}$$

Specific vorticity energies:

$$\mathbb{V}_N(t,u) \doteq \sup_{t'\in[0,t]} \|\sqrt{\upmu}\mathcal{P}^N\varOmega\|^2_{L^2(\Sigma^u_{t'})} + \sup_{u'\in[0,u]} \|\mathcal{P}^N\varOmega\|^2_{L^2(\mathcal{F}^t_{u'})}, \tag{3-3a}$$

$$\mathbb{C}_N(t,u) \doteq \sup_{t'\in[0,t]} \|\sqrt{\upmu}\mathcal{P}^N\mathcal{C}\|^2_{L^2(\Sigma^u_{t'})} + \sup_{u'\in[0,u]} \|\mathcal{P}^N\mathcal{C}\|^2_{L^2(\mathcal{F}^t_{u'})}. \tag{3-3b}$$

Entropy gradient energies:

$$\mathbb{S}_N(t,u) \doteq \sup_{t'\in[0,t]} \|\sqrt{\upmu}\mathcal{P}^N S\|^2_{L^2(\Sigma^u_{t'})} + \sup_{u'\in[0,u]} \|\mathcal{P}^N S\|^2_{L^2(\mathcal{F}^t_{u'})}, \tag{3-4a}$$

$$\mathbb{D}_N(t,u) \doteq \sup_{t'\in[0,t]} \|\sqrt{\upmu}\mathcal{P}^N\mathcal{D}\|^2_{L^2(\Sigma^u_{t'})} + \sup_{u'\in[0,u]} \|\mathcal{P}^N\mathcal{D}\|^2_{L^2(\mathcal{F}^t_{u'})}. \tag{3-4b}$$

Definition 3.5 (important conventions for energies).

(1) We define the following convention for sums (cf. Definition 3.4(3)):

$$\mathbb{E}_{\le N}(t,u) \doteq \sum_{N'=0}^{N} \mathbb{E}_{N'}(t,u), \qquad \mathbb{E}_{[1,N]}(t,u) \doteq \sum_{N'=1}^{N} \mathbb{E}_{N'}(t,u),$$

and similarly for other energies.

(2) Abusing notation slightly, if we write an energy as a function of only t (instead of a function of (t,u)), then it is understood that we take supremum in u, e.g.,

$$\mathbb{E}_N(t) \doteq \sup_{u\in\mathbb{R}} \mathbb{E}_N(t,u).$$

4. Assumptions on the data and statement of the main theorems

4A. ***Assumptions on the data of the fluid variables.*** We now introduce the assumptions on the data for our main theorem. We have five parameters (see Theorem 1.1), denoted by $\mathring{\upsigma}$, $\mathring{\updelta}_*$, $\mathring{\updelta}$, $\mathring{\upalpha}$ and $\mathring{\upepsilon}$:

- $\mathring{\upsigma}$ measures the size of the initial support in x^1.
- $\mathring{\updelta}_*$ gives a lower bound on the quantity that controls the blowup, and in particular determines the time interval for which we need to control our solution before a singularity forms.

- $\mathring{\delta}$, $\mathring{\alpha}$ and $\mathring{\epsilon}$ are parameters that control the sizes of various norms of the solution. The parameter $\mathring{\delta}$ measures the L^∞ size of the transversal derivatives of $\mathcal{R}_{(+)}$, and it *can be large*, while $\mathring{\alpha}$ limits the size of the amplitude of the fluid variables, is *small* depending on the equation of state and the background density $\bar{\varrho} > 0$, and is used to control basic features of the Lorentzian geometry. The parameter $\mathring{\epsilon}$ is *small* depending on the equation of state and all the other parameters. In particular, $\mathring{\epsilon}$ controls the size of solution in "directions that break the simple plane symmetry," and it provides the most crucial smallness that we exploit in the analysis.
- We assume that $\mathring{\epsilon}^{1/2} \leq \mathring{\alpha}$.

Here are the assumptions on the initial data.[39]

In what follows, N_{top} and M_* denote large positive integers that are constrained in particular by $N_{\mathrm{top}} \geq 2M_* + 10$. In our proof of Proposition 12.1, we will show that our estimates close with M_* chosen to be a universal positive integer. The restriction $N_{\mathrm{top}} \geq 2M_* + 10$ is further explained in Remark 6.1. See also the discussion in Section 2A.

Compact support assumptions:

$$\text{If } |x^1| \geq \mathring{\sigma}, \text{ then } (\rho, v, s) = (0, 0, 0). \tag{4-1}$$

By (2-13), when $t = 0$, the data are supported on the set where $u \in [0, 2\mathring{\sigma}]$. This explains why in some of the data assumptions stated below, we only consider regions in which $u \in [0, 2\mathring{\sigma}]$.

Lower bound for the quantity that controls the blowup-time:[40]

$$\mathring{\delta}_* \doteq \sup_{\Sigma_0} \tfrac{1}{2}[c^{-1}(c^{-1}c_{;\rho} + 1)(\breve{X}\mathcal{R}_{(+)})]_+ > 0. \tag{4-2}$$

Remark 4.1 (nondegeneracy assumption on the factor $c^{-1}(c^{-1}c_{;\rho}+1)$). Recall the factor $c^{-1}(c^{-1}c_{;\rho}+1)$ in (4-2) can be viewed as a function of (ρ, s). For the solutions under study, we are assuming that $c^{-1}(c^{-1}c_{;\rho} + 1)$ is nonvanishing when evaluated at the trivial background solution $(\rho, s) \equiv (0, 0)$ (recall that this background corresponds to a state with constant density $\varrho \equiv \bar{\varrho} > 0$). One can check that for any smooth equation of state except that of a Chaplygin gas, there are always open sets of $\bar{\varrho} > 0$ such that the nonvanishing condition holds; see the end of [36, Section 2.16] for further discussion. We also point out that for the Chaplygin gas, it is not expected that shocks will form.

Assumptions on the amplitude and transversal derivatives of the wave variables:

$$\|\mathcal{R}_{(+)}\|_{L^\infty(\Sigma_0)} \leq \mathring{\alpha}, \tag{4-3a}$$

$$\|\breve{X}^{[1,3]}\mathcal{R}_{(+)}\|_{L^\infty(\Sigma_0)} \leq \mathring{\delta}, \tag{4-3b}$$

$$\|\breve{X}^{\leq 3}(\mathcal{R}_{(-)}, v^2, v^3, s)\|_{L^\infty(\Sigma_0)} \leq \mathring{\epsilon}, \tag{4-3c}$$

$$\|L\breve{X}\breve{X}\breve{X}\Psi\|_{L^\infty(\Sigma_0)} \leq \mathring{\epsilon}. \tag{4-3d}$$

[39] Of course, we are only allowed to prescribe (ϱ, v^i, s) without explicitly specifying their derivatives transversal to Σ_0. Nevertheless, using (1-1)–(1-3), we can compute the traces of all derivatives on Σ_0. The derivative assumptions that we specify here are to be understood in this sense. Notice that all the assumptions are satisfied by the data of exactly simple plane-symmetric solutions with $\mathring{\epsilon} = 0$. Thus, they are also satisfied by small perturbations of them.

[40] Here, $z_+ \doteq \max\{z, 0\}$.

Smallness assumptions for good derivatives of the wave variables:

$$\begin{aligned}&\|\mathcal{P}^{[1,N_{\text{top}}-M_*-2]}\Psi\|_{L^\infty(\Sigma_0)}, \quad \|\mathcal{P}^{[1,N_{\text{top}}-M_*-4]}\breve{X}\Psi\|_{L^\infty(\Sigma_0)}, \quad \|\mathcal{P}^{[1,2]}\breve{X}\breve{X}\Psi\|_{L^\infty(\Sigma_0)},\\ &\qquad \sup_{u\in[0,2\mathring{\sigma}]}\|\mathcal{P}^{[1,N_{\text{top}}-M_*]}\Psi\|_{L^2(\ell_{0,u})}, \quad \|\mathcal{P}^{[1,N_{\text{top}}+1]}\Psi\|_{L^2(\Sigma_0)}, \quad \|\mathcal{P}^{[1,N_{\text{top}}]}\breve{X}\Psi\|_{L^2(\Sigma_0)} \leq \mathring{\epsilon}. \end{aligned} \tag{4-4}$$

Smallness assumptions for the specific vorticity and entropy gradient:

$$\|\mathcal{P}^{\leq N_{\text{top}}-M_*-2}(\varOmega, S)\|_{L^\infty(\Sigma_0)}, \quad \sup_{u\in[0,2\mathring{\sigma}]}\|\mathcal{P}^{\leq N_{\text{top}}-M_*}(\varOmega, S)\|_{L^2(\ell_{0,u})}, \quad \|\mathcal{P}^{\leq N_{\text{top}}}(\varOmega, S)\|_{L^2(\Sigma_0)} \leq \mathring{\epsilon}^{3/2}. \tag{4-5}$$

Smallness assumptions for the modified fluid variables:

$$\|\mathcal{P}^{\leq N_{\text{top}}-M_*-3}(\mathcal{C}, \mathcal{D})\|_{L^\infty(\Sigma_0)}, \quad \sup_{u\in[0,2\mathring{\sigma}]}\|\mathcal{P}^{\leq N_{\text{top}}-M_*-1}(\mathcal{C}, \mathcal{D})\|_{L^2(\ell_{0,u})}, \quad \|\mathcal{P}^{\leq N_{\text{top}}}(\mathcal{C}, \mathcal{D})\|_{L^2(\Sigma_0)} \leq \mathring{\epsilon}^{3/2}. \tag{4-6}$$

4B. ***Statement of the main theorem.*** We are now ready to give a precise statement of Theorem 1.1 (see Theorems 4.2 and 4.3 below), as well as the corollaries in interesting subregimes of solutions discussed in Remark 1.5 (see Corollaries 4.4 and 4.5).

We first discuss Theorem 1.1. It will be convenient to think of Theorem 1.1 as two theorems. The first, which is the much harder theorem, is a *regularity* statement, stating — with precise estimates — that in the region under study, the only possible singularity is that of a shock, i.e., one that is associated with the vanishing of μ. This is the content of Theorem 4.2. Once Theorem 4.2 has been proved, the proof that a shock indeed occurs is much easier. This is the content of Theorem 4.3.

Theorem 4.2 (regularity unless shock occurs). *Let $\mathring{\sigma}$, $\mathring{\delta}$, $\mathring{\delta}_* > 0$. There exists a large integer M_* that is absolute in the sense that it is independent of the equation of state, $\bar{\varrho}$, $\mathring{\sigma}$, $\mathring{\delta}$, and $\mathring{\delta}_*^{-1}$ such that the following hold. Assume that*:

- *The integer N_{top} satisfies $N_{\text{top}} \geq 2M_* + 10$ (see Remark 6.1 regarding the size of N_{top}).*
- *$\mathring{\alpha} > 0$ is sufficiently small in a manner that depends only on the equation of state and $\bar{\varrho}$.*
- *$\mathring{\epsilon} > 0$ satisfies*[41] *$\mathring{\epsilon}^{1/2} \leq \mathring{\alpha}$ and is sufficiently small in a manner that depends only on the equation of state, N_{top}, $\bar{\varrho}$, $\mathring{\sigma}$, $\mathring{\delta}$, and $\mathring{\delta}_*^{-1}$.*
- *The initial data satisfy the support-size and norm-size assumptions*[42] (4-1)–(4-6).

Then the corresponding solution $(\varrho, v^1, v^2, v^3, s)$ to the compressible Euler equations (1-1)–(1-3) *exhibits the following properties.*

Suppose $T \in (0, 2\mathring{\delta}_*^{-1}]$, and assume that there is a smooth solution such that the following two conditions hold:

- *The change of variables map $(t, u, x^2, x^3) \to (t, x^1, x^2, x^3)$ from geometric to Cartesian coordinates is a diffeomorphism from $[0, T) \times \mathbb{R} \times \mathbb{T}^2$ onto $[0, T) \times \Sigma$.*
- *$\mu > 0$ in $[0, T) \times \Sigma$.*

[41]The assumption $\mathring{\epsilon}^{1/2} \leq \mathring{\alpha}$ allows us to simplify the presentation of various estimates, for example, by allowing us to write $\mathcal{O}(\mathring{\alpha})$ instead of $\mathcal{O}(\mathring{\epsilon}^{1/2}) + \mathcal{O}(\mathring{\alpha})$.

[42]Note that our plane-symmetric background solutions satisfy these assumptions with $\mathring{\epsilon} = 0$.

Then the following estimates hold for every $t \in [0, T)$, where the implicit constants in $\lesssim_\blacklozenge$ depend only on the equation of state and $\bar{\varrho}$, while the implicit constants in $\lesssim$ depend only on the equation of state, N_{top}, $\bar{\varrho}$ $\mathring{\sigma}$, $\mathring{\delta}$, and $\mathring{\delta}_*^{-1}$ (in particular, all implicit constants are independent of t and T).

(1) *The following energy estimates hold (where the energies are defined in (3-2a)–(3-4b) and $\mu_\star(t)$ is as in Definition 2.16):*

$$\mathbb{W}_N(t) \lesssim \mathring{\epsilon}^2 \max\{1, \mu_\star^{-2M_*+2N_{\mathrm{top}}-2N+1.8}(t)\} \quad \text{for } 1 \le N \le N_{\mathrm{top}}, \tag{4-7a}$$

$$\mathbb{V}_N(t),\ \mathbb{S}_N(t) \lesssim \mathring{\epsilon}^3 \max\{1, \mu_\star^{-2M_*+2N_{\mathrm{top}}-2N+2.8}(t)\} \quad \text{for } 0 \le N \le N_{\mathrm{top}}, \tag{4-7b}$$

$$\mathbb{C}_N(t),\ \mathbb{D}_N(t) \lesssim \mathring{\epsilon}^3 \max\{1, \mu_\star^{-2M_*+2N_{\mathrm{top}}-2N+0.8}(t)\} \quad \text{for } 0 \le N \le N_{\mathrm{top}}. \tag{4-7c}$$

(2) *The following L^∞ estimates hold:*

$$\|\mathcal{P}^{[1,N_{\mathrm{top}}-M_*-2]}\Psi\|_{L^\infty(\Sigma_t)},\quad \|\mathcal{P}^{[1,N_{\mathrm{top}}-M_*-4]}\breve{X}\Psi\|_{L^\infty(\Sigma_t)} \lesssim \mathring{\epsilon}, \tag{4-8a}$$

$$\|\mathcal{R}_{(+)}\|_{L^\infty(\Sigma_t)} \lesssim_\blacklozenge \mathring{\alpha},\quad \|(\mathcal{R}_{(-)}, v^2, v^3, s)\|_{L^\infty(\Sigma_t)} \lesssim \mathring{\epsilon}, \tag{4-8b}$$

$$\|\breve{X}\mathcal{R}_{(+)}\|_{L^\infty(\Sigma_t)} \le 2\mathring{\delta},\quad \|\breve{X}(\mathcal{R}_{(-)}, v^2, v^3, s)\|_{L^\infty(\Sigma_t)} \lesssim \mathring{\epsilon}, \tag{4-8c}$$

$$\begin{aligned}\|\mathcal{P}^{[1,N_{\mathrm{top}}-M_*-2]}(\varOmega, S)\|_{L^\infty(\Sigma_t)},\quad \|\mathcal{P}^{[1,N_{\mathrm{top}}-M_*-3]}(\mathcal{C},\mathcal{D})\|_{L^\infty(\Sigma_t)},&\\ \|\mathcal{P}^{[1,N_{\mathrm{top}}-M_*-4]}\breve{X}(\varOmega, S)\|_{L^\infty(\Sigma_t)} \lesssim \mathring{\epsilon}^{3/2}.&\end{aligned} \tag{4-8d}$$

In addition, the solution can be smoothly extended to $[0, T] \times \mathbb{R} \times \mathbb{T}^2$ as a function of the geometric coordinates (t, u, x^2, x^3).

Finally, if $\inf_{t\in[0,T)} \mu_\star(t) > 0$, then the solution can be smoothly extended to a Cartesian slab $[0, T+\epsilon] \times \Sigma$ for some $\epsilon > 0$ such that the map $(t, u, x^2, x^3) \to (t, x^1, x^2, x^3)$ is a diffeomorphism from $[0, T+\epsilon] \times \mathbb{R} \times \mathbb{T}^2$ onto $[0, T+\epsilon] \times \Sigma$. In particular, on the extended region, the solution is a smooth function of the geometric coordinates and the Cartesian coordinates.

Theorem 4.3 (complete description of the shock formation at the first singular time). *Under the assumptions of Theorem 4.2 — perhaps taking $\mathring{\alpha}$ and $\mathring{\epsilon}$ smaller in a manner that depends on the same quantities stated in the theorem — there exists $T_{(\mathrm{sing})} \in [0, 2\mathring{\delta}_*^{-1}]$ satisfying the estimate*[43]

$$T_{(\mathrm{sing})} = \{1 + \mathcal{O}_\blacklozenge(\mathring{\alpha}) + \mathcal{O}(\mathring{\epsilon})\}\mathring{\delta}_*^{-1} \tag{4-9}$$

such that the following hold:

(1) *The solution variables are smooth functions of the Cartesian coordinates (t, x^1, x^2, x^3) in $[0, T_{(\mathrm{sing})}) \times \Sigma$.*

(2) *The solution variables extend as smooth functions of the geometric coordinates (t, u, x^2, x^3) to $[0, T_{(\mathrm{sing})}] \times \mathbb{R} \times \mathbb{T}^2$.*

(3) *The inverse foliation density tends to zero at $T_{(\mathrm{sing})}$, i.e., $\liminf_{t\uparrow T^-_{(\mathrm{sing})}} \mu_\star(t) = 0$.*

(4) *$\partial_1\mathcal{R}_{(+)}$ blows up as $t \uparrow T^-_{(\mathrm{sing})}$, i.e., $\limsup_{t\uparrow T^-_{(\mathrm{sing})}} \sup_{\Sigma_t} |\partial_1\mathcal{R}_{(+)}| = \infty$.*

[43] See Section 2A regarding our use of the notation $\mathcal{O}_\blacklozenge(\,\cdot\,)$, $\mathcal{O}(\,\cdot\,)$, etc.

(5) *Moreover, let*[44]

$$\mathscr{S}_{\text{blowup}} \doteq \Big\{(u, x^2, x^3) \in \mathbb{R} \times \mathbb{T}^2 : \limsup_{(\tilde{t},\tilde{u},\tilde{x}^2,\tilde{x}^3)\to(T_{(\text{sing})}^-,u,x^2,x^3)} |\partial_1 \mathcal{R}_{(+)}|(\tilde{t}, \tilde{u}, \tilde{x}^2, \tilde{x}^3) = \infty\Big\},$$

$$\mathscr{S}_{\text{vanish}} \doteq \{(u, x^2, x^3) \in \mathbb{R} \times \mathbb{T}^2 : \mu(T_{(\text{sing})}, u, x^2, x^3) = 0\},$$

and

$$\mathscr{S}_{\text{regular}} \doteq \Big\{(u, x^2, x^3) \in \mathbb{R} \times \mathbb{T}^2 : \textit{all solution variables extend to be } C^1 \textit{ functions of the geometric and Cartesian coordinates in a neighborhood of the point with geometric coordinates } (T_{(\text{sing})}, u, x^2, x^3), \textit{ intersected with the half-space} \{t \le T_{(\text{sing})}\}\Big\}.$$

Then $\mathscr{S}_{\text{blowup}}$ *and* $\mathscr{S}_{\text{vanish}}$ *are nonempty, and*

$$\mathscr{S}_{\text{blowup}} = \mathscr{S}_{\text{vanish}} = \mathbb{R} \times \mathbb{T}^2 \setminus \mathscr{S}_{\text{regular}}.$$

The proofs of both Theorems 4.2 and 4.3 are located in Section 14B.

The next two corollaries concern some refined conclusions one can make with *additional* assumptions on the initial data.

Corollary 4.4 (nonvanishing of the vorticity and entropy at the blowup-points). *Assume the hypotheses and conclusions of Theorem 4.2, but perhaps taking* $\mathring{\alpha}$ *and* $\mathring{\epsilon}$ *smaller in a manner that depends on the same quantities stated in the theorem. Assume in addition that,*[45] *for all* $(x^2, x^3) \in \mathbb{T}^2$,

$$\tfrac{1}{2}[c^{-1}(c^{-1}c_{;\rho} + 1)(\breve{X}\mathcal{R}_{(+)})]_+(t = 0, u, x^2, x^3) \le \tfrac{1}{2}\mathring{\delta}_*^{-1} \quad \textit{when } |u - \mathring{\sigma} + \mathring{\delta}_*^{-1}| \ge 3\mathring{\alpha}\mathring{\delta}_*^{-1}, \tag{4-10}$$

and

$$\tfrac{1}{2}\mathring{\epsilon}^2 \le |\Omega(t = 0, u, x^2, x^3)| \le \mathring{\epsilon}^2, \quad \tfrac{1}{2}\mathring{\epsilon}^3 \le |S(t = 0, u, x^2, x^3)| \le \mathring{\epsilon}^3 \quad \textit{when } |u - \mathring{\sigma}| \le \mathring{\alpha}^{1/2}. \tag{4-11}$$

Then Ω *and* S *are nonvanishing near the singular set; i.e., for any* $(u, x^2, x^3) \in \mathscr{S}_{\text{blowup}}$ *(as in Theorem 4.3), we have* $\Omega(T_{(\text{sing})}, u, x^2, x^3) \ne 0$ *and* $S(T_{(\text{sing})}, u, x^2, x^3) \ne 0$.

The proof of Corollary 4.4 is located in Section 14C.

Corollary 4.5 (the spatial Hölder regularity of the solution relative to the Cartesian coordinates). *Let* $\mathring{\beta} > 0$ *be a constant, and assume that the following hold*:

(1) *For all u such that* $|u - \mathring{\sigma}| \ge \mathring{\sigma}/4$ *and all* $(x^2, x^3) \in \mathbb{T}^2$, *we have*

$$\tfrac{1}{2}[c^{-1}(c^{-1}c_{;\rho} + 1)(\breve{X}\mathcal{R}_{(+)})]_+(t = 0, u, x^2, x^3) \le \tfrac{1}{4}\mathring{\delta}_*.$$

(2) *For all*[46] $u \in [\mathring{\sigma}/2, 3\mathring{\sigma}/2]$ *and all* $(x^2, x^3) \in \mathbb{T}^2$,

$$\tfrac{1}{2}\breve{X}\breve{X}\{(c^{-1}c_{;\rho} + 1)(\breve{X}\mathcal{R}_{(+)})\}(t = 0, u, x^2, x^3) \le -3\mathring{\delta}_*\mathring{\beta} < 0. \tag{4-12}$$

[44] For definiteness, in the definition of the subset $\mathscr{S}_{\text{regular}}$, we have made statements only about the boundedness of the solution's C^1 norm. However, our proof shows that on $\mathscr{S}_{\text{regular}}$, the solution inherits the full regularity enjoyed by the initial data.

[45] Recall the initial condition (2-13) for u, which shows that $u \restriction_{\Sigma_0} = \mathring{\sigma} - x^1$.

[46] This is a nondegeneracy condition in the sense that it guarantees that for every $(x^2, x^3) \in \mathbb{T}^2$, the quantity $(c^{-1}c_{;\rho} + 1)(\breve{X}\mathcal{R}_{(+)}) \restriction_{\Sigma_0}$, when viewed as a one-variable function of u, has a nondegenerate maximum. (Note also that $(c^{-1}c_{;\rho} + 1)(\breve{X}\mathcal{R}_{(+)}) \restriction_{\Sigma_0}$ is related to the quantity in (4-2), whose reciprocal controls the blowup-time.)

Also assume the hypotheses and conclusions of Theorems 4.2 and 4.3, but perhaps taking $\mathring{\alpha}$ *and* $\mathring{\varepsilon}$ *smaller in a manner that depends on* $\mathring{\beta}$ *and the same quantities stated in Theorem 4.2. Then the spatial* $C^{1/3}$ *norms (i.e., the standard* $C^{1/3}$ *Hölder norms with respect to the Cartesian spatial coordinates) of all of the fluid variables and higher-order variables* ρ, v^i, Ω^i, S^i, $\mathcal{C}^i$ *and* $\mathcal{D}$ *are uniformly bounded up to the first singular time.*

The proof of Corollary 4.5 is located in Section 14D.

5. Reformulation of the equations and the remarkable null structure

We recall in this section the main result in [50], which is of crucial importance for our analysis.

Theorem 5.1 (the geometric wave-transport-divergence-curl formulation of the compressible Euler equations). *Consider a smooth solution to the compressible Euler equations* (1-1)–(1-3) *under an equation of state* $p = p(\varrho, s)$ *and constant* $\bar{\varrho} > 0$ *such that the normalization condition* (2-1) *holds. Then the scalar-valued functions* v^i, $\mathcal{R}_{(\pm)}$, Ω^i, s, S^i, $\operatorname{div}\Omega$, $\mathcal{C}^i$, $\mathcal{D}$, *and* $(\operatorname{curl} S)^i$, $i = 1, 2, 3$, *(see Definitions 2.3 and 2.7) obey the following system of equations (where the Cartesian component functions* v^i *are treated as scalar-valued functions under covariant differentiation on the left-hand side of* (5-1a)):

<u>Covariant wave equations</u>:

$$\square_g v^i = -c^2\exp(2\rho)\mathcal{C}^i + \mathfrak{Q}^i_{(v)} + \mathfrak{L}^i_{(v)}, \tag{5-1a}$$

$$\square_g \mathcal{R}_{(\pm)} = -c^2\exp(2\rho)\mathcal{C}^1 \pm \left\{F_{;s}c^2\exp(2\rho) - c\exp(\rho)\frac{p_{;s}}{\bar{\varrho}}\right\}\mathcal{D} + \mathfrak{Q}_{(\pm)} + \mathfrak{L}_{(\pm)}, \tag{5-1b}$$

$$\square_g s = c^2\exp(2\rho)\mathcal{D} + \mathfrak{L}_{(s)}. \tag{5-1c}$$

<u>Transport equations</u>:

$$B\Omega^i = \mathfrak{L}^i_{(\Omega)}, \tag{5-2a}$$

$$Bs = 0, \tag{5-2b}$$

$$BS^i = \mathfrak{L}^i_{(S)}. \tag{5-2c}$$

<u>Transport-divergence-curl system for the specific vorticity</u>:

$$\operatorname{div}\Omega = \mathfrak{L}_{(\operatorname{div}\Omega)}, \tag{5-3a}$$

$$B\mathcal{C}^i = \mathfrak{M}^i_{(\mathcal{C})} + \mathfrak{Q}^i_{(\mathcal{C})} + \mathfrak{L}^i_{(\mathcal{C})}. \tag{5-3b}$$

<u>Transport-divergence-curl system for the entropy gradient</u>:

$$B\mathcal{D} = \mathfrak{M}_{(\mathcal{D})} + \mathfrak{Q}_{(\mathcal{D})}, \tag{5-4a}$$

$$(\operatorname{curl} S)^i = 0. \tag{5-4b}$$

Above, the main terms in the transport equations for the modified fluid variables take the form

$$\begin{aligned}\mathfrak{M}^i_{(\mathcal{C})} \doteq{}& -2\delta_{jk}\epsilon_{iab}\exp(-\rho)(\partial_a v^j)\partial_b\Omega^k + \epsilon_{ajk}\exp(-\rho)(\partial_a v^i)\partial_j\Omega^k \\ &+ \exp(-3\rho)c^{-2}\frac{p_{;s}}{\bar{\varrho}}\{(BS^a)\partial_a v^i - (Bv^i)\partial_a S^a\} + \exp(-3\rho)c^{-2}\frac{p_{;s}}{\bar{\varrho}}\{(Bv^a)\partial_a S^i - (\partial_a v^a)BS^i\},\end{aligned} \tag{5-5a}$$

$$\mathfrak{M}_{(\mathcal{D})} \doteq 2\exp(-2\rho)\{(\partial_a v^a)\partial_b S^b - (\partial_a v^b)\partial_b S^a\} + \exp(-\rho)\delta_{ab}(\operatorname{curl}\varOmega)^a S^b. \tag{5-5b}$$

The terms $\mathfrak{Q}^i_{(v)}$, $\mathfrak{Q}_{(\pm)}$, $\mathfrak{Q}^i_{(\mathcal{C})}$, *and* $\mathfrak{Q}_{(\mathcal{D})}$ *are the null forms relative to* g *defined by*

$$\mathfrak{Q}^i_{(v)} \doteq -\{1 + c^{-1}c_{;\rho}\}(g^{-1})^{\alpha\beta}(\partial_\alpha\rho)\partial_\beta v^i, \tag{5-6a}$$

$$\mathfrak{Q}_{(\pm)} \doteq \mathfrak{Q}^i_{(v)} \mp 2c_{;\rho}(g^{-1})^{\alpha\beta}\partial_\alpha\rho\partial_\beta\rho \pm c\{(\partial_a v^a)(\partial_b v^b) - (\partial_a v^b)\partial_b v^a\}, \tag{5-6b}$$

$$\begin{aligned}
\mathfrak{Q}^i_{(\mathcal{C})} \doteq{}& \exp(-3\rho)c^{-2}\frac{p_{;s}}{\bar{\varrho}}S^i\{(\partial_a v^b)\partial_b v^a - (\partial_a v^a)\partial_b v^b\} \\
&+ \exp(-3\rho)c^{-2}\frac{p_{;s}}{\bar{\varrho}}\{(\partial_a v^a)S^b\partial_b v^i - (S^a\partial_a v^b)\partial_b v^i\} \\
&+ 2\exp(-3\rho)c^{-2}\frac{p_{;s}}{\bar{\varrho}}\{(S^a\partial_a\rho)Bv^i - (B\rho)S^a\partial_a v^i\} \\
&+ 2\exp(-3\rho)c^{-3}c_{;\rho}\frac{p_{;s}}{\bar{\varrho}}\{(S^a\partial_a\rho)Bv^i - (B\rho)S^a\partial_a v^i\} \\
&+ \exp(-3\rho)c^{-2}\frac{p_{;s;\rho}}{\bar{\varrho}}\{(B\rho)S^a\partial_a v^i - (S^a\partial_a\rho)Bv^i\} \\
&+ \exp(-3\rho)c^{-2}\frac{p_{;s;\rho}}{\bar{\varrho}}S^i\{(Bv^a)\partial_a\rho - (\partial_a v^a)B\rho\} \\
&+ 2\exp(-3\rho)c^{-2}\frac{p_{;s}}{\bar{\varrho}}S^i\{(\partial_a v^a)B\rho - (Bv^a)\partial_a\rho\} \\
&+ 2\exp(-3\rho)c^{-3}c_{;\rho}\frac{p_{;s}}{\bar{\varrho}}S^i\{(\partial_a v^a)B\rho - (Bv^a)\partial_a\rho\},
\end{aligned} \tag{5-6c}$$

$$\mathfrak{Q}_{(\mathcal{D})} \doteq 2\exp(-2\rho)\{(S^a\partial_a v^b)\partial_b\rho - (\partial_a v^a)S^b\partial_b\rho\}. \tag{5-6d}$$

In addition, the terms $\mathfrak{L}^i_{(v)}$, $\mathfrak{L}_{(\pm)}$, $\mathfrak{L}_{(s)}$, $\mathfrak{L}^i_{(\varOmega)}$, $\mathfrak{L}^i_{(S)}$, $\mathfrak{L}_{(\operatorname{div}\varOmega)}$, *and* $\mathfrak{L}^i_{(\mathcal{C})}$, *which are at most linear in the derivatives of the unknowns, are defined as*

$$\begin{aligned}
\mathfrak{L}^i_{(v)} \doteq{}& 2\exp(\rho)\epsilon_{iab}(Bv^a)\varOmega^b - \frac{p_{;s}}{\bar{\varrho}}\epsilon_{iab}\varOmega^a S^b \tfrac{1}{2}\exp(-\rho)\frac{p_{;\rho;s}}{\bar{\varrho}}S^a\partial_a v^i \\
&- 2\exp(-\rho)c^{-1}c_{;\rho}\frac{p_{;s}}{\bar{\varrho}}(B\rho)S^i + \exp(-\rho)\frac{p_{;s;\rho}}{\bar{\varrho}}(B\rho)S^i,
\end{aligned} \tag{5-7a}$$

$$\begin{aligned}
\mathfrak{L}_{(\pm)} \doteq{}& \mathfrak{L}^i_{(v)} \pm F_{;s}\mathfrak{L}_{(s)} \mp \tfrac{5}{2}c\exp(-\rho)\frac{p_{;s;\rho}}{\bar{\varrho}}S^a\partial_a\rho \pm 2c^2c_{;s}S^a\partial_a\rho \\
&\mp c\exp(-\rho)\frac{p_{;s;s}}{\bar{\varrho}}\delta_{ab}S^aS^b \pm F_{;s;s}c^2\delta_{ab}S^aS^b,
\end{aligned} \tag{5-7b}$$

$$\mathfrak{L}_{(s)} \doteq c^2S^a\partial_a\rho - cc_{;\rho}S^a\partial_a\rho - cc_{;s}\delta_{ab}S^aS^b, \tag{5-7c}$$

$$\mathfrak{L}^i_{(\varOmega)} \doteq \varOmega^a\partial_a v^i - \exp(-2\rho)c^{-2}\frac{p_{;s}}{\bar{\varrho}}\epsilon_{iab}(Bv^a)S^b, \tag{5-7d}$$

$$\mathfrak{L}^i_{(S)} \doteq -S^a\partial_a v^i + \epsilon_{iab}\exp(\rho)\varOmega^aS^b, \tag{5-7e}$$

$$\mathfrak{L}_{(\operatorname{div}\varOmega)} \doteq -\varOmega^a\partial_a\rho, \tag{5-7f}$$

$$\begin{aligned}
\mathfrak{L}^i_{(\mathcal{C})} \doteq{}& 2\exp(-3\rho)c^{-3}c_{;s}\frac{p_{;s}}{\bar{\varrho}}(Bv^i)\delta_{ab}S^aS^b - 2\exp(-3\rho)c^{-3}c_{;s}\frac{p_{;s}}{\bar{\varrho}}\delta_{ab}S^a(Bv^b)S^i \\
&+ \exp(-3\rho)c^{-2}\frac{p_{;s;s}}{\bar{\varrho}}(Bv^i)\delta_{ab}S^aS^b - \exp(-3\rho)c^{-2}\frac{p_{;s;s}}{\bar{\varrho}}\delta_{ab}(Bv^a)S^bS^i.
\end{aligned} \tag{5-7g}$$

Proof. The equations are copied from [50, Theorem 1], except we have replaced the wave equations for ρ, v^1 from [50, Theorem 1] with equivalent wave equations for $\mathcal{R}_{(\pm)}$ with the help of the identity

$$\square_g\mathcal{R}_{(\pm)} = \square_g v^1 \pm \{c\,\square_g\,\rho + c_{;\rho}(g^{-1})^{\alpha\beta}\partial_\alpha\rho\partial_\beta\rho + 2c^2c_{;s}S^a\partial_a\rho + F_{;s;s}c^2\delta_{ab}S^aS^b + F_{;s}\,\square_g\,s\},$$

which follows from (2-2), the chain rule, the expression (2-8) for g^{-1}, and the transport equation $Bs = 0$, i.e., (1-3). □

6. The bootstrap assumptions and statement of the main a priori estimates

We prove our theorem with a bootstrap argument. In this section, we state the precise bootstrap assumptions, as well as a theorem that features our main a priori estimates. The proof of the theorem occupies Sections 7–14A.

6A. *Bootstrap assumptions.* We now introduce our bootstrap assumptions. In the context of Theorem 6.3 below, we assume that the bootstrap assumptions in the next two subsubsections hold for $t \in [0, T_{(\mathrm{Boot})})$, where $T_{(\mathrm{Boot})} \in [0, 2\mathring{\delta}_*^{-1}]$ is a "bootstrap time."

6A1. *Soft bootstrap assumptions.*

(1) We assume that the change-of-variables map $(t, u, x^2, x^3) \to (t, x^1, x^2, x^3)$ from geometric to Cartesian coordinates is a C^1 diffeomorphism from $[0, T_{(\mathrm{Boot})}) \times \mathbb{R} \times \mathbb{T}^2$ onto $[0, T_{(\mathrm{Boot})}) \times \Sigma$.

(2) We assume that $\mu > 0$ on $[0, T_{(\mathrm{Boot})}) \times \mathbb{R} \times \mathbb{T}^2$.

The first of these "soft bootstrap assumptions" allows us, in particular, to switch back and forth between viewing tensorfields as a function of the geometric coordinates (which is the dominant view we take throughout the analysis) and the Cartesian coordinates. The second soft bootstrap assumption guarantees that there are no shocks present in the bootstrap region (though it allows for the possibility that a shock will form precisely at time $T_{(\mathrm{Boot})}$).

6A2. *Quantitative bootstrap assumptions.* Let $M_* \in \mathbb{N}$ be the absolute constant appearing in the statements of Theorem 4.2 above and Proposition 12.1 below. Moreover, as we stated already in Section 4A, N_{top} denotes any fixed positive integer satisfying $N_{\mathrm{top}} \geq 2M_* + 10$.

Remark 6.1 (rationale behind our choice $N_{\mathrm{top}} \geq 2M_*+10$). Later on, our assumption $N_{\mathrm{top}} \geq 2M_*+10$ and the bootstrap assumptions will allow us to control $\leq N_{\mathrm{top}}$ derivatives of nonlinear products by bounding all terms in L^∞ except perhaps the one factor hit by the most derivatives. Roughly,[47] the reason is that our derivative count will be such that any factor that is hit by $\leq N_{\mathrm{top}} - M_* - 4$ or fewer derivatives is bounded in L^∞. We will often avoid explicitly pointing out this aspect of our derivative count.

<u>L^2 bootstrap assumptions for the wave variables</u>: For $N_{\mathrm{top}} - M_* + 1 \leq N \leq N_{\mathrm{top}}$,[48] we assume the following bounds, where the energies $\mathbb{W}_N$ are defined in Section 3B2 and $\mu_\star(t)$ is defined in Definition 2.16:

$$\mathbb{W}_N(t) \leq \mathring{\epsilon}\mu_\star^{-2M_*+2N_{\mathrm{top}}-2N+1.8}(t). \tag{6-1}$$

For $1 \leq N \leq N_{\mathrm{top}} - M_*$,

$$\mathbb{W}_N(t) \leq \mathring{\epsilon}. \tag{6-2}$$

[47] In reality, the different solution variables that we have to track, such as Ψ, Ω^i, L^i, μ, etc., exhibit slightly different amounts of L^∞ regularity.

[48] Equivalently, for $0 \leq K \leq M_* - 1$, we have $\mathbb{W}_{N_{\mathrm{top}}-K}(t) \leq \mathring{\epsilon}\mu_\star^{-2M_*+2K+1.8}(t)$.

L^∞ bootstrap assumptions for the wave variables:

$$\|\mathcal{R}_{(+)}\|_{L^\infty(\Sigma_t)} \leq \mathring{\alpha}^{1/2}, \quad \|\breve{X}\mathcal{R}_{(+)}\|_{L^\infty(\Sigma_t)} \leq 3\mathring{\delta}, \tag{6-3}$$

$$\|(\mathcal{R}_{(-)}, v^2, v^3, s)\|_{L^\infty(\Sigma_t)}, \quad \|\breve{X}(\mathcal{R}_{(-)}, v^2, v^3, s)\|_{L^\infty(\Sigma_t)} \leq \mathring{\varepsilon}^{1/2}, \tag{6-4}$$

$$\|\mathcal{P}^{[1,N_{\mathrm{top}}-M_*-2]}\Psi\|_{L^\infty(\Sigma_t)} \leq \mathring{\varepsilon}^{1/2}, \quad \|\mathcal{P}^{[1,N_{\mathrm{top}}-M_*-4]}\breve{X}\Psi\|_{L^\infty(\Sigma_t)} \leq \mathring{\varepsilon}^{1/2}. \tag{6-5}$$

L^∞ bootstrap assumptions for the specific vorticity:

$$\|\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-2}\Omega\|_{L^\infty(\Sigma_t)} + \|\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-4}\breve{X}\Omega\|_{L^\infty(\Sigma_t)} \leq \mathring{\varepsilon}. \tag{6-6}$$

L^∞ bootstrap assumptions for the entropy gradient:

$$\|\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-2}S\|_{L^\infty(\Sigma_t)} + \|\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-4}\breve{X}S\|_{L^\infty(\Sigma_t)} \leq \mathring{\varepsilon}. \tag{6-7}$$

L^∞ bootstrap assumptions for the modified fluid variables:

$$\|\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-3}(\mathcal{C}, \mathcal{D})\|_{L^\infty(\Sigma_t)} \leq \mathring{\varepsilon}. \tag{6-8}$$

Remark 6.2 (the main large quantity in the problem). From the discussion of the parameters at the beginning of Section 4A and (6-3)–(6-8) we see that the main large quantity in the problem is $\breve{X}\mathcal{R}_{(+)}$; all other terms exhibit smallness that is controlled by $\mathring{\alpha}$ and $\mathring{\varepsilon}$. This, of course, is tied to the kind of initial data we treat here.

6B. *Statement of the main a priori estimates.* We now state the theorem that yields our main a priori estimates. Its proof will be the content of Sections 7–14A.

Theorem 6.3 (the main a priori estimates). *Let $T_{(\mathrm{Boot})} \in [0, 2\mathring{\delta}_*^{-1}]$. Suppose that:*

(1) *The assumptions on the initial data stated in Section 4A hold.* (*Note that these assumptions involve N_{top}, M_*, $\mathring{\sigma}$, $\mathring{\delta}_*$, $\mathring{\delta}$, $\mathring{\alpha}$, and $\mathring{\varepsilon}$.*)

(2) *The bootstrap assumptions* (6-1)–(6-8) *all hold for all $t \in [0, T_{(\mathrm{Boot})})$* (*where we recall that in the bootstrap assumptions, N_{top} is any integer satisfying $N_{\mathrm{top}} \geq 2M_* + 10$, where $M_* \in \mathbb{N}$ is the absolute constant appearing in the statements of Theorem 4.2 and Proposition 12.1*).

(3) *In* (6-3), *the parameter $\mathring{\alpha}$ is sufficiently small in a manner that depends only on the equation of state and $\bar{\varrho}$.*

(4) *The parameter $\mathring{\varepsilon} > 0$ in* (6-1)–(6-8) *satisfies $\mathring{\varepsilon}^{1/2} \leq \mathring{\alpha}$ and is sufficiently small in a manner that depends only on the equation of state, N_{top}, $\bar{\varrho}$, $\mathring{\sigma}$, $\mathring{\delta}$, and $\mathring{\delta}_*^{-1}$.*

(5) *The soft bootstrap assumptions stated in Section 6A1 hold* (*including $\mu > 0$ in $[0, T_{(\mathrm{Boot})}) \times \mathbb{R} \times \mathbb{T}^2$*).

Then there exists a constant $C_\blacklozenge > 0$ depending only on the equation of state and $\bar{\varrho}$, and a constant $C > 0$ depending on the equation of state, N_{top}, $\bar{\varrho}$, $\mathring{\sigma}$, $\mathring{\delta}$, and $\mathring{\delta}_*^{-1}$ such that the following holds for all $t \in [0, T_{(\mathrm{Boot})})$:

(1) (6-1) *and* (6-2) *hold with $\mathring{\varepsilon}$ replaced by $C\mathring{\varepsilon}^2$.*

(2) *The two inequalities in* (6-3) *hold with $\mathring{\alpha}^{1/2}$ replaced by $C_\blacklozenge\mathring{\alpha}$ and $3\mathring{\delta}$ replaced by $2\mathring{\delta}$ respectively.*

(3) *The inequalities in* (6-4) *and* (6-5) *hold with* $\mathring{\varepsilon}^{1/2}$ *replaced by* $C\mathring{\varepsilon}$.

(4) *The inequalities* (6-6)–(6-8) *all hold with* $\mathring{\varepsilon}$ *replaced by* $C\mathring{\varepsilon}^{3/2}$.

Sections 7–13 will be devoted to the proof of Theorem 6.3. See Section 14A for the conclusion of the proof.

From now on, we will use the conventions for constants stated in Section 2A and Theorem 6.3.

7. A localization lemma via finite speed of propagation

We work under the assumptions of Theorem 6.3.

Lemma 7.1 (a localization lemma). *Let* $U_0 \doteq 2\mathring{\sigma} + 4\mathring{\delta}_*^{-1}$. *Then, for all* $t \in [0, T_{(\mathrm{Boot})})$,

$$(\rho, v, s) = (0, 0, 0), \quad \textit{whenever } u \notin (0, U_0).$$

Proof. Recall that we have normalized (see (2-1)) $c(0, 0) = 1$, and (by (4-1)) the data are compactly supported in the region where $|x^1| \leq \mathring{\sigma}$. Hence, by a standard finite speed of propagation argument, we see that $(\rho, v, s) = (0, 0, 0)$ whenever $|x^1| \geq \mathring{\sigma} + t$. More precisely, this can be proved by applying standard energy methods to the first-order formulation of the compressible Euler equations provided by [16, equation (1.201)], where the relevant energy identities can be obtained with the help of the "energy current" vectorfields defined by [16, equations (1.204), (1.205)]. Since $t < T_{(\mathrm{Boot})} \leq 2\mathring{\delta}_*^{-1}$,

$$\{(t, x) \in [0, T_{(\mathrm{Boot})}) \times \Sigma : t - x^1 \geq \mathring{\sigma} + 4\mathring{\delta}_*^{-1}\} \subseteq \overbrace{\{(t, x) \in [0, T_{(\mathrm{Boot})}) \times \Sigma : x^1 \leq -\mathring{\sigma} - t\}}^{\text{solution is trivial here}}.$$

In particular, this implies

$$(\rho, v, s) = (0, 0, 0) \quad \text{unless } -\mathring{\sigma} < t - x^1 < \mathring{\sigma} + 4\mathring{\delta}_*^{-1}. \tag{7-1}$$

Observe now that since $u \restriction_{\{t=0\}} = \mathring{\sigma} - x^1$, in the set $\{(t, x) \in [0, T_{(\mathrm{Boot})}) \times \Sigma : |x^1| \geq \mathring{\sigma} + t\}$ (where the solution is trivial), we have $u = t + \mathring{\sigma} - x^1$. In particular, $\{u = 0\} = \{t - x^1 = -\mathring{\sigma}\}$ and $\{u = U_0\} = \{t - x^1 = \mathring{\sigma} + 4\mathring{\delta}_*^{-1}\}$. The conclusion thus follows from (7-1). ☐

For the rest of the paper, $U_0 > 0$ *denotes the constant appearing in the statement of Lemma 7.1.*

8. Estimates for the geometric quantities associated to the acoustical metric

We continue to work under the assumptions of Theorem 6.3.

In this section, we collect some estimates of the geometric quantities μ, $L^i_{(\mathrm{small})}$ (see Definition 2.21), under the bootstrap assumptions on the fluid variables. These estimates are the same as those appearing in [36; 52]. Our analysis will therefore be somewhat brief in some spots, and we will refer the reader to [36; 52] for details.

We highlight the following point, which is crucial for the subsequent analysis: the bounds for μ, $L^i_{(\mathrm{small})}$ and the wave variables Ψ control all the other geometric quantities, including the transformation coefficients between different sets of vectorfields, as well as the commutators of vectorfields.

8A. ***Some preliminary geoanalytic identities.*** In this section, we provide some geoanalytic identities that we will use throughout our analysis.

We start by recalling the definition of a null form with respect to the acoustical metric ("g-null form" for short).

Definition 8.1 (g-null forms). Let $\phi^{(1)}$ and $\phi^{(2)}$ be scalar functions. We use the notation $\mathcal{Q}^{(g)}(\partial\phi^{(1)}, \partial\phi^{(2)})$ to denote any derivative-quadratic term of the form

$$\mathcal{Q}^{(g)}(\partial\phi^{(1)}, \partial\phi^{(2)}) = \mathrm{f}(L^i, \Psi)(g^{-1})^{\alpha\beta}\partial_\alpha\phi^{(1)}\partial_\beta\phi^{(2)}, \tag{8-1}$$

where $\mathrm{f}(\,\cdot\,)$ is a smooth function.

We use the notation $\mathcal{Q}_{\alpha\beta}(\partial\phi^{(1)}, \partial\phi^{(2)})$ to denote any derivative-quadratic term of the form

$$\mathcal{Q}_{\alpha\beta}(\partial\phi^{(1)}, \partial\phi^{(2)}) = \mathrm{f}(L^i, \Psi)\{\partial_\alpha\phi^{(1)}\partial_\beta\phi^{(2)} - \partial_\beta\phi^{(1)}\partial_\alpha\phi^{(2)}\}, \tag{8-2}$$

where $\mathrm{f}(\,\cdot\,)$ is a smooth function.

Lemma 8.2 (crucial structural properties of null forms). *Let $\mathcal{Q}(\partial\phi^{(1)}, \partial\phi^{(2)})$ be a g-null form of type* (8-1) *or* (8-2). *Then there exist smooth functions, all schematically denoted by* f (*and which are different from the* f *in Definition 8.1*), *such that the following identity holds*:

$$\mu\mathcal{Q}(\partial\phi^{(1)}, \partial\phi^{(2)}) = \mathrm{f}(L^i, \Psi)\breve{X}\phi^{(1)} \cdot \mathcal{P}\phi^{(2)} + \mathrm{f}(L^i, \Psi)\breve{X}\phi^{(2)} \cdot \mathcal{P}\phi^{(1)} + \mu\mathrm{f}(L^i, \Psi)\mathcal{P}\phi^{(1)} \cdot \mathcal{P}\phi^{(2)}. \tag{8-3}$$

In particular, decomposing all differentiations in the null form with respect to the $\{L, X, Y, Z\}$ frame leads to the absence of all $X\phi^{(1)} \cdot X\phi^{(2)}$ terms on the right-hand side of (8-3).

Proof. For null forms of type (8-2), (8-3) follows from Lemma 2.22 and the fact that the Cartesian component functions X^1, X^2, X^3 are smooth functions of the L^i and Ψ (see (2-23)). For null forms of type (8-1), (8-3) follows from the basic identity $g^{-1} = -L \otimes L - (L \otimes X + X \otimes L) + \not{g}^{-1}$ (see, e.g., [52, (2.40b)]) and Lemma 2.32. □

Lemma 8.3 (expressions for the transversal derivatives of the transport variables in terms of tangential derivatives). *There exist smooth functions, all schematically denoted by* "f", *such that the following identities hold*:

$$\breve{X}\Omega^i = -\mu L\Omega^i + (\Omega, S) \cdot \mathrm{f}(\Psi, L^i, \mu, \breve{X}\Psi, \mathcal{P}\Psi), \tag{8-4}$$

$$\breve{X}S^i = -\mu LS^i + (\Omega, S) \cdot \mathrm{f}(\Psi, L^i, \mu, \breve{X}\Psi, \mathcal{P}\Psi), \tag{8-5}$$

$$\breve{X}\mathcal{C}^i = -\mu L\mathcal{C}^i + (\Omega, S, \mathcal{P}\Omega, \mathcal{P}S) \cdot \mathrm{f}(\Psi, L^i, \mu, \breve{X}\Psi, \mathcal{P}\Psi), \tag{8-6}$$

$$\breve{X}\mathcal{D}^i = -\mu L\mathcal{D}^i + (\Omega, S, \mathcal{P}\Omega, \mathcal{P}S) \cdot \mathrm{f}(\Psi, L^i, \mu, \breve{X}\Psi, \mathcal{P}\Psi). \tag{8-7}$$

Proof. Equations (8-4) and (8-5) follow from the transport equations (5-2a) and (5-2c), (2-23) (which implies that $\mu B = \breve{X} + \mu L$), and Lemma 2.22.

Equations (8-6) and (8-7) follow from a similar argument based the transport equations (5-3b) and (5-4a), where we use Lemma 8.2 to decompose the null form source terms and (8-4)–(8-5) to re-express all $\breve{X}$ derivatives of (Ω, S). □

Lemma 8.4 (identity for $\breve{X}L^i$). *There exist smooth functions, all schematically denoted by* f, *such that*

$$\breve{X}L^i = \mathrm{f}(\Psi, L^i)\breve{X}\Psi + \mu\mathrm{f}(\Psi, L^i)\mathcal{P}\Psi + \mathrm{f}(\Psi, L^i)\mathcal{P}\mu. \tag{8-8}$$

Proof. This was proved as [52, (2.71)] (which holds in the present context with obvious modifications such as replacing $G_{LL}\breve{X}\Psi$ with $\vec{G}_{LL} \diamond \breve{X}\vec{\Psi}$, etc.), where we have used that the Cartesian component functions X^1, X^2, X^3 are smooth functions of the L^i and Ψ (see (2-23)). □

Lemma 8.5 (simple commutator identities). *For each pair $\mathcal{P}_1, \mathcal{P}_2 \in \{L, Y, Z\}$, there exist smooth functions, all schematically denoted by "f", such that the following identity holds*:

$$[\mathcal{P}_1, \mathcal{P}_2] = \mathrm{f}(L^i, \Psi, \mathcal{P}L^i, \mathcal{P}\Psi)Y + \mathrm{f}(L^i, \Psi, \mathcal{P}L^i, \mathcal{P}\Psi)Z. \tag{8-9}$$

Moreover, for each $\mathcal{P} \in \{L, Y, Z\}$, there exist smooth functions, all schematically denoted by "f", such that the following identity holds:

$$[\mathcal{P}, \breve{X}] = \mathrm{f}(\mu, L^i, \Psi, \mathcal{P}\mu, \breve{X}\Psi, \mathcal{P}\Psi)Y + \mathrm{f}(\mu L^i, \Psi, \mathcal{P}\mu, \breve{X}\Psi, \mathcal{P}\Psi)Z. \tag{8-10}$$

Proof. We first prove (8-10). Lemma 2.23 implies that $[\mathcal{P}, \breve{X}]$ is $\ell_{t,u}$-tangent, i.e., that $[\mathcal{P}, \breve{X}]t = [\mathcal{P}, \breve{X}]u = 0$. Hence, (2-28b)–(2-28c) imply that this commutator can be written as a linear combination of Y, Z. Since the Cartesian component functions X^1, X^2, X^3 are smooth functions of the L^i and Ψ (see (2-23)), the same holds for the component functions $\mathcal{P}^0$, $\mathcal{P}^1$, $\mathcal{P}^2$, $\mathcal{P}^3$ (this is obvious for $\mathcal{P} = L$, while see Lemmas 2.23–2.24 for $\mathcal{P} = Y, Z$). Also using that $\breve{X}^i = \mu\breve{X}^i$, we conclude (8-10) by computing relative to the Cartesian coordinates, using Lemma 2.22 to express Cartesian coordinate partial derivatives in terms of derivatives with respect to Y, Z (the X- and L-derivative components of the commutator must vanish since $[\mathcal{P}, \breve{X}]$ is $\ell_{t,u}$-tangent), and using (8-8) to substitute for $\breve{X}L^i$ factors.

The identity (8-9) can be proved through similar but simpler arguments that do not involve factors of μ or $\breve{X}$ differentiations. □

8B. *The easy L^∞ estimates.*

Proposition 8.6 (L^∞ estimates for the acoustical geometry). *The following estimates hold for all $t \in [0, T_{(\text{Boot})})$*:

$$\|\mu\|_{L^\infty(\Sigma_t)} + \|L\mu\|_{L^\infty(\Sigma_t)} \lesssim 1, \quad \|L^i_{(\text{small})}\|_{L^\infty(\Sigma_t)} \lesssim_{\blacklozenge} \mathring{\alpha}, \quad \|Y\mu\|_{L^\infty(\Sigma_t)} + \|Z\mu\|_{L^\infty(\Sigma_t)} \lesssim \mathring{\epsilon}^{1/2},$$
$$\|\mathcal{P}^{[2,N_{\text{top}}-M_*-4]}\mu\|_{L^\infty(\Sigma_t)} + \|\mathcal{P}^{[1,N_{\text{top}}-M_*-3]}L^i\|_{L^\infty(\Sigma_t)} \lesssim \mathring{\epsilon}^{1/2}.$$

Proof. These can be proved using the transport equations (2-40) and (2-41) (commuted with $\mathcal{P}^N$), the initial data size-assumptions (4-3a)–(4-4), and the bootstrap assumptions (6-3)–(6-5). See [52, Proposition 8.10] for details of this argument. We note these estimates lose a slight amount of regularity compared to Ψ because the transport equations (2-40) and (2-41) depend on the derivatives of Ψ. □

Our analysis also relies on the following L^∞ estimates.

Proposition 8.7 (L^∞ estimates for other geometric quantities). *The following estimates hold for all $t \in [0, T_{(\text{Boot})})$, where c denotes the speed of sound*:

$$\|X^i_{(\text{small})}\|_{L^\infty(\Sigma_t)} \lesssim_{\blacklozenge} \mathring{\alpha}^{1/2}, \qquad \|c-1\|_{L^\infty(\Sigma_t)} \lesssim_{\blacklozenge} \mathring{\alpha}^{1/2},$$
$$\|\mathcal{P}^{[1,N_{\text{top}}-M_*-3]}X^i\|_{L^\infty(\Sigma_t)} \lesssim \mathring{\epsilon}^{1/2}, \qquad \|\mathcal{P}^{[1,N_{\text{top}}-M_*-2]}c\|_{L^\infty(\Sigma_t)} \lesssim \mathring{\epsilon}^{1/2}.$$

Proof. The estimates for $X_{(\mathrm{small})}$ follow from (2-25a)–(2-25b), (2-26a), the bootstrap assumptions (6-3)–(6-5), and Proposition 8.6.

The estimates for c follow from the bootstrap assumptions (6-3)–(6-5) and the fact that c is a smooth function of ρ and s with $c(0,0)=1$ (see (2-1)). $\square$

The estimates in Propositions 8.6 and 8.7 also imply the following bounds for the commutators.

Proposition 8.8 (pointwise bounds for vectorfield commutators). *All the commutators* $[L,\breve{X}]$, $[L,Y]$, $[L,Z]$, $[\breve{X},Y]$, $[\breve{X},Z]$ *and* $[Y,Z]$ *are* $\ell_{t,u}$*-tangent.*

Moreover, if ϕ *is a scalar function, then for* $0\le N\le N_{\mathrm{top}}$ *iterated commutators can be bounded pointwise as follows:*

$$\begin{aligned}|[L,\mathcal{P}^N]\phi| &\lesssim \mathring{\epsilon}^{1/2}|\mathcal{P}^{[1,N]}\phi| + \sum_{\substack{N_1+N_2\le N+1\\ N_1,N_2\le N}} |\mathcal{P}^{[2,N_1]}(L^i,\Psi)|\,|\mathcal{P}^{[1,N_2]}\phi|,\\ |[\breve{X},\mathcal{P}^N]\phi| &\lesssim |\mathcal{P}^{[1,N]}\phi| + \sum_{\substack{N_1+N_2\le N+1\\ N_1,N_2\le N}} |\mathcal{P}^{[2,N_1]}(\mu,L^i,\Psi)|\,|\mathcal{P}^{[1,N_2]}\phi| + \sum_{\substack{N_1+N_2\le N\\ N_1\le N-1}} |\mathcal{P}^{[2,N_1]}\breve{X}\Psi|\,|\mathcal{P}^{[1,N_2]}\phi|.\end{aligned} \tag{8-11}$$

In particular,

$$\begin{aligned}|[L,\mathcal{P}^N]\phi| &\lesssim \mathring{\epsilon}^{1/2}|\mathcal{P}^{[1,N]}\phi| \quad \text{if } 0\le N\le N_{\mathrm{top}}-M_*-3,\\ |[\breve{X},\mathcal{P}^N]\phi| &\lesssim |\mathcal{P}^{[1,N]}\phi| \qquad\quad \text{if } 0\le N\le N_{\mathrm{top}}-M_*-4.\end{aligned} \tag{8-12}$$

Proof. All the commutators can be read off from Lemma 2.23 (and using that coordinate vectorfields commute). In particular, since the coefficient of ∂_t in L and the coefficient of ∂_u in $\breve{X}$ both are equal to 1, all the stated commutators are $\ell_{t,u}$-tangent.

We first prove (8-11) for $|[L,\mathcal{P}^N]\phi|$. By Lemma 2.23 and the fact $L^i+X^i-v^i=0$ (by (2-26a)),

$$|[L,\mathcal{P}^N]\phi| \lesssim \sum_{k=2}^{N} \sum_{\substack{N_1+\cdots+N_k=N+1\\ 1\le N_k\le N}} \underbrace{|\mathcal{P}^{N_1}(L^i,\Psi)|\cdots|\mathcal{P}^{N_{k-1}}(L^i,\Psi)|\,|\mathcal{P}^{[1,N_k]}\phi|}_{\doteq(*)}. \tag{8-13}$$

By (6-3)–(6-5), Propositions 8.6, 8.7 (and $N\le N_{\mathrm{top}}$), either $|\mathcal{P}^{N_j}(L^i,\Psi)|\lesssim \mathring{\epsilon}^{1/2}$ for $1\le j\le k-1$ (in which case $(*)\lesssim \mathring{\epsilon}^{1/2}|\mathcal{P}^{[1,N]}\phi|$), or else there is exactly one factor $|\mathcal{P}^{N_j}(L^i,\Psi)|$ with $N_j>N_{\mathrm{top}}-M_*-3$ not bounded by $\lesssim \mathring{\epsilon}^{1/2}$, in which case

$$(*) \lesssim \sum_{\substack{N_1+N_2\le N+1\\ N_1,N_2\le N}} |\mathcal{P}^{[2,N_1]}(L^i,\Psi)|\,|\mathcal{P}^{[1,N_2]}\phi|.$$

Hence, (8-13) is bounded above by the right-hand side of the first inequality in (8-11).

To bound $[\breve{X},\mathcal{P}^N]\phi$, we note that according to Lemma 2.23, there is, in addition to (8-13), the terms[49]

$$\sum_{k=2}^{N} \sum_{\substack{N_1+\cdots+N_k=N\\ N_{k-1}\le N-1\\ 1\le N_k\le N}} |\mathcal{P}^{N_1}(L^i,\Psi)|\cdots|\mathcal{P}^{N_{k-2}}(L^i,\Psi)|\,|\mathcal{P}^{N_{k-1}}\breve{X}(L^i,\Psi)|\,|\mathcal{P}^{N_k}\phi|, \tag{8-14}$$

[49]Importantly, one checks from Lemma 2.23 that there are no terms of the form $|\mathcal{P}^{N_{k-1}}\breve{X}\mu|$!

$$\sum_{k=2}^{N} \sum_{\substack{N_1+\cdots+N_k=N+1 \\ 1\le N_k\le N}} |\mathcal{P}^{N_1}(L^i,\Psi)|\cdots|\mathcal{P}^{N_{k-2}}(L^i,\Psi)||\mathcal{P}^{N_{k-1}}\upmu||\mathcal{P}^{N_k}\phi|. \tag{8-15}$$

Hence, with the help of (8-8), we can substitute for the terms $\breve{X}L^i$ on the right-hand side of (8-14), and thus the right-hand side of (8-14) can be bounded above by the right-hand side of (8-13) plus (8-15) and

$$\sum_{k=2}^{N} \sum_{\substack{N_1+\cdots+N_k=N \\ N_{k-1}\le N-1 \\ 1\le N_k\le N}} |\mathcal{P}^{N_1}(L^i,\Psi)|\cdots|\mathcal{P}^{N_{k-2}}(L^i,\Psi)||\mathcal{P}^{N_{k-1}}\breve{X}\Psi||\mathcal{P}^{N_k}\phi|, \tag{8-16}$$

both of which, by arguments similar to the ones we used to prove (8-13), can be bounded above by the right-hand side of the second inequality in (8-11).

To get from (8-11) to (8-12), we use the L^∞ bounds in (6-3)–(6-5) and Propositions 8.6 and 8.7, which are applicable in the sense that they control a sufficient number of derivatives of all relevant quantities in L^∞. □

In the rest of the paper, we will often silently use the following simple lemma.

Lemma 8.9 (the norm of the $\ell_{t,u}$-tangent commutator vectorfields and simple comparison estimates). *The $\ell_{t,u}$-tangent commutator vectorfields $\{Y, Z\}$ satisfy the following pointwise bounds on $\mathcal{M}_{T_{(\mathrm{Boot})},U_0}$:*

$$|Y| \lesssim 1, \quad |Z| \lesssim 1. \tag{8-17}$$

Moreover, for any $\ell_{t,u}$-tangent tensorfield ξ, the following pointwise bounds hold on $\mathcal{M}_{T_{(\mathrm{Boot})},U_0}$:

$$|\nabla\!\!\!/\,\xi| \approx |\nabla\!\!\!/_Y\xi| + |\nabla\!\!\!/_Z\xi|. \tag{8-18}$$

Proof. To prove (8-17), we use Lemmas 2.23 and 2.32 and the fact that the Cartesian component functions X^1, X^2, X^3 are smooth functions of the L^i and Ψ (see (2-23)) to deduce that $|Y|^2 = g\!\!\!/_{AB}Y^AY^B = \mathrm{f}(L^i,\Psi)$, where f is a smooth function. Similar remarks hold for $|Z|^2$. The desired estimates in (8-17) therefore follow from the bootstrap assumptions (6-3)–(6-4) and Proposition 8.6.

To prove (8-18), we note that the $g\!\!\!/$-Cauchy–Schwarz inequality and (8-17) imply that $|\nabla\!\!\!/_Y\xi| + |\nabla\!\!\!/_Z\xi| \lesssim |\nabla\!\!\!/\,\xi|$. We will show how to obtain the reverse inequality when ξ is a scalar function; the case of an arbitrary $\ell_{t,u}$-tangent tensorfield can be handled using the same arguments, which will complete the proof. To proceed, we note that for scalar functions ξ we have $|\nabla\!\!\!/\,\xi|^2 = (g\!\!\!/^{-1})^{AB}(\partial\!\!\!/_A\xi)(\partial\!\!\!/_B\xi)$. We now use Lemmas 2.24 and 2.32 and the fact that X^1, X^2, X^3 are smooth functions of L^i and Ψ (as noted above) to deduce that there exist smooth functions, all schematically denoted by f, such that $(g\!\!\!/^{-1})^{AB}(\partial\!\!\!/_A\xi)(\partial\!\!\!/_B\xi) = \mathrm{f}(L^i,\Psi)(Y\xi)^2 + \mathrm{f}(L^i,\Psi)(Y\xi)(Z\xi) + \mathrm{f}(L^i,\Psi)(Z\xi)^2$. Also using the bootstrap assumptions (6-3)–(6-4), Young's inequality, and Proposition 8.6, we conclude that $|\nabla\!\!\!/\,\xi|^2 \lesssim |Y\xi|^2 + |Z\xi|^2 = |\nabla\!\!\!/_Y\xi|^2 + |\nabla\!\!\!/_Z\xi|^2$ as desired. □

8C. *L^∞ estimates involving higher transversal derivatives.* Some aspects of our main results rely on having L^∞ estimates for the higher transversal derivatives of various solution variables. We provide these estimates in the next proposition. The proofs are similar to the proofs of related estimates in [52].

Proposition 8.10 (L^∞ estimates involving higher transversal derivatives). *The following estimates hold*[50] *for all* $t \in [0, T_{(\text{Boot})})$ *and* $u \in [0, U_0]$, *where in* (8-22b), $\not{\mathcal{P}} \in \{Y, Z\}$:

L^∞ estimates involving two or three transversal derivatives of the wave variables:

$$\|L\mathcal{P}^{\leq 2}\breve{X}\breve{X}\Psi\|_{L^\infty(\mathcal{M}_{t,u})} \leq C\mathring{\epsilon}^{1/2}, \tag{8-19a}$$

$$\|\mathcal{P}^{[1,2]}\breve{X}\breve{X}\Psi\|_{L^\infty(\mathcal{M}_{t,u})} \leq C\mathring{\epsilon}^{1/2}, \tag{8-19b}$$

$$\|\breve{X}\breve{X}\mathcal{R}_{(+)}\|_{L^\infty(\mathcal{M}_{t,u})} \leq \|\breve{X}\breve{X}\mathcal{R}_{(+)}\|_{L^\infty(\Sigma_0)} + C\mathring{\epsilon}^{1/2}, \tag{8-19c}$$

$$\|\breve{X}\breve{X}(\mathcal{R}_{(-)}, v^1, v^2, s)\|_{L^\infty(\mathcal{M}_{t,u})} \leq C\mathring{\epsilon}^{1/2}, \tag{8-19d}$$

$$\|L\breve{X}\breve{X}\breve{X}\Psi\|_{L^\infty(\mathcal{M}_{t,u})} \leq C\mathring{\epsilon}^{1/2}, \tag{8-20a}$$

$$\|\breve{X}\breve{X}\breve{X}\mathcal{R}_{(+)}\|_{L^\infty(\mathcal{M}_{t,u})} \leq \|\breve{X}\breve{X}\breve{X}\mathcal{R}_{(+)}\|_{L^\infty(\Sigma_0)} + C\mathring{\epsilon}^{1/2}, \tag{8-20b}$$

$$\|\breve{X}\breve{X}\breve{X}(\mathcal{R}_{(-)}, v^1, v^2, s)\|_{L^\infty(\mathcal{M}_{t,u})} \leq C\mathring{\epsilon}^{1/2}. \tag{8-20c}$$

L^∞ estimates involving one or two transversal derivatives of μ:

$$\|L\breve{X}\mu\|_{L^\infty(\mathcal{M}_{t,u})} \leq \tfrac{1}{2}\|\breve{X}(\vec{G}_{LL} \diamond \breve{X}\vec{\Psi})\|_{L^\infty(\Sigma_0)} + C\mathring{\epsilon}^{1/2}, \tag{8-21a}$$

$$\|\breve{X}\mu\|_{L^\infty(\mathcal{M}_{t,u})} \leq \|\breve{X}\mu\|_{L^\infty(\Sigma_0)} + \mathring{\delta}_*^{-1}\|\breve{X}(\vec{G}_{LL} \diamond \breve{X}\vec{\Psi})\|_{L^\infty(\Sigma_0)} + C\mathring{\epsilon}^{1/2}, \tag{8-21b}$$

$$\|L\breve{X}\mathcal{P}\mu\|_{L^\infty(\mathcal{M}_{t,u})}, \quad \|L\breve{X}\mathcal{P}^2\mu\|_{L^\infty(\mathcal{M}_{t,u})} \leq C\mathring{\epsilon}^{1/2}, \tag{8-22a}$$

$$\|\breve{X}\not{\mathcal{P}}\mu\|_{L^\infty(\mathcal{M}_{t,u})}, \quad \|\breve{X}\mathcal{P}^2\mu\|_{L^\infty(\mathcal{M}_{t,u})} \leq C\mathring{\epsilon}^{1/2}, \tag{8-22b}$$

$$\|LL\breve{X}\breve{X}\mu\|_{L^\infty(\mathcal{M}_{t,u})} \leq C\mathring{\epsilon}^{1/2}, \tag{8-23a}$$

$$\|L\breve{X}\breve{X}\mu\|_{L^\infty(\mathcal{M}_{t,u})} \leq \tfrac{1}{2}\|\breve{X}\breve{X}(\vec{G}_{LL} \diamond \breve{X}\vec{\Psi})\|_{L^\infty(\Sigma_0)} + C\mathring{\epsilon}^{1/2}, \tag{8-23b}$$

$$\|\breve{X}\breve{X}\mu\|_{L^\infty(\mathcal{M}_{t,u})} \leq \|\breve{X}\breve{X}\mu\|_{L^\infty(\Sigma_0)} + \mathring{\delta}_*^{-1}\|\breve{X}\breve{X}(\vec{G}_{LL} \diamond \breve{X}\vec{\Psi})\|_{L^\infty(\Sigma_0)} + C\mathring{\epsilon}^{1/2}. \tag{8-23c}$$

L^∞ estimates involving one or two transversal derivatives of L^i:

$$\|\mathcal{P}^{[1,N_{\text{top}}-M_*-5]}\breve{X}L^i\|_{L^\infty(\mathcal{M}_{t,u})} \leq C\mathring{\epsilon}^{1/2}, \tag{8-24a}$$

$$\|\breve{X}L^i\|_{L^\infty(\mathcal{M}_{t,u})} \leq C, \tag{8-24b}$$

$$\|L\mathcal{P}\breve{X}\breve{X}L^i\|_{L^\infty(\mathcal{M}_{t,u})} \leq C\mathring{\epsilon}^{1/2}, \tag{8-25a}$$

$$\|\mathcal{P}\breve{X}\breve{X}L^i\|_{L^\infty(\mathcal{M}_{t,u})} \leq C\mathring{\epsilon}^{1/2}, \tag{8-25b}$$

$$\|\breve{X}\breve{X}L^i\|_{L^\infty(\mathcal{M}_{t,u})} \leq C. \tag{8-25c}$$

L^∞ estimates involving transversal derivatives of the transported variables:

$$\begin{aligned}\|\mathcal{P}^{\leq 3}\breve{X}^{\leq 1}(\Omega, S)\|_{L^\infty(\mathcal{M}_{t,u})} + \|\mathcal{P}^{\leq 2}\breve{X}\breve{X}(\Omega, S)\|_{L^\infty(\mathcal{M}_{t,u})} + \|\breve{X}^{\leq 3}(\Omega, S)\|_{L^\infty(\mathcal{M}_{t,u})} \\ + \|\mathcal{P}^{\leq 2}\breve{X}^{\leq 1}(\mathcal{C}, \mathcal{D})\|_{L^\infty(\mathcal{M}_{t,u})} + \|\breve{X}^{\leq 2}(\mathcal{C}, \mathcal{D})\|_{L^\infty(\mathcal{M}_{t,u})} \leq C\mathring{\epsilon}.\end{aligned} \tag{8-26}$$

[50] Based on our assumptions on the data (see Section 4A), we could obtain L^∞ control over additional $\mathcal{F}_u$-tangential derivatives of the quantities stated in the proposition — but not! additional $\breve{X}$ differentiations. However, for convenience, in the proposition, we have only derived control of a sufficient number of derivatives so that the estimates close and so that we can use the results in our proof of Lemma 14.2 and in the Appendix.

Finally,

we can permute the vectorfield operators on the left-hand sides of (8-19a)–(8-25c) *up to error terms of L^∞ size $\mathcal{O}(\mathring{\varepsilon}^{1/2})$*, (8-27)

and on the left-hand side of (8-26) *up to error terms of L^∞ size $\mathcal{O}(\mathring{\varepsilon})$*. (8-28)

Proof. To prove the lemma, we make the "new bootstrap assumption" that the estimates in (8-26) hold for $t \in [0, T_{(\mathrm{Boot})})$ with the $C\mathring{\varepsilon}$-term on the right-hand side replaced by $\mathring{\varepsilon}^{1/2}$, and also that (8-28) holds with $\mathcal{O}(\mathring{\varepsilon})$ replaced by $\mathring{\varepsilon}^{1/2}$. Given this new bootstrap assumption, to obtain (8-19a)–(8-25c) and (8-27), we can simply repeat[51] the proof of [52, Lemma 9.3], which relies on transport-type estimates that lose derivatives (in particular, one uses the transport equations (2-40)–(2-41) and also treats the wave equation as a derivative-losing transport equation $L\breve{X}\Psi = \cdots$ by using (13-13)). The only difference between the estimates derived in [52, Lemma 9.3] and the estimates we need to derive is that our wave equations (5-1a)–(5-1c), when weighted with a factor of μ (so that the decomposition (13-13) of $\mu\square_g$ can be employed), feature some new inhomogeneous terms compared to [52, Lemma 9.3], specifically, some of the ones depending on $(\mathcal{C}, \mathcal{D}, \Omega, S)$ and the first derivatives of (Ω, S). The key point is that our new bootstrap assumption implies that the new inhomogeneous terms are all bounded in L^∞ by $\lesssim \mathring{\varepsilon}^{1/2}$, which is compatible with the $\mathcal{O}(\mathring{\varepsilon}^{1/2})$-size bounds that one is aiming to prove; i.e., our new $\mathcal{O}(\mathring{\varepsilon}^{1/2})$-sized error terms are harmless in the context of the proof. From this logic, it follows that the estimates (8-19a)–(8-25c) and (8-27) hold for all $t \in [0, T_{(\mathrm{Boot})})$. We clarify that the estimates (8-23a) and (8-25a) were not explicitly stated in [52, Lemma 9.3]. However (8-23a) follows from commuting the transport equation (2-40) with $L\breve{X}\breve{X}$ via Lemma 8.5 and bounding the resulting algebraic expression for $LL\breve{X}\breve{X}\mu$ using the fact that the Cartesian component functions X^1, X^2, X^3 are smooth functions of the L^i and Ψ (see (2-23)), the bootstrap assumptions (6-3)–(6-7), Proposition 8.6, and the estimates in (8-19a)–(8-25c) and (8-27) besides (8-23a) and (8-25a) . Similarly, (8-25a) follows from commuting the transport equation (2-41) with $\mathcal{P}\breve{X}\breve{X}$.

To complete the proof, it only remains for us to prove (8-26) and (8-28) (with the help of the already established bounds (8-19a)–(8-25c) and (8-27)); for if $\mathring{\varepsilon}$ is sufficiently small, this yields a strict improvement of the new bootstrap assumption mentioned at the beginning of the proof, and the conclusions of the proposition then follow from a standard continuity argument. We start by noting that the bounds in (8-26) for the pure $\mathcal{F}_u$-tangential derivatives of (Ω, S) are included in the bootstrap assumptions (6-6)–(6-7), as are the bounds

$$\|\mathcal{P}^{\leq 3}\breve{X}(\Omega, S)\|_{L^\infty(\mathcal{M}_{t,u})} \lesssim \mathring{\varepsilon}. \tag{8-29}$$

Next, we use Lemma 8.5, the bootstrap assumptions (6-3)–(6-7), Proposition 8.6, the estimates (8-19a)–(8-25c) and (8-27), and the bounds (8-29) to deduce that the estimate (8-29) also holds for all permutations of the vectorfield operators on the left-hand side.

[51]We clarify that the bootstrap parameter "ε" from [52] should be identified with the quantity $\mathring{\varepsilon}^{1/2}$ in our bootstrap assumptions (6-4)–(6-8).

We next show that

$$\|\mathcal{P}^{\leq 2}\breve{X}\breve{X}(\varOmega, S)\|_{L^\infty(\mathcal{M}_{t,u})} \lesssim \mathring{\varepsilon}. \tag{8-30}$$

This estimate follows from differentiating the identities (8-4)–(8-5) with $\mathcal{P}^{\leq 2}\breve{X}$ and using the bootstrap assumptions (6-3)–(6-7), Proposition 8.6, the estimates (8-19a)–(8-25c) and (8-27), the estimate (8-29), and the analog of (8-29) for all permutations of the vectorfield operators on the left-hand side. (Notice that we can indeed prove (8-30) with a strict improvement of our new bootstrap assumptions because the terms arising from differentiating (8-4)–(8-5) by $\mathcal{P}^{\leq 2}\breve{X}$ contain at least one factor of $(\varOmega, S)$ differentiated with at most one $\breve{X}$ derivative, and such factors have already been shown to bounded in the norm $\|\cdot\|_{L^\infty(\mathcal{M}_{t,u})}$ by $\lesssim \mathring{\varepsilon}$.) Again using Lemma 8.5 to commute vectorfield derivatives, we also deduce that the estimate (8-30) also holds for all permutations of the vectorfield operators on the left-hand side.

We next show that

$$\|\breve{X}\breve{X}\breve{X}(\varOmega, S)\|_{L^\infty(\mathcal{M}_{t,u})} \lesssim \mathring{\varepsilon}. \tag{8-31}$$

This estimate follows from differentiating the identities (8-4)–(8-5) with $\breve{X}\breve{X}$ and using the bootstrap assumptions (6-3)–(6-7), Proposition 8.6, the estimates (8-19a)–(8-25c) and (8-27), the estimates (8-29)–(8-30), and the analogs of (8-29)–(8-30) for all permutations of the vectorfield operators on the left-hand sides.

Similarly, we can first prove

$$\|\mathcal{P}^{\leq 2}\breve{X}^{\leq 1}(\mathcal{C}, \mathcal{D})\|_{L^\infty(\mathcal{M}_{t,u})} \lesssim \mathring{\varepsilon} \tag{8-32}$$

and then

$$\|\breve{X}^{\leq 2}(\mathcal{C}, \mathcal{D})\|_{L^\infty(\mathcal{M}_{t,u})} \lesssim \mathring{\varepsilon} \tag{8-33}$$

(and that (8-32) holds for all permutations of the vectorfield operators on the left-hand side all permutations of the vectorfield operators on the left-hand side) by using the identities (8-6)–(8-7) and arguing as above, using in addition the bootstrap assumption (6-8) and the already proven estimates for $(\varOmega, S)$.

We have therefore established (8-26) and (8-28), which completes the proof of the proposition. □

8D. ***Sharp estimates for*** $\mu_\star$**.** Recall the definition of $\mu_\star(t)$ in Definition 2.16. In this subsection, in Propositions 8.11 and 8.12, we provide some estimates for $\mu_\star(t)$ that were proved in [52]. We will simply cite the relevant estimates, noting that their proof relies only on the L^∞ bounds for (lower-order derivatives of) the wave variables and the geometric quantities that we have already established. Moreover, we remark that these estimates capture that $\mu_\star(t)$ tends to 0 linearly, a fact that is crucial for bounding the maximum possible singularity strength of our high-order geometric energies (i.e., for controlling the blowup-rate of the energies in, for example, (6-1)).

Thanks to our bootstrap assumptions and the estimates of Proposition 8.6, the following estimates for $\mu_\star(t)$ can be proved exactly as in [52, (10.36), (10.39)]:

Proposition 8.11 (control of integrals of $\mu_\star$). *Let $M_* \in \mathbb{N}$ be the absolute constant appearing in the statements of Theorem 4.2 and Proposition 12.1 below. For $1 < b \leq 100M_*$, the quantities $\mu_\star(t, u)$ and $\mu_\star(t)$ from Definition 2.16 obey the following estimates for every $(t, u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$:*

$$\int_{t'=0}^{t'=t} \mu_\star^{-b}(t', u)\, dt' \lesssim \left(1 + \frac{1}{b-1}\right)\mu_\star^{-b+1}(t, u), \qquad \int_{t'=0}^{t'=t} \mu_\star^{-b}(t')\, dt' \lesssim \left(1 + \frac{1}{b-1}\right)\mu_\star^{-b+1}(t). \tag{8-34}$$

Moreover, for all $t \in [0, T_{(\mathrm{Boot})})$,

$$\int_{t'=0}^{t'=t} \mu_\star^{-0.9}(t', u)\, dt' \lesssim 1, \qquad \int_{t'=0}^{t'=t} \mu_\star^{-0.9}(t')\, dt' \lesssim 1. \tag{8-35}$$

Thanks to our bootstrap assumptions and the estimates of Proposition 8.6, the following "almost-monotonicity" of $\mu_\star$ can be proved as in [52, (10.23)]:

Proposition 8.12 (the approximate monotonicity of $\mu_\star$). *For* $0 \le s_1 \le s_2 < T_{(\mathrm{Boot})}$,

$$\mu_\star^{-1}(s_1) \le 2\mu_\star^{-1}(s_2).$$

8E. *L^2 estimates for the geometric quantities.* We start with a simple lemma that provides L^2 estimates for solutions to transport equations along the integral curves of L.

Lemma 8.13 (L^2 estimate for solutions to L-transport equations). *Let F and f be smooth scalar functions on* $[0, T_{(\mathrm{Boot})}) \times [0, U_0] \times \mathbb{T}^2$. *Assume that* $LF(t, u, x^2, x^3) = f(t, u, x^2, x^3)$ *with initial data* $F(0, u, x^2, x^3)$ *for every* $(t, u, x^2, x^3) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0] \times \mathbb{T}^2$. *Then the following estimate holds for every* $(t, u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$:

$$\|F\|_{L^2(\Sigma_t^u)} \le (1 + C\mathring{\epsilon}^{1/2})\|F\|_{L^2(\Sigma_0^u)} + (1 + C\mathring{\epsilon}^{1/2}) \int_{t'=0}^{t'=t} \|f\|_{L^2(\Sigma_{t'}^u)}\, dt'. \tag{8-36}$$

Proof. Thanks to our bootstrap assumptions and the estimates of Proposition 8.6, (8-36) can be proved using essentially the same arguments used in the proof of [52, Lemmas 12.2, 12.3, 13.2]. The only differences are that we have to use the bootstrap assumptions (6-3)–(6-8) in place of the similar bootstrap assumptions from [52], and that different coordinates along $\ell_{t,u}$ were used in [52] (this is irrelevant in the sense that the estimate (8-36) is independent of the coordinates on $\ell_{t,u}$). We clarify that the bootstrap parameter "ε" from [52] should be identified with the quantity $\mathring{\epsilon}^{1/2}$ in our bootstrap assumptions (6-3)–(6-8). □

Proposition 8.14 (easy L^2 estimates for the acoustical geometry). *For* $1 \le N \le N_{\mathrm{top}}$, *the following estimates hold for all* $t \in [0, T_{(\mathrm{Boot})})$:

$$\|\mathcal{P}^{[2,N]}\mu\|_{L^2(\Sigma_t)}^2,\ \|\mathcal{P}^{[1,N]}L^i\|_{L^2(\Sigma_t)}^2 \lesssim \mathring{\epsilon} \max\{1, \mu_\star^{-2M_*+2N_{\mathrm{top}}-2N+2.8}(t)\}.$$

Proof. In an identical manner as [52, Lemma 14.3], based on the transport equations (2-40)–(2-41) and (8-36), we obtain

$$\|\mathcal{P}^{[2,N]}\mu\|_{L^2(\Sigma_t)},\ \|\mathcal{P}^{[1,N]}L^i\|_{L^2(\Sigma_t)} \lesssim \mathring{\epsilon} + \int_{s=0}^{s=t} \frac{\mathbb{W}_{[1,N]}^{1/2}(s)}{\mu_\star^{1/2}(s)}\, ds.$$

(Recall our notation in Definition 3.4, (3-2e) and Definition 3.5.) Also using our bootstrap assumptions (6-1) and (6-2) and Proposition 8.11, we arrive at the desired conclusion. □

In the next proposition, with the help of Proposition 8.14, we derive L^2 estimates for commutators.

Proposition 8.15 (L^2 estimates for commutator terms). *Let ϕ be a scalar function. For $1 \le N \le N_{\text{top}}$, the following estimates hold for all $(t,u) \in [0, T_{(\text{Boot})}) \times [0, U_0]$:*

$$\begin{aligned}\|[L, \mathcal{P}^N]\phi\|^2_{L^2(\Sigma_t^u)},\ \|[\breve{X}, \mathcal{P}^N]\phi\|^2_{L^2(\Sigma_t^u)},\ \|[\mu B, \mathcal{P}^N]\phi\|^2_{L^2(\Sigma_t^u)} \\ \lesssim \|\mathcal{P}^{[1,N]}\phi\|^2_{L^2(\Sigma_t^u)} + \mathring{\epsilon}\max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+2.8}(t)\}\|\mathcal{P}^{[1,N_{\text{top}}-M_*-5]}\phi\|^2_{L^\infty(\Sigma_t^u)}.\end{aligned} \tag{8-37}$$

Moreover, we also have

$$\|\mathcal{P}^N \breve{X}\Psi\|^2_{L^2(\Sigma_t^u)} \lesssim \mathring{\epsilon}\max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+1.8}(t)\}. \tag{8-38}$$

Proof. Recall the pointwise estimate (8-11). For each of the sums in (8-11), either $N_2 > N_1$, in which case by (6-5) and Proposition 8.6, we have $|\mathcal{P}^{[2,N_1]}(\mu, L^i, \Psi)|,\ |\mathcal{P}^{[2,N_1]}\breve{X}\Psi| \lesssim 1$; or else $N_2 \le N_1$, in which case (since $N \le N_{\text{top}}$) $|\mathcal{P}^{[1,N_2]}\phi| \lesssim |\mathcal{P}^{[1,N_{\text{top}}-M_*-5]}\phi|$. Hence,

$$\begin{aligned}&|[L, \mathcal{P}^N]\phi|,\ |[\breve{X}, \mathcal{P}^N]\phi| \\ &\quad\lesssim |\mathcal{P}^{[1,N]}\phi| + \big\{|\mathcal{P}^{[2,N]}(\mu, L^i, \Psi)| + |\mathcal{P}^{[2,N-1]}\breve{X}\Psi|\big\}|\mathcal{P}^{[1,N_{\text{top}}-M_*-5]}\phi| \\ &\quad\lesssim |\mathcal{P}^{[1,N]}\phi| + \big\{|\mathcal{P}^{[2,N]}(\mu, L^i, \Psi)| + |\breve{X}\mathcal{P}^{[2,N-1]}\Psi| + |[\breve{X}, \mathcal{P}^{[2,N-1]}]\Psi|\big\}|\mathcal{P}^{[1,N_{\text{top}}-M_*-5]}\phi|.\end{aligned} \tag{8-39}$$

We first apply (8-39) to $\phi = \Psi$. Taking the $L^2(\Sigma_t^u)$ norm and introducing an induction argument in N which uses (6-1)–(6-5) and Proposition 8.14, we obtain

$$\|[\breve{X}, \mathcal{P}^N]\Psi\|^2_{L^2(\Sigma_t^u)} \lesssim \mathring{\epsilon}\max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+2.8}(t)\}. \tag{8-40}$$

Taking the $L^2(\Sigma_t^u)$ norm in (8-39), plugging in the estimate (8-40), and using (6-1), (6-2), and Proposition 8.14, we deduce the desired estimates in (8-37) for $[L, \mathcal{P}^N]\phi$ and $[\breve{X}, \mathcal{P}^N]\phi$.

To obtain the $[\mu B, \mathcal{P}^N]\phi$ estimate in (8-37), we first note that, by (2-23),

$$[\mu B, \mathcal{P}^N]\phi = \mu[L, \mathcal{P}^N]\phi + [\mu, \mathcal{P}^N]L\phi + [\breve{X}, \mathcal{P}^N]\phi.$$

The first and last terms can be controlled by combining the commutator estimates we just established with the simple bound $\|\mu\|_{L^\infty(\Sigma_t)} \lesssim 1$ from Proposition 8.6, while the second term can be controlled simply using the product rule and Propositions 8.6 and 8.14. We have therefore established (8-37).

Finally, we have (8-38) thanks to (6-1), (6-2) and (8-40). □

9. Transport estimates for the specific vorticity and the entropy gradient

We continue to work under the assumptions of Theorem 6.3.

In this section, we use the transport equations (5-2a) and (5-2c) to bound $\mathcal{P}^N\Omega$ and $\mathcal{P}^N S$ for $N \le N_{\text{top}}$. We clarify that the "true" top-order estimates for the vorticity and entropy are found in Section 11; those estimates are more involved and rely on the modified fluid variables as well as elliptic estimates.

We will start by deriving energy estimates for general transport equations (which will also be useful in the next section). In particular, this will reduce the derivation of the energy estimates for $\mathcal{P}^N\Omega$ and $\mathcal{P}^N S$ to controlling the inhomogeneous terms in the transport equations and their derivatives, which we will carry out in Section 9B. The final estimates for $\mathcal{P}^N\Omega$ and $\mathcal{P}^N S$ are located in Section 9C.

9A. ***Estimates for general transport equations.***

Proposition 9.1 (L^2 estimates for solutions to B-transport equations). *Let ϕ be a scalar function satisfying*

$$\upmu B\phi = \mathfrak{F},$$

with both ϕ and $\mathfrak{F}$ being compactly supported in $[0, U_0] \times \mathbb{T}^2$ for every $t \in [0, T_{(\mathrm{Boot})})$.

Then the following estimate holds for every $(t, u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$:

$$\sup_{t' \in [0,t)} \|\sqrt{\upmu}\phi\|^2_{L^2(\Sigma^u_{t'})} + \sup_{u' \in [0,u)} \|\phi\|^2_{L^2(\mathcal{F}^t_{u'})} \lesssim \|\sqrt{\upmu}\phi\|^2_{L^2(\Sigma^u_0)} + \|\mathfrak{F}\|^2_{L^2(\mathcal{M}_{t,u})}.$$

Proof. In an identical manner as [36, Proposition 3.5], we have, for any $(t', u') \in [0, t) \times [0, u)$, the identity

$$\int_{\Sigma^{u'}_{t'}} \upmu\phi^2 \, d\underline{\varpi} + \int_{\mathcal{F}^{t'}_{u'}} \phi^2 \, d\overline{\varpi} = \int_{\Sigma^{u'}_0} \upmu\phi^2 \, d\underline{\varpi} + \underbrace{\int_{\mathcal{F}^{t'}_0} \phi^2 \, d\overline{\varpi}}_{0 \text{ by support assumptions}} + \int_{\mathcal{M}_{t',u'}} \{2\phi\mathfrak{F} + (L\upmu + \upmu \operatorname{tr}_{\not g} \not k)\phi^2\} \, d\varpi. \tag{9-1}$$

Using (2-38c), (2-40), Lemma 2.32, (6-3)–(6-5), and Propositions 8.6 and 8.7, we have $|L\upmu|$, $|\upmu \operatorname{tr}_{\not g} \not k| \lesssim 1$. Thus, applying also the Cauchy–Schwarz inequality to the $2\phi\mathfrak{F}$ term, we have

$$\sup_{t' \in [0,t)} \|\sqrt{\upmu}\phi\|^2_{L^2(\Sigma^u_{t'})} + \sup_{u' \in [0,u)} \|\phi\|^2_{L^2(\mathcal{F}^t_{u'})} \lesssim \|\sqrt{\upmu}\phi\|^2_{L^2(\Sigma^u_0)} + \int_{u'=0}^{u'=u} \|\phi\|^2_{L^2(\mathcal{F}^t_{u'})} \, du' + \|\mathfrak{F}\|^2_{L^2(\mathcal{M}_{t,u})}.$$

The conclusion follows from applying Grönwall's inequality in u. □

Proposition 9.2 (higher-order L^2 estimates for solutions to transport equations). *Let ϕ be a scalar function satisfying*

$$\upmu B\phi = \mathfrak{F},$$

with both ϕ and $\mathfrak{F}$ being compactly supported in $[0, U_0] \times \mathbb{T}^2$ for every $t \in [0, T_{(\mathrm{Boot})})$.

Then the following estimate holds for every $(t, u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$ and $0 \le N \le N_{\mathrm{top}}$:

$$\begin{aligned} &\sup_{t' \in [0,t)} \|\sqrt{\upmu}\mathcal{P}^{\le N}\phi\|^2_{L^2(\Sigma^u_{t'})} + \sup_{u' \in [0,u)} \|\mathcal{P}^{\le N}\phi\|^2_{L^2(\mathcal{F}^t_{u'})} \\ &\quad \lesssim \|\mathcal{P}^{\le N}\phi\|^2_{L^2(\Sigma_0)} + \|\mathcal{P}^{\le N}\mathfrak{F}\|^2_{L^2(\mathcal{M}_{t,u})} + \mathring{\epsilon} \max\{1, \upmu_\star^{-2M_*+2N_{\mathrm{top}}-2N+3.8}(t)\} \|\mathcal{P}^{[1,N_{\mathrm{top}}-M_*-5]}\phi\|^2_{L^\infty(\mathcal{M}_{t,u})}. \end{aligned}$$

Proof. Take $0 \le N' \le N$. We write

$$\upmu B\mathcal{P}^{N'}\phi = \mathcal{P}^{N'}\mathfrak{F} + [\upmu B, \mathcal{P}^{N'}]\phi.$$

Therefore, by Proposition 9.1,

$$\begin{aligned} &\sup_{t' \in [0,t)} \|\sqrt{\upmu}\mathcal{P}^{N'}\phi\|^2_{L^2(\Sigma^u_{t'})} + \sup_{u' \in [0,u)} \|\mathcal{P}^{N'}\phi\|^2_{L^2(\mathcal{F}^t_{u'})} \\ &\qquad \lesssim \|\mathcal{P}^{N'}\phi\|^2_{L^2(\Sigma_0)} + \|\mathcal{P}^{N'}\mathfrak{F}\|^2_{L^2(\mathcal{M}_{t,u})} + \|[\upmu B, \mathcal{P}^{N'}]\phi\|^2_{L^2(\mathcal{M}_{t,u})}. \end{aligned} \tag{9-2}$$

Using Proposition 8.15 and then Proposition 8.11, we obtain

$$
\begin{aligned}
&\|[\upmu B, \mathcal{P}^{N'}]\phi\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\quad\lesssim \int_{t'=0}^{t'=t} \|[\upmu B, \mathcal{P}^{N'}]\phi\|^2_{L^2(\Sigma_{t'}^u)}\, dt' \\
&\quad\lesssim \|\mathcal{P}^{[1,N']}\phi\|^2_{L^2(\mathcal{M}_{t,u})} + \mathring{\epsilon}\|\mathcal{P}^{[1,N_{\mathrm{top}}-M_*-5]}\phi\|^2_{L^\infty(\mathcal{M}_{t,u})} \int_{t'=0}^{t'=t} \max\{1, \upmu_\star^{-2M_*+2N_{\mathrm{top}}-2N'+2.8}(t')\}\, dt' \\
&\quad\lesssim \int_{u'=0}^{u'=u} \|\mathcal{P}^{\le N}\phi\|^2_{L^2(\mathcal{F}_{u'}^t)}\, du' + \mathring{\epsilon}\max\{1, \upmu_\star^{-2M_*+2N_{\mathrm{top}}-2N'+3.8}(t)\}\|\mathcal{P}^{[1,N_{\mathrm{top}}-M_*-5}\phi\|^2_{L^\infty(\mathcal{M}_{t,u})}. \qquad (9\text{-}3)
\end{aligned}
$$

Plugging (9-3) into (9-2) and summing over all $0 \le N' \le N$, we obtain

$$
\begin{aligned}
&\sup_{t'\in[0,t)} \|\sqrt{\upmu}\mathcal{P}^{\le N}\phi\|^2_{L^2(\Sigma_{t'}^u)} + \sup_{u'\in[0,u)} \|\mathcal{P}^{\le N}\phi\|^2_{L^2(\mathcal{F}_{u'}^t)} \\
&\quad\lesssim \|\mathcal{P}^{\le N}\phi\|^2_{L^2(\Sigma_0)} + \|\mathcal{P}^{\le N}\mathfrak{F}\|^2_{L^2(\mathcal{M}_{t,u})} + \int_{u'=0}^{u'=u} \|\mathcal{P}^{\le N}\phi\|^2_{L^2(\mathcal{F}_{u'}^t)}\, du' \\
&\qquad\qquad + \mathring{\epsilon}\max\{1, \upmu_\star^{-2M_*+2N_{\mathrm{top}}-2N+3.8}(t)\}\|\mathcal{P}^{[1,N_{\mathrm{top}}-M_*-5}\phi\|^2_{L^\infty(\mathcal{M}_{t,u})}. \qquad (9\text{-}4)
\end{aligned}
$$

Applying Grönwall's inequality in u, we arrive at the desired estimate. □

9B. *Controlling the inhomogeneous terms.*

Proposition 9.3 (estimates tied to the inhomogeneous terms in the transport equations for $\varOmega$ and S). *For $0 \le N \le N_{\mathrm{top}}$, the following hold for every $(t, u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$:*

$$
\begin{aligned}
&\|\mathcal{P}^N(\upmu B\varOmega)\|^2_{L^2(\mathcal{M}_{t,u})} + \|\mathcal{P}^N(\upmu BS)\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\qquad\lesssim \mathring{\epsilon}^3\max\{1, \upmu_\star^{-2M_*+2N_{\mathrm{top}}-2N+2.8}(t)\} + \int_{u'=0}^{u'=u} (\mathbb{V}_{\le N}(t,u') + \mathbb{S}_{\le N}(t,u'))\, du' \qquad (9\text{-}5)
\end{aligned}
$$

and

$$
\begin{aligned}
&\|(\mathcal{P}^N\varOmega)\mathcal{P}^N(\upmu B\varOmega)\|_{L^1(\mathcal{M}_{t,u})} + \|(\mathcal{P}^N S)\mathcal{P}^N(\upmu BS)\|_{L^1(\mathcal{M}_{t,u})} \\
&\qquad\lesssim \mathring{\epsilon}^3\max\{1, \upmu_\star^{-2M_*+2N_{\mathrm{top}}-2N+2.8}(t)\} + \int_{u'=0}^{u'=u} (\mathbb{V}_{\le N}(t,u') + \mathbb{S}_{\le N}(t,u'))\, du'. \qquad (9\text{-}6)
\end{aligned}
$$

Proof. Step 1: basic pointwise estimates. We claim that the derivatives of the $\upmu$-weighted inhomogeneous terms $\upmu\mathfrak{L}^i_{(\varOmega)}$ and $\upmu\mathfrak{L}^i_{(S)}$, which are defined respectively in (5-7d) and (5-7e), obey the following pointwise bounds:

$$
\begin{aligned}
&|\mathcal{P}^N(\upmu\mathfrak{L}^i_{(\varOmega)})| + |\mathcal{P}^N(\upmu\mathfrak{L}^i_{(S)})| \\
&\qquad\lesssim \underbrace{|\mathcal{P}^{\le N}(\varOmega, S)|}_{\doteq I} + \underbrace{\mathring{\epsilon}(|\upmu\mathcal{P}^{[2,N+1]}\Psi| + |\mathcal{P}^{[1,N]}\breve{X}\Psi|)}_{\doteq II} + \underbrace{\mathring{\epsilon}|\mathcal{P}^{[2,N]}(\upmu, L^i, \Psi)|}_{\doteq III}. \qquad (9\text{-}7)
\end{aligned}
$$

Since this is the first instance of these kind of estimates (and we will derive similar estimates later), we give some details on how to obtain (9-7).

(1) By Lemma 2.22 and the fact that the Cartesian component functions X^1, X^2, X^3 are smooth functions of the L^i and Ψ (see (2-23)), the weighted Cartesian coordinate vectorfield $\mu\partial_i$ and the transport vectorfield μB can be decomposed regularly (i.e., with coefficients being smooth functions of μ, L^i and Ψ) in terms of $\breve{X}$, μY, μZ and μL.

(2) Therefore, $\mathcal{P}^N(\mu\mathfrak{L}^i_{(\Omega)})$ and $\mathcal{P}^N(\mu\mathfrak{L}^i_{(S)})$ can be bounded as follows:

$$
\begin{aligned}
&|\mathcal{P}^N(\mu\mathfrak{L}^i_{(\Omega)})|+|\mathcal{P}^N(\mu\mathfrak{L}^i_{(S)})| \\
&\quad\lesssim \sum_{k=0}^{N} \sum_{N_1+\cdots+N_k+n_1+n_2=N} (1+|\mathcal{P}^{N_1}(\mu, L^i, \Psi)|)\cdots(1+|\mathcal{P}^{N_k}(\mu, L^i, \Psi)|)\times|\mathcal{P}^{n_1}(\Omega, S)|\times|\mathcal{P}^{n_2}(\mu\mathcal{P}\Psi, \breve{X}\Psi)| \\
&\quad\doteq \sum_{k=0}^{N} \mathrm{Error}_{N_1,\dots,N_k,n_1,n_2}.
\end{aligned}
\tag{9-8}
$$

We now bound the right-hand side of (9-8).

(3) If $N_1, \dots, N_k \le N_{\mathrm{top}} - M_* - 5$ and $n_2 \le N_{\mathrm{top}} - M_* - 5$, we bound the terms $(1+|\mathcal{P}^{N_j}(\mu, L^i, \Psi)|)$ (for all $j = 1, \dots, k$) and $|\mathcal{P}^{n_2}(\mu\mathcal{P}\Psi, \breve{X}\Psi)|$ in L^∞ by $\lesssim 1$ using (6-3)–(6-5) and Proposition 8.6, which yields

$$\mathrm{Error}_{N_1,\dots,N_k,n_1,n_2} \lesssim |\mathcal{P}^{\le N}(\Omega, S)|. \tag{9-9}$$

(4) If $N_j > N_{\mathrm{top}} - M_* - 5$ for some[52] j, then all the terms $(1+|\mathcal{P}^{N_{j'}}(\mu, L^i, \Psi)|)$, when $j' \ne j$, and $|\mathcal{P}^{n_2}(\mu\mathcal{P}\Psi, \breve{X}\Psi)|$ can be bounded in L^∞ by $\lesssim 1$ using (6-3)–(6-5) and Proposition 8.6. Moreover, since it must also hold that $n_1 \le N_{\mathrm{top}} - M_* - 5$, we also have $|\mathcal{P}^{n_1}(\Omega, S)| \lesssim \mathring{\epsilon}$ by the bootstrap assumptions (6-6) and (6-7). Hence, we have

$$
\begin{aligned}
\mathrm{Error}_{N_1,\dots,N_k,n_1,n_2} &\lesssim (1+|\mathcal{P}^{[2,N_j]}(\mu,\ L^i,\ \Psi)|)|\mathcal{P}^{\le n_1}(\Omega, S)| \\
&\lesssim |\mathcal{P}^{\le N}(\Omega, S)| + \mathring{\epsilon}|\mathcal{P}^{[2,N]}(\mu,\ L^i,\ \Psi)|.
\end{aligned}
\tag{9-10}
$$

(5) When $n_2 > N_{\mathrm{top}} - M_* - 5$, we can argue as above to see that $(1+|\mathcal{P}^{N_j}(\mu, L^i, \Psi)|) \lesssim 1$ for all j, and $|\mathcal{P}^{n_1}(\Omega, S)| \lesssim \mathring{\epsilon}$. Notice further that since $n_2 > N_{\mathrm{top}} - M_* - 5$, by (6-5) and Proposition 8.6 we have

$$|\mathcal{P}^{n_2}(\mu\mathcal{P}\Psi)| \lesssim |\mu\mathcal{P}^{[2,n_2+1]}\Psi| + |\mathcal{P}^{[2,n_2]}\Psi| + |\mathcal{P}^{[2,n_2]}\mu|.$$

Hence, we have

$$\mathrm{Error}_{N_1,\dots,N_k,n_1,n_2} \lesssim \mathring{\epsilon}(\mu|\mathcal{P}^{N+1}\Psi| + |\mathcal{P}^{[2,N]}\Psi| + |\mathcal{P}^{[1,N]}\breve{X}\Psi| + |\mathcal{P}^{[2,n_2]}\mu|). \tag{9-11}$$

Finally, it is easy to check that (9-9)–(9-11) are all bounded above by the right-hand side of (9-7).

<u>Step 2</u>: proof of (9-5). To derive (9-5), we control each term in (9-7) in the $L^2(\mathcal{M}_{t,u})$ norm.

We begin with the term I in (9-7), which we estimate using the definition of the $\mathbb{V}_{\le N}$ and $\mathbb{S}_{\le N}$ energies (see Section 3B2):

$$\|\mathcal{P}^{\le N}(\Omega, S)\|^2_{L^2(\mathcal{M}_{t,u})} \lesssim \int_{u'=0}^{u'=u} [\mathbb{V}_{\le N} + \mathbb{S}_{\le N}](t, u')\, du'. \tag{9-12}$$

[52] Note that there can be at most one such j.

We control term II in (9-7) by the $\mathbb{E}_{[1,N]}$ norm, and use the bootstrap assumptions (6-1), (6-2), the bound (8-38), and Proposition 8.11 to obtain

$$\begin{aligned}
&\mathring{\epsilon}^2 \|\mu \mathcal{P}^{[2,N+1]}\Psi\|^2_{L^2(\mathcal{M}_{t,u})} + \mathring{\epsilon}^2 \|\mathcal{P}^{[1,N]}\breve{X}\Psi\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\qquad \lesssim \mathring{\epsilon}^2 \int_{t'=0}^{t'=t} \big[\mathbb{E}_{[1,N]}(t') + \mathring{\epsilon}^2 \max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+1.8}(t')\}\big]\, dt' \\
&\qquad \lesssim \mathring{\epsilon}^3 \max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+2.8}(t)\}.
\end{aligned} \tag{9-13}$$

Finally, for the term III, we use the control for $\mathbb{K}_{[1,N-1]}$ and $\mathbb{F}_{[1,N-1]}$ provided by the bootstrap assumptions (6-1) and (6-2), the bounds in Proposition 8.14, and Proposition 8.11 to obtain

$$\begin{aligned}
&\mathring{\epsilon}^2 \|\mathcal{P}^{[2,N]}(\mu, L^i, \Psi)\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\quad \lesssim \mathring{\epsilon}^2 \mathbb{K}_{[1,N-1]}(t,u) + \mathring{\epsilon}^2 \int_{u'=0}^{u'=u} \mathbb{F}_{[1,N-1]}(t,u')\, du' + \mathring{\epsilon}^3 \int_{t'=0}^{t'=t} \max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+2.8}(t')\}\, dt' \\
&\quad \lesssim \mathring{\epsilon}^3 \max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+3.8}(t)\}.
\end{aligned} \tag{9-14}$$

Combining (9-7) with (9-12)–(9-14), we arrive at the desired bound (9-5).

Step 3: proof of (9-6). The estimate (9-6) follows as a simple consequence of the already obtained bound (9-5) and the Cauchy–Schwarz inequality. □

9C. *Putting everything together.*

Proposition 9.4 (estimates for the specific vorticity and entropy gradient). *For $0 \le N \le N_{\text{top}}$, the following holds for all $t \in [0, T_{(\text{Boot})}) \times [0, U_0]$:*

$$\mathbb{V}_N(t,u) + \mathbb{S}_N(t,u) \lesssim \mathring{\epsilon}^3 \max\{1, \ \mu_\star^{-2M_*+2N_{\text{top}}-2N+2.8}(t)\}.$$

Proof. Using Proposition 9.2 for $\phi = \Omega^i, S^i$, the initial data size assumptions in (4-5), the bootstrap assumptions (6-6)–(6-7), and the inhomogeneous term estimates in Proposition 9.3 for the terms on right-hand sides of the transport equations (5-2a) and (5-2c), we deduce

$$\mathbb{V}_{\le N}(t,u) + \mathbb{S}_{\le N}(t,u) \lesssim \mathring{\epsilon}^3 \max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+2.8}(t)\} + \int_{u'=0}^{u'=u} (\mathbb{V}_{\le N}(t,u') + \mathbb{S}_{\le N}(t,u'))\, du'.$$

The desired estimate now follows from applying Grönwall's inequality in u. □

10. Lower-order transport estimates for the modified fluid variables

We continue to work under the assumptions of Theorem 6.3.

In this section, we derive the energy estimates for the modified fluid variables $\mathcal{C}$ and $\mathcal{D}$ *except for the top-order*. (We will derive the top-order estimates in the next section.) Thanks to Proposition 9.2, to obtain the desired estimates, it remains only for us to bound the inhomogeneous terms in the transport equations (5-3b) and (5-4a). Before we estimate the inhomogeneous terms, we will first control the $\breve{X}$ derivative of Ω and S in Section 10A, and give general bounds for null forms in Section 10B. (The null forms will also be useful later on, in Section 12.) We will combine these results to control the inhomogeneous terms in Section 10C. We provide the final estimate in Section 10D.

10A. ***Preliminaries.*** A priori, the norms $\mathbb{V}_N$ and $\mathbb{S}_N$ do not control the $\breve{X}$ derivatives of Ω or S. Nonetheless, we can obtain such control in terms of the norms $\mathbb{V}_N$ and $\mathbb{S}_N$ by using the transport equations (5-2a) and (5-2c).

Proposition 10.1 (L^2 control of the transversal derivatives of the Ω and S). *For* $1 \le N \le N_{\text{top}}$, *the following holds for all* $(t,u) \in [0, T_{(\text{Boot})}) \times [0, U_0]$:

$$\|\mathcal{P}^{N-1}\breve{X}(\Omega, S)\|^2_{L^2(\mathcal{M}_{t,u})} \lesssim \mathring{\epsilon}^3 \max\{1, \upmu_\star^{-2M_*+2N_{\text{top}}-2N+2.8}(t)\}.$$

Proof. Recalling (2-23), we have

$$\mathcal{P}^{N-1}\breve{X}\Omega = \mathcal{P}^{N-1}(\upmu B\Omega) - \mathcal{P}^{N-1}(\upmu L\Omega), \quad \mathcal{P}^{N-1}\breve{X}S = \mathcal{P}^{N-1}(\upmu BS) - \mathcal{P}^{N-1}(\upmu LS). \tag{10-1}$$

The terms $\mathcal{P}^{N-1}(\upmu B\Omega)$ and $\mathcal{P}^{N-1}(\upmu BS)$ can be bounded as follows using (9-5) and Proposition 9.4:

$$\|\mathcal{P}^{N-1}(\upmu B\Omega)\|^2_{L^2(\mathcal{M}_{t,u})} + \|\mathcal{P}^{N-1}(\upmu BS)\|^2_{L^2(\mathcal{M}_{t,u})} \lesssim \mathring{\epsilon}^3 \max\{1, \upmu_\star^{-2M_*+2N_{\text{top}}-2N+2.8}(t)\}. \tag{10-2}$$

By (6-6), (6-7), Propositions 8.6, 8.12, 8.14, and 9.4, we have

$$\begin{aligned}
&\|\mathcal{P}^{N-1}(\upmu L\Omega)\|^2_{L^2(\mathcal{M}_{t,u})} + \|\mathcal{P}^{N-1}(\upmu LS)\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\qquad \lesssim \|\mathcal{P}^{\le N}(\Omega, S)\|^2_{L^2(\mathcal{M}_{t,u})} + \mathring{\epsilon}^2\|\mathcal{P}^{[2,N-1]}\upmu\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\qquad \lesssim \int_0^u [\mathbb{V}_{\le N} + \mathbb{S}_{\le N}](t,u')\,du' + \mathring{\epsilon}^3 \int_{t'=0}^{t'=t} \max\{1, \upmu_\star^{-2M_*+2N_{\text{top}}-2N+4.8}(t')\}\,dt' \\
&\qquad \lesssim \mathring{\epsilon}^3 \max\{1, \upmu_\star^{-2M_*+2N_{\text{top}}-2N+2.8}(t)\}.
\end{aligned} \tag{10-3}$$

Therefore, combining (10-1)–(10-3), we obtain the desired conclusion. ☐

10B. ***General estimates for null forms.***

Lemma 10.2 (pointwise estimates for null forms). *Suppose*

(1) $\mathcal{Q}(\partial\phi^{(1)}, \partial\phi^{(2)})$ *is a* g*-null form, as in Definition 8.1*; *and*

(2) $\phi^{(1)}$ *and* $\phi^{(2)}$ *obey the following* L^∞ *estimates for some* $\mathfrak{d}^{(1,1)} \gtrsim \mathfrak{d}^{(1,2)}$, $\mathfrak{d}^{(2,1)} \gtrsim \mathfrak{d}^{(2,2)}$ *for all* $t \in [0, T_{(\text{Boot})})$:

$$\begin{aligned}
\|\mathcal{P}^{\le N_{\text{top}}-M_*-5}\breve{X}\phi^{(1)}\|_{L^\infty(\Sigma_t)} &\le \mathfrak{d}^{(1,1)}, \\
\|\mathcal{P}^{[1,N_{\text{top}}-M_*-5]}\phi^{(1)}\|_{L^\infty(\Sigma_t)} &\le \mathfrak{d}^{(1,2)}, \\
\|\mathcal{P}^{\le N_{\text{top}}-M_*-5}\breve{X}\phi^{(2)}\|_{L^\infty(\Sigma_t)} &\le \mathfrak{d}^{(2,1)}, \\
\|\mathcal{P}^{[1,N_{\text{top}}-M_*-5]}\phi^{(2)}\|_{L^\infty(\Sigma_t)} &\le \mathfrak{d}^{(2,2)}.
\end{aligned} \tag{10-4}$$

Then, for any $0 \le N \le N_{\text{top}}$, *the following pointwise estimate holds on* $[0, T_{(\text{Boot})}) \times \Sigma$:

$$\begin{aligned}
&|\mathcal{P}^N[\upmu\mathcal{Q}(\partial\phi^{(1)}, \partial\phi^{(2)})]| \\
&\quad \lesssim \mathfrak{d}^{(2,1)}|\mathcal{P}^{[1,N+1]}\phi^{(1)}| + \mathfrak{d}^{(2,2)}|\mathcal{P}^{[1,N]}\breve{X}\phi^{(1)}| + \mathfrak{d}^{(1,1)}|\mathcal{P}^{[1,N+1]}\phi^{(2)}| + \mathfrak{d}^{(1,2)}|\mathcal{P}^{[1,N]}\breve{X}\phi^{(2)}| \\
&\qquad + \max\{\mathfrak{d}^{(1,1)}\mathfrak{d}^{(2,2)}, \mathfrak{d}^{(1,2)}\mathfrak{d}^{(2,1)}\}|\mathcal{P}^{[2,N]}(\upmu, L^i, \Psi)|,
\end{aligned} \tag{10-5}$$

and, *for any* $1 \leq N \leq N_{\text{top}}$, *we have*

$$
\begin{aligned}
&|\mathcal{P}^{[1,N]}[\upmu \mathcal{Q}(\partial\phi^{(1)}, \partial\phi^{(2)})]| \\
&\quad \lesssim \mathfrak{d}^{(2,1)}|\mathcal{P}^{[2,N+1]}\phi^{(1)}| + \mathfrak{d}^{(2,2)}|\mathcal{P}^{[1,N]}\breve{X}\phi^{(1)}| + \mathfrak{d}^{(1,1)}|\mathcal{P}^{[2,N+1]}\phi^{(2)}| + \mathfrak{d}^{(1,2)}|\mathcal{P}^{[1,N]}\breve{X}\phi^{(2)}| \\
&\qquad + \underbrace{\mathring{\epsilon}^{1/2}(\mathfrak{d}^{(2,1)}|\mathcal{P}\phi^{(1)}| + \mathfrak{d}^{(1,1)}|\mathcal{P}\phi^{(2)}|)}_{\doteq\mathscr{A}} + \underbrace{\mathfrak{d}^{(2,2)}|\mathcal{P}\phi^{(1)}| + \mathfrak{d}^{(1,2)}|\mathcal{P}\phi^{(2)}|}_{\doteq\mathscr{B}} \\
&\qquad + \max\{\mathfrak{d}^{(1,1)}\mathfrak{d}^{(2,2)},\ \mathfrak{d}^{(1,2)}\mathfrak{d}^{(2,1)}\}|\mathcal{P}^{[2,N]}(\upmu, L^i, \Psi)|. \qquad (10\text{-}6)
\end{aligned}
$$

Proof. Throughout this proof, $\mathrm{f}(\cdot)$ denotes a smooth function of its arguments that is free to vary from line to line. By (8-3), we need to control

$$
\underbrace{\mathcal{P}^N[\mathrm{f}(L^i, \Psi)\upmu(\mathcal{P}\phi^{(1)})(\mathcal{P}\phi^{(2)})]}_{\doteq I}, \quad \underbrace{\mathcal{P}^N[\mathrm{f}(L^i, \Psi)(\mathcal{P}\phi^{(1)})(\breve{X}\phi^{(2)})]}_{\doteq II}, \quad \underbrace{\mathcal{P}^N[\mathrm{f}(L^i, \Psi)(\breve{X}\phi^{(1)})(\mathcal{P}\phi^{(2)})]}_{\doteq III}.
$$

We first prove (10-5). Consider term II. Arguing as in the proof of (9-7) and then using (10-4), we obtain

$$
\begin{aligned}
&|\mathcal{P}^N[\mathrm{f}(L^i, \Psi, \upmu)(\mathcal{P}\phi^{(1)})(\breve{X}\phi^{(2)})]| \\
&\quad \lesssim |\mathcal{P}^{[1,N_{\text{top}}-M_*-5]}\phi^{(1)}||\mathcal{P}^{[1,N]}\breve{X}\phi^{(2)}| + |\mathcal{P}^{[1,N+1]}\phi^{(1)}||\mathcal{P}^{\leq N_{\text{top}}-M_*-5}\breve{X}\phi^{(2)}| \\
&\qquad + |\mathcal{P}^{\leq N_{\text{top}}-M_*-5}\phi^{(1)}||\mathcal{P}^{\leq N_{\text{top}}-M_*-5}\breve{X}\phi^{(2)}||\mathcal{P}^{[2,N]}(\upmu, L^i, \Psi)| \\
&\quad \lesssim \mathfrak{d}^{(1,2)}|\mathcal{P}^{[1,N]}\breve{X}\phi^{(2)}| + \mathfrak{d}^{(2,1)}|\mathcal{P}^{[1,N+1]}\phi^{(1)}| + \mathfrak{d}^{(1,2)}\mathfrak{d}^{(2,1)}|\mathcal{P}^{[2,N]}(\upmu, L^i, \Psi)|,
\end{aligned}
$$

which is bounded from above by the right-hand side of (10-5).

Next, we observe that the term III can be handled just like term II, after we interchange the roles of $\phi^{(1)}$ and $\phi^{(2)}$. Moreover, the term I is even easier to handle because $\mathfrak{d}^{(1,1)} \gtrsim \mathfrak{d}^{(1,2)}$ and $\mathfrak{d}^{(2,1)} \gtrsim \mathfrak{d}^{(2,2)}$.

We finally turn to the proof of (10-6), in which we need to show an improvement compared to (10-5) using the fact that on the left-hand side of the estimate, the $\upmu$-weighted null form is differentiated by *at least one* $\mathcal{P}$. More precisely, we need to improve $\mathfrak{d}^{(1,1)}|\mathcal{P}^{[1,N+1]}\phi^{(2)}|$ and $\mathfrak{d}^{(1,2)}|\mathcal{P}^{[1,N]}\breve{X}\phi^{(2)}|$ to $\mathfrak{d}^{(1,1)}|\mathcal{P}^{[2,N+1]}\phi^{(2)}|$ and $\mathfrak{d}^{(1,2)}|\mathcal{P}^{[2,N]}\breve{X}\phi^{(2)}|$, at the expense of incurring terms of the type $\mathscr{A}$ and $\mathscr{B}$ in (10-6).

It is straightforward to use the arguments given in the previous paragraph to confirm that if $N \geq 2$, then $\mathfrak{d}^{(1,1)}|\mathcal{P}^{[1,N+1]}\phi^{(2)}|$ and $\mathfrak{d}^{(1,2)}|\mathcal{P}^{[1,N]}\breve{X}\phi^{(2)}|$ on the right-hand side of (10-5) can be replaced by $\mathfrak{d}^{(1,1)}|\mathcal{P}^{[2,N+1]}\phi^{(2)}|$ and $\mathfrak{d}^{(1,2)}|\mathcal{P}^{[2,N]}\breve{X}\phi^{(2)}|$. We are thus only concerned with the following terms in the case when $N = 1$:

$$
\underbrace{[\mathcal{P}(\mathrm{f}(L^i, \Psi)\upmu)](\mathcal{P}\phi^{(1)})(\mathcal{P}\phi^{(2)})}_{\doteq I'}, \quad \underbrace{[\mathcal{P}\mathrm{f}(L^i, \Psi)](\mathcal{P}\phi^{(1)})(\breve{X}\phi^{(2)})}_{\doteq II'}, \quad \underbrace{[\mathcal{P}\mathrm{f}(L^i, \Psi)](\breve{X}\phi^{(1)})(\mathcal{P}\phi^{(2)})}_{\doteq III'}.
$$

Next, we observe that for the terms II' and III', when the $\mathcal{P}$ derivative falls on $\mathrm{f}(L^i, \Psi)$, (6-5) and Proposition 8.6 yield a smallness factor of $\mathring{\epsilon}^{1/2}$. Thus, II' and III' can be bounded by $\mathscr{A}$. Finally, to handle the term I', we can control either $\mathcal{P}\phi^{(1)}$ or $\mathcal{P}\phi^{(2)}$ in L^∞, which allows us to bound I' by $\mathscr{B}$. $\square$

10C. ***Estimates of the inhomogeneous terms in the transport equations for $\mathcal{C}$ and $\mathcal{D}$.***

Proposition 10.3 (below-top-order estimates for the main inhomogeneous terms in the transport equations for the modified fluid variables). *For*[53] $0 \le N \le N_{\mathrm{top}} - 1$, *the main terms* $\mathfrak{M} \in \{\mathfrak{M}^i_{(\mathcal{C})}, \mathfrak{M}_{(\mathcal{D})}\}$ (*see* (5-5a)–(5-5b)) *can be estimated as follows for every* $(t, u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$:

$$\|\mathcal{P}^N(\upmu\mathfrak{M})\|^2_{L^2(\mathcal{M}_{t,u})} \lesssim \mathring{\epsilon}^3 \max\{1, \upmu_\star^{-2M_*+2N_{\mathrm{top}}-2N+0.8}(t)\}. \tag{10-7}$$

Proof. Note that $\mathfrak{M}^i_{(\mathcal{C})}$ consists of null forms (see Definition 8.1) $\mathcal{Q}(\partial\Psi, \partial\Omega)$, $\mathcal{Q}(\partial\Psi, \partial S)$. Therefore, by Lemma 10.2 (with $\phi^{(1)} = \Omega^i, S^i$, $\phi^{(2)} = \Psi$, $\mathfrak{d}^{(1,1)} = \mathfrak{d}^{(1,2)} \doteq \mathring{\epsilon}$, $\mathfrak{d}^{(2,2)} \doteq \mathring{\epsilon}^{1/2}$, and $\mathfrak{d}^{(2,1)} = \mathcal{O}(1)$ by virtue of the bootstrap assumptions (6-3)–(6-7)),[54] we have

$$|\mathcal{P}^N(\upmu\mathfrak{M}^i_{(\mathcal{C})})| \lesssim \mathring{\epsilon}^{3/2} + \underbrace{|\mathcal{P}^{\le N+1}(\Omega, S)| + |\mathcal{P}^{\le N}\breve{X}(\Omega, S)|}_{\doteq I} + \underbrace{\mathring{\epsilon}(|\mathcal{P}^{[2,N+1]}\Psi| + |\mathcal{P}^{[1,N]}\breve{X}\Psi|)}_{\doteq II} + \underbrace{\mathring{\epsilon}|\mathcal{P}^{[2,N]}(\upmu, L^i, \Psi)|}_{\doteq III}. \tag{10-8}$$

We recall the expression for $\mathfrak{M}_{(\mathcal{D})}$ given by (5-5b). The term $2\exp(-2\uprho)\{(\partial_a v^a)\partial_b S^b - (\partial_a v^b)\partial_b S^a\}$ is a null form of type $\mathcal{Q}(\partial\Psi, \partial S)$. Thus, using the same arguments we gave when handling $\mathfrak{M}^i_{(\mathcal{C})}$, we can pointwise bound its $\mathcal{P}^N(\upmu\,\cdot)$ derivatives by the right-hand side of (10-8).

Moreover, using the same arguments given below (9-7), we see that the $\mathcal{P}^N$ derivatives of the term $\upmu\exp(-\uprho)\delta_{ab}(\mathrm{curl}\,\Omega)^a S^b$ can be pointwise bounded by the right-hand side of (10-8). From now on, it therefore suffices to consider the terms on the right-hand side of (10-8).

The term I can be controlled using Propositions 9.4 and 10.1 so that

$$\|\mathcal{P}^{\le N+1}(\Omega, S)\|^2_{L^2(\mathcal{M}_{t,u})} + \|\mathcal{P}^{\le N}\breve{X}(\Omega, S)\|^2_{L^2(\mathcal{M}_{t,u})} \lesssim \mathring{\epsilon}^3 \max\{1, \upmu_\star^{-2M_*+2N_{\mathrm{top}}-2N+0.8}(t)\}. \tag{10-9}$$

For the term II in (10-8), we use the bootstrap assumptions (6-1), (6-2), and (6-5) and the estimates of Propositions 8.12 and 8.15 to obtain

$$\begin{aligned}
&\mathring{\epsilon}^2\|\mathcal{P}^{[2,N+1]}\Psi\|^2_{L^2(\mathcal{M}_{t,u})} + \mathring{\epsilon}^2\|\mathcal{P}^{[1,N]}\breve{X}\Psi\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\lesssim \mathring{\epsilon}^2\mathbb{K}_{[1,N]}(t,u) + \mathring{\epsilon}^2\int_{u'=0}^{u'=u}\mathbb{F}_{[1,N]}(t,u')\,du' + \mathring{\epsilon}^2\int_{t'=0}^{t'=t}\mathbb{E}_{[1,N]}(t',u)\,dt' + \mathring{\epsilon}^3\max\{1, \upmu_\star^{-2M_*+2N_{\mathrm{top}}-2N+1.8}(t)\} \\
&\lesssim \mathring{\epsilon}^3\max\{1, \upmu_\star^{-2M_*+2N_{\mathrm{top}}-2N+1.8}(t)\}.
\end{aligned} \tag{10-10}$$

The term III in (10-8) is the same as the term III in (9-7), and can be bounded as in the proof of Proposition 9.3, which, when combined with Proposition 9.4, implies that it is bounded by $\lesssim \mathring{\epsilon}^3\max\{1, \upmu_\star^{-2M_*+2N_{\mathrm{top}}-2N+1.8}(t)\}$.

Combining the above estimates, we conclude the desired estimate (10-7). □

[53] Note that in the case $N = N_{\mathrm{top}}$, the error terms on the right-hand side involving $\mathbb{V}_{\le N+1}$ and $\mathbb{S}_{\le N+1}$ have *not* been estimated in Section 9A. It is for this reason that we only consider $0 \le N \le N_{\mathrm{top}} - 1$ at this point.

[54] Note that by Lemma 10.2, there is also a term $\mathring{\epsilon}|\mathcal{P}\Psi|$, which we bound by $\lesssim \mathring{\epsilon}^{3/2}$ using (6-5).

Proposition 10.4 (L^2 control of some null forms in the modified fluid variable transport equations). *For* $0 \leq N \leq N_{\text{top}}$, *the terms* $\mathfrak{Q} \in \{\mathfrak{Q}^i_{(\mathcal{C})}, \mathfrak{Q}_{(\mathcal{D})}\}$ (*see* (5-6c)–(5-6d)) *can be estimated as follows for all* $(t, u) \in [0, T_{(\text{Boot})}) \times [0, U_0]$:

$$\|\mathcal{P}^N(\mu\mathfrak{Q})\|^2_{L^2(\mathcal{M}_{t,u})} \lesssim \mathring{\epsilon}^3 \max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+0.8}(t)\}. \tag{10-11}$$

Proof. The $\mathfrak{Q}$ terms can all be expressed as S multiplied by a null form $\mathcal{Q}(\partial\Psi, \partial\Psi)$. We control the null form using (10-5) with $\mathfrak{d}^{(1,1)}, \mathfrak{d}^{(1,2)}, \mathfrak{d}^{(2,1)}, \mathfrak{d}^{(2,2)} \lesssim 1$ (justified by (6-3)–(6-5)) so that

$$\begin{aligned}|\mathcal{P}^N(\mu\mathfrak{Q})| &\lesssim \sum_{N_1+N_2\leq N} |\mathcal{P}^{\leq N_1}(\Omega, S)|(|\mathcal{P}^{[1,N_2+1]}\Psi| + |\mathcal{P}^{[1,N_2]}\breve{X}\Psi| + |\mathcal{P}^{[2,N_2]}(\mu, L^i, \Psi)|) \\ &\lesssim \underbrace{|\mathcal{P}^{\leq N}(\Omega, S)|}_{\doteq I} + \underbrace{\mathring{\epsilon}(|\mathcal{P}^{[2,N+1]}\Psi| + |\mathcal{P}^{[1,N]}\breve{X}\Psi|)}_{\doteq II} + \underbrace{\mathring{\epsilon}|\mathcal{P}^{[2,N]}(\mu, L^i, \Psi)|}_{\doteq III},\end{aligned} \tag{10-12}$$

where in the last line, we used the L^∞ estimates (6-6), (6-7) for (Ω, S) if $N_1 \leq N_{\text{top}} - M_* - 5$, and otherwise, we used the L^∞ estimates (6-3)–(6-5) and Proposition 8.6 for Ψ, μ, and L^i.

Next, we observe that the terms II and III are exactly the same as II and III in (10-8) in Proposition 10.3. We can therefore argue exactly as in Proposition 10.3 to show that these terms in $\|\cdot\|^2_{L^2(\mathcal{M}_{t,u})}$ are bounded above by $\mathring{\epsilon}^3 \max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+1.8}(t)\}$. Notice in particular that while Proposition 10.3 was only stated for $0 \leq N \leq N_{\text{top}} - 1$, the bounds for these two terms in fact also hold (and can be proved in the same way) for $N = N_{\text{top}}$.

It thus remains to consider the term I in (10-12). Importantly, notice that term I in (10-12) is *better* than the corresponding term I in (10-8) because it has up to N, as opposed to $N+1$ derivatives. We control this term using the definition of $\mathbb{V}_{\leq N}$, $\mathbb{S}_{\leq N}$ and Proposition 9.4 as follows:

$$\|\mathcal{P}^{\leq N}(\Omega, S)\|^2_{L^2(\mathcal{M}_{t,u})} \lesssim \int_{u'=0}^{u'=u} [\mathbb{V}_{\leq N} + \mathbb{S}_{\leq N}](t, u')\, du' \lesssim \mathring{\epsilon}^3 \max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+2.8}(t)\}.$$

Combining the above estimates, we conclude the proposition. □

Proposition 10.5 (L^2 control of some easy terms in the transport equation for $\mathcal{C}$). *For* $0 \leq N \leq N_{\text{top}}$, *the term* $\mathfrak{L}^i_{(\mathcal{C})}$ (*see* (5-7g)) *can be estimated as follows for all* $(t, u) \in [0, T_{(\text{Boot})}) \times [0, U_0]$:

$$\|\mathcal{P}^N(\mu\mathfrak{L}^i_{(\mathcal{C})})\|_{L^2(\mathcal{M}_{t,u})} \lesssim \mathring{\epsilon}^3 \max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+0.8}(t)\}.$$

Proof. We begin with the pointwise estimate

$$|\mathcal{P}^{\leq N}(\mu\mathfrak{L}^i_{(\mathcal{C})})| \lesssim \mathring{\epsilon}|\mathcal{P}^{\leq N}(\Omega, S)| + \mathring{\epsilon}^2(|\mathcal{P}^{[2,N+1]}\Psi| + |\mathcal{P}^{[1,N]}\breve{X}\Psi|) + \mathring{\epsilon}^2|\mathcal{P}^{[2,N]}(\mu, L^i)|,$$

which can be derived by using the same arguments we used to obtain (9-7). Notice that all the above terms can be bounded above by the right-hand of (10-12). They can therefore be bounded in the norm $\|\cdot\|_{L^2(\mathcal{M}_{t,u})}$ via exactly the same arguments we used in the proof of Proposition 10.4. This yields the desired conclusion. □

10D. ***Below top-order estimates for*** $\mathbb{C}$ ***and*** $\mathbb{D}$.

Proposition 10.6 (below top-order estimates for the modified fluid variables). *For* $0 \le N \le N_{\text{top}} - 1$, *the following holds for* $(t, u) \in [0, T_{(\text{Boot})}) \times [0, U_0]$:

$$\mathbb{C}_N(t, u) + \mathbb{D}_N(t, u) \lesssim \mathring{\epsilon}^3 \max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+0.8}(t)\}.$$

Proof. This follows from combining Proposition 9.2 for $\phi = \mathcal{C}^i, \mathcal{D}^i$ with the initial data size assumptions in (4-6), the bootstrap assumptions (6-8), and the inhomogeneous term estimates (in Propositions 10.3–10.5) for the terms on the right-hand sides of the transport equations (5-3b) and (5-4a). □

11. Top-order transport and elliptic estimates for the specific vorticity and the entropy gradient

We continue to work under the assumptions of Theorem 6.3.

In this section, we derive top-order estimates for the modified fluid variables $\mathcal{C}$ and $\mathcal{D}$. The key difference with the lower-order estimates (which we derived in Proposition 10.6) is that we *cannot* bound the top-order derivatives of Ω and S using the $\mathbb{V}$ and $\mathbb{S}$ norms; that approach would lead to a loss of a derivative, which is not permissible at the top-order. To avoid losing a derivative, we rely on the following additional ingredient: *weighted* elliptic estimates for the specific vorticity and entropy gradient (recall Sections 1A6, 1A7).

In Section 11A, we derive top-order transport estimates. The estimates are similar to the ones we derived in Section 10, except there are some top-order inhomogeneous terms. We derive the elliptic estimates in Sections 11B and 11C. For the final estimate, see Section 11D.

In our analysis, we rely on elliptic estimates relative to the *Cartesian* spatial coordinates. In deriving these estimates, we will use the "Cartesian pointwise norms" from the following definition.

Definition 11.1. Denote by $\underline{\partial}$ the gradient with respect to the Cartesian spatial coordinates. For a scalar function f and a one-form ϕ, define respectively

$$|\underline{\partial} f|^2 \doteq \sum_{i=1}^{3} |\partial_i f|^2, \quad |\underline{\partial}\phi|^2 \doteq \sum_{i,j=1}^{3} |\partial_i \phi_j|^2.$$

11A. ***Top-order transport estimates for*** $\mathbb{C}_{N_{\text{top}}}$ ***and*** $\mathbb{D}_{N_{\text{top}}}$.

Proposition 11.2 (preliminary top-order L^2 estimates for the modified fluid variables). *Let* $\varsigma \in (0, 1]$. *There exists a constant* $C > 0$ *independent of* ς *and a constant* $\mathfrak{c}_\varsigma > 0$ (*depending on* ς) *such that whenever* $\mathfrak{c} \ge \mathfrak{c}_\varsigma$ *the following estimate holds for every* $(t, u) \in [0, T_{(\text{Boot})}) \times [0, U_0]$ (*with* u' *denoting the* u*-value of the integrand*):

$$\begin{aligned} &\|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\mathcal{P}^{N_{\text{top}}}(\mathcal{C}, \mathcal{D})\|^2_{L^2(\Sigma^u_t)} + \|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\text{top}}}(\mathcal{C}, \mathcal{D})\|^2_{L^2(\mathcal{F}^t_u)} + \frac{\mathfrak{c}}{2}\|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\text{top}}}(\mathcal{C}, \mathcal{D})\|^2_{L^2(\mathcal{M}_{t,u})} \\ &\qquad \le C\mathring{\epsilon}^3\mu_\star^{-2M_*+0.8}(t) + \varsigma\int_{t'=0}^{t'=t} \frac{1}{\mu_\star(t')}\|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\underline{\partial}\mathcal{P}^{N_{\text{top}}}(\Omega, S)\|^2_{L^2(\Sigma^u_{t'})}\, dt'. \end{aligned} \tag{11-1}$$

Proof. Let $\varsigma', \mathfrak{c} > 0$ be constants to be specified later. It is crucial that all explicit constants $C > 0$ and implicit constants in this proof are *independent* of ς' and $\mathfrak{c}$. At the end of the proof, there will be a large constant C such that we will choose ς' to satisfy $\varsigma = C\varsigma'$, where $\varsigma > 0$ is the constant from the statement of the proposition.

Step 1: transport estimate in the weighted norms. Since $\mu B u = 1$ by (2-21), (2-23), we have

$$\mu B(e^{-\mathfrak{c}u/2}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{C}) = -\frac{\mathfrak{c}}{2}e^{-\mathfrak{c}u/2}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{C} + e^{-\mathfrak{c}u/2}\mu B(\mathcal{P}^{N_{\mathrm{top}}}\mathcal{C}), \tag{11-2}$$

$$\mu B(e^{-\mathfrak{c}u/2}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D}) = -\frac{\mathfrak{c}}{2}e^{-\mathfrak{c}u/2}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D} + e^{-\mathfrak{c}u/2}\mu B(\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D}). \tag{11-3}$$

Starting with (11-2) and (11-3), we now argue using the identity (9-1) with $\phi \doteq (\mathcal{C}^i, \mathcal{D}^i)$, except now, unlike in the proof of Proposition 9.1, we do not use Grönwall's inequality but instead take advantage of the good terms associated with the terms $-(\mathfrak{c}/2)e^{-\mathfrak{c}u/2}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{C}$ and $-(\mathfrak{c}/2)e^{-\mathfrak{c}u/2}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D}$ on the right-hand sides (11-2)–(11-3). We thus obtain, for any $\varsigma' > 0$ (here, u' denotes the u-value of the integrand),

$$\begin{aligned}
&\|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{C}\|^2_{L^2(\Sigma^u_t)} + \|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{C}\|^2_{L^2(\mathcal{F}^t_u)} + \mathfrak{c}\|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{C}\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\qquad + \|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D}\|^2_{L^2(\Sigma^u_t)} + \|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D}\|^2_{L^2(\mathcal{F}^t_u)} + \mathfrak{c}\|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D}\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\quad \lesssim \|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\mathcal{P}^{N_{\mathrm{top}}}(\mathcal{C},\mathcal{D})\|^2_{L^2(\Sigma^u_0)} + \|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}}(\mathcal{C},\mathcal{D})\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\qquad\qquad + \|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}}(\mathcal{C},\mathcal{D})\|_{L^2(\mathcal{M}_{t,u})}\|e^{-\mathfrak{c}u'/2}\mu B\mathcal{P}^{N_{\mathrm{top}}}(\mathcal{C},\mathcal{D})\|_{L^2(\mathcal{M}_{t,u})} \\
&\quad \lesssim \|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\mathcal{P}^{N_{\mathrm{top}}}(\mathcal{C},\mathcal{D})\|^2_{L^2(\Sigma^u_0)} + (1+(\varsigma')^{-1})\|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}}(\mathcal{C},\mathcal{D})\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\qquad\qquad + \varsigma'\|e^{-\mathfrak{c}u'/2}\mu B\mathcal{P}^{N_{\mathrm{top}}}(\mathcal{C},\mathcal{D})\|^2_{L^2(\mathcal{M}_{t,u})}.
\end{aligned} \tag{11-4}$$

Step 2: estimating the easy terms. We now consider the terms on the right-hand side of (11-4). First, the assumptions (4-6) on the initial data and the simple bound $\|\mu\|_{L^\infty(\Sigma_0)} \lesssim 1$ from Proposition 8.6 give

$$\|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\mathcal{P}^{N_{\mathrm{top}}}(\mathcal{C},\mathcal{D})\|^2_{L^2(\Sigma^u_0)} \lesssim \mathring{\epsilon}^3. \tag{11-5}$$

Recalling the transport equations (5-3b), (5-4a), we notice that the terms $\|e^{-\mathfrak{c}u'/2}\mu B\mathcal{P}^{N_{\mathrm{top}}}\mathcal{C}\|_{L^2(\mathcal{M}_{t,u})}$ and $\|e^{-\mathfrak{c}u'/2}\mu B\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D}\|_{L^2(\mathcal{M}_{t,u})}$ have essentially been estimated in Propositions 10.3–10.5 (using $e^{-\mathfrak{c}u'/2} \leq 1$). Crucially, however, unlike in Proposition 10.3, we have not yet bounded the following terms in (10-9):

$$\|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}+1}(\varOmega, S)\|^2_{L^2(\mathcal{M}_{t,u})} + \|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}}\breve{X}(\varOmega, S)\|^2_{L^2(\mathcal{M}_{t,u})}$$

(since this is one more derivative than $\mathbb{V}_{N_{\mathrm{top}}}$ and $\mathbb{S}_{N_{\mathrm{top}}}$ control). In other words, simply repeating the argument in Propositions 10.3–10.5 and separating the error terms that depend on $N_{\mathrm{top}}+1$ derivatives of $(\varOmega, S)$, we obtain

$$\begin{aligned}
&\|e^{-\mathfrak{c}u'/2}\mu B\mathcal{P}^{N_{\mathrm{top}}}\mathcal{C}\|^2_{L^2(\mathcal{M}_{t,u})} + \|e^{-\mathfrak{c}u'/2}\mu B\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D}\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\qquad \lesssim \mathring{\epsilon}^3\mu_\star^{-2M_*+0.8}(t) + \|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}+1}(\varOmega, S)\|^2_{L^2(\mathcal{M}_{t,u})} + \|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}}\breve{X}(\varOmega, S)\|^2_{L^2(\mathcal{M}_{t,u})}.
\end{aligned} \tag{11-6}$$

Step 3: controlling the top-order terms. We now consider the terms on the right-hand side of (11-6). First, using the commutator estimates (8-37), Proposition 9.4, and the bootstrap assumptions (6-6)–(6-7) to control $[\mathcal{P}^{N_{\mathrm{top}}}, \breve{X}](\Omega, S)$ (and using $e^{-\mathfrak{c}u'/2} \le 1$), we see that

$$
\begin{aligned}
&\|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}+1}(\Omega, S)\|^2_{L^2(\mathcal{M}_{t,u})} + \|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}}\breve{X}(\Omega, S)\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\quad\lesssim \mathring{\epsilon}^3\mu_\star^{-2M_*+2.8}(t) + \|e^{-\mathfrak{c}u'/2}\breve{X}\mathcal{P}^{N_{\mathrm{top}}}(\Omega, S)\|^2_{L^2(\mathcal{M}_{t,u})} + \|e^{-\mathfrak{c}u'/2}L\mathcal{P}^{N_{\mathrm{top}}}(\Omega, S)\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\qquad\qquad + \|e^{-\mathfrak{c}u'/2}Y\mathcal{P}^{N_{\mathrm{top}}}(\Omega, S)\|^2_{L^2(\mathcal{M}_{t,u})} + \|e^{-\mathfrak{c}u'/2}Z\mathcal{P}^{N_{\mathrm{top}}}(\Omega, S)\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\quad\lesssim \mathring{\epsilon}^3\mu_\star^{-2M_*+2.8}(t) + \|e^{-\mathfrak{c}u'/2}B\mathcal{P}^{N_{\mathrm{top}}}(\Omega, S)\|^2_{L^2(\mathcal{M}_{t,u})} + \|e^{-\mathfrak{c}u'/2}\underline{\partial}\mathcal{P}^{N_{\mathrm{top}}}(\Omega, S)\|^2_{L^2(\mathcal{M}_{t,u})},
\end{aligned}
\tag{11-7}
$$

where we have replaced $L\mathcal{P}^{N_{\mathrm{top}}}(\Omega, S) = B\mathcal{P}^{N_{\mathrm{top}}}(\Omega, S) - X\mathcal{P}^{N_{\mathrm{top}}}(\Omega, S)$ (by (2-23)) and also used Lemmas 2.23 and 2.24 to express (X, Y, Z) in terms of the Cartesian coordinate spatial partial derivative vectorfields, and Propositions 8.6 and 8.7 to bound the coefficients in the expressions by $\lesssim 1$. Moreover, using the commutator identity $B\mathcal{P}^{N_{\mathrm{top}}}(\Omega, S) = \mu^{-1}\mathcal{P}^{N_{\mathrm{top}}}[\mu B(\Omega, S)] + \mu^{-1}[\mu B, \mathcal{P}^{N_{\mathrm{top}}}](\Omega, S)$, the commutator estimates of Proposition 8.15 with $\phi \doteq (\Omega^i, S^i)$, the bootstrap assumptions (6-6)–(6-7), Proposition 9.4, the estimate (9-5), and Proposition 8.11, we deduce (also using $e^{-\mathfrak{c}u'/2} \le 1$) that $\|e^{-\mathfrak{c}u'/2}B\mathcal{P}^{N_{\mathrm{top}}}(\Omega, S)\|^2_{L^2(\mathcal{M}_{t,u})} \lesssim \mathring{\epsilon}^3\mu_\star^{-2M_*+0.8}(t)$. Combining the above results, we deduce

$$
\begin{aligned}
&\|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}+1}(\Omega, S)\|^2_{L^2(\mathcal{M}_{t,u})} + \|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}}\breve{X}(\Omega, S)\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\qquad\lesssim \mathring{\epsilon}^3\mu_\star^{-2M_*+0.8}(t) + \int_{t'=0}^{t'=t} \frac{1}{\mu_\star(t')}\|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\underline{\partial}\mathcal{P}^{N_{\mathrm{top}}}(\Omega, S)\|^2_{L^2(\Sigma^u_{t'})}\, dt'.
\end{aligned}
\tag{11-8}
$$

Step 4: putting everything together. Using (11-5), (11-6) and (11-8) to control the terms on the right-hand side of (11-4), we see that there is a $C > 0$ such that

$$
\begin{aligned}
&\|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{C}\|^2_{L^2(\Sigma^u_t)} + \|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{C}\|^2_{L^2(\mathcal{F}^t_u)} + \mathfrak{c}\|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{C}\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\qquad + \|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D}\|^2_{L^2(\Sigma^u_t)} + \|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D}\|^2_{L^2(\mathcal{F}^t_u)} + \mathfrak{c}\|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D}\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\quad\le C(1+\varsigma')\mathring{\epsilon}^3\mu_\star^{-2M_*+0.8}(t) + C(1+(\varsigma')^{-1})\|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}}(\mathcal{C}, \mathcal{D})\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\qquad\qquad + C\varsigma' \int_{t'=0}^{t'=t} \frac{1}{\mu_\star(t')}\|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\underline{\partial}\mathcal{P}^{N_{\mathrm{top}}}(\Omega, S)\|^2_{L^2(\Sigma^u_{t'})}\, dt'.
\end{aligned}
\tag{11-9}
$$

Finally, relabeling the coefficients $C\varsigma'$ on the right-hand side of (11-9) by setting $\varsigma \doteq C\varsigma'$, bounding the data term $C(1+\varsigma')\mathring{\epsilon}^3\mu_\star^{-2M_*+0.8}(t)$ by a new constant C times $\mathring{\epsilon}^3\mu_\star^{-2M_*+0.8}(t)$ via the assumption $\varsigma \in (0, 1]$, taking $\mathfrak{c}_\varsigma$ sufficiently large (depending on ς) so that

$$
C(1+(\varsigma')^{-1})\|e^{-\mathfrak{c}u'/2}\mathcal{P}^{N_{\mathrm{top}}}(\mathcal{C}, \mathcal{D})\|^2_{L^2(\mathcal{M}_{t,u})} \le \frac{\mathfrak{c}_\varsigma}{2}\int_{\mathcal{M}_{t,u}} e^{-\mathfrak{c}u'}[|\mathcal{P}^{N_{\mathrm{top}}}\mathcal{C}|^2 + |\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D}|^2]\, d\varpi,
$$

now allowing $\mathfrak{c}$ to be any constant such that $\mathfrak{c} \ge \mathfrak{c}_\varsigma$, and subtracting $(\mathfrak{c}/2)\int_{\mathcal{M}_{t,u}} e^{-\mathfrak{c}u'}[|\mathcal{P}^{N_{\mathrm{top}}}\mathcal{C}|^2 + |\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D}|^2]\, d\varpi$ from both sides of (11-9), we obtain the desired inequality (11-1). □

11B. ***General elliptic estimates on*** $\mathbb{R}\times\mathbb{T}^2$**.** We begin with a standard weighted *Euclidean* elliptic estimate on $\mathbb{R}\times\mathbb{T}^2$ in Proposition 11.3. We then apply this in our geometric setting for general one-forms in Proposition 11.4.

