ANALYSIS & PDE

Volume 17 No. 4 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. 4, pp. 1127–1173
DOI: 10.2140/apde.2024.17.1127

msp

THE SINGULAR STRATA OF A FREE-BOUNDARY PROBLEM FOR HARMONIC MEASURE

SEAN MCCURDY

We obtain *quantitative* estimates on the fine structure of the singular set of the mutual boundary $\partial\Omega^{\pm}$ for pairs of complementary domains $\Omega^{+}, \Omega^{-} \subset \mathbb{R}^n$ which arise in a class of two-sided free boundary problems for harmonic measure. These estimates give new insight into the structure of the mutual boundary $\partial\Omega^{\pm}$.

1. Introduction

The focus of this paper is the study of a class of two-phase free boundary problems for harmonic measure. For $n \geq 3$, let $\Omega^{+} \subset \mathbb{R}^n$ and $\Omega^{-} = \overline{\Omega}^{+c}$ be unbounded nontangentially accessible (NTA) domains (see Definition 2.1), let $\omega^{\pm}$ be their associated harmonic measures, and let $u^{\pm}$ be the associated Green's functions with poles at infinity. Let $\omega^{-} \ll \omega^{+} \ll \omega^{-}$, and let $h = d\omega^{-}/d\omega^{+}$ satisfy $\ln(h) \in C^{0,\alpha}$ for some $0 < \alpha < 1$. We obtain new results on the structure of the geometric singular set of the boundary $\partial\Omega^{\pm}$.

This problem was introduced without the regularity assumption on $\omega^{\pm}$ by Kenig, Preiss, and Toro [Kenig et al. 2009], with other work under the assumption that $\ln(h) \in \mathrm{VMO}(\partial\Omega^{\pm})$ by Kenig and Toro [2006], Badger [2011; 2013], and Badger, Engelstein, and Toro [Badger et al. 2017]. Questions about the structure of the free boundary and the singular set when $\ln(h) \in C^{0,\alpha}$ for $0 < \alpha < 1$ have been addressed by Engelstein [2016] and Badger, Engelstein, Toro [Badger et al. 2020], respectively. Engelstein [2016]

MSC2010: 31B05, 31B25, 35R35.
Keywords: free-boundary, harmonic measure, partial regularity, singular strata.

shows that under the additional assumption that the boundary is sufficiently flat in the sense of Reifenberg, the boundary is locally $C^{1,\alpha}$. In [Badger et al. 2020], the authors remove the assumption of flatness and prove that the geometric singular set is contained in countably many $C^{1,\beta}$ submanifolds of the appropriate dimension. See [Kenig et al. 2009] for an overview of this problem in lower dimensions and [Badger et al. 2017; 2020] for further background.

Until recently, almost all work on the two-sided free boundary problem for harmonic measure in higher dimensions has operated under the assumption that $\Omega^{\pm}$ are NTA domains because the NTA conditions allow for scale-invariant estimates of harmonic measure. However, Azzam, Mourgoglou, Tolsa, and Volberg [Azzam et al. 2019] proved, among other things, that if we relax the assumption that the domains are NTA, then $\omega^- \ll \omega^+ \ll \omega^-$ on $G \subset \partial\Omega^{\pm}$ implies that G can be decomposed into $G = R \cup B$, where R is $(n-1)$-rectifiable and $\omega^{\pm}(B) = 0$. However, we shall work under the assumption that $\Omega^{\pm}$ are NTA domains.

Based upon [Badger et al. 2020], we know that when $\ln(h) \in C^{0,\alpha}$, the singular set of $\partial\Omega^{\pm}$ is countably $C^{1,\beta}$-rectifiable where β depends on but is not equal to α. This leaves open the question of whether or not the singular set is dense, or more generally how it sits in space. In this paper, we answer the question of how the singular set "sits in space". In particular, we provide upper Minkowski content bounds upon the quantitative strata of the singular set (see Theorem 2.15). The main approach will be to follow [Engelstein 2016] and consider jump functions $v = u^+ - u^-$ which are *almost* harmonic, and employ the Almgren frequency function and geometric techniques as in [Cheeger et al. 2015; Han and Lin 1994] in conjunction with the powerful quantitative differentiation techniques of [De Lellis et al. 2018; Naber and Valtorta 2017]. While these tools are common for problems in calculus of variations, it is important to note that the jump functions v are not minimizers of any energy, nor do they satisfy any global PDE.

2. Definitions and statement of main results

2A. *Domains and their Green's functions.* Nontangentially accessible (NTA) domains were formally introduced by Jerison and Kenig [1982] to study the boundary behavior of PDEs on nonsmooth domains.

Definition 2.1. A domain $\Omega \subset \mathbb{R}^n$ is a nontangentially accessible (NTA) domain if there exist constants $M > 1$ and $R_0 > 0$ such that the following holds:

(1) Ω satisfies the *corkscrew condition*. That is, for any $Q \in \partial\Omega$ and $0 < r < R_0$, there exists a point $A_r(Q) \in \Omega$ with the following two properties:

$$|A_r(Q) - Q| < r \quad \text{and} \quad B_{r/M}(A_r(Q)) \subset \Omega.$$

(2) $\overline{\Omega}^c$ also satisfies the *corkscrew condition*.

(3) Ω satisfies the *Harnack chain condition*. That is, for any $\epsilon > 0$ and $Q \in \partial\Omega$, if

$$x_1, x_2 \in \Omega \cap B_{R_0/4}(Q) \setminus B_\epsilon(\partial\Omega) \quad \text{and} \quad |x_1 - x_2| \leq 2^k\epsilon,$$

then there exists a "Harnack chain" of balls $\{B_{r_i}(y_i)\}_{i=1}^N$ satisfying:

(a) $x_1 \in B_{r_i}(y_1)$ and $x_1 \in B_{r_N}(y_N)$.

(b) $B_{r_i}(y_i) \subset \Omega$ for all $i = 1, \dots, N$.
(c) $B_{r_i}(y_i) \cap B_{r_{i+1}}(y_{i+1}) \neq \varnothing$ for all $i = 1, \dots, N-1$.
(d) $N \leq Mk$.
(e) For all $i = 1, \dots, N$,

$$\frac{1}{2M} \min_{i=1,2,\dots,N} \{\mathrm{dist}(x_i, \partial\Omega)\} \leq r_i \leq \mathrm{dist}(y_i, \partial\Omega).$$

Note that by increasing the radii if necessary, we may assume that $r_i \sim_M \mathrm{dist}(y_i, \partial\Omega)$.

We say that Ω^+ is a *two-sided NTA domain* if both Ω^+ and $\Omega^- := \overline{\Omega}^c$ are NTA domains. We shall refer to the complementary pair $\Omega^\pm$ of domains as complementary two-sided NTA domains and denote their mutual boundary by $\partial\Omega^\pm$.

In this paper, we shall only deal with unbounded two-sided NTA domains. That is, we shall assume that $R_0 = \infty$. However, the results are essentially local.

Definition 2.2 (Green's functions). For $\Omega^\pm \subset \mathbb{R}^n$ a pair of complementary two-sided NTA domains, we shall use $u^\pm$ to denote the *Green's function with pole at infinity* corresponding to $\Omega^\pm$, respectively.

Recall that $u^\pm$ are unique up to scalar multiplication and that to each $u^\pm$ is associated the *harmonic measure* $\omega^\pm$, defined by the property that, for all $\phi \in C_c^\infty(\mathbb{R}^n)$,

$$\int \Delta\phi u^\pm \, dV = \int \phi \, d\omega^\pm.$$

See [Garnett and Marshall 2005] for more details about harmonic measures.

Observe that if ω^+ is the harmonic measure associated to u^+, then $c\omega^+$ is the harmonic measure associated to cu^+ for any $c > 0$.

If $f \in C^{0,\alpha}(\mathbb{R}^n)$, we shall use $\|f\|_\alpha$ to denote the local norm:

$$\|f\|_\alpha := \sup_{B_2(0)} |f| + \sup_{x \neq y \in B_2(0)} \frac{|f(x) - f(y)|}{|x-y|^\alpha}.$$

Definition 2.3. We define the class $\mathcal{D}(n, \alpha, M_0)$ to be the collection of domains $\Omega^\pm \subset \mathbb{R}^n$ such that $\Omega^\pm$ are complementary unbounded two-sided NTA domains for which $M < M_0$, $\omega^- \ll \omega^+ \ll \omega^-$, the Radon–Nikodym derivative $h = d\omega^-/d\omega^+$ satisfies $\ln(h) \in C^{0,\alpha}(\partial\Omega)$, and $0 \in \partial\Omega^\pm$.

Note that if $\Omega^\pm \in \mathcal{D}(n, \alpha, M_0)$ and $Q \in \partial\Omega^\pm$, then $\Omega^\pm - Q \in \mathcal{D}(n, \alpha, M_0)$.

2B. *A class of functions and their rescalings.*

Definition 2.4. Let $\Omega^\pm \subset \mathbb{R}^n$ be a pair of complementary two-sided NTA domains with mutual boundary $\partial\Omega^\pm$. For any $Q \in \partial\Omega^\pm$ and any Green's functions $u^\pm$ we define the *jump function*

$$v^Q(x) := h(Q)u^+(x) - u^-(x). \tag{2-1}$$

The scaling $h(Q)u^+$ normalizes the Radon–Nikodym derivative of the harmonic measure associated to $h(Q)u^+$ and u^- at $Q \in \partial\Omega^\pm$.

Definition 2.5. Let $\Omega^\pm \in \mathcal{D}(n, \alpha, M_0)$ and $Q \in \partial\Omega^\pm$. For scales $0 < r$, we define the rescaling of the function v^Q at scale r at the point $Q' \in \partial\Omega^\pm$ by

$$v^Q_{Q',r}(x) := v^Q(rx + Q')\frac{r^{n-2}}{\omega^-(B_r(Q'))}$$

and the corresponding rescaled measure as

$$\omega^\pm_{Q',r}(E) := \frac{\omega^\pm(rE + Q')}{\omega^\pm(B_r(Q'))}. \tag{2-2}$$

The rescalings $v^Q_{Q,r}$ were first introduced by Kenig and Toro [2006]. In this paper, we shall employ the following results by Kenig, Toro, Badger, and Engelstein.

Theorem 2.6 [Badger 2011; Engelstein 2016; Kenig and Toro 2006]. *For $v^Q_{Q,r}$, $\omega^\pm_{Q,r}$ as in Definition 2.5*:

(1) *Subsequential limits as $r \to 0$ of the functions $v^Q_{Q,r}$ converge to harmonic polynomials. Furthermore, the degree of these polynomials is bounded and depends only upon the NTA constant, M_0.* [Kenig and Toro 2006]

(2) *Subsequential limits as $r \to 0$ of the functions $v^Q_{Q,r}$ converge to* ***homogeneous*** *harmonic polynomials. Furthermore, the degree of homogeneity is unique along blow-ups.* [Badger 2011]

(3) *The rescalings $v^Q_{Q,r}$ are uniformly locally Lipschitz with Lipschitz constant that only depends upon M_0.* [Engelstein 2016]

(4) *The measures $\omega^\pm_{Q,r}$ are locally uniformly bounded.* [Engelstein 2016]

In addition to the $v^Q_{Q',r}$ rescalings, we shall also use a different kind of rescaling.

Definition 2.7 [Cheeger et al. 2015]. Let $f : B_1(0) \to \mathbb{R}$ be a function in $C(\mathbb{R}^n)$. We define the rescaled function $T_{x,r}f$ of f at a point $x \in B_{1-r}(0)$ at scale $0 < r < 1$ by

$$T_{x,r}f(y) := \frac{f(x + ry) - f(x)}{\left(⨍_{\partial B_1(0)}(f(x + rz) - f(x))^2 \, d\sigma(z)\right)^{1/2}}. \tag{2-3}$$

In the case that the denominator is zero, we define $T_{x,r}f = \infty$. We denote the limit as $r \to 0$ by

$$T_x f(y) := \lim_{r\to 0} T_{x,r}f(y).$$

Definition 2.8. Let $\mathcal{A}(n, \alpha, M_0)$ be the set of functions $v : \mathbb{R}^n \to \mathbb{R}$ such that

$$v := v^0 = h(0)u^+ - u^-,$$

where $u^\pm$ are the Green's functions with poles at infinity associated to a two-sided NTA domain $\Omega^\pm \in \mathcal{D}(n, \alpha, M_0)$ and $h = d\omega^-/d\omega^+$, where $\omega^\pm$ are the harmonic measures associated to $u^\pm$.

Remark 2.9. For any fixed domain $\Omega^\pm \in \mathcal{D}(n, \alpha, M_0)$ there is a one-parameter family of associated functions $v \in \mathcal{A}(n, \alpha, M_0)$ with $\{v = 0\} = \partial\Omega^\pm$. Indeed, $cv^0_{0,1} \in \mathcal{A}(n, \alpha, M_0)$ for all $c > 0$. To avoid degeneracy because of this degree of freedom within the family $\mathcal{A}(n, \alpha, M_0)$, we shall make extensive use of the normalizations in Definitions 2.5 and 2.7 in the arguments to come.

Finally, note that in general the functions $v^Q_{Q',r}$ will not belong to $\mathcal{A}(n, \alpha, M_0)$ if $Q \neq 0$ and/or $Q' \neq Q$.

2C. *Quantitative symmetry.* The geometry we wish to capture with the blow-ups $T_x f$ is encoded in their translational symmetries.

Definition 2.10. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function. We say f is 0-*symmetric* if

$$f(x) := cP^+(x) - P^-(x) \tag{2-4}$$

for some $c > 0$, where $P^\pm$ are the positive and negative parts of a homogeneous harmonic polynomial P. We will say that f is k-*symmetric* if f is 0-symmetric and there exists a k-dimensional subspace V such that $f(x+y) = f(x)$ for all $x \in \mathbb{R}^n$ and all $y \in V$.

The constant $c > 0$ is there to allow for the function to "hinge" along its zero set. We must allow this kind of "hinging" to accommodate for the "nonalignment" issue in the blow-ups at $Q \in \partial\Omega^\pm \setminus \{0\}$. See Remark 3.2.

We now define a quantitative version of symmetry.

Definition 2.11. For any $f \in C(\mathbb{R}^n)$, f will be called (k, ϵ, r, p)-*symmetric* if there exists a k-symmetric function P such that

(1) $⨍_{\partial B_1(0)} |P|^2 \, dV = 1$,

(2) $⨍_{B_1(0)} |T_{p,r} f - P|^2 \, dV < \epsilon$.

Sometimes, we shall refer to a function f as being (k, ϵ)-symmetric in the ball $B_r(p)$ to mean f is (k, ϵ, r, p)-symmetric.

This quantitative control allows us to define a quantitative stratification following [Cheeger and Naber 2013].

Definition 2.12 (quantitative singular strata). Let $v \in \mathcal{A}(n, \alpha, M_0)$ and $0 < r \le 1$. We denote the (k, ϵ, r)-*singular stratum* of v by $\mathcal{S}^k_{\epsilon,r}(v)$, and we define it by

$$\mathcal{S}^k_{\epsilon,r}(v) := \{x \in \partial\Omega^\pm : v \text{ is not } (k+1, \epsilon, s, x)\text{-symmetric for all } r \le s \le 1\}. \tag{2-5}$$

We shall also use the notation $\mathcal{S}^k_\epsilon(v)$ for $\mathcal{S}^k_{\epsilon,0}(v)$. It is immediate from the definitions that $\mathcal{S}^k_{\epsilon,r}(v) \subset \mathcal{S}^{k'}_{\epsilon',r'}(v)$ if $k \le k'$, $\epsilon' \le \epsilon$, $r \le r'$.

We can recover the *qualitative* stratification

$$\mathcal{S}^k(v) := \{x \in \partial\Omega^\pm : T_x v \text{ is not } (k{+}1)\text{-symmetric}\} = \bigcup_\eta \bigcap_r \mathcal{S}^k_{\eta,r}(v).$$

The set $\mathcal{S}^k(v)$ is the k-th stratum of $\mathcal{S}^{n-2}(v) = \operatorname{sing}(\partial\Omega^\pm)$. Furthermore, if $x \in \mathcal{S}^k(v)$, then there exists an $\epsilon > 0$ such that $x \in \mathcal{S}^k_\epsilon(v)$.

Remark 2.13. Note that the singular set and its strata are all stable under the operations

$$\mathcal{S}^k(v) = \mathcal{S}^k(cv) \quad \text{and} \quad \mathcal{S}^k(v) = \mathcal{S}^k(cv^+ - v^-)$$

for all $c \neq 0$. The former is a trivial consequence of the fact that $T_{p,r} f = T_{p,r}(cf)$. The latter follows from Definition 2.10 and Theorem 2.6. In particular, for all $v \in \mathcal{A}(n, \alpha, M_0)$, we have $\mathcal{S}^k(v) = \mathcal{S}^k(v^Q)$ for all $Q \in \partial\Omega^\pm$.

Previous results on the singular set are summed up in the following theorem.

Theorem 2.14 [Badger et al. 2017; 2020; Engelstein 2016]. *For $v \in \mathcal{A}(n, \alpha, M_0)$, the following hold*:

(1) $(\mathcal{S}^{n-2}(v) \setminus \mathcal{S}^{n-3}(v)) \cap \partial\Omega^{\pm} = \varnothing$. [Badger et al. 2020, Remark 7.2]

(2) *There exists an* $\epsilon > 0$ *such that* $\operatorname{sing}(\partial\Omega^{\pm}) = \mathcal{S}^{n-3}(v) \cap \partial\Omega^{\pm} \subset \mathcal{S}^{n-2}_{\epsilon}(v)$. [Engelstein 2016, Theorem 1.1]

(3) $\overline{\dim}_{\mathcal{M}}(\operatorname{sing}(\partial\Omega^{\pm})) \leq n-3$. [Badger et al. 2017, Theorem 7.5]

2D. *Main results and outline of the proof.* In this paper, we prove volume bounds on tubular neighborhoods around the $S^k_{\epsilon,r}(v)$. We are able to show the following estimates.

Theorem 2.15. *Let $v \in \mathcal{A}(n, \alpha, M_0)$ with $\|\ln(h)\|_{\alpha} \leq \Gamma$. For every $0 < \epsilon$ and $0 \leq k \leq n-2$ there is an $r_0(n, \alpha, M_0, \Gamma, \epsilon) > 0$ such that, for all $0 < r < r_0$ and any $r \leq R \leq 1$,*

$$\operatorname{Vol}(B_R(B_{1/4}(0) \cap \mathcal{S}^k_{\epsilon,r}(v))) \leq C(n, \alpha, M_0, \Gamma, \epsilon) R^{n-k}. \tag{2-6}$$

We have the following immediate corollary.