Proposition 11.3 (weighted Euclidean elliptic estimates). *Let $w : \mathbb{R} \times \mathbb{T}^2 \to \mathbb{R}_{>0}$ be a smooth, strictly positive, bounded weight function.*

The following inequality holds for all one-forms $\phi = \phi_a dx^a \in C_c^2(\mathbb{R} \times \mathbb{T}^2)$:

$$\begin{aligned}\|\sqrt{w}\underline{\partial}\phi\|^2_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)} &\\ \le 4\|\sqrt{w}\operatorname{curl}\phi\|^2_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)} &+ 4\|\sqrt{w}\operatorname{div}\phi\|^2_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)} + 3\|\underline{\partial}\log w\|^2_{L^\infty(\mathbb{R}\times\mathbb{T}^2)}\|\sqrt{w}\phi\|^2_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)},\end{aligned}$$

where $\underline{\partial}$ is as in Definition 11.1, $\|\xi\|^2_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)} \doteq \int_{\mathbb{R}\times\mathbb{T}^2} |\xi|_e^2\, dx$ for tensorfields ξ, $|\xi|_e$ denotes the standard Euclidean pointwise norm of ξ, and $dx = dx^1\, dx^2\, dx^3$.

Proof. Integrating by parts and using Hölder's inequality, we find that

$$\begin{aligned}
&\|\sqrt{w}\underline{\partial}\phi\|^2_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)} \\
&= \sum_{i,j=1}^{3}\int_{\mathbb{R}\times\mathbb{T}^2} w(\partial_i\phi_j)^2\, dx \\
&= -\sum_{i,j=1}^{3}\left\{\int_{\mathbb{R}\times\mathbb{T}^2} w\phi_j(\partial^2_{ii}\phi_j)\, dx + \int_{\mathbb{R}\times\mathbb{T}^2}(\partial_i w)\phi_j(\partial_i\phi_j)\, dx\right\} \\
&= -\sum_{i,j=1}^{3}\int_{\mathbb{R}\times\mathbb{T}^2} w\phi_j\partial^2_{ij}\phi_i\, dx + \sum_{i,j=1}^{3}\int_{\mathbb{R}\times\mathbb{T}^2} w\phi_j\partial_i(\partial_j\phi_i - \partial_i\phi_j)\, dx - \sum_{i,j=1}^{3}\int_{\mathbb{R}\times\mathbb{T}^2}(\partial_i w)\phi_j(\partial_i\phi_j)\, dx \\
&= \sum_{i,j=1}^{3}\int_{\mathbb{R}\times\mathbb{T}^2} w(\partial_j\phi_j)(\partial_i\phi_i)\, dx - \sum_{i,j=1}^{3}\int_{\mathbb{R}\times\mathbb{T}^2} w(\partial_i\phi_j)(\partial_j\phi_i - \partial_i\phi_j)\, dx \\
&\quad + \sum_{i,j=1}^{3}\int_{\mathbb{R}\times\mathbb{T}^2}(\partial_j w)\phi_j(\partial_i\phi_i)\, dx - \sum_{i,j=1}^{3}\int_{\mathbb{R}\times\mathbb{T}^2}(\partial_i w)\phi_j(\partial_j\phi_i - \partial_i\phi_j)\, dx - \sum_{i,j=1}^{3}\int_{\mathbb{R}\times\mathbb{T}^2}(\partial_i w)\phi_j(\partial_i\phi_j)\, dx \\
&\le \|\sqrt{w}\ \operatorname{div}\ \phi\|^2_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)} + \|\sqrt{w}\ \underline{\partial}\phi\|_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)}\|\sqrt{w}\operatorname{curl}\ \phi\|_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)} \\
&\quad + \|\underline{\partial}\log w\|_{L^\infty(\mathbb{R}\times\mathbb{T}^2)}\|\sqrt{w}\phi\|_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)}\{\|\sqrt{w}\ \operatorname{div}\ \phi\|_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)} + \|\sqrt{w}\operatorname{curl}\ \phi\|_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)}\} \\
&\qquad + \|\underline{\partial}\log w\|_{L^\infty(\mathbb{R}\times\mathbb{T}^2)}\|\sqrt{w}\phi\|_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)}\|\sqrt{w}\ \underline{\partial}\phi\|_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)}.
\end{aligned} \tag{11-10}$$

Using $|ab| \le a^2/4 + b^2$, we find that

$$\begin{aligned}\|\sqrt{w}\underline{\partial}\phi\|^2_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)} \le \tfrac{1}{2}\|\sqrt{w}\ \underline{\partial}\phi\|^2_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)} &+ 2\|\sqrt{w}\operatorname{div}\phi\|^2_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)} \\ + 2\|\sqrt{w}\operatorname{curl}\phi\|^2_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)} &+ \tfrac{3}{2}\|\underline{\partial}\log w\|^2_{L^\infty(\mathbb{R}\times\mathbb{T}^2)}\|\sqrt{w}\phi\|^2_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)}.\end{aligned} \tag{11-11}$$

The conclusion of the lemma follows from subtracting $\frac{1}{2}\|\sqrt{w}\ \underline{\partial}\phi\|^2_{L^2(\mathbb{R}\times\mathbb{T}^2,dx)}$ from both sides of (11-11). □

Proposition 11.4 (Euclidean elliptic estimates with u-weights). *Let $\phi = \phi_a dx^a$ be a smooth compactly supported one-form on Σ_t. Then for each $\mathfrak{c} > 0$ and each $t \in [0, T_{(\mathrm{Boot})})$, the following elliptic estimate holds, where the implicit constants are independent of $\mathfrak{c}$:*

$$\|e^{-\mathfrak{c}u/2}\sqrt{\mu}\underline{\partial}\phi\|_{L^2(\Sigma_t)} \lesssim \|e^{-\mathfrak{c}u/2}\sqrt{\mu}\operatorname{div}\phi\|_{L^2(\Sigma_t)} + \|e^{-\mathfrak{c}u/2}\sqrt{\mu}\operatorname{curl}\phi\|_{L^2(\Sigma_t)} + \mathfrak{c}\mu_\star^{-1}(t)\|e^{-\mathfrak{c}u/2}\sqrt{\mu}\phi\|_{L^2(\Sigma_t)}.$$

Proof. In this proof, the implicit constants in $\lesssim$ are *independent* of $\mathfrak{c}$.

We apply Proposition 11.3 with $w \doteq e^{-\mathfrak{c}u}$. By Lemma 2.22, (2-21), and Proposition 8.7, we have $\|\underline{\partial} \log w\|_{L^\infty(\mathbb{R}\times\mathbb{T}^2)} \lesssim \mathfrak{c}\mu_\star^{-1}(t)$. Hence,

$$\|e^{-\mathfrak{c}u/2}\underline{\partial}\phi\|_{L^2(\Sigma_t,dx)} \lesssim \|e^{-\mathfrak{c}u/2} \operatorname{div} \phi\|_{L^2(\Sigma_t,dx)} + \|e^{-\mathfrak{c}u/2} \operatorname{curl} \phi\|_{L^2(\Sigma_t,dx)} + \mathfrak{c}\mu_\star^{-1}(t)\|e^{-\mathfrak{c}u/2} \phi\|_{L^2(\Sigma_t,dx)}.$$

The conclusion thus follows from the fact that the volume measures $\mu\, dx$ and $d\underline{\varpi}$ are comparable, which in turn follows from (3-1) and Proposition 8.7. □

11C. ***Top-order elliptic estimates for Ω and S.*** In this section, we derive top-order elliptic estimates for Ω and S.

There are four main steps. Ultimately, our goal is to exploit the preliminary energy inequality for $(\mathcal{P}^{N_{\text{top}}}\mathcal{C}, \mathcal{P}^{N_{\text{top}}}\mathcal{D})$ that we derived in Proposition 11.2, and to do this, we have to control the integrand term $\|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\underline{\partial}\mathcal{P}^{N_{\text{top}}}(\Omega, S)\|^2_{L^2(\Sigma^u_{t'})}$ on the right-hand side of (11-1) with the help of elliptic estimates. To achieve this, we first commute the top-order operators $\mathcal{P}^{N_{\text{top}}}$ through the Euclidean operators div and curl. To avoid uncontrollable commutator terms, we introduce a μ weight into the commutators. In the second step, we have to control $(\operatorname{div} \mathcal{P}^{N_{\text{top}}}\Omega, \operatorname{div} \mathcal{P}^{N_{\text{top}}}S)$ and $(\operatorname{curl} \mathcal{P}^{N_{\text{top}}}\Omega, \operatorname{curl} \mathcal{P}^{N_{\text{top}}}S)$ in terms of the modified fluid variables $(\mathcal{P}^{N_{\text{top}}}\mathcal{C}, \mathcal{P}^{N_{\text{top}}}\mathcal{D})$ from (2-5a)–(2-5b) plus simpler error terms. The first and second steps are carried out in Lemmas 11.6–11.9.

Next, in Proposition 11.10, we use the weighted elliptic estimates on Σ_t provided by Proposition 11.4 and the results of the first two steps to obtain

$$\|e^{-\mathfrak{c}u/2}\sqrt{\mu}(\underline{\partial}\mathcal{P}^{N_{\text{top}}}\Omega, \underline{\partial}\mathcal{P}^{N_{\text{top}}}S)\|^2_{L^2(\Sigma_t)} \lesssim \|e^{-\mathfrak{c}u/2}\sqrt{\mu}(\mathcal{P}^{N_{\text{top}}}\mathcal{C}, \mathcal{P}^{N_{\text{top}}}\mathcal{D})\|^2_{L^2(\Sigma_t)} + \cdots,$$

where "$\cdots$" denotes simpler error terms for which we already have an independent bound. Finally, in Proposition 11.11, we combine all of these results to obtain our main L^2 estimate[55] for $(\mathcal{P}^{N_{\text{top}}}\mathcal{C}, \mathcal{P}^{N_{\text{top}}}\mathcal{D})$.

11C1. *Controlling* $\operatorname{curl} \mathcal{P}^{N_{\text{top}}}\Omega$ *and* $\operatorname{div} \mathcal{P}^{N_{\text{top}}}\Omega$. We start with a simple commutation lemma.

Lemma 11.5 (commuting geometric vectorfields with μ-weighted Cartesian vectorfields). *Let ϕ be a smooth function such that*

$$\|\mathcal{P}^{\le N_{\text{top}}-M_*-5}\phi\|_{L^\infty(\Sigma_t)} \le \mathring{\varepsilon}, \quad \|\mathcal{P}^{\le N_{\text{top}}-M_*-5}\breve{X}\phi\|_{L^\infty(\Sigma_t)} \le \mathring{\varepsilon}$$

for all $t \in [0, T_{(\text{Boot})})$.

Then, for $0 \le N \le N_{\text{top}}$, the following holds in $\mathcal{M}_{T_{(\text{Boot})},U_0}$:

$$|[\mu\partial_i, \mathcal{P}^N]\phi| \lesssim |\mathcal{P}^{[1,N]}\phi| + |\mathcal{P}^{\le N-1}\breve{X}\phi| + \mathring{\varepsilon}(|\mathcal{P}^{[2,N]}(\mu, L^i, \Psi)| + |\mathcal{P}^{[2,N-1]}\breve{X}\Psi|).$$

Proof. We first use Lemma 2.22 to express $\mu\partial_i$ in terms of the geometric vectorfields and then argue as in Proposition 8.8. □

[55] We clarify that although the estimate for $(\mathcal{P}^{N_{\text{top}}}\mathcal{C}, \mathcal{P}^{N_{\text{top}}}\mathcal{D})$ and the aforementioned estimates $\|e^{-\mathfrak{c}u/2}\sqrt{\mu}(\underline{\partial}\mathcal{P}^{N_{\text{top}}}\Omega, \underline{\partial}\mathcal{P}^{N_{\text{top}}}S)\|^2_{L^2(\Sigma_t)} \lesssim \|e^{-\mathfrak{c}u/2}\sqrt{\mu}(\mathcal{P}^{N_{\text{top}}}\mathcal{C}, \mathcal{P}^{N_{\text{top}}}\mathcal{D})\|^2_{L^2(\Sigma_t)} + \cdots$ together imply a top-order L^2 estimate for $(\underline{\partial}\mathcal{P}^{N_{\text{top}}}\Omega, \underline{\partial}\mathcal{P}^{N_{\text{top}}}S)$, we do not explicitly state such an estimate in the paper because we do not need it for our main results.

Lemma 11.6 (L^2 estimates for the Euclidean curl of the derivatives of $\varOmega$ in terms of the derivatives of $\mathcal{C}$). *Let $\mathfrak{c} \geq 0$ be a real number. The following estimate holds for all $t \in [0, T_{(\text{Boot})})$, where the implicit constants are independent of $\mathfrak{c}$:*

$$\|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\ \text{curl}\ \mathcal{P}^{N_{\text{top}}}\varOmega\|^2_{L^2(\Sigma_t)} \lesssim \mathring{\epsilon}^3\mu_\star^{-2M_*+0.8}(t) + \|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\,\mathcal{P}^{N_{\text{top}}}\mathcal{C}\|^2_{L^2(\Sigma_t)}.$$

Proof. We first compute the commutator $[\mu\,\text{curl}, \mathcal{P}^{N_{\text{top}}}]$ using Lemma 11.5 and the bootstrap assumption (6-6):

$$|[\mu\ \text{curl}, \mathcal{P}^{N_{\text{top}}}]\varOmega| \lesssim |\mathcal{P}^{\leq N_{\text{top}}}\varOmega| + |\mathcal{P}^{\leq N_{\text{top}}-1}\breve{X}\varOmega| + \mathring{\epsilon}(|\mathcal{P}^{[2,N_{\text{top}}]}(\mu, L^i, \Psi)| + |\mathcal{P}^{[2,N_{\text{top}}-1]}\breve{X}\Psi|). \tag{11-12}$$

On the other hand, by (2-5a), Lemma 2.22, the bootstrap assumptions (6-3)–(6-8), and Propositions 8.6 and 8.7, we have

$$\begin{aligned}|\mathcal{P}^{N_{\text{top}}}(\mu\ \text{curl}\ \varOmega)| &= \left|\mathcal{P}^{N_{\text{top}}}\left\{\mu\left[\exp(\rho)\mathcal{C} - \exp(-2\rho)c_s^{-2}\frac{p_{;s}}{\bar{\varrho}}S^a\partial_a v + \exp(-2\rho)c_s^{-2}\frac{p_{;s}}{\bar{\varrho}}(\partial_a v^a)S\right]\right\}\right| \\ &\lesssim \mu|\mathcal{P}^{N_{\text{top}}}\mathcal{C}| + |\mathcal{P}^{\leq N_{\text{top}}-1}\mathcal{C}| + |\mathcal{P}^{\leq N_{\text{top}}}S| \\ &\quad + \mathring{\epsilon}\big(|\mathcal{P}^{[2,N_{\text{top}}]}(\mu, L^i)| + \mu|\mathcal{P}^{N_{\text{top}}+1}\Psi| + |\mathcal{P}^{[2,N_{\text{top}}]}\Psi| + |\mathcal{P}^{[1,N_{\text{top}}]}\breve{X}\Psi|\big).\end{aligned} \tag{11-13}$$

We stress that on the right-hand side of (11-13), it is important that the top-order terms $\mathcal{P}^{N_{\text{top}}}\mathcal{C}$ and $\mathcal{P}^{N_{\text{top}}+1}\Psi$ are accompanied by a factor of μ.

We can therefore use (11-12) and (11-13) (to write $\mu\,\text{curl}\,\mathcal{P}^{N_{\text{top}}}\varOmega = [\mu\ \text{curl}, \mathcal{P}^{N_{\text{top}}}]\varOmega + \mathcal{P}^{N_{\text{top}}}(\mu\ \text{curl}\ \varOmega)$), multiply by $e^{-\mathfrak{c}u/2}\mu^{-1/2}$, take the $L^2(\Sigma_t)$ norm, and then use $e^{-\mathfrak{c}u/2} \leq 1$ to obtain

$$\begin{aligned}&\|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\,\text{curl}(\mathcal{P}^{N_{\text{top}}}\varOmega)\|_{L^2(\Sigma_t)} \\ &\quad\lesssim \|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\mathcal{P}^{N_{\text{top}}}\mathcal{C}\|_{L^2(\Sigma_t)} + \mu_\star^{-1}(t)\|\sqrt{\mu}\mathcal{P}^{\leq N_{\text{top}}-1}\mathcal{C}\|_{L^2(\Sigma_t)} + \mu_\star^{-1}(t)\|\sqrt{\mu}\mathcal{P}^{\leq N_{\text{top}}}(\varOmega, S)\|_{L^2(\Sigma_t)} \\ &\qquad + \mu_\star^{-1}(t)\|\sqrt{\mu}\mathcal{P}^{\leq N_{\text{top}}-1}\breve{X}\varOmega\|_{L^2(\Sigma_t)} + \mathring{\epsilon}\mu_\star^{-1/2}(t)\|\mathcal{P}^{[2,N_{\text{top}}]}(\mu, L^i)\|_{L^2(\Sigma_t)} \\ &\qquad\quad + \mathring{\epsilon}(\|\sqrt{\mu}\mathcal{P}^{N_{\text{top}}+1}\Psi\|_{L^2(\Sigma_t)} + \mu_\star^{-1/2}(t)\|\mathcal{P}^{[1,N_{\text{top}}]}\breve{X}\Psi\|_{L^2(\Sigma_t)} + \mu_\star^{-1}(t)\|\sqrt{\mu}\mathcal{P}^{[2,N_{\text{top}}]}\Psi\|_{L^2(\Sigma_t)}) \\ &\quad\lesssim \|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\mathcal{P}^{N_{\text{top}}}\mathcal{C}\|_{L^2(\Sigma_t)} + \mathring{\epsilon}^{3/2}\mu_\star^{-M_*+0.4}(t),\end{aligned} \tag{11-14}$$

where we have used Proposition 10.6 to bound $\mu_\star^{-1}(t)\|\sqrt{\mu}\mathcal{P}^{\leq N_{\text{top}}-1}\mathcal{C}\|_{L^2(\Sigma_t)}$, Proposition 9.4 to bound $\mu_\star^{-1}(t)\|\sqrt{\mu}\mathcal{P}^{\leq N_{\text{top}}}(\varOmega, S)\|_{L^2(\Sigma_t)}$, Proposition 10.1 to bound $\mu_\star^{-1}(t)\|\sqrt{\mu}\mathcal{P}^{\leq N_{\text{top}}-1}\breve{X}\varOmega\|_{L^2(\Sigma_t)}$, Proposition 8.14 to bound $\mathring{\epsilon}\mu_\star^{-1/2}(t)\|\mathcal{P}^{[2,N_{\text{top}}]}(\mu, L^i)\|_{L^2(\Sigma_t)}$, and the bootstrap assumptions (6-1), (6-2), and (8-38) to estimate all the remaining terms. (We remark that the worst terms are $\mu_\star^{-1}(t)\|\sqrt{\mu}\mathcal{P}^{\leq N_{\text{top}}-1}\mathcal{C}\|_{L^2(\Sigma_t)}$, $\mu_\star^{-1}(t)\|\sqrt{\mu}\mathcal{P}^{\leq N_{\text{top}}}(\varOmega, S)\|_{L^2(\Sigma_t)}$, $\mu_\star^{-1}(t)\|\sqrt{\mu}\mathcal{P}^{\leq N_{\text{top}}-1}\breve{X}\varOmega\|_{L^2(\Sigma_t)}$, and $\mu_\star^{-1/2}(t)\|\mathcal{P}^{[1,N_{\text{top}}]}\breve{X}\Psi\|_{L^2(\Sigma_t)}$, which determine the blowup-exponent $-M_* + 0.4$ for $\mu_\star$ on the right-hand side of (11-14)). Squaring (11-14), we arrive at the desired result. □

Lemma 11.7 (L^2 estimates for the Euclidean divergence of the derivatives of $\varOmega$). *Let $\mathfrak{c} \geq 0$ be a real number. The following estimate holds for all $t \in [0, T_{(\text{Boot})})$, where the implicit constant is independent of $\mathfrak{c}$:*

$$\|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\,\text{div}\ \mathcal{P}^{N_{\text{top}}}\varOmega\|^2_{L^2(\Sigma_t)} \lesssim \mathring{\epsilon}^3\mu_\star^{-2M_*+0.8}(t).$$

Proof. The commutator $[\upmu \operatorname{div}, \mathcal{P}^{N_{\text{top}}}]\Omega$ can be computed exactly as (11-12). Thus, we have

$$|[\upmu \operatorname{div}, \mathcal{P}^{N_{\text{top}}}]\Omega| \lesssim \text{the right-hand side of (11-12).} \tag{11-15}$$

We also use Lemma 2.22, the fact that the Cartesian component functions X^1, X^2, X^3 are smooth functions of the L^i and Ψ (see (2-23)), (5-3a), and the L^∞ bounds in (6-3)–(6-6) and Proposition 8.6 to deduce

$$\begin{aligned} |\mathcal{P}^{N_{\text{top}}}(\upmu \operatorname{div} \Omega)| &= |\mathcal{P}^{N_{\text{top}}}(\upmu\, \Omega^a \partial_a \rho)| \\ &\lesssim |\mathcal{P}^{\leq N_{\text{top}}}\Omega| + \mathring{\epsilon}(|\mathcal{P}^{[2,N_{\text{top}}]}\upmu| + \upmu|\mathcal{P}^{N_{\text{top}}+1}\Psi| + |\mathcal{P}^{[2,N_{\text{top}}]}\Psi| + |\mathcal{P}^{[1,N_{\text{top}}]}\breve{X}\Psi|). \end{aligned} \tag{11-16}$$

Notice that every term on the right-hand side of (11-16) has already appeared on the right-hand sides of (11-12) and (11-13). Hence, with the help of the simple identity

$$\upmu \operatorname{div} \mathcal{P}^{N_{\text{top}}}\Omega = \mathcal{P}^{N_{\text{top}}}(\upmu \operatorname{div} \Omega) + [\upmu \operatorname{div}, \mathcal{P}^{N_{\text{top}}}]\Omega$$

and the estimates obtained above, we can argue exactly as in Lemma 11.6 to obtain the same estimate. (Note that here there are no $\mathcal{C}$ terms and so we do not have the term $\|e^{-\mathfrak{c}u'/2}\sqrt{\upmu}\,\mathcal{P}^{N_{\text{top}}}\mathcal{C}\|^2_{L^2(\Sigma_t)}$.) ☐

11C2. *Controlling* $\operatorname{curl} \mathcal{P}^{N_{\text{top}}} S$ *and* $\operatorname{div} \mathcal{P}^{N_{\text{top}}} S$.

Lemma 11.8 (L^2 estimates for the Euclidean curl of the derivatives of S). *Let* $\mathfrak{c} \geq 0$ *be a real number. The following estimate holds for all* $t \in [0, T_{(\text{Boot})})$, *where the implicit constant is independent of* $\mathfrak{c}$:

$$\|e^{-\mathfrak{c}u'/2}\sqrt{\upmu} \operatorname{curl} \mathcal{P}^{N_{\text{top}}} S\|^2_{L^2(\Sigma_t)} \lesssim \mathring{\epsilon}^3 \upmu_\star^{-2M_*+0.8}(t).$$

Proof. By (5-4b), $\operatorname{curl} S = 0$. Hence, using Lemma 11.5 and the bootstrap assumption (6-7),

$$\begin{aligned} |\upmu \operatorname{curl} \mathcal{P}^{N_{\text{top}}} S| &= |[\upmu \operatorname{curl}, \mathcal{P}^{N_{\text{top}}}]S| \\ &\lesssim |\mathcal{P}^{\leq N_{\text{top}}} S| + |\mathcal{P}^{\leq N_{\text{top}}-1}\breve{X} S| + \mathring{\epsilon}(|\mathcal{P}^{[2,N_{\text{top}}]}(\upmu, L^i, \Psi)| + |\mathcal{P}^{[2,N_{\text{top}}-1]}\breve{X}\Psi|). \end{aligned} \tag{11-17}$$

The only new terms here compared to (11-12) and (11-13) are $|\mathcal{P}^{\leq N_{\text{top}}} S|$ and $|\mathcal{P}^{\leq N_{\text{top}}-1}\breve{X} S|$, which can be handled using Propositions 9.4 and 10.1 in the same way that we handled the corresponding terms $\upmu_\star^{-1}\|\sqrt{\upmu}\mathcal{P}^{\leq N_{\text{top}}}\Omega\|_{L^2(\Sigma_t)}$ and $\upmu_\star^{-1}\|\sqrt{\upmu}\mathcal{P}^{\leq N_{\text{top}}-1}\breve{X}\Omega\|_{L^2(\Sigma_t)}$ in the proof of Lemma 11.6. ☐

Lemma 11.9 (L^2 estimates for the Euclidean divergence of the derivatives of S in terms of the derivatives of $\mathcal{D}$). *Let* $\mathfrak{c} \geq 0$ *be a real number. The following estimate holds for all* $t \in [0, T_{(\text{Boot})})$, *where the implicit constants are independent of* $\mathfrak{c}$:

$$\|e^{-\mathfrak{c}u'/2}\sqrt{\upmu} \operatorname{div} \mathcal{P}^{N_{\text{top}}} S\|^2_{L^2(\Sigma_t)} \lesssim \mathring{\epsilon}^3 \upmu_\star^{-2M_*+0.8} + \|e^{-\mathfrak{c}u'/2}\sqrt{\upmu}\,\mathcal{P}^{N_{\text{top}}}\mathcal{D}\|^2_{L^2(\Sigma_t)}.$$

Proof. Using Lemma 11.5 and the bootstrap assumption (6-7), we find that

$$|[\upmu \operatorname{div}, \mathcal{P}^{N_{\text{top}}}]S| \lesssim \text{the right-hand side of (11-17).}$$

Therefore, we can therefore handle $|[\upmu \operatorname{div}, \mathcal{P}^{N_{\text{top}}}]S|$ by using the same arguments we gave in the proof of Lemma 11.8.

We then express div S in terms of $\mathcal{D}$ using (2-5b) and use Lemma 2.22, the fact that the Cartesian component functions X^1, X^2, X^3 are smooth functions of the L^i and Ψ (see (2-23)), and the L^∞ bounds in (6-3)–(6-5), (6-7), (6-8), and Proposition 8.6 to deduce

$$\begin{aligned}|\mathcal{P}^{N_{\mathrm{top}}}(\mu \operatorname{div} S)| &\le |\mathcal{P}^{N_{\mathrm{top}}}(\mu\exp(2\rho)\mathcal{D})| + |\mathcal{P}^{N_{\mathrm{top}}}(\mu\exp(2\rho)S^a\partial_a\rho)| \\ &\lesssim \mu|\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D}| + |\mathcal{P}^{\le N_{\mathrm{top}}-1}\mathcal{D}| + |\mathcal{P}^{\le N_{\mathrm{top}}}S| \\ &\qquad + \mathring{\epsilon}(|\mathcal{P}^{[2,N_{\mathrm{top}}]}(\mu, L^i)| + \mu|\mathcal{P}^{N_{\mathrm{top}}+1}\Psi| + |\mathcal{P}^{[2,N_{\mathrm{top}}]}\Psi| + |\mathcal{P}^{[1,N_{\mathrm{top}}]}\breve{X}\Psi|).\end{aligned}$$

The new terms here compared to (11-12) and (11-13) are $|\mathcal{P}^{\le N_{\mathrm{top}}}S|$, which we handled just below (11-17), and $\mu|\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D}|$ and $|\mathcal{P}^{\le N_{\mathrm{top}}-1}\mathcal{D}|$, which can be treated using the same arguments we used to handle the terms $\mu|\mathcal{P}^{N_{\mathrm{top}}}\mathcal{C}|$ and $|\mathcal{P}^{\le N_{\mathrm{top}}-1}\mathcal{C}|$ in our proof of Lemma 11.6. Hence, the weighted, squared $L^2(\Sigma_t)$ norms corresponding to these new terms are bounded above by $\mathring{\epsilon}^3\mu_\star^{-2M_*+0.8} + \|e^{-\mathfrak{c}u'/2}\sqrt{\mu}\,\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D}\|^2_{L^2(\Sigma_t)}$. □

11C3. *Proving the elliptic estimates.* We now combine Lemmas 11.6–11.9 and the elliptic estimates in Proposition 11.4 to obtain the following proposition.

Proposition 11.10 (preliminary top-order elliptic estimates for Ω and S). *Let $\mathfrak{c} \ge 0$ be a real number. The following estimates hold for all $t \in [0, T_{(\mathrm{Boot})})$, where the implicit constants are independent of $\mathfrak{c}$:*

$$\|e^{-\mathfrak{c}u/2}\sqrt{\mu}\underline{\partial}\mathcal{P}^{N_{\mathrm{top}}}\Omega\|^2_{L^2(\Sigma_t)} \lesssim \mathring{\epsilon}^3(1+\mathfrak{c}^2)\mu_\star^{-2M_*+0.8}(t) + \|e^{-\mathfrak{c}u/2}\sqrt{\mu}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{C}\|^2_{L^2(\Sigma_t)}, \tag{11-18}$$
$$\|e^{-\mathfrak{c}u/2}\sqrt{\mu}\underline{\partial}\mathcal{P}^{N_{\mathrm{top}}}S\|^2_{L^2(\Sigma_t)} \lesssim \mathring{\epsilon}^3(1+\mathfrak{c}^2)\mu_\star^{-2M_*+0.8}(t) + \|e^{-\mathfrak{c}u/2}\sqrt{\mu}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{D}\|^2_{L^2(\Sigma_t)}. \tag{11-19}$$

Proof. Applying first Proposition 11.4, and then Lemmas 11.6, 11.7, Proposition 9.4 (and using $e^{-\mathfrak{c}u/2} \le 1$), we obtain

$$\begin{aligned}&\|e^{-\mathfrak{c}u/2}\sqrt{\mu}\underline{\partial}\mathcal{P}^{N_{\mathrm{top}}}\Omega\|^2_{L^2(\Sigma_t)} \\ &\quad\lesssim \|e^{-\mathfrak{c}u/2}\sqrt{\mu}\operatorname{div}\mathcal{P}^{N_{\mathrm{top}}}\Omega\|^2_{L^2(\Sigma_t)} + \|e^{-\mathfrak{c}u/2}\sqrt{\mu}\operatorname{curl}\mathcal{P}^{N_{\mathrm{top}}}\Omega\|^2_{L^2(\Sigma_t)} + \mathfrak{c}^2\mu_\star^{-2}(t)\|e^{-\mathfrak{c}u/2}\sqrt{\mu}\mathcal{P}^{N_{\mathrm{top}}}\Omega\|^2_{L^2(\Sigma_t)} \\ &\quad\lesssim \mathring{\epsilon}^3(1+\mathfrak{c}^2)\mu_\star^{-2M_*+0.8}(t) + \|e^{-\mathfrak{c}u/2}\sqrt{\mu}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{C}\|^2_{L^2(\Sigma_t)},\end{aligned}$$

which proves (11-18). The proof of (11-19) is similar, except we use Lemmas 11.8, 11.9 instead of Lemmas 11.6, 11.7. □

11D. *Putting everything together.*

Proposition 11.11 (the main top-order estimates for the modified fluid variables). *The following estimate holds for every $(t, u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$:*

$$\mathbb{C}_{N_{\mathrm{top}}}(t,u) + \mathbb{D}_{N_{\mathrm{top}}}(t,u) \lesssim \mathring{\epsilon}^3\mu_\star^{-2M_*+0.8}(t).$$

Proof. Step 1: controlling $\|e^{-\mathfrak{c}u/2}\sqrt{\mu}\underline{\partial}\mathcal{P}^{N_{\mathrm{top}}}(\Omega, S)\|^2_{L^2(\Sigma_t)}$ via Grönwall-type argument. Given $\varsigma > 0$, we first apply Proposition 11.10 and then use[56] Proposition 11.2 (for[57] $u = U_0$) to deduce that if $\mathfrak{c} > 0$ is

[56] Here, we again relabeled the ς from Proposition 11.2

[57] Note that in view of the fact that Ω, S are compactly supported in $u \in [0, U_0]$ (by Lemma 7.1), it follows that the integral on $\Sigma_t^{U_0}$ is the same as the integral on Σ_t.

sufficiently large (depending on ς), then the following estimate holds, where the constants $C > 0$ and $C_* > 0$ are independent of $\mathfrak{c}$ and ς:

$$\begin{aligned}\|e^{-\mathfrak{c}u/2}\sqrt{\upmu}\underline{\partial}\mathcal{P}^{N_{\mathrm{top}}}(\varOmega, S)\|^2_{L^2(\Sigma_t)} &\\ \leq C\mathring{\epsilon}^3(1+\mathfrak{c}^2)\upmu_\star^{-2M_*+0.8}(t) + C\|e^{-\mathfrak{c}u/2}\sqrt{\upmu}\mathcal{P}^{N_{\mathrm{top}}}(\mathcal{C}, \mathcal{D})\|^2_{L^2(\Sigma_t)} &\\ \leq C_*\mathring{\epsilon}^3(1+\mathfrak{c}^2)\upmu_\star^{-2M_*+0.8}(t) + \varsigma\int_{t'=0}^{t'=t}\frac{1}{\upmu_\star(t')}\|e^{-\mathfrak{c}u/2}\sqrt{\upmu}\underline{\partial}\mathcal{P}^{N_{\mathrm{top}}}(\varOmega, S)\|^2_{L^2(\Sigma_{t'})}\,dt'. &\end{aligned} \tag{11-20}$$

We clarify that it is only for notational convenience for the argument in (11-21)–(11-23) below that we have used the symbol $C_* > 0$ to denote the fixed constant on the last line of (11-20).

We now argue by a continuity argument to show that, after choosing ς smaller and $\mathfrak{c}$ larger if necessary, (11-20) implies the estimate

$$\|e^{-\mathfrak{c}u/2}\sqrt{\upmu}\underline{\partial}\mathcal{P}^{N_{\mathrm{top}}}(\varOmega, S)\|^2_{L^2(\Sigma_t)} \leq 2C_*\mathring{\epsilon}^3(1+\mathfrak{c}^2)\upmu_\star^{-2M_*+0.8}(t). \tag{11-21}$$

If it is not the case that (11-21) holds on $[0, T_{(\mathrm{Boot})})$, then by continuity, there exists $T_* \in [0, T_{(\mathrm{Boot})})$ such that (11-21) holds for all $t \in [0, T_*]$ and such that

$$\|e^{-\mathfrak{c}u/2}\sqrt{\upmu}\underline{\partial}\mathcal{P}^{N_{\mathrm{top}}}(\varOmega, S)\|^2_{L^2(\Sigma_{T_*})} = 2C_*\mathring{\epsilon}^3(1+\mathfrak{c}^2)\upmu_\star^{-2M_*+0.8}(T_*). \tag{11-22}$$

However, plugging the estimate (11-21) (which by assumption holds for $t \in [0, T_*]$) into the integral in (11-20), using Proposition 8.11 (and $M_* \geq 1$) to integrate away a negative power of $\upmu_\star$, and finally choosing ς sufficiently small, we obtain that for $t \in [0, T_*]$, we have

$$\|e^{-\mathfrak{c}u/2}\sqrt{\upmu}\underline{\partial}\mathcal{P}^{N_{\mathrm{top}}}(\varOmega, S)\|^2_{L^2(\Sigma_t)} \leq \tfrac{3}{2}C_*\mathring{\epsilon}^3(1+\mathfrak{c}^2)\upmu_\star^{-2M_*+0.8}(t), \tag{11-23}$$

which obviously contradicts (11-22) when $t = T_*$. It therefore follows that our desired estimate (11-21) holds for all $t \in [0, T_{(\mathrm{Boot})})$.

Step 2: deducing the estimates for $\mathbb{C}_{N_{\mathrm{top}}}(t, u)$ and $\mathbb{D}_{N_{\mathrm{top}}}(t, u)$. At this point, we can fix the constants $\mathfrak{c}$, ς, which we will absorb into the ensuring generic constants C. Moreover, since $u \in [0, U_0]$ on the support of $\varOmega$ and S (by Lemma 7.1), we will also absorb the weights $e^{-\mathfrak{c}u/2}$ into the constants. Hence, plugging (11-21) into the right-hand side of (11-1) and then using Proposition 8.11, we obtain

$$\begin{aligned}\mathbb{C}_{N_{\mathrm{top}}}(t, u) + \mathbb{D}_{N_{\mathrm{top}}}(t, u) &\lesssim \mathring{\epsilon}^3\upmu_\star^{-2M_*+0.8}(t) + \int_{t'=0}^{t'=t}\frac{1}{\upmu_\star(t')}\|e^{-\mathfrak{c}u'/2}\sqrt{\upmu}\underline{\partial}\mathcal{P}^{N_{\mathrm{top}}}(\varOmega, S)\|^2_{L^2(\Sigma_{t'})}\,dt' \\ &\lesssim \mathring{\epsilon}^3\upmu_\star^{-2M_*+0.8}(t) + \mathring{\epsilon}^3\int_{t'=0}^{t'=t}\upmu_\star^{-2M_*-0.2}(t')\,dt' \lesssim \mathring{\epsilon}^3\upmu_\star^{-2M_*+0.8}(t),\end{aligned} \tag{11-24}$$

as desired. □

12. Wave estimates for the fluid variables

We continue to work under the assumptions of Theorem 6.3.

In this section, we derive a priori energy estimates for the wave variables, which will in particular yield strict improvements of the bootstrap assumptions (6-1)–(6-2). In Section 12A, we start by providing a somewhat general[58] "auxiliary" proposition, which yields energy estimates for solutions to *inhomogeneous* quasilinear wave equations *in terms of norms of the inhomogeneity*. The difficult aspect of the proof is that we have to close the estimates even though μ can be tending towards 0, that is, even though the shock may be forming. We delay discussing the proof of the auxiliary proposition until the Appendix; as we will explain, modulo small modifications based on established techniques, the proposition was proved as [36, Proposition 14.1] (see also [52, Proposition 14.1]). Then, in Section 12B, we bound the specific inhomogeneous terms that are relevant for our main results, that is, the inhomogeneous terms on the right-hand sides of the fluid wave equations (5-1a)–(5-1c). Finally, in Section 12C, we prove the final a priori energy estimates.

12A. ***The main estimates for inhomogeneous covariant wave equations.*** In this section, we state the "auxiliary" Proposition 12.1, which yields energy estimates for solutions to the fluid wave equations. In this section, we ignore the precise structure of the inhomogeneous terms and simply denote them by $\mathfrak{G}$. That is, we state the estimates of Proposition 12.1 in terms of various norms of $\mathfrak{G}$. Later on, in Proposition 12.7, we will control the relevant norms of $\mathfrak{G}$ to obtain our final a priori energy estimates for the wave variables. Proposition 12.1 is of independent interest in the sense that with small modifications, it could be used to study shock formation for compressible Euler flow with given smooth forcing terms.

Proposition 12.1 (the main estimates for the inhomogeneous geometric wave equations). *Let* $\vec{\Psi} \doteq (\Psi_1, \Psi_2, \Psi_3, \Psi_4, \Psi_5) \doteq (\mathcal{R}_{(+)}, \mathcal{R}_{(-)}, v^2, v^3, s)$, *as in* (2-3). *Recall that the* Ψ_ι *are solutions to the inhomogeneous covariant wave system*

$$\mu \square_{g(\vec{\Psi})} \Psi_\iota = \mathfrak{G}_\iota,$$

where $\vec{\mathfrak{G}} = (\mathfrak{G}_1, \mathfrak{G}_2, \mathfrak{G}_3, \mathfrak{G}_4, \mathfrak{G}_5)$ *is the array whose entries are the product of* μ *and the inhomogeneous terms on the right-hand sides of the five scalar wave equations* (5-1a)–(5-1c). *Assume that the following smallness bound holds*:[59]

$$\|\mathcal{P}^{\leq \lceil N_{\mathrm{top}}/2 \rceil} \vec{\mathfrak{G}}\|_{L^\infty(\mathcal{M}_{t,u})} \leq \mathring{\epsilon}^{1/2}. \tag{12-1}$$

Then there exists an absolute constant $M_* \in \mathbb{N}$, *independent of the equation of state and all other parameters in the problem, such that the following hold. As in Theorem 6.3, let* $T_{(\mathrm{Boot})} \in [0, 2\mathring{\delta}_*^{-1}]$, *and assume that*:

(1) *The bootstrap assumptions* (6-1)–(6-8) *all hold for all* $t \in [0, T_{(\mathrm{Boot})})$, *where we recall that in the bootstrap assumptions*, N_{top} *is any integer satisfying* $N_{\mathrm{top}} \geq 2M_* + 10$.

(2) *In* (6-3), *the parameter* $\mathring{\alpha}$ *is sufficiently small in a manner only on the equation of state and* $\bar{\varrho}$.

[58] Using a slight reorganization of the paper, these estimates could be upgraded so that they are "black box" estimates for inhomogeneous wave equations. Given the setup of this paper, they are not quite black box estimates because the proofs rely on the estimates of Section 8, some of which (e.g., some of the estimates in Proposition 8.10) depend on the structure of the inhomogeneous terms in the wave equations.

[59] We clarify that in our main results, in the proof of Proposition 12.7, we will show that the smallness assumption (12-1) is satisfied for the particular inhomogeneous terms $\vec{\mathfrak{G}}$ stated in the hypotheses of the proposition. However, for the purposes of proving Proposition 12.1, the precise structure of $\vec{\mathfrak{G}}$ is not important.

(3) *The parameter* $\mathring{\epsilon} > 0$ *in* (6-1)–(6-8) *satisfies* $\mathring{\epsilon}^{1/2} \leq \mathring{\alpha}$ *and is sufficiently small in a manner that depends only on the equation of state*, N_{top}, $\bar{\varrho}$, $\mathring{\sigma}$, $\mathring{\delta}$, *and* $\mathring{\delta}_*^{-1}$.

(4) *The soft bootstrap assumptions stated in Section 6A1 hold* (*including* $\mu > 0$ *in* $[0, T_{(\text{Boot})}) \times \mathbb{R} \times \mathbb{T}^2$).

Then the following estimates hold for every $(t, u) \in [0, T_{(\text{Boot})}) \times [0, U_0]$, *where* $\mu_\star$ *is defined in Definition 2.16*:

(1) *The top- and penultimate-order wave energies defined in* (3-2e) *obey the estimates*

$$\begin{aligned}\sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-1.8}(\hat{t})\mathbb{W}_{[1,N_{\text{top}}]}(\hat{t}, u) + \sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-3.8}(\hat{t})\mathbb{W}_{[1,N_{\text{top}}-1]}(\hat{t}, u) \\ \lesssim \mathring{\epsilon}^2 + \sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-1.8}(\hat{t}) \int_{t'=0}^{t'=\hat{t}} \mu_\star^{-3/2}(t')\left\{\int_{s=0}^{s=t'} \|\mathcal{P}^{[1,N_{\text{top}}]}\vec{\mathfrak{G}}\|_{L^2(\Sigma_s^u)}\, ds\right\}^2 dt' \\ + \sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-1.8}(\hat{t})\|(|L\mathcal{P}^{[1,N_{\text{top}}]}\Psi| + |\breve{X}\mathcal{P}^{[1,N_{\text{top}}]}\Psi|)|\mathcal{P}^{[1,N_{\text{top}}]}\vec{\mathfrak{G}}|\|_{L^1(\mathcal{M}_{\hat{t},u})} \\ + \sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-3.8}(\hat{t})\|(|L\mathcal{P}^{[1,N_{\text{top}}-1]}\Psi| + |\breve{X}\mathcal{P}^{[1,N_{\text{top}}-1]}\Psi|)|\mathcal{P}^{[1,N_{\text{top}}-1]}\vec{\mathfrak{G}}|\|_{L^1(\mathcal{M}_{\hat{t},u})}. \end{aligned} \tag{12-2}$$

(2) *For* $1 \leq N \leq N_{\text{top}} - 1$, *the lower-order wave energies* $\mathbb{W}_{[1,N]}$ *defined in* (3-2c) *obey the estimates*

$$\begin{aligned}\mathbb{W}_{[1,N]}(t, u) \lesssim \mathring{\epsilon}^2 + \max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+1.8}(t)\}\Big(\sup_{s\in[0,t]} \min\{1, \mu_\star^{2M_*-2N_{\text{top}}+2N+0.2}(s)\}\mathbb{Q}_{[1,N+1]}(s)\Big) \\ + \|(|L\mathcal{P}^{[1,N]}\Psi| + |\breve{X}\mathcal{P}^{[1,N]}\Psi|)|\mathcal{P}^{[1,N]}\vec{\mathfrak{G}}|\|_{L^1(\mathcal{M}_{t,u})}. \end{aligned} \tag{12-3}$$

Remark 12.2. The proof of Proposition 12.1 follows from almost exactly the same arguments used in the proof of [36, Proposition 14.1]. The only differences are the following two changes:

(1) We have to track the influence of the inhomogeneous terms $\mathfrak{G}$ on the estimates.

(2) In three dimensions, the second fundamental form of the null hypersurfaces of the acoustical metric has three (as opposed to one) independent components. This necessitates an additional elliptic estimate that was not needed in the two-dimensional case treated in [36]. This elliptic estimate is standard; see [15; 17; 33].

These differences necessitate minor modifications to the proof of [36, Proposition 14.1]. We will sketch them in the Appendix.

Remark 12.3 (additional term in the top-order estimate). In Proposition 12.1, the inhomogeneous term $\vec{\mathfrak{G}}$ makes an additional appearance in the top- and penultimate-order estimates as compared to the estimates of all the lower orders. By "additional appearance," we are referring to the double time integral, which comes from a difficult top-order commutator term that depends on the acoustic geometry; this difficult term has to be controlled by first integrating a transport equation, which explains the double time-integration; see the Appendix.

12B. ***Estimates for the inhomogeneous terms.*** We start by controlling the null forms in the wave equations.

Proposition 12.4 (control of wave equation error terms involving null forms). *For* $\mathfrak{Q} \in \{\mathfrak{Q}^i_{(v)}, \mathfrak{Q}_{(\pm)}\}$ (*see* (5-6a), (5-6b)) *and* $1 \leq N \leq N_{\text{top}}$, *the following hold for all* $(t, u) \in [0, T_{(\text{Boot})}) \times [0, U_0]$ *and for all* $\varsigma \in (0, 1]$, *where the implicit constants are independent of* ς:

$$\begin{aligned}
&\|(|L\mathcal{P}^{[1,N]}\Psi| + |\breve{X}\mathcal{P}^{[1,N]}\Psi|)\mathcal{P}^{[1,N]}(\upmu\mathfrak{Q})\|_{L^1(\mathcal{M}_{t,u})} \\
&\quad \lesssim \mathring{\epsilon}^2 \max\{1, \upmu_\star^{-2M_*+2N_{\text{top}}-2N+1.8}(t)\} \\
&\qquad\qquad + \varsigma\mathbb{K}_{[1,N]}(t,u) + (1+\varsigma^{-1})\left(\int_{u'=0}^{u'=u} \mathbb{F}_{[1,N]}(t,u')\,du' + \int_{t'=0}^{t'=t} \mathbb{E}_{[1,N]}(t',u)\,dt'\right)
\end{aligned} \tag{12-4}$$

and

$$\begin{aligned}
&\int_{t'=0}^{t'=t} \upmu_\star^{-3/2}(t')\left\{\int_{s=0}^{s=t'} \|\mathcal{P}^{[1,N_{\text{top}}]}(\upmu\mathfrak{Q})\|_{L^2(\Sigma_s^u)}\,ds\right\}^2 dt' \\
&\qquad\qquad \lesssim \mathring{\epsilon}^2\upmu_\star^{-2M_*+1.8}(t) + \int_{t'=0}^{t'=t} \upmu_\star^{-3/2}(t')\left\{\int_{s=0}^{s=t'} \upmu_\star^{-1/2}(s)\mathbb{E}^{1/2}_{[1,N_{\text{top}}]}(s)\,ds\right\}^2 dt'.
\end{aligned} \tag{12-5}$$

Proof. Step 1: proof of (12-4). To bound the left-hand side of (12-4), we use the Cauchy–Schwarz and the Young inequalities to obtain, for any $\varsigma > 0$,

$$\begin{aligned}
&\|(|L\mathcal{P}^{[1,N]}\Psi| + |\breve{X}\mathcal{P}^{[1,N]}\Psi|)\mathcal{P}^{[1,N]}(\upmu\mathfrak{Q})\|_{L^1(\mathcal{M}_{t,u})} \\
&\quad \lesssim (1+\varsigma^{-1})\left(\int_{u'=0}^{u'=u} \mathbb{F}_{[1,N]}(t,u')\,du' + \int_{t'=0}^{t'=t} \mathbb{E}_{[1,N]}(t',u)\,dt'\right) + \varsigma\|\mathcal{P}^{[1,N]}(\upmu\mathfrak{Q})\|^2_{L^2(\mathcal{M}_{t,u})}.
\end{aligned} \tag{12-6}$$

By inspection, it can be checked that $\mathfrak{Q}$ is a g-null form (see Definition 8.1) that is quadratic in the wave variables. Hence, applying (10-6) with $\phi^{(1)}, \phi^{(2)} = \Psi$, $\mathfrak{d}^{(1,1)}, \mathfrak{d}^{(2,1)} \lesssim 1$, $\mathfrak{d}^{(1,2)}, \mathfrak{d}^{(2,2)} \lesssim \mathring{\epsilon}^{1/2}$ (which is justified by the bootstrap assumptions (6-3)–(6-5)), we obtain

$$|\mathcal{P}^{[1,N]}(\upmu\mathfrak{Q})| \lesssim |\mathcal{P}^{[2,N+1]}\Psi| + \mathring{\epsilon}^{1/2}\{|\mathcal{P}^{[1,N]}\breve{X}\Psi| + |\mathcal{P}\Psi| + |\mathcal{P}^{[2,N]}(\upmu, L^i)|\}. \tag{12-7}$$

To bound (12-7) in $L^2(\mathcal{M}_{t,u})$, we control $|\mathcal{P}^{[2,N+1]}\Psi|$ by the energies (3-2a)–(3-2c), control $|\mathcal{P}^{[1,N]}\breve{X}\Psi|$ by (8-38), bound $|\mathcal{P}\Psi|$ by (6-5), and $|\mathcal{P}^{[2,N]}(\upmu, L^i)|$ by Proposition 8.14. We thus obtain the following bound for any $\varsigma \in (0, 1]$, where the implicit constants are independent of ς:

$$\begin{aligned}
\varsigma\|\mathcal{P}^{[1,N]}(\upmu\mathfrak{Q})\|^2_{L^2(\mathcal{M}_{t,u})} &\lesssim \varsigma\left\{\mathbb{K}_{[1,N]}(t,u) + \int_{u'=0}^{u'=u} \mathbb{F}_{[1,N]}(t,u')\,du' + \int_{t'=0}^{t'=t} \mathbb{E}_{[1,N]}(t',u)\,dt'\right\} \\
&\qquad + \mathring{\epsilon}^2 \int_{t'=0}^{t'=t} \max\{1, \upmu_\star^{-2M_*+2N_{\text{top}}-2N+1.8}(t')\}\,dt' + \mathring{\epsilon}^2 \\
&\lesssim \mathring{\epsilon}^2 \max\{1, \upmu_\star^{-2M_*+2N_{\text{top}}-2N+1.8}(t)\} \\
&\qquad + \varsigma\left\{\mathbb{K}_{[1,N]}(t,u) + \int_{u'=0}^{u'=u} \mathbb{F}_{[1,N]}(t,u')\,du' + \int_{t'=0}^{t'=t} \mathbb{E}_{[1,N]}(t',u)\,dt'\right\},
\end{aligned} \tag{12-8}$$

where in the last line, we have used Proposition 8.12.

Putting (12-6)–(12-8) together, we obtain (12-4).

Step 2: proof of (12-5). We begin with (12-7) when $N = N_{\text{top}}$. Notice that unlike in Step 1, we now have to control $|\mathcal{P}^{[2,N+1]}\Psi|$ only with the $\mathbb{E}$ (but not $\mathbb{F}$ and $\mathbb{K}$) energy (since we need an estimate on a fixed-t

hypersurface). This gives a $\upmu_\star^{-1/2}$ degeneration. The other terms can be controlled by using arguments similar to the ones we used in Step 1. In total, for $0 \leq s \leq t' \leq t$, we have

$$\|\mathcal{P}^{[1,N_{\mathrm{top}}]}(\upmu\mathfrak{Q})\|_{L^2(\Sigma_s^u)} \lesssim \upmu_\star^{-1/2}(s)\mathbb{E}^{1/2}_{[1,N_{\mathrm{top}}]}(s) + \mathring{\epsilon}\max\{1, \upmu_\star^{-M_*+0.9}(s)\}. \tag{12-9}$$

Finally, integrating with respect to time and using Proposition 8.11, we obtain (12-5). □

Next, we control the easy linear terms in the wave equations.

Proposition 12.5 (control of wave equation error terms involving easy linear inhomogeneous terms). *For* $\mathfrak{L} \in \{\mathfrak{L}^i_{(v)}, \mathfrak{L}_{(\pm)}, \mathfrak{L}_{(s)}\}$ *(see* (5-7a), (5-7b), (5-7c)*) and* $1 \leq N \leq N_{\mathrm{top}}$, *the following holds for all* $(t,u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$:

$$\|(|L\mathcal{P}^{[1,N]}\Psi| + |\breve{X}\mathcal{P}^{[1,N]}\Psi|)\mathcal{P}^{[1,N]}(\upmu\mathfrak{L})\|_{L^1(\mathcal{M}_{t,u})} \lesssim \textit{the right-hand side of } (12\text{-}4), \tag{12-10}$$

and

$$\int_{t'=0}^{t'=t} \upmu_\star^{-3/2}(t')\left\{\int_{s=0}^{s=t'} \|\mathcal{P}^{[1,N_{\mathrm{top}}]}(\upmu\mathfrak{L})\|_{L^2(\Sigma_s^u)}\,ds\right\}^2 dt' \lesssim \textit{the right-hand side of } (12\text{-}5). \tag{12-11}$$

Proof. We first pointwise bound $\upmu\mathfrak{L} \in \{\upmu\mathfrak{L}^i_{(v)}, \upmu\mathfrak{L}_{(\pm)}, \upmu\mathfrak{L}_{(s)}\}$ in a similar manner[60] to (12-7):

$$|\mathcal{P}^{[1,N]}(\upmu\mathfrak{L})| \lesssim |\mathcal{P}^{\leq N}(\Omega, S)| + \text{terms already in (12-7)}. \tag{12-12}$$

<u>Proof of (12-10)</u>. The terms in (12-12) that are already in (12-7) can of course be controlled as in Proposition 12.4. We therefore focus on $|\mathcal{P}^{\leq N}(\Omega, S)|$, for which we have the following estimate using the Cauchy–Schwarz and Hölder inequalities and Proposition 9.4:

$$\begin{aligned}
&\|(|L\mathcal{P}^{[1,N]}\Psi| + |\breve{X}\mathcal{P}^{[1,N]}\Psi|)\mathcal{P}^{\leq N}(\Omega, S)\|_{L^1(\mathcal{M}_{t,u})} \\
&\quad\lesssim \|L\mathcal{P}^{[1,N]}\Psi\|^2_{L^2(\mathcal{M}_{t,u})} + \|\breve{X}\mathcal{P}^{[1,N]}\Psi\|^2_{L^2(\mathcal{M}_{t,u})} + \int_0^u \|\mathcal{P}^{\leq N}(\Omega, S)\|^2_{L^2(\mathcal{F}^t_{u'})}\,du' \\
&\quad\lesssim \|L\mathcal{P}^{[1,N]}\Psi\|^2_{L^2(\mathcal{M}_{t,u})} + \|\breve{X}\mathcal{P}^{[1,N]}\Psi\|^2_{L^2(\mathcal{M}_{t,u})} + \mathring{\epsilon}^3\max\{1, \upmu_\star^{-2M_*+2N_{\mathrm{top}}+2N+2.8}(t)\},
\end{aligned} \tag{12-13}$$

which can indeed be bounded above by the right-hand side of (12-4) as claimed.

<u>Proof of (12-11)</u>. Again, we only focus on the $|\mathcal{P}^{\leq N_{\mathrm{top}}}(\Omega, S)|$ term in (12-12). Using the definitions of the $\mathbb{V}$ and $\mathbb{S}$ norms and Propositions 8.11 and 9.4, we deduce

$$\begin{aligned}
&\int_{t'=0}^{t'=t} \upmu_\star^{-3/2}(t')\left\{\int_{s=0}^{s=t'} \|\mathcal{P}^{\leq N_{\mathrm{top}}}(\Omega, S)\|_{L^2(\Sigma_s^u)}\,ds\right\}^2 dt' \\
&\qquad\lesssim \int_{t'=0}^{t'=t} \upmu_\star^{-3/2}(t')\left\{\int_{s=0}^{s=t'} \upmu_\star^{-1/2}(s)[\mathbb{V}^{1/2}_{\leq N_{\mathrm{top}}}(s) + \mathbb{S}^{1/2}_{\leq N_{\mathrm{top}}}(s)]\,ds\right\}^2 dt' \\
&\qquad\lesssim \mathring{\epsilon}^3 \int_{t'=0}^{t'=t} \upmu_\star^{-3/2}(t')\left\{\int_{s=0}^{s=t'} \upmu_\star^{-M_*+0.9}(s)\,ds\right\}^2 dt' \\
&\qquad\lesssim \mathring{\epsilon}^3\max\{1, \upmu_\star^{-2M_*+3.3}(t)\} \lesssim \mathring{\epsilon}^2\max\{1, \upmu_\star^{-2M_*+1.8}(t)\},
\end{aligned} \tag{12-14}$$

which can indeed be bounded above by the right-hand side of (12-5) as claimed. □

[60] In fact, we can even do better than terms in (12-7) because of the extra smallness in $\mathring{\epsilon}$ we have from the bootstrap assumptions. However, we do not need this improvement for our proof.

Finally, we consider the linear terms involving $\mathcal{C}$ and $\mathcal{D}$.

Proposition 12.6 (control of wave equation error terms involving $\mathcal{C}$ and $\mathcal{D}$). *For*

$$\mathfrak{M} \in \left\{ c^2 \exp(2\rho)\mathcal{C}^i,\ c\exp(\rho)\frac{p_{;s}}{\bar{\varrho}}\mathcal{D},\ c^2\exp(2\rho)\mathcal{D},\ F_{;s}c^2\exp(2\rho)\mathcal{D} \right\}$$

(*cf. main terms in* (5-1a)–(5-1c)) *and* $1 \le N \le N_{\text{top}}$, *the following hold for all* $(t,u) \in [0, T_{(\text{Boot})}) \times [0, U_0]$:

$$\|(|L\mathcal{P}^{[1,N]}\Psi| + |\breve{X}\mathcal{P}^{[1,N]}\Psi|)\mathcal{P}^{[1,N]}(\mu\mathfrak{M})\|_{L^1(\mathcal{M}_{t,u})} \lesssim \textit{the right-hand side of } (12\text{-}4), \tag{12-15}$$

$$\int_{t'=0}^{t'=t} \mu_\star^{-3/2}(t') \left\{ \int_{s=0}^{s=t'} \|\mathcal{P}^{[1,N_{\text{top}}]}(\mu\mathfrak{M})\|_{L^2(\Sigma_s^u)}\, ds \right\}^2 dt' \lesssim \textit{the right-hand side of } (12\text{-}5). \tag{12-16}$$

Proof. We first use the bootstrap assumptions (6-3)–(6-5) and (6-8) and Proposition 8.6 to deduce

$$|\mathcal{P}^N(\mu\mathfrak{M})| \lesssim \underbrace{\mu|\mathcal{P}^N(\mathcal{C},\mathcal{D})|}_{\doteq I} + \underbrace{|\mathcal{P}^{\le N-1}(\mathcal{C},\mathcal{D})|}_{\doteq II} + \text{terms already in (12-7)}. \tag{12-17}$$

Step 1: proof of (12-15). The terms already in (12-7) were handled in the proof of (12-4), so we only have to handle I and II in (12-17). We will use slightly different arguments for each of these two terms. For I, we have[61] by the Cauchy–Schwarz inequality, Propositions 10.6, 11.11, the bootstrap assumptions (6-1), (6-2), and Propositions 8.6 and 8.11 that

$$\begin{aligned}
&\|(|L\mathcal{P}^{[1,N]}\Psi| + |\breve{X}\mathcal{P}^{[1,N]}\Psi|)\mu\mathcal{P}^N(\mathcal{C},\mathcal{D})\|_{L^1(\mathcal{M}_{t,u})} \\
&\qquad \lesssim \int_{t'=0}^{t'=t} \mathbb{E}_{[1,N]}^{1/2}(t',u)[\mathbb{C}_{\le N}^{1/2} + \mathbb{D}_{\le N}^{1/2}](t',u)\,dt' \\
&\qquad \lesssim \mathring{\epsilon}^{1/2}\mathring{\epsilon}^{3/2} \int_{t'=0}^{t'=t} \max\{1, \mu_\star^{-M_*+N_{\text{top}}-N+0.9}(t')\} \max\{1, \mu_\star^{-M_*+N_{\text{top}}-N+0.4}(t')\}\,dt' \\
&\qquad \lesssim \mathring{\epsilon}^2 \max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+2.3}(t)\}.
\end{aligned} \tag{12-18}$$

For II in (12-17), we use Cauchy–Schwarz and Proposition 10.6 to obtain

$$\begin{aligned}
&\|(|L\mathcal{P}^{[1,N]}\Psi| + |\breve{X}\mathcal{P}^{[1,N]}\Psi|)\mathcal{P}^{\le N-1}(\mathcal{C},\mathcal{D})\|_{L^1(\mathcal{M}_{t,u})} \\
&\qquad \lesssim \|L\mathcal{P}^{[1,N]}\Psi\|^2_{L^2(\mathcal{M}_{t,u})} + \|\breve{X}\mathcal{P}^{[1,N]}\Psi\|^2_{L^2(\mathcal{M}_{t,u})} + \|\mathcal{P}^{\le N-1}(\mathcal{C},\mathcal{D})\|^2_{L^2(\mathcal{M}_{t,u})} \\
&\qquad \lesssim \|L\mathcal{P}^{[1,N]}\Psi\|^2_{L^2(\mathcal{M}_{t,u})} + \|\breve{X}\mathcal{P}^{[1,N]}\Psi\|^2_{L^2(\mathcal{M}_{t,u})} + \int_{u'=0}^{u'=u} [\mathbb{C}_{\le N-1} + \mathbb{D}_{\le N-1}](t,u')\,du' \\
&\qquad \lesssim \|L\mathcal{P}^{[1,N]}\Psi\|^2_{L^2(\mathcal{M}_{t,u})} + \|\breve{X}\mathcal{P}^{[1,N]}\Psi\|^2_{L^2(\mathcal{M}_{t,u})} + \mathring{\epsilon}^3 \max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+2.8}(t)\}.
\end{aligned} \tag{12-19}$$

Finally, we observe that the right-hand side of (12-18) and the right-hand side of (12-19) are less than or equal to the right-hand side of (12-4). We have therefore proved (12-15).

Step 2: proof of (12-16). Returning to (12-17), we again note that we only have to consider terms not already controlled in Proposition 12.4. Applying Propositions 8.6, 8.11, 10.6, and 11.11, we have

[61] Note that it is only at the top $N = N_{\text{top}}$ level that $\mathbb{C}_{\le N}^{1/2}$ and $\mathbb{D}_{\le N}^{1/2}$ is only bounded by $\mu_\star^{-M_*+N_{\text{top}}-N+0.4}(t')$. For $N < N_{\text{top}}$, we have the stronger estimates in Proposition 10.6, which in principle would allow us to avoid controlling the term I separately.

$$\begin{aligned}
&\int_{t'=0}^{t'=t} \mu_\star^{-3/2}(t')\left\{\int_{s=0}^{s=t'} [\|\mu\mathcal{P}^{\le N_{\text{top}}}(\mathcal{C},\mathcal{D})\|_{L^2(\Sigma_s)} + \|\mathcal{P}^{\le N_{\text{top}}-1}(\mathcal{C},\mathcal{D})\|_{L^2(\Sigma_s)}]\,ds\right\}^2 dt' \\
&\qquad \lesssim \int_{t'=0}^{t'=t} \mu_\star^{-3/2}(t')\left\{\int_{s=0}^{s=t'} [\mathbb{C}^{1/2}_{\le N_{\text{top}}} + \mathbb{D}^{1/2}_{\le N_{\text{top}}}](s) + \frac{1}{\mu_\star^{1/2}(s)}[\mathbb{C}^{1/2}_{\le N_{\text{top}}-1} + \mathbb{D}^{1/2}_{\le N_{\text{top}}-1}](s)\,ds\right\}^2 dt' \\
&\qquad \lesssim \mathring{\epsilon}^3 \int_{t'=0}^{t'=t} \mu_\star^{-3/2}(t')\left\{\int_{s=0}^{s=t'} \mu_\star^{-M_*+0.4}(s)\,ds\right\}^2 dt' \lesssim \mathring{\epsilon}^3\mu_\star^{-2M_*+2.3}(t) \lesssim \mathring{\epsilon}^2\mu_\star^{-2M_*+1.8}(t),
\end{aligned}$$

which is therefore bounded above by the right-hand side of (12-5). □

12C. *Putting everything together.*

Proposition 12.7 (main L^2 estimates for the wave variables). *For $1 \le N \le N_{\text{top}}$, the following holds for all $(t,u) \in [0, T_{(\text{Boot})}) \times [0, U_0]$:*

$$\mathbb{W}_{[1,N]}(t,u) \lesssim \mathring{\epsilon}^2 \max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+1.8}(t)\}. \tag{12-20}$$

Proof. We first use the pointwise bounds (12-7), (12-12), (12-17), the bootstrap assumptions (6-5)–(6-8), and Proposition 8.6 to deduce that the assumption (12-1) in Proposition 12.1 on the inhomogeneous terms $\vec{\mathfrak{G}}$, i.e., the terms on the right-hand sides of (5-1a)–(5-1c), is satisfied. Hence, the results of Proposition 12.1 are valid, and we will use them throughout the rest of this proof. We will also silently use the basic fact that $\mu_\star(t,u) \le 1$ and $\mu_\star(t) \le 1$; see Definition 2.16.

Step 1: $N = N_{\text{top}}$. By the top- and penultimate-order general wave estimates (12-2) in Proposition 12.1, the initial data assumptions in (4-1), (4-3a)–(4-4), and the bounds for the inhomogeneous terms in Propositions 12.4–12.6, we obtain the following bound for any $\varsigma \in (0,1]$ (with implicit constants that are independent of ς):

$$\begin{aligned}
&\sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-1.8}(\hat{t})\big(\mathbb{E}_{[1,N_{\text{top}}]}(\hat{t},u)+\mathbb{F}_{[1,N_{\text{top}}]}(\hat{t},u)+\mathbb{K}_{[1,N_{\text{top}}]}(\hat{t},u)\big) \\
&\qquad\qquad + \sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-3.8}(\hat{t})\big(\mathbb{E}_{[1,N_{\text{top}}-1]}(\hat{t},u)+\mathbb{F}_{[1,N_{\text{top}}-1]}(\hat{t},u)+\mathbb{K}_{[1,N_{\text{top}}-1]}(\hat{t},u)\big) \\
&\lesssim \mathring{\epsilon}^2 + \sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-1.8}(\hat{t}) \int_{t'=0}^{t'=\hat{t}} \mu_\star^{-3/2}(t')\left\{\int_{s=0}^{s=t'} \|\mathcal{P}^{[1,N_{\text{top}}]}\vec{\mathfrak{G}}\|_{L^2(\Sigma_s)}\,ds\right\}^2 dt' \\
&\quad + \sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-1.8}(\hat{t}) \|(|L\mathcal{P}^{[1,N_{\text{top}}]}\Psi| + |\breve{X}\mathcal{P}^{[1,N_{\text{top}}]}\Psi|)|\mathcal{P}^{[1,N_{\text{top}}]}\vec{\mathfrak{G}}|\|_{L^1(\mathcal{M}_{\hat{t},u})} \\
&\quad + \sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-3.8}(\hat{t}) \|(|L\mathcal{P}^{[1,N_{\text{top}}-1]}\Psi| + |\breve{X}\mathcal{P}^{[1,N_{\text{top}}-1]}\Psi|)|\mathcal{P}^{[1,N_{\text{top}}-1]}\vec{\mathfrak{G}}|\|_{L^1(\mathcal{M}_{\hat{t},u})} \\
&\lesssim \mathring{\epsilon}^2 + \sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-1.8}(\hat{t}) \int_{t'=0}^{t'=t} \mu_\star^{-3/2}(t',u)\left\{\int_{s=0}^{s=t'} \mu_\star^{-1/2}(s)\mathbb{E}^{1/2}_{[1,N_{\text{top}}]}(s)\,ds\right\}^2 dt' \\
&\quad + \sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-1.8}(\hat{t}) \left\{\varsigma\mathbb{K}_{[1,N_{\text{top}}]}(\hat{t},u) + (1+\varsigma^{-1})\left(\int_{t'=0}^{t'=\hat{t}} \mathbb{E}_{[1,N_{\text{top}}]}(t',u)\,dt' + \int_{u'=0}^{u'=u} \mathbb{F}_{[1,N_{\text{top}}]}(\hat{t},u')\,du'\right)\right\} \\
&\quad + \sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-3.8}(\hat{t}) \left\{\varsigma\mathbb{K}_{[1,N_{\text{top}}-1]}(\hat{t},u) + (1+\varsigma^{-1})\left(\int_{t'=0}^{t'=\hat{t}} \mathbb{E}_{[1,N_{\text{top}}-1]}(t',u)\,dt' + \int_{u'=0}^{u'=u} \mathbb{F}_{[1,N_{\text{top}}-1]}(\hat{t},u')\,du'\right)\right\}.
\end{aligned} \tag{12-21}$$

We now argue as follows using (12-21):

- We choose $\varsigma > 0$ sufficiently small and absorb the terms

$$\varsigma \sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-1.8}(\hat{t})\mathbb{K}_{[1,N_{\text{top}}]}(\hat{t},u), \quad \varsigma \sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-3.8}(\hat{t})\mathbb{K}_{[1,N_{\text{top}}-1]}(\hat{t},u)$$

appearing on the right-hand side by the terms

$$\sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-1.8}(\hat{t})\mathbb{K}_{[1,N_{\text{top}}]}(\hat{t},u), \quad \sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-3.8}(\hat{t})\mathbb{K}_{[1,N_{\text{top}}-1]}(\hat{t},u)$$

on the left-hand side.

- We then apply Proposition 8.12 (using that the exponents $2M_*-1.8$ and $2M_*-3.8$ are positive) and Grönwall's inequality to handle the terms involving the integrals of $\mathbb{E}$ and $\mathbb{F}$.

This leads to the following estimate (where on the left-hand side, we have dropped the below-top-order energies):

$$\begin{aligned}\sup_{\hat{t}\in[0,t]} &\mu_\star^{2M_*-1.8}(\hat{t})\big(\mathbb{E}_{[1,N_{\text{top}}]}(\hat{t},u)+\mathbb{F}_{[1,N_{\text{top}}]}(\hat{t},u)+\mathbb{K}_{[1,N_{\text{top}}]}(\hat{t},u)\big)\\ &\lesssim \mathring{\epsilon}^2+\sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-1.8}(\hat{t})\int_{t'=0}^{t'=t}\mu_\star^{-3/2}(t',u)\left\{\int_{s=0}^{s=t'}\mu_\star^{-1/2}(s)\mathbb{E}^{1/2}_{[1,N_{\text{top}}]}(s)\,ds\right\}^2 dt'. \end{aligned}\tag{12-22}$$

We will now apply a further Grönwall-type argument to (12-22). Define

$$\iota(t) \doteq \exp\left(\int_{s=0}^{s=t}\mu_\star^{-0.9}(s)\,ds\right),$$

and, for a large $\mathfrak{C}>0$ to be chosen later,

$$H(t) \doteq \sup_{\hat{t}\in[0,t]} \iota^{-2\mathfrak{C}}(\hat{t})\mu_\star^{2M_*-1.8}(\hat{t})\mathbb{E}_{[1,N_{\text{top}}]}(\hat{t}).$$

From the definitions of $\mathbb{E}_{[1,N_{\text{top}}]}$, ι, and H, the fact that ι is increasing, and the estimate (12-22), we find that there exists a constant[62] $C_{**}>0$ *independent of* $\mathfrak{C}>0$ so that

$$H(t)\le C_{**}\left(\mathring{\epsilon}^2+\sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-1.8}(\hat{t})\iota^{-2\mathfrak{C}}(\hat{t})\int_{t'=0}^{t'=\hat{t}}\mu_\star^{-3/2}(t')\left\{\int_{s=0}^{s=t'}\mu_\star^{-1/2}(s)\mathbb{E}^{1/2}_{[1,N_{\text{top}}]}(s)\,ds\right\}^2 dt'\right). \tag{12-23}$$

Before we proceed, note that for $n=1,2$ an easy change of variables gives

$$\int_{s=0}^{s=t'}\iota^{n\mathfrak{C}}(s)\mu_\star^{-0.9}(s)\,ds=\int_{y=0}^{y=\int_{\tau=0}^{\tau=t'}\mu_\star^{-0.9}(\tau)\,d\tau} e^{n\mathfrak{C}y}\,dy\le\frac{1}{n\mathfrak{C}}\iota^{n\mathfrak{C}}(t'). \tag{12-24}$$

Fix $t\in[0,T_{(\text{Boot})})$ and $\hat{t}\in[0,t]$. Since $\iota^{-\mathfrak{C}}$ is decreasing and $\mu_\star$ is almost decreasing by Proposition 8.12, we have, using (12-24) and the estimate (12-23) for H, the following bound for the terms under the sup

[62] We call the constant C_{**} so as to make the notation clearer later in the proof.

on right-hand side of (12-23):

$$
\begin{aligned}
&\upmu_\star^{2M_*-1.8}(\hat{t})\iota^{-2\mathfrak{C}}(\hat{t})\int_{t'=0}^{t'=\hat{t}}\upmu_\star^{-3/2}(t')\left\{\int_{s=0}^{s=t'}\upmu_\star^{-1/2}(s)\mathbb{E}^{1/2}_{[1,N_{\mathrm{top}}]}(s)\,ds\right\}^2 dt' \\
&\quad\le \upmu_\star^{2M_*-1.8}(\hat{t})\iota^{-2\mathfrak{C}}(\hat{t})\int_{t'=0}^{t'=\hat{t}}\upmu_\star^{-3/2}(t')\Biggl\{\int_{s=0}^{s=t'}\upmu_\star^{-M_*+1.3}(s) \\
&\qquad\qquad\qquad\times[\iota^{\mathfrak{C}}(s)\upmu_\star^{-0.9}(s)][\iota^{-\mathfrak{C}}(s)\upmu_\star^{M_*-0.9}(s)\mathbb{E}^{1/2}_{[1,N_{\mathrm{top}}]}(s)]\,ds\Biggr\}^2 dt' \\
&\quad\le 2^{2M_*-2.6}\upmu_\star^{2M_*-1.8}(\hat{t})\iota^{-2\mathfrak{C}}(\hat{t})\int_{t'=0}^{t'=\hat{t}}\upmu_\star^{-2M_*+1.1}(t')\left\{\int_{s=0}^{s=t'}\iota^{\mathfrak{C}}(s)\upmu_\star^{-0.9}(s)H^{1/2}(s)\,ds\right\}^2 dt' \\
&\quad\le 2^{2M_*-2.6}\upmu_\star^{2M_*-1.8}(\hat{t})\iota^{-2\mathfrak{C}}(\hat{t})H(t)\int_{t'=0}^{t'=\hat{t}}\upmu_\star^{-2M_*+1.1}(t')\left\{\int_{s=0}^{s=t'}\iota^{\mathfrak{C}}(s)\upmu_\star^{-0.9}(s)\,ds\right\}^2 dt' \\
&\quad\le 2^{2M_*-2.6}\upmu_\star^{2M_*-1.8}(\hat{t})\iota^{-2\mathfrak{C}}(\hat{t})\frac{H(t)}{\mathfrak{C}^2}\int_{t'=0}^{t'=\hat{t}}[\iota^{2\mathfrak{C}}(t')\upmu_\star^{-0.9}(t')]\upmu_\star^{-2M_*+2}(t')\,dt' \\
&\quad\le 2^{4M_*-4.6}\upmu_\star^{0.2}(\hat{t})\iota^{-2\mathfrak{C}}(\hat{t})\frac{H(t)}{\mathfrak{C}^2}\int_{t'=0}^{t'=\hat{t}}\iota^{2\mathfrak{C}}(t')\upmu_\star^{-0.9}(t')\,dt' \\
&\quad\le 2^{4M_*-5.6}\upmu_\star^{0.2}(\hat{t})\frac{H(t)}{\mathfrak{C}^3}\le 2^{4M_*-5.6}\frac{H(t)}{\mathfrak{C}^3}.
\end{aligned}
\tag{12-25}
$$

Plugging (12-25) into (12-23), we obtain

$$
H(t)\le C_{**}\left\{\mathring{\epsilon}^2+2^{4M_*-5.6}\frac{H(t)}{\mathfrak{C}^3}\right\}.
\tag{12-26}
$$

Choosing $\mathfrak{C}>0$ sufficiently large such that $2^{4M_*-5.6}/\mathfrak{C}^3\le\frac{1}{2}$, we immediately infer from (12-26) that $H(t)\le 2C_{**}\mathring{\epsilon}^2$. From this estimate, (12-25) the definition of $\iota(t)$, and the estimate (8-35), we find that the right-hand side of (12-22) is at most $C\mathring{\epsilon}^2$, where C is allowed to depend on $\mathfrak{C}$. From this estimate and the definition of $\mathbb{W}_{[1,N]}(t,u)$, we conclude (12-20) in the case $N=N_{\mathrm{top}}$.

Step 2: $1\le N\le N_{\mathrm{top}}-1$. Let $1\le N\le N_{\mathrm{top}}-1$. Arguing like we did at the beginning of Step 1, except for using (12-3) instead of (12-2), we obtain

$$
\begin{aligned}
&\mathbb{E}_{[1,N]}(t,u)+\mathbb{F}_{[1,N]}(t,u)+\mathbb{K}_{[1,N]}(t,u) \\
&\qquad\lesssim \mathring{\epsilon}^2\max\{1,\upmu_\star^{-2M_*+2N_{\mathrm{top}}-2N+1.8}(t)\} \\
&\qquad\quad+\max\{1,\upmu_\star^{-2M_*+2N_{\mathrm{top}}-2N+1.8}(t)\}\bigl(\sup_{s\in[0,t]}\min\{1,\upmu_\star^{2M_*-2N_{\mathrm{top}}+2N+0.2}(s)\}\mathbb{Q}_{[1,N+1]}(s)\bigr) \\
&\qquad\qquad+\varsigma\mathbb{K}_{[1,N]}+(1+\varsigma^{-1})\left(\int_{t'=0}^{t'=t}\mathbb{E}_{[1,N]}(t',u)\,dt'+\int_{u'=0}^{u'=u}\mathbb{F}_{[1,N]}(t,u')\,du'\right) \\
&\qquad\lesssim \mathring{\epsilon}^2\max\{1,\upmu_\star^{-2M_*+2N_{\mathrm{top}}-2N+1.8}(t)\} \\
&\qquad\quad+\max\{1,\upmu_\star^{-2M_*+2N_{\mathrm{top}}-2N+1.8}(t)\}\bigl(\sup_{s\in[0,t]}\min\{1,\upmu_\star^{2M_*-2N_{\mathrm{top}}+2N+0.2}(s)\}\mathbb{Q}_{[1,N+1]}(s)\bigr),
\end{aligned}
\tag{12-27}
$$

where to obtain the last inequality, we first took ς to be sufficiently small to absorb $\varsigma\mathbb{K}_{[1,N]}$, and then used Grönwall's inequality.

Using (12-27), we easily obtain (12-20) by induction in decreasing N. Notice in particular that the base case $N = N_{\mathrm{top}}$ has already been proven in Step 1. □

13. Proving the L^∞ estimates

We continue to work under the assumptions of Theorem 6.3.

In this section, we derive L^∞ estimates that in particular yield an improvement over the bootstrap assumptions we made in Section 6A. This is the final section in which we derive PDE estimates that are needed for the proof of Theorem 6.3; aside from the Appendix, the rest of the paper (i.e., Section 14) entails deriving consequences of the estimates and assembling the logic of the proof.

We first bound (in Propositions 13.2, 13.3) the L^∞ norm of the fluid variables, specific vorticity, entropy gradient and modified fluid variables and their $\mathcal{P}$ derivatives using the energy estimates we have already obtained and Sobolev embedding (Lemma 13.1). Then, in Propositions 13.3 and 13.4, we control derivatives of these variables that involve one factor of $\breve{X}$ by combining the just-obtained L^∞-estimates for $\mathcal{P}$-derivatives with the (wave or transport) equations.

Lemma 13.1 (Sobolev embedding estimates). *Suppose ϕ is a smooth function with u-support in* $[0, U_0]$. *Then*, *for every* $t \in [0, T_{(\mathrm{Boot})})$, *we have the estimate*

$$\|\phi\|_{L^\infty(\Sigma_t)} \lesssim \sup_{u\in[0,U_0]} \|\mathcal{P}^{\le 2}\phi\|_{L^2(\ell_{0,u})} + \sup_{u\in[0,U_0]} \|L\mathcal{P}^{\le 2}\phi\|_{L^2(\mathcal{F}_u^t)}. \tag{13-1}$$

Proof. First, using standard Sobolev embedding on $\mathbb{T}^2$, using (2-28b)–(2-28c) to express $\not\partial_2, \not\partial_3$ in terms of derivatives with respect to $\{Y, Z\}$, comparing the volume forms using Definition 3.1, and using the estimates of Proposition 8.7, we deduce

$$\|\phi\|_{L^\infty(\ell_{t,u})} \lesssim \sum_{i+j\le 2} \left(\int_{\ell_{t,u}} |\not\partial_2^i \not\partial_3^j \phi|^2 \, dx^2\, dx^3\right)^{\frac{1}{2}} \lesssim \left(\int_{\ell_{t,u}} |\mathcal{P}^{\le 2}\phi|^2 \, d\lambda_{\not g}\right)^{\frac{1}{2}} \doteq \|\mathcal{P}^{\le 2}\phi\|_{L^2(\ell_{t,u})}. \tag{13-2}$$

To complete the proof of (13-1), it remains only for us to control the right-hand side of (13-2) by showing that for any smooth function φ (where the role of φ will be played by $\mathcal{P}^{\le 2}\phi$), we have

$$\|\varphi\|_{L^2(\ell_{t,u})} \le C\|\varphi\|_{L^2(\ell_{0,u})} + C\|L\varphi\|_{L^2(\mathcal{F}_u^t)}. \tag{13-3}$$

To prove (13-3), we start by using the identity $\not\partial_t = L - L^A\not\partial_A$ (see (2-27a)) to deduce that

$$\begin{aligned}\frac{\partial}{\partial t}\int_{\ell_{t,u}} \varphi^2\, dx^2\, dx^3 &= 2\int_{\ell_{t,u}} \varphi\not\partial_t\varphi\, dx^2\, dx^3 = 2\int_{\ell_{t,u}} \varphi L\varphi\, dx^2\, dx^3 - 2\int_{\ell_{t,u}} \varphi L^A\not\partial_A\varphi\, dx^2\, dx^3 \\ &= 2\int_{\ell_{t,u}} \varphi L\varphi\, dx^2\, dx^3 + \int_{\ell_{t,u}} \varphi^2(\not\partial_A L^A)\, dx^2\, dx^3,\end{aligned} \tag{13-4}$$

where in the last step, we integrated the geometric coordinate partial derivatives $\not\partial_A$ by parts (and we recall that capital Latin indices vary over 2, 3). Again using (2-28b)–(2-28c) to express $\not\partial_2, \not\partial_3$ in terms of derivatives with respect to $\{Y, Z\}$, and using the estimates of Propositions 8.6 and 8.7, we find that

$|\not\partial_A L^A| \le C$. From this estimate, (13-4), and Young's inequality, we deduce that

$$\left| \frac{\partial}{\partial t} \int_{\ell_{t,u}} \varphi^2 \, dx^2 \, dx^3 \right| \le C \int_{\ell_{t,u}} |L\varphi|^2 \, dx^2 \, dx^3 + C \int_{\ell_{t,u}} \varphi^2 \, dx^2 \, dx^3. \tag{13-5}$$

Integrating (13-5) with respect to time, using the fundamental theorem of calculus, and then applying Grönwall's inequality, we find that

$$\int_{\ell_{t,u}} \varphi^2 \, dx^2 \, dx^3 \le C \int_{\ell_{0,u}} \varphi^2 \, dx^2 \, dx^3 + C \int_{t'=0} \int_{\ell_{t',u}} |L\varphi|^2 \, dx^2 \, dx^3 \, dt'. \tag{13-6}$$

Again comparing the volume forms using Definition 3.1 and using the estimates of Proposition 8.7, we arrive at the desired bound (13-3). □

Proposition 13.2. *The following L^∞ estimates hold for all $t \in [0, T_{(\mathrm{Boot})})$:*

$$\|\mathcal{P}^{[1, N_{\mathrm{top}} - M_* - 2]} \Psi\|_{L^\infty(\Sigma_t)} \lesssim \mathring{\epsilon}, \tag{13-7}$$

$$\|\mathcal{P}^{\le N_{\mathrm{top}} - M_* - 2}(\Omega, S)\|_{L^\infty(\Sigma_t)} + \|\mathcal{P}^{\le N_{\mathrm{top}} - M_* - 3}(\mathcal{C}, \mathcal{D})\|_{L^\infty(\Sigma_t)} \lesssim \mathring{\epsilon}^{3/2}. \tag{13-8}$$

Proof. These two estimates follow as immediate consequences of the energy estimates (respectively for $(\mathbb{V}, \mathbb{S})$, $(\mathbb{C}, \mathbb{D})$ and $\mathbb{W}$) in Propositions 9.4, 10.6, and 12.7, Lemma 13.1, and the initial data size-assumptions (4-4)–(4-6). □

Proposition 13.3. *The following L^∞ estimates hold for all $t \in [0, T_{(\mathrm{Boot})})$:*

$$\|\mathcal{R}_{(+)}\|_{L^\infty(\Sigma_t)} \lesssim_{\blacklozenge} \mathring{\alpha}, \qquad \|(\mathcal{R}_{(-)}, v^2, v^3, s)\|_{L^\infty(\Sigma_t)} \lesssim \mathring{\epsilon}, \tag{13-9a}$$

$$\|\breve{X}\mathcal{R}_{(+)}\|_{L^\infty(\Sigma_t)} \le 2\mathring{\delta}, \quad \|\breve{X}(\mathcal{R}_{(-)}, v^2, v^3, s)\|_{L^\infty(\Sigma_t)} \lesssim \mathring{\epsilon}, \tag{13-9b}$$

$$\|\mathcal{P}^{[1, N_{\mathrm{top}} - M_* - 4]} \breve{X}\Psi\|_{L^\infty(\Sigma_t)} \lesssim \mathring{\epsilon}. \tag{13-9c}$$

Proof. Step 1: proof of (13-9a). Since $Lt = 1$, we can apply the fundamental theorem of calculus along the integral curves of L to deduce that for any scalar function ϕ, we have

$$\|\phi\|_{L^\infty(\Sigma_t)} \le \|\phi\|_{L^\infty(\Sigma_0)} + \int_{t'=0}^{t} \|L\phi\|_{L^\infty(\Sigma_{t'})} \, dt'. \tag{13-10}$$

By Proposition 13.2, we have $\|L\Psi\|_{L^\infty(\Sigma_t)} \lesssim \mathring{\epsilon}$. From this estimate, the data assumptions (4-3a) and (4-3c), and (13-10) with $\phi \doteq \Psi$, we conclude the desired bounds in (13-9a).

Step 2: an auxiliary estimate for $\mathrm{tr}_{\not g}\,\chi$. We need an auxiliary estimate before proving (13-9b). To start, we note that the same arguments used to prove Proposition 8.6, based on the transport equation[63] (2-41), but now with the estimate (13-7) in place of the L^∞ bootstrap assumptions for $\|\mathcal{P}^{[1, N_{\mathrm{top}} - M_* - 2]} \Psi\|_{L^\infty(\Sigma_t)}$ in (4-4), yield the estimate

$$\|\mathcal{P}^{[1, N_{\mathrm{top}} - M_* - 3]} L^i\|_{L^\infty(\Sigma_t)} \lesssim \mathring{\epsilon}. \tag{13-11}$$

We next use Lemmas 2.23 and 2.32, and the fact that the Cartesian component functions X^1, X^2, X^3 are smooth functions of the L^i and Ψ (see (2-23)) to write the identity (2-38b) in the following form, where f

[63] Note importantly that the right-hand side of (2-41) does not contain an $\breve{X}\Psi$ term!

schematically denotes smooth functions: $\mathrm{tr}_{\not g}\chi = \mathrm{f}(L^i, \Psi)\mathcal{P}L^i + \mathrm{f}(L^i, \Psi)\mathcal{P}\Psi$. From this equation, the estimates of Proposition 13.2, (13-9a), and (13-11), we obtain the desired auxiliary estimate:

$$\|\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-4}\,\mathrm{tr}_{\not g}\chi\|_{L^\infty(\Sigma_t)} \lesssim \mathring{\epsilon}. \tag{13-12}$$

Step 3: controlling $\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-4}L\breve{X}\Psi$. By [52, Proposition 2.16], the wave operator is given by[64]

$$\mu\Box_{g(\vec{\Psi})} f = -L(\mu L f + 2\breve{X} f) + \mu \not\Delta f - \mathrm{tr}_{\not g}\chi \breve{X} f - \mu\,\mathrm{tr}_{\not g}\not k L f - 2\mu\zeta^{\#}\cdot \not d f. \tag{13-13}$$

Consider now the wave equations (5-1a)–(5-1c). We will now bound the inhomogeneous terms in these equations. For $\mathfrak{Q} \in \{\mathfrak{Q}^i_{(v)}, \mathfrak{Q}_{(\pm)}, \mathfrak{Q}_{(s)}\}$, we first apply (10-5) with $\phi^{(1)}$, $\phi^{(2)} = \Psi$, $\mathfrak{d}^{(1,1)}$, $\mathfrak{d}^{(2,1)} \lesssim 1$, $\mathfrak{d}^{(1,2)}$, $\mathfrak{d}^{(2,2)} \lesssim \mathring{\epsilon}$ (which is justified by Proposition 13.2 and the bootstrap assumptions (6-3)–(6-5)), and then use (6-3)–(6-5) and Propositions 8.6 and 13.2 to obtain

$$|\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-4}(\mu\mathfrak{Q})| \lesssim |\mathcal{P}^{[1,N_{\mathrm{top}}-M_*-3]}\Psi| + \mathring{\epsilon}\{|\mathcal{P}^{[1,N_{\mathrm{top}}-M_*-4]}\breve{X}\Psi| + |\mathcal{P}^{[2,N_{\mathrm{top}}-M_*-4]}(\mu, L^i)|\} \lesssim \mathring{\epsilon}. \tag{13-14}$$

For $\mathfrak{L} \in \{\mathfrak{L}^i_{(v)}, \mathfrak{L}_{(\pm)}, \mathfrak{L}_{(s)}\}$ and

$$\mathfrak{M} \in \left\{c^2\exp(2\rho)\mathcal{C}^i,\ c\exp(\rho)\frac{p_{;s}}{\bar{\varrho}}\mathcal{D},\ c^2\exp(2\rho)\mathcal{D}\right\},$$

we use the pointwise bounds (12-12), (12-17) together with (6-3)–(6-8) and Propositions 8.6 and 13.2 to obtain

$$|\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-4}(\mu\mathfrak{L})| + |\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-4}(\mu\mathfrak{M})| \lesssim \mathring{\epsilon}. \tag{13-15}$$

Combining (13-14) and (13-15), we thus obtain

$$|\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-4}(\mu\Box_g\Psi)| \lesssim \mathring{\epsilon}. \tag{13-16}$$

We now use (13-16) together with (13-13) to control $\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-4}L\breve{X}\Psi$. The key point is that every term in $\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-4}$(13-13) except for $\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-4}(-2L\breve{X}\Psi)$ is already known to be bounded in L^∞ by $\mathcal{O}(\mathring{\epsilon})$. More precisely, we express the Ricci coefficients on the right-hand side of (13-13) using (2-38b)–(2-38d) and $\not\Delta$ using Lemmas 2.24 and 2.32. We also use the transport equation (2-40) to express[65] the factor of $L\mu$ on the right-hand side of (13-13) as the right-hand side of (2-40). Then using Propositions 8.6, 8.7, and 13.2, the estimates (13-9a) and (13-11)–(13-12), and the bootstrap assumptions (6-3)–(6-5) (to control all $\breve{X}\Psi$-involving products on the right-hand side of (13-13) except $-2L\breve{X}\Psi$), we obtain $|\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-4}L\breve{X}\Psi| \lesssim \mathring{\epsilon}$. Also using the first commutator estimate in (8-12) with $\phi \doteq \breve{X}\Psi$ and the bootstrap assumption (6-5), we further deduce that

$$\|L\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-4}\breve{X}\Psi\|_{L^\infty(\Sigma_t)} \lesssim |\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-4}L\breve{X}\Psi| + \mathring{\epsilon}^{1/2}|\mathcal{P}^{[1,N_{\mathrm{top}}-M_*-4]}\breve{X}\Psi| \lesssim \mathring{\epsilon}. \tag{13-17}$$

Step 4: proof of (13-9b) and (13-9c). We finally conclude (13-9b) and (13-9c) using (13-10) and (13-17), together with the initial data bounds (4-3b), (4-3c) and (4-4). □

[64] Here, $\not\Delta$ is the Laplace–Beltrami operator on $\ell_{t,u}$, which can be expressed as a second order differential operator in Y and Z with regular coefficients.

[65] This step is needed to avoid having to control $N_{\mathrm{top}} - M_* - 3$ $\mathcal{P}$-derivatives of μ in L^∞, since Proposition 8.6 does not yield L^∞ control of that many derivatives of μ.

Proposition 13.4. *The following L^∞ estimates hold for all $t \in [0, T_{(\mathrm{Boot})})$:*

$$\|\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-4}\breve{X}(\Omega, S)\|_{L^\infty(\mathcal{M}_{t,u})} \lesssim \mathring{\epsilon}^{3/2}.$$

Proof. We apply $\mathcal{P}^{\leq N_{\mathrm{top}}-M_*-4}$ to (8-4)–(8-5) and then bound all terms on the right-hand side in L^∞ using Propositions 8.6, 13.2, and 13.3. □

14. Putting everything together

This is the concluding section. First, in Section 14A, we use the estimates derived in Sections 7–13 to conclude our main a priori estimates, i.e., to prove Theorem 6.3.

With the help of Theorem 6.3, all of the main results stated in Section 4B are quite easy to prove. We will prove Theorems 4.2 and 4.3 in Section 14B, Corollary 4.4 in Section 14C, and finally, Corollary 4.5 in Section 14D.

14A. *Proof of the main a priori estimates.*

Proof of Theorem 6.3. We prove each of the four conclusions asserted by Theorem 6.3.

(1) By Proposition 12.7, for $1 \leq N \leq N_{\mathrm{top}}$, the following wave estimates hold:

$$\mathbb{W}_N(t) \lesssim \mathring{\epsilon}^2 \max\{1, \mu_\star^{-2M_*+2N_{\mathrm{top}}-2N+1.8}(t)\}.$$

Hence, the inequalities in (6-1)–(6-2) hold with $\mathring{\epsilon}$ replaced by $C\mathring{\epsilon}^2$.

(2) By (13-9a)–(13-9b), the inequalities in (6-3) hold with $\mathring{\alpha}^{1/2}$ replaced by $C_\blacklozenge\mathring{\alpha}$ and $3\mathring{\delta}$ replaced by $2\mathring{\delta}$.

(3) By (13-7) and (13-9a)–(13-9c), the inequalities in (6-4)–(6-5) hold with $\mathring{\epsilon}^{1/2}$ replaced by $C\mathring{\epsilon}$.

(4) By (13-8) and Proposition 13.4, the inequalities (6-6)–(6-8) hold with $\mathring{\epsilon}$ replaced by $C\mathring{\epsilon}^{3/2}$. □

14B. *Proof of the main theorems.*

Proof of the regularity theorem (*Theorem 4.2*). By the main a priori estimates (Theorem 6.3) and a standard continuity argument, all the estimates established in the proof of Theorem 6.3 hold on $[0, T) \times \Sigma$. As a consequence, the energy estimates (4-7a), (4-7b) and (4-7c) follow from Propositions 12.7, 9.4, 10.6, and 11.11. As for the L^∞ estimates, (4-8a) holds thanks to (13-7) and (13-9c); (4-8b) and (4-8c) hold thanks to (13-9a) and (13-9b) respectively; and (4-8d) holds thanks to (13-8) and Proposition 13.4.

Moreover, Lemma 2.24, the identity $\not\partial_t = L - L^A \not\partial_A$ (see (2-27a)), and the L^∞ estimates mentioned above, together with those of Propositions 8.6, 8.7, and 8.10, imply that the solution can be smoothly extended[66] to $[0, T] \times \mathbb{R} \times \mathbb{T}^2$ as a function of the geometric coordinates (t, u, x^2, x^3).

It remains for us to show that the solution can be extended as a smooth solution of both the geometric and the Cartesian coordinates as long as $\inf_{t\in[0,T)} \mu_\star(t) > 0$. Now the estimates (4-8a)–(4-8c), Lemma 2.22, and the assumed lower bound on $\mu_\star$ together imply that the fluid variables and their first partial derivatives with respect to the Cartesian coordinates remain bounded. Standard local existence results/continuation criteria

[66]Note that these estimates imply that the $\not\partial_t$ derivatives of many geometric coordinate partial derivatives of the solution are uniformly bounded on $[0, T] \times \mathbb{R} \times \mathbb{T}^2$, which leads to their extendibility to $[0, T] \times \mathbb{R} \times \mathbb{T}^2$.

then imply that the solution can be smoothly extended in the Cartesian coordinates to a Cartesian slab $[0, T+\epsilon] \times \Sigma$ for some $\epsilon > 0$. Finally, within this Cartesian slab, one can solve the eikonal equation (2-13) such that the map $(t, u, x^2, x^3) \to (t, x^1, x^2, x^3)$ is a diffeomorphism from $[0, T+\epsilon] \times \mathbb{R} \times \mathbb{T}^2$ onto $[0, T+\epsilon] \times \Sigma$; the diffeomorphism property of this map follows easily from the identity $\partial\!\!\!/_u x^1 = \upmu c^2/X^1$ (see (2-28a)) and the fact that $\upmu c^2/X^1 < 0$ in $[0, T+\epsilon] \times \mathbb{R} \times \mathbb{T}^2$ whenever ϵ is small enough, thanks to $\upmu > 0$, (2-25b), and the estimates of Proposition 8.7 for $X^i_{(\text{small})}$ and $c - 1$. This implies that the solution can also be smoothly extended in the geometric coordinates (t, u, x^2, x^3). □

Proof of the shock formation theorem (*Theorem 4.3*). Step 1: vanishing of $\upmu_\star$. First, we will show that

$$\upmu_\star(t) = 1 + \mathcal{O}_\blacklozenge(\mathring{\upalpha}) + \mathcal{O}(\mathring{\upepsilon}) - \mathring{\updelta}_* t. \tag{14-1}$$

To prove (14-1), we start by using (2-40), (2-42), and the L^∞ estimates established in Propositions 8.6 and 8.7 and Theorem 4.2 to deduce that

$$L\upmu = -\tfrac{1}{2} c^{-1}(c^{-1} c_{;\uprho} + 1)\check{X}\mathcal{R}_{(+)} + \mathcal{O}(\mathring{\upepsilon}) \tag{14-2}$$

and

$$L\left\{\tfrac{1}{2} c^{-1}(c^{-1} c_{;\uprho} + 1)\right\} = \mathcal{O}(\mathring{\upepsilon}), \quad L\left\{\tfrac{1}{2} c^{-1}(c^{-1} c_{;\uprho} + 1)\check{X}\mathcal{R}_{(+)}\right\} = \mathcal{O}(\mathring{\upepsilon}). \tag{14-3}$$

Moreover, from (2-13), (2-14), and our data assumptions (4-3a) and (4-3c), we have the following initial condition estimate for $\upmu$:

$$\upmu \restriction_{\Sigma_0} = 1 + \mathcal{O}_\blacklozenge(\mathring{\upalpha}) + \mathcal{O}(\mathring{\upepsilon}). \tag{14-4}$$

From (14-2)–(14-4), (4-2), and the fundamental theorem of calculus along the integral curves of L (and recalling that $Lt = 1$), we conclude (14-1).

Step 2: proof of (1), (2), and (3). Define

$$T_{(\text{sing})} \doteq \sup\{T \in [0, 2\mathring{\updelta}_*^{-1}] : \text{a smooth solutions exists with } \upmu > 0 \text{ on } [0, T) \times \Sigma\}. \tag{14-5}$$

From Theorem 4.2, it follows that either $T_{(\text{sing})} = 2\mathring{\updelta}_*^{-1}$ or $\liminf_{t \to T_{(\text{sing})}^-} \upmu_\star(t) = 0$.

Using (14-2), we infer that $\upmu_\star(t)$ first vanishes at a time equal to $\{1 + \mathcal{O}_\blacklozenge(\mathring{\upalpha}) + \mathcal{O}(\mathring{\upepsilon})\}\mathring{\updelta}_*^{-1}$. From this fact, the definition of $T_{(\text{sing})}$, and the above discussion, it follows that this time of first vanishing of $\upmu_\star(t)$ is equal to $T_{(\text{sing})}$, which implies (4-9). Using Theorem 4.2 again, we have therefore proved parts (1), (2) and (3) of Theorem 4.3.

Step 3: proof of (4). In the next step, we will show that the vanishing of $\upmu_\star$ along $\Sigma_{T_{(\text{sing})}}$ coincides with the blowup of $|\partial_1 \mathcal{R}_{(+)}|$ at one or more points in $\Sigma_{T_{(\text{sing})}}$; that will show that $T_{(\text{sing})}$ is indeed the time of first singularity formation and in particular yields the conclusion (4) stated in Theorem 4.3.

Step 4: proof of (5). We now prove that $\mathscr{S}_{\text{blowup}} = \mathscr{S}_{\text{vanish}}$. This in particular also implies the blowup-claim in conclusion (4) of Theorem 4.3. We first prove $\mathscr{S}_{\text{blowup}} \subseteq \mathscr{S}_{\text{vanish}}$. If $(u, x^2, x^3) \notin \mathscr{S}_{\text{vanish}}$, then $\upmu$ has a lower bound away from 0 near $(T_{(\text{sing})}, u, x^2, x^3)$ and thus the estimates in Theorem 4.2 (and Lemma 2.22) imply that the fluid variables are C^1 functions of the geometric coordinates and the Cartesian coordinates near the point with geometric coordinates $(T_{(\text{sing})}, u, x^2, x^3)$, i.e., $(u, x^2, x^3) \notin \mathscr{S}_{\text{blowup}}$.

To show $\mathscr{S}_{\text{blowup}} \supseteq \mathscr{S}_{\text{vanish}}$, suppose $(u, x^2, x^3) \in \mathscr{S}_{\text{vanish}}$. Let $\beta(t)$ denote the t-parametrized integral curve of L emanating from $(T_{(\text{sing})}, u, x^2, x^3)$. Note in particular that $\upmu \circ \beta(T_{(\text{sing})}) \doteq \upmu(T_{(\text{sing})}, u, x^2, x^3) = 0$, and recall that $Lt = 1$. We next use (14-2)–(14-4), (4-2), (4-9), and the fundamental theorem of calculus along the integral curve $\beta(t)$ to deduce that, for $0 \leq t \leq T_{(\text{sing})}$, we have

$$\tfrac{1}{2}|c^{-1}(c^{-1}c_{;\rho}+1)| \circ \beta(0) \times |\breve{X}\mathcal{R}_{(+)}| \circ \beta(t) \geq \tfrac{1}{4}3\mathring{\delta}_*$$

(for otherwise, $\upmu \circ \beta(T_{(\text{sing})}) = 0$ would not be possible). Also using (2-26b), Propositions 8.6, 8.7, and the L^∞ estimates of Theorem 6.3, we find that the following estimate holds for $0 \leq t \leq T_{(\text{sing})}$:

$$\tfrac{1}{2}|c^{-1}(c^{-1}c_{;\rho}+1)| \circ \beta(0) \times |\upmu\partial_1\mathcal{R}_{(+)}| \circ \beta(t) \geq \tfrac{1}{2}\mathring{\delta}_*.$$

In particular, also considering Remark 4.1, we deduce that

$$\limsup_{t \uparrow T^-_{(\text{sing})}} |\partial_1\mathcal{R}_{(+)})| \circ \beta(t) \geq \frac{\mathring{\delta}_*}{2|c^{-1}(c^{-1}c_{;\rho}+1)| \circ \beta(0)} \limsup_{t \uparrow T^-_{(\text{sing})}} \frac{1}{\upmu \circ \beta(t)} = \infty.$$

Hence $(u, x^2, x^3) \in \mathscr{S}_{\text{blowup}}$, which finishes the proof that $\mathscr{S}_{\text{blowup}} = \mathscr{S}_{\text{vanish}}$.

Finally, we prove that $\mathscr{S}_{\text{vanish}} = \mathbb{R} \times \mathbb{T}^2 \setminus \mathscr{S}_{\text{regular}}$. The direction $\subseteq$ holds since $\mathscr{S}_{\text{vanish}} = \mathscr{S}_{\text{blowup}}$ and obviously $\mathscr{S}_{\text{blowup}} \subseteq \mathbb{R} \times \mathbb{T}^2 \setminus \mathscr{S}_{\text{regular}}$. We now show the direction $\supseteq$. Suppose that $(u, x^2, x^3) \notin \mathscr{S}_{\text{blowup}}$, i.e., $\upmu(T_{(\text{sing})}, u, x^2, x^3) > 0$. Then the estimates with respect to the geometric vectorfields established in Theorem 4.2 and Lemma 2.22 imply that in a neighborhood of $(T_{(\text{sing})}, u, x^2, x^3)$ intersected with $\{t \leq T_{(\text{sing})}\}$, the fluid variables remain C^1 functions of the geometric coordinates and Cartesian coordinates. We have therefore proved part (5) of Theorem 4.3, which completes its proof. □

14C. *Nontriviality of Ω and S (Proof of Corollary 4.4).*

Proof of Corollary 4.4. Using equations (14-2)–(14-4), we deduce (recalling that $\mathring{\epsilon}^{1/2} \leq \mathring{\alpha}$ by assumption) that along any t-parametrized integral curve $\beta(t)$ of L emanating from Σ_0 (i.e., $\beta^0(0) = 0$, where β^α denotes the Cartesian components of β), we have $\upmu \circ \beta(t) = 1 - \frac{1}{2}t[c^{-1}(c^{-1}c_{;\rho}+1)\breve{X}\mathcal{R}_{(+)}] \circ \beta(0) + \mathcal{O}_\blacklozenge(\mathring{\alpha})$. From this bound, (4-9) (which implies that $0 \leq t \leq T_{(\text{sing})} = \{1 + \mathcal{O}_\blacklozenge(\mathring{\alpha}) + \mathcal{O}(\mathring{\epsilon})\}\mathring{\delta}_*^{-1}$), (4-2), and the assumption (4-10), we see that if $|u \circ \beta(0) - \mathring{\sigma} + \mathring{\delta}_*^{-1}| \geq 3\mathring{\alpha}\mathring{\delta}_*^{-1}$ (where $u \circ \beta(0)$ is the value of the u-coordinate at $\beta(0)$), then $\upmu \circ \beta(t) \geq \frac{3}{8}$ for $0 \leq t \leq T_{(\text{sing})}$ (assuming that $\mathring{\alpha}$ and $\mathring{\epsilon}$ are sufficiently small).

Now fix any $(u_*, x_*^2, x_*^3) \in \mathscr{S}_{\text{vanish}}$ (that is, $\upmu(T_{(\text{sing})}, u_*, x_*^2, x_*^3) = 0$). We will show that under the assumptions of the corollary, there is a constant $C > 1$ such that

$$C^{-1}\mathring{\epsilon}^3 \leq |S(T_{(\text{sing})}, u_*, x_*^2, x_*^3)| \leq C\mathring{\epsilon}^3, \quad C^{-1}\mathring{\epsilon}^2 \leq |\Omega(T_{(\text{sing})}, u_*, x_*^2, x_*^3)| \leq C\mathring{\epsilon}^2. \tag{14-6}$$

Clearly, the bounds (14-6) imply the desired conclusion of the corollary.

To initiate the proof of (14-6), we let $\beta_{(\text{sing})}(t)$ denote the t-parametrized integral curve of L passing through $(T_{(\text{sing})}, u_*, x_*^2, x_*^3)$. Then since (2-21) implies that the coordinate function u is constant along $\beta_{(\text{sing})}$ (and thus $u \circ \beta_{(\text{sing})}(0) = u_*$), the results derived two paragraphs above guarantee that $|u_* - \mathring{\sigma} + \mathring{\delta}_*^{-1}| \leq 3\mathring{\alpha}\mathring{\delta}_*^{-1}$. In particular, in view of the initial condition (2-13) for u along Σ_0, we see that $|\beta^1_{(\text{sing})}(0) - \mathring{\delta}_*^{-1}| \leq 3\mathring{\alpha}\mathring{\delta}_*^{-1}$, where $\beta^1_{(\text{sing})}(0) \doteq x^1 \circ \beta_{(\text{sing})}(0)$ is the x^1-coordinate of the point $\beta_{(\text{sing})}(0) \in \Sigma_0$.

Then, since Proposition 8.6 yields that $(d/dt)\beta^1 = L\beta^1 = L^1 = 1 + L^1_{(\mathrm{small})} = 1 + \mathcal{O}_{\blacklozenge}(\mathring{\alpha})$, we can integrate in time and use (4-9) to deduce that

$$\beta^1(T_{(\mathrm{sing})}) = \beta^1(0) + T_{(\mathrm{sing})} + \mathcal{O}_{\blacklozenge}(\mathring{\alpha})T_{(\mathrm{sing})} = -\mathring{\delta}_*^{-1} + T_{(\mathrm{sing})} + \mathcal{O}_{\blacklozenge}(\mathring{\alpha})T_{(\mathrm{sing})} = \mathcal{O}_{\blacklozenge}(\mathring{\alpha})\mathring{\delta}_*^{-1}.$$

That is, the x^1-coordinate of the singular point $(T_{(\mathrm{sing})}, u_*, x^2_*, x^3_*)$ is of size $\mathcal{O}_{\blacklozenge}(\mathring{\alpha})\mathring{\delta}_*^{-1}$.

Let now $\gamma_{(\mathrm{sing})}$ be the integral curve of B passing through the singular point $(T_{(\mathrm{sing})}, u_*, x^2_*, x^3_*)$ as above. Since (2-23) and (4-8b) imply that $B = \partial_t + \mathcal{O}_{\blacklozenge}(\mathring{\alpha})\partial$, we can integrate with respect to time along $\gamma_{(\mathrm{sing})}$ and use (4-9) and the bound on the x^1-coordinate of the singular point $(T_{(\mathrm{sing})}, u_*, x^2_*, x^3_*)$ proved above to deduce that $\gamma_{(\mathrm{sing})}$ intersects Σ_0 at a point q with x^1-coordinate q^1 of size $q^1 = \mathcal{O}_{\blacklozenge}(\mathring{\alpha})\mathring{\delta}_*^{-1}$. In view of the initial condition (2-13) for u along Σ_0, we see that the u-coordinate of q^1, which we denote by $u|_q$, satisfies $|u|_q - \mathring{\sigma}| = \mathcal{O}_{\blacklozenge}(\mathring{\alpha})\mathring{\delta}_*^{-1}$. From this bound and the assumption (4-11), we see that

$$\tfrac{1}{2}\mathring{\epsilon}^2 \le |\Omega|_q| \le \mathring{\epsilon}^2, \quad \tfrac{1}{2}\mathring{\epsilon}^3 \le |S|_q| \le \mathring{\epsilon}^3. \tag{14-7}$$

To complete the proof, we need to use (14-7) to prove (14-6). To this end, we find it convenient to parametrize $\gamma_{(\mathrm{sing})}$ by the eikonal function. Since (2-23) and (2-21) guarantee that $\mu B u = 1$, this is equivalent to studying integral curves of μB. That is, we slightly abuse notation by denoting the reparametrized integral curve by the same symbol $\gamma_{(\mathrm{sing})}$; i.e., $\gamma_{(\mathrm{sing})}$ solves the integral curve ODE $(d/du)\gamma_{(\mathrm{sing})}(u) = \mu B \circ \gamma_{(\mathrm{sing})}(u)$. To proceed, we multiply the transport equations (5-2a) and (5-2c) by μ and use (2-23), (2-21), Lemma 2.22, Propositions 8.6, 8.7, and the L^∞ estimates of Theorem 6.3 to deduce that along $\gamma_{(\mathrm{sing})}$, (5-2a) and (5-2c) imply the following evolution equations, expressed in schematic form:

$$\frac{d}{du}\Omega \circ \gamma_{(\mathrm{sing})}(u) = \mathcal{O}(1)\Omega \circ \gamma_{(\mathrm{sing})}(u) + \mathcal{O}(1)S \circ \gamma_{(\mathrm{sing})}(u), \tag{14-8}$$

$$\frac{d}{du}S \circ \gamma_{(\mathrm{sing})}(u) = \mathcal{O}(1)S \circ \gamma_{(\mathrm{sing})}(u). \tag{14-9}$$

From the evolution equations (14-8)–(14-9), the initial conditions (14-7), and the fact that $0 \le u \le U_0$ in the support of the solution (see Section 7), we conclude that if $\mathring{\epsilon}$ is sufficiently small, then there is a $C > 1$ such that (14-6) holds. □

14D. ***Hölder estimates (proof of Corollary 4.5).*** Throughout this section, we work under the assumptions of Corollary 4.5.

Lemma 14.1 (a simple calculus lemma). *Let $J \subseteq \mathbb{R}$ be an interval. Suppose $f : J \to \mathbb{R}$ is a C^3 function such that*:

(1) *f is increasing, i.e., $f' \ge 0$.*

(2) *There exists $\mathring{b} > 0$ such that $f^{(3)}(y) \ge \mathring{b}$ for every $y \in J$, where $f^{(3)}$ denotes the third derivative of f.*

Then for any $y_1,\ y_2 \in J$, the following estimate holds:

$$|f(y_1) - f(y_2)| \ge \frac{\mathring{b}}{48}|y_1 - y_2|^3.$$

Proof. First, note that the assumption on $f^{(3)}$ implies that f'' is strictly increasing. In particular, f'' can at most change sign once.

Without loss of generality, assume $y_1 \neq y_2$. We consider three cases: the first two are such that $f''(y_1)$ and $f''(y_2)$ are of the same sign, while the third is such that they have opposite sign.

Case 1: $y_1 < y_2$ and $f''(y_1) < f''(y_2) \leq 0$. By Taylor's theorem,

$$f(y_1) = f(y_2) - f'(y_2)(y_2 - y_1) + \frac{1}{2} f''(y_2)(y_2 - y_1)^2 \\ - \frac{1}{2}(y_2 - y_1)^3 \int_0^1 (1-\tau)^2 f^{(3)}(y_2 + \tau(y_1 - y_2))\, d\tau \leq f(y_2) - \frac{\mathring{b}}{6}(y_2 - y_1)^3,$$

where we have used $f'(y_2) \geq 0$, $f''(y_2) \leq 0$ and $f^{(3)}(y) \geq \mathring{b}$.

Therefore,

$$|f(y_1) - f(y_2)| = f(y_2) - f(y_1) \geq \frac{\mathring{b}}{6}(y_2 - y_1)^3.$$

Case 2: $y_2 < y_1$ and $f''(y_1) > f''(y_2) \geq 0$. This can be treated in the same way as Case 1 so that we have

$$|f(y_1) - f(y_2)| = f(y_1) - f(y_2) \geq \frac{\mathring{b}}{6}(y_1 - y_2)^3.$$

Case 3: $y_1 < y_2$, $f''(y_1) < 0 < f''(y_2)$. Since f'' is strictly increasing, there exists a unique $z \in (y_1, y_2)$ such that $f''(z) = 0$. Therefore, using Case 1 (for y_1 and z) and Case 2 (for y_2 and z), we have

$$|f(y_1) - f(y_2)| = f(y_2) - f(z) + f(z) - f(y_1) \geq \frac{\mathring{b}}{6}(|y_2 - z|^3 + |y_1 - z|^3) \geq \frac{\mathring{b}}{2^3 \cdot 6}(y_2 - y_1)^3,$$

where in the very last inequality we have used $y_2 - y_1 \leq 2 \max\{|y_1 - z|,\ |y_2 - z|\}$.

Combining all three cases, we conclude the desired inequality. □

Lemma 14.2 (quantitative negativity of $\breve{\partial}_u^3 x^1$). *Under the assumptions of Corollary 4.5, the following holds at all points such that* $(t, u) \in [3T_{(\mathrm{sing})}/4, T_{(\mathrm{sing})}) \times [\mathring{\sigma}/2, 3\mathring{\sigma}/2]$:

$$\breve{\partial}_u^3 x^1 \leq -\mathring{\beta}.$$

Proof. In this proof, we will silently use the fact that the Cartesian component functions X^1, X^2, X^3 are smooth functions of the L^i and Ψ (see (2-23)) and the fact that c is a smooth function of Ψ.

By (2-29), to prove the lemma, we need to estimate $\breve{\partial}_u^3 x^1 = \breve{\partial}_u^2(\mu c^2/X^1)$. To proceed, we use (2-28a) (in particular, the fact that $\breve{\partial}_u - \breve{X}$ is $\ell_{t,u}$-tangent) and the L^∞ estimates of Propositions 8.6, 8.7, and 8.10 and Theorem 6.3 to deduce that

$$\breve{\partial}_u^3 x^1 = \breve{X}\breve{X}\left(\frac{\mu c^2}{X^1}\right) + \mathcal{O}(\mathring{\varepsilon}). \tag{14-10}$$

We will now estimate the term $\breve{X}\breve{X}(\mu c^2/X^1)$ on the right-hand side of (14-10). We start by noting that the L^∞ estimates of Propositions 8.6, 8.7, and 8.10 and Theorem 6.3 together imply that $|LL\breve{X}\breve{X}(\mu c^2/X^1)| = \mathcal{O}(\mathring{\varepsilon})$. Therefore, letting $\gamma(t)$ be any integral curve of L parametrized by Cartesian time t (with $\gamma(0) \in \Sigma_0$) and recalling that $Lt = 1$, we integrate this estimate twice in time to deduce that for $t \in [0, T_{(\text{sing})})$, we have

$$\begin{aligned}\breve{X}\breve{X}\left(\frac{\mu c^2}{X^1}\right) \circ \gamma(t) &= \left[\breve{X}\breve{X}\left(\frac{\mu c^2}{X^1}\right)\right] \circ \gamma(0) + t\left[L\breve{X}\breve{X}\left(\frac{\mu c^2}{X^1}\right)\right] \circ \gamma(0) + \mathcal{O}(\mathring{\epsilon}) \\ &= \left[\breve{X}\breve{X}\left(\frac{\mu c^2}{X^1}\right)\right] \circ \gamma(0) + t\left[\breve{X}\breve{X}L\left(\frac{\mu c^2}{X^1}\right)\right] \circ \gamma(0) + \mathcal{O}(\mathring{\epsilon}), \end{aligned} \tag{14-11}$$

where to deduce the last equality, we used in particular (8-27).

Next, using the transport equation (2-40), (2-42), the fact that $X \restriction_{\Sigma_0} = -c\partial_1$ (by (2-7), (2-13), (2-26b), and the normalization condition $g(X, X) = 1$), and the L^∞ estimates mentioned above, we deduce that

$$\begin{aligned}\left[\breve{X}\breve{X}L\left(\frac{\mu c^2}{X^1}\right)\right] \circ \gamma(0) &= \left[\breve{X}\breve{X}\left\{(L\mu)\frac{c^2}{X^1}\right\}\right] \circ \gamma(0) + \mathcal{O}(\mathring{\epsilon}) \\ &= \frac{1}{2}[\breve{X}\breve{X}\{(c^{-1}c_{;\rho} + 1)(\breve{X}\mathcal{R}_{(+)})\}] \circ \gamma(0) + \mathcal{O}(\mathring{\epsilon}). \end{aligned} \tag{14-12}$$

Next, using that $X \restriction_{\Sigma_0} = -c\partial_1$, and using that $\mu \restriction_{\Sigma_0} = 1/c$ (this follows from the initial condition in (2-13) and the fact that (2-21) implies that $Xu = 1/\mu$), we deduce

$$\breve{X}\breve{X}\left(\frac{\mu c^2}{X^1}\right) \restriction_{\Sigma_0} = -\breve{X}\breve{X}(1) = 0. \tag{14-13}$$

Combining (14-11)–(14-13), we find that

$$\breve{X}\breve{X}\left(\frac{\mu c^2}{X^1}\right) \circ \gamma(t) = \frac{t}{2}[\breve{X}\breve{X}\{(c^{-1}c_{;\rho} + 1)(\breve{X}\mathcal{R}_{(+)})\}] \circ \gamma(0) + \mathcal{O}(\mathring{\epsilon}). \tag{14-14}$$

From (14-14) and our assumption (4-12), we deduce that at any point whose corresponding u-coordinate[67] satisfies $u \in [\mathring{\sigma}/2, 3\mathring{\sigma}/2]$, we have

$$\breve{X}\breve{X}\left(\frac{\mu c^2}{X^1}\right) \circ \gamma(t) \le -2t\mathring{\delta}_*\mathring{\beta} + \mathcal{O}(\mathring{\epsilon}). \tag{14-15}$$

In particular, for points whose corresponding u- and t-coordinates satisfy, respectively, $u \in [\mathring{\sigma}/2, 3\mathring{\sigma}/2]$ and $t \in [3T_{(\text{sing})}/4, T_{(\text{sing})})$, we have, in view of (4-9), the estimate

$$\breve{X}\breve{X}\left(\frac{\mu c^2}{X^1}\right) \circ \gamma(t) \le -\frac{3\mathring{\beta}}{2} + \mathcal{O}_{\blacklozenge}(\mathring{\alpha})\mathring{\delta}_*\mathring{\beta} + \mathcal{O}(\mathring{\epsilon}). \tag{14-16}$$

Combining (14-10) and (14-16), we conclude the lemma. □

Lemma 14.3 (the main Hölder estimate for the eikonal function). *Under the assumptions of Corollary 4.5, the following holds for $t \in [3T_{(\text{sing})}/4, T_{(\text{sing})})$:*

$$\sup_{\substack{p_1, p_2 \in \Sigma_t,\, p_1 \neq p_2 \\ u(p_i) \in [\mathring{\sigma}/2, 3\mathring{\sigma}/2]}} \frac{|u(p_1) - u(p_2)|}{\text{dist}_{\text{Euc}}(p_1, p_2)^{1/3}} \le 5\mathring{\beta}^{-1/3}.$$

Above, $u(p_i)$ denotes the value of the eikonal function at p_i, $x(p_i)$ denotes the Cartesian spatial coordinates of p_i, and $\text{dist}_{\text{Euc}}(p_1, p_2)$ denotes the Euclidean distance in Σ_t between p_1 and p_2.

[67] Recall that $u \restriction_{\Sigma_0} = \mathring{\sigma} - x^1$ and the u-value is constant along the integral curves of L by virtue of the first equation in (2-21).

Proof. Step 1: estimating $\min_{u(p_i)=u_i} \mathrm{dist}_{\mathrm{Euc}}(p_1, p_2)$ by carefully choosing two points. Consider two distinct values u_1, u_2 which obey $u_i \in [\mathring{\sigma}/2, 3\mathring{\sigma}/2]$. By compactness of the constant-u hypersurfaces in Σ_t, there exist points $\boldsymbol{p}_1, \boldsymbol{p}_2 \in \Sigma_t$ with $u(\boldsymbol{p}_i) = u_i$ and $\mathrm{dist}_{\mathrm{Euc}}(\boldsymbol{p}_1, \boldsymbol{p}_2) = \min_{u(p_i)=u_i} \mathrm{dist}_{\mathrm{Euc}}(p_1, p_2)$. In particular, $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are connected by a Euclidean straight line $L_{\boldsymbol{p}_1,\boldsymbol{p}_2}$ which is Euclidean-perpendicular to $\{u = u_i\}$ at the point $\boldsymbol{p}_i$ for $i = 1, 2$.

Now by Lemma 2.22 and (2-21), the Euclidean gradient of u satisfies

$$\upmu \partial_i u = c^{-2} X^i, \quad i = 1, 2, 3. \tag{14-17}$$

Recalling (by Proposition 8.7 and conclusions (2) and (3) of Theorem 6.3) that $c^{-2}X^1 = -1 + \mathcal{O}_{\blacklozenge}(\mathring{\alpha})$, $c^{-2}X^2$, $c^{-2}X^3 = \mathcal{O}_{\blacklozenge}(\mathring{\alpha})$, we deduce from (14-17) that $L_{\boldsymbol{p}_1,\boldsymbol{p}_2}$ makes a Euclidean angle of $\mathcal{O}_{\blacklozenge}(\mathring{\alpha})$ with respect to ∂_1. Therefore, using (14-17) again (which implies that constant-u hypersurfaces in Σ_t make an angle $\mathcal{O}(\mathring{\alpha})$ with constant-x^1 planes), we infer that there exist[68] $\mathfrak{p}_1, \mathfrak{p}_2$ such that:

(1) $u(\mathfrak{p}_i) = u_i$.

(2) ∂_1 is tangent to the Euclidean line $\mathfrak{L}$ connecting $\mathfrak{p}_1$ and $\mathfrak{p}_2$.

(3) $\min_{u(p_i)=u_i} \mathrm{dist}_{\mathrm{Euc}}(p_1, p_2) = \mathrm{dist}_{\mathrm{Euc}}(\boldsymbol{p}_1, \boldsymbol{p}_2) \geq \frac{1}{2} \mathrm{dist}_{\mathrm{Euc}}(\mathfrak{p}_1, \mathfrak{p}_2) = \frac{1}{2}|x^1(\mathfrak{p}_1) - x^1(\mathfrak{p}_2)|$.

We fix such a choice of $(\mathfrak{p}_1, \mathfrak{p}_2)$ for any given (u_1, u_2) (with $u_1 \neq u_2$).

Step 2: estimating $|x^1(\mathfrak{p}_1) - x^1(\mathfrak{p}_2)|$. By (2-29), Proposition 8.7, and conclusions (2) and (3) of Theorem 6.3, we have

$$\breve{X}_u x^1 = \upmu(-1 + \mathcal{O}_{\blacklozenge}(\mathring{\alpha})).$$

Hence, for every fixed (x^2, x^3), x^1 is a strictly decreasing function in u. Moreover, by Lemma 14.2, $\breve{X}_u^3 x^1 \leq -\mathring{\upbeta}$. Hence, we are exactly in the setting to apply Lemma 14.1 (for the one-variable function $f(u) = -x^1(u)$, where (x^2, x^3) is fixed, and $\mathring{b} = \mathring{\upbeta}$) to obtain

$$|x^1(\mathfrak{p}_1) - x^1(\mathfrak{p}_2)| \geq \frac{\mathring{\upbeta}}{48}|u_1 - u_2|^3. \tag{14-18}$$

In view of our choice of $\mathfrak{p}_1$ and $\mathfrak{p}_2$ in Step 1, we conclude from (14-18) that

$$\begin{aligned}
\sup_{\substack{p_1,p_2\in\Sigma_t, p_1\neq p_2\\ u(p_i)\in[\mathring{\sigma}/2,3\mathring{\sigma}/2]}} \frac{|u(p_1)-u(p_2)|}{\mathrm{dist}_{\mathrm{Euc}}(p_1,p_2)^{1/3}} &\leq \sup_{\substack{u_1\neq u_2\\ u_i\in[\mathring{\sigma}/2,3\mathring{\sigma}/2]}} \frac{|u_1-u_2|}{\inf_{p_1,p_2\in\Sigma_t, u(p_i)=u_i} \mathrm{dist}_{\mathrm{Euc}}(p_1,p_2)^{1/3}} \\
&\leq 2^{1/3} \sup_{\substack{p_1,p_2\in\Sigma_t, p_1\neq p_2\\ u(p_i)\in[\mathring{\sigma}/2,3\mathring{\sigma}/2]}} \frac{|u_1-u_2|}{|x^1(\mathfrak{p}_1)-x^1(\mathfrak{p}_2)|^{1/3}} \leq 96^{1/3}\mathring{\upbeta}^{-1/3} \leq 5\mathring{\upbeta}^{-1/3}. \quad \square
\end{aligned}$$

We are now ready to conclude the proof of Corollary 4.5.

Proof of Corollary 4.5. Our starting point is the observation that the estimates in Theorem 4.2 guarantee that, for at each fixed t with $0 \leq t \leq T_{(\mathrm{sing})}$, the fluid variables and higher-order variables ρ, v^i, Ω^i, S^i, $\mathcal{C}^i$, and $\mathcal{D}$ are all uniformly Lipschitz *when viewed as functions of the* (u, x^2, x^3)*-coordinates*. Therefore, the key to proving Corollary 4.5 is to understand the regularity of the map $(x^1, x^2, x^3) \mapsto (u, x^2, x^3)$.

[68] We can, for instance, take $\mathfrak{p}_1 = \boldsymbol{p}_1$ and let $\mathfrak{p}_2$ be the unique point in both the level set $\{u = u_2\}$ and the line passing through $\mathfrak{p}_1$ with tangent vector everywhere equal to ∂_1.

To this end, we first note that by the assumption (1) in Corollary 4.5, the equations (14-2)–(14-4), (4-9), and the arguments given in the proof of Corollary 4.4, it follows that away from $u \in [3\mathring{\sigma}/4, 5\mathring{\sigma}/4]$, we have $\mu > \frac{1}{2}$. From this lower bound, Lemma 2.22, and the estimates of Proposition 8.7, we see that when $u \notin [3\mathring{\sigma}/4, 5\mathring{\sigma}/4]$, the map $(x^1, x^2, x^3) \mapsto (u, x^2, x^3)$ remains uniformly Lipschitz (in fact, we could prove that it is even more regular). Combined with the aforementioned fact that ρ, v^i, Ω^i, S^i, $\mathcal{C}^i$ and $\mathcal{D}$ are uniformly Lipschitz in the (u, x^2, x^3)-coordinates, we see that at each fixed t, with $0 \leq t \leq T_{(\mathrm{sing})}$, ρ, v^i, Ω^i, S^i, $\mathcal{C}^i$, and $\mathcal{D}$ are also uniformly Lipschitz in the (x^1, x^2, x^3)-coordinates away from $u \in [3\mathring{\sigma}/4, 5\mathring{\sigma}/4]$. Moreover, (14-1) guarantees that in the region $\{0 \leq t \leq 3T_{(\mathrm{sing})}/4\}$, we have $\mu > \frac{1}{8}$. Thus, for the same reasons given above, the map $(x^1, x^2, x^3) \mapsto (u, x^2, x^3)$ is uniformly Lipschitz in $\{0 \leq t \leq 3T_{(\mathrm{sing})}/4\}$, and thus ρ, v^i, Ω^i, S^i, $\mathcal{C}^i$, and $\mathcal{D}$ also remain uniformly Lipschitz in the (x^1, x^2, x^3)-coordinates in this region.

It remains for us to consider the difficult region in which $u \in [3\mathring{\sigma}/4, 5\mathring{\sigma}/4] \subseteq [\mathring{\sigma}/2, 3\mathring{\sigma}/2]$ and $t \in [3T_{(\mathrm{sing})}/4, T_{(\mathrm{sing})})$. Using Lemma 14.3, we see that the map $(x^1, x^2, x^3) \mapsto (u, x^2, x^3)$ is uniformly $C^{1/3}$ in this difficult region. Hence, ρ, v^i, Ω^i, S^i, $\mathcal{C}^i$, and $\mathcal{D}$ all have uniformly bounded Cartesian spatial $C^{1/3}$ norms in this region as well. □

Appendix: Proof of the wave estimates

In this appendix, we sketch the proof of the wave equation estimates, that is, of Proposition 12.1. As we already discussed in Section 12A, although the wave equation estimates that we need are almost identical to the ones derived in [36], there are two differences:

(1) The wave equations in Proposition 12.1 feature the inhomogeneous terms $\vec{\mathfrak{G}}$, and we need to track the influence of these inhomogeneous terms on the estimates. Recall that the precise inhomogeneous terms are located on the right-hand sides of (5-1a)–(5-1c), but for purposes of proving Proposition 12.1, we do not need to know their precise structure.

(2) Recall that our commutation vectorfields $\{L, Y, Z\}$ are constructed out of the acoustic eikonal function u, and hence the commuted wave equations feature error terms that depend on the acoustic geometry. In three dimensions, some additional arguments are needed (compared to the two-dimensional case treated in [36]) to control the top-order derivatives of some of these error terms.

The issue (2) is tied to the fact that the null second fundamental form of null hypersurfaces in 1+3 dimensions has now three independent components, which stands in contrast to the case of 1+2 dimensions, where it has only a single component (i.e., it is trace-free in 1+2 dimensions). This issue is by now very well-understood, and it can be resolved by using an elliptic estimate. For completeness, we will nonetheless sketch the main points needed for the argument in this appendix.

We now further discuss the issue (2). In 1+2 dimensions, $\mathrm{tr}_{\not{g}}\,\chi$ satisfies a transport equation known as the Raychaudhuri equation[69] (see [52, (6.2.5)]):

$$\mu L\,\mathrm{tr}_{\not{g}}\,\chi = (L\mu)\,\mathrm{tr}_{\not{g}}\,\chi - \mu(\mathrm{tr}_{\not{g}}\,\chi)^2 - \mu\,\mathrm{Ric}_{LL}, \tag{A-1}$$

[69]Note that this is a purely differential geometric identity that is independent of the compressible Euler equations.

where Ric is the Ricci curvature of the acoustical metric g and $\mathrm{Ric}_{LL} \doteq \mathrm{Ric}_{\alpha\beta}\, L^\alpha L^\beta$. In contrast, in 1+3 dimensions, the right-hand side of (A-1) features some additional terms. Specifically, in 1+3 dimensions, the Raychaudhuri equation takes the following form (see [48, (11.23)]):

$$\mu L\, \mathrm{tr}_{\not g}\, \chi = (L\mu)\, \mathrm{tr}_{\not g}\, \chi - \mu|\chi|^2 - \mu\, \mathrm{Ric}_{LL} = (L\mu)\, \mathrm{tr}_{\not g}\, \chi - \mu(\mathrm{tr}_{\not g}\, \chi)^2 - \mu|\hat{\chi}|^2 - \mu\, \mathrm{Ric}_{LL}, \tag{A-2}$$

where $\hat{\chi}$ is the traceless part of χ, i.e., it can be defined by imposing the identity $\chi = \hat{\chi} + \frac{1}{2}(\mathrm{tr}_{\not g}\, \chi)\not g$. In other words, (A-2) has an additional $-\mu|\hat{\chi}|^2$ term compared to (A-1), and this additional term cannot be bounded using the only the transport equation (A-2), (since the left-hand side of (A-2) features a transport operator acting only on the component $\mathrm{tr}_{\not g}\, \chi$, as opposed to the full second fundamental form χ). The saving grace, however, as already noticed in [15] (see also [17; 33]), is that one can use geometric identities (specifically, the famous Codazzi equation) and elliptic estimates to control $\not\nabla\hat{\chi}$ in terms of $\not d\, \mathrm{tr}_{\not g}\, \chi$ plus simpler error terms. A top-order version of this kind of argument allows one to control the difficult top-order derivatives of the term $-\mu|\hat{\chi}|^2$ on the right-hand side of (A-2); see Section A5 for the details. We remark that for the solutions under study, the $-\mu|\hat{\chi}|^2$ term is quadratically small and, as it turns out, it does not have much effect on the dynamics.

A1. ***Running assumptions in the appendix and the dependence of constants and parameters.*** Throughout the entire appendix, we work in the setting of Proposition 12.1. In particular, we make the same assumptions as we did in Theorem 6.3 (which provides the main a priori estimates), as well as the smallness assumption (12-1) for the inhomogeneous terms $\vec{\mathfrak{G}}$.

Our analysis involves various constants and parameters that play distinct roles in the proof. We have already introduced these quantities earlier in the article. For the reader's convenience, we again provide a brief description of these quantities in order to help the reader understand their role in our subsequent arguments in the appendix.

- The background density constant $\bar{\varrho} > 0$ was fixed at the beginning of the paper. The parameters $\mathring{\sigma}$, $\mathring{\delta}_*$, $\mathring{\delta}$, $\mathring{\alpha}$ and $\mathring{\epsilon}$ measure the size of the x^1-support and various norms of the initial data; see Section 4A.
- As in the rest of the paper, the positive integer N_{top} denotes the maximum number of times that we commute the equations for the purpose of obtaining L^2-type energy estimates.
- M_{abs} denotes an *absolute constant*, that is, a constant that can be chosen to be independent of N_{top}, the equation of state, $\bar{\varrho}$, $\mathring{\sigma}$, $\mathring{\delta}$, and $\mathring{\delta}_*^{-1}$, as long as $\mathring{\alpha}$ and $\mathring{\epsilon}$ are sufficiently small. The constants M_{abs} arise as numerical coefficients that multiply the borderline energy error integrals; see in particular the right-hand side of (A-37). The universality of the M_{abs} is crucial since, as the next two points clarify, they drive the blowup-rate of the top-order energies, which in turn controls the size of largeness of N_{top} needed to close the proof.
- As in the rest of the paper, the positive integer M_* controls the blowup-rate of the high-order energies. The following point is crucial: *for the proof to close we need to choose M_* to be sufficiently large in a manner that depends only on the absolute constants M_{abs}. In particular, M_* does not depend on N_{top}.*
- Once M_* has been chosen to be sufficiently large (as described in the previous point), for the proof to close *we need to choose N_{top} to be sufficiently large in a manner that depends only on the integer M_* fixed in the previous step.*

• Once N_{top} has been chosen to be sufficiently large (as described in the previous point), to close the proof, we must choose $\mathring{\varepsilon}$ to be sufficiently small in a manner that is allowed to depend on all other parameters and constants. We must also choose $\mathring{\alpha}$ to be sufficiently small in a manner that depends only on the equation of state and $\bar{\varrho}$. We always assume that $\mathring{\varepsilon}^{1/2} \leq \mathring{\alpha}$.

• In contrast to M_{abs}, the constants C' are less delicate and are allowed to depend on the equation of state, $\bar{\varrho}$, $\mathring{\sigma}$, $\mathring{\delta}$, and $\mathring{\delta}_*^{-1}$. We use the notation C' to emphasize that these constants multiply difficult, borderline energy estimate error terms, but we could have just as well denoted these constants by C (where C has the properties described in the next point), and the proof would go through.

• Unless otherwise stated, "general" constants C are allowed to depend on N_{top}, M_{abs}, the equation of state, $\bar{\varrho}$, $\mathring{\sigma}$, $\mathring{\delta}$, and $\mathring{\delta}_*^{-1}$. When we write $A \lesssim B$, it means that there exists a $C > 0$ with the above dependence properties such that $A \leq CB$. Moreover, $A \approx B$ means that $A \lesssim B$ and $B \lesssim A$.

A2. ***An outline of the rest of the appendix.*** In Sections A3–A8, we will derive the estimates we need to prove Proposition 12.1. The conclusion of the proof of Proposition 12.1 is located in Section A9.

Proposition 12.1 is an analog of the similar result [36, Proposition 14.1]. In fact, in our proof of the proposition, we will exactly follow the strategy from [36]. For this reason, we will only focus on terms which did not already appear in [36]. We begin by identifying the most difficult wave equation error terms in Section A3. As in [36; 52], these hardest terms are commutator terms involving the top-order derivatives of $\operatorname{tr}_{\not{g}} \chi$, which we control using the following steps:

• In Section A4, we write down the transport equations satisfied by the important *modified quantities*. The modified quantities are special combinations of solution variables involving $\operatorname{tr}_{\not{g}} \chi$. With the help of the Raychaudhuri equation (A-2), the modified quantities will allow us to avoid the loss of a derivative at the top order and/or allow us to avoid fatal borderline error integrals.

• In Section A5, we use elliptic estimates on $\ell_{t,u}$ to control the top-order derivatives of $\hat{\chi}$ in terms of the modified quantities.

• In Section A6, we define partial energies, which are similar to the energies we defined in Section 3B, but they control all wave variables *except for* the "difficult" one $\mathcal{R}_{(+)}$ (which is such that $|\partial_1 \mathcal{R}_{(+)}|$ blows up as the shock forms). As in [36], the partial energies play an important role in allowing us to close the proof using a universal number of derivatives, that is, a number N_{top} that is independent of the equation of state and all parameters in the problem; the role of these partial energies will be made clear in Section A9.

• In Section A7, we use the transport equations in Section A4 and the estimates in Section A5 to obtain the bounds for the top-order derivatives of $\operatorname{tr}_{\not{g}} \chi$.

At this point in the proof, we will have obtained all of the main new estimates we need to prove Proposition 12.1. In Section A8, we use our estimates for the top-order derivatives of $\operatorname{tr}_{\not{g}} \chi$ to derive preliminary energy integral inequalities for the wave equation solutions. These are the same integral inequalities that were derived in [36, Proposition 14.3], except they include the new terms generated by the inhomogeneous terms $\vec{\mathfrak{G}}$ featured in the statement of Proposition 12.1. Finally, in Section A9, we use these integral inequalities and a slightly modified version of the Grönwall-type argument used in the proof of [36, Proposition 14.1], carefully tracking the different kinds of constants, thereby obtaining a priori estimates for the energies and concluding the proof of Proposition 12.1.

We close this section with three remarks to help the reader understand how we use cite/use results that were proved in [36].

Remark A.1 (implicit reliance on results we have already proved). The estimates in this appendix rely, in addition to the bootstrap assumptions, on many of the estimates that we independently derived in Section 8, such as the results of Propositions 8.6, 8.7, 8.10, 8.11, 8.12, and 8.14. Many of the results that we cite from [36] rely on these propositions, and we will not always explicitly indicate the dependence of the results of [36] on these propositions.

Remark A.2 (ε vs. $\mathring{\epsilon}^{1/2}$). The bootstrap smallness parameter ε from [36] should be identified with the quantity $\mathring{\epsilon}^{1/2}$ in our bootstrap assumptions (6-4)–(6-8). For this reason, various error terms from [36] reappear in the present paper, but with the factors of ε replaced by $\mathring{\epsilon}^{1/2}$. This minor point has no substantial effect on our analysis, and we will often avoid explicitly pointing out that the error terms from [36] need to be modified as such.

Remark A.3 (vorticity terms have been absorbed into $\vec{\mathfrak{G}}$). Many error terms in the estimates of [36] involve vorticity terms that are generated by the vorticity terms on the right-hand side of the wave equations. However, in this appendix, we have absorbed these error terms into our definition of the inhomogeneous terms $\vec{\mathfrak{G}}$ in Proposition 12.1. For this reason, it is to be understood that many of the estimates cited from [36] have to be modified so that these vorticity terms are absent and are instead replaced with analogous error terms that depend on $\vec{\mathfrak{G}}$ (where throughout the appendix, we carefully explain how the term $\vec{\mathfrak{G}}$ appears in various estimates).

A3. ***The top-order commutator terms that require the modified quantities.*** To begin, we recall that $\{Y, Z\}$ denotes the commutation vectorfields tangent to $\ell_{t,u}$, and that we use the notation $\not{\mathcal{P}}$ to denote a generic element of this set. In the following proposition, we identify the most difficult error terms in the top-order commuted wave equations.

Proposition A.4 (identifying the most difficult commutator terms). *Let $\mathfrak{G}$ denote the inhomogeneous terms in the wave equations from Proposition 12.1. Then solutions to the wave equations of Proposition 12.1 satisfy the following top-order wave equations* (*which identify the most difficult commutator terms*):

$$\mu\,\square_g\,(\not{\mathcal{P}}^{N_{\mathrm{top}}-1}L\Psi) = (\not{d}^{\#}\Psi)(\mu\not{d}\not{\mathcal{P}}^{N_{\mathrm{top}}-1}\operatorname{tr}_{\not{g}}\chi) + \not{\mathcal{P}}^{N_{\mathrm{top}}-1}L\mathfrak{G} + \textit{Harmless}, \tag{A-3}$$

$$\mu\,\square_g\,(\not{\mathcal{P}}^{N_{\mathrm{top}}-1}Y\Psi) = (\breve{X}\Psi)(\not{\mathcal{P}}^{N_{\mathrm{top}}-1}Y\operatorname{tr}_{\not{g}}\chi) + c^{-2}X^2(\not{d}^{\#}\Psi)(\mu\not{d}\not{\mathcal{P}}^{N_{\mathrm{top}}-1}\operatorname{tr}_{\not{g}}\chi) + \not{\mathcal{P}}^{N_{\mathrm{top}}-1}Y\mathfrak{G} + \textit{Harmless}, \tag{A-4}$$

$$\mu\,\square_g\,(\not{\mathcal{P}}^{N_{\mathrm{top}}-1}Z\Psi) = (\breve{X}\Psi)(\not{\mathcal{P}}^{N_{\mathrm{top}}-1}Z\operatorname{tr}_{\not{g}}\chi) + c^{-2}X^3(\not{d}^{\#}\Psi)(\mu\not{d}\not{\mathcal{P}}^{N_{\mathrm{top}}-1}\operatorname{tr}_{\not{g}}\chi) + \not{\mathcal{P}}^{N_{\mathrm{top}}-1}Z\mathfrak{G} + \textit{Harmless}. \tag{A-5}$$

Above, the terms "Harmless" are precisely the $\mathrm{Harmless}^{\le N_{\mathrm{top}}}_{(\mathrm{Wave})}$ *terms defined in* [36, Definition 13.1], *except here we do not need to allow for the presence of vorticity-involving terms in the definition of* $\mathrm{Harmless}^{\le N_{\mathrm{top}}}_{(\mathrm{Wave})}$ *because we have absorbed these terms into our definition of the wave equation inhomogeneous term* $\mathfrak{G}$.

Moreover, for any other top-order operator $\mathcal{P}^{N_{\text{top}}}$ (*i.e., a top-order operator featuring at least two copies of* L *or featuring only a single* L *but in an order different from* (A-3)), *there are no difficult commutator terms in the sense that the following equation holds*:

$$\mu\square_g(\mathcal{P}^{N_{\text{top}}}\Psi) = \mathcal{P}^{N_{\text{top}}}\mathfrak{G} + Harmless. \tag{A-6}$$

Proof. This is exactly the same as [36, Proposition 13.2] with the obvious modifications: we have $\{L, Y, Z\}$ (as opposed to just $\{L, Y\}$) as commutation vectorfields, and we have accounted for the presence of the inhomogeneous terms $\mathfrak{G}$. We stress that even in three spatial dimensions, the top-order derivatives of χ that appear on the right-hand sides of (A-3)–(A-5) only involve its trace-part $\operatorname{tr}_{\not{g}}\chi$, as opposed to involving the full tensor χ. Roughly speaking, this follows from three basic facts: all of these top-order terms are generated when all $N_{\text{top}}+1$ derivatives (including the two coming from $\square_g$) on the left-hand sides fall on the components $\mathcal{P}^i$ (where $\mathcal{P} \in \{L, Y, Z\}$); all $\mathcal{P}^i$ can be expressed as functions Ψ and L^1, L^2, L^3; and Lemma 2.19 and (13-13) with $f \doteq u$ together imply that $\mu\square_g u = -\operatorname{tr}_{\not{g}}\chi$. Hence, considering also (2-14), we have, schematically, that $\mu\square_g \partial u = -\partial\operatorname{tr}_{\not{g}}\chi + \cdots$, where "$\cdots$" denotes terms that involve lower-order derivatives (i.e., up to second-order derivatives) of the eikonal function u and/or derivatives of Ψ. Thus, (2-14), (2-16), (2-17) imply that the scalar functions $\mathcal{P}^i$ satisfy, schematically,[70] $\square_g\mathcal{P}^i = \partial\operatorname{tr}_{\not{g}}\chi + \cdots$. □

Remark A.5. Notice that in [36, Proposition 13.2], there is an additional difficult commutator term coming from (in the language of the present paper) the commutation with $\breve{X}$. Since in this paper, we use only the subset of energy estimates in [36] that *avoid commutations with* $\breve{X}$, an added benefit of our approach here is that we do not need to handle these additional terms.[71]

A4. ***The modified quantities and the additional terms in the transport equations.*** In order to control the top-order commutator terms from Proposition A.4, the idea from [15] is to introduce modified quantities, which are corrected versions of $\operatorname{tr}_{\not{g}}\chi$. The "fully modified quantities" solve transport equations with source terms that enjoy improved regularity, thus allowing us to avoid a loss of regularity at the top order. The "partially modified quantities" lead to cancellations in the energy identities that allow us to avoid error integrals whose singularity strength would have been too severe for us to control.

Definition A.6 (modified versions of the derivatives of $\operatorname{tr}_{\not{g}}\chi$). We define, for every[72] fixed string of order-N commutators $\mathcal{P}^N \in \mathscr{P}^{(N)}$, the *fully modified quantity* ${}^{(\mathcal{P}^N)}\mathscr{X}$ as

$$ {}^{(\mathcal{P}^N)}\mathscr{X} \doteq \mu\mathcal{P}^N\operatorname{tr}_{\not{g}}\chi + \mathcal{P}^N\mathfrak{X}, \tag{A-7a}$$

$$\mathfrak{X} \doteq -\vec{G}_{LL}\diamond\breve{X}\vec{\Psi} - \tfrac{1}{2}\mu\operatorname{tr}_{\not{g}}\vec{\not{G}}\diamond L\vec{\Psi} - \tfrac{1}{2}\mu\vec{G}_{LL}\diamond L\vec{\Psi} + \mu\vec{\not{G}}_L^{\#}\diamond\not{d}\vec{\Psi}. \tag{A-7b}$$

[70] Of course, careful geometric decompositions are needed to obtain the precise form of the terms on the right-hand sides of (A-3)–(A-5); here we are simply emphasizing that the dependence of the top-order terms is through the derivatives of $\operatorname{tr}_{\not{g}}\chi$.

[71] Of course, even if these terms had been present in our work here, we could have handled them in the same way they were handled in [36].

[72] In practice, we need these quantities only to handle the difficult terms from Proposition A.4, which involve purely $\ell_{t,u}$-tangential derivatives of $\operatorname{tr}_{\not{g}}\chi$. Put differently, in practice, we only need to use the quantities ${}^{(\mathcal{P}^N)}\mathscr{X}$.

We define, for every[73] fixed string of order $\mathcal{P}^N \in \mathscr{P}^{(N)}$, the *partially modified quantity* ${}^{(\mathcal{P}^N)}\widetilde{\mathscr{X}}$ as

$$ {}^{(\mathcal{P}^N)}\widetilde{\mathscr{X}} \doteq \mathcal{P}^N \operatorname{tr}_{\not g} \chi + {}^{(\mathcal{P}^N)}\widetilde{\mathfrak{X}}, \tag{A-8a} $$

$$ {}^{(\mathcal{P}^N)}\widetilde{\mathfrak{X}} \doteq -\tfrac{1}{2} \operatorname{tr}_{\not g} \vec{\not G} \diamond L\mathcal{P}^N \vec{\Psi} + \vec{\not G}_L^{\#} \diamond \not d\mathcal{P}^N \vec{\Psi}. \tag{A-8b} $$

Proposition A.7 (transport equations satisfied by the modified quantities). *The fully modified quantities solve the following modified version of equation* [36, (6.9)], *where* $\vec{\mathfrak{G}}$ *denotes the array of inhomogeneous terms in the wave equations from Proposition 12.1*:

$$ L^{(\mathcal{P}^{N_{\mathrm{top}}})}\mathscr{X} - \left(2\frac{L\mu}{\mu} - 2\operatorname{tr}_{\not g}\chi\right){}^{(\mathcal{P}^{N_{\mathrm{top}}})}\mathscr{X} $$
$$ = \textit{non-vorticity-involving terms in } [36, (6.9)] - \mathcal{P}^{N_{\mathrm{top}}}(\mu|\hat{\chi}|^2) + \tfrac{1}{2}\mathcal{P}^{N_{\mathrm{top}}}(\vec{G}_{LL} \diamond \vec{\mathfrak{G}}). \tag{A-9} $$

Moreover, the partially modified quantities solve the following modified version of equation [36, (6.10)]:

$$ L^{(\mathcal{P}^{N_{\mathrm{top}}-1})}\widetilde{\mathscr{X}} = \textit{terms in } [36, (6.10)] - \mathcal{P}^{N_{\mathrm{top}}-1}(|\hat{\chi}|^2). \tag{A-10} $$

Remark A.8. We clarify that the vorticity-involving terms in [36, (6.9)] are absent from the right-hand side of (A-9) because we have absorbed these terms into our definition of the wave equation inhomogeneous term $\vec{\mathfrak{G}}$.

Proof of Proposition A.7. The key point is that the derivations of both [36, (6.9), (6.10)] used the Raychaudhuri transport equation satisfied by $\operatorname{tr}_{\not g}\chi$, and thus we need to take into account the additional $-\mu|\hat{\chi}|^2$ term in (A-2) as compared to (A-1).

The derivation of [36, (6.9)] consists of two steps. First, in [36, Lemma 6.1], one expresses $\mu\,\mathrm{Ric}_{LL}$ in terms of a sum of two terms: one term is a total L derivative, and the other term is of lower order; see [36, (6.1)]. Step 1 in particular uses the wave equations $\mu\square_{g(\vec{\Psi})}\Psi_\iota = \cdots$. In the second step, one combines the result of [36, Lemma 6.1] with the 1+2-dimensional Raychaudhuri equation (A-1) and then commutes the resulting equation to obtain [36, (6.9)]. In our setting, each step requires a small modification.

- In the first step, instead of $\mu\square_{g(\vec{\Psi})}\Psi_\iota = \cdots$, we have $\mu\square_{g(\vec{\Psi})}\Psi_\iota = \mathfrak{G}_\iota$. Thus, we get an additional term $\frac{1}{2}\mathcal{P}^{N_{\mathrm{top}}}(\vec{G}_{LL} \diamond \vec{\mathfrak{G}})$ on the right-hand side of (A-9).
- In the second step, we need to use the 1+3-dimensional Raychaudhuri equation (A-2) instead of (A-1) and get the extra term $-\mathcal{P}^{N_{\mathrm{top}}}(\mu|\hat{\chi}|^2)$ on the right-hand side of (A-9).

We thus obtain (A-9).

The derivation of [36, (6.10)] is simpler because its proof relies only on the 1+2-dimensional Raychaudhuri equation (A-1) (in particular, it does not rely on the wave equations $\mu\square_{g(\vec{\Psi})}\Psi_\iota = \cdots$). Thus, to obtain (A-10), we simply replace the application of (A-1) from [36, (6.10)] by an application of (A-2). The additional term in (A-10) is a result of the extra $-\mu|\hat{\chi}|^2$ term in (A-2) compared to (A-1). □

[73]As in footnote 72, in practice, we only need to use the quantities ${}^{(\mathcal{P}^N)}\widetilde{\mathscr{X}}$.

A5. ***Control of the geometry of $\ell_{t,u}$ and the elliptic estimates for $\hat{\chi}$.*** The following elliptic estimate is standard; see [15, Lemma 8.8].

Lemma A.9 (elliptic estimate for symmetric, trace-free tensorfields). *Let $(\mathcal{M}_2, \gamma\!\!\!/)$ be a closed, orientable Riemannian manifold, and let μ be a nonnegative function on $\mathcal{M}_2$. Then the following estimate holds for all trace-free symmetric covariant 2-tensorfields ξ belonging to $W^{1,2}(\mathcal{M}_2, \gamma\!\!\!/)$*:

$$\int_{\mathcal{M}_2} \mu^2\left(\tfrac{1}{2}|\nabla\!\!\!\!/\,\xi|_{\gamma\!\!\!/}^2 + 2\mathfrak{K}_{\gamma\!\!\!/}|\xi|_{\gamma\!\!\!/}^2\right) d\mathrm{A}_{\gamma\!\!\!/} \leq 3\int_{\mathcal{M}_2} \mu^2 |\mathrm{d}\!\!\!/\mathrm{iv}_{\gamma\!\!\!/}\,\xi|_{\gamma\!\!\!/}^2 \, d\mathrm{A}_{\gamma\!\!\!/} + 3\int_{\mathcal{M}_2} |\nabla\!\!\!\!/\,\mu|_{\gamma\!\!\!/}^2 |\xi|_{\gamma\!\!\!/}^2 \, d\mathrm{A}_{\gamma\!\!\!/}, \tag{A-11}$$

where $\nabla\!\!\!\!/$, $\mathrm{d}\!\!\!/\mathrm{iv}_{\gamma\!\!\!/}$, $\mathfrak{K}_{\gamma\!\!\!/}$ and $d\mathrm{A}_{\gamma\!\!\!/}$ are respectively the Levi-Civita connection, divergence operator, Gaussian curvature and induced area measure associated with $\gamma\!\!\!/$.

In order to use Lemma A.9, we need an L^∞ estimate for the Gaussian curvature of the tori $(\ell_{t,u}, g\!\!\!/)$. We provide this basic estimate in the following proposition.

Proposition A.10. *The Gaussian curvature $\mathfrak{K}_{g\!\!\!/}$ of $(\ell_{t,u}, g\!\!\!/)$ satisfies the following estimate for every $(t,u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$*:

$$\|\mathfrak{K}_{g\!\!\!/}\|_{L^\infty(\mathcal{M}_{t,u})} \lesssim \mathring{\epsilon}^{1/2}.$$

Proof. It is a standard fact that at fixed (t,u), $\mathfrak{K}_{g\!\!\!/}$ can be expressed in terms of the components of $g\!\!\!/$, $g\!\!\!/^{-1}$ with respect to the coordinate system (x^2, x^3) on $\ell_{t,u}$ and their first and second partial derivatives with respect to the geometric coordinate vectorfields $\partial\!\!\!/_2$, $\partial\!\!\!/_3$. Schematically, we have

$$\mathfrak{K}_{g\!\!\!/} = g\!\!\!/^{-1} \cdot g\!\!\!/^{-1} \cdot \partial\!\!\!/^2 g\!\!\!/ + g\!\!\!/^{-1} \cdot g\!\!\!/^{-1} \cdot g\!\!\!/^{-1} \cdot \partial\!\!\!/ g\!\!\!/ \cdot \partial\!\!\!/ g\!\!\!/,$$

where $\partial\!\!\!/ \in \{\partial\!\!\!/_2, \partial\!\!\!/_3\}$.

Recalling the expression for the induced metric $g\!\!\!/$ in Lemma 2.32 and the relations between the vectorfields in Lemma 2.24, we see that the desired estimate for $\mathfrak{K}_{g\!\!\!/}$ follows from Proposition 8.7. □

We now apply the elliptic estimate in Lemma A.9 to control the top-order derivatives of $\hat{\chi}$ in terms of the top-order pure $\ell_{t,u}$-tangential derivatives of $\mathrm{tr}_{g\!\!\!/}\,\chi$.

Proposition A.11. *The following estimate holds for[74] the N_{top}-th $\ell_{t,u}$-tangential derivatives of $\hat{\chi}$ for every $(t,u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$*:

$$\|\mu(\not{\mathcal{L}}_{\mathcal{P}\!\!\!/})^{N_{\mathrm{top}}}\hat{\chi}\|_{L^2(\Sigma_t^u)} \lesssim \|\mu \mathcal{P}\!\!\!/^{N_{\mathrm{top}}}\,\mathrm{tr}_{g\!\!\!/}\,\chi\|_{L^2(\Sigma_t^u)} + \mathring{\epsilon}^{1/2}\mu_\star^{-M_*+0.9}(t).$$

Proof. Step 0: preliminaries. Throughout the proof, we will silently use the following observations, valid for $\mathcal{P} \in \{L, Y, Z\}$ and $\mathcal{P}\!\!\!/ \in \{Y, Z\}$, where $\mathrm{f}(\cdot)$ denotes a generic smooth function of its arguments that is allowed to vary from line to line.

- The component functions X^1, X^2, X^3 are smooth functions of the L^i and Ψ; see (2-23). The same holds for the component functions $\mathcal{P}^0, \mathcal{P}^1, \mathcal{P}^2, \mathcal{P}^3$; this is obvious for $\mathcal{P} = L$, while see Lemma 2.23 for $\mathcal{P} = Y, Z$. Similarly, the geometric coordinate component functions $g\!\!\!/_{AB}$ and $(g\!\!\!/^{-1})^{AB}$ are smooth functions of the L^i and Ψ; see Lemma 2.32.

[74] Recall that $\not{\mathcal{L}}_{\mathcal{P}\!\!\!/}$ denotes Lie differentiation with respect to elements $\mathcal{P}\!\!\!/ \in \{Y, Z\}$, followed by projection onto $\ell_{t,u}$.

- For $\not\partial \in \{\not\partial_2, \not\partial_3\}$, we have the following schematic identity: $\not\partial = \mathrm{f}(L^i, \Psi)Y + \mathrm{f}(L^i, \Psi)Z$; see Lemma 2.24.
- For $\ell_{t,u}$-tangent one-forms ξ, we have $|\xi| \approx \sum_{A=2,3} |\xi_A| \approx |\xi_Y| + |\xi_Z| \doteq |\xi(Y)| + |\xi(Z)|$; this follows from the discussion in the previous two points, the bootstrap assumptions (6-3)–(6-5), and the L^∞ estimates for $L^i_{(\mathrm{small})}$ from Proposition 8.6. In particular, for scalar functions ϕ, we have $|\not\nabla \phi| \approx \sum_{A=2,3} |\not\partial_A \phi| \approx |Y\phi| + |Z\phi|$. Analogous estimates hold for $\ell_{t,u}$-tangent tensorfields of any order.
- For type-$\binom{0}{n}$ tensorfields, we have the following covariant identity, expressed schematically: $[\not\nabla, \not\mathcal{L}_{\mathcal{P}}]\xi = (\not\nabla \not\mathcal{L}_{\mathcal{P}} \not g) \cdot \xi$. It is straightforward to check that $\not\mathcal{L}_{\mathcal{P}} \not g$ is in fact equal to the $\ell_{t,u}$-projection of the deformation tensor of $\mathcal{P}$ (the deformation tensor itself is equal to $\mathcal{L}_{\mathcal{P}} g$, where g is the acoustical metric).
- Relative to the geometric coordinates (t, u, x^2, x^3), we have $\not\mathcal{L}_{\mathcal{P}} \not g = \mathrm{f}(L^i, \Psi)(\mathcal{P}L^i, \mathcal{P}\Psi)$ (where the $\mathcal{P}$'s on the left- and right-hand sides can be different).
- For $\ell_{t,u}$-tangent tensorfields ξ, we have the following schematic identity, valid relative to the geometric coordinates, where $\not\partial \in \{\not\partial_2, \not\partial_3\}$: $\not\nabla \xi - \not\partial \xi = \mathrm{f}(L^i, \Psi)\xi \cdot (\not\mathcal{P} L^i, \not\mathcal{P}\Psi)$; this follows from expressing $\not\nabla$ in terms of geometric coordinate partial derivatives and the Christoffel symbols of $\not g$ and then expressing $\not\partial = \mathrm{f}(L^i, \Psi)\not\mathcal{P}$ on the right-hand side.
- For $\ell_{t,u}$-tangent tensorfields ξ, we have the following schematic identity, valid relative to the geometric coordinates, where $\not\partial \in \{\not\partial_2, \not\partial_3\}$: $\not\mathcal{L}_{\not\mathcal{P}} \xi = \not\mathcal{P}^A \not\partial_A \xi + \mathrm{f}(L^i, \Psi)\xi \cdot (\not\mathcal{P} L^i, \not\mathcal{P}\Psi)$ (where the $\not\mathcal{P}$'s on the left- and right-hand sides can be different); this formula is straightforward to verify relative to the geometric coordinates.
- For $\ell_{t,u}$-tangent tensorfields ξ, we have the following schematic identity, valid relative to the geometric coordinates, where $\not\partial \in \{\not\partial_2, \not\partial_3\}$: $\not\nabla_{\not\mathcal{P}} \xi - \not\mathcal{L}_{\not\mathcal{P}} \xi = \mathrm{f}(L^i, \vec{\Psi})\xi \cdot (\not\mathcal{P} L^i, \not\mathcal{P}\Psi)$ (where the $\not\mathcal{P}$'s on the left- and right-hand sides can be different); this formula is straightforward to verify relative to the geometric coordinates.
- If f is a scalar function, then $\mathcal{L}_{\mathcal{P}} \not d f = \not d \mathcal{P} f$, where $\not d$ denotes $\ell_{t,u}$-gradient of f; this formula is straightforward to verify relative to the geometric coordinates.

Step 1: Codazzi equation.[75] We compute $(\not\mathcal{L}_{\mathcal{P}})^{N_{\mathrm{top}}-1} \not\nabla^A \chi_{BA}$ by differentiating (2-38a) with the operator $(\not\mathcal{L}_{\mathcal{P}})^{N_{\mathrm{top}}-1} \mathrm{d}\not{\mathrm{i}}\mathrm{v}_{\not g}$ and treating all capital Latin indices as tensorial indices, while treating all lowercase Latin indices as corresponding to scalar functions. We clarify that the tensor on the left-hand sides of (2-38a) is symmetric, while the first, third, and fourth products on the right-hand side of (2-38a) are not. Hence, for clarity, we emphasize that when we write "differentiating (2-38a) with $(\not\mathcal{L}_{\mathcal{P}})^{N_{\mathrm{top}}-1} \not\nabla^A \chi_{BA}$," it is to be understood that the corresponding first term on the right-hand side is an $\ell_{t,u}$-tangent one-form with index B whose top-order part (in the sense of the number of derivatives that fall on L^a) is $(\not\mathcal{L}_{\mathcal{P}})^{N_{\mathrm{top}}-1}(g_{ab}((\not g^{-1})^{AC} \not\nabla_C \not\partial_B L^a) \otimes \not\partial_A x^b)) = (\not\mathcal{L}_{\mathcal{P}})^{N_{\mathrm{top}}-1}(g_{ab}((\not g^{-1})^{AC} \not\nabla_B \not\partial_C L^a) \otimes \not\partial_A x^b))$, where to obtain the last equality, we used the commutation identity $\not\nabla_C \not\partial_B L^a = \not\nabla_B \not\partial_C L^a$, which is a consequence of the torsion-free property of $\not\nabla$ and the fact that we are viewing the Cartesian components L^a as scalar functions. Notice that unless all the N_{top} derivatives fall on the factor $\not d L^a$ in the first product on the

[75] We use the phrase "Codazzi equation" because the equations we use in this analysis are closely related to the classical Codazzi equation, which links $\mathrm{d}\not{\mathrm{i}}\mathrm{v}\chi$, $\not\nabla \operatorname{tr}_{\not g} \chi$, and the curvature components of the acoustical metric.

right-hand side of (2-38a), the expression involves at most N_{top} derivatives on L and Ψ, and we can control such terms using the bounds we have obtained thus far. In total, using the symmetry property $\chi_{BA}=\chi_{AB}$, isolating the terms featuring the top-order derivatives of the components L^a, and estimating the remaining terms with (6-1)–(6-5) and Propositions 8.6 and 8.14, we obtain

$$\begin{aligned}\|\mu(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\not{\nabla}^A\chi_{AB}-\mu(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\{g_{ab}(\not{g}^{-1})^{AC}(\not{\nabla}_C\not{d}_B L^a)(\not{d}_A x^b)\}\|_{L^2(\Sigma_t^u)}&\\ \lesssim \|\mu\mathcal{P}^{N_{\mathrm{top}}+1}\Psi\|_{L^2(\Sigma_t^u)}+\|\mu\mathcal{P}^{[1,N_{\mathrm{top}}]}(L^i,\Psi)\|_{L^2(\Sigma_t^u)}\lesssim \mathring{\epsilon}^{1/2}\mu_\star^{-M_*+0.9}(t).&\quad\text{(A-12)}\end{aligned}$$

We then compute $(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\not{d}_B\,\mathrm{tr}_{\not{g}}\,\chi$ in a similar manner using (2-38b) to obtain

$$\|\mu(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\not{d}_B\,\mathrm{tr}_{\not{g}}\,\chi-\mu(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\{g_{ab}(\not{g}^{-1})^{AC}(\not{\nabla}_B\not{d}_C L^a)(\not{d}_A x^b)\}\|_{L^2(\Sigma_t^u)}\lesssim \mathring{\epsilon}^{1/2}\mu_\star^{-M_*+0.9}(t).\quad\text{(A-13)}$$

In view of the commutation identity $\not{\nabla}_C\not{d}_B L^a=\not{\nabla}_B\not{d}_C L^a$ mentioned above (which implies that the second terms on left-hand sides of (A-12) and (A-13) coincide), we can use (A-12), (A-13), and the triangle inequality to obtain

$$\begin{aligned}\|\mu(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\not{\nabla}^A\chi_{AB}\|_{L^2(\Sigma_t^u)}&\lesssim\|\mu(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\not{d}_B\,\mathrm{tr}_{\not{g}}\,\chi\|_{L^2(\Sigma_t^u)}+\mathring{\epsilon}^{1/2}\mu_\star^{-M_*+0.9}(t)\\&\lesssim\|\mu\not{\mathcal{P}}^{N_{\mathrm{top}}}\,\mathrm{tr}_{\not{g}}\,\chi\|_{L^2(\Sigma_t^u)}+\mathring{\epsilon}^{1/2}\mu_\star^{-M_*+0.9}(t),\end{aligned}\quad\text{(A-14)}$$

where to obtain the last line, we used the commutation identity $(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\not{d}_B\,\mathrm{tr}_{\not{g}}\,\chi=\not{d}_B\not{\mathcal{P}}^{N_{\mathrm{top}}-1}\,\mathrm{tr}_{\not{g}}\,\chi$ (in which we are thinking of both sides as $\ell_{t,u}$-tangent one-forms with components corresponding to the index B), the schematic identity $\not{d}=\mathrm{f}(L^i,\Psi)Y+\mathrm{f}(L^i,\Psi)Z$, and Proposition 8.6.

Now since $(\not{\mathrm{div}}_{\not{g}}\hat{\chi})_B=(\not{\mathrm{div}}_{\not{g}}\chi)_B-\frac{1}{2}\not{d}_B\,\mathrm{tr}_{\not{g}}\,\chi=\not{\nabla}^A\chi_{AB}-\frac{1}{2}\not{d}_B\,\mathrm{tr}_{\not{g}}\,\chi$, we deuce from the estimate (A-14) that

$$\|\mu(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\not{\mathrm{div}}_{\not{g}}\hat{\chi}\|_{L^2(\Sigma_t^u)}\lesssim\|\mu\not{\mathcal{P}}^{N_{\mathrm{top}}}\,\mathrm{tr}_{\not{g}}\,\chi\|_{L^2(\Sigma_t^u)}+\mathring{\epsilon}^{1/2}\mu_\star^{-M_*+0.9}(t).\quad\text{(A-15)}$$

<u>Step 2</u>: commuting $\not{\mathrm{div}}_{\not{g}}$ with $\not{\mathcal{L}}_{\not{\mathcal{P}}}$ derivatives. We now deduce from (A-15) an estimate for $\not{\mathrm{div}}_{\not{g}}(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\hat{\chi}$. For this, we simply note that the commutator $[\not{\mathrm{div}}_{\not{g}},(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}]\hat{\chi}$ can be controlled by up to N_{top} $\not{\mathcal{P}}$ derivatives of Ψ and L^i. Hence, by (A-15), (6-1)–(6-5), and Propositions 8.6 and 8.14, we have

$$\begin{aligned}\|\mu\not{\mathrm{div}}_{\not{g}}(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\hat{\chi}\|_{L^2(\Sigma_t^u)}&\lesssim\|\mu(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\not{\mathrm{div}}_{\not{g}}\hat{\chi}\|_{L^2(\Sigma_t^u)}+\|\mu\not{\mathcal{P}}^{[1,N_{\mathrm{top}}]}(\Psi,L^i)\|_{L^2(\Sigma_t^u)}\\&\lesssim\|\mu\not{\mathcal{P}}^{N_{\mathrm{top}}}\,\mathrm{tr}_{\not{g}}\,\chi\|_{L^2(\Sigma_t^u)}+\mathring{\epsilon}^{1/2}\mu_\star^{-M_*+0.9}(t).\end{aligned}\quad\text{(A-16)}$$

<u>Step 3</u>: bounding the trace-part of $(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\hat{\chi}$. By definition, $\mathrm{tr}_{\not{g}}\,\hat{\chi}=0$. Note that the commutator $[\not{g}^{-1},(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}]\hat{\chi}$ can be controlled by up to $N_{\mathrm{top}}-1$ $\not{\mathcal{P}}$ derivatives of Ψ and L^i. Hence, this commutator can be treated in the same way we treated the commutator term in Step 2, which yields the bound

$$\|\,\mathrm{tr}_{\not{g}}(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\hat{\chi}\|_{L^2(\Sigma_t^u)}\lesssim\|[\not{g}^{-1},(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}]\hat{\chi}\|_{L^2(\Sigma_t^u)}\lesssim\mathring{\epsilon}^{1/2}\mu_\star^{-M_*+0.9}(t).\quad\text{(A-17)}$$

Moreover, we can take a further $\not{\mathcal{P}}$-derivative of $\mathrm{tr}_{\not{g}}(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\hat{\chi}$, and the resulting term can be controlled by up to N_{top} $\not{\mathcal{P}}$ derivatives of Ψ and L^i. Therefore, using (6-1)–(6-5) and Propositions 8.6 and 8.14, we obtain

$$\|\mu\not{\nabla}(\mathrm{tr}_{\not{g}}(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\hat{\chi})\|_{L^2(\Sigma_t^u)}\lesssim\|\mu\not{\mathcal{P}}^{[1,N_{\mathrm{top}}]}(\Psi,L^i)\|_{L^2(\Sigma_t^u)}\lesssim\mathring{\epsilon}^{1/2}\mu_\star^{-M_*+0.9}(t).\quad\text{(A-18)}$$

Step 4: elliptic estimates. Define ξ to be the $g\!\!\!/$-trace-free part of $(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\hat{\chi}$, i.e.,

$$\xi_{AB} \doteq (\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\hat{\chi}_{AB} - \tfrac{1}{2}g\!\!\!/_{AB}\,\mathrm{tr}_{g\!\!\!/}(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\hat{\chi}. \tag{A-19}$$

The term $(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\hat{\chi}_{AB}$ on the right-hand side of (A-19) can be written using (2-38a), (2-38b) as an expression of up to N_{top} $\mathcal{P}$ derivatives of Ψ and L^i. Hence, by (2-38a), (2-38b), (6-1)–(6-5), and Propositions 8.6 and 8.14, we obtain

$$\|(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\hat{\chi}\|_{L^2(\Sigma_t^u)} \lesssim \|\mathcal{P}^{[1,N_{\mathrm{top}}]}(\Psi, L^i)\|_{L^2(\Sigma_t^u)} \lesssim \mathring{\epsilon}^{1/2}\mu_\star^{-M_*+0.9}(t). \tag{A-20}$$

Combining (A-20) with (A-17), we find that

$$\|\xi\|_{L^2(\Sigma_t^u)} \lesssim \mathring{\epsilon}^{1/2}\mu_\star^{-M_*+0.9}(t). \tag{A-21}$$

Moreover, in view of the algebraic relation

$$\mathrm{d}\!\!\!/\mathrm{iv}_{g\!\!\!/}\,\xi = \mathrm{d}\!\!\!/\mathrm{iv}_{g\!\!\!/}(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\hat{\chi} - \tfrac{1}{2}\nabla\!\!\!\!/(\mathrm{tr}_{g\!\!\!/}(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\hat{\chi})$$

and the estimates (A-16) and (A-18), we have

$$\|\mu\,\mathrm{d}\!\!\!/\mathrm{iv}_{g\!\!\!/}\,\xi\|_{L^2(\Sigma_t^u)} \lesssim \|\mu\not{\mathcal{P}}^{N_{\mathrm{top}}}\,\mathrm{tr}_{g\!\!\!/}\,\chi\|_{L^2(\Sigma_t^u)} + \mathring{\epsilon}^{1/2}\mu_\star^{-M_*+0.9}(t). \tag{A-22}$$

Therefore, applying the elliptic estimates in Lemma A.9 on $\ell_{t,u}$ with ξ as in (A-19) and $\mu = \mu$, integrating over $u \in [0, U_0]$, and using (A-18), (A-20), (A-21), and (A-22), as well as the Gauss curvature estimate in Proposition A.10 and the estimates of Proposition 8.6 (including the bound $|\nabla\!\!\!\!/\mu| \lesssim \mathring{\epsilon}^{1/2}$ that it implies), we obtain

$$\begin{aligned}
\|\mu(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}}\hat{\chi}\|_{L^2(\Sigma_t^u)} &\lesssim \|\mu\nabla\!\!\!\!/(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\hat{\chi}\|_{L^2(\Sigma_t^u)} + \|\mu(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\mathrm{top}}-1}\hat{\chi}\|_{L^2(\Sigma_t^u)} \\
&\lesssim \|\mu\nabla\!\!\!\!/\xi\|_{L^2(\Sigma_t^u)} + \mathring{\epsilon}^{1/2}\mu_\star^{-M_*+0.9}(t) \\
&\lesssim \|\mu\,\mathrm{d}\!\!\!/\mathrm{iv}_{g\!\!\!/}\,\xi\|_{L^2(\Sigma_t^u)} + (\|\mathfrak{K}_{g\!\!\!/}\|^{1/2}_{L^\infty(\Sigma_t)} + \|\nabla\!\!\!\!/\mu\|_{L^\infty(\Sigma_t)})\|\xi\|_{L^2(\Sigma_t^u)} + \mathring{\epsilon}^{1/2}\mu_\star^{-M_*+0.9}(t) \\
&\lesssim \|\mu\not{\mathcal{P}}^{N_{\mathrm{top}}}\,\mathrm{tr}_{g\!\!\!/}\,\chi\|_{L^2(\Sigma_t^u)} + \mathring{\epsilon}^{1/2}\mu_\star^{-M_*+0.9}(t),
\end{aligned}$$

which is what we wanted to prove. □

A6. ***The partial energies.*** To derive our top-order energy estimates for the wave equations, we will use the approach of [36], which relies on distinguishing the "full energies" featured in definitions (3-2a)–(3-2e) (which control all wave variables) from the "partial energies," which are captured by the next definition. The main point is that the partial energies do not control the difficult almost Riemann invariant $\mathcal{R}_{(+)}$ (it is difficult in the sense that the shock formation is driven by the relative largeness of $|\breve{X}\mathcal{R}_{(+)}|$), and it turns out that this leads to easier error terms in the corresponding energy identities. Importantly, we need to distinguish the partial energies from the full energies in order to close the proof using a uniform number of derivatives[76] N_{top}, that is, a number derivatives that does not depend on the equation of state or any parameters in the problem; see the arguments in Section A9 for clarification on the role played by the

[76]We could close the proof without introducing the partial energies, but those simpler, less precise arguments would allow for the possibility that the number of derivatives needed to close the estimates might depend on the equation of state, $\bar{\varrho}$, $\mathring{\sigma}$, $\mathring{\delta}$ and $\mathring{\delta}_*^{-1}$.

partial energies in allowing us to close the proof using a number of derivatives that is independent of the equation of state and all parameters in the problem.

Definition A.12 (the partial energies). At the top-order, we define the partial energy by

$$\mathbb{E}^{(\text{Partial})}_{N_{\text{top}}}(t,u) \doteq \sup_{t'\in[0,t)} \sum_{\widetilde{\Psi}\in\{\mathcal{R}_{(-)},v^2,v^3,s\}} \left(\|\breve{X}\mathcal{P}^{N_{\text{top}}}\widetilde{\Psi}\|^2_{L^2(\Sigma^u_{t'})} + \|\sqrt{\mu}\mathcal{P}^{N_{\text{top}}+1}\widetilde{\Psi}\|^2_{L^2(\Sigma^u_{t'})}\right).$$

Similarly, we separate the contribution of $\mathcal{R}_{(+)}$ from that of other components of Ψ and define $\mathbb{F}^{(\text{Partial})}_{N_{\text{top}}}$, $\mathbb{K}^{(\text{Partial})}_{N_{\text{top}}}$, $\mathbb{Q}^{(\text{Partial})}_{N_{\text{top}}}$ in an analogous way, that is, as in Section 3B, but without the $\mathcal{R}_{(+)}$-involving terms.

A7. *L^2 estimates for the top-order derivatives of* $\mathrm{tr}_{g\!\!/}\,\chi$ *tied to the modified quantities.*

Proposition A.13 (L^2 estimates for the top-order derivatives of $\mathrm{tr}_{g\!\!/}\,\chi$ tied to the fully modified quantities). *There exists an absolute positive constant $M_{\text{abs}} \in \mathbb{N}$, a positive constant $C' \in \mathbb{N}$, and a constant $C > 0$ (each having the properties described in Section A1) such that the following estimates (whose right-hand sides involve the wave energies* (3-2a)–(3-2e) *as well as the partial energies of Definition A.12) holds for every $(t,u) \in [0, T_{(\text{Boot})}) \times [0, U_0]$:*

$$\begin{aligned}
&\|(\breve{X}\mathcal{R}_{(+)})\mathcal{P\!\!\!/}^{N_{\text{top}}}\,\mathrm{tr}_{g\!\!/}\,\chi\|_{L^2(\Sigma^u_t)} \\
&\quad \le \textit{non-vorticity-involving terms on the right-hand side of } [36, (14.27)] \\
&\qquad \textit{with the boxed constants replaced by } \boxed{M_{\text{abs}}} \textit{ and the constant } C_* \textit{ replaced by } \boxed{C'} \\
&\qquad\qquad + C\mathring{\epsilon}\mu_\star^{-M_*+0.9}(t) + C\mu_\star^{-1}(t)\int_{t'=0}^{t'=t} \|\mathcal{P}^{[1,N_{\text{top}}]}\vec{\mathfrak{G}}\|_{L^2(\Sigma^u_{t'})}\,dt', \qquad \text{(A-23)}
\end{aligned}$$

and

$$\begin{aligned}
&\|\mu\mathcal{P\!\!\!/}^{N_{\text{top}}}\,\mathrm{tr}_{g\!\!/}\,\chi\|_{L^2(\Sigma^u_t)} \\
&\quad \lesssim \textit{non-vorticity-involving terms on the right-hand side of } [36, (14.28)] \\
&\qquad\qquad + \mathring{\epsilon}\mu_\star^{-M_*+1.9}(t) + \int_{t'=0}^{t'=t} \|\mathcal{P}^{[1,N_{\text{top}}]}\vec{\mathfrak{G}}\|_{L^2(\Sigma^u_{t'})}\,dt'. \qquad \text{(A-24)}
\end{aligned}$$

Remark A.14. We clarify that in the proofs of [36, (14.27)] and [36, (14.28)], the vorticity-involving inhomogeneous terms in the wave equations led to error integrals on the right-hand sides of [36, (14.27)] and [36, (14.28)] that involved the vorticity energies; in contrast, on the right-hand sides of (A-23)–(A-24), the vorticity-involving terms are not explicitly indicated because we have absorbed them into our definition of the wave equation inhomogeneous term $\vec{\mathfrak{G}}$.

Proof. The proofs of both estimates are similar. We first discuss the proof of (A-24) in Steps 1–2. In Step 3, we describe the changes we need in order to obtain (A-23). Throughout this proof, we freely use the observations made in Step 0 of the proof of Proposition A.11.

Following [36; 52], in order to bound $\mu\mathcal{P\!\!\!/}^{N_{\text{top}}}\,\mathrm{tr}_{g\!\!/}\,\chi$, we first control the fully modified quantity (recall the definition in (A-7a)), and then bound the difference of $\mu\mathcal{P\!\!\!/}^{N_{\text{top}}}\,\mathrm{tr}_{g\!\!/}\,\chi$ and the fully modified quantity. See the corresponding estimates in [36, Lemma 13.9, Proposition 13.11, Lemma 14.14].

Step 1: controlling the inhomogeneous terms in (A-9). We first estimate the two new terms on the right-hand sides of (A-9) in the following norms (recall that here we are assuming that in (A-9), $\mathcal{P}^{N_{\text{top}}}$ is equal to a pure $\ell_{t,u}$-tangential operator $\not{\mathcal{P}}^{N_{\text{top}}}$):

$$\int_{t'=0}^{t'=t} \|\not{\mathcal{P}}^{N_{\text{top}}}(\mu|\hat{\chi}|^2)\|_{L^2(\Sigma_{t'}^u)}\,dt', \qquad \int_{t'=0}^{t'=t} \|\not{\mathcal{P}}^{N_{\text{top}}}(\vec{G}_{LL}\diamond\vec{\mathfrak{G}})\|_{L^2(\Sigma_{t'}^u)}\,dt'. \tag{A-25}$$

Step 1(a): the $\not{\mathcal{P}}^{N_{\text{top}}}(\mu|\hat{\chi}|^2)$ term. For the first term in (A-25), the most (and indeed only) difficult contribution arises when all operators $\not{\mathcal{P}}^{N_{\text{top}}}$ fall on one factor of $\hat{\chi}$. For the lower-order terms, we use the identities (2-38a), (2-38b), and $\hat{\chi}_{AB} = \chi_{AB} - \frac{1}{2}\not{g}_{AB}\,\mathrm{tr}_{\not{g}}\,\chi$, (6-3)–(6-5), and Proposition 8.6 to obtain the pointwise estimates

$$|\not{\mathcal{P}}^{N_{\text{top}}}(\mu|\hat{\chi}|^2) - 2\mu\hat{\chi}^{\sharp\sharp}(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\text{top}}}\hat{\chi}| \lesssim \mathring{\epsilon}^{1/2}|\not{\mathcal{P}}^{[1,N_{\text{top}}]}(\Psi, L^i, \mu)|. \tag{A-26}$$

From (6-1), (6-2) and Proposition 8.14, and the estimate (A-26), we see that

$$\|\not{\mathcal{P}}^{N_{\text{top}}}(\mu|\hat{\chi}|^2) - 2\mu\hat{\chi}^{\sharp\sharp}(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\text{top}}}\hat{\chi}\|_{L^2(\Sigma_t^u)} \lesssim \mathring{\epsilon}^{1/2}\|\not{\mathcal{P}}^{[1,N_{\text{top}}]}(\Psi, L^i, \mu)\|_{L^2(\Sigma_t^u)} \lesssim \mathring{\epsilon}\mu_\star^{-M_*+1.4}(t). \tag{A-27}$$

On the other hand, the top-order derivative $\mu(\not{\mathcal{L}}_{\not{\mathcal{P}}})^{N_{\text{top}}}\hat{\chi}$ can be bounded using Proposition A.11, while the low-order factor $\hat{\chi}^{\sharp\sharp}$ can be bounded in L^∞ by $\lesssim \mathring{\epsilon}^{1/2}$ by virtue of the bootstrap assumptions (6-3)–(6-5) and the estimates of Proposition 8.6. Therefore, combining (A-27) and Proposition A.11, and then using Proposition 8.11, we bound the first term in (A-25) as

$$\begin{aligned}\int_{t'=0}^{t'=t} \|\not{\mathcal{P}}^{N_{\text{top}}}(\mu|\hat{\chi}|^2)\|_{L^2(\Sigma_{t'}^u)}\,dt' &\lesssim \mathring{\epsilon}^{1/2}\int_{t'=0}^{t'=t}\|\mu\not{\mathcal{P}}^{N_{\text{top}}}\,\mathrm{tr}_{\not{g}}\,\chi\|_{L^2(\Sigma_{t'}^u)}\,dt' + \mathring{\epsilon}\int_{t'=0}^{t'=t}\mu_\star^{-M_*+0.9}(t')\,dt' \\ &\lesssim \mathring{\epsilon}^{1/2}\int_{t'=0}^{t'=t}\|\mu\not{\mathcal{P}}^{N_{\text{top}}}\,\mathrm{tr}_{\not{g}}\,\chi\|_{L^2(\Sigma_{t'}^u)}\,dt' + \mathring{\epsilon}\mu_\star^{-M_*+1.9}(t).\end{aligned} \tag{A-28}$$

Step 1(b): the $\not{\mathcal{P}}^{N_{\text{top}}}(\vec{G}_{LL}\diamond\vec{\mathfrak{G}})$ term. To handle the second term in (A-25), we simply use Hölder's inequality together with (6-1)–(6-5), Propositions 8.6, 8.14, the assumption (12-1), and Proposition 8.11 to obtain the bound

$$\begin{aligned}&\int_{t'=0}^{t'=t}\|\not{\mathcal{P}}^{N_{\text{top}}}(\vec{G}_{LL}\diamond\vec{\mathfrak{G}})\|_{L^2(\Sigma_{t'}^u)}\,dt' \\ &\quad\lesssim \int_{t'=0}^{t'=t}\left\{\|\not{\mathcal{P}}^{[2,N_{\text{top}}]}(\Psi, L^i)\|_{L^2(\Sigma_{t'}^u)}\|\not{\mathcal{P}}^{\leq\lceil N_{\text{top}}/2\rceil}\vec{\mathfrak{G}}\|_{L^\infty(\Sigma_{t'})} + \|\not{\mathcal{P}}^{[1,N_{\text{top}}]}\vec{\mathfrak{G}}\|_{L^2(\Sigma_{t'}^u)}\right\}dt' \\ &\quad\lesssim \int_{t'=0}^{t'=t}\left\{\mathring{\epsilon}\mu_\star^{-M_*+1.4}(t') + \|\not{\mathcal{P}}^{[1,N_{\text{top}}]}\vec{\mathfrak{G}}\|_{L^2(\Sigma_{t'}^u)}\right\}dt' \\ &\quad\lesssim \mathring{\epsilon}\mu_\star^{-M_*+2.4}(t) + \int_{t'=0}^{t'=t}\|\mathcal{P}^{[1,N_{\text{top}}]}\vec{\mathfrak{G}}\|_{L^2(\Sigma_{t'}^u)}\,dt'.\end{aligned} \tag{A-29}$$

Step 2: bounding the fully modified quantity. The fully modified quantity ${}^{(\not{\mathcal{P}}^{N_{\text{top}}})}\mathscr{X}$ satisfies the transport equation (A-9) in the L-direction. We use the arguments given in [36, Proposition 13.11] to integrate the transport equation to obtain a pointwise estimate for ${}^{(\not{\mathcal{P}}^{N_{\text{top}}})}\mathscr{X}$. On the right-hand side of the pointwise estimate there appears, in particular, the time integral of the new terms $\mathscr{P}^{N_{\mathrm{top}}}(\mu|\hat{\chi}|^2)$ and $\frac{1}{2}\mathscr{P}^{N_{\mathrm{top}}}(\vec{G}_{LL}\diamond\vec{\mathfrak{G}})$ on the right-hand side of (A-9). We then take the $L^2(\Sigma_t^u)$ norm of the resulting pointwise inequality, as in the proof of [36, Lemma 14.14]. This yields an $L^2(\Sigma_t^u)$ estimate for ${}^{(\mathscr{P}^{N_{\mathrm{top}}})}\mathscr{X}$. We next use (A-7a) to algebraically express $\mu\mathscr{P}^{N_{\mathrm{top}}}\operatorname{tr}_{\not{g}}\chi$ in terms of ${}^{(\mathscr{P}^{N_{\mathrm{top}}})}\mathscr{X}$ plus a remainder term, and then use the triangle inequality to obtain an $L^2(\Sigma_t^u)$ estimate for $\mu\mathscr{P}^{N_{\mathrm{top}}}\operatorname{tr}_{\not{g}}\chi$. One of the remainder terms is $\mathscr{P}^{N_{\mathrm{top}}}\mathfrak{X}$, and it can be estimated exactly as in [36, Lemma 14.14]. In total, we find that

$$
\begin{aligned}
&\|\mu\mathscr{P}^{N_{\mathrm{top}}}\operatorname{tr}_{\not{g}}\chi\|_{L^2(\Sigma_t^u)}\\
&\quad\lesssim \|{}^{(\mathscr{P}^{N_{\mathrm{top}}})}\mathscr{X}\|_{L^2(\Sigma_t^u)}+\|\mathscr{P}^{N_{\mathrm{top}}}\mathfrak{X}\|_{L^2(\Sigma_t^u)}\\
&\quad\lesssim \text{terms on the right-hand side of [36, (14.28)]}\\
&\qquad\qquad+\int_{t'=0}^{t'=t}\|\mathscr{P}^{N_{\mathrm{top}}}(\mu|\hat{\chi}|^2)\|_{L^2(\Sigma_{t'}^u)}\,dt'+\int_{t'=0}^{t'=t}\|\mathscr{P}^{N_{\mathrm{top}}}(\vec{G}_{LL}\diamond\vec{\mathfrak{G}})\|_{L^2(\Sigma_{t'}^u)}\,dt'\\
&\quad\lesssim \text{terms on the right-hand side of [36, (14.28)]}\\
&\qquad\qquad+\mathring{\epsilon}\mu_\star^{-M_*+1.9}(t)+\mathring{\epsilon}^{1/2}\int_{t'=0}^{t'=t}\|\mu\mathscr{P}^{N_{\mathrm{top}}}\operatorname{tr}_{\not{g}}\chi\|_{L^2(\Sigma_{t'}^u)}\,dt'+\int_{t'=0}^{t'=t}\|\mathscr{P}^{[1,N_{\mathrm{top}}]}\vec{\mathfrak{G}}\|_{L^2(\Sigma_{t'}^u)}\,dt'\\
&\quad\lesssim \text{terms on the right-hand side of [36, (14.28)]}+\mathring{\epsilon}\mu_\star^{-M_*+1.9}(t)+\int_{t'=0}^{t'=t}\|\mathcal{P}^{[1,N_{\mathrm{top}}]}\vec{\mathfrak{G}}\|_{L^2(\Sigma_{t'}^u)}\,dt',
\end{aligned}
$$

where to obtain the next-to-last line, we used the estimates (A-28) and (A-29), and to obtain the last line, we used Grönwall's inequality to eliminate the factor $\mathring{\epsilon}^{1/2}\int_{t'=0}^{t'=t}\|\mu\mathscr{P}^{N_{\mathrm{top}}}\operatorname{tr}_{\not{g}}\chi\|_{L^2(\Sigma_{t'}^u)}\,dt'$ on the right-hand side. We have therefore proved (A-24).

Step 3: proof of (A-23). Estimate (A-23) can be proved using arguments that are very similar to the ones we used in the proof of (A-24), except that we need to keep track of the constants in the borderline terms, i.e., the absolute constant M_{abs} (whose precise value we do not bother to estimate here) and the parameter-dependent constant C'. This can be done exactly as in the proof of [36, (14.27)]. The only terms which are not already present in [36, (14.27)] are exactly those we encountered already in Steps 1–2. These new terms can be treated exactly as in the proof of (A-24), since we do not have to keep track of the sharp constants for these new terms (we instead allow a general constant C). □

Proposition A.15 (L^2 estimates for the partially modified quantities). *There exists an absolute positive constant $M_{\mathrm{abs}}\in\mathbb{N}$, a positive constant $C'\in\mathbb{N}$, and a constant $C>0$ (each having the properties described in Section A1) such that the partially modified quantity ${}^{(\mathscr{P}^{N_{\mathrm{top}}-1})}\widetilde{\mathscr{X}}$ obeys the following estimates (whose right-hand sides involve the wave energies* (3-2a)–(3-2e) *as well as the partial energies of Definition A.12) for every $(t,u)\in[0,T_{(\mathrm{Boot})})\times[0,U_0]$:*

$$
\begin{aligned}
&\left\|\frac{1}{\sqrt{\mu}}(\breve{X}\mathcal{R}_{(+)})L^{(\mathscr{P}^{N_{\mathrm{top}}-1})}\widetilde{\mathscr{X}}\right\|_{L^2(\Sigma_t^u)}\\
&\quad\le \textit{terms on the right-hand side of } \text{[36, (14.32a)]} \textit{ with the boxed constants replaced by } \boxed{M_{\mathrm{abs}}}\\
&\qquad\qquad\qquad\textit{and the constant } C_* \textit{ replaced by } \boxed{C'}+C\mathring{\epsilon}\mu_\star^{-M_*+0.9}(t), \qquad \text{(A-30)}
\end{aligned}
$$

$$\left\| \frac{1}{\sqrt{\mu}} (\breve{X}\mathcal{R}_{(+)})^{(\mathcal{P}^{N_{\mathrm{top}}-1})}\widetilde{\mathscr{X}} \right\|_{L^2(\Sigma_t^u)}$$

$\leq$ *terms on the right-hand side of* [36, (14.32b)] *with the boxed constants replaced by* $\boxed{M_{\mathrm{abs}}}$ *and the constant* C_* *replaced by* $\boxed{C'} + C\mathring{\epsilon}\mu_\star^{-M_*+1.9}(t)$, (A-31)

$$\|L^{(\mathcal{P}^{N_{\mathrm{top}}-1})}\widetilde{\mathscr{X}}\|_{L^2(\Sigma_t^u)} \lesssim \textit{terms in } [36, (14.33a)] + \mathring{\epsilon}\mu_\star^{-M_*+1.4}(t), \tag{A-32}$$

$$\|^{(\mathcal{P}^{N_{\mathrm{top}}-1})}\widetilde{\mathscr{X}}\|_{L^2(\Sigma_t^u)} \lesssim \textit{terms in } [36, (14.33b)] + \mathring{\epsilon}\mu_\star^{-M_*+2.4}(t). \tag{A-33}$$

Proof. To control $L^{(\mathcal{P}^{N_{\mathrm{top}}-1})}\widetilde{\mathscr{X}}$, we bound the terms on the right-hand side of the transport equation (A-10). Note that for this estimate, the only term not already found in [36] is the term $-\mathcal{P}^{N_{\mathrm{top}}-1}(|\hat{\chi}|^2)$. Compared to the estimates for the fully modified quantity that we derived in Proposition A.13, the estimates for the partially modified quantity is simpler in two ways: the transport equation (A-10) does not feature the wave equation inhomogeneous term $\vec{\mathfrak{G}}$, and the additional term only has up to $N_{\mathrm{top}}-1$ derivatives of $\hat{\chi}$, and thus elliptic estimates are not necessary to control this term.

We now estimate $-\mathcal{P}^{N_{\mathrm{top}}-1}(|\hat{\chi}|^2)$. By (2-38a), (6-1)–(6-5), and Propositions 8.6 and 8.14, we have

$$\|\mathcal{P}^{N_{\mathrm{top}}-1}(|\hat{\chi}|^2)\|_{L^2(\Sigma_t^u)} \lesssim \mathring{\epsilon}^{1/2}\|\mathcal{P}^{[1,N_{\mathrm{top}}]}(\Psi, L^i)\|_{L^2(\Sigma_t^u)} \lesssim \mathring{\epsilon}\mu_\star^{-M_*+1.4}(t). \tag{A-34}$$

We now recall (A-10). The terms that are already in terms in [36, (6.10)] can be treated using the same arguments that were used to prove [36, (14.32a)] and [36, (14.33a)], except here we do not bother to estimate the absolute constant M_{abs} that arises in the arguments, and we have renamed the constant C_* as $\boxed{C'}$. From this fact, the estimate (A-34), and the bootstrap assumption (6-3) for $\breve{X}\mathcal{R}_{(+)}$, we deduce (A-30) and (A-32).

To obtain (A-33), we use the transport equation estimate provided by Lemma 8.13, the estimate (A-32) for the source term, Proposition 8.11, and the initial data bound $\|^{(\mathcal{P}^{N_{\mathrm{top}}-1})}\widetilde{\mathscr{X}}(0,\cdot)\|_{L^2(\Sigma_t^u)} \lesssim \mathring{\epsilon}$ obtained in the proof of [36, (14.33b)].

Similarly, (A-31) can be proved using the same arguments used in the proof of [36, (14.32b)]. The estimate is based on integrating the transport equation (A-10) along the integral curves of L and using Lemma 8.13. The only new term we have to handle comes from the $-\mathcal{P}^{N_{\mathrm{top}}-1}(|\hat{\chi}|^2)$ term on the right-hand side of (A-10), and by Lemma 8.13, this term leads to the following additional term that has to be controlled:

$$\frac{1}{\sqrt{\mu_\star(t)}}\|\breve{X}\mathcal{R}_{(+)}\|_{L^\infty(\Sigma_t^u)}\int_{t'=0}^{t'=t}\|\mathcal{P}^{N_{\mathrm{top}}-1}(|\hat{\chi}|^2)\|_{L^2(\Sigma_{t'}^u)}\,dt'.$$

In view of the bootstrap assumption (6-3), the estimate (A-34), and Proposition 8.11, we bound this additional term by $\lesssim \mu_\star^{-M_*+1.9}(t)$, which is less than or equal to the right-hand side of (A-31) as desired. □

A8. ***The main integral inequalities for the energies.*** Our main goal in this section is to prove Proposition A.17, which provides integral inequalities for the various wave energies at various derivative levels. Most of the analysis is the same as in [36]. In the next definition, we highlight the error terms in the energy estimates that are new in the present paper compared to [36]. The new terms stem from the inhomogeneous term $\vec{\mathfrak{G}}$ in the wave equations as well as the $-\mu|\hat{\chi}|^2$ term on the right-hand side of the three-dimensional Raychaudhuri equation (A-2).

Definition A.16 (new energy estimate error terms). We use the notation $\mathrm{NewError}_{N_{\mathrm{top}}}^{(\mathrm{Top})}(t,u)$ to denote any term that obeys the following bound for every $(t,u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$:

$$\begin{aligned} \mathrm{NewError}_{N_{\mathrm{top}}}^{(\mathrm{Top})}(t,u) \le C\mathring{\epsilon}^2 \mu_\star^{-2M_*+1.8}(t) + C\int_{t'=0}^{t'=t} \mu_\star^{-3/2}(t') \left\{ \int_{s=0}^{s=t'} \|\mathcal{P}^{[1,N_{\mathrm{top}}]}\vec{\mathfrak{G}}\|_{L^2(\Sigma_s^u)}\, ds \right\}^2 dt' \\ + C\|(|L\mathcal{P}^{[1,N_{\mathrm{top}}]}\Psi| + |\breve{X}\mathcal{P}^{[1,N_{\mathrm{top}}]}\Psi|)|\mathcal{P}^{[1,N_{\mathrm{top}}]}\vec{\mathfrak{G}}|\|_{L^1(\mathcal{M}_{t,u})}, \end{aligned} \tag{A-35}$$

where $C > 0$ is a constant of the type described in Section A1.

Similarly, we use the notation $\mathrm{NewError}_{N-1}^{(\mathrm{Below-Top})}(t,u)$ to denote any term that obeys the following bound for every $(t,u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$:

$$\mathrm{NewError}_{N-1}^{(\mathrm{Below-Top})}(t,u) \le C\|(|L\mathcal{P}^{[1,N-1]}\Psi| + |\breve{X}\mathcal{P}^{[1,N-1]}\Psi|)|\mathcal{P}^{[1,N-1]}\vec{\mathfrak{G}}|\|_{L^1(\mathcal{M}_{t,u})}. \tag{A-36}$$

Proposition A.17 (the main integral inequalities for the energies). *Let* $\mathbb{Q}_{[1,N]}(t,u)$, $\mathbb{K}_{[1,N]}(t,u)$ *be the wave energies from Section 3B2, and let* $\mathbb{Q}_{[1,N]}^{(\mathrm{Partial})}(t,u)$, $\mathbb{K}_{[1,N]}^{(\mathrm{Partial})}(t,u)$ *be the partial wave energies from Section A6. There exist an absolute constant* $M_{\mathrm{abs}} \in \mathbb{N}$ *and a constant* $C' \in \mathbb{N}$ *depending on the equation of state,* $\bar{\varrho}$, $\mathring{\sigma}$, $\mathring{\delta}$ *and* $\mathring{\delta}_*^{-1}$ *such that the following estimate, which is a modified version of* [36, (14.3)], *hold for every* $(t,u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$:

$$\begin{aligned} &\max\{\mathbb{Q}_{[1,N_{\mathrm{top}}]}(t,u), \mathbb{K}_{[1,N_{\mathrm{top}}]}(t,u)\} \\ &\le \boxed{M_{\mathrm{abs}}} \int_{t'=0}^{t} \frac{\|[L\mu]_-\|_{L^\infty(\Sigma_{t'}^u)}}{\mu_\star(t',u)} \mathbb{Q}_{[1,N_{\mathrm{top}}]}(t',u)\, dt' \\ &\quad + \boxed{M_{\mathrm{abs}}} \int_{t'=0}^{t'=t} \frac{\|[L\mu]_-\|_{L^\infty(\Sigma_{t'}^u)}}{\mu_\star(t',u)} \sqrt{\mathbb{Q}_{[1,N_{\mathrm{top}}]}}(t',u) \int_{s=0}^{s=t'} \frac{\|[L\mu]_-\|_{L^\infty(\Sigma_s^u)}}{\mu_\star(s,u)} \sqrt{\mathbb{Q}_{[1,N_{\mathrm{top}}]}}(s,u)\, ds\, dt' \\ &\quad + \boxed{M_{\mathrm{abs}}} \frac{\|L\mu\|_{L^\infty({}^{(-)}\Sigma_{t;t}^u)}}{\mu_\star^{1/2}(t,u)} \sqrt{\mathbb{Q}_{[1,N_{\mathrm{top}}]}}(t,u) \int_{t'=0}^{t'=t} \frac{1}{\mu_\star^{1/2}(t',u)} \sqrt{\mathbb{Q}_{[1,N_{\mathrm{top}}]}}(t',u)\, dt' \\ &\quad + \boxed{C'} \int_{t'=0}^{t} \frac{\|[L\mu]_-\|_{L^\infty(\Sigma_{t'}^u)}}{\mu_\star(t',u)} \sqrt{\mathbb{Q}_{[1,N_{\mathrm{top}}]}}(t',u) \sqrt{\mathbb{Q}_{[1,N_{\mathrm{top}}]}^{(\mathrm{Partial})}}(t',u)\, dt' \\ &\quad + \boxed{C'} \int_{t'=0}^{t'=t} \frac{\|[L\mu]_-\|_{L^\infty(\Sigma_{t'}^u)}}{\mu_\star(t',u)} \sqrt{\mathbb{Q}_{[1,N_{\mathrm{top}}]}}(t',u) \int_{s=0}^{s=t'} \frac{\|[L\mu]_-\|_{L^\infty(\Sigma_s^u)}}{\mu_\star(s,u)} \sqrt{\mathbb{Q}_{[1,N_{\mathrm{top}}]}^{(\mathrm{Partial})}}(s,u)\, ds\, dt' \\ &\quad + \boxed{C'} \frac{\|L\mu\|_{L^\infty({}^{(-)}\Sigma_{t;t}^u)}}{\mu_\star(t,u)^{1/2}} \sqrt{\mathbb{Q}_{[1,N_{\mathrm{top}}]}}(t,u) \int_{t'=0}^{t'=t} \frac{1}{\mu_\star^{1/2}(t',u)} \sqrt{\mathbb{Q}_{[1,N_{\mathrm{top}}]}^{(\mathrm{Partial})}}(t',u)\, dt' \\ &\qquad + \textit{the error terms } \mathrm{Error}_{N_{\mathrm{top}}}^{(\mathrm{Top})}(t,u) \textit{ defined by } [36, (14.4)] \\ &\qquad\quad + \textit{the error terms } \mathrm{NewError}_{N_{\mathrm{top}}}^{(\mathrm{Top})}(t,u) \textit{ defined by } (\text{A-35}). \end{aligned} \tag{A-37}$$

The set ${}^{(-)}\Sigma_{t;t}^u$ *appearing on the right-hand side of* (A-37) *is defined in*[77] [36, Definition 10.4].

[77] We have no need to state its precise definition here; later, we will simply quote the relevant estimates from [36] that are tied to this set.

Moreover, the partial wave energies obey the following estimate, which is a modified version of [36, (14.5)]:

$$\max\{\mathbb{Q}^{(\mathrm{Partial})}_{[1,N_{\mathrm{top}}]}(t,u), \mathbb{K}^{(\mathrm{Partial})}_{[1,N_{\mathrm{top}}]}(t,u)\} \leq \textit{the error terms } \mathrm{Error}^{(\mathrm{Top})}_{N_{\mathrm{top}}}(t,u) \textit{ defined by } [36, (14.4)] \\ + \textit{the error terms } \mathrm{NewError}^{(\mathrm{Top})}_{N_{\mathrm{top}}}(t,u) \textit{ defined by } (\text{A-35}). \quad (\text{A-38})$$

Finally, we have the following below-top-order estimate, which is a modified version[78] *of* [36, (14.6)]:

$$\max\{\mathbb{Q}_{[1,N-1]}(t,u), \mathbb{K}_{[1,N-1]}(t,u)\} \\ \leq C\int_{t'=0}^{t} \frac{1}{\mu_\star^{1/2}(t',u)}\sqrt{\mathbb{Q}_{[1,N-1]}}(t',u)\int_{s=0}^{t'}\frac{1}{\mu_\star^{1/2}(s,u)}\sqrt{\mathbb{Q}_{[1,N]}}(s,u)\,ds\,dt' \\ + \textit{the error terms } \mathrm{Error}^{(\mathrm{Below-Top})}_{N-1}(t,u) \textit{ defined by } [36, (14.7)] \\ + \textit{the error terms } \mathrm{NewError}^{(\mathrm{Below-Top})}_{N-1}(t,u) \textit{ defined by } (\text{A-36}). \quad (\text{A-39})$$

Proof. Step 1: proof of (A-39). We begin with (A-39), which is the easier estimate since it is below top-order. Here, we use that [36, (14.6)] is proved by differentiating the wave equation $\mu\,\square_{g(\vec{\Psi})}\Psi = \cdots$ with $\mathcal{P}^{N'}$, computing the commutator $[\mu\square_{g(\vec{\Psi})}, \mathcal{P}^{N'}]$, multiplying the commuted equation by $(1+2\mu)L\mathcal{P}^{N'}\Psi + \breve{X}\mathcal{P}^{N'}\Psi$, and then integrating (with respect to the volume form $d\varpi$ Definition 3.1) by parts over the spacetime region $\mathcal{M}_{t,u}$ (for $1 \leq N' \leq N-1$). Hence, to prove (A-39), we repeat the argument in [36], except that here we simply denote all of the inhomogeneous terms in the wave equations as $\mathfrak{G}$. That is, we start with the wave equations $\mu\square_{g(\vec{\Psi})}\Psi_\iota = \mathfrak{G}_\iota$ and commute them to obtain the wave equations $\mu\square_{g(\vec{\Psi})}\mathcal{P}^{N'}\Psi_\iota = [\mu\square_{g(\vec{\Psi})}, \mathcal{P}^{N'}]\Psi_\iota + \mathcal{P}^{N'}\mathfrak{G}_\iota$. The main point is that for the below-top-order estimates, all commutator terms $[\mu\square_{g(\vec{\Psi})}, \mathcal{P}^{N'}]\Psi_\iota$ can be handled exactly as in [36]. These commutator terms lead to the presence of the first term on the right-hand side of (A-39), as well as the error term $\mathrm{Error}^{(\mathrm{Below-Top})}_{N-1}(t,u)$ on the right-hand side of (A-39). We clarify that in the proof of [36, (14.6)], the vorticity-involving inhomogeneous terms in the wave equation led to error integrals on the right-hand side of [36, (14.6)] that involved the vorticity energies; in contrast, on the right-hand side of (A-39), the vorticity-involving terms are not explicitly indicated because we have absorbed them into our definition of $\mathfrak{G}_\iota$. Thus, to complete the proof of (A-39), we only have to discuss the contribution of the inhomogeneous term $\mathfrak{G}_\iota$. From the above discussion, it follows that we only have to show that the following energy identity error integrals are bounded above in magnitude by the right-hand side of (A-39) when $1 \leq N' \leq N-1$ and $(t,u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$:

$$\int_{\mathcal{M}_{t,u}} \{(1+2\mu)L\mathcal{P}^{N'}\Psi + \breve{X}\mathcal{P}^{N'}\Psi\}\mathcal{P}^{N'}\mathfrak{G}\,d\varpi.$$

[78] Note that the lower-order estimate [36, (14.6)] is easier and has fewer additional terms. This is because to obtain the top-order estimates [36, (14.3), (14.5)], one needs to bound all of the commutator terms, including the difficult ones identified in Proposition A.4, without losing derivatives. In contrast, to obtain the lower-order estimates [36, (14.6)], one is allowed to lose a derivative, as is manifested by the double-time-integral term on the right-hand side of (A-39). This double-time-integral will eventually be responsible for the coupling between the energies of different orders; see in particular the estimates (12-3) in the statement of Proposition 12.1.

The desired estimate is simple; in view of the L^∞ estimates for $\upmu$ provided by Proposition 8.6, we see that these error integrals are all bounded by $C\|(|L\mathcal{P}^{[1,N-1]}\Psi| + |\breve{X}\mathcal{P}^{[1,N-1]}\Psi|)|\mathcal{P}^{[1,N-1]}\mathfrak{G}|\|_{L^1(\mathcal{M}_{t,u})}$, which are exactly the error terms we have defined in (A-36).

Step 2: proof of (A-37).

Step 2(a): preliminaries. As in our proof of (A-39), to prove (A-37), the only new step compared to [36] is tracking the contribution of the wave equation inhomogeneous terms $\mathfrak{G}_\iota$ to the energy estimates. As in Step 1, one way in which this inhomogeneous term contributes to the energy estimates is through the error terms $C\|(|L\mathcal{P}^{[1,N_{\mathrm{top}}]}\Psi| + |\breve{X}\mathcal{P}^{[1,N_{\mathrm{top}}]}\Psi|)|\mathcal{P}^{[1,N_{\mathrm{top}}]}\vec{\mathfrak{G}}|\|_{L^1(\mathcal{M}_{t,u})}$, which are found on the right-hand side of (A-35). However, in the top-order case, there is a second way in which $\mathfrak{G}_\iota$ contributes to the top-order energy estimates. To explain this contribution, we first note that, as in the proof of [36, (14.3)], we have to handle some additional difficult top-order commutator terms involving the top-order derivatives of $\mathrm{tr}_{g\!\!/}\,\chi$. Specifically, these difficult top-order commutator terms are explicitly listed on the right-hand sides of (A-3)–(A-5). Recalling that we multiply the wave equation by $(1+2\upmu)L\mathcal{P}^{N'}\Psi + \breve{X}\mathcal{P}^{N'}\Psi$ to derive the wave equation energy estimates at level N', we see that up to harmless factors that are $\mathcal{O}(1)$ by virtue of the estimates of Proposition 8.7, these difficult commutator terms lead to the following three error integrals in the top-order energy estimates:

$$\int_{\mathcal{M}_{t,u}} (\breve{X}\mathcal{P}^{N_{\mathrm{top}}}\Psi)(\breve{X}\Psi){\mathcal{P}\!\!\!/}^{N_{\mathrm{top}}}\,\mathrm{tr}_{g\!\!/}\,\chi\,d\varpi,$$
$$\int_{\mathcal{M}_{t,u}} \{(1+2\upmu)L\mathcal{P}^{N_{\mathrm{top}}}\Psi\}(\breve{X}\Psi){\mathcal{P}\!\!\!/}^{N_{\mathrm{top}}}\,\mathrm{tr}_{g\!\!/}\,\chi\,d\varpi,$$
$$\int_{\mathcal{M}_{t,u}} \{(1+2\upmu)L\mathcal{P}^{N_{\mathrm{top}}}\Psi + \breve{X}\mathcal{P}^{N_{\mathrm{top}}}\Psi\}({\mathcal{P}\!\!\!/}\Psi)\upmu{\mathcal{P}\!\!\!/}^{N_{\mathrm{top}}}\,\mathrm{tr}_{g\!\!/}\,\chi\,d\varpi.$$

We will control these three terms, respectively, in Steps 2(b)–(d) below.

Step 2(b): contributions from $\int_{\mathcal{M}_{t,u}} (\breve{X}\mathcal{P}^{N_{\mathrm{top}}}\Psi)(\breve{X}\Psi){\mathcal{P}\!\!\!/}^{N_{\mathrm{top}}}\,\mathrm{tr}_{g\!\!/}\,\chi\,d\varpi$. We first consider the case $\Psi = \mathcal{R}_{(+)}$, which is by far the most difficult case. Using Hölder's inequality and the estimate (A-23) in Proposition A.13, we deduce that

$$\begin{aligned}
&\left|\int_{\mathcal{M}_{t,u}} (\breve{X}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{R}_{(+)})(\breve{X}\mathcal{R}_{(+)}){\mathcal{P}\!\!\!/}^{N_{\mathrm{top}}}\,\mathrm{tr}_{g\!\!/}\,\chi\,d\varpi\right| \\
&\le \int_{t'=0}^{t'=t} \|\breve{X}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{R}_{(+)}\|_{L^2(\Sigma_{t'}^u)} \|(\breve{X}\mathcal{R}_{(+)}){\mathcal{P}\!\!\!/}^{N_{\mathrm{top}}}\,\mathrm{tr}_{g\!\!/}\,\chi\|_{L^2(\Sigma_{t'}^u)}\,dt' \\
&\le \text{terms on the right-hand sides of [36, (14.3)] with the boxed constants replaced by } \boxed{M_{\mathrm{abs}}} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{and the constant } C_* \text{ replaced by } \boxed{C'} \\
&\quad + \underbrace{C\int_{t'=0}^{t'=t} \|\breve{X}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{R}_{(+)}\|_{L^2(\Sigma_{t'}^u)}\mathring{\epsilon}\upmu_\star^{-M_*+0.9}(t')\,dt'}_{\doteq \mathrm{I}} \\
&\quad + C\overbrace{\int_{t'=0}^{t'=t} \upmu_\star^{-1}(t)\|\breve{X}\mathcal{P}^{N_{\mathrm{top}}}\mathcal{R}_{(+)}\|_{L^2(\Sigma_{t'}^u)}\left\{\int_{s=0}^{s=t'} \|\mathcal{P}^{[1,N_{\mathrm{top}}]}\mathfrak{G}\|_{L^2(\Sigma_s^u)}\,ds\right\}dt'}^{\doteq \mathrm{II}}.
\end{aligned} \tag{A-40}$$

We clarify that Remark A.14 also applies to the terms on the right-hand sides of [36, (14.3)] (some of which also appear on the right-hand side of (A-40)).

To handle the term I in (A-40), we use Cauchy–Schwarz inequality in t' and Proposition 8.11 to deduce

$$\begin{aligned}\int_{t'=0}^{t'=t} &\|\breve{X}\mathcal{P}^{N_{\text{top}}}\mathcal{R}_{(+)}\|_{L^2(\Sigma_{t'}^u)}\mathring{\epsilon}\mu_\star^{-M_*+0.9}(t')\,dt' \\ &\lesssim \int_{t'=0}^{t'=t} \mu_\star^{-1/2}(t)\|\breve{X}\mathcal{P}^{N_{\text{top}}}\mathcal{R}_{(+)}\|^2_{L^2(\Sigma_{t'}^u)}\,dt' + \mathring{\epsilon}^2\int_{t'=0}^{t'=t}\mu_\star^{-2M_*+2.3}(t')\,dt' \\ &\lesssim \int_{t'=0}^{t'=t} \mu_\star^{-1/2}(t)\|\breve{X}\mathcal{P}^{N_{\text{top}}}\mathcal{R}_{(+)}\|^2_{L^2(\Sigma_{t'}^u)}\,dt' + \mathring{\epsilon}^2\mu_\star^{-M_*+3.3}(t). \qquad \text{(A-41)}\end{aligned}$$

For the term II in (A-40), we apply first the Cauchy–Schwarz inequality in t' and then Young's inequality to obtain

$$\begin{aligned}&\int_{t'=0}^{t'=t} \mu_\star^{-1}(t)\|\breve{X}\mathcal{P}^{N_{\text{top}}}\mathcal{R}_{(+)}\|_{L^2(\Sigma_{t'}^u)}\left\{\int_{s=0}^{s=t'}\|\mathcal{P}^{[1,N_{\text{top}}]}\mathfrak{G}\|_{L^2(\Sigma_s^u)}\,ds\right\}dt' \\ &\lesssim \int_{t'=0}^{t'=t} \mu_\star^{-1/2}(t)\|\breve{X}\mathcal{P}^{N_{\text{top}}}\mathcal{R}_{(+)}\|^2_{L^2(\Sigma_{t'}^u)}\,dt' + \int_{t'=0}^{t'=t}\mu_\star^{-3/2}(t')\left\{\int_{s=0}^{s=t'}\|\mathcal{P}^{[1,N_{\text{top}}]}\mathfrak{G}\|_{L^2(\Sigma_s^u)}\,ds\right\}^2 dt'. \qquad \text{(A-42)}\end{aligned}$$

Notice that the term

$$\int_{t'=0}^{t'=t} \mu_\star^{-1/2}(t)\|\breve{X}\mathcal{P}^{N_{\text{top}}}\mathcal{R}_{(+)}\|^2_{L^2(\Sigma_{t'}^u)}\,dt'$$

appearing on the right-hand sides of both (A-41) and (A-42) is bounded above by

$$\int_{t'=0}^{t'=t} \mu_\star^{-1/2}(t',u)\mathbb{Q}_{N_{\text{top}}}(t',u)\,dt',$$

which is among the error terms $\text{Error}^{(\text{Top})}_{N_{\text{top}}}(t,u)$ defined by [36, (14.4)]. Therefore, combining (A-40)–(A-42) and taking into account (A-35), we obtain that

$$\|(\breve{X}\mathcal{P}^{N_{\text{top}}}\mathcal{R}_{(+)})(\breve{X}\mathcal{R}_{(+)})\not{\mathcal{P}}^{N_{\text{top}}}\operatorname{tr}_{\not{g}}\chi\|_{L^1(\mathcal{M}_{t,u})} \leq \text{the right-hand side of (A-37)} \qquad \text{(A-43)}$$

as desired.

We also need to bound the integral

$$\int_{\mathcal{M}_{t,u}} (\breve{X}\mathcal{P}^{N_{\text{top}}}\Psi)(\breve{X}\Psi)\not{\mathcal{P}}^{N_{\text{top}}}\operatorname{tr}_{\not{g}}\chi\,d\varpi$$

in the remaining cases $\Psi \in \{R_{(-)}, v^2, v^3, s\}$. As we further explain below in Step 3, a similar argument allows us to bound these error integrals by exploiting one crucial simplifying feature: these error integrals are bounded by the right-hand side of (A-37), but *without the difficult boxed-constant-involving integrals* on the right-hand side. The difference is that we can take advantage of the smallness of the factor $\|\breve{X}\Psi\|_{L^\infty(\Sigma_t)} \leq \mathring{\epsilon}^{1/2}$ (valid for $\Psi \in \{\mathcal{R}_{(-)}, v^2, v^3, s\}$ — but not for $\mathcal{R}_{(+)}$!), which is provided by the bootstrap assumption (6-4); this allows us to avoid the error terms with large boxed constants and thus allows us to relegate the contribution of these error integrals to the error term $\text{Error}^{(\text{Top})}_{N_{\text{top}}}(t,u)$ on the right-hand side of (A-37); we refer to [36, pg. 154] for further details.

Step 2(c): contributions from $\int_{\mathcal{M}_{t,u}}\{(1+2\mu)L\mathcal{P}^{N_{\text{top}}}\Psi\}(\breve{X}\Psi)\not{\mathcal{P}}^{N_{\text{top}}}\operatorname{tr}_{\not{g}}\chi\,d\varpi$. We first consider the case $\Psi = \mathcal{R}_{(+)}$, which is by far the most difficult case. Unlike the error integral we controlled in Step 2(b), as in [36], this error integral can be controlled by first using the definition (A-8a) of the partially modified quantities to algebraically replace the factor $\not{\mathcal{P}}^{N_{\text{top}}} \operatorname{tr}_{\not{g}} \chi$ with a $\not{\mathcal{P}}$ derivative of ${}^{(\not{\mathcal{P}}^{N_{\text{top}}-1})}\widetilde{\mathscr{X}}$ plus remainder terms (that one controls separately), and then using integration by parts to swap the L and $\not{\mathcal{P}}$ derivatives. Notice that by Proposition A.15, the partially modified quantity obeys the same bounds as in [36, Lemma 14.19], except the estimates of Proposition A.15 feature $\mathring{\epsilon}$-multiplied terms such as $C\mathring{\epsilon}\mu_{\star}^{-M_*+1.4}(t)$ on the right-hand sides, which can be handled using arguments of the type we used to control the error term (A-41). In particular, the right-hand sides of the estimates in Proposition A.15 do not involve the wave equation inhomogeneity $\mathfrak{G}$. Hence, the error integral $\int_{\mathcal{M}_{t,u}} \{(1+2\mu)L\mathcal{P}^{N_{\text{top}}}\mathcal{R}_{(+)}\}(\breve{X}\mathcal{R}_{(+)})\not{\mathcal{P}}^{N_{\text{top}}} \operatorname{tr}_{\not{g}} \chi \, d\varpi$ can be bounded using exactly the same arguments given in [36, Lemma 14.17] and [52, Lemma 14.12], except with the boxed constants from [36] replaced by $\boxed{M_{\text{abs}}}$ and the constant C_* from [36] replaced by $\boxed{C'}$. As a consequence, the error integral under consideration is bounded above in magnitude by the right-hand side of (A-37).

To bound the integral

$$\int_{\mathcal{M}_{t,u}} \{(1+2\mu)L\mathcal{P}^{N_{\text{top}}}\Psi\}(\breve{X}\Psi)\not{\mathcal{P}}^{N_{\text{top}}} \operatorname{tr}_{\not{g}} \chi \, d\varpi$$

in the remaining cases $\Psi \in \{R_{(-)}, v^2, v^3, s\}$, we can again (as in Step 2(b)) take advantage of the smallness $\|\breve{X}\Psi\|_{L^\infty(\Sigma_t)} \le \mathring{\epsilon}^{1/2}$ (valid for $\Psi \in \{\mathcal{R}_{(-)}, v^2, v^3, s\}$ — but not for $\mathcal{R}_{(+)}$!), which is provided by the bootstrap assumption (6-4). This again allows us to relegate the contribution of these integrals to the error term $\text{Error}_{N_{\text{top}}}^{(\text{Top})}(t,u)$ on the right-hand side of (A-37); see [36, p. 154] for further details.

<u>Step 2(d)</u>: contributions from $\int_{\mathcal{M}_{t,u}} \{(1+2\mu)L\mathcal{P}^{N_{\text{top}}}\Psi + \breve{X}\mathcal{P}^{N_{\text{top}}}\Psi\}(\not{\mathcal{P}}\Psi)\mu\not{\mathcal{P}}^{N_{\text{top}}} \operatorname{tr}_{\not{g}} \chi \, d\varpi$. This error integral is similar to the one we treated in Step 2(b), but easier. Here are the differences:

- There is an additional μ factor.
- There is a $L\mathcal{P}^{N_{\text{top}}}\Psi$ term, in addition to a $\breve{X}\mathcal{P}^{N_{\text{top}}}\Psi$ term.
- There is a factor of $\not{\mathcal{P}}\Psi$ instead of $\breve{X}\Psi$.

Notice that due to the additional factor of μ, we can control the $L^2(\Sigma_t^u)$ norm of $\sqrt{\mu}L\mathcal{P}^{N_{\text{top}}}\Psi$ by the $\mathbb{Q}_{N_{\text{top}}}$ energy (recall the definition (3-2a) for the energy). Moreover, comparing (6-5) with (6-3), we see that the factor $\not{\mathcal{P}}\Psi$ gives an additional $\mathring{\epsilon}^{1/2}$ L^∞-smallness factor compared to $\breve{X}\mathcal{R}_{(+)}$. Therefore, we can use Hölder's inequality, (6-5), the L^∞ bound for μ in Proposition 8.6, (A-24) in Proposition A.13, and Proposition 8.11 and argue as in Step 2(b) (taking into account (A-35)) to obtain

$$\begin{aligned}
&\left|\int_{\mathcal{M}_{t,u}} \{(1+2\mu)L\mathcal{P}^{N_{\text{top}}}\Psi + \breve{X}\mathcal{P}^{N_{\text{top}}}\Psi\}(\not{\mathcal{P}}\Psi)\mu\not{\mathcal{P}}^{N_{\text{top}}} \operatorname{tr}_{\not{g}} \chi \, d\varpi\right| \\
&\quad\lesssim \mathring{\epsilon}^{1/2}\int_{t'=0}^{t'=t} \mu_{\star}^{-1/2}(t',u)\big(\|\breve{X}\mathcal{P}^{N_{\text{top}}}\Psi\|_{L^2(\Sigma_{t'}^u)} + \|\sqrt{\mu}L\mathcal{P}^{N_{\text{top}}}\Psi\|_{L^2(\Sigma_{t'}^u)}\big)\|\mu\not{\mathcal{P}}^{N_{\text{top}}} \operatorname{tr}_{\not{g}} \chi\|_{L^2(\Sigma_{t'}^u)}\, dt' \\
&\quad\lesssim \mathring{\epsilon}^{1/2}\int_{t'=0}^{t'=t} \mu_{\star}^{-1/2}(t',u)\mathbb{Q}_{N_{\text{top}}}(t',u)\, dt' + \mathring{\epsilon}^{5/2}\mu_{\star}^{-2M_*+4.3}(t) \\
&\qquad\qquad + \mathring{\epsilon}^{1/2}\int_{t'=0}^{t'=t} \mu_{\star}^{-1/2}(t',u)\left\{\int_{s=0}^{s=t'} \|\mathcal{P}^{[1,N_{\text{top}}]}\mathfrak{G}\|_{L^2(\Sigma_s^u)}\, ds\right\}^2 dt' \\
&\quad\le \text{non-boxed-constant-involving terms on the right-hand side of (A-37).}
\end{aligned}$$

Combining Steps 2(a)–2(d), we arrive at the desired bound (A-37).

Step 3: proof of (A-38). In this step, we only have to derive top-order energy estimates for $\mathcal{R}_{(-)}, v^2, v^3, s$. This is in contrast to Step 2, in which we also had to derive energy estimates for $\mathcal{R}_{(+)}$. The proof of (A-38) is the same as the proof of [36, (14.5)], except we have to account for the contribution of the inhomogeneous terms $\mathfrak{G}_\iota$ in the wave equations satisfied by $\widetilde{\Psi} \in \{\mathcal{R}_{(-)}, v^2, v^3, s\}$. For the same reason as in Step 2, these inhomogeneous terms lead to error integrals that are controlled by the terms $\text{NewError}_{N_{\text{top}}}^{(\text{Top})}(t,u)$ on the right-hand side of (A-38). We clarify that the proof of (A-38) requires that we control the difficult error integrals

$$\int_{\mathcal{M}_{t,u}} (\breve{X}\mathcal{P}^{N_{\text{top}}}\widetilde{\Psi})(\breve{X}\widetilde{\Psi})\not{\mathcal{P}}^{N_{\text{top}}} \operatorname{tr}_{\not{g}} \chi \, d\varpi,$$

$$\int_{\mathcal{M}_{t,u}} \{(1+2\mu)L\mathcal{P}^{N_{\text{top}}}\widetilde{\Psi}\}(\breve{X}\widetilde{\Psi})\not{\mathcal{P}}^{N_{\text{top}}} \operatorname{tr}_{\not{g}} \chi \, d\varpi,$$

$$\int_{\mathcal{M}_{t,u}} \{(1+2\mu)L\mathcal{P}^{N_{\text{top}}}\widetilde{\Psi} + \breve{X}\mathcal{P}^{N_{\text{top}}}\widetilde{\Psi}\}(\mathcal{P}\widetilde{\Psi})\mu\not{\mathcal{P}}^{N_{\text{top}}} \operatorname{tr}_{\not{g}} \chi \, d\varpi,$$

as in Step 2. In Step 2, the first two of these error integrals led to error terms that are controlled by the boxed-constant-involving terms on the right-hand side of (A-37). However, in Step 3, we can take advantage of the smallness of the factors $\breve{X}\widetilde{\Psi}$ in these integrals. That is, we can exploit the smallness estimate $\|\breve{X}\widetilde{\Psi}\|_{L^\infty(\Sigma_t)} \le \mathring{\epsilon}^{1/2}$ (valid for $\widetilde{\Psi} \in \{\mathcal{R}_{(-)}, v^2, v^3, s\}$ — but not for $\mathcal{R}_{(+)}$!), which is provided by the bootstrap assumption (6-4); this allows us to avoid the error terms with large boxed constants (which are found on the right-hand of (A-37)), and allow us to relegate the contribution of the corresponding error integrals to the error term $\text{Error}_{N_{\text{top}}}^{(\text{Top})}(t,u)$ on the right-hand side of (A-38). See [36, p. 154] for further details. □

A9. *Sketch of the proof of Proposition 12.1.* The argument here is the same as in the proof of [36, Proposition 14.1], except we have to handle the additional terms in Proposition A.17.

Sketch of proof of Proposition 12.1. Step 1: the top- and penultimate- orders (proof of (12-2)). It turns out that the top-order energies are heavily coupled to the penultimate-order energies. In turn, this forces us to perform a Grönwall-type argument that simultaneously handles the top- and penultimate-order energy estimates at the same time. For these reasons, we follow the notation of [36, Proposition 14.1] and define[79]

$$F(t,u) \doteq \sup_{(\hat{t},\hat{u})\in[0,t]\times[0,u]} \iota_F^{-1}(\hat{t},\hat{u}) \max\{\mathbb{Q}_{[1,N_{\text{top}}]}(\hat{t},\hat{u}),\ \mathbb{K}_{[1,N_{\text{top}}]}(\hat{t},\hat{u})\}, \tag{A-44}$$

$$G(t,u) \doteq \sup_{(\hat{t},\hat{u})\in[0,t]\times[0,u]} \iota_G^{-1}(\hat{t},\hat{u}) \max\{\mathbb{Q}_{[1,N_{\text{top}}]}^{(\text{Partial})}(\hat{t},\hat{u}),\ \mathbb{K}_{[1,N_{\text{top}}]}^{(\text{Partial})}(\hat{t},\hat{u})\}, \tag{A-45}$$

$$H(t,u) \doteq \sup_{(\hat{t},\hat{u})\in[0,t]\times[0,u]} \iota_H^{-1}(\hat{t},\hat{u}) \max\{\mathbb{Q}_{[1,N_{\text{top}}-1]}(\hat{t},\hat{u}),\ \mathbb{K}_{[1,N_{\text{top}}-1]}(\hat{t},\hat{u})\}, \tag{A-46}$$

where

$$\iota_1(t) \doteq \int_{t'=0}^{t'=t} \frac{1}{\sqrt{T_{(\text{Boot})}-t'}}\,dt', \quad \iota_F(t,u) = \iota_G(t,u) \doteq \mu_\star^{-2M_*+1.8}(t)\iota_1^c(t)\iota_2^c(t)e^{ct}e^{cu},$$

$$\iota_2(t) \doteq \int_{t'=0}^{t'=t} \mu_\star^{-0.9}(t')\,dt', \qquad \iota_H(t,u) \doteq \mu_\star^{-2M_*+3.8}(t)\iota_1^c(t)\iota_2^c(t)e^{ct}e^{cu}.$$

[79]For easy comparisons with the proof of [36, Proposition 14.1], we are using the notation F, G, and H here. The reader should be careful to distinguish these functions from the different functions F and G in Definitions 2.3 and 2.12.

Following exactly the same[80] argument[81] used in the proof of [36, Proposition 14.1] (see in particular [36, (14.64)–(14.66)]), but taking into account the additional terms in Proposition A.17, we can choose $M_* \in \mathbb{N}$ and $c > 0$ sufficiently large depending on the absolute constant M_{abs} in Proposition A.17 so that the following hold[82] for every $(\hat{t}, \hat{u}) \in [0, t] \times [0, u]$:

$$\begin{aligned} F(\hat{t}, \hat{u}) &\leq C \mathring{\epsilon}^2 + \alpha_1 F(t, u) + \alpha_2 H(t, u) + \alpha_3 G(t, u) \\ &\quad + C \iota_F^{-1}(\hat{t}, \hat{u}) \int_{t'=0}^{t'=\hat{t}} \|(|L \mathcal{P}^{[1, N_{\text{top}}]} \Psi| + |\breve{X} \mathcal{P}^{[1, N_{\text{top}}]} \Psi|) \mathcal{P}^{[1, N_{\text{top}}]} \vec{\mathfrak{G}}\|_{L^1(\Sigma_{t'}^u)} \, dt' \\ &\quad + C \iota_F^{-1}(\hat{t}, \hat{u}) \int_{t'=0}^{t'=\hat{t}} \mu_\star^{-3/2}(t') \left\{ \int_{s=0}^{s=t'} \|\mathcal{P}^{[1, N_{\text{top}}]} \vec{\mathfrak{G}}\|_{L^2(\Sigma_s^u)} \, ds \right\}^2 dt', \end{aligned} \tag{A-47}$$

$$\begin{aligned} G(\hat{t}, \hat{u}) &\leq C \mathring{\epsilon}^2 + \beta_1 F(t, u) + \beta_2 H(t, u) \\ &\quad + C \iota_G^{-1}(\hat{t}, \hat{u}) \int_{t'=0}^{t'=\hat{t}} \|(|L \mathcal{P}^{[1, N_{\text{top}}]} \Psi| + |\breve{X} \mathcal{P}^{[1, N_{\text{top}}]} \Psi|) \mathcal{P}^{[1, N_{\text{top}}]} \mathfrak{G}\|_{L^1(\Sigma_{t'}^u)} \, dt' \\ &\quad + C \iota_G^{-1}(\hat{t}, \hat{u}) \int_{t'=0}^{t'=\hat{t}} \mu_\star^{-3/2}(t') \left\{ \int_{s=0}^{s=t'} \|\mathcal{P}^{[1, N_{\text{top}}]} \mathfrak{G}\|_{L^2(\Sigma_s^u)} \, ds \right\}^2 dt', \end{aligned} \tag{A-48}$$

$$\begin{aligned} H(\hat{t}, \hat{u}) &\leq C \mathring{\epsilon}^2 + \gamma_1 F(t, u) + \gamma_2 H(t, u) \\ &\quad + C \iota_H^{-1}(\hat{t}, \hat{u}) \int_{t'=0}^{t'=\hat{t}} \|(|L \mathcal{P}^{[1, N_{\text{top}}-1]} \Psi| + |\breve{X} \mathcal{P}^{[1, N_{\text{top}}-1]} \Psi|) \mathcal{P}^{[1, N_{\text{top}}-1]} \mathfrak{G}\|_{L^1(\Sigma_{t'}^u)} \, dt', \end{aligned} \tag{A-49}$$

where $C > 0$ is a constant, while α_1, α_2, α_3, β_1, β_2, γ_1 and γ_2 are constants that obey the following smallness conditions (as long as $M_* \in \mathbb{N}$ and $c > 0$ are sufficiently large):

$$\begin{aligned} \alpha_1 + 4\alpha_2\gamma_1 + \alpha_3\beta_1 + 4\alpha_3\beta_2\gamma_1 &< 1, \\ \gamma_2 &< \tfrac{3}{4}. \end{aligned} \tag{A-50}$$

At this point we fix $c > 0$ and $M_* \in \mathbb{N}$. From now on, we allow the general constants $C > 0$ to depend on these particular fixed choices of c and M_*.

For each of the three integrals on the right-hand sides of (A-47)–(A-49), we absorb $\iota_1^c(\hat{t})\iota_2^c(\hat{t})e^{c\hat{t}}e^{c\hat{u}}$ into the general constant C, and then take the supremum with respect to $\hat{t}$. For instance, for the first

[80] Here we note one minor difference compared to [36, Proposition 14.1]: that proposition was more precise with respect to u in the sense that it yielded a priori estimates in terms of powers of $\mu_\star(t, u)$, rather than $\mu_\star(t)$ (see Definition 2.16). For this reason, in the proof [36, Proposition 14.1], the definition of the analog of ι_2 involved $\mu_\star(t, u)$, and similarly for the $\mu_\star$-dependent factors on the right-hand sides of the analogs of ι_F, ι_G, and ι_H. The change we have made in this paper has no substantial effect on the analysis; at the relevant points in the proof of [36, Proposition 14.1], all of the needed estimates hold true with $\mu_\star(t)$ in place of $\mu_\star(t, u)$.

[81] The detailed argument relies on some extensions and sharpened versions of the estimates of Proposition 8.11. Given the estimates of Section 8, such as Propositions 8.6, 8.7, and 8.10, the needed estimates can be proved using the same arguments given in [36].

[82] The inequality [36, (14.64)] featured a term $CF^{1/2}(t, u)G^{1/2}(t, u)$ on the right-hand side. We used Young's inequality to bound this term by $\leq aF(t, u) + \alpha_3 G(t, u)$, where $\alpha_3 \doteq C^2/a$ and we have chosen a to be small, which allows us to absorb $aF(t, u)$ into the term $\alpha_1 F(t, u)$.

integral on the right-hand side of (A-47), we deduce that for $(\hat{t},\hat{u})\in[0,t]\times[0,u]$, we have

$$\iota_F^{-1}(\hat{t},\hat{u})\int_{t'=0}^{t'=\hat{t}}\|(|L\mathcal{P}^{[1,N_{\mathrm{top}}]}\Psi|+|\breve{X}\mathcal{P}^{[1,N_{\mathrm{top}}]}\Psi|)\mathcal{P}^{[1,N_{\mathrm{top}}]}\vec{\mathfrak{G}}\|_{L^1(\Sigma_{t'}^u)}\,dt' \\ \lesssim \sup_{\hat{t}'\in[0,t]}\mu_\star^{2M_*-1.8}(\hat{t}')\int_{t'=0}^{t'=\hat{t}'}\|(|L\mathcal{P}^{[1,N_{\mathrm{top}}]}\Psi|+|\breve{X}\mathcal{P}^{[1,N_{\mathrm{top}}]}\Psi|)\mathcal{P}^{[1,N_{\mathrm{top}}]}\vec{\mathfrak{G}}\|_{L^1(\Sigma_{t'}^u)}\,dt'.$$

We perform the same operation on the other integrals. Since we have taken a supremum, the right-hand sides are independent of $(\hat{t},\hat{u})$. We then take supremum over $(\hat{t},\hat{u})\in[0,t]\times[0,u]$ on the left-hand sides of (A-47)–(A-49) to obtain, with the same constants α_1, α_2, α_3, β_1, β_2, γ_1 and γ_2, but with a different constant C, the inequalities

$$\begin{aligned} F(t,u)\leq{}& C\mathring{\epsilon}^2+\alpha_1F(t,u)+\alpha_2H(t,u)+\alpha_3G(t,u)\\ &+C\sup_{\hat{t}\in[0,t]}\mu_\star^{2M_*-1.8}(\hat{t})\int_{t'=0}^{t'=\hat{t}}\|(|L\mathcal{P}^{[1,N_{\mathrm{top}}]}\Psi|+|\breve{X}\mathcal{P}^{[1,N_{\mathrm{top}}]}\Psi|)\mathcal{P}^{[1,N_{\mathrm{top}}]}\vec{\mathfrak{G}}\|_{L^1(\Sigma_{t'}^u)}\,dt'\\ &+C\sup_{\hat{t}\in[0,t]}\mu_\star^{2M_*-1.8}(\hat{t})\int_{t'=0}^{t'=\hat{t}}\mu_\star^{-3/2}(t')\left\{\int_{s=0}^{s=t'}\|\mathcal{P}^{[1,N_{\mathrm{top}}]}\vec{\mathfrak{G}}\|_{L^2(\Sigma_s^u)}\,ds\right\}^2dt', \end{aligned}\tag{A-51}$$

$$\begin{aligned} G(t,u)\leq{}& C\mathring{\epsilon}^2+\beta_1F(t,u)+\beta_2H(t,u)\\ &+C\sup_{\hat{t}\in[0,t]}\mu_\star^{2M_*-1.8}(\hat{t})\int_{t'=0}^{t'=\hat{t}}\|(|L\mathcal{P}^{[1,N_{\mathrm{top}}]}\Psi|+|\breve{X}\mathcal{P}^{[1,N_{\mathrm{top}}]}\Psi|)\mathcal{P}^{[1,N_{\mathrm{top}}]}\vec{\mathfrak{G}}\|_{L^1(\Sigma_{t'}^u)}\,dt'\\ &+C\sup_{\hat{t}\in[0,t]}\mu_\star^{2M_*-1.8}(\hat{t})\int_{t'=0}^{t'=\hat{t}}\mu_\star^{-3/2}(t')\left\{\int_{s=0}^{s=t'}\|\mathcal{P}^{[1,N_{\mathrm{top}}]}\vec{\mathfrak{G}}\|_{L^2(\Sigma_s^u)}\,ds\right\}^2dt', \end{aligned}\tag{A-52}$$

$$\begin{aligned} H(t,u)\leq{}& C\mathring{\epsilon}^2+\gamma_1F(t,u)+\gamma_2H(t,u)\\ &+C\sup_{\hat{t}\in[0,t]}\mu_\star^{2M_*-3.8}(\hat{t})\int_{t'=0}^{t'=\hat{t}}\|(|L\mathcal{P}^{[1,N_{\mathrm{top}}-1]}\Psi|+|\breve{X}\mathcal{P}^{[1,N_{\mathrm{top}}-1]}\Psi|)\mathcal{P}^{[1,N_{\mathrm{top}}-1]}\vec{\mathfrak{G}}\|_{L^1(\Sigma_{t'}^u)}\,dt'. \end{aligned}\tag{A-53}$$

The main point is the smallness conditions (A-50) on the constants $\alpha_1,\dots,\gamma_2$ allow us to solve the inequalities (A-51)–(A-53) using a reductive approach. More precisely, using that $\gamma_2<\frac{3}{4}$, we absorb the $\gamma_2H(t,u)$ term on the right-hand side of (A-53) back into the left-hand side to isolate $H(t,u)$, at the expense of enlarging C and replacing γ_1 with $4\gamma_1$. We then insert this estimate for $H(t,u)$ into the right-hand side of (A-52) to obtain an estimate for $G(t,u)$, and then insert these estimates for $H(t,u)$ and $G(t,u)$ into the right-hand side of (A-51) to obtain the inequality

$$\begin{aligned} &F(t,u)\\ &\leq C\mathring{\epsilon}^2+\{\alpha_1+4\alpha_2\gamma_1+\alpha_3\beta_1+4\alpha_3\beta_2\gamma_1\}F(t,u)\\ &\quad+C\sup_{\hat{t}\in[0,t]}\mu_\star^{2M_*-1.8}(\hat{t})\int_{t'=0}^{t'=\hat{t}}\|(|L\mathcal{P}^{[1,N_{\mathrm{top}}]}\Psi|+|\breve{X}\mathcal{P}^{[1,N_{\mathrm{top}}]}\Psi|)\mathcal{P}^{[1,N_{\mathrm{top}}]}\vec{\mathfrak{G}}\|_{L^1(\Sigma_{t'}^u)}\,dt'\\ &\quad+C\sup_{\hat{t}\in[0,t]}\mu_\star^{2M_*-1.8}(\hat{t})\int_{t'=0}^{t'=\hat{t}}\mu_\star^{-3/2}(t')\left\{\int_{s=0}^{s=t'}\|\mathcal{P}^{[1,N_{\mathrm{top}}]}\vec{\mathfrak{G}}\|_{L^2(\Sigma_s^u)}\,ds\right\}^2dt'\\ &\quad+C\sup_{\hat{t}\in[0,t]}\mu_\star^{2M_*-3.8}(\hat{t})\int_{t'=0}^{t'=\hat{t}}\|(|L\mathcal{P}^{[1,N_{\mathrm{top}}-1]}\Psi|+|\breve{X}\mathcal{P}^{[1,N_{\mathrm{top}}-1]}\Psi|)\mathcal{P}^{[1,N_{\mathrm{top}}-1]}\vec{\mathfrak{G}}\|_{L^1(\Sigma_{t'}^u)}\,dt'. \end{aligned}\tag{A-54}$$

From the smallness condition $\alpha_1 + 4\alpha_2\gamma_1 + \alpha_3\beta_1 + 4\alpha_3\beta_2\gamma_1 < 1$ featured in (A-50), it follows that we can absorb the terms $\{\alpha_1 + 4\alpha_2\gamma_1 + \alpha_3\beta_1 + 4\alpha_3\beta_2\gamma_1\}F(t,u)$ on the right-hand side of (A-54) back into the left-hand side of (A-54) to isolate $F(t,u)$, at the expense of increasing the constant C. We therefore deduce the inequality

$$\begin{aligned} F(t,u) \lesssim \mathring{\epsilon}^2 &+ \sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-1.8}(\hat{t}) \int_{t'=0}^{t'=\hat{t}} \|(|L\mathcal{P}^{[1,N_{\text{top}}]}\Psi| + |\breve{X}\mathcal{P}^{[1,N_{\text{top}}]}\Psi|)\mathcal{P}^{[1,N_{\text{top}}]}\vec{\mathfrak{G}}\|_{L^1(\Sigma_{t'}^u)}\,dt' \\ &+ \sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-1.8}(\hat{t}) \int_{t'=0}^{t'=\hat{t}} \mu_\star^{-3/2}(t') \left\{ \int_{s=0}^{s=t'} \|\mathcal{P}^{[1,N_{\text{top}}]}\vec{\mathfrak{G}}\|_{L^2(\Sigma_s^u)}\,ds \right\}^2 dt' \\ &+ \sup_{\hat{t}\in[0,t]} \mu_\star^{2M_*-3.8}(\hat{t}) \int_{t'=0}^{t'=\hat{t}} \|(|L\mathcal{P}^{[1,N_{\text{top}}-1]}\Psi| + |\breve{X}\mathcal{P}^{[1,N_{\text{top}}-1]}\Psi|)\mathcal{P}^{[1,N_{\text{top}}-1]}\vec{\mathfrak{G}}\|_{L^1(\Sigma_{t'}^u)}\,dt'. \end{aligned} \quad \text{(A-55)}$$

Then from (A-55) and the arguments given above, we deduce that $G(t,u)$ and $H(t,u)$ are also bounded above by the right-hand side of (A-55) (where we enlarge C if necessary).

Recalling the definitions of F, G, and H in (A-44)–(A-46), we see that (A-55) and the similar bounds for $G(t,u)$ and $H(t,u)$ collectively imply (12-2).

Step 2: the lower orders (proof of (12-3)). To prove the lower-order energy estimates, we start by considering the energy inequality given by the below-top-order estimate from Proposition A.17, i.e., the estimate (A-39), which features the additional term (A-36) compared to [36, (14.6)].

Observe that on the right-hand side of (A-39), except for

$$\int_{t'=0}^{t'=t} \frac{\mathbb{Q}_{[1,N-1]}^{1/2}(t',u)}{\mu_\star^{1/2}(t',u)} \left\{ \int_{s=0}^{s=t'} \frac{\mathbb{Q}_{[1,N]}^{1/2}(s,u)}{\mu_\star^{1/2}(s,u)}\,ds \right\} dt',$$

every other term can be treated directly by Grönwall's inequality (using Proposition 8.11), as in [36]. It thus follows that

$$\begin{aligned} &\sup_{t'\in[0,t]} \max\{\mathbb{Q}_{[1,N-1]}(t',u),\ \mathbb{K}_{[1,N-1]}(t',u)\} \\ &\quad \le C\mathring{\epsilon}^2 + C\int_{t'=0}^{t'=t} \frac{\mathbb{Q}_{[1,N-1]}^{1/2}(t',u)}{\mu_\star^{1/2}(t',u)} \left\{ \int_{s=0}^{s=t'} \frac{\mathbb{Q}_{[1,N]}^{1/2}(s,u)}{\mu_\star^{1/2}(s,u)}\,ds \right\} dt' \\ &\qquad\qquad + C\|(|L\mathcal{P}^{[1,N-1]}\Psi| + |\breve{X}\mathcal{P}^{[1,N-1]}\Psi|)|\mathcal{P}^{[1,N-1]}\vec{\mathfrak{G}}|\|_{L^1(\mathcal{M}_{t,u})}. \end{aligned} \quad \text{(A-56)}$$

To proceed, we analyze the double time-integral term on the right-hand side of (A-56). For any $\varsigma > 0$, we have

$$\begin{aligned} &\int_{t'=0}^{t'=t} \frac{\mathbb{Q}_{[1,N-1]}^{1/2}(t',u)}{\mu_\star^{1/2}(t',u)} \left\{ \int_{s=0}^{s=t'} \frac{\mathbb{Q}_{[1,N]}^{1/2}(s,u)}{\mu_\star^{1/2}(s,u)}\,ds \right\} dt' \\ &\quad \le \Big(\sup_{t'\in[0,t]} \mathbb{Q}_{[1,N-1]}^{1/2}(t')\Big) \times \Big(\sup_{s\in[0,t]} \min\{1, \mu_\star^{M_*-N_{\text{top}}+N-0.9}(s)\}\mathbb{Q}_{[1,N]}^{1/2}(s)\Big) \\ &\qquad\qquad \times \int_{t'=0}^{t'=t} \frac{1}{\mu_\star^{1/2}(t')} \left\{ \int_{s=0}^{s=t'} \frac{\max\{1, \mu_\star^{-M_*+N_{\text{top}}-N+0.9}(s)\}}{\mu_\star^{1/2}(s)}\,ds \right\} dt' \\ &\quad \le \varsigma \sup_{t'\in[0,t]} \mathbb{Q}_{[1,N-1]}(t') \\ &\qquad + C\varsigma^{-1} \max\{1, \mu_\star^{-2M_*+2N_{\text{top}}-2N+3.8}(t)\}\Big(\sup_{s\in[0,t]} \min\{1, \mu_\star^{2M_*-2N_{\text{top}}+2N-1.8}(s)\}\mathbb{Q}_{[1,N]}(s)\Big), \end{aligned} \quad \text{(A-57)}$$

where to obtain the last inequality, we have used Young's inequality and the following estimate, which follows from Proposition 8.11:

$$\int_{t'=0}^{t'=t} \frac{1}{\mu_\star^{1/2}(t')} \left\{ \int_{s=0}^{s=t'} \frac{\max\{1, \mu_\star^{-M_*+N_{\mathrm{top}}-N+0.9}(s)\}}{\mu_\star^{1/2}(s)}\, ds \right\} dt' \lesssim \int_{t'=0}^{t'=t} \frac{\max\{1, \mu_\star^{-M+N_{\mathrm{top}}-N+1.4}(t')\}}{\mu_\star^{1/2}(t')}\, dt' \lesssim \max\{1, \mu_\star^{-M_*+N_{\mathrm{top}}-N+1.9}(t)\}.$$

Inserting (A-57) into (A-56) and fixing $\varsigma > 0$ to be sufficiently small, we can absorb the term $C\varsigma(\sup_{t'\in[0,t]} \mathbb{Q}_{[1,N-1]}(t'))$ back into the left-hand side of (A-56). Thus, for this fixed value of ς, we obtain

$$\begin{aligned} &\sup_{t'\in[0,t]} \max\{\mathbb{Q}_{[1,N-1]}(t',u),\ \mathbb{K}_{[1,N-1]}(t',u)\} \\ &\quad\lesssim \mathring{\epsilon}^2 + \max\{1, \mu_\star^{-2M+2N_{\mathrm{top}}-2N+3.8}(t)\}\Big(\sup_{s\in[0,t]} \min\{1, \mu_\star^{2M_*-2N_{\mathrm{top}}+2N-1.8}(s)\}\mathbb{Q}_{[1,N]}(s)\Big) \\ &\qquad\qquad + \|(|L\mathcal{P}^{[1,N-1]}\Psi| + |\breve{X}\mathcal{P}^{[1,N-1]}\Psi|)|\mathcal{P}^{[1,N-1]}\vec{\mathfrak{G}}|\|_{L^1(\mathcal{M}_{t,u})}. \end{aligned}$$

After changing the index N to $N+1$, we conclude the estimate (12-3). □

References

[1] L. Abbrescia and J. Speck, "Remarkable localized integral identities for 3D compressible Euler flow and the double-null framework", preprint, 2020. arXiv 2003.02815

[2] S. Alinhac, "Temps de vie des solutions régulières des équations d'Euler compressibles axisymétriques en dimension deux", *Invent. Math.* **111**:3 (1993), 627–670. MR Zbl

[3] S. Alinhac, "Blowup of small data solutions for a class of quasilinear wave equations in two space dimensions, II", *Acta Math.* **182**:1 (1999), 1–23. MR Zbl

[4] S. Alinhac, "Blowup of small data solutions for a quasilinear wave equation in two space dimensions", *Ann. of Math.* (2) **149**:1 (1999), 97–127. MR Zbl

[5] S. Alinhac, "A minicourse on global existence and blowup of classical solutions to multidimensional quasilinear wave equations", pp. art. id. 1 in *Journées "Équations aux Dérivées Partielles"* (Forges-les-Eaux, France, 2002), Univ. Nantes, 2002. MR

[6] A. Bressan, G. Crasta, and B. Piccoli, *Well-posedness of the Cauchy problem for n × n systems of conservation laws*, Mem. Amer. Math. Soc. **694**, Amer. Math. Soc., Providence, RI, 2000. MR Zbl

[7] T. Buckmaster and S. Iyer, "Formation of unstable shocks for 2D isentropic compressible Euler", *Comm. Math. Phys.* **389**:1 (2022), 197–271. MR Zbl

[8] T. Buckmaster, T. D. Drivas, S. Shkoller, and V. Vicol, "Simultaneous development of shocks and cusps for 2D Euler with azimuthal symmetry from smooth data", *Ann. PDE* **8**:2 (2022), art. id. 26. MR Zbl

[9] T. Buckmaster, S. Shkoller, and V. Vicol, "Formation of shocks for 2D isentropic compressible Euler", *Comm. Pure Appl. Math.* **75**:9 (2022), 2069–2120. MR Zbl

[10] T. Buckmaster, S. Shkoller, and V. Vicol, "Formation of point shocks for 3D compressible Euler", *Comm. Pure Appl. Math.* **76**:9 (2023), 2073–2191. MR Zbl

[11] T. Buckmaster, S. Shkoller, and V. Vicol, "Shock formation and vorticity creation for 3D Euler", *Comm. Pure Appl. Math.* **76**:9 (2023), 1965–2072. MR Zbl

[12] G.-Q. Chen and D. Wang, "The Cauchy problem for the Euler equations for compressible fluids", pp. 421–543 in *Handbook of mathematical fluid dynamics*, *I*, edited by S. Friedlander and D. Serre, North-Holland, Amsterdam, 2002. MR Zbl

[13] K. R. Chickering, R. C. Moreno-Vasquez, and G. Pandya, "Asymptotically self-similar shock formation for 1D fractal Burgers' equation", *SIAM J. Math. Anal.* **55**:6 (2023), 7328–7360. MR Zbl

[14] D. Christodoulou, "The Euler equations of compressible fluid flow", *Bull. Amer. Math. Soc.* (*N.S.*) **44**:4 (2007), 581–602. MR Zbl

[15] D. Christodoulou, *The formation of shocks in* 3*-dimensional fluids*, Eur. Math. Soc., Zürich, 2007. MR Zbl

[16] D. Christodoulou, *The shock development problem*, Eur. Math. Soc., Zürich, 2019. MR Zbl

[17] D. Christodoulou and S. Klainerman, *The global nonlinear stability of the Minkowski space*, Princeton Math. Ser. **41**, Princeton Univ. Press, 1993. MR Zbl

[18] D. Christodoulou and A. Lisibach, "Shock development in spherical symmetry", *Ann. PDE* **2**:1 (2016), art. id. 3. MR Zbl

[19] D. Christodoulou and S. Miao, *Compressible flow and Euler's equations*, Surv. Modern Math. **9**, Int. Press, Somerville, MA, 2014. MR Zbl

[20] D. Christodoulou and D. R. Perez, "On the formation of shocks of electromagnetic plane waves in non-linear crystals", *J. Math. Phys.* **57**:8 (2016), art. id. 081506. MR Zbl

[21] C. Collot, T.-E. Ghoul, and N. Masmoudi, "Singularities and unsteady separation for the inviscid two-dimensional Prandtl system", *Arch. Ration. Mech. Anal.* **240**:3 (2021), 1349–1430. MR Zbl

[22] C. Collot, T.-E. Ghoul, S. Ibrahim, and N. Masmoudi, "On singularity formation for the two-dimensional unsteady Prandtl system around the axis", *J. Eur. Math. Soc.* **24**:11 (2022), 3703–3800. MR Zbl

[23] C. Collot, T.-E. Ghoul, and N. Masmoudi, "Singularity formation for Burgers' equation with transverse viscosity", *Ann. Sci. École Norm. Sup.* (4) **55**:4 (2022), 1047–1133. MR Zbl

[24] C. M. Dafermos, *Hyperbolic conservation laws in continuum physics*, Grundlehren der Math. Wissenschaften **325**, Springer, 2000. MR Zbl

[25] M. M. Disconzi and J. Speck, "The relativistic Euler equations: remarkable null structures and regularity properties", *Ann. Henri Poincaré* **20**:7 (2019), 2173–2270. MR Zbl

[26] M. M. Disconzi, C. Luo, G. Mazzone, and J. Speck, "Rough sound waves in 3D compressible Euler flow with vorticity", *Selecta Math.* (*N.S.*) **28**:2 (2022), art. id. 41. MR Zbl

[27] J. Eggers and M. A. Fontelos, "The role of self-similarity in singularities of partial differential equations", *Nonlinearity* **22**:1 (2009), R1–R44. MR Zbl

[28] J. Glimm, "Solutions in the large for nonlinear hyperbolic systems of equations", *Comm. Pure Appl. Math.* **18** (1965), 697–715. MR Zbl

[29] G. Guderley, "Starke kugelige und zylindrische Verdichtungsstösse in der Nähe des Kugelmittelpunktes bzw. der Zylinderachse", *Luftfahrtforschung* **19** (1942), 302–311. MR Zbl

[30] G. Holzegel, S. Klainerman, J. Speck, and W. W.-Y. Wong, "Small-data shock formation in solutions to 3D quasilinear wave equations: an overview", *J. Hyperbolic Differ. Equ.* **13**:1 (2016), 1–105. MR Zbl

[31] F. John, "Formation of singularities in one-dimensional nonlinear wave propagation", *Comm. Pure Appl. Math.* **27** (1974), 377–405. MR Zbl

[32] S. Klainerman, "Long time behaviour of solutions to nonlinear wave equations", pp. 1209–1215 in *Proceedings of the International Congress of Mathematicians*, *II* (Warsaw, 1983), edited by Z. Ciesielski and C. Olech, PWN, Warsaw, 1984. MR Zbl

[33] S. Klainerman and I. Rodnianski, "Improved local well-posedness for quasilinear wave equations in dimension three", *Duke Math. J.* **117**:1 (2003), 1–124. MR Zbl

[34] P. D. Lax, "Development of singularities of solutions of nonlinear hyperbolic partial differential equations", *J. Math. Phys.* **5** (1964), 611–613. MR Zbl

[35] T. P. Liu, "Development of singularities in the nonlinear waves for quasilinear hyperbolic partial differential equations", *J. Differential Equations* **33**:1 (1979), 92–111. MR Zbl

[36] J. Luk and J. Speck, "Shock formation in solutions to the 2D compressible Euler equations in the presence of non-zero vorticity", *Invent. Math.* **214**:1 (2018), 1–169. MR Zbl

[37] J. Luk and J. Speck, "The hidden null structure of the compressible Euler equations and a prelude to applications", *J. Hyperbolic Differ. Equ.* **17**:1 (2020), 1–60. MR Zbl

[38] F. Merle, P. Raphaël, I. Rodnianski, and J. Szeftel, "On blow up for the energy super critical defocusing nonlinear Schrödinger equations", *Invent. Math.* **227**:1 (2022), 247–413. MR Zbl

[39] F. Merle, P. Raphaël, I. Rodnianski, and J. Szeftel, "On the implosion of a compressible fluid, I: Smooth self-similar inviscid profiles", *Ann. of Math.* (2) **196**:2 (2022), 567–778. MR Zbl

[40] F. Merle, P. Raphaël, I. Rodnianski, and J. Szeftel, "On the implosion of a compressible fluid, II: Singularity formation", *Ann. of Math.* (2) **196**:2 (2022), 779–889. MR Zbl

[41] S. Miao, "On the formation of shock for quasilinear wave equations with weak intensity pulse", *Ann. PDE* **4**:1 (2018), art. id. 10. MR Zbl

[42] S. Miao and P. Yu, "On the formation of shocks for quasilinear wave equations", *Invent. Math.* **207**:2 (2017), 697–831. MR Zbl

[43] S.-J. Oh and F. Pasqualotto, "Gradient blow-up for dispersive and dissipative perturbations of the Burgers equation", preprint, 2021. arXiv 2107.07172

[44] J. Rauch, "BV estimates fail for most quasilinear hyperbolic systems in dimensions greater than one", *Comm. Math. Phys.* **106**:3 (1986), 481–484. MR Zbl

[45] B. Riemann, "Über die Fortpflanzung ebener Luftwellen von endlicher Schwingungsweite", *Abh. König. Gesellschaft Wissen. Göttingen* **8** (1860), 43–66.

[46] L. I. Sedov, *Similarity and dimensional methods in mechanics*, Mir, Moscow, 1982. MR Zbl

[47] T. C. Sideris, "Formation of singularities in three-dimensional compressible fluids", *Comm. Math. Phys.* **101**:4 (1985), 475–485. MR Zbl

[48] J. Speck, *Shock formation in small-data solutions to* 3*D quasilinear wave equations*, Math. Surv. Monogr. **214**, Amer. Math. Soc., Providence, RI, 2016. MR Zbl

[49] J. Speck, "Shock formation for 2D quasilinear wave systems featuring multiple speeds: blowup for the fastest wave, with non-trivial interactions up to the singularity", *Ann. PDE* **4**:1 (2018), art. id. 6. MR Zbl

[50] J. Speck, "A new formulation of the 3D compressible Euler equations with dynamic entropy: remarkable null structures and regularity properties", *Arch. Ration. Mech. Anal.* **234**:3 (2019), 1223–1279. MR Zbl

[51] J. Speck, "Stable ODE-type blowup for some quasilinear wave equations with derivative-quadratic nonlinearities", *Anal. PDE* **13**:1 (2020), 93–146. MR Zbl

[52] J. Speck, G. Holzegel, J. Luk, and W. Wong, "Stable shock formation for nearly simple outgoing plane symmetric waves", *Ann. PDE* **2**:2 (2016), art. id. 10. MR Zbl

[53] Q. Wang, "Rough solutions of the 3-D compressible Euler equations", *Ann. of Math.* (2) **195**:2 (2022), 509–654. MR Zbl

[54] R. Yang, "Shock formation of the Burgers–Hilbert equation", *SIAM J. Math. Anal.* **53**:5 (2021), 5756–5802. MR Zbl

[55] H. Yin, "Formation and construction of a shock wave for 3-D compressible Euler equations with the spherical initial data", *Nagoya Math. J.* **175** (2004), 125–164. MR Zbl

[56] S. Yu, "Rough solutions of the relativistic Euler equations", preprint, 2022. arXiv 2203.11746

Received 5 Sep 2021. Revised 30 Jun 2022. Accepted 3 Aug 2022.

JONATHAN LUK: `jluk@stanford.edu`
Stanford University, Stanford, CA, United States

JARED SPECK: `jared.speck@vanderbilt.edu`
Vanderbilt University, Nashville, TN, United States

ANALYSIS AND PDE
Vol. 17 (2024), No. 3, pp. 943–979
DOI: 10.2140/apde.2024.17.943

FAMILIES OF FUNCTIONALS REPRESENTING SOBOLEV NORMS

HAÏM BREZIS, ANDREAS SEEGER, JEAN VAN SCHAFTINGEN AND PO-LAM YUNG

We obtain new characterizations of the Sobolev spaces $\dot{W}^{1,p}(\mathbb{R}^N)$ and the bounded variation space $\dot{\mathrm{BV}}(\mathbb{R}^N)$. The characterizations are in terms of the functionals $\nu_\gamma(E_{\lambda,\gamma/p}[u])$, where

$$E_{\lambda,\gamma/p}[u] = \left\{ (x, y) \in \mathbb{R}^N \times \mathbb{R}^N : x \neq y,\ \frac{|u(x) - u(y)|}{|x - y|^{1+\gamma/p}} > \lambda \right\}$$

and the measure ν_γ is given by $\mathrm{d}\nu_\gamma(x, y) = |x - y|^{\gamma - N}\,\mathrm{d}x\,\mathrm{d}y$. We provide characterizations which involve the $L^{p,\infty}$-quasinorms $\sup_{\lambda>0} \lambda \nu_\gamma(E_{\lambda,\gamma/p}[u])^{1/p}$ and also exact formulas via corresponding limit functionals, with the limit for $\lambda \to \infty$ when $\gamma > 0$ and the limit for $\lambda \to 0^+$ when $\gamma < 0$. The results unify and substantially extend previous work by Nguyen and by Brezis, Van Schaftingen and Yung. For $p > 1$ the characterizations hold for all $\gamma \neq 0$. For $p = 1$ the upper bounds for the $L^{1,\infty}$ quasinorms fail in the range $\gamma \in [-1, 0)$; moreover, in this case the limit functionals represent the L^1 norm of the gradient for C_c^∞-functions but not for generic $\dot{W}^{1,1}$-functions. For this situation we provide new counterexamples which are built on self-similar sets of dimension $\gamma + 1$. For $\gamma = 0$ the characterizations of Sobolev spaces fail; however, we obtain a new formula for the Lipschitz norm via the expressions $\nu_0(E_{\lambda,0}[u])$.

1. Introduction

We are concerned with various ways in which we can recover the Sobolev seminorm $\|\nabla u\|_{L^p(\mathbb{R}^N)}$ via positive nonconvex functionals involving differences $u(x) - u(y)$.

We begin by mentioning two relevant results already in the literature. A theorem of H.-M. Nguyen [2006] (see also [Brezis and Nguyen 2018; 2020]) states that, for $1 < p < \infty$ and u in the inhomogeneous Sobolev space $W^{1,p}(\mathbb{R}^N)$,

$$\lim_{\lambda \searrow 0} \lambda^p \iint_{|u(x)-u(y)|>\lambda} |x - y|^{-p-N}\,\mathrm{d}x\,\mathrm{d}y = \frac{\kappa(p, N)}{p} \|\nabla u\|^p_{L^p(\mathbb{R}^N)}, \tag{1-1}$$

with

$$\kappa(p, N) := \int_{\mathbb{S}^{N-1}} |e \cdot \omega|^p\,\mathrm{d}\omega = \frac{2\Gamma((p+1)/2)\pi^{(N-1)/2}}{\Gamma((N+p)/2)}, \tag{1-2}$$

and e is any unit vector in $\mathbb{R}^N$. As shown in [Brezis and Nguyen 2018], (1-1) still holds for all $u \in C_c^1(\mathbb{R}^N)$ when $p = 1$ but fails for general $u \in W^{1,1}(\mathbb{R}^N)$. The limit formula (1-1) may be compared to a theorem of [Brezis et al. 2021b], which states that, for all $u \in C_c^\infty(\mathbb{R}^N)$ and $1 \le p < \infty$, one has

$$\lim_{\lambda \to \infty} \lambda^p \mathcal{L}^{2N}(\{(x, y) \in \mathbb{R}^N \times \mathbb{R}^N : |u(x) - u(y)| > \lambda |x - y|^{1+N/p}\}) = \frac{\kappa(p, N)}{N} \|\nabla u\|^p_{L^p(\mathbb{R}^N)}, \tag{1-3}$$

MSC2020: primary 26D10, 26A33; secondary 35A23, 42B25, 42B35, 46E30, 46E35.
Keywords: Sobolev norms, nonconvex functionals, nonlocal functionals, Marcinkiewicz spaces, Cantor sets and functions.

where $\mathcal{L}^{2N}$ denotes the Lebesgue measure on $\mathbb{R}^N \times \mathbb{R}^N$. Our first result, namely Theorem 1.1 below, provides an extension of (1-1) and (1-3) that unifies the two statements. Before we state the theorem, we introduce some notation that will be used throughout the paper.

First, for Lebesgue measurable subsets E of $\mathbb{R}^{2N} = \mathbb{R}^N \times \mathbb{R}^N$ and $\gamma \in \mathbb{R}$, we define

$$\nu_\gamma(E) := \iint_{\substack{(x,y)\in E\\ x\neq y}} |x-y|^{\gamma-N}\,\mathrm{d}x\,\mathrm{d}y. \tag{1-4}$$

In particular, when $\gamma = N$, ν_N is just the Lebesgue measure on $\mathbb{R}^{2N}$. If u is a measurable function on $\mathbb{R}^N$ and $b \in \mathbb{R}$, we define, for $(x, y) \in \mathbb{R}^N \times \mathbb{R}^N$ with $x \neq y$, a difference quotient

$$Q_b u(x,y) := \frac{u(x)-u(y)}{|x-y|^{1+b}}; \tag{1-5}$$

moreover, we define, for $\lambda > 0$, the superlevel set of $Q_b u$ at height λ by

$$E_{\lambda,b}[u] := \{(x,y) \in \mathbb{R}^N \times \mathbb{R}^N : x \neq y,\ |Q_b u(x,y)| > \lambda\}. \tag{1-6}$$

We will denote by $\dot{W}^{1,p}(\mathbb{R}^N)$, $p \geq 1$, the homogeneous Sobolev space, i.e., the space of $L^1_{\mathrm{loc}}(\mathbb{R}^N)$ functions for which the distributional gradient ∇u belongs to $L^p(\mathbb{R}^N)$, with the seminorm $\|u\|_{\dot{W}^{1,p}} := \|\nabla u\|_{L^p(\mathbb{R}^N)}$. The inhomogeneous Sobolev space $W^{1,p}$ is the subspace of $\dot{W}^{1,p}$-functions u for which $u \in L^p$, and we set $\|u\|_{W^{1,p}} := \|u\|_{L^p} + \|\nabla u\|_{L^p}$. For $p = 1$ we will also consider the space $\dot{\mathrm{BV}}(\mathbb{R}^N)$ of functions of bounded variations, i.e., locally integrable functions u for which the gradient $\nabla u \in \mathcal{M}$ belongs to the space $\mathcal{M}$ of $\mathbb{R}^N$-valued bounded Borel measures and we put $\|u\|_{\dot{\mathrm{BV}}} := \|\nabla u\|_{\mathcal{M}}$; furthermore, let $\mathrm{BV} := \dot{\mathrm{BV}} \cap L^1$. In the dual formulation, with C^1_c denoting the space of C^1 functions with compact support,

$$\|u\|_{\dot{\mathrm{BV}}} := \sup\left\{ \left| \int_{\mathbb{R}^N} u \operatorname{div}(\phi) \right| : \phi \in C^1_c(\mathbb{R}^N, \mathbb{R}^N),\ \|\phi\|_\infty \leq 1 \right\}.$$

For general background material on Sobolev spaces, see [Brezis 2011; Stein 1970].

Theorem 1.1. *Suppose* $N \geq 1$, $1 \leq p < \infty$, $\gamma \in \mathbb{R} \setminus \{0\}$.

(a) *If* $\gamma > 0$, *then, for all* $u \in \dot{W}^{1,p}(\mathbb{R}^N)$,

$$\lim_{\lambda\to+\infty} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) = \frac{\kappa(p,N)}{|\gamma|} \|\nabla u\|^p_{L^p(\mathbb{R}^N)}. \tag{1-7}$$

(b) *If either* $\gamma < 0$, $p > 1$ *or* $\gamma < -1$, $p = 1$ *then, for all* $u \in \dot{W}^{1,p}(\mathbb{R}^N)$,

$$\lim_{\lambda\searrow 0} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) = \frac{\kappa(p,N)}{|\gamma|} \|\nabla u\|^p_{L^p(\mathbb{R}^N)}. \tag{1-8}$$

(c) *If* $p = 1$ *and* $-1 \leq \gamma < 0$ *then* (1-8) *remains true for all* $u \in C^1_c(\mathbb{R}^N)$ *but fails for generic* $u \in \dot{W}^{1,1}(\mathbb{R}^N)$. *However, we still have, for all* $u \in \dot{W}^{1,1}(\mathbb{R}^N)$,

$$\liminf_{\lambda\searrow 0} \lambda \nu_\gamma(E_{\lambda,\gamma}[u]) \geq \frac{\kappa(1,N)}{|\gamma|} \|\nabla u\|_{L^1(\mathbb{R}^N)}. \tag{1-9}$$

Formula (1-1) is the special case of (1-8) with $\gamma = -p$, and formula (1-3) is the special case of (1-7) with $\gamma = N$. Note that our result concerns functions in the homogeneous Sobolev space $\dot{W}^{1,p}$; we do not require u to be in L^p.

Remarks. (i) The reader will note the resemblance of (1-8) and (1-7) and may wonder why in (1-8), for $\gamma < 0$, one is concerned with the limit as $\lambda \searrow 0$ and in (1-7), for $\gamma > 0$, one takes the limit as $\lambda \to \infty$. In the proofs of these formulas one relates limits involving $\lambda \nu_\gamma (E_{\lambda,\gamma/p}[u])^{1/p}$ to (the absolute value of) limits of directional difference quotients $\delta^{-1}(u(x+\delta\theta) - u(x))$ with increment $\delta = \lambda^{-p/\gamma}$, and in order to recover the directional derivative $\langle \theta, \nabla u(x)\rangle$ we need to let $\delta \to 0$, which suggests that we need to take $\lambda \to \infty$ or $\lambda \searrow 0$ depending on the sign of γ. For the calculations see the proofs of Lemmas 3.2 and 3.3 below.

(ii) The failure of (1-8) for $p = 1$, $\gamma \in [-1, 0)$ and $u \in \dot{W}^{1,1}(\mathbb{R}^N)$ is generic in the sense of Baire category. It may happen that $\lim_{\lambda \searrow 0} \lambda \nu_\gamma (E_{\lambda,\gamma}[u]) = \infty$. This phenomenon was originally revealed when $\gamma = -1$ by A. Ponce and is presented in [Nguyen 2006]; see also [Brezis and Nguyen 2018, Pathology 1]. For stronger statements and more information, see Theorem 1.8. For $\gamma \in (-1, 0)$ we provide new examples based on self-similarity considerations. For discussion of failure in the case $\gamma = 0$, see Theorem 1.5 below. The special case of (1-9) for $\gamma = -1$ was already established in [Brezis and Nguyen 2018, Proposition 1].

When $p = 1$ we can also consider what happens if one allows functions in $\dot{\mathrm{BV}}(\mathbb{R}^N)$ in (1-7) and (1-8). For $\gamma = N$ in particular Poliakovsky [2022] asked whether the limit formulas remain valid in this generality (with $\|\nabla u\|_{L^1}$ replaced by $\|\nabla u\|_{\mathcal{M}}$). We provide a negative answer:

Proposition 1.2. (i) *The analogues of the limiting formulas* (1-7) *for* $\gamma > 0$, $p = 1$ *and* (1-8) *for* $\gamma < 0$, $p = 1$, *with* $\|\nabla u\|_{\mathcal{M}}$ *on the right-hand side, fail for suitable* $u \in \dot{\mathrm{BV}}$.

(ii) *Specifically, let* $\Omega \subset \mathbb{R}^N$ *be a bounded convex domain with smooth boundary and let u be the characteristic function of* Ω. *The limits* $\lim_{\lambda\to\infty} \lambda \nu_\gamma (E_{\lambda,\gamma}[u])$ *for* $\gamma > 0$ *and* $\lim_{\lambda \to 0+} \lambda \nu_\gamma (E_{\lambda,\gamma}[u])$ *for* $\gamma < -1$ *exist, but they are not equal to* $|\gamma|^{-1}\kappa(1, N)\|\nabla u\|_{\mathcal{M}}$.

For a more detailed discussion we refer to Section 3F. See also Section 7B for a discussion about some related open problems.

Motivated by [Brezis et al. 2021b], we will also be interested in what happens to the larger quantity obtained by replacing the limits on the left-hand sides of (1-7) and (1-8) by $\sup_{\lambda>0}$. This will be formulated in terms of the Marcinkiewicz space $L^{p,\infty}(\mathbb{R}^{2N}, \nu_\gamma)$ (a.k.a. weak-type L^p) defined by the condition

$$[F]^p_{L^{p,\infty}(\mathbb{R}^{2N},\nu_\gamma)} := \sup_{\lambda>0} \lambda^p \nu_\gamma \big(\{(x, y) \in \mathbb{R}^N \times \mathbb{R}^N : |F(x, y)| > \lambda\}\big) < \infty. \tag{1-10}$$

As an immediate consequence of Theorem 1.1 we have, for $N \geq 1$, $1 \leq p < \infty$, $\gamma \neq 0$ and all $u \in C_c^\infty(\mathbb{R}^N)$,

$$[Q_{\gamma/p} u]^p_{L^{p,\infty}(\mathbb{R}^{2N},\nu_\gamma)} \geq C(N, p, \gamma)\|\nabla u\|^p_{L^p(\mathbb{R}^N)}, \tag{1-11}$$

where $C(N, p, \gamma)$ is a positive constant depending only on N, p and γ. Moreover, the same conclusion holds for all $u \in \dot{W}^{1,p}(\mathbb{R}^N)$ when $p > 1$, with any $\gamma \neq 0$, and when $p = 1$, with any $\gamma \notin [-1, 0]$. We shall show that the conditions in the last statement can in fact be relaxed; see the inequalities (1-14) and

(1-16) below. In addition we have the important upper bounds for $Q_{\gamma/p}u$, extending the case $\gamma = N$ already dealt with in [Brezis et al. 2021b] for $u \in C_c^\infty(\mathbb{R}^N)$. The result in [Brezis et al. 2021b] states that, for every $N \geq 1$, there exists a constant $C(N)$ such that

$$[Q_{N/p}u]^p_{L^{p,\infty}(\mathbb{R}^{2N},\nu_N)} \leq C(N)\|\nabla u\|^p_{L^p(\mathbb{R}^N)} \tag{1-12}$$

for all $u \in C_c^\infty(\mathbb{R}^N)$ and all $1 \leq p < \infty$. In light of Theorem 1.1, it is natural to ask whether one can replace the limits on the left-hand sides of (1-7) and (1-8) by $\sup_{\lambda>0}$ and still obtain a quantity that is comparable to $\|\nabla u\|^p_{L^p(\mathbb{R}^N)}$. As suggested by Theorem 1.1 the answer to our question is sensitive to the values of γ and p.

Theorem 1.3. *Suppose that $N \geq 1$, $1 < p < \infty$ and $\gamma \in \mathbb{R}$. Then the following hold*:

(i) *The inequality*

$$[Q_{\gamma/p}u]_{L^{p,\infty}(\mathbb{R}^{2N},\nu_\gamma)} \leq C(N,p,\gamma)\|\nabla u\|_{L^p(\mathbb{R}^N)} \tag{1-13}$$

holds for all $u \in C_c^\infty(\mathbb{R}^N)$ if and only if $\gamma \neq 0$. In this case (1-13) *extends to all $u \in \dot W^{1,p}(\mathbb{R}^N)$.*

(ii) *Suppose that $u \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$ and $Q_{\gamma/p}u \in L^{p,\infty}(\mathbb{R}^{2N},\nu_\gamma)$. Then $u \in \dot W^{1,p}(\mathbb{R}^N)$ and we have the inequality*

$$\|\nabla u\|_{L^p(\mathbb{R}^N)} \leq C_{N,p,\gamma}[Q_{\gamma/p}u]_{L^{p,\infty}(\mathbb{R}^{2N},\nu_\gamma)}. \tag{1-14}$$

There is a new phenomenon for $p = 1$, namely the upper bounds for $Q_\gamma u$ only hold for the more restrictive range $\gamma \in (-\infty,-1) \cup (0,\infty)$. Here it is also natural to replace $\dot W^{1,1}$ with $\dot{\mathrm{BV}}$.

Theorem 1.4. *Suppose that $N \geq 1$ and $\gamma \in \mathbb{R}$. Then the following hold*:

(i) *The inequality*

$$[Q_\gamma u]_{L^{1,\infty}(\mathbb{R}^{2N},\nu_\gamma)} \leq C(N,\gamma)\|\nabla u\|_{L^1(\mathbb{R}^N)} \tag{1-15}$$

holds for all $u \in C_c^\infty(\mathbb{R}^N)$ if and only if $\gamma \notin [-1,0]$. In this case (1-15) *extends to all $u \in \dot W^{1,1}(\mathbb{R}^N)$, and, if $\|\nabla u\|_{L^1(\mathbb{R}^N)}$ is replaced by $\|\nabla u\|_{\mathcal{M}}$, to all $u \in \dot{\mathrm{BV}}(\mathbb{R}^N)$.*

(ii) *Suppose that $u \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$ and $Q_\gamma u \in L^{1,\infty}(\mathbb{R}^{2N},\nu_\gamma)$. Then $u \in \dot{\mathrm{BV}}(\mathbb{R}^N)$ and we have the inequality*

$$\|\nabla u\|_{\mathcal{M}} \leq C_{N,\gamma}[Q_\gamma u]_{L^{1,\infty}(\mathbb{R}^{2N},\nu_\gamma)}. \tag{1-16}$$

We note that the quantitative bounds (1-13) and (1-15) in Theorems 1.3 and 1.4 are crucial tools for establishing the limiting relations for all $\dot W^{1,p}$ functions in Theorem 1.1. Note that there is no restriction on γ in (1-14) and (1-16). The constants in the inequalities will be quantified further later in the paper. In particular, $C(N,p,\gamma)$ in (1-13) remains bounded as $p \searrow 1$ only in the range $\gamma \in (0,\infty) \cup (-\infty,-1)$ (see Theorem 2.2 and Proposition 6.1).

Historical comments. Some special cases of the above quantitative estimates have been known. Estimate (1-13) for $\gamma = -p$ and $1 < p < \infty$ was discovered independently by H.-M. Nguyen [2006], and by A. Ponce and J. Van Schaftingen (unpublished communication to H. Brezis and H.-M. Nguyen), both relying on the Hardy–Littlewood maximal inequality. A. Poliakovsky [2022] recently proved generalizations of results in [Brezis et al. 2021b] to Sobolev spaces on domains; moreover, he obtained Theorems 1.3 and 1.4 in the special case $\gamma = N$ under the additional assumption that $u \in L^p$. Other far-reaching generalizations to one-parameter families of operators were obtained by Ó. Domínguez and M. Milman [2022].

The case $\gamma = 0$. We shall now return to the necessity of the assumption $\gamma \notin [-1, 0]$ in parts of Theorems 1.1, 1.3 and 1.4. When $\gamma = 0$, the bounds for $[Q_{\gamma/p}u]_{L^{p,\infty}(\mathbb{R}^{2N},\nu_\gamma)}$ fail in a striking way. We begin by formulating a result illustrating this failure, which also gives a characterization of the seminorm in the Lipschitz space $\dot{W}^{1,\infty}$.

Theorem 1.5. *Suppose* $N \geq 1$, *u is locally integrable on* $\mathbb{R}^N$ *and* $\nabla u \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$. *Then*

$$\|\nabla u\|_{L^\infty(\mathbb{R}^N)} = \inf\{\lambda > 0 : \nu_0(E_{\lambda,0}[u]) < \infty\}. \tag{1-17}$$

Indeed in Proposition 5.1 we shall prove the stronger statement that $\nu_0(E_{\lambda,0}[u]) = 0$ for $\lambda > \|\nabla u\|_\infty$, and $\nu_0(E_{\lambda,0}[u]) = \infty$ for $\lambda < \|\nabla u\|_\infty$. As an immediate consequence of Theorem 1.5 we get:

Corollary 1.6. *Let u be locally integrable on* $\mathbb{R}^N$. *If* $\nabla u \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$ *and if* $\nu_0(E_{\lambda,0}[u])$ *is finite for all* $\lambda > 0$, *then u is almost everywhere equal to a constant function.*

In view of other known results [Brezis 2002; Brezis et al. 2021a] on how to recognize constant functions, a natural question arises whether the hypothesis on the local integrability of ∇u in the corollary could be relaxed; one can ask whether the constancy conclusion holds for all locally integrable functions satisfying $\nu_0(E_{\lambda,0}[u]) < \infty$ for all $\lambda > 0$. However, the following example shows that such an extension fails (for details, see Lemma 5.2).

Example 1.7. Let $\Omega \subset \mathbb{R}^N$ be a bounded Lipschitz domain and let u be the characteristic function of Ω. Then $u \in \mathrm{BV}(\mathbb{R}^N) \setminus \dot{W}^{1,1}(\mathbb{R}^N)$ and $\sup_{\lambda>0} \lambda\, \nu_0(E_{\lambda,0}[u]) < \infty$.

More on counterexamples. We now make more explicit the exclusion of the parameters $\gamma \in [-1, 0)$ in part (c) of Theorem 1.1 and in (1-15). We shall show in Section 6B that for $\gamma \in (-1, 0)$ these negative results can be related to self-similar Cantor subsets of $\mathbb{R}$, of dimension $1 + \gamma$.

Theorem 1.8. *Suppose* $N \geq 1$. *Then the following hold*:

(i) *Let* $-1 \leq \gamma < 0$. *There exists a* C^∞ *function* $u \in \dot{W}^{1,1}(\mathbb{R}^N)$, *rapidly decreasing as* $|x| \to \infty$ *and such that*

$$\lim_{\lambda \searrow 0} \lambda \nu_\gamma(E_{\lambda,\gamma}[u]) = \infty. \tag{1-18}$$

(ii) *Let* $-1 \leq \gamma < 0$. *There exists a compactly supported* $u \in W^{1,1}(\mathbb{R}^N)$ *for which* (1-18) *holds. The set*

$$\Big\{u \in W^{1,1}(\mathbb{R}^N) : \limsup_{\lambda \searrow 0} \lambda \nu_\gamma(E_{\lambda,\gamma}[u]) < \infty\Big\}$$

is meager in $W^{1,1}(\mathbb{R}^N)$, *i.e., of first category in the sense of Baire.*

(iii) *Let* $-1 \leq \gamma < 0$, $N \geq 2$ *or* $-1 < \gamma < 0$, $N = 1$. *There exists a compactly supported* $u \in W^{1,1}(\mathbb{R}^N)$ *such that* $\nu_\gamma(E_{\lambda,\gamma}[u]) = \infty$ *for all* $\lambda > 0$; *moreover, the set*

$$\{u \in W^{1,1}(\mathbb{R}^N) : \nu_\gamma(E_{\lambda,\gamma}[u]) < \infty \text{ for some } \lambda \in (0, \infty)\}$$

is meager in $W^{1,1}(\mathbb{R}^N)$.

The case $N = 1 = -\gamma$ plays a special role and is excluded in the strongest statement (iii) since for all compactly supported $u \in \dot{W}^{1,1}(\mathbb{R})$ one has $\nu_{-1}(E_{\lambda,-1}[u]) < \infty$ for all $\lambda > 0$ (see Lemma 6.5 below). The proofs of existence of counterexamples are constructive and the Baire category statements will be obtained as rather straightforward consequences of the constructions.

Outline of the paper. In Section 2 we provide the upper bounds for $[Q_{\gamma/p}u]_{L^{p,\infty}(\mathbb{R}^{2N},\nu_\gamma)}$, i.e., the proof of inequalities (1-13) and (1-15) in Theorems 1.3 and 1.4. We first derive these for a dense subclass, relying on covering lemmas, and then extend in Sections 2C and 2D to general $\dot{W}^{1,p}$ and $\dot{\mathrm{BV}}$-functions. In Section 3 we derive the limit formulas of Theorem 1.1; specifically in Section 3B we prove the sharp lower bounds involving a $\liminf \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u])$ for general functions in $\dot{W}^{1,p}$ and in Section 3C we obtain the sharp upper bounds for $\limsup \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u])$, under the assumption that $u \in C^1$ is compactly supported. Then in Section 3D we extend these limits to general $\dot{W}^{1,p}$ functions. In Section 3F we show that the limit formulas for $\dot{W}^{1,1}$ do not extend to general $\dot{\mathrm{BV}}$ functions and prove Proposition 1.2. In Section 4 we prove the reverse inequalities (1-14) and (1-16) in Theorems 1.3 and 1.4. In Section 5 we prove Theorem 1.5 on a characterization of the Lipschitz norm and also discuss Example 1.7. In Section 6 we provide various constructions of counterexamples and in particular prove Theorem 1.8. We discuss some further perspectives and open problems in Section 7.

2. Bounding $[Q_{\gamma/p}u]_{L^{p,\infty}(\mathbb{R}^{2N},\nu_\gamma)}$ by the Sobolev norm

In this section we prove inequalities (1-13) and (1-15) in Theorems 1.3 and 1.4.

2A. ***The bound* (1-13) *via the Hardy–Littlewood maximal operator.*** Following [Brezis et al. 2021b], one can prove the result of Theorem 1.3 for $p > 1$ by an elementary argument involving the Hardy–Littlewood maximal function $M|\nabla u|$ of $|\nabla u|$; however, the behavior of the constants as $p \searrow 1$ will only be sharp in the range $-1 \le \gamma < 0$.

Proposition 2.1. *Let $N \ge 1$ and $1 < p < \infty$. There exists a constant C_N such that, for all $\gamma \ne 0$ and all $u \in \dot{W}^{1,p}(\mathbb{R}^N)$,*

$$\sup_{\lambda>0} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) \le \frac{C_N}{|\gamma|}\left(\frac{p}{p-1}\right)^p \|\nabla u\|^p_{L^p(\mathbb{R}^N)}. \tag{2-1}$$

Proof. We assume first that $u \in C^1$ and that ∇u is compactly supported. As in [Brezis et al. 2021b, Remark 2.3], one uses the Lusin–Lipschitz inequality

$$\frac{|u(x)-u(y)|}{|x-y|} \le C[M(|\nabla u|)(x) + M(|\nabla u|)(y)] \tag{2-2}$$

and observes that (2-2) implies

$$E_{\lambda,\gamma/p}[u] \subseteq \{|x-y|^{\gamma/p} < 2C\lambda^{-1}M(|\nabla u|)(x)\} \cup \{|x-y|^{\gamma/p} < 2C\lambda^{-1}M(|\nabla u|)(y)\}.$$

As a consequence

$$\nu_\gamma(E_{\lambda,\gamma/p}[u]) \le 2\int_x \int_{|h|^\gamma < 2C[\lambda^{-1}M(|\nabla u|)(x)]^p} |h|^{\gamma-N}\,\mathrm{d}h\,\mathrm{d}x.$$

Direct computation of the inner integral (distinguishing the cases $\gamma > 0$ and $\gamma < 0$) yields

$$\nu_\gamma(E_{\lambda,\gamma/p}[u]) \lesssim_N C^p |\gamma|^{-1} \lambda^{-p} \int_{\mathbb{R}^N} [M(|\nabla u|)(x)]^p \, \mathrm{d}x.$$

Inequality (2-1) follows then from the standard maximal inequality $\|Mf\|_p^p \le [C(N)p']^p \|f\|_p^p$ for $p > 1$; see [Stein 1970] (here $p' = p/(p-1)$). The extension to general $\dot{W}^{1,p}$ functions will be taken up in Section 2C. □

2B. *The case $\gamma \in \mathbb{R} \setminus [-1, 0]$.* We shall prove the following more precise versions of the estimates (1-13) and (1-15) when $\gamma \notin [-1, 0]$, with constants that stay bounded as $p \searrow 1$; indeed we cover all $p \in [1, \infty)$. We denote by σ_{N-1} the surface area of the sphere $\mathbb{S}^{N-1}$. In the proof of the following theorem we will first establish the estimates for functions $u \in C^1(\mathbb{R}^N)$ whose gradient is compactly supported. The extension to $\dot{W}^{1,p}$ and $\dot{\mathrm{BV}}$ will be taken up in Sections 2C and 2D.