Corollary 2.16. *Let $v \in \mathcal{A}(n, \alpha, M_0)$ and $0 \leq k \leq n-2$. For every $0 < \epsilon$,*

$$\overline{\dim}_{\mathcal{M}}(\mathcal{S}^k_{\epsilon}(v)) \leq k, \tag{2-7}$$

and there exists a constant such that

$$\mathcal{M}^{*,k}(\mathcal{S}^k_{\epsilon}(v) \cap B_{1/4}(0)) \leq C(n, \alpha, M_0, \Gamma, \epsilon). \tag{2-8}$$

Thanks to an ϵ-regularity result due to [Engelstein 2016] we are able to strengthen the conclusion of Theorem 2.15 when we consider the full singular set.

Corollary 2.17. *Let $v \in \mathcal{A}(n, \alpha, M_0)$ with $\|\ln(h)\|_{\alpha} \leq \Gamma$. Recall that*

$$\operatorname{sing}(\partial\Omega^{\pm}) = \mathcal{S}^{n-3} \cap \partial\Omega^{\pm}.$$

There exists $0 < \epsilon = \epsilon(M_0, \Gamma, \alpha)$ such that $\operatorname{sing}(\partial\Omega^{\pm}) \subset \mathcal{S}^{n-3}_{\epsilon}$; see Lemma 12.1. Thus, there is a constant $C = C(n, \alpha, M_0, \Gamma) < \infty$ such that

$$\mathcal{M}^{*,n-3}(\operatorname{sing}(\partial\Omega^{\pm}) \cap B_{1/4}(0)) \leq C(n, \alpha, M_0, \Gamma). \tag{2-9}$$

Proof. This follows immediately from Lemma 12.1 and Theorem 2.15. □

2E. *Outline of the proof of Theorem 2.15.* In order to prove a theorem of this kind, we must build a cover of $\mathcal{S}^k_{\epsilon,r}(v)$, and we must count how many balls we use. Therefore two things are critical: getting geometric information about $\mathcal{S}^k_{\epsilon,r}(v)$ and keeping track of how the balls pack.

The overall strategy of proof is similar to that of [De Lellis et al. 2018; Edelen and Engelstein 2019]. However, there are several major differences. First, the functions $v \in \mathcal{A}(n, \alpha, M_0)$ considered here are not harmonic functions or minimizers of an energy. Sections 3–5 are devoted to showing that the relevant analogs of harmonic results (e.g., compactness, almost monotonicity of the Almgren frequency, local uniform boundedness of the Almgren frequency, quantitative rigidity for the Almgren frequency, cone-splitting, etc.) hold for $v \in \mathcal{A}(n, \alpha, M_0)$. In particular, we prove an estimate on the nondegeneracy of the almost monotonicity for Almgren frequency in Lemma 4.9. Local geometric control on $\mathcal{S}^k_\epsilon(v)$ is obtained in Section 6.

However, geometric control is not enough to obtain Theorem 2.15. To obtain finite upper Minkowski content bounds we need the discrete Reifenberg theorem from [Naber and Valtorta 2017]; see Theorem 9.1. This requires that we prove a "frequency pinching" result (Lemma 8.2) in which we connect the drop in the Almgren frequency over small scales with the β-numbers. The main challenge is to connect the lower bound on the derivative of the Almgren frequency (Lemma 4.9) and employ the techniques of [De Lellis et al. 2018] to obtain the necessary estimates on $N(Q, r, v) - N(Q', r, v)$; see Section 7.

In Section 9, we obtain the necessary packing estimates, following the framework of [Naber and Valtorta 2017] to accommodate the estimates of Section 8. Sections 10 and 11 construct the covering which proves the theorem according to the program laid out by [Naber and Valtorta 2017]. These are included for completeness.

3. Compactness

The main goal of this section is to show that $\mathcal{A}(n, \alpha, M_0)$ enjoys sufficient compactness to allow for limit-compactness arguments. Namely, we wish to establish that, for any sequence $v_i \in \mathcal{A}(n, \alpha, M_0)$, we can extract a subsequence which converges to a function v_∞ and that $N(p, r, v_i) \to N(p, r, v_\infty)$; see Corollary 4.3. This requires strong convergence in $W^{1,2}_{\mathrm{loc}}(\mathbb{R}^n)$; see Lemmas 3.10 and 3.6.

On a technical level, we must extend the compactness implied by Theorem 2.6 for $v^Q_{Q,r}$ to $v^Q_{Q',r}$ and $T_{Q',r}v$. Throughout, we shall make essential use of "standard NTA results" such as the doubling of harmonic measure and various comparability results, all of which may be found in [Jerison and Kenig 1982].

Remark 3.1. Recall that for $E \subset \partial\Omega^\pm$

$$\omega^+(E) = \int \chi_E \, d\omega^+ \quad \text{and} \quad \omega^-(E) = \int \chi_E h \, d\omega^+.$$

Furthermore, if $\ln(h) \in C^{0,\alpha}$ with $\|\ln(h)\|_\alpha \le \Gamma$, then for all $Q, Q' \in \partial\Omega^\pm$

$$e^{-\Gamma|Q-Q'|^\alpha} h(Q') \le h(Q) \le e^{\Gamma|Q-Q'|^\alpha} h(Q'). \tag{3-1}$$

Using (3-1) in the above integral equations implies that in any compact set K, if $v \in \mathcal{A}(n, \alpha, M_0)$ with $\|\ln(h)\|_\alpha \le \Gamma$, there is a constant $C(K, \Gamma, \alpha) > 1$ such that for any $E \subset K \cap \partial\Omega^\pm$

$$C^{-1} \le \frac{\omega^-(E)}{\omega^+(E)} \le C.$$

Remark 3.2. By Theorem 2.6, we know that subsequential limits as $r \to 0$ of the functions $v^Q_{Q,r}$ converge to homogeneous harmonic polynomials. However, for $Q, Q' \in \partial\Omega^\pm$ and $Q \neq Q'$, it is not true in general that $v^Q_{Q',r}$ converges to a homogeneous harmonic polynomial. As $r \to 0$, the function $v^Q_{Q',r}$ will converge to a 0-symmetric function (see Definition 2.10) where $c = h(Q)$.

Definition 3.3. We shall abuse the notation $T_{Q,r}$ from Definition 2.7 to denote translated and scaled versions of various objects. For example, for sets this is the usual push-forward

$$T_{Q,r}\Omega^{\pm} := \frac{\Omega^{\pm} - Q}{r}, \quad T_{Q,r}\partial\Omega^{\pm} := \frac{\partial\Omega^{\pm} - Q}{r}.$$

However, for the measures $\omega^{\pm}$, we will denote by $T_{Q,r}\omega^{\pm}$ the harmonic measures associated to the positive and negative parts of $T_{Q,r}v$. The corkscrew points $A^{\pm}_R(Q)$ will always denote the corkscrew point associated to Q at scale R in the appropriate domain $\Omega^{\pm}$. We shall use $T_{Q,r}A^{\pm}_{r'}(Q')$ to denote the corkscrew point associated to $T_{Q,r}Q' = (Q' - Q)/r \in T_{Q,r}\partial\Omega^{\pm}$ at the scale r'/r.

Lemma 3.4 (local Lipschitz bounds). *Let $v \in \mathcal{A}(n, \alpha, M_0)$ with $\|\ln(h)\|_\alpha \le \Gamma$. For all $Q \in \partial\Omega^{\pm} \cap B_2(0)$ and all radii $0 < r \le 2$, the function $T_{Q,r}v$ is locally Lipschitz with uniform constants depending only upon M_0, Γ, α.*

Proof. Recall that by Definition 2.5,

$$v_{Q,r} = v^0_{Q,r} = \frac{r^{n-2}}{\omega^-(B_r(Q))} v(rx + Q).$$

By NTA estimates, for all $0 < r$, we have $|v(A^-_r(Q))| \sim \omega^-(B_r(Q))/r^{n-2}$ by constants which only depend upon M_0. Thus, $v_{Q,r}(T_{Q,r}A^-_r(Q))$ is bounded above and below by constants which only depend upon M_0. By constructing Harnack chains from $T_{Q,r}A^-_r(Q)$ to $T_{Q,r}A^-_{M_0 r}(Q)$ we can find a point $y \in \partial B_1(0)$ such that $y \in B_{r_i}(y_i) \subset T_{Q,r}\Omega^-$ and $\operatorname{dist}(y_i, T_{Q,r}\partial\Omega^{\pm}) \ge (2M_0^2)^{-1}$. Applying Harnack's inequality to the function $-v$ in a chain of balls which connect $T_{Q,r}A^-_r(Q)$ and y in Ω^-, we have $|v_{Q,r}(y)| \sim_{M_0} |v_{Q,r}(T_{Q,r}A^-_r(Q))|$. That is, $|v_{Q,r}(y)|$ is bounded above and below by constants that only depend upon M_0. Thus, by the uniform Lipschitz property of $v_{Q,r}$ guaranteed by Theorem 2.6, we can find a ball of radius $0 < c$ such that $|v_{Q,r}| \ge c(M_0)$ on $\partial B_1(0) \cap B_c(y)$. Thus, $H(0, 1, v_{Q,r}) \ge c(M_0)$. Now, recalling Definition 2.7 and the fact that $T_{0,1}v = T_{0,1}(cv)$ for any constant $c > 0$, we have $T_{Q,r}v = T_{0,1}v_{Q,r}$. Since we assumed $\|\ln(h)\|_\alpha \le \Gamma$, $Q \in B_2(0)$, and $0 < r \le 2$, the $v_{Q,r}$ are locally uniformly Lipschitz by Theorem 2.6. Thus $H(0, 1, v_{Q,r}) \ge c(M_0)$ implies $T_{0,1}v_{Q,r} = T_{Q,r}v$ is also locally uniformly Lipschitz. □

Lemma 3.5 (local nondegeneracy). *Let $Q \in \partial\Omega^{\pm}$ and $0 < r < \infty$. Let $v \in \mathcal{A}(n, \alpha, M_0)$ be such that $\|\ln(h)\|_\alpha \le \Gamma$. The rescaling $T_{Q,r}v$ satisfies the following minimum growth conditions. For all $0 < \epsilon$, there is a constant $C = C(M_0, \alpha, \Gamma, \epsilon, R)$ such that, if $p \in B_R(0)$ with $\operatorname{dist}(p, \{T_{Q,r}\partial\Omega^{\pm}\} \cap B_R(0)) > \epsilon$,*

$$|T_{Q,r}v(p)| > C.$$

Proof. As in Lemma 3.4, $T_{Q,r}v(T_{Q,r}A^-_r(Q))$ is bounded above and below by constants that only depend upon the NTA constant M_0, Γ, and R. Thus, by Harnack chains between $T_{Q,r}A^-_r(Q)$ and $p \in T_{Q,r}\Omega^- \cap B_R(0)$ such that $\operatorname{dist}(p, T_{Q,r}\partial\Omega^{\pm} \cap B_R(0)) > \epsilon$, Harnack's inequality applied to $-T_{Q,r}v$ implies that $|T_{Q,r}v(p)| > C$. Note that C only depends upon R, M_0, and ϵ.

To get the same inequality for $p \in T_{Q,r}\Omega^+ \cap B_R(0)$, we recall that standard NTA results compare $T_{Q,r}v(T_{Q,r}A^+_r(Q))$ to $T_{Q,r}\omega^+(B_1(0))$. By Remark 3.1, $T_{Q,r}\omega^+(B_1(0)) \sim T_{Q,r}\omega^-(B_1(0))$ by constants which only depend upon R, Γ, α, and the NTA constants in the definition of the class $\mathcal{A}(n, \alpha, M_0)$. Applying the same Harnack chain and Harnack inequality argument as above gives the lemma. □

Lemma 3.6 (compactness). *Let $\{v_i\}$ be a sequence of functions in $\mathcal{A}(n, \alpha, M_0)$ such that $\|\ln(h)\|_\alpha \leq \Gamma$. Let $\{Q_i\} \subset \partial\Omega_i^\pm \cap B_1(0)$ and $0 < r_i \leq 1$. There is a subsequence $\{v_j\}$ and a Lipschitz function $v_\infty \in W^{1,2}_{\text{loc}}$ such that $T_{Q_j,r_j} v_j \to v_\infty$ in the following senses*:

(1) $T_{Q_j,r_j} v_j \to v_\infty$ *in* $C_{\text{loc}}(\mathbb{R}^n)$.

(2) $T_{Q_j,r_j} v_j \to v_\infty$ *in* $L^2_{\text{loc}}(\mathbb{R}^n)$.

(3) $\nabla T_{Q_j,r_j} v_j \rightharpoonup \nabla v_\infty$ *in* $L^2_{\text{loc}}(\mathbb{R}^n; \mathbb{R}^n)$.

Proof. To see (1), we recall Lemma 3.4 and the fact that $T_{Q_i,r_i} v_i(0) = 0$. By the Arzelà–Ascoli theorem there exists a subsequence such that $T_{Q_j,r_j} v_j \to v_\infty$ in $C_{\text{loc}}(\mathbb{R}^n)$. This implies convergence in $L^2_{\text{loc}}(\mathbb{R}^n)$. Being uniformly locally Lipschitz and uniformly bounded also implies that the functions $\{T_{Q_j,r_j} v_i\}$ are bounded in $W^{1,2}_{\text{loc}}(\mathbb{R}^n)$. By Rellich compactness, there exists a further subsequence such that $\nabla T_{Q_j,r_j} v_j \rightharpoonup \nabla v_\infty$ in $L^2_{\text{loc}}(\mathbb{R}^n; \mathbb{R}^n)$. □

Before we can prove the strong convergence $T_{Q_j,r_j} v_j \to v_\infty$ in $W^{1,2}_{\text{loc}}(\mathbb{R}^n)$, we need to control the upper Minkowski dimension of $\{v_\infty = 0\}$.

Lemma 3.7. *Under the assumptions of Lemma 3.6, if $T_{Q_i,r_i} v_i \to v_\infty$ in $C_{\text{loc}}(\mathbb{R}^n)$, then $T_{Q_i,r_i} \partial\Omega^\pm \to \{v_\infty = 0\}$ locally in the Hausdorff metric on compact subsets.*

Proof. We argue by contradiction. Suppose that there exists an $\epsilon > 0$, a radius $0 < R$, and a sequence of functions $T_{Q_i,r_i} v_i$ for which we can find a sequence of points $x_i \in B_R(0) \cap \{T_{Q_i,r_i} v_i = 0\}$ such that $\text{dist}(x_i, \{v_\infty = 0\}) > \epsilon$. Taking a subsequence which converges in $C_{\text{loc}}(\mathbb{R}^n)$, we may assume that $x_i \to x_\infty \in \overline{B_R(0)} \setminus B_\epsilon(\{v_\infty = 0\})$. Now, convergence in $C_{\text{loc}}(\mathbb{R}^n)$ implies that $T_{Q_i,r_i} v_i(x_\infty) \to v_\infty(x_\infty)$. Furthermore, since the $T_{Q_i,r_i} v_i$ are uniformly locally Lipschitz, $x_i \to x_\infty$, and $x_i \in \{T_{Q_i,r_i} v_i = 0\}$, we have

$$T_{Q_i,r_i} v_i(x_\infty) \to 0.$$

This implies $x_\infty \in \{v_\infty = 0\}$, which contradicts our previous assertion that $x_\infty \in \overline{B_R(0)} \setminus B_\epsilon(\{v_\infty = 0\})$.

The other direction goes the same way. Suppose that we could find a subsequence of $T_{Q_i,r_i} v_i \to v_\infty$ such that there was a point, $x \in \{v_\infty = 0\} \cap B_R(0)$, for which

$$\text{dist}(x, \{T_{Q_i,r_i} v_i = 0\} \cap B_R(0)) > \epsilon$$

for all $i = 1, 2, \ldots$. By Lemma 3.5, we know that $T_{Q_i,r_i} v_i(x) > C$. This contradicts convergence in $C_{\text{loc}}(\mathbb{R}^n)$, however, since $v_\infty(x) = 0$. □

Theorem 3.8 [Kenig and Toro 2006, Theorem 4.1]. *In general, if $\partial\Omega_i^\pm \in \mathcal{D}(n, \alpha, M_0)$ converge to a closed set A locally in the Hausdorff metric on compact subsets, then A divides $\mathbb{R}^n$ into two unbounded, two-sided NTA domains with NTA constant bounded by $2M_0$.*

We must now bound the upper Minkowski dimension of $A = \{v_\infty = 0\}$. We do so crudely, using only that A is the mutual boundary of a pair of two-sided NTA domains. That is, using the machinery of porous sets we are able to prove the following lemma.

Lemma 3.9. *Let $\Sigma \subset \mathbb{R}^n$ be the mutual boundary of a pair of unbounded two-sided NTA domains with NTA constant $1 < M_0$. Then, there exists $0 < \epsilon = \epsilon(M_0, n)$ such that $\overline{\dim}_{\mathcal{M}}(E) \leq n - \epsilon$.*

This is an elementary fact which seems to be omitted in the literature. We defer the proof to the Appendix. We now prove strong convergence.

Lemma 3.10 (strong compactness). *Let $\{v_i\}$ be a sequence of functions in $\mathcal{A}(n,\alpha,M_0)$ such that $\|\ln(h)\|_\alpha \le \Gamma$. Let $\{Q_i\} \subset \partial\Omega_i^{\pm} \cap B_1(0)$ and $0 < r_i < 1$. There is a subsequence $\{v_j\}$ and a Lipschitz function $v_\infty \in W^{1,2}_{\mathrm{loc}}$ such that $T_{Q_j,r_j} v_j \to v_\infty$ in the following senses*:

(1) *$T_{Q_j,r_j} v_j \to v_\infty$ in $C_{\mathrm{loc}}(\mathbb{R}^n)$.*

(2) *$T_{Q_j,r_j} v_j \to v_\infty$ in $W^{1,2}_{\mathrm{loc}}(\mathbb{R}^n)$.*

Proof. The only new claim is that $\nabla T_{Q_j,r_j} v_j \to \nabla v_\infty$ in $L^2_{\mathrm{loc}}(\mathbb{R}^n;\mathbb{R}^n)$. By Lemma 3.7, Theorem 3.8, and Lemma 3.9, we have that $\overline{\dim}_{\mathcal{M}}(\{v_\infty = 0\}) \le n - \epsilon$. In particular, then, $\mathcal{H}^n(B_r(\{v_\infty = 0\} \cap B_R(0))) \to 0$ as $r \to 0$ (see [Mattila 1995] for fundamental facts about Minkowski content, dimension and Hausdorff measure). Thus, for any $\theta > 0$ we can find an $r(\theta) > 0$ such that $\mathcal{H}^n(B_r(\{v_\infty = 0\} \cap B_R(0))) \le \theta$. This allows us to estimate

$$\begin{aligned}
\limsup_{j\to\infty} \|\nabla T_{Q_j,r_j} v_j\|^2_{L^2(B_R(0))} \\
&= \limsup_{j\to\infty} \left(\int_{B_R(0)\cap B_r(\{v_\infty=0\})} |\nabla T_{Q_j,r_j} v_j|^2 \, dV + \int_{B_R(0)\setminus B_r(\{v_\infty=0\})} |\nabla T_{Q_j,r_j} v_j|^2 \, dV \right) \\
&\le \lim_{j\to\infty} \int_{B_R(0)\setminus B_r(\{v_\infty=0\})} |\nabla T_{Q_j,r_j} v_j|^2 \, dV + C\theta \\
&\le \|\nabla v_\infty\|^2_{L^2(B_R(0))} + C\theta,
\end{aligned}$$

where the penultimate inequality uses the fact that v_j are uniformly Lipschitz, and the last equality follows from convergence in $C(B_{R+r}(0) \setminus B_{r/2}(\{v_\infty = 0\}))$ implying $C^\infty(B_R(0) \setminus B_r(\{v_\infty = 0\}))$ convergence because the $T_{Q_j,r_j} v_j$ are harmonic functions in this region. Since $\theta > 0$ was arbitrary, we have that $\limsup_{j\to\infty} \|\nabla T_{Q_j,r_j} v_j\|^2_{B_R(0)} \le \|\nabla v_\infty\|^2_{L^2(B_R(0))}$. The other inequality follows from the same trick or from lower semicontinuity. Therefore, we have the equality

$$\lim_{j\to\infty} \|\nabla T_{Q_j,r_j} v_j\|^2_{L^2(B_R(0))} = \|\nabla v_\infty\|^2_{L^2(B_R(0))}.$$

Thus, by weak convergence and norm convergence we have

$$\begin{aligned}
\lim_j \|\nabla T_{Q_j,r_j} v_j - \nabla v_\infty\|^2_{L^2(B_R(0))} \\
&= \lim_j \int_{B_R(0)} |\nabla T_{Q_j,r_j} v_j - \nabla v_\infty|^2 \, dV \\
&= \lim_j \|\nabla T_{Q_j,r_j} v_j\|^2_{L^2(B_R(0))} + \|\nabla v_\infty\|^2_{L^2(B_R(0))} - 2 \lim_j \langle \nabla T_{Q_j,r_j} v_j, \nabla v_\infty \rangle_{L^2(B_R(0))} \\
&= 2\|\nabla v_\infty\|^2_{L^2(B_R(0))} - 2\|\nabla v_\infty\|^2_{L^2(B_R(0))} = 0.
\end{aligned}$$

□

Because the functions $v^Q_{p,r}$ are merely Lipschitz, we will often need to work with a mollified version of them. We will use the convention that $v_\epsilon = v \star \phi_\epsilon$ for $\phi \in C^\infty$ a mollifying function (meaning $\mathrm{spt}(\phi) \subset B_1$ and $\int \phi \, dV = 1$).

Corollary 3.11. *Let $v \in \mathcal{A}(n, \alpha, M_0)$ and $v_\epsilon = v \star \phi_\epsilon$ be a mollification of v. By standard mollification results,*

$$v_\epsilon \to v \quad \text{in } W^{1,2}_{\text{loc}}(\mathbb{R}^n),\ C_{\text{loc}}(\mathbb{R}^n) \qquad \text{as } \epsilon \to 0.$$

4. Almost monotonicity of the Almgren frequency function

One of the key tools of this paper will be the Almgren frequency function (introduced in [Almgren 1979]).

Definition 4.1 (Almgren frequency function). For any Lipschitz function $v : \mathbb{R}^n \to \mathbb{R}$, radius $r > 0$, and point $Q \in \partial\Omega^\pm$, the *Almgren frequency function* is defined as

$$N(p, r, v) = r\frac{D(p, r, v)}{H(p, r, v)}, \tag{4-1}$$

where

$$H(p, r, v) = \int_{\partial B_r(p)} |v|^2\, d\sigma, \quad D(p, r, v) = \int_{B_r(p)} |\nabla v|^2\, dV. \tag{4-2}$$

Remark 4.2. The Almgren frequency function is invariant in the following senses. For $a, b \in \mathbb{R}$ with $a, b \neq 0$, if $w(x) = av(bx)$, then $N(0, r, v) = N(0, b^{-1}r, w)$.

If u is harmonic then $N(p, r, u)$ is monotonically nondecreasing. If additionally one assumes that $u(p) = 0$ then $\lim_{r\to 0} N(p, r, u) = N(p, 0, u) \geq 1$ is the degree of the leading homogeneous harmonic polynomial in the Taylor expansion of u at the point p.

4A. ***Consequences of Section 3 for the Almgren frequency function.*** Before turning to the main results of this section, we note that the results of Section 3 immediately imply the following corollaries.

Corollary 4.3. *Under the hypotheses of Lemma 3.6, there exists a subsequence such that, for all $r \in (0, 2]$,*

$$N(0, r, T_{Q_j, r_j} v_j) \to N(0, r, v_\infty).$$

Moreover, if $v_\epsilon = v \star \phi$ for a mollifier ϕ as in Corollary 3.11 then, for all $Q \in B_1(0) \cap \partial\Omega^\pm$ and $0 < r \leq 1$,

$$\lim_{\epsilon\to 0} N(Q, r, v_\epsilon) = N(Q, r, v).$$

Proof. This follows from the convergence of the numerator and the denominator; the former follows from Lemma 3.6 (2) and the latter from Lemma 3.6 (1). For the convolution, both follow from Corollary 3.11. □

Corollary 4.4. *Let $v \in \mathcal{A}(n, \alpha, M_0)$ as above. There is a function $C(\alpha, \Gamma, M_0)$ such that, if $\|\ln(h)\|_\alpha \leq \Gamma$ then for all $Q \in B_1(0) \cap \partial\Omega^\pm$ and all $r \in (0, 1]$,*

$$N(Q, r, v) \leq C(\Gamma, \alpha, M_0). \tag{4-3}$$

Proof. We recall that the Almgren frequency function is invariant under rescalings of the function v. Therefore, $N(0, 1, v_{Q,r}) = D(0, 1, T_{Q,r} v)$ is bounded by Lemma 3.4 and the constant only depends upon M_0, Γ, and α. □

4B. *Quantitative almost monotonicity.* This section is dedicated to providing a quantitative version of the following result of Engelstein [2016].

Lemma 4.5 [Engelstein 2016]. *Let $v \in \mathcal{A}(n, \alpha, M_0)$ and $Q \in K \Subset \partial\Omega^{\pm}$. There exists a constant $C < \infty$ (which can be taken uniformly over K and $r \in (0, 1]$) such that*

$$\liminf_{\epsilon \to 0} N(Q, r, v_\epsilon^Q) - N(Q, 0, v_\epsilon^Q) > -Cr^\alpha.$$

The quantitative version of this result which we prove below in Lemma 4.9 is essential for connecting the Almgren frequency to Jones' beta numbers in the "frequency pinching" result later in Lemma 8.2. It comes from examining the derivative of the Almgren frequency function in the r variable.

Throughout this section, we shall use the notation $(v_\epsilon)_\nu(y) = \nabla v_\epsilon(y) \cdot \nu(y)$, where $\nu(y)$ is the unit normal to $\partial B_r(Q)$ at y. By differentiation (see [Engelstein 2016, Section 5.1] for details of the derivation),

$$\begin{aligned} H(Q, r, v_\epsilon)^2 \frac{d}{dr} N(Q, r, v_\epsilon) = 2r\biggl(\int_{\partial B_r(Q)} (v_\epsilon)_\nu^2 \, d\sigma \int_{\partial B_r(Q)} |v_\epsilon|^2 \, d\sigma - \biggl[\int_{\partial B_r(Q)} v_\epsilon (v_\epsilon)_\nu \, d\sigma\biggr]^2\biggr) \\ + 2r\biggl(\int_{B_r(Q)} v_\epsilon \Delta v_\epsilon \, dV\biggr)\biggl(\int_{\partial B_r(Q)} v_\epsilon (v_\epsilon)_\nu \, d\sigma\biggr) \\ - 2H(Q, r, v_\epsilon) \int_{B_r(Q)} \langle x - Q, \nabla v_\epsilon \rangle \Delta v_\epsilon \, dV. \end{aligned} \tag{4-4}$$

We write the decomposition $\frac{d}{dr} N(Q, r, v_\epsilon) = N_1'(Q, r, v_\epsilon) + N_2'(Q, r, v_\epsilon)$ with

$$N_1'(Q, r, v_\epsilon) := H(Q, r, v_\epsilon)^{-2} 2r\biggl(\int_{\partial B_r(Q)} (v_\epsilon)_\nu^2 \, d\sigma \int_{\partial B_r(Q)} |v_\epsilon|^2 \, d\sigma - \biggl[\int_{\partial B_r(Q)} v_\epsilon (v_\epsilon)_\nu \, d\sigma\biggr]^2\biggr).$$

We call what remains $N_2'(Q, r, v_\epsilon)$:

$$\begin{aligned} N_2'(Q, r, v_\epsilon) := H(Q, r, v_\epsilon)^{-2}\biggl[2r\biggl(\int_{B_r(Q)} v_\epsilon \Delta v_\epsilon \, dV\biggr)\biggl(\int_{\partial B_r(Q)} v_\epsilon (v_\epsilon)_\nu \, d\sigma\biggr) \\ -2H(Q, r, v_\epsilon) \int_{B_r(Q)} \langle x - Q, \nabla v_\epsilon \rangle \Delta v_\epsilon \, dV(x)\biggr]. \end{aligned}$$

Note that by the Cauchy–Schwarz inequality, $N_1'(Q, r, v_\epsilon) \geq 0$.