Theorem 2.2. *There exists an absolute constant $C > 0$ such that, for every $N \ge 1$, every $1 \le p < \infty$, and every $u \in \dot{W}^{1,p}(\mathbb{R}^N)$:*

(i) *If $\gamma > 0$, then*

$$\sup_{\lambda>0} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) \le C\sigma_{N-1} \frac{5^\gamma}{\gamma} \|\nabla u\|^p_{L^p(\mathbb{R}^N)}. \tag{2-3}$$

(ii) *If $\gamma < -1$, then*

$$\sup_{\lambda>0} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) \le \frac{C\sigma_{N-1}}{|\gamma|}\left(1 + \frac{1}{|\gamma+1|}\right) \|\nabla u\|^p_{L^p(\mathbb{R}^N)}. \tag{2-4}$$

When $p = 1$ the above assertions hold for all $u \in \dot{\mathrm{BV}}(\mathbb{R}^N)$ provided that $\|\nabla u\|_{L^1(\mathbb{R}^N)}$ is replaced by $\|\nabla u\|_{\mathcal{M}}$.

The proof of Theorem 2.2 relies on the following proposition, in which $[x, y] \subset \mathbb{R}^N$ denotes the closed line segment connecting two points $x, y \in \mathbb{R}^N$.

Proposition 2.3. *Let*

$$E(f, \gamma) := \left\{(x, y) \in \mathbb{R}^N \times \mathbb{R}^N : x \neq y, \int_{[x,y]} |f| \, \mathrm{d}s > |x-y|^{\gamma+1}\right\} \tag{2-5}$$

for $f \in C_c(\mathbb{R}^N)$. There exists an absolute constant $C > 0$ such that, for every $N \ge 1$ and every $f \in C_c(\mathbb{R}^N)$:

(i) *If $\gamma > 0$, then*

$$\iint_{E(f,\gamma)} |x-y|^{\gamma-N} \, \mathrm{d}x \, \mathrm{d}y \le C\sigma_{N-1} \frac{5^\gamma}{\gamma} \|f\|_{L^1(\mathbb{R}^N)}. \tag{2-6}$$

(ii) *If $\gamma < -1$, then*

$$\iint_{E(f,\gamma)} |x-y|^{\gamma-N} \, \mathrm{d}x \, \mathrm{d}y \le \frac{C\sigma_{N-1}}{|\gamma|}\left(1 + \frac{1}{|\gamma+1|}\right) \|f\|_{L^1(\mathbb{R}^N)}. \tag{2-7}$$

Indeed, to deduce Theorem 2.2 from Proposition 2.3 one argues as in the proof of (1-12) in [Brezis et al. 2021b]; for $u \in C^1(\mathbb{R}^N)$ and $1 \le p < \infty$, one has

$$|u(x) - u(y)|^p \le \left[\int_{[x,y]} |\nabla u| \, \mathrm{d}s\right]^p \le \int_{[x,y]} |\nabla u|^p \, \mathrm{d}s \, |x-y|^{p-1}$$

for all $x, y \in \mathbb{R}^N$, which implies

$$E_{\lambda,\gamma/p}[u] \subseteq E(\lambda^{-p}|\nabla u|^p, \gamma).$$

Hence for $u \in C^1(\mathbb{R}^N)$ whose gradient is compactly supported, one establishes Theorem 2.2 by applying Proposition 2.3 with $f := \lambda^{-p}|\nabla u|^p$. The extension to $u \in \dot{W}^{1,p}$ will be taken up in Section 2C.

Proof of Proposition 2.3. As in the proof of [Brezis et al. 2021b, Proposition 2.2], using the method of rotation, we only need to prove Proposition 2.3 for $N = 1$. Indeed,

$$\iint_{E(f,\gamma)} |x-y|^{\gamma-N}\,\mathrm{d}x\,\mathrm{d}y = \frac{1}{2}\int_{\mathbb{S}^{N-1}}\int_{\omega^\perp}\iint_{E(f_{\omega,x'},\gamma)} |r-s|^{\gamma-1}\,\mathrm{d}r\,\mathrm{d}s\,\mathrm{d}x'\,\mathrm{d}\omega,$$

where for every $\omega \in \mathbb{S}^{N-1}$ and every $x' \in \omega^\perp$, $f_{\omega,x'}$ is a function of one real variable defined by

$$f_{\omega,x'}(t) := f(x' + t\omega).$$

The innermost double integral can be estimated by the case $N = 1$ of Proposition 2.3, and

$$\int_{\mathbb{S}^{N-1}}\int_{\omega^\perp}\int_{\mathbb{R}} |f_{\omega,x'}(t)|\,\mathrm{d}t\,\mathrm{d}x'\,\mathrm{d}\omega = \sigma_{N-1}\|f\|_{L^1(\mathbb{R}^N)}.$$

Thus from now on, we assume $N = 1$ and $f \in C_c(\mathbb{R})$.

If $\gamma > 0$, the desired estimate (2-6) is the content of [Brezis et al. 2021b, Proposition 2.1]. On the other hand, suppose now $\gamma < -1$. Without loss of generality, assume $f \geq 0$ on $\mathbb{R}$. In addition, we may assume that f is not identically zero, for otherwise there is nothing to prove.

Let

$$E_+(f,\gamma) := \{(x,y) \in E(f,\gamma) : y < x\}.$$

Then by symmetry,

$$\iint_{E(f,\gamma)} |x-y|^{\gamma-1}\,\mathrm{d}x\,\mathrm{d}y = 2\iint_{E_+(f,\gamma)} |x-y|^{\gamma-1}\,\mathrm{d}x\,\mathrm{d}y,$$

and it suffices to estimate the latter integral.

In what follows we will need to always keep in mind that in view of our assumption $\gamma < -1$ we have $-(\gamma+1) = |\gamma| - 1 > 0$. We will now use a simple stopping-time argument based on the fact that for all $c \in \mathbb{R}$ the continuous function

$$x \mapsto (x-c)^{-(\gamma+1)}\int_c^x f(s)\,\mathrm{d}s, \quad x \geq c,$$

increases from 0 to ∞ on $[c, \infty)$.

Assume that $\operatorname{supp} f \subseteq [a,b]$. We construct a finite sequence of intervals $I_1, \ldots, I_K$, that are disjoint up to endpoints, that cover $\operatorname{supp} f = [a,b]$, and that satisfy

$$|I_i|^{-(\gamma+1)}\int_{I_i} f = \frac{1}{2} \quad \text{for } 1 \leq i \leq K. \tag{2-8}$$

Indeed, we may take $a_1 := a$, and $a_2 > a_1$ to be the unique number for which

$$(a_2 - a_1)^{-(\gamma+1)}\int_{a_1}^{a_2} f = \frac{1}{2},$$

and set $I_1 := [a_1, a_2]$. If $a_2 < b$, we may now repeat, and take $I_2 := [a_2, a_3]$, where $a_3 > a_2$ is the unique number for which $(a_3 - a_2)^{-(\gamma+1)} \int_{a_2}^{a_3} f = \frac{1}{2}$. Note that the a_i's chosen as such satisfy

$$(a_{i+1} - a_i)^{-(\gamma+1)} \geq \frac{1}{2} \| f \|_{L^1(\mathbb{R})}^{-1},$$

so that $a_{i+1} - a_i \geq (2\| f \|_{L^1(\mathbb{R})})^{1/(\gamma+1)}$. This shows that in finitely many steps, we would reach $a_{K+1} \geq b$ for some $K \geq 1$, with $a_K < b$ if $1 \leq K$. Then we have our sequence of disjoint (up to endpoints) intervals $I_1, \dots, I_K$ that cover $[a, b]$ and satisfy (2-8). We also write $I_0 := (-\infty, a_1]$ and $I_{K+1} := [a_{K+1}, +\infty)$.

We now claim that $I_i \times I_i \cap E_+(f, \gamma) = \varnothing$ for every $0 \leq i \leq K+1$. This being trivially the case when $i \in \{0, K+1\}$, we consider the case $i \in \{1, \dots, K\}$: any $x, y \in I_i$ satisfy

$$|x-y|^{-(\gamma+1)} \left| \int_y^x f \right| \leq |I_i|^{-(\gamma+1)} \int_{I_i} f = \frac{1}{2} < 1.$$

It follows thus that

$$E_+(f, \gamma) = \bigcup_{i=1}^{K+1} E_+(f, \gamma) \cap ((a_i, +\infty) \times (-\infty, a_i)). \tag{2-9}$$

Furthermore, for $i \in \{2, \dots, K\}$, if $y < a_i < x$ and $x - y < \min\{|I_i|, |I_{i-1}|\}$, then

$$\begin{aligned}
|x-y|^{-(\gamma+1)} \left| \int_y^x f \right| &< \min\{|I_i|, |I_{i-1}|\}^{-(\gamma+1)} \left(\int_{I_{i-1}} f + \int_{I_i} f \right) \\
&\leq |I_{i-1}|^{-(\gamma+1)} \int_{I_{i-1}} f + |I_i|^{-(\gamma+1)} \int_{I_i} f \leq \frac{1}{2} + \frac{1}{2} = 1
\end{aligned}$$

(again we used $\gamma < -1$ so that $-(\gamma+1) > 0$ here), from which it follows that $(x, y) \notin E_+(f, \gamma)$. Combining this with a similar argument for $i \in \{1, K+1\}$, we get that if $(x, y) \in E_+(f, \gamma) \cap (a_i, +\infty) \times (-\infty, a_i)$, then $|x - y| \geq \min\{|I_i|, |I_{i-1}|\}$, and thus

$$\begin{aligned}
\int_{E_+(f,\gamma) \cap (a_i,+\infty) \times (-\infty, a_i)} |x-y|^{\gamma-1} \, \mathrm{d}x \, \mathrm{d}y &\leq \int_{a_i}^{\infty} \int_{-\infty}^{\min\{a_i, x - \min\{|I_i|, |I_{i-1}|\}\}} |x-y|^{\gamma-1} \, \mathrm{d}y \, \mathrm{d}x \\
&= \frac{1}{|\gamma|} \int_{a_i}^{\infty} (\max\{x - a_i, \min\{|I_i|, |I_{i-1}|\}\})^{\gamma} \, \mathrm{d}x \\
&= \frac{1}{|\gamma|} \left(1 + \frac{1}{|\gamma+1|} \right) \min\{|I_i|, |I_{i-1}|\}^{\gamma+1} \\
&\leq \frac{2}{|\gamma|} \left(1 + \frac{1}{|\gamma+1|} \right) \int_{I_{i-1} \cup I_i} f.
\end{aligned}$$

(The computation of these integrals uses our assumption $\gamma + 1 < 0$.) Summing the estimates, we get in view of (2-9)

$$\int_{E_+(f,\gamma)} |x-y|^{\gamma-1} \, \mathrm{d}x \, \mathrm{d}y \leq \frac{4}{|\gamma|} \left(1 + \frac{1}{|\gamma+1|} \right) \int_{\mathbb{R}} f.$$

We have thus completed the proof of (2-7) under the assumption $\gamma < -1$ and $N = 1$. □

2C. ***Proof of Proposition 2.1 and Theorem 2.2 for general $\dot{W}^{1,p}$ functions.*** We use a limiting argument, together with the following fact: if $u \in \dot{W}^{1,p}(\mathbb{R}^N)$, $N \geq 1$, and $1 \leq p < \infty$, then there exists a Lebesgue measurable set $X \subset \mathbb{R}^{2N}$, with $\mathcal{L}^{2N}(X) = 0$, so that, for every $(x, h) \in \mathbb{R}^{2N} \setminus X$, we have

$$u(x+h) - u(x) = \int_0^1 \langle h, \nabla u(x+th) \rangle \,\mathrm{d}t. \tag{2-10}$$

Indeed, both sides are measurable functions of $(x, h) \in \mathbb{R}^{2N}$, and if X is the set of all (x, h) where the two sides are not equal, then X is a measurable subset of $\mathbb{R}^{2N}$, and the assertion will follow from Fubini's theorem if, for every fixed $h \in \mathbb{R}^N$, we have $\mathcal{L}^N(\{x \in \mathbb{R}^N : (x, h) \in X\}) = 0$, i.e., (2-10) holds for $\mathcal{L}^N$ almost every x. This follows since for every $\phi \in C_c^\infty(\mathbb{R}^N)$, one has

$$\begin{aligned}\int_{\mathbb{R}^N} [u(x+h)-u(x)]\phi(x)\,\mathrm{d}x &= \int_{\mathbb{R}^N} u(x)[\phi(x-h)-\phi(x)]\,\mathrm{d}x = -\int_{\mathbb{R}^N} u(x) \int_0^1 \langle h, \nabla\phi(x-th)\rangle \,\mathrm{d}t\,\mathrm{d}x \\ &= \int_{\mathbb{R}^N} \int_0^1 \langle h, \nabla u(x)\rangle \phi(x-th)\,\mathrm{d}t\,\mathrm{d}x = \int_{\mathbb{R}^N} \int_0^1 \langle h, \nabla u(x+th)\rangle \,\mathrm{d}t\,\phi(x)\,\mathrm{d}x.\end{aligned}$$

Now given $u \in \dot{W}^{1,p}(\mathbb{R}^N)$, there exists a sequence $u_n \in C^\infty(\mathbb{R}^N)$ such that ∇u_n are compactly supported, and

$$\|\nabla(u_n - u)\|_{L^p(\mathbb{R}^N)} \to 0. \tag{2-11}$$

Indeed if $N > 1$ and $p \geq 1$, or if $N = 1$ and $p > 1$, then this follows from the density of $C_c^\infty(\mathbb{R}^N)$ in $\dot{W}^{1,p}(\mathbb{R}^N)$ as asserted in [Hajłasz and Kałamajska 1995] (in this case one may choose $u_n \in C_c^\infty(\mathbb{R}^N)$). The density of $C_c^\infty(\mathbb{R}^N)$ in $\dot{W}^{1,p}$ *fails* when $N = p = 1$ (again see [Hajłasz and Kałamajska 1995]); the issue is that if ∇u is supported in a convex set in $\mathbb{R}^N$, $N \geq 2$, then u is constant in the complement of the set, but this fails for $N = 1$ since the complement of a bounded interval has two connected components. On the other hand, in the anomalous case $N = 1$ and $p = 1$, one can choose an approximation of the identity to get a sequence v_n of C_c^∞ functions on $\mathbb{R}$ such that $\|v_n - u'\|_{L^1(\mathbb{R})} \to 0$. One can then take $u_n(x) := \int_0^x v_n(t)\,\mathrm{d}t$, and (2-11) follows with $u_n' = v_n$ being compactly supported (even though u_n may not be compactly supported).

Let, for $R > 1$,

$$K_R = \{(x, y) \in \mathbb{R}^{2N} : |x| \leq R,\ |y| \leq R \text{ and } R^{-1} \leq |x-y|\}.$$

By monotone convergence it suffices to prove

$$\nu_\gamma(E_{\lambda,\gamma/p}[u] \cap K_R) \leq C \frac{\|\nabla u\|^p_{L^p(\mathbb{R}^N)}}{\lambda^p}. \tag{2-12}$$

with C independent of R.

Under the assumptions of Proposition 2.1 and Theorem 2.2 on p and γ, since $u_n \in C_c^\infty(\mathbb{R}^N)$, we already know

$$\nu_\gamma(E_{\lambda,\gamma/p}[u_n]) \leq C \frac{\|\nabla u_n\|^p_{L^p(\mathbb{R}^N)}}{\lambda^p}.$$

Moreover, the sequence $Q_{\gamma/p}u_n$ converges to $Q_{\gamma/p}u$ in $L^p(K_R)$ as $n\to\infty$. Indeed, using (2-10) we may write

$$Q_{\gamma/p}u(x,y)=\frac{1}{|x-y|^{\gamma/p}}\int_0^1\left\langle\frac{x-y}{|x-y|},\nabla u((1-t)y+tx)\right\rangle \mathrm{d}t$$

for $\mathcal{L}^{2N}$ a.e. $(x,y)\in\mathbb{R}^{2N}$, and similarly for u_n in place of u, which allows us to estimate

$$\begin{aligned}\left(\iint_{K_R}|Q_{\gamma/p}u_n(x,y)-Q_{\gamma/p}u(x,y)|^p\,\mathrm{d}x\,\mathrm{d}y\right)^{1/p}&\\ &\le R^{\gamma/p}\int_0^1\left(\int_{|x|\le R}\int_{|y|\le R}|\nabla(u_n-u)((1-s)x+sy)|^p\,\mathrm{d}x\,\mathrm{d}y\right)^{1/p}\mathrm{d}s\\ &\le 2^{N/p}(2R)^{N/p}R^{\gamma/p}\|\nabla(u_n-u_{n+1})\|_p\to 0.\end{aligned}$$

By passing to a subsequence if necessary, we may assume that $Q_{\gamma/p}u_n$ converges $\mathcal{L}^{2N}$-a.e. to $Q_{\gamma/p}u$ on K_R as $n\to\infty$. Thus

$$K_R\cap E_{\lambda,\gamma/p}[u]\subseteq K_R\cap\left(\bigcup_{n\in\mathbb{N}}\bigcap_{\ell\ge n}E_{\lambda,\gamma/p}[u_\ell]\right),$$

which implies

$$\begin{aligned}\nu_\gamma(K_R\cap E_{\lambda,\gamma/p}[u])&\le\lim_{n\to\infty}\nu_\gamma\left(K_R\cap\bigcap_{\ell\ge n}E_{\lambda,\gamma/p}[u_\ell]\right)\le\liminf_{n\to\infty}\nu_\gamma(K_R\cap E_{\lambda,\gamma/p}[u_n])\\ &\le C\liminf_{n\to\infty}\frac{\|\nabla u_n\|^p_{L^p(\mathbb{R}^N)}}{\lambda^p}\le C\frac{\|\nabla u\|^p_{L^p(\mathbb{R}^N)}}{\lambda^p}.\end{aligned}$$

2D. ***Proof of Theorem 2.2 for*** $\dot{\mathbf{B}}\mathbf{V}$***-functions.*** We choose a sequence $\rho_n\in C_c^\infty(\mathbb{R}^N)$, with $\rho_n=2^{nN}\rho(2^n\,\cdot)$ and $\int_{\mathbb{R}^N}\rho\,\mathrm{d}x=1$, and set $u_n:=u*\rho_n$. Then $u_n\in\dot{W}^{1,1}(\mathbb{R}^N)$ and $u_n\to u$ almost everywhere. This means if $G_L:=\{(x,h)\in\mathbb{R}^N\times\mathbb{R}^N:|x|\le L,\ L^{-1}\le|h|\le L\}$ then

$$\lim_{n\to\infty}\nu_\gamma(E_{\lambda,\gamma}[u_n]\cap G_L)=\nu_\gamma(E_{\lambda,\gamma}[u]\cap G_L),$$

by dominated convergence. Also

$$\|\nabla u_n\|_{L^1(\mathbb{R}^N)}=\sup_{\substack{\vec\phi\in C_c^\infty\\ \|\phi\|_\infty\le 1}}\left|\int u_n(x)\ \operatorname{div}\vec\phi(x)\,\mathrm{d}x\right|=\sup_{\substack{\vec\phi\in C_c^\infty\\ \|\phi\|_\infty\le 1}}\left|\int u(x)\ \operatorname{div}(\rho_n*\vec\phi)(x)\,\mathrm{d}x\right|\le\|\nabla u\|_{\mathcal{M}};$$

here we used $\|\rho_n*\vec\phi\|_\infty\le\|\vec\phi\|_\infty$ for the last inequality. Combining these two limiting identities with Theorem 2.2 we get the desired inequalities with $E_{\lambda,\gamma}[u]$ replaced by $E_{\lambda,\gamma}[u]\cap G_L$. By monotone convergence we may finish the proof letting $L\to\infty$. ☐

3. Proof of Theorem 1.1

We extend and refine arguments from [Brezis and Nguyen 2018; Brezis et al. 2021b], which are partially inspired by techniques developed in [Bourgain et al. 2001].

3A. ***A Lebesgue differentiation lemma.*** Our argument uses the following standard variant of the Lebesgue differentiation theorem. For lack of a proper reference, a proof is provided for the convenience of the reader.

Lemma 3.1. *Let $u \in \dot{W}^{1,1}(\mathbb{R}^N)$ and let $\{\delta_n\}$ be a sequence of positive numbers with $\lim_{n\to\infty} \delta_n = 0$. Then*

$$\lim_{n\to\infty} \frac{u(x+\delta_n h) - u(x)}{\delta_n} = \langle h, \nabla u(x)\rangle$$

for almost every $(x, h) \in \mathbb{R}^N \times \mathbb{R}^N$.

Proof. If $u \in C^1$ with compact support the limit relation clearly holds for all (x, h). We shall below consider for each $\theta \in \mathbb{S}^{N-1}$ the maximal function

$$\mathfrak{M}_\theta F(x) = \sup_{t>0} \frac{1}{t} \int_0^t |F(x+r\theta)| \,\mathrm{d}r,$$

which is well-defined for all θ, a measurable function on $\mathbb{R}^N \times \mathbb{S}^{N-1}$, and satisfies a weak-type (1, 1) inequality

$$\mathcal{L}^N(\{x \in \mathbb{R}^N : \mathfrak{M}_\theta F(x) > a\}) \le 5a^{-1}\|F\|_1.$$

Let $u \in \dot{W}^{1,1}(\mathbb{R}^N)$ and $\mathcal{A}_M = \{h \in \mathbb{R}^N : 2^{-M} \le |h| \le 2^M\}$. It suffices to prove the limit relation for almost every $(x, h) \in \mathbb{R}^N \times \mathcal{A}_M$. From (2-10) we get that, for every $n \ge 1$,

$$\frac{u(x+\delta_n h) - u(x)}{\delta_n} = \frac{1}{\delta_n |h|} \int_0^{\delta_n |h|} \left\langle h, \nabla u\left(x + r\frac{h}{|h|}\right)\right\rangle \mathrm{d}r$$

for $\mathcal{L}^{2N}$ almost every $(x, h) \in \mathbb{R}^N \times \mathcal{A}_M$; as a result, there exist representatives of u, ∇u and a null set $\mathcal{N} \in \mathbb{R}^N \times \mathcal{A}_M$ such that the identity holds for all $(x, h) \in \mathcal{N}^\complement$ and all $n \ge 1$. It suffices to show that, for every $\alpha > 0$, $\varepsilon > 0$,

$$\mathcal{L}^{2N}\left(\left\{(x, h) \in \mathbb{R}^N \times \mathcal{A}_M : \limsup_{n\to\infty} \left| \frac{1}{\delta_n |h|} \int_0^{\delta_n |h|} \langle h, \nabla u(x+rh)\rangle \,\mathrm{d}r - \langle h, \nabla u(x)\rangle \right| > \alpha \right\}\right) \le \varepsilon. \quad (3\text{-}1)$$

Let $v \in C_c^1$ so that $\|\nabla(v-u)\|_1 \le \alpha\varepsilon/(12\mathcal{L}^N(\mathcal{A}_M))$. Let $g = u - v$. Since the asserted limiting relation holds for v, we see that the expression on the left-hand side of (3-1) is dominated by

$$\begin{aligned}\mathcal{L}^{2N}&\left(\left\{(x, h) \in \mathbb{R}^N \times \mathcal{A}_M : |\nabla g(x)| + \sup_{n>0} \frac{1}{\delta_n |h|} \int_0^{\delta_n |h|} \left|\nabla g\left(x + r\frac{h}{|h|}\right)\right| \mathrm{d}r > \alpha \right\}\right) \\ &\le 2\mathcal{L}^N(\mathcal{A}_M)\alpha^{-1}\|\nabla g\|_1 + \int_{\mathcal{A}_M} \mathcal{L}^N\left(\left\{x : \mathfrak{M}_{h/|h|}|\nabla g|(x) > \frac{\alpha}{2}\right\}\right) \mathrm{d}h \\ &\le 12\mathcal{L}^N(\mathcal{A}_M)\alpha^{-1}\|\nabla g\|_1 \le \varepsilon\end{aligned}$$

since $\|\nabla g\|_1 \le \alpha\varepsilon/(12\mathcal{L}^N(\mathcal{A}_M))$. □

3B. ***The lower bounds for*** $\liminf \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u])$**.** We use Lemma 3.1 to establish lower bounds, relying on an idea in [Brezis and Nguyen 2018], where the case $\gamma = -1$ was considered.

Lemma 3.2. *Let* $1 \le p < \infty$ *and* $u \in \dot{W}^{1,p}(\mathbb{R}^N)$. *Then*:

(i) *For* $\gamma > 0$,
$$\liminf_{\lambda\to\infty} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) \ge \frac{\kappa(p,N)}{|\gamma|}\|\nabla u\|^p_{L^p(\mathbb{R}^N)}.$$

(ii) *For* $\gamma < 0$,
$$\liminf_{\lambda\searrow 0} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) \ge \frac{\kappa(p,N)}{|\gamma|}\|\nabla u\|^p_{L^p(\mathbb{R}^N)}.$$

Proof. We write, for $\lambda > 0$ and $\delta > 0$,
$$\begin{aligned}\lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) &= \lambda^p \iint_{|u(x+h)-u(x)|/|h|^{1+\gamma/p}>\lambda} |h|^{\gamma-N}\,\mathrm{d}h\,\mathrm{d}x\\ &= \lambda^p\delta^\gamma \iint_{\left|(u(x+\delta h)-u(x))/(\delta|h|)\right|^p>\lambda^p\delta^\gamma|h|^\gamma} |h|^{\gamma-N}\,\mathrm{d}h\,\mathrm{d}x;\end{aligned}$$
here we have changed variables replacing h by δh. Hence
$$\lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) = \iint \mathbb{1}_{(|h|^\gamma,\infty)}\left(\left|\frac{u(x+\delta h)-u(x)}{\delta|h|}\right|^p\right)|h|^{\gamma-N}\,\mathrm{d}h\,\mathrm{d}x, \quad \text{with } \delta = \lambda^{-p/\gamma}. \tag{3-2}$$
We now take a sequence $\{\lambda_n\}$ of positive numbers, set $\delta_n = \lambda_n^{-p/\gamma}$ and note that
$$\lim_{n\to\infty}\delta_n = 0 \quad \text{if } \begin{cases}\lim_{n\to\infty}\lambda_n = \infty \text{ and } \gamma > 0,\\ \lim_{n\to\infty}\lambda_n = 0 \quad \text{and } \gamma < 0.\end{cases} \tag{3-3}$$
Also observe that
$$\liminf_{n\to\infty} \mathbb{1}_{(|h|^\gamma,\infty)}(s_n) \ge \mathbb{1}_{(|h|^{-\gamma},\infty)}(t) \quad \text{if } \lim_{n\to\infty} s_n = t.$$

Now assume that $\lambda_n \to \infty$ if $\gamma > 0$ and $\lambda_n \to 0^+$ if $\gamma < 0$ and stay with $\delta_n = \lambda_n^{-p/\gamma}$, a sequence which converges to 0 in both cases. Use Fatou's lemma in (3-2) and combine it with Lemma 3.1 to get
$$\begin{aligned}\liminf_{n\to\infty}\lambda_n^p \nu_\gamma(E_{\lambda_n,\gamma/p}[u]) &\ge \iint \liminf_{n\to\infty}\mathbb{1}_{(|h|^\gamma,\infty)}\left(\left|\frac{u(x+\delta_n h)-u(x)}{\delta_n|h|}\right|^p\right)|h|^{\gamma-N}\,\mathrm{d}h\,\mathrm{d}x\\ &\ge \iint \mathbb{1}_{(|h|^\gamma,\infty)}\left(\lim_{n\to\infty}\left|\frac{u(x+\delta_n h)-u(x)}{\delta_n|h|}\right|^p\right)|h|^{\gamma-N}\,\mathrm{d}h\,\mathrm{d}x\\ &= \iint_{|h|^\gamma<|\langle h/|h|,\nabla u(x)\rangle|^p} |h|^{\gamma-N}\,\mathrm{d}h\,\mathrm{d}x =: J_\gamma.\end{aligned}$$

We use polar coordinates $h = r\theta$ and write the last expression as
$$\begin{aligned}J_\gamma &= \iint_{\mathbb{R}^N\times\mathbb{S}^{N-1}}\int_{r^\gamma<|\langle\theta,\nabla u(x)\rangle|^p} r^{\gamma-1}\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}x\\ &= \frac{1}{|\gamma|}\iint_{\mathbb{R}^N\times S^{N-1}} |\langle\theta,\nabla u(x)\rangle|^p\,\mathrm{d}\theta\,\mathrm{d}x = \frac{\kappa(p,N)}{|\gamma|}\|\nabla u\|^p_{L^p(\mathbb{R}^N)},\end{aligned}$$
with the calculation valid in both cases $\gamma > 0$ and $\gamma < 0$. ☐

3C. ***Upper bounds for*** $\limsup \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u])$***, for*** C_c^1 ***functions.*** We assume that $u \in C^1$ is compactly supported and obtain the sharp upper bounds for $\limsup_{\lambda\to\infty} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u])$ when $\gamma > 0$ and $\limsup_{\lambda\to 0} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u])$ when $\gamma < 0$.

Lemma 3.3. *Suppose $u \in C_c^1(\mathbb{R}^N)$ and $1 \le p < \infty$. Then the following hold*:

(i) *If $\gamma > 0$ then*

$$\limsup_{\lambda\to\infty} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) \le \frac{\kappa(p,N)}{|\gamma|} \|\nabla u\|^p_{L^p(\mathbb{R}^N)}.$$

(ii) *If $\gamma < 0$ then*

$$\limsup_{\lambda\searrow 0} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) \le \frac{\kappa(p,N)}{|\gamma|} \|\nabla u\|^p_{L^p(\mathbb{R}^N)}.$$

(iii) *The statement in part* (i) *continues to hold for $u \in C^1(\mathbb{R}^N)$ whose gradient is compactly supported.*

Remark 3.4. The subtlety in part (iii) above is only relevant in dimension $N = 1$, since if $N \ge 2$, then any function in $C^1(\mathbb{R}^N)$ with a compactly supported gradient is constant outside a compact set.

Proof of Lemma 3.3. We distinguish the cases $\gamma > 0$ and $\gamma < 0$.

The case $\gamma > 0$. We assume that ∇u is compactly supported. To prove part (iii) (and thus part (i)) assume

$$\lambda \ge L := \left\| \left(\sum_{i=1}^N |\partial_i u|^2 \right)^{1/2} \right\|_{L^\infty(\mathbb{R}^N)}. \tag{3-4}$$

Then

$$(x,y) \in E_{\lambda,\gamma/p}[u] \quad\Longrightarrow\quad \lambda|x-y|^{\gamma/p} \le L \quad\Longrightarrow\quad |x-y| \le 1. \tag{3-5}$$

Furthermore, if $(x,y) \in E_{\lambda,\gamma/p}[u]$, then writing $y = x + r\omega$ with $r > 0$ and $\omega \in \mathbb{S}^{N-1}$, we have

$$\lambda r^{\gamma/p} \le |\nabla u(x)\cdot\omega| + \rho(r), \quad \text{with } \rho(r) := \sup_{x\in\mathbb{R}^N} \sup_{|h|\le r} |\nabla u(x+h) - \nabla u(x)|; \tag{3-6}$$

since ∇u is uniformly continuous on $\mathbb{R}^N$, we have $\rho(r) \searrow 0$ as $r \searrow 0$. This, together with the first implication of (3-5), shows

$$\lambda r^{\gamma/p} \le |\nabla u(x)\cdot\omega| + \rho\left(\left(\frac{L}{\lambda}\right)^{p/\gamma}\right). \tag{3-7}$$

Let B be a ball centered at the origin containing $\operatorname{supp}(\nabla u)$, and let $\widetilde{B}$ be the expanded ball with radius $1 + \operatorname{rad}(B)$. Then for $x \notin \widetilde{B}$, we have $Q_{\gamma/p}u(x,y) = 0$ for every y with $|x-y| \le 1$, and (3-5) shows $(x,y) \notin E_{\lambda,\gamma/p}[u]$ for every y with $|x-y| > 1$, so $E_{\lambda,\gamma/p}[u] \subseteq \widetilde{B} \times \mathbb{R}^N$. Define, for $x \in \widetilde{B}$, $\omega \in \mathbb{S}^{N-1}$, and $\lambda > 0$

$$\bar{R}(x,\omega,\lambda) := \left(\lambda^{-1}\left(|\nabla u(x)\cdot\omega| + \rho\left(\left(\frac{L}{\lambda}\right)^{p/\gamma}\right)\right)\right)^{p/\gamma}.$$

Then by (3-7),

$$\begin{aligned}\lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) &\le \lambda^p \int_{\widetilde{B}} \int_{\mathbb{S}^{N-1}} \int_0^{\bar{R}(x,\omega,\lambda)} r^{\gamma-1}\,\mathrm{d}r\,\mathrm{d}\omega\,\mathrm{d}x \\ &= \gamma^{-1} \int_{\widetilde{B}} \int_{\mathbb{S}^{N-1}} \left(|\nabla u(x)\cdot\omega| + \rho\left(\left(\frac{L}{\lambda}\right)^{p/\gamma}\right)\right)^p \mathrm{d}\omega\,\mathrm{d}x.\end{aligned}$$

Letting $\lambda \to \infty$ we get

$$\limsup_{\lambda\to\infty} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) \le \gamma^{-1}\kappa(p,N)\int_{\widetilde{B}} |\nabla u(x)|^p \,\mathrm{d}x$$

and hence the assertion.

The case $\gamma < 0$. We first note that if $(x, y) \in E_{\lambda,\gamma/p}[u]$, then writing $y = x + r\omega$, we have again (3-6).

Now let $\varepsilon > 0$, and let $\delta(\varepsilon) > 0$ be such that $\rho(r) \le \varepsilon$ for $0 < r \le \delta(\varepsilon)$. Let

$$r_\lambda(x,\omega,\varepsilon) = \min\left\{\delta(\varepsilon), \left(\frac{\lambda}{|\nabla u(x)\cdot\omega| + \varepsilon}\right)^{-p/\gamma}\right\}.$$

Note that $r_\lambda(x, \omega, \varepsilon) > 0$ for $\lambda > 0$. Also if $(x, x + r\omega) \in E_{\lambda,\gamma/p}[u]$ then $r \ge r_\lambda(x, \omega, \varepsilon)$; indeed, either $r_\lambda(x, \omega, \varepsilon) \ge \delta(\varepsilon)$ already, or else $r_\lambda(x, \omega, \varepsilon) < \delta(\varepsilon)$, in which case (3-6) shows

$$r_\lambda(x,\omega,\varepsilon) \ge \left(\frac{\lambda}{|\nabla u(x)\cdot\omega| + \varepsilon}\right)^{-p/\gamma}.$$

Finally let B be any ball in $\mathbb{R}^N$ containing the support of u, and let $\widetilde{B}$ be the double ball. Then

$$\begin{aligned}
\limsup_{\lambda\searrow 0} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u] \cap (\widetilde{B}\times\mathbb{R}^N)) &\le \limsup_{\lambda\searrow 0} \lambda^p \int_{\widetilde{B}}\int_{\mathbb{S}^{N-1}}\int_{r_\lambda(x,\omega,\varepsilon)}^\infty r^{\gamma-1}\,\mathrm{d}r\,\mathrm{d}\omega\,\mathrm{d}x\\
&= \limsup_{\lambda\searrow 0} \lambda^p \int_{\widetilde{B}}\int_{\mathbb{S}^{N-1}} \frac{1}{|\gamma|}[r_\lambda(x,\omega,\varepsilon)]^\gamma\,\mathrm{d}\omega\,\mathrm{d}x\\
&= \limsup_{\lambda\searrow 0} \frac{1}{|\gamma|}\int_{\widetilde{B}}\int_{\mathbb{S}^{N-1}} \max\{\lambda^p\delta(\varepsilon)^\gamma, (|\nabla u(x)\cdot\omega| + \varepsilon)^p\}\,\mathrm{d}\omega\,\mathrm{d}x\\
&= \frac{1}{|\gamma|}\int_{\widetilde{B}}\int_{\mathbb{S}^{N-1}} (|\nabla u(x)\cdot\omega| + \varepsilon)^p\,\mathrm{d}\omega\,\mathrm{d}x.
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary we obtain

$$\limsup_{\lambda\searrow 0} \lambda^p \nu_\gamma\big(E_{\lambda,\gamma/p}[u] \cap (\widetilde{B}\times\mathbb{R}^N)\big) \le \frac{1}{|\gamma|}\kappa(p,N)\|\nabla u\|^p_{L^p(\mathbb{R}^N)}. \tag{3-8}$$

Since $u = 0$ in $\mathbb{R}^N \setminus B$, if $(x, y) \in E_{\lambda,\gamma/p}[u] \cap ((\mathbb{R}^N \setminus \widetilde{B}) \times \mathbb{R}^N)$ then $y \in B$. Therefore

$$\limsup_{\lambda\searrow 0} \lambda^p \nu_\gamma\big(E_{\lambda,\gamma/p}[u] \cap ((\mathbb{R}^N\setminus\widetilde{B})\times\mathbb{R}^N)\big) \le \limsup_{\lambda\searrow 0} \lambda^p \int_B\int_{\mathbb{R}^N\setminus\widetilde{B}} |x-y|^{\gamma-N}\,\mathrm{d}x\,\mathrm{d}y = 0.$$

This finishes the proof of part (ii). □

In dimension $N = 1$, when $\gamma < -1$, one can also weaken the hypothesis $u \in C^1_c(\mathbb{R})$ in Lemma 3.3 to $u \in C^1(\mathbb{R})$ and u' is compactly supported:

Lemma 3.5. *Suppose $u \in C^1(\mathbb{R})$, u' is compactly supported, and $1 \le p < \infty$. If $\gamma < -1$ then*

$$\limsup_{\lambda\searrow 0} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) \le \frac{\kappa(p,N)}{|\gamma|}\|u'\|^p_{L^p(\mathbb{R})}.$$

Proof. Let $\operatorname{supp}(u') \subset B := (-\beta, \beta)$. By (3-8) we have

$$\limsup_{\lambda \searrow 0} \nu_\gamma(E_{\lambda,\gamma/p}[u] \cap (-2\beta, 2\beta) \times \mathbb{R}) \le \frac{1}{|\gamma|} \kappa(p,1) \|u'\|^p_{L^p(\mathbb{R})}.$$

Moreover, since u is constant on (β, ∞) and constant on $(-\infty, -\beta)$, if $(x, y) \in E_{\lambda,\gamma/p}[u]$ and $x < -2\beta$ then $y > -\beta$, and if $(x, y) \in E_{\lambda,\gamma/p}[u]$ and $x > 2\beta$ then $y < \beta$. Since $\gamma < -1$,

$$\nu_\gamma(E_{\lambda,\gamma/p}[u] \cap (\mathbb{R} \setminus (-2\beta, 2\beta)) \times \mathbb{R}) \le \int_{2\beta}^{\infty} \int_{-\infty}^{\beta} (x-y)^{\gamma-1} \, \mathrm{d}y \, \mathrm{d}x + \int_{-\infty}^{-2\beta} \int_{-\beta}^{\infty} (y-x)^{\gamma-1} \, \mathrm{d}y \, \mathrm{d}x < \infty.$$

We conclude

$$\limsup_{\lambda \searrow 0} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u] \cap (\mathbb{R} \setminus (-2\beta, 2\beta)) \times \mathbb{R}) = 0. \qquad \square$$

3D. ***Upper bounds for*** $\limsup \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u])$***, for general*** $\dot W^{1,p}$ ***functions.*** Let $N \ge 1$, $1 \le p < \infty$ and $u \in \dot W^{1,p}(\mathbb{R}^N)$. In light of Lemma 3.2, to prove the limiting relations (1-7) and (1-8) in Theorem 1.1, we need only show that

$$\limsup_{\lambda \to \infty} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) \le \frac{\kappa(p,N)}{|\gamma|} \|\nabla u\|^p_{L^p(\mathbb{R}^N)} \tag{3-9}$$

if $\gamma > 0$ and

$$\limsup_{\lambda \searrow 0} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) \le \frac{\kappa(p,N)}{|\gamma|} \|\nabla u\|^p_{L^p(\mathbb{R}^N)} \tag{3-10}$$

if $\gamma < 0$ and $p > 1$, or $\gamma < -1$ and $p = 1$. Lemma 3.3(i)–(ii) asserts that these desired inequalities hold for functions in $C^1_c(\mathbb{R}^N)$. When $N \ge 2$ or $p > 1$, a general $\dot W^{1,p}(\mathbb{R}^N)$ function can be approximated in $\dot W^{1,p}(\mathbb{R}^N)$ by functions in $C^1_c(\mathbb{R}^N)$; by [Hajłasz and Kałamajska 1995], there exists a sequence $\{u_n\}$ in $C^\infty_c(\mathbb{R}^N)$ such that $\lim_{n\to\infty} \|\nabla(u_n - u)\|_{L^p(\mathbb{R}^N)} = 0$. If further $\gamma > 0$, or $\gamma < 0$ and $p > 1$, or $\gamma < -1$ and $p = 1$, then by parts (i) of Theorems 1.3 and 1.4 (proved in Section 2), we have

$$\sup_{\lambda > 0} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u_n - u]) \le C^p_{N,p,\gamma} \|\nabla(u_n - u)\|^p_{L^p(\mathbb{R}^N)}. \tag{3-11}$$

It follows that, for every n and every $\delta \in (0, 1)$,

$$\begin{aligned}\limsup_{\lambda \to \infty} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) &\le \limsup_{\lambda \to \infty} \lambda^p \nu_\gamma(E_{(1-\delta)\lambda,\gamma/p}[u_n]) + \sup_{\lambda > 0} \lambda^p \nu_\gamma(E_{\delta\lambda,\gamma/p}[u_n - u]) \\ &\le \frac{\kappa(p,N)}{|\gamma|(1-\delta)} \|\nabla u_n\|^p_{L^p(\mathbb{R}^N)} + \frac{C^p_{N,p,\gamma} \|\nabla(u_n - u)\|^p_{L^p(\mathbb{R}^N)}}{\delta^p}\end{aligned} \tag{3-12}$$

if $\gamma > 0$, and a similar inequality holds with $\limsup_{\lambda\to\infty}$ replaced by $\limsup_{\lambda\searrow 0}$ if $\gamma < 0$, $p > 1$ or $\gamma < -1$, $p = 1$. Letting first $n \to \infty$ and then $\delta \to 0$, we get the desired conclusions (3-9) and (3-10) under the corresponding conditions on γ and p.

It remains to tackle the case $N = p = 1$, in which case we only need to prove (3-9) when $\gamma > 0$ and (3-10) when $\gamma < -1$. Using (2-11), we approximate u by finding a sequence $\{u_n\}$ in $C^\infty(\mathbb{R})$ so that u'_n are compactly supported for each n, and $\lim_{n\to\infty} \|u'_n - u'\|_{L^1(\mathbb{R})} = 0$. Since the desired inequalities hold for u_n in place of u by Lemma 3.3(iii) and Lemma 3.5, and since part (i) of Theorem 1.4 applies to give

(3-11) when $\gamma > 0$ or $\gamma < -1$, our earlier argument in (3-12) can be repeated to yield (3-9) when $\gamma > 0$ and (3-10) when $\gamma < -1$. This completes our proof of parts (a) and (b) of Theorem 1.1.

3E. *Conclusion of the proof of Theorem 1.1.* In Section 3D we proved parts (a) and (b) of Theorem 1.1. The lower bound for the lim inf in part (c) has been established in Lemma 3.2(ii), and the limiting equality for $u \in C_c^1(\mathbb{R}^N)$ when $p = 1$ and $-1 \le \gamma < 0$ follows by combining that with the upper bound for the lim sup in part (ii) of Lemma 3.3. The proof of the negative result in part (c) of the theorem (generic failure for $p = 1$, $-1 \le \gamma < 0$) will be given in Proposition 6.6 below. □

3F. *On limit formulas for* $\dot{\mathrm{BV}}(\mathbb{R})$*-functions: the proof of Proposition 1.2.* When $p = 1$, Poliakovsky [2022] asked whether (1-7) still holds for $u \in \dot{\mathrm{BV}}(\mathbb{R}^N)$ instead of $\dot{W}^{1,1}(\mathbb{R}^N)$ if $\gamma = N$. More generally, one may wonder whether it is possible that, for all $u \in \dot{\mathrm{BV}}(\mathbb{R}^N)$, one has

$$\lim_{\lambda\to\infty} \lambda \nu_\gamma(E_{\lambda,\gamma}[u]) = \frac{\kappa(1, N)}{|\gamma|} \|\nabla u\|_{\mathcal{M}} \quad \text{when } \gamma > 0, \tag{3-13}$$

$$\lim_{\lambda\to 0^+} \lambda \nu_\gamma(E_{\lambda,\gamma}[u]) = \frac{\kappa(1, N)}{|\gamma|} \|\nabla u\|_{\mathcal{M}} \quad \text{when } \gamma < 0. \tag{3-14}$$

We show that this is *not* the case.

First, when $-1 \le \gamma < 0$, Theorem 1.8(i) (proved in Proposition 6.3 below) shows that even if $u \in \dot{W}^{1,1}(\mathbb{R}^N)$, it may happen that $\lim_{\lambda\to 0^+} \lambda \nu_\gamma(E_{\lambda,\gamma}[u]) = \infty$. So (3-14) cannot hold for all $u \in \dot{\mathrm{BV}}(\mathbb{R}^N)$ for such γ.

The following lemma provides examples of failure of (3-13) and (3-14) when $\gamma \in \mathbb{R} \setminus [-1, 0]$, since $|\gamma + 1| \ne |\gamma|$ unless $\gamma = -\frac{1}{2}$:

Lemma 3.6. *Suppose $N \ge 1$ and $u = \mathbb{1}_\Omega$, where Ω is any bounded convex domain in $\mathbb{R}^N$ with smooth boundary. Then $u \in \dot{\mathrm{BV}}(\mathbb{R}^N)$ and*

$$\lim_{\lambda\to\infty} \lambda \nu_\gamma(E_{\lambda,\gamma}[u]) = \frac{\kappa(1, N)}{|\gamma + 1|} \|\nabla u\|_{\mathcal{M}} \quad \textit{for all } \gamma > -1,$$

while

$$\lim_{\lambda\to 0^+} \lambda \nu_\gamma(E_{\lambda,\gamma}[u]) = \frac{\kappa(1, N)}{|\gamma + 1|} \|\nabla u\|_{\mathcal{M}} \quad \textit{for all } \gamma < -1.$$

Proof. First consider the case $N = 1$. If $u = \mathbb{1}_{[0,\infty)}$ (so that $\|u'\|_{\mathcal{M}(\mathbb{R})} = 1$), then, for every $\gamma \in \mathbb{R} \setminus \{-1\}$ and $\lambda > 0$, one has

$$\nu_\gamma(E_{\lambda,\gamma}[u]) = 2\nu_\gamma(\{(x, y) \in \mathbb{R} : x \ge 0,\ y < 0,\ |x - y|^{-(\gamma+1)} \ge \lambda\}) = \frac{2}{|\gamma + 1|} \frac{1}{\lambda}, \tag{3-15}$$

which follows from a change of variables $s = x - y,\ t = x + y$: when $\gamma > -1$, one has

$$\nu_\gamma(E_{\lambda,\gamma}[u]) = \int_0^{\lambda^{-1/(\gamma+1)}} \int_{-s}^{s} \mathrm{d}t\, s^{\gamma-1}\, \mathrm{d}s = 2 \int_0^{\lambda^{-1/(\gamma+1)}} s^\gamma\, \mathrm{d}s = \frac{2}{\gamma + 1} \frac{1}{\lambda},$$

while when $\gamma < -1$, one has

$$\nu_\gamma(E_{\lambda,\gamma}[u]) = \int_{\lambda^{-1/(\gamma+1)}}^{\infty} \int_{-s}^{s} \mathrm{d}t\, s^{\gamma-1}\, \mathrm{d}s = 2 \int_{\lambda^{-1/(\gamma+1)}}^{\infty} s^\gamma\, \mathrm{d}s = \frac{2}{|\gamma + 1|} \frac{1}{\lambda}.$$

A similar calculation shows that if $u = \mathbb{1}_I$ is a characteristic function of a bounded open interval (so that $\|u'\|_{\mathcal{M}(\mathbb{R})} = 2$), then

$$\lim_{\lambda\to\infty} \lambda\nu_\gamma(E_{\lambda,\gamma}[u]) = \frac{2}{|\gamma+1|}\|u'\|_{\mathcal{M}(\mathbb{R})} \quad \text{for all } \gamma > -1, \tag{3-16}$$

while

$$\lim_{\lambda\to 0^+} \lambda\nu_\gamma(E_{\lambda,\gamma}[u]) = \frac{2}{|\gamma+1|}\|u'\|_{\mathcal{M}(\mathbb{R})} \quad \text{for all } \gamma < -1; \tag{3-17}$$

we also have

$$\sup_{\lambda>0} \lambda\nu_\gamma(E_{\lambda,\gamma}[u]) \le \frac{2}{|\gamma+1|}\|u'\|_{\mathcal{M}(\mathbb{R})} \quad \text{for all } \gamma \in \mathbb{R}\setminus\{-1\}. \tag{3-18}$$

Now consider the case $N \ge 2$. Let Ω be a bounded convex domain in $\mathbb{R}^N$ with smooth boundary and $u = \mathbb{1}_\Omega$. Then $u \in \dot{\mathrm{BV}}(\mathbb{R}^N)$ with $\|\nabla u\|_{\mathcal{M}} = \mathcal{L}^{N-1}(\partial\Omega)$. The method of rotation shows

$$\lambda\nu_\gamma(E_{\lambda,\gamma}[u]) = \frac{1}{2}\int_{\mathbb{S}^{N-1}}\int_{\omega^\perp} \lambda\nu_\gamma(E_{\lambda,\gamma}[u_{\omega,x'}])\,\mathrm{d}x'\,\mathrm{d}\omega,$$

where $u_{\omega,x'}(t) := u(x'+t\omega)$ for $\omega \in \mathbb{S}^{N-1}$ and $x' \in \omega^\perp$. Note that $\|u'_{\omega,x'}\|_{\mathcal{M}(\mathbb{R})} \le 2$ for all $\omega \in \mathbb{S}^{N-1}$ and all $x' \in \omega^\perp$, since Ω is convex and every line only meets $\partial\Omega$ at at most two points. Thus (3-16), (3-18) and the dominated convergence theorem allow one to show that

$$\lim_{\lambda\to\infty} \lambda\nu_\gamma(E_{\lambda,\gamma}[u]) = \frac{1}{|\gamma+1|}\int_{\mathbb{S}^{N-1}}\int_{\omega^\perp} \|u'_{\omega,x'}\|_{\mathcal{M}(\mathbb{R})}\,\mathrm{d}x'\,\mathrm{d}\omega \quad \text{for all } \gamma > -1,$$

and using (3-17) in place of (3-16) we obtain the same conclusion with $\lim_{\lambda\to\infty}$ replaced by $\lim_{\lambda\to 0^+}$ if $\gamma < -1$. It remains to observe that

$$\int_{\mathbb{S}^{N-1}}\int_{\omega^\perp} \|u'_{\omega,x'}\|_{\mathcal{M}(\mathbb{R})}\,\mathrm{d}x'\,\mathrm{d}\omega = \kappa(1,N)\|\nabla u\|_{\mathcal{M}}. \tag{3-19}$$

This holds by Fubini's theorem if $u = \mathbb{1}_\Omega$ is replaced by $u_\varepsilon := u * \rho_\varepsilon$, where ρ_ε is a suitable family of mollifiers, because the left-hand side is then just

$$\int_{\mathbb{S}^{N-1}}\int_{\omega^\perp}\int_{\mathbb{R}} \left|\frac{d}{dt}u_\varepsilon(x'+t\omega)\right| \mathrm{d}t\,\mathrm{d}x'\,\mathrm{d}\omega = \int_{\mathbb{S}^{N-1}}\int_{\mathbb{R}^N} |\omega\cdot\nabla u_\varepsilon(x)|\,\mathrm{d}x\,\mathrm{d}\omega,$$

which equals $\kappa(1,N)\|\nabla u_\varepsilon\|_{L^1(\mathbb{R}^N)}$. One then just needs to let $\varepsilon \to 0$ to obtain (3-19): in fact, a standard argument shows that

$$\lim_{\varepsilon\to 0^+} \|\nabla u_\varepsilon\|_{L^1(\mathbb{R}^N)} = \|\nabla u\|_{\mathcal{M}(\mathbb{R}^N)}.$$

so it remains to prove that

$$\lim_{\varepsilon\to 0^+}\int_{\mathbb{S}^{N-1}}\int_{\omega^\perp}\int_{\mathbb{R}} \left|\frac{d}{dt}u_\varepsilon(x'+t\omega)\right| \mathrm{d}t\,\mathrm{d}x'\,\mathrm{d}\omega = \int_{\mathbb{S}^{N-1}}\int_{\omega^\perp} \|u'_{\omega,x'}\|_{\mathcal{M}(\mathbb{R})}\,\mathrm{d}x'\,\mathrm{d}\omega. \tag{3-20}$$

But for every $\omega \in \mathbb{S}^{N-1}$, and almost every $x' \in \omega^\perp$ (as long as $t \mapsto x'+t\omega$ parametrizes a line $L_{\omega,x'}$ that is either disjoint from Ω, or intersects $\partial\Omega$ transversely at two different points), we have

$$\lim_{\varepsilon\to 0^+}\int_{\mathbb{R}} \left|\frac{d}{dt}u_\varepsilon(x'+t\omega)\right| \mathrm{d}t = \|u'_{\omega,x'}\|_{\mathcal{M}(\mathbb{R})}. \tag{3-21}$$

The validity of (3-21) is clear if $L_{\omega,x'}$ does not intersect Ω, while if $L_{\omega,x'}$ intersects $\partial\Omega$ transversely at two different points, then we can choose a coordinate system so that $\omega = (0, \dots, 0, 1)$, and assume that for some open neighborhood U of x' in $\omega^\perp$, the intersection of $U \times L_{\omega,x'}$ with Ω takes the form

$$\{(y', y_N) : y' \in U, \phi_1(y') < y_N < \phi_2(y')\}$$

for some smooth functions ϕ_1 and ϕ_2 of $y' \in U$. Then, for $\varepsilon > 0$ sufficiently small,

$$\begin{aligned}\int_{\mathbb{R}}\left|\frac{d}{dt}u_\varepsilon(x'+t\omega)\right| \mathrm{d}t &= \int_{\mathbb{R}}\left|\int_{\mathbb{R}^N} \mathbb{1}_\Omega(y)\partial_N\rho_\varepsilon(x'-y', t-y_N)\,\mathrm{d}y\right| \mathrm{d}t\\ &= \int_{\mathbb{R}}\left|-\int_{\mathbb{R}^N} \mathbb{1}_{\phi_1(y')<y_N<\phi_2(y')}\frac{\partial}{\partial y_N}[\rho_\varepsilon(x'-y', t-y_N)]\,\mathrm{d}y\right| \mathrm{d}t\\ &= \int_{\mathbb{R}}\left|\int_{\mathbb{R}^{N-1}} \rho_\varepsilon(x'-y', t-\phi_1(y')) - \rho_\varepsilon(x'-y', t-\phi_2(y'))\,\mathrm{d}y'\right| \mathrm{d}t\\ &= \int_{\mathbb{R}}\left(\int_{\mathbb{R}^{N-1}} \rho_\varepsilon(x'-y', t-\phi_1(y'))\,\mathrm{d}y' + \int_{\mathbb{R}^{N-1}} \rho_\varepsilon(x'-y', t-\phi_2(y'))\,\mathrm{d}y'\right) \mathrm{d}t\\ &= 2\int_{\mathbb{R}^{N-1}}\int_{\mathbb{R}} \rho_\varepsilon(x'-y', t)\,\mathrm{d}t\,\mathrm{d}y'\\ &= 2 = \|u'_{\omega,x'}\|_{\mathcal{M}(\mathbb{R})}.\end{aligned}$$

This proves (3-21), and then the dominated convergence theorem allows one to conclude the proof of (3-20). ☐

Remark. The identity (3-19) for $u = \mathbb{1}_\Omega$ can be derived from Crofton's formula for rather general (not necessarily convex) domains Ω. See [Federer 1969, Chapter 3.2.26], which showed that when $\partial\Omega$ is rectifiable, then its $(N-1)$-dimensional Hausdorff measure $\mathcal{H}^{N-1}(\partial\Omega)$ is equal to $\mathscr{I}_1^{N-1}(\partial\Omega)$, where $\mathscr{I}_1^{N-1}(\partial\Omega)$ is given by [Federer 1969, Chapter 2.10.15] as

$$\frac{1}{\beta_1(N, N-1)}\int_{p\in\mathbf{O}^*(N,N-1)}\int_{y\in\mathbb{R}^{N-1}} N(p|_{\partial\Omega}, y)\,\mathrm{d}y\,\mathrm{d}p;$$

here $\mathbf{O}^*(N, N-1)$ is the space of all orthogonal projections p from $\mathbb{R}^N$ onto $\mathbb{R}^{N-1}$, $\mathrm{d}p$ is the right-$\mathbf{O}(N)$-invariant measure on $\mathbf{O}^*(N, N-1)$ normalized so that $\int_{\mathbf{O}^*(N,N-1)}\mathrm{d}p = 1$, $N(p|_{\partial\Omega}, y)$ is the number of points $x \in \partial\Omega$ so that $px = y$, and

$$\beta_1(N, N-1) = \frac{\Gamma(N/2)}{\Gamma((N+1)/2)\Gamma(1/2)}$$

according to [Federer 1969, Chapter 3.2.13]. It follows that, for $u = \mathbb{1}_\Omega$,

$$\begin{aligned}\int_{\mathbb{S}^{N-1}}\int_{\omega^\perp}\|u'_{\omega,x'}\|_{\mathcal{M}(\mathbb{R})}\,\mathrm{d}x'\,\mathrm{d}\omega &= \mathcal{H}^{N-1}(\mathbb{S}^{N-1})\int_{p\in\mathbf{O}^*(N,N-1)}\int_{y\in\mathbb{R}^{N-1}} N(p|_{\partial\Omega}, y)\,\mathrm{d}y\,\mathrm{d}p\\ &= \frac{2\pi^{N/2}}{\Gamma(N/2)}\beta_1(N, N-1)\mathcal{H}^{N-1}(\partial\Omega)\\ &= \frac{2\pi^{(N-1)/2}}{\Gamma((N+1)/2)}\|\nabla u\|_{\mathcal{M}} = \kappa(1, N)\|\nabla u\|_{\mathcal{M}},\end{aligned}$$

as asserted in (3-19).

4. From weak-type bounds on quotients to $\dot{W}^{1,p}$ and $\dot{\mathrm{BV}}$

In this section we complete the proofs of Theorems 1.3 and 1.4 proving part (ii) of these theorems. We use as a key tool the BBM formula discovered in [Bourgain et al. 2001] (see also [Dávila 2002] for additional information for the BV case), in a way that is reminiscent of the proof of [Nguyen 2006, Theorem 2], and we apply duality for Lorentz spaces to control the double integral arising in the BBM formula. The BBM formula stated in [Bourgain et al. 2001] is quite flexible, involving a bounded smooth domain Ω and a sequence of nonnegative radial mollifiers $\rho_n(|x|)$, with $\int_0^\infty \rho_n(r) r^{N-1}\,\mathrm{d}r = 1$ and $\lim_{n\to\infty} \int_\delta^\infty \rho_n(r) r^{N-1}\,\mathrm{d}r = 0$ for every $\delta > 0$; we will apply it in the case when $\Omega = B_R$, the ball of radius R centered at 0, and $\rho_n(r) = s_n p (2R)^{-s_n p} r^{-N+s_n p} \mathbb{1}_{[0,2R]}(r)$, where $\{s_n\}$ is a sequence of positive numbers tending to 0. As a result, we conclude that if $R > 0$, $1 \le p < \infty$, $u \in L^p(B_R)$ and

$$\liminf_{s\to 0^+} s \iint_{B_R\times B_R} \frac{|u(x)-u(y)|^p}{|x-y|^{N+p-sp}}\,\mathrm{d}x\,\mathrm{d}y < \infty,$$

then for $p = 1$ we have $u \in \dot{\mathrm{BV}}(B_R)$ with $\|\nabla u\|_{\mathcal{M}(B_R)}$ being bounded by $\kappa(1, N)$ times the above liminf, and, for $1 < p < \infty$ we have $u \in \dot{W}^{1,p}(B_R)$ and $\|\nabla u\|_{L^p(B_R)}$ being bounded by $\kappa(p, N)/p$ times the above liminf. The assumption $u \in L^p(B_R)$ can easily be relaxed to $u \in L^1(B_R)$, via an observation of Stein as explained in [Brezis 2002, proof of Theorem 2]: if $u \in L^1(B_R)$ and the above liminf is finite for some $1 < p < \infty$, then, for any $\delta > 0$ and any $\varepsilon \in (0, \delta)$, we may consider $u_\varepsilon := u * \phi_\varepsilon(x)$, where $\phi_\varepsilon(x) := \varepsilon^{-N}\phi(\varepsilon^{-1}x)$ and $\phi \in C_c^\infty(B_1)$ is nonnegative and has integral 1. Then u_ε is C^∞ on the closure of the ball $B_{R-\delta}$, so the above formulation of BBM applies, and $\|\nabla u_\varepsilon\|_{L^p(B_{R-\delta})}$ is uniformly bounded independent of $\varepsilon \in (0, \delta)$; indeed Jensen's inequality implies

$$\iint_{B_{R-\delta}\times B_{R-\delta}} \frac{|u_\varepsilon(x)-u_\varepsilon(y)|^p}{|x-y|^{N+p-sp}}\,\mathrm{d}x\,\mathrm{d}y \le \iint_{B_R\times B_R} \frac{|u(x)-u(y)|^p}{|x-y|^{N+p-sp}}\,\mathrm{d}x\,\mathrm{d}y$$

for every ε. This shows that a subsequence of $\{\nabla u_\varepsilon\}$ converges weakly in $L^p(B_{R-\delta})$ to the distributional gradient ∇u on $B_{R-\delta}$, and a desired bound on $\|\nabla u\|_{L^p(B_{R-\delta})}$ follows for every $\delta > 0$.

Suppose now $N \ge 1$, $1 \le p < \infty$, $\gamma \in \mathbb{R}$, $u \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$ and $Q_{\gamma/p}u \in L^{p,\infty}(\mathbb{R}^{2N}, \nu_\gamma)$. Let

$$A := \sup_{R>0} \liminf_{s\to 0^+} s \iint_{B_R\times B_R} \frac{|u(x)-u(y)|^p}{|x-y|^{N+p-sp}}\,\mathrm{d}x\,\mathrm{d}y. \tag{4-1}$$

Suppose A is finite. If $p = 1$, then the BBM formula above implies $u \in \dot{\mathrm{BV}}(B_R)$ for every $R > 0$, with $\|\nabla u\|_{\mathcal{M}(B_R)} \le \kappa(1, N)A$ independent of R; as a result, $u \in \dot{\mathrm{BV}}(\mathbb{R}^N)$, with $\|\nabla u\|_{\mathcal{M}(\mathbb{R}^N)} \le \kappa(1, N)A$. Similarly, if $1 < p < \infty$, the above BBM formula (applicable for $u \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$) implies $u \in \dot{W}^{1,p}(\mathbb{R}^N)$, with $\|\nabla u\|_{L^p(\mathbb{R}^N)} \le (\kappa(1, N)A/p)^{1/p}$.

It remains to prove that $A < \infty$. By considering truncations of u we may assume additionally that $u \in L^\infty(\mathbb{R}^N)$; the reduction is based on the pointwise bound

$$Q_{\gamma/p}u_n(x, y) \le Q_{\gamma/p}u(x, y), \quad \text{where } u_n(x) = \begin{cases} u(x) & \text{if } |u(x)| < n, \\ nu(x)/|u(x)| & \text{if } |u(x)| \ge n. \end{cases}$$

Using the definition of weak derivative we see by a limiting argument that the conclusion $\sup_n \|\nabla u_n\|_p \le C$ implies $\|\nabla u\|_p \le C$ if $p > 1$ and $\sup_n \|\nabla u_n\|_{\mathcal{M}} \le C$ implies $\|\nabla u\|_{\mathcal{M}} \le C$.

In order to establish our estimate for bounded functions we will use Lorentz duality in the following form: if F, G are measurable functions on $\mathbb{R}^{2N}$, then, for any $1 < q < \infty$, we have

$$\iint_{\mathbb{R}^N\times\mathbb{R}^N} F(x,y)G(x,y)\,\mathrm{d}\nu_\gamma \le q'[F]_{L^{q,\infty}(\mathbb{R}^{2N},\nu_\gamma)}[G]_{L^{q',1}(\mathbb{R}^{2N},\nu_\gamma)}, \tag{4-2}$$

where $1/q + 1/q' = 1$,

$$[F]_{L^{q,\infty}(\mathbb{R}^{2N},\nu_\gamma)} := \sup_{\lambda>0} \lambda\nu_\gamma(\{|F|>\lambda\})^{1/q} = \sup_{t>0} t^{1/q}F^*(t),$$
$$[G]_{L^{q',1}(\mathbb{R}^{2N},\nu_\gamma)} := \int_0^\infty \nu_\gamma(\{|G|>\lambda\})^{1/q'}\,\mathrm{d}\lambda = \frac{1}{q'}\int_0^\infty t^{1/q'}G^*(t)\frac{\mathrm{d}t}{t};$$

here $F^*(t) := \inf\{s>0 : \nu_\gamma(\{|F|>\lambda\}) \le s\}$ is the nonincreasing rearrangement of F, and similarly for $G^*(t)$; see [Hunt 1966; Stein and Weiss 1971]. Indeed, (4-2) follows by noticing that

$$\iint_{\mathbb{R}^N\times\mathbb{R}^N} F(x,y)G(x,y)\,\mathrm{d}\nu_\gamma \le \int_0^\infty F^*(t)G^*(t)\,\mathrm{d}t = \int_0^\infty [t^{1/q}F^*(t)][t^{1/q'}G^*(t)]\frac{\mathrm{d}t}{t},$$

which is clearly $\le q'[F]_{L^{q,\infty}(\mathbb{R}^{2N},\nu_\gamma)}[G]_{L^{q',1}(\mathbb{R}^{2N},\nu_\gamma)}$.

First we consider the case $\gamma > 0$. For sufficiently small $s > 0$, define

$$\theta := \frac{s}{1+\gamma/p}$$

so that $\theta \in (0,1)$ and $p - sp = p(1-\theta)(1+\gamma/p) - \gamma$. Then, for every $R > 0$,

$$\begin{aligned}\iint_{B_R\times B_R} \frac{|u(x)-u(y)|^p}{|x-y|^{N+p-sp}}\,\mathrm{d}x\,\mathrm{d}y &= \iint_{\mathbb{R}^N\times\mathbb{R}^N} (Q_{\gamma/p}u(x,y))^{p(1-\theta)}(|u(x)-u(y)|\mathbb{1}_{B_R\times B_R}(x,y))^{p\theta}\,\mathrm{d}\nu_\gamma \\ &\le \frac{1}{\theta}[(Q_{\gamma/p}u)^{p(1-\theta)}]_{L^{1/(1-\theta),\infty}(\mathbb{R}^{2N},\nu_\gamma)}[|u(x)-u(y)|^{p\theta}]_{L^{1/\theta,1}(B_R\times B_R,\nu_\gamma)}\end{aligned}$$

by (4-2). But

$$[(Q_{\gamma/p}u)^{p(1-\theta)}]_{L^{1/(1-\theta),\infty}(\mathbb{R}^{2N},\nu_\gamma)} = [Q_{\gamma/p}u]^{p(1-\theta)}_{L^{p,\infty}(\mathbb{R}^{2N},\nu_\gamma)}$$

and

$$\begin{aligned}[|u(x)-u(y)|^{p\theta}]_{L^{1/\theta,1}(B_R\times B_R,\nu_\gamma)} &\le (2\|u\|_{L^\infty(\mathbb{R}^N)})^{p\theta}[\mathbb{1}_{B_R\times B_R}]_{L^{1/\theta,1}(\mathbb{R}^N\times\mathbb{R}^N,\nu_\gamma)} \\ &= (2\|u\|_{L^\infty(\mathbb{R}^N)})^{p\theta}\nu_\gamma(B_R\times B_R)^\theta,\end{aligned}$$

from which it follows that

$$s\iint_{B_R\times B_R} \frac{|u(x)-u(y)|^p}{|x-y|^{N+p-sp}}\,\mathrm{d}x\,\mathrm{d}y \le \frac{s}{\theta}[Q_{\gamma/p}u]^{p(1-\theta)}_{L^{p,\infty}(\mathbb{R}^{2N},\nu_\gamma)}(2\|u\|_{L^\infty(\mathbb{R}^N)})^{p\theta}\nu_\gamma(B_R\times B_R)^\theta.$$

Furthermore, since $\gamma > 0$, we have

$$\nu_\gamma(B_R\times B_R) \le |B_R|\int_{B_{2R}} \frac{1}{|h|^{N-\gamma}}\,\mathrm{d}h < \infty.$$

Recall $\theta = s/(1+\gamma/p)$. Thus as $s \to 0^+$, we have

$$\limsup_{s\to 0^+} s \iint_{B_R\times B_R} \frac{|u(x)-u(y)|^p}{|x-y|^{N+p-sp}}\,\mathrm{d}x\,\mathrm{d}y \le \left(1+\frac{\gamma}{p}\right)[Q_{\gamma/p}u]^p_{L^{p,\infty}(\mathbb{R}^{2N},\nu_\gamma)} < \infty.$$

Since this upper bound holds uniformly over all $R > 0$, this concludes the argument for the case $\gamma > 0$.

Next we turn to the case $\gamma \le 0$. We then observe that, for $0 < s < 1$ and every $R > 0$,

$$\begin{aligned}\iint_{B_R\times B_R} &\frac{|u(x)-u(y)|^p}{|x-y|^{N+p-sp}}\,\mathrm{d}x\,\mathrm{d}y \\ &= \iint_{\mathbb{R}^N\times\mathbb{R}^N} (Q_{\gamma/p}u(x,y))^{p(1-s/2)}(|u(x)-u(y)|\,|x-y|^{1-\gamma/p}\mathbb{1}_{B_R\times B_R})^{ps/2}\,\mathrm{d}\nu_\gamma \\ &\le \frac{2}{s}[(Q_{\gamma/p}u)^{p(1-s/2)}]_{L^{1/(1-s/2),\infty}(\mathbb{R}^{2N},\nu_\gamma)}[(|u(x)-u(y)|\,|x-y|^{1-\gamma/p})^{ps/2}]_{L^{2/s,1}(B_R\times B_R,\nu_\gamma)}.\end{aligned}$$

Again

$$[(Q_{\gamma/p}u)^{p(1-s/2)}]_{L^{1/(1-s/2),\infty}(\mathbb{R}^{2N},\nu_\gamma)} = [Q_{\gamma/p}u]^{p(1-s/2)}_{L^{p,\infty}(\mathbb{R}^{2N},\nu_\gamma)}$$

and

$$\begin{aligned}[(|u(x)-u(y)|\,&|x-y|^{1-\gamma/p})^{ps/2}]_{L^{2/s,1}(B_R\times B_R,\nu_\gamma)} \\ &\le (2\|u\|_{L^\infty(\mathbb{R}^N)})^{ps/2}[|x-y|^{(p-\gamma)s/2}]_{L^{2/s,1}(B_R\times B_R,\nu_\gamma)}. \qquad (4\text{-}3)\end{aligned}$$

We will show that

$$\limsup_{s\to 0^+}[|x-y|^{(p-\gamma)s/2}]_{L^{2/s,1}(B_R\times B_R,\nu_\gamma)} \le 1-\frac{\gamma}{p} \qquad (4\text{-}4)$$

when $\gamma \le 0$. We then see that

$$\limsup_{s\to 0^+} s \iint_{B_R\times B_R} \frac{|u(x)-u(y)|^p}{|x-y|^{N+p-sp}}\,\mathrm{d}x\,\mathrm{d}y \le 2\left(1-\frac{\gamma}{p}\right)[Q_{\gamma/p}u]^p_{L^{p,\infty}(\mathbb{R}^{2N},\nu_\gamma)},$$

which concludes the argument in this case since this bound is uniform in $R > 0$.

It remains to prove (4-4) when $\gamma \le 0$. Note that in this case $p-\gamma > 0$, so $|x-y|^{(p-\gamma)s/2} \le (2R)^{(p-\gamma)s/2}$ on $B_R \times B_R$. Thus

$$[|x-y|^{(p-\gamma)s/2}]_{L^{2/s,1}(B_R\times B_R,\nu_\gamma)} = \int_0^{(2R)^{(p-\gamma)s/2}} \nu_\gamma\{(x,y)\in B_R\times B_R : |x-y|^{(p-\gamma)s/2} > \lambda\}^{s/2}\,\mathrm{d}\lambda.$$

If $\gamma < 0$, then

$$\nu_\gamma\{(x,y)\in B_R\times B_R : |x-y|^{(p-\gamma)s/2} > \lambda\} \le |B_R|\int_{|h|>\lambda^{2/(s(p-\gamma))}} \frac{1}{|h|^{N-\gamma}}\,\mathrm{d}h \le \sigma_{N-1}|B_R|\frac{1}{|\gamma|}\lambda^{2\gamma/(s(p-\gamma))},$$

where σ_{N-1} is the surface area of $\mathbb{S}^{N-1}$. Hence in this case,

$$\begin{aligned}[|x-y|^{(p-\gamma)s/2}]_{L^{2/s,1}(B_R\times B_R,\nu_\gamma)} &\le \left(\sigma_{N-1}|B_R|\frac{1}{|\gamma|}\right)^{s/2}\int_0^{(2R)^{(p-\gamma)s/2}} \lambda^{\gamma/(p-\gamma)}\,\mathrm{d}\lambda \\ &= \left(1-\frac{\gamma}{p}\right)\left(\sigma_{N-1}|B_R|\frac{1}{|\gamma|}\right)^{s/2}(2R)^{ps/2}.\end{aligned}$$

(Here we used $\gamma/(p-\gamma) = -1/(1-\gamma/p) \in (-1, 0)$ whenever $\gamma < 0$.) This proves (4-4) when $\gamma < 0$. Next, suppose $\gamma = 0$. Then

$$\begin{aligned}
[|x-y|^{(p-\gamma)s/2}]_{L^{2/s,1}(B_R\times B_R,\nu_\gamma)} &= \int_0^{(2R)^{ps/2}} \nu_0\{(x,y)\in B_R\times B_R : |x-y|^{ps/2} > \lambda\}^{s/2}\,\mathrm{d}\lambda \\
&\le \int_0^{(2R)^{ps/2}} \biggl(|B_R| \int_{\lambda^{2/(sp)}\le|h|\le 2R} \frac{1}{|h|^N}\,\mathrm{d}h\biggr)^{s/2}\mathrm{d}\lambda \\
&= \int_0^{(2R)^{ps/2}} \biggl(|B_R|\omega_{N-1}\frac{2}{ps}\log\biggl(\frac{(2R)^{ps/2}}{\lambda}\biggr)\biggr)^{s/2}\mathrm{d}\lambda \\
&= (2R)^{ps/2}\int_0^1 \biggl(|B_R|\omega_{N-1}\frac{2}{ps}\log\biggl(\frac{1}{\lambda}\biggr)\biggr)^{s/2}\mathrm{d}\lambda,
\end{aligned}$$

which shows (4-4) remains valid when $\gamma = 0$ by the dominated convergence theorem. □

5. Finiteness of $\nu_0(E_{\lambda,0}[u])$ and the Lipschitz norm

In this section we prove Theorem 1.5, which we put in the following more precise form.

Proposition 5.1. *Let u be locally integrable on $\mathbb{R}^N$ and $\nabla u \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$. Then*

$$\nu_0(E_{\lambda,0}[u]) = \begin{cases} 0 & \text{if } \lambda > \|\nabla u\|_\infty, \\ \infty & \text{if } \lambda < \|\nabla u\|_\infty. \end{cases}$$

Proof. First assume $\nabla u \in L^\infty$ and $\lambda > \|\nabla u\|_\infty$. Then for every $h \in \mathbb{R}^N$ we have $|u(x+h)-u(x)|/|h| \le \lambda$ for almost every $x \in \mathbb{R}^N$. This immediately implies $\nu_0(E_{\lambda,0}[u]) = 0$.

For the more substantial part assume $\lambda < \|\nabla u\|_\infty$, where $\|\nabla u\|_\infty$ may be finite or infinite. We need to show that $\nu_0(E_{\lambda,0}[u]) = \infty$. We pick λ_1, λ_2 such that

$$\lambda < \lambda_1 < \lambda_2 < \|\nabla u\|_\infty.$$

Let $B_R = \{x \in \mathbb{R}^N : |x| < R\}$ and assume that $R > 1$ is so large that $\|\nabla u\|_{L^\infty(B_R)} > \lambda_2$. Let $\chi \in C_c^\infty$ such that $\chi(x) = 1$ in a neighborhood of $\overline{B}_{2R}$ and set $u_\circ = \chi u$. Then $\nabla u_\circ = \nabla u$ as integrable functions on B_{2R}. There is a measurable set $F_0 \subset B_R$ of positive measure such that $|\nabla u(x)| > \lambda_2$ for all $x \in F_0$.

Fix $0 < \varepsilon \ll 1 - \lambda_1/\lambda_2$. We now consider the set $\mathfrak{S}_\varepsilon$ of all spherical balls $S \subset \mathbb{S}^{N-1}$ with positive radius and the property that $\langle\theta_1, \theta_2\rangle > 1 - \varepsilon$ for all $\theta_1, \theta_2 \in S$. By pigeonholing there exists a spherical ball $S \in \mathfrak{S}_\varepsilon$ and a Lebesgue measurable subset $F \subset F_0$ such that $\mathcal{L}^N(F) > 0$ and $\nabla u(x)/|\nabla u(x)| \in S$ for all $x \in F$. For the remainder of the argument we fix this spherical ball S; we denote by $\sigma(S)$ its spherical measure.

We first note that, for $|h| \le 1$ and for almost every $|x| \le R$,

$$\frac{u(x+h)-u(x)}{|h|} = \frac{u_\circ(x+h)-u_\circ(x)}{|h|} = \biggl\langle \frac{h}{|h|}, \int_0^1 \nabla u_\circ(x+sh)\,\mathrm{d}s \biggr\rangle. \tag{5-1}$$

Secondly since the translation operator is continuous in the strong operator topology of L^1 we see that there exists $\delta_0 < 1$ such that

$$\|\nabla u_\circ(\cdot + w) - \nabla u_\circ\|_{L^1(\mathbb{R}^N)} < \frac{\mathcal{L}^N(F)(\lambda_1 - \lambda)}{10} \quad \text{for } |w| \le \delta_0. \tag{5-2}$$

In what follows we let $\delta \ll \delta_0$ and set

$$S(\delta, \delta_0) = \left\{ h \in \mathbb{R}^N : \delta \le |h| \le \delta_0,\ \frac{h}{|h|} \in S \right\}.$$

Let

$$\mathcal{E}_0 = \left\{ (x, h) : x \in F,\ h \in S(\delta, \delta_0),\ \frac{|u(x+h) - u(x)|}{|h|} > \lambda \right\}$$

so that $(x, h) \in \mathcal{E}_0$ implies $(x, x+h) \in E_{\lambda,0}[u]$. We then have by (5-1)

$$\begin{aligned} \nu_0(E_{\lambda,0}[u]) \ge \nu_0(\mathcal{E}_0) &= \nu_0\left(\left\{ (x, h) : x \in F,\ h \in S(\delta, \delta_0),\ \left| \left\langle \frac{h}{|h|}, \int_0^1 \nabla u_\circ(x + sh)\, \mathrm{d}s \right\rangle \right| > \lambda \right\} \right) \\ &\ge \nu_0(\mathcal{E}_1) - \nu_0(\mathcal{E}_2), \end{aligned} \tag{5-3}$$

where

$$\begin{aligned} \mathcal{E}_1 &= \left\{ (x, h) : x \in F,\ h \in S(\delta, \delta_0),\ \left| \left\langle \frac{h}{|h|}, \nabla u_\circ(x) \right\rangle \right| > \lambda_1 \right\}, \\ \mathcal{E}_2 &= \left\{ (x, h) : x \in F, h \in S(\delta, \delta_0),\ \int_0^1 |\nabla u_\circ(x + sh) - \nabla u_\circ(x)|\, \mathrm{d}s > \lambda_1 - \lambda \right\}. \end{aligned}$$

Indeed, if $(x, h) \notin \mathcal{E}_0 \cup \mathcal{E}_2$ then

$$\left| \left\langle \frac{h}{|h|}, \nabla u_\circ(x) \right\rangle \right| \le \left| \left\langle \frac{h}{|h|}, \int_0^1 \nabla u_\circ(x + sh)\, \mathrm{d}s \right\rangle \right| + \int_0^1 |\nabla u_\circ(x + sh) - \nabla u_\circ(x)|\, \mathrm{d}s,$$

which is then $\le \lambda_1$, so $(x, h) \notin \mathcal{E}_1$, establishing $\mathcal{E}_1 \subset \mathcal{E}_0 \cup \mathcal{E}_2$ and thus (5-3).

The set $\mathcal{E}_1$ does not change if we replace $u_\circ$ by u in its definition. Since

$$\left\langle \frac{h}{|h|}, \nabla u(x) \right\rangle \ge (1 - \varepsilon)|\nabla u(x)| > (1 - \varepsilon)\lambda_2 > \lambda_1 \quad \text{for } x \in F,\ \frac{h}{|h|} \in S,$$

we get

$$\nu_0(\mathcal{E}_1) \ge \int_F \mathrm{d}x \int_{S(\delta,\delta_0)} \frac{\mathrm{d}h}{|h|^N} = \mathcal{L}^N(F)\sigma(S) \log\left(\frac{\delta_0}{\delta} \right).$$

Moreover, using (5-2) and Chebyshev's inequality we see that

$$\begin{aligned} \nu_0(\mathcal{E}_2) &\le \int_{S(\delta,\delta_0)} \frac{\int_0^1 \|\nabla u_\circ(\cdot + sh) - \nabla u_\circ\|_{L^1(\mathbb{R}^N)}\, \mathrm{d}s}{\lambda_1 - \lambda} \frac{\mathrm{d}h}{|h|^N} \\ &\le \int_{S(\delta,\delta_0)} \frac{\mathcal{L}^N(F)(\lambda_1 - \lambda)/10}{\lambda_1 - \lambda} \frac{\mathrm{d}h}{|h|^N} = \frac{\mathcal{L}^N(F)}{10} \sigma(S) \log\left(\frac{\delta_0}{\delta} \right), \end{aligned}$$

and hence putting pieces together we obtain for $\delta < \delta_0$

$$\nu_0(E_{\lambda,0}[u]) \ge \nu_0(\mathcal{E}_1) - \nu_0(\mathcal{E}_2) > \frac{\mathcal{L}^N(F)}{2}\sigma(S)\log\left(\frac{\delta_0}{\delta}\right).$$

Here $\delta < \delta_0$ was arbitrary and by letting $\delta \to 0$ we conclude that $\nu_0(E_{\lambda,0}[u]) = \infty$. □

We now give a more precise version of Example 1.7.

Lemma 5.2. *Let $\Omega \subset \mathbb{R}^N$ be a bounded domain with Lipschitz boundary and let $u = \mathbb{1}_\Omega$. Then $u \in \mathrm{BV}(\mathbb{R}^N) \setminus \dot{W}^{1,1}(\mathbb{R}^N)$, with*

$$\nu_0(E_{\lambda,0}[u]) \le C_\Omega \times \begin{cases} \log(2/\lambda) & \text{if } \lambda \le 1, \\ \lambda^{-1} & \text{if } \lambda > 1; \end{cases}$$

in particular we have $\sup_{\lambda>0} \lambda\, \nu_0(E_{\lambda,0}[u]) < \infty$.

Proof. Let

$$E(r,\lambda) = \{(x,y) \in E_{\lambda,0}[u] : r \le |x-y| \le 2r\}.$$

We begin with the observation that $r\lambda \le 2$ if $\nu_0(E(r,\lambda)) > 0$. Furthermore, if $(x,y) \in E(r,\lambda)$ for some $y \in \mathbb{R}^N$, then x belongs to the $2r$-neighborhood of $\partial\Omega$. The Lebesgue measure of such a neighborhood is $O(r)$ if $r \le r_0$, where r_0 is some positive constant depending on Ω (because the boundary of a bounded Lipschitz domain can be covered by finitely many Lipschitz graphs, and the $2r$-neighborhood of such graphs can be approximated by a union of $O(r)$ neighborhoods of suitable hyperplanes). Hence for $r \le r_0$ we have $\nu_0(E(r,\lambda)) \le Cr$ if $r \le 2/\lambda$ and $\nu_0(E(r,\lambda)) = 0$ if $r > 2/\lambda$. As a result, if $2/\lambda \le r_0$ we get

$$\nu_0(E_{\lambda,0}[u]) \le \sum_{j\in\mathbb{Z}:2^j\le 2/\lambda} \nu_0(E(2^j,\lambda)) \lesssim \lambda^{-1}$$

and if $2/\lambda > r_0$ we get

$$\nu_0(E_{\lambda,0}[u]) \le \sum_{j\in\mathbb{Z}:2^j\le r_0} \nu_0(E(2^j,\lambda)) + 2\int_\Omega \int_{r_0\le|x-y|\le 2/\lambda} \frac{\mathrm{d}y}{|x-y|^N}\,\mathrm{d}x \lesssim 1 + \log(\lambda^{-1}).$$ □

6. When the upper bound (1-15) fails

In this section we make various constructions demonstrating the failure of (1-15) in the range $-1 \le \gamma < 0$, and give the proof of Theorem 1.8. We first establish:

Proposition 6.1. *Suppose $N \ge 1$ and $-1 \le \gamma < 0$.*

(i) *For every $m > 0$, there exists $u \in C_c^\infty(\mathbb{R}^N)$ such that*

$$\nu_\gamma(E_{1,\gamma}[u]) > m\|\nabla u\|_{L^1(\mathbb{R}^N)}. \tag{6-1}$$

(ii) *There exists $C = C(N,\gamma) > 0$ and $p_0 = p_0(N,\gamma) > 1$ such that, for all $1 < p < p_0$,*

$$\sup_{\substack{u\in C_c^\infty(\mathbb{R}^N)\\ \|\nabla u\|_{L^p}\le 1}} \nu_\gamma(E_{1,\gamma/p}[u]) \ge C\frac{p}{p-1}. \tag{6-2}$$

6A. ***Proof of Proposition 6.1: the case $\gamma = -1$.*** Here we may choose, for $m > 1$,

$$v_m = 2\,\eta_m * \mathbb{1}_{B_1} \in C_c^\infty(\mathbb{R}^N), \tag{6-3}$$

where $\eta_m(x) := 2^{mN}\eta(2^m x)$ for some nonnegative, radially decreasing $\eta \in C_c^\infty(B_1)$, with $\int_{\mathbb{R}^N}\eta = 1$. Then when $1 \le p < \infty$ and $m \le p' = p/(p-1)$ (which is no restriction on m if $p = 1$), we have $\|\nabla v_m\|_p \lesssim 2^{m/p'} \lesssim 1$, while $E_{1,-1/p}[v_m] \supseteq \{|x| \le 1-2^{-m},\ 1+2^{-m} \le |y| \le 2\}$ (because for (x, y) in the latter set, $|v_m(x) - v_m(y)| = 2$ and $|x-y|^{1-1/p} \le 2^{1-1/p}$, which means $|Q_{-1/p}v_m(x, y)| \ge 2/2^{1-1/p} = 2^{1/p} > 1$). Hence

$$\begin{aligned}\nu_{-1}(E_{1,-1/p}[v_m]) &\ge \int_{|x|\le 1-2^{-m}} \int_{1+2^{-m}\le|y|\le 2} |x-y|^{-1-N}\,\mathrm{d}x\,\mathrm{d}y \\ &\ge c_N \int_{|x|\le 1-2^{-m}} (1+2^{-m}-|x|)^{-1} - (2-|x|)^{-1}\,\mathrm{d}x \ge c_N' m.\end{aligned}$$

This proves both (i) and (ii) of Proposition 6.1 in the case $\gamma = -1$. □

6B. ***The case $-1 < \gamma < 0$: examples of Cantor–Lebesgue-type on the real line.*** We now discuss some examples related to self-similar Cantor sets of dimension $\beta = 1+\gamma$. Recall the definition of ν_γ, Q_γ in (1-4), (1-5) and observe the behavior under dilations:

$$\nu_\gamma(tE) = t^{1+\gamma}\nu_\gamma(E). \tag{6-4}$$

We have:

Lemma 6.2. *Let $-1 < \gamma < 0$. There exist constants $c_\gamma > 0$, $C_\gamma > 0$, and a sequence of functions $g_m \in C^\infty(\mathbb{R})$, with $g_m(x) = 0$ for $x \le 0$ and $g_m(x) = 1$ for $x \ge 1$, such that, for all $1 \le p < \infty$,*

$$\|g_m'\|_p \le c_\gamma 2^{m|\gamma|/(1+\gamma)(1-1/p)} \tag{6-5}$$

and if

$$m - 1 \le \frac{\gamma+1}{|\gamma|}\frac{p}{p-1},$$

then

$$\nu_\gamma\Big(\Big\{(x, y) \in [0, 1]^2 : |Q_{\gamma/p}g_m(x, y)| > \frac{1}{4}\Big\}\Big) \ge \frac{m}{C_\gamma}. \tag{6-6}$$

Proof. For $-1 < \gamma < 0$ let

$$\rho = 2^{-1/(1+\gamma)} \tag{6-7}$$

so that $0 < \rho < \frac{1}{2}$. We construct g_m such that its derivative is supported on the m-th step of the construction of symmetric Cantor sets of dimension $\beta = 1+\gamma = \log 2/\log(1/\rho)$, with an equal variation on each of its 2^m components [Mattila 2015, Chapter 8.1].

Let $g_0 \in C^\infty(\mathbb{R})$ be such that $0 \le g_0 \le 1$, $g_0(x) = 0$ for $x \le \rho$ and $g_0(x) = 1$ for $x \ge 1-\rho$. Set, for $m \in \mathbb{N}$,

$$g_{m+1}(x) := \frac{1}{2}g_m\Big(\frac{x}{\rho}\Big) + \frac{1}{2}g_m\Big(1 - \frac{1-x}{\rho}\Big).$$

Since $\rho < \frac{1}{2}$, we have, for $p \in [1, \infty)$, $\|g_{m+1}'\|_{L^p(\mathbb{R})}^p = 2 \times (2\rho)^{-p}\rho\|g_m'\|_{L^p(\mathbb{R})}^p$, and thus

$$\|g_m'\|_{L^p(\mathbb{R})} = (2\rho)^{(1/p-1)m}\|g_0'\|_{L^p(\mathbb{R})} = 2^{(1-1/p)m|\gamma|/(\gamma+1)}\|g_0'\|_{L^p(\mathbb{R})}.$$

Fix now $1 \le p < \infty$, and for $m \in \mathbb{N}$, $\lambda > 0$ define

$$A_{m,\lambda} := \nu_\gamma(\{(x, y) \in [0, 1]^2 : |Q_{\gamma/p} g_m(x, y)| > \lambda\}).$$

Our goal is to estimate $A_{m,1/4}$, which we do by deriving a recursive estimate for $A_{m,\lambda}$. We have the decomposition

$$\begin{aligned} A_{m+1,\lambda} \ge\ & \nu_\gamma(\{(x, y) \in [0, \rho]^2 : |Q_{\gamma/p} g_{m+1}(x, y)| > \lambda\}) \\ &+ \nu_\gamma(\{(x, y) \in [1-\rho, 1]^2 : |Q_{\gamma/p} g_{m+1}(x, y)| > \lambda\}) \\ &+ \nu_\gamma(\{(x, y) \in [0, \rho] \times [1-\rho, 1] : |Q_{\gamma/p} g_{m+1}(x, y)| > \lambda\}). \end{aligned} \tag{6-8}$$

Using the definition of g_{m+1}, (6-7) and (6-4), we compute the first term in the right-hand side of (6-8) as

$$\begin{aligned} &\nu_\gamma(\{(x, y) \in [0, \rho]^2 : |Q_\gamma g_{m+1}(x, y)| > \lambda\}) \\ &\qquad = \nu_\gamma(\{(\rho w, \rho z) : (w, z) \in [0, 1]^2,\ |Q_\gamma g_m(w, z)| > 2\rho^{1+\gamma/p}\lambda\}) \\ &\qquad = \rho^{\gamma+1} \nu_\gamma(\{(w, z) \in [0, 1]^2 : |Q_\gamma g_m(w, z)| > 2^{|\gamma|/(p'(\gamma+1))}\lambda\}) = \tfrac{1}{2} A_{m,s\lambda}, \end{aligned} \tag{6-9}$$

where $s := 2\rho^{1+\gamma/p} = 2^{|\gamma|/(p'(\gamma+1))}$, and similarly the second term as

$$\nu_\gamma\big(\big\{(x, y) \in [1-\rho, 1]^2 : |Q_\gamma g_{m+1}(x, y)| > \tfrac{1}{2}\big\}\big) = \tfrac{1}{2} A_{m,s\lambda}. \tag{6-10}$$

Thus

$$A_{m+1,\lambda} \ge A_{m,s\lambda} + \nu_\gamma(\{(x, y) \in [0, \rho] \times [1-\rho, 1] : |Q_{\gamma/p} g_{m+1}(x, y)| > \lambda\}),$$

which iterates to give

$$A_{m,1/4} \ge A_{0,s^m/4} + \sum_{j=1}^{m} \nu_\gamma\big(\big\{(x, y) \in [0, \rho] \times [1-\rho, 1] : |Q_{\gamma/p} g_j(x, y)| > \tfrac{1}{4} s^{m-j}\big\}\big).$$

We drop the first term, and note that as long as

$$m - 1 \le \frac{\gamma+1}{|\gamma|} \frac{p}{p-1},$$

we have $\frac{1}{4} s^{m-j} \le \frac{1}{2}$ for all $j = 1, \dots, m$. Moreover, for every $x \in [0, \rho^2] \times [1-\rho^2, 1]$ and every $j \ge 1$, we have $g_j(x) \le \frac{1}{4}$ and $g_j(y) \ge \frac{3}{4}$, so $|Q_{\gamma/p} g_j(x, y)| > \frac{1}{2}$. Thus we obtain the desired conclusion

$$A_{m,1/4} \ge m \nu_\gamma([0, \rho^2] \times [1-\rho^2, 1]) = \frac{m}{C_\gamma}. \qquad \square$$

6C. *Conclusion of the proof of Proposition 6.1.* We continue with the case $-1 < \gamma < 0$. Let $\eta_1 \in C_c^\infty(\mathbb{R})$ supported in $(-1, 2)$ such that $\eta_1(s) = 1$ on $\left(-\frac{1}{2}, \frac{3}{2}\right)$ and $0 \le \eta_1(s) \le 1$ for all $s \in \mathbb{R}$.

We split $x = (x_1, x')$ with $x' \in \mathbb{R}^{N-1}$, where the variable x' should simply be dropped in the case $N = 1$. Set $\eta(x) = \prod_{i=1}^{N} \eta_1(x_i)$ and define

$$u_m(x_1, x') = 16 g_m(x_1) \eta(x), \tag{6-11}$$

where g_m is as in Lemma 6.2. Then $u_m \in C_c^\infty(\mathbb{R}^N)$, and if $1 \le p < \infty$ and

$$m - 1 \le \frac{\gamma+1}{|\gamma|} \frac{p}{p-1},$$

we have $\|\nabla u_m\|_p \lesssim 1$. Both parts of Proposition 6.1 will follow, if we can prove that under the same hypotheses on p and m, we have

$$\nu_\gamma(E_{1,\gamma/p}[u_m]) \geq c(N,\gamma)m - C(N,\gamma)^p. \tag{6-12}$$

We aim to reduce to the one-dimensional situation in Lemma 6.2 and split $Q_{\gamma/p}u_m(x,y)$ as

$$Q_{\gamma/p}u_m(x,y) = 16\eta(x)\frac{g_m(x_1)-g_m(y_1)}{|x-y|^{1+\gamma/p}} + 16 g_m(y_1)\frac{\eta(x)-\eta(y)}{|x-y|^{1+\gamma/p}} = I_m(x,y) + II_m(x,y),$$

so that

$$\begin{aligned}\nu_\gamma(E_{1,\gamma/p}[u_m]) &\geq \iint_{\substack{x_1,y_1\in[0,1]\\ |I_m(x,y)+II_m(x,y)|>1}} |x-y|^{\gamma-N}\,\mathrm{d}x\,\mathrm{d}y \\ &\geq \iint_{\substack{x\in[0,1]^N,\,y_1\in[0,1]\\ |x_1-y_1|\geq|x'-y'|\\ |I_m(x,y)|>2}} |x-y|^{\gamma-N}\,\mathrm{d}x\,\mathrm{d}y - \iint_{|II_m(x,y)|>1} |x-y|^{\gamma-N}\,\mathrm{d}x\,\mathrm{d}y. \end{aligned} \tag{6-13}$$

Clearly if B_2 is the ball in $\mathbb{R}^N$ of radius 2 centered at the origin then

$$|II_m(x,y)| \leq c_N|x-y|^{-\gamma/p}(\mathbb{1}_{B_2}(x)+\mathbb{1}_{B_2}(y)),$$

and it follows immediately (since $-\gamma>0$) that

$$\iint_{|II_m(x,y)|>1} |x-y|^{\gamma-N}\,\mathrm{d}x\,\mathrm{d}y \leq |\gamma|^{-1}C(N)^p.$$

For the first term in (6-13), we prove a lower bound and estimate by integrating in y'

$$\begin{aligned}\iint_{\substack{x\in[0,1]^N,\,y_1\in[0,1]\\ |x_1-y_1|\geq|x'-y'|\\ |I_m(x,y)|>2}} |x-y|^{\gamma-N}\,\mathrm{d}x\,\mathrm{d}y &\geq \iint_{\substack{x\in[0,1]^N,\,y_1\in[0,1]\\ |x_1-y_1|\geq|x'-y'|\\ |16g_m(x_1)-16g_m(y_1)|/|x_1-y_1|^{1+\gamma/p}>4}} |x-y|^{\gamma-N}\,\mathrm{d}x\,\mathrm{d}y \\ &\geq c_N \iint_{\substack{x_1,y_1\in[0,1]\\ |Q_{\gamma/p}g_m(x_1,y_1)|>1/4}} |x_1-y_1|^{\gamma-1}\,\mathrm{d}x_1\,\mathrm{d}y_1,\end{aligned}$$

but by Lemma 6.2 the last expression is bounded below for large m by $c_N m/C_\gamma$ under our hypothesis on m. This concludes the proof of (6-12). □

For later purposes, note the inequality (6-13) (with $p=1$) and the argument that follows proved also that for all sufficiently large $m > m(N,\gamma)$, we have

$$\nu_\gamma(E_{1,\gamma}[u_m]\cap([0,1]\times\mathbb{R}^{N-1})^2) \geq c(N,\gamma)m. \tag{6-14}$$

6D. ***Examples related to Theorems 1.1 and 1.8.*** We now consider the limit (1-8) in the range $-1\leq\gamma<0$ and provide counterexamples for cases where u is no longer required to be a C_c^∞ function. The following proposition covers part (i) of Theorem 1.8.

Proposition 6.3. *Let $-1\leq\gamma<0$. Let $s\mapsto\omega(s)$ be any decreasing function on $[0,\infty)$, with $\omega(0)\leq 1$ and $\omega(s)>0$ for all $s\geq 0$. Then there exists a C^∞ function $u\in\dot W^{1,1}(\mathbb{R}^N)$ such that*

$$|u(x)| \leq C\omega(|x|) \quad \textit{for all } x\in\mathbb{R}^N \tag{6-15}$$

and

$$\lim_{\lambda\searrow 0}\lambda\nu_\gamma(E_{\lambda,\gamma}[u]) = \infty. \tag{6-16}$$

Proof. We consider the case $-1 < \gamma < 0$. Let $u_m \in C_c^\infty(\mathbb{R}^N)$ be as in (6-11) and define

$$f_m(x) = u_m(x_1 - 2, x') \tag{6-17}$$

so that $f_m(x) = 0$ if $x_1 \notin [1, 4]$. Let, for $n \in \mathbb{N}$,

$$R_n = 2^{2n}, \quad \lambda_n = R_n^{-(N+\gamma)}\omega(R_{n+1}), \quad m(n) \geq 4\frac{\lambda_n}{\lambda_{n+1}}\omega(R_{n+1})^{-1}n^3. \tag{6-18}$$

We also assume $m(n) > m(N, \gamma)$ so that by (6-14) in Section 6C,

$$\nu_\gamma(\{(x, y) : x_1, y_1 \in [2, 3], |Q_\gamma f_{m(n)}(x, y)| > 1\}) \geq c(N, \gamma)m(n) \tag{6-19}$$

for all $n \in \mathbb{N}$. Finally let

$$u(x) = \sum_{n=2}^{\infty} \frac{\omega(R_{n+1})}{R_n^{N-1}n^2} f_{m(n)}\left(\frac{x}{R_n}\right). \tag{6-20}$$

Since $\|f_m\|_{\dot{W}^{1,1}} \leq C$, and ω is bounded, it is easy to see that the sum converges in $\dot{W}^{1,1}(\mathbb{R}^N)$, and that u is in $\dot{W}^{1,1}(\mathbb{R}^N)$. Also, the supports of $f_{m(n)}(R_n^{-1}\,\cdot)$, namely $[R_n, 4R_n] \times [-4R_n, 4R_n]^{N-1}$, are disjoint as n varies, so clearly $u \in C^\infty(\mathbb{R}^N)$. Since $\|f_m\|_{L^\infty} \leq C$, we have

$$|u(x)| \leq \omega(R_{n+1})R_n^{-(N-1)}n^{-2} \quad \text{for } |x| \geq R_n,$$

so $|u(x)| \leq C'|x|^{-N+1}\omega(|x|)$ for $|x| \geq 2$. In particular $|u(x)| \leq C\omega(|x|)$.

For $\lambda \in ((n+1)^{-2}\lambda_{n+1}, n^{-2}\lambda_n]$ we estimate

$$\lambda\nu_\gamma(E_{\lambda,\gamma}[u]) \geq (n+1)^{-2}\lambda_{n+1}\nu_\gamma(E_{n^{-2}\lambda_n,\gamma}[u]) \geq \frac{\lambda_{n+1}}{4\lambda_n}\, n^{-2}\lambda_n\nu_\gamma(\mathcal{E}_n),$$

where $\mathcal{E}_n := E_{n^{-2}\lambda_n,\gamma}[u] \cap ([2R_n, 3R_n] \times \mathbb{R}^{N-1})^2$. Moreover, for $(x, y) \in ([2R_n, 3R_n] \times \mathbb{R}^{N-1})^2$, we have

$$u(x) - u(y) = R_n^{1-N}n^{-2}\omega(R_{n+1})(f_{m(n)}(R_n^{-1}x) - f_{m(n)}(R_n^{-1}y)),$$

so

$$|Q_\gamma u(x, y)| > n^{-2}\lambda_n \quad \Longleftrightarrow \quad \frac{|f_{m(n)}(R_n^{-1}x) - f_{m(n)}(R_n^{-1}y)|}{|R_n^{-1}x - R_n^{-1}y|^{1+\gamma}} > \frac{R_n^{N+\gamma}}{\omega(R_{n+1})}\lambda_n = 1,$$

where the last equality follows from (6-18). Hence rescaling using (6-4) yields

$$\begin{aligned} n^{-2}\lambda_n\nu_\gamma(\mathcal{E}_n) &= n^{-2}\lambda_n R_n^{\gamma+N}\nu_\gamma(\{(x, y) : x_1, y_1 \in [2, 3], |Q_\gamma f_{m(n)}(x, y)| > 1\}) \\ &\geq c(N, \gamma)m(n)\omega(R_{n+1})n^{-2}, \end{aligned} \tag{6-21}$$

with $c(N, \gamma) > 0$, by (6-19). Thus we have shown

$$\inf_{\lambda \in ((n+1)^{-2}\lambda_{n+1}, n^{-2}\lambda_n]} \lambda\nu_\gamma(E_{\lambda,\gamma}[u]) \geq c(N, \gamma)\frac{\lambda_{n+1}}{4\lambda_n}\omega(R_{n+1})m(n)n^{-2} \geq c(N, \gamma)n,$$

where for the last inequality we have used our assumption (6-18) on $m(n)$. The assertion follows for $-1 < \gamma < 0$.

Finally consider the case $\gamma = -1$. We now choose v_m as in (6-3) and

$$R_n = 2^{2n}, \quad \lambda_n = R_n^{-(N-1)}\omega(R_{n+1}), \quad m(n) \geq 4\frac{\lambda_n}{\lambda_{n+1}}\frac{n^3}{\omega(R_{n+1})}. \tag{6-22}$$

In analogy to (6-20) we now use

$$u(x) = \sum_{n=2}^{\infty} \frac{\omega(R_{n+1})}{R_n^{N-1}n^2} v_{m(n)}\left(\frac{x}{R_n}\right). \tag{6-23}$$

Since ω is bounded it is immediate that $u \in \dot{W}^{1,1}(\mathbb{R}^N)$ and also that $|u(x)| \lesssim \omega(|x|)$. We need to check that $\lambda\nu_{-1}(E_{\lambda,-1}[u]) \to \infty$ as $\lambda \to 0^+$. If $|x| \leq R_n(1-2^{m(n)})$ and $|y| \geq R_n(1+2^{m(n)})$, then

$$u(x) - u(y) \geq \frac{\omega(R_{n+1})}{R_n^{N-1}n^2} v_{m(n)}\left(\frac{x}{R_n}\right) = 2\frac{\omega(R_{n+1})}{R_n^{N-1}n^2} = 2n^{-2}\lambda_n > n^{-2}\lambda_n,$$

so $(x, y) \in E_{n^{-2}\lambda_n,-1}[u]$. Hence we get

$$\begin{aligned} n^{-2}\lambda_n \nu_{-1}(E_{n^{-2}\lambda_n,-1}[u]) &\geq n^{-2}\lambda_n \iint_{\substack{|x|\leq R_n(1-2^{m(n)}) \\ |y|\geq R_n(1+2^{m(n)})}} |x-y|^{-1-N}\,\mathrm{d}x\,\mathrm{d}y \\ &\geq n^{-2}\lambda_n R_n^{N-1} \iint_{\substack{|x|\leq 1-2^{m(n)} \\ |y|\geq 1+2^{m(n)}}} |x-y|^{-1-N}\,\mathrm{d}x\,\mathrm{d}y \\ &\geq c_N m(n)\omega(R_{n+1})n^{-2} \end{aligned}$$

(using (6-22) in the last inequality). This, together with our assumption on $m(n)$, implies that

$$\inf_{\lambda\in((n+1)^{-2}\lambda_{n+1},\, n^{-2}\lambda_n]} \lambda\nu_{-1}(E_{\lambda,-1}[u]) \geq c_N n \to \infty$$

when $n \to \infty$, as desired. □

The next proposition is relevant for part (ii) of Theorem 1.8.

Proposition 6.4. *Suppose $-1 \leq \gamma < 0$. Then there exists a compactly supported $u \in W^{1,1}(\mathbb{R}^N)$ such that u is C^∞ for $x \neq 0$,*

$$|u(x)| \leq \frac{C}{|x|^{N-1}[\log(2+|x|^{-1})]^2} \tag{6-24}$$

and

$$\lim_{\lambda\searrow 0} \lambda\nu_\gamma(E_{\lambda,\gamma}[u]) = \infty. \tag{6-25}$$

If in addition $N \geq 2$ or $-1 < \gamma < 0$ there exists u with the above properties and

$$\nu_\gamma(E_{\lambda,\gamma}[u]) = \infty \quad \textit{for all } \lambda > 0. \tag{6-26}$$

Proof. Consider first the case $-1 < \gamma < 0$. We choose for $n \in \mathbb{N}$

$$R_n = 2^{-2n}, \quad m(n) \geq 2^{2^n}, \tag{6-27}$$

and with these choices of R_n and $m(n)$ and f_m as in (6-17) and (6-11) we define again

$$u(x) = \sum_{n=2}^{\infty} \frac{1}{n^2 R_n^{N-1}} f_{m(n)}\left(\frac{x}{R_n}\right).$$

The sum converges in $W^{1,1}$ to a function supported in $[-4,4]^N$. We have $|u(x)| \le C 2^{2n(N-1)} n^{-2}$ for $0 < x_1 \le 2^{-2n}$; moreover, $|x'| \lesssim |x_1|$ on the support of u. This implies $|u(x)| \le C'[|x|^{1-N} \log(1/|x|)]^{-2}$ for small x. Also, because of the choices of R_n, we see that u is smooth away from 0.

Fix $\lambda > 0$. Since $\lim_{n\to\infty} R_n^{N+\gamma} n^2 = 0$, we may choose n_0 such that

$$\lambda R_n^{N+\gamma} n^2 \le 1 \quad \text{for all } n \ge n_0. \tag{6-28}$$

Now $\nu_\gamma(E_{\lambda,\gamma}[u]) \ge \nu_\gamma(E_{\lambda,\gamma}[u] \cap ([2R_n, 3R_n] \times \mathbb{R}^{N-1})^2)$, and again $f_{m(n)}(R_n^{-1}\,\cdot)$ is supported in $\mathcal{R}(n) = [R_n, 4R_n] \times [-4R_n, 4R_n]^{N-1}$. Hence by the same rescaling argument as in (6-21), we obtain

$$\nu_\gamma(E_{\lambda,\gamma}[u]) \ge R_n^{N+\gamma} \nu_\gamma(\{(x, y) : x_1, y_1 \in [2,3], |Q_\gamma f_{m(n)}(x, y)| > \lambda R_n^{N+\gamma} n^2\}).$$

If $n \ge n_0$ then this gives

$$\begin{aligned}\nu_\gamma(E_{\lambda,\gamma}[u]) &\ge R_n^{N+\gamma} \nu_\gamma(\{(x, y) : x_1, y_1 \in [2,3], |Q_\gamma f_{m(n)}(x, y)| > 1\}) \\ &\ge c(N,\gamma) m(n) R_n^{N+\gamma}\end{aligned}$$

by (6-19). Since $\lim_{n\to\infty} m(n) R_n^{N+\gamma} = \infty$, by (6-27) we conclude $\nu_\gamma(E_{\lambda,\gamma}[u]) = \infty$.

For the case $\gamma = -1$ and $N \ge 2$, define u as in (6-23) but with the choice of the parameters R_n, $m(n)$ as in (6-27) to obtain a compactly supported $u \in W^{1,1}$ satisfying (6-24). We now fix $\lambda > 0$ and note that when $N \ge 2$ we have $\lambda R_n^{N-1} n^2 \to 0$ as $n \to \infty$. The above calculation gives $\nu_{-1}(E_{\lambda,-1}[u]) \ge c(N) m(n) R_n^{N-1}$ provided that $\lambda R_n^{N-1} n^2 \le 1$ and thus the conclusion $\nu_{-1}(E_{\lambda,-1}[u]) = \infty$.

Finally, clearly (6-25) follows from (6-26), and the latter was proved if $-1 < \gamma < 0$ or $N \ge 2$. It remains to consider the case $N = 1$, $\gamma = -1$. We define u as in the previous paragraph. The above calculation shows that $\nu_{-1}(E_{\lambda,-1}[u]) \ge c m(n)$ provided that $\lambda < 1/n^2$ which establishes (6-25) in this last case. □

The case $N = 1$, $\gamma = -1$ plays a special role. The following lemma shows that the conclusion (6-26) in Proposition 6.4 fails in this case.

Lemma 6.5. *Let $u \in \dot W^{1,1}(\mathbb{R})$ be compactly supported. Then $\nu_{-1}(E_{\lambda,-1}[u]) < \infty$ for all $\lambda > 0$.*

Proof. Let $u \in \dot W^{1,1}(\mathbb{R})$ be compactly supported in $[-R, R]$. Then given any $\lambda \in (0,1)$, there exists $\delta(\lambda) > 0$ such that $\int_I |u'| \le \lambda/2$ for every interval $I \subset \mathbb{R}$ with length $\le \delta(\lambda)$. As a result, u is uniformly continuous on $\mathbb{R}$, with $\sup_{x\in\mathbb{R}} |u(x+h) - u(x)| \le \lambda/2$ for $|h| \le \delta(\lambda)$. Thus

$$\begin{aligned}\nu_{-1}(E_{\lambda,-1}[u]) &= 2\int_{-\infty}^{\infty} \int_{\substack{h>0 \\ |u(x+h)-u(x)|>\lambda}} \frac{\mathrm{d}h}{h^2}\,\mathrm{d}x \\ &\le \int_{-2R}^{2R} \int_{\delta(\lambda)}^{\infty} \frac{\mathrm{d}h}{h^2}\,\mathrm{d}x + \int_{\mathbb{R}\setminus[-2R,2R]} \int_{|x|-R}^{|x|+R} \frac{\mathrm{d}h}{h^2}\,\mathrm{d}x \\ &\le 4R(\delta(\lambda))^{-1} + 4.\end{aligned}$$

□

6E. ***Generic failure in $W^{1,1}$ for the case $-1 \le \gamma < 0$.***

Proposition 6.6. *Let $-1 \le \gamma < 0$, $N \ge 2$ or $-1 < \gamma < 0$, $N \ge 1$. Let*

$$\mathcal{V} = \{f \in W^{1,1}(\mathbb{R}^N) : \nu_\gamma(E_{\lambda,\gamma}[f]) < \infty \text{ for some } > 0.\} \tag{6-29}$$

Then $\mathcal{V}$ is of first category in $W^{1,1}(\mathbb{R}^N)$, in the sense of Baire.

Let

$$\begin{aligned} U_k &= \{(x, y) \in \mathbb{R}^{2N} : 2^{k-1} \le |x-y| \le 2^k\}, \\ \Omega_\ell &= \bigcup_{k=1-\ell}^{\ell} U_k. \end{aligned} \tag{6-30}$$

For the proof of Proposition 6.6 we use an elementary estimate for the intersections $E_{\lambda,\gamma}[u] \cap \Omega_\ell$.

Lemma 6.7. *For all $\gamma \in \mathbb{R}$, $u \in W^{1,1}(\mathbb{R}^N)$, $\ell > 0$ and Ω_ℓ as in* (6-30),

$$\sup_{\lambda>0} \lambda \nu_\gamma(E_{\lambda,\gamma}[u] \cap \Omega_\ell) \le C(N, \gamma)\ell \|\nabla u\|_1.$$

Proof. For $u \in C^1$ we use the Lusin–Lipschitz inequality (2-2) to see that

$$\begin{aligned} \lambda \iint_{E_{\lambda,\gamma}[u]\cap U_k} |x-y|^{\gamma-N}\,\mathrm{d}x\,\mathrm{d}y &\le C(\gamma)\lambda 2^{k\gamma} \mathcal{L}^N\{x \in \mathbb{R}^N : M(|\nabla u|)(x) > c2^{k\gamma}\lambda\} \\ &\le C(N, \gamma)\|\nabla u\|_1 \end{aligned}$$

by the Hardy–Littlewood maximal inequality. Now sum in $1-\ell \le k \le \ell$. The extension to general $u \in W^{1,1}$ is obtained as in the limiting argument of Section 2C. □

Proof of Proposition 6.6. Let, for $m \in \mathbb{N}$ and $j \in \mathbb{Z}$,

$$\mathcal{V}(m, j) = \{u \in W^{1,1}(\mathbb{R}^N) : \nu_\gamma(E_{\lambda,\gamma}[u]) \le m \text{ for all } \lambda > 2^j\}.$$

Since $\lambda \mapsto \nu_\gamma(E_{\lambda,\gamma}[u])$ is decreasing, we see that $\mathcal{V}$ is contained in $\bigcup_{m\ge1}\bigcup_{j\in\mathbb{Z}} \mathcal{V}(m, j)$. To show that $\mathcal{V}$ is of first category in $W^{1,1}(\mathbb{R}^N)$, we need to show that for every $m \in \mathbb{N}$, $j \in \mathbb{Z}$, the set $\mathcal{V}(m, j)$ is nowhere dense.

We first show that $\mathcal{V}(m, j)$ is closed in $W^{1,1}(\mathbb{R}^N)$. Let $u_n \in \mathcal{V}(m, j)$ and $u \in W^{1,1}(\mathbb{R}^N)$ such that $\lim_{n\to\infty} \|u-u_n\|_{W^{1,1}(\mathbb{R}^N)} = 0$. It suffices to show that given $\varepsilon > 0$, we have $\nu_\gamma(E_{\lambda,\gamma}[u]) \le m+\varepsilon$ for all $\lambda > 2^j$. By the monotone convergence theorem, we have

$$\lim_{\ell\to\infty} \nu_\gamma(E_{\lambda,\gamma}[u] \cap \Omega_\ell) = \nu_\gamma(E_{\lambda,\gamma}[u]),$$

and it suffices to verify that

$$\nu_\gamma(E_{\lambda,\gamma}[u] \cap \Omega_\ell) \le m+\varepsilon \quad \text{for } \lambda > 2^j, \tag{6-31}$$

for all $\ell \in \mathbb{N}$. Now let $\delta > 0$ such that $(1-\delta)\lambda > 2^j$. Then

$$\nu_\gamma(E_{\lambda,\gamma}[u] \cap \Omega_\ell) \le \nu_\gamma(E_{(1-\delta)\lambda,\gamma}[u_n] \cap \Omega_\ell) + \nu_\gamma(E_{\delta\lambda,\gamma}[u-u_n] \cap \Omega_\ell)$$

and using that $u_n \in \mathcal{V}(m, j)$ together with $(1-\delta)\lambda > 2^j$, and Lemma 6.7, we see that for $\lambda > 2^j$

$$\nu_\gamma(E_{\lambda,\gamma}[u] \cap \Omega_\ell) \le m + C(N, \gamma)\ell \frac{1+\delta}{\delta 2^j} \|\nabla(u_n - u)\|_1.$$

Since $\delta > 0$ was arbitrary and since $\|\nabla(u_n - u)\|_{L^1(\mathbb{R}^N)} \to 0$ by assumption, we obtain (6-31).

To show that the closed set $\mathcal{V}(m, j)$ is nowhere dense when $-1 \le \gamma < 0$, we need to verify that for every $u \in \mathcal{V}(m, j)$ and $\varepsilon_1 > 0$ there exists $f \in W^{1,1}(\mathbb{R}^N)$ such that $\|f - u\|_{W^{1,1}(\mathbb{R}^N)} < \varepsilon_1$ and $f \notin \mathcal{V}(m, j)$. To see this we use Proposition 6.4, according to which there exists a compactly supported $W^{1,1}$ function f_0 for which $\nu_\gamma(E_{\lambda,\gamma}[f_0]) = \infty$ for all $\lambda > 0$. It is then clear that

$$f = u + \frac{\varepsilon_1}{2} \frac{f_0}{\|f_0\|_{W^{1,1}}}$$

satisfies $\|f - u\|_{W^{1,1}} \le \varepsilon_1/2$ and also,

$$\nu_\gamma(E_{\lambda,\gamma}[f]) \ge \nu_\gamma\left(E_{2\lambda,\gamma}\left[\frac{\varepsilon_1}{2} \frac{f_0}{\|f_0\|_{W^{1,1}}}\right]\right) - \nu_\gamma(E_{\lambda,\gamma}[u]) = \infty$$

for every $\lambda > 2^j$, for all $j \in \mathbb{Z}$. The proposition is proved. □

To include a result of generic failure of the limiting relation in the case $N = 1$, $\gamma = -1$ we give

Proposition 6.8. *Let* $-1 \le \gamma < 0$. *Let*

$$\mathcal{W} = \left\{ f \in W^{1,1}(\mathbb{R}) : \limsup_{R \to 0} \sup_{\lambda > R} R\nu_\gamma(E_{\lambda,\gamma}[f]) < \infty \right\}.$$

Then $\mathcal{W}$ is of first category in $W^{1,1}$, in the sense of Baire.

Proof. Clearly $\mathcal{W} \subset \mathcal{V}$, where $\mathcal{V}$ is defined in (6-29). We define

$$\mathcal{W}(m, j) = \left\{ u \in W^{1,1}(\mathbb{R}) : \sup_{0 < R \le 2^{-j}} \sup_{\lambda > R} R\nu_\gamma(E_{\lambda,\gamma}[u]) \le m \right\}$$

and note that

$$\mathcal{W} \subset \cup_{j \ge 1} \bigcup_{m \ge 1} \mathcal{W}(m, j). \tag{6-32}$$

The arguments in the proof of Proposition 6.6 that were used to show that the sets $\mathcal{V}(m, j)$ are closed in $W^{1,1}(\mathbb{R}^N)$ also show that the sets $\mathcal{W}(m, j)$ are closed in $W^{1,1}(\mathbb{R})$.

Let $u \in \mathcal{W}(m, j)$, and let $\varepsilon_1 > 0$. By Proposition 6.4 there is $f_0 \in W^{1,1}(\mathbb{R})$ such that

$$\lim_{\lambda \searrow 0} \lambda \nu_\gamma(E_{\lambda,\gamma}[f_0]) = \infty.$$

We may normalize so that $\|f_0\|_{W^{1,1}(\mathbb{R})} = 1$. Pick $R \in (0, 2^{-j}]$ so that $\lambda\nu_\gamma(E_{\lambda,\gamma}[f_0]) > 16m/\varepsilon_1$ for $\lambda \le 8R/\varepsilon_1$. Let $f = u + (\varepsilon_1/2) f_0$ so that $\|f - u\|_{W^{1,1}(\mathbb{R})} \le \varepsilon_1/2$. Moreover if $\lambda = 2R$, then $\lambda > R$ and

$$\begin{aligned} R\nu_\gamma(E_{\lambda,\gamma}[f]) &\ge R\nu_\gamma\left(E_{2\lambda,\gamma}\left[\frac{\varepsilon_1}{2} f_0\right]\right) - R\nu_\gamma(E_{\lambda,\gamma}[u]) \\ &= \frac{\varepsilon_1}{8} \frac{8R}{\varepsilon_1} \nu_\gamma(E_{8R/\varepsilon_1,\gamma}[f_0]) - R\nu_\gamma(E_{\lambda,\gamma}[u]) > \frac{\varepsilon_1}{8} \frac{16m}{\varepsilon_1} - m = m, \end{aligned}$$

and we see that $f \notin \mathcal{W}(m, j)$. Thus we have shown that $\mathcal{W}(m, j)$ is nowhere dense in $W^{1,1}(\mathbb{R})$. By (6-32) the proof is concluded. □

7. Perspectives and open problems

7A. ***Subspaces of $\dot{W}^{1,1}$ and $\dot{\mathrm{BV}}$ and related spaces.*** The failure of the upper bounds for $[Q_\gamma u]_{L^{1,\infty}(\mathbb{R}^{2N},\nu_\gamma)}$ for $\gamma \in [-1, 0)$ raises a number of interesting questions. Consider the space $\dot{\mathrm{BV}}(\gamma)$ consisting of all $\dot{\mathrm{BV}}$ functions satisfying

$$\|u\|_{\dot{\mathrm{BV}}(\gamma)} := \|\nabla u\|_{\mathcal{M}} + \sup_{\lambda>0} \lambda \nu_\gamma(E_{\lambda,\gamma}[u]) < \infty \tag{7-1}$$

and the corresponding subspace $\dot{W}^{1,1}(\gamma)$ of $\dot{W}^{1,1}$.

Embeddings. We proved in this paper that for $\gamma \notin [-1, 0]$ we have $\dot{\mathrm{BV}}(\gamma) = \dot{\mathrm{BV}}$ and $\dot{W}^{1,1}(\gamma) = \dot{W}^{1,1}$. It is natural to ask how in the range $-1 \le \gamma < 0$ the proper subspaces $\dot{\mathrm{BV}}(\gamma)$ and $\dot{W}^{1,1}(\gamma)$ relate to other families of function spaces, in particular to the Hardy–Sobolev space $\dot{F}^1_{1,2}$, another subspace of $\dot{W}^{1,1}$.

Triangle inequalities. The spaces $\dot{W}^{1,1}(\gamma)$ and $\dot{\mathrm{BV}}(\gamma)$ are defined via $L^{1,\infty}$-quasinorms, and the space $L^{1,\infty}$ is not normable (unlike $L^{p,\infty}$ for $1 < p < \infty$, which is normable [Hunt 1966]). However Theorem 1.4 tells us that $\dot{W}^{1,1}(\gamma)$ and $\dot{\mathrm{BV}}(\gamma)$ are normable for $\gamma \notin [-1, 0]$. Are these spaces normable in the range $\gamma \in [-1, 0)$?

Related quasinorms. Consider for $0 < s \le 1$

$$\|u\|_{(p,s,\gamma)} = \left[\frac{u(x) - u(y)}{|x - y|^{\gamma/p+s}}\right]_{L^{p,\infty}(\mathbb{R}^{2N},\nu_\gamma)}.$$

It is an obvious consequence of Theorem 1.3 that for $s = 1$ and fixed $p > 1$, these expressions define equivalent (semi/quasi)-norms on C^∞_c as γ varies over $\mathbb{R}\setminus\{0\}$. It would be interesting to find a more direct proof of this observation which does not involve the relation with $\dot{W}^{1,p}$. We note that the equivalence for varying γ breaks down for $0 < s < 1$. This result, and more about the spaces for which $\|u\|_{(p,s,\gamma)} < \infty$ with $0 < s < 1$, such as their connection to Besov spaces and interpolation, can be found in [Domínguez et al. 2023].

7B. ***Other limit functionals.*** Our results, combined with the various developments presented in [Brezis and Nguyen 2018; 2020; Nguyen 2007; 2011], suggest several possible directions of research.

Can one prove a generalization of (1-14), (1-16) where the supremum is replaced by the $\liminf_{\lambda\to\infty}$ when $\gamma > 0$ and by a $\liminf_{\lambda\to 0^+}$ when $\gamma < 0$? More precisely, for $1 < p < \infty$, is there a positive constant $C(N, \gamma, p)$ such that, for all $u \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$,

$$\|\nabla u\|^p_{L^p} \le C(N, \gamma, p) \liminf_{\lambda\to\infty} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) \quad \text{if } \gamma > 0, \tag{7-2a}$$

$$\|\nabla u\|^p_{L^p} \le C(N, \gamma, p) \liminf_{\lambda\searrow 0} \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]) \quad \text{if } \gamma < 0, \tag{7-2b}$$

in the sense that $\|\nabla u\|_{L^p} = \infty$ if $u \in L^1_{\mathrm{loc}} \setminus \dot{W}^{1,p}$?

For $p = 1$ we can also ask: is there a positive constant $C(N, \gamma)$ such that, for all $u \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$,

$$\|\nabla u\|_{\mathcal{M}} \le C(N, \gamma) \liminf_{\lambda \to \infty} \lambda \nu_\gamma(E_{\lambda,\gamma}[u]) \quad \text{if } \gamma > 0, \tag{7-3a}$$

$$\|\nabla u\|_{\mathcal{M}} \le C(N, \gamma) \liminf_{\lambda \searrow 0} \lambda \nu_\gamma(E_{\lambda,\gamma}[u]) \quad \text{if } \gamma < 0, \tag{7-3b}$$

in the sense that $\|\nabla u\|_{\mathcal{M}} = \infty$ if $u \in L^1_{\mathrm{loc}} \setminus \dot{\mathrm{BV}}$?

Theorem 1.1 gives (7-2a) and (7-2b) if we additionally assume $u \in \dot{W}^{1,p}(\mathbb{R}^N)$. It also gives (7-3a) and (7-3b) if we additionally assume that $u \in \dot{W}^{1,1}(\mathbb{R}^N)$. It would already be interesting to establish (7-3a), (7-3b) for all $\dot{\mathrm{BV}}$ functions.

When $\gamma = -1$, $p = 1$, (7-3b) holds for all $u \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$ as established in [Nguyen 2008, Theorem 2] and [Brezis and Nguyen 2018, Section 3.4]. For $\gamma = -p$, $1 < p < \infty$, inequality (7-2b) was proved in [Bourgain and Nguyen 2006]. For $\gamma = N$, Poliakovsky [2022] proved weaker versions of (7-2a) and (7-3a) where the lim inf is replaced by a lim sup.

7C. Γ-*convergence.* This is a far-reaching generalization of the questions raised in Section 7B. For fixed $p \ge 1$ and $\gamma \in \mathbb{R} \setminus \{0\}$ consider the functionals

$$\Phi_\lambda[u] := \lambda^p \nu_\gamma(E_{\lambda,\gamma/p}[u]), \quad \lambda \in (0, \infty),$$

defined for all $u \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$. It would be very interesting to study the Γ-limit of Φ_λ in $L^1_{\mathrm{loc}}(\mathbb{R}^N)$, in the sense of De Giorgi, as $\lambda \to \infty$ when $\gamma > 0$ and as $\lambda \searrow 0$ when $\gamma < 0$. More specifically, if $p > 1$, define on $L^1_{\mathrm{loc}}(\mathbb{R}^N)$

$$\Phi_{*,c}[u] = \begin{cases} c\|\nabla u\|^p_{L^p} & \text{if } u \in \dot{W}^{1,p}(\mathbb{R}^N), \\ \infty & \text{otherwise,} \end{cases}$$

and for $p = 1$ define

$$\Phi_{*,c}[u] = \begin{cases} c\|\nabla u\|_{\mathcal{M}} & \text{if } u \in \dot{\mathrm{BV}}(\mathbb{R}^N), \\ \infty & \text{otherwise.} \end{cases}$$

A challenging question is whether there exists a constant $c = c(p, \gamma, N) > 0$ such that $\Phi_\lambda \to \Phi_{*,c}$ in the sense of Γ-convergence, meaning

(1) whenever $u_\lambda \to u$ in L^1_{loc} then $\liminf \Phi_\lambda[u_\lambda] \ge \Phi_{*,c}[u]$, and

(2) for each $u \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$ there exist (v_λ) with $v_\lambda \in L^1_{\mathrm{loc}}(\mathbb{R}^N)$, $v_\lambda \to u$ in L^1_{loc} and $\limsup \Phi_\lambda[v_\lambda] \le \Phi_{*,c}[u]$.

This question is especially meaningful in the case $p = 1$ where the pointwise limit behaves somewhat pathologically. Indeed, recall that for $p = 1$, $-1 \le \gamma < 0$ there is no universal upper bound for $\Phi_\lambda[u]$ in terms of $\|\nabla u\|_{L^1}$. Also when $p = 1$ and $\gamma \in \mathbb{R} \setminus [-1, 0]$ the examples in Section 3F show that the pointwise limit in $\dot{W}^{1,1}$ and on $\dot{\mathrm{BV}} \setminus \dot{W}^{1,1}$ may differ (by a multiplicative constant). A remarkable result of Nguyen [2007; 2011] states that $\Phi_\lambda \to \Phi_{*,c}$ as $\lambda \to 0$, in the sense of Γ–convergence, when $p \ge 1$, and $\gamma = -p$ for some appropriate constant $c = c(p, N)$; see also [Brezis and Nguyen 2020] (note, however, that $\dot{W}^{1,p}$ and $\dot{\mathrm{BV}}$ are replaced in these papers by $W^{1,p}$ and BV).

7D. ***More general families of functionals.*** Consider a monotone nondecreasing function $\varphi : [0, \infty) \to [0, \infty)$ and set (inspired by [Brezis and Nguyen 2018; 2020])

$$\Psi_\lambda[u] := \lambda^p \iint_{\mathbb{R}^N \times \mathbb{R}^N} \varphi\left(\frac{|u(x) - u(y)|}{\lambda |x-y|^{1+\gamma/p}}\right) |x-y|^{\gamma - N} \, \mathrm{d}x \, \mathrm{d}y.$$

The family Φ_λ in Section 7C corresponds to $\varphi = \mathbb{1}_{(1,\infty)}$. It is an interesting generalization of the above problems to study the limit of Ψ_λ as $\lambda \searrow 0$ when $\gamma < 0$ and the limit of Ψ_λ as $\lambda \to \infty$ when $\gamma > 0$, both in the sense of pointwise convergence or in the sense of Γ-convergence. A formal computation suggests that our Theorem 1.1 should go over modulo a factor $\int_0^\infty \varphi(s)/s^{p+1} \, \mathrm{d}s$; see [Brezis and Nguyen 2020]. We refer to [Brezis and Nguyen 2018] for a further discussion of applications.

Acknowledgements

Seeger and Yung would like to thank the Hausdorff Research Institute of Mathematics and the organizers of the trimester program “Harmonic analysis and analytic number theory” for a pleasant working environment in the summer of 2021. The research was supported in part by NSF grants DMS-1764295, DMS-2054220 (Seeger) and by a Future Fellowship FT200100399 from the Australian Research Council (Yung). We thank the referee for a careful reading and numerous valuable suggestions.

References

[Bourgain and Nguyen 2006] J. Bourgain and H.-M. Nguyen, “A new characterization of Sobolev spaces”, *C. R. Math. Acad. Sci. Paris* **343**:2 (2006), 75–80. MR Zbl

[Bourgain et al. 2001] J. Bourgain, H. Brezis, and P. Mironescu, “Another look at Sobolev spaces”, pp. 439–455 in *Optimal control and partial differential equations*, edited by J. L. Menaldi et al., IOS, Amsterdam, 2001. MR Zbl

[Brezis 2002] H. Brezis, “How to recognize constant functions: a connection with Sobolev spaces”, *Uspekhi Mat. Nauk* **57**:4(346) (2002), 59–74. In Russian; translated in *Russian Math. Surv.* **57**:4 (2002), 693–708. MR Zbl

[Brezis 2011] H. Brezis, *Functional analysis, Sobolev spaces and partial differential equations*, Springer, 2011. MR Zbl

[Brezis and Nguyen 2018] H. Brezis and H.-M. Nguyen, “Non-local functionals related to the total variation and connections with image processing”, *Ann. PDE* **4**:1 (2018), art. id. 9. MR Zbl

[Brezis and Nguyen 2020] H. Brezis and H.-M. Nguyen, “Non-local, non-convex functionals converging to Sobolev norms”, *Nonlinear Anal.* **191** (2020), art. id. 111626. MR Zbl

[Brezis et al. 2021a] H. Brezis, J. Van Schaftingen, and P.-L. Yung, “Going to Lorentz when fractional Sobolev, Gagliardo and Nirenberg estimates fail”, *Calc. Var. Partial Differential Equations* **60**:4 (2021), art. id. 129. MR Zbl

[Brezis et al. 2021b] H. Brezis, J. Van Schaftingen, and P.-L. Yung, “A surprising formula for Sobolev norms”, *Proc. Natl. Acad. Sci. USA* **118**:8 (2021), art. id. e2025254118. MR

[Dávila 2002] J. Dávila, “On an open question about functions of bounded variation”, *Calc. Var. Partial Differential Equations* **15**:4 (2002), 519–527. MR Zbl

[Domínguez and Milman 2022] Ó. Domínguez and M. Milman, “New Brezis–Van Schaftingen–Yung–Sobolev type inequalities connected with maximal inequalities and one parameter families of operators”, *Adv. Math.* **411**:A (2022), art. id. 108774. MR Zbl

[Domínguez et al. 2023] Ó. Domínguez, A. Seeger, B. Street, J. Van Schaftingen, and P.-L. Yung, “Spaces of Besov–Sobolev type and a problem on nonlinear approximation”, *J. Funct. Anal.* **284**:4 (2023), art. id. 109775. MR Zbl

[Federer 1969] H. Federer, *Geometric measure theory*, Grundlehren der Math. Wissenschaften **153**, Springer, 1969. MR Zbl

[Hajłasz and Kałamajska 1995] P. Hajłasz and A. Kałamajska, "Polynomial asymptotics and approximation of Sobolev functions", *Studia Math.* **113**:1 (1995), 55–64. MR Zbl

[Hunt 1966] R. A. Hunt, "On $L(p, q)$ spaces", *Enseign. Math.* (2) **12** (1966), 249–276. MR Zbl

[Mattila 2015] P. Mattila, *Fourier analysis and Hausdorff dimension*, Cambridge Stud. Adv. Math. **150**, Cambridge Univ. Press, 2015. MR Zbl

[Nguyen 2006] H.-M. Nguyen, "Some new characterizations of Sobolev spaces", *J. Funct. Anal.* **237**:2 (2006), 689–720. MR Zbl

[Nguyen 2007] H.-M. Nguyen, "Γ-convergence and Sobolev norms", *C. R. Math. Acad. Sci. Paris* **345**:12 (2007), 679–684. MR Zbl

[Nguyen 2008] H.-M. Nguyen, "Further characterizations of Sobolev spaces", *J. Eur. Math. Soc.* **10**:1 (2008), 191–229. MR Zbl

[Nguyen 2011] H.-M. Nguyen, "Γ-convergence, Sobolev norms, and BV functions", *Duke Math. J.* **157**:3 (2011), 495–533. MR Zbl

[Poliakovsky 2022] A. Poliakovsky, "Some remarks on a formula for Sobolev norms due to Brezis, Van Schaftingen and Yung", *J. Funct. Anal.* **282**:3 (2022), art. id. 109312. MR Zbl

[Stein 1970] E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton Math. Series **30**, Princeton Univ. Press, 1970. MR Zbl

[Stein and Weiss 1971] E. M. Stein and G. Weiss, *Introduction to Fourier analysis on Euclidean spaces*, Princeton Math. Series **32**, Princeton Univ. Press, 1971. MR Zbl

Received 24 Sep 2021. Revised 5 Jul 2022. Accepted 11 Aug 2022.

HAÏM BREZIS: brezis@math.rutgers.edu
Department of Mathematics, Rutgers University, Piscataway, NJ, United States

and

Departments of Mathematics and Computer Science, Technion, Israel Institute of Technology, Haifa, Israel

and

Laboratoire Jacques-Louis Lions, Sorbonne Universités, UPMC Université Paris-6, Paris, France

ANDREAS SEEGER: seeger@math.wisc.edu
Department of Mathematics, University of Wisconsin, Madison, WI, United States

JEAN VAN SCHAFTINGEN: jean.vanschaftingen@uclouvain.be
Institut de Recherche en Mathématique et Physique, Université catholique de Louvain, Louvain-la-Neuve, Belgium

PO-LAM YUNG: polam.yung@anu.edu.au
Mathematical Sciences Institute, Australian National University, Canberra, Australia

ANALYSIS AND PDE
Vol. 17 (2024), No. 3, pp. 981–1003
DOI: 10.2140/apde.2024.17.981

SCHWARZ–PICK LEMMA FOR HARMONIC MAPS WHICH ARE CONFORMAL AT A POINT

FRANC FORSTNERIČ AND DAVID KALAJ

We obtain a sharp estimate on the norm of the differential of a harmonic map from the unit disc $\mathbb{D}$ in $\mathbb{C}$ into the unit ball $\mathbb{B}^n$ of $\mathbb{R}^n$, $n \geq 2$, at any point where the map is conformal. For $n = 2$ this generalizes the classical Schwarz–Pick lemma, and for $n \geq 3$ it gives the optimal Schwarz–Pick lemma for conformal minimal discs $\mathbb{D} \to \mathbb{B}^n$. This implies that conformal harmonic maps $M \to \mathbb{B}^n$ from any hyperbolic conformal surface are distance decreasing in the Poincaré metric on M and the Cayley–Klein metric on the ball $\mathbb{B}^n$, and the extremal maps are the conformal embeddings of the disc $\mathbb{D}$ onto affine discs in $\mathbb{B}^n$. Motivated by these results, we introduce an intrinsic pseudometric on any Riemannian manifold of dimension at least three by using conformal minimal discs, and we lay foundations of the corresponding hyperbolicity theory.

1. Introduction

In this paper, we establish precise estimates of derivatives and the rate of growth of conformal harmonic maps from hyperbolic conformal surfaces into the unit ball $\mathbb{B}^n$ of $\mathbb{R}^n$ for any $n \geq 3$; see Theorem 2.6. Such maps parametrize minimal surfaces, objects of high interest in geometry. To motivate the discussion, we begin with the following special case of one of our main results, Theorem 2.1. This generalizes the classical Schwarz–Pick lemma, due to H. A. Schwarz [1890, Band II, p. 108], H. Poincaré [1884], C. Carathéodory [1912], and G. A. Pick [1915], to a substantially larger class of maps.

Theorem 1.1. *Let $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ denote the unit disc. If $f : \mathbb{D} \to \mathbb{D}$ is a harmonic map which is conformal at a point $z \in \mathbb{D}$, then at this point we have*

$$\|df_z\| \leq \frac{1 - |f(z)|^2}{1 - |z|^2}, \tag{1-1}$$

with equality if and only if f is a conformal diffeomorphism of the disc $\mathbb{D}$.

The classical Schwarz–Pick lemma gives the same conclusion under the much stronger hypothesis that the map f is holomorphic or antiholomorphic, which means that it is conformal at every noncritical point; see, e.g., [Dineen 1989; Kobayashi 2005; Royden 1988]. This fundamental rigidity result in complex analysis leads to the notion of Kobayashi hyperbolic manifolds [1967; 1976; 2005] and provides a connection to complex differential geometry via the Ahlfors lemma (see [Ahlfors 1938; Kobayashi 2005, Theorem 2.1; Royden 1988]) and its generalizations by S.-T. Yau [1978] and others.

MSC2020: primary 53A10; secondary 30C80, 31A05, 32Q45.
Keywords: harmonic map, minimal surface, Schwarz–Pick lemma, Cayley–Klein metric.

The conditions in Theorem 1.1 are invariant under precompositions by holomorphic automorphisms of $\mathbb{D}$, so the proof reduces to the case $z = 0$. On the other hand, postcompositions of harmonic maps into $\mathbb{D}$ by holomorphic automorphisms of $\mathbb{D}$ need not be harmonic, so we cannot exchange $f(0)$ and 0. Hence, the standard proof of the classical Schwarz–Pick lemma breaks down. The estimate (1-1) fails for some nonconformal harmonic diffeomorphisms of $\mathbb{D}$ (see Example 4.1), as well as for harmonic maps $\mathbb{D} \to D$ to more general domains which are conformal at a point (see Example 4.2 and Problem 4.3).

Our main results are precise estimates of the differential and the rate of growth of conformal harmonic maps $M \to \mathbb{B}^n$ from an open conformal surface M to the unit ball $\mathbb{B}^n$ of $\mathbb{R}^n$ for any $n \geq 3$. It is classical that such maps parametrize minimal surfaces. Indeed, a smooth conformal map $f : M \to \mathbb{R}^n$ from an open conformal surface M into $\mathbb{R}^n$ with the Euclidean metric parametrizes a minimal surface in $\mathbb{R}^n$ if and only if f is a harmonic map; see [Alarcón et al. 2021, Chapter 2; Duren 2004; Osserman 1969], among many other sources. Note that an oriented conformal surface is a Riemann surface.

The focal point of the paper is Theorem 2.1, which gives a precise upper bound on the norm $\|df_z\|$ of the differential df_z of a harmonic map $f : \mathbb{D} \to \mathbb{B}^n$ at any point $z \in \mathbb{D}$ where the map is conformal. The estimate is similar to the one in Theorem 1.1, except that, for $n \geq 3$, it also involves the angle θ between the position vector $f(z) \in \mathbb{B}^n$ and the 2-plane $df_z(\mathbb{R}^2) \subset \mathbb{R}^n$. A related result (see Theorem 2.2) shows that the worst case estimate, which occurs for $\theta = \frac{\pi}{2}$ (i.e., when the vector $f(z)$ is orthogonal to the plane $df_z(\mathbb{R}^2)$), holds for all harmonic maps $f : \mathbb{D} \to \mathbb{B}^n$ provided that $\|df_z\|$ is replaced by $\sqrt{2}^{-1}|\nabla f(z)|$; these quantities coincide if f is conformal at z.

We then give a differential geometric formulation and an extension of Theorem 2.1. Let $\mathcal{CK}$ denote the Cayley–Klein metric on the ball $\mathbb{B}^n$ ($n \geq 2$), also called the Beltrami–Klein metric; see (2-6) and the footnote on page 985. This metric is one of the classical models of hyperbolic geometry. It coincides with the restriction of the Kobayashi metric on the complex ball $\mathbb{B}^n_{\mathbb{C}} \subset \mathbb{C}^n$ (2-5) (which is the same as $1/\sqrt{n+1}$ times the Bergman metric on $\mathbb{B}^n_{\mathbb{C}}$) to points of the real ball $\mathbb{B}^n$ and real tangent vectors. Theorem 2.1 implies that any conformal harmonic map $f : M \to \mathbb{B}^n$, $n \geq 3$, from a hyperbolic conformal surface is metric and distance decreasing in the Poincaré metric on M and the Cayley–Klein metric on $\mathbb{B}^n$; see Theorem 2.6. Furthermore, if the differential df_p has the operator norm equal to 1 at some point $p \in M$ in this pair of metrics, or if f preserves the distance between a pair of distinct points in M, then M is necessarily the disc $\mathbb{D}$ and f is a conformal diffeomorphism of $\mathbb{D}$ onto a proper affine disc in $\mathbb{B}^n$. In particular, a conformal harmonic disc $f : \mathbb{D} \to \mathbb{B}^n$ with $f(0) = 0$ satisfies $|f(z)| \leq |z|$ for all $z \in \mathbb{D}$ (see Corollary 2.7).

In Section 2 we give precise statements of the mentioned results. Theorem 2.1 is proved in Section 3. We introduce a new idea into the subject, connecting it to Lempert's seminal work [1981] on complex geodesics of the Kobayashi metric on bounded convex domains in $\mathbb{C}^n$. Theorem 2.2 is proven in Section 4. In Section 5 we apply Theorem 1.1 to estimate the gradient of a quasiconformal harmonic self-map of the disc in terms of its second Beltrami coefficient at the reference point; see Theorem 5.1.

Motivated by these result, we introduce in Section 6 an intrinsic pseudometric on any domain in $\mathbb{R}^n$, $n \geq 3$ (and more generally on any Riemannian manifold of dimension at least three) in terms of conformal minimal discs, in analogy to Kobayashi's definition of his pseudometric on complex manifolds in terms of holomorphic discs. This provides the basis for a new hyperbolicity theory of such domains and of Riemannian manifolds.

2. The main results

Given a differentiable map $f : \mathbb{D} \to \mathbb{R}^n$, we denote by f_x and f_y its partial derivatives with respect to x and y, where $z = x + \mathrm{i}y \in \mathbb{D}$. The gradient $\nabla f = (f_x, f_y)$ is an $n \times 2$ matrix representing the differential df. The map f is said to be *conformal* at $z \in \mathbb{D}$ if

$$|f_x(z)| = |f_y(z)| \quad \text{and} \quad f_x(z) \cdot f_y(z) = 0. \tag{2-1}$$

Here, the dot stands for the Euclidean inner product on $\mathbb{R}^n$, and $|\boldsymbol{x}|$ is the Euclidean norm of $\boldsymbol{x} \in \mathbb{R}^n$. If f is an immersion at z then (2-1) holds if and only if df_z preserves angles. It follows from (2-1) that f has rank zero at any branch point. We denote by $|\nabla f|$ the Euclidean norm of the gradient:

$$|\nabla f(z)|^2 = |f_x(z)|^2 + |f_y(z)|^2, \quad z \in \mathbb{D}.$$

If f is conformal at z then clearly $\|df_z\| = \sqrt{2}^{-1} |\nabla f(z)| = |f_x(z)| = |f_y(z)|$. The map $f = (f_1, \dots, f_n) : \mathbb{D} \to \mathbb{R}^n$ is harmonic if and only if every component f_k is a harmonic function on $\mathbb{D}$, meaning that the Laplacian $\Delta f_k = \partial^2 f_k / \partial x^2 + \partial^2 f_k / \partial y^2$ vanishes identically.

We denote by $\mathbb{B}^n$ the unit ball of $\mathbb{R}^n$:

$$\mathbb{B}^n = \left\{ \boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{R}^n : |\boldsymbol{x}|^2 = \sum_{k=1}^{n} x_k^2 < 1 \right\}. \tag{2-2}$$

Our first main result is the following; it is proved in Section 3.

Theorem 2.1. *Let $f : \mathbb{D} \to \mathbb{B}^n$ for $n \geq 2$ be a harmonic map. If f is conformal at a point $z \in \mathbb{D}$ and $\theta \in \left[0, \frac{\pi}{2}\right]$ denotes the angle between the vector $f(z)$ and the plane $\Lambda = df_z(\mathbb{R}^2) \subset \mathbb{R}^n$, then*

$$\|df_z\| = \frac{1}{\sqrt{2}} |\nabla f(z)| \leq \frac{1 - |f(z)|^2}{1 - |z|^2} \frac{1}{\sqrt{1 - |f(z)|^2 \sin^2 \theta}}, \tag{2-3}$$

with equality if and only if f is a conformal diffeomorphism of $\mathbb{D}$ onto the affine disc $\Sigma = (f(z) + \Lambda) \cap \mathbb{B}^n$. (When $f(z) = 0$ or $df_z = 0$, the angle θ does not matter.)

Note that the number $R = \sqrt{1 - |f(z)|^2 \sin^2 \theta}$ is the radius of the affine disc Σ. In dimension $n = 2$ we have $\theta = 0$, so Theorem 1.1 is a special case of Theorem 2.1. Without assuming that f is conformal at z or that $f(z) = 0$, inequality (2-3) fails for some harmonic diffeomorphisms of the disc as shown by Example 4.1.