Lemma 4.6. *Let $v \in \mathcal{A}(n, \alpha, M_0)$, $Q \in \partial\Omega^{\pm} \cap B_1(0)$ and $0 < r \leq 1$. Then, if $C = \mathrm{Lip}(v|_{B_2(0)})$,*

$$\begin{aligned} N_1'(Q, r, v) &= 2\int_{\partial B_r(Q)} \frac{|\nabla v \cdot (y - Q) - N(Q, r, v)v|^2}{H(Q, |y - Q|, v)|y - Q|} \, d\sigma(y) \\ &\geq \frac{2}{C} \int_{\partial B_r(Q)} \frac{|\nabla v \cdot (y - Q) - N(Q, r, v)v|^2}{|y - Q|^{n+2}} \, d\sigma(y). \end{aligned} \tag{4-5}$$

Proof. Recall that for the Cauchy–Schwarz inequality, we have, for $\lambda = \langle u, v\rangle / \|v\|^2$,

$$\|v\|^2 \|u - \lambda v\|^2 = \|u\|^2 \|v\|^2 - |\langle u, v\rangle|^2.$$

Choosing

$$u = \nabla v_\epsilon \cdot \left(\frac{y-Q}{|y-Q|}\right) \quad \text{and} \quad v = v_\epsilon,$$

using the divergence theorem on λ, and letting $\epsilon \to 0$, we have

$$N_1'(Q,r,v) = H(Q,r,v)^{-1} 2r \left(\int_{\partial B_r(Q)} \left| (v)_\nu - \frac{1}{r} N(Q,r,v) v \right|^2 d\sigma \right).$$

This proves the equality. To prove the lower bound, we let $C = \mathrm{Lip}(T_{0,1}v|_{B_2(0)})$ and observe that $H(Q,r,T_{0,1}v_\epsilon) \le Cr^{n+1}$. Plugging this into the above equation, we get the desired inequality

$$N_1'(Q,r,v) \ge \frac{2}{C} \int_{\partial B_r(Q)} \frac{|\nabla T_{0,1}v(y)\cdot(y-Q) - N(Q,r,v)T_{0,1}v(y)|^2}{|y-Q|^{n+2}}\, d\sigma(y). \qquad \square$$

In order to bound the parts of $N_2'(Q,r,v_\epsilon)$, we recall some estimates from [Engelstein 2016].

Lemma 4.7 [Engelstein 2016, Lemmata 5.4, 5.5, and 5.6]. *Let $v \in \mathcal{A}(n,\alpha,M_0)$ with $\|\ln(h)\|_\alpha \le \Gamma$, and let $Q \in B_1(0) \cap \partial\Omega^\pm$. For any $0 < s$ and $\epsilon \ll s$,*

$$\int_{\partial B_s(Q)} |v_\epsilon|^2\, d\sigma \ge C(M_0) \frac{\omega^-(B_s(Q))^2}{s^{n-3}}, \tag{4-6}$$

$$\left| \int_{B_s(Q)} v_\epsilon \Delta v_\epsilon\, dV \right| \le C\|\ln(h)\|_\alpha s^\alpha \frac{\omega^-(B_s(Q))^2}{s^{n-2}}, \tag{4-7}$$

$$\left| \int_{B_s(Q)} \langle \nabla v_\epsilon, x-Q\rangle \Delta v_\epsilon\, dV(x) \right| \le C\|\ln(h)\|_\alpha s^\alpha \frac{\omega^-(B_s(Q))^2}{s^{n-2}}, \tag{4-8}$$

$$\left| \int_{\partial B_s(Q)} v_\epsilon (v_\epsilon)_\nu\, d\sigma \right| \le C \frac{\omega^-(B_s(Q))^2}{s^{n-1}}, \tag{4-9}$$

where $C = C(\alpha, M_0, \Gamma)$.

Proof. Let $v \in \mathcal{A}(n,\alpha,M_0)$ be given. Recall that $v = v^0$. Engelstein [2016, Lemmata 5.4, 5.5, and 5.6] proves the claim for the functions $v^Q_{Q,1}$. Hence, for any such v and any such Q, the integral estimates hold for $u(x) = v^Q(x+Q)$ as well. However, in general, such $v^0(\cdot + Q)$ are not in $\mathcal{A}(n,\alpha,M_0)$ because $h(0)$ may not be 0. But,

$$v^0(x+Q) = cu^+(x+Q) - u^-(x+Q)$$

is an element of $\mathcal{A}(n,\alpha,M_0)$ for some constant $e^{-\Gamma|Q|^\alpha} \le c \le e^{\Gamma|Q|^\alpha}$ as in (3-1). Using this identity and following the proofs of [Engelstein 2016, Lemmata 5.4, 5.5, and 5.6] gives the claim. $\square$

Remark 4.8. Recalling our expansion of $\frac{d}{dr}N(r,p,v_\epsilon)$ in (4-4) and the bounds contained in Lemma 4.7 we have that, for $v \in \mathcal{A}(n,\alpha,M_0)$ with $\|\ln(h)\|_\alpha \le \Gamma$, $\epsilon \ll r$, and $Q \in B_1(0) \cap \partial\Omega^\pm$,

$$|N_2'(Q,r,v_\epsilon)| \le C_1 \|\ln(h)\|_\alpha r^{\alpha-1}, \tag{4-10}$$

where $C_1 = C(\alpha, M_0, \Gamma)$.

We now state the main result of this section.

Lemma 4.9. *Let $v \in \mathcal{A}(n, \alpha, M_0)$ with $\|\ln(h)\|_\alpha \le \Gamma$, and let $Q \in B_1(0) \cap \partial\Omega^{\pm}$. For any $0 \le s < S \le 1$,*

$$\begin{aligned}\frac{2}{C}\int_{A_{s,S}(Q)} \frac{|\nabla T_{0,1}v(y)\cdot(y-Q) - N(Q,|y-Q|,T_{0,1}v)T_{0,1}v(y)|^2}{|y-Q|^{n+2}}\,dV(y)& \\ \le 2\int_{A_{s,S}(Q)} \frac{|\nabla v(y)\cdot(y-Q) - N(Q,|y-Q|,v)v(y)|^2}{H(Q,|y-Q|,v)|y-Q|}\,dV(y)& \\ \le N(Q,S,v) - N(Q,s,v) + C_1\|\ln(h)\|_\alpha S^\alpha,& \qquad (4\text{-}11)\end{aligned}$$

where $C_1 = C_1(\alpha, M_0, \Gamma)$ and $C(M_0, \Gamma, \alpha) = \mathrm{Lip}(T_{0,1}v|_{B_2(0)})$.

Proof. We begin by normalizing v. Since $N(r, p, v) = N(r, p, cv)$ for any $c \neq 0$, we may work with $T_{0,1}v$. Note that by Remark 4.8 and (4-5), $N(Q, r, v)$ is continuous in r and hence we may find an $0 \le s < s_1$ such that

$$|N(Q,s,v) - N(Q,s_1,v)| \le \|\ln(h)\|_\alpha S^\alpha.$$

By Corollaries 3.11 and 4.3 we can find an $\epsilon \ll s$ small enough that

$$|N(Q,s_1,v_\epsilon) - N(Q,s_1,v)| < \|\ln(h)\|_\alpha S^\alpha \quad \text{and} \quad |N(Q,S,v_\epsilon) - N(Q,S,v)| < \|\ln(h)\|_\alpha S^\alpha.$$

Thus, we reduce to estimating $N(Q, S, T_{0,1}v_\epsilon) - N(Q, s_1, T_{0,1}v_\epsilon)$:

$$\begin{aligned}N(Q,S,T_{0,1}v_\epsilon) - N(Q,s_1,T_{0,1}v_\epsilon) &= \int_{s_1}^{S} \frac{d}{dr}N(Q,r,T_{0,1}v_\epsilon)\,dr \\ &= \int_{s_1}^{S} N_1'(Q,r,T_{0,1}v_\epsilon)\,dr + \int_{s_1}^{S} N_2'(Q,r,T_{0,1}v_\epsilon)\,dr.\end{aligned}$$

Recalling Remark 4.8, Lemma 4.6, and letting $\epsilon \to 0$ gives the lemma. □

Using these estimates it is possible to control the drop across scales from the total drop.

Lemma 4.10. *If $v \in \mathcal{A}(n, \alpha, M_0)$ with $\|\ln(h)\|_\alpha \le \Gamma$ and $Q \in B_1(0) \cap \partial\Omega^{\pm}$, then for any $0 \le r \le s < S \le R$*

$$N(Q,S,v) - N(Q,s,v) \le 2C_1\|\ln(h)\|_\alpha R^\alpha + |N(Q,R,v) - N(Q,r,v)|.$$

Proof. This is essentially a "rays of the sun" argument. To wit,

$$\begin{aligned}N(Q,S,v) - N(Q,s,v) &= \int_s^S N_1'(Q,\rho,v) + N_2'(Q,\rho,v)\,d\rho \\ &\le \int_s^S N_1'(Q,\rho,v) + |N_2'(Q,\rho,v)|\,d\rho \\ &\le \int_r^R N_1'(Q,\rho,v) + |N_2'(Q,\rho,v)|\,d\rho \\ &\le 2\int_r^R |N_2'(Q,\rho,v)|\,d\rho + |N(Q,R,v) - N(Q,r,v)|.\end{aligned}$$

The bounds in Remark 4.8 give the desired statement. □

We now turn our attention to proving a "doubling" property for $H(p, r, v)$. This is an analog of classical harmonic results for the Almgren frequency function, modified for our almost harmonic functions $v \in \mathcal{A}(n, \alpha, M_0)$.