For a fixed value of $|f(z)| \in [0, 1)$, the maximum of the right-hand side of (2-3) over angles $\theta \in \left[0, \frac{\pi}{2}\right]$ equals $\sqrt{1 - |f(z)|^2} / (1 - |z|^2)$ and is attained precisely at $\theta = \frac{\pi}{2}$, i.e., when the vector $f(z)$ is orthogonal to $\Lambda = df_z(\mathbb{R}^2)$, unless $f(z) = 0$ when it is independent of θ. It turns out that this weaker estimate holds for all harmonic maps $\mathbb{D} \to \mathbb{B}^n$ without any conformality assumption. The following result is proved in Section 4.

Theorem 2.2. *For every harmonic map $f : \mathbb{D} \to \mathbb{B}^n$ $(n \geq 2)$ we have that*

$$\frac{1}{\sqrt{2}}|\nabla f(z)| \leq \frac{\sqrt{1-|f(z)|^2}}{1-|z|^2}, \quad z \in \mathbb{D}. \tag{2-4}$$

Equality holds for some $z \in \mathbb{D}$ if $f(z)$ is orthogonal to the two-plane $\Lambda = df_z(\mathbb{R}^2)$ and f is a conformal diffeomorphism onto the affine disc $(f(z)+\Lambda) \cap \mathbb{B}^n$. In particular, if $f(z) = 0$ then

$$|\nabla f(z)| \leq \frac{\sqrt{2}}{1-|z|^2},$$

with equality if and only if f is a conformal diffeomorphism onto the linear disc $\Lambda \cap \mathbb{B}^n$.

The proof of estimate (2-4) relies on Parseval's inequality, using the hypothesis that the L^1-norm of $|f|^2 = \sum_{k=1}^n f_k^2$ on the circles $\{|z| = r\}$ for $0 < r < 1$ is bounded by 1. We find it surprising that this simple approach gives an optimal estimate in certain cases indicated in the theorem. Except in these cases, we do not know whether there exist harmonic maps $\mathbb{D} \to \mathbb{B}^n$ reaching (near) equality in (2-4).

The precise upper bound on the size of the gradient $\|df_0\|$ of a nonconformal harmonic map $f : \mathbb{D} \to \mathbb{B}^n$ with a given center $f(0) = \boldsymbol{x} \in \mathbb{B}^n \setminus \{0\}$ for $n \geq 2$ in terms of the distortion of f at 0 is unknown; see [Brevig et al. 2021; Kovalev and Yang 2020] for $n = 2$. On the other hand, for $n = 1$ the *harmonic Schwarz lemma* (see [Axler et al. 2001, Theorem 6.26]) says that any harmonic function $f : \mathbb{B}^m \to (-1, +1)$ for $m \geq 2$ satisfies the sharp estimate

$$|\nabla f(0)| \leq \frac{2\,\mathrm{Vol}(\mathbb{B}^{m-1})}{\mathrm{Vol}(\mathbb{B}^m)}.$$

For $m = 2$ the inequality reads $|\nabla f(0)| \leq \frac{4}{\pi}$, and a simple proof in this case was given by Kalaj and Vuorinen [2012, Theorem 1.8].

Let us mention a consequence of Theorem 2.1 related to the Schwarz lemma for holomorphic discs in the ball of the complex Euclidean space,

$$\mathbb{B}^n_{\mathbb{C}} = \left\{ \boldsymbol{z} = (z_1, \ldots, z_n) \in \mathbb{C}^n : |\boldsymbol{z}|^2 = \sum_{k=1}^n |z_k|^2 < 1 \right\} \tag{2-5}$$

(see [Rudin 1980, Section 8.1]). The following corollary to Theorem 2.1 shows that the extremal holomorphic discs in $\mathbb{B}^n_{\mathbb{C}}$ are precisely those extremal orientation-preserving conformal harmonic discs $\mathbb{D} \to \mathbb{B}^n_{\mathbb{C}}$ which parametrize affine complex discs.

Corollary 2.3. *Let $f : \mathbb{D} \to \mathbb{B}^n_{\mathbb{C}}$ be a harmonic map which is conformal at a point $z \in \mathbb{D}$. If $\Lambda = df_z(\mathbb{R}^2)$ is a complex line in $\mathbb{C}^n$, then equality holds in* (2-3) *for this z if and only if f is a biholomorphic or antibiholomorphic map onto the affine complex disc $(f(z)+\Lambda) \cap \mathbb{B}^n_{\mathbb{C}}$.*

The Cayley–Klein metric. A differential geometric interpretation of the classical Schwarz–Pick lemma is that holomorphic maps $\mathbb{D} \to \mathbb{D}$ are distance decreasing in the Poincaré metric on $\mathbb{D}$, and isometries coincide with holomorphic and antiholomorphic automorphisms of $\mathbb{D}$ (see [Kobayashi 2005]). The analogous conclusion holds for holomorphic maps $\mathbb{D} \to \mathbb{B}^n_{\mathbb{C}}$ with the Kobayashi metric on the complex ball $\mathbb{B}^n_{\mathbb{C}}$ (2-5), where orientation-preserving isometric embeddings are precisely holomorphic embeddings onto affine complex discs in $\mathbb{B}^n_{\mathbb{C}}$.

In the same spirit, we shall now interpret Theorem 2.1 as the distance-decreasing property of conformal harmonic maps $\mathbb{D} \to \mathbb{B}^n$ with respect to the *Cayley–Klein metric*[1] on $\mathbb{B}^n$:

$$\mathcal{CK}(\boldsymbol{x}, \boldsymbol{v}) = \frac{\sqrt{1-|\boldsymbol{x}|^2 \sin^2\phi}}{1-|\boldsymbol{x}|^2}|\boldsymbol{v}|, \quad \boldsymbol{x} \in \mathbb{B}^n,\ \boldsymbol{v} \in \mathbb{R}^n, \tag{2-6}$$

where $\phi \in \left[0, \frac{\pi}{2}\right]$ is the angle between the vector $\boldsymbol{x}$ and the line $\mathbb{R}\boldsymbol{v}$. Equivalently,

$$\mathcal{CK}(\boldsymbol{x}, \boldsymbol{v})^2 = \frac{(1-|\boldsymbol{x}|^2)|\boldsymbol{v}|^2 + |\boldsymbol{x}\cdot\boldsymbol{v}|^2}{(1-|\boldsymbol{x}|^2)^2} = \frac{|\boldsymbol{v}|^2}{1-|\boldsymbol{x}|^2} + \frac{|\boldsymbol{x}\cdot\boldsymbol{v}|^2}{(1-|\boldsymbol{x}|^2)^2}. \tag{2-7}$$

Let $G_2(\mathbb{R}^n)$ denote the Grassmann manifold of two-planes in $\mathbb{R}^n$. We define a Finsler pseudometric $\mathcal{M} : \mathbb{B}^n \times G_2(\mathbb{R}^n) \to \mathbb{R}_+$ by

$$\mathcal{M}(\boldsymbol{x}, \Lambda) = \frac{\sqrt{1-|\boldsymbol{x}|^2 \sin^2\theta}}{1-|\boldsymbol{x}|^2}, \quad \boldsymbol{x} \in \mathbb{B}^n,\ \Lambda \in G_2(\mathbb{R}^n), \tag{2-8}$$

where $\theta \in \left[0, \frac{\pi}{2}\right]$ is the angle between $\boldsymbol{x}$ and Λ. At $\boldsymbol{x} = \boldsymbol{0}$ we have $\mathcal{M}(\boldsymbol{0}, \Lambda) = 1$ for all $\Lambda \in G_2(\mathbb{R}^n)$. Assume now that $\boldsymbol{x} \neq \boldsymbol{0}$. Let $\boldsymbol{v} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ be a vector having angle $\phi \in \left[0, \frac{\pi}{2}\right]$ with the line $\mathbb{R}\boldsymbol{x}$. The angle θ between $\boldsymbol{x}$ and any 2-plane Λ containing $\boldsymbol{v}$ satisfies $0 \le \theta \le \phi$, and the maximum of θ over all such Λ equals ϕ. Hence, (2-6) gives

$$\frac{\mathcal{CK}(\boldsymbol{x}, \boldsymbol{v})}{|\boldsymbol{v}|} = \min\{\mathcal{M}(\boldsymbol{x}, \Lambda) : \Lambda \in G_2(\mathbb{R}^n),\ \boldsymbol{v} \in \Lambda\}, \tag{2-9}$$

$$\mathcal{M}(\boldsymbol{x}, \Lambda) = \max\left\{\frac{\mathcal{CK}(\boldsymbol{x}, \boldsymbol{v})}{|\boldsymbol{v}|} : \boldsymbol{v} \in \Lambda\right\}. \tag{2-10}$$

Inequality (2-3) in Theorem 2.1 is obviously equivalent to

$$\mathcal{M}(f(z), df_z(\mathbb{R}^2))|df_z(\xi)| = \frac{\sqrt{1-|f(z)|^2 \sin^2\theta}}{1-|f(z)|^2}|df_z(\xi)| \le \frac{|\xi|}{1-|z|^2}, \tag{2-11}$$

where $\theta \in \left[0, \frac{\pi}{2}\right]$ is the angle between $f(z)$ and the 2-plane $\Lambda = df_z(\mathbb{R}^2)$. By (2-9) the left-hand side of (2-11) is bigger than or equal to $\mathcal{CK}(f(z), df_z(\xi))$. Equality holds if and only if the angle ϕ between the line $f(z)\mathbb{R}$ and the vector $df_z(\xi) \in \Lambda$ equals θ; clearly this holds if and only if $df_z(\xi)$ is tangent to the diameter of the affine disc $\Sigma = (f(z)+\Lambda) \cap \mathbb{B}^n$ through the point $f(z)$. This and the addition concerning equality in (2-3) give the following corollary to Theorem 2.1. Note that $\mathcal{P}_{\mathbb{D}}(z, \xi) := |\xi|/(1-|z|^2)$ is the Poincaré metric on the disc.

[1]The Beltrami–Calvin–Klein model of hyperbolic geometry was introduced by Arthur Cayley [1859] and Eugenio Beltrami [1868], and it was developed by Felix Klein [1871; 1873]. The underlying space is the n-dimensional unit ball, geodesics are straight line segments with ideal endpoints on the boundary sphere, and the distance between points on a geodesic is given by a cross ratio. This is a special case of the Hilbert metric on convex domains in $\mathbb{R}^n$ and $\mathbb{RP}^n$, introduced by David Hilbert [1895]. These are examples of projectively invariant metrics discussed by many authors; see the surveys by S. Kobayashi [1977; 1984], W. M. Goldman [2019], and J. G. Ratcliffe [1994].

Corollary 2.4. *If $f : \mathbb{D} \to \mathbb{B}^n$ is a conformal harmonic map then for every point $z \in \mathbb{D}$ and tangent vector $\xi \in \mathbb{R}^2$ we have*

$$\mathcal{CK}(f(z), df_z(\xi)) \le \frac{|\xi|}{1-|z|^2} = \mathcal{P}_{\mathbb{D}}(z, \xi). \tag{2-12}$$

Equality holds for some $z \in \mathbb{D}$ and $\xi \in \mathbb{R}^2 \setminus \{0\}$ if and only if f is a conformal diffeomorphism onto the affine disc

$$\Sigma = (f(z) + df_z(\mathbb{R}^2)) \cap \mathbb{B}^n$$

and the vector $df_z(\xi)$ is tangent to the diameter of Σ through the point $f(z)$.

This shows in particular that every linear conformal embedding $f : \mathbb{D} \to \Sigma$ onto a proper affine disc in $\mathbb{B}^n$ is geodesic on each diameter $(-1, +1) \ni r \mapsto f(re^{it}) \in \Sigma$ for every fixed $t \in \mathbb{R}$. However, distances between points of different rays strictly decrease from the Poincaré metric on $\mathbb{D}$ to the Cayley–Klein metric on the disc $\Sigma \subset \mathbb{B}^n$.

Remark 2.5. The Cayley–Klein metric (2-7) is the restriction of the Kobayashi metric on the unit ball $\mathbb{B}^n_{\mathbb{C}} \subset \mathbb{C}^n$ to points $\boldsymbol{x} \in \mathbb{B}^n = \mathbb{B}^n_{\mathbb{C}} \cap \mathbb{R}^n$ of the real ball and tangent vectors in $T_{\boldsymbol{x}}\mathbb{R}^n \cong \mathbb{R}^n$. A direct geometric argument was given by Lempert [1993, proof of Theorem 3.1]. The Cayley–Klein metric also equals $1/\sqrt{n+1}$ times the Bergman metric on $\mathbb{B}^n_{\mathbb{C}}$ restricted to $\mathbb{B}^n$ and real tangent vectors; see [Krantz 1992, Proposition 1.4.22]. (On the ball of $\mathbb{C}^n$, most holomorphically invariant metrics coincide up to scalar factors.) The Cayley–Klein metric equals the Poincaré metric $|\boldsymbol{v}|/(1-|\boldsymbol{x}|^2)$ on $\mathbb{B}^n$ on vectors $\boldsymbol{v}$ parallel to the base point $\boldsymbol{x} \in \mathbb{B}^n$, but is strictly smaller on other vectors. While the Poincaré metric on $\mathbb{B}^n$ is conformally equivalent to the Euclidean metric, the Cayley–Klein metric is not.

We now extend Corollary 2.4 to more general minimal surfaces. A *conformal surface* is a topological surface M together with a *conformal atlas*, i.e., an atlas whose transition maps between charts are conformal diffeomorphisms between plane domains. Every surface admits a conformal structure. Indeed, every topological surface admits a smoothing, and a conformal structure on a smooth surface is determined by the choice of a Riemannian metric in view of the existence of local isothermal coordinates (see [Osserman 1969] or [Alarcón et al. 2021, Theorem 1.8.6]). Oriented conformal surfaces are Riemann surfaces. There is a well-defined notion of a harmonic function on a conformal surface. Indeed, a Riemannian metric g defines the metric Laplacian Δ_g and hence g-harmonic functions satisfying $\Delta_g h = 0$. The Laplacians associated to any two Riemannian metrics in the same conformal class on a surface differ by a positive multiplicative function (see [Alarcón et al. 2021, Corollary 1.8.2]), and hence the notion of a harmonic function is independent of the choice of metric in a given conformal class.

A conformal surface M is said to be *hyperbolic* if its universal conformal covering space is the disc $\mathbb{D}$. Let $h : \mathbb{D} \to M$ be a universal conformal covering map. Since conformal automorphisms of $\mathbb{D}$ are isometries of the Poincaré metric $\mathcal{P}_{\mathbb{D}} = |dz|/(1-|z|^2)$, there is a unique Riemannian metric $\mathcal{P}_M$ on M (a Kähler metric if M is a Riemann surface) such that h is a local isometry. This Poincaré metric $\mathcal{P}_M$ is a complete metric of constant Gaussian curvature -4 (see [Kobayashi 2005, p. 48, Example 2]), which agrees with the Kobayashi metric if M is a Riemann surface. This leads to the following generalization of Corollary 2.4.

Theorem 2.6 (metric and distance decreasing property of conformal harmonic maps). *Let M be a connected hyperbolic conformal surface endowed with the Poincaré metric $\mathcal{P}_M$. Every conformal harmonic map $f : M \to \mathbb{B}^n$ $(n \geq 3)$ satisfies the estimate*

$$\mathcal{CK}(f(p), df_p(\xi)) \leq \mathcal{P}_M(p, \xi), \quad p \in M,\ \xi \in T_pM. \tag{2-13}$$

If equality holds in (2-13) *for some point $p \in M$ and vector $0 \neq \xi \in T_pM$, or if f preserves the distance on a pair of distinct points in M, then $M = \mathbb{D}$ and f is a conformal diffeomorphism onto an affine disc in $\mathbb{B}^n$.*

Proof. Assume first that M is orientable and hence a Riemann surface. Choose a holomorphic covering map $h : \mathbb{D} \to M$ and a point $z \in \mathbb{D}$ with $h(z) = p$. The conformal harmonic map $\tilde{f} = f \circ h : \mathbb{D} \to \mathbb{B}^n$ then satisfies $\tilde{f}(z) = f(p)$ and $d\tilde{f}_z = df_p \circ dh_z$. Let $\eta \in \mathbb{R}^2$ be such that $dh_z(\eta) = \xi$. Then $\mathcal{P}_M(p, \xi) = \mathcal{P}_{\mathbb{D}}(z, \eta)$ by the definition of the metric $\mathcal{P}_M$, and $d\tilde{f}_z(\eta) = df_p(\xi)$. From (2-12) it follows that

$$\mathcal{CK}(f(p), df_p(\xi)) = \mathcal{CK}(\tilde{f}(z), d\tilde{f}_z(\eta)) \leq \frac{|d\tilde{f}_z(\eta)|}{1 - |\tilde{f}(z)|^2} = \frac{|df_p(\xi)|}{1 - |f(p)|^2},$$

which gives (2-13). If $\xi \neq 0$ and equality holds, then by Corollary 2.4 the map $\tilde{f} = f \circ h : \mathbb{D} \to \mathbb{B}^n$ is a conformal diffeomorphism onto an affine disc in $\mathbb{B}^n$, and hence $h : \mathbb{D} \to M$ is a biholomorphism.

For a nonorientable hyperbolic conformal surface M we obtain the same conclusion by passing to its orientable two-sheeted conformal cover. The statement concerning distances is an immediate consequence. Note that if the distances agree for a pair of distinct points in M and their images in $\mathbb{B}^n$, then the differential df_p has operator norm 1 at some point $p \in M$ in the given pair of metrics. □

On the disc with the Poincaré metric $\mathcal{P}_{\mathbb{D}} = |dz|/(1 - |z|^2)$, the Poincaré distance equals

$$\mathrm{dist}_{\mathcal{P}}(z, w) = \frac{1}{2} \log\left(\frac{|1 - z\overline{w}| + |z - w|}{|1 - z\overline{w}| - |z - w|}\right), \quad z, w \in \mathbb{D}. \tag{2-14}$$

The Cayley–Klein distance function on the ball $\mathbb{B}^n$ coincides up to a scalar factor $\sqrt{n+1}$ with the restriction to $\mathbb{B}^n$ of the Bergman distance function on the complex ball $\mathbb{B}^n_{\mathbb{C}}$ or, equivalently, with the restriction to $\mathbb{B}^n$ of the Kobayashi distance function on $\mathbb{B}^n_{\mathbb{C}}$. The following explicit formula for the Kobayashi distance between a pair of points $\boldsymbol{z}, \boldsymbol{w} \in \mathbb{B}^n_{\mathbb{C}}$ can be found in [Krantz 1992, p. 437]; here, $\boldsymbol{z} \cdot \overline{\boldsymbol{w}} = \sum_{k=1}^n z_k \overline{w}_k$:

$$\mathrm{dist}(\boldsymbol{z}, \boldsymbol{w}) = \frac{1}{2} \log\left(\frac{|1 - \boldsymbol{z} \cdot \overline{\boldsymbol{w}}| + \sqrt{|\boldsymbol{z} - \boldsymbol{w}|^2 + |\boldsymbol{z} \cdot \overline{\boldsymbol{w}}|^2 - |\boldsymbol{z}|^2 |\boldsymbol{w}|^2}}{|1 - \boldsymbol{z} \cdot \overline{\boldsymbol{w}}| - \sqrt{|\boldsymbol{z} - \boldsymbol{w}|^2 + |\boldsymbol{z} \cdot \overline{\boldsymbol{w}}|^2 - |\boldsymbol{z}|^2 |\boldsymbol{w}|^2}}\right). \tag{2-15}$$

As said before, the same formula applied to points in $\mathbb{B}^n$ gives the Cayley–Klein distance. Taking $w = 0$ and $\boldsymbol{w} = 0$ in the above formulas, we obtain

$$\mathrm{dist}_{\mathcal{P}}(z, 0) = \frac{1}{2} \log\left(\frac{1 + |z|}{1 - |z|}\right) \ (z \in \mathbb{D}), \quad \mathrm{dist}(\boldsymbol{z}, \boldsymbol{0}) = \frac{1}{2} \log\left(\frac{1 + |\boldsymbol{z}|}{1 - |\boldsymbol{z}|}\right) \ (\boldsymbol{z} \in \mathbb{B}^n).$$

Together with Theorem 2.6 this implies the following corollary.

Corollary 2.7. *If $f : \mathbb{D} \to \mathbb{B}^n$, $n \geq 3$, is a conformal harmonic map with $f(0) = \mathbf{0}$, then $|f(z)| \leq |z|$ for all $z \in \mathbb{D}$. Equality at one point $z \in \mathbb{D} \setminus \{0\}$ implies that f is a conformal parametrization of a linear disc obtained by intersecting $\mathbb{B}^n$ with a plane through the origin, and hence equality holds at all points.*

3. Proof of Theorem 2.1

It suffices to prove Theorem 2.1 for $z = 0$. Indeed, with f and z as in the theorem, let $\phi_z \in \mathrm{Aut}(\mathbb{D})$ be such that $\phi_z(0) = z$. The harmonic map $g = f \circ \phi_z : \mathbb{D} \to \mathbb{B}^n$ is then conformal at the origin. Since $|\phi_z'(0)| = 1 - |z|^2$, inequality (2-3) follows from the same estimate for g at $z = 0$. On the image side, the hypotheses and the statement of the theorem are invariant under postcomposition of maps $\mathbb{D} \to \mathbb{B}^n$ by elements of the orthogonal group O_n.

We begin with an explicit description of conformal parametrizations of proper affine discs in $\mathbb{B}^n$. Fix a point $\boldsymbol{q} \in \mathbb{B}^n$ and a linear two-plane $0 \in \Lambda \subset \mathbb{R}^n$, and consider the affine disc $\Sigma = (\boldsymbol{q} + \Lambda) \cap \mathbb{B}^n$. Let us identify conformal parametrizations $\mathbb{D} \to \Sigma$ sending 0 to $\boldsymbol{q}$. Let $\boldsymbol{p} \in \Sigma$ be the closest point to the origin. If $n = 2$ then $\boldsymbol{p} = 0$ and $\Sigma = \mathbb{D}$. Suppose now that $n \geq 3$. Up to an orthogonal rotation, we may assume

$$\boldsymbol{p} = (0, 0, p, 0, \dots, 0) \quad \text{and} \quad \Sigma = \{(x, y, p, 0, \dots, 0) : x^2 + y^2 < 1 - p^2\}. \tag{3-1}$$

Let $\boldsymbol{q} = (b_1, b_2, p, 0, \dots, 0) \in \Sigma$, and let θ denote the angle between $\boldsymbol{q}$ and Σ. Set

$$c = \sqrt{1 - p^2} = \sqrt{1 - |\boldsymbol{q}|^2 \sin^2\theta}, \quad a = \frac{b_1 + \mathrm{i} b_2}{c} \in \mathbb{D}, \quad |a| = \frac{|\boldsymbol{q}| \cos\theta}{c}. \tag{3-2}$$

We orient Σ by the tangent vectors ∂_x, ∂_y in the parametrization (3-1). Every orientation-preserving conformal parametrization $f : \mathbb{D} \to \Sigma$ with $f(0) = \boldsymbol{q}$ is then of the form

$$f(z) = \left(c\Re \frac{\mathrm{e}^{\mathrm{i}t} z + a}{1 + \bar{a}\mathrm{e}^{\mathrm{i}t} z}, c\Im \frac{\mathrm{e}^{\mathrm{i}t} z + a}{1 + \bar{a}\mathrm{e}^{\mathrm{i}t} z}, p, 0, \dots, 0 \right), \quad z \in \mathbb{D}, \tag{3-3}$$

for some $t \in \mathbb{R}$. (Here, $\Re$ and $\Im$ stand for the real and imaginary parts of a complex number. If $n = 2$ then $p = 0$ and $c = 1$, and the same holds if we drop all coordinates except the first two. Orientation-reversing conformal parametrizations are obtained by replacing $z = x + \mathrm{i}y$ with $\bar{z} = x - \mathrm{i}y$. By a rotation in the (x, y)-plane, we may further assume that $b_2 = 0$ and $f(0) = (b_1, 0, p, 0, \dots, 0)$; in this case $a \in [0, 1)$. By also allowing rotations on the disc $\mathbb{D}$, we can take $t = 0$ in (3-3).) Using the complex coordinate $x + \mathrm{i}y$ in the plane $df_0(\mathbb{R}^2) = \mathbb{R}^2 \times \{0\}^{n-2}$, the map (3-3) can be written in the form

$$f(z) = \left(c \frac{\mathrm{e}^{\mathrm{i}t} z + a}{1 + \bar{a}\mathrm{e}^{\mathrm{i}t} z}, p, 0, \dots, 0 \right) = (h(z), p, 0, \dots, 0).$$

From (3-2) it follows that

$$\begin{aligned} |h'(0)| = c(1 - |a|^2) = \frac{c^2 - c^2|a|^2}{c} &= \frac{1 - |\boldsymbol{q}|^2 \sin^2\theta - |\boldsymbol{q}|^2 \cos^2\theta}{c} \\ &= \frac{1 - |\boldsymbol{q}|^2}{\sqrt{1 - |\boldsymbol{q}|^2 \sin^2\theta}} = \frac{1 - |f(0)|^2}{\sqrt{1 - |f(0)|^2 \sin^2\theta}}. \end{aligned}$$

Since $\|df_0\| = |h'(0)|$, this gives equality in (2-3) at $z = 0$.

Theorem 2.1 now follows immediately from the following lemma.

Lemma 3.1. *Let* $f : \mathbb{D} \to \mathbb{B}^n$ $(n \geq 2)$ *be the disc* (3-3). *If* $g : \mathbb{D} \to \mathbb{B}^n$ *is a harmonic disc such that* $g(0) = f(0)$, g *is conformal at* 0, *and* $dg_0(\mathbb{R}^2) = df_0(\mathbb{R}^2)$, *then* $\|dg_0\| \leq \|df_0\|$, *with equality if and only if* $g(z) = f(\mathrm{e}^{\mathrm{i}s}z)$ *or* $g(z) = f(\mathrm{e}^{\mathrm{i}s}\bar{z})$ *for some* $s \in \mathbb{R}$ *and all* $z \in \mathbb{D}$.

The proof of Lemma 3.1 uses ideas from Lempert's seminal paper [1981] concerning complex geodesics of the Kobayashi metric in convex domains in $\mathbb{C}^n$; see Remark 3.2.

Proof. Let p, c and a be as in (3-2) related to the map f in (3-3), where $\boldsymbol{q} = f(0)$. Precomposing f by a rotation in $\mathbb{C}$, we may assume that $t = 0$ in (3-3). For simplicity of notation we assume that $n = 3$; the proof for $n \neq 3$ is exactly the same. If $n = 2$, we delete the remaining components and take $c = 1$.

Consider the holomorphic disc $F : \mathbb{D} \to \Omega = \mathbb{B}^3 \times \mathrm{i}\mathbb{R}^3$ given by

$$F(z) = \left(c\frac{z+a}{1+\bar{a}z}, -c\mathrm{i}\frac{z+a}{1+\bar{a}z}, p \right), \quad z \in \mathbb{D}. \tag{3-4}$$

Then, $f = \Re F$. Suppose that $g : \mathbb{D} \to \mathbb{B}^3$ is as in the lemma. Up to replacing $g(z)$ by $g(\mathrm{e}^{\mathrm{i}s}z)$ or $g(\mathrm{e}^{\mathrm{i}s}\bar{z})$ for a suitable $s \in \mathbb{R}$, we may assume that

$$dg_0 = r\,df_0 \quad \text{for some } r > 0. \tag{3-5}$$

We must prove that $r \leq 1$, and that $r = 1$ if and only if $g = f$.

Let $G : \mathbb{D} \to \Omega$ be the unique holomorphic map with $\Re G = g$ and $G(0) = F(0)$. In view of the Cauchy–Riemann equations, condition (3-5) implies

$$G'(0) = rF'(0), \tag{3-6}$$

where the prime denotes the complex derivative. It follows that the map $(F(z) - G(z))/z$ is holomorphic on $\mathbb{D}$, and its value at $z = 0$ equals

$$\lim_{z\to 0} \frac{F(z) - G(z)}{z} = F'(0) - G'(0) = (1 - r)F'(0). \tag{3-7}$$

The bounded harmonic map $g : \mathbb{D} \to \mathbb{B}^3$ has a nontangential boundary value at almost every point of the circle $\mathbb{T} = b\mathbb{D}$. Since the Hilbert transform is an isometry on the Hilbert space $L^2(\mathbb{T})$, the same is true for its holomorphic extension G; see [Garnett 1981].

Denote by $\langle \cdot , \cdot \rangle$ the complex bilinear form on $\mathbb{C}^n$ given by $\langle \boldsymbol{z}, \boldsymbol{w} \rangle = \sum_{i=1}^n z_i w_i$ for $\boldsymbol{z}, \boldsymbol{w} \in \mathbb{C}^n$. Note that on vectors in $\mathbb{R}^n$ this is the Euclidean inner product. For each $z = \mathrm{e}^{\mathrm{i}t} \in b\mathbb{D}$ the vector $f(z) \in b\mathbb{B}^3$ is the unit normal vector to the sphere $b\mathbb{B}^3$ at the point $f(z)$. Since $\mathbb{B}^3$ is strongly convex and f is real-valued, we have

$$\Re\langle F(z) - G(z), f(z)\rangle = \langle f(z) - g(z), f(z)\rangle \geq 0 \quad a.e.\ z \in b\mathbb{D}, \tag{3-8}$$

and the value is positive for almost every $z \in b\mathbb{D}$ if and only if $g \neq f$. It is at this point that strong convexity of the ball $\mathbb{B}^3$ is used in an essential way.

We now consider the map $\tilde{f}$ on the circle $b\mathbb{D}$ given by

$$\tilde{f}(z) = z|1+\bar{a}z|^2 f(z), \quad |z| = 1. \tag{3-9}$$

An explicit calculation, taking into account $z\bar{z} = 1$, shows that

$$\tilde{f}(z) = \begin{pmatrix} \frac{1}{2}c(1+a^2+4(\Re a)z+(1+\bar{a}^2)z^2) \\ \frac{1}{2}c(\mathrm{i}(1-a^2)+4(\Im a)z+\mathrm{i}(\bar{a}^2-1)z^2) \\ p(z+a)(1+\bar{a}z) \end{pmatrix}. \tag{3-10}$$

We extend $\tilde{f}$ to all $z \in \mathbb{C}$ by letting it equal the quadratic holomorphic polynomial map on the right-hand side above. Since $|1+\bar{a}z|^2 > 0$ for $z \in \overline{\mathbb{D}}$, (3-8) implies

$$h(z) := \Re\langle F(z) - G(z), |1+\bar{a}z|^2 f(z)\rangle = \langle f(z) - g(z), |1+\bar{a}z|^2 f(z)\rangle \geq 0 \quad a.e.\ z \in b\mathbb{D},$$

and $h > 0$ almost everywhere on $b\mathbb{D}$ if and only if $g \neq f$. From (3-9) we see that

$$h(z) = \Re\left\langle \frac{F(z) - G(z)}{z}, \tilde{f}(z)\right\rangle \quad a.e.\ z \in b\mathbb{D}. \tag{3-11}$$

Since the maps $(F(z) - G(z))/z$ and $\tilde{f}(z)$ are holomorphic on $\mathbb{D}$, formula (3-11) provides an extension of h from $b\mathbb{D}$ to a nonnegative harmonic function on $\mathbb{D}$ which is positive on $\mathbb{D}$ unless $f = g$. Inserting the value (3-7) into (3-11) gives

$$h(0) = \Re\langle F'(0) - G'(0), \tilde{f}(0)\rangle = (1-r)\Re\langle F'(0), \tilde{f}(0)\rangle \geq 0,$$

with equality if and only if $f = g$. Applying this argument to the linear map $g(z) = f(0) + r\,df_0(z)$ ($z \in \mathbb{D}$) for a small $r > 0$ we get $\Re\langle F'(0), \tilde{f}(0)\rangle > 0$. It follows that $r \leq 1$, with equality if and only if $g = f$. □

Remark 3.2. The main point in the above proof is that the complexification of a conformal proper affine disc in $\mathbb{B}^n$ is a *stationary disc* in the tube $\mathcal{T}_{\mathbb{B}^n} = \mathbb{B}^n \times \mathrm{i}\mathbb{R}^n$. In Lempert's terminology [1981], a proper holomorphic disc $F : \mathbb{D} \to \Omega$ in a smoothly bounded convex domain $\Omega \subset \mathbb{C}^n$, extending continuously to $\overline{\mathbb{D}}$, is a stationary disc if, denoting by $\nu : b\mathbb{D} \to \mathbb{C}^n$ the unit normal vector field to $b\Omega$ along the boundary circle $F(b\mathbb{D}) \subset b\Omega$, there is a positive continuous function $q > 0$ on $b\mathbb{D}$ such that the function $zq(z)\overline{\nu(z)}$ extends from the circle $|z| = 1$ to a holomorphic function $\tilde{f}(z)$ on $\mathbb{D}$. Lempert [1981] showed that every stationary disc F in a bounded strongly convex domain is the unique Kobayashi extremal disc through the point $F(a)$ in the tangent direction $F'(a)$ for every $a \in \mathbb{D}$. In our case, a suitable holomorphic function $\tilde{f}$ is given by (3-9) and (3-10). Lempert's theory also works on tubes over bounded strongly convex domains (see [Jarnicki and Pflug 2013, Section 11.1]); however, our proof of Theorem 2.1 does not depend on this information.

4. Proof of Theorem 2.2

Precomposing the given harmonic map $f : \mathbb{D} \to \mathbb{B}^n$ in Theorem 2.2 by a holomorphic automorphism of the disc $\mathbb{D}$, we see that it suffices to prove estimate (2-4) for $z = 0$.

Assume first that $f:\mathbb{D}\to\mathbb{R}$ is a harmonic function on $\mathbb{D}$. Let $F(z)=a_0+a_1z+a_2z^2+\cdots$ be the holomorphic function on $\mathbb{D}$ with $\Re F=f$ and $F(0)=f(0)\in\mathbb{R}$. Writing $z=re^{\mathrm{i}t}$ with $0\le r<1$ and $t\in\mathbb{R}$, we have

$$\begin{aligned} f(re^{\mathrm{i}t})^2 &= \tfrac{1}{4}(a_0+a_1re^{\mathrm{i}t}+r^2e^{2\mathrm{i}t}+\cdots+a_0+\bar a_1re^{-\mathrm{i}t}+\bar a_2r^2e^{-2\mathrm{i}t}+\cdots)^2 \\ &= a_0^2+\frac{1}{2}\sum_{k=1}^{\infty}r^{2k}|a_k|^2+\cdots, \end{aligned}$$

where each of the remaining terms in the series contains a power $e^{m\mathrm{i}t}$ for some $m\in\mathbb{Z}\setminus\{0\}$. Integrating around the circle $|z|=r$ for $0<r<1$ annihilates all such terms and yields

$$\int_0^{2\pi} f(re^{\mathrm{i}t})^2\,\frac{dt}{2\pi}=a_0^2+\frac{1}{2}\sum_{k=1}^{\infty}r^{2k}|a_k|^2.$$

Clearly, $a_0=f(0)$. Writing $z=x+\mathrm{i}y$, we have that $a_1=F'(0)=F_x(0)=f_x(0)-\mathrm{i}f_y(0)$ by the Cauchy–Riemann equations. Therefore,

$$a_0^2=f(0)^2\quad\text{and}\quad |a_1|^2=f_x(0)^2+f_y(0)^2=|\nabla f(0)|^2,$$

and hence

$$\int_0^{2\pi} f(re^{\mathrm{i}t})^2\,\frac{dt}{2\pi}=|f(0)|^2+\frac{1}{2}|\nabla f(0)|^2r^2+\frac{1}{2}\sum_{k=2}^{\infty}r^{2k}|a_k|^2. \tag{4-1}$$

Suppose now that $f=(f_1,\dots,f_n):\mathbb{D}\to\mathbb{B}^n$ is a harmonic map. Then, $\sum_{j=1}^n f_j(re^{\mathrm{i}t})^2<1$ for all $0\le r<1$ and $t\in\mathbb{R}$. Integrating this inequality and taking into account the identity (4-1) for each component f_j of f gives

$$\int_0^{2\pi} |f(re^{\mathrm{i}t})|^2\,\frac{dt}{2\pi}=|f(0)|^2+\frac{1}{2}|\nabla f(0)|^2r^2+\frac{1}{2}\sum_{k=2}^{\infty}r^{2k}|a_k|^2<1.$$

Letting r increase to 1 gives $|f(0)|^2+\frac{1}{2}|\nabla f(0)|^2\le 1$, with equality if and only if all higher-order coefficients in the Fourier expansion of f vanish. The latter holds if and only if f is a linear disc. This gives the estimate (2-4).

Note that (2-4) holds if the L^2-Hardy norm of f is at most 1. This does not necessarily imply that there is a harmonic disc in $\mathbb{B}^n$ reaching equality in (2-4). However, equality is attained if $f(0)$ is orthogonal to the two-plane $df_0(\mathbb{R}^2)$. In this case we may assume that $f(0)=(0,0,p,0,\dots,0)$ for some $0\le p<1$ and $df_0(\mathbb{R}^2)=\mathbb{R}^2\times\{0\}^{n-2}$. The affine disc

$$\Sigma=\{(x,y,p,0,\dots,0):x^2+y^2<1-p^2\}$$

of radius $c=\sqrt{1-p^2}$ is then orthogonal to $f(0)$, proper in $\mathbb{B}^n$, and its conformal linear parametrization f has gradient of size $c\sqrt{2}$ at the origin, so $|f(0)|^2+\frac{1}{2}|\nabla f(0)|^2=p^2+c^2=1$. (Compare with (3-1) and (3-3).) This completes the proof of Theorem 2.2. □

We now show by examples that the inequality (2-3) fails in general for some nonconformal harmonic maps, and even for harmonic diffeomorphisms of the disc.

Example 4.1. Let U be the harmonic function on the disc $\mathbb{D}$ given by

$$U(z) = \Im \frac{2}{\pi} \log \frac{1+z}{1-z} = \frac{2}{\pi} \arctan \frac{2y}{1-x^2-y^2}. \tag{4-2}$$

This is the extremal harmonic function whose boundary value equals $+1$ on the upper unit semicircle and -1 on the lower semicircle, and we have that $\nabla U(0) = \frac{4}{\pi}(0, 1)$ and $|\nabla U(0)| = \frac{4}{\pi}$. For every $c \in \mathbb{R}$ the harmonic map

$$f(z) = \frac{1}{\sqrt{1+|c|^2}}(c + \mathrm{i}U(z)), \qquad z \in \mathbb{D},$$

clearly takes the unit disc into itself. For $c = 1$ we have

$$f(0) = \frac{1}{\sqrt{2}}, \quad \nabla f(0) = \frac{2\sqrt{2}}{\pi}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad |\nabla f(0)| = \frac{2\sqrt{2}}{\pi} \approx 0.9, \quad \sqrt{2}(1 - |f(0)|^2) = \frac{\sqrt{2}}{2} \approx 0.7.$$

Hence, inequality (2-3) fails in this example. On the other hand, $\sqrt{2}\sqrt{1-|f(0)|^2} = 1$, so inequality (2-4) holds, as it should by Theorem 2.2.

With some more effort we can show that inequality (2-3) fails for harmonic diffeomorphisms of the unit disc onto itself. Consider the sequence φ_n, $n \in \mathbb{N}$, of orientation-preserving homeomorphisms of the interval $[0, 2\pi]$ onto itself, defined by

$$\varphi_n(t) = \begin{cases} \dfrac{\pi}{2\pi - 1/n} t & \text{if } t \in \left[0, 2\pi - \frac{1}{n}\right], \\ 2(\pi - n\pi^2) + n\pi t & \text{if } t \in \left[2\pi - \frac{1}{n}, 2\pi\right]. \end{cases}$$

Let $\phi_n : \mathbb{T} \to \mathbb{T}$ be the associated sequence of homeomorphisms of the circle $\mathbb{T} = b\mathbb{D}$ given by $\phi_n(\mathrm{e}^{\mathrm{i}t}) = \mathrm{e}^{\mathrm{i}\varphi_n(t)}$ for $t \in [0, 2\pi]$. Denote by

$$f_n(z) = P[\phi_n](z) = \frac{1}{2\pi}\int_0^{2\pi} \frac{1-|z|^2}{|\mathrm{e}^{\mathrm{i}t} - z|^2}\phi_n(\mathrm{e}^{\mathrm{i}t})\, dt, \qquad z \in \mathbb{D},$$

the Poisson extension of ϕ_n. By the Radó–Kneser–Choquet theorem (see [Duren 2004, Section 3.1]), f_n is a harmonic diffeomorphism of $\mathbb{D}$ for every $n \in \mathbb{N}$. As $n \to \infty$, the sequence f_n converges uniformly on compacts in $\mathbb{D}$ to the harmonic map $f = P[\phi_0]$, where $\phi_0(\mathrm{e}^{\mathrm{i}t}) = \lim_{n\to\infty} \phi_n(\mathrm{e}^{\mathrm{i}t}) = \mathrm{e}^{\mathrm{i}t/2}$ for $t \in [0, 2\pi)$. Further, we have

$$\lim_{n\to\infty} \frac{|\nabla f_n(0)|}{1 - |f_n(0)|^2} = \frac{|\nabla f(0)|}{1-|f(0)|^2}.$$

A calculation shows that

$$\frac{1}{\sqrt{2}}\frac{|\nabla f(0)|}{1-|f(0)|^2} = \frac{\sqrt{|A|^2 + |B^2|}}{1-|C|^2},$$

where

$$A = \frac{1}{\pi}\int_0^{2\pi} \mathrm{e}^{\mathrm{i}t/2}\cos t\, dt = -\frac{4\mathrm{i}}{3\pi}, \quad B = \frac{1}{\pi}\int_0^{2\pi} \mathrm{e}^{\mathrm{i}t/2}\sin t\, dt = \frac{8}{3\pi}, \quad C = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{e}^{\mathrm{i}t/2}\, dt = \frac{2\mathrm{i}}{\pi}.$$

Hence,

$$\frac{1}{\sqrt{2}}\frac{|\nabla f(0)|}{1-|f(0)|^2} = \frac{2\sqrt{10}}{3\pi(1-4/\pi^2)} \approx 1.1.$$

This shows that (2-3) fails for harmonic diffeomorphisms of the unit disc onto itself.

Example 4.2. Let $U(x, y)$ be the function (4-2). The harmonic map $f(x, y) = (U(y, x), U(x, y))$ takes the disc $\mathbb{D}$ onto the square $P = \{(x, y) \in \mathbb{R}^2 : |x| < 1,\ |y| < 1\}$ and $df_0(0, 0) = \frac{4}{\pi}\mathrm{Id}$. In particular, f is conformal at $(0, 0)$ and $\|df_0\| = \frac{4}{\pi} \approx 1.27$. On the other hand, a conformal diffeomorphism of $\mathbb{D}$ onto P mapping the origin to itself has the derivative at the origin of absolute value ≈ 1.08. Hence, the Schwarz–Pick lemma in Theorem 1.1 fails for maps from the disc to more general domains in $\mathbb{C}$. In particular, while the domain $\mathbb{C} \setminus \{0, 1\}$ is Kobayashi hyperbolic, one can find nonconstant harmonic maps $\mathbb{C} \to \mathbb{C} \setminus \{0, 1\}$ whose differential is nonvanishing and conformal at the origin.

Problem 4.3. Assume that $D \subsetneq \mathbb{R}^2$ is a simply connected domain such that, for some point $p \in D$, the supremum of the norm $\|df_0\|$ of the differential of f at $0 \in \mathbb{D}$ over all harmonic maps $f : \mathbb{D} \to D$ with $f(0) = p$ which are conformal at 0 is attained by a conformal diffeomorphism of $\mathbb{D}$ onto D. Does it follow that D is a disc?

5. A Schwarz–Pick lemma for quasiconformal harmonic maps

In this section we apply the Schwarz–Pick lemma for harmonic self-maps of the disc, given by Theorem 1.1, to provide an estimate of the gradient of an orientation-preserving harmonic map $f : \mathbb{D} \to \mathbb{D}$ in terms of its *second Beltrami coefficient*

$$\omega(z) = \frac{\overline{(f_{\bar z})}}{f_z}, \quad z \in \mathbb{D}. \tag{5-1}$$

Here, $f_z = \frac{1}{2}(f_x - \mathrm{i} f_y)$ and $f_{\bar z} = \frac{1}{2}(f_x + \mathrm{i} f_y)$. If the map f is harmonic then ω is a holomorphic function; see (5-2). This is not the case for the Beltrami coefficient μ from the Beltrami equation $f_{\bar z} = \mu(z) f_z$. The number $|\mu(z)| = |\omega(z)|$ measures the dilatation of df_z; in particular, $\mu(z) = \omega(z) = 0$ if and only if f is conformal at z. We refer to [Ahlfors 1966; Duren 2004; Lehto and Virtanen 1973; Hengartner and Schober 1986] for background on the theory of quasiconformal maps.

The main question is to find the optimal estimate on $\|df_0\|$ for a harmonic map $f : \mathbb{D} \to \mathbb{D}$ with $f(0) = 0$ and with a given value of $|\omega(0)| = |\mu(0)|$. A related problem was studied by Kovalev and Yang [2020] and Brevig et al. [2021], where the reader can find references to earlier works. Here we prove the following result.

Theorem 5.1. *Assume that f is an orientation-preserving harmonic map of the unit disc into itself, and let $\omega(z)$ denote its second Beltrami coefficient* (5-1). *Then we have the inequality*

$$\|df_z\| \le \frac{2(|\omega(z) f(z)^2| + \Re(\omega(z) f(z)^2))}{(1 - |\omega(z)|^2)(1 - |z|^2)} + \frac{1 + |\omega(z)|}{1 - |\omega(z)|}\frac{1 - |f(z)|^2}{1 - |z|^2}, \quad z \in \mathbb{D}.$$

If f is conformal at a point z, i.e., $\omega(z) = 0$, this estimate coincides with the Schwarz–Pick inequality (1-1) in Theorem 1.1.

Proof. It suffices to prove the inequality in the theorem for $z = 0$. For other points, we obtain it replacing f by $f \circ \phi_z$ for $\phi_z \in \mathrm{Aut}(\mathbb{D})$. However, we cannot reduce to the case $f(0) = 0$ since postcompositions by automorphisms of $\mathbb{D}$ are not allowed. The main idea is to construct from f a new harmonic map $\tilde f : \mathbb{D} \to \mathbb{D}$ which is conformal at 0, to which we then apply the Schwarz–Pick lemma given by Theorem 1.1.

Let us write $f = g + \bar{h}$, where g and h are holomorphic functions on $\mathbb{D}$. Then,

$$f_z(z) = g'(z), \quad f_{\bar{z}}(z) = \overline{h'(z)}, \quad \omega(z) = \frac{h'(z)}{g'(z)}. \tag{5-2}$$

Recall that the second Beltrami coefficient ω (5-1) is holomorphic. It follows that

$$\|df\| = |g'|(1 + |\omega|). \tag{5-3}$$

Since f is orientation preserving, we have that $|g'(z)| \geq |h'(z)|$ for all $z \in \mathbb{D}$. Let

$$a = g'(0) \quad \text{and} \quad b = h'(0). \tag{5-4}$$

We may assume that $|a| + |b| > 0$, for otherwise the estimate is trivial. Since $|f(z)| < 1$ for all $z \in \mathbb{D}$, the complex harmonic function

$$\tilde{f}(z) := \frac{\bar{a}f - \bar{b}\bar{f}}{|a| + |b|}, \quad z \in \mathbb{D}, \tag{5-5}$$

clearly maps the unit disc into itself. We have $\tilde{f} = \tilde{g} + \bar{\tilde{h}}$, where

$$\tilde{g} = \frac{\bar{a}g - \bar{b}h}{|a| + |b|} \quad \text{and} \quad \tilde{h} = \frac{ah - bg}{|a| + |b|} \tag{5-6}$$

are holomorphic functions on $\mathbb{D}$. Since

$$\tilde{h}'(0) = \frac{ah'(0) - bg'(0)}{|a| + |b|} = 0, \tag{5-7}$$

the second Beltrami coefficient $\tilde{\omega}$ of $\tilde{f}$ vanishes at $z = 0$, and hence $\tilde{f}$ is conformal at 0. Our Schwarz–Pick lemma (see Theorem 1.1) gives

$$\|d\tilde{f}_0\| \leq 1 - |\tilde{f}(0)|^2. \tag{5-8}$$

Taking into account (5-3), (5-4), (5-6), and (5-7), we have

$$\|d\tilde{f}_0\| = |\tilde{g}'(0)| = |g'(0)| - |h'(0)|.$$

Together with (5-5), (5-4), and (5-8) this gives the estimate

$$\begin{aligned} |g'(0)| - |h'(0)| &\leq 1 - \frac{|g'(0)\overline{f(0)} - h'(0)f(0)|^2}{(|g'(0)| + |h'(0)|)^2} \\ &= \frac{2|g'(0)| \cdot |h'(0)| \cdot |f(0)|^2 + 2\Re(\overline{g'(0)}h'(0)f(0)^2)}{(|g'(0)| + |h'(0)|)^2} + 1 - |f(0)|^2. \end{aligned}$$

In view of (5-2), this inequality can be written in the form

$$(1 - |\omega(0)|)|g'(0)| \leq \frac{2(|\omega(0)|\,|f(0)|^2 + \Re(\omega(0)f(0)^2))}{(1 + |\omega(0)|)^2} + 1 - |f(0)|^2.$$

From (5-3) we see that

$$|g'(0)| = \frac{\|df_0\|}{1 + |\omega(0)|}.$$

Inserting this into the expression on the left-hand side of the previous inequality gives

$$\|df_0\|\frac{1-|\omega(0)|}{1+|\omega(0)|} \le \frac{2|\omega(0)f(0)^2| + 2\Re(\omega(0)f(0)^2)}{(1+|\omega(0)|)^2} + 1 - |f(0)|^2,$$

which is clearly equivalent to

$$\|df_0\| \le \frac{2|\omega(0)f(0)^2| + 2\Re(\omega(0)f(0)^2)}{1-|\omega(0)|^2} + \frac{1+|\omega(0)|}{1-|\omega(0)|}(1-|f(0)|^2). \qquad \square$$

6. An intrinsic pseudometric defined by conformal harmonic discs

In this section we introduce an intrinsic Finsler pseudometric g_D on any domain D in $\mathbb{R}^n$, $n \ge 3$, and more generally on any Riemannian manifold of dimension at least three, in terms of conformal minimal discs $\mathbb{D} \to D$. The definition is modeled on Kobayashi's definition of his pseudometric on complex manifolds, which uses holomorphic discs. The pseudometric g_D and the associated pseudodistance $\rho_D : D \times D \to \mathbb{R}_+$ are the largest ones having the property that any conformal harmonic map $M \to D$ from a hyperbolic conformal surface with the Poincaré metric is metric and distance decreasing. On the ball $\mathbb{B}^n$, we have that $g_{\mathbb{B}^n}$ coincides with the Cayley–Klein metric; see Theorem 6.2. The same definition of g_D applies in any Riemannian manifold of dimension at least three; see Remark 6.6. This provides the basis for hyperbolicity theory of domains in Euclidean spaces and, more generally, of Riemannian manifolds, in terms of minimal surfaces.

We begin by introducing a Finsler pseudometric on the bundle of two-planes over a domain $D \subset \mathbb{R}^n$, analogous to the metric $\mathcal{M}$ on the ball; see (2-8). A *conformal frame* in $\mathbb{R}^n$ is a pair $(\boldsymbol{u}, \boldsymbol{v}) \in \mathbb{R}^n \times \mathbb{R}^n$ such that $|\boldsymbol{u}| = |\boldsymbol{v}|$ and $\boldsymbol{u} \cdot \boldsymbol{v} = 0$. We denote by CF_n the space of all conformal frames on $\mathbb{R}^n$, including $(\mathbf{0}, \mathbf{0})$. Given a domain $D \subset \mathbb{R}^n$, let $\mathrm{CH}(\mathbb{D}, D)$ denote the space of conformal harmonic maps $\mathbb{D} \to D$ (i.e., such that (2-1) holds at every point of $\mathbb{D}$). Define the function $\mathcal{M}_D : D \times \mathrm{CF}_n \to \mathbb{R}_+$ by

$$\mathcal{M}_D(\boldsymbol{x}, (\boldsymbol{u}, \boldsymbol{v})) = \inf\left\{\frac{1}{r} : \exists f \in \mathrm{CH}(\mathbb{D}, D),\ f(0) = \boldsymbol{x},\ f_x(0) = r\boldsymbol{u},\ f_y(0) = r\boldsymbol{v}\right\}. \tag{6-1}$$

Clearly, $\mathcal{M}_D$ is homogeneous and rotation-invariant, in the sense that for any $c \in \mathbb{R}$ and orthogonal rotation R in the two-plane $\Lambda = \mathrm{span}\{\boldsymbol{u}, \boldsymbol{v}\}$ we have for every $\boldsymbol{x} \in D$ that

$$\mathcal{M}_D(\boldsymbol{x}, (c\boldsymbol{u}, c\boldsymbol{v})) = |c|\mathcal{M}_D(\boldsymbol{x}, (\boldsymbol{u}, \boldsymbol{v})), \quad \mathcal{M}_D(\boldsymbol{x}, (R\boldsymbol{u}, R\boldsymbol{v})) = \mathcal{M}_D(\boldsymbol{x}, (\boldsymbol{u}, \boldsymbol{v})). \tag{6-2}$$

Thus, $\mathcal{M}_D$ is determined by its values on unitary conformal frames $(\boldsymbol{u}, \boldsymbol{v})$ with $|\boldsymbol{u}| = |\boldsymbol{v}| = 1$ and hence on $D \times G_2(\mathbb{R}^n)$, where $G_2(\mathbb{R}^n)$ is the Grassmann manifold of two-planes in $\mathbb{R}^n$. Precisely, for a two-plane $\Lambda \in G_2(\mathbb{R}^n)$ we set $\mathcal{M}_D(\boldsymbol{x}, \Lambda) = \mathcal{M}_D(\boldsymbol{x}, (\boldsymbol{u}, \boldsymbol{v}))$, where $(\boldsymbol{u}, \boldsymbol{v})$ is any unitary conformal frame spanning Λ. Note that

$$\mathcal{M}_D(\boldsymbol{x}, \Lambda) = \inf\left\{\frac{1}{\|df_0\|} : f \in \mathrm{CH}(\mathbb{D}, D),\ f(0) = \boldsymbol{x},\ df_0(\mathbb{R}^2) = \Lambda\right\}. \tag{6-3}$$

By shrinking the disc $\mathbb{D}$ and using rotations and translations on $\mathbb{R}^n$, we see that the function $\mathcal{M}_D$ is upper semicontinuous on $D \times \mathrm{CF}_n$. Obviously, $\mathcal{M}_{\mathbb{R}^n} \equiv 0$. On the ball $\mathbb{B}^n$, we have that $\mathcal{M}_{\mathbb{B}^n}(\boldsymbol{x}, \Lambda)$ is given by (2-8) according to Theorem 2.1.

We also introduce a Finsler pseudometric $g_D : D \times \mathbb{R}^n \to \mathbb{R}_+$, called the *minimal metric* on D, whose value at a point $\boldsymbol{x} \in D$ on a tangent vector $\boldsymbol{u} \in T_{\boldsymbol{x}} D = \mathbb{R}^n$ is given by

$$\begin{aligned} g_D(\boldsymbol{x}, \boldsymbol{u}) &= \inf\left\{\frac{1}{r} > 0 : \exists f \in \mathrm{CH}(\mathbb{D}, D),\ f(0) = \boldsymbol{x},\ f_x(0) = r\boldsymbol{u}\right\} \\ &= |\boldsymbol{u}| \cdot \inf\{\mathcal{M}_D(\boldsymbol{x}, \Lambda) : \Lambda \in G_2(\mathbb{R}^n),\ \boldsymbol{u} \in \Lambda\}. \end{aligned} \tag{6-4}$$

It follows that every conformal harmonic map $f : \mathbb{D} \to D$ satisfies

$$g_D(f(z), df_z(\xi)) \le \mathcal{P}(z, \xi) = \frac{|\xi|}{1 - |z|^2}, \quad z \in \mathbb{D},\ \xi \in \mathbb{R}^2, \tag{6-5}$$

and g_D is the biggest pseudometric on D with this property. For $z = 0$ this follows directly from the definition, and for any other point $z \in \mathbb{D}$ we precompose f by a conformal automorphism of $\mathbb{D}$ mapping 0 to z. The same holds if $\mathbb{D}$ is replaced by any hyperbolic conformal surface; see the proof of Theorem 2.6.

By integrating g_D we get the *minimal pseudodistance* $\rho_D : \Omega \times \Omega \to \mathbb{R}_+$:

$$\rho_D(\boldsymbol{x}, \boldsymbol{y}) = \inf_\gamma \int_0^1 g_D(\gamma(t), \dot\gamma(t))\, dt, \quad \boldsymbol{x}, \boldsymbol{y} \in \Omega. \tag{6-6}$$

The infimum is over all piecewise smooth paths $\gamma : [0, 1] \to \Omega$ with $\gamma(0) = \boldsymbol{x}$ and $\gamma(1) = \boldsymbol{y}$. Obviously, ρ_Ω satisfies the triangle inequality, but it need not be a distance function. In particular, $\rho_{\mathbb{R}^n}$ vanishes identically.

There is another natural procedure to obtain the pseudodistance ρ_D in (6-6), which is motivated by Kobayashi's definition of his pseudodistance on complex manifolds [1967]. Fix a pair of points $\boldsymbol{x}, \boldsymbol{y} \in D$. To any finite chain of conformal harmonic discs $f_i : \mathbb{D} \to D$ and points $a_i \in \mathbb{D}$ $(i = 1, \dots, k)$ such that

$$f_1(0) = \boldsymbol{x}, \quad f_{i+1}(0) = f_i(a_i) \text{ for } i = 1, \dots, k-1, \quad f_k(a_k) = \boldsymbol{y}, \tag{6-7}$$

we associate the number

$$\sum_{i=1}^k \frac{1}{2} \log \frac{1 + |a_i|}{1 - |a_i|} \ge 0.$$

The i-th term in the sum is the Poincaré distance from 0 to a_i in $\mathbb{D}$. The pseudodistance $\rho_D(\boldsymbol{x}, \boldsymbol{y})$ is defined to be the infimum of the numbers obtained in this way. The proof that the two definitions yield the same result is similar to the one given for the Kobayashi pseudodistance by Royden [1971, Theorem 1]; see [Drinovec Drnovšek and Forstnerič 2023, Theorem 3.1] for the details.

The following proposition says that the minimal pseudodistance ρ_D gives an upper bound for growth of conformal minimal surfaces in the domain D.

Proposition 6.1. *Every conformal harmonic map $M \to D$ from a hyperbolic conformal surface is distance decreasing in the Poincaré distance on M and the pseudodistance ρ_D, and ρ_D is the biggest pseudodistance on D for which this holds.*

Proof. Let M be a hyperbolic conformal surface and $h : \mathbb{D} \to M$ be a conformal universal covering. Choose a conformal harmonic map $f : M \to D$ and a pair of points $p, q \in M$. Let $a, b \in \mathbb{D}$ be such that $h(a) = p$ and $h(b) = q$. Precomposing h by an automorphism of the disc, we may assume that $a = 0$.

Then, $g := f \circ h : \mathbb{D} \to D$ is a conformal harmonic disc with $g(0) = f(p)$ and $g(b) = f(q)$, and it follows from the definition of ρ_D that

$$\rho_D(f(p), f(q)) = \rho_D(g(0), g(b)) \le \frac{1}{2}\log\frac{1+|b|}{1-|b|}.$$

The infimum of the right-hand side over all points $b \in \mathbb{D}$ with $h(b) = q$ equals the Poincaré distance between p and q in M, so we see that f is distance decreasing.

Suppose now that τ is a pseudodistance on D such that every conformal harmonic map $\mathbb{D} \to D$ is distance decreasing with the Poincaré metric on $\mathbb{D}$. Let $f_i : \mathbb{D} \to D$ and $a_i \in \mathbb{D}$ for $i = 1, \dots, k$ be a chain as in (6-7) connecting the points $\boldsymbol{x}, \boldsymbol{y} \in D$. Then,

$$\tau(\boldsymbol{x}, \boldsymbol{y}) \le \sum_{i=1}^{k} \tau(f_i(0), f_i(a_i)) \le \sum_{i=1}^{k} \frac{1}{2}\log\frac{1+|a_i|}{1-|a_i|}.$$

Taking the infimum over all such chains gives $\tau(\boldsymbol{x}, \boldsymbol{y}) \le \rho_D(\boldsymbol{x}, \boldsymbol{y})$. □

We have already observed that, on the ball $\mathbb{B}^n$ ($n \ge 3$), the Finsler metric $\mathcal{M}_{\mathbb{B}^n}$ is given by (2-8). From (2-9) and (6-4) it follows that $g_{\mathbb{B}^n}$ equals the Cayley–Klein metric $\mathcal{CK}$ (2-6):

Theorem 6.2. *On the ball* $\mathbb{B}^n$, $n \ge 3$, *we have*

$$g_{\mathbb{B}^n} = \mathcal{CK}, \qquad \rho_{\mathbb{B}^n} = \mathrm{dist}_{\mathcal{CK}}.$$

Hilbert [1895] defined a metric on any convex domain in $\mathbb{RP}^n$ that generalizes the Cayley–Klein metric on the ball. Hilbert metrics are examples of projectively invariant metrics which have been studied by many authors; see the surveys by Kobayashi [1977; 1984] and Goldman [2019]. Kobayashi [1977] discussed the analogy between his metric and Hilbert's metric. Lempert [1987] established an explicit connection, and then in [Lempert 1993, Theorem 3.1] proved that the Hilbert metric $\mathcal{H}_D$ on any bounded convex domain $D \subset \mathbb{R}^n$ is the restriction to D of the Kobayashi metric on the elliptic tube $D^* \subset D \times \mathrm{i}\mathbb{R}^n \subset \mathbb{C}^n$ obtained as follows; see [Lempert 1993, p. 441]. Every affine line segment $L \subset D$ with endpoints on bD is the diameter of a unique complex disc in $D \times \mathrm{i}\mathbb{R}^n$, and D^* is the union of all such discs. The elliptic tube over the ball $\mathbb{B}^n$ is the complex ball $\mathbb{B}^n_{\mathbb{C}}$, and the metric $g_{\mathbb{B}^n}$ agrees with the Hilbert metric $\mathcal{H}_{\mathbb{B}^n} = \mathcal{CK}$ according to Theorem 6.2.

While Hilbert's metric is invariant under projective linear transformations, the minimal metric is invariant (at least in an obvious way) only under the conformal group (see Proposition 6.5); hence it is expected that the two metrics differ on most convex domains. We give an explicit example on ellipsoids.

Example 6.3. Let (x, y, z) be coordinates on $\mathbb{R}^3$. Consider the ellipsoid

$$D_a = \left\{(x, y, z) \in \mathbb{R}^3 : x^2 + \frac{1}{a^2}(y^2 + z^2) < 1\right\}, \qquad a > 0.$$

Note that $D_a \subset \mathbb{B}^3$ if and only if $0 < a \le 1$, and $D_1 = \mathbb{B}^3$. We will show that for $0 < a < 1$ the Hilbert metric on D_a does not agree with the minimal metric at the origin $\mathbf{0} \in \mathbb{R}^3$. Since the x-axis intersects D_a in the interval $(-1, 1)$, the Hilbert length of the vector $\boldsymbol{e}_1 = (1, 0, 0)$ equals 1. Pick a two-plane $\Lambda \subset \mathbb{R}^3$ containing the vector $\boldsymbol{e}_1$. Due to rotational symmetry of D_a in the (y, z)-coordinates the value of $\mathcal{M}_{D_a}(\mathbf{0}, \Lambda)$ (6-3) does not depend on the choice of Λ, so we may take $\Lambda = \{z = 0\}$. Let $f = (f_1, f_2, f_3) : \mathbb{D} \to D_a$ be a conformal harmonic disc with $f(0) = \mathbf{0}$ and $df_0(\mathbb{R}^2) = \{z = 0\}$. Replacing f by $f(\mathrm{e}^{\mathrm{i}t} z)$ for a suitable $t \in \mathbb{R}$ gives $f_x(0) = r\boldsymbol{e}_1$ and $f_y(0) = \pm r\boldsymbol{e}_2$ with $r = \|df_0\| > 0$. The projection $h = (f_1, f_2) : \mathbb{D} \to \mathbb{R}^2$ maps $\mathbb{D}$ into the ellipse $E_a = \{x^2 + y^2/a^2 < 1\}$, $h(0) = \mathbf{0}$, and h is conformal at 0. For $0 < a < 1$ we have $E_a \subsetneq \mathbb{D}$. Theorem 1.1 implies that $r = \|dh_0\| < 1$; equality is excluded since in that case we would have $h(\mathbb{D}) = \mathbb{D}$. By a normal families argument we also have that $\sup_f \|df_0\| < 1$. It follows that $\mathcal{M}_{D_a}(\mathbf{0}, \Lambda) > 1$ for every such Λ, and hence

$$g_{D_a}(\mathbf{0}, \boldsymbol{e}_1) = \mathcal{M}_{D_a}(\mathbf{0}, \Lambda) > 1 = \mathcal{H}_{D_a}(\mathbf{0}, \boldsymbol{e}_1) \quad \text{if } 0 < a < 1.$$

Problem 6.4. On which bounded convex domains $D \subset \mathbb{R}^n$, $n \geq 3$ (besides the ball) does the Hilbert metric coincide with the minimal metric g_D? Is the ball the only such domain?

Denote by $\mathcal{R}_n$ the Lie group of transformations $\mathbb{R}^n \to \mathbb{R}^n$ generated by the orthogonal group O_n, translations, and dilations by positive numbers. Elements of $\mathcal{R}_n$ are called *rigid transformations* of $\mathbb{R}^n$. Postcomposition of any conformal harmonic map $f : M \to \mathbb{R}^n$ by a rigid transformation of $\mathbb{R}^n$ is again a conformal harmonic map, and it is well known that $\mathcal{R}_n$ is the largest group of diffeomorphisms of $\mathbb{R}^n$ having this property. This gives the following.

Proposition 6.5. *Given a domain $D \subset \mathbb{R}^n$, $n \geq 3$, and a map $R \in \mathcal{R}_n$, the restriction $R|_D : D \to D' = R(D)$ is an isometry of pseudometric spaces $(D, \rho_D) \to (D', \rho_{D'})$.*

Remark 6.6. The intrinsic pseudometric g_D and the associated pseudodistance ρ_D can be defined in the very same way on an arbitrary Riemannian manifold $(D, \tilde{g})$ of dimension at least three. The Riemannian metric $\tilde{g}$ determines the class of conformal harmonic maps $\mathbb{D} \to D$, which coincide with conformal minimal discs in D.

***Hyperbolic domains in* $\mathbb{R}^n$.** We now introduce the notion of (complete) hyperbolic domains in $\mathbb{R}^n$, in analogy with Kobayashi hyperbolic complex manifolds.

Definition 6.7. A domain $D \subset \mathbb{R}^n$ $(n \geq 3)$ is *hyperbolic* if the pseudodistance ρ_D is a distance function on D, and is *complete hyperbolic* if (D, ρ_D) is a complete metric space.

Example 6.8. (a) The ball $\mathbb{B}^n \subset \mathbb{R}^n$ $(n \geq 3)$ is complete hyperbolic. Indeed, the Cayley–Klein metric (2-6) is complete, so the conclusion follows from Theorem 6.2.

(b) Every bounded domain $D \subset \mathbb{R}^n$ is hyperbolic. Indeed, if B is a ball containing D then $\rho_D(\boldsymbol{x}, \boldsymbol{y}) \geq \rho_B(\boldsymbol{x}, \boldsymbol{y})$ for any pair $\boldsymbol{x}, \boldsymbol{y} \in D$, and B is complete hyperbolic by (a). However, a bounded domain need not be complete hyperbolic. For example, if bD is strongly concave at $\boldsymbol{p} \in bD$, there is a conformal linear disc $\Sigma \subset D \cup \{\boldsymbol{p}\}$ containing $\boldsymbol{p}$, and it is easily seen that $\boldsymbol{p}$ is at finite ρ-distance from D.

(c) The half-space $\mathbb{H}^n = \{\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n : x_n > 0\}$ is not hyperbolic, and the pseudodistance $\rho_{\mathbb{H}^n}$ vanishes on all planes $x_n = \text{const}$. However, every point on $b\mathbb{H}^n = \{x_n = 0\}$ is at infinite minimal distance from points in $\mathbb{H}^n$ [Drinovec Drnovšek and Forstnerič 2023, Lemma 5.2].

By using the expression for the metric (2-8) on the ball we can determine the asymptotic rate of growth of the Finsler metric $\mathcal{M}_D$, and hence of the distance function ρ_D, on any bounded strongly convex domain $D \subset \mathbb{R}^n$ with $\mathscr{C}^2$ boundary. Let $\delta = \delta(\boldsymbol{x}) = 1 - |\boldsymbol{x}|$ denote the distance from a point $\boldsymbol{x} \in \mathbb{B}^n \setminus \{0\}$ to the sphere $b\mathbb{B}^n$, and let $\Lambda \subset \mathbb{R}^n$ be a 2-plane forming an angle θ with $\boldsymbol{x}$. As $\boldsymbol{x}$ approaches the sphere radially, we have

$$\mathcal{M}_{\mathbb{B}^n}(\boldsymbol{x}, \theta) := \mathcal{M}_{\mathbb{B}^n}(\boldsymbol{x}, \Lambda) \approx \frac{\sqrt{\cos^2\theta + 2\delta \sin^2\theta}}{2\delta},$$

in the sense that the quotient of the two sides converges to 1 as $\delta \to 0$. In particular,

$$\mathcal{M}_{\mathbb{B}^n}\Big(\boldsymbol{x}, \frac{\pi}{2}\Big) \approx \frac{1}{\sqrt{2\delta}} \qquad \text{and} \qquad \mathcal{M}_{\mathbb{B}^n}(\boldsymbol{x}, \theta) \approx \frac{\cos\theta}{2\delta} \quad \text{for } \theta \in \Big[0, \frac{\pi}{2}\Big).$$

Assume now that $D \subset \mathbb{R}^n$ is a bounded strongly convex domain with $\mathscr{C}^2$ boundary. There is a collar $U \subset \mathbb{R}^n$ around bD such that every point $\boldsymbol{x} \in U \cap D$ has a unique closest point $\pi(\boldsymbol{x}) \in bD$. Comparison with inscribed and circumscribed balls to D passing through the point $\pi(\boldsymbol{x})$ shows that there are constants $0 < c < C$ such that

$$c\,\frac{\sqrt{\cos^2\theta + 2\delta \sin^2\theta}}{2\delta} \leq \mathcal{M}_D(\boldsymbol{x}, \Lambda) \leq C\,\frac{\sqrt{\cos^2\theta + 2\delta \sin^2\theta}}{2\delta} \tag{6-8}$$

for $\boldsymbol{x} \in U \cap D$, where $\delta = |\boldsymbol{x} - \pi(\boldsymbol{x})| = \operatorname{dist}(\boldsymbol{x}, bD)$ and θ is the angle between the 2-plane Λ and the normal vector $N_{\boldsymbol{x}} = \delta^{-1}(\pi(\boldsymbol{x}) - \boldsymbol{x})$ to bD at $\pi(\boldsymbol{x}) \in bD$. The upper bound uses comparison with inscribed balls, so it holds on any domain with $\mathscr{C}^2$ boundary, while the lower bound uses comparison with circumscribed ball, and hence it depends on strong convexity of D. These estimates are analogous to the asymptotic boundary estimates of the Kobayashi metric in bounded strongly pseudoconvex domains in $\mathbb{C}^n$ and are due to Graham [1975]. (There is a large subsequent literature on this subject.) These estimates show in particular that the distance function ρ_D induced by $\mathcal{M}_D$ is complete, thereby giving the following.

Theorem 6.9. *Every bounded strongly convex domain in $\mathbb{R}^n$, $n \geq 3$, with $\mathscr{C}^2$ boundary is complete hyperbolic in the minimal metric.*

Remark 6.10. Since the first version of this paper was posted on arXiv in February 2021, progress on the subject of minimal hyperbolicity was made by Drinovec Drnovšek and Forstnerič [2023], whose paper we will henceforth abbreviate as [DDF 2023]. Besides establishing basic characterizations of (complete) hyperbolicity, they proved that a convex domain in $\mathbb{R}^n$ is hyperbolic if and only if it is complete hyperbolic if and only if it does not contain any affine 2-plane [DDF 2023, Theorem 5.1]. They also showed that every bounded strongly minimally convex domain in $\mathbb{R}^n$, $n \geq 3$, is complete hyperbolic [DDF 2023, Theorem 9.2]. This is a considerable generalization of Theorem 6.9, whose proof relies on the lower bound for $\mathcal{M}_\Omega$ (and hence g_Ω) given by another Finsler pseudometric $\mathcal{F}_\Omega : \Omega \times G_2(\mathbb{R}^n) \to \mathbb{R}_+$ defined in terms of minimal plurisubharmonic functions; see [DDF 2023, Section 7]. A discussion of this class of domains and functions can be found in [Alarcón et al. 2019; 2021, Chapter 8]. Finally, they established a localization theorem for the minimal pseudometric analogous to the results for the Kobayashi pseudometric; see [DDF 2023, Section 8].

The following problem remains open; an affirmative answer is known for the case when M is a plane; see [DDF 2023, Lemma 5.2].

Problem 6.11. Let M be an embedded minimal surface in $\mathbb{R}^3$. Is the minimal distance from $\mathbb{R}^3 \setminus M$ to M infinite? Is the complement of a catenoid in $\mathbb{R}^3$ complete hyperbolic?

Extremal minimal discs. Another important and natural question is the following.

Problem 6.12. Let $D \subset \mathbb{R}^n$ be a bounded strongly convex domain with smooth boundary. Is there a unique (up to a conformal reparametrization) extremal conformal harmonic disc through any given point $\boldsymbol{x} \in D$ tangent to a given two-plane $\Lambda \in G_2(\mathbb{R}^n)$ at $\boldsymbol{x}$?

Theorem 2.1 gives an affirmative answer on the ball, and this is the only domain for which the answer seems to be known. By the seminal result of Lempert [1981; 1987], the analogous result holds for the extremal holomorphic discs for the Kobayashi metric in any smoothly bounded strongly convex domain $D \subset \mathbb{C}^n$.

We now describe a condition which implies an affirmative answer to this problem. It explores a comparison between the Finsler pseudometric $\mathcal{M}_D$ (6-1) on a domain $D \subset \mathbb{R}^n$ and a Kobayashi-type pseudometric on the tube $\mathcal{T}_D = \mathbb{D} \times \mathrm{i}\mathbb{R}^n \subset \mathbb{C}^n$. To this end, we recall a few basic facts from the theory of minimal surfaces; see [Alarcón et al. 2021, Chapter 2] or [Osserman 1969].

A holomorphic map $F = (F_1, \dots, F_n) : \mathbb{D} \to \mathbb{C}^n$ satisfying

$$\sum_{i=1} F_i'(z)^2 = 0 \quad \text{for all } z \in \mathbb{D}$$

is called a *holomorphic null map*. The analogous definition applies with the disc replaced by any open Riemann surface, considering the above equation in local holomorphic coordinates. (The map F need not be an immersion.) The complex cone

$$\boldsymbol{A}^{n-1} = \left\{ \boldsymbol{z} = (z_1, \dots, z_n) \in \mathbb{C}^n : \sum_{i=1}^n z_i^2 = 0 \right\} \tag{6-9}$$

is called the *null cone*, and its elements are *null vectors*. Hence, a holomorphic map F is null if and only if the complex derivative $F'(z)$ lies in $\boldsymbol{A}^{n-1}$ for every z. It is a basic fact that the real and imaginary parts of a holomorphic null map $M \to \mathbb{C}^n$ are conformal harmonic maps $M \to \mathbb{R}^n$; conversely, every conformal harmonic map $\mathbb{D} \to \mathbb{R}^n$ from the disc is the real part of a holomorphic null map $\mathbb{D} \to \mathbb{C}^n$; see [Alarcón et al. 2021, Theorem 2.3.4]. Given a domain $D \subset \mathbb{R}^n$, we denote by $\mathrm{HN}(\mathbb{D}, \mathcal{T}_D)$ the space of all holomorphic null maps $F = (F_1, \dots, F_n) : \mathbb{D} \to \mathcal{T}_D$. Define a pseudometric on $(\boldsymbol{z}, \boldsymbol{w}) \in \mathcal{T}_D \times \boldsymbol{A}^{n-1}$ by

$$\mathcal{N}_D(\boldsymbol{z}, \boldsymbol{w}) = \inf\left\{ \frac{1}{|a|} : \exists F \in \mathrm{HN}(\mathbb{D}, \mathcal{T}_D),\ F(0) = \boldsymbol{z},\ F'(0) = a\boldsymbol{w} \right\}. \tag{6-10}$$

Here, a may be a complex number. Clearly, $\mathcal{N}_D(\boldsymbol{z}, \boldsymbol{w})$ is bigger than or equal to the Kobayashi pseudonorm of the vector $\boldsymbol{w} \in T_{\boldsymbol{z}}(\mathcal{T}_D)$, since in the definition of the latter one uses all holomorphic discs as opposed to just null discs. Note that for each conformal frame $(\boldsymbol{u}, \boldsymbol{v}) \in \mathrm{CF}_n$ the vectors $\boldsymbol{u} \pm \mathrm{i}\boldsymbol{v} \in \mathbb{C}^n$ are null vectors; conversely, the real and imaginary component of a null vector $\boldsymbol{w} \in \boldsymbol{A}^{n-1}$ form a conformal frame. The aforementioned correspondence between conformal harmonic discs in D and holomorphic null discs in $\mathcal{T}_D$ shows that for all $\boldsymbol{x} \in D$, $\boldsymbol{y} \in \mathbb{R}^n$, and $(\boldsymbol{u}, \boldsymbol{v}) \in \mathrm{CF}_n$ we have

$$\mathcal{N}_D(\boldsymbol{x} + \mathrm{i}\boldsymbol{y}, \boldsymbol{u} \pm \mathrm{i}\boldsymbol{v}) = \mathcal{M}_D(\boldsymbol{x}, (\boldsymbol{u}, \boldsymbol{v})). \tag{6-11}$$

This shows in particular that every extremal conformal harmonic disc in D is the real part of an extremal holomorphic null disc in the tube $\mathcal{T}_D$. Therefore, the correspondence between the extremal conformal minimal discs in the ball $\mathbb{B}^n \subset \mathbb{R}^n$ and the Kobayashi geodesics in the tube $\mathcal{T}_{\mathbb{B}^n}$, used in the proof of Lemma 3.1, extends to any bounded strongly convex domain $D \subset \mathbb{R}^n$ with $\mathscr{C}^2$ boundary satisfying the following condition. The notion of a stationary holomorphic disc was explained in Remark 3.2.

Definition 6.13. A domain $D \subset \mathbb{R}^n$ satisfies *Condition N* if for every point $\boldsymbol{x} \in D$ and null vector $0 \neq \boldsymbol{w} \in \boldsymbol{A}^{n-1}$ there is a stationary holomorphic null disc in the tube $\mathcal{T}_D$ through the point $\boldsymbol{x} + \mathrm{i}0$ in the direction $\boldsymbol{w}$.

Our proof of Theorem 2.1 implies the following.

Theorem 6.14. *If D is a bounded strongly convex domain in $\mathbb{R}^n$, $n \geq 3$, with smooth boundary and satisfying Condition N, then for every point $\boldsymbol{x} \in D$ and two-plane $\Lambda \in G_2(\mathbb{R}^n)$ there exists an extremal conformal harmonic disc $f : \mathbb{D} \to D$ with $f(0) = \boldsymbol{x}$ and $df_0(\mathbb{R}^2) = \Lambda$. Such an f is unique up to a rotation of $\mathbb{D}$.*

Proof. Let $0 \neq \boldsymbol{w} = \boldsymbol{u} - \mathrm{i}\boldsymbol{v} \in \boldsymbol{A}^{n-1}$ be such that $\Lambda = \mathrm{span}\{\boldsymbol{u}, \boldsymbol{v}\}$. By Condition N there is a stationary holomorphic null disc $F : \mathbb{D} \to \mathcal{T}_D$ with $F(0) = \boldsymbol{x} + \mathrm{i}0$ and $F'(0) = \alpha\boldsymbol{w}$ for $\alpha \in \mathbb{C}$, and F is unique up to rotations of $\mathbb{D}$ by Lempert's theorem [1981, Theorem 2]. The real part $f = \Re F : \mathbb{D} \to D$ is then a conformal harmonic disc as in the theorem. □

Problem 6.15. Which bounded strongly convex domains in $\mathbb{R}^n$, besides the ball, satisfy Condition N?

Complex geodesics of the Kobayashi metric in tubes over convex domains $D \subset \mathbb{R}^n$ were studied by Zając [2015; 2016], Pflug and Zwonek [2018], and Zwonek [2022]. It would be of interest to see whether these works can be used to give information on the validity of Condition N. The fact that Condition N holds on the ball $\mathbb{B}^n$ may simply be a lucky coincidence which makes our analysis work on this most symmetric domain.

Acknowledgements

Forstnerič was partially supported by the research program P1-0291 and grants J1-9104, J1-3005, and N1-0237 from ARRS, Republic of Slovenia. Kalaj was partially supported by a research fund of the University of Montenegro. We wish to thank Antonio Alarcón for remarks which improved the presentation, Dmitry Khavinson for information regarding harmonic self-maps of the disc, László Lempert and Stefan Nemirovski for information regarding projectively invariant metrics, Joaquín Pérez for consultation on the state of the art concerning growth of conformal minimal surfaces, and Sylwester Zając and Włodzimierz Zwonek for communications concerning geodesics of the Kobayashi metric in tube domains. Forstnerič also thanks Miodrag Mateljević with whom he discussed this problem in 2016. Finally, we thank the referee for a careful reading of the paper and suggestions which improved the presentation.

References

[Ahlfors 1938] L. V. Ahlfors, "An extension of Schwarz's lemma", *Trans. Amer. Math. Soc.* **43**:3 (1938), 359–364. MR Zbl

[Ahlfors 1966] L. V. Ahlfors, *Lectures on quasiconformal mappings*, Van Nostrand Math. Stud. **10**, Van Nostrand, Toronto, 1966. MR Zbl

[Alarcón et al. 2019] A. Alarcón, B. Drinovec Drnovšek, F. Forstnerič, and F. J. López, "Minimal surfaces in minimally convex domains", *Trans. Amer. Math. Soc.* **371**:3 (2019), 1735–1770. MR Zbl

[Alarcón et al. 2021] A. Alarcón, F. Forstnerič, and F. J. López, *Minimal surfaces from a complex analytic viewpoint*, Springer, 2021. MR Zbl

[Axler et al. 2001] S. Axler, P. Bourdon, and W. Ramey, *Harmonic function theory*, 2nd ed., Grad. Texts in Math. **137**, Springer, 2001. MR Zbl

[Beltrami 1868] E. Beltrami, "Teoria fondamentale degli spazii di curvatura costante", *Ann. Mat. Pura Appl.* **2** (1868), 232–255. Zbl

[Brevig et al. 2021] O. F. Brevig, J. Ortega-Cerdà, and K. Seip, "A converse to the Schwarz lemma for planar harmonic maps", *J. Math. Anal. Appl.* **497**:2 (2021), art. id. 124908. MR Zbl

[Carathéodory 1912] C. Carathéodory, "Untersuchungen über die konformen Abbildungen von festen und veränderlichen Gebieten", *Math. Ann.* **72**:1 (1912), 107–144. MR Zbl

[Cayley 1859] A. Cayley, "A sixth memoir upon quantics", *Philos. Trans. Roy. Soc.* **149** (1859), 61–90. Zbl

[Dineen 1989] S. Dineen, *The Schwarz lemma*, Oxford Univ. Press, 1989. MR Zbl

[Drinovec Drnovšek and Forstnerič 2023] B. Drinovec Drnovšek and F. Forstnerič, "Hyperbolic domains in real Euclidean spaces", *Pure Appl. Math. Q.* **19**:6 (2023), 2689–2735. MR Zbl

[Duren 2004] P. Duren, *Harmonic mappings in the plane*, Cambridge Tracts in Math. **156**, Cambridge Univ. Press, 2004. MR Zbl

[Garnett 1981] J. B. Garnett, *Bounded analytic functions*, Pure Appl. Math. **96**, Academic Press, New York, 1981. MR Zbl

[Goldman 2019] W. M. Goldman, "Flat affine, projective and conformal structures on manifolds: a historical perspective", pp. 515–552 in *Geometry in history*, edited by S. G. Dani and A. Papadopoulos, Springer, 2019. MR Zbl

[Graham 1975] I. Graham, "Boundary behavior of the Carathéodory and Kobayashi metrics on strongly pseudoconvex domains in C^n with smooth boundary", *Trans. Amer. Math. Soc.* **207** (1975), 219–240. MR Zbl

[Hengartner and Schober 1986] W. Hengartner and G. Schober, "Harmonic mappings with given dilatation", *J. Lond. Math. Soc.* (2) **33**:3 (1986), 473–483. MR Zbl

[Hilbert 1895] D. Hilbert, "Ueber die gerade Linie als kürzeste Verbindung zweier Punkte", *Math. Ann.* **46** (1895), 91–96. Zbl

[Jarnicki and Pflug 2013] M. Jarnicki and P. Pflug, *Invariant distances and metrics in complex analysis*, 2nd ed., de Gruyter Expo. Math. **9**, de Gruyter, Berlin, 2013. MR Zbl

[Kalaj and Vuorinen 2012] D. Kalaj and M. Vuorinen, "On harmonic functions and the Schwarz lemma", *Proc. Amer. Math. Soc.* **140**:1 (2012), 161–165. MR Zbl

[Klein 1871] F. Klein, "Ueber die sogenannte Nicht-Euklidische Geometrie", *Math. Ann.* **4**:4 (1871), 573–625. Zbl

[Klein 1873] F. Klein, "Ueber die sogenannte Nicht-Euklidische Geometrie: Zweiter Aufsatz", *Math. Ann.* **6**:2 (1873), 112–145. MR Zbl

[Kobayashi 1967] S. Kobayashi, "Invariant distances on complex manifolds and holomorphic mappings", *J. Math. Soc. Japan* **19** (1967), 460–480. MR Zbl

[Kobayashi 1976] S. Kobayashi, "Intrinsic distances, measures and geometric function theory", *Bull. Amer. Math. Soc.* **82**:3 (1976), 357–416. MR Zbl

[Kobayashi 1977] S. Kobayashi, "Intrinsic distances associated with flat affine or projective structures", *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **24**:1 (1977), 129–135. MR Zbl

[Kobayashi 1984] S. Kobayashi, "Projectively invariant distances for affine and projective structures", pp. 127–152 in *Differential geometry* (Warsaw, 1979), edited by W. Waliszewski et al., Banach Center Publ. **12**, PWN, Warsaw, 1984. MR Zbl

[Kobayashi 2005] S. Kobayashi, *Hyperbolic manifolds and holomorphic mappings*: *an introduction*, 2nd ed., World Sci., Hackensack, NJ, 2005. MR Zbl

[Kovalev and Yang 2020] L. V. Kovalev and X. Yang, "Near-isometric duality of Hardy norms with applications to harmonic mappings", *J. Math. Anal. Appl.* **487**:2 (2020), art. id. 124040. MR Zbl

[Krantz 1992] S. G. Krantz, *Function theory of several complex variables*, 2nd ed., Wadsworth & Brooks/Cole, Pacific Grove, CA, 1992. MR Zbl

[Lehto and Virtanen 1973] O. Lehto and K. I. Virtanen, *Quasiconformal mappings in the plane*, 2nd ed., Grundl. Math. Wissen. **126**, Springer, 1973. MR Zbl

[Lempert 1981] L. Lempert, "La métrique de Kobayashi et la représentation des domaines sur la boule", *Bull. Soc. Math. France* **109**:4 (1981), 427–474. MR Zbl

[Lempert 1987] L. Lempert, "Complex geometry in convex domains", pp. 759–765 in *Proceedings of the International Congress of Mathematicians*, *I* (Berkeley, CA, 1986), edited by A. M. Gleason, Amer. Math. Soc., Providence, RI, 1987. MR Zbl

[Lempert 1993] L. Lempert, "Elliptic and hyperbolic tubes", pp. 440–456 in *Several complex variables* (Stockholm, 1987/1988), edited by J. E. Fornæss, Math. Notes **38**, Princeton Univ. Press, 1993. MR Zbl

[Osserman 1969] R. Osserman, *A survey of minimal surfaces*, Van Nostrand, New York, 1969. MR Zbl

[Pflug and Zwonek 2018] P. Pflug and W. Zwonek, "Regularity of complex geodesics and (non-)Gromov hyperbolicity of convex tube domains", *Forum Math.* **30**:1 (2018), 159–170. MR Zbl

[Pick 1915] G. Pick, "Über die Beschränkungen analytischer Funktionen, welche durch vorgegebene Funktionswerte bewirkt werden", *Math. Ann.* **77**:1 (1915), 7–23. MR Zbl

[Poincaré 1884] H. Poincaré, "Sur les groupes des équations linéaires", *Acta Math.* **4**:1 (1884), 201–312. MR Zbl

[Ratcliffe 1994] J. G. Ratcliffe, *Foundations of hyperbolic manifolds*, Grad. Texts in Math. **149**, Springer, 1994. MR Zbl

[Royden 1971] H. L. Royden, "Remarks on the Kobayashi metric", pp. 125–137 in *Several complex variables*, *II* (College Park, MD, 1970), edited by J. Horváth, Lecture Notes in Math. **185**, Springer, 1971. MR Zbl

[Royden 1988] H. L. Royden, "Hyperbolicity in complex analysis", *Ann. Acad. Sci. Fenn. Ser. A I Math.* **13**:3 (1988), 387–400. MR Zbl

[Rudin 1980] W. Rudin, *Function theory in the unit ball of* $\mathbb{C}^n$, Grundl. Math. Wissen. **241**, Springer, 1980. MR Zbl

[Schwarz 1890] H. A. Schwarz, "Zur Theorie der Abbildung", pp. 108–132 in *Gesammelte mathematische Abhandlungen*, Band 2, Springer, Berlin, 1890. Zbl

[Yau 1978] S. T. Yau, "A general Schwarz lemma for Kähler manifolds", *Amer. J. Math.* **100**:1 (1978), 197–203. MR Zbl

[Zając 2015] S. Zając, "Complex geodesics in convex tube domains", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **14**:4 (2015), 1337–1361. MR Zbl

[Zając 2016] S. Zając, "Complex geodesics in convex tube domains, II", *Ann. Mat. Pura Appl.* (4) **195**:6 (2016), 1865–1887. MR Zbl

[Zwonek 2022] W. Zwonek, "Complex geodesics in tube domains and their role in the study of harmonic mappings in the disc", *Anal. Math.* **48**:2 (2022), 601–617. MR Zbl

Received 1 Dec 2021. Revised 5 Jul 2022. Accepted 11 Aug 2022.

FRANC FORSTNERIČ: franc.forstneric@fmf.uni-lj.si
Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia

and

Institute of Mathematics, Physics, and Mechanics, Ljubljana, Slovenia

DAVID KALAJ: davidkalaj@gmail.com
Faculty of Natural Sciences and Mathematics, University of Montenegro, Podgorica, Montenegro

ANALYSIS AND PDE
Vol. 17 (2024), No. 3, pp. 1005–1018
DOI: 10.2140/apde.2024.17.1005

AN IMPROVED REGULARITY CRITERION AND ABSENCE OF SPLASH-LIKE SINGULARITIES FOR G-SQG PATCHES

JUNEKEY JEON AND ANDREJ ZLATOŠ

We prove that splash-like singularities cannot occur for sufficiently regular patch solutions to the generalized surface quasi-geostrophic equation on the plane or half-plane with parameter $\alpha \le \frac{1}{4}$. This includes potential touches of more than two patch boundary segments in the same location, an eventuality that has not been excluded previously and presents nontrivial complications (in fact, if we do a priori exclude it, then our results extend to all $\alpha \in (0, 1)$). As a corollary, we obtain an improved global regularity criterion for H^3 patch solutions when $\alpha \le \frac{1}{4}$, namely that finite time singularities cannot occur while the H^3 norms of patch boundaries remain bounded.

1. Introduction

The *g-SQG* (*generalized surface quasi-geostrophic*) *equation* is the active scalar PDE

$$\partial_t \omega + u \cdot \nabla \omega = 0, \tag{1-1}$$

where the scalar $\omega\colon \mathbb{R}^2 \times (0, \infty) \to \mathbb{R}$ is advected by the velocity field

$$u := \nabla^{\perp} (-\Delta)^{-1+\alpha} \omega. \tag{1-2}$$

Here $\nabla^{\perp} := (-\partial_{x_2}, \partial_{x_1})$ and $\alpha \in (0, 1)$ is a given parameter. Note that (1-1) is the vorticity form of the (incompressible) two-dimensional Euler equation when $\alpha = 0$, which models the motion of ideal fluids, with u the fluid velocity and $\omega := \nabla^{\perp} \cdot u$ its vorticity. When $\alpha = \frac{1}{2}$, it is the SQG equation, which is used in atmospheric science models [Pedlosky 1979] and was first analyzed rigorously by Constantin, Majda, and Tabak [Constantin et al. 1994]. The g-SQG equation with $\alpha \in (0, 1)$ is its generalization and has also been studied in both geophysical and mathematical literature, including in [Chae et al. 2012; Constantin et al. 2008; Córdoba et al. 2005; Gancedo 2008; Kiselev and Luo 2023; Kiselev et al. 2016; 2017; Pierrehumbert et al. 1994; Smith et al. 2002].

Global regularity for (smooth or bounded) solutions has been known in the Euler case $\alpha = 0$ since the works of Hölder [1933], Wolibner [1933], and Yudovich [1963], but it is still an open problem in the g-SQG case with any $\alpha \in (0, 1)$. In this work we consider so-called *patch solutions* to (1-1), that is, weak solutions that are linear combinations of characteristic functions of some time-dependent sets $\Omega_n(t) \subseteq \mathbb{R}^2$ (often only a single such set/patch is considered but the extension to multiple sets is typically

Jeon acknowledges partial support by NSF grant DMS-1900943. Zlatoš acknowledges partial support by NSF grant DMS-1900943 and by a Simons Fellowship.
MSC2020: 35Q35.
Keywords: generalized SQG equation, patch solutions, splash singularity.

straightforward). The main question now is that of global well-posedness for these solutions: if the boundary of each initial patch $\Omega_n(0)$ is a simple closed curve of some prescribed regularity (H^k or $C^{k,\gamma}$) and these curves are pairwise disjoint, does this setup persist forever or may it cease existing in finite time? This of course involves not only the required regularity of each $\partial\Omega_n(t)$, but also that they all remain pairwise disjoint simple closed curves.

Chemin [1993] showed the answer to be in the affirmative when $\alpha = 0$, but the question remains open for any $\alpha \in (0, 1)$. Local existence for these models was proved for $\alpha \in \left(0, \frac{1}{2}\right]$ and H^3 patches by Gancedo [2008], who also obtained uniqueness for $\alpha \in \left(0, \frac{1}{2}\right)$ and those solutions that satisfy a related *contour equation* (well-posedness for $\alpha = \frac{1}{2}$ in a special class of patches was earlier proved by Rodrigo [2005]). Local existence was also proved for $\alpha \in \left[\frac{1}{2}, 1\right)$ and H^4 patches by Chae, Constantin, Córdoba, Gancedo, and Wu [Chae et al. 2012]. Kiselev, Yao, and Zlatoš [Kiselev et al. 2017] later proved full local well-posedness for $\alpha \in \left(0, \frac{1}{2}\right)$ and H^3 patches (they also considered the related half-plane case, in which global well-posedness was proved to fail by Kiselev, Yao, Ryzhik, and Zlatoš [Kiselev et al. 2016]). In addition, Córdoba, Córdoba, and Gancedo achieved this for $\alpha = \frac{1}{2}$ and H^3 patches [Córdoba et al. 2018], Gancedo and Patel for $\alpha \in \left(0, \frac{1}{2}\right)$ and H^2 patches as well as for $\alpha \in \left(\frac{1}{2}, 1\right)$ and H^3 patches [Gancedo and Patel 2021], and Gancedo, Nguyen, and Patel for $\alpha = \frac{1}{2}$ and $H^{2+\gamma}$ patches [Gancedo et al. 2022].

The singularity-formation mechanism on the half-plane from [Kiselev et al. 2016], which was motivated by numerical simulations for the three-dimensional Euler equation due to Luo and Hou [2014a; 2014b] and by the related proof of double exponential growth of gradients for smooth solutions to the two-dimensional Euler equation on a bounded domain by Kiselev and Šverák [2014], does not seem to extend to the whole plane case. It is therefore still unknown whether global well-posedness holds on $\mathbb{R}^2$ for any $\alpha \in (0, 1)$. Nevertheless, the local well-posedness result in [Kiselev et al. 2017] does show that, at least for $\alpha \in \left(0, \frac{1}{2}\right)$ and H^3 patches, finite time singularity can only occur if either a patch boundary loses H^3 regularity or a touch happens. The latter might involve two or more patch boundary segments, which might belong to different patches or to a single patch.

The main result of this paper is that, for $\alpha \in \left(0, \frac{1}{4}\right]$, a touch cannot occur without the loss of $C^{1,2\alpha/(1-2\alpha)}$ (and hence also H^3) regularity of a patch boundary at the same time (this is also suggested by numerical simulations of Córdoba, Fontelos, Mancho, and Rodrigo [Córdoba et al. 2005]). If it did occur and the $C^{1,\gamma}$ norm of the patch boundary would stay uniformly bounded for some $\gamma > 0$, the resulting singularity would be called a *splash*. One might think that its existence for the free boundary Euler equation, demonstrated by Castro, Córdoba, Fefferman, Gancedo, and Gómez-Serrano [Castro et al. 2013] and Coutand and Shkoller [2014], would suggest its possibility for g-SQG patches as well. But these two cases are very different: the converging boundary segments are separated by vacuum in the free boundary case, while for (1-1) they are separated by the (incompressible) fluid medium, which must be "squeezed out" of the region between them before a touch can occur.

One might also think that impossibility of general splash singularities was already proved by Gancedo and Strain [2014] for the SQG case $\alpha = \frac{1}{2}$ and smooth patches, who showed that a touch of two patch boundary segments (which we call a *simple splash*) is indeed impossible at any specific location without a loss of boundary smoothness; their argument extends to all $\alpha \in \left(0, \frac{1}{2}\right)$. However, they proved this assuming that no singularity occurs elsewhere, and the result also does not exclude simultaneous touches of three or more boundary segments. Crucially, their proof does not extend to this case either. In it, they place the two segments in a coordinate system in which both are close to horizontal, and use the fact that normal vectors at two points that minimize the vertical distance of the two boundary segments (at any given time) are automatically parallel. This causes important cancellations in the integral evaluating the approach velocity of the two points, which bound this velocity by a multiple of the product of the distance of the two points and the log of this distance. Grönwall's inequality then yields at most double exponential in time approach rate of the two segments.

We can even obtain a simple exponential bound for $C^{2,\gamma}$ patches with $\gamma > 0$ by instead minimizing the distance (rather than vertical distance) of the two boundary segments, in which case the normals at the closest points both lie on the line connecting these points. The resulting computation then bounds the approach velocity by only a multiple of the distance, and it even extends to all $\alpha \in (0, 1)$ with appropriate γ (see Section 2E).

However, when a third boundary segment is present nearby, its normal vector at the point where it intersects the above line need not lie on that line, which significantly compromises the cancellations involved. One then needs to obtain very precise bounds on the resulting errors in this case, which we will achieve by using the uniform $C^{1,2\alpha/(1-2\alpha)}$ bound on the patch boundary to estimate the angle between this normal vector and the line, in terms of the distance of the third segment from the two closest points on the first two segments. When this distance is small, the error will be controlled because the angle must be small; this control worsens when the distance is larger, but then the effect of the third segment on the two points decreases as well. This will yield the needed bound on the approach velocity of the closest points, and this estimate will even extend to the case of arbitrarily many boundary segments folded on top of each other and attempting to create a *complex splash* singularity.

As a result, we will obtain an improved regularity criterion for H^3 patch solutions to (1-1), requiring only a uniform bound on the $C^{1,2\alpha/(1-2\alpha)}$ norm of the patch boundaries. Nevertheless, this approach only works when $\alpha \in \left(0, \frac{1}{4}\right]$, and the obtained estimates are insufficient for larger α (specifically, Lemma 2.5). The reason for this is not just technical, and simply assuming higher boundary regularity will not suffice to overcome the new complications involved. We believe that a different (dynamical) approach will be needed for $\alpha > \frac{1}{4}$ (if the result extends to this range at all), which likely makes it a very difficult problem.

Let us now state rigorously the definition of patch solutions to (1-1) from [Kiselev et al. 2017] (which even allows patches to be nested) and our main result. Below we let $\mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$.

Definition 1.1. Let $\Omega \subseteq \mathbb{R}^2$ be a bounded open set whose boundary $\partial\Omega$ is a simple closed C^1 curve with arc-length $|\partial\Omega|$. We call a *constant-speed parametrization* of $\partial\Omega$ any counterclockwise parametrization $z\colon \mathbb{T} \to \mathbb{R}^2$ of $\partial\Omega$ with $|z'| \equiv |\partial\Omega|/(2\pi)$ on $\mathbb{T}$ (these are all translations of each other), and we define $\|\Omega\|_{C^{k,\gamma}} := \|z\|_{C^{k,\gamma}(\mathbb{T})}$ and $\|\Omega\|_{H^k} := \|z\|_{H^k}$ for $(k, \gamma) \in \mathbb{N}_0 \times [0, 1]$.

Next we note that, when $\alpha \in \left(0, \frac{1}{2}\right)$, the velocity u from (1-2) satisfies the explicit formula

$$u(x, t) := c_\alpha \int_{\mathbb{R}^2} \frac{(x - y)^\perp}{|x - y|^{2+2\alpha}} \omega(y, t)\, dy \tag{1-3}$$

for bounded ω, with $v^{\perp} := (-v_2, v_1)$ and $c_\alpha > 0$ an appropriate constant; see Section 2E for the necessary adjustments when $\alpha \in \left[\frac{1}{2}, 1\right)$. For any $\Gamma \subseteq \mathbb{R}^2$, vector field $v\colon \Gamma \to \mathbb{R}^2$, and $h \in \mathbb{R}$, we let the set to which Γ is advected by v in time h be

$$X_v^h[\Gamma] := \{x + hv(x) \mid x \in \Gamma\}.$$

Definition 1.2. Let $\theta_1, \dots, \theta_N \in \mathbb{R} \setminus \{0\}$, and for each $t \in [0, T)$, let $\Omega_1(t), \dots, \Omega_N(t) \subseteq \mathbb{R}^2$ be bounded open sets whose boundaries are pairwise disjoint simple closed curves such that each $\partial\Omega_n(t)$ is also continuous in $t \in [0, T)$ with respect to Hausdorff distance d_H of sets. Define $\partial\Omega(t) := \bigcup_{n=1}^N \partial\Omega_n(t)$ and $\|\Omega(t)\|_Y := \sum_{n=1}^N \|\Omega_n(t)\|_Y$ for $Y \in \{C^{k,\gamma}, H^k\}$, and let

$$\omega(\,\cdot\,, t) := \sum_{n=1}^N \theta_n \chi_{\Omega_n(t)}. \tag{1-4}$$

If for each $t \in (0, T)$ we have

$$\lim_{h\to 0} \frac{d_H(\partial\Omega(t+h), X^h_{u(\,\cdot\,,t)}[\partial\Omega(t)])}{h} = 0, \tag{1-5}$$

with u from (1-3), then ω is a *a patch solution* to (1-1)–(1-2) on the time interval $[0, T)$. If we also have $\sup_{t\in[0,T']} \|\Omega(t)\|_Y < \infty$ for some $Y \in \{C^{k,\gamma}, H^k\}$ and each $T' \in (0, T)$, then ω is a Y *patch solution* to (1-1)–(1-2) on $[0, T)$.

While (1-5) is stated for each single time t (akin to the definition of strong or classical solutions to a PDE), it agrees with the usual flow-map based definition of solutions to the two-dimensional Euler equation; see the remarks after Definition 1.2 in [Kiselev et al. 2017]. Since u is only Hölder continuous at the patch boundaries when $\alpha > 0$ (and hence the flow map may not be unique), this definition is more appropriate in the g-SQG case.

Theorem 1.5 in [Kiselev et al. 2017] shows that for any $\theta_1, \dots, \theta_N \in \mathbb{R} \setminus \{0\}$ and any bounded open sets $\Omega_1(0), \dots, \Omega_N(0) \subseteq \mathbb{R}^2$ whose boundaries are pairwise disjoint simple closed H^3 curves, there is a time $T \in (0, \infty]$ such that a unique H^3 patch solution $\omega = \sum_{n=1}^N \theta_n \chi_{\Omega_n(\,\cdot\,)}$ to (1-1)–(1-2) exists on $[0, T)$. And if the maximal such T is finite, then either $\sup_{t\in[0,T)} \|\Omega(t)\|_{H^3} = \infty$ or

$$\sup_{t\in[0,T)} \sup_{\substack{(n,\xi),(j,\eta)\in\{1,\dots,N\}\times\mathbb{T} \\ (n,\xi)\neq(j,\eta)}} \frac{|n-j| + |\xi - \eta|}{|z_n(\xi, t) - z_j(\eta, t)|} = \infty, \tag{1-6}$$

where $z_n(\,\cdot\,, t)$ is a constant-speed parametrization of $\partial\Omega_n(t)$ and $|\xi - \eta|$ is distance on $\mathbb{T}$. Note that if (1-6) holds with $n = j$ — which means that the *arc-chord ratio* for some $\Omega_n(\,\cdot\,)$ becomes unbounded as $t \to T$ — then this must be realized by a touch of "distinct" segments (which we call *folds*) of $\partial\Omega_n(\,\cdot\,)$ whenever $\sup_{t\in[0,T)} \|\Omega(t)\|_{C^{1,\gamma}} < \infty$ for some $\gamma > 0$. Indeed, since $|\partial_\xi z_n(\xi, t)|$ is then bounded below by a positive constant uniformly in (n, ξ, t) (see (2-4)), it follows that the fraction in (1-6) is uniformly bounded above when $n = j$ and $|\xi - \eta|$ is small enough. Therefore (1-6) is the correct definition of a *splash-like singularity* at time T (i.e., a touch of either boundaries of distinct patches or folds of the same patch boundary, including both simple and complex splashes) when $\sup_{t\in[0,T)} \|\Omega(t)\|_{C^{1,\gamma}} < \infty$ for some $\gamma > 0$.

The following theorem is now our main result.

Theorem 1.3. *If* $\alpha \in \left(0, \frac{1}{4}\right]$ *and a* $C^{1,2\alpha/(1-2\alpha)}$ *patch solution to* (1-1)–(1-2) *on the time interval* $[0, T)$ *with* $T < \infty$ *satisfies* $\sup_{t\in[0,T)} \|\Omega(t)\|_{C^{1,2\alpha/(1-2\alpha)}} < \infty$, *then* (1-6) *fails* (*so no splash-like singularity can occur*). *In particular*, *if the maximal time* T *of existence of an* H^3 *patch solution from* [Kiselev et al. 2017, Theorem 1.5] *is finite and* $\alpha \in \left(0, \frac{1}{4}\right]$, *then* $\sup_{t\in[0,T)} \|\Omega(t)\|_{H^3} = \infty$.

Remark. (1) Our proof shows that the left-hand side of (1-6) with $\sup_{t\in[0,T)}$ removed can grow at most exponentially in time (up to time T) if $\sup_{t\in[0,T)} \|\Omega(t)\|_{C^{1,2\alpha/(1-2\alpha)}} < \infty$. Hence boundaries of distinct patches, as well as folds of the same patch boundary, can only approach each other exponentially quickly in this case.

(2) While we do not know whether this result extends to some $\alpha > \frac{1}{4}$, in Section 2E we provide an extension to all $\alpha \in (0, 1)$ when one a priori requires that only simple splashes can occur (i.e., no more than two segments of $\partial\Omega$ are allowed to touch in the same location) and $\sup_{t\in[0,T)} \|\Omega(t)\|_{C^{k,\gamma}} < \infty$ holds for $k = 1$ and some $\gamma \geq 2\alpha$, when $\alpha \in \left(0, \frac{1}{2}\right]$, or for $k = 2$ and some $\gamma \geq 2\alpha - 1$, when $\alpha \in \left[\frac{1}{2}, 1\right)$. The obtained bound on the approach rate of two patches/folds is now double exponential when γ is equal to the minimal value above (2α or $2\alpha - 1$), and exponential otherwise. We note that when the potential simple splash is assumed to have a predetermined location and development of singularities elsewhere is a priori excluded, then this was also proved for $\alpha = \frac{1}{2}$ in [Gancedo and Strain 2014] (for smooth patches and with a double exponential bound on the approach rate), and for all $\alpha \in (0, 1)$ in [Kiselev and Luo 2023] (this work was done contemporaneously with and independently of ours).

Finally, here is an extension to the half-plane; see Section 3 for the relevant adjustments.

Theorem 1.4. *Theorem 1.3 extends to patch solutions on the half-plane*, *with the second claim involving* H^3 *patch solutions from* [Kiselev et al. 2017, Theorem 1.4] *and* $\alpha \in \left(0, \frac{1}{24}\right)$, *or* H^2 *patch solutions from* [Gancedo and Patel 2021, Theorem 1.1] *and* $\alpha \in \left(0, \frac{1}{6}\right)$.

It was proved in [Kiselev et al. 2016] that, for any $\alpha \in \left(0, \frac{1}{24}\right)$, there are H^3 patch solutions on the half-plane that become singular in finite time. For $\alpha \in \left(0, \frac{1}{6}\right)$ and H^2 patch solutions this was proved in [Gancedo and Patel 2021]. Theorem 1.4 shows that this cannot happen only via a splash-like singularity and always involves blow-up of their H^3 and H^2 norms, respectively.

2. Proof of Theorem 1.3

2A. ***The single patch case.*** For the sake of simplicity of notation, let us first consider the case of a single patch on which $\omega \equiv 1$; that is, $\omega(\cdot, t) = \chi_{\Omega(t)}$. Then (1-3) becomes

$$u(x, t) := \int_{\Omega(t)} \frac{(x - y)^\perp}{|x - y|^{2+2\alpha}}\, dy \tag{2-1}$$

after rescaling (1-1) in time by c_α (which we do in order to remove the constant).

We will not assume $\alpha \leq \frac{1}{4}$ until it is needed, so that it is clear where this hypothesis enters into our argument. We will therefore consider a $C^{1,\gamma}$ patch solution with any $\gamma \in (0, 1]$ below. If now $z(\cdot, t)$ is any constant-speed parametrization of $\partial\Omega(t)$ for $t \in [0, T)$, we assume that

$$M := \sup_{t\in[0,T)} \|z(\,\cdot\,, t)\|_{C^{1,\gamma}} < \infty. \tag{2-2}$$

We now want to show that this implies

$$\sup_{t\in[0,T)} \sup_{\substack{\xi,\eta\in\mathbb{T}\\ \xi\neq\eta}} \frac{|\xi-\eta|}{|z(\xi,t)-z(\eta,t)|} < \infty. \tag{2-3}$$

Since a C^1 patch solution is also a weak solution to (1-1)–(1-2) with $|\Omega(t)|$ being conserved (see Remark 3 after Definition 1.2 in [Kiselev et al. 2017]), the isoperimetric inequality shows that

$$M' := \inf_{(\xi,t)\in\mathbb{T}\times[0,T)} |\partial_\xi z(\xi,t)| > 0. \tag{2-4}$$

Now for any $t \in [0, T)$ and $\xi, \eta \in \mathbb{T}$, there are $\xi_1, \xi_2 \in \mathbb{T}$ between ξ and η such that

$$|z(\xi,t)-z(\eta,t)| = |\xi-\eta|\,|(\partial_\xi z_1(\xi_1,t), \partial_\xi z_2(\xi_2,t))| \geq |\xi-\eta|(|\partial_\xi z(\xi,t)| - 2M|\xi-\eta|^\gamma).$$

Hence if we let $\delta := (M'/4M)^{1/\gamma}$, then

$$|z(\xi,t)-z(\eta,t)| \geq \tfrac{1}{2}M'|\xi-\eta| \tag{2-5}$$

whenever $|\xi-\eta| \leq \delta$. To conclude (2-3), it now suffices to show

$$\inf_{t\in[0,T)} \min_{\substack{\xi,\eta\in\mathbb{T}\\ |\xi-\eta|\geq\delta}} |z(\xi,t)-z(\eta,t)| > 0. \tag{2-6}$$

We therefore let

$$m(t) := \min_{\substack{\xi,\eta\in\mathbb{T}\\ |\xi-\eta|\geq\delta}} |z(\xi,t)-z(\eta,t)| \geq 0, \tag{2-7}$$

and let $\xi_t, \eta_t \in \mathbb{T}$ be such that $|z(\xi_t,t)-z(\eta_t,t)| = m(t)$. If (2-6) fails, then clearly for all $t < T$ close enough to T we have $m(t) < \frac{1}{2}M'\delta$, which shows that $|\xi_t-\eta_t| > \delta$ for these t because (2-5) holds. It suffices to consider only such t. Then, following an argument in [Constantin and Escher 1998], one can easily see that $m(t)$ is locally Lipschitz (and so differentiable at almost all such t) and we have

$$m'(t) = \frac{z(\xi_t,t)-z(\eta_t,t)}{m(t)} \cdot (u(z(\xi_t,t),t) - u(z(\eta_t,t),t)) \tag{2-8}$$

for almost every such t. Hence Grönwall's inequality shows that it suffices to prove

$$-(u(z(\xi_t,t),t) - u(z(\eta_t,t),t)) \cdot n_t \leq Cm(t) \tag{2-9}$$

for some t-independent $C < \infty$ and all t such that $m(t) \in \left(0, \frac{1}{2}M'\delta\right)$, where $n_t := (z(\xi_t,t)-z(\eta_t,t))/m(t)$ is the unit vector in the direction $z(\xi_t,t)-z(\eta_t,t)$. Of course, the definition of ξ_t, η_t shows that n_t is also normal to $\partial\Omega(t)$ at both $z(\xi_t,t)$ and $z(\eta_t,t)$.

Since (2-9) only involves quantities at a single time, we will now assume that t is close to T and drop the dependence of Ω, z, u, and m on t from our notation. The above also shows that after a translation and rotation we can assume:

(1) $z(\eta_t) = (0, 0)$ and $z(\xi_t) = (0, m)$, with $m \in \left(0, \frac{1}{2}M'\delta\right)$.

(2) $\partial_\xi z(\xi_t), \partial_\xi z(\eta_t) \perp (0, 1)$.

We will do so, and then (2-9) becomes just

$$u_2(0, 0) - u_2(0, m) \le Cm. \tag{2-10}$$

We will prove this in the next three subsections. We note that all constants below may depend on α, γ, M, M' (recall that δ also depends on these), but will be independent of m and t.

2B. *Some geometric lemmas.* We first state some geometric lemmas that will be used throughout. The first of these is a trivial consequence of $C^{1,\gamma}$-regularity of $\partial\Omega$, which says that near any $z(\xi)$, the curve z is the graph of some function $f : \mathbb{R} \to \mathbb{R}$ defined with respect to the coordinate system centered at $z(\xi)$ and with the horizontal axis not too far from $\partial_\xi z(\xi)$.

Lemma 2.1. *There are $A \ge 1$ and $R_0 > 0$ such that, for any $\xi \in \mathbb{T}$ and any $v \in \mathbb{S}^1$ with $|\partial_\xi z(\xi) \cdot v| \ge \frac{1}{2}|\partial_\xi z(\xi)|$, there is $f : [-R_0, R_0] \to \mathbb{R}$ with $\|f\|_{C^{1,\gamma}} \le A$ such that*

$$\{z(\xi) + hv + f(h)v^\perp \mid h \in [-R_1, R_1]\} = z([\xi - \xi_1, \xi + \xi_2])$$

for each $R_1 \in [0, R_0]$ and some $\xi_1, \xi_2 \in [R_1/M, 3R_1/M']$. Then

$$f(0) = 0 \quad \textit{and} \quad f'(0) = \frac{\partial_\xi z(\xi) \cdot v^\perp}{\partial_\xi z(\xi) \cdot v}.$$

The next lemma shows that when two folds of $\partial\Omega$ are close to each other, the angles between their tangent lines are controlled by their distance.

Lemma 2.2. *There are $B, R > 0$ such that, for any $\xi, \eta \in \mathbb{T}$ with $|z(\xi) - z(\eta)| \le R$ we have*

$$|\tan\theta| \le B|z(\xi) - z(\eta)|^{\gamma/(1+\gamma)}, \tag{2-11}$$

where θ is the angle between $\partial_\xi z(\xi)$ and $\partial_\xi z(\eta)$.

Proof. Let A, R_0 be from Lemma 2.1. First note that it suffices to prove

$$|\sin\theta| \le B|z(\xi) - z(\eta)|^{\gamma/(1+\gamma)} \tag{2-12}$$

instead of (2-11). Indeed, we then only need to replace R by $\min\{R, (2B)^{-(1+\gamma)/\gamma}\}$, which yields $|\cos\theta| \ge \frac{1}{2}$, and then double B.

Take $C := 9A$, and let $R := \min\left\{\frac{1}{2}C^{-(2+2\gamma)/\gamma}, (R_0/(3C))^2\right\}$ and $B := 3C^2$. Without loss assume that $z(\eta) = 0$ and $\partial_\xi z(\eta)/|\partial_\xi z(\eta)| = (1, 0)$, and then let $r := |z(\xi)| \le R$ and $r' := Cr^{1/(1+\gamma)} \le CR^{1/2} \le \frac{1}{3}R_0$. Then Lemma 2.1 with $v := (1, 0)$ shows that z near the origin is a curve connecting the vertical sides of the rectangle $Q := [-3r', 3r'] \times [-C^3r, C^3r]$ because

$$A(3r')^\gamma(3r') \le 9AC^2r \le C^3r$$

(note that the definition of R shows that $C^3r < r'$, so the vertical sides are the shorter ones).

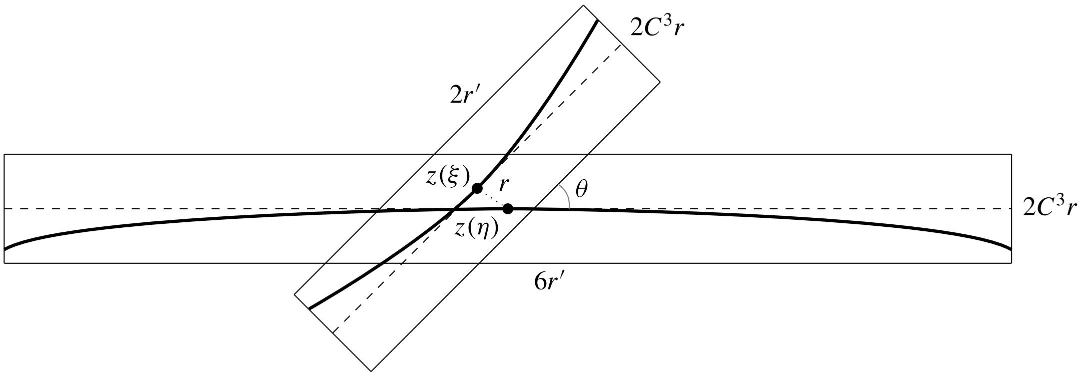

Figure 1. The curve z inside the rectangles Q and Q'.

Apply the same argument with $v := \partial_\xi z(\xi)/|\partial_\xi z(\xi)|$ and the rectangle Q' centered at $z(\xi)$ whose longer axis connects the points $z(\xi) \pm vr'$ and whose shorter sides have again length $2C^3r$. It shows that z near $z(\xi)$ is a curve connecting the shorter sides of Q'. If (2-12) is violated, then one of these sides lies fully in $(-3r', 3r') \times (C^3r, \infty)$ and the other in $(-3r', 3r') \times (-\infty, -C^3r)$ (see Figure 1) because $r + r' + C^3r < 3r'$ and

$$r' \sin\theta > BCr \geq 3C^3r > 2C^3r + r.$$

But this means that the two curves must intersect, a contradiction with our assumption that no touch has occurred before time T. $\square$

We can now combine Lemmas 2.1 and 2.2 to obtain the following constraint on the geometry of $\partial\Omega$ near the origin.

Lemma 2.3. *In the setting of* (1) *and* (2), *there are* $A, B, R > 0$ *with*

$$B(3R)^{\gamma/(1+\gamma)} \leq \tfrac{1}{2} \quad \textit{and} \quad M(4R)^\gamma \leq (M')^{1+\gamma}$$

such that, for any $\xi \in \mathbb{T}$ *with* $z(\xi) \in [-R, R] \times [-2R, 2R]$, *there is* $f : [-R, R] \to \mathbb{R}$ *with*

$$\|f\|_{C^{1,\gamma}} \leq A \quad \textit{and} \quad |f'(z_1(\xi))| \leq B|z(\xi)|^{\gamma/(1+\gamma)}$$

such that the graph of f is a segment of the curve z around $z(\xi)$. In particular,

$$|f(h) - z_2(\xi) - f'(z_1(\xi))(h - z_1(\xi))| \leq A|h - z_1(\xi)|^{1+\gamma}$$

for all $h \in [-R, R]$. *And if* $|f(h')| > 2R$ *for some* $h' \in [-R, R]$, *then* $|f(h)| > R$ *for all* $h \in [-R, R]$.

Proof. The first statement is an immediate consequence of Lemmas 2.1 and 2.2, with A from Lemma 2.1, B from Lemma 2.2, and R being the minimum of one third of R from Lemma 2.2 and

$$\min\left\{\tfrac{1}{3}(2B)^{-(1+\gamma)/\gamma}, \tfrac{1}{4}(M')^{(1+\gamma)/\gamma}M^{-1/\gamma}\right\}$$

(Lemma 2.1 is applied with $v := (1, 0)$ and $z_2(\xi)$ is added to the obtained f). The second statement is its immediate consequence, while the third holds by $B(3R)^{\gamma/(1+\gamma)} \leq \frac{1}{2}$. $\square$

The third claim shows that any connected component of $\partial\Omega \cap ([-R, R] \times [-2R, 2R])$ that intersects $[-R, R]^2$ is a graph of a function $f : [-R, R] \to [-2R, 2R]$ that satisfies the lemma. Note also that since the arc-length of any such component is at least $2R$ and the arc-length of $\partial\Omega$ is uniformly bounded above because so is $\|\partial\Omega\|_{C^1}$, it follows that the number of such components is bounded above by some constant K. That is, if we assume (2-2), only a finite number of folds of $\partial\Omega$ might potentially create a single touch (splash) at time T; we will show below that this is in fact not possible when $\alpha \le \frac{1}{4}$.

2C. ***Reduction to regions near individual boundary segments.*** Take A, B, R from Lemma 2.3, and K from the above discussion. From (2-1) we see that the left-hand side of (2-10) is the sum of the terms

$$I := \int_{\Omega\cap[-R,R]^2} \left(\frac{y_1}{|y|^{2+2\alpha}} - \frac{y_1}{|y-(0,m)|^{2+2\alpha}} \right) dy$$

and

$$I' := \int_{\Omega\setminus[-R,R]^2} \left(\frac{y_1}{|y|^{2+2\alpha}} - \frac{y_1}{|y-(0,m)|^{2+2\alpha}} \right) dy.$$

To prove (2-10), it clearly suffices to assume that $m \le \frac{1}{2}R$, in which case clearly $|I'| \le Cm$ for some constant C. Hence we only need to show that $I \le Cm$.

Assume that $f_1, \dots, f_k \colon [-R, R] \to [-2R, 2R]$ are distinct functions whose graphs are all the connected components of $\partial\Omega$ from the paragraph after Lemma 2.3 (so $k \le K$), and order them so that $f_1(0) < \cdots < f_k(0)$. Let

$$g_i := \operatorname{sgn}(f_i) \min\{|f_i|, R\},$$

so that

$$I \le \sum_{i=1}^{k+1} \left| \int_{-R}^{R} \int_{g_{i-1}(h)}^{g_i(h)} \left(\frac{h}{(h^2+v^2)^{1+\alpha}} - \frac{h}{(h^2+(v-m)^2)^{1+\alpha}} \right) dv\, dh \right|, \tag{2-13}$$

where $g_0 \equiv -R$ and $g_{k+1} \equiv R$. Since the integrand is odd in h, its integral on any region symmetric with respect to the vertical axis is zero. Since $[-R, R] \times [g_{i-1}(0), g_i(0)]$ is such a region we can replace the integral $\int_{g_{i-1}(h)}^{g_i(h)}$ in (2-13) by the sum of integrals $\int_{g_{i-1}(h)}^{g_{i-1}(0)}$ and $\int_{g_i(0)}^{g_i(h)}$ (with the same integrand). We therefore obtain $|I| \le 2\sum_{i=1}^{k} |I_i|$, where

$$I_i := \int_{-R}^{R} \int_{g_i(0)}^{g_i(h)} \left(\frac{h}{(h^2+v^2)^{1+\alpha}} - \frac{h}{(h^2+(v-m)^2)^{1+\alpha}}\, dv \right) dh.$$

Hence, we are left with showing $|I_i| \le Cm$ for each i. When doing this, we can just assume that $g_i = f_i$ because the error we incur by this involves only integration over $\Omega \setminus [-R, R]^2$ and therefore is no more than Cm (similarly to I').

2D. ***Estimating the individual integrals.*** We thus consider any $f \colon [-R, R] \to [-2R, 2R]$ whose graph is a segment of the curve z passing through a point in $[-R, R]^2$, let

$$J := \int_{-R}^{R} \int_{f(0)}^{f(h)} \left(\frac{h}{(h^2+v^2)^{1+\alpha}} - \frac{h}{(h^2+(v-m)^2)^{1+\alpha}} \right) dv\, dh, \tag{2-14}$$

and need to show that $|J| \le Cm$. We further divide this integral into two pieces:

$$J_1 := \int_{-R}^{R} \int_{f(0)}^{f(0)+hf'(0)} \left(\frac{h}{(h^2+v^2)^{1+\alpha}} - \frac{h}{(h^2+(v-m)^2)^{1+\alpha}} \right) dv\, dh,$$
$$J_2 := \int_{-R}^{R} \int_{f(0)+hf'(0)}^{f(h)} \left(\frac{h}{(h^2+v^2)^{1+\alpha}} - \frac{h}{(h^2+(v-m)^2)^{1+\alpha}} \right) dv\, dh,$$

and estimate J_2 first.

Lemma 2.4. *We have $|J_2| \le Cm$ when $\gamma > 2\alpha$, and $|J_2| \le Cm(1+\ln_- m)$ when $\gamma = 2\alpha$, for some constant C.*

Proof. By Lemma 2.3, we have

$$|J_2| \le \int_{-R}^{R} \int_{f(0)+hf'(0)-A|h|^{1+\gamma}}^{f(0)+hf'(0)+A|h|^{1+\gamma}} \left| \frac{h}{(h^2+v^2)^{1+\alpha}} - \frac{h}{(h^2+(v-m)^2)^{1+\alpha}} \right| dv\, dh.$$

The mean value theorem yields

$$\left| \frac{h}{(h^2+v^2)^{1+\alpha}} - \frac{h}{(h^2+(v-m)^2)^{1+\alpha}} \right| = \frac{(2+2\alpha)|h|(h^2+\bar{v}^2)^{\alpha}|\bar{v}|}{(h^2+v^2)^{1+\alpha}(h^2+(v-m)^2)^{1+\alpha}} m$$

for some $\bar{v} \in [v-m, v]$. Hence

$$\left| \frac{h}{(h^2+v^2)^{1+\alpha}} - \frac{h}{(h^2+(v-m)^2)^{1+\alpha}} \right| \le \frac{3m}{|h|^{2+2\alpha}},$$

and if $V := \max\{|v|, |v-m|\} \ge |h|$, then we also have

$$\left| \frac{h}{(h^2+v^2)^{1+\alpha}} - \frac{h}{(h^2+(v-m)^2)^{1+\alpha}} \right| \le \frac{3m|h|2^{\alpha}V^{1+2\alpha}}{|h|^{2+2\alpha}V^{2+2\alpha}} \le \frac{6m}{|h|^{1+2\alpha}V}.$$

Since $V \ge \frac{1}{2}m$ and we assume that $m \le \frac{1}{2}R$ (see the start of Section 2C), we obtain

$$|J_2| \le 2\int_0^{m/2} \frac{12}{|h|^{1+2\alpha}} 2A|h|^{1+\gamma}\, dh + 2\int_{m/2}^{R} \frac{6m}{|h|^{2+2\alpha}} 2A|h|^{1+\gamma}\, dh.$$

This is less than Cm if $\gamma > 2\alpha$ and less than $Cm(1+\ln_- m)$ if $\gamma = 2\alpha$ (for some C). □

To estimate J_1, it suffices to assume that $f(0) \notin [0, m]$. Indeed, if $f(0) \in \{0, m\}$, then the graph of f contains either $(0, 0)$ or $(0, m)$, so (1) and (2) above (2-10) imply $f'(0) = 0$ and therefore $J_1 = 0$. And if $f(0) \in (0, m)$, then the definition of m shows that there must be $\eta \in \mathbb{T}$ with $|\eta - \eta_t| < \delta$ such that $z(\eta) = (0, f(0))$. Here (2-5) yields $|\eta - \eta_t| \le 4R/M'$, and hence for all ξ between η and η_t we have

$$|\partial_\xi z(\xi) - \partial_\xi z(\eta_t)| \le M\left(\frac{4R}{M'}\right)^{\gamma} \le M' \le |\partial_\xi z(\eta_t)|$$

by Lemma 2.3. This shows that $\partial_\xi z(\xi) \cdot \partial_\xi z(\eta_t) \ge 0$ for all these ξ, which clearly contradicts

$$(z(\eta) - z(\eta_t)) \cdot \partial_\xi z(\eta_t) = 0.$$

So $f(0) \notin [0, m]$, and we define $a := -f(0) > 0$ when $f(0) < 0$, and $a := f(0) - m > 0$ when $f(0) > m$. In both cases Lemma 2.3 yields

$$|f'(0)| \le Ba^{\gamma/(1+\gamma)} \le BR^{\gamma/(1+\gamma)} \le \tfrac{1}{2}. \tag{2-15}$$

Lemma 2.5. *We have $|J_1| \le Ca^{\gamma/(1+\gamma)}(a+m)^{-2\alpha}m$ for some constant C.*

Proof. Note that the definition of m shows that $a \ge m$, so we could replace $a+m$ by a. We will not use this so that this result also applies in Section 3. We will assume $f(0) < 0$ since the proof for the other case is virtually identical. We can then rewrite J_1 as

$$J_1 = \int_{-R}^{R} \int_0^{hf'(0)} \left(\frac{h}{(h^2+(v-a)^2)^{1+\alpha}} - \frac{h}{(h^2+(v-a-m)^2)^{1+\alpha}} \right) dv\, dh.$$

We split the integral into two parts:

$$J_3 := \int_{|h|<a+m} \int_0^{hf'(0)} \left(\frac{h}{(h^2+(v-a)^2)^{1+\alpha}} - \frac{h}{(h^2+(v-a-m)^2)^{1+\alpha}} \right) dv\, dh,$$
$$J_4 := \int_{a+m\le|h|\le R} \int_0^{hf'(0)} \left(\frac{h}{(h^2+(v-a)^2)^{1+\alpha}} - \frac{h}{(h^2+(v-a-m)^2)^{1+\alpha}} \right) dv\, dh.$$

For J_3, note that for any v in the domain of integration, we have

$$v-a-m \le |(a+m)f'(0)| - a - m \le -\tfrac{1}{2}(a+m).$$

This also shows that $v-a \le \frac{1}{2}(m-a)$, so $|v-a| \le |v-a-m|$. The mean value theorem then gives, for some $\bar{v} \in [v-a-m, v-a]$,

$$\left| \frac{h}{(h^2+(v-a)^2)^{1+\alpha}} - \frac{h}{(h^2+(v-a-m)^2)^{1+\alpha}} \right| = \frac{(2+2\alpha)|h|(h^2+\bar{v}^2)^{\alpha}|\bar{v}|}{(h^2+(v-a)^2)^{1+\alpha}(h^2+(v-a-m)^2)^{1+\alpha}} m$$
$$\le \frac{6m}{|h|^{1+2\alpha}(a+m)}$$

because $\max\bigl\{|\bar{v}|, \frac{1}{2}(a+m)\bigr\} \le |v-a-m|$. This and (2-15) yield

$$|J_3| \le \int_{-a-m}^{a+m} \frac{6m}{|h|^{1+2\alpha}(a+m)} |hf'(0)|\, dh \le \frac{12m}{1-2\alpha} \frac{|f'(0)|}{(a+m)^{2\alpha}} \le \frac{12B}{1-2\alpha} a^{\gamma/(1+\gamma)} \frac{m}{(a+m)^{2\alpha}}.$$

As for J_4, the mean value theorem yields

$$\left| \frac{h}{(h^2+(v-a)^2)^{1+\alpha}} - \frac{h}{(h^2+(v-a-m)^2)^{1+\alpha}} \right| \le \frac{3m}{|h|^{2+2\alpha}}.$$

From this and (2-15) we obtain

$$|J_4| \le 2\int_{a+m}^{R} \frac{3m}{|h|^{2+2\alpha}} |hf'(0)|\, dh \le \frac{3m}{\alpha} \frac{|f'(0)|}{(a+m)^{2\alpha}} \le \frac{3B}{\alpha} a^{\gamma/(1+\gamma)} \frac{m}{(a+m)^{2\alpha}}. \qquad \square$$

The last two lemmas, together with the estimate $|I'| \le Cm$ above and $k \le K$, show that (2-10) holds when $\gamma > 2\alpha$ and $\gamma/(1+\gamma) \ge 2\alpha$. Since $\gamma = 2\alpha/(1-2\alpha)$ satisfies this (and (2-2) holds for it by the hypothesis), the proof of the single-patch case of Theorem 1.3 (and of Remark (1) after it) is finished.

2E. ***Absence of simple splashes for all $\alpha \in (0, 1)$.*** Let us now assume that only simple splashes can happen for a $C^{1,\gamma}$ patch and $\alpha \in \left(0, \frac{1}{2}\right)$. That is, there is $R > 0$ such that, for all t close enough to T and any $\xi_t, \eta_t \in \mathbb{T}$ satisfying $|z(\xi_t, t) - z(\eta_t, t)| = m(t)$, there is no $\xi \in \mathbb{T}$ such that $\min\{|\xi - \xi_t|, |\xi - \eta_t|\} \ge \delta$ and also $|z(\xi, t) - z(\eta_t, t)| \le R$. This essentially means that any potential splash only involves two folds of $\partial\Omega$, although this requirement is in fact weaker than that: multiple folds are allowed but not near minimizers of (2-7). Then in Lemma 2.5 we have $f'(0) = 0$ and so $J_1 = 0$. Hence Lemma 2.4 shows that a simple splash cannot occur by time T if $\sup_{t\in[0,T)} \|\Omega(t)\|_{C^{1,\gamma}} < \infty$ for some $\gamma \in [2\alpha, 1]$, and $m(t)$ can decrease at most exponentially when $\gamma > 2\alpha$ and at most double exponentially when $\gamma = 2\alpha$.

In fact, one can extend this result to all $\alpha \in \left[\frac{1}{2}, 1\right)$. In this case one must replace u in (1-5) (which becomes infinite on $\partial\Omega(t)$) by its normal "component"

$$u_n(x, t) := \text{p.v.}\int_{\Omega(t)} c_\alpha \frac{(x-y)^\perp \cdot n_{x,t}}{|x-y|^{2+2\alpha}}\, dy\, n_{x,t}$$

(which is finite), with $n_{x,t}$ the unit outer normal vector to $\Omega(t)$ at $x \in \partial\Omega(t)$; see also [Kiselev et al. 2017, Remark 2 after Definition 1.2] or [Kiselev and Luo 2023]. If we now assume $\sup_{t\in[0,T)} \|\Omega(t)\|_{C^{2,\gamma}} < \infty$, one can use (2) above (2-10) to show that $\left| f(h) - f(0) - \frac{1}{2} f''(0)h^2 \right| \le A|h|^{2+\gamma}$ in (2-14). Then the oddness of the integrand in h will yield the estimate

$$|J_2| \le 2\int_0^{m/2} \frac{16}{|h|^{1+2\alpha}} 2A|h|^{2+\gamma}\, dh + 2\int_{m/2}^{R} \frac{8m}{|h|^{2+2\alpha}} 2A|h|^{2+\gamma}\, dh$$

in the proof of Lemma 2.4 whenever $\gamma \in [2\alpha - 1, 1]$. Hence no finite time simple splash can happen by time T in this case either, and we again obtain an exponential (resp. double exponential) lower bound on $m(t)$ when $\gamma > 2\alpha - 1$ (resp. $\gamma = 2\alpha - 1$). Note also that for $\alpha = \frac{1}{2}$ it even suffices to assume $\sup_{t\in[0,T)} \|\Omega(t)\|_{C^{1,1}} < \infty$, with $2+\gamma$ replaced by 2 and with a double exponential lower bound on $m(t)$.

2F. ***The multiple patches case.*** In the general multiple patches case, (2-3) becomes

$$\sup_{t\in[0,T)} \sup_{\substack{(n,\xi),(j,\eta)\in Z_N\times\mathbb{T} \\ (n,\xi)\ne(j,\eta)}} \frac{|n-j| + |\xi - \eta|}{|z_n(\xi, t) - z_j(\eta, t)|} < \infty,$$

where $Z_N := \{1, \dots, N\}$ and $z_n(\,\cdot\,, t)$ is a constant-speed parametrization of $\partial\Omega_n(t)$. We choose the same δ (with all z_n included in the definitions of M and M'), and then

$$m(t) := \min_{\substack{(n,\xi),(j,\eta)\in Z_N\times\mathbb{T} \\ |(n,\xi)-(j,\eta)|\ge\delta}} |z_n(\xi, t) - z_j(\eta, t)| \ge 0. \tag{2-16}$$

The points ξ_t and η_t may now be on the boundaries of distinct patches, but that does not change our analysis, which only deals with the individual patch segments in a small rectangle centered at η_t. The geometric lemmas are unchanged; the estimates on integrals I' and I_i in Section 2C only change by the factor $|\theta_1|+\cdots+|\theta_N|$ and hence so does the rest of the argument. This finishes the proof of Theorem 1.3 (and of Remark (1) after it) as stated. The claim in Remark (2) after Theorem 1.3 also extends to this case.

3. Proof of Theorem 1.4

Let us now turn to the half-plane case $D := \mathbb{R}\times\mathbb{R}^+$, when the proof is essentially identical to Theorem 1.3 (and the H^3 and H^2 local well-posedness results from [Gancedo and Patel 2021; Kiselev et al. 2017] require $\alpha\in\big(0,\frac{1}{24}\big)$ and $\alpha\in\big(0,\frac{1}{6}\big)$, respectively).

Let us first recall the definition of patch solutions in this setting from [Kiselev et al. 2017]. Equation (1-1) is unchanged, and Δ in (1-2) is the Dirichlet Laplacian on D. If we assume that $\alpha\in\big(0,\frac{1}{2}\big)$, this means that for an appropriate constant $c_\alpha>0$ we have

$$u(x,t)=c_\alpha\int_D\left(\frac{(x-y)^\perp}{|x-y|^{2+2\alpha}}-\frac{(x-\bar{y})^\perp}{|x-\bar{y}|^{2+2\alpha}}\right)\omega(y,t)\,dy \tag{3-1}$$

for each $x\in\overline{D}$, where $\bar{y}:=(y_1,-y_2)$. Definition 1.2 is as before, but with the patches $\Omega_1(t),\dots,\Omega_N(t)$ now contained in D instead of $\mathbb{R}^2$, and with u from (3-1) instead of (1-3).

We define M, M', and δ as before and $m(t)$ via (2-16). We also consider the reflected patches $\overline{\Omega}_n(t):=\{y\in\mathbb{R}^2\setminus D\mid \bar{y}\in\Omega_n(t)\}$, which allows us to write (after dropping c_α via rescaling)

$$u(x,t)=\sum_{n=1}^N\theta_n\int_{\Omega_n(t)}\frac{(x-y)^\perp}{|x-y|^{2+2\alpha}}\,dy-\sum_{n=1}^N\theta_n\int_{\overline{\Omega}_n(t)}\frac{(x-y)^\perp}{|x-y|^{2+2\alpha}}\,dy.$$

Theorem 1.4 will now follow once we show (2-9) with this u. This is proved in the same way as on $\mathbb{R}^2$, but now the boundary segments defining functions f_i in Section 2C can belong to both the original and the reflected patches. Note that the distance of $\partial\Omega_n(t)$ and $\partial\overline{\Omega}_n(t)$ can be less than $m(t)$ — even 0 because they can touch at ∂D, in which case their normal vectors coincide at any point of touch. But they obviously cannot cross — this is why we did not assume $a\geq m$ in Lemma 2.5 — which allows us to use the same estimates as in Section 2, modulo a factor of 2 due to the number of patches now being doubled.

References

[Castro et al. 2013] A. Castro, D. Córdoba, C. Fefferman, F. Gancedo, and J. Gómez-Serrano, "Finite time singularities for the free boundary incompressible Euler equations", *Ann. of Math.* (2) **178**:3 (2013), 1061–1134. MR Zbl

[Chae et al. 2012] D. Chae, P. Constantin, D. Córdoba, F. Gancedo, and J. Wu, "Generalized surface quasi-geostrophic equations with singular velocities", *Comm. Pure Appl. Math.* **65**:8 (2012), 1037–1066. MR Zbl

[Chemin 1993] J.-Y. Chemin, "Persistance de structures géométriques dans les fluides incompressibles bidimensionnels", *Ann. Sci. École Norm. Sup.* (4) **26**:4 (1993), 517–542. MR Zbl

[Constantin and Escher 1998] A. Constantin and J. Escher, "Wave breaking for nonlinear nonlocal shallow water equations", *Acta Math.* **181**:2 (1998), 229–243. MR Zbl

[Constantin et al. 1994] P. Constantin, A. J. Majda, and E. Tabak, "Formation of strong fronts in the 2-D quasigeostrophic thermal active scalar", *Nonlinearity* **7**:6 (1994), 1495–1533. MR Zbl

[Constantin et al. 2008] P. Constantin, G. Iyer, and J. Wu, "Global regularity for a modified critical dissipative quasi-geostrophic equation", *Indiana Univ. Math. J.* **57**:6 (2008), 2681–2692. MR Zbl

[Córdoba et al. 2005] D. Córdoba, M. A. Fontelos, A. M. Mancho, and J. L. Rodrigo, "Evidence of singularities for a family of contour dynamics equations", *Proc. Natl. Acad. Sci. USA* **102**:17 (2005), 5949–5952. MR Zbl

[Córdoba et al. 2018] A. Córdoba, D. Córdoba, and F. Gancedo, "Uniqueness for SQG patch solutions", *Trans. Amer. Math. Soc. Ser. B* **5** (2018), 1–31. MR Zbl

[Coutand and Shkoller 2014] D. Coutand and S. Shkoller, "On the finite-time splash and splat singularities for the 3-D free-surface Euler equations", *Comm. Math. Phys.* **325**:1 (2014), 143–183. MR Zbl

[Gancedo 2008] F. Gancedo, "Existence for the α-patch model and the QG sharp front in Sobolev spaces", *Adv. Math.* **217**:6 (2008), 2569–2598. MR Zbl

[Gancedo and Patel 2021] F. Gancedo and N. Patel, "On the local existence and blow-up for generalized SQG patches", *Ann. PDE* **7**:1 (2021), art. id. 4. MR Zbl

[Gancedo and Strain 2014] F. Gancedo and R. M. Strain, "Absence of splash singularities for surface quasi-geostrophic sharp fronts and the Muskat problem", *Proc. Natl. Acad. Sci. USA* **111**:2 (2014), 635–639. MR Zbl

[Gancedo et al. 2022] F. Gancedo, H. Q. Nguyen, and N. Patel, "Well-posedness for SQG sharp fronts with unbounded curvature", *Math. Models Methods Appl. Sci.* **32**:13 (2022), 2551–2599. MR

[Hölder 1933] E. Hölder, "Über die unbeschränkte Fortsetzbarkeit einer stetigen ebenen Bewegung in einer unbegrenzten inkompressiblen Flüssigkeit", *Math. Z.* **37**:1 (1933), 727–738. MR Zbl

[Kiselev and Luo 2023] A. Kiselev and X. Luo, "On nonexistence of splash singularities for the α-SQG patches", *J. Nonlinear Sci.* **33**:2 (2023), art. id. 37. MR

[Kiselev and Šverák 2014] A. Kiselev and V. Šverák, "Small scale creation for solutions of the incompressible two-dimensional Euler equation", *Ann. of Math.* (2) **180**:3 (2014), 1205–1220. MR Zbl

[Kiselev et al. 2016] A. Kiselev, L. Ryzhik, Y. Yao, and A. Zlatoš, "Finite time singularity for the modified SQG patch equation", *Ann. of Math.* (2) **184**:3 (2016), 909–948. MR Zbl

[Kiselev et al. 2017] A. Kiselev, Y. Yao, and A. Zlatoš, "Local regularity for the modified SQG patch equation", *Comm. Pure Appl. Math.* **70**:7 (2017), 1253–1315. MR Zbl

[Luo and Hou 2014a] G. Luo and T. Y. Hou, "Potentially singular solutions of the 3D axisymmetric Euler equations", *Proc. Natl. Acad. Sci. USA* **111**:36 (2014), 12968–12973. Zbl

[Luo and Hou 2014b] G. Luo and T. Y. Hou, "Toward the finite-time blowup of the 3D axisymmetric Euler equations: a numerical investigation", *Multiscale Model. Simul.* **12**:4 (2014), 1722–1776. MR Zbl

[Pedlosky 1979] J. Pedlosky, *Geophysical fluid dynamics*, Springer, 1979. Zbl

[Pierrehumbert et al. 1994] R. T. Pierrehumbert, I. M. Held, and K. L. Swanson, "Spectra of local and nonlocal two-dimensional turbulence", *Chaos Solitons Fractals* **4**:6 (1994), 1111–1116. Zbl

[Rodrigo 2005] J. L. Rodrigo, "On the evolution of sharp fronts for the quasi-geostrophic equation", *Comm. Pure Appl. Math.* **58**:6 (2005), 821–866. MR Zbl

[Smith et al. 2002] K. S. Smith, G. Boccaletti, C. C. Henning, I. Marinov, C. Y. Tam, I. M. Held, and G. K. Vallis, "Turbulent diffusion in the geostrophic inverse cascade", *J. Fluid Mech.* **469** (2002), 13–48. MR Zbl

[Wolibner 1933] W. Wolibner, "Un theorème sur l'existence du mouvement plan d'un fluide parfait, homogène, incompressible, pendant un temps infiniment long", *Math. Z.* **37**:1 (1933), 698–726. MR Zbl

[Yudovich 1963] V. I. Yudovich, "Non-stationary flows of an ideal incompressible fluid", *Zh. Vychisl. Mat. Mat. Fiz.* **3**:6 (1963), 1032–1066. In Russian; translated in *USSR Comput. Math. Math. Phys.* **3**:6 (1963), 1407–1456. MR

Received 5 Dec 2021. Revised 15 May 2022. Accepted 11 Aug 2022.

JUNEKEY JEON: j6jeon@ucsd.edu
Department of Mathematics, University of California San Diego, La Jolla, CA, United States

ANDREJ ZLATOŠ: zlatos@ucsd.edu
Department of Mathematics, University of California San Diego, La Jolla, CA, United States

mathematical sciences publishers msp

ANALYSIS AND PDE
Vol. 17 (2024), No. 3, pp. 1019–1126
DOI: 10.2140/apde.2024.17.1019

SPECTRAL GAP FOR OBSTACLE SCATTERING IN DIMENSION 2

LUCAS VACOSSIN

We study the problem of scattering by several strictly convex obstacles, with smooth boundary and satisfying a noneclipse condition. We show, in dimension 2 only, the existence of a spectral gap for the meromorphic continuation of the Laplace operator outside the obstacles. The proof of this result relies on a reduction to an *open hyperbolic quantum map*, achieved by Nonnenmacher et al. (*Ann. of Math.* (2) **179**:1 (2014), 179–251). In fact, we obtain a spectral gap for this type of object, which also has applications in potential scattering. The second main ingredient of this article is a fractal uncertainty principle. We adapt the techniques of Dyatlov et al. (*J. Amer. Math. Soc.* **35**:2 (2022), 361–465) to apply this fractal uncertainty principle in our context.

1. Introduction 1019
2. Main theorem and applications 1030
3. Preliminaries 1036
4. Construction of a refined quantum partition 1062
5. Reduction to a fractal uncertainty principle via microlocalization properties 1092
6. Application of the fractal uncertainty principle 1111
Appendix 1115
Acknowledgment 1125
References 1125

1. Introduction

Scattering by convex obstacles and spectral gap. We are interested by the problem of scattering by strictly convex obstacles in the plane; see Figure 1. Assume

$$\mathcal{O} = \bigcup_{j=1}^{J} \mathcal{O}_j,$$

where $\mathcal{O}_j$ are open, strictly convex connected obstacles in $\mathbb{R}^2$ having smooth boundary and satisfying the *Ikawa condition*: for $i \neq j \neq k$, $\overline{\mathcal{O}}_i$ does not intersect the convex hull of $\overline{\mathcal{O}}_j \cup \overline{\mathcal{O}}_k$. Let

$$\Omega = \mathbb{R}^2 \setminus \overline{\mathcal{O}}.$$

It is known that the resolvent of the Dirichlet Laplacian in Ω continues meromorphically to the logarithmic cover of $\mathbb{C}$; see for instance [Dyatlov and Zworski 2019]. More precisely, suppose that $\chi \in C_c^\infty(\mathbb{R}^2)$ is equal to 1 in a neighborhood of $\overline{\mathcal{O}}$. The map

$$\chi(-\Delta - \lambda^2)^{-1}\chi : L^2(\Omega) \to L^2(\Omega)$$

MSC2020: 35P05, 35P25, 35Q40, 35S30, 35J05, 35J10, 37D20.
Keywords: scattering resonances, spectral gap, fractal uncertainty principle.

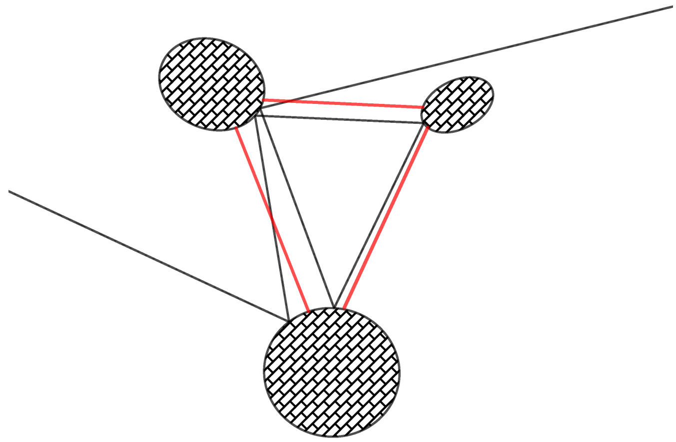

Figure 1. Scattering by three obstacles in the plane.

is holomorphic in the region $\{\operatorname{Im}\lambda > 0\}$ and it continues meromorphically to the logarithmic cover of $\mathbb{C}$. Its poles are the *scattering resonances*. We are interested in the problem of the existence of a spectral gap in the first sheet of the logarithmic cover (i.e., $\mathbb{C} \setminus i\mathbb{R}^-$). We prove the following theorem:

Theorem A. *There exist $\gamma > 0$ and $\lambda_0 > 0$ such that there is no resonance in the region*

$$[\lambda_0, +\infty[+ i[-\gamma, 0],$$

This problem has a long history in the physics and mathematics literature. The spectral gap was for instance studied by [Ikawa 1988] in dimension 3. It was experimentally investigated in [Barkhofen et al. 2013] for three- and five-disk systems. In this study, the author brings experimental evidence of the presence of a spectral gap, no matter how thin the trapped set is. For related problems concerning the distribution of scattering resonances for such systems, here is a nonexhaustive list of papers in which the reader can find pointers to a larger literature: [Gaspard and Rice 1989] for the three-disk problem, [Gérard 1988; Ikawa 1982] for the two-obstacle problem, [Petkov and Stoyanov 2010] for a link with dynamical zeta functions, [Bardos et al. 1987; Hargé and Lebeau 1994] for the diffraction by one convex obstacle, [Sjöstrand and Zworski 1999] among others papers of the two authors concerning the distribution of the scattering resonances. We will also widely use the presentation and the arguments of [Nonnenmacher et al. 2014].

The spectral gap problem is a high-frequency problem and justifies the introduction of a small parameter h, where $1/h$ corresponds to a large frequency scale. Under this rescaling, we are interested in the semiclassical operator

$$P(h) = -h^2\Delta - 1, \quad h \le h_0,$$

and spectral parameter $z \in D(0, Ch)$ for some $C > 0$.

In the semiclassical limit, the classical dynamics associated to this quantum problem is the billiard flow in $\Omega \times \mathbb{S}^1$, that is to say, the free motion outside the obstacles with normal reflection on their boundaries. A relevant dynamical object is the trapped set corresponding to the points $(x, \xi) \in \Omega \times \mathbb{S}^1$ that do not escape to infinity in the backward and forward direction of the flow. In the case of two obstacles, it is a single closed geodesic. As soon as more obstacles are involved, the structure of the trapped set becomes complex and exhibits a fractal structure. This is a consequence of the hyperbolicity of the billiard flow. It is known that the structure of the trapped set plays a crucial role in the spectral gap problem.

A good dynamical object to study this structure is the topological pressure associated to the unstable Jacobian ϕ_u. This dynamical quantity is a strictly decreasing function $s \mapsto P(s)$ which measures the instability of the flow (see Section 2 for definitions and references given there). In dimension 2, Bowen's formula shows that the Hausdorff and upper-box dimensions of the trapped set are $2s_0$, where s_0 is the unique root of the equation $P(s) = 0$. In [Nonnenmacher and Zworski 2009], the existence of a spectral gap for such systems has been proved under the pressure condition

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

Their result holds in any dimension, with a quantitative spectral gap. Our result doesn't need this assumption anymore. In fact, it relies on the weaker pressure condition

$$P(1) < 0.$$

It is known that this condition is always satisfied in the scattering problem we consider since the trapped set is not an attractor [Bowen and Ruelle 1975]. Due to Bowen's formula, this condition can be interpreted as a fractal condition. This is this fractal property that will be crucial in the analysis.

Open hyperbolic systems and spectral gaps. The problem of scattering by obstacles falls into the wider class of spectral problems for open hyperbolic systems; see [Nonnenmacher 2011]. In these open systems, the spectral problems concern the resonances; these are generalized eigenvalues which exhibit some resonant states. Among the problems which widely interest mathematicians and physicists, resonance counting and spectral gaps are on the top of the list. Spectral gaps are known to be important to give resonance expansion (see for instance [Dyatlov and Zworski 2019]) and local energy decay (see for instance [Ikawa 1982; 1988] concerning local energy decay in the exterior of two or more obstacles in $\mathbb{R}^3$). It was conjectured in [Zworski 2017, Conjecture 3] that such systems might exhibit a spectral gap as soon as the trapped set has a fractal structure.

Potential scattering. Scattering by a compactly supported potential falls in the class of open systems. It consists of studying the semiclassical operator $P(h) = -h^2\Delta + V(x)$, where $V \in C_c^\infty(\mathbb{R}^2)$; see Figure 2. In this framework, the spectral gap problem consists of exhibiting bands in the complex plane of the form

$$[a, b] - i \times [0, h\gamma],$$

where $P(h)$ has no resonance for h small enough. In the semiclassical limit, the behavior of $P(h)$ is linked to the classical flow of the system, that is, the Hamiltonian flow generated by $p(x, \xi) = |\xi|^2 + V(x)$. Note that in potential scattering, one has to focus on some energy shell $\{p = E\}$, where $E \in \mathbb{R}$ is independent of h, with $\operatorname{Re} z$ sufficiently close to E. This specification is not necessary in obstacle scattering (implicitly, we have already decided to work with $E = 1$). The properties of the resonant states u_h, which are

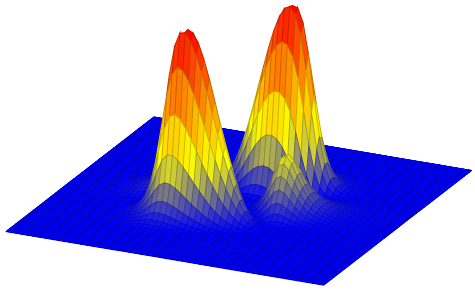

Figure 2. Scattering by a smooth compactly supported potential V.

generalized solutions of the equation $(P(h) - z)u_h = 0$, are linked to the trapped set of the flow at energy E. This trapped set K_E corresponds to all the trajectories which stay bounded for the backward and forward evolution of the flow on the energy shell $\{p = E\}$. When the flow is hyperbolic on the trapped set, this trapped set is known to exhibit a fractal structure.

In fact, a by-product of our method is that we can obtain a spectral gap in potential scattering, under the dynamical assumptions of [Nonnenmacher et al. 2011], recalled in Section 2B:

Theorem B. *Assume that the Hamiltonian flow is hyperbolic on K_E and that K_E is topologically one-dimensional. Then, there exists $\delta > 0$ such that for any $C > 0$, there exists $h_0 > 0$ such that, for $0 < h \leq h_0$, $P(h) = -h^2\Delta + V - E$ has no resonance in*

$$D(0, Ch) \cap \{\operatorname{Im} z \in [-\delta h, 0]\}.$$

It is possible to obtain a spectral gap for the more general quantum Hamiltonian presented in [Nonnenmacher et al. 2011, Section 2.1] for manifolds with Euclidean ends.

Convex cocompact hyperbolic surfaces. Another class of open hyperbolic systems exhibiting a fractal trapped set consists of the convex cocompact hyperbolic surfaces, which can be obtained as the quotient of the hyperbolic plane $\mathbb{H}^2$ by Schottky groups Γ. The spectral problem concerns the Laplacian on these surfaces and its classical counterpart is the geodesic flow on the cosphere bundle, which is known to be hyperbolic due to the negative curvature of these surfaces. In this context, it is common to write the energy variable $\lambda^2 = s(1-s)$ and study

$$(-\Delta - s(1-s))^{-1}.$$

The trapped set is linked to the limit set of Γ and the dimension δ of this limit set influences the spectrum. The Patterson–Sullivan theory (see for instance [Borthwick 2007]) tells that there is a resonance at $s = \delta$ and that the other resonances are located in $\{\operatorname{Re}(s) < \delta\}$. In particular, it gives an essential spectral gap of size $\max\big(0, \frac{1}{2} - \delta\big)$. This is consistent with the pressure condition $P(s) < \frac{1}{2}$ since in that situation, $P(s)$ is simply given by $P(s) = \delta - s$. Results where obtained by Naud [2005], where he improves the gap given by Patterson–Sullivan theory in the case $\delta \leq \frac{1}{2}$. Recent results, initiated by [Dyatlov and Zahl 2016], have improved this gap. In [Bourgain and Dyatlov 2018], the authors show that there exists an essential spectral gap for any convex cocompact hyperbolic surface. In particular, the pressure condition $\delta < \frac{1}{2}$

is no longer a necessary assumption. The new idea in these papers is the use of a fractal uncertainty principle. It will be a crucial tool of our analysis.

Reduction to open hyperbolic quantum maps. An important aspect of our analysis to prove Theorem A relies on previous results of [Nonnenmacher et al. 2014]. Their Theorem 5 (found in Section 6 of that work) reduces the study of the scattering poles to the study of the cancellation of

$$z \mapsto \det(\mathrm{I} - M(z)),$$

where

$$M(z) : L^2(\partial\mathcal{O}) \to L^2(\partial\mathcal{O}) \tag{1-1}$$

is a family of *hyperbolic open quantum maps* (see below Section 2A). The family $z \mapsto M(z)$ depends holomorphically on $z \in D(0, Ch)$ for some $C > 0$ and is sometimes called a *hyperbolic quantum monodromy operator*. The construction of this operator relies on the study of the operators $M_0(z)$ defined as follows: For $1 \le j \le J$, let $H_j(z) : C^\infty(\partial\mathcal{O}_i) \to C^\infty(\mathbb{R}^2 \setminus \mathcal{O}_j)$ be the resolvent of the problem

$$\begin{cases} (-h^2\Delta - 1 - z)(H_j(z)v) = 0, \\ H_j(z)v \text{ is outgoing,} \\ H_j(z)v = v \text{ on } \partial\mathcal{O}_j. \end{cases}$$

Let γ_j be the restriction of a smooth function $u \in C^\infty(\mathbb{R}^2)$ to $C^\infty(\partial\mathcal{O}_j)$ and define $M_0(z)$ by

$$M_0(z) = \begin{cases} 0 & \text{if } i = j, \\ -\gamma_i H_j(z) & \text{otherwise.} \end{cases}$$

Due to results of [Gérard 1988, Appendix II], this matrix is a Fourier integral operator associated with a Lagrangian relation related to the billiard flow. A priori, it excludes neither the glancing rays nor the shadow region. Ikawa's condition ensures that they do not play a role when considering the trapped set and allows the author to neglect the effects of these regions; see Section 6 in [Nonnenmacher et al. 2014]. A consequence of their analysis is that $M(z)$ is associated with a simpler Lagrangian relation $\mathcal{B}$, which is the restriction of the billiard map to a domain excluding the glancing rays. To be more precise, let us introduce

$$\begin{gathered} S^*_{\partial\mathcal{O}_j} = \{(x, \xi) \in T^*\mathbb{R}^2 : x \in \partial\mathcal{O}_j,\ |\xi| = 1\}, \\ B^*\partial\mathcal{O}_j = \{(y, \eta) \in T^*\partial\mathcal{O}_j : |\eta| \le 1\}, \\ \pi_j : S^*_{\partial\mathcal{O}_j} \to B^*\partial\mathcal{O}_j \text{ the orthogonal projection on each fiber.} \end{gathered}$$

$\mathcal{B}$ is then the union of the relations $\mathcal{B}_{ij}$ corresponding to the reflection on two obstacles: for $(\rho_i, \rho_j) \in B^*\partial\mathcal{O}_i \times B^*\partial\mathcal{O}_j$,

$$\begin{gathered} (\rho_i, \rho_j) \in \mathcal{B}_{ij} \quad \Longleftrightarrow \quad \text{there exists } t > 0 \text{ such that } \xi \in \mathbb{S}^1,\ x \in \partial\mathcal{O}_j, \\ \pi_j(x, \xi) = \rho_j, \quad \pi_i(x + t\xi, \xi) = \rho_i, \quad \nu_j(x) \cdot \xi > 0,\ \nu_i(x + t\xi) \cdot \xi < 0. \end{gathered}$$

See Figure 3. It is a standard fact in the study of chaotic billiards (see for instance [Chernov and Markarian 2006]) that the billiard map is hyperbolic due to the strict convexity assumption. Ikawa's condition ensures that the restriction of the dynamical system to the trapped set has a symbolic representation [Morita 1991].

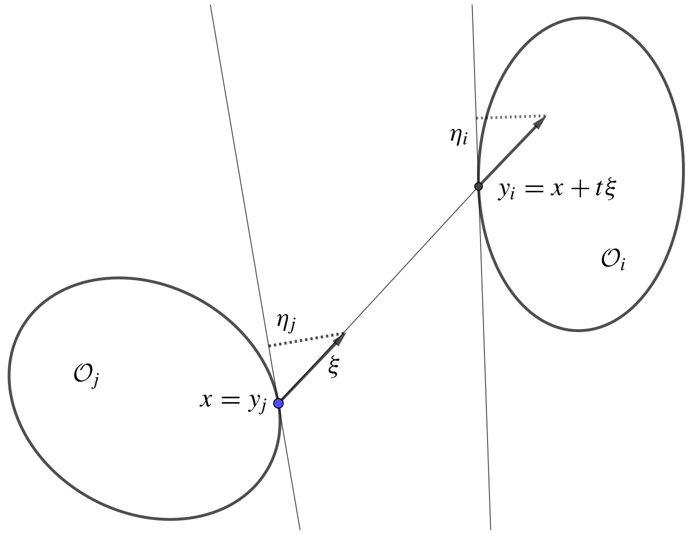

Figure 3. Description of the Lagrangian relation $\mathcal{B}_{ij}$.

Spectral gap for hyperbolic open quantum maps. Using this reduction, Theorem A will be proved once we are able to show that the spectral radius of $M(z)$ is strictly smaller than 1 for $z \in D(0, Ch) \cap \{\operatorname{Im} z \in [-\delta h, 0]\}$ for some $\delta > 0$. This will be a consequence of the following statement, which will be demonstrated in this paper (see Section 2 below for a more precise version).

Theorem C. *Let $(M(z))_z$ be the family introduced in* (1-1), *that is, a hyperbolic quantum monodromy operator associated with the open Lagrangian relation $\mathcal{B}$. Then, there exist $h_0 > 0$, $\gamma > 0$ and $\tau_{\max} > 0$ such that the spectral radius of $M(z)$, $\rho_{\text{spec}}(z)$, satisfies, for all $h \le h_0$ and all $z \in D(0, Ch)$,*

$$\rho_{\text{spec}}(z) \le e^{-\gamma - \tau_{\max} \operatorname{Im} z}.$$

When $z \in \mathbb{R}$, the operator $M(z)$ is microlocally unitary near the trapped set and its L^2 norm is essentially 1. Then, we have the trivial bound

$$\rho_{\text{spec}}(z) \le 1.$$

The bound given by the theorem is a spectral gap since we obtain

$$\rho_{\text{spec}}(z) \le e^{-\gamma} < 1.$$

The dependence of the bound with the parameter z is related to the symbol of the open quantum map $M(z)$.

The link between open quantum maps and the resonances of open quantum systems has also been established in [Nonnenmacher et al. 2011] for the case of potential scattering and this is why we will also obtain a spectral gap in this context. We review this reduction both in obstacle and potential scattering in Section 2 and show how it implies the spectral gap. This correspondence between open quantum maps and open quantum systems leads to a heuristic: to a resonance z for the open quantum systems, it corresponds an eigenvalue $e^{-i\tau z/h}$ of an open quantum map. Here, τ is a return time associated with the

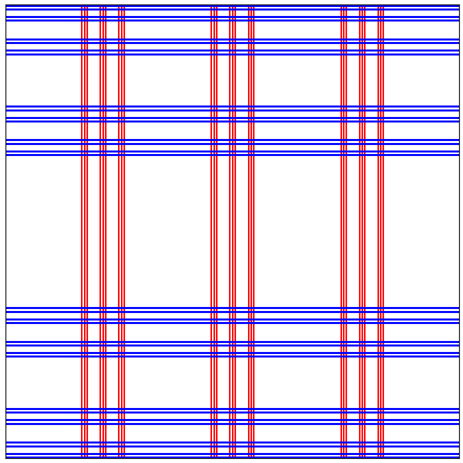

Figure 4. The fractal uncertainty principle asserts that no state can be microlocalized both in frequencies (in blue) and positions (in red) near fractal sets.

classical dynamics of the open system. In particular, the spectral gap for open quantum maps given by the theorem heuristically implies that the resonances of the open systems might satisfy $\operatorname{Im} z < -h\gamma/\tau$.

Resolvent estimates. In this paper, we use the results of [Nonnenmacher et al. 2011; 2014] as a black box. In particular, we apply directly their main theorem establishing a correspondence between scattering resonances and eigenvalues of open quantum maps. This allows us to get information on the locations of the resonances, but cannot transfer resolvent estimates from open quantum maps to the scattering resolvent directly. The main estimate of this paper (see Proposition 4.2) can be used to obtain resolvent estimates for open quantum maps. In an ongoing work, we analyze precisely the proofs in [Nonnenmacher et al. 2011; 2014] so as to explain how to deduce polynomial estimates for the cut-off resolvent both in obstacle and potential scattering. It seems to us that it should be possible to use the gluing method of [Datchev and Vasy 2012] to obtain the same kind of results (spectral gap and polynomial resolvent estimates) with other types of infinite ends, when the trapped set is hyperbolic for the flow and topologically one-dimensional.

On the fractal uncertainty principle. The fractal uncertainty principle is a recent tool in harmonic analysis in one dimension developed by Dyatlov and several collaborators. For a large survey on this topic, we refer the reader to [Dyatlov 2018]. We do not enter into the details in this introduction and give the precise definitions and statements in Section 6. We rather explain here the general idea of this principle in the spirit of our use; see Figure 4. Roughly speaking, it says that no function can be concentrated both in frequencies and positions near a fractal set. Suppose that $X, Y \subset \mathbb{R}$ are fractal sets. To fix the ideas, let's say that X and Y have upper-box dimensions δ_X and δ_Y strictly smaller than 1. For $c > 0$, we write $X(c) = X + [-c, +c]$ and the same for Y. Also denote by $\mathcal{F}_h$ the h-Fourier transform

$$\mathcal{F}_h u(\xi) = \frac{1}{(2\pi h)^{1/2}} \int_{\mathbb{R}} e^{-ix\xi/h} u(x)\, dx.$$

The fractal uncertainty principle then states that there exists $\beta > 0$ depending on X and Y (see Proposition 6.5 for the precise dependence) such that, for h small enough,

$$\|\mathbb{1}_{X(h)}\mathcal{F}_h\mathbb{1}_{Y(h)}\|_{L^2(\mathbb{R})\to L^2(\mathbb{R})} \le h^\beta.$$

Actually, one can change the scales and look for the sets $X(h^{\alpha_X})$ and $Y(h^{\alpha_Y})$, where α_X and α_Y are positive exponents. The result will stay true when these exponents satisfy the saturation condition

$$\alpha_X + \alpha_Y > 1.$$

It will be a key ingredient in the proof of the main theorem of this paper. It has been successfully used to show spectral gaps for convex cocompact hyperbolic surfaces [Dyatlov and Zahl 2016; Bourgain and Dyatlov 2017; Dyatlov and Jin 2018; Dyatlov and Zworski 2020]. A discrete version of the fractal uncertainty principle is also the main ingredient of [Dyatlov and Jin 2017], where the author proved a spectral gap for open quantum maps in a toy model case. Their results concerning the open baker's map on the torus $\mathbb{T}^2$ partly motivates our theorem on open quantum maps.

The fractal uncertainty principle has also given new results in quantum chaos on negatively curved compact surfaces. It was first successfully used for compact hyperbolic surfaces in [Dyatlov and Jin 2017], where the authors proved that semiclassical measures have full support. The hyperbolic case was treated using quantization procedures developed in [Dyatlov and Zahl 2016], which allow one to have a good semiclassical calculus for symbols very irregular in the stable direction, but smooth in the unstable one (or conversely). In [Schwartz 2021], the same ideas lead to a full delocalization of eigenstates for quantum cat maps. The quantization procedures used in these papers rely on the smoothness of the unstable and stable distributions. This smoothness is not possible for general negatively curved surfaces. However, in [Dyatlov et al. 2022], the authors bypassed this obstacle and succeeded in extending these results to the case of negatively curved surfaces. It is mainly from this paper that we borrow techniques and we adapt them in our setting.

A model example. To explain the main ideas of the proof of Theorem C, let us show how it works in an example where the trapped set is the smallest possible, a single point. In this context, we only need a simpler uncertainty principle. We focus on the case $z = 0$ in Theorem C and focus on a single open quantum map.

We consider the hyperbolic map

$$F : (x, \xi) \in \mathbb{R}^2 \mapsto (2^{-1}x, 2\xi) \in \mathbb{R}^2.$$

It has a unique hyperbolic fixed point $\rho_0 = 0$ and the stable (resp. unstable) manifold at 0 is given by $\{\xi = 0\}$ (resp. $\{x = 0\}$). The scaling operator

$$U : v \in L^2(\mathbb{R}) \mapsto \sqrt{2}v(2x)$$

is a quantum map quantizing F. To open it, consider a cut-off function $\chi \in C_c^\infty(\mathbb{R}^2)$ such that $\chi \equiv 1$ in $B\left(0, \frac{1}{2}\right)$ and $\operatorname{supp} \chi \Subset B(0, 1)$ and we consider the open quantum map

$$M = M(h) = \mathrm{Op}_h(\chi)U,$$

where Op_h is in this example (and only in this example) the left quantization

$$\mathrm{Op}_h(\chi)u(x) = \frac{1}{2\pi h}\int_{\mathbb{R}^2} \chi(x,\xi)e^{i(x-y)\xi/h}u(y)\,dy\,d\xi.$$

One easily checks that Egorov's property for U is true without remainder term:

$$U^*\,\mathrm{Op}_h(\chi)U = \mathrm{Op}_h(\chi\circ F), \quad U\,\mathrm{Op}_h(\chi)U^* = \mathrm{Op}_h(\chi\circ F^{-1}).$$

To show a spectral gap for M, we study M^n with

$$n = n(h) \sim -\frac{3}{4}\frac{\log h}{\log 2}.$$

This time is longer than the Ehrenfest time $-\log h/\log 2$. We write

$$M^n = U^n\,\mathrm{Op}_h(\chi\circ F^n)\cdots\mathrm{Op}_h(\chi\circ F^1).$$

The formula $[\mathrm{Op}_h(a), \mathrm{Op}_h(b)] = O(h^{1-2\delta})$ is valid for a, b symbols in S_δ (we recall the definitions of symbol classes in Section 3) and $\delta < \frac{1}{2}$. The problem here is that, for $1 \le k \le n$, $\chi\circ F^k$ are uniformly in $S_{3/4}$; this is not a good symbol class. To bypass this difficulty, we observe that the symbols $\chi\circ F^k$ are uniformly in $S_{3/8}$ for $k \in \{-n/2, \dots, n/2\}$. As a consequence, for $j \in \{1, \dots, n\}$, we write

$$\begin{aligned}
[\mathrm{Op}_h(\chi\circ F^n), \mathrm{Op}_h(\chi\circ F^j)] &= U^{-n/2}[\mathrm{Op}_h(\chi\circ F^{n/2}), \mathrm{Op}_h(\chi\circ F^{j-n/2})]U^{n/2}\\
&= U^{-n/2}O(h^{1/4})U^{n/2}\\
&= O(h^{1/4}),
\end{aligned}$$

where the constants in O are uniform in j and depend only on χ. Applying this formula recursively to move the term $\mathrm{Op}_h(\chi\circ F^n)$ to the right, we get

$$M^n = U^n\,\mathrm{Op}_h(\chi\circ F^{n-1})\cdots\mathrm{Op}_h(\chi\circ F^1)\,\mathrm{Op}_h(\chi\circ F^n) + O(h^{1/4}\log h).$$

Similarly, we can write

$$M^{n+1} = \mathrm{Op}_h(\chi\circ F^{-n})\,\mathrm{Op}_h(\chi)\cdots\mathrm{Op}_h(\chi\circ F^{-n+1})U^{n+1} + O(h^{1/4}\log h).$$

Hence, we have

$$M^{2n+1} = A\,\mathrm{Op}_h(\chi\circ F^n)\,\mathrm{Op}_h(\chi\circ F^{-n})B + O(h^{1/4}\log h),$$

with

$$\begin{aligned}
A = A(h) &= U^n\,\mathrm{Op}_h(\chi\circ F^{n-1})\cdots\mathrm{Op}_h(\chi\circ F^1) = O(1),\\
B = B(h) &= \mathrm{Op}_h(\chi)\cdots\mathrm{Op}_h(\chi\circ F^{-n+1})U^{n+1} = O(1).
\end{aligned}$$

We have the following properties on the supports:

$$\mathrm{supp}\,\chi\circ F^n \subset \{|\xi| \le 2^{-n}\}, \quad \mathrm{supp}\,\chi\circ F^n \subset \{|x| \le 2^{-n}\}.$$

Assuming $n(h) \geq -\frac{3}{4}(\log h/\log 2)$, we observe that

$$\mathrm{Op}_h(\chi \circ F^n) = \mathrm{Op}_h(\chi \circ F^n)\mathbb{1}_{[-h^{3/4},h^{3/4}]}(hD_x),$$
$$\mathrm{Op}_h(\chi \circ F^{-n}) = \mathbb{1}_{[h^{-3/4},h^{3/4}]}(x)\,\mathrm{Op}_h(\chi \circ F^{-n}).$$

Finally, we have

$$M^{2n+1} = A\,\mathrm{Op}_h(\chi \circ F^n)\mathbb{1}_{[-h^{3/4},h^{3/4}]}(hD_x)\mathbb{1}_{[h^{-3/4},h^{3/4}]}(x)\,\mathrm{Op}_h(\chi \circ F^{-n})B + O(h^{1/4}\log h).$$

This is where we need an uncertainty principle:

$$\begin{aligned}\|\mathbb{1}_{[-h^{3/4},h^{3/4}]}(hD_x)\mathbb{1}_{[h^{-3/4},h^{3/4}]}(x)\|_{L^2\to L^2} &= \|\mathbb{1}_{[-h^{3/4},h^{3/4}]}\mathcal{F}_h\mathbb{1}_{[-h^{3/4},h^{3/4}]}\|_{L^2\to L^2}\\ &\leq \|\mathbb{1}_{[-h^{3/4},h^{3/4}]}\|_{L^\infty\to L^2} \times \|\mathcal{F}_h\|_{L^1\to L^\infty} \times \|\mathbb{1}_{[-h^{3/4},h^{3/4}]}\|_{L^2\to L^1}\\ &\leq Ch^{3/8} \times h^{-1/2} \times h^{3/8} = Ch^{1/4}.\end{aligned}$$

Here, the bound can be understood as a volume estimate; the box in phase space of size $h^{3/4}$ is smaller than a "quantum box". Gathering all the computations together, we see that

$$\|M^{2n+1}\|_{L^2\to L^2} = O(h^{1/4}\log h).$$

Elevating this to the power $1/(2n+1)$, we see that, for every $\varepsilon > 0$, we can find h_ε such that, for $h \leq h_\varepsilon$,

$$\rho(M) \leq (1+\varepsilon)2^{-1/6}.$$

Remark. What matters in this example is the strategy we use, and not particularly the bound, which is in fact not optimal.

Sketch of proof. The strategy presented in this simple model case is the guideline, but its direct application will encounter major pitfalls that we'll have to bypass.

- Since the trapped set is a more complex fractal set, we'll need the general fractal uncertainty principle developed by Dyatlov and his collaborators.
- Even in small coordinate charts, the trapped set cannot be written has a product of fractal sets in the unstable and stable directions. To tackle this difficulty, we build adapted coordinate charts (see Section 3E) in which we straighten the unstable manifolds. The existence of such coordinate charts is made possible by Theorem 5, in which we prove that the unstable (and stable) distribution can be extended in a neighborhood of the trapped set to a $C^{1+\beta}$ vector field.
- In the model case, there is only one point and hence one unstable Jacobian to consider which gives the Lyapouvov exponent of the map $\log J_u^1(0) = \log 2$. Generally, the growth rate of the unstable Jacobian differs from one point to another (see Section 4C) and the choice of the integer $n(h)$ is not as simple. In fact, we prefer to break the symmetry $2n(h) = n(h) + n(h)$ and split $2n(h)$ into a small logarithmic time $N_0(h)$ and a long logarithmic time $N_1(h)$ (see Section 4A). The first one is supposed to be smaller than the Ehrenfest time and allows us to use semiclassical calculus to handle M^{N_0}. As a matter of fact, the major technical difficulties concern the study of M^{N_1}.

- The study of M^{N_1} requires fine microlocal techniques. The trick used in the model case to have the commutator estimate is not possible and we have to use propagation results up to twice the Ehrenfest time. This is what we do in Section 4D but this study has to be made locally and we need to split M^{N_1} into a sum of many terms $U_{\boldsymbol{q}}$.

- We could use the fractal uncertainty principle to get the decay for single terms $M^{N_0}U_{\boldsymbol{q}}$. However, a simple triangle inequality to handle their sum will not give a decay for $M^{N_0+N_1}$ since the number of terms in the sum grows like a negative power of h. To bypass this problem, we need a more careful analysis and we gather them into clouds (see Section 4G). These clouds are supposed to interact with a few other ones, so that a Cotlar–Stein-type estimate reduces the study of the norm of the sum to the norm of each cloud. The elements of a single cloud are supposed to be close to each other, so that the fractal uncertainty principle can be applied to all of them in the same time and gives the required decay for a single cloud.

Our strategy follows the main lines of the proof of [Dyatlov et al. 2022]. In particular, their strategy allows us to apply the fractal uncertainty principle of [Bourgain and Dyatlov 2018] in a case where the unstable foliation is not smooth (and in fact, a priori defined only in a fractal set). Their strategy relies on the existence of adapted charts based on C^{2^-} regularity of the unstable foliations in negatively curved surfaces. It is based on results of [Katok and Hasselblatt 1995] for Anosov flows. We needed to prove the existence of such adapted charts in this different context. To do so, we prove that the unstable lamination can be extended into a $C^{1+\beta}$ foliation (see Section 3E). Another aspect which changes from [Dyatlov et al. 2022] is the proof of porosity. In their study, the porous sets arise as iterations of artifical "holes", and they had to control the evolution of such holes. In our context, this study is easier since we already know that the trapped set has a fractal structure, characterized by its Hausdorff dimension. In this paper, we will rather use the upper-box dimension (but these two dimensions are equal in this context).

Restrictions. The main restriction of our theorem is that it only applies to quantum maps with two-dimensional phase space. In terms of open systems, it only concerns problems with physical space of dimension 2. Several points explain this restriction:

- The fractal uncertainty principle works in dimension 1. In higher dimensions, the result is currently not well understood and the only known cases require strong assumptions on the fractal sets; see [Dyatlov 2018, Section 6].

- Our proof strongly relies on the regularity of the stable and unstable laminations.

- The growth of the unstable Jacobian controls the contraction (resp. expansion) rate in the unique stable (resp. unstable) direction.

Plan of the paper. The paper is organized as follows:

- In Section 2, we present the main theorem of this paper and show how it gives a spectral gap in some open quantum systems.

- In Section 3, we give some background material in semiclassical analysis (pseudodifferential operators and Fourier integral operators). We also recall some standard facts about hyperbolic dynamical systems and give further results. In particular, in Theorem 5, we show that the unstable and stable distribution have $C^{1+\beta}$ regularity.

- The proof of Theorem 1 starts in Section 4, where we introduce the main ingredients needed for the proof and give several technical results.

- In Section 5, we use fine microlocal methods to microlocalize the operators we work with in small regions where the dynamic is well understood and we reduce the proof of Theorem 1 to a fractal uncertainty principle with the techniques of [Dyatlov et al. 2022].

- In Section 6, we conclude the proof of this theorem by applying the fractal uncertainty principle of [Bourgain and Dyatlov 2018], and more precisely, the version stated in [Dyatlov et al. 2022].

2. Main theorem and applications

2A. ***Hyperbolic open quantum maps.*** We introduce the main tools needed to state the main theorem of this paper. The following long definition is based on the definitions in the works of Nonnenmacher, Sjöstrand and Zworski [Nonnenmacher et al. 2011; 2014] specialized to the two-dimensional phase space. Consider open intervals $Y_1, \dots, Y_J$ of $\mathbb{R}$ and set

$$Y = \bigsqcup_{j=1}^{J} Y_j \subset \bigsqcup_{j=1}^{J} \mathbb{R}$$

and consider

$$U = \bigsqcup_{j=1}^{J} U_j \subset \bigsqcup_{j=1}^{J} T^*\mathbb{R}^d, \quad U_j \Subset T^*Y_j.$$

The Hilbert space $L^2(Y)$ is the orthogonal sum $\bigoplus_{i=1}^J L^2(Y_i)$.

Then, we introduce a smooth Lagrangian relation $F \subset U \times U$. It is a disjoint union of symplectomorphisms. For $j = 1, \dots, J$, consider open disjoint subsets $\widetilde{D}_{ij} \Subset U_j$, $1 \le i \le J$, and similarly, for $i = 1, \dots, J$, consider open disjoint subsets $\tilde{A}_{ij} \Subset U_i$, $1 \le j \le J$. We consider a family of smooth symplectomorphisms

$$F_{ij} : \widetilde{D}_{ij} \to F_{ij}(\widetilde{D}_{ij}) = \tilde{A}_{ij} \tag{2-1}$$

and define the relation F as the disjoint union of the relation F_{ij}, namely,

$$(\rho', \rho) \in F \quad \Longleftrightarrow \quad \text{there exist } 1 \le i, j \le J \text{ such that } \rho' = F_{ij}(\rho).$$

In particular, F and F^{-1} are single-valued. We will identify F with a smooth map and write by abuse of notation $\rho' = F(\rho)$ or $\rho = F^{-1}(\rho')$ instead of $(\rho', \rho) \in F$.

We let

$$\pi_L(F) = \tilde{A} = \bigsqcup_{i=1}^{J} \bigcup_{j=1}^{J} \tilde{A}_{ij}, \quad \pi_R(F) = \widetilde{D} = \bigsqcup_{j=1}^{J} \bigcup_{i=1}^{J} \widetilde{D}_{ij}.$$

We define the outgoing (resp. incoming) tail by $\mathcal{T}_+ := \{\rho \in U : F^{-n}(\rho) \in U \text{ for all } n \in \mathbb{N}\}$ (resp. $\mathcal{T}_- := \{\rho \in U : F^n(\rho) \in U \text{ for all } n \in \mathbb{N}\}$). We assume that they are closed subsets of U and that the *trapped set*

$$\mathcal{T} = \mathcal{T}_+ \cap \mathcal{T}_- \tag{2-2}$$

is compact. We denote by $f : \mathcal{T} \to \mathcal{T}$ the restriction of F to $\mathcal{T}$. For $i, j \in \{1, \dots, J\}$, we write $\mathcal{T}_i = \mathcal{T} \cap U_i$,

$$D_{ij} = \{\rho \in \mathcal{T}_j : f(\rho) \in \mathcal{T}_i\} \subset \widetilde{D}_{ij},$$
$$A_{ij} = \{\rho \in \mathcal{T}_i : f^{-1}(\rho) \in \mathcal{T}_j\} \subset \tilde{A}_{ij}.$$

Remark. F is an open canonical transformation since F (resp. F^{-1}) is defined only in $\widetilde{D}$ (resp. $\tilde{A}$). The sets $U \setminus \widetilde{D}$ (resp. $U \setminus \tilde{A}$) can be seen as holes in which a point ρ can fall in the future (resp. in the past).

We then make the following hyperbolic assumption:

$$\mathcal{T} \text{ is a hyperbolic set for } F. \tag{Hyp}$$

Namely, for every $\rho \in \mathcal{T}$, we assume that there exist stable and unstable tangent spaces $E^s(\rho)$ and $E^u(\rho)$ such that:

- $\dim E^s(\rho) = \dim E^u(\rho) = 1$.
- $T_\rho U = E^s(\rho) \oplus E^u(\rho)$.
- There exist $\lambda > 0$, $C > 0$ such that, for every $v \in E^\star(\rho)$ ($\star$ stands for u or s) and any $n \in \mathbb{N}$,

$$v \in E^s(\rho) \quad \implies \quad \|d_\rho F^n(v)\| \leq Ce^{-n\lambda}\|v\|, \tag{2-3}$$
$$v \in E^u(\rho) \quad \implies \quad \|d_\rho F^{-n}(v_\star)\| \leq Ce^{-n\lambda}\|v\|, \tag{2-4}$$

  where $\|\cdot\|$ is a fixed Riemannian metric on U.

The decomposition of $T_\rho U$ into stable and unstable spaces is assumed to be continuous.

Remark. • The definition is valid for any Riemannian metric and we can of course suppose that is it the standard Euclidean metric on $\mathbb{R}^2$.

• It is a standard fact (see [Mather 1968]) that there exists a smooth Riemannian metric on U, which is said to be adapted to the dynamics, such that (2-3) and (2-4) hold with $C = 1$.

• It is known that the map $\rho \mapsto E_{u/s}(\rho)$ is in fact β-Hölder for some $\beta > 0$ [Katok and Hasselblatt 1995]. We will show further an improved regularity. This will be an essential property for the proof of the main theorem.

The last assumption we'll make on $\mathcal{T}$ is a fractal assumption. To state it, we introduce the map $\phi_u : \rho \in \mathcal{T} \mapsto -\log \|d_\rho F|_{E_u(\rho)}\|$ associated with the bijection f. We suppose that

$$-\gamma_{\text{cl}} := -P(-\log \|d_\rho F|_{E_u(\rho)}\|, f) > 0. \tag{Fractal}$$

Here, in terms of thermodynamics formalism, P denotes the topological pressure of the map ϕ_u. The norm $\|\cdot\|$ is associated with any Riemannian metric on U. For instance, a possible formula for the definition of the pressure is

$$P(\phi) = \lim_{\varepsilon \to 0} \limsup_{n \to +\infty} \frac{1}{n} \log \sup_E \sum_{\rho \in E} \exp^{\sum_{k=0}^{n-1} \phi(f^k \rho)},$$

where the supremum ranges over all the (n, ε)-separated subsets $E \subset \mathcal{T}$ (E is said to be (n, ε)-separated if, for every $\rho, \rho' \in E$, there exists $k \in \{0, \dots, n-1\}$ such that $d(f^k(\rho), f^k(\rho')) > \varepsilon$).

Remark. • γ_{cl} is the classical decay rate of the dynamical system. It has the following physical interpretation: Fix a point $\rho_0 \in \mathcal{T}$ and consider the set $B_m(\rho_0, \varepsilon)$ of points $\rho \in U$ such that $|F^k(\rho) - F^k(\rho_0)| < \varepsilon$ for $0 \le k \le m-1$. Then, its Lebesgue measure if of order $e^{-m\gamma_{\mathrm{cl}}}$.

• In Section A4, we recall arguments showing that $\mathcal{T}$ is indeed "fractal". More precisely, the trace of $\mathcal{T}$ along the unstable and stable manifolds (see Lemma 3.11 for the definitions of these manifolds) have upper-box dimension strictly smaller than 1. In fact, Bowen's formula (see for instance [Barreira 2008]) gives that this upper-box dimension corresponds to the Hausdorff dimension d_H and it is the unique solution of the equation

$$P(s\phi_u, f) = 0, \quad s \in \mathbb{R}.$$

The Hausdorff dimension of the trapped set is then $2d_H$.

• This condition has to be compared with the pressure condition $P\big(\frac{1}{2}\phi_u\big) < 0$ in [Nonnenmacher and Zworski 2009], which ensured a spectral gap for chaotic systems. This condition required that $\mathcal{T}$ was sufficiently "thin", i.e., with Hausdorff dimension strictly smaller than 1. Our condition allows to go up to the limit $\dim_H \mathcal{T} = 2^-$.

We then associate to F *hyperbolic open quantum maps*, which are its quantum counterpart.

Definition 2.1. Fix $\delta \in \big[0, \frac{1}{2}\big[$. We say that $T = T(h)$ is a semiclassical Fourier integral operator associated with F, and we let $T = T(h) \in I_\delta(Y \times Y, F')$ if, for each couple $(i, j) \in \{1, \dots, J\}^2$, there exists a semiclassical Fourier integral operator $T_{ij} = T_{ij}(h) \in I_\delta(Y_j \times Y_i, F'_{ij})$ associated with F_{ij} in the sense of Definition 3.9, such that

$$T = (T_{ij})_{1 \le i, j \le J} : \bigoplus_{i=1}^{J} L^2(Y_i) \to \bigoplus_{i=1}^{J} L^2(Y_i).$$

In particular $\mathrm{WF}_h(T) \subset \tilde{A} \times \widetilde{D}$. We define $I_{0^+}(Y \times Y, F') = \bigcap_{\delta > 0} I_\delta(Y \times Y, F')$.

We will say that T is *microlocally unitary near* $\mathcal{T}$ if the two following conditions hold:

- $\|TT^*\| \le 1 + O(h^\varepsilon)$ for some $\varepsilon > 0$.
- There exists a neighborhood $\Omega \subset U$ of $\mathcal{T}$ such that, for every $u = (u_1, \dots, u_J) \in \bigoplus_{j=1}^{J} L^2(Y_j)$,

$$\text{for all } j \in \{1, \dots, J\}, \ \ \mathrm{WF}_h(u_j) \subset \Omega \cap U_j \implies TT^* u = u + O(h^\infty)\|u\|_{L^2}, \ \ T^*Tu = u + O(h^\infty)\|u\|_{L^2}.$$

Let us now briefly see what the second condition implies for the components of T^*T. First focus on the off-diagonal entries

$$(T^*T)_{ij} = \sum_{k=1}^{J} (T^*)_{ik} T_{kj} = \sum_{k=1}^{J} (T_{ki})^* T_{kj}.$$

If $k \in \{1, \dots, J\}$ and $i \neq j$, $(T_{ki})^* T_{kj} = O(h^\infty)$ since

$$\mathrm{WF}_h(T_{ki}^*) \subset \widetilde{D}_{ki} \times \tilde{A}_{ki}, \quad \mathrm{WF}_h(T_{kj}) \subset \tilde{A}_{kj} \times \widetilde{D}_{kj} \quad \text{and} \quad \tilde{A}_{kj} \cap \tilde{A}_{ki} = \varnothing.$$

As a consequence, the off-diagonal terms are always $O(h^\infty)$. For the diagonal entries,

$$(T^*T)_{ii} = \sum_{k=1}^{J} (T_{ki})^* T_{ki}.$$

Each term of this sum is a pseudodifferential operator with wavefront set

$$\mathrm{WF}_h(T_{ki}^* T_{ki}) \subset \widetilde{D}_{ki}.$$

Since the $\widetilde{D}_{ki}$ are pairwise disjoint, $T^*T = \mathrm{Id}_{L^2(Y)} + O(h^\infty)$ microlocally near $\mathcal{T}$ if and only if, for all k, i, $T_{ki}^* T_{ki} = \mathrm{Id}_{L^2(Y_i)} + O(h^\infty)$ microlocally near D_{ki}. The same computations apply to TT^*. As a consequence, T is microlocally unitary near $\mathcal{T}$ if and only if, for all (k, i), T_{ki} is a Fourier integral operator associated with F_{ki}, microlocally unitary near $D_{ki} \times A_{ki}$ (see the paragraph below Definition 3.9).

Notation. An element of $S_\delta^{\mathrm{comp}}(U)$ is a J-tuple $\alpha = (\alpha_1, \dots, \alpha_J)$, where each α_j is an element of $S^\delta_{\mathrm{comp}}(\mathbb{R}^2)$ such that $\operatorname{ess\,supp} \alpha_j \subset U_j$ (this notation is recalled in the next section).

We fix a smooth function $\Psi_Y = (\Psi_1, \dots, \Psi_J)$ such that, for $1 \leq j \leq J$, $\Psi_j \in C_c^\infty(Y_j, [0,1])$ satisfies $\Psi_j = 1$ on $\pi(U_j)$ (recall that $U_j \Subset T^*Y_j$).

For $\alpha \in S_\delta^{\mathrm{comp}}(U)$, we also denote by $\mathrm{Op}_h(\alpha)$ the diagonal operator-valued matrix

$$\mathrm{Op}_h(\alpha) = \mathrm{Diag}(\Psi_1 \mathrm{Op}_h(\alpha_1) \Psi_1, \dots, \Psi_J \mathrm{Op}_h(\alpha_J) \Psi_J) : \bigoplus_{j=1}^{J} L^2(Y_j) \to \bigoplus_{j=1}^{J} L^2(Y_j).$$

Note that as operators on $L^2(\mathbb{R})$, $\mathrm{Op}_h(\alpha_j)$ and $\Psi_j \mathrm{Op}_h(\alpha_j) \Psi_j$ are equal modulo $O(h^\infty)$.

We can now state the main theorem of this paper, namely a spectral gap for hyperbolic open quantum maps. We denote by $\rho_{\mathrm{spec}}(A)$ the spectral radius of a bounded operator $A : L^2(Y) \to L^2(Y)$.

Theorem 1. *Suppose that the above assumptions on F,* (Hyp) *and* (Fractal) *are satisfied. Then, there exists $\gamma > 0$ such that the following holds*:

Let $T = T(h) \in I_{0^+}(Y \times Y, F')$ be a semiclassical Fourier integral operator associated with F in the sense of Definition 2.1 and $\alpha \in S_\delta^{\mathrm{comp}}(U)$. Assume that T is microlocally unitary in a neighborhood of $\mathcal{T}$. Then, there exists $h_0 > 0$ such that,

$$\textit{for all } 0 < h \leq h_0, \quad \rho_{\mathrm{spec}}(T(h) \mathrm{Op}_h(\alpha)) \leq e^{-\gamma} \|\alpha\|_\infty,$$

where h_0 depends on (U, F), T and seminorms of α in S_δ.

For applications, we will need the following corollary (it is in fact rather a corollary of the method used to prove Theorem 1):

Corollary 1. *With the same notations and assumptions as in Theorem 1, if $R(h)$ is a family of bounded operators on $L^2(Y)$ satisfying $\|R(h)\| = O(h^\eta)$ for some $\eta > 0$, then the there exists γ' depending only on γ and η such that, for $0 < h \leq h_0$,*

$$\rho_{\text{spec}}(T(h)\operatorname{Op}_h(\alpha) + R(h)) \leq e^{-\gamma'}\|\alpha\|_\infty.$$

Remark. • If the value h_0 depends on T and α, this is not the case of γ which depends on (U, F).

• This is a spectral gap; it has to be compared with the easy bound we could have

$$\rho_{\text{spec}}(T\operatorname{Op}_h(\alpha)) \leq \|\alpha\|_\infty + o(1).$$

In particular, if $\alpha \equiv 1$ in a neighborhood of $\mathcal{T}$ and $|\alpha| \leq 1$ everywhere, $\rho_{\text{spec}}(T(h)) \leq e^{-\gamma} < 1$.

• $T\operatorname{Op}_h(\alpha)$ is the way we've chosen to write our Fourier integral operator with "gain" (or absorption depending on the modulus of α) factor α. $T\operatorname{Op}_h(\alpha)$ transforms a wave packet u_0 microlocalized near ρ_0 lying in a small neighborhood of $\mathcal{T}$ into a wave packet microlocalized near $F(\rho_0)$, with norm essentially changed by a factor $|\alpha(\rho_0)|$.

• The proof will actually show that if η is strictly bigger than some threshold, then $\gamma' = \gamma$.

Notation. Throughout the paper, the meaning of the constants C can change from line to line but these constants will only depend on our dynamical system (U, F). If there is another dependence, it will be specified.

2B. ***Applications of the theorem.*** This theorem has applications in the study of open quantum systems. We refer the reader to [Nonnenmacher 2011] for a survey on this topic. The spectral gap given by Theorem 1 will actually give a spectral gap for the resonances of semiclassical operators $P(h)$ in $\mathbb{R}^2$, or for the resonances of the Dirichlet Laplacian in the exterior of strictly convex obstacles satisfying the Ikawa noneclipse condition. We refer the reader to the review [Zworski 2017] for more background on scattering resonances or to the book [Dyatlov and Zworski 2019]. The results we will obtain from Theorem 1 give a positive answer (in dimension 2) to Conjecture 3 in [Zworski 2017], under a fractal assumption.

Scattering by strictly convex obstacles in the plane. As already explained in the Introduction the main problem motivating Theorem 1 is the problem of scattering by obstacles in the plane $\mathbb{R}^2$. It leads to:

Theorem 2. *Assume that $\mathcal{O} = \bigcup_{i=1}^J \mathcal{O}_j$, where $\mathcal{O}_j$ are open, strictly convex connected obstacles in $\mathbb{R}^2$ having smooth boundary and satisfying the* ***Ikawa condition****: for $i \neq j \neq k$, $\overline{\mathcal{O}}_i$ does not intersect the convex hull of $\overline{\mathcal{O}}_j \cup \overline{\mathcal{O}}_k$. Let*

$$\Omega = \mathbb{R}^2 \setminus \overline{\mathcal{O}}.$$

There exist $\gamma > 0$ and $\lambda_0 > 1$ such that the Dirichlet Laplacian $-\Delta$ on $L^2(\Omega)$ has no scattering resonance in the region

$$[\lambda_0, +\infty[+ i[-\gamma, 0].$$

Let us give the arguments to see why Theorem 1 implies this theorem. After a semiclassical reparametrization, it is enough to show that there exist $\delta > 0$ and $h_0 > 0$ such that $P(h) := -h^2\Delta - 1$ has no resonance in $D(0, Ch) \cap \{\operatorname{Im} z \in [-\delta h, 0]\}$ for any $h \leq h_0$. As already explained, the implication relies on [Nonnenmacher et al. 2014, Theorem 5, Section 6]. There they prove the existence of a family of

$$(\mathcal{M}(z))_{z \in D(0,Ch)} = (\mathcal{M}(z, h)) \tag{2-5}$$

such that:

- $\mathcal{M}(z) = \Pi_h M(z) \Pi_h + O(h^L)$, where Π_h is a finite-rank projector, of rank comparable to h^{-1}, $L > 0$ is a fixed constant (which can in fact be chosen as big as we want) and $M(z)$ is described below and satisfies $\Pi_h M(z) \Pi_h = M(z) + O(h^L)$.
- $M(0)$ is an open quantum map associated with a Lagrangian relation $\mathcal{B}$ presented in the Introduction, which is microlocally unitary near $\mathcal{T}$. $\mathcal{B}$ and $M(0)$ play the roles of F and T in Theorem 1 and satisfy its assumptions.
- $M(z) = M(0) \operatorname{Op}_h(e^{iz\tau/h}) + O(h^{1-\varepsilon})$ uniformly in $D(0, Ch)$, where $\varepsilon > 0$ can be chosen arbitrarily close to zero and $\tau \in C_c^\infty(U)$ is a smooth function (which has to be seen as a return time).
- The resonances of $P(h)$ in $D(0, Ch)$ are the roots, with multiplicities, of the equation

$$\det(I - \mathcal{M}(z)) = 0.$$

Hence, to prove the theorem, it is enough to show that the spectral radius of $\mathcal{M}(z)$ is strictly smaller than 1 for $z \in D(0, Ch) \cap \{\operatorname{Im} z \in [-\delta h, 0]\}$ for some $\delta > 0$ and for h small enough. To see that, we write

$$\mathcal{M}(z) = M(0) \operatorname{Op}_h(e^{iz\tau/h}) + R(h),$$

with $R(h) = O(h^\eta)$ for any $\eta < \min(1, L)$. We apply Theorem 1 and find some γ' such that

$$\rho_{\text{spec}}(\mathcal{M}(z)) \leq e^{-\gamma'} \|e^{iz\tau/h}\|_\infty \leq e^{-\gamma'} e^{\delta \tau_{\max}}, \quad z \in D(0, Ch) \cap \{\operatorname{Im} z \in [-\delta h, 0]\},$$

where $\tau_{\max} = \|\tau\|_\infty$. This ensures a spectral gap of size

$$\delta < \frac{\gamma'}{\tau_{\max}}.$$

Schrödinger operators. Actually, the obstacles, seen as infinite potential barriers, can be smoothened with a potential $V \in C_c^\infty(\mathbb{R}^2)$ and we can consider the Schrödinger operators $P_0(h) = -h^2\Delta + V(x)$.

Unlike the obstacle problem, a simple rescaling does not allow to pass from energy 1 to any energy E and the behavior of the classical flow can drastically change from an energy shell to another. To study the problem at energy $E > 0$, independent of h, we rather consider

$$P(h) = P_0(h) - E.$$

The resolvent $(P(h) - z)^{-1}$ continues meromorphically from $\operatorname{Im} z > 0$ to $D(0, Ch)$ (as previously in the sense that $\chi(P(h) - z)^1 \chi$ extends meromorphically with $\chi \in C_c^\infty(\mathbb{R}^2)$) and we are interested in the existence of a spectral gap.

The classical Hamiltonian flow associated with $P(h)$ is the Hamiltonian flow Φ^t generated by $p_0(x,\xi) = |\xi|^2 + V(x)$ on the energy shell $p_0^{-1}(E)$. The trapped set is defined as above by

$$K_E := \{(x,\xi) \in T^*\mathbb{R}^2 : p_0(x,\xi) = E,\ \Phi^t(x,\xi) \text{ stays bounded as } t \to \pm\infty\}.$$

We assume that the flow is hyperbolic on K_E and that the trapped set is topologically one-dimensional. Equivalently, we assume that transversely to the flow, K_E is zero-dimensional. Under these assumptions, the authors proved (see Theorem 1 in [Nonnenmacher et al. 2011]) the existence of a family of monodromy operators associated with a Lagrangian relation F_E which is a Poincaré map of the flow on different Poincaré sections $\Sigma_1, \dots, \Sigma_J \subset p_0^{-1}(E)$. The assumption on the dimension of K_E implies that the assumption (Fractal) is satisfied since K_E cannot be an attractor [Bowen and Ruelle 1975]. Hence, Theorem 1 applies and we can prove as done in the case of obstacles:

Theorem 3. *Under the above assumptions, there exists $\delta > 0$ such that $P(h)$ has no resonances in*

$$D(0, Ch) \cap \{\operatorname{Im} z \in [-\delta h, 0]\}.$$

3. Preliminaries

3A. ***Pseudodifferential operators and Weyl quantization.*** We recall some basic notions and properties of the Weyl quantization on $\mathbb{R}^n$. We refer the reader to [Zworski 2012] for the proofs of the statements and further considerations on semiclassical analysis and quantizations. We start by defining classes of h-dependent symbols.

Definition 3.1. Let $0 \leq \delta \leq \frac{1}{2}$. We say that an h-dependent family $a := (a(\cdot\,; h))_{0<h\leqslant 1}$ is in the class $S_\delta(T^*\mathbb{R}^n)$ (or simply S_δ if there is no ambiguity) if, for every $\alpha \in \mathbb{N}^{2n}$, there exists $C_\alpha > 0$ such that,

$$\text{for all } 0 < h \leq 1, \qquad \sup_{(x,\xi)\in\mathbb{R}^n} |\partial^\alpha a(x,\xi;h)| \leq C_\alpha h^{-\delta|\alpha|}.$$

In this paper, we will mostly be concerned with $\delta < \frac{1}{2}$. We will also use the notation $S_{0^+} = \bigcap_{\delta>0} S_\delta$. We write $a = O(h^N)_{S_\delta}$ to mean that, for every $\alpha \in \mathbb{N}^{2n}$, there exists $C_{\alpha,N}$ such that,

$$\text{for all } 0 < h \leq 1, \qquad \sup_{(x,\xi)\in\mathbb{R}^n} |\partial^\alpha a(x,\xi;h)| \leq C_{\alpha,N} h^{-\delta|\alpha|} h^N.$$

If $a = O(h^N)_{S_\delta}$ for all $N \in N$, we'll write $a = O(h^\infty)_{S_\delta}$. A priori, the constants $C_{\alpha,N}$ depend on the symbol a. However, in this paper, we will often make them depend on different parameters but not directly on a. This will be specified when needed.

For a given symbol $a \in S_\delta(T^*\mathbb{R}^n)$, we say that a has a compact essential support if there exists a compact set K such that,

$$\text{for all } \chi \in C_c^\infty(\Omega), \quad \operatorname{supp}\chi \cap K = \varnothing \quad \Longrightarrow \quad \chi a = O(h^\infty)_{\mathcal{S}(T^*\mathbb{R}^n)}$$

(here $\mathcal{S}$ stands for the Schwartz space). We let $\operatorname{ess\,supp} a \subset K$ and say that a belongs to the class $S_\delta^{\mathrm{comp}}(T^*\mathbb{R}^n)$. The essential support of a is then the intersection of all such compact K's. In particular, the class S_δ^{comp} contains all the symbols supported in an h-independent compact set and these symbols correspond, modulo $O(h^\infty)_{\mathcal{S}(T^*\mathbb{R})}$, to all symbols of S_δ^{comp}. For this reason, we will adopt the notation $a \in S_\delta^{\mathrm{comp}}(\Omega) \iff \operatorname{ess\,supp} a \Subset \Omega$.

For a symbol $a \in S_\delta(T^*\mathbb{R}^n)$, we'll quantize it using Weyl's quantization procedure. It is informally written as

$$(\mathrm{Op}_h(a)u)(x) = (a^W(x, hD_x)u)(x) = \frac{1}{(2\pi h)^n}\int_{\mathbb{R}^{2n}} a\left(\frac{x+y}{2}, \xi\right)u(y)e^{i((x-y)\cdot\xi)/h}\,dy\,d\xi.$$

We will denote by $\Psi_\delta(\mathbb{R}^n)$ the corresponding classes of pseudodifferential operators. By definition, the wavefront set of $A = \mathrm{Op}_h(a)$ is $\mathrm{WF}_h(A) = \operatorname{ess\,supp} a$.

We say that a family $u = u(h) \in \mathcal{D}'(\mathbb{R}^n)$ is h-tempered if, for every $\chi \in C_c^\infty(\mathbb{R}^n)$, there exist $C > 0$ and $N \in \mathbb{N}$ such that $\|\chi u\|_{H_h^{-N}} \le Ch^{-N}$. For a h-tempered family u, we say that a point $\rho \in T^*\mathbb{R}^n$ does *not* belong to the wavefront set of u if there exists $a \in S^{\mathrm{comp}}(T^*\mathbb{R}^n)$ such that $a(\rho) \neq 0$ and $\mathrm{Op}_h(a)u = O(h^\infty)_{\mathcal{S}}$. We denote by $\mathrm{WF}_h(u)$ the wavefront set of u.

We say that a family of operators $B = B(h) : C_c^\infty(\mathbb{R}^{n_2}) \to \mathcal{D}'(\mathbb{R}^{n_1})$ is h-tempered if its Schwartz kernel $\mathcal{K}_B \in \mathcal{D}'(\mathbb{R}^{n_1} \times \mathbb{R}^{n_2})$ is h-tempered. We define

$$\mathrm{WF}_h'(B) = \{(x, \xi, y, -\eta) \in T^*\mathbb{R}^{n_1} \times T^*\mathbb{R}^{n_2} : (x, \xi, y, \eta) \in \mathrm{WF}_h(\mathcal{K}_B)\}.$$

Let us now recall standard results in semiclassical analysis concerning the L^2-boundedness of pseudodifferential operators and their composition. We'll use the following version of the Calderón–Vaillancourt theorem [Zworski 2012, Theorem 4.23].

Theorem 4. *There exists $C_n > 0$ such that the following holds. For every $0 \le \delta < \frac{1}{2}$ and $a \in S_\delta(T^*\mathbb{R}^n)$, $\mathrm{Op}_h(a)$ is a bounded operator on L^2 and*

$$\|\mathrm{Op}_h(a)\|_{L^2(\mathbb{R}^n)\to L^2(\mathbb{R}^n)} \le C_n \sum_{|\alpha|\le 8n} h^{|\alpha|/2}\|\partial^\alpha a\|_{L^\infty}.$$

As a consequence of the sharp Gårding inequality (see [Zworski 2012, Theorem 4.32]), we also have the precise estimate of L^2 norms of pseudodifferential operator,

Proposition 3.2. *Assume that $a \in S_\delta(\mathbb{R}^{2n})$. Then, there exists C_a depending on a finite number of seminorms of a such that*

$$\|\mathrm{Op}_h(a)\|_{L^2\to L^2} \le \|a\|_\infty + C_a h^{1/2-\delta}.$$

We recall that the Weyl quantizations of real symbols are self-adjoint in L^2. The composition of two pseudodifferential operators in Ψ_δ is still a pseudodifferential operator. More precisely (see [Zworski 2012, Theorems 4.11 and 4.18]), if $a, b \in S_\delta$, then $\mathrm{Op}_h(a) \circ \mathrm{Op}_h(b)$ is given by $\mathrm{Op}_h(a \# b)$, where $a \# b$ is the Moyal product of a and b. It is given by

$$a \# b(\rho) = e^{ihA(D)}(a \otimes b)|_{\rho=\rho_1=\rho_2},$$

where $a \otimes b(\rho_1, \rho_2) = a(\rho_1)b(\rho_2)$, $e^{ihA(D)}$ is a Fourier multiplier acting on functions on $\mathbb{R}^{4n}$ and, writing $\rho_i = (x_i, \xi_i)$,

$$A(D) = \tfrac{1}{2}(D_{\xi_1} \circ D_{x_2} - D_{x_1} \circ D_{\xi_2}).$$

We can estimate the Moyal product by a quadratic stationary phase and get the following expansion: for all $N \in \mathbb{N}$,

$$a \# b(\rho) = \sum_{k=0}^{N-1} \frac{i^k h^k}{k!} A(D)^k (a \otimes b)|_{\rho=\rho_1=\rho_2} + r_N,$$

where, for all $\alpha \in \mathbb{N}^{2n}$, there exists C_α, independent of a and b, such that

$$\|\partial^\alpha r_N\|_\infty \leq C_\alpha h^N \|a \otimes b\|_{C^{2N+4n+1+|\alpha|}}.$$

As a consequence of this asymptotic expansion, we have the precise product formula:

Lemma 3.3. *For every $N \in \mathbb{N}$, there exists $C_N > 0$ such that, for every $a, b \in S_\delta(\mathbb{R}^n)$,*

$$\mathrm{Op}_h(a) \circ \mathrm{Op}_h(b) = \mathrm{Op}_h\left(\sum_{k=0}^{N-1} \frac{i^k h^k}{k!} A(D)^k (a \otimes b)|_{\rho=\rho_1=\rho_2}\right) + R_N, \tag{3-1}$$

where

$$\|R_N\|_{L^2(\mathbb{R}) \to L^2(\mathbb{R})} \leq C_N h^N \|a \otimes b\|_{C^{2N+12n+1}}. \tag{3-2}$$

Remark. It will be important in the sequel to understand the derivatives of a and b involved in the k-th term of the previous expansion. A quick recurrence using the precise form of the operator $A(D)$ shows that $A(D)^k(a \otimes b)(\rho_1, \rho_2)$ is of the form

$$\sum_{|\alpha|=k, |\beta|=k} \lambda_{\alpha,\beta} \partial^\alpha a(\rho_1) \partial^\beta b(\rho_2).$$

This can be rewritten $l_k(d^k a(\rho_1), d^k b(\rho_2))$, where l_k is a bilinear form on the spaces of k-symmetric forms on $\mathbb{R}^{2n}$. Of course, we make use of the identifications $T_{\rho_1} T^*\mathbb{R}^n \simeq T_{\rho_2} T^*\mathbb{R}^n \simeq \mathbb{R}^{2n}$.

As a simple corollary, we get an expression for the commutator of pseudodifferential operators.

Corollary 3.4. *For every $N \in \mathbb{N}$, there exists $C_N > 0$ such that, for every $a, b \in S_\delta(\mathbb{R}^n)$,*

$$[\mathrm{Op}_h(a), \mathrm{Op}_h(b)] = \mathrm{Op}_h\left(\frac{h}{i}\{a, b\} + \sum_{k=2}^{N-1} h^k L_k(d^k a, d^k b)\right) + R_N,$$

where

$$\|R_N\|_{L^2(\mathbb{R}) \to L^2(\mathbb{R})} \leq C_N h^N \|a \otimes b\|_{C^{2N+12n+1}},$$

where the L_k are bilinear forms on the spaces of k-symmetric forms on $\mathbb{R}^{2n}$.

3B. ***Fourier Integral operators.*** We now review some aspects of the theory of Fourier integral operators. We follow [Zworski 2012, Chapter 11] and [Nonnenmacher et al. 2014]. We refer the reader to [Guillemin and Sternberg 2013] for further details. Finally, we will give the precise definition needed to understand Definition 2.1.

3B1. *Local symplectomorphisms and their quantization.* We momentarily work in dimension n. Let us denote by $\mathcal{K}$ the set of symplectomorphisms $\kappa : T^*\mathbb{R}^n \to T^*\mathbb{R}^n$ such that the following holds: there exist continuous and piecewise smooth families of smooth functions $(\kappa_t)_{t\in[0,1]}$, $(q_t)_{t\in[0,1]}$ such that:

- For all $t \in [0, 1]$, $\kappa_t : T^*\mathbb{R}^n \to T^*\mathbb{R}^n$ is a symplectomorphism.
- $\kappa_0 = \mathrm{Id}_{T^*\mathbb{R}^n}$, $\kappa_1 = \kappa$.
- For all $t \in [0, 1]$, $\kappa_t(0) = 0$.
- There exists $K \Subset T^*\mathbb{R}^n$ compact such that, for all $t \in [0, 1]$, $q_t : T^*\mathbb{R}^n \to \mathbb{R}$ and $\operatorname{supp} q_t \subset K$.
- $(d/dt)\kappa_t = (\kappa_t)^* H_{q_t}$.

If $\kappa \in \mathcal{K}$, we denote by $C = \mathrm{Gr}'(\kappa) = \{(x, \xi, y, -\eta) : (x, \xi) = \kappa(y, \eta)\}$ the twisted graph of κ. We recall [Zworski 2012, Lemma 11.4], which asserts that local symplectomorphisms can be seen as elements of $\mathcal{K}$, as soon as we have some geometric freedom.

Lemma 3.5. *Let U_0, U_1 be open and precompact subsets of $T^*\mathbb{R}^n$. Assume that $\kappa : U_0 \to U_1$ is a local symplectomorphism fixing* 0 *and which extends to $V_0 \Supset U_0$ an open star-shaped neighborhood of* 0. *Then, there exists $\tilde{\kappa} \in \mathcal{K}$ such that $\tilde{\kappa}|_{U_0} = \kappa$.*

If $\kappa \in \mathcal{K}$ and if (q_t) denotes the family of smooth functions associated with κ in its definition, we let $Q(t) = \mathrm{Op}_h(q_t)$. It is a continuous and piecewise smooth family of operators. Then the Cauchy problem

$$\begin{cases} hD_t U(t) + U(t)Q(t) = 0, \\ U(0) = \mathrm{Id} \end{cases} \tag{3-3}$$

is globally well-posed.

Following [Nonnenmacher et al. 2014, Definition 3.9], we adopt the definition:

Definition 3.6. Fix $\delta \in \left[0, \tfrac{1}{2}\right[$. We say that $U \in I_\delta(\mathbb{R}^n \times \mathbb{R}^n; C)$ if there exist $a \in S_\delta(T^*\mathbb{R}^n)$ and a path (κ_t) from Id to κ satisfying the above assumptions such that $U = \mathrm{Op}_h(a)U(1)$, where $t \mapsto U(t)$ is the solution of the Cauchy problem (3-3).

The class $I_{0^+}(\mathbb{R} \times \mathbb{R}, C)$ is by definition $\bigcap_{\delta>0} I_\delta(\mathbb{R} \times \mathbb{R}, C)$.

It is a standard result, known as Egorov's theorem (see [Zworski 2012, Theorem 11.1]) that if $U(t)$ solves the Cauchy problem (3-3) and if $a \in S_\delta$, then $U^{-1}\,\mathrm{Op}_h(a)U$ is a pseudodifferential operator in Ψ_δ and if $b = a \circ \kappa$, then $U^{-1}\,\mathrm{Op}_h(a)U - \mathrm{Op}_h(b) \in h^{1-2\delta}\Psi_\delta$.

Remark. Applying Egorov's theorem and Beal's theorem, it is possible to show that if (κ_t) is a closed path from Id to Id, and $U(t)$ solves (3-3), then $U(1) \in \Psi_0(\mathbb{R}^n)$. In other words, $I_\delta(\mathbb{R} \times \mathbb{R}, \mathrm{Gr}'(\mathrm{Id})) \subset \Psi_\delta(\mathbb{R}^n)$. But the other inclusion is trivial. Hence, this in an equality:

$$I_\delta(\mathbb{R}^n \times \mathbb{R}^n, \mathrm{Gr}'(\mathrm{Id})) = \Psi_\delta(\mathbb{R}^n).$$

The notation $I(\mathbb{R}^n \times \mathbb{R}^n, C)$ comes from the fact that the Schwartz kernel of such operators are Lagrangian distributions associated with C, and in particular have wavefront set included in C. As a consequence, if $T \in I_\delta(\mathbb{R}^n \times \mathbb{R}^n, C)$, then $\mathrm{WF}'_h(T) \subset \mathrm{Gr}(T)$.

Let us state a simple proposition concerning the composition of Fourier integral operators:

Proposition 3.7. *Let $\kappa_1, \kappa_2 \in \mathcal{K}$ and $U_1 \in I_\delta(\mathbb{R} \times \mathbb{R}, \mathrm{Gr}'(\kappa_1))$, $U_2 \in I_\delta(\mathbb{R} \times \mathbb{R}, \mathrm{Gr}'(\kappa_1))$. Then,*

$$U_1 \circ U_2 \in I_\delta(\mathbb{R} \times \mathbb{R}, \mathrm{Gr}'(\kappa_1 \circ \kappa_2)).$$

Proof. Let's write $U_1 = \mathrm{Op}_h(a_1)U_1(1)$, $U_2 = \mathrm{Op}_h(a_2)U_2(1)$ with the obvious notation associated with the Cauchy problems (3-3) for κ_1 and κ_2. Egorov's theorem asserts that $U_1(1)\,\mathrm{Op}_h(a_2)U_1(1)^{-1} = \mathrm{Op}_h(b_2)$ for some $b_2 \in S_\delta$ and $\mathrm{Op}_h(a_1)\,\mathrm{Op}_h(b_2) = \mathrm{Op}_h(a_1 \# b_2)$. It is then enough to focus on the case $a_1 = a_2 = 1$. We set

$$U_3(t) := \begin{cases} U_1(2t) & \text{for } 0 \le t \le \frac{1}{2}, \\ U_1(1) \circ U_2(2t-1) & \text{for } \frac{1}{2} \le t \le 1. \end{cases}$$

It solves the Cauchy problem

$$\begin{cases} hD_t U_3(t) + U_3(t) Q_3(t) = 0, \\ U(0) = \mathrm{Id}, \end{cases}$$

with

$$Q_3(t) := \begin{cases} 2Q_1(2t) & \text{for } 0 \le t \le \frac{1}{2}, \\ 2Q_2(2t-1) & \text{for } \frac{1}{2} \le t \le 1. \end{cases}$$

To conclude the proof, it is enough to notice that this Cauchy problem is associated with the path $\kappa_3(t)$ between $\kappa(0) = \mathrm{Id}$ and $\kappa_3(1) = \kappa_1 \circ \kappa_2$, where

$$\kappa_3(t) := \begin{cases} \kappa_1(2t) & \text{for } 0 \le t \le \frac{1}{2}, \\ \kappa_1 \circ \kappa_2(2t-1) & \text{for } \frac{1}{2} \le t \le 1. \end{cases}$$ □

3B2. *Precise version of Egorov's theorem.* We will need a more quantitative version of Egorov's theorem, similar to the one in [Dyatlov et al. 2022, Lemma A.7]. The result does not show that $U(1)^{-1}\,\mathrm{Op}_h(a)U(1)$ is a pseudodifferential operator (one would need Beal's theorem to say that) but it gives a precise estimate on the remainder, depending on the seminorms of a. We now specialize to the case of dimension $n = 1$ but the following result holds in any dimension but changing the constant 15 to something of the form Mn.

Proposition 3.8. *Consider $\kappa \in \mathcal{K}$ and denote by $U(t)$ the solution of* (3-3)*. There exists a family of differential operators $(D_j)_{j\in\mathbb{N}}$ of order j such that, for all $a \in S_\delta$ and all $N \in \mathbb{N}$,*

$$U(1)^{-1}\,\mathrm{Op}_h(a)U(1) = \mathrm{Op}_h\Big(a \circ \kappa + \sum_{j=1}^{N-1} h^j (D_{j+1}a) \circ \kappa\Big) + O_\kappa(h^N \|a\|_{C^{2N+15}}). \tag{3-4}$$

Proof. We keep the notation introduced previously. Let us first define

$$A_0(t) = U(t)\,\mathrm{Op}_h(a \circ \kappa_t)U(t)^{-1}$$

and compute

$$\begin{aligned} &U(t)^{-1}\partial_t A_0(t)U(t) \\ &\quad= -\frac{i}{h}[Q(t), \mathrm{Op}_h(a\circ\kappa_t)] + \mathrm{Op}_h(\{q_t, a\circ\kappa_t\}) \\ &\quad= \mathrm{Op}_h(\{q_t, a\circ\kappa_t\}) - \frac{i}{h}\bigg(\mathrm{Op}_h\Big(\frac{h}{i}\{q_t, a\circ\kappa_t\} + \sum_{j=2}^{N} h^j L_j(d^j q_t, d^j(a\circ\kappa_t))\Big)\bigg) + O(h^N \|q_t \otimes (a\circ\kappa_t)\|_{C^{2(N+1)+13}}) \\ &\quad= \mathrm{Op}_h\Big(\sum_{j=1}^{N-1} -ih^j L_{j+1}(d^{j+1}q_t, d^{j+1}(a\circ\kappa_t))\Big) + O_{\kappa_t}(h^N \|a\|_{C^{2N+15}}). \end{aligned}$$

We now define by induction a family of functions $a_j(t),\ j=0,\dots,N-1$, by

$$a_0(t)=a, \quad a_k(t)=\sum_{m=0}^{k-1}\int_0^t iL_{k+1-m}(d^{k+1-m}q_s, d^{k+1-m}(a_m(s)\circ\kappa_s))\circ\kappa_s^{-1}\,ds,$$

and set $A_k(t)=U(t)\operatorname{Op}_h\bigl(\sum_{j=0}^k h^j a_j(t)\circ\kappa_t\bigr)U(t)^{-1}$. We first remark by an easy induction on k, that $a_k(t)$ is of the form $D_{k+1}(t)a$, where $D_{k+1}(t)$ is a differential operator of order at most $k+1$, with coefficients depending continuously on t and on $(\kappa_t)_t$. We now check the following by induction:

$$U(t)^{-1}\partial_t A_k(t)U(t)=-i\operatorname{Op}_h\Biggl(\sum_{j=k+1}^{N-1}\sum_{m=0}^{k}h^jL_{j+1-m}(d^{j+1-m}q_t, d^{j+1-m}(a_m(t)\circ\kappa_t))\Biggr)+O_\kappa(h^N\|a\|_{C^{2N+15}}).$$

We've already done it for $k=0$. Let's assume that the equality holds for $k-1$ and let's prove it for $k\geq 1$:

$$U(t)^{-1}\partial_t A_k(t)U(t)=U(t)^{-1}\partial_t A_{k-1}(t)U(t)+h^kU(t)^{-1}\partial_t\operatorname{Op}_h(a_k(t)\circ\kappa_t)U(t).$$

Let's compute the second part of the right-hand side:

$$\begin{aligned}
&U(t)^{-1}\partial_t\operatorname{Op}_h(a_k(t)\circ\kappa_t)U(t)\\
&\quad=-\frac{i}{h}[Q(t),\operatorname{Op}_h(a_k(t)\circ\kappa_t)]+\operatorname{Op}_h(\{q_t,a_k(t)\circ\kappa_t\})+\operatorname{Op}_h(\partial_t a_k(t)\circ\kappa_t)\\
&\quad=-i\operatorname{Op}_h\Biggl(\sum_{l=1}^{N-1-k}h^jL_{l+1}(d^{l+1}q_t, d^{l+1}(a_k(t)\circ\kappa_t))\Biggr)+O_\kappa(h^{N-k}\|a_k(t)\|_{C^{2(N+1-k)+13}})+\operatorname{Op}_h(\partial_t a_k(t)\circ\kappa_t).
\end{aligned}$$

We can estimate the remainder by

$$O_\kappa(h^{N-k}\|a_k(t)\|_{C^{2(N+1-k)+13}})=O_\kappa(h^{N-k}\|a\|_{C^{2(N+1-k)+13+k+1}})=O_\kappa(h^{N-k}\|a\|_{C^{2N+15}}).$$

We now combine this with the value of

$$U(t)^{-1}\partial_t A_{k-1}(t)U(t)=-i\operatorname{Op}_h\Biggl(\sum_{j=k}^{N-1}\sum_{m=0}^{k-1}h^jL_{j+1-m}(d^{j+1-m}q_t, d^{j+1-m}(a_m(t)\circ\kappa_t))\Biggr)+O_\kappa(h^N\|a\|_{C^{2N+15}}).$$

By the definition of $a_k(t)$, the term $h^k\operatorname{Op}_h(\partial_t a_k(t)\circ\kappa_t)$ cancels the term corresponding to $j=k$ in the sum. Moreover, for every $j>k$, writing $j=k+l,\ l\in\{1,\dots,N-1-k\}$, the term $h^{k+l}L_{l+1}(d^{l+1}q_t, d^{l+1}(a_k(t)\circ\kappa_t))$ gives the missing term $h^jL_{j+1-k}(d^{j+1-k}q_t, d^{j+1-k}(a_k(t)\circ\kappa_t))$. This gives the required equality for $A_k(t)$.

In particular, $\partial_t A_{N-1}(t)=O_\kappa(h^N\|a\|_{C^{2N+15}})$. We now use the fact that at $t=0$, $a_0(0)=a$, $a_k(0)=0$, $k=1,\dots,N-1$, $U(0)=\mathrm{Id}$, $\kappa_0=\mathrm{Id}$, and hence $A_{N-1}(0)=\operatorname{Op}_h(a)$. Integrating between 0 and 1, we have

$$A_{N-1}(t)-\operatorname{Op}_h(a)=O_\kappa(h^N\|a\|_{C^{2N+15}}).$$

Conjugating by $U(1)$, we finally have

$$U(1)^{-1}\operatorname{Op}_h(a)U(1)=\operatorname{Op}_h\Biggl(a\circ\kappa+\sum_{k=1}^{N-1}h^ka_k(1)\circ\kappa\Biggr)+O_\kappa(h^N\|a\|_{C^{2N+15}}),$$

which is what we wanted, since $a_k(1)=D_{k+1}(1)a$. $\square$

3B3. *An important example.* Let us focus on a particular case of canonical transformations. Suppose that $\kappa : T^*\mathbb{R}^n \to T^*\mathbb{R}^n$ is a canonical transformation such that

$$(x, \xi, y, \eta) \in \mathrm{Gr}(\kappa) \mapsto (x, \eta)$$

is a local diffeomorphism near $(x_0, \xi_0, y_0, \eta_0)$. Then, there exists a phase function $\psi \in C^\infty(\mathbb{R}^n \times \mathbb{R}^n)$, Ω_x, Ω_η open sets of $\mathbb{R}^n$ and Ω a neighborhood of $(x_0, \xi_0, y_0, \eta_0)$ such that

$$\mathrm{Gr}'(\kappa) \cap \Omega = \{(x, \partial_x \psi(x, \eta), \partial_\eta \psi(x, \eta), -\eta) : x \in \Omega_x, \eta \in \Omega_\eta\}.$$

One says that ψ generates $\mathrm{Gr}'(\kappa)$. Suppose that $\alpha \in S_\delta^{\mathrm{comp}}(\Omega_x \times \Omega_\eta)$. Then, modulo a smoothing operator $O(h^\infty)$, the following operator T is an element of $I_\delta^{\mathrm{comp}}(\mathbb{R}^n \times \mathbb{R}^n, \mathrm{Gr}'(\kappa))$:

$$Tu(x) = \frac{1}{(2\pi h)^n} \int_{\mathbb{R}^{2n}} e^{(i/h)(\psi(x,\eta) - y\cdot\eta)} \alpha(x, \eta)\, u(y)\, dy\, d\eta,$$

and if $T^*T = \mathrm{Id}$ microlocally near (y_0, η_0) then $|\alpha(x, \eta)|^2 = |\det D^2_{x\eta} \psi(x, \eta)| + O(h^{1-2\delta})_{S_\delta}$ near $(x_0, \xi_0, y_0, \eta_0)$. The converse statement holds: microlocally near $(x_0, \xi_0, y_0, \eta_0)$ and modulo $O(h^\infty)$, the elements of $I_\delta(\mathbb{R}^n \times \mathbb{R}^n, \mathrm{Gr}'(\kappa))$ can be written under this form.

3B4. *Lagrangian relations.* Recall that the Lagrangian relation F we consider is the union of local Lagrangian relations $F_{ij} \subset U_i \times U_j$. We fix a compact set $W \subset \pi_L(F)$ containing some neighborhood of $\mathcal{T}$. Our definition will depend on W. Following [Nonnenmacher et al. 2014, Section 3.4.2], we now focus on the definition of the elements of $I_\delta(Y \times Y; F')$. An element $T \in I_\delta(Y \times Y; F')$ is a matrix of operators

$$T = (T_{ij})_{1 \le i, j \le J} : \bigoplus_{j=1}^{J} L^2(Y_j) \to \bigoplus_{i=1}^{J} L^2(Y_i).$$

Each T_{ij} is an element of $I_\delta(Y_i \times Y_j, F'_{ij})$. Let's now describe the recipe to construct elements of $I_\delta(Y_i \times Y_j, F'_{ij})$. We fix $i, j \in \{1, \dots, J\}$.

- Fix some small $\varepsilon > 0$ and two open covers of U_j, $U_j \subset \bigcup_{l=1}^{L} \Omega_l$, $\Omega_l \Subset \widetilde{\Omega}_l$, with $\widetilde{\Omega}_l$ star-shaped and having diameter smaller than ε. We denote by $\mathcal{L}$ the sets of indices l such that $\Omega_l \subset \pi_R(F_{ij}) = \widetilde{D}_{ij} \subset U_j$ and we require (this is possible if ε is small enough)

$$F^{-1}(W) \cap U_j \subset \bigcup_{l \in \mathcal{L}} \Omega_l.$$

- Introduce a smooth partition of unity associated with the cover (Ω_l), $(\chi_l)_{1 \le l \le L} \in C_c^\infty(\Omega_l, [0, 1])$, $\mathrm{supp}\, \chi_l \subset \Omega_l$, $\sum_l \chi_l = 1$ in a neighborhood of $\overline{U}_j$.

- For each $l \in \mathcal{L}$, we denote by F_l the restriction to $\widetilde{\Omega}_l$ of F_{ij}, seen as a symplectomorphism $F_{ij} : \widetilde{D}_{ij} \subset U \to V$. By Lemma 3.5, there exists $\kappa_l \in \mathcal{K}$ which coincides with F_l on Ω_l.

- We consider $T_l = \mathrm{Op}_h(\alpha_i) U_l(1)$, where $U_l(t)$ is the solution of the Cauchy problem (3-3) associated with κ_l and $\alpha_i \in S_\delta^{\mathrm{comp}}(T^*\mathbb{R})$.

- We set

$$T^{\mathbb{R}} = \sum_{l \in \mathcal{L}} T_l \operatorname{Op}_h(\chi_l) : L^2(\mathbb{R}) \to L^2(\mathbb{R}). \tag{3-5}$$

$T^{\mathbb{R}}$ is a globally defined Fourier integral operator. We will write $T^{\mathbb{R}} \in I_\delta(\mathbb{R} \times \mathbb{R}, F'_{ij})$. Its wavefront set is included in $\tilde{A}_{ij} \times \widetilde{D}_{ij}$.

- Finally, we fix cut-off functions $(\Psi_i, \Psi_j) \in C_c^\infty(Y_i, [0,1]) \times C_c^\infty(Y_j, [0,1])$ such that $\Psi_i \equiv 1$ on $\pi(U_i)$ and $\Psi_j \equiv 1$ on $\pi(U_j)$ (here, $\pi : (x, \xi) \in T^*Y_\cdot \mapsto x \in Y_\cdot$ is the natural projection) and we adopt the following definitions:

Definition 3.9. We say that $T : \mathcal{D}'(Y_j) \to C^\infty(\overline{Y}_i)$ is a Fourier integral operator in the class $I_\delta(Y_i \times Y_j, F'_{ij})$ if there exists $T^{\mathbb{R}} \in I_\delta(\mathbb{R} \times \mathbb{R}, F')$ as constructed above such that

- $T - \Psi_i T \Psi_j = O(h^\infty)_{\mathcal{D}'(Y) \to C^\infty(\bar{Z})}$,
- $\Psi_i T \Psi_j = \Psi_i T^{\mathbb{R}} \Psi_j$.

For $U'_j \subset U_j$ and $U'_i = F(U'_j) \subset U_i$, we say that T (or $T^{\mathbb{R}}$) is microlocally unitary in $U'_i \times U'_j$ if $TT^* = \mathrm{Id}$ microlocally in U'_i and $T^*T = \mathrm{Id}$ microlocally in U'_j.

Remark. The definition of this class is not canonical since it depends in fact on the compact set W through the partition of unity.

Another version of Egorov's theorem. The precise version of Egorov's theorem in Proposition 3.8 is only stated for globally unitary Fourier integral operator defined using the Cauchy equation (3-3). We extend it here to microlocally unitary and globally defined Fourier integral operators. We fix $i, j \in \{1, \dots, J\}$.

Lemma 3.10. *Let $T \in I_\delta(\mathbb{R} \times \mathbb{R}, F'_{ij})$. Suppose that $B(\rho, 4\varepsilon) \subset U_j$ and that T is microlocally unitary in $F_{ij}(B(\rho, 4\varepsilon)) \times B(\rho, 4\varepsilon)$. Then, there exists a family $(D_k)_{k \in \mathbb{N}}$ of differential operators of order k, compactly supported in $B(\rho, 3\varepsilon)$ such that the following holds: For every $N \in \mathbb{N}$ and for all $b \in C_c^\infty(B(\rho, 2\varepsilon))$,*

$$T \operatorname{Op}_h(b) = \operatorname{Op}_h\left(b \circ \kappa^{-1} + \sum_{k=1}^{N-1} h^k (D_{k+1} b) \circ \kappa^{-1}\right) T + O(h^N \|b\|_{C^{2N+15}})_{L^2(\mathbb{R}) \to L^2(\mathbb{R})}.$$

The constants in O depend on T and F.

Proof. First, introduce some cut-off function χ such that $\chi \equiv 1$ in a neighborhood of $B(\rho, 2\varepsilon)$ and $\operatorname{supp} \chi \subset B(\rho, 3\varepsilon)$. Due to these properties and Lemma 3.3, we have

$$\operatorname{Op}_h(b) = \operatorname{Op}_h(\chi) \operatorname{Op}_h(b) \operatorname{Op}_h(\chi) \operatorname{Op}_h(\chi) + O(h^N \|b\|_{C^{2N+13}})_{L^2(\mathbb{R}) \to L^2(\mathbb{R})}.$$

Moreover, $\operatorname{Op}_h(\chi) T^* T = \operatorname{Op}_h(\chi) + O(h^\infty)$, and hence

$$T \operatorname{Op}_h(b) = T \operatorname{Op}_h(\chi) \operatorname{Op}_h(b) \operatorname{Op}_h(\chi) \operatorname{Op}_h(\chi) T^* T + O(h^N \|b\|_{C^{2N+13}})_{L^2 \to L^2} + O(h^\infty) \| \operatorname{Op}_h(b)\|_{L^2 \to L^2}.$$

The term $O(h^\infty) \| \operatorname{Op}_h(b)\|_{L^2 \to L^2}$ can be absorbed in $O(h^N \|b\|_{C^{2N+13}})_{L^2 \to L^2}$. Consider $\tilde{\kappa} \in \mathcal{K}$ extending $\kappa|_{B(\rho, 3\varepsilon)}$ and construct $U = U(1)$ by solving the Cauchy problem (3-3) associated with $\tilde{\kappa}$. Due to the properties on composition of Fourier integral operators (Proposition 3.7), $T\,\mathrm{Op}_h(\chi)U^{-1}$ and $U\,\mathrm{Op}_h(\chi)T^*$ are pseudodifferential operators, and we denote them by $\mathrm{Op}_h(a_1)$, $\mathrm{Op}_h(a_2)$. Now write

$$\begin{aligned} T\,\mathrm{Op}_h(b) &= [T\,\mathrm{Op}_h(\chi)U^{-1}]U\,\mathrm{Op}_h(b)\,\mathrm{Op}_h(\chi)U^{-1}[U\,\mathrm{Op}_h(\chi)T^*]T + O(h^N\|b\|_{C^{2N+13}})_{L^2\to L^2} \\ &= \mathrm{Op}_h(a_1)[U\,\mathrm{Op}_h(b)\,\mathrm{Op}_h(\chi)U^{-1}]\,\mathrm{Op}_h(a_2)T + O(h^N\|b\|_{C^{2N+13}})_{L^2\to L^2}. \end{aligned}$$

By using the precise version in Proposition 3.8, one can write

$$U\,\mathrm{Op}_h(b)\,\mathrm{Op}_h(\chi)U^{-1} = \mathrm{Op}_h\biggl(b\circ\kappa^{-1} + \sum_{k=1}^{N-1}(L_{k+1}b)\circ\kappa^{-1}\biggr) + O(h^N\|b\|_{C^{2N+15}})_{L^2\to L^2}.$$

Applying Lemma 3.3, we see that we can write

$$T\,\mathrm{Op}_h(b) = \mathrm{Op}_h\biggl(b_0\circ\kappa^{-1} + \sum_{k=1}^{N-1}(D_{k+1}b)\circ\kappa^{-1}\biggr)T + O(h^N\|b\|_{C^{2N+15}})_{L^2\to L^2},$$

where $b_0 = a_1\times b\circ\kappa^{-1}\times a_2$. Since T is microlocally unitary in $B(\rho, 4\varepsilon)$, the product a_1a_2 is equal to 1 in $B(\rho, 2\varepsilon)$, and hence, the lemma is proved. □

3C. *Hyperbolic dynamics.* We assumed that F is hyperbolic on the trapped set $\mathcal{T}$. As already mentioned, we can fix an adapted Riemannian metric on U such that the following stronger version of the hyperbolic estimates are satisfied for some $\lambda_0 > 0$: for every $\rho\in\mathcal{T}$, $n\in\mathbb{N}$,

$$v\in E_u(\rho) \quad\Longrightarrow\quad \|d_\rho F^{-n}(v)\| \le e^{-\lambda_0 n}\|v\|, \tag{3-6}$$

$$v\in E_s(\rho) \quad\Longrightarrow\quad \|d_\rho F^{n}(v)\| \le e^{-\lambda_0 n}\|v\|. \tag{3-7}$$

Notation. We now use the induced Riemannian distance on U and denote it by d.

We also use the same notation $\|\cdot\|$ to denote the subordinate norm on the space of linear maps between tangent spaces of U; namely, if $F(\rho_1) = \rho_2$,

$$\|d_{\rho_1}F\| = \sup_{v\in T_{\rho_1}U,\ \|v\|_{\rho_1}=1} \|d_{\rho_1}F(v)\|_{\rho_2}.$$

If $\rho\in\mathcal{T}$, $n\in\mathbb{Z}$, we use this Riemannian metric to define the unstable Jacobian $J_n^u(\rho)$ and stable Jacobian $J_n^s(\rho)$ at ρ by

$$v\in E_u(\rho) \quad\Longrightarrow\quad \|d_\rho F^{n}(v)\| = J_n^u(\rho)\|v\|, \tag{3-8}$$

$$v\in E_s(\rho) \quad\Longrightarrow\quad \|d_\rho F^{n}(v)\| = J_n^s(\rho)\|v\|. \tag{3-9}$$

These Jacobians quantify the local hyperbolicity of the map.

Notation. Suppose that f and g are some real-valued functions depending on the same family of parameters $\mathcal{P}$. For instance, for $J_n^u(\rho)$, $\mathcal{P} = \{n, \rho\}$. We will write $f\sim g$ to mean that there exists a constant $C\ge 1$ depending only on (U, F), but not on $\mathcal{P}$, such that $C^{-1}g\le f\le Cg$.

For instance, if we define unstable and stable Jacobians $\tilde{J}_n^u$ and $\tilde{J}_n^s$ using another Riemannian metric, then, for every $n\in\mathbb{Z}$ and $\rho\in\mathcal{T}$,

$$\tilde{J}_n^u(\rho)\sim J_n^u(\rho), \qquad \tilde{J}_n^s(\rho)\sim J_n^s(\rho).$$

From the compactness of $\mathcal{T}$, there exist $\lambda_1 \geq \lambda_0$ which satisfy

$$e^{n\lambda_0} \leq J_n^u(\rho) \leq e^{n\lambda_1} \quad \text{and} \quad e^{-n\lambda_1} \leq J_n^s(\rho) \leq e^{-n\lambda_0}, \quad n \in \mathbb{N},\ \rho \in \mathcal{T}, \tag{3-10}$$

$$e^{n\lambda_0} \leq J_{-n}^s(\rho) \leq e^{n\lambda_1} \quad \text{and} \quad e^{-n\lambda_1} \leq J_{-n}^u(\rho) \leq e^{-n\lambda_0}, \quad n \in \mathbb{N},\ \rho \in \mathcal{T}. \tag{3-11}$$

We cite here standard facts about the stable and unstable manifolds; see for instance [Katok and Hasselblatt 1995, Chapter 6].

Lemma 3.11. *For any $\rho \in \mathcal{T}$, there exist local stable and unstable manifolds $W_s(\rho)$, $W_u(\rho) \subset U$ satisfying, for some $\varepsilon_1 > 0$ (only depending on F) ($\star$ will denote a letter in $\{u, s\}$ and the use of $\pm$ with $\star$ has to be read with the convention $u \to -$, $s \to +$):*

1. *$W_s(\rho)$, $W_u(\rho)$ are C^∞-embedded curves, with the C^∞ norms of the embedding uniformly bounded in ρ.*
2. *The boundary of $W_\star(\rho)$ do not intersect $\overline{B(\rho, \varepsilon_1)}$.*[1]
3. *$W_s(\rho) \cap W_u(\rho) = \{\rho\}$ and $T_\rho W_\star(\rho) = E_\star(\rho)$.*
4. *$F^{\pm}(W_\star(\rho)) \subset W_\star(F(\rho))$.*
5. *For each $\rho' \in W_\star(\rho)$, we have $d(F^{\pm n}(\rho), F^{\pm n}(\rho')) \to 0$.*
6. *Let $\theta > 0$ satisfying $e^{-\lambda_0} < \theta < 1$. If $\rho' \in U$ satisfies $d(F^{\pm i}(\rho), F^{\pm i}(\rho')) \leq \varepsilon_1$ for all $i = 0, \dots, n$ then $d(\rho', W_\star(\rho)) \leq C\theta^n \varepsilon_1$ for some $C > 0$.*
7. *If $\rho, \rho' \in \mathcal{T}$ satisfy $d(\rho, \rho') \leq \varepsilon_1$, then $W_u(\rho) \cap W_s(\rho')$ consists of exactly one point in $\mathcal{T}$.*

Since we work with the local unstable and stable manifolds, we may assume that $W_\star(\rho) \subset B(\rho, 2\varepsilon_1)$.

For our purpose, we will need a more precise version of these results. The following lemmas are an adaptation of Lemma 2.1 in [Dyatlov et al. 2022] to our setting.

Lemma 3.12. *There exists a constant $C > 0$ depending only on (U, F), such that, for all $\rho, \rho' \in U$:*

1. *If $\rho \in \mathcal{T}$ and $\rho' \in W_s(\rho)$ then*
$$d(F^n(\rho), F^n(\rho')) \leq C J_n^s(\rho) d(\rho, \rho') \quad \textit{for all } n \in \mathbb{N}. \tag{3-12}$$
2. *If $\rho \in \mathcal{T}$ and $\rho' \in W_u(\rho)$ then*
$$d(F^{-n}(\rho), F^{-n}(\rho')) \leq C J_{-n}^u(\rho) d(\rho, \rho') \quad \textit{for all } n \in \mathbb{N}. \tag{3-13}$$

Proof. We prove (1). Part (2) is proved in a similar way by inverting the time direction. Let $\rho \in \mathcal{T}$, $\rho' \in W_s(\rho)$. Since $T_\rho(W_s(\rho)) = E_s(\rho)$ and $d_\rho F(E_s(\rho)) = E_s(F(\rho))$, the Taylor development of F along $W_s(\rho)$ gives

$$d(F(\rho), F(\rho')) \leq J_1^s(\rho) d(\rho, \rho') + C d(\rho, \rho')^2 \leq J_1^s(\rho) d(\rho, \rho')(1 + C d(\rho, \rho')) \tag{3-14}$$

[1] In other words, there exists a smooth curve $\gamma : [-\delta, \delta] \to U$ such that $\overline{B(\rho, \varepsilon_1)} \cap W_\star(\rho) = \operatorname{Im} \gamma$, with $\gamma(0) = \rho$; it means that the size of the (un-)stable manifolds is bounded from below uniformly.

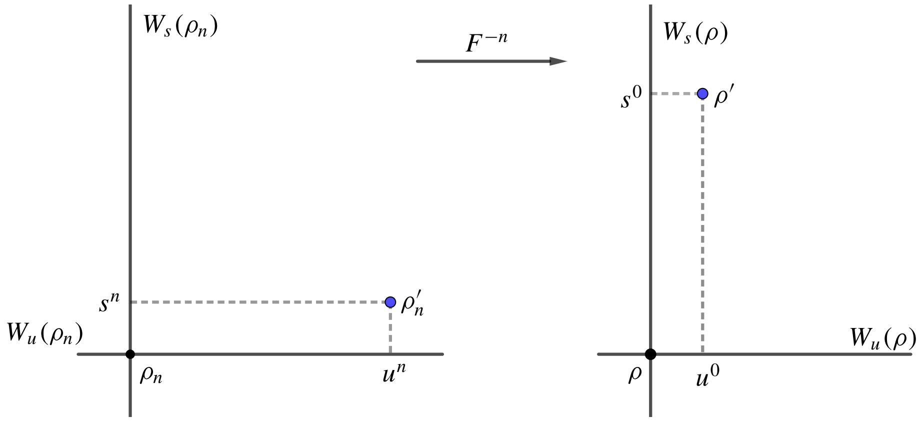

Figure 5. Framework for the proof of Lemma 3.13.

since $J_1^s \geq C^{-1}$. Applying this inequality with $F^k(\rho)$ and $F^k(\rho')$ instead of ρ and ρ', and recalling that, by Lemma 3.11, $d(F^k(\rho), F^k(\rho')) \leq C\theta^k d(\rho, \rho')$, we can write

$$d(F^{k+1}(\rho), F^{k+1}(\rho')) \leq J_1^s(F^k(\rho))\, d(F^k(\rho), F^k(\rho'))(1 + C\theta^k). \tag{3-15}$$

By this last inequality and the chain rule, we have

$$d(F^n(\rho), F^n(\rho')) \leq J_n^s(\rho) d(\rho, \rho') \prod_{k=0}^{n-1} (1 + C\theta^k) \leq C J_n^s(\rho) d(\rho, \rho'), \tag{3-16}$$

completing the proof. ☐

The following lemma gives a stronger version of (6) in Lemma 3.11.

Lemma 3.13. *There exist $C > 0$ and $\varepsilon_1 > 0$, depending only on (U, F), such that, for all $\rho, \rho' \in U$ and $n \in \mathbb{N}$:*

(1) *If $\rho \in \mathcal{T}$ and $d(F^i(\rho), F^i(\rho')) \leq \varepsilon_1$ for all $i \in \{0, \dots, n\}$ then*

$$d(\rho', W_s(\rho)) \leq \frac{C}{J_n^u(\rho)}, \tag{3-17}$$

$$\|d_{\rho'} F^n\| \leq C J_n^u(\rho). \tag{3-18}$$

(2) *If $\rho \in \mathcal{T}$ and $d(F^{-i}(\rho), F^{-i}(\rho')) \leq \varepsilon_1$ for all $i \in \{0, \dots, n\}$ then*

$$d(\rho', W_u(\rho)) \leq \frac{C}{J_{-n}^s(\rho)}, \tag{3-19}$$

$$\|d_{\rho'} F^{-n}\| \leq C J_{-n}^s(\rho). \tag{3-20}$$

Proof. We prove (1). Part (2) is proved in a similar way by inverting the time direction. Let $\rho \in \mathcal{T}$ and $\rho' \in U$ be such that $d(F^i(\rho), F^i(\rho')) \leq \varepsilon_1$ for $0 \leq i \leq n$ with ε_1 to be determined. Define $\rho_k = F^k(\rho)$. The first condition on ε_1 is that it is smaller than the one of Lemma 3.11 so that we ensure the following estimates: for $k \in \{0, \dots, n\}$,

$$d(F^k(\rho'), W_s(F^k(\rho))) \le C\theta^{n-k}\varepsilon_1, \tag{3-21}$$

$$d(F^k(\rho'), W_s(F^k(\rho))) \le C\theta^{k}\varepsilon_1. \tag{3-22}$$

We will use coordinates charts $\kappa_k : \hat{\rho} \in U_k \mapsto (u^k, s^k) \in V_k$ adapted to the dynamical system; see, in [Katok and Hasselblatt 1995], Theorem 6.2.3, the explanations below and Theorem 6.2.8 for the existence of this chart. More precisely, we want these charts to satisfy:

- $\kappa_k(\rho_k) = (0, 0)$.
- $\kappa_k(W_s(\rho_k) \cap U_k) = \{(0, s) : s \in \mathbb{R}\} \cap V_k$.
- $\kappa_k(W_u(\rho_k) \cap U_k) = \{(u, 0) : u \in \mathbb{R}\} \cap V_k$.
- For $\hat{\rho} \in U_k$, we have $|u^k| \sim d(\hat{\rho}, W_s(\rho_k))$, $|s^k| \sim d(\hat{\rho}, W_u(\rho_k))$ and $|s^k|^2 + |u^k|^2 \sim d(\rho_k, \hat{\rho})^2$.
- $(\kappa_k)_{0 \le k \le n}$ are uniformly bounded in the C^N topology for all N, with constant independent of ρ_0 and n. In particular, we may assume that ε_1 is chosen small enough so that $B(\rho_k, \varepsilon_1) \subset U_k$ for all $0 \le k \le n$.
- Up to changing the metric we work with (which is not problematic), we may assume that the restrictions of $d\kappa_k(\rho)$ to $E_s(\rho)$ and $E_u(\rho)$ are isometries for the metrics $|\cdot|_s$ and $|\cdot|_u$.

If we write $\widetilde{F}_k = \kappa_k \circ F \circ \kappa_{k-1}^{-1}$, we can check that in this pair of coordinates charts, the action of F^{-1} is given by

$$\widetilde{F}_k^{-1}(u^k, s^k) = (\pm J_{-1}^u(\rho_k)u^k + \alpha_k(u^k, s^k), \pm J_{-1}^s(\rho_k)s^k + \beta_k(u^k, s^k)), \tag{3-23}$$

where α_k, β_k are smooth functions, uniformly bounded in k for the C^2 topology and such that $\alpha_k(0, s^k) = 0$, $\beta_k(u^k, 0) = 0$, $d\alpha_k(0, 0) = 0$, $d\beta_k(0, 0) = 0$.

With these properties, one can check that

$$\alpha_k(u^k, s^k) \le C|u^k|\,\|(u^k, s^k)\|. \tag{3-24}$$

Let's now define $\rho_k' = F^k(\rho')$ and $(u^k, s^k) = \kappa_k(\rho_k')$. By (3-21), (3-22), (3-23), (3-24), we can write

$$\begin{aligned}
|u^{k-1}| &\le J_{-1}^u(\rho_k)|u^k| + C|u^k|\|(u^k, s^k)\| \\
&\le J_{-1}^u(F^k(\rho))|u^k|(1 + C\varepsilon_1(\theta_1^k + \theta_1^{n-k})) \\
&\le J_{-1}^u(F^k(\rho))|u^k|(1 + C\varepsilon_1\theta^{\min(k,n-k)}).
\end{aligned}$$

Then, using the chain rule, one has

$$d(\rho', W_s(\rho)) \le C|u^0| \le C J_{-n}^u(F^n(\rho)) \prod_{k=0}^{n-1}(1 + C\varepsilon_1\theta^{\min(k,n-k)}). \tag{3-25}$$

Finally, we can estimate

$$\prod_{k=0}^{n}(1 + C\varepsilon_1\theta^{\min(k,n-k)}) \le \prod_{k=0}^{\lceil n/2 \rceil}(1 + C\varepsilon_1\theta^k)^2 \le C,$$

which gives

$$d(\rho', W_s(\rho)) \le C J^u_{-n}(F^n(\rho)) = \frac{C}{J^u_n(\rho)}. \tag{3-26}$$

This proves (3-17).

To prove (3-18), we first construct a metric which simplifies the computations. If $\rho \in \mathcal{T}$, we pick $v_\star(\rho) \in E_\star(\rho)$[2] such that $\|v_\star(\rho)\| = 1$. There exists a Riemannian metric $|\cdot|$ on $\mathcal{T}$ such that, for every $\rho \in \mathcal{T}$, $(v_u(\rho), v_s(\rho))$ is an orthonormal basis of $T_\rho U$. This metric is γ-Hölder in $\rho \in \mathcal{T}$ since stable and unstable distributions are γ-Hölder for some $\gamma \in (0, 1)$.

If $\rho \in \mathcal{T}$ and $n \in \mathbb{Z}$, we denote by $\tilde{J}^{u/s}_n(\rho) \in \mathbb{R}$ the numbers such that

$$d_\rho(F^n)(v_u(\rho)) = \tilde{J}^u_n(\rho) v_u(F^n(\rho)), \quad d_\rho(F^n)(v_s(\rho)) = \tilde{J}^s_n(\rho) v_s(F^n(\rho)).$$

As already observed, $|\tilde{J}^u_n(\rho)| \sim J^u_n(\rho)$ for all n (with constants independent of n). We can also assume that $|\tilde{J}^u_1(\rho)| > |\tilde{J}^s_1(\rho)|$ for all ρ. In the orthonormal basis $(v_u(\rho), v_s(\rho))$ and $(v_u(F^n(\rho), v_s(F^n(\rho)))$, $d_\rho F^n$ has the form

$$\begin{pmatrix} \tilde{J}^u_n(\rho) & 0 \\ 0 & \tilde{J}^s_n(\rho) \end{pmatrix}.$$

Due to the orthonormality of these basis, we have that for the subordinate norms, $\|d_\rho F^n\| = |\tilde{J}^u_n(\rho)|$. Hence, the chain rule implies the following equality for this particular Riemannian metric defined on $\mathcal{T}$:

$$\text{for all } \rho \in \mathcal{T}, \quad \|d_\rho(F^n)\| = |\tilde{J}^u_n(\rho)| = \prod_{i=0}^{n-1} |\tilde{J}^u_1(F^i(\rho)| = \prod_{i=0}^{n-1} \|d_{F^i(\rho)} F\|. \tag{3-27}$$

We now claim that we can extend $|\cdot|$ to a relatively compact neighborhood V of $\mathcal{T}$ such that $\rho \in V \mapsto |\cdot|_\rho$ is still γ-Hölder. To do so, it is enough to extend the coefficients of the metric in a coordinate chart in a γ-Hölder way, which is possible (for instance, by virtue of Corollary 1 in [McShane 1934]), which still defines a nondegenerate 2-form in a sufficiently small neighborhood of $\mathcal{T}$.

We now aim at proving (3-18) for this particular metric. (3-18) will hold in the general case since two continuous metric are always uniformly equivalent in a compact neighborhood of $\mathcal{T}$.

In the following, we assume that ε_1 is small enough so that ρ belongs to the neighborhood of $\mathcal{T}$ in which $|\cdot|$ is defined. Since $\rho \mapsto \|d_\rho F\|_{T_\rho U \to T_{F(\rho)} U}$ is γ-Hölder (in the following, we will drop the subscript in the norm) we have, for all $i \in \{0, \dots, n\}$,

$$\big| \|d_{F^i(\rho')} F\| - \|d_{F^i(\rho)} F\| \big| \le C d(F^i(\rho'), F^i(\rho))^\gamma \le C\varepsilon_1 \theta^{\gamma \min(i, n-i)}. \tag{3-28}$$

Using the chain rule and the submultiplicativity of $\|\cdot\|$, we have

$$\|d_{\rho'} F^n\| \le \prod_{i=0}^{n} \|d_{F^i(\rho')} F\| \le \prod_{i=0}^{n} \|d_{F^i(\rho)} F\| (1 + C\varepsilon_1 \theta^{\gamma \min(i, n-i)}). \tag{3-29}$$

Eventually, by (3-27) and the fact that $\prod_{i=0}^n (1 + C\varepsilon_1 \theta^{\gamma \min(i,n-i)})$ is convergent, (3-18) holds. □

As an immediate consequence of this lemma, we get:

[2] Here, we are not concerned by the orientation. It is simply a matter of direction.

Corollary 3.14. *There exist $C > 0$ and $\varepsilon_1 > 0$ (depending only on (U, F)) such that, for all $\rho, \rho' \in \mathcal{T}$ and $n \in \mathbb{N}$:*

(1) *If $d(F^i(\rho), F^i(\rho')) \leq \varepsilon_1$ for all $i \in \{0, \dots, n\}$ then*

$$C^{-1} J_n^u(\rho) \leq J_n^u(\rho') \leq C J_n^u(\rho). \tag{3-30}$$

(2) *If $d(F^{-i}(\rho), F^{-i}(\rho')) \leq \varepsilon_1$ for all $i \in \{0, \dots, n\}$ then*

$$C^{-1} J_{-n}^s(\rho) \leq J_{-n}^s(\rho') \leq C J_{-n}^s(\rho). \tag{3-31}$$

Proof. This is a consequence of the previous lemma and of the fact that, uniformly in ρ and $n \in \mathbb{N}$,

$$\|d_\rho F^n\| \sim J_n^u(\rho),$$
$$\|d_\rho F^{-n}\| \sim J_{-n}^s(\rho). \qquad \square$$

3D. *Regularity of the invariant splitting.* It is known for Anosov diffeomorphisms that stable and unstable distributions are in fact $C^{2-\varepsilon}$ in dimension 2; see [Hurder and Katok 1990]. For our purpose, we need to extend this result to our setting, where the hyperbolic invariant set $\mathcal{T}$ is not the full phase space, but a fractal subset of it. In fact, we will show that one can extend the stable and unstable distributions to an open neighborhood of $\mathcal{T}$ and that these extensions are $C^{1,\beta}$ for some $\beta > 0$. Actually, since what happens outside a fixed neighborhood of $\mathcal{T}$ is irrelevant (one can always use cut-offs), we will prove the following theorem which might be of independent interest.

Theorem 5. *Let us denote by $\mathcal{G}_1(U)$ the Grassmannian bundle of 1-plane in TU. There exists $\beta > 0$ and sections $E_u, E_s : U \to \mathcal{G}_1(U)$ such that:*

- *For every $\rho \in \mathcal{T}$, $E_u(\rho)$ (resp. $E_s(\rho)$) is the unstable (resp. stable) distribution at ρ.*
- *E_u and E_s have regularity $C^{1,\beta}$.*

Remark. Our proof relies on the techniques of [Hirsch and Pugh 1969]. In fact, in [Katok and Hasselblatt 1995, Chapter 19, Section 1.d] the authors show how one can obtain C^1 regularity of the map $\rho \in \mathcal{T} \mapsto E_u(\rho)$ and explain how to prove $C^{1,\beta}$ regularity. Their notion of differentiability on the set $\mathcal{T}$ (which is clearly not open in our case) relies on the existence of linear approximations. Here, we choose to show a slightly different version of this regularity by proving that $\rho \in \mathcal{T} \mapsto E_u(\rho)$ can be obtained as the restriction of a $C^{1,\beta}$ map defined in an open neighborhood of $\mathcal{T}$.

3D1. *Proof of the $C^{1,\beta}$ regularity.*

Preliminaries. We recall that $\mathcal{T}$ is an invariant hyperbolic set for F. Hence, there exists a continuous splitting of $T_{\mathcal{T}}U$ into stable and unstable spaces $\rho \in \mathcal{T} \mapsto E_s(\rho)$, $\rho \in \mathcal{T} \mapsto E_u(\rho)$. We use a continuous Riemannian metric on $T_{\mathcal{T}}U$ such that $d_\rho F$ is a contraction from $E_s(\rho) \to E_s(F(\rho))$ and expanding from $E_u(\rho) \to E_u(F(\rho))$, and making $E_u(\rho)$ and $E_s(\rho)$ orthogonal.

Let $\rho \in \mathcal{T} \mapsto e_u(\rho) \in TU$ and $\rho \in \mathcal{T} \mapsto e_s(\rho) \in TU$ be two continuous sections[3] such that, for every $\rho \in \mathcal{T}$,

- $e_u(\rho)$ spans $E_u(\rho)$,
- $e_s(\rho)$ spans $E_s(\rho)$,
- $\|e_u(\rho)\| = 1$, $\|e_s(\rho)\| = 1$.

The matrix representation of $d_\rho F$[4] in these basis is

$$d_\rho F = \begin{pmatrix} \tilde{J}^u(\rho) & 0 \\ 0 & \tilde{J}^s(\rho) \end{pmatrix},$$

with $\nu := \sup_{\rho \in \mathcal{T}} \max[(|\tilde{J}^u(\rho)|)^{-1}, |\tilde{J}^s(\rho)|] < 1$.

We can extend e_u and e_s to U to continuous functions, still denoted by e_u and e_s. Let us consider smooth vector fields v_u and v_s on U approximating e_u and e_s and a smooth Riemannian metric approximating the one considered above. By slightly modifying this vector field, we can assume that, for this new metric, $(v_u(\rho), v_s(\rho))$ is an orthonormal basis for all $\rho \in U$. In these new basis, we now write

$$d_\rho F = \begin{pmatrix} a(\rho) & b(\rho) \\ c(\rho) & d(\rho) \end{pmatrix}.$$

We assume that v_u and v_s are sufficiently close to e_u and e_s to ensure that, for some $\eta > 0$ small enough,

$$\begin{aligned} \sup_{\rho \in \mathcal{T}} \max(|b(\rho)|, |c(\rho)|) &\leq \eta, \\ \sup_{\rho \in \mathcal{T}} |d(\rho)| &\leq \nu + \eta \leq 1 - 4\eta, \\ \inf_{\rho \in \mathcal{T}} |a(\rho)| &\geq \nu^{-1} - \eta \geq 1 + 4\eta. \end{aligned}$$

We consider an open neighborhood Ω of $\mathcal{T}$ such that the following hold:

$$\begin{aligned} \sup_{\rho \in \Omega} \max(|b(\rho)|, |c(\rho)|) &\leq 2\eta, \\ \sup_{\rho \in \Omega} |d(\rho)| &\leq \nu + 2\eta \leq 1 - 3\eta, \\ \inf_{\rho \in \Omega} |a(\rho)| &\geq \nu^{-1} - 2\eta \geq 1 + 3\eta. \end{aligned}$$

Our method relies on different uses of the contraction map theorem. We state the fiber contraction theorem of [Hirsch and Pugh 1969, Section 1], which will be used further. We recall that a fixed point x_0 of a continuous map $f : X \to X$ is said to be *attractive* if, for every $x \in X$, $f^n(x) \to x_0$.

[3] Note that there is no problem of orientation in constructing such global sections. Indeed, $\mathcal{T}$ is totally disconnected and hence, one can cover $\mathcal{T}$ by a disjoint union of open sets small enough so that it is possible to construct local sections in each such sets. Since these open sets are disjoint, these local sections allow us to build a global continuous section.

[4] The definition of $\tilde{J}^{u/s}$ may differ from the one of $J_1^{u/s}$ above since we don't work a priori with the same metric.

Theorem 6 (fiber contraction theorem). *Let (X, d) be a metric space and $h : X \to X$ be a map having an attractive fixed point x_0. Let us consider Y another metric space and a family of maps $(g_x : Y \to Y)_{x \in X}$ and denote by H the map*

$$H : (x, y) \in X \times Y \mapsto (h(x), g_x(y)) \in X \times Y.$$

Assume that:

- *H is continuous.*
- *For all $x \in X$, $\limsup_{n \to +\infty} L(g_{h^n(x)}) < 1$, where $L(g_{h^n(x)})$ denotes the best Lipschitz constant for $g_{h^n(x)}$.*
- *y_0 is an attractive fixed point for g_{x_0}.*

Then (x_0, y_0) is an attractive fixed point for H.

In the following, we study the regularity of the unstable distribution. The same holds for the stable distribution by changing the roles of F^{-1} and F.

E_u is a fixed point of a contraction. By our assumption on v_u and v_s, there exists a continuous function $\lambda : U \to \mathbb{R}$ such that

$$\mathbb{R}e_u(\rho) = \mathbb{R}(v_u(\rho) + \lambda(\rho) v_s(\rho)).$$

Hence, we will represent the extension of the unstable distribution by a continuous map $\lambda : \Omega \to \mathbb{R}$. Our aim is to show that we can find λ regular enough such that, for $\rho \in \mathcal{T}$,

$$E_u(\rho) = \mathbb{R}(v_u(\rho) + \lambda(\rho) v_s(\rho)).$$

To do so, we will start by constructing λ as a fixed point of a contraction in a nice space. This contraction will be related to invariance properties of the unstable distribution.

First of all, if $\rho' = F(\rho) \in \Omega \cap F(\Omega)$, and if $v = v_u(\rho) + \lambda v_s(\rho)$, then $d_\rho F$ maps v to

$$w = (a(\rho) + \lambda b(\rho)) v_u(\rho') + (c(\rho) + \lambda d(\rho)) v_s(\rho').$$

Hence, the line of $T_\rho U$ represented by λ is sent to the line represented by $t(\rho, \lambda)$ in $T_{\rho'} U$, where

$$t(\rho, \lambda) = \frac{\lambda d(\rho) + c(\rho)}{a(\rho) + \lambda b(\rho)}. \tag{3-32}$$

Set $\Omega_1 = \Omega \cap F(\Omega)$ and let us consider a cut-off function $\chi \in C_c^\infty(\Omega_1)$ such that $0 \le \chi \le 1$ and $\chi \equiv 1$ in a neighborhood of $\mathcal{T}$. Let us introduce the complete metric space

$$X = \{\lambda \in C(\Omega : \mathbb{R}) : \|\lambda\|_\infty \le 1\}$$

and consider the map $T : X \to X$ defined, for $\lambda \in X$ and $\rho' \in \Omega$,

$$(T\lambda)(\rho') = \chi(\rho') t\big(F^{-1}(\rho'), \lambda(F^{-1}(\rho'))\big). \tag{3-33}$$

To see that this is well-defined, first note that F^{-1} is well-defined on $\operatorname{supp} \chi$ and $F^{-1}(\operatorname{supp} \chi) \subset \Omega$. It is clear that if $\lambda \in X$, then $T\lambda$ is continuous. To see that $\|T\lambda\|_\infty \le 1$, it is enough to note that if $\rho \in \Omega$

and $|\lambda| \le 1$,

$$|t(\rho,\lambda)| \le \frac{|d(\rho)|+|c(\rho)|}{|a(\rho)|-|b(\rho)|} \le \frac{1-3\eta+2\eta}{1+3\eta-2\eta} \le \frac{1-\eta}{1+\eta} < 1.$$

Let us now prove the following.

Proposition 3.15. • *T is a contraction.*

- *If λ_u denotes its unique fixed point, then, for every $\rho \in \mathcal{T}$, we have $E_u(\rho) = \mathbb{R}(v_u(\rho)+\lambda_u(\rho)v_s(\rho))$.*

Proof. Let $\lambda, \mu \in X$. If $\rho' \in \Omega \setminus \operatorname{supp}\chi$, we have $T\mu(\rho') = T\lambda(\rho') = 0$. Now assume that $\rho' \in \operatorname{supp}\chi$ and write $\rho' = F(\rho)$ with $\rho \in \Omega$. Then

$$|T\lambda(\rho') - T\mu(\rho')| = |\chi(\rho')|\,|t(\rho,\lambda(\rho)) - t(\rho,\mu(\rho))| \le |t(\rho,\lambda(\rho)) - t(\rho,\mu(\rho))|.$$

The map $\lambda \in [-1,1] \mapsto t(\rho,\lambda)$ is smooth, so we can write

$$\|T\lambda - T\mu\|_\infty \le \sup_{\rho' \in \operatorname{supp}\chi} |T\lambda(\rho') - T\mu(\rho')| \le \sup_{\Omega\times[-1,1]} |\partial_\lambda t| \times \|\lambda-\mu\|_\infty.$$

It is then enough to show that $\sup_{\Omega\times[-1,1]} |\partial_\lambda t| < 1$. For $(\rho,\lambda) \in \Omega \times [-1,1]$, we have

$$\partial_\lambda t(\rho,\lambda) = \frac{d(\rho)}{a(\rho)+\lambda b(\rho)} - b(\rho)\frac{\lambda d(\rho)+c(\rho)}{(a(\rho)+\lambda b(\rho))^2}. \tag{3-34}$$

Hence, we can control

$$|\partial_\lambda t(\rho,\lambda)| \le \frac{1-3\eta}{1+\eta} + \eta\frac{1-\eta}{(1+\eta)^2} = \kappa_\eta < 1$$

if η is small enough. This demonstrates that T is a contraction.

As a consequence, T has a unique fixed point, λ_u. We let $v(\rho) = v_u(\rho)+\lambda_u(\rho)v_s(\rho)$. We want to show that $v(\rho) \in \mathbb{R}e_u(\rho)$ for $\rho \in \mathcal{T}$ (recall that $e_u : U \to TU$ is continuous and that $e_u(\rho)$ spans $E_u(\rho)$ if $\rho \in \mathcal{T}$). Since $\chi = 1$ on $\mathcal{T}$, we see by the definition of T that, for every $\rho \in \mathcal{T}$,

$$d_\rho F(v(\rho)) \in \mathbb{R}v(F(\rho)). \tag{3-35}$$

If v_u is sufficiently close to e_u, we can find a continuous and bounded function μ such that

$$\mathbb{R}v(x) = \mathbb{R}(e_u(x)+\mu(x)e_s(x)).$$

From (3-35), if $\rho' = F(\rho) \in \mathcal{T}$,

$$d_\rho F(e_u(\rho)+\mu(\rho)e_s(\rho)) = \tilde{J}_1^u(\rho)\left(e_u(\rho')+\mu(\rho)\frac{\tilde{J}_1^s(\rho)}{\tilde{J}_1^u(\rho)}e_s(\rho')\right) \in \mathbb{R}(e_u(\rho')+\mu(\rho')e_s(\rho')).$$

This implies the equality

$$\mu(\rho') = \mu(\rho)\frac{\tilde{J}_1^s(\rho)}{\tilde{J}_1^u(\rho)}. \tag{3-36}$$

This equality implies that $\mu = 0$ on $\mathcal{T}$, and hence $v = e_u$ on $\mathcal{T}$, as expected. ☐

Remark. As long as $\rho' \in \{\chi = 1\}$, the vector field $v(\rho') = v_u(\rho') + \lambda(\rho') v_s(\rho')$ is invariant by dF. When $\rho' \in W_u(\rho) \cap \{\chi = 1\}$ for some $\rho \in \mathcal{T}$, we will see below that the direction given by $v(\rho')$ coincides with the tangent space to $W_u(\rho)$, namely $T_{\rho'} W_u(\rho) = \mathbb{R} v(\rho')$. When $\rho' \notin \bigcup_{\rho \in \mathcal{T}} W_u(\rho)$, there exists $n \in \mathbb{N}$ such that $F^{-n}(\rho') \notin \operatorname{supp} \chi$. Hence, $\lambda_u(\rho')$ is given by an explicit expression obtained by iterating the fixed-point formula.

Differentiability of $\boldsymbol{\lambda_u}$. We go on by showing that λ is C^1 by adapting the method of [Hirsch and Pugh 1969]. We now introduce the Banach space Y of bounded continuous sections $\alpha : \Omega \to T^*\Omega$. We will use the norm on $T^*\Omega$ adapted to the metric on $T\Omega$; namely, if $\alpha \in Y$,

$$\|\alpha\|_Y = \sup_{\rho \in \Omega} \sup_{v \in T_\rho \Omega, v \neq 0} \frac{|\alpha(\rho)(v)|}{\|v\|_{T_\rho \Omega}}.$$

For $\lambda \in X$, let us introduce the map $G_\lambda : Y \to Y$, defined as follows. For $\alpha \in Y$ and $\rho' \in \Omega$,

$$(G_\lambda \alpha)(\rho') = \chi(\rho')[d_\rho t(\rho, \lambda(\rho)) + \partial_\lambda t(\rho, \lambda(\rho)) \alpha(\rho)] \circ (d_\rho F)^{-1} + t(\rho, \lambda(\rho))\, d_{\rho'} \chi, \tag{3-37}$$

with $\rho = F^{-1}(\rho')$, which is well-defined since $\rho \in \Omega$ if $\rho' \in \operatorname{supp}(\chi)$. G_λ is constructed to satisfy, for $\lambda \in X$, if λ is C^1, then the following relation holds:

$$G_\lambda(d\lambda) = d(T\lambda). \tag{3-38}$$

Let us first state the key tool to show the differentiability of λ_u.

Proposition 3.16. *For every $\lambda \in X$, G_λ is a contraction with Lipschitz constant L_λ satisfying*

$$\sup_{\lambda \in X} L_\lambda < 1.$$

Before proving it, let us show how it leads us to:

Proposition 3.17. *We know λ_u is C^1.*

Proof. We use the contraction fiber theorem. Let α_u be the unique fixed point of G_{λ_u}. The map

$$H : (\lambda, \alpha) \in X \times Y \mapsto (T\lambda, G_\lambda \alpha) \in X \times Y$$

is continuous and the previous proposition shows that, for every $\lambda \in X$, $\sup_n L(G_{T^n \lambda}) < 1$. The contraction fiber theorem implies that (λ_u, α_u) is an attractive fixed point for H.

Let $\lambda \in X$ be C^1. Hence, $H^n(\lambda, d\lambda) \to (\lambda_u, \alpha_u)$. But $H^n(\lambda, d\lambda) = (T^n \lambda, \alpha_n)$, with

$$\alpha_n = G_{T^{n-1}\lambda} \circ \cdots \circ G_\lambda d\lambda.$$

It is clear that if $\lambda \in C^1$, so is $T\lambda$ and an iterative use of (3-38) implies that $\alpha_n = d(T^n \lambda)$. This shows that λ_u is C^1 and $d\lambda_u = \alpha_u$. □

Let us now prove Proposition 3.16.

Proof. Let $\lambda \in X$ and fix $\alpha, \beta \in Y$. It is of course enough to control $\|G_\lambda \alpha(\rho') - G_\lambda \beta(\rho')\|$ for $\rho' \in \operatorname{supp}(\chi)$ since both $G_\lambda \alpha$ and $G_\lambda \beta$ vanish outside. Let us fix $\rho' = F(\rho) \in \operatorname{supp}(\chi)$.

$G_\lambda \alpha(\rho') - G_\lambda \beta(\rho')$ is given by

$$\chi(\rho')\partial_\lambda t(\rho, \lambda(\rho))[\alpha(\rho) - \beta(\rho)] \circ (d_\rho F)^{-1},$$

so it is enough to control $\partial_\lambda t(\rho, \lambda(\rho))\gamma(\rho) \circ (d_\rho F)^{-1}$ for $\gamma = \alpha - \beta$. With the precise expression of $\partial_\lambda t(\rho, \lambda(\rho))$ given by (3-34), we can estimate

$$|\partial_\lambda t(\rho, \lambda(\rho))| = \frac{|d(\rho)|}{|a(\rho) + \lambda(\rho) b(\rho)|} + O_\nu(\eta) = \frac{|d(\rho)|}{|a(\rho)|} + O_\nu(\eta).$$

(By the notation $O_\nu(\eta)$, we mean that this term is bounded by $C\eta$ where C is a constant depending only on ν and (F, U)).

Moreover, we have

$$\|(d_\rho F)^{-1}\| = \max\left(\frac{1}{a(\rho)}, \frac{1}{d(\rho)}\right) + O_\nu(\eta) = \frac{1}{d(\rho)} + O_\nu(\eta).$$

Hence,

$$\|\partial_\lambda t(\rho, \lambda(\rho))\gamma(\rho) \circ (d_\rho F)^{-1}\| \le \left(\frac{1}{a(\rho)} + O_\nu(\eta)\right)\|\gamma(\rho)\| \le (\nu + O_\nu(\eta))\|\gamma\|_Y.$$

Hence, if η is small enough, the proposition is proved. □

Hölder regularity of α_u. In fact, as explained at the end of [Katok and Hasselblatt 1995, Chapter 19, Section 1.d], we can improve the C^1 regularity.

To deal with Hölder regularity of sections $\alpha : \Omega \to T^*\Omega$, we will simply evaluate the distance between $\alpha(\rho_1)$ and $\alpha(\rho_2)$ for $\rho_1, \rho_2 \in \Omega$ using the natural identification $T^*\Omega = \Omega \times (\mathbb{R}^2)^*$, where we see $\alpha(\rho_1)$ as an element of $(\mathbb{R}^2)^*$. This allows us to write $\alpha(\rho_1) - \alpha(\rho_2)$ and compute $\|\alpha(\rho_1) - \alpha(\rho_2)\|$, where $\|\cdot\|$ is a norm on $(\mathbb{R}^2)^*$. There exists $C > 0$ such that, for every $\alpha \in Y$, $\sup_{\rho \in \Omega} \|\alpha(\rho)\| \le C\|\alpha\|_Y$.

Let us introduce μ a Lipschitz constant for F^{-1} on Ω and an exponent $\beta > 0$ such that

$$\nu\mu^\beta < 1. \tag{3-39}$$

This condition is called a *bunching condition* in [Katok and Hasselblatt 1995, Chapter 19, Section1.d]. Such a β exists. We will then show the following, which finally concludes the proof of Theorem 5.

Proposition 3.18. *α_u is β-Hölder, that is to say, λ_u is $C^{1,\beta}$.*

Proof. Let us introduce

$$Y^\beta := \{\alpha \in Y : \alpha \text{ is } \beta\text{-Hölder}\}.$$

Let us consider some $\varepsilon > 0$ to be determined later and we equip Y^β with the norm

$$\|\alpha\|_{Y^\beta} = \|\alpha\|_Y + \varepsilon\|\alpha\|_\beta, \quad \|\alpha\|_\beta = \sup_{\rho_1 \ne \rho_2} \frac{\|\alpha(\rho_1) - \alpha(\rho_2)\|}{d(\rho_1, \rho_2)^\beta}.$$

The map $T : X \to X$ defined by (3-33) actually maps $X \cap C^1(\Omega, \mathbb{R})$ to $X \cap C^1(\Omega, \mathbb{R})$. Moreover, our previous results have proved that λ_u is an attractive fixed point for T in $X \cap C^1(\Omega, \mathbb{R})$, where $X \cap C^1(\Omega, \mathbb{R})$ is now equipped with the C^1 norm. For $\lambda \in X$ and $\alpha \in Y$, we can write

$$G_\lambda \alpha = \gamma_\lambda + \widetilde{G}_\lambda \alpha,$$

where, for $\rho' = F(\rho) \in \operatorname{supp} \chi$,

$$\begin{aligned} \gamma_\lambda(\rho') &= \chi(\rho') d_\rho t(\rho, \lambda(\rho)) + t(\rho, \lambda(\rho)) d_{\rho'} \chi, \\ \widetilde{G}_\lambda \alpha(\rho') &= \chi(\rho') \partial_\lambda t(\rho, \lambda(\rho)) \alpha(\rho) \circ (d_\rho F)^{-1}. \end{aligned}$$

We state here some obvious facts on γ_λ and $\widetilde{G}_\lambda$:

- $C_1 := \sup_{\lambda \in X} \|\gamma_\lambda\|_\infty < +\infty$.
- If $\lambda \in X \cap C^1(\Omega, \mathbb{R})$, γ_λ is also C^1.
- According to Proposition 3.16; $\widetilde{G}_\lambda : Y \to Y$ is a contraction with Lipschitz constant L_λ and $\nu_1 := \sup_{\lambda \in X} L_\lambda < 1$.
- If $\lambda \in X \cap C^1(\Omega, \mathbb{R})$ and α is β-Hölder, then $\widetilde{G}_\lambda \alpha$ is β-Hölder.

If $M > C_1/(1 - \nu_1)$ and $\lambda \in X \cap C^1(\Omega, \mathbb{R})$, then $\|d\lambda\|_Y \le M$ implies $\|d(T\lambda)\|_Y \le M$. Indeed, we have

$$\|d(T\lambda)\|_Y = \|G_\lambda(d\lambda)\|_Y = \|\gamma_\lambda + \widetilde{G}_\lambda d\lambda\|_Y \le C_1 + \nu_1 M \le M.$$

Hence, we introduce the complete metric space

$$X' = \{\lambda \in X \cap C^1(\Omega, \mathbb{R}) : \|d\lambda\|_Y \le M\}, \tag{3-40}$$

$T(X') \subset X'$ and λ_u is an attractive fixed point for (X', T).

We now wish to apply the fiber contraction theorem to

$$H_\beta : (\lambda, \alpha) \in X' \times Y^\beta \mapsto (T\lambda, G_\lambda \alpha) \in X' \times Y^\beta.$$

To do so, we need to show that, for every $\lambda \in X'$, $G_\lambda : Y^\beta \to Y^\beta$ is a contraction and find a uniform estimate for the Lipschitz constants.

Let's consider $\alpha_1, \alpha_2 \in Y^\beta$ and set $\gamma = \alpha_1 - \alpha_2$. We want to estimate the Y^β norm of $\widetilde{G}_\lambda \gamma$. We already know that $\|\widetilde{G}_\lambda \gamma\|_Y \le \nu_1 \|\gamma\|_Y$. Take $\rho_1', \rho_2' \in \Omega$ and let's estimate $\|\widetilde{G}_\lambda \gamma(\rho_1') - \widetilde{G}_\lambda \gamma(\rho_2')\|$. We distinguish three cases:

- $\rho_1', \rho_2' \notin \operatorname{supp} \chi$. There is nothing to write.
- $\rho_1' \in \operatorname{supp} \chi$, $\rho_2' \notin \Omega \cap F(\Omega)$. In this case, $d(\rho_1', \rho_2') \ge \delta > 0$, where δ is the distance between $\operatorname{supp} \chi$ and $(\Omega \cap F(\Omega))^c$. Hence,

$$\frac{\|\widetilde{G}_\lambda \gamma(\rho_1') - \widetilde{G}_\lambda(\rho_2')\|}{d(\rho_1', \rho_2')^\beta} \le \delta^{-\beta} \|\widetilde{G}_\lambda \gamma(\rho_1')\| \le \delta^{-\beta} C \|\widetilde{G}_\lambda \gamma\|_Y \le \nu_1 \delta^{-\beta} C \|\gamma\|_Y.$$

- $\rho_1', \rho_2' \in \Omega \cap F(\Omega)$. Let's write $\rho_1' = F(\rho_1)$, $\rho_2' = F(\rho_2)$ and note that $d(\rho_1, \rho_2) \le \mu d(\rho_1', \rho_2')$. Then

$$\begin{aligned}\widetilde{G}_\lambda \gamma(\rho_1') - \widetilde{G}_\lambda \gamma(\rho_2') = \chi(\rho_1')\partial_\lambda t(\rho_1, \lambda(\rho_1))[\gamma(\rho_1) - \gamma(\rho_2)] \circ (d_{\rho_1} F)^{-1} \quad & (*)\\ + [\chi(\rho_1')\partial_\lambda t(\rho_1, \lambda(\rho_1)) - \chi(\rho_2')\partial_\lambda t(\rho_2, \lambda(\rho_2))]\gamma(\rho_2) \circ (d_{\rho_1} F)^{-1} \quad & (**)\\ + \chi(\rho_2')\partial_\lambda t(\rho_2, \lambda(\rho_2))\gamma(\rho_2) \circ [(d_{\rho_1} F)^{-1} - (d_{\rho_2} F)^{-1}]. \quad & (***)\end{aligned}$$

To handle the last two terms $(**)$ and $(***)$, we notice that $\rho' \in \Omega \cap F(\Omega) \mapsto \chi(\rho')\partial_\lambda t(\rho, \lambda(\rho))$ is Lipschitz since λ is C^1, with Lipschitz constant which can be chosen uniform for $\lambda \in X'$. The same is true for $\rho \mapsto d_\rho F^{-1}$. Hence, there exists a uniform constant $C > 0$ such that

$$\|(**) + (***)\| \le C d(\rho_1', \rho_2')^\beta \|\gamma\|_Y.$$

To deal with the first term $(*)$, we recall that by previous computations,

$$|\chi(\rho')\partial_\lambda t(\rho, \lambda(\rho))| \cdot \|(d_\rho F)^{-1}\| \le \nu + O_\nu(\eta).$$

As consequence, we have

$$\|(*)\| \le (\nu + O_\nu(\eta))\|\gamma\|_\beta\, d(\rho_1, \rho_2)^\beta \le (\nu + O_\nu(\eta))\mu^\beta \|\gamma\|_\beta\, d(\rho_1', \rho_2')^\beta.$$

Henceforth, if η is small enough, so that $\nu_2 := (\nu + O_\nu(\eta))\mu^\beta < 1$,

$$\|H_\lambda \gamma\|_\beta \le \nu_2 \|\gamma\|_\beta + C\|\gamma\|_Y.$$

Eventually,

$$\begin{aligned}\|\widetilde{G}_\lambda \gamma\|_{Y^\beta} &\le \nu_1 \|\gamma\|_Y + \varepsilon(\nu_2\|\gamma\|_\beta + C\|\gamma\|_Y)\\ &\le (\nu_1 + \varepsilon C)\|\gamma\|_Y + \nu_2 \varepsilon \|\gamma\|_\beta \le \nu_3 \|\gamma\|_{Y^\beta},\end{aligned}$$

where $\nu_3 = \max(\nu_1 + \varepsilon C, \nu_2) < 1$ if ε is small enough.

The fiber contraction theorem applies and says that (λ_u, α_u) is an attractive fixed point for H_β. We conclude as previously: Consider $\lambda \in C^{1,\beta}(\Omega, \mathbb{R}) \cap X'$ so that $(\lambda, d\lambda) \in X' \times Y^\beta$. Then $H_\beta^n(\lambda, d\lambda) = (T^n \lambda, dT^n \lambda) \to (\lambda_u, \alpha_u)$ in $X' \times Y^\beta$, which ensures that α_u is β-Hölder. $\square$

3D2. *Regularity of the stable and unstable leaves.* Once we've extended the unstable distribution to an open neighborhood of $\mathcal{T}$, we take advantage of the fact that these distributions are one-dimensional to integrate the vector field defined by their unit vector.

We set $E_u(\rho) = \mathbb{R}(v_u(\rho) + \lambda_u(\rho) v_s(\rho))$. Recall that in a compact neighborhood of $\mathcal{T}$, the relation $d_\rho F(E_u(\rho)) = E_u(F(\rho))$ is valid due to the definition of λ_u as the fixed point of T defined in (3-33). T^*U is equipped with a smooth Riemannian metric such that dF^{-1} is a contraction on $E_u(\rho)$ for $\rho \in \mathcal{T}$ and hence, in a compact neighborhood of $\mathcal{T}$, this is also true. We can consider the vector field

$$\rho \in U \mapsto e_u(\rho),$$

where $e_u(\rho)$ is a unit vector spanning $E_u(\rho)$. By our previous result, this vector field is $C^{1,\beta}$ and if ρ lies in a sufficiently small neighborhood of $\mathcal{T}$, then $d_\rho(F^{-1})(e_u(\rho)) = \tilde{J}^u(\rho) e_u(F^{-1}(\rho))$, where $|\tilde{J}^u(\rho)| \le \nu < 1$.

We denote by $\varphi_u^t(\rho)$ the flow generated by $e_u(\rho)$ and we will show that one can identify the unstable manifolds and the flow lines of e_u in a small neighborhood of $\mathcal{T}$.

Proposition 3.19. *There exists t_0 such that, for every $\rho \in \mathcal{T}$, we have $\{\varphi_u^t(\rho) : |t| \le t_0\} \subset W_u(\rho)$.*

Proof. Consider t_0 sufficiently small that $|\tilde{J}^u(\varphi_u^t(\rho))| \le \nu < 1$ for $\rho \in \mathcal{T}$, $t \in [-t_0, t_0]$. For $(t, \rho) \in \mathbb{R} \times U$, set $\mu(t, \rho) = \int_0^t \tilde{J}^u(\varphi_u^s(\rho))\, ds$ and we claim that for t_0 small enough, if $|t| \le t_0$,

$$F^{-1}(\varphi_u^t(\rho)) = \varphi_u^{\mu(t,\rho)}(F^{-1}(\rho)).$$

Indeed, in $t = 0$, both are equal to $F^{-1}(\rho)$ and a quick computation shows that both satisfy the ODE

$$\frac{d}{dt} Y(t) = J^u(\varphi_u^t(\rho)) e_u(Y(t)).$$

As a consequence, by induction, we see that one can write, for $n \in \mathbb{N}$,

$$F^{-n}(\varphi_u^t(\rho)) = \varphi_u^{\mu_n(t,\rho)}(F^{-n}(\rho)),$$

where μ_n is defined by induction by $\mu_{n+1}(t, \rho) = \mu(\mu_n(t, \rho), F^{-n}(\rho))$. Hence, if $|t| \le t_0$ and $\rho \in \mathcal{T}$, we see that $\mu_n(t, \rho)$ stays in $[-t_0, t_0]$ and moreover $|\mu_n(t, \rho)| \le \nu^n |t|$. We then see that if $|t| \le t_0$ and $\rho \in \mathcal{T}$,

$$d\big(F^{-n}(\varphi_u^t(\rho)), F^{-n}(\rho)\big) = d\big(\varphi_u^{\mu_n(t,\rho)}(F^{-n}(\rho)), F^{-n}(\rho)\big) \le C|\mu_n(t, \rho)| \le C\nu^n.$$

This shows that $\varphi_u^t(\rho)$ belongs to the global unstable manifold at ρ, and hence, if t_0 is small enough, $\varphi_u^t(\rho)$ belongs to the local manifold $W_u(\rho)$ and t_0 can be chosen uniformly with respect to $\rho \in \mathcal{T}$. □

Since the regularity of the unstable distributions implies the same regularity for the flow φ_u^t (see Lemma A.1 in the Appendix), we deduce that, up to reducing the size of the local unstable manifolds, these local unstable manifolds $W_u(\rho)$ depend $C^{1,\beta}$ on the base point $\rho \in \mathcal{T}$. We'll also use this proposition to show the same regularity for holonomy maps. Suppose that ε_0 is small enough. We know that if $\rho_1, \rho_2 \in \mathcal{T}$ satisfy $d(\rho_1, \rho_2) \le \varepsilon_0$, then $W_u(\rho_2) \cap W_s(\rho_1)$ consists of exactly one point. Let's denote it by $H^u_{\rho_1}(\rho_2)$.

Finally, we define the holonomy map

$$H^u_{\rho_1,\rho_2} : \rho_3 \in W_s(\rho_2) \cap \mathcal{T} \mapsto H^u_{\rho_1}(\rho_3) \in W_s(\rho_1) \cap \mathcal{T}.$$

Lemma 3.20. *If ε_0 is small enough, for every $\rho_1 \in \mathcal{T}$, the map*

$$H^u_{\rho_1} : \mathcal{T} \cap B(\rho_1, \varepsilon_0) \to W_s(\rho_1) \cap \mathcal{T}$$

is the restriction of a map $\widetilde{H}^u_{\rho_1} : B(\rho_1, \varepsilon_0) \to W_s(\rho_1)$ which is $C^{1,\beta}$.

Proof. Let $\rho_1 \in \mathcal{T}$. As in the proof of Lemma 3.13, consider a smooth chart $\kappa : U_1 \to V_1 \subset \mathbb{R}^2$, $\rho_1 \in U_1$, $0 \in V_1$ such that:

- $\kappa(\rho_1) = (0, 0)$.
- $\kappa(W_s(\rho_1) \cap U_1) = \{(0, s) : s \in \mathbb{R}\} \cap V_1$.
- $\kappa(W_u(\rho_1) \cap U_1) = \{(u, 0) : u \in \mathbb{R}\} \cap V_1$.
- $d_{\rho_1}\kappa(e_u(\rho_1)) = (1, 0)$.

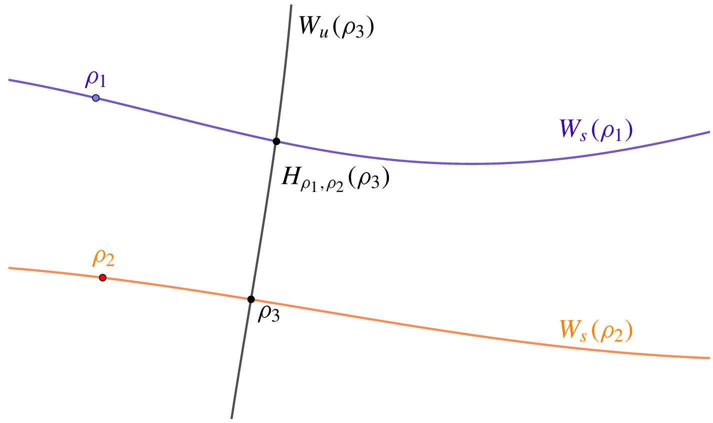

Figure 6. The holonomy map.

We now work in this chart V_1 and denote by $\Phi^t = \kappa \circ \varphi_u^t \circ \kappa^{-1}$ the flow in this chart, well-defined for t small enough. Consider the map

$$\psi(u, s) = \Phi^u(0, s);$$

ψ is $C^{1,\beta}$ and $d_0\psi = \mathrm{I}_2$. By the inverse function theorem, ψ is a local diffeomorphism between neighborhoods of 0:

$$\psi : V_2 \to V_2'.$$

Since $d_{(u,s)}(\psi^{-1}) = (d_{\psi^{-1}(u,s)}\psi)^{-1}$, we know ψ^{-1} is $C^{1,\beta}$. We now consider

$$\kappa_0 = \psi^{-1} \circ \kappa : \kappa^{-1}(V_2) := U_2 \to V_2'$$

and observe that:

- $\kappa_0(W_s(\rho_1) \cap U_2) = \{(0, s), s \in \mathbb{R}\} \cap V_2'$.
- $\kappa_0 \circ \varphi_u^t \circ \kappa_0^{-1}(u, s)) = (u + t, s)$. In other words κ_0 rectifies the unstable manifolds.