Lemma 4.11 ($H(r, p, v_\epsilon)$ is almost doubling). *Let $v \in \mathcal{A}(n, \alpha, M_0)$ with $\|\ln(h)\|_\alpha \le \Gamma$, and let $Q \in B_{1/2}(0) \cap \partial\Omega^\pm$. For any $0 < s < S \le 1$, if $\epsilon \ll s$ is sufficiently small,*

$$H(Q, S, v_\epsilon) \le \left(\frac{S}{s}\right)^{(n-1)+2(N(Q,S,v_\epsilon)+CS^\alpha)} \exp\left(\frac{2C}{\alpha}[S^\alpha - s^\alpha]\right) H(Q, s, v_\epsilon), \tag{4-12}$$

where $C = \|\ln(h)\|_\alpha C_1(M_0, \alpha, \Gamma)$ and C_1 is as in Remark 4.8.

Proof. First, observe that

$$\frac{d}{dr} H(Q, r, v_\epsilon) = \frac{n-1}{r} \int_{\partial B_r(Q)} |v_\epsilon|^2 \, d\sigma + 2 \int_{B_r(Q)} |\nabla v_\epsilon|^2 \, dV + 2 \int_{B_r(Q)} v_\epsilon \Delta v_\epsilon \, dV.$$

Next, we consider the identity

$$\begin{aligned} \ln\left(\frac{H(Q, S, v_\epsilon)}{H(Q, s, v_\epsilon)}\right) &= \ln(H(Q, S, v_\epsilon)) - \ln(H(Q, s, v_\epsilon)) \\ &= \int_s^S \frac{H'(Q, r, v_\epsilon)}{H(Q, r, v_\epsilon)} \, dr = \int_s^S \frac{n-1}{r} + \frac{2}{r} N(Q, r, v_\epsilon) + 2\left(\frac{\int_{B_r(Q)} v_\epsilon \Delta v_\epsilon dV}{\int_{\partial B_r(Q)} (v_\epsilon)^2 d\sigma}\right) dr. \end{aligned}$$

We bound $N(r, Q, v_\epsilon)$ by Lemma 4.9. We bound the last term using Lemma 4.7. Plugging in these bounds, we have, for $\epsilon \ll s$,

$$\ln\left(\frac{H(Q, S, v_\epsilon)}{H(Q, s, v_\epsilon)}\right) \le [(n-1) + 2(N(Q, S, v_\epsilon) + CS^\alpha)] \ln(r)\Big|_s^S + \frac{2C}{\alpha} r^\alpha \Big|_s^S.$$

Evaluating and exponentiating gives the desired result. □

Remark 4.12. Because $H(Q, r, v_\epsilon) \to H(Q, r, v)$ as $\epsilon \to 0$ and $N(Q, r, v_\epsilon) \to N(Q, r, v)$ as $\epsilon \to 0$ (a consequence of Corollary 3.11), we have the following inequality. For all $v \in \mathcal{A}(n, \alpha, M_0)$ with $\|\ln(h)\|_\alpha \le \Gamma$, $Q \in B_{1/2}(0) \cap \partial\Omega^\pm$, and $0 < s < S \le 1$,

$$H(Q, S, v) \le \left(\frac{S}{s}\right)^{(n-1)+2(N(Q,S,v)+CS^\alpha)} \exp\left(\frac{2C}{\alpha}[S^\alpha - s^\alpha]\right) H(Q, s, v). \tag{4-13}$$

5. Quantitative rigidity

Throughout the rest of the paper, we shall need to use limit-compactness arguments. The key will be that $v \to u$ for some harmonic function u as $\|\ln(h)\|_\alpha \to 0$. We make this rigorous in the following lemma.

Lemma 5.1 (convergence to harmonic functions). *Let $v_i \in \mathcal{A}(n, \alpha, M_0)$ with $\|\ln(h_i)\|_\alpha \to 0$. Assume that $Q_i \in B_1(0) \cap \partial\Omega_i^\pm$ and $\{r_i\} \subset (0, 1]$. Then there exists a function v_∞ and a subsequence v_j such that $T_{Q_j, r_j} v_j \to v_\infty$ in the sense of Lemma 3.10 and v_∞ is harmonic.*

Proof. Lemma 3.10 gives a subsequence $T_{Q_j,r_j}v_j$ which converges strongly in $W^{1,2}_{\text{loc}}(\mathbb{R}^n)$ to a function v_∞. We claim that v_∞ is harmonic. To see this, we investigate the behavior of its mollifications $v_{\infty,\epsilon} = v_\infty \star \phi_\epsilon$. Observe that by Young's inequality,

$$\|T_{Q_j,r_j}v_{j,\epsilon} - v_{\infty,\epsilon}\|_{L^2(B_2(0))} \leq \|\phi_\epsilon\|_{L^1(B_2(0))}\|T_{Q_j,r_j}v_j - v_\infty\|_{L^2(B_2(0))}.$$

Thus, for any $\epsilon > 0$ we have $T_{Q_j,r_j}v_{j,\epsilon} \to v_{\infty,\epsilon}$ as $j \to \infty$ strongly in $L^2(B_2(0))$. By a similar argument applied to $\nabla T_{Q_j,r_j}v_{j,\epsilon}$, we also have that $\nabla T_{Q_j,r_j}v_{j,\epsilon} \to \nabla v_{\infty,\epsilon}$ in $L^2(B_2(0); \mathbb{R}^n)$ as $j \to \infty$. Furthermore, by our uniform Lipschitz bounds, $T_{Q_j,r_j}v_{j,\epsilon} \to v_{\infty,\epsilon}$ as $j \to \infty$ in $C(B_2(0))$ as well.

We will show that for $\epsilon \ll 1$ the function $v_{\infty,\epsilon}$ is harmonic. First, for any test function $\xi \in C_c^\infty(B_2(0))$, we have

$$\left|\int_{B_2(0)} \xi(\Delta T_{Q_j,r_j}v_{j,\epsilon} - \Delta v_{\infty,\epsilon})\,dV\right| = \left|\int_{B_2(0)} \Delta\xi(T_{Q_j,r_j}v_{j,\epsilon} - v_{\infty,\epsilon})\,dV\right| \leq \|\Delta\xi\|_{L^2(B_2(0))}\|T_{Q_j,r_j}v_{j,\epsilon} - v_{\infty,\epsilon}\|_{L^2(B_2(0))}.$$

Since $T_{Q_j,r_j}v_{j,\epsilon} \to v_{\infty,\epsilon}$ strongly in $L^2(B_2(0))$, we have $\Delta T_{Q_j,r_j}v_{j,\epsilon} \rightharpoonup \Delta v_{\infty,\epsilon}$ in $L^2(B_2(0))$.

However, by assumption, we also have

$$\left|\int_{B_2(0)} \xi\,\Delta T_{Q_j,r_j}v_{j,\epsilon}\,dV\right| \leq \int_{B_2(0)} |\xi_\epsilon|\left|\frac{h_j(0)}{h_j(x)} - 1\right| dT_{Q_j,r_j}\omega^- \leq C\max_{B_2(0)}|\xi|\cdot\|\ln(h_j)\|_\alpha T_{Q_j,r_j}\omega^-(B_3(0)),$$

where $T_{Q_j,r_j}\omega^\pm$ are the interior and exterior harmonic measures associated to $T_{Q_j,r_j}v_j$. Note that $T_{Q_j,r_j}\omega^- \neq \omega^-_{Q_j,r_j}$, but, by Definitions 2.4 and 2.7 and Lemma 3.4, there is a constant $c' = c'(M_0)$ such that $T_{Q_j,r_j}\omega^- = c\omega^-_{Q_j,r_j}$ and $c \leq c'$. Since $\omega^-_{r_j,Q_j}(B_3(0))$ are uniformly bounded by Theorem 2.6, the $T_{Q_j,r_j}\omega^-(B_3(0))$ are, too. Thus, as $j \to \infty$, we have that $\Delta T_{Q_j,r_j}v_{j,\epsilon} \rightharpoonup 0$ in $L^2(B_2(0))$ as well. Thus, $\Delta v_{\infty,\epsilon} = 0$ weakly in $L^2(B_2(0))$. Since $v_{\infty,\epsilon} \in C^\infty(B_2(0))$, we have that $v_{\infty,\epsilon}$ is harmonic.

Since v_∞ is Lipschitz continuous, $v_{\infty,\epsilon} \to v_\infty$ in $C(B_R(0))$ as $\epsilon \to 0$. Thus, for all $x \in B_R(0)$ we have both that $v_{\infty,\epsilon}(x) \to v_\infty(x)$ as $\epsilon \to 0$ and that

$$\fint_{B_r(x)} v_{\infty,\epsilon}(y)\,dV(y) \to \fint_{B_r(x)} v_\infty(y)\,dV(y)$$

as $\epsilon \to 0$. Thus, v_∞ must satisfy the mean value property and is therefore harmonic. □

Now that we have Lemma 5.1, we can prove a quantitative rigidity result. Loosely speaking, it says that if a function $v \in \mathcal{A}(n, \alpha, M_0)$ behaves like a homogeneous harmonic polynomial with respect to the Almgren frequency (in the sense that it has small drop across scales), then it must be close to being a homogeneous harmonic polynomial. This will connect the behavior of the Almgren frequency to our quantitative stratification.