Armed with these facts, we define

$$\widetilde{H}^u_{\rho_1} : U_2 \to W_s(\rho_1), \qquad \widetilde{H}^u_{\rho_1} = \kappa_0^{-1} \circ \pi_s \circ \kappa_0,$$

where $\pi_s(u, s) = (0, s)$. $\widetilde{H}^u_{\rho_1}$ is $C^{1,\beta}$. We assume that $B(0, \varepsilon_0) \subset U_1$. Let us check that $\widetilde{H}^u_{\rho_1}$ extends the holonomy map in $B(\rho_1, \varepsilon_0)$ (if ε_0 is small enough). Let $\rho_2 \in \mathcal{T} \cap B(\rho_1, \varepsilon_0)$ and let $\rho_2' = \widetilde{H}^u_{\rho_1}(\rho_2)$. By the definition of $\widetilde{H}^u_{\rho_1}$, ρ_2' can be written $\rho_2' = \varphi_u^t(\rho_1)$ and hence, if ε_0 is small enough, $\rho_2' \in W_u(\rho_1)$. Since, $\rho_2' \in W_s(\rho_2)$, we see that $\rho_2' = H^u_{\rho_1}(\rho_2)$. $\square$

Note that by compactness, ε_0 can be chosen uniformly in $\rho_1 \in \mathcal{T}$ and the $C^{1,\beta}$ norms of $\widetilde{H}^u_{\rho_1}$ are uniform. As a corollary, we get the following:

Corollary 3.21. *Suppose that ε_0 is small enough. Then, the holonomy maps, defined for $\rho_1, \rho_2 \in \mathcal{T}$ with $d(\rho_1, \rho_2) \le \varepsilon_0$,*

$$H^u_{\rho_1,\rho_2} : W_s(\rho_2) \cap \mathcal{T} \to W_s(\rho_1) \cap \mathcal{T}$$

are the restrictions of $C^{1,\beta}$: $\widetilde{H}^u_{\rho_1,\rho_2} : W_s(\rho_1) \to W_s(\rho_2)$, with $C^{1,\beta}$ norms uniform in ρ_1, ρ_2. See Figure 6.

3E. ***Adapted charts.*** We construct charts in which the unstable manifolds are close to horizontal lines. These charts will be used at different places and their existence relies on the $C^{1+\beta}$ regularity of the unstable distribution.

Weak version. We start with a weak version of these charts.

Lemma 3.22. *Suppose that $C > 0$ is a fixed global constant and ε_0 is chosen small enough. For every $\rho_0 \in \mathcal{T}$, there exists a canonical transformation*

$$\kappa_0 : U'_{\rho_0} \to V'_{\rho_0} \subset \mathbb{R}^2$$

satisfying (*we denote by* (y, η) *the variable in* $\mathbb{R}^2$):

(1) $B(\rho_0, C\varepsilon_0) \subset U'_{\rho_0}$.

(2) $\kappa_0(\rho_0) = 0,\ d_{\rho_0}\kappa_0(E_u(\rho_0)) = \mathbb{R} \times \{0\},\ d_{\rho_0}\kappa_0(E_s(x)) = \{0\} \times \mathbb{R}$.

(3) *The image of the unstable manifold $W_u(\rho_0) \cap U'_{\rho_0}$ is exactly $\{(y, 0) : y \in \mathbb{R}\} \cap V'_{\rho_0}$.*

Moreover, for every N, the C^N norms of κ_0 are bounded uniformly with respect to $\rho_0 \in \mathcal{T}$.

Remark. The difference with the charts used in the proof of Lemma 3.13 is that we require κ_0 to be a smooth canonical transformation.

Proof. $W_u(\rho_0)$ is a C^∞ manifold; hence there exists a C^∞ defining function η defined in a neighborhood ρ_0; namely, $d_{\rho_0}\eta \neq 0$ and $W_u(\rho_0) = \{\eta = 0\}$ locally near ρ_0. Darboux's theorem gives a function y defined in a neighborhood of ρ_0 such that (y, η) forms a system of symplectic coordinates. We can assume that $y(\rho_0) = 0$. If $\kappa(\rho) = (y, \eta)$, the third point is satisfied by assumption on η and we need to ensure that $d_{\rho_0}\kappa(E_s(\rho_0)) = \{0\} \times \mathbb{R}$ by modifying η in a symplectic way.

Assume that $d_{\rho_0}\kappa(E_s(\rho_0)) = \mathbb{R}^t(a, 1)$. The symplectic matrix

$$A = \begin{pmatrix} 1 & -a \\ 0 & 1 \end{pmatrix}$$

maps the basis $({}^t(1, 0), {}^t(a, 1))$ to the canonical basis of $\mathbb{R}^2$ and we can set $\kappa_0 := A \circ \kappa$, which is the required canonical transformation, defined in a small neighborhood U'_{ρ_0} of ρ_0.

We can ensure that $B(\rho_0, C\varepsilon_0) \subset U'_{\rho_0}$ for ε_0 small enough and the uniformity of the C^N norms of κ thanks to the compactness of $\mathcal{T}$ and the fact that the unstable distribution depends continuously on $\rho_0 \in \mathcal{T}$. □

Straightened version. We now straighten the unstable manifolds in a stronger version of the previous charts. The construction and the use of these charts is similar to [Dyatlov et al. 2022, Lemma 2.3].

Lemma 3.23. *Suppose that ε_0 is chosen small enough. For every $\rho_0 \in \mathcal{T}$ there exists a canonical transformation*

$$\kappa = \kappa_{\rho_0} : U_{\rho_0} \subset U \to V_{\rho_0} \subset \mathbb{R}^2$$

satisfying (*we denote by* (y, η) *the variable in* $\mathbb{R}^2$):

(1) $B(\rho_0, 2\varepsilon_0) \subset U_{\rho_0}$.

(2) $\kappa(\rho_0) = 0,\ d_{\rho_0}\kappa(E_u(\rho_0)) = \mathbb{R} \times \{0\},\ d_{\rho_0}\kappa(E_s(\rho_0)) = \{0\} \times \mathbb{R}$.

(3) *The images of the unstable manifolds $W_u(\rho)$, $\rho \in U_{\rho_0} \cap \mathcal{T}$, are described by*

$$\kappa(W_u(\rho) \cap U_{\rho_0}) = \{(y, g(y, \zeta(\rho))) : y \in \Omega\}, \tag{3-41}$$

where $\Omega \subset \mathbb{R}$ is an open set, $\zeta : U_{\rho_0} \to \mathbb{R}$ is $C^{1+\beta}$, $g : \Omega \times I \to \mathbb{R}$ is $C^{1+\beta}$ (where I is a neighborhood of $\zeta(U_{\rho_0})$) and they satisfy:

(a) *ζ is constant on the unstable manifolds.*

(b) $\zeta(\rho_0) = 0,\ g(y, 0) = 0$.

(c) $g(0, \zeta) = \zeta$.

(d) $\partial_\zeta g(y, 0) = 1$.

The derivatives of κ_{ρ_0} and the $C^{1+\beta}$ norms of g, ζ are bounded uniformly in ρ_0.

Remark. The most important condition, which will be used later on, is the last one: it makes the unstable manifolds very close to horizontal lines. The model situation we expect is when the unstable distribution is constant and horizontal.

Proof. Around a point $\rho_0 \in \mathcal{T}$, we work in the charts given by Lemma 3.22: $\kappa_0 : U'_{\rho_0} \to V'_{\rho_0}$. We recall that the unstable distribution is given by the restriction of a $C^{1+\beta}$ vector field e_u. If U'_{ρ_0} is a sufficiently small neighborhood of ρ_0, we can write, for $\rho \in U'_{\rho_0}$,

$$d_\rho\kappa_0(e_u(\rho)) \in \mathbb{R}\tilde{e}_u(\rho), \quad \text{with } \tilde{e}_u(\rho) = {}^t(1, f_0(\rho)), \tag{3-42}$$

where $f_0 : U'_{\rho_0} \to \mathbb{R}$ is a $C^{1+\beta}$ function which is nothing but the slope of the unstable direction in the chart κ_0. In the (y, η)-variable, we still write $f_0(\rho) = f_0(y, \eta)$ and we observe that due to the assumption on κ_0, we have

$$f_0(y, 0) = 0, \quad (y, 0) \in V'_{\rho_0}.$$

We consider $\Phi^t(y, \eta)$, the flow generated by the vector field $\tilde{e}_u$. Due to the form of $\tilde{e}_u$, we can write

$$\Phi^t(y, \eta) = (y + t, Z^t(y, \eta)).$$

The reparametrization made in (3-42) does not change the flow lines of the vector field $(\kappa_0)_* e_u$. In particular, by virtue of Proposition 3.19, they coincide locally with the unstable manifolds. More precisely, if we set

$$g_0(y, \eta) := Z^y(0, \eta)$$

(see Figure 7) then, for $(0, \eta) = \kappa_0(\rho) \in \kappa_0(\mathcal{T} \cap W_s(\rho_0))$,

$$\kappa_0(W_u(\rho)) \cap \{|y| < y_0\} = \{(y, g_0(y, \eta)) : |y| < y_0\}$$

for some y_0 small enough (which can be chosen uniformly in ρ_0). To define ζ, we go back up the flow: Suppose that $\rho \in U'_{\rho_0}$ and write $\kappa_0(\rho) = (y, \eta)$ and assume $|y| < y_0$. We set

$$\zeta(\rho) := Z^{-y}(y, \eta).$$

To say it differently, $\kappa_0(W_u(\rho)$ intersects the axis $\{y = 0\}$ at $(0, \zeta(\rho))$.

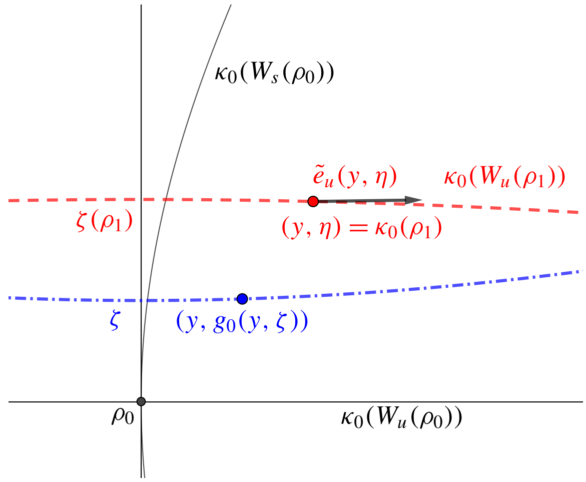

Figure 7. The definitions of g_0 and ζ use the flow generated by $\tilde{e}_u$.

We know ζ and g_0 are $C^{1+\beta}$, their $C^{1+\beta}$ norms depend uniformly on ρ_0 and they satisfy:

- By definition, ζ is constant on the flow lines, and hence, on the unstable manifolds $W_u(\rho)$ if $\rho \in \mathcal{T} \cap U'_{\rho_0} \cap \{|y| < y_0\}$.
- $\zeta(\rho_0) = 0$.
- Since $f_0(y, 0) = 0$, we have $Z^y(0, 0) = 0$ and hence $g_0(y, 0) = 0$.
- Since $Z^0(0, \eta) = \eta$, we have $g_0(0, \eta) = \eta$.

However, at this stage, the last condition ($\partial_\zeta g_0(y, 0) = 1$) is not satisfied by g_0 and we need to modify the chart. To do so, we'll make use of the following lemma, which is proved in Section A2 in the Appendix.

Lemma 3.24. *The map $y \in \{|y| < y_0\} \mapsto \partial_\eta f_0(y, 0)$ is smooth, with C^N norms bounded uniformly in ρ_0.*

We first show that this lemma implies that $y \in \{|y| < y_0\} \mapsto \partial_\eta g_0(y, 0)$ is smooth. Indeed, due to the $C^{1+\beta}$ regularity of E_u, $(t, y, \eta) \mapsto Z^t(y, \eta)$ is C^1 and satisfies

$$\frac{d}{dt}\partial_\eta Z^t(y, \eta) = \partial_\eta f_0(y + t, Z^t(y, \eta)).$$

Setting $(y, \eta) = (0, 0)$, we have

$$\frac{d}{dt}\partial_\eta Z^t(0, 0) = \partial_\eta f_0(t, 0).$$

This exactly says that $y \mapsto \partial_\eta g_0(y, 0)$ is C^1 and has $\partial_\eta f_0(y, 0)$ as derivative with respect to y and hence $y \mapsto \partial_\eta g_0(y, 0)$ is smooth, as required.

Due to the relation $g_0(0, \eta) = \eta$, we have $\partial_\eta g_0(0, 0) = 1$. As a consequence, if y_0 is small enough, we can assume that $\partial_\eta g_0(y, 0) > 0$ for $|y| < y_0$ and consider the smooth diffeomorphism defined in $\{|y| < y_0\}$

$$\psi : y \mapsto \int_0^y \partial_\eta g_0(s, 0)\, ds.$$

We then use the canonical transformation

$$\Psi : (y, \eta) \mapsto \left(\psi(y), \frac{\eta}{\psi'(y)} \right).$$

We finally consider the chart $\kappa_{\rho_0} = \Psi \circ \kappa_0$ defined in $U_{\rho_0} = U'_{\rho_0} \cap \{|y| < y_0\}$ and if ε_0 is small enough, we can ensure that $B(\rho_0, 2\varepsilon_0) \subset U_{\rho_0}$. In this chart, the graph of $g_0(\cdot\,, \zeta)$ is sent to the graph of the function

$$g : y \in \Omega := \psi((-y_0, y_0)) \mapsto \frac{g_0(\psi^{-1}(y), \zeta)}{\psi'(\psi^{-1}(y))}.$$

We eventually check that:

- $g(y, 0) = 0$ since $g_0(y, 0) = 0$.
- $g(0, \zeta) = \zeta$ since $\psi(0) = 0,\ \psi'(0) = 1$ and $g(0, \zeta) = \zeta$.
- $\partial_\eta g(y, 0) = 1$.
- The $C^{1+\beta}$ norm of g is bounded uniformly in ρ_0.
- The C^N norms of κ_{ρ_0} are bounded uniformly in ρ_0. ☐

4. Construction of a refined quantum partition

We start the proof of Theorem 1. We consider $T = T(h) \in I_{0^+}(Y \times Y, F')$ a semiclassical Fourier integral operator associated with F, microlocally unitary in a neighborhood of $\mathcal{T}$, and a symbol $\alpha \in S_{0^+}(U)$. We want to show a bound for the spectral radius of $M(h) = T(h)\,\mathrm{Op}_h(\alpha)$, independent of h.

4A. ***Numerology.*** We'll use the standard fact

$$\|M^n\|_{L^2 \to L^2} \le \rho \quad \Longrightarrow \quad \rho_{\text{spec}}(M) \le \rho^{1/n}.$$

The trivial lemma which follows reduces the theorem to the study of $\|M^n\|$ with $n = n(h) \sim \delta|\log h|$.

Lemma 4.1. *Let $\delta > 0$ and $N(h) \in \mathbb{N}$ satisfy $N(h) \sim \delta|\log h|$. Suppose that there exists $h_0 > 0$ and $\gamma > 0$ such that,*

$$\text{for all } 0 < h < h_0, \quad \|M(h)^{N(h)}\| \le h^\gamma \|\alpha\|_\infty^{N(h)}. \tag{4-1}$$

Then, for every $\varepsilon > 0$, there exists h_ε such that, for $h \le h_\varepsilon$,

$$\rho_{\text{spec}}(M(h)) \le e^{-\gamma/\delta+\varepsilon}\|\alpha\|_\infty.$$

Proof. It suffices to observe that under the assumption (4-1), we have $\rho_{\text{spec}}(M(h)) \le e^{\gamma \log h/N(h)}\|\alpha\|_\infty$ and use the equivalence for $N(h)$. ☐

Remark. If we use the bound $\|M\| \le \|\alpha\|_\infty + O(h^{1/2-\varepsilon})$, one get the obvious bound $\|M^N\| \le \|\alpha\|_\infty^N(1 + o(1))$. Hence, (4-1) is a decay bound.

The proof of Theorem 1 is then reduced to the proof of the following proposition.

Proposition 4.2. *There exists $\delta > 0$, a family of integer $N(h) \sim \delta|\log(h)|$ and $\gamma > 0$ such that, for h small enough, (4-1) holds.*

Actually, this proposition is enough to show Corollary 1 concerning perturbed operators, by virtue of:

Corollary 4.3. *Suppose that $R(h) : L^2(Y) \to L^2(Y)$ is a family of bounded operators such that $R(h) = O(h^\eta)$ for some $\eta > 0$. Then, there exists $\gamma' = \gamma'(\gamma, \eta)$ such that, for h small enough,*

$$\|(M(h) + R(h))^{N(h)}\| \le h^{\gamma'} \|\alpha\|_\infty^{N(h)}.$$

Proof. We write

$$(M+R)^N = M^N + \sum_{\substack{\varepsilon\in\{0,1\}^N \\ \varepsilon\neq(1,\dots,1)}} (\varepsilon_1 M + (1-\varepsilon_1)R)\cdots(\varepsilon_N M + (1-\varepsilon_N)R).$$

Using this, we can estimate

$$\begin{aligned}
\|(M+R)^N\| &\le h^\gamma \|\alpha\|_\infty^N + ((\|M\| + \|R\|)^N - \|M\|^N) \\
&\le h^\gamma \|\alpha\|_\infty^N + N\|R\|(\|M\| + \|R\|)^{N-1} \\
&\le h^\gamma \|\alpha\|_\infty^N + C|\log h| h^\eta \|\alpha\|_\infty^{N-1}(1 + O(h^\eta)) \\
&= O((h^\gamma + h^{\eta-})\|\alpha\|_\infty^N).
\end{aligned}$$

This gives the desired bound for any $\gamma' < \min(\gamma, \eta)$. ☐

Actually, the precise value of $N(h)$ we'll use is rather explicit and we now describe it. We set

$$\mathfrak{b} = \frac{1}{1+\beta}, \tag{4-2}$$

where β is the one appearing in Theorem 5 concerning the regularity of the unstable distribution. We now choose $\delta_0 \in \left(0, \frac{1}{2}\right)$ such that

$$\mathfrak{b} + \delta_0 < 1. \tag{4-3}$$

For instance, let us set

$$\delta_0 = \frac{1-\mathfrak{b}}{2} = \frac{\beta}{2(1+\beta)}.$$

Recalling the definitions of the exponent $\lambda_0 \le \lambda_1$ in (3-10) and (3-11), we introduce the notation

$$N(h) = N_0(h) + N_1(h), \quad N_0(h) = \left\lceil \frac{\delta_0}{\lambda_1} |\log(h)| \right\rceil, \quad N_1(h) = \left\lceil \frac{1}{\lambda_0} |\log(h)| \right\rceil, \tag{4-4}$$

where $N_0(h)$ (resp. $N_1(h)$) corresponds to a short (resp. long) logarithmic time. We will omit the dependence on h in the following.

To be complete with the numerology, we introduce another number $\tau < 1$ such that

$$\mathfrak{b} < \tau < 1 \quad \text{and} \quad \delta_0 \frac{\lambda_0}{\lambda_1} + \tau > 1. \tag{4-5}$$

The meaning of these conditions will be clear in the core of the proof and we will indicate where they are used. For instance, we set

$$\tau = 1 - \frac{\lambda_0}{\lambda_1} \frac{1-\mathfrak{b}}{4}. \tag{4-6}$$

An important remark. If two operators $M_1(h)$ and $M_2(h)$ are equal modulo $O(h^\infty)$, this is also the case for $M_1(h)^{N(h)}$ and $M_2(h)^{N(h)}$ as long as

- $N(h) = O(\log h)$.
- $M_1(h), M_2(h) = O(h^{-K})$ for some K.

This will be widely used in the following. In particular, recall that we work with operators acting on $L^2(Y)$ but these operators take the form $M_1(h) = \Psi_Y M_2(h) \Psi_Y$, where $\Psi_Y \in C_c^\infty(Y, [0,1])$ and $M_2(h)$ is a bounded operator on $\bigoplus_{j=1}^J L^2(\mathbb{R})$ such that $M_2(h) = \Psi_Y M_2(h) \Psi_Y + O(h^\infty)_{L^2}$. As a consequence, modulo $O(h^\infty)$, it is enough to focus on $M_2(h)^{N(h)}$. For this reason, from now on and even if we keep the same notation, we work with

$$M(h) = T(h) \operatorname{Op}_h(\alpha) : \bigoplus_{j=1}^J L^2(\mathbb{R}) \to \bigoplus_{j=1}^J L^2(\mathbb{R}),$$

where $T(h) = (T_{ij}(h))$, with $T_{ij} \in I_{0+}(\mathbb{R} \times \mathbb{R}, F'_{ij})$ and

$$\operatorname{Op}_h(\alpha) = \operatorname{Diag}(\operatorname{Op}_h(\alpha_1), \dots, \operatorname{Op}_h(\alpha_J)).$$

4B. ***Microlocal partition of unity and notations.*** We consider some $\varepsilon_0 > 0$, which is supposed small enough to satisfy all the assumptions which will appear in the following.

We consider a cover of $\mathcal{T}$ by a finite number of balls of radius ε_0,

$$\mathcal{T} \subset \bigcup_{q=1}^Q B(\rho_q, \varepsilon_0), \quad \rho_q \in \mathcal{T},$$

and we assume that for all $q \in \{1, \dots, Q\}$, there exist $j_q, l_q, m_q \in \{1, \dots, J\}$ such that

$$B(\rho_q, 2\varepsilon_0) \subset \tilde{A}_{j_q l_q} \cap \widetilde{D}_{m_q j_q} \subset U_{j_q}.$$

We also assume that T is microlocally unitary in $B(\rho_q, 4\varepsilon_0)$. We then let

$$\mathcal{V}_q = B(\rho_q, 2\varepsilon_0). \tag{4-7}$$

See Figure 8.

Remark. In the case of obstacle scattering, with obstacles satisfying the noneclipse condition, it is possible to choose a simple partition of unity, related to the coding of the trapped set according to the sequence of obstacles hit by a trajectory. Indeed, due to a result of [Morita 1991], there is a homeomorphism between the trapped set and the admissible — that is, two consecutive obstacles are different — sequences of obstacles. As a consequence, if the obstacles are numbered from 1 to J, we can partition the trapped set by open subsets $U_{\vec{\alpha}}$ indexed by

$$\{\vec{\alpha} = (\alpha_{-N}, \dots, \alpha_N) \in \{1, \dots, J\}^{2N+1} : \alpha_{i+1} \neq \alpha_i\}.$$

The diameter of such partition goes to 0 as N goes to $+\infty$ and we could get the required partition $(\mathcal{V}_q)_q$, with the additional property of being disjoint open subsets of U. This would simplify the study in this particular setting.

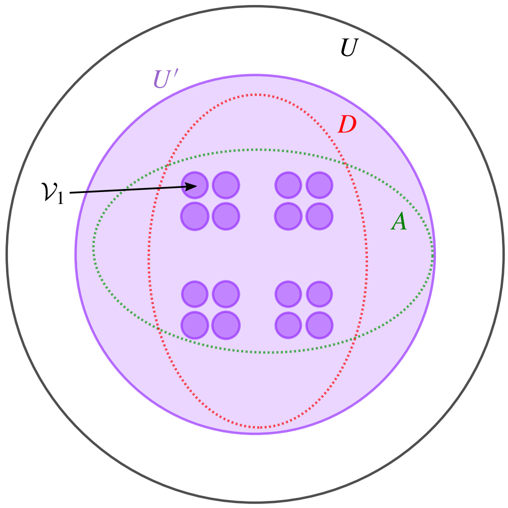

Figure 8. The partition $(\mathcal{V}_q)_{q\in\mathcal{A}_\infty}$ is made by small neighborhoods of $\mathcal{T}$ (small purple disks) and a big open set included in U'.

We complete this cover with

$$\mathcal{V}_\infty = U' \setminus \bigcup_{q=1}^{Q} \overline{B(\rho_q, \varepsilon_0)}. \tag{4-8}$$

$U' \Subset U$ is an open set such that $\mathrm{WF}_h(M) \Subset U' \times U'$. We denote by U'_j the component of U' inside U_j.

We let $\mathcal{A} = \{1, \dots, Q\}$ and $\mathcal{A}_\infty = \mathcal{A} \cup \{\infty\}$.

We then consider a partition of unity associated with the cover $\mathcal{V}_1, \dots, \mathcal{V}_Q, \mathcal{V}_\infty$, namely a family of smooth functions $\chi_q \in C_c^\infty(U)$ for $q \in \mathcal{A}_\infty$ such that:

- $\operatorname{supp} \chi_q \subset \mathcal{V}_q$.
- $0 \le \chi_q \le 1$.
- $1 = \sum_{q\in\mathcal{A}_\infty} \chi_q$ in $\bigcup_{q\in\mathcal{A}_\infty} \mathcal{V}_q$.

More precisely, if $q \in \mathcal{A}$, $\chi_q \in C^\infty(U_{j_q})$ and, for every $j \in \{1, \dots, J\}$, there exists $b_j \in C_c^\infty(U_j)$ such that on U'_j, then $1 = b_j + \sum_{q\in\mathcal{A}, j_q=j} \chi_q$. Thus, $\chi_\infty = \sum_{j=1}^{J} b_j$.

We can then quantize these symbols so as to get a pseudodifferential partition of unity. More precisely, to respect the matrix structure, we may write this quantization in a diagonal operator-valued matrix, still denoted by Op_h:

- For $q \in \mathcal{A}$, $A_q = \mathrm{Op}_h(\chi_q)$ is the diagonal matrix $\mathrm{Diag}(0, \dots, \mathrm{Op}_h(\chi_q), 0, \dots, 0)$, where the block $\mathrm{Op}_h(\chi_q)$ is in the j_q-th position.
- $\mathrm{Op}_h(\chi_\infty) = \mathrm{Diag}(\mathrm{Op}_h(b_1), \dots, \mathrm{Op}_h(b_J))$.

The family $(A_q)_{q\in\mathcal{A}_\infty}$ satisfies the properties

$$\sum_{q\in\mathcal{A}_\infty} A_q = \mathrm{Id} \text{ microlocally in } U' \quad \text{for all } q \in \mathcal{A}_\infty, \quad \|A_q\| \le 1 + O(h^{1/2}). \tag{4-9}$$

Since $M = \sum_{q \in \mathcal{A}_\infty} M A_q + O(h^\infty)$, we may write

$$M^n = \sum_{\boldsymbol{q} \in \mathcal{A}_\infty^n} U_{\boldsymbol{q}} + O(h^\infty),$$

where, for $\boldsymbol{q} = q_0 \cdots q_{n-1} \in \mathcal{A}_\infty^n$,

$$U_{\boldsymbol{q}} := M A_{q_{n-1}} \cdots M A_{q_0}. \tag{4-10}$$

For $\boldsymbol{q} = q_0 \cdots q_{n-1} \in \mathcal{A}_\infty^n$, we also define a family of refined neighborhoods, forming a refined cover of $\mathcal{T}$,

$$\mathcal{V}_{\boldsymbol{q}}^- = \bigcap_{i=0}^{n-1} F^{-i}(\mathcal{V}_{q_i}), \quad \mathcal{V}_{\boldsymbol{q}}^+ = F^n(\mathcal{V}_{\boldsymbol{q}}^-) = \bigcap_{i=0}^{n-1} F^{n-i}(\mathcal{V}_{q_i}). \tag{4-11}$$

This definition implies that a point $\rho \in \mathcal{V}_{\boldsymbol{q}}^-$ lies in $\mathcal{V}_{q_i}$ at time i (i.e., $F^i(\rho) \in \mathcal{V}_{q_i}$) for $0 \le i \le n-1$ and a point $\rho \in \mathcal{V}_{\boldsymbol{q}}^+$ lies in $\mathcal{V}_{q_{n-i}}$ at time $-i$ for $1 \le i \le n$. Roughly speaking, we expect that each operator $U_{\boldsymbol{q}}$ acts from $\mathcal{V}_{\boldsymbol{q}}^-$ to $\mathcal{V}_{\boldsymbol{q}}^+$ and is negligible (in some sense to be specified later on) elsewhere. Combining (4-9) and the bound on M, the following bound is valid (for any $\varepsilon > 0$):

$$\|U_{\boldsymbol{q}}\|_{L^2 \to L^2} \le (\|\alpha\|_\infty + O(h^{1/2-\varepsilon}))^n. \tag{4-12}$$

As soon as $|n| \le C_0 |\log h|$, we have $\|U_{\boldsymbol{q}}\|_{L^2 \to L^2} \le C \|\alpha\|_\infty^n$ for some C depending on C_0 and a finite number of seminorms of α.

Reduction to words in $\mathcal{A}$. We can find a uniform $T_0 \in \mathbb{N}$ such that if $\rho \in \mathcal{V}_\infty$, there exists $k \in \{-T_0, \dots, T_0\}$ such that $F^k(\rho)$ "falls" in the hole. By standard properties of the Fourier integral operators, each component $(M^{T_0})_{ij}$ of M^{T_0} is a Fourier integral operator associated with the component $(F^{T_0})_{ij}$ of F^{T_0}. In particular, $\mathrm{WF}_h'(M^{T_0}) \subset \mathrm{Gr}'(F^{T_0})$.

Let us study $M^{2T_0 + N(h)}$. If $\boldsymbol{q} = q_0 \cdots q_{N-1} \in \mathcal{A}_\infty^N$ and if there exists an index $i \in \{0, \dots, N-1\}$ such that $q_i = \infty$, one can isolate this index i and trap A_{q_i} between two Fourier integral operators M_1, M_2, belonging to a finite family of FIO associated with F^{T_0}, so that we can write

$$M^{T_0} U_{\boldsymbol{q}} M^{T_0} = B_1 M_1 A_\infty M_2 B_2,$$

where B_1, B_2 satisfy the L^2-bound

$$\|B_1\| \times \|B_2\| \le (\|\alpha\|_\infty + O(h^{1/4}))^{N-1} = O(h^{-K})$$

for some integer K. Since

$$\mathrm{WF}_h'(M_1 A_\infty M_2) \subset \{(F^{T_0}(\rho), F^{-T_0}(\rho)) : \rho \in \mathrm{WF}_h(A_\infty)\} = \varnothing,$$

we have $M_1 A_\infty M_2 = O(h^\infty)$, with constants that can be chosen independent of $\boldsymbol{q}$. Hence, the same is true for $M^{T_0} U_{\boldsymbol{q}} M^{T_0}$. $|\mathcal{A}^N|$ is bounded by a negative power of h. So, we can write

$$M^{N+2T_0} = \sum_{\boldsymbol{q} \in \mathcal{A}_\infty^N} M^{T_0} U_{\boldsymbol{q}} M^{T_0} = \sum_{\boldsymbol{q} \in \mathcal{A}^N} M^{T_0} U_{\boldsymbol{q}} M^{T_0} + O(h^\infty) = M^{T_0} \Bigg(\sum_{\boldsymbol{q} \in \mathcal{A}^N} U_{\boldsymbol{q}} \Bigg) M^{T_0} + O(h^\infty).$$

We can then replace M by

$$\mathfrak{M} = \sum_{q\in\mathcal{A}} MA_q = M(\mathrm{Id} - A_\infty) + O(h^\infty)_{L^2\to L^2}. \tag{4-13}$$

The decay bound

$$\|\mathfrak{M}(h)^{N(h)}\| \le h^\gamma \|\alpha\|_\infty^{N(h)} \tag{4-14}$$

will imply the required decay bound (4-1) for M with $N(h)$ replaced by $N(h)+2T_0$. We are hence reduced to proving the decay bound (4-14).

4C. *Local Jacobian.*

A first definition. Following [Dyatlov et al. 2022], we introduce local unstable and stable Jacobians and we then state several properties. For $n \in \mathbb{N}^*$ and $\boldsymbol{q} \in \mathcal{A}^n$, let us define its local stable and unstable Jacobian:

$$J_{\boldsymbol{q}}^- := \inf_{\rho\in\mathcal{T}\cap\mathcal{V}_{\boldsymbol{q}}^-} J_n^u(\rho), \qquad J_{\boldsymbol{q}}^+ := \inf_{\rho\in\mathcal{T}\cap\mathcal{V}_{\boldsymbol{q}}^+} J_{-n}^s(\rho). \tag{4-15}$$

By the chain rule, we have, for $\rho \in \mathcal{T}\cap\mathcal{V}_{\boldsymbol{q}}^-$,

$$J_n^u(\rho) = \prod_{i=0}^{n-1} J_1^u(F^i(\rho)).$$

A similar formula is true for $\rho \in \mathcal{T}\cap\mathcal{V}_{\boldsymbol{q}}^+$:

$$J_{-n}^s(\rho) = \prod_{i=0}^{n-1} (J_1^s(F^{i-n}(\rho)))^{-1} = \prod_{i=0}^{n-1} J_{-1}^s(F^{-i}(\rho)).$$

Hence, we've got the basic estimates

$$\mathcal{T}\cap\mathcal{V}_{\boldsymbol{q}}^- \ne \varnothing \quad\Longrightarrow\quad e^{\lambda_0 n} \le J_{\boldsymbol{q}}^- \le e^{\lambda_1 n}, \tag{4-16}$$

$$\mathcal{T}\cap\mathcal{V}_{\boldsymbol{q}}^+ \ne \varnothing \quad\Longrightarrow\quad e^{\lambda_0 n} \le J_{\boldsymbol{q}}^+ \le e^{\lambda_1 n}. \tag{4-17}$$

If $\boldsymbol{q} = q_0\cdots q_{n-1}$ and $\boldsymbol{q}_- = q_0\cdots q_{n-2}$, then $\mathcal{V}_{\boldsymbol{q}}^- \subset \mathcal{V}_{\boldsymbol{q}_-}^-$ and thus

$$J_{\boldsymbol{q}}^- \ge e^{\lambda_0} J_{\boldsymbol{q}_-}^-. \tag{4-18}$$

Similarly, if $\boldsymbol{q}_+ = q_1\cdots q_{n-1}$, then $\mathcal{V}_{\boldsymbol{q}}^+ \subset \mathcal{V}_{\boldsymbol{q}_+}^+$ and

$$J_{\boldsymbol{q}}^+ \ge e^{\lambda_0} J_{\boldsymbol{q}_+}^+. \tag{4-19}$$

As a consequence of Corollary 3.14, if ε_0 is small enough, the local stable and unstable Jacobians give the expansion rate of the flow at every point of $\mathcal{T}\cap\mathcal{V}_{\boldsymbol{q}}^\pm$. If $\mathcal{T}\cap\mathcal{V}_{\boldsymbol{q}}^\pm \ne \varnothing$,

$$\text{for all } \rho \in \mathcal{T}\cap\mathcal{V}_{\boldsymbol{q}}^-, \quad J_n^u(\rho) \sim J_{\boldsymbol{q}}^-, \tag{4-20}$$

$$\text{for all } \rho \in \mathcal{T}\cap\mathcal{V}_{\boldsymbol{q}}^+, \quad J_{-n}^s(\rho) \sim J_{\boldsymbol{q}}^+. \tag{4-21}$$

This definition is slightly unsatisfactory since $J_{\boldsymbol{q}}^\pm = +\infty$ as soon as $\mathcal{V}_{\boldsymbol{q}}^\pm \cap \mathcal{T} = \varnothing$. However, when $\mathcal{V}_{\boldsymbol{q}}^\pm \ne \varnothing$, this set can still stay relevant. For this purpose, we will give a definition of local stable and unstable Jacobian for such words with help of the shadowing lemma [Katok and Hasselblatt 1995, Section 18.1].

Enlarged definition. Let $n \in \mathbb{N}$ and $\boldsymbol{q} = q_0 \cdots q_{n-1} \in \mathcal{A}^n$. We focus on $\mathcal{V}_{\boldsymbol{q}}^-$, with the case of $\mathcal{V}_{\boldsymbol{q}}^+$ handled similarly by considering F^{-1} instead of F.

If $\mathcal{V}_{\boldsymbol{q}}^- \cap \mathcal{T} \neq \varnothing$, we keep the definition given in (4-15). Assume now that $\mathcal{V}_{\boldsymbol{q}}^- \neq \varnothing$ but $\mathcal{V}_{\boldsymbol{q}}^- \cap \mathcal{T} = \varnothing$. Fix $\rho \in \mathcal{V}_{\boldsymbol{q}}^-$. By definition of $\mathcal{V}_{q_i}$, for $i \in \{0, \dots, n-1\}$, we have $d(\rho_i, F^i(\rho)) \leq 2\varepsilon_0$. Hence,

$$d(F(\rho_i), \rho_{i+1}) \leq d(F(\rho_i), F^{i+1}(\rho)) + d(F^{i+1}(\rho), \rho_{i+1}) \leq C\varepsilon_0$$

for a constant C only depending on F. That is to say, $(\rho_0, \dots, \rho_{n-1})$ is a $C\varepsilon_0$ pseudo-orbit. Assume that $\delta_0 > 0$ is a small fixed parameter. By virtue of the shadowing lemma, if ε_0 is sufficiently small and $(\rho_0, \dots, \rho_{n-1})$ is δ_0 shadowed by an orbit of F, then there exists $\rho' \in \mathcal{T}$ such that, for $i \in \{0, \dots, n-1\}$, $d(\rho_i, F(\rho')) \leq \delta_0$. Consequently, $d(F^i(\rho), F^i(\rho')) \leq \delta_0 + C\varepsilon_0$. If ρ_2 is another point in $\mathcal{V}_{\boldsymbol{q}}^-$, for $i = 0, \dots, n-1$, $d(F^i(\rho_2), F^i(\rho')) \leq 2\varepsilon_0 + C\varepsilon_0 + \delta_0$. For convenience, set $\varepsilon_2 = 2\varepsilon_0 + \delta_0 + C\varepsilon_0$ and note that ε_2 can be arbitrarily small depending on ε_0. As a consequence, we have proven the following:

Lemma 4.4. *If $\mathcal{V}_{\boldsymbol{q}}^- \neq \varnothing$, then there exists $\rho' \in \mathcal{T}$ such that, for all $i \in \{0, \dots, n-1\}$ and $\rho \in \mathcal{V}_{\boldsymbol{q}}^-$, $d(F^i(\rho), F^i(\rho')) \leq \varepsilon_2$.*

Fix any ρ' satisfying the conclusions of this lemma and we arbitrarily set

$$J_{\boldsymbol{q}}^- = J_n^u(\rho'). \tag{4-22}$$

If ρ_1' is another point satisfying this conclusion, we have $d(F^i(\rho'), F^i(\rho_1')) \leq 2\varepsilon_2$ for $i \in \{0, \dots, n-1\}$ and by virtue of Corollary 3.14,

$$J_n^u(\rho') \sim J_n^u(\rho_1').$$

Hence, up to global multiplicative constants, the definition of this unstable Jacobian is independent of the choice of ρ'. Notice that if $\mathcal{V}_{\boldsymbol{q}}^- \cap \mathcal{T} \neq \varnothing$, any $\rho' \in \mathcal{T} \cap \mathcal{V}_{\boldsymbol{q}}^-$ satisfies the conclusions of Lemma 4.4 and $J_{\boldsymbol{q}}^- \sim J_n^u(\rho')$.

To define $J_{\boldsymbol{q}}^+$, we can argue similarly and show that there exists ρ' satisfying $d(F^i(\rho'), F^i(\rho)) \leq \varepsilon_2$ for $i \in \{-n, \dots, -1\}$ and $\rho \in \mathcal{V}_{\boldsymbol{q}}^+$. We can assume that this is the same ε_2 as before and we set $J_{\boldsymbol{q}}^+ = J_{-n}^s(\rho')$ for any ρ'.

Behavior of the local Jacobian. See Figure 9. The following three lemmas are crucial to understand the behavior of the evolution of points in the sets $\mathcal{V}_{\boldsymbol{q}}^\pm$. The first one gives estimates to handle these quantities.

Lemma 4.5. *Let $n \in \mathbb{N}$ and $\boldsymbol{q}$, $\boldsymbol{p}$ in $\mathcal{A}^n$. If ε_0 is chosen small enough, then the following hold:*

(1) $\mathcal{V}_{\boldsymbol{q}}^+ \neq \varnothing \iff \mathcal{V}_{\boldsymbol{q}}^- \neq \varnothing$ *and in that case* $J_{\boldsymbol{q}}^- \sim J_{\boldsymbol{q}}^+$.

(2) *If two propagated neighborhoods intersect, the local Jacobians are comparable:*

$$\mathcal{V}_{\boldsymbol{q}}^\pm \cap \mathcal{V}_{\boldsymbol{p}}^\pm \neq \varnothing \quad \Longrightarrow \quad J_{\boldsymbol{q}}^\pm \sim J_{\boldsymbol{p}}^\pm. \tag{4-23}$$

(3) *If $\boldsymbol{q}$ can be written as the concatenation of $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ of lengths n_1 and n_2 such that $n_1 + n_2 = n$ and if $\mathcal{V}_{\boldsymbol{q}}^\pm \neq \varnothing$, then*

$$J_{\boldsymbol{q}}^\pm \sim J_{\boldsymbol{q}_1}^\pm J_{\boldsymbol{q}_2}^\pm. \tag{4-24}$$

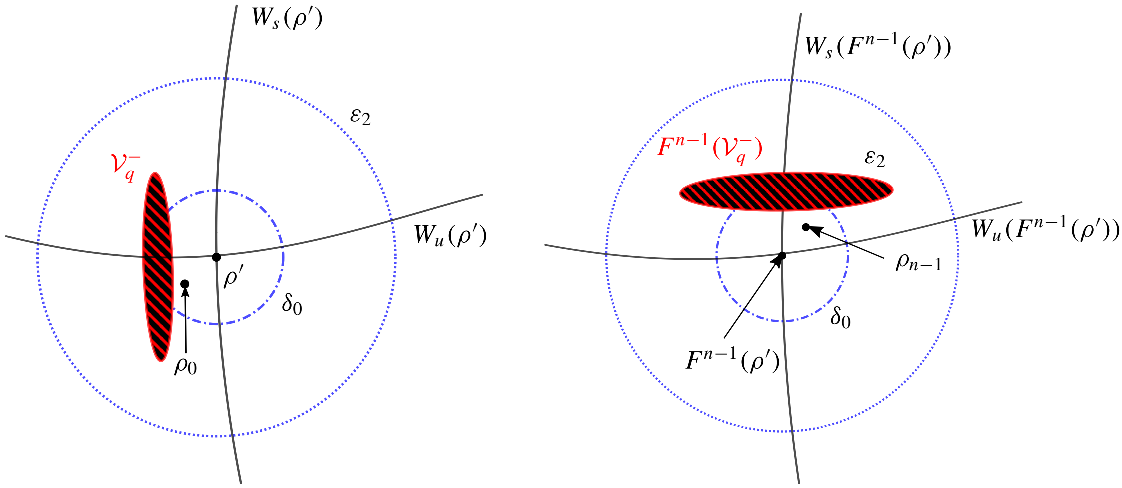

Figure 9. Evolution of the set $\mathcal{V}_{\boldsymbol{q}}^-$ (the red hatched set) at time 0 and $n-1$. The points ρ_i, $F^i(\rho')$ are represented at these times, so as the balls $B(F^i(\rho'), \varepsilon_2)$ and $B(F^i(\rho'), \delta_0)$ (their boundaries are the blue dotted lines). We've also represented the stable (resp. unstable) manifold at $F^i(\rho')$ to show the directions in which F contracts (resp. expands).

Notation. The constants in $\sim$ are independent of ρ and n. They depend on F but also on the partition $(\mathcal{V}_q)_q$. In the following, we'll be lead to use constants with the same kind of dependence. These constants will be allowed to depend also on the partition of unity $(\chi_q)_q$ and on M. Constants with such dependence will be called *global* constants.

Proof. (1) The equivalence is obvious. From the fact that F is a volume-preserving canonical transformation, we have, for some $C > 0$,

$$\text{for all } \rho \in \mathcal{T}, \text{ for all } n \in \mathbb{N}, \quad C^{-1} \le J_n^u(\rho) J_n^s(\rho) \le C,$$

and we write $J_n^u(\rho) \sim J_n^s(\rho)^{-1}$. From $F^{-n} \circ F^n(\rho) = \rho$, we also get $J_n^s(\rho)^{-1} = J_{-n}^s(F^n(\rho))$. Eventually, if $\rho' \in \mathcal{T}$ satisfies $d(F^i(\rho), F^i(\rho') \le \varepsilon_2$ for $i \in \{0, \dots, n-1\}$ and $\rho \in \mathcal{V}_{\boldsymbol{q}}^-$, $F^n(\rho') = \rho^+$ satisfies $d(F^i(\rho), F^i(\rho^+)) \le \varepsilon_2$ for $i \in \{-n, \dots, -1\}$ and $\rho \in \mathcal{V}_{\boldsymbol{q}}^+$. Hence

$$J_{\boldsymbol{q}}^+ \sim J_{-n}^s(\rho^+) \sim J_n^u(\rho') \sim J_{\boldsymbol{q}}^-.$$

Thanks to this first point, it is enough to show the remaining point only for $-$.

(2) Pick $\rho_{\boldsymbol{q}} \in \mathcal{T}$ (resp. $\rho_{\boldsymbol{p}}$) satisfying the conclusions of Lemma 4.4 for $\mathcal{V}_{\boldsymbol{q}}^-$ (resp. $\mathcal{V}_{\boldsymbol{p}}^-$). We have $d(F^i(\rho_{\boldsymbol{q}}), F^i(\rho_{\boldsymbol{p}})) \le 2\varepsilon_2$ and hence, by virtue of Corollary 3.14, $J_n^u(\rho_{\boldsymbol{q}}) \sim J_n^u(\rho_{\boldsymbol{p}})$. This gives (2).

(3) Pick $\rho \in \mathcal{T}$ satisfying the conclusions of Lemma 4.4 for $\mathcal{V}_{\boldsymbol{q}}^-$. By the chain rule, we have $J_n^u(\rho) = J_{n_2}^u(F^{n_1}(\rho)) J_{n_1}^u(\rho)$. Note that

$$\mathcal{V}_{\boldsymbol{q}}^- = \mathcal{V}_{\boldsymbol{q}_1}^- \cap F^{-n_1}(\mathcal{V}_{\boldsymbol{q}_2}^-).$$

Hence, ρ satisfies the conclusions of Lemma 4.4 for $\boldsymbol{q}_1$ with ε_2 and the same is true for $F^{n_1}(\rho)$ and $\boldsymbol{q}_2$. It follows that $J_{\boldsymbol{q}_1}^- \sim J_{n_1}^u(\rho)$ and $J_{\boldsymbol{q}_2}^- \sim J_{n_2}^u(F^{n_1}(\rho))$. This gives (3). $\square$

Remark. The first point of the previous lemma shows that we could consider only one of the two quantities. Nevertheless, we prefer keeping track of it. The reason is that a priori J^+ and J^- support two different kind of information: $J_{\boldsymbol{q}}^+$ controls the growth of F^n, whereas $J_{\boldsymbol{q}}^-$ controls the growth of F^{-n}. The fact that the two dynamics (in the past and in the future) have similar behaviors is a consequence of the fact that F is volume-preserving.

The next lemmas relate the local Jacobian to the expansion rates of the flow in the $\mathcal{V}_{\boldsymbol{q}}^{\pm}$. It will be important in our semiclassical study of operators microlocally supported in $\mathcal{V}_{\boldsymbol{q}}^{\pm}$.

Lemma 4.6 (control of expansion rate by unstable Jacobian). *If ε_0 is small enough, there exists a global constant $C > 0$ satisfying the following inequalities*:

For every $n \in \mathbb{N}^*$ and $\boldsymbol{q} \in \mathcal{A}^n$ such that $\mathcal{V}_{\boldsymbol{q}}^- \neq \varnothing$ we have

$$\sup_{\rho \in \mathcal{V}_{\boldsymbol{q}}^-} \|d_\rho F^n\| \leq C J_{\boldsymbol{q}}^-, \tag{4-25}$$

$$\sup_{\rho \in \mathcal{V}_{\boldsymbol{q}}^+} \|d_\rho F^{-n}\| \leq C J_{\boldsymbol{q}}^+. \tag{4-26}$$

Proof. This is a consequence of (3-18). Indeed, if $\mathcal{V}_{\boldsymbol{q}}^- \neq \varnothing$ and if $\rho' \in \mathcal{T}$ satisfies the conclusions of Lemma 4.4, then for every $\rho \in \mathcal{V}_{\boldsymbol{q}}^-$, $\|d_\rho F^n\| \leq C J_n^u(\rho)$ with C a global constant depending only on ε_2. □

This third lemma emphasizes that $\mathcal{V}_{\boldsymbol{q}}^-$ lies in a small neighborhood of a stable manifold and $\mathcal{V}_{\boldsymbol{q}}^+$ lies in a small neighborhood of an unstable manifold, with the size of this neighborhood controlled by the local Jacobian. It is a direct consequence of Lemma 3.13.

Lemma 4.7 (localization of the $\mathcal{V}_{\boldsymbol{q}}^{\pm}$). *There exists a global constant $C > 0$ such that for all $n \in \mathbb{N}$ and $\boldsymbol{q} \in \mathcal{A}^n$*:

(1) *If $\mathcal{V}_{\boldsymbol{q}}^- \neq \varnothing$ and $\rho' \in \mathcal{T}$ satisfies the conclusion of Lemma 4.4, then, for all $\rho \in \mathcal{V}_{\boldsymbol{q}}^-$*,

$$d(\rho, W_s(\rho')) \leq \frac{C}{J_{\boldsymbol{q}}^-}. \tag{4-27}$$

(2) *If $\mathcal{V}_{\boldsymbol{q}}^+ \neq \varnothing$ and $\rho' \in \mathcal{T}$ satisfies the conclusion of Lemma 4.4 in the future* (*namely*, $d(F^i(\rho), F^i(\rho')) \leq \varepsilon_2$ *for all $\rho \in \mathcal{V}_{\boldsymbol{q}}^+$ and $i \in \{-n, \dots, -1\}$*), *then, for all $\rho \in \mathcal{V}_{\boldsymbol{q}}^+$*,

$$d(\rho, W_u(\rho')) \leq \frac{C}{J_{\boldsymbol{q}}^+}. \tag{4-28}$$

4D. ***Propagation up to local Ehrenfest time.*** In this section, we show that under some control of the local Jacobian defined above, one can handle the operators $U_{\boldsymbol{q}}$ and prove the existence of symbols $a_{\boldsymbol{q}}^{\pm}$ (in exotic classes S_δ) such that

$$U_{\boldsymbol{q}} = \mathrm{Op}_h(a_{\boldsymbol{q}}^+) T^{|\boldsymbol{q}|} + O(h^\infty), \tag{4-29}$$

$$U_{\boldsymbol{q}} = T^{|\boldsymbol{q}|} \mathrm{Op}_h(a_{\boldsymbol{q}}^-) + O(h^\infty), \tag{4-30}$$

with symbols $a_{\boldsymbol{q}}^{\pm}$ supported in $\mathcal{V}_{\boldsymbol{q}}^{\pm}$. We recall that $U_{\boldsymbol{q}} = M A_{q_{n-1}} \cdots M A_{q_0}$, with $M = T\,\mathrm{Op}_h(\alpha)$. Let us state the precise statement we will prove.

Proposition 4.8. *Fix $0 < \delta < \delta_1 < \frac{1}{2}$ and $C_0 > 0$.*

(1) *For every $n \in \mathbb{N}$ and for all $\boldsymbol{q} \in \mathcal{A}^n$ satisfying*

$$J_{\boldsymbol{q}}^+ \leq C_0 h^{-\delta}, \tag{4-31}$$

there exists $a_{\boldsymbol{q}}^+ \in \|\alpha\|_\infty^n S_{\delta_1}^{\mathrm{comp}}$ such that

$$U_{\boldsymbol{q}} = \mathrm{Op}_h(a_{\boldsymbol{q}}^+) T^n + O(h^\infty)_{L^2 \to L^2}, \tag{4-32}$$

$$\operatorname{supp} a_{\boldsymbol{q}}^+ \subset \mathcal{V}_{\boldsymbol{q}}^+. \tag{4-33}$$

(2) *For every $n \in \mathbb{N}$ and for all $\boldsymbol{q} \in \mathcal{A}^n$ satisfying*

$$J_{\boldsymbol{q}}^- \leq C_0 h^{-\delta}, \tag{4-34}$$

there exists $a_{\boldsymbol{q}}^- \in \|\alpha\|_\infty^n S_{\delta_1}^{\mathrm{comp}}$ such that

$$U_{\boldsymbol{q}} = T^n \, \mathrm{Op}_h(a_{\boldsymbol{q}}^-) + O(h^\infty)_{L^2 \to L^2}, \tag{4-35}$$

$$\operatorname{supp} a_{\boldsymbol{q}}^- \subset \mathcal{V}_{\boldsymbol{q}}^-. \tag{4-36}$$

Remark. • The implied constants appearing in the $O(h^\infty)$ are quasiglobal; they have the same dependence as global constants but depend also on C_0, δ, δ_1. What is important is that they are independent of n and $\boldsymbol{q}$ as soon as the assumption (4-31) is satisfied.

• (4-31) implies that $\mathcal{V}_{\boldsymbol{q}}^+ \neq \varnothing$. In particular, if $\boldsymbol{q}$ satisfies this assumption, there exists a sequence $(i_0, \dots, i_n)$ such that, for all $p \in \{0, \dots, n-1\}$, $\mathcal{V}_{q_p} \subset \widetilde{D}_{i_{p+1}, i_p} \subset U_{i_p}$.

• In fact, $\operatorname{supp} a_{\boldsymbol{q}}^+ \subset F(\mathcal{V}_{q_{n-1}}) \subset U_{i_n}$. Hence, the operator $\mathrm{Op}_h(a_{\boldsymbol{q}}^+)$ acting on $\bigoplus_{i=1}^J L^2(\mathbb{R})$ is the diagonal matrix $\mathrm{Diag}(0, \dots, \mathrm{Op}_h(a_{\boldsymbol{q}}^+), \dots, 0)$.

• The symbol $a_{\boldsymbol{q}}^+$ has an asymptotic expansion in power of h. The principal symbol is given by

$$(a_{\boldsymbol{q}}^+)_0 = \prod_{p=1}^{n} a_{q_{n-p}} \circ F^{-p}, \tag{4-37}$$

where $a_q = \chi_q \times \alpha$. Note that if the functions $a_{q_{n-p}} \circ F^{-p}$ are not necessarily well-defined, the product is well-defined thanks to the assumptions on the supports of χ_q, namely $\operatorname{supp} \chi_q \Subset \mathcal{V}_q$. Indeed, such a symbol can be constructed inductively as the n-th term b_n of the sequence of functions $b_1 = a_{q_0} \circ F^{-1}$ and b_{i+1} is obtained from a_i by

$$b_{i+1} = (a_{q_i} \times a_i) \circ F^{-1}.$$

If we assume that $\operatorname{supp} b_i \Subset \mathcal{V}_{q_0 \cdots q_{i-1}}^+$, then $\operatorname{supp}(a_{q_i} \times b_i) \Subset F^{-1}(\mathcal{V}_{q_0 \cdots q_i}^+)$. This property allows us to define b_{i+1} and $\operatorname{supp} b_{i+1} \Subset \mathcal{V}_{q_0 \cdots q_i}^+$.

• The same holds for $a_{\boldsymbol{q}}^-$ with principal symbol

$$(a_{\boldsymbol{q}}^-)_0 = \prod_{p=0}^{n-1} a_{q_p} \circ F^p. \tag{4-38}$$

• Our proof follows the sketch of proof of [Dyatlov et al. 2022, Section 5] and [Rivière 2010, Section 7].

In the end of this section, we focus on proving this proposition. We only prove the first point. The second point can be proved similarly by using the same techniques.

4D1. *Iterative construction of the symbols.* Let us start by a lemma combining the precise versions of the expansion of the Moyal product (Lemma 3.3) and of Egorov theorem (Proposition 3.8). This lemma is the key ingredient for the iterative formulas below.

Lemma 4.9. *Let $q \in \mathcal{A}$ and let $a \in S_{\delta_1}^{\mathrm{comp}}$ such that* $\operatorname{supp} a \Subset U_j$ *for some* $j \in \{1, \dots, J\}$. *Then, there exists a family of differential operators $L_{k,q}$ of order $2k$, with smooth coefficients compactly supported in $\mathcal{V}_q$, such that, for every $N \in \mathbb{N}$, we have the expansion*

$$M A_q \operatorname{Op}_h(a) = \operatorname{Op}_h\left(\sum_{k=0}^{N-1} h^k (L_{k,q} a) \circ F^{-1}\right) T + O(\|a\|_{C^{2N+15}} h^N)_{L^2 \to L^2}. \tag{4-39}$$

Moreover, one has $L_{0,q} = \chi_q \times \alpha := a_q$.

Remark. • Again, since $\operatorname{supp} a \subset U_j$, $\operatorname{Op}_h(a)$ is a diagonal matrix with only one nonzero block equal to $\operatorname{Op}_h(a)$.

• Recall that we've supposed that $\mathcal{V}_q \subset \widetilde{D}_{m_q j_q}$. As a consequence, the symbols

$$a_1^{(k)} := L_{k,q} a \circ F^{-1}$$

are equal to $L_{k,q} a \circ (F_{m_q j_q})^{-1}$ and are supported in U_{m_q}; $\operatorname{Op}_h(a_1^{(k)})$ is still a diagonal matrix.

Proof. Let us first work at the order of operators $L^2(\mathbb{R}) \to L^2(\mathbb{R})$ and let us study

$$M_{m_q j_q} \operatorname{Op}_h(\chi_q) \operatorname{Op}_h(a) = T_{m_q j_q} \operatorname{Op}_h(\alpha_{j_q}) \operatorname{Op}_h(\chi_q) \operatorname{Op}_h(a).$$

Using Lemma 3.3, we write

$$\operatorname{Op}_h(\chi_q) \operatorname{Op}_h(a) = \operatorname{Op}_h\left(\sum_{k=0}^{N-1} \frac{i^k h^k}{k!} A(D)^k (\chi_q \otimes a)|_{\rho=\rho_1=\rho_2}\right) + O(h^N \|\chi_q \otimes a\|_{C^{2N+13}}),$$

the principal term of the expansion being $\chi_q a$. Set $a_{q,k}(\rho) = A(D)^k (\chi_q \otimes a)|_{\rho=\rho_1=\rho_2}$ and use Lemma 3.3 to write

$$\operatorname{Op}_h(\alpha_{j_q}) \operatorname{Op}_h(\chi_q) \operatorname{Op}_h(a) = \sum_{k_1+k_2<N} \frac{i^{k_1+k_2} h^{k_1+k_2}}{k_1!\, k_2!} \operatorname{Op}_h(A(D)^{k_2} (\alpha_{j_q} \otimes a_{q,k_1})|_{\rho=\rho_1=\rho_2}) + O(h^N \|a\|_{C^{2N+13}}).$$

The principal term in the expansion is $\alpha_{j_q} \chi_q a$. We note that

$$a \mapsto \sum_{k_1+k_2=k} A(D)^{k_2} (\alpha_{j_q} \otimes a_{q,k_1})|_{\rho=\rho_1=\rho_2}$$

is a differential operator of order $2k$. Using the precise version of Egorov theorem in Lemma 3.10, we see that, for any b with $\operatorname{supp}(b) \subset \mathcal{V}_q$,

$$T_{m_q j_q} \operatorname{Op}_h(b) = \operatorname{Op}_h\left(b \circ (F_{m_q j_q})^{-1} + \sum_{k=1}^{N-1} h^k (D_k b) \circ (F_{m_q j_q})^{-1}\right) + O(h^N \|b\|_{C^{2N+15}}),$$

where D_k are differential of order $2k$ compactly supported in $\mathcal{V}_q$. Applying this to the previous expansion, we see that we can write

$$T_{m_q j_q} \operatorname{Op}_h(\alpha_{j_q}) \operatorname{Op}_h(\chi_q) \operatorname{Op}_h(a) = \operatorname{Op}_h\left((\alpha_{j_q}\chi_q a)\circ F^{-1} + \sum_{k=1}^{N-1} k^k (L_{k,q}a)\circ F^{-1}\right) + O(h^N \|a\|_{C^{2N+15}}).$$

We now come to the entire matrix operator. Note that the matrix $M\operatorname{Op}_h(\chi_q)\operatorname{Op}_h(a)$ is of the form

$$\begin{pmatrix} 0 & \cdots & M_{1j_q}\operatorname{Op}_h(\chi_q) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & M_{Jj_q}\operatorname{Op}_h(\chi_q) & \cdots & 0 \end{pmatrix} \operatorname{Op}_h(a).$$

Recall that $\mathrm{WF}_h(\operatorname{Op}_h(\chi_q)) \subset \widetilde{D}_{m_q j_q}$ and $\mathrm{WF}_h'(M_{m_q j_q}) \subset \mathrm{Gr}'(F_{m_q j_q})$. Hence, for $m \neq m_q$, $M_{mj_q}\operatorname{Op}_h(\chi_q) = O(h^\infty)$ and the previous matrix can be written

$$\begin{pmatrix} 0 & \cdots & O(h^\infty) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & M_{m_q j_q}\operatorname{Op}_h(\chi_q) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & O(h^\infty) & \cdots & 0 \end{pmatrix} \operatorname{Op}_h(a) = \begin{pmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & M_{m_q j_q}\operatorname{Op}_h(\chi_q)\operatorname{Op}_h(a) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} + O(h^\infty)\|\operatorname{Op}_h(a)\|_{L^2}.$$

With constant in $O(h^\infty)$ depending on χ_q, M and $\|\operatorname{Op}_h(a)\|_{L^2\to L^2} = O(\|a\|_{C^8})$. Let's write

$$a_1^{(k)} = L_{k,q}a\circ F^{-1}$$

and observe that $\operatorname{supp}(a_1^{(k)}) \subset F(\operatorname{supp}\chi_q) \Subset \tilde{A}_{m_q j_q}$. Consider a cut-off function $\tilde{\chi}_q$ such that $\tilde{\chi}_q \equiv 1$ in a neighborhood of $F(\operatorname{supp}\chi_q)$ and $\operatorname{supp}\tilde{\chi}_q \subset \tilde{A}_{m_q j_q}$. Using Lemma 3.3 and the support properties of $\tilde{\chi}_q$, one has

$$\operatorname{Op}_h(a_1^{(k)}) = \operatorname{Op}_h(a_1^{(k)})\operatorname{Op}_h(\tilde{\chi}_q) + O(h^{N-k}\|a_1^{(k)}\|_{C^{2(N-k)+13}}) = \operatorname{Op}_h(a_1^{(k)})\operatorname{Op}_h(\tilde{\chi}_q) + O(h^{N-k}\|a\|_{C^{2N+13}}).$$

Then, one can write $\operatorname{Op}_h(a_1^{(k)})T$ on the form

$$\begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \vdots & \vdots \\ \operatorname{Op}_h(a_1^{(k)})\operatorname{Op}_h(\tilde{\chi}_q)T_{m_q 1} & \cdots & \operatorname{Op}_h(a_1^{(k)})\operatorname{Op}_h(\tilde{\chi}_q)T_{m_q J} \\ \vdots & \vdots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} + O(h^{N-k}\|a\|_{C^{2N+13}})$$

and, for $j \neq j_q$, $\operatorname{Op}_h(\tilde{\chi}_q)T_{m_q j} = O(h^\infty)$. We can conclude that

$$\operatorname{Op}_h(a_1^{(k)})T = \begin{pmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ 0 & \cdots & \operatorname{Op}_h(a_1^{(k)})\operatorname{Op}_h(\tilde{\chi}_q)T_{m_q j_q} & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix} + O(h^\infty)\|\operatorname{Op}_h(a_1^{(k)})\|_{L^2\to L^2} + O(h^{N-k}\|a\|_{C^{2N+13}})$$

$$= \begin{pmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ 0 & \cdots & \mathrm{Op}_h(a_1^{(k)})T_{m_q j_q} & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix} + O(h^{N-k}\|a\|_{C^{2N+13}}).$$

Combining this with the version obtained with $M_{m_q j_q}$, we get (4-39). □

Let us now start the iterative construction of the symbols. Fix $N \in \mathbb{N}$ which can be taken arbitrarily large. Recall that we want to write

$$U_{\boldsymbol{q}} = \mathrm{Op}_h(a_{\boldsymbol{q}}^+)T^{|\boldsymbol{q}|} + O(h^\infty)_{L^2 \to L^2}. \tag{4-40}$$

Note $U_r = U_{q_0 \cdots q_{r-1}}$. We want to write

$$U_r = \mathrm{Op}_h\left(\sum_{k=0}^{N-1} h^k a_r^{(k)}\right)T^r + R_r^{(N)}. \tag{4-41}$$

We start by writing

$$U_1 = \mathrm{Op}_h\left(\sum_{k=0}^{N-1} h^k a_1^{(k)}\right)T + R_1^{(N)}, \tag{4-42}$$

which is possible by virtue of (4-39). To pass from U_r to U_{r+1}, we have the relation

$$U_{r+1} = MA_{q_r}U_r = \sum_{k=0}^{N-1} h^k MA_{q_r}\,\mathrm{Op}_h(a_r^{(k)})T^r + MA_{q_r}R_r^{(N)}.$$

So, we will construct inductively our symbols by setting

$$a_{r+1}^{(k)} = \sum_{p=0}^{k}(L_{p,q_r}a_r^{(k-p)}) \circ (F_{i_{r+1},i_r})^{-1}, \tag{4-43}$$

$$R_{r+1}^{(N)} = MA_{q_r}R_r^{(N)} + \sum_{k=0}^{N-1} O(\|a_r^{(k)}\|_{C^{2(N-k)+15}}). \tag{4-44}$$

The O encompasses the remainder terms in (4-39). The constants in the O only depend on M and the χ_q, $q \in \mathcal{A}$, but not on $\boldsymbol{q}$.

To make this construction work, we will have to prove that the symbols $a_r^{(k)}$ lie in a good symbol class $S_{\delta_1}^{\mathrm{comp}}$.

Before reaching this step, let us just note that by induction one sees that:

- $$\|R_r^{(N)}\| \le C_N h^N\left(1 + \sum_{k=0}^{N-1}\sum_{l=0}^{r-1} \|a_l^{(k)}\|_{C^{2(N-k)+15}}\right), \tag{4-45}$$

with C_N depending on N, M and the a_q, but neither on r nor $\boldsymbol{q}$.

- Since L_{p,q_r} has coefficient supported in $\mathcal{V}_{q_r}$, we see by induction that $\operatorname{supp} a_{r+1}^{(k)} \subset \mathcal{V}_{q_0 \cdots q_r}^+$ as announced.
- $a_{r+1}^{(0)} = \prod_{p=1}^{r+1} a_{q_{r+1-p}} \circ F^{-p}$.

4D2. *Control of the symbols.* We aim at estimating the seminorms $\|a_r^{(k)}\|_{C^m}$ for $k < N,\ 1 \le r \le n$ and $m \in \mathbb{N}$. We will show the following:

Proposition 4.10. *For every* $r \in \{1, \dots, n\},\ k \in \{0, \dots, N-1\}$ *and* $m \in \mathbb{N}$, *there exists* $C(k, m)$, *such that*, *with* $\Gamma_{k,m} = (k+1)(m+k+1)$,

$$\|a_r^{(k)}\|_{C^m} \le C(k, m) r^{\Gamma_{k,m}} (J^+_{q_0 \cdots q_{r-1}})^{2k+m} \|\alpha\|_\infty^r. \tag{4-46}$$

Remark. • What is important in this result is the way in which the bound depends on r and $\boldsymbol{q}$. Up to the term $r^{\Gamma_{k,m}}$, which is supposed to behave like $O(|\log h|^{\Gamma_{k,m}})$, the significant part of the estimate is that we can control the symbols by the local Jacobian.

• Since $\operatorname{supp} a_r^{(k)} \subset \mathcal{V}^+_{q_0 \cdots q_{r-1}}$, we need to focus on points $\rho \in \mathcal{V}^+_{q_0 \cdots q_{r-1}}$.

• Our method is very close to the ones developed in [Rivière 2010; Dyatlov et al. 2022]. However, we've changed a few things at the cost of being less precise on the exponent $\Gamma_{k,m}$. Our aim was to treat our problem as if we wanted to control the product of r triangular matrices.

Let us pick $\rho \in \mathcal{V}^+_{q_0 \cdots q_{r-1}}$. With (4-43), one sees that if $k, m \in \mathbb{N}$, then $d^m a_{r+1}^{(k)}$ depends on $d^{m'} a_r^{(k')}(F^{-1}(\rho))$ for several m', k'. Before going deeper in the analysis of this dependence, let us note two obvious facts:

- This dependence is linear, with coefficients smoothly depending on ρ.
- If $d^m a_{r+1}^{(k)}$ depends effectively on $d^{m'} a_r^{(k')}(F^{-1}(\rho))$, then $k' \le k$ and $2k' + m' \le 2k + m$.

Precise analysis of the dependence. That being said, let us pick $m_0, k_0 \in \mathbb{N}$. Set $N_0 = 2k_0 + m_0$ and consider the (column) vector

$$A_r(\rho) := (d^m a_r^{(k)}(\rho))_{k \le k_0, 2k+m \le N_0} \in \bigoplus_{k \le k_0, 2k+m \le N_0} S^m T^*_\rho U. \tag{4-47}$$

Here $S^m T^*_\rho U$ is the space of m-linear symmetric forms on $T_\rho U$. To define a norm on the fibers $S^m T^*_\rho U$, we can use, for $f \in S^m T^*_\rho U$,

$$\|f\|_{m,\rho} = \sup_{v_1, \dots, v_m \in T_\rho U} \frac{f(v_1, \dots, v_m)}{\|v_1\|_\rho \cdots \|v_m\|_\rho}, \tag{4-48}$$

where $\|v\|_\rho$ for $v \in T_\rho U$ is the norm induced by the Riemannian metric used to define J_1^u in (3-8). Note that, for any fixed neighborhood of $\mathcal{T}$, there exists a global constant $C > 0$ such that, for each $a \in C_c^\infty(U)$ supported in this neighborhood, one has

$$C^{-1} \|a\|_{C^m} \le \sup_{m' \le m} \sup_{\rho \in U} \|d^{m'} a\|_{m',\rho} \le C \|a\|_{C^m}.$$

We will denote by γ_1, γ_2, etc. elements of $\mathcal{I} := \mathcal{I}(k_0, m_0) = \{(k, m) \in \mathbb{N}^2 : k \le k_0,\ 2k+m \le N_0\}$. We equip $\mathcal{I}$ with the lexicographic order $\prec$ and write $\#\mathcal{I} := \Gamma_{k_0, m_0}$ (see Figure 10). We order the indices of $A_r(\rho)$ with $\prec$. $A_r(\rho)$ depends linearly on $A_{r-1}(F^{-1}(\rho))$ and this dependence can be made explicit by a matrix

$$P^{(r)}(\rho) = (P^{(r)}_{\gamma_1 \gamma_2}(\rho))_{\gamma_1, \gamma_2 \in \mathcal{I}}, \quad \text{where } P^{(r)}_{\gamma_1 \gamma_2}(\rho) \in L(S^{m'} T^*_{F^{-1}(\rho)} U, S^m T^*_\rho U) \quad \text{if } \gamma_1 = (k, m),\ \gamma_2 = (k', m'),$$

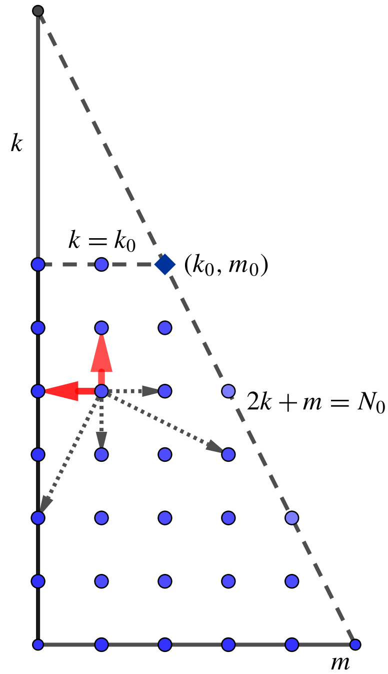

Figure 10. The starting point (k_0, m_0) is represented by a diamond. The set $\mathcal{I}$ corresponds to the couple $(k, m) \in \mathbb{N}^2$ in the region under the dotted lines $k = k_0$ and $2k + m = N_0$. We've represented a family of arrows starting from a point $\gamma_1 \in \mathcal{I}$. The dotted arrows points toward β such that $\gamma_2 \prec \gamma_1$. The big red arrows points toward points γ_2 such that $P^{(r)}_{\gamma_1\gamma_2} = 0$.

so that

$$A_r(\rho) = P^{(r)}(\rho)\, A_{r-1}(F^{-1}(\rho)). \tag{4-49}$$

Notation. If $\gamma_1 = (k, m),\ \gamma_2 = (m', k'),\ \rho, \rho' \in U$ and if $A : S^{m'} T^*_{\rho'} U \to S^m T^*_\rho U$ is a linear operator, we will denote by

$$\| \cdot \|_{\gamma_1, \rho, \gamma_2, \rho'}$$

its subordinate norm for the norms defined by (4-48).

Analyzing (4-43), it turns out that if $\gamma_1 = (k, m),\ \gamma_2 = (k', m') \in \mathcal{I}$, then:

- If $k' > k$, then $P^{(r)}_{\gamma_1\gamma_2}(\rho) = 0$.
- If $k = k'$, the contribution to $d^m a^{(k)}_r(\rho)$ of $a^{(k)}_{r-1}$ comes from

$$\begin{aligned}
&d^m((a_{q_{r-1}} a^{(k)}_{r-1}) \circ F^{-1})(\rho) \\
&= a_{q_{r-1}}(F^{-1}(\rho)) \times d^m(a^{(k)}_{r-1} \circ F^{-1})(\rho) + (\text{derivatives of order strictly less than m for } a^{(k)}_{r-1}) \\
&= a_{q_{r-1}}(F^{-1}(\rho)) \times ({}^t dF^{-1}(\rho))^{\otimes m} d^m a^{(k)}_{r-1}(F^{-1}(\rho)) + (\text{derivatives of order strictly less than m for } a^{(k)}_{r-1}).
\end{aligned}$$

In particular, if $\gamma_1 = (k, m) \prec \gamma_2 = (k, m')$ doesn't hold, we see that $P^{(r)}_{\gamma_1\gamma_2}(\rho) = 0$.

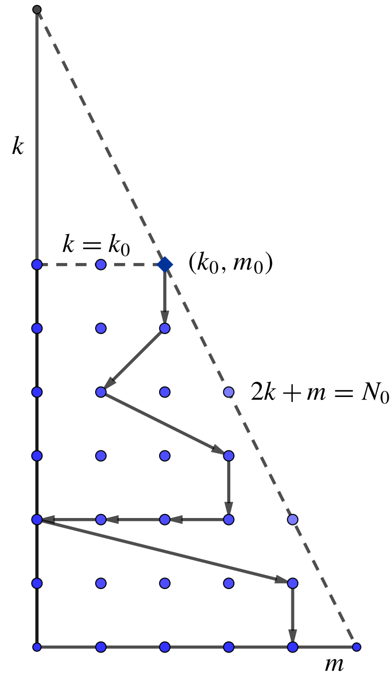

Figure 11. We've represented the reduction of an element $\vec{\gamma} \in \mathcal{E}_r(k_0, m_0)$, i.e., the arrows between γ_i and γ_{i+1} when $\gamma_i \neq \gamma_{i+1}$. During the descent, the value of m can only increase when k decreases strictly.

- If $k' < k$, we can have $P^{(r)}_{\gamma_1\gamma_2}(\rho) \neq 0$ with $m' > m$. But, the use of the lexicographic order ensures that $\gamma_1 \prec \gamma_2$ in that case.

Hence, $P^{(r)}(\rho)$ is a lower triangular matrix and the diagonal coefficients for the index $\gamma_1 = (k, m)$ are given by

$$P^{(r)}_{\gamma_1\gamma_1}(\rho) : f \in S^m T^*_{F^{-1}(\rho)} U \mapsto a_{q_{r-1}}(F^{-1}(\rho)) \times ({}^t dF^{-1}(\rho))^{\otimes m} f \in S^m T^*_\rho U. \tag{4-50}$$

Iterating (4-49), we have

$$A_r(\rho) = P^{(r)}(\rho) P^{(r-1)}(F^{-1}(\rho)) \cdots P^{(2)}(F^{-(r-2)}(\rho)) A_1(F^{1-r}(\rho)).$$

For $\gamma \in \mathcal{I}$, we define, see Figure 11,

$$\mathcal{E}_r(\gamma) = \{\vec{\gamma} = (\gamma_1, \dots, \gamma_r) \in \mathcal{I}^r : \gamma_r = \gamma,\ \gamma_i \prec \gamma_{i+1}\}.$$

The triangular property of P allows us to write

$$(A_r(\rho))_\gamma = \sum_{\vec{\gamma} \in \mathcal{E}_r(\gamma)} P^{(r)}_{\gamma_r\gamma_{r-1}}(\rho) \cdots P^{(2)}_{\gamma_2\gamma_1}(F^{-(r-2)}(\rho))(A_1(F^{1-r}(\rho)))_{\gamma_1}.$$

Control of individual terms. Let us fix $\gamma = (k, m)$ and pick $\vec{\gamma} \in \mathcal{E}_r(\gamma)$. We wish to analyze the operator

$$P_{\vec{\gamma}}(\rho) := P^{(r)}_{\gamma_r\gamma_{r-1}}(\rho) \cdots P^{(2)}_{\gamma_2\gamma_1}(F^{-(r-2)}(\rho)).$$

First of all, $\#\{i \in \{1,\dots,r-1\} : \gamma_{i+1} \neq \gamma_i\} \leq \Gamma_{k_0,m_0}$. So let us write

$$\{i \in \{1,\dots,r-1\} : \gamma_{i+1} \neq \gamma_i\} = \{t_1 < \cdots < t_d\},$$

with $d \leq \Gamma_{k_0,m_0}$. We can set $t_{d+1} = r,\ t_0 = 0$ and we can rewrite

$$\vec{\gamma} = (\underbrace{\beta_1,\dots,\beta_1}_{t_1}, \underbrace{\beta_2,\dots,\beta_2}_{t_2-t_1}, \dots, \underbrace{\beta_d,\dots,\beta_d}_{t_d-t_{d-1}}, \underbrace{\beta_{d+1},\dots,\beta_{d+1}}_{t_{d+1}-t_d}).$$

For $p \in \{1,\dots,d+1\}$, we introduce the operator

$$D_p(\rho) = P^{(t_p)}_{\beta_p\beta_p}(F^{-(r-t_p)}(\rho)) \cdots P^{(t_{p-1}+2)}_{\beta_p\beta_p}(F^{-(r-t_{p-1}-2)}(\rho)),$$

and for $p \in \{1,\dots,d\}$

$$T_p(\rho) = P^{t_p+1}_{\beta_{p+1}\beta_p}(F^{-(r-t_p-1)}(\rho))$$

so that we can write

$$P_{\vec{\gamma}}(\rho) = D_{d+1}(\rho)T_d(\rho)D_d(\rho)\cdots T_1(\rho)D_1(\rho).$$

For $p \in \{1,\dots,d+1\}$, if $\beta_p = (k,m)$, we can see that

$$\begin{aligned} D_p(\rho) &= \Bigg[\prod_{j=t_{p-1}+1}^{t_p-1} a_{q_j} \circ F^{-(r-j)}(\rho)\Bigg][({}^t dF^{-1}(F^{-(r-t_p)}(\rho)))^{\otimes m} \circ \cdots \circ ({}^t dF^{-1}(F^{-(r-t_{p-1}-2)}(\rho)))^{\otimes m}] \\ &= \Bigg[\prod_{j=t_{p-1}+1}^{t_p-1} a_{q_j} \circ F^{-(r-j)}(\rho)\Bigg]({}^t dF^{-(t_p-t_{p-1}-1)}(F^{-(r-t_p)}(\rho)))^{\otimes m}. \end{aligned}$$

We introduce the word

$$\boldsymbol{q}_p = q_{t_{p-1}} \cdots q_{t_p-1},$$

and set $\rho_p = F^{-(r-t_p)}(\rho),\ \rho'_p = F^{-(t_p-t_{p-1}-1)}(\rho_p)$. To estimate the subordinate norm of $D_p(\rho)$, we use Lemma 4.6. Since $\rho \in \mathcal{V}_{\boldsymbol{q}^+}$, $\rho_p \in \mathcal{V}^+_{\boldsymbol{q}_p}$ and we have

$$\begin{aligned} \|D_p(\rho)\|_{\beta_p,\rho_p,\beta_p,\rho'_p} &\leq \Bigg|\prod_{j=t_{p-1}+1}^{t_p-1} a_{q_j} \circ F^{-(r-j)}(\rho)\Bigg| \sup_{\rho_p \in \mathcal{V}^+_{q_p}} \|dF^{-(t_p-t_{p-1}-1)}(\rho_p)\|^m \\ &\leq (CJ^+_{\boldsymbol{q}_p})^m \Bigg|\prod_{j=t_{p-1}+1}^{t_p-1} a_{q_j} \circ F^{-(r-j)}(\rho)\Bigg| \leq C_{k_0,m_0}(J^+_{\boldsymbol{q}_p})^{N_0}\Bigg|\prod_{j=t_{p-1}+1}^{t_p-1} a_{q_j} \circ F^{-(r-j)}(\rho)\Bigg|. \end{aligned}$$

To estimate the norms of $T_p(\rho)$, we simply note that they depend smoothly on ρ_p, which lies in a compact set, so we can bound them by a uniform constant C_1. This is not a problem since they appear d times in $P_{\vec{\gamma}}$ with $d \leq \Gamma_{k_0,m_0}$. Consequently, we can estimate $\|P_{\vec{\gamma}}(\rho)\|_{\gamma,\rho,\gamma_1,F^{-(r-1)}(\rho)}$,

$$\|P_{\vec{\gamma}}(\rho)\|_{\gamma,\rho,\gamma_1,F^{-(r-1)}(\rho)} \leq C_{k_0,m_0}(J^+_{\boldsymbol{q}_1}\cdots J^+_{\boldsymbol{q}_{d+1}})^{N_0}|a_{\boldsymbol{q},\vec{\gamma}}(\rho)| \leq C_{k_0,m_0}(J^+_{\boldsymbol{q}})^{N_0}|a_{\boldsymbol{q},\vec{\gamma}}(\rho)|, \tag{4-51}$$

where

$$a_{\boldsymbol{q},\vec{\gamma}} = \prod_{p=1}^{d+1}\prod_{j=t_{p-1}+1}^{t_p-1} a_{q_j} \circ F^{-(r-j)}. \tag{4-52}$$

Here, the last inequality holds by applying d times (4-24), with $d \le \Gamma_{k_0,m_0}$, once we've noted that

$$\boldsymbol{q} = \boldsymbol{q}_1 \cdots \boldsymbol{q}_{d+1}.$$

Finally, if $\gamma_1 = (k_1, m_1)$, to estimate $\|(A_1(F^{1-r}(\rho)))_{\gamma_1}\|_{m_1,F^{1-r}(\rho)}$, we simply note that it depends smoothly on $F^{1-r}(\rho)$, so that we can bound it by a uniform constant. Hence, we have

$$\|P_{\vec{\gamma}}(\rho)A_1(F^{1-r}(\rho))\|_{m,\rho} \le C_{k_0,m_0}(J^+_{\boldsymbol{q}})^{N_0}|a_{\boldsymbol{q},\vec{\gamma}}(\rho)|. \tag{4-53}$$

Cardinality of $\mathcal{E}_r(\gamma)$. The bound we will provide is far from optimal but it will turn out to be enough for our purpose. To count the number of elements in $\mathcal{E}_r(\gamma)$, we remark that it is similar to counting the number of decreasing sequences of length r starting from γ. This number is smaller than the number of increasing sequences of length r in $\{1, \dots, \Gamma_{k_0,m_0}\}$. Recalling that the number of sequences $u_1 \le u_2 \le \cdots \le u_r$ satisfying $u_1 = 1$ and $u_r = b$ is equal to $\binom{b+r-2}{r-2}$, one can estimate

$$\#\mathcal{E}_r(\gamma) \le \sum_{b=1}^{\Gamma_{k_0,m_0}} \binom{b+r-2}{r-2} \le \Gamma_{k_0,m_0}(r-1)^{\Gamma_{k_0,m_0}}. \tag{4-54}$$

Finally, we can compute explicitly Γ_{k_0,m_0} and we find $\Gamma_{k_0,m_0} = (k_0+1)(m_0+1+k_0)$.

Conclusion. We finally combine (4-54) and (4-53) to prove Proposition 4.10 (recall $|a_q| = |\alpha|\chi_q \le \|\alpha\|_\infty$):

$$\begin{aligned}
\sup_{\rho \in \mathcal{V}_{q_0\cdots q_{r-1}}} \|d^{m_0}a_r^{(k_0)}\|_{m_0,\rho} &= \sup_{\rho \in \mathcal{V}_{q_0\cdots q_{r-1}}} \|(A_r(\rho))_{(k_0,m_0)}\|_{m_0,\rho} \\
&\le \sum_{\vec{\gamma} \in \mathcal{E}_r(k_0,m_0)} \|P_{\vec{\gamma}}(\rho)A_1(F^{1-r}(\rho))\|_{m_0,\rho} \\
&\le \Gamma_{k_0,m_0} r^{\Gamma_{k_0,m_0}} C_{k_0,m_0}(J^+_{\boldsymbol{q}})^{N_0}|a_{\boldsymbol{q},\vec{\gamma}}(\rho)| \\
&\le C_{k_0,m_0} r^{\Gamma_{k_0,m_0}}(J^+_{\boldsymbol{q}})^{N_0}\|\alpha\|^r_\infty.
\end{aligned}$$

Finally, we get as expected

$$\|a_r^{(k_0)}\|_{C^{m_0}} \le C_{k_0,m_0} r^{\Gamma_{k_0,m_0}}(J^+_{\boldsymbol{q}})^{N_0}\|\alpha\|^r_\infty.$$

4D3. *End of proof of Proposition 4.8.* Armed with these estimates, we are now able to conclude the proof of Proposition 4.8 under the assumptions (4-31). Assume that this assumption is satisfied and construct inductively the symbols $a_r^{(k)}$ with the formula (4-43). Since $J^+_{\boldsymbol{q}} \le Ch^{-\delta}$, it implies that $n = O(\log h)$. Hence, we have, for $r \le n$,

$$\|a_r^{(k)}\|_{C^m} \le C_{k,m}h^{-\delta m}h^{-2k\delta}|\log h|^{\Gamma_{k,m}}\|\alpha\|^r_\infty \le C_{k,m}h^{-\delta_1 m}h^{-2k\delta_1}\|\alpha\|^r_\infty.$$

The symbol $h^{2\delta_1 k}a_r^{(k)}$ lies in $\|\alpha\|^r_\infty S^{\mathrm{comp}}_{\delta_1}(T^*\mathbb{R})$. Using Borel's theorem with the parameter $h' = h^{1-2\delta_1}$, we can construct a symbol

$$a^+_{q_0\cdots q_{r-1}} \sim \sum_{k=0}^{\infty}(h')^k h^{2\delta_1 k}a_r^{(k)} = \sum_{k=0}^{\infty} h^k a_r^{(k)} \in \|\alpha\|^r_\infty S^{\mathrm{comp}}_{\delta_1},$$

that is, for every $N \in \mathbb{N}$,

$$a^+_{q_0\cdots q_{r-1}} - \sum_{k=0}^{N-1} h^k a_r^{(k)} = O(h^{(1-2\delta_1)N}\|\alpha\|_\infty^r).$$

By construction of the $a_r^{(k)}$, for every $N \in \mathbb{N}$, we have

$$U_{\boldsymbol{q}}^+ - \mathrm{Op}_h(a_{\boldsymbol{q}}^+)T^{|\boldsymbol{q}|} = R_n^{(N)} + O(h^{(1-2\delta_1)}\|\alpha\|_\infty^r).$$

Fix some $K \geq 0$ such that $\min(1, \|\alpha\|_\infty^n) = O(h^{-K})$, so that $\|\alpha\|_\infty^r = O(k^{-K})$. With (4-45) and our estimates, we can control

$$\|R_n^{(N)}\| \leq C_N h^N (1 + |\log h|^{\Gamma_{k,m}+1} h^{-\delta(2N+15)} h^{-K}) \leq C_N h^{-15\delta_1 + N(1-2\delta_1)-K}.$$

Since we can choose N as large as we want, we have finally proved that

$$U_{\boldsymbol{q}}^+ - \mathrm{Op}_h(a_{\boldsymbol{q}}^+)T^{|\boldsymbol{q}|} = O(h^\infty). \qquad \square$$

4D4. *Norm of sums over many words.* We'll make use of the tools and notation developed in this subsection to prove the following proposition. To state it, we introduce the notation

$$\mathcal{Q}(n, \tau, C_0) := \{\boldsymbol{q} \in \mathcal{A}^n : J_{\boldsymbol{q}}^+ \leq C_0 h^{-\tau}\}. \tag{4-55}$$

Proposition 4.11. *There exists $C = C(C_0, \tau)$ such that, for every $\mathcal{Q} \subset \mathcal{Q}(n, \tau, C_0)$, the following bound holds*:

$$\left|\sum_{\boldsymbol{q} \in \mathcal{Q}} U_{\boldsymbol{q}}\right|_{L^2 \to L^2} \leq C\|\alpha\|^n |\log h|. \tag{4-56}$$

Proof. Throughout the proof, we'll denote by C quasiglobal constants, i.e., constants depending on C_0, τ and the same other parameters as global constants. We will also be led to use a constant C_1: it has the same dependence.

Step 1: First note that, since $J_{\boldsymbol{q}}^+ \leq C_0 h^{-\tau}$, n satisfies the bound $n = O(\log h)$.

Step 2: If $\boldsymbol{q} \in \mathcal{Q}(n, \tau, C_0)$, denote by $l(\boldsymbol{q}) = l$ the largest integer such that

$$J^+_{q_0\cdots q_{l-1}} \leq h^{-\tau/2}.$$

Since $J_{q_0\cdots q_l} > h^{-\tau/2}$, $J^+_{q_0\cdots q_{l-1}} > C h^{-\tau/2}$ and hence

$$J^+_{q_l\cdots q_{n-1}} \leq C \frac{h^{-\tau}}{J^+_{q_0\cdots q_{l-1}}} \leq C_1 h^{-\tau/2}.$$

We can then write $\boldsymbol{q} = \boldsymbol{s}\boldsymbol{r}$ with $\boldsymbol{s} \in \mathcal{Q}(l, \tau/2, 1)$, $\boldsymbol{r} \in \mathcal{Q}(n-l, \tau/2, C_1)$. It follows that we can write

$$\sum_{\boldsymbol{q} \in \mathcal{Q}} U_{\boldsymbol{q}} = \sum_{l=1}^{n} \sum_{\substack{\boldsymbol{s} \in \mathcal{Q}(l,\tau/2,1) \\ \boldsymbol{r} \in \mathcal{Q}(n-l,\tau/2,C_1)}} F_l(\boldsymbol{s}, \boldsymbol{r}) U_{\boldsymbol{r}} U_{\boldsymbol{s}},$$

with $F_l(\boldsymbol{s},\boldsymbol{r}) = \mathbb{1}_{\boldsymbol{sr}\in\mathcal{Q}}$. It is then enough to show the bound

$$\max_{1\le l\le n}\left|\sum_{\substack{\boldsymbol{s}\in\mathcal{Q}(l,\tau/2,1)\\ \boldsymbol{r}\in\mathcal{Q}(n-l,\tau/2,C_1)}} F_l(s,r)U_{\boldsymbol{r}}U_{\boldsymbol{s}}\right| \le C\|\alpha\|_\infty^n. \tag{4-57}$$

In the following, we fix some $1\le l\le n$ and we'll simply write $\sum_{\boldsymbol{s},\boldsymbol{r}}$ to alleviate the notation. Note that the number of terms in the sum is bounded by

$$|\mathcal{Q}(l,\tau/2,1)\times\mathcal{Q}(n-l,\tau/2,C_1)| \le |\mathcal{A}|^l\times|\mathcal{A}|^{n-l}\le|\mathcal{A}|^n\le h^{-Q},$$

where $Q = C\log|\mathcal{A}|$.

<u>Step 3</u>: We fix some large $N\in\mathbb{N}$ and $\delta_1\in(\tau/2,1/2)$. Recall that we can write

$$U_{\boldsymbol{s}} = \left(\mathrm{Op}_h\left(\sum_{k=0}^{N-1}h^k a_{\boldsymbol{s}}^{(k)}\right) + O_{L^2\to L^2}(h^{(1-2\delta_1)N-15\delta_1}\|\alpha\|_\infty^l)\right)T^l,$$

$$U_{\boldsymbol{r}} = T^{n-l}\left(\mathrm{Op}_h\left(\sum_{k=0}^{N-1}h^k a_{\boldsymbol{r}}^{(k)}\right) + O_{L^2\to L^2}(h^{(1-2\delta_1)N-15\delta_1}\|\alpha\|_\infty^{n-l})\right),$$

with bounds on $a_{\boldsymbol{s}}^{(k)}$ and $a_{\boldsymbol{r}}^{(k)}$ given by Proposition 4.8.

We then use the formula for the composition of operators in $\Psi_{\delta_1}^{\mathrm{comp}}(T^*\mathbb{R})$ (Lemma 3.3) and for simplicity, we write $\mathcal{L}_k(a,b)(\rho) = (i^k/k!)(A(D))^k(a\otimes b)(\rho,\rho)$. For $0\le k\le N-1$, we set

$$a_{\boldsymbol{s},\boldsymbol{r},k} = \sum_{j+k_-+k_+=k}\mathcal{L}_j(a_{\boldsymbol{r}}^{(k_-)}, a_{\boldsymbol{s}}^{(k_+)}).$$

Note that if $j+k_-+k_+\ge N$,

$$\begin{aligned}\|a_{\boldsymbol{r}}^{(k_-)}\otimes a_{\boldsymbol{s}}^{(k_+)}\|_{C^{2j+13}} &\le C_j\sup_{m_++m_-=2j+13}\|a_{\boldsymbol{r}}^{(k_-)}\|_{C^{m_-}}\|a_{\boldsymbol{s}}^{(k_+)}\|_{C^{m_+}}\\ &\le C_{j,k_-,k_+}h^{-(2k_-+m_-)\delta_1}h^{-(2k_-+m_+)\delta_1}\|\alpha\|_\infty^n\\ &\le C_{j,k_-,k_+}h^{-2\delta_1(j+k_-+k_+)-13\delta_1}\|\alpha\|_\infty^n\\ &\le C_{j,k_-,k_+}h^{-2\delta_1N-13\delta_1}\|\alpha\|_\infty^n\end{aligned}$$

and henceforth,

$$O(h^{j+k_-+k_+}\|a_{\boldsymbol{r}}^{(k_-)}\otimes a_{\boldsymbol{s}}^{(k_+)}\|_{C^{2j+13}}) = O(h^{(1-2\delta_1)N-15\delta_1}\|\alpha\|_\infty^n).$$

As a consequence, we can write

$$U_{\boldsymbol{r}}U_{\boldsymbol{s}} = T^{n-l}\left(\mathrm{Op}_h\left(\sum_{k=0}^{N-1}h^k a_{\boldsymbol{s},\boldsymbol{r},k}\right)\right)T^l + O_{L^2\to L^2}(h^{(1-2\delta_1)N-15\delta_1}\|\alpha\|_\infty^n).$$

It follows that

$$\sum_{\boldsymbol{s},\boldsymbol{r}}U_{\boldsymbol{r}}U_{\boldsymbol{s}} = T^{n-l}\left(\mathrm{Op}_h\left(\sum_{k=0}^{N-1}h^k a^{(k)}\right)\right)T^l + O_{L^2\to L^2}(h^{(1-2\delta_1)N-15\delta_1-Q}\|\alpha\|_\infty^n),$$

where

$$a^{(k)} = \sum_{\boldsymbol{s},\boldsymbol{r}} F(\boldsymbol{s},\boldsymbol{r}) a_{\boldsymbol{s},\boldsymbol{r},k}. \tag{4-58}$$

Suppose that N has been chosen such that

$$(1-2\delta_1)N > 15\delta_1 + Q.$$

The remainder term is thus controlled by the desired bound since it is of order $O(\|\alpha\|_\infty^n)$.

Step 4: C^0 norm of $a^{(0)}$. We have

$$a^{(0)} = \sum_{\boldsymbol{s},\boldsymbol{r}} F(\boldsymbol{s},\boldsymbol{r}) a_{\boldsymbol{s}}^{(0)} a_{\boldsymbol{r}}^{(0)},$$

where, by virtue of (4-37) and (4-38),

$$a_{\boldsymbol{s}}^{(0)} = \prod_{p=1}^{l} a_{s_{l-p}} \circ F^{-p}, \quad a_{\boldsymbol{r}}^{(0)} = \prod_{p=0}^{n-l-1} a_{r_p} \circ F^{p}.$$

As a consequence, we can estimate

$$|a^{(0)}| \le \sum_{\boldsymbol{s},\boldsymbol{r}} |a_{\boldsymbol{s}}^{(0)}|\,|a_{\boldsymbol{r}}^{(0)}| \le \prod_{p=1}^{l}\biggl(\sum_{q\in\mathcal{A}} |a_q|\biggr) \circ F^{-p} \times \prod_{p=0}^{n-l-1}\biggl(\sum_{q\in\mathcal{A}} |a_q|\biggr) \circ F^{p} \le \|\alpha\|_\infty^n.$$

Step 5: C^m norms of $a^{(k)}$. We will show there exist constants $C_{k,m}$ (depending only on C_0, δ_1, τ and m, k) such that, for all $0 \le k \le N-1$ and $m \in \mathbb{N}$,

$$\|a^{(k)}\|_{C^m} \le C_{k,m} h^{-(2k+m)\delta_1} \|\alpha\|_\infty^n. \tag{4-59}$$

Let's compute

$$\begin{aligned}
\|a^{(k)}\|_{C^m} &\le \sum_{\boldsymbol{s},\boldsymbol{r}} \|a_{\boldsymbol{s},\boldsymbol{r},k}\|_{C^m} \le \sum_{\boldsymbol{s},\boldsymbol{r}} \sum_{j+k_++k_-=k} \|\mathcal{L}_j(a_{\boldsymbol{r}}^{(k_-)}, a_{\boldsymbol{s}}^{(k_+)})\|_{C^m} \\
&\le \sum_{\boldsymbol{s},\boldsymbol{r}} \sum_{j+k_++k_-=k} \|a_{\boldsymbol{r}}^{(k_-)} \otimes a_{\boldsymbol{s}}^{(k_+)}\|_{C^{2j+m}} \\
&\le \sum_{\boldsymbol{s},\boldsymbol{r}} \sum_{\substack{j+k_++k_-=k \\ m_++m_-\le m+2j}} \|a_{\boldsymbol{r}}^{(k_-)}\|_{C^{m_-}} \|a_{\boldsymbol{s}}^{(k_+)}\|_{C^{m_+}},
\end{aligned}$$

and hence

$$\|a^{(k)}\|_{C^m} \le C_{k,m} \sup_{\substack{j+k_++k_-=k \\ m_++m_-\le m+2j}} \sum_{\boldsymbol{s},\boldsymbol{r}} \|a_{\boldsymbol{r}}^{(k_-)}\|_{C^{m_-}} \|a_{\boldsymbol{s}}^{(k_+)}\|_{C^{m_+}}. \tag{4-60}$$

Let us fix j, k_+, k_-, m_+, m_- satisfying $j+k_++k_-=k$, $m_-+m_+\le m+2j$ and let us estimate

$$\sum_{\boldsymbol{s}} \|a_{\boldsymbol{s}}^{(k_+)}\|_{C^{m_+}} \times \sum_{\boldsymbol{r}} \|a_{\boldsymbol{r}}^{(k_-)}\|_{C^{m_-}}.$$

We estimate the sum over $\boldsymbol{s}$. The same kind of estimates will hold for $\boldsymbol{r}$ with the same methods. We reuse the tools developed in the last subsections. Namely, we set $N_+ = 2k_+ + m_+$, $\gamma_+ = (k_+, m_+)$, $\mathcal{I} = \mathcal{I}(\gamma_+)$ and

$$(A_{\boldsymbol{s}}(\rho)) = (d^m a_{\boldsymbol{s}}^{(k)})_{k\le k_+,\, 2k+m\le N_+}.$$

We have shown that there exists a global constant $C > 0$ such that

$$\|a_{\boldsymbol{s}}^{(k_+)}\|_{C^{m_+}} \leq \sup_{\rho} \|A_{\boldsymbol{s}}(\rho)\| \leq C \sum_{\vec{\gamma}\in\mathcal{E}_l(\gamma_+)} \|P_{\vec{\gamma}}(\rho)\| \leq \sum_{\vec{\gamma}\in\mathcal{E}_l(\gamma_+)} C_{N_+,k_+}(J_{\boldsymbol{s}}^+)^{N_+}|a_{\boldsymbol{s},\vec{\gamma}}(\rho)|$$
$$\leq C_{N_+,k_+} h^{-\tau N_+/2} \sum_{\vec{\gamma}\in\mathcal{E}_l(\gamma_+)} |a_{\boldsymbol{s},\vec{\gamma}}(\rho)|,$$

where C_{N_+,k_+} depends on C_0, τ, N_+, k_+ and global parameters. We hence have to estimate

$$\sum_{\boldsymbol{s}} \sum_{\vec{\gamma}\in\mathcal{E}_l(\gamma_+)} |a_{\boldsymbol{s},\vec{\gamma}}(\rho)|.$$

Fix $\vec{\gamma} \in \mathcal{E}_l(\alpha_+)$ and write it

$$\vec{\gamma} = (\underbrace{\beta_1,\dots,\beta_1}_{t_1}, \underbrace{\beta_2,\dots,\beta_2}_{t_2-t_1}, \dots, \underbrace{\beta_d,\dots,\beta_d}_{t_d-t_{d-1}}, \underbrace{\beta_{d+1},\dots,\beta_{d+1}}_{t_{d+1}-t_d}), \quad \text{where } d \leq \Gamma_{k_+,m_+},$$

and recall that

$$a_{\boldsymbol{s},\vec{\gamma}} = \prod_{p=1}^{d+1} \prod_{j=t_{p-1}+1}^{t_p-1} a_{s_j} \circ F^{-(l-j)}.$$

When one sums over $\boldsymbol{s} \in \mathcal{A}^l$, the values of $\boldsymbol{s}$ at the indices t_i, $1 \leq i \leq d$, do not play a role and we write

$$\sum_{\boldsymbol{s}} |a_{\boldsymbol{s},\vec{\gamma}}| = \sum_{s_{t_1}\in\mathcal{A}} \cdots \sum_{s_{t_d}\in\mathcal{A}} \prod_{p=1}^{d+1} \prod_{j=t_{p-1}+1}^{t_p-1} \Bigl(\sum_{s\in\mathcal{A}} |a_s|\Bigr) \circ F^{-(l-j)}$$
$$\leq |\mathcal{A}|^d \sup_{\rho} \Bigl(\sum_{s\in\mathcal{A}} |a_s|\Bigr)^l \leq K^{\Gamma_{k_+,m_+}} \|\alpha\|_\infty^l \leq C_{k_+,m_+} \|\alpha\|_\infty^l.$$

As a consequence,

$$\sum_{\boldsymbol{s}} \sum_{\vec{\gamma}\in\mathcal{E}_l(\gamma_+)} |a_{\boldsymbol{s},\vec{\gamma}}| \leq \#\mathcal{E}_l(\gamma_+) C_{k_+,m_+} \|\alpha\|_\infty^l \leq C_{k_+,m_+} (l-1)^{\Gamma_{k_+,m_+}} \|\alpha\|_\infty^l,$$

which gives

$$\sum_{\boldsymbol{s}} \|a_{\boldsymbol{s}}^{(k_+)}\|_{C^{m_+}} \leq C_{k_+,m_+} h^{-\tau N_+/2} (l-1)^{\Gamma_{k_+,m_+}} \|\alpha\|_\infty^l \leq C_{k_+,m_+} h^{-\delta_1 N_+} \|\alpha\|_\infty^l,$$

where the last inequality (with a different value of C_{k_+,m_+}) follows from the fact that $l = O(\log h)$ and $\delta_1 > \tau/2$. The same kind of estimates holds for the sum over $\boldsymbol{r}$:

$$\sum_{\boldsymbol{r}} \|a_{\boldsymbol{r}}^{(k_-)}\|_{C^{m_-}} \leq C_{k_-,m_-} h^{-\delta_1 N_-} \|\alpha\|_\infty^{n-l}.$$

Eventually, using (4-60), we get (4-59) since

$$N_+ + N_- = 2k_+ + m_+ + 2k_- + m_- \leq 2(k_+ + k_- + j) + m = 2k + m.$$

Step 6: Conclusion. We can conclude the proof of the Proposition 4.11. The bound (4-59) shows that, for $0 \leq k \leq N-1$, $a^{(k)} \in h^{-2k\delta_1} \|\alpha\|_\infty^n S_{\delta_1}^{\mathrm{comp}}$ and thus $\sum_{k=0}^{N-1} h^k a^{(k)} \in S_{\delta_1}^{\mathrm{comp}} \|\alpha\|_\infty^n$. From the L^2-boundedness of pseudodifferential operators with symbol in S_{δ_1},

$$\left\| \mathrm{Op}_h\left(\sum_{k=0}^{N-1} h^k a^{(k)}\right)\right\| \leq \sum_{k=0}^{N-1}\sum_{m\leq M} h^{k+m/2}\|a^{(k)}\|_{C^m} \leq \sum_{k=0}^{N-1}\sum_{m\leq M} C_{k,m} h^{(k+2m)(1/2-\delta_1)}\|\alpha\|_\infty^n \leq C\|\alpha\|_\infty^n,$$

where C depends only on C_0, τ, δ_1. Since $\|T\| \leq 1$, we get

$$\left\|\sum_{\boldsymbol{s},\boldsymbol{r}} F(\boldsymbol{s},\boldsymbol{r}) U_{\boldsymbol{r}} U_{\boldsymbol{s}}\right\| \leq C\|\alpha\|_\infty^n,$$

which concludes the proof of Proposition 4.11. □

4E. *Manipulations of the $U_{\boldsymbol{q}}$.*

4E1. *First consequences.* We now make use of Proposition 4.8 to deduce several important facts. We go on following [Dyatlov et al. 2022]. In the whole subsection, we fix $0 \leq \delta < \delta_1 < \frac{1}{2}$ and $C_0 > 0$. We define $\mathcal{A}^{\rightarrow} = \bigcup_{n\in\mathbb{N}} \mathcal{A}^n$.

Remark. The constants in $O(h^\infty)$ depend on $\boldsymbol{p}$ and $\boldsymbol{q}$ only through C_0, δ, δ_1, not on the precise values of $\boldsymbol{p}$ and $\boldsymbol{q}$. It will always be the case in the following and we won't mention it anymore. As already done, all the quasiglobal constants (i.e., depending on global parameters and $C_0, \delta, \tau, \delta_1$) will be noted by the letter C.

Lemma 4.12. *Let $\boldsymbol{q}, \boldsymbol{p} \in \mathcal{A}^{\rightarrow}$ satisfying $\mathcal{V}_{\boldsymbol{q}}^+ \cap \mathcal{V}_{\boldsymbol{p}}^- = \varnothing$ and $\max(J_{\boldsymbol{q}}^+, J_{\boldsymbol{p}}^-) \leq C_0 h^{-\delta}$. Then*

$$U_{\boldsymbol{p}} U_{\boldsymbol{q}} = O(h^\infty)_{L^2\to L^2}.$$

Proof. By virtue of Proposition 4.8, we can write

$$U_{\boldsymbol{p}} = T^{|\boldsymbol{p}|}\,\mathrm{Op}_h(a_{\boldsymbol{p}}^-) + O(h^\infty),$$
$$U_{\boldsymbol{q}} = \mathrm{Op}_h(a_{\boldsymbol{q}}^+) T^{|\boldsymbol{q}|} + O(h^\infty).$$

With $a_{\boldsymbol{q}}^+ \in \|\alpha\|_\infty^{|\boldsymbol{q}|} S_{\delta_1}^{\mathrm{comp}}$, $a_{\boldsymbol{p}}^- \in \|\alpha\|_\infty^{|\boldsymbol{p}|} S_{\delta_1}^{\mathrm{comp}}$ and $\operatorname{supp} a_{\boldsymbol{p}}^- \subset \mathcal{V}_{\boldsymbol{p}}^-$, $\operatorname{supp} a_{\boldsymbol{q}}^+ \subset \mathcal{V}_{\boldsymbol{q}}^+$. Since $\mathcal{V}_{\boldsymbol{q}}^+ \cap \mathcal{V}_{\boldsymbol{p}}^- = \varnothing$, $\mathrm{Op}_h(a_{\boldsymbol{p}}^-)\,\mathrm{Op}_h(a_{\boldsymbol{q}}^+) = O(h^\infty)$ as a consequence of the composition of two symbols of S_{δ_1}. The constants in $O(h^\infty)$ depend on seminorms of these symbols, themselves depending on C_0, τ, δ_1. Since $T^n = O(1)$, the result is proved. □

Lemma 4.12 will have interesting consequences, starting with the following lemma which enables us to get rid (that is to say to control by $O(h^\infty)$) of words $\boldsymbol{q}$ where $\mathcal{V}_{\boldsymbol{q}}^{\pm} = \varnothing$, under some assumptions. In particular, it can be applied without trouble to words of "small" lengths $N \leq |\log h|/(2\lambda_1)$, which could also be deduced from applying Egorov's theorem up to the global Ehrenfest time $|\log h|/(2\lambda_1)$.

Lemma 4.13. *Let $\boldsymbol{q} \in \mathcal{A}^{\rightarrow}$ such that $n = |\boldsymbol{q}| \leq C_0|\log h|$ and assume that $\mathcal{V}_{\boldsymbol{q}}^- = \varnothing$. We suppose that one of the following assumptions is satisfied*:

(i) *If $m = \max\{k \in \{1, \dots, n\} : \mathcal{V}_{q_0\cdots q_{k-1}}^- \neq \varnothing\}$, then $J_{q_0\cdots q_{m-1}}^- \leq C_0 h^{-2\delta}$.*

(ii) *If $m = \min\{k \in \{0, \dots, n-1\} : \mathcal{V}_{q_m\cdots q_{n-1}}^- \neq \varnothing\}$, then $J_{q_m\cdots q_{n-1}}^- \leq C_0 h^{-2\delta}$.*

Then, $U_{\boldsymbol{q}} = O(h^\infty)$.

Proof. We prove this lemma under assumption (i). This is similar under (ii). We let $m=\max\{k\in\{1,\dots,n\}: \mathcal{V}^-_{q_0\cdots q_{m-1}}\neq\varnothing\}$ and assume $J^-_{q_0\cdots q_{m-1}}\le C_0h^{-2\delta}$. Due to (4-12), it is enough to show that $U_{q_0\cdots q_m}=O(h^\infty)$. Let us define $l=\max\{k\in\{1,\dots,m\}: J^-_{q_0\cdots q_{l-1}}\le h^{-\delta}\}$ and notice that $l<m$ (if h is small enough). By maximality of l, it is clear that $J^-_{q_0\cdots q_l}\ge h^{-\delta}$. According to the third point of Lemma 4.5,

$$J^-_{q_{l+1}\cdots q_{m-1}}\sim\frac{J^-_{q_0\cdots q_{m-1}}}{J^-_{q_0\cdots q_l}}\le Ch^{-\delta}.$$

Set $\boldsymbol{p}=q_l\cdots q_m$. We distinguish now between two cases:

- $\mathcal{V}^-_{\boldsymbol{p}}\neq\varnothing$: We set $\boldsymbol{r}=q_0\cdots q_{l-1}$. It follows that

$$\max(J^-_{\boldsymbol{p}},J^-_{\boldsymbol{r}})\le Ch^{-\delta}.$$

Moreover,

$$\mathcal{V}^-_{\boldsymbol{p}}\cap\mathcal{V}^+_{\boldsymbol{r}}=F^l(\mathcal{V}^-_{q_0\cdots q_m})=\varnothing.$$

By Lemma 4.12, $U_{\boldsymbol{p}}U_{\boldsymbol{r}}=U_{q_0\cdots q_m}=O(h^\infty)$.

- $\mathcal{V}^-_{\boldsymbol{p}}=\varnothing$: This time, we have $\max(J^-_{q_l\cdots q_{m-1}},J^-_{q_m})\le Ch^{-\delta}$ and $\mathcal{V}^-_{q_m}\cap\mathcal{V}^+_{q_l\cdots q_{m-1}}=\varnothing$. According to Lemma 4.12, $U_{q_l\cdots q_m}=U_{q_m}U_{q_l\cdots q_{m-1}}=O(h^\infty)$. It follows that $U_{q_0\cdots q_m}=O(h^\infty)$. ☐

4E2. *Orthogonality of the $U_{\boldsymbol{q}}$.* We now focus on terms $U_{\boldsymbol{q}}U^*_{\boldsymbol{p}}$ and $U^*_{\boldsymbol{q}}U_{\boldsymbol{p}}$ when $\mathcal{V}^+_{\boldsymbol{q}}$ and $\mathcal{V}^+_{\boldsymbol{p}}$ are disjoint, under growth conditions of the Jacobian. The following result shows that the operators $U_{\boldsymbol{q}}$ and $U_{\boldsymbol{p}}$ are (up to $O(h^\infty)$) orthogonal. These estimates will turn out to be important to apply Cotlar–Stein-type estimates.

Proposition 4.14. *Assume that $\boldsymbol{q},\boldsymbol{p}\in\mathcal{A}^\to$ are two words of same length $|\boldsymbol{q}|=|\boldsymbol{p}|=n$ satisfying $\mathcal{V}^+_{\boldsymbol{q}}\cap\mathcal{V}^+_{\boldsymbol{p}}=\varnothing$ and $\max(J^+_{\boldsymbol{q}},J^+_{\boldsymbol{p}})\le C_0h^{-2\delta}$. Then,*

$$U_{\boldsymbol{q}}U^*_{\boldsymbol{p}}=O(h^\infty),$$
$$U^*_{\boldsymbol{q}}U_{\boldsymbol{p}}=O(h^\infty).$$

Before proving it, we need the following lemma, whose proof relies on the iterative construction of the symbols $a^\pm_{\boldsymbol{q}}$.

Lemma 4.15. *Assume $\boldsymbol{q},\boldsymbol{p}\in\mathcal{A}^\to$ are two words of same length $|\boldsymbol{q}|=|\boldsymbol{p}|=n$ satisfying $\max(J^+_{\boldsymbol{q}},J^+_{\boldsymbol{p}})\le C_0h^{-\delta}$. Then,*

$$U_{\boldsymbol{q}}U^*_{\boldsymbol{p}}=\mathrm{Op}_h(a^+_{\boldsymbol{q}})\,\mathrm{Op}_h(a^+_{\boldsymbol{p}})^*+O(h^\infty),$$
$$U^*_{\boldsymbol{q}}U_{\boldsymbol{p}}=\mathrm{Op}_h(a^-_{\boldsymbol{q}})^*\,\mathrm{Op}_h(a^-_{\boldsymbol{p}})+O(h^\infty).$$

Proof of Lemma 4.15. We prove the first equality. The second one could be treated similarly. Recall the construction procedure of Section 4D. We adopt the same notation. We will show by induction on $r\in\{0,\dots,n-1\}$ that

$$V_r:=U_{q_0\cdots q_{r-1}}U^*_{p_0\cdots p_{r-1}}=\mathrm{Op}_h(a^+_{q_0\cdots q_{r-1}})\,\mathrm{Op}_h(a^+_{p_0\cdots p_{r-1}})^*+O(h^\infty).$$

The case $r=1$ follows from

$$MA_{q_0}A^*_{p_0}M^*=\mathrm{Op}_h(a^+_{q_0})TT^*\,\mathrm{Op}_h(a^+_{p_0})^*+O(h^\infty)=\mathrm{Op}_h(a^+_{q_0})\,\mathrm{Op}_h(a^+_{p_0})^*+O(h^\infty),$$

where we use the fact that $TT^* = I + O(h^\infty)$ microlocally in $\mathcal{V}^+_{p_0}$, Assume that the assumption is satisfied for r, namely

$$V_r = \mathrm{Op}_h(a^+_{q_0\cdots q_{r-1}})\,\mathrm{Op}_h(a^+_{p_0\cdots p_{r-1}}) + O(h^\infty),$$

and let's prove it for $r+1$:

$$\begin{aligned} V_{r+1} &= MA_{q_r}V_rA^*_{p_r}M^* \\ &= MA_{q_r}\,\mathrm{Op}_h(a^+_{q_0\cdots q_{r-1}})\,\mathrm{Op}_h(a^+_{p_0\cdots p_{r-1}})^*A^*_{p_r}M^*r + O(h^\infty) \\ &= \mathrm{Op}_h(a^+_{q_0\cdots q_r})TT^*\,\mathrm{Op}_h(a^+_{p_0\cdots p_r})^* + O(h^\infty) \\ &= \mathrm{Op}_h(a^+_{q_0\cdots q_r})\,\mathrm{Op}_h(a^+_{p_0\cdots p_r})^* + O(h^\infty). \end{aligned}$$

The last equality follows from $TT^* = I + O(h^\infty)$ microlocally in $\mathcal{V}^+_{p_r}$ and the one before is due to the recursive construction of the symbols $a^+_{q_0\cdots q_r}$ in the Section 4D. □

Proof of Proposition 4.14. Let us begin with the first equality. Consider the largest integer l such that

$$\max(J^+_{q_0\cdots q_{l-1}}, J^+_{p_0\cdots p_{l-1}}) \le h^{-\delta}.$$

We set $\boldsymbol{q}_\leftarrow = q_0\cdots q_{l-1}$ and $\boldsymbol{q}_\rightarrow = q_l\cdots q_{n-1}$, and define similar notation for $\boldsymbol{p}$. We obviously have

$$U_{\boldsymbol{q}}U^*_{\boldsymbol{p}} = U_{\boldsymbol{q}_\rightarrow}U_{\boldsymbol{q}_\leftarrow}U^*_{\boldsymbol{p}_\leftarrow}U^*_{\boldsymbol{p}_\rightarrow}.$$

We then consider two cases:

- $\mathcal{V}^+_{\boldsymbol{q}_\leftarrow} \cap \mathcal{V}^+_{\boldsymbol{p}_\leftarrow} = \varnothing$: we may write

$$U_{\boldsymbol{q}_\leftarrow}U^*_{\boldsymbol{p}_\leftarrow} = T^l\,\mathrm{Op}_h(a^-_{\boldsymbol{q}_\leftarrow})\,\mathrm{Op}_h(a^-_{\boldsymbol{q}_\leftarrow})^*T^l + O(h^\infty).$$

Since, $\mathcal{V}^-_{\boldsymbol{q}_\leftarrow} \cap \mathcal{V}^-_{\boldsymbol{p}_\leftarrow} = \varnothing$, we can use the composition formula in $S^{\mathrm{comp}}_{\delta_1}$ to conclude $\mathrm{Op}_h(a^-_{\boldsymbol{q}_\leftarrow})\,\mathrm{Op}_h(a^-_{\boldsymbol{q}_\leftarrow})^* = O(h^\infty)$, which gives the desired result, recalling that $U_{\boldsymbol{q}} = O(1)$.

- $\mathcal{V}^+_{\boldsymbol{q}_\leftarrow} \cap \mathcal{V}^+_{\boldsymbol{p}_\leftarrow} \neq \varnothing$: In this case, we use the previous lemma and we can write

$$U_{\boldsymbol{q}_\leftarrow}U^*_{\boldsymbol{p}_\leftarrow} = \mathrm{Op}_h(a^+_{\boldsymbol{q}_\leftarrow})\,\mathrm{Op}_h(a^+_{\boldsymbol{p}_\leftarrow})^* + O(h^\infty).$$

By virtue of the second point of Lemma 4.5, $J^+_{\boldsymbol{q}_\leftarrow} \sim J^+_{\boldsymbol{p}_\leftarrow}$. Moreover, by maximality of l, either $J^+_{\boldsymbol{q}_\leftarrow q_l} > h^{-\delta}$ or $J^+_{\boldsymbol{p}_\leftarrow p_l} > h^{-\delta}$. But

$$J^+_{\boldsymbol{q}_\leftarrow q_l} \sim J^+_{\boldsymbol{q}_\leftarrow}.$$

Hence, $J^+_{\boldsymbol{q}_\leftarrow} \sim h^{-\delta}$. Using now the third point of Lemma 4.5, we conclude that

$$J^+_{\boldsymbol{q}_\rightarrow} \sim J^+_{\boldsymbol{p}_\rightarrow} \sim h^{-\delta}.$$

This estimate allows us to write

$$U_{\boldsymbol{q}}U^*_{\boldsymbol{p}} = T^{n-l}\,\mathrm{Op}_h(a^-_{\boldsymbol{q}_\rightarrow})\,\mathrm{Op}_h(a^+_{\boldsymbol{q}_\leftarrow})\,\mathrm{Op}_h(a^+_{\boldsymbol{p}_\leftarrow})^*\,\mathrm{Op}_h(a^-_{\boldsymbol{p}_\rightarrow})^*(T^*)^{n-l} + O(h^\infty),$$

with all the symbols in $h^{-M}S^{\mathrm{comp}}_{\delta_1}$ for some $M > 0$. To conclude, we use the composition formula in this symbol class, noting that

$$\mathcal{V}^+_{\boldsymbol{q}_\leftarrow} \cap \mathcal{V}^-_{\boldsymbol{q}_\rightarrow} \cap \mathcal{V}^+_{\boldsymbol{p}_\leftarrow} \cap \mathcal{V}^-_{\boldsymbol{p}_\rightarrow} = F^l(\mathcal{V}^-_{\boldsymbol{q}} \cap \mathcal{V}^-_{\boldsymbol{p}}) = \varnothing.$$

To deal with the second equality, we consider the smallest integer l such that

$$\max(J^+_{q_l\cdots q_{n-1}}, J^+_{p_l\cdots p_{n-1}}) \le h^{-\delta}.$$

As before, we write $\boldsymbol{q}_\leftarrow = q_0\cdots q_{l-1}$ and $\boldsymbol{q}_\rightarrow = q_l\cdots q_{n-1}$, and define similar notation for $\boldsymbol{p}$. We obviously have

$$U^*_{\boldsymbol{q}} U_{\boldsymbol{p}} = U^*_{\boldsymbol{q}_\leftarrow} U^*_{\boldsymbol{q}_\rightarrow} U_{\boldsymbol{p}_\rightarrow} U_{\boldsymbol{p}_\leftarrow}.$$

We distinguish the cases $\mathcal{V}^+_{\boldsymbol{q}_\rightarrow} \cap \mathcal{V}^+_{\boldsymbol{p}_\rightarrow} = \varnothing$ and $\mathcal{V}^+_{\boldsymbol{q}_\rightarrow} \cap \mathcal{V}^+_{\boldsymbol{p}_\rightarrow} \ne \varnothing$ and argue similarly. ☐

4F. *Reduction to subwords with precise growth of their Jacobian.* Recall that we are interested in a decay bound for $\|\mathfrak{M}^{N_0+N_1}\|$, where $\mathfrak{M} = M(\mathrm{Id} - A_\infty) = \sum_{q\in\mathcal{A}} MA_q$. For this purpose, we take the decomposition $\mathfrak{M}^{N_1} = \sum_{\boldsymbol{q}\in\mathcal{A}^{N_1}} U_{\boldsymbol{q}}$.

If $\boldsymbol{q} \in \mathcal{A}^{N_1}$, either $\mathcal{V}^+_{\boldsymbol{q}} = \varnothing$, and in this case $J^+_{\boldsymbol{q}} = +\infty$, or $\mathcal{V}^+_{\boldsymbol{q}} \ne \varnothing$, which implies that $J^+_{\boldsymbol{q}} \ge e^{\lambda_1 N_1} \ge h^{-1} \gg h^{-\tau}$. In both cases, the following integer is well-defined:

$$n(\boldsymbol{q}) = \max\{k \in \{1, N_1\} : J^+_{q_{N_1-k}\cdots q_{N_1-1}} \le h^{-\tau}\}. \tag{4-61}$$

We then set $\boldsymbol{q}_\tau = q_{N_1-n(q)-1}\cdots q_{N_1-1}$. The case $\mathcal{V}_{\boldsymbol{q}_\tau} = \varnothing$ is irrelevant. Indeed, if $\boldsymbol{q} \in \mathcal{A}^{N_1}$ and if $\mathcal{V}_{\boldsymbol{q}_\tau} = \varnothing$, then $U_{\boldsymbol{q}} = O(h^\infty)$, as an obvious consequence of Lemma 4.13. Then, we set

$$Q = \{\boldsymbol{q} \in \mathcal{A}^{N_1} : \mathcal{V}_{\boldsymbol{q}_\tau} \ne \varnothing\} \tag{4-62}$$

so that, due to the fact that $|\mathcal{A}^{N_1}| = O(h^{-M})$, for some $M > 0$, we have

$$\mathfrak{M}^{N_1} = \sum_{\boldsymbol{q}\in Q} U_{\boldsymbol{q}} + O(h^\infty).$$

We partition Q in function of the length of $\boldsymbol{q}_\tau$ and the value of q_{N_1-1}. Namely, we set

$$Q_0(n,a) = \{\boldsymbol{q} \in Q : |\boldsymbol{q}_\tau| = n,\ q_{N_1-1} = a\}.$$

We finally set $Q(n,a) = \{\boldsymbol{q}_\tau : \boldsymbol{q} \in Q_0(n,a)\}$, which is simply the set of words $\boldsymbol{q} \in \mathcal{A}^n$ such that $q_{n-1} = a$ and $J^+_{q_1\cdots q_{n-1}} \le h^{-\tau} < J^+_{\boldsymbol{q}}$. Note that every word $\boldsymbol{q} \in Q_0(n,a)$ can be written in the form $\boldsymbol{q} = \boldsymbol{r}\boldsymbol{p}$, with $\boldsymbol{p} \in Q(n,a)$ and $\boldsymbol{r} \in \mathcal{A}^{N_1-n}$. We deduce that, *modulo* $O(h^\infty)$,

$$\mathfrak{M}^{N_1} = \sum_{n=1}^{N_1}\sum_{a\in\mathcal{A}}\sum_{\boldsymbol{q}\in Q_0(n,a)} U_{\boldsymbol{q}} = \sum_{n=1}^{N_1}\sum_{a\in\mathcal{A}}\sum_{\substack{\boldsymbol{p}\in Q(n,a)\\ \boldsymbol{r}\in\mathcal{A}^{N_1-n}}} U_{\boldsymbol{p}} U_{\boldsymbol{r}} = \sum_{n=1}^{N_1}\sum_{a\in\mathcal{A}}\Bigg(\sum_{\boldsymbol{q}\in Q(n,a)} U_{\boldsymbol{q}}\Bigg)\mathfrak{M}^{N_1-n}.$$

As a consequence, we get

$$\|\mathfrak{M}^{N_0+N_1}\| \le CN_1|\mathcal{A}| \sup_{\substack{1\le n\le N_1\\ a\in\mathcal{A}}} \|\mathfrak{M}^{N_0} U_{Q(n,a)}\| (\|\alpha\|_\infty)^{N_1-n}, \tag{4-63}$$

where

$$U_{Q(n,a)} = \sum_{\boldsymbol{q}\in Q(n,a)} U_{\boldsymbol{q}}. \tag{4-64}$$

Since $N_1 = O(\log h)$, the proof of (4-14) is reduced to proving:

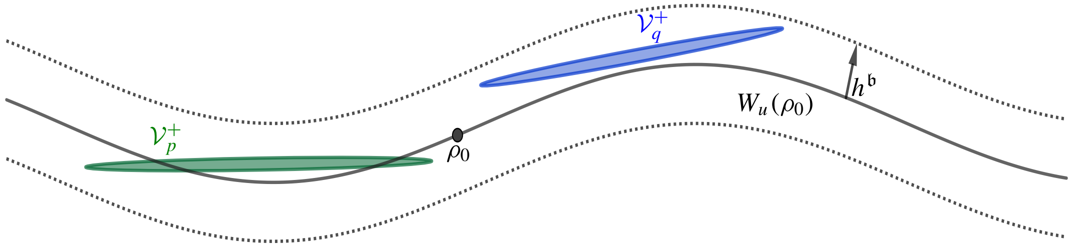

Figure 12. Two words $\boldsymbol{q}, \boldsymbol{p} \in Q(n, a)$ are close to each other if $\mathcal{V}_{\boldsymbol{q}}^+$ and $\mathcal{V}_{\boldsymbol{p}}^+$ lie in the $h^{\mathfrak{b}}$-neighborhood of the same unstable leaves, as stated in Definition 4.17.

Proposition 4.16. *There exists $\gamma > 0$ such that, for h small enough, we have*

$$\sup_{\substack{1 \leq n \leq N_1 \\ a \in \mathcal{A}}} \frac{\|\mathfrak{M}^{N_0} U_{Q(n,a)}\|}{\|\alpha\|_\infty^{n+N_0}} \leq h^\gamma. \tag{4-65}$$

4G. *Partition into clouds.* We fix $1 \leq n \leq N_1$ and $a \in \mathcal{A}$. We aim at gathering pieces of $\mathfrak{M}^{N_0} U_{Q(n,a)}$ into clouds and we want these clouds to interact (with a meaning we will define further) with only a finite and uniform number of other clouds, so that the global norm of $\|\mathfrak{M}^{N_0} U_{Q(n,a)}\|$ can be deduced from a uniform bound for each cloud.

Recall that δ_0 and τ (see (4-2), (4-3) and (4-5)) have be chosen such that

$$\mathfrak{b} + \delta_0 < 1, \quad \mathfrak{b} < \tau.$$

We start by defining a notion of closeness between two words $\boldsymbol{q}, \boldsymbol{p} \in Q(n, a)$. We choose ε_2 as in Lemma 4.4.

Definition 4.17. Let $\boldsymbol{q}, \boldsymbol{p} \in Q(n, a)$. We say that these two words are *close to each other* if there exists $\rho_0 \in \mathcal{T} \cap F(\mathcal{V}_a(\varepsilon_2))$ such that,

$$\text{for all } \rho \in \mathcal{V}_{\boldsymbol{q}}^+ \cup \mathcal{V}_{\boldsymbol{p}}^+, \quad d(\rho, W_u(\rho_0)) \leq h^{\mathfrak{b}}.$$

Otherwise, we say that $\boldsymbol{q}$ and $\boldsymbol{p}$ are *far from each other*. See Figure 12.

Remark. By the definition of $\mathcal{V}_{\boldsymbol{q}}^+$, if $\boldsymbol{q} \in \mathcal{Q}(n, a)$ and if $\rho \in \mathcal{V}_{\boldsymbol{q}}^+$, then ρ does not lie in $\mathcal{V}_a$, but $F^{-1}(\rho)$ does. Hence, we work with $F(\mathcal{V}_a)$ instead of $\mathcal{V}_a$. Moreover, the set $F(\mathcal{V}_a(\varepsilon_2))$ is chosen to fit well in the computations below and in particular in the proof of Lemma 4.19. We could replace it by $\mathcal{V}_a^+(C\varepsilon_2)$, where C is any Lipschitz constant for F.

The important fact on words $\boldsymbol{p}, \boldsymbol{q}$ far from each other is that the associated operators $\mathfrak{M}^{N_0} U_{\boldsymbol{p}}$, $\mathfrak{M}^{N_0} U_{\boldsymbol{q}}$ are almost orthogonal:

Proposition 4.18. *Assume that $\boldsymbol{q}, \boldsymbol{p} \in Q(n, a)$ are far from each other. Then,*

$$(\mathfrak{M}^{N_0} U_{\boldsymbol{q}})^* (\mathfrak{M}^{N_0} U_{\boldsymbol{p}}) = O(h^\infty), \tag{4-66}$$

$$(\mathfrak{M}^{N_0} U_{\boldsymbol{q}}) (\mathfrak{M}^{N_0} U_{\boldsymbol{q}})^* = O(h^\infty). \tag{4-67}$$

We will need the following lemma.

Lemma 4.19. *If $\boldsymbol{q}, \boldsymbol{p} \in Q(n,a)$ are far from each other, there exist words $\boldsymbol{p}_1, \boldsymbol{q}_1, \boldsymbol{p}_2, \boldsymbol{q}_2$ such that*

- $|\boldsymbol{p}_1| = |\boldsymbol{q}_1|, |\boldsymbol{p}_2| = |\boldsymbol{q}_2|$.
- $\boldsymbol{q} = \boldsymbol{q}_1\boldsymbol{q}_2, \ \boldsymbol{p} = \boldsymbol{p}_1\boldsymbol{p}_2$.
- $\mathcal{V}^+_{\boldsymbol{q}_2} \cap \mathcal{V}^+_{\boldsymbol{p}_2} = \varnothing$.
- $\max(J^+_{\boldsymbol{q}_2}, J^+_{\boldsymbol{p}_2}) \le Ch^{-\mathfrak{b}}$ (*for some global constant $C > 0$*).

In particular, $\mathcal{V}^+_{\boldsymbol{q}} \cap \mathcal{V}^+_{\boldsymbol{p}} = \varnothing$.

Let's momentarily admit it and prove the proposition.

Proof of Proposition 4.18. Fix $\boldsymbol{q}, \boldsymbol{p} \in Q(n,a)$ far from each other. Since $\mathcal{V}^+_{\boldsymbol{q}} \cap \mathcal{V}^+_{\boldsymbol{p}} = \varnothing$, we have $U_{\boldsymbol{q}} U^*_{\boldsymbol{p}} = O(h^\infty)$ by virtue of Proposition 4.14. Hence, using the polynomial bounds $\|\mathfrak{M}^{N_0}\| = O(h^{-M})$ (for some $M > 0$), we have

$$(\mathfrak{M}^{N_0} U_{\boldsymbol{q}})(\mathfrak{M}^{N_0} U_{\boldsymbol{p}})^* = O(h^\infty).$$

To prove the first point, we write

$$(\mathfrak{M}^{N_0} U_{\boldsymbol{q}})^*(\mathfrak{M}^{N_0} U_{\boldsymbol{p}}) = \sum_{\boldsymbol{s},\boldsymbol{t} \in \mathcal{A}^{N_0}} U^\star_{\boldsymbol{q}_1} U^*_{\boldsymbol{q}_2} U^*_{\boldsymbol{s}} U_{\boldsymbol{t}} U_{\boldsymbol{p}_2} U_{\boldsymbol{p}_1}.$$

Hence, it is enough to show that $U^*_{\boldsymbol{q}_2} U^*_{\boldsymbol{s}} U_{\boldsymbol{t}} U_{\boldsymbol{p}_2} = O(h^\infty)$ uniformly in $\boldsymbol{s}, \boldsymbol{t}$. To do so, we note that

$$\mathcal{V}^+_{\boldsymbol{q}_2\boldsymbol{s}} \cap \mathcal{V}^+_{\boldsymbol{p}_2\boldsymbol{t}} \subset F^{N_0}(\mathcal{V}^+_{\boldsymbol{q}_2} \cap \mathcal{V}^+_{\boldsymbol{p}_2}) = \varnothing,$$
$$J^+_{\boldsymbol{q}_2\boldsymbol{s}} \le C J^+_{\boldsymbol{s}} J^+_{\boldsymbol{q}_2} \le C e^{\lambda_1 N_0} h^{-\mathfrak{b}} \le C h^{-(\delta_0 + \mathfrak{b})},$$
$$J^+_{\boldsymbol{p}_2\boldsymbol{t}} \le C h^{-(\delta_0+\mathfrak{b})}$$

and apply Proposition 4.14, with $\delta = (\delta_0 + \mathfrak{b})/2 < \frac{1}{2}$ (here we use condition (4-3)). □

We now prove the lemma.

Proof of Lemma 4.19. Consider $\boldsymbol{q}, \boldsymbol{p} \in Q(n,a)$ far from each other. Consider the smallest integer m such that $\mathcal{V}^+_{q_m \cdots q_{n-1}} \cap \mathcal{V}^+_{p_m \cdots p_{n-1}} \neq \varnothing$. We will show that $m > 0$ and set $\boldsymbol{p}_2 = p_{m-1} \cdots p_{n-1}$, $\boldsymbol{q}_2 = q_{m-1} \cdots q_{n-1}$. Pick $\rho \in \mathcal{V}^+_{q_m \cdots q_{n-1}} \cap \mathcal{V}^+_{p_m \cdots p_{n-1}}$. By choice of ε_2 after Lemma 4.4, there exists $\rho_0 \in \mathcal{T}$ such that $d(F^{-i}(\rho), F^{-i}(\rho_0)) \le \varepsilon_2$ for $i \in \{1, \dots, n-m\}$. In particular, $d(F^{-1}(\rho), F^{-1}(\rho_0)) \le \varepsilon_2$ and $F^{-1}(\rho) \in \mathcal{V}_a$, so that $\rho_0 \in F(\mathcal{V}_a(\varepsilon_2))$. Since, $\boldsymbol{q}, \boldsymbol{p}$ are far from each other, there exists $\rho_1 \in \mathcal{V}^+_{\boldsymbol{q}} \cup \mathcal{V}^+_{\boldsymbol{p}}$ such that $d(\rho_1, W_u(\rho_0)) > h^{\mathfrak{b}}$ (otherwise, it would contradict Definition 4.17).

Suppose for instance that $\rho_1 \in \mathcal{V}^+_{\boldsymbol{q}} \subset \mathcal{V}^+_{q_m \cdots q_{n-1}}$. Hence, $d(F^{-i}(\rho_0), F^{-i}(\rho_1)) \le 2\varepsilon_0 + \varepsilon_2$ for $i \in \{1, \dots, n-m\}$. From (3-17), $d(\rho_1, W_u(\rho_0)) \le C(J^{n-m}_s(\rho_0))^{-1}$ and hence, $J^{n-m}_s(\rho_0) \le Ch^{-\mathfrak{b}}$.

But, $J^{n-m}_s(\rho_0) \sim J^+_{p_m \cdots p_{n-1}} \sim J^+_{q_m \cdots q_{n-1}}$, which gives

$$\max(J^+_{p_m \cdots p_{n-1}}, J^+_{q_m \cdots q_{n-1}}) \le Ch^{-\mathfrak{b}}.$$

Since $\min(J^+_{\boldsymbol{q}}, J^+_{\boldsymbol{p}}) > h^{-\tau} \gg h^{-\mathfrak{b}}$ (here we use (4-5)), we cannot have $m = 0$ (if h small enough). Thus, we can set $\boldsymbol{p}_2 = p_{m-1} \cdots p_{n-1}$, $\boldsymbol{q}_2 = q_{m-1} \cdots q_{n-1}$, which satisfy the required properties by minimality of m. □

We now decompose $U_{Q(n,a)}$ into a sum of operators, each of them corresponding to a *cloud* of words. In the following, we'll use the term *cloud* to mean a subset $\mathcal{Q} \subset Q(n,a)$ and we'll adopt the notation

$$\mathcal{V}^+_{\mathcal{Q}} = \bigcup_{\boldsymbol{q} \in \mathcal{Q}} \mathcal{V}^+_{\boldsymbol{q}}$$

and the definition:

Definition 4.20. We say that two clouds $\mathcal{Q}_1, \mathcal{Q}_2$ *do not interact* if, for all pairs $(\boldsymbol{q}_1, \boldsymbol{q}_2) \in \mathcal{Q}_1 \times \mathcal{Q}_2$, $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ are far from each other.

The existence of such a decomposition follows from the key proposition (see Figure 13):

Proposition 4.21. *Suppose ε_0 is small enough. There exists a partition of $Q(n,a)$ into clouds $\mathcal{Q}_1, \dots, \mathcal{Q}_r$ and a global constant $C > 0$ such that, for $i = 1, \dots, r$:*

(i) *There exists $\rho_i \in \mathcal{T}$ such that, for all $\rho \in \mathcal{V}^+_{\mathcal{Q}_i}$, $d(\rho, W_u(\rho_i)) \le Ch^{\mathfrak{b}}$.*

(ii) *If $\mathcal{Q}_i$ interacts with exactly c_i clouds, then $c_i \le C$.*

Remark. Actually, r and the clouds $\mathcal{Q}_i$ depend on n and a. We do not write this dependence explicitly here to make the notation lighter. The second point is relevant since a priori, the only obvious bound on $r = r(n,a)$ is $|r| \le |\mathcal{A}|^n$, where $n = O(\log h)$.

Proof. Keeping in mind that, for all $\boldsymbol{q} \in Q(n,a)$, we have $\mathcal{V}^+_{\boldsymbol{q}} \subset \mathcal{V}^+_a$, we fix $\rho_a \in \mathcal{V}^+_a$. If ε_0 is small enough, $\mathcal{V}^+_a$ does not intersect the boundaries of $W_s(\rho_a)$ and $W_u(\rho_a)$.

For $\boldsymbol{q} \in Q(n,a)$, there exists $\rho_{\boldsymbol{q}} \in \mathcal{T}$ such that $d(F^{-i}(\rho), F^{-i}(\rho_{\boldsymbol{q}})) \le \varepsilon_2$ for all $\rho \in \mathcal{V}^+_{\boldsymbol{q}}$ and for $i = 1, \dots, n$, according to Lemma 4.4 and since $J^+_{\boldsymbol{q}} \sim h^{\tau}$,

$$d(\rho, W_u(\rho_{\boldsymbol{q}})) \le Ch^{-\tau},$$

$d(\rho_a, \rho_{\boldsymbol{q}}) \le C(\varepsilon_2 + \varepsilon_0)$ and hence, if ε_0 is small enough, $z_{\boldsymbol{q}} := H^u_{\rho_a}(\rho_{\boldsymbol{q}})$ (here, $H^u_{\rho_a} : B(\rho_a, \varepsilon_0') \to W_s(\rho_a)$) is the unstable holonomy map defined before Lemma 3.20) is well-defined, and depends Lipschitz-continuously on $\rho_{\boldsymbol{q}}$ (with global Lipschitz constant).

Next, consider a maximal subset $\{z_1, \dots, z_r\} \subset \{z_{\boldsymbol{q}}, \boldsymbol{q} \in Q(n,a)\}$ which is $h^{\mathfrak{b}}$ separated. By maximality, for every $\boldsymbol{q} \in Q(n,a)$, there exists $i \in \{1, \dots, r\}$ such that $|z_i - z_{\boldsymbol{q}}| \le h^{\mathfrak{b}}$ and we use these z_i to partition $Q(n,a)$ into clouds $\mathcal{Q}_i$, where for $i \in \{1, \dots, r\}$, $|z_i - z_{\boldsymbol{q}}| \le h^{\mathfrak{b}}$ for all $\boldsymbol{q} \in \mathcal{Q}_i$. We now show that this partition satisfies the required properties.

Let $i \in \{1, \dots, r\}$, $\boldsymbol{q} \in \mathcal{Q}_i$ and $\rho \in \mathcal{V}^+_{\boldsymbol{q}}$. By local uniqueness of the unstable leaves, we may assume that ε_0 is small enough so that $W_u(\rho_{\boldsymbol{q}}) \cap \mathcal{V}^+_a = W_u(z_{\boldsymbol{q}}) \cap \mathcal{V}^+_a$. Hence,

$$d(\rho, W_u(z_{\boldsymbol{q}})) \le Ch^{-\tau}.$$

Since the unstable leaves depend Lipschitz-continuously on $\rho \in \mathcal{T}$, we have

$$d(\rho, W_u(z_i)) \le C|z_i - z_{\boldsymbol{q}}| + Cd(\rho, W_u(z_{\boldsymbol{q}})) \le Ch^{\mathfrak{b}} + Ch^{\tau} \le Ch^{\mathfrak{b}}.$$

This gives (i).

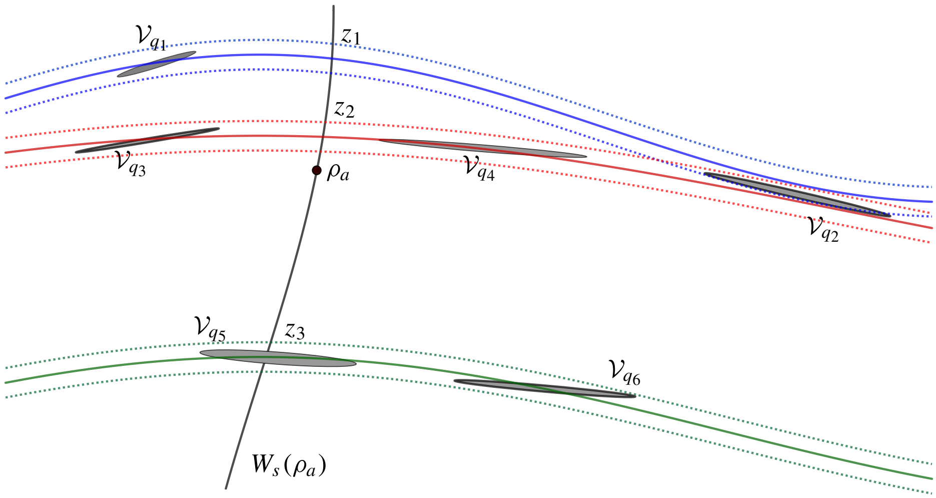

Figure 13. We gather the six small sets $\mathcal{V}_{\boldsymbol{q}}$ into three clouds corresponding to z_1, z_2 and z_3. Here, $\mathcal{Q}_1 = \{\boldsymbol{q}_1\}$, $\mathcal{Q}_2 = \{\boldsymbol{q}_2, \boldsymbol{q}_3, \boldsymbol{q}_4\}$ and $\mathcal{Q}_3 = \{\boldsymbol{q}_5, \boldsymbol{q}_6\}$. The clouds $\mathcal{Q}_1$ and $\mathcal{Q}_2$ interact. The dotted lines draw tubes of width $Ch^{\flat}$ around the unstable leaves $W_u(z_i)$. The sets $\mathcal{V}_{\boldsymbol{q}}$ have width of order h^{τ}.

To show (ii), suppose that $\mathcal{Q}_i$ and $\mathcal{Q}_j$ interact. Then, there exist $(\boldsymbol{q}, \boldsymbol{p}) \in \mathcal{Q}_i \times \mathcal{Q}_j$ and $\rho_0 \in \mathcal{T}$ such that, for all $\rho \in \mathcal{V}_{\boldsymbol{q}}^+ \cup \mathcal{V}_{\boldsymbol{p}}^+$, $d(\rho, W_u(\rho_0)) \le h^{\flat}$. It follows that $d(z_{\boldsymbol{q}}, W_u(\rho_0)) \le Ch^{\tau} + h^{\flat} \le Ch^{\flat}$ and if we denote by $z_0 = H^u_{\rho_a}(\rho_0)$ the unique point in $W_u(\rho_0) \cap W_s(\rho_a)$ then $|z_0 - z_{\boldsymbol{q}}| \le Ch^{\flat}$. The same is true for $\boldsymbol{p}$ and we have $|z_{\boldsymbol{q}} - z_{\boldsymbol{p}}| \le Ch^{\flat}$ and eventually, $|z_i - z_j| \le Ch^{\flat}$. Since $z_1, \dots, z_r$ are $h^{\flat}$-separated, we see after rescaling that the number of j such that $\mathcal{Q}_i$ and $\mathcal{Q}_j$ interact is smaller than the maximal number of points in $B(0, C)$ which are 1-separated (one can for instance bound it by $(2C+1)^2$, but what matters is that it is a global constant). □

This partition into clouds allows us to decompose $\mathfrak{M}^{N_0} U_{Q(n,a)}$ into a sum of operators

$$B_i = \mathfrak{M}^{N_0} U_{\mathcal{Q}_i} = \sum_{\boldsymbol{q} \in \mathcal{Q}_i} \mathfrak{M}^{N_0} U_{\boldsymbol{q}}, \quad \mathfrak{M}^{N_0} U_{Q(n,a)} = \sum_{i=1}^{r} B_i. \tag{4-68}$$

The use of Cotlar–Stein theorem [Zworski 2012, Theorem C.5] reduces the control of the sum by the control of individual clouds:

Lemma 4.22. *With the above notation, there exists a global constant $C > 0$ such that*

$$\|\mathfrak{M}^{N_0} U_{Q(n,a)}\| \le C \sup_{1 \le i \le r} \|B_i\| + O(h^{\infty}). \tag{4-69}$$

Proof. Cotlar–Stein theorem reduces to control

$$\max_i \sum_j \|B_i^* B_j\|^{1/2}, \quad \max_i \sum_j \|B_j B_i^*\|^{1/2}.$$

Fix $i \in \{1, \dots, r\}$.

If $\mathcal{Q}_i$ and $\mathcal{Q}_j$ do not interact, then $\|B_i^* B_j\|^{1/2}$ (resp. $\|B_j B_i^*\|^{1/2}$) is a sum of terms of the form $(\mathfrak{M}^{N_0} U_{\boldsymbol{q}})^* (\mathfrak{M}^{N_0} U_{\boldsymbol{p}})$ (resp. $(\mathfrak{M}^{N_0} U_{\boldsymbol{q}})(\mathfrak{M}^{N_0} U_{\boldsymbol{p}})^*$), where $\boldsymbol{p}$ and $\boldsymbol{q}$ are far from each other. By virtue of Proposition 4.14, these terms are uniformly $O(h^\infty)$ and since the number of terms in the sum grows at most polynomially with h, we can gather all these terms in a single uniform $O(h^\infty)$. As a consequence, we have

$$\begin{aligned}\sum_j \|B_i^* B_j\|^{1/2} &\leq \sum_{\mathcal{Q}_i \text{ and } \mathcal{Q}_j \text{ interact}} \|B_i^* B_j\|^{1/2} + O(h^\infty) \\ &\leq \sum_{\mathcal{Q}_i \text{ and } \mathcal{Q}_j \text{ interact}} \max_k \|B_k\| + O(h^\infty) \leq C \max_k \|B_k\| + O(h^\infty),\end{aligned}$$

and the same holds for the second sum. This gives the desired inequalities. □

The proof of (4-14) and, as a consequence, of Proposition 4.2 is then reduced to the proof of:

Proposition 4.23. *There exists $\gamma > 0$ such that the following holds for h small enough. Assume that $\mathcal{Q} \subset \mathcal{Q}(n, a)$ satisfies, for some global constant $C > 0$,*

$$\text{there exists } \rho_0 \in \mathcal{T} \text{ such that for all } \rho \in \mathcal{V}_{\mathcal{Q}}^+, \quad d(\rho, W_u(\rho_0)) \leq C h^{\mathfrak{b}},$$

where $\mathfrak{b} = 1/(1+\beta)$ is defined in (4-2). *Then,*

$$\frac{\|\mathfrak{M}^{N_0} U_{\mathcal{Q}}\|}{\|\alpha\|_\infty^{N_0+n}} \leq h^\gamma.$$

5. Reduction to a fractal uncertainty principle via microlocalization properties

In this section, we reduce the proof of Proposition 4.23 to a fractal uncertainty principle. To do so, we aim at showing microlocalization properties of the operators involved. The dissymmetry between N_0 and N_1 in the decomposition $N = N_0 + N_1$ will appear clearly in this section. Since N_0 is below the Ehrenfest time, we can actually use semiclassical tools. By contrast, things are more complicated for operators $U_{\boldsymbol{q}}$, with $\boldsymbol{q} \in \mathcal{Q}(n, a)$, and we'll use methods of propagation of Lagrangian leaves. These methods are inspired by [Anantharaman and Nonnenmacher 2007a; 2007b; Nonnenmacher and Zworski 2009] and are also used in [Dyatlov et al. 2022].

5A. *Microlocalization of* $\mathfrak{M}^{N_0}$.

We first state a microlocalization result for $\mathfrak{M}^{N_0}$. This is the easiest one to obtain since N_0 is below the Ehrenfest time. We recall the definition of $\mathcal{T}_-$, the set of the future trapped points

$$\mathcal{T}_- = \bigcap_{n \in \mathbb{N}} F^{-n}(U)$$

and focus on $\mathcal{T}_-^{\text{loc}} := \mathcal{T}_- \cap \mathcal{T}(4\varepsilon_0)$. The set $\mathcal{T}_-$ is laminated by the weak global stable leaves. Hence, if ε_0 is small enough, ensuring that the boundaries of the local stable leaves $W_s(\rho)$, $\rho \in \mathcal{T}$, do not intersect $\mathcal{T}(4\varepsilon_0)$, we have

$$\mathcal{T}_-^{\text{loc}} \subset \bigcup_{\rho \in \mathcal{T}} W_s(\rho).$$

When $\boldsymbol{q} \in \mathcal{A}^{N_0}$ and $\mathcal{V}_{\boldsymbol{q}}^- \neq \varnothing$, $\mathcal{V}_{\boldsymbol{q}}^-$ lies in an $O(h^{\delta_0\lambda_0/\lambda_1})$ neighborhood of a stable leaves, as stated in the following lemma. In the following, we write

$$\delta_2 = \delta_0 \frac{\lambda_0}{\lambda_1}. \tag{5-1}$$

We recall that we have defined $\mathfrak{b}$ in (4-2) and τ in (4-6) such that $\alpha < \tau < 1$ and $\delta_2 + \tau > 1$ (see (4-5)). Moreover, $N_0 = \lceil (\delta_0/\lambda_1)|\log h| \rceil$.

Lemma 5.1. *There exists a global constant $C_2 > 0$ such that, for all $\boldsymbol{q} \in \mathcal{A}^{N_0}$ satisfying $\mathcal{V}_{\boldsymbol{q}}^- \neq \varnothing$,*

$$d(\mathcal{V}_{\boldsymbol{q}}^-, \mathcal{T}_-^{\mathrm{loc}}) \leq C_2 h^{\delta_2}.$$

Remark. In the end of this section, the use of C_2 will always refer to the constant appearing in this lemma. On other places, we keep our convention on global constants, denoting them always by C.

Proof. We already know by Lemma 4.7 that there exists $C > 0$ such that if $\mathcal{V}_{\boldsymbol{q}}^- \neq \varnothing$, there exists $\rho_0 \in \mathcal{T}$ such that

$$d(\mathcal{V}_{\boldsymbol{q}}^-, W_s(\rho_0)) \leq \frac{C}{J_{\boldsymbol{q}}^-}.$$

But $J_{\boldsymbol{q}}^- \geq e^{\lambda_0 N_0} \geq C^{-1} h^{-\delta_0\lambda_0/\lambda_1}$. Finally, $d(\mathcal{V}_{\boldsymbol{q}}^-, \mathcal{T}_-^{\mathrm{loc}}) \leq C h^{\delta_2}$, as required. □

The following lemma allows us to construct symbols in nice symbol classes with supports in h^δ neighborhood. Its proof can be found in [Dyatlov and Zahl 2016, Lemma 3.3].

Lemma 5.2. *Let $\varepsilon > 0$ and $\delta \in \left[0, \frac{1}{2}\right[$. Let $V_0(h) \subset V_1(h) \subset \mathbb{R}^d$ be sets depending on h and assume that, for $0 \leq h \leq 1$, $d(V_0(h), V_1(h)^c) > \varepsilon h^\delta$. Then, there exists a family $\chi_h \in C_c^\infty(\mathbb{R}^d)$ such that, for all $h \leq 1$:*

- *$\chi_h = 1$ on $V_0(h)$.*
- supp $\chi \subset V_1(h)$.
- *For every $\alpha \in \mathbb{N}^d$, there exists C_α depending only on ε such that, for all $x \in \mathbb{R}^d$ and for all $0 < h \leq 1$,*

$$|\partial^\alpha \chi_h(x)| \leq C_\alpha h^{-\delta|\alpha|}.$$

Applying this lemma with $V_0(h) = \mathcal{T}_-^{\mathrm{loc}}(2C_2 h^{\delta_2})$, $V_1(h) = \mathcal{T}_-^{\mathrm{loc}}(4C_2 h^{\delta_2})$ with $\varepsilon = 2C_2$, we consider a family of smooth cut-offs $\chi_h \in S_{\delta_2}^{\mathrm{comp}}$ and we can consider it as an element of $S_{\delta_2}^{\mathrm{comp}}(U)$ since (at least for h small enough) the support of χ_h is included in U. We are now ready to state the microlocalization property of $\mathfrak{M}^{N_0}$.

Proposition 5.3.

$$\mathfrak{M}^{N_0} = \mathfrak{M}^{N_0} \operatorname{Op}_h(\chi_h) + O(h^\infty)_{L^2(Y) \to L^2(Y)}. \tag{5-2}$$

Proof. We need to show that $\mathfrak{M}^{N_0}(\operatorname{Op}_h(1 - \chi_h)) = O(h^\infty)$. To do so, we take the decomposition $\mathfrak{M}^{N_0} = \sum_{\boldsymbol{q} \in \mathcal{A}^{N_0}} U_{\boldsymbol{q}}$. Since the number of terms in this sum grows polynomially with h, it is enough to show that,

$$\text{for all } \boldsymbol{q} \in \mathcal{A}^{N_0}, \quad U_{\boldsymbol{q}}(\operatorname{Op}_h(1 - \chi_h)) = O(h^\infty),$$

with bounds uniform in $\boldsymbol{q}$. We then consider two cases:

- $\mathcal{V}_{\boldsymbol{q}}^{-} = \varnothing$: Lemma 4.13 applies. Indeed, if $m \leq N_0$ and $\mathcal{V}_{q_0\cdots q_{m-1}}^{-} \neq \varnothing$, we have

$$J_{q_0\cdots q_{m-1}}^{-} \leq e^{m\lambda_1} \leq e^{N_0\lambda_1} \leq Ch^{-\delta_0}.$$

Hence, $U_{\boldsymbol{q}} = O(h^{\infty})$, with global constants in the $O(h^{\infty})$.

- $\mathcal{V}_{\boldsymbol{q}}^{-} \neq \varnothing$: We apply Proposition 4.8. Since $J_{\boldsymbol{q}}^{-} \leq Ce^{\lambda_1 N_0} \leq Ch^{-\delta_0}$, we take some $\delta_1 \in \left]\delta_0, \frac{1}{2}\right[$ (in particular, $\delta_2 < \delta_1$) and we can write $U_{\boldsymbol{q}} = T^{N_0} \operatorname{Op}_h(a_{\boldsymbol{q}}^{-}) + O(h^{\infty})$, with $a_{\boldsymbol{q}}^{-} \in S_{\delta_1}^{\text{comp}}(U)$ and $\operatorname{supp} a_{\boldsymbol{q}}^{-} \subset \mathcal{V}_{\boldsymbol{q}}^{-}$. Noticing that $\chi_h = 1$ on $\mathcal{V}_{\boldsymbol{q}}^{-} \subset \mathcal{T}_{-}^{\text{loc}}(2C_2 h^{\delta_2})$, the composition formula in $S_{\delta_1}^{\text{comp}}$ implies that $\operatorname{Op}_h(a_{\boldsymbol{q}}^{-}) \operatorname{Op}_h(1 - \chi_h) = O(h^{\infty})$. Since the seminorms of $a_{\boldsymbol{q}}^{-}$ are uniformly bounded in $\boldsymbol{q}$, the constants appearing in $O(h^{\infty})$ are uniform in $\boldsymbol{q}$. □

5B. ***Propagation of Lagrangian leaves and Lagrangian states.*** To study the microlocalization of $U_{\boldsymbol{q}}$ we'll use the same strategy as in [Dyatlov et al. 2022], the authors themselves inspired by [Anantharaman and Nonnenmacher 2007a; 2007b; Nonnenmacher and Zworski 2009]. We cannot show that $U_{\boldsymbol{q}}$ is a Fourier integral operator since the propagation goes behind the Ehrenfest time. Instead, we show a weaker result which will be enough for our purpose. The idea is to decompose a state u in a sum of Lagrangian states associated with Lagrangian leaves almost parallel to unstable leaves, what we will call horizontal leaves (because we will consider them in charts where the unstable leaves are close to be horizontal). Studying the precise behavior of these states, we can get fine information on the microlocalization of $U_{\boldsymbol{q}}u$. Roughly speaking, we'll show that if u is a Lagrangian state associated with an original horizontal Lagrangian $\mathcal{L}_{q_0,\theta} \subset \mathcal{V}_{q_0}$, then $U_{\boldsymbol{q}}u$ is a Lagrangian state associated with the piece of the evolved Lagrangian $F^n(\mathcal{L}_{q_0,\theta})$ inside $\mathcal{V}_{\boldsymbol{q}}^{+}$.

To define "horizontal" Lagrangian leaves, we need to work in adapted coordinate charts in which the notion of horizontality (thinking $W_u(\rho)$ as the reference) makes sense. For this purpose, for $q \in \mathcal{A}$, we consider charts centered around the points ρ_q, associated with the fixed macroscopic partition of $\mathcal{T}$ by the $\mathcal{V}_q = B(\rho_q, 2\varepsilon_0)$. First, we consider symplectic maps

$$\kappa_q : W_q \subset U_{k_q} \to V_q \subset \mathbb{R}^2$$

satisfying (we denote by (x, ξ) the variable in U and (y, η) in $\mathbb{R}^2$):

(1) $B(\rho_q, C\varepsilon_0) \subset W_q$ for some global constant $C \gg 2$.

(2) $\kappa(\rho_q) = 0,\ d\kappa(\rho_q)(E_u(\rho_q)) = \mathbb{R} \times \{0\} : d\kappa(\rho_q)(E_s(\rho_q)) = \{0\} \times \mathbb{R}$.

(3) The image of the unstable leave $W_u(\rho_q)$ is exactly $\{(y, 0) : y \in \mathbb{R}\} \cap \widetilde{V}_q$.

Theses charts are for instance given by Lemma 3.22 (at this stage, the strong straightening property is not necessary). In these adapted charts where $W_u(\rho_q)$ coincides with $\mathbb{R} \times \{0\}$, the horizontal Lagrangian leaves will be the of the form

$$\mathcal{C}_\theta := \{(y, \theta) : y \in \mathbb{R}\}. \tag{5-3}$$

Finally, we fix unit vectors on $E_u(\rho_q)$ and $E_s(\rho_q)$, $e_u(\rho_q)$ and $e_s(\rho_q)$, used to defined the unstable and stable Jacobians in Section 3C. Let's write

$$d\kappa_q(e_u(\rho_q)) = (\lambda_{q,u}, 0), \quad d\kappa_q(e_s(\rho_q)) = (0, \lambda_{q,s}).$$

Note

$$D_q = \begin{pmatrix} \lambda_{q,u} & 0 \\ 0 & \lambda_{q,s} \end{pmatrix}.$$

We dilate the chart $\tilde{\kappa}_q$ and define

$$\tilde{\kappa}_q : \rho \in W_q \mapsto D_q \kappa_q(\rho) \in \widetilde{V}_q := D_q(V_q).$$

5B1. *Horizontal Lagrangian and their evolution.* Let us fix a word $\boldsymbol{q} \in \mathcal{A}^n$ and let us define

$$\mathcal{L}_{q_0,\theta} = \kappa_{q_0}^{-1}(\mathcal{C}_\theta \cap V_{q_0}) \cap \mathcal{V}_{q_0}. \tag{5-4}$$

Then, let's define inductively

$$\mathcal{L}_{q_0 \cdots q_j,\theta} = F(\mathcal{L}_{q_0 \cdots q_{j-1},\theta}) \cap \mathcal{V}_{q_j}, \tag{5-5}$$

which allows us to define $\mathcal{L}_{\boldsymbol{q},\theta}$. One can check that

$$\mathcal{L}_{\boldsymbol{q},\theta} = F^{-1}(\mathcal{V}_{\boldsymbol{q}}^+) \cap F^{n-1}(\mathcal{L}_{q_0,\theta}). \tag{5-6}$$

The term F^{-1} comes from the definition of $\mathcal{V}_{\boldsymbol{q}}^+$:

$$\rho \in \mathcal{V}_{\boldsymbol{q}}^+ \iff \text{for all } 1 \le i \le n, \quad F^{-i}(\rho) \in \mathcal{V}_{q_{n-i}}.$$

Finally, let's define

$$\mathcal{C}_{\boldsymbol{q},\theta} = \kappa_{q_{n-1}}(\mathcal{L}_{\boldsymbol{q},\theta}). \tag{5-7}$$

We first focus on one step of the iterative process.

In $\widetilde{V}_q \subset \mathbb{R}^2$, we use the notation $\widetilde{B}_q(0,r)$ for the cube$]-r,r[\times]-r,r[$. We keep the subscript q to keep track of the chart in which this cube is supposed to live. Finally, we set

$$B_q(0,r) = D_q^{-1}(\widetilde{B}_q(0,r)) \subset V_q.$$

$B_q(0,r)$ is simply a rectangle centered at zero with size only depending on q (this is also a ball for some norm in $\mathbb{R}^2$). The advantage of $\widetilde{B}_q$ and $\tilde{\kappa}_q$ compared with B_q and κ_q will appear below. However, $\tilde{\kappa}_q$ is not symplectic, and for further use, it is not possible to use $\tilde{\kappa}_q$ as a symplectic change of coordinates.

Let $q, p \in \mathcal{A}$ and suppose that $\mathcal{V}_q \cap F^{-1}(\mathcal{V}_p) \neq \varnothing$. As a consequence there exists a global constant $C' > 0$ such that $d(F(\rho_q), \rho_p) \le C'\varepsilon_0$ and if C in (1) of Lemma 3.22 is large enough, we can assume that, for some global constant $C_1 > 0$,

$$\kappa_q(\mathcal{V}_q) \subset B_q(0, C_1\varepsilon_0) \subset V_q, \quad \kappa_p \circ F \circ \kappa_q^{-1}(B_q(0, C_1\varepsilon_0)) \subset V_p. \tag{5-8}$$

The following map is hence well-defined:

$$\tau_{p,q} := \kappa_p \circ F \circ \kappa_q^{-1} : B_q(0, C_1\varepsilon_0) \to \tau_{p,q}(B_q(0, C_1\varepsilon_0)) \subset V_p;$$

$\tau_{p,q}$ is nothing but the writing of F between the charts V_q and V_p. Note that since the number of possible transitions is finite, we can assume that C_1 is uniform for all $q, p \in \mathcal{A}$ such that $\mathcal{V}_q \cap F^{-1}(\mathcal{V}_p) \neq \varnothing$.

We also adopt the following definitions and notation:

Definition 5.4. Let $G_q :]-C_1\varepsilon_0, C_1\varepsilon_0[\to]-C_1\varepsilon_0, C_1\varepsilon_0[$ be a smooth map. It represents the horizontal Lagrangian

$$\mathcal{L}_{G_q} := D_q^{-1}\big(\{(y, G_q(y)) : y \in]-C_1\varepsilon_0, C_1\varepsilon_0[\}\big) \subset B_q(0, C_1\varepsilon_0) \subset V_q.$$

We say that such a Lagrangian lies in the γ-unstable cone if

$$\|G_q'\|_\infty \leq \gamma,$$

and we write $G_q \in \mathcal{C}_q^u(C_1\varepsilon_0, \gamma)$.

Remark. This is where the use of $\tilde{\kappa}_q$ and $\widetilde{B}_q$ turns out to be useful; to represent horizontal Lagrangian in V_q, we use the cube $\widetilde{B}_q(0, C_1\varepsilon_0) \subset \widetilde{V}_q$ of fixed size.

With this definition, we show in the following lemma an invariance property of the γ-unstable cones:

Lemma 5.5. *There exist global constants $C > 0$, $C_1 > 0$ such that if ε_0 is sufficiently small, then the following holds*:

For every $G_q \in \mathcal{C}_q^u(C_1\varepsilon_0, C\varepsilon_0)$, there exists $G_p \in \mathcal{C}_p^u(C_1\varepsilon_0, C\varepsilon_0)$ such that:

(i) $\tau_{p,q}(\mathcal{L}_{G_q}) \cap B_p(0, C_1\varepsilon_0) = \mathcal{L}_{G_p}$.

(ii) *For some global constants C_l, $l \geq 2$, we have $\|G_q\|_{C^l} \leq C_l \Longrightarrow \|G_p\|_{C^l} \leq C_l$.*

Moreover, let's define $\phi_{qp} :]-C_1\varepsilon_0, C_1\varepsilon_0[\to \mathbb{R}$ by

$$y_q = \phi_{qp}(y_p) \quad \Longleftrightarrow \quad (y_p, G_p(y_p)) = D_p \circ \tau_{pq} \circ D_q^{-1}(\phi_{qp}(y_p), G_q \circ \phi_{qp}(y_p)).$$

Then, ϕ_{pq} is smooth contracting diffeomorphism onto its image. In particular, there exists a global constant $\nu < 1$ such that $\|\phi_{pq}'\|_\infty \leq \nu$.

Proof. Take C_1 large but fixed (with conditions further imposed) and assume that ε_0 is small enough so that (5-8) holds. Let us define $\lambda_q = J_1^u(\rho_q) > 1$ and $\mu_q = J_1^s(\rho_q) < 1$ and let us fix some global ν satisfying,

$$\text{for all } q \in \mathcal{A}, \quad \max(\lambda_q^{-1}, \mu_q) < \nu < 1.$$

Recall that e_u and e_s are $C^{1,\varepsilon}$ in ρ. We write ∂_y and ∂_η to denote the unit vector of $\mathbb{R} \times \{0\}$ and $\{0\} \times \mathbb{R}$ respectively. We fix a constant $C > 0$ with conditions imposed further and we assume that $\|G_p'\|_\infty \leq C\varepsilon_0$. We let $\tilde{\tau} = D_p \circ \tau_{p,q} \circ D_q^{-1}$ (we drop the subscript for $\tilde{\tau}$ to alleviate the notation). In the computations below, the implied constants in the O are global constants (depending also on the choices on κ_q):

- $\tilde{\tau}(0) = \tilde{\kappa}_p \circ F(\rho_q) = O(\varepsilon_0)$.
- $d\tilde{\tau}(0) = d\tilde{\kappa}_p(F(\rho_q)) \circ dF(\rho_q) \circ [d\tilde{\kappa}_q(\rho_q)]^{-1}$.
- $d\tilde{\tau}(0)(\partial_y) = d\tilde{\kappa}_p(F(\rho_q))(\lambda_q e_u(F(\rho_q))) = \lambda_q(d\tilde{\kappa}_p(\rho_p) + O(\varepsilon_0))(e_u(\rho_p) + O(\varepsilon_0)) = \lambda_q\partial_y + O(\varepsilon_0)$, where we use the Lipschitz regularity of $\rho \mapsto e_u(\rho)$ in the second equality.
- Similarly, $d\tilde{\tau}(0)(\partial_\eta) = \mu_q\partial_\eta + O(\varepsilon_0)$.

(It is here that we use the renormalization of κ_q into $\tilde{\kappa}_q$). Eventually, we use the fact that $\tilde{\tau}-\tilde{\tau}(0)-d\tilde{\tau}(0)=O(C_1\varepsilon_0)_{C^1(B(0,C_1\varepsilon_0))}$ and we get

$$\tilde{\tau}(y,\eta)=(\lambda_q y+y_r(y,\eta),\mu_q\eta+\eta_r(y,\eta)),\quad (y,\eta)\in\widetilde{B}_q(0,C_1\varepsilon_0), \tag{5-9}$$

where $y_r(y,\eta)$ and $\eta_r(y,\eta)$ are $O(C_1\varepsilon_0)_{C^1}$. Before going further, let us show that we can fix C_1 such that

$$(y,\eta)\in\widetilde{B}_q(0,C_1\varepsilon_0)\quad\Longrightarrow\quad|\mu_q\eta+\eta_r(y,\eta))|\le C_1\varepsilon_0. \tag{5-10}$$

To do so, let us note that in fact $\tilde{\tau}-\tilde{\tau}(0)-d\tilde{\tau}(0)=O((C_1\varepsilon_0)^2)_{C^0(B(0,C_1\varepsilon_0))}$ and hence if $(y,\eta)\in\widetilde{B}_q(0,C_1\varepsilon_0)$, we have

$$|\eta_r(y,\eta)|=O(\varepsilon_0)+O((C_1\varepsilon_0)^2)_{C^0(B(0,C_1\varepsilon_0))}\le C'\varepsilon_0(1+C_1^2\varepsilon_0).$$

Assume that C_1 is large enough such that $\nu C_1+C'<C_1(\nu+1)/2$. If $(y,\eta)\in\widetilde{B}_q(0,C_1\varepsilon_0)$, we have

$$|\mu_q\eta+\eta_r(y,\eta)|\le\nu C_1\varepsilon_0+C'\varepsilon_0(1+C_1^2\varepsilon_0)\le\Big(C_1\frac{\nu+1}{2}+C_1^2\varepsilon_0\Big)\varepsilon_0.$$

This fixes C_1. Since C_1 is now a global fixed parameter, we can remove it from the O in the estimates. If ε_0 is small enough, depending on our choice of C_1, (5-10) holds.

To write the image of the leaf as a graph, we observe that, if ε_0 is small enough (depending only on global parameters) the map

$$\psi: y\in\,]-C_1\varepsilon_0,C_1\varepsilon_0[\,\mapsto\lambda_q y+y_r(y,G_q(y))$$

is expanding and we can impose $|\psi'|\ge\nu^{-1}$. In particular, $\operatorname{Im}\psi$ contains an interval of size $2\nu^{-1}C_1\varepsilon_0$. Moreover, $\psi(0)=y_r(0,G_q(0))\le\|y_r\|_{C^1}|G_q(y)|=O(\varepsilon_0^2)$. We claim that if ε_0 is small enough, $\operatorname{Im}\psi$ contains $]-C_1\varepsilon_0,C_1\varepsilon_0[$. Indeed, it suffices to have

$$\nu^{-1}C_1\varepsilon_0-|\psi(0)|\ge C_1\varepsilon_0.$$

But we have

$$C_1\varepsilon_0+|\psi(0)|\le C_1\varepsilon_0(1+O(\varepsilon_0))\le C_1\varepsilon_0\nu^{-1}$$

if $1+O(\varepsilon_0)\le\nu^{-1}$, a condition that can be satisfied if ε_0 is small enough. Hence, $\phi:=\phi_{pq}=\psi^{-1}_{|]-C_1\varepsilon_0,C_1\varepsilon_0[}$ is well-defined and we set

$$G_p(y)=\mu_q G_q(\phi(y))+\eta_r(\phi(y),G_q(\phi(y))),\quad y\in\,]-C_1\varepsilon_0,C_1\varepsilon_0[. \tag{5-11}$$

By definition, it is clear that $\tau_{p,q}(\mathcal{L}_{G_q})\cap B_p(0,C_1\varepsilon_0)=\mathcal{L}_{G_p}$ and $(y,G_p(y))=\tilde{\tau}(\phi(y),G_q(\phi(y)))$. The map ϕ is obviously a smooth contracting diffeomorphism and $\|\phi'\|\le 1/\inf|\psi'(y)|\le\nu$. Moreover, due to (5-10), $|G_p(y)|\le C_1\varepsilon_0$. To prove that $G_p\in\mathcal{C}_p^u(C_1\varepsilon_0,C\varepsilon_0)$, we compute

$$G_p'(y)=\mu_q G_q'(\phi(y))\times\phi'(y)+(\partial_y\eta_r+\partial_\eta\eta_r\times G_q'(\phi(y)))\phi'(y),$$
$$|G_p'(y)|\le\nu^2C\varepsilon_0+O(\varepsilon_0(1+C\varepsilon_0))\nu\le[\nu^2C+\nu C'(1+C\varepsilon_0)]\varepsilon_0$$

for some global $C' > 0$. If we assume $\nu^2 + \varepsilon_0 C'\nu < 1$, which is possible if ε_0 is small enough, then we can choose C large enough satisfying

$$C \times (\nu^2 + \nu C'\varepsilon_0) + \nu C' \leq C.$$

This ensures that $\|G'_p\|_\infty \leq C\varepsilon_0$.

Finally, we prove (ii) by induction on l: The case $l = 1$ is done. Assume that there exists a constant C_l such that $\|G_q\|_{C^l} \leq C_l \implies \|G_p\|_{C^l} \leq C_l$. We want to find a constant C_{l+1} fitting for the C^{l+1} norm. Using (5-11), we see by induction that the $(l+1)$-th derivative of G_p has the form

$$G_p^{(l+1)}(y) = \phi'(y)^{l+1} \times G_q^{(l+1)}(y) \times (1 + \partial_\eta \eta_r(y, \phi(y))) + P_y(G_q(y), \dots, G_q^{(l)}(y)),$$

where $P_y(\tau_0, \dots, \tau_l)$ is a polynomial with smooth coefficients in y. Hence, there exists a constant $M(C_l)$ such that for $y \in \left]-C_1\varepsilon_0, C_1\varepsilon_0\right[$, $|P_y(G_q(y), \dots, G_q^{(l)}(y))| \leq M(C_l)$. Since

$$|\phi'(y)^{l+1}(1 + \partial_\eta \eta_r(y, \phi(y)))| \leq \nu(1 + \varepsilon_0 C') := \nu_1$$

if ε_0 is small enough ensuring that $\nu_1 < 1$, we can take

$$C_{l+1} = \max\left(C_l, \frac{M(C_l)}{1 - \nu_1}\right).$$

Indeed, with such a constant, assuming that $\|G_q\|_{C^{l+1}} \leq C_{l+1}$, we have

$$|G_p^{(l+1)}(y)| \leq C_{l+1}\nu_1 + M(C_l) \leq C_{l+1}. \qquad \square$$

Armed with this lemma, we can now iterate the process and get the following proposition describing the evolution of the Lagrangian $\mathcal{C}_{\boldsymbol{q},\theta}$.

Proposition 5.6. *Assume that ε_0 is small enough. Then, for every $n \in \mathbb{N}^*$, $\boldsymbol{q} \in \mathcal{A}^n$, and $\theta \in \mathbb{R}$, there exists an open subset $I_{\boldsymbol{q},\theta} \subset \mathbb{R}$ and a smooth map $G_{\boldsymbol{q},\theta}$ such that:*

- $\mathcal{C}_{\boldsymbol{q},\theta} = \{(y, G_{\boldsymbol{q},\theta}(y)) : y \in I_{\boldsymbol{q},\theta}\}$.
- $\|G'_{\boldsymbol{q},\theta}\|_\infty \leq C\varepsilon_0$ *for some global constant C.*
- *For every $l \geq 2$, $\|G_{\boldsymbol{q},\theta}\|_{C^l} \leq C_l$ for some global C_l.*
- *If $\phi_{\boldsymbol{q},\theta} : I_{\boldsymbol{q},\theta} \to \mathbb{R}$ is defined by*

$$\kappa_{q_{n-1}} \circ F^{n-1} \circ \kappa_{q_0}^{-1}(\phi_{\boldsymbol{q},\theta}(y), \theta) = (y, G_{\boldsymbol{q},\theta}(y)).$$

Then, for some global constants $C > 0$ and $0 < \nu < 1$, $\|\phi'_{\boldsymbol{q},\theta}\| \leq C\nu^{n-1}$.

Proof. Assume that $\mathcal{L}_{\boldsymbol{q},\theta} \neq \varnothing$; otherwise, there is nothing to prove. In particular, we can restrict our attention to small θ, $|\theta| \leq C_1\varepsilon_0$. As a consequence, for every $i \in \{1, \dots, n\}$, $F(\mathcal{V}_{q_{i-1}}) \cap \mathcal{V}_{q_i} \neq \varnothing$. Hence, we can consider the maps $\tau_i := \tau_{q_i,q_{i-1}}$ and since we assume that $\kappa_{q_i}(\mathcal{V}_{q_i}) \subset B_{q_i}(0, C_1\varepsilon_0)$,

$$C_{q_0 \cdots q_i,\theta} = \tau_i(C_{q_0 \cdots q_{i-1},\theta}) \cap \kappa_{q_i}(\mathcal{V}_{q_i}).$$

We start with a constant function $G_0 \in \mathcal{C}_0^u(C_1\varepsilon_0, 0)$ such that $\mathcal{L}_{G_0} = \mathcal{C}_\theta$ (it suffices to take $G_0 = \lambda_{q_0,s}\theta$) and we inductively apply the previous lemma to show the existence of a family $G_j \in \mathcal{C}_{q_j}^u(C_1\varepsilon_0, C\varepsilon_0)$, $0 \le j \le n-1$, such that:

(i) $\tau_i(\mathcal{L}_{G_i}) \cap B_{q_i}(0, C_1\varepsilon_0) = \mathcal{L}_{G_{i+1}}$.

(ii) $\|G_i\|_{C^l} \le C_l$.

(iii) If we define $\phi_i :]-C_1\varepsilon_0, C_1\varepsilon_0[\to]-C_1\varepsilon_0, C_1\varepsilon_0[$ by

$$(y, G_i(y)) = D_{q_i} \circ \tau_i \circ D_{q_{i-1}}^{-1}(\phi_i(y), G_{i-1} \circ \phi_i(y))$$

then there exists $\nu < 1$ such that $\|\phi_i'\|_\infty \le \nu$.

(iv) $\mathcal{C}_{q_0 \cdots q_i, \theta}$ is an open subset of $\mathcal{L}_{G_i}$.

We have

$$\mathcal{L}_{G_{n-1}} = D_{q_{n-1}}^{-1}(\{(y, G_{n-1}(y)) : y \in]-C_1\varepsilon_0, C_1\varepsilon_0[\}).$$

This can be also written

$$\mathcal{L}_{G_{n-1}} = \{(y, \lambda_{q_{n-1},s}^{-1} G_{n-1}(\lambda_{q_{n-1},u} y)) : |y| < \lambda_{q_{n-1},u}^{-1} C_1\varepsilon_0\}.$$

It suffices to consider

$$\begin{aligned} G_{\boldsymbol{q},\theta}(y) &= \lambda_{q_{n-1},s}^{-1} G_{n-1}(\lambda_{q_{n-1},u} y), \\ I_{\boldsymbol{q},\theta} &= \{y \in]-\lambda_{q_{n-1},u}^{-1} C_1\varepsilon_0, \lambda_{q_{n-1},u}^{-1} C_1\varepsilon_0[: (y, G_{\boldsymbol{q},\theta}(y)) \in \mathcal{C}_{\boldsymbol{q},\theta}\}, \\ \phi_{\boldsymbol{q},\theta}(y) &= \lambda_{q_1,u}^{-1} \phi_1 \circ \cdots \circ \phi_{n-1}(\lambda_{q_{n-1},u} y). \end{aligned}$$ □

5B2. *Evolution of Lagrangian states.* Once we've studied the evolution of the Lagrangian leaves starting from $\mathcal{C}_\theta$, we can study the evolution of the corresponding Lagrangian states. In our case, since the leaves stay rather horizontal, the form of the Lagrangian states we'll consider is the simplest:

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

where a is an amplitude and ψ a generating phase function. It is associated with the Lagrangian,

$$\mathcal{L} = \{(y, \psi'(y)) : y \in \operatorname{supp} a\}.$$

For $q \in \mathcal{A}$, we quantize κ_q. Remind that we denote by k_q the integer such that $\mathcal{V}_q \Subset U_{k_q}$. There exist Fourier integral operators $B_q, B_q' \in I_0^{\mathrm{comp}}(\kappa_q) \times I_0^{\mathrm{comp}}(\kappa_q^{-1})$,

$$B_q : L^2(Y_{k_q}) \to L^2(\mathbb{R}), \quad B_q' : L^2(\mathbb{R}) \to L^2(Y_{k_q})$$

such that they quantize κ_q in a neighborhood of $\kappa_q(\overline{\mathcal{V}}_q) \times \overline{\mathcal{V}}_q$. Moreover, we impose that $\mathrm{WF}_h(B_q B_q')$ is a compact subset of $\mathbb{R}^2$. We will still denote by B_q and B_q' the operators

$$B_q = (0, \dots, \underbrace{B_q}_{k_q}, \dots, 0) : L^2(Y) \to L^2(\mathbb{R}), \quad B_q' = {}^t(0, \dots, \underbrace{B_q'}_{k_q}, \dots, 0) : L^2(\mathbb{R}) \to L^2(Y).$$

If $\operatorname{supp}(c_q) \subset \mathcal{V}_q$ and if C denotes the operator-valued matrix with only one nonzero entry $\mathrm{Op}_h(c_q)$ in position (k_q, k_q), then as operators $L^2(Y) \to L^2(Y)$,

$$B_q' B_q C = C + O(h^\infty), \quad C B_q' B_q = C + O(h^\infty).$$

The proposition we aim at proving is the following:

Proposition 5.7. *Fix $C_0 > 0$. For every $n \in \mathbb{N}$, $\boldsymbol{q} \in \mathcal{A}^n$ and $\theta \in \mathbb{R}$ satisfying*

$$n \le C_0 |\log h|, \quad |\theta| \le C_0 \tag{5-12}$$

and, for every $N \in \mathbb{N}$, there exists a symbol $a_{\boldsymbol{q},\theta,N} \in C_c^\infty(I_{\boldsymbol{q},\theta})$ such that

(i) $U_{\boldsymbol{q}}(B'_{q_0} e^{i(\theta\cdot)/h}) = M A_{q_{n-1}} B'_{q_{n-1}} (e^{i\psi_{\boldsymbol{q}}/h} a_{\boldsymbol{q},\theta,N}) + O(h^N)_{L^2}$,

(ii) $\|a_{\boldsymbol{q},\theta,N}\|_{C_l} \le C_{l,N} h^{-C_0 \log B}$,

(iii) *there exists $\delta > 0$ such that $d(\mathrm{supp}(a_{\boldsymbol{q},\theta,N}), \mathbb{R} \setminus I_{\boldsymbol{q},N,\theta}) \ge \delta$,*

where $\psi_{\boldsymbol{q},\theta}$ is a primitive of $G_{\boldsymbol{q},\theta}$ and $B > 0$ is a global constant.

Remark. • As usual, δ, $C_{l,N}$ and C_N depend only on F, A_q, B_q, B'_q, κ_q and the indices indicated in their notation.

• In other words, the Lagrangian state $e^{i(\theta\cdot)/h}$ is changed to a Lagrangian state associated with $\mathcal{C}_{\boldsymbol{q},\theta}$.

The end of this subsection is devoted to the proof of Proposition 5.7. In the rest of this section, we fix a constant $C_0 > 0$ and we work with a fixed word $\boldsymbol{q} \in \mathcal{A}^n$ with length $n \le C_0 |\log h|$ and a fixed momentum $|\theta| \le C_0$. From now on and until the end of the proof, the constants below will always be uniform in $\boldsymbol{q}, \theta$ satisfying the previous assumption. They will depend on global parameters and on C_0. If they depend on other parameters, we will specify it with subscripts. This is also the case for implicit constants in O (such as in $O(h^\infty)$).

Preparatory work. We first note the following fact: if $\mathcal{V}_q \cap F^{-1}(\mathcal{V}_p) = \varnothing$, $A_p M A_q = O(h^\infty)$. As a consequence, if $\mathcal{V}_{q_{i-1}} \cap F^{-1}(\mathcal{V}_{q_i}) = \varnothing$ for some i, then $U_{\boldsymbol{q}} = O(h^\infty)$. In the sequel, it is enough to consider words $\boldsymbol{q}$ for which $\mathcal{V}_{q_{i-1}} \cap F^{-1}(\mathcal{V}_{q_i}) \ne \varnothing$ for $1 \le i \le n-1$.

We consider symbols $\tilde{a}_q$ such that $\mathrm{supp}(\tilde{a}_q) \subset \mathcal{V}_q$ and $\tilde{a}_q \equiv 1$ on $\mathrm{supp}(\chi_q)$. We denote by $\tilde{A}_q = \mathrm{Op}_h(\tilde{a}_q)$ (as usual thought of as a diagonal operator-valued matrix). The following computations holds since $n = O(\log h)$ and $\|M A_q\| \le \|\alpha\|_\infty + o(1)$ uniformly in q:

$$\begin{aligned} U_{\boldsymbol{q}} B'_{q_0} &= M A_{q_{n-1}} \tilde{A}_{q_{n-1}} M A_{q_{n-2}} \tilde{A}_{q_{n-2}} \cdots M A_{q_1} \tilde{A}_{q_1} M A_{q_0} B'_{q_0} + O(h^\infty) \\ &= M A_{q_{n-1}} B'_{q_{n-1}} B_{q_{n-1}} \tilde{A}_{q_{n-1}} M \cdots M A_{q_1} B'_{q_1} B_{q_1} \tilde{A}_{q_1} M A_{q_0} B'_{q_0} + O(h^\infty). \end{aligned}$$

We set $T_{p,q} = B_p \tilde{A}_p M A_q B'_q$ and $M_q = M A_q B'_q$, which allows us to write

$$U_{\boldsymbol{q}} B'_{q_0} = M_{q_{n-1}} T_{q_{n-1},q_{n-2}} \cdots T_{q_1,q_0} + O(h^\infty).$$

For $p, q \in \mathcal{A}$ with $\mathcal{V}_q \cap F^{-1}(\mathcal{V}_p) \ne \varnothing$, we have $T_{q,p} \in I_0^{\mathrm{comp}}(\tau_{p,q})$. Moreover, the previous computations have shown that $\tau_{p,q}$ has the form

$$\tau_{p,q}(y,\eta) = (\lambda_{p,q} y + y_r(y,\eta), \mu_{p,q}\eta + \eta_r(y,\eta)), \quad (y,\eta) \in B_q(0, C_1\varepsilon_0),$$

where $y_r(y,\eta)$ and $\eta_r(y,\eta)$ are $O(\varepsilon_0)_{C^1}$. This time, $\lambda_{p,q}$, $\mu_{p,q}$ are simply constants uniformly bounded from below and from above for $p, q \in \mathcal{A}$ (recall that $B_q(0, C_1\varepsilon_0)$ is a rectangle in $\mathbb{R}^2$, built from the cube $\widetilde{B}_q(0, C_1\varepsilon_0)$ adapted to the definition of the unstable Jacobian). If ε_0 small enough, the projection $\pi : (y, \eta, x, \xi) \in \mathcal{L}_{q,p} \mapsto (y, \xi) \in \mathbb{R}^2$ is a diffeomorphism onto its image, where

$$\mathcal{L}_{q,p} = \{(\tau_{q,p}(x, \xi), x, -\xi) : (x, \xi) \in B_q(0, C_1\varepsilon_0)\}$$

is the twisted graph of $\tau_{p,q}$. As a consequence, there exists a smooth phase function $S_{p,q}$ defined in an open set $\Omega_{p,q}$ of $\mathbb{R}^2$, generating $\mathcal{L}_{p,q}$ locally, i.e.,

$$\mathcal{L}_{p,q} \cap \tau_{p,q}(B_q(0, C_1\varepsilon_0)) \times B_q(0, C_1\varepsilon_0) = \{(y, \partial_y S_{p,q}(y, \xi), \partial_\xi S_{p,q}(y, \xi), -\xi) : (y, \xi) \in \Omega_{q,p}\}.$$

Hence, $T_{p,q}$ can be written in the following form, up to a $O(h^\infty)$ remainder and for some symbol $\alpha_{p,q}(\,\cdot\,; h) \in C_c^\infty(\Omega_{p,q})$:

$$T_{p,q}u(y) = \frac{1}{2\pi h} \int_{\mathbb{R}^2} e^{(i/h)(S_{p,q}(y,\xi)-x\xi)} \alpha_{p,q}(y, \xi; h) u(x)\, dx\, d\xi. \tag{5-13}$$

Moreover, due to the operators $\tilde{A}_p$ and A_q in the definition of $T_{p,q}$, we can assume that

$$(y, \xi) \in \operatorname{supp}(\alpha_{p,q}) \quad \Longrightarrow \quad (\partial_\xi S_{p,q}(y, \xi), \xi) \in \kappa_q(\operatorname{supp} a_q), \quad (y, \partial_y S_{p,q}(y, \xi)) \in \kappa_p(\operatorname{supp} \tilde{a}_p).$$

In the sequel, we write

$$\mathcal{C}_i = \mathcal{C}_{q_0 \cdots q_i, \theta}$$

and we change the subscripts (q_{i-1}, q_i) to i in all the objects T, α, S, τ. Due to the previous results, we can write $\mathcal{C}_i = \{(y, G_i(y)) : y \in I_i\}$, with $I_i := I_{q_0 \cdots q_i, \theta}$ and $G_i := G_{q_0 \cdots q_i, \theta}$. We also have projection maps $\Phi_i : I_i \to \mathbb{R}$ defined by

$$\tau_i \circ \cdots \circ \tau_1(\Phi_i(y), \theta) = (y, G_i(y))$$

satisfying $\|\Phi_i'\|_\infty \le C\nu^i < 1$. Moreover, if we define the intermediate corresponding projection $\phi_i := \Phi_i \circ \Phi_{i-1}^{-1} : I_i \to I_{i-1}$, we observe that ϕ_i is constructed using the properties of F and G_{i-1} (see the proof of Proposition 5.6) and hence, for every l, $\|\phi_i\|_{C^l} \le C_l$ for some C_l not depending on $\boldsymbol{q}, \theta$ nor i.

For $0 \le i \le n-1$, we consider a primitive ψ_i of G_i so that $\mathcal{C}_i$ is generated by ψ_i, i.e.,

$$\mathcal{C}_i = \{(y, \psi_i'(y) : y \in I_i\}.$$

The following lemma can be found in [Nonnenmacher and Zworski 2009, Lemma 4.1]. We state it without proof, since it is the reference but it is a direct application of the stationary phase theorem.

Lemma 5.8. *Pick $i \in \{1, \dots, n-1\}$. For any $a \in C_c^\infty(I_{i-1})$, the application of T_i to the Lagrangian state $ae^{i\psi_{i-1}/h}$ associated with $\mathcal{C}_{i-1}$ gives a Lagrangian state associated with $\mathcal{C}_i$ and satisfies*

$$T_i(ae^{i\psi_{i-1}/h})(y) = e^{i\beta_i/h} e^{i\psi_i(y)/h} \left(\sum_{j=0}^{N-1} b_j(y) h^j + h^N r_N(y; h) \right), \tag{5-14}$$

where, if we let $x = \phi_i(y)$, then $b_j(y) = (L_{j,i}(x, D_x)a)(x)$ for some differential operator $L_{j,i}$ of order $2j$ with smooth coefficients supported in I_{i-1} and $\beta_i \in \mathbb{R}$. Moreover, one has:

- *$b_0(y) = |\phi_i'(y)|^{1/2} a(x) \alpha_i(y, \xi) / |\det D^2_{y,\xi} S_i(y, \xi)|^{1/2}$, with $\xi = \psi_{i-1}'(x)$.*

- $\|b_j\|_{C^l(I_i)} \le C_{l,j} \|a\|_{C^{l+2j}(I_{i-1})}$, $l \in \mathbb{N}$, $0 \le j \le N-1$.
- $\|r_N\|_{C^l(I_i)} \le C_N \|a\|_{C^{l+1+2N}(I_{i-1})}$.

The constants C_N and $C_{l,j}$ depend on τ_i, α_i, $\|\psi_i^{(m)}\|_{\infty,I_i}$.

Remark. • In particular, by virtue of Proposition 5.6, the constants $C_{l,j}$ and C_N can be chosen uniform in $\boldsymbol{q}, \theta$ as soon as they satisfy the required assumptions: $|q| \le C_0 |\log h|$, $\theta \le C_0$.

• Without loss of generality, we can replace ψ_i by $\beta_i + \psi_i$ (this actually corresponds to fixing an antiderivative on $\mathcal{C}_{i+1}$) and hence we can assume that $\beta_i = 0$.

• The properties on the support of α_i imply the following ones on the support of the differential operators $L_{j,i}$:

$$y \in \operatorname{supp} L_{j,i} \quad \Longrightarrow \quad (y, \psi_i'(y)) \in \kappa_{q_i}(\operatorname{supp} \tilde{a}_{q_i}) \cap \tau_{i-1} \circ \kappa_{q_{i-1}}(\operatorname{supp} a_{q_{i-1}}). \tag{5-15}$$

Iteration formulas and analysis of the symbols. Then, we iterate this lemma starting from $\psi_0(x) = x \cdot \theta$, in the spirit of Proposition 4.1 in [Nonnenmacher and Zworski 2009]. In the sequel, we adopt the following convention: we denote by x_k the variable in I_k and we naturally define $(x_k, x_{k-1}, \dots, x_1, x_0)$, the sequence defined by $x_{i-1} = \phi_i(x_i)$. We also let

$$\beta_i(x_i) = \frac{\alpha_i(x_i, \xi)}{|\det D^2_{x_i,\xi} S_i(x_i, \xi)|^{1/2}}, \quad \xi = \psi_{i-1}'(x_{i-1}),$$
$$f_i(x_i) = \beta(x_i) |\phi_i'(x_i)|^{1/2}.$$

We fix a constant $B > 0$ (depending only on F, A_q, B_q, B_q', C_0) satisfying, for all $1 \le i \le n-1$,

$$\sup_{x_i \in I_i} |\beta_i(x_i)| \le B, \quad \|T_i\| \le B.$$

Roughly speaking, B is of order $\|\alpha\|_\infty$, but in this part, the precise value of B is not relevant. Finally, note that there exists $\nu < 1$ (again depending only on F, A_q, B_q, B_q') such that $|\phi_i'(x_i)| \le \nu$ for $x_i \in I_i$. Fix $N \in \mathbb{N}$ and define

$$\widetilde{N} = 1 + \lceil N + C_0 \log B \rceil. \tag{5-16}$$

We iteratively define a sequence of symbols $a_{i,j}$, $0 \le i \le n-1$, $0 \le j \le \widetilde{N} - 1$ by $a_{0,0} = 1$, $a_{0,j} = 0$ and for $0 \le j \le \widetilde{N} - 1$

$$a_{i,j}(x_i) = \sum_{p=0}^{j} L_{j-p,i}(a_{i-1,p})(x_{i-1}). \tag{5-17}$$

The following lemma controls the growth of the symbols. The proof is a precise analysis of the iteration formula (5-17) and is rather technical. We write the detailed proof in the Appendix (see Section A3) and refer the reader to [Nonnenmacher and Zworski 2009, Proposition 4.1], where the author carried out the same analysis (but in the case $B = 1$).

Lemma 5.9. *For all $j \in \{0, \dots, \widetilde{N} - 1\}$, $l \in \mathbb{N}$, there exists $C_{j,l} > 0$ such that, for all $i \in \{0, \dots, n-1\}$, one has*

$$\|a_{i,j}\|_{C^l(I_i)} \le C_{j,l} (B \nu^{1/2})^i (i+1)^{l+3j}. \tag{5-18}$$

Remark. Again, what is important is the fact that $C_{j,l}$ does *not* depend on $\boldsymbol{q}$, n, θ nor i: it depends on C_0 and global parameters.

Control of the remainder. Let us call $r_{i,N}(a)$ the remainder appearing in Lemma 5.8. Define inductively $(R_{i,\widetilde{N}})$ by $R_{0,\widetilde{N}} = 0$ and

$$R_{i+1,\widetilde{N}} = e^{-(i\psi_{i+1})/h} T_{i+1}(e^{i\psi_i/h} R_{i,\widetilde{N}}) + \sum_{j=0}^{\widetilde{N}-1} r_{i+1,\widetilde{N}-j}(a_{i,j}). \tag{5-19}$$

This definition ensures that, for all $1 \le i \le n$,

$$T_i \cdots T_1(e^{i\psi_0/h}) = e^{i\psi_i(y)/h}\left(\sum_{j=0}^{\widetilde{N}-1} h^j a_{i,j} + h^{\widetilde{N}} R_{i,\widetilde{N}}\right). \tag{5-20}$$

Lemma 5.10. *There exists $C_{\widetilde{N}}$ depending only on $\widetilde{N}$, C_0 and global parameters such that, for all $1 \le i \le n-1$,*

$$\|R_{i,\widetilde{N}}\|_{L^2(\mathbb{R})} \le C_{\widetilde{N}} B^i.$$

Proof. Recalling that $\|T_i\|_{L^2\to L^2} \le B$ and the bound on the remainder in Lemma 5.8, the recursive definition of $R_{i,\widetilde{N}}$ gives the bound

$$\|R_{i,\widetilde{N}}\|_{L^2} \le B\|R_{i-1,\widetilde{N}}\|_{L^2} + \sum_{j=0}^{\widetilde{N}-1} C_{\widetilde{N}-j}\|a_{i-1,j}\|_{C^{1+2(\widetilde{N}-j)}}.$$

By induction and using the previous bounds on $\|a_{i,j}\|_{C^l}$, we get

$$\begin{aligned}
\|R_{\widetilde{N},i}\|_{L^2} &\le \sum_{p=0}^{i-1} B^{i-1-p} \sum_{j=0}^{\widetilde{N}-1} C_{\widetilde{N}-j}\|a_{p,j}\|_{C^{1+2(\widetilde{N}-j)}} \\
&\le \sum_{p=0}^{i-1} B^{i-1-p} \sum_{j=0}^{N_1-1} C_{\widetilde{N}-j} C_{\widetilde{N}-j,0} (B\nu^{1/2})^p (p+1)^{1+2\widetilde{N}+j} \\
&\le C_{\widetilde{N}} B^i \sum_{p=0}^{i-1} \nu^{p/2}(p+1)^{1+3N_1} \le C_{\widetilde{N}} B^i,
\end{aligned}$$

using that the sum is absolutely convergent. □

End of proof of Proposition 5.7. We've got now all the elements to conclude the proof. We set

$$a_{\boldsymbol{q},\theta,N} := \sum_{j=0}^{\widetilde{N}-1} h^j a_{n-1,j}.$$

We know that

$$U_{\boldsymbol{q}} B'_{q_0}(e^{i\theta/h}) = M_{q_{n-1}}(e^{i\psi_{\boldsymbol{q}}\cdot/h} a_{\boldsymbol{q},\theta,N}) + M_{q_{n-1}}(h^{\widetilde{N}} R_{n-1,\widetilde{N}}).$$

Since M_q are uniformly bounded in q and $R_{n-1,\widetilde{N}} \le C_{\widetilde{N}} B^{n-1} \le C_{N_1} h^{-C_0 \log B}$, we have

$$\|M_{q_{n-1}}(h^{\widetilde{N}} R_{n-1,\widetilde{N}})\|_{L^2} \le C_N h^{\widetilde{N} - C_0 \log B} \le C_N h^N.$$

Concerning the bounds on $a_{\boldsymbol{q},\theta,N}$, we have

$$\|a_{\boldsymbol{q},\theta,N}\|_{C^l} \le \sum_{j=0}^{\widetilde{N}-1} h^j \|a_{n-1,j}\|_{C^l} \le \sum_{j=0}^{\widetilde{N}-1} C_{j,l}(B\nu^{1/2})^{n-1} n^{l+3j} h^j$$
$$\le C_{l,N} n^{l+3\widetilde{N}} (B\nu^{1/2})^{n-1} \le C_{l,N} h^{-C_0 \log B} n^{l+3\widetilde{N}} \nu^{(n-1)/2} \le C_{l,N} h^{-C_0 \log B},$$

where we use the fact that $n \le C_0 |\log h|$ and bound $n^{l+3\widetilde{N}} \nu^{(n-1)/2}$ by some $C_{l,\widetilde{N}}$ since $\nu < 1$.

Finally, we need to prove the property on the support of $a_{\boldsymbol{q},\theta,N}$. To do so, let us introduce, for $q \in \mathcal{A}$, an open set $\mathcal{W}_q$ satisfying

$$\operatorname{supp} \tilde{a}_q \Subset \mathcal{W}_q \subset \mathcal{V}_q.$$

This allows us to define new objects replacing $\mathcal{V}_q$ by $\mathcal{W}_q$ in the definitions

$$\mathcal{W}_{\boldsymbol{q}}^+ = \bigcap_{i=0}^{n-1} F^{n-i}(\mathcal{W}_{q_i}) \Subset \mathcal{V}_{\boldsymbol{q}}^+,$$
$$\mathcal{D}_{\boldsymbol{q},\theta} = \kappa_{q_{n-1}}(F^{-1}(\mathcal{W}_{\boldsymbol{q}}^+) \cap F^{n-1}(\mathcal{L}_{q_0,\theta})) \Subset \mathcal{C}_{\boldsymbol{q},\theta},$$

and the associated subinterval $J_{\boldsymbol{q},\theta} \Subset I_{\boldsymbol{q},\theta}$ built thanks to Proposition 5.6 such that

$$\mathcal{D}_{\boldsymbol{q},\theta} = \{(y, G_{\boldsymbol{q},\theta}(y)) : y \in J_{\boldsymbol{q},\theta}\}.$$

Let us fix $\delta > 0$ small (with further conditions imposed). We will show the stronger statement

$$d(\operatorname{supp}(a_{\boldsymbol{q},\theta,N}), \mathbb{R} \setminus J_{\boldsymbol{q},\theta}) \ge \delta.$$

Suppose this is not the case. We can find $x_{n-1} \in \operatorname{supp} a_{\boldsymbol{q},\theta,N}$, $y_{n-1} \in I_{\boldsymbol{q},\theta} \setminus J_{\boldsymbol{q},\theta}$ such that $|x_{n-1} - y_{n-1}| \le \delta$. As already done, we denote by x_i (resp. y_i) the points defined by $x_{i-1} = \phi_i(x_i)$ (resp. $y_{i-1} = \phi_i(y_i)$). Since ϕ_i are contractions, we have $|x_i - y_i| \le \delta$ for $1 \le i \le n-1$. If we define

$$\rho_i = \kappa_{q_i}^{-1}(x_i, \psi_i'(x_i)), \quad \zeta_i = \kappa_{q_i}^{-1}(y_i, \psi_i'(y_i)),$$

we have, for some $C > 0$, $d(\rho_i, \zeta_i) \le C\delta$. By definition, one also has

$$F^{-i}(\rho_{n-1}) = \rho_{n-1-i}, \quad F^{-i}(\zeta_{n-1}) = \zeta_{n-1-i}.$$

By the support property (5-15) of the operators $L_{j,i}$, $\rho_i \in \operatorname{supp} \tilde{a}_{q_i}$ for $0 \le i \le n-1$. Let's assume that δ is small enough so that, for all $q \in \mathcal{A}$,

$$d(\operatorname{supp} \tilde{a}_q, (\mathcal{W}_q)^c) \ge 2C\delta.$$

Hence,

$$\rho_i \in \operatorname{supp} \tilde{a}_{q_i} \text{ and } d(\rho_i, \zeta_i) \le C\delta \quad \Longrightarrow \quad \zeta_i \in \mathcal{W}_{q_i}.$$

As a consequence, for all $0 \le i \le n-1$, $F^{i+1-n}(\zeta_{n-1}) \in \mathcal{W}_{q_i}$, or equivalently $\zeta_{n-1} \in F^{-1}(\mathcal{W}_{\boldsymbol{q}}^+)$. Hence,

$$(y_{n-1}, \psi_{n-1}'(y_{n-1})) \in \mathcal{C}_{\boldsymbol{q},\theta} \cap \kappa_{q_{n-1}}(F^{-1}(\mathcal{W}_{\boldsymbol{q}}^+)) \subset \mathcal{D}_{\boldsymbol{q},\theta}$$

showing that $y_{n-1} \in J_{\boldsymbol{q},\theta}$, and giving a contradiction with $y_{n-1} \in I_{\boldsymbol{q},\theta} \setminus J_{\boldsymbol{q},\theta}$.

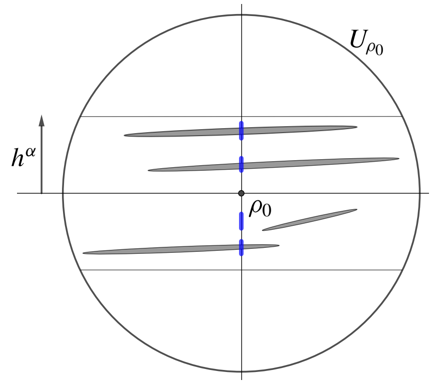

Figure 14. The definition of the sets $\Gamma^+_{\boldsymbol{q}}$. They are represented by the blue segments on the η-axis and are the projections on the η variable of the sets $\mathcal{V}^+_{\boldsymbol{q}}$ (the shaded sets). They are of width of order h^τ.

5C. *Microlocalization of $U_{\mathcal{Q}}$.* We now fix a cloud $\mathcal{Q} \subset \mathcal{Q}(n, a)$, centered at a point $\rho_0 \in \mathcal{T}$, namely satisfying the condition of Proposition 4.23:

$$\text{for all } \rho \in \bigcup_{\boldsymbol{q}\in\mathcal{Q}} \mathcal{V}^+_{\boldsymbol{q}}, \quad d(\rho, W_u(\rho_0)) \le Ch^{\mathfrak{b}}.$$

Let us define

$$U_{\mathcal{Q}} = \sum_{\boldsymbol{q}\in\mathcal{Q}} U_{\boldsymbol{q}} \tag{5-21}$$

and

$$\mathcal{V}^+_{\mathcal{Q}} = \bigcup_{\boldsymbol{q}\in\mathcal{Q}} \mathcal{V}^+_{\boldsymbol{q}}. \tag{5-22}$$

We fix an adapted chart $\kappa := \kappa_{\rho_0} : U_0 \to V_0$ around ρ_0 as permitted by the Lemma 3.23. We can assume that $\mathcal{V}^+_a \Subset U_0$ (if ε_0 is small enough and since the local unstable leaf $W_u(\rho_0)$ is close to points in $\mathcal{V}^+_a$)). We consider a cut-off function $\tilde{\chi}_a \in C_c^\infty(U_0)$ such that $\tilde{\chi}_a \equiv 1$ on $F(\operatorname{supp}\chi_a)$ and $\operatorname{supp}\tilde{\chi}_a \subset \mathcal{V}^+_a$. Let us write $\Xi_a = \operatorname{Op}_h(\tilde{\chi}_a)$. Since $\Xi_a M A_a = M A_a + O(h^\infty)$, $|\mathcal{Q}| = O(h^{-K})$ and $\|U_{\boldsymbol{q}}\| = O(h^{-K})$ for some $K > 0$, we have

$$\mathfrak{M}^{N_0} U_{\mathcal{Q}} = \mathfrak{M}^{N_0}\,\Xi_a U_{\mathcal{Q}} + O(h^\infty).$$

Let us introduce Fourier integral operators B, B' quantizing κ in $\operatorname{supp}(\chi_a)$:

$$B'B = I + O(h^\infty) \text{ microlocally in } \operatorname{supp}(\chi_a).$$

Hence

$$\mathfrak{M}^{N_0} U_{\mathcal{Q}} = \mathfrak{M}^{N_0}\,\Xi_a B' B U_{\mathcal{Q}} + O(h^\infty).$$

We introduce the sets

$$\Gamma^+ = \eta(\kappa(\mathcal{V}^+_{\mathcal{Q}})), \quad \Omega^+ = \Gamma^+(h^\tau), \tag{5-23}$$

and, for $\boldsymbol{q} \in \mathcal{Q}$,

$$\Gamma^+_{\boldsymbol{q}} = \eta(\kappa(\mathcal{V}^+_{\boldsymbol{q}})). \tag{5-24}$$

We will prove in the following lemma that the pieces $U_{\boldsymbol{q}}$ are microlocalized in thin horizontal rectangles (see Figure 14).

Lemma 5.11. *For every $\boldsymbol{q} \in \mathcal{Q}$,*

$$\mathbb{1}_{\Gamma^+_{\boldsymbol{q}}(h^\tau)}(hD_y)BU_{\boldsymbol{q}} = BU_{\boldsymbol{q}} + O(h^\infty)_{L^2\to L^2}, \tag{5-25}$$

with uniform bounds in the $O(h^\infty)$.

Using the polynomial bounds $|\mathcal{Q}| = O(h^{-C})$ and $\|U_{\boldsymbol{q}}\| = O(h^{-C})$, we immediately deduce:

Proposition 5.12.
$$\mathbb{1}_{\Omega^+}(hD_y)BU_{\mathcal{Q}} = BU_{\mathcal{Q}} + O(h^\infty)_{L^2\to L^2}. \tag{5-26}$$

5C1. *Proof of Lemma 5.11.* We fix a word $\boldsymbol{q} = q_0\cdots q_{n-2}a \in \mathcal{Q}$. Since $\mathrm{WF}_h(A_{q_0})$ is compact, we can find $\chi \in C^\infty_c(\mathbb{R})$ such that

$$A_{q_0} = A_{q_0}B'_{q_0}\chi(hD_y)B_{q_0} + O(h^\infty).$$

Since there is a finite number of symbols in $\mathcal{A}$, we can choose one single χ for all the possible symbols q_0. We are hence reduced to proving that

$$\underbrace{\mathbb{1}_{\mathbb{R}\setminus\Gamma^+_{\boldsymbol{q}}(h^\tau)}(hD_y)BU_{\boldsymbol{q}}B'_{q_0}}_{T}\chi(hD_y) = O(h^\infty)_{L^2\to L^2}. \tag{5-27}$$

If $u \in L^2(\mathbb{R})$, writing

$$(\chi(hD_y)u)(y) = \frac{1}{(2\pi h)^{1/2}}\int_{\mathbb{R}}\chi(\theta)\mathcal{F}_hu(\theta)e^{i(\theta y)/h}\,d\theta,$$

we have

$$T(\chi(hD_y)u) = \frac{1}{(2\pi h)^{1/2}}\int_{\mathbb{R}}\chi(\theta)\mathcal{F}_hu(\theta)(Te^{i(\theta\cdot)/h})\,d\theta.$$

Hence,

$$\begin{aligned}\|T(\chi(hD_y)u)\|_{L^2} &\le \frac{1}{(2\pi h)^{1/2}}\int_{\mathbb{R}}|\chi(\theta)\mathcal{F}_hu(\theta)|\|Te^{i(\theta\cdot)/h}\|_{L^2}\,d\theta\\ &\le \frac{1}{(2\pi h)^{1/2}}\int_{\mathbb{R}}|\chi(\theta)\mathcal{F}_hu(\theta)|\sup_{\theta\in\operatorname{supp}\chi}\|Te^{i(\theta\cdot)/h}\|_{L^2}\\ &\le \frac{C_\chi}{h^{1/2}}\|\mathcal{F}_hu\|_{L^2}\sup_{\theta\in\operatorname{supp}\chi}\|Te^{i(\theta\cdot)/h}\|_{L^2}\\ &\le \frac{C_\chi}{h^{1/2}}\|u\|_{L^2}\sup_{\theta\in\operatorname{supp}\chi}\|Te^{i(\theta\cdot)/h}\|_{L^2}.\end{aligned}$$

As a consequence, we are lead to estimate $\sup_{\theta\in\operatorname{supp}\chi}\|Te^{i(\theta\cdot)/h}\|_{L^2}$. We fix $\theta \in \operatorname{supp}\chi$. Writing that $\operatorname{supp}\chi \subset [-C_0, C_0]$ and recalling $|\boldsymbol{q}| = n \le C_0|\log h|$ for some global C_0, we are in the framework of Proposition 5.7.

We fix $N \in \mathbb{N}$ and we aim at proving that $Te^{i\theta\cdot/(h)} = O(h^N)$. By Proposition 5.7, there exists $a_{\boldsymbol{q},N,\theta} \in C^\infty_c(I_{\boldsymbol{q},\theta})$ such that

$$U_{\boldsymbol{q}}B'_{q_0}(e^{i(\theta\cdot)/h}) = MA_aB'_a(a_{\boldsymbol{q},N,\theta}e^{i\Phi_{\boldsymbol{q},\theta}/h}) + O(h^N).$$

Set $S := BMA_aB'_a$. Then S is a Fourier integral operator associated with $s := \kappa\circ F\circ\kappa_a^{-1}$. Recall that the definitions and the description of the Lagrangian

$$C_{\boldsymbol{q},\theta} = \kappa_a(F^{-1}(\mathcal{V}^+_{\boldsymbol{q}})\cap F^{n-1}(\mathcal{L}_{q_0,\theta})) = \{(y, \Phi'_{\boldsymbol{q},\theta}(y)) : y \in I_{\boldsymbol{q},\theta}\},$$

with $\Phi_{\boldsymbol{q},\theta} \in C^\infty(I_{\boldsymbol{q},\theta})$, $\|\Phi_{\boldsymbol{q},\theta}\|_{C^1} \le C\varepsilon_0$, $\|\Phi_{\boldsymbol{q},\theta}\|_{C^l} \le C_l$.

If ε_0 is small enough, we can assume that:

- s is well-defined on $B_a(0, C_1\varepsilon_0)$ and satisfies the conclusion of Lemma 5.5. As a consequence, the Lagrangian line
$$s(\mathcal{C}_{\boldsymbol{q},\theta}) = \kappa(\mathcal{V}_{\boldsymbol{q}}^+) \cap \kappa \circ F^n(\mathcal{L}_{q_0,\theta})$$
can be written $\{(y, \Psi'(y)) : y \in I\}$ for some open $I \subset \mathbb{R}$ and some function $\Psi \in C^\infty(I)$ satisfying
$$\|\Psi\|_{C^1} \le C\varepsilon_0, \quad \|\Psi\|_{C^l} \le C_l,$$
with global constants C and C_l.
- S has the form (5-13) with a phase function and a symbol having C^l norms bounded by global constants (depending on l).

Hence, we can apply Lemma 5.8 to see that there exists $b \in C_c^\infty(I)$ such that
$$S(a_{\boldsymbol{q},N,\theta} e^{i\Phi_{q,\theta}/h}) = be^{i\Psi/h} + O(h^N)_{L^2},$$
and b satisfies the same type of bounds as $a_{\boldsymbol{q},N,\theta}$; namely,
$$\|b\|_{C^l} \le C_{l,N} h^{-C_0 \log B}.$$
Moreover, since $d(\operatorname{supp} a_{\boldsymbol{q},N,\theta}, \mathbb{R} \setminus I_{\boldsymbol{q},\theta}) \ge \delta$, there exists $\delta' > 0$ such that $d(\operatorname{supp} b, \mathbb{R} \setminus I) \ge \delta'$. The constants $C_{l,N}$ and δ' are global constants. Since N is arbitrary, to conclude the proof of Lemma 5.11, it remains to show that
$$\mathbb{1}_{\mathbb{R} \setminus \Gamma_{\boldsymbol{q}}^+(h^\tau)}(hD_y)(be^{i\Psi/h}) = O(h^N). \tag{5-28}$$

To do so, we make use of the fine Fourier localization statement from Proposition 2.7 in [Dyatlov et al. 2022]. We state it for convenience but refer the reader to the quoted paper for the proof.

Proposition 5.13. *Let $U \subset \mathbb{R}^n$ open, $K \subset U$ compact, $\Phi \in C^\infty(U)$ and $a \in C_c^\infty(U)$ with $\operatorname{supp} a \subset K$. Assume that there is a constant C_0 and constants C_N, $N \in \mathbb{N}^*$, such that*
$$\operatorname{vol}(K) \le C_0, \tag{5-29}$$
$$d(K, \mathbb{R}^n \setminus U) \ge C_0^{-1}, \tag{5-30}$$
$$\max_{0<|\alpha|\le N} \sup_U |\partial^\alpha \Phi| \le C_N, \quad N \ge 1, \tag{5-31}$$
$$\max_{0\le|\alpha|\le N} \sup_U |\partial^\alpha a| \le C_N, \quad N \ge 1. \tag{5-32}$$
Finally, assume that the projection of the Lagrangian $\{(x, \Phi'(x)) : x \in U\}$ on the momentum variable has a diameter of order h^τ; namely,
$$\operatorname{diam}(\Omega_\Phi) \le C_0 h^\tau, \quad \textit{where } \Omega_\Phi = \{\Phi'(x) : x \in U\}. \tag{5-33}$$
Define the Lagrangian state
$$u(x) = a(x)e^{i\Phi(x)/h} \in C_c^\infty(U) \subset C_c^\infty(\mathbb{R}^n).$$

Then, for every $N \geq 1$, *there exists* C'_N *such that*

$$\|\mathbb{1}_{\mathbb{R}^n \setminus \Omega_\Phi(h^\tau)} u\| \leq C'_N h^N, \tag{5-34}$$

where C'_N *depends on* $\tau, n, N, C_0, C_{N'}$ *for some* $N'(n, N, \tau)$.

When $U = I,\ K = \operatorname{supp} b,\ a = h^{C_0 \log B} b,\ \Phi = \Psi$, the assumptions (5-29) to (5-32) are satisfied for some global constants C_0, C_N. In this case,

$$\Omega_\Psi = \{\Psi'(y) : y \in I\} = \eta\big(\kappa(\mathcal{V}_{\boldsymbol{q}}^+) \cap \kappa \circ F^n(\mathcal{L}_{q_0,\theta})\big).$$

Since $\Omega_\Psi \subset \Gamma_{\boldsymbol{q}}^+$, to prove (5-28), it is enough to prove it with $\Gamma_{\boldsymbol{q}}^+$ replaced by Ω_Ψ and to apply the last proposition, it remains to check that the last point (5-33) is satisfied. Since who can do more, can do less, we will show that

$$\operatorname{diam}(\Gamma_{\boldsymbol{q}}^+) \leq C_0 h^\tau.$$

This is where the strong assumption on the adapted charts will play a role. To insist on this role, we state the following lemma:

Lemma 5.14. *Let* $C_0 > 0$. *Assume that* $\rho_1 \in \mathcal{T} \cap U_{\rho_0}$ *satisfies* $d(\rho_1, W_u(\rho_0)) \leq C_0 h^{\mathfrak{b}}$. *If* $\rho_2 \in W_u(\rho_1)$, *then, for some global constant* $C > 0$,

$$|\eta(\kappa(\rho_1)) - \eta(\kappa(\rho_2))| \leq C C_0^{1+\beta} h. \tag{5-35}$$

Proof. Recall that the chart (κ, U_{ρ_0}) is the one centered at ρ_0, given by Lemma 3.23. In this chart, $\kappa(W_u(\rho_1))$ is almost horizontal; we have

$$\kappa(W_u(\rho_1)) = \{y : g(y, \zeta(\rho_1)),\ y \in \Omega\},$$

where Ω is some open bounded set of $\mathbb{R}$, with g and ζ satisfying the properties of Lemma 3.23. Hence, to prove the lemma, it is enough to estimate $|g(y, \zeta(\rho_1)) - g(0, \zeta(\rho_1))|,\ y \in \Omega$. Since $\zeta(\rho_0) = 0$ and ζ is Lipschitz, $|\zeta(\rho_1)| \leq C_0 h^{\mathfrak{b}}$. Indeed, if $\rho'_0 \in W_u(\rho_0)$ satisfies $d(\rho'_0, \rho_1) \leq 2C_0 h^{\mathfrak{b}}$,

$$|\zeta(\rho_1)| = |\zeta(\rho_1) - \zeta(\rho'_0)| \leq C d(\rho_1, \rho'_0) \leq C C_0 h^{\mathfrak{b}}.$$

Then, we have

$$\begin{aligned} |g(y, \zeta(\rho_1)) - g(0, \zeta(\rho_1))| &= |g(y, \zeta(\rho_1)) - g(y, 0) - \partial_\zeta g(y, 0)\zeta(\rho_1)| \\ &= \left| \int_0^{\zeta(\rho_1)} (\partial_\zeta g(y, \zeta) - \partial_\zeta g(y, 0))\, d\zeta \right| \\ &\leq \left| \int_0^{\zeta(\rho_1)} C\zeta^\beta\, d\zeta \right| \leq C\zeta(\rho_1)^{1+\beta} \leq C C_0^{1+\beta} h^{\mathfrak{b}(1+\beta)}. \end{aligned}$$

In the first equality, we've used the facts that $g(0, \zeta) = \zeta,\ \partial_\zeta g(y, 0) = 1$ and $g(y, 0) = 0$. This concludes the proof since, by definition (see (4-2)), $\mathfrak{b}(1 + \beta) = 1$. □

Remark. This lemma explains our definition of $\mathfrak{b}$.

From this lemma, we can deduce (5-33). Indeed, recall that there exists $\rho_{\boldsymbol{q}} \in \mathcal{T}$ such that $\mathcal{V}_{\boldsymbol{q}}^+ \subset W_u(\rho_{\boldsymbol{q}})(Ch^\tau)$. If $\rho_1, \rho_2 \in \mathcal{V}_{\boldsymbol{q}}^+$, there exists $\rho_1', \rho_2' \in W_u(\rho_{\boldsymbol{q}})$ such that

$$d(\rho_i, \rho_i') \le Ch^\tau, \quad i = 1, 2.$$

Hence, one can estimate

$$|\eta(\kappa(\rho_1)) - \eta(\kappa(\rho_2))| \le \underbrace{|\eta(\kappa(\rho_1)) - \eta(\kappa(\rho_1'))|}_{\le Ch^\tau} + \underbrace{|\eta(\kappa(\rho_1')) - \eta(\kappa(\rho_2'))|}_{\le Ch} + \underbrace{|\eta(\kappa(\rho_2)) - \eta(\kappa(\rho_2'))|}_{\le Ch^\tau}.$$

The inequality in the middle is a consequence of the previous lemma. Indeed, $\rho_1', \rho_2' \in W_u(\rho_1')$, where (recall that $\tau > \mathfrak{b}$)

$$d(\rho_1', W_u(\rho_0)) \le d(\rho_1, \rho_1') + d(\rho_1, W_u(\rho_0)) \le Ch^\tau + Ch^{\mathfrak{b}} \le 2Ch^{\mathfrak{b}}.$$

5D. ***Reduction to a fractal uncertainty principle.*** We go on the work started in the last subsection and we keep the same notation. By Propositions 5.3 and 5.12, we can write

$$\mathfrak{M}^{N_0} U_{\mathcal{Q}} = \mathfrak{M}^{N_0} B' B \operatorname{Op}_h(\chi_h) \Xi_a B' \mathbb{1}_{\Omega^+}(hD_y) B U_{\mathcal{Q}} + O(h^\infty)_{L^2 \to L^2}, \tag{5-36}$$

where

- $\chi_h \in S_{\delta_2}^{\text{comp}}$, $\chi_h \equiv 1$ on $\mathcal{T}_-^{\text{loc}}(2C_2h^{\delta_2})$ and $\operatorname{supp} \chi_h \in \mathcal{T}_-^{\text{loc}}(4C_2h^{\delta_2})$ (see Proposition 5.3 and before).
- $\Xi_a = \operatorname{Op}_h(\tilde{\chi}_a)$, where $\tilde{\chi}_a \in C_c^\infty(U_0)$ is a cut-off function such that $\tilde{\chi}_a \equiv 1$ on $F(\operatorname{supp} \chi_a)$ and $\operatorname{supp} \tilde{\chi}_a \subset \mathcal{V}_a^+$ (see the beginning of Section 5C).
- $\Omega^+ = \eta(\kappa(\mathcal{V}_{\mathcal{Q}}^+))(h^\tau)$ (see (5-23) and Proposition 5.12).

In V_{ρ_0}, $U_{\mathcal{Q}}$ is microlocalized in a region $\{|\eta| \le Ch^{\mathfrak{b}}\}$. To work with symbols in usual symbol classes, we will rather consider a bigger region $\{|\eta| \le h^{\delta_0}\}$. For this purpose, let us define

$$\Gamma^- = y\big(\kappa(\mathcal{V}_a^+ \cap \mathcal{T}_-^{\text{loc}}(4C_2h^{\delta_2})) \cap \{|\eta| \le h^{\delta_0}\}\big), \quad \Omega^- = \Gamma^-(h^{\delta_0}). \tag{5-37}$$

Since $\mathcal{V}_{\mathcal{Q}}^+ \subset W_u(\rho_0)(Ch^{\mathfrak{b}})$, we have $\Omega_+ \subset [-C_0h^{\mathfrak{b}}, C_0h^{\mathfrak{b}}] \subset [-h^{\delta_0}, h^{\delta_0}]$ for h small enough. By Lemma 5.2, there exists $\chi_+(\eta) := \chi_+(\eta; h) \in C_c^\infty(\mathbb{R})$ such that

- $\chi_+ \equiv 1$ on Ω^+,
- $\operatorname{supp} \chi_+ \subset [-h^{\delta_0}, h^{\delta_0}]$,
- for all $k \in \mathbb{N}$ and $\eta \in \mathbb{R}$, $|\chi_+^{(k)}(\eta)| \le C_k h^{-\delta_0 k}$ for some global constants C_k,

and χ_+ satisfies

$$\mathbb{1}_{\Omega^+}(hD_y) = \chi_+(hD_y)\mathbb{1}_{\Omega^+}(hD_y).$$

Let's now consider the following subset of Γ^-:

$$\widetilde{\Gamma}^- = y\big(\kappa(\mathcal{V}_a^+ \cap \mathcal{T}_-^{\text{loc}}(4C_2h^{\delta_2})) \cap \{\eta \in \operatorname{supp} \chi_+\}\big).$$

The inclusion $\widetilde{\Gamma}^- \subset \Gamma^-$ comes from the support property of χ_+.

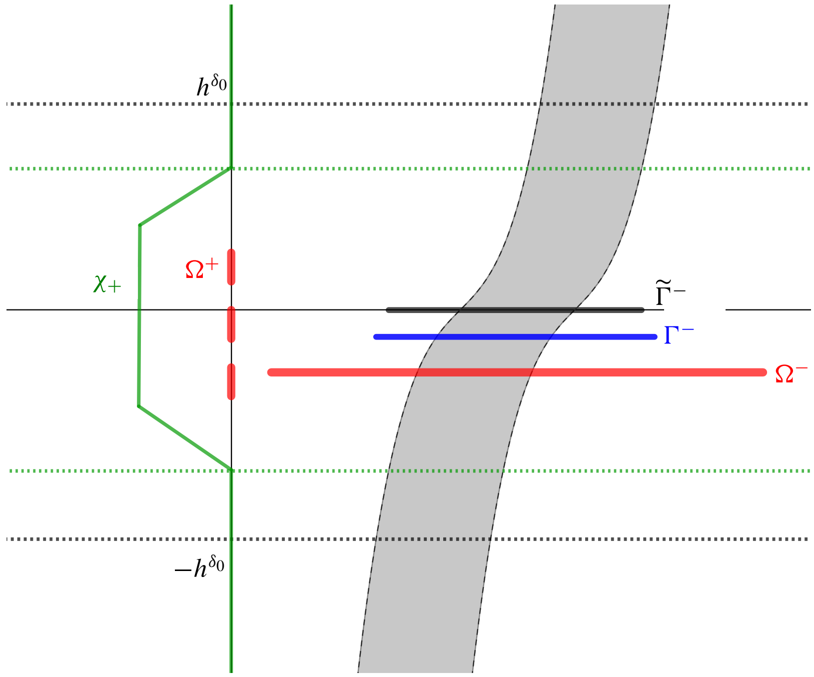

Figure 15. The set Ω^+ is represented on the η-axis, with the support of the function χ_+. On the y-axis, we project the gray set $\kappa(\mathcal{V}_a^+ \cap \mathcal{T}_-^{\mathrm{loc}}(4C_2h^{\delta_2}))$ to obtain both Γ^- and $\widetilde{\Gamma}^-$ depending on the size of the η-window. The larger set Ω^- is also represented in red.

Using again Lemma 5.2, we construct a family $\chi_-(y) := \chi_-(y;h) \in C_c^\infty(\mathbb{R})$ such that

- $\chi_- \equiv 1$ on $\widetilde{\Gamma}^-$,
- $\operatorname{supp} \chi_- \subset \Omega^- = \Gamma^-(h^{\delta_0})$,
- for all $k \in \mathbb{N}$ and $y \in \mathbb{R}$, $|\chi_-^{(k)}(y)| \le C_k h^{-\delta_0 k}$,

and χ_- allows us to write

$$\chi_-(y)\mathbb{1}_{\Omega^-}(y) = \chi_-(y).$$

We encourage the reader to use Figure 15 to fix the ideas. We now claim that

$$\mathfrak{M}^{N_0} U_{\mathcal{Q}} = \mathfrak{M}^{N_0} \operatorname{Op}_h(\chi_h)\Xi_a B' \chi_-(y)\mathbb{1}_{\Omega^-}(y)\mathbb{1}_{\Omega^+}(hD_y)BU_{\mathcal{Q}} + O(h^\infty)_{L^2\to L^2}. \tag{5-38}$$

Due to the polynomial bounds on $\|\mathfrak{M}^{N_0}\|$ and $\|U_{\mathcal{Q}}\|$, it is then enough to show that

$$\operatorname{Op}_h(\chi_h)\Xi_a B'(1-\chi_-(y))\chi_+(hDy) = O(h^\infty).$$

Using Egorov's theorem in $\Psi_{\delta_2}(\mathbb{R})$, we see that $\Xi_0 := B \operatorname{Op}_h(\chi_h)\Xi_a B'$ is in $\Psi_{\delta_2}(\mathbb{R})$ and $\mathrm{WF}_h(\Xi_0) \subset \kappa(\operatorname{supp} \chi_a \cap \operatorname{supp} \chi_h)$. We now observe that

$$(y,\eta) \in \mathrm{WF}_h(\Xi_0) \cap \mathrm{WF}_h(1-\chi_-(y)) \cap \mathrm{WF}_h(\chi_+(hD_y))$$
$$\implies \quad (y,\eta) \in \kappa(\operatorname{supp} \chi_a \cap \operatorname{supp} \chi_h), \quad \eta \in \operatorname{supp} \chi_+,\ y \notin \widetilde{\Gamma}^-,$$

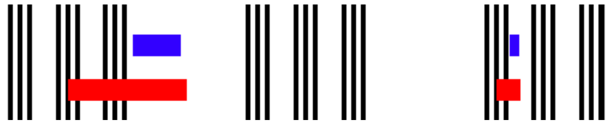

Figure 16. Example of a porous set. Its construction is based on a Cantor-like set. Red intervals correspond to choices of I, blue ones correspond to J.

But the first two conditions imply that $y \in \widetilde{\Gamma}^-$. Hence,

$$\mathrm{WF}_h(\Xi_0) \cap \mathrm{WF}_h(1-\chi_-(y)) \cap \mathrm{WF}_h(\chi_+(hD_y)) = \varnothing.$$

By the composition formulas in $\Psi_{\delta_0}(\mathbb{R})$, we have $\Xi_0(1-\chi_-(y))\chi_+(hD_y) = O(h^\infty)$. Note that the constants in $O(h^\infty)$ depend on the seminorms of $\chi_\pm$, χ_h and χ_a. Due to their construction, the seminorms of $\chi_\pm$ and χ_h are bounded by global constants. As a consequence, the constants $O(h^\infty)$ are global constants.

This proves (5-38). Recalling the bound

$$\|\mathfrak{M}^{N_0}\|_{L^2\to L^2} \le \|\alpha\|^{N_0}(1+o(1)), \quad \|U_{\mathcal{Q}}\|_{L^2\to L^2} \le C|\log h|\,\|\alpha\|_\infty^{N_1},$$

we see that the proof of Proposition 4.23 and hence of Proposition 4.2, has been reduced to proving the following proposition.

Proposition 5.15. *With the above notation, There exist* $\gamma > 0$ *and* $h_0 > 0$ *such that,*

$$\text{for all } h \le h_0, \quad \|\mathbb{1}_{\Omega^-}(y)\mathbb{1}_{\Omega^+}(hD_y)\|_{L^2\to L^2} \le h^\gamma. \tag{5-39}$$

Remark. Note γ and h_0 are global; they do not depend on the particular $\mathcal{Q} \subset \mathcal{Q}(n,a)$ satisfying the conditions of Proposition 4.23, nor on n.

The proof of this proposition is the aim of the next section and relies on a fractal uncertainty principle.

6. Application of the fractal uncertainty principle

The fractal uncertainty principle, first introduced in [Dyatlov and Zahl 2016] and further proved in full generality in [Bourgain and Dyatlov 2018], is the key tool for our decay estimate. We'll use the slightly more general version proved and used in [Dyatlov et al. 2022].

6A. ***Porous sets.*** See for instance Figure 16 for an example. We start by recalling the definition of porous sets and then we state the version of the fractal uncertainty principle we'll use.

Definition 6.1. Let $\nu \in (0,1)$ and $0 \le \alpha_0 \le \alpha_1$. We say that a subset $\Omega \subset \mathbb{R}$ is ν-porous on a scale from α_0 to α_1 if, for every interval $I \subset \mathbb{R}$ of size $|I| \in [\alpha_0, \alpha_1]$, there exists a subinterval $J \subset I$ of size $|J| = \nu|I|$ such that $J \cap \Omega = \varnothing$.

The following simple lemma shows that when one fattens a porous set, one gets another porous set. For its (very elementary) proof, see [Dyatlov et al. 2022, Lemma 2.12].

Lemma 6.2. *Let $\nu \in (0,1)$ and $0 \le \alpha_0 < \alpha_1$. Assume that $\alpha_2 \in]0, \nu\alpha_1/3]$ and $\Omega \subset \mathbb{R}$ is ν-porous on a scale from α_0 to α_1. Then, the neighborhood $\Omega(\alpha_2) = \Omega + [-\alpha_2, \alpha_2]$ is $(\nu/3)$-porous on a scale from* $\max(\alpha_0, 3\alpha_2/\nu)$ *to* α_1.

The notion of porosity can be related to the different notions of fractal dimensions. Let us recall the definition of the upper-box dimension of a metric space (X, d). We denote by $N_X(\varepsilon)$ the minimal number of open balls of radius ε needed to cover X. Then, the upper-box dimension of X is defined by

$$\overline{\dim}\, X := \limsup_{\varepsilon \to 0} \frac{\log N_X(\varepsilon)}{-\log \varepsilon}. \tag{6-1}$$

In particular, if $\delta > \overline{\dim}_X$, there exists $\varepsilon_0 > 0$ such that for every $\varepsilon \le \varepsilon_0$, $N_X(\varepsilon) \le \varepsilon^{-\delta}$. This observation motivates the following lemma:

Lemma 6.3. *Let $\Omega \subset \mathbb{R}$. Suppose that there exist $0 < \delta < 1$, $C > 0$ and $\varepsilon_0 > 0$ such that,*

$$\textit{for all } \varepsilon \le \varepsilon_0, \quad N_\Omega(\varepsilon) \le C\varepsilon^{-\delta}.$$

Then, there exists $\nu = \nu(\delta, \varepsilon_0, C)$ such that Ω is ν-porous on a scale from 0 *to* 1.

Remark. The proof will give an explicit value of ν. This quantitative statement will be important in the sequel to ensure the same porosity for all the sets $W_{u/s}(\rho_0) \cap \mathcal{T}$.

Proof. Let us set $T = \lfloor \max((6\varepsilon_0)^{-1}, (6^\delta C)^{1/(1-\delta)}) \rfloor + 1$ and $\nu = (3T)^{-1}$. We will show that Ω is ν-porous on a scale from 0 to 1. Let $I \subset \mathbb{R}$ be an interval of size $|I| \in]0, 1]$. Cut I into $3T$ consecutive closed intervals of size ν: $J_0, \dots, J_{3T-1}$. We argue by contradiction and assume that each of these intervals does intersect Ω. Let us show that

$$N_\Omega(\nu/2) \ge T. \tag{6-2}$$

Assume that $U_1, \dots, U_k$ is a family of open intervals of size ν covering Ω. For $i = 0, \dots, T-1$, there exists $x_i \in J_{3i+1}$ and $j_i \in \{1, \dots, k\}$ such that $x_i \in U_{j_i}$. It follows that $U_{j_i} \subset J_{3i} \cup J_{3i+1} \cup J_{3i+2}$ and hence $i \ne l \Longrightarrow U_{j_i} \cap U_{j_l} = \varnothing$. The map $i \in \{0, \dots, T-1\} \mapsto j_i \in \{1, \dots, k\}$ is one-to-one, and it gives (6-2). Since $T \ge 1/(6\varepsilon_0)$, we have $\nu/2 \le \varepsilon_0$. As a consequence,

$$T \le N(\nu/2) \le C(6T)^\delta,$$

which implies $T^{1-\delta} \le C6^\delta$. This contradicts the definition of T. ☐

In Section A5 of the Appendix, we give a result in the other way, namely, porous sets down to scale 0 have an upper-box dimension strictly smaller than 1.

For further use, we also record the easy lemma:

Lemma 6.4. *Assume (X, d), (Y, d') are metric spaces and $f : X \to Y$ is C-Lipschitz. Then, for every $\varepsilon > 0$,*

$$N_{f(X)}(\varepsilon) \le N_X(\varepsilon/C).$$

In particular, if $N_X(\varepsilon) \le C_1^\delta \varepsilon^\delta$ for $\varepsilon \le \varepsilon_0$, then, for $\varepsilon \le C\varepsilon_0$, we have $N_{f(X)}(\varepsilon) \le (C_1 C)^\delta \varepsilon^{-\delta}$.

6B. ***Fractal uncertainty principle.*** We state here the version of the fractal uncertainty principle we'll use. This version is stated in Proposition 2.11 in [Dyatlov et al. 2022]. The difference with the original version in [Bourgain and Dyatlov 2018] is that it relaxes the assumption regarding the scales on which the sets are porous. We refer the reader to [Dyatlov 2019] to an overview on the fractal uncertainty principle with other references and applications.

Proposition 6.5 (fractal uncertainty principle). *Fix numbers $\gamma_0^\pm$, $\gamma_1^\pm$ such that*

$$0 \le \gamma_1^\pm < \gamma_0^\pm \le 1, \quad \gamma_1^+ + \gamma_1^- < 1 < \gamma_0^+ + \gamma_0^-$$

and define

$$\gamma := \min(\gamma_0^+, 1-\gamma_1^-) - \max(\gamma_1^+, 1-\gamma_0^-).$$

Then, for each $\nu > 0$, there exists $\beta = \beta(\nu) > 0$ and $C = C(\nu)$ such that the estimate

$$\|\mathbb{1}_{\Omega_-}\mathcal{F}_h\mathbb{1}_{\Omega_+}\|_{L^2(\mathbb{R})\to L^2(\mathbb{R})} \le Ch^{\gamma\beta} \tag{6-3}$$

holds for all $0 < h \le 1$ and all h-dependent sets $\Omega_\pm \subset \mathbb{R}$ which are ν-porous on a scale from $h^{\gamma_0^\pm}$ to $h^{\gamma_1^\pm}$.

Remark. In the sequel, we will use this result with $\gamma_1^\pm = 0$. In this case, the condition on $\gamma_0^\pm$ becomes $\gamma_0^- + \gamma_0^+ > 1$ and the exponent γ is $\gamma_0^- + \gamma_0^+ - 1$. This condition can be interpreted as a condition of saturation of the standard uncertainty principle: a rectangle of size $h^{\gamma_0^+} \times h^{\gamma_0^-}$ will be subplanckian.

6C. ***Porosity of Ω^+ and Ω^-.*** Since we want to apply Proposition 6.5 to prove Proposition 5.15, we need to show the porosity of the sets $\Omega^\pm$ defined in (5-23) and (5-37). The main tool is the following proposition.

Proposition 6.6. *There exist $\delta \in [0, 1[$, $C > 0$ and $\varepsilon_0 > 0$ such that, for every $\rho_0 \in \mathcal{T}$, if $X = W_{u/s}(\rho_0) \cap \mathcal{T} \cap U_{\rho_0}$,*

$$N_X(\varepsilon) \le C\varepsilon^{-\delta} \quad \text{for all } \varepsilon \le \varepsilon_0.$$

Remark. Recall that $W_{u/s}(\rho_0)$ is a local unstable (resp. stable) manifold at ρ_0, and in particular a single smooth curve. U_{ρ_0} is the domain of the chart adapted κ_{ρ_0} (see Lemma 3.23).

Roughly speaking, this proposition says that the upper-box dimension of the sets $W_{u/s}(\rho) \cap \mathcal{T}$, the trace of $\mathcal{T}$ along the stable and unstable manifolds, is strictly smaller than 1. This condition on the upper-box dimension is a fractal condition. In our case, we need uniform estimates on the numbers $N_X(\varepsilon)$ for $X = W_{u/s}(\rho) \cap \mathcal{T}$. This uniformity is a consequence of the fact that the holonomy maps are C^1 with uniform C^1 bounds (and thus Lipschitz, which is enough to conclude). This result is clearly linked with Bowen's formula, which has been proved in different contexts and links the dimension of X with the topological pressure of the map $\phi_u = -\log|J_u^1|$. This is where the assumption (Fractal) is used. This proposition is proved in Section A4 of the Appendix where we borrow the arguments of [Barreira 2008, Section 4.3] to get the required bounds.

From the Proposition 6.6, we get:

Corollary 6.7. *There exists $\nu > 0$ such that, for every $\rho_0 \in \mathcal{T}$, the sets $y \circ \kappa(W_u(\rho_0) \cap \mathcal{T} \cap U_{\rho_0})$ and $\zeta(W_s(\rho_0) \cap \mathcal{T} \cap U_{\rho_0})$ are ν-porous on a scale from 0 to 1.*

Proof. The maps $y \circ \kappa$ and ζ are C-Lipschitz for a global constant C. As a consequence, the previous lemma and Lemma 6.4 give,

$$\text{for all } \varepsilon \leq \varepsilon_0/C, \quad N_\Omega(\varepsilon) \leq C^\delta \varepsilon^{-\delta}, \quad \text{where } \Omega = y \circ \kappa(W_u(\rho_0) \cap \mathcal{T} \cap U_{\rho_0}) \text{ or } \zeta(W_s(\rho_0) \cap \mathcal{T} \cap U_{\rho_0}).$$

Applying Lemma 6.3, the ν-porosity is proved for some $\nu = \nu(\delta, C, \varepsilon_0)$. ☐

To conclude, we use this corollary to show the porosity of $\Omega^\pm$. We start by studying Ω^+.

Lemma 6.8. *There exists a global constant $C > 0$ such that*

$$\Omega^+ \subset \zeta(W_s(\rho_0) \cap \mathcal{T} \cap U_{\rho_0})(Ch^\tau).$$

Proof. Since $\Omega^+ = \Gamma^+(h^\tau)$, it is enough to show the same statement for $\Gamma^+ = \eta \circ \kappa_{\rho_0}(\mathcal{V}_\mathcal{Q}^+)$.

Let $\rho \in \mathcal{V}_\mathcal{Q}^+$. By assumption on $\mathcal{Q}$ and ρ_0, $d(\rho, W_u(\rho_0)) \leq Ch^\mathfrak{b}$. Since $\rho \in \mathcal{V}_{\boldsymbol{q}}$ for some $\boldsymbol{q} \in \mathcal{Q}$, there exists $\rho_1 \in \mathcal{T}$ such that $d(\rho, W_u(\rho_1)) \leq C/J_{\boldsymbol{q}}^+(\rho_1) \leq Ch^\tau$. Fix $\rho_2 \in W_u(\rho_1)$ such that $d(\rho, \rho_2) \leq Ch^\tau$. Then

$$|\eta \circ \kappa(\rho) - \zeta(\rho_1)| = |\eta \circ \kappa(\rho) - \zeta(\rho_2)| \leq |\eta \circ \kappa(\rho) - \eta \circ \kappa(\rho_2)| + |\eta \circ \kappa(\rho_2) - \zeta(\rho_2)|.$$

Since $\eta \circ \kappa$ is Lipschitz, we can control the first term by

$$|\eta \circ \kappa(\rho) - \eta \circ \kappa(\rho_2)| \leq Cd(\rho, \rho_2) \leq Ch^\tau.$$

To estimate the second term, the same arguments used after Lemma 5.14 show that

$$|\eta \circ \kappa(\rho_2) - \zeta(\rho_2)| \leq \operatorname{diam}[\eta \circ \kappa(W_u(\rho_2) \cap U_{\rho_0})] \leq Ch.$$

It gives $|\eta \circ \kappa(\rho) - \zeta(\rho_1)| \leq Ch^\tau$. To conclude, note that there exists a unique point $\rho_1' \in W_s(\rho_0) \cap W_u(\rho_1)$ and $\zeta(\rho_1) = \zeta(\rho_1')$. ☐

As a simple corollary of this lemma and of Lemma 6.2, we get:

Corollary 6.9. *Ω^+ is $\nu/3$-porous on a scale from $(3/\nu)Ch^\tau$ to* 1.

We now turn to the study of Ω^-. We can state and prove similar results with different scales of porosity. Recall that $\delta_2 = (\lambda_0/\lambda_1)\delta_0$.

Lemma 6.10. *There exists a global constant $C > 0$ such that*

$$\Omega^- \subset y \circ \kappa(W_u(\rho_0) \cap \mathcal{T} \cap U_{\rho_0})(Ch^{\delta_2}).$$

Proof. Since $\Omega^- = \Gamma^-(h^{\delta_0})$ with $\delta_0 > \delta_2$, it is enough to prove it for

$$\Gamma^- = y \circ \kappa\big(\mathcal{V}_a^+ \cap \mathcal{T}_-^{\mathrm{loc}}(4C_2h^{\delta_2}) \cap \{|\eta| \leq h^{\delta_0}\}\big).$$

Recall that $\mathcal{T}_-^{\mathrm{loc}} \subset \bigcup_{\rho \in \mathcal{T}} W_s(\rho)$. Since in $\mathcal{V}_a^+$ all the local stable leaves intersect $W_u(\rho_0)$, we have

$$\mathcal{V}_a^+ \cap \mathcal{T}_-^{\mathrm{loc}}(4C_2h^{\delta_2}) \subset \bigcup_{\rho \in W_u(\rho_0) \cap \mathcal{T}} W_s(\rho)(4C_2h^{\delta_2}).$$

Fix $\rho \in W_u(\rho_0) \cap \mathcal{T}$. Since $d\kappa(E_s(\rho_0)) = \mathbb{R}\partial_\eta$, if ε_0 is small enough, we can write $\kappa(W_s(\rho)) = \{(G_\rho(\eta), \eta) : \eta \in O\}$, where O is some open subset of $\mathbb{R}$ and $G_\rho : O \to \mathbb{R}$ is C^∞. In particular, it is Lipschitz with a global

Lipschitz constant C. If $|\eta| \leq h^{\delta_0}$, then $|G_\rho(\eta) - G_\rho(0)| \leq Ch^{\delta_0}$. Recall that $\kappa(W_u(\rho_0) \cap U_{\rho_0}) \subset \mathbb{R} \times \{0\}$ and hence, $G_\rho(0) = y \circ \kappa(\rho)$. As a consequence, if $\rho_1 \in W_s(\rho) \cap \{|\eta| \leq h^{\delta_0}\}$, writing $\kappa(\rho_1) = (G_\rho(\eta), \eta)$, we have

$$|y \circ \kappa(\rho_1) - y \circ \kappa(\rho)| = |G_\rho(\eta) - G_\rho(0)| \leq Ch^{\delta_0}.$$

Then, if $\rho_2 \in W_s(\rho)(4C_2 h^{\delta_2})$, since κ is Lipschitz with global Lipschitz constant,

$$|y \circ \kappa(\rho_2) - y \circ \kappa(\rho)| \leq Ch^{\delta_2} + Ch^{\delta_0} \leq Ch^{\delta_2}.$$

This shows that $y \circ \kappa(\rho_2) \in y \circ \kappa(W_u(\rho_0) \cap \mathcal{T})(Ch^{\delta_2})$ and concludes the proof. □

As a corollary, using Lemma 6.2, we get:

Corollary 6.11. *Ω^- is $\nu/3$-porous on a scale from $(3/\nu)Ch^{\delta_2}$ to 1.*

We can now prove the last Proposition 5.15 needed to end the proof of Proposition 4.2. This is a consequence of the porosity of $\Omega^\pm$ and the fractal uncertainty principle. To apply Proposition 6.5, we need to ensure that the scale condition is satisfied, that is to say

$$\delta_2 + \tau > 1,$$

which has been supposed when defining τ in (4-5) and (4-6). Proposition 4.2 then comes with any $0 < \gamma < (\delta_2 + \tau - 1)\beta(\nu/3)$.

Appendix

A1. *Holder regularity for flows.*

Lemma A.1. *Let $U \subset \mathbb{R}^n$ be open and $Y : U \to \mathbb{R}^n$ be a complete $C^{1+\beta}$ vector field. We denote by $\phi^t(x)$ the flow generated by Y. Then, for any $T \in \mathbb{R}$ and $K \subset U$ compact, the map*

$$(t, x) \in [-T, T] \times K \mapsto \phi^t(x)$$

is $C^{1+\beta}$.

Proof. We fix T, K as in the statement. We'll use the same constants C, C' at different places, with different meaning. In addition to Y, they will depend on T, K.

Since Y is C^1, Cauchy–Lipschitz theorem gives the local existence and uniqueness of the flow. It is standard that the flow is also C^1 and satisfies

$$\partial_t d\phi^t(x) = dY(\phi^t(x)) \circ d\phi^t(x). \tag{A-1}$$

Let's define $A^t(x) = d\phi^t(x)$ and $\Xi(t, x) = dY(\phi^t(x))$. The assumption on Y implies that Ξ is β-Hölder.

Fix $(t_0, x_0), (t_1, x_1) \in [-T, T] \times K$ and let's estimate $\|A^{t_1}(x_1) - A^{t_0}(x_0)\|$. We split it into two pieces and control it with the triangle inequality:

$$\|A^{t_1}(x_1) - A^{t_0}(x_0)\| \leq \|A^{t_1}(x_1) - A^{t_0}(x_1)\| + \|A^{t_0}(x_1) - A^{t_0}(x_0)\|.$$

It is not hard to control the first term of the right-hand side using (A-1) since

$$\|A^{t_1}(x_1)-A^{t_0}(x_1)\| = \left|\int_{t_0}^{t_1} \Xi(s,x_1)\circ A^s(x_1)\,ds\right| \leq C|t_1-t_0|.$$

To estimate the second term, we estimate

$$\begin{aligned}\|\partial_t(A^t(x_1)-A^t(x_0))\| &\leq \|(\Xi(t,x_1)-\Xi(t,x_0))\circ A^t(x_1)+\Xi(t,x_0)\circ(A^t(x_1)-A^t(x_0))\|\\ &\leq Cd(x_0,x_1)^\beta + C'\|A^t(x_1)-A^t(x_0)\|.\end{aligned}$$

By Gronwall's lemma,

$$\|A^{t_0}(x_1)-A^{t_0}(x_0)\| \leq Cd(x_0,x_1)^\beta e^{C't_0} \leq Cd(x_0,x_1)^\beta. \qquad \square$$

A2. *Proof of Lemma 3.24.* We give the missing proof of Lemma 3.24 and widely use the notation of the Section 3E. Its proof uses the construction of e_u in the proof of Theorem 5. It is inspired by techniques usually used to show the unstable manifold theorem; see for instance [Dyatlov 2018]. In fact, the smoothness of $y\mapsto f_0(y,0)$ is a direct consequence of the smoothness of the unstable manifold $W_u(\rho_0)$. It was not clear for us if it was possible to easily deduce from this the required smoothness of $y\mapsto \partial_\eta f_0(y,0)$. This is why we decided to give a proof of this proposition. It uses the fact that e_u has been constructed to satisfy $\mathbb{R}d_\rho F(e_u(\rho))=\mathbb{R}e_u(F(\rho))$ for ρ in a small neighborhood of $\mathcal{T}$. To show the lemma, we need information along all the orbit of ρ_0. For this purpose, we introduce the following, for $m\in\mathbb{Z}$:

- $\rho_m = F^m(\rho_0)$.
- $\kappa_m : U_m \to V_m \subset \mathbb{R}^2$ the chart given by Lemma 3.22 centered at ρ_m and we assume that the relation $\mathbb{R}d_\rho F(e_u(\rho))=\mathbb{R}e_u(F(\rho))$ holds for $\rho\in U_m$. We will denote by (y_m,η_m) the variable in V_m.
- $G_m = \kappa_{m+1}\circ F\circ\kappa_m^{-1} : V_m\to V_{m+1}$.
- A reparametrization of the vector field $(\kappa_m)_*e_u$: $\mathbb{R}(\kappa_m)_*e_u=\mathbb{R}e_m$, where $e_m(y_m,\eta_m)={}^t(1,s_m(y_m,\eta_m))$, where s_m is a slope function which is known to be $C^{1+\beta}$.

Note that $s_m(y_m,0)=0$ due to the fact that $\kappa_m(W_u(\rho_m))\subset\mathbb{R}\times\{0\}$. The hyperbolicity assumption on F and the properties of κ_m allow us to write

$$G_m(y_m,\eta_m) = (\lambda_m y_m + \alpha_m(y_m,\eta_m),\ \mu_m\eta_m+\beta_m(y_m,\eta_m)),$$

where

- For some $\nu<1$, $0\leq|\mu_m|\leq\nu$, $|\lambda_m|\geq\nu^{-1}$ for all $m\in\mathbb{N}$.
- $\alpha_m(0,0)=\beta_m(0,0)=0$.
- $\beta_m(y_m,0)=0$ for $(y_m,0)\in V_m$.
- $d\alpha_m(0,0)=d\beta_m(0,0)=0$.
- We can assume that U_m are sufficiently small neighborhoods of ρ_m so that $\beta_m,\alpha_m=O(\delta_0)_{C^1(U_m)}$ for some small $\delta_0>0$.

The property $d_\rho F(e_u(\rho)) \in \mathbb{R}e_u(F(\rho))$ implies that $d_{(y_m,\eta_m)} G_m(e_m(y_m,\eta_m)) \in \mathbb{R}e_{m+1}(G_m(y_m,\eta_m))$. As a consequence, the transformation of the slopes gives an equation satisfied by the family of slopes $(s_m)_{m\in\mathbb{Z}}$:

$$s_{m+1}(G_m(y_m,\eta_m)) = Q_m(y_m,\eta_m,s_m(y_m,\eta_m)), \tag{A-2}$$

where Q_m is the smooth function

$$Q_m(y_m,\eta_m,s) = \frac{s\times(\mu_m+\partial_{\eta_m}\beta_m(y_m,\eta_m)) + \partial_{y_m}\beta_m(y_m,\eta_m)}{\lambda_m + \partial_{y_m}\alpha_m(y_m,\eta_m) + s\times\partial_{\eta_m}\alpha_m(y_m,\eta_m)}.$$

Writing $G_m(y_m,\eta_m) = (y_{m+1},\eta_{m+1})$, we deduce by differentiation of (A-2) with respect to η_{m+1} (we omit the point of evaluation of the maps involved in the right-hand side to alleviate the line)

$$\begin{aligned}\partial_{\eta_{m+1}} s_{m+1}(y_{m+1},\eta_{m+1}) = \partial_{y_m} Q_m \times \partial_{\eta_{m+1}} y_m + \partial_{\eta_m} Q_m \times \partial_{\eta_{m+1}}\eta_m \\ + \partial_s Q_m \times (\partial_{y_m} s_m \times \partial_{\eta_{m+1}} y_m + \partial_{\eta_m} s_m \times \partial_{\eta_{m+1}}\eta_m).\end{aligned} \tag{A-3}$$

This last equation gives the transformation of vertical derivative of the slope. We now evaluate this identity at the point $(y_{m+1},0)$. In the following lines, when the variables y_m and y_{m+1} appear in the same equation, we implicitly assume that they are related by $(y_{m+1},0) = G_m(y_m,0)$, namely $y_{m+1} = \lambda_m y_m + \alpha_m(y_m,0)$. We remark that due to the fact that $\beta_m(y_m,0)=0$, we have $Q_m(y_m,0,0)=0$ and the first term of the right-hand side vanishes. The term $\partial_{y_m} s_m$ also vanishes at $(y_m,0)$. We will write

$$\begin{aligned}\sigma_m(y_m) &= \partial_{\eta_m} s_m(y_m,0),\\ h_m(y_m) &= \partial_{\eta_m} Q_m(y_m,0,0)\times\partial_{\eta_{m+1}}\eta_m(y_{m+1},0),\\ c_m(y_m) &= \partial_s Q_m(y_m,0,0)\times\partial_{\eta_{m+1}}\eta_m(y_{m+1},0).\end{aligned}$$

This notation allows us to rewrite (A-3) at $(y_{m+1},0)$:

$$\sigma_{m+1}(y_{m+1}) = h_m(y_m) + c_m(y_m)\times\sigma_m(y_m). \tag{A-4}$$

We observe that $|\partial_{\eta_{m+1}}\eta_m(y_m,0)| = |\mu_m^{-1} + O(\delta_0)_{C^0}|$ and after some computations, we see that

$$\partial_s Q_m(y_m,0,0) = \frac{\mu_m}{\lambda_m} + O(\delta_0)_{C^0}.$$

As a consequence,

$$|c_m(y_m)| = |\lambda_m^{-1}| + O(\delta_0)_{C^0} \le \nu_1, \tag{A-5}$$

where, if δ_0 is small enough, we can fix $\nu_1 < 1$. Moreover, c_m and h_m are smooth functions and their C^N norms are bounded uniformly in m, and actually by global constants depending only on F. Furthermore, $y_m \mapsto y_{m+1}$ is given by $y_m \mapsto \lambda_m y + \alpha_m(y_m,0)$ and is an expanding diffeomorphism provided δ_0 is small enough.

We fix some small ε such that $(-\varepsilon,\varepsilon)\times\{0\}\subset U_m$ for all m. Let's define $I=(-\varepsilon,\varepsilon)$. We will make use of the fiber contraction theorem to show that $y_m\in I\mapsto\sigma_m(y_m)$ is smooth for every m, with uniform C^N norms. For this purpose, let us introduce the following notation:

- $C_0 \le C_1 \le \cdots \le C_N \le \cdots$ a family of constants which will be specified in the sequel.

- The complete metric space $X_N = \{\gamma \in C^N(I) : \|\gamma\|_{C^k} \leq C_k,\ 0 \leq k \leq N\}$ equipped with the C^N norm.
- The auxiliary metric space $X_N^{\text{aux}} = \{\gamma \in C^0(I) : \|\gamma\|_\infty \leq C_N\}$ equipped with the C^0 norm.
- The complete metric space $E_N = (X_N)^{\mathbb{Z}}$ equipped with the metric

$$d(\gamma_1, \gamma_2) = \sup_{m\in\mathbb{Z}} \|(\gamma_1)_m - (\gamma_2)_m\|_{C^N}.$$

- Its auxiliary counterpart $E_N^{\text{aux}} = (X_N^{\text{aux}})^{\mathbb{Z}}$ equipped with the metric

$$d(\gamma_1, \gamma_2) = \sup_{m\in\mathbb{Z}} \|(\gamma_1)_m - (\gamma_2)_m\|_{C^0}.$$

For $\gamma \in E_N$, let's define $T\gamma$ with the formula (A-4):

$$(T\gamma)_{m+1}(y_{m+1}) = (h_m + c_m\gamma_m)(y_m).$$

Since $y_m \mapsto y_{m+1}$ is expanding, we see that $y_{m+1} \in I \implies y_m \in I$. Hence, $(T\gamma)_{m+1}$ is well-defined on I. Our aim is to show by induction on N that for every $N \in \mathbb{N}$, $\sigma := (\sigma_m)_{m\in\mathbb{Z}}$ is in E_N and is an attractive fixed point of $T : E_N \to E_N$.

We start with the case $N = 0$. We need to check that $T(E_0) \subset E_0$. It will be the case as soon as

$$C_0\nu_1 + \sup_m \|h_m\|_\infty \leq C_0.$$

For instance, take $C_0 = 2\sup_m \|h_m\|_\infty/(1-\nu_1)$. Due to the fact that $\|c_m\|_{C^0(I)} \leq \nu_1$, T is a contraction with contraction rate ν_1 and hence $T : E_0 \to E_0$ has a unique attractive fixed point. This fixed point is necessarily σ since σ satisfies (A-4).

Arguing by induction, we assume that $\sigma \in E_N$, $T(E_N) \subset E_N$ and σ is an attractive fixed point for T and we want to show that the same is true for $N+1$. For this purpose, suppose that $\gamma \in E_N$ is of class C^{N+1}. Analyzing the formula defining T, we see that can write, for $m \in \mathbb{Z}$,

$$\begin{aligned}(T\gamma)_m^{(N+1)}(y_{m+1}) = h_m^{(N+1)}(y_m) + c_m(y_m) \times \left(\frac{\partial y_{m+1}}{\partial y_m}(y_m)\right)^{-N-1} \times \gamma_m^{(N+1)}(y_m)& \\ + R_{N,m}(y_m, \gamma_m(y_m), \dots, \gamma_m^{(N)}(y_m)),& \quad \text{(A-6)}\end{aligned}$$

where $R_{N,m} : I \times [-C_0, C_0] \times \cdots \times [-C_N, C_N] \to \mathbb{R}$ is a polynomial in the last $N+1$ variables with smooth coefficients in y_m, uniformly bounded in m. As a consequence, there exists a global constant C'_{N+1} such that

$$\sup_m \sup_{I\times[-C_0,C_0]\times\cdots\times[-C_N,C_N]} |R_{N,m}(y_m, \tau_0, \dots, \tau_N)| \leq C'_{N+1}.$$

We can then choose $C_{N+1} \geq C_N$ such that

$$\sup_m \|h_m\|_{C^{N+1}} + C'_{N+1} + \nu_1 C_{N+1} \leq C_{N+1},$$

which ensures that $T : E_{N+1} \to E_{N+1}$. We now wish to use the fiber contraction theorem (Theorem 6). If $\gamma \in E_N$, we define the map $S_\gamma : E_{N+1}^{\text{aux}} \to E_{N+1}^{\text{aux}}$ by

$$(S_\gamma\theta)_{m+1}(y_{m+1}) = h_m^{(N+1)}(y_m) + c_m(y_m) \times \left(\frac{\partial y_{m+1}}{\partial y_m}(y_m)\right)^{-N-1} \times \theta_m(y_m) + R_{N,m}(y_m, \gamma_m(y_m), \dots, \gamma_m^N(y_m)).$$

Due to the choice of C_{N+1}, we see that S_γ is well-defined and since we have

$$\left|\frac{\partial y_{m+1}}{\partial y_m}(y_m)\right| \geq 1$$

and $\|c_m\|_{C^0(I)} \leq \nu_1$, we know S_γ is a contraction with contraction rate ν_1 for every $\gamma \in E_N$. In particular, the map S_σ has a unique fixed point $\sigma_{N+1} \in E^{\mathrm{aux}}_{N+1}$.

The fiber contraction theorem (Theorem 6) applies to the continuous map

$$T_N : (\gamma, \theta) \in E_N \times E^{\mathrm{aux}}_{N+1} \mapsto (T\gamma, S_\gamma \theta) \in E_N \times E^{\mathrm{aux}}_{N+1}$$

and (σ, σ_{N+1}) is an attractive fixed point of T_N in $E_N \times E^{\mathrm{aux}}_{N+1}$.

In particular, if $\gamma \in E_{N+1}$, then $\tilde{\gamma} := (\gamma, \gamma^{(N+1)}) \in E_N \times E^{\mathrm{aux}}_{N+1}$ and

$$\lim_{p \to +\infty} T_N^p \tilde{\gamma} = (\sigma, \sigma_{N+1}) \quad \text{in } E_N \times E^{\mathrm{aux}}_{N+1}.$$

However, by the definition of S_γ,

$$T_N^p \tilde{\gamma} = (T^p \gamma, (T^p \gamma)^{(N+1)}).$$

Hence, for every fixed m, we know $(T^p\gamma)_m$ converges to σ_m in X_N and $(T^p\gamma)_m^{(N+1)}$ converges uniformly on I to σ_{N+1}. This proves that σ is C^{N+1} and $\sigma^{(N+1)} = \sigma_{N+1}$. We conclude that $\sigma \in E_{N+1}$ is then an attractive fixed point of $T : E_{N+1} \to E_{N+1}$, which proves the induction and concludes the proof of Lemma 3.24.

A3. *Proof of Lemma 5.9.* We give the missing proof of Lemma 5.9. The proof is a precise analysis of the iteration formula (5-17). We adopt the notation introduced for Lemma 5.9. We argue by induction on J to show the property $\mathcal{P}_J$: the bound (5-18) is valid for all $j \leq J$ and, for all $1 \leq i \leq n-1$, $l \in \mathbb{N}$, with some constants $C_{j,l}$.

1. *Base case.* Let us start with $\mathcal{P}_0$. The iteration formula (5-17) implies

$$a_{i,0}(x_i) = \prod_{l=1}^{i} f_l(x_l).$$

Hence, the bound $\|a_{i,0}\|_{C^0} \leq (B\nu^{1/2})^i$ is obvious and we can set $C_{0,0} = 1$. We now argue by induction on i and prove the property $\mathcal{P}_{0,i}$: the bound (5-18) is valid for $j = 0$, i and for all $l \in \mathbb{N}$, for some constants $C_{j,l}$. These bounds are trivially true for $i = 0$ and are direct consequences of Lemma 5.8 for $i = 1$. Suppose that the property holds for $i-1$ for some $i \geq 1$ and let's show it for i.

1.1. *Case $l = 1$.* Let us first deal with $l = 1$ and compute the derivative of $a_{i,0}$, using the formula $a_{i,0}(x_i) = f_i(x_i)a_{i-1,0}(x_{i-1})$:

$$a'_{i,0}(x_i) = f'(x_i)a_{i-1,0}(x_{i-1}) + f_i(x_i)a'_{i-1,0}(x_{i-1})\left(\frac{\partial x_{i-1}}{\partial x_i}\right).$$

We use the (weak) bound $|\partial x_{i-1}/\partial x_i| \leq 1$ and the property $\mathcal{P}_{0,i-1}$ to show that

$$\|a_{i,0}\|_{C^1} \leq C(B\nu^{1/2})^{i-1} + C_{0,1}(B\nu^{1/2}) \times (B\nu^{1/2})^{i-1} i \leq C_{0,1}(B\nu^{1/2})^i (i+1),$$

assuming that $C_{0,1} > C(B\nu^{1/2})^{-1}$.

1.2. ***General case for $l > 0$.*** We now come back to the general case $l > 0$. By using the formula $a_{i,0}(x_i) = f_i(x_i)a_{i-1,0}(x_{i-1})$, one sees that we can write $a_{i,0}^{(l)}$ in the form

$$a_{i,0}^{(l)}(x_i) = f_i(x_i)a_{i-1,0}^{(l)}(x_{i-1})\left(\frac{\partial x_{i-1}}{\partial x_i}\right)^l + O(\|a_{i-1,0}\|_{C^{l-1}}).$$

The constants appearing in the O depend on C^l norms of f_i and ϕ_i, which, by assumption are controlled by some uniform C_l'. Hence, using the assumption $\mathcal{P}_{0,i-1}$,

$$\begin{aligned}|a_{i,0}^{(l)}(x_i)| &\le (B\nu^{1/2})\|a_{i-1,0}\|_{C^l}\left(\frac{\partial x_{i-1}}{\partial x_i}\right)^l + C_l'\|a_{i-1,0}\|_{C^{l-1}}\\ &\le C_{0,l}(B\nu^{1/2})(B\nu^{1/2})^{i-1}i^l + C_l'C_{0,l-1}(B\nu^{1/2})^{i-1}i^{l-1}\\ &\le C_{0,l}(B\nu^{1/2})^i(i+1)^l,\end{aligned}$$

assuming that $C_{0,l}$ is chosen bigger than $(1/l)C_l'C_{0,l-1}(B\nu^{1/2})^{-1}$. As a consequence, we can build constants satisfying these conditions by defining inductively

$$C_{0,l} = \max\Big(C_{0,l-1}, \frac{1}{l}C_l'C_{0,l-1}(B\nu^{1/2})^{-1}\Big).$$

This ends the proof of $\mathcal{P}_{0,i}$ and hence of $\mathcal{P}_0$.

2. ***Induction step.*** We now assume that $\mathcal{P}_{j-1}$ is true for some $j \ge 1$ and aim at proving $\mathcal{P}_j$. Again, we do it by induction on i by proving the properties $\mathcal{P}_{j,i}$: the bound (5-18) is true for j, i and all $l \in \mathbb{N}$. These bounds are trivially true for $i = 0$ and are direct consequences of Lemma 5.8 for $i = 1$. Suppose that the property holds for $i - 1$ for some $i \ge 2$ and let's show it for i.

2.1 ***Case $l = 0$.*** Let's start with $l = 0$. The iteration formula shows that

$$a_{i,j}(x_i) = f_i(x_i)a_{i-1,j}(x_{i-1}) + \sum_{p=0}^{j-1} L_{j-p,i}(a_{i-1,p})(x_{i-1}).$$

By Lemma 5.8, there exist constants $C_{p,m}' > 0$ such that

$$\|L_{p,i}a\|_{C^m(I_i)} \le C_{p,m}'\|a\|_{C^{2p+m}(I_{i-1})}.$$

Hence, assuming that (5-18) holds for $a_{i-1,j}$ with $l = 0$,

$$\begin{aligned}\|a_{i,j}\|_\infty &\le C_{j,0}(B\nu^{1/2})(B\nu^{1/2})^{i-1}i^{3j} + \sum_{p=0}^{j-1} C_{j-p,0}'\|a_{i-1,p}\|_{C^{2(j-p)}}\\ &\le C_{j,0}(B\nu^{1/2})^i i^{3j} + \sum_{p=0}^{j-1} C_{j-p,0}'C_{p,2(j-p)}(B\nu^{1/2})^{i-1}i^{2(j-p)+3p}\\ &\le C_{j,0}(B\nu^{1/2})^i i^{3j} + i^{2j}(B\nu^{1/2})^{i-1}\sum_{p=0}^{j-1} C_{j-p,0}'C_{p,2(j-p)}i^p\\ &\le C_{j,0}(B\nu^{1/2})^i i^{3j} + i^{2j}(B\nu^{1/2})^{i-1}\Big[\sup_{0\le p\le j-1} C_{j-p,0}'C_{p,2(j-p)}\Big]\frac{i^j-1}{i-1}\end{aligned}$$

$$\begin{aligned}&\le C_{j,0}(B\nu^{1/2})^i i^{3j} + i^{3j-1}(B\nu^{1/2})^{i-1}\big[\sup_{0\le p\le j-1} C'_{j-p,0}C_{p,2(j-p)}\big]\widetilde{C}_j, \quad \text{where } \frac{i^j-1}{i-1}\le \widetilde{C}_j i^{j-1},\\ &\le C_{j,0}(B\nu^{1/2})^i(i+1)^{3j},\end{aligned}$$

assuming that $C_{j,0}$ is chosen bigger than

$$K_j := \frac{1}{3j}(B\nu^{1/2})^{-1}\big[\sup_{0\le p\le j-1} C'_{j-p,0}C_{p,2(j-p)}\big]\widetilde{C}_j.$$

As a consequence, the bounds hold for $l=0$ and i, j if we set $C_{j,0}=\max(1, K_j)$.

2.2. ***Case $l > 0$.*** Consider now $l>0$. As already done, one can write

$$a^{(l)}_{i,j}(x_i) = f_i(x_i)a^{(l)}_{i-1,j}(x_{i-1})\left(\frac{\partial x_{i-1}}{\partial x_i}\right)^l + O(\|a_{i-1,j}\|_{C^{l-1}}) + \sum_{p=0}^{j-1}(L_{j-p,i}(a_{i-1,p}))^{(l)}(x_{i-1}).$$

As usual, the constants in O depend on l, j but not on i and we denote by $C''_{l,j}$ the constant in this O. Hence, we can control

$$\begin{aligned}\|a^{(l)}_{i,j}\|_\infty &\le C_{j,l}(B\nu^{1/2})(B\nu^{1/2})^{i-1}i^{l+3j} + C''_{l,j}C_{j,l-1}(B\nu^{1/2})^{i-1}i^{l+3j-1} + \sum_{p=0}^{j-1}\|L_{j-p,i}(a_{i-1,p})\|_{C^l}\\ &\le C_{j,l}(B\nu^{1/2})^i i^{l+3j} + C''_{l,j}C_{j,l-1}(B\nu^{1/2})^{i-1}i^{l+3j-1} + \sum_{p=0}^{j-1}C'_{j-p,l}\|a_{i-1,p}\|_{C^{l+2(j-p)}}\\ &\le C_{j,l}(B\nu^{1/2})^i i^{l+3j} + C''_{l,j}C_{j,l-1}(B\nu^{1/2})^{i-1}i^{l+3j-1} + \sum_{p=0}^{j-1}C'_{j-p,l}C_{p,l+2(j-p)}(B\nu^{1/2})^{i-1}i^{l+2(j-p)+3p}\\ &\le C_{j,l}(B\nu^{1/2})^i\Big(i^{l+3j} + i^{l+3j-1}\frac{1}{C_{j,l}}\underbrace{(B\nu^{1/2})^{-1}\big(C''_{l,j}C_{j,l-1} + \sup_{0\le p\le j-1}C'_{j-p,l}C_{p,l+2(j-p)}\widetilde{C}_j\big)}_{\widetilde{C}_{j,l}}\\ &\le C_{j,l}(B\nu^{1/2})^i(i+1)^{l+3j}\end{aligned}$$

if $C_{j,l}\ge \widetilde{C}_{j,l}$. Eventually, we define by induction on l the constants $C_{j,l}$ by setting $C_{j,l}=\max(C_{j,l-1}, \widetilde{C}_{j,l})$, achieving the proof of $\mathcal{P}_j$. This concludes the proof of the lemma.

A4. ***Upper box dimension for hyperbolic set.*** This subsection is devoted to the proof of Proposition 6.6. We will simply recall some arguments which lead to give an upper bound to the upper-box dimension. We borrow these arguments from [Barreira 2008, Section 4.3] and refer the reader to this book for the definitions and properties of topological pressure (Definition 2.3.1), Markov partition (Definition 4.2.6) and other references on this theory.

We'll show that the pressure condition (Fractal) implies Proposition 6.6. We prove it for the unstable manifolds. The proof is similar in the case of stable manifolds by changing F into F^{-1}. We first begin by fixing a Markov partition for $\mathcal{T}$ with diameter at most η_0. This is possible by virtue of Theorem 18.7.3 in [Katok and Hasselblatt 1995]. We denote by $R_1, \dots, R_p \subset \mathcal{T}$ this Markov partition. Here, η_0 is smaller than the diameter of the local stable and unstable manifolds and the holonomy maps $H^{u/s}_{\rho,\rho'}$ are well-defined for $d(\rho, \rho') \leq \eta_0$:

$$H^{u/s}_{\rho,\rho'} : W_{s/u}(\rho) \to W_{s/u}(\rho'), \quad \zeta \mapsto \text{the unique point in } W_u(\zeta) \cap W_s(\rho').$$

Due to our results on the regularity of the stable and unstable distributions, these maps are Lipschitz with global Lipschitz constants. In particular, if an inequality of the kind

$$N_{W_u(\rho)\cap\mathcal{T}}(\varepsilon) \leq C\varepsilon^{-\delta}$$

holds for some ρ, it holds for ρ' if $d(\rho, \rho') \leq \eta_0$ with C replaced by $K^\delta C$ where K is a Lipschitz constant for the holonomy maps. We fix $(\rho_1, \dots, \rho_p)$ in $(R_1, \dots, R_p)$ and we set $V = \bigcup_{i=1}^p W_u(\rho_i) \cap R_i$. It is then enough to show that

$$\overline{\dim}\, V < 1.$$

Indeed, if $\overline{\dim}\, V < 1$ for $\delta \in (\overline{\dim}\, V, 1)$, there exists $\varepsilon_0 > 0$ such that,

$$\text{for all } \varepsilon \leq \varepsilon_0, \quad N_V(\varepsilon) \leq \varepsilon^{-\delta},$$

and we conclude the proof of Section A4 with the above considerations on the holonomy maps.

Note $\delta := \overline{\dim}\, V$ satisfies the equation $P(\delta\phi_u) = 0$. We will actually show that $P(\delta\phi_u) \geq 0$. Since $s \mapsto P(s\phi_u)$ is strictly decreasing and has a unique root, the assumption $P(\phi_u) < 0$ will give $\delta < 1$. We will denote by

$$R_{i_0,\dots,i_n} = \bigcap_{k=0}^{n} F^{-i}(R_{i_k}), \quad V_{i_0,\dots,i_n} = R_{i_0,\dots,i_n} \cap V$$

the elements of the refined partition at time n. Similarly to the definitions of $J^+_{\boldsymbol{q}}$, we will write

$$J_{i_0,\dots,i_n} = \inf\{J^n_u(\rho), \rho \in R_{i_0,\dots,i_n}\}$$

and write

$$c_n(s) = \sum_{i_0,\dots,i_n} J^{-s}_{i_0,\dots,i_n} = \sum_{i_0,\dots,i_n} \exp \max_{R_{i_0,\dots,i_n}} \left(s \sum_{k=0}^{n-1} \phi_u \circ F^k \right)$$

(the last equality follows from the chain rule). Properties of Markov partitions ensure that

$$P(s\phi_u) = \lim_{n\to\infty} \frac{1}{n} \log c_n(s).$$

Fix $s > \delta$. Hence, there exists ε_1 such that, for all $\varepsilon \leq \varepsilon_1$, $N_V(\varepsilon) \leq \varepsilon^{-s}$.

Fix $n \in \mathbb{N}^*$. By writing $V = \bigcup_{i_0,\dots,i_n} V_{i_0,\dots,i_n}$ we have

$$N_V(\varepsilon) \leq \sum_{i_0,\dots,i_n} N_{V_{i_0,\dots,i_n}}(\varepsilon).$$

Note that

$$F^n(V_{i_0,\dots,i_n}) \subset W_u(F^n(\rho_{i_0})) \cap R_{i_n}$$

and

$$H^s_{F^n(\rho_{i_0}),\rho_{i_n}}(F^n(V_{i_0,\dots,i_n})) \subset V_{i_n}.$$

Hence, if we cover V_{i_n} by N sets of diameter at most ε, $U_1, \dots, U_N$, the sets $F^{-n} \circ H^s_{\rho_{i_n}, F^n(\rho_{i_0})}(U_i)$, $1 \le i \le N$, cover $V_{i_0,\dots,i_n}$ and have diameters at most $K\varepsilon J^{-1}_{i_0,\dots,i_n}$. Hence,

$$N_{V_{i_n}}(\varepsilon) \ge N_{V_{i_0,\dots,i_n}}(K\varepsilon J^{-1}_{i_0,\dots,i_n}),$$

which gives

$$N_V(\varepsilon) \le \sum_{i_0,\dots,i_n} N_{V_{i_n}}(\varepsilon K^{-1} J_{i_0,\dots,i_n}).$$

As a consequence, if $\varepsilon < \varepsilon_1 K J_n^{-1}$, where $J_n = \sup_{i_0,\dots,i_n} J_{i_0,\dots,i_n}$, we have

$$N_V(\varepsilon) \le \sum_{i_0,\dots,i_n} K^s J^{-s}_{i_0,\dots,i_n} \varepsilon^{-s} = K^s \varepsilon^{-s} c_n(s).$$

By iterating this process, we see that, for all $m \in \mathbb{N}$, if $\varepsilon < \varepsilon_1 (K J_n^{-1})^m$,

$$N_V(\varepsilon) \le \varepsilon^{-s} K^{ms} c_n(s)^m.$$

Hence,

$$\frac{\log N_V(\varepsilon)}{-\log \varepsilon} \le s + m\frac{\log(K^s c_n(s))}{-\log \varepsilon} \le s + m\frac{\log(K^s c_n(s))}{-\log(\varepsilon_1 (K J_n^{-1})^m)}.$$

We then take the lim sup as $\varepsilon \to 0$ first and then pass to the limit as $m \to +\infty$ and find that

$$\overline{\dim}\, V \le s + \frac{\log K^s c_n(s)}{-\log K J_n^{-1}}.$$

Then, we pass to the limit $s \to \delta$ and find that $\log(K^\delta c_n(\delta)) \ge 0$. Hence,

$$P(\delta\phi_u) = \lim_{n\to\infty} \frac{1}{n} \log c_n(\delta) \ge \lim_{n\to\infty} \frac{-\delta \log K}{n} = 0.$$

This ends the proof of the required inequality and gives that $\overline{\dim}\, V < 1$.

A5. ***From porosity to upper-box dimension.*** We have shown that sets with upper-box dimension strictly smaller than 1 are porous. In this appendix, we show a result in the other way, namely, porous sets down to scale 0 have an upper-box dimension strictly smaller than 1. The following lemma gives a quantitative version of this statement. This is not useful for our use (we only needed the first implication) but we found that it could be of independent interest. Our proof is based on the proof of Lemma 5.4 in [Dyatlov and Jin 2018]. We adopt the same notation as in Section 6A.

Lemma A.2. *Let $M \in \mathbb{N}$, $\nu > 0$, $\alpha_1 > 0$. Let $X \subset [-M, M]$ be a closed set and assume that X is ν-porous on a scale from* 0 *to α_1. Then, there exists $C = C(\nu, \alpha_1, M) > 0$, $\varepsilon_0 = \varepsilon_0(\nu, \alpha_1, M)$ and $\delta = \delta(\nu) \in [0, 1[$ such that*,

$$\textit{for all } \varepsilon \le \varepsilon_0, \quad N_X(\varepsilon) \le C\varepsilon^{-\delta}.$$

In particular,

$$\overline{\dim}\, X \le \delta.$$

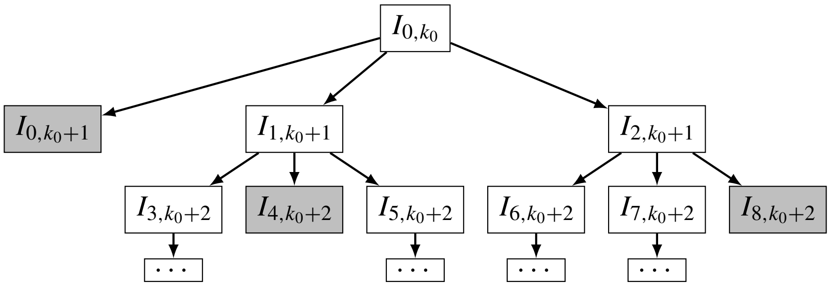

Figure 17. It illustrates the tree structure of the family of intervals $I_{k,m}$ with $L = 3$. The porosity allows us to withdraw at least one child to any parent. The missing children are shaded in gray.

Proof. We define $L = \lceil 2/\nu \rceil$ and denote by k_0 the unique integer such that

$$L^{-k_0} \le \alpha_1 < L^{-k_0+1}.$$

We will let $I_{m,k} = [mL^{-k}, (m+1)L^{-k}]$ for $k \in \mathbb{N},\ m \in \mathbb{Z}$.

We now show by induction on $k \ge k_0$ that there exists $Y_k \subset \mathbb{Z}$ such that

$$\# Y_k \le 2ML^{k_0}(L-1)^{k-k_0}, \quad \Omega \subset \bigcup_{m \in Y_k} I_{m,k}, \tag{A-7}$$

namely, at each level $k \ge k_0$, one new interval $I_{m,k}$ does not intersect Ω. See Figure 17.

The case $k = k_0$ is trivial since we simply cover Ω by the intervals I_{m,k_0} for $ML^{k_0} \le m < ML^{k_0}$.

We now assume that the result is proved for $k \ge k_0$ and we prove it for $k+1$. Fix $m \in Y_k$. We write $I = \bigcup_{j=0}^{L-1} I_{mL+j,k+1}$. We claim that among the intervals $I_{mL+j,k+1}$, at least one does not intersect Ω. Indeed, since $|I| \le L^{-k_0} \le \alpha_1$, the porosity of Ω implies the existence of an interval $J \subset I$ of size $\nu|I| = \nu L^{-k} \ge 2L^{-k-1}$ such that $J \cap \Omega = \varnothing$. Since $|J| \ge 2L^{-k-1}$, J contains at least one of the intervals $I_{mL+j,k+1}$. We denote this index by j_m. We now set

$$Y_{k+1} = \bigcup_{m \in Y_k} \{mL + j : j \in \{0, \dots, L_1\} \setminus j_m\}.$$

By the property of j_m, we have $\Omega \subset \bigcup_{m \in Y_{k+1}} I_{m,k+1}$ and $\# Y_{k+1} \le (L-1)\, \# Y_k \le (L-1)^{k+1-k_0} 2ML^{k_0}$.

We now consider $\varepsilon \le \frac{1}{2} L^{-k_0}$ and write k the unique integer such that

$$L^{-k} \le 2\varepsilon < L^{-k+1} \quad \text{i.e.,} \quad k = \left\lceil \frac{-\log(2\varepsilon)}{\log L} \right\rceil.$$

Since we can cover Ω by $2ML^{k_0}(L-1)^{k-k_0}$ closed intervals of size L^{-k}, we can also cover Ω by $4ML^{k_0}(L-1)^{k-k_0}$ open intervals of size 2ε. Hence,

$$N_\Omega(\varepsilon) \le 4ML^{k_0}(L-1)^{k-k_0} \le 4M\left(\frac{L}{L-1}\right)^{k_0} (L-1)^{-\log(2\varepsilon)/\log L + 1} \le C\varepsilon^{-\delta},$$

with $\delta = \log(L-1)/\log L \in [0, 1[$ and $C = 4M(L/(L-1))^{k_0}(L-1)^{1-\log 2/\log L}$. □

Acknowledgment

The author would like to warmly thank Stéphane Nonnenmacher for his careful reading and helpful discussions which contributed a lot to the completion of this work. He also thanks the anonymous referee for helpful suggestions and comments.

References

[Anantharaman and Nonnenmacher 2007a] N. Anantharaman and S. Nonnenmacher, "Entropy of semiclassical measures of the Walsh-quantized baker's map", *Ann. Henri Poincaré* **8**:1 (2007), 37–74. MR Zbl

[Anantharaman and Nonnenmacher 2007b] N. Anantharaman and S. Nonnenmacher, "Half-delocalization of eigenfunctions for the Laplacian on an Anosov manifold", *Ann. Inst. Fourier* (*Grenoble*) **57**:7 (2007), 2465–2523. MR Zbl

[Bardos et al. 1987] C. Bardos, G. Lebeau, and J. Rauch, "Scattering frequencies and Gevrey 3 singularities", *Invent. Math.* **90**:1 (1987), 77–114. MR Zbl

[Barkhofen et al. 2013] S. Barkhofen, T. Weich, A. Potzuweit, H.-J. Stöckmann, U. Kuhl, and M. Zworski, "Experimental observation of the spectral gap in microwave n-disk systems", *Phys. Rev. Lett.* **110**:16 (2013), art. id. 164102.

[Barreira 2008] L. Barreira, *Dimension and recurrence in hyperbolic dynamics*, Progr. Math. **272**, Birkhäuser, Basel, 2008. MR Zbl

[Borthwick 2007] D. Borthwick, *Spectral theory of infinite-area hyperbolic surfaces*, Progr. Math. **256**, Birkhäuser, Boston, 2007. MR Zbl

[Bourgain and Dyatlov 2017] J. Bourgain and S. Dyatlov, "Fourier dimension and spectral gaps for hyperbolic surfaces", *Geom. Funct. Anal.* **27**:4 (2017), 744–771. MR Zbl

[Bourgain and Dyatlov 2018] J. Bourgain and S. Dyatlov, "Spectral gaps without the pressure condition", *Ann. of Math.* (2) **187**:3 (2018), 825–867. MR Zbl

[Bowen and Ruelle 1975] R. Bowen and D. Ruelle, "The ergodic theory of Axiom A flows", *Invent. Math.* **29**:3 (1975), 181–202. MR Zbl

[Chernov and Markarian 2006] N. Chernov and R. Markarian, *Chaotic billiards*, Math. Surv. Monogr. **127**, Amer. Math. Soc., Providence, RI, 2006. MR Zbl

[Datchev and Vasy 2012] K. Datchev and A. Vasy, "Gluing semiclassical resolvent estimates via propagation of singularities", *Int. Math. Res. Not.* **2012**:23 (2012), 5409–5443. MR Zbl

[Dyatlov 2018] S. Dyatlov, "Notes on hyperbolic dynamics", preprint, 2018. arXiv 1805.11660

[Dyatlov 2019] S. Dyatlov, "An introduction to fractal uncertainty principle", *J. Math. Phys.* **60**:8 (2019), art. id. 081505. MR Zbl

[Dyatlov and Jin 2017] S. Dyatlov and L. Jin, "Resonances for open quantum maps and a fractal uncertainty principle", *Comm. Math. Phys.* **354**:1 (2017), 269–316. MR Zbl

[Dyatlov and Jin 2018] S. Dyatlov and L. Jin, "Semiclassical measures on hyperbolic surfaces have full support", *Acta Math.* **220**:2 (2018), 297–339. MR Zbl

[Dyatlov and Zahl 2016] S. Dyatlov and J. Zahl, "Spectral gaps, additive energy, and a fractal uncertainty principle", *Geom. Funct. Anal.* **26**:4 (2016), 1011–1094. MR Zbl

[Dyatlov and Zworski 2019] S. Dyatlov and M. Zworski, *Mathematical theory of scattering resonances*, Grad. Stud. Math. **200**, Amer. Math. Soc., Providence, RI, 2019. MR Zbl

[Dyatlov and Zworski 2020] S. Dyatlov and M. Zworski, "Fractal uncertainty for transfer operators", *Int. Math. Res. Not.* **2020**:3 (2020), 781–812. MR Zbl

[Dyatlov et al. 2022] S. Dyatlov, L. Jin, and S. Nonnenmacher, "Control of eigenfunctions on surfaces of variable curvature", *J. Amer. Math. Soc.* **35**:2 (2022), 361–465. MR Zbl

[Gaspard and Rice 1989] P. Gaspard and S. A. Rice, "Scattering from a classically chaotic repellor", *J. Chem. Phys.* **90**:4 (1989), 2225–2241. MR

[Gérard 1988] C. Gérard, *Asymptotique des pôles de la matrice de scattering pour deux obstacles strictement convexes*, Mém. Soc. Math. France (N.S.) **31**, Soc. Math. France, Paris, 1988. MR Zbl

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

[Hargé and Lebeau 1994] T. Hargé and G. Lebeau, "Diffraction par un convexe", *Invent. Math.* **118**:1 (1994), 161–196. MR Zbl

[Hirsch and Pugh 1969] M. W. Hirsch and C. C. Pugh, "Stable manifolds for hyperbolic sets", *Bull. Amer. Math. Soc.* **75** (1969), 149–152. MR Zbl

[Hurder and Katok 1990] S. Hurder and A. Katok, "Differentiability, rigidity and Godbillon–Vey classes for Anosov flows", *Inst. Hautes Études Sci. Publ. Math.* **72** (1990), 5–61. MR Zbl

[Ikawa 1982] M. Ikawa, "Decay of solutions of the wave equation in the exterior of two convex obstacles", *Osaka Math. J.* **19**:3 (1982), 459–509. MR Zbl

[Ikawa 1988] M. Ikawa, "Decay of solutions of the wave equation in the exterior of several convex bodies", *Ann. Inst. Fourier* (*Grenoble*) **38**:2 (1988), 113–146. MR Zbl

[Katok and Hasselblatt 1995] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Encycl. Math. Appl. **54**, Cambridge Univ. Press, 1995. MR Zbl

[Mather 1968] J. N. Mather, "Characterization of Anosov diffeomorphisms", *Indag. Math.* **30** (1968), 479–483. MR Zbl

[McShane 1934] E. J. McShane, "Extension of range of functions", *Bull. Amer. Math. Soc.* **40**:12 (1934), 837–842. MR Zbl

[Morita 1991] T. Morita, "The symbolic representation of billiards without boundary condition", *Trans. Amer. Math. Soc.* **325**:2 (1991), 819–828. MR Zbl

[Naud 2005] F. Naud, "Expanding maps on Cantor sets and analytic continuation of zeta functions", *Ann. Sci. École Norm. Sup.* (4) **38**:1 (2005), 116–153. MR Zbl

[Nonnenmacher 2011] S. Nonnenmacher, "Spectral problems in open quantum chaos", *Nonlinearity* **24**:12 (2011), R123–R167. Zbl

[Nonnenmacher and Zworski 2009] S. Nonnenmacher and M. Zworski, "Quantum decay rates in chaotic scattering", *Acta Math.* **203**:2 (2009), 149–233. MR Zbl

[Nonnenmacher et al. 2011] S. Nonnenmacher, J. Sjöstrand, and M. Zworski, "From open quantum systems to open quantum maps", *Comm. Math. Phys.* **304**:1 (2011), 1–48. MR Zbl

[Nonnenmacher et al. 2014] S. Nonnenmacher, J. Sjöstrand, and M. Zworski, "Fractal Weyl law for open quantum chaotic maps", *Ann. of Math.* (2) **179**:1 (2014), 179–251. MR Zbl

[Petkov and Stoyanov 2010] V. Petkov and L. Stoyanov, "Analytic continuation of the resolvent of the Laplacian and the dynamical zeta function", *Anal. PDE* **3**:4 (2010), 427–489. MR Zbl

[Rivière 2010] G. Rivière, "Entropy of semiclassical measures in dimension 2", *Duke Math. J.* **155**:2 (2010), 271–336. MR Zbl

[Schwartz 2021] N. Schwartz, "The full delocalization of eigenstates for the quantized cat map", preprint, 2021. arXiv 2103.06633

[Sjöstrand and Zworski 1999] J. Sjöstrand and M. Zworski, "Asymptotic distribution of resonances for convex obstacles", *Acta Math.* **183**:2 (1999), 191–253. MR Zbl

[Zworski 2012] M. Zworski, *Semiclassical analysis*, Grad. Stud. Math. **138**, Amer. Math. Soc., Providence, RI, 2012. MR Zbl

[Zworski 2017] M. Zworski, "Mathematical study of scattering resonances", *Bull. Math. Sci.* **7**:1 (2017), 1–85. MR Zbl

Received 13 Jan 2022. Revised 10 Jun 2022. Accepted 11 Aug 2022.

LUCAS VACOSSIN: lucas.vacossin@universite-paris-saclay.fr
Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay, UMR 8628 du CNRS, Orsay, France

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at msp.org/apde.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in APDE are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Analysis & PDE

Volume 17 No. 3 2024

Schauder estimates for equations with cone metrics, II 757
Bin Guo and Jian Song

The stability of simple plane-symmetric shock formation for three-dimensional compressible Euler flow with vorticity and entropy 831
Jonathan Luk and Jared Speck

Families of functionals representing Sobolev norms 943
Haïm Brezis, Andreas Seeger, Jean Van Schaftingen and Po-Lam Yung

Schwarz–Pick lemma for harmonic maps which are conformal at a point 981
Franc Forstnerič and David Kalaj

An improved regularity criterion and absence of splash-like singularities for g-SQG patches 1005
Junekey Jeon and Andrej Zlatoš

Spectral gap for obstacle scattering in dimension 2 1019
Lucas Vacossin