Lemma 5.2 (quantitative rigidity). *Let $v \in \mathcal{A}(n, \alpha, M_0)$, as above. Let $Q \in B_1(0) \cap \partial\Omega^\pm$. For every $\delta > 0$, there is an $\gamma = \gamma(n, \alpha, M_0, \delta) > 0$ such that if $\|\ln(h)\|_\alpha \leq \gamma$ and*

$$N(Q, 1, v) - N(Q, \gamma, v) \leq \gamma,$$

then v is $(0, \delta, 1, Q)$-symmetric.

Proof. We argue by contradiction. Assume that there exists a $\delta > 0$ such that there is a sequence of functions $v_i \in \mathcal{A}(n, \alpha, M_0)$ with $\|\ln(h_i)\|_\alpha \le 2^{-i}$ for which there exists a point $Q_i \in B_1(0) \cap \partial\Omega_i^\pm$ with

$$N(Q_i, 1, v_i) - N(Q_i, 2^{-i}, v_i) \le 2^{-i},$$

but where no v_i is $(0, \delta, 1, Q_i)$-symmetric.

By Lemma 5.1 there exists a subsequence $T_{Q_j,1}v_j$ which converges strongly in $W^{1,2}_{\mathrm{loc}}$ to a harmonic function v_∞. Therefore $N(Q, r, v_\infty)$ is monotone increasing. Further, by Corollary 4.3 we know that $\lim_{j\to\infty} N(0, r, T_{Q_j,1}v_j) = N(0, r, v_\infty)$ for all $0 < r \le 1$. By Lemma 4.10 and the aforementioned convergence, we have that

$$N(0, 1, v_\infty) - N(0, 0, v_\infty) = 0.$$

This implies that v_∞ is a homogeneous harmonic polynomial (see, for example, the proof of [Han and Lin 1994, Theorem 2.2.3]). Thus, we arrive at our contradiction, since the $T_{Q_j,1}v_j$ were assumed to stay away from all such functions in $L^2(B_1(0))$. □

Remark 5.3. Since $N(Q, r, v)$ is scale-invariant, Lemma 5.2 is also scale-invariant in the sense that if $N(Q, r, v) - N(Q, \gamma r, v) \le \gamma$ and $\|\ln(h)\|_\alpha \le \gamma$, then v is $(0, \delta, r, Q)$-symmetric.

6. A dichotomy

The proof technique in the rest of the paper is an adaptation of techniques developed by Naber and Valtorta [2017].

This section is dedicated to proving a lemma that gives us geometric information on the quantitative strata. Roughly, it says that if we can find $(k+1)$ points that are well-separated and the Almgren frequency has very small drop at these points, then the quantitative strata is contained in a neighborhood of the affine k-plane which contains them and we have control on the Almgren frequency for all points in that neighborhood. This is a quantitative analog of the following classical result.

Proposition 6.1. *Let $P : \mathbb{R}^n \to \mathbb{R}$ be a homogeneous harmonic polynomial. Let $0 \le k \le n-2$. If P is translation-invariant with respect to some k-dimensional subspace V and P is homogeneous with respect to some point $x \notin V$, then P is $(k+1)$-symmetric with respect to* $\operatorname{span}\{x, V\}$.

See [Cheeger et al. 2015, Proposition 2.11] or [Han and Lin 1994, proof of Theorem 4.1.3].

We shall use the notation $\langle y_0, \dots, y_k\rangle$ to denote the k-dimensional affine linear subspace which passes through $y_0, \dots, y_k$.

Lemma 6.2. *Let $v \in \mathcal{A}(n, \alpha, M_0)$ and $0 < \epsilon$ be fixed. Let $\gamma, \eta', \rho > 0$ be fixed, then there exist constants $0 < \eta_0(n, \alpha, E_0, \epsilon, \eta', \gamma, \rho) \ll \rho$ and $0 < \beta(n, \alpha, E_0, \epsilon, \eta', \rho) < 1$ such that, if*

(1) $E = \sup_{Q\in B_1(0)\cap\partial\Omega^\pm} N(Q, 2, v) \in [0, E_0]$,

(2) *there exist points $\{y_0, y_1, \dots, y_k\} \subset B_1(0) \cap \partial\Omega^\pm$ satisfying $y_i \notin B_\rho(\langle y_0, \dots, y_{i-1}, y_{i+1}, \dots, y_k\rangle)$ and*

$$N(y_i, \gamma\rho, v) \ge E - \eta_0$$

for all $i = 0, 1, \dots, k$, and

(3) $\|\ln(h)\|_\alpha \le \eta_0$,

then, writing $\langle y_0, \dots, y_k\rangle = L$, *for all* $Q \in B_\beta(L) \cap B_1(0) \cap \partial\Omega^\pm$,

$$N(Q, \gamma\rho, v) \ge E - \eta'$$

and

$$\mathcal{S}^k_{\epsilon,\eta_0} \cap B_1(0) \subset B_\beta(L).$$

Proof. There are two conclusions. We argue by contradiction for both. Suppose that the first claim fails. That is, assume that there exist constants $\gamma, \rho, \eta' > 0$ for which there exists a sequence $v_i \in \mathcal{A}(n, \alpha, M_0)$ with $\sup_{Q\in B_1(0)} N(Q, 2, v_i) = E_i \in [0, E_0]$ and points $\{y_{i,j}\}_j$ satisfying (2) above, with $\|\ln(h_i)\|_\alpha \le 2^{-i}$, $\eta_0 < 2^{-i}$, and a sequence $\beta_i \le 2^{-i}$ such that, for each i, there exists a point $x_i \in B_{\beta_i}(L_i) \cap B_1(0) \cap \partial\Omega^\pm_i$ for which $N(x_i, \gamma\rho, v_i) < E - \eta'$.

By Lemma 5.1, there exists a subsequence v_i such that $T_{0,1}v_i$ converges to a harmonic function v_∞ in the senses outlined in the lemma. Further, by the compactness of $[0, E_0]$, $\overline{B_1(0)}$, and the Grassmannian, we may assume that

$$E_i \to E, \quad y_{i,j} \to y_j, \quad L_i \to L, \quad x_i \to x_\infty \in \overline{B_1(0)} \cap \partial\Omega^\pm_\infty,$$

where $\partial\Omega^\pm_\infty = \{v_\infty = 0\}$ is a two-sided NTA domain with constant $2M_0$ by Theorem 3.8. Note that the convergence given by Lemma 5.1 implies

$$\sup_{Q\in B_1(0)} N(Q, 2, v_\infty) \le E, \qquad N(x_\infty, \gamma\rho, v_\infty) < E - \eta',$$

and

$$N(y_j, \gamma\rho, v_\infty) \ge E$$

for all $j = 0, 1, \dots, k$. Because v_∞ is harmonic, $N(p, r, v_\infty)$ is nondecreasing in r for all $p \in B_2(0)$. Therefore, $N(y_i, r, v_\infty) = E$ for all y_i and all $r \in [\gamma\rho, 2]$. Thus, v_∞ is a 0-symmetric function in $B_2(y_j) \setminus B_{\gamma\rho}(y_j)$ for each y_j. By unique continuation, v_∞ is 0-symmetric with respect to y_j for each j. Because the $y_j \in \overline{B_1(0)}$ are in general position, by Proposition 6.1, v_∞ is translation-invariant along L in $B_2(0)$. Since $x_\infty \in L \cap \overline{B_1(0)}$, this implies that $N(x_\infty, 0, v_\infty) = E$. But this contradicts $N(x_\infty, \gamma\rho, v_\infty) < E - \eta'$, since $N(x_\infty, r, v_\infty)$ must be nondecreasing in r. This proves the first claim.

Now assume that the second claim fails. That is, fix $\beta > 0$ and assume that there is a sequence $v_i \in \mathcal{A}(n, \alpha, M_0)$ with $\sup_{Q\in B_1(0)} N(Q, 2, v_i) = E_i \in [0, E_0]$ and points $\{y_{i,j}\}_j$ satisfying (2) above, with $\|\ln(h_i)\|_\alpha \le 2^{-i}$ and a sequence $\eta_i \to 0$ such that for each i there exists a point $x_i \in \mathcal{S}^k_{\epsilon,\eta_i}(v_i) \cap B_1(0) \setminus B_\beta(L_i)$.

Again, we extract a subsequence as above. The function v_∞ will be harmonic and k-symmetric in $B_{1+\delta}(0)$, as above, and $x_i \to x \in \overline{B_1(0)} \setminus B_\beta(L)$. Note that by our definition of $\mathcal{S}^k_{\epsilon,\eta_i}(v_i)$ and the convergence in Lemma 5.1, $x \in \mathcal{S}^k_{\epsilon/2}(v_\infty)$.

Since v_∞ is k-symmetric and L is its k-dimensional spine, every blow-up at a point in $\overline{B_1(0)} \setminus B_\beta(L)$ will be $(k+1)$-symmetric. Thus, there must exist a radius r for which v_∞ is $\left(k+1, \frac{1}{4}\epsilon, r, x\right)$-symmetric. This contradicts the conclusion that $x \in \mathcal{S}^k_{\epsilon/2}(v_\infty)$. □

Consider the following dichotomy: either we can find $(k+1)$ well-separated points y_{ij} with very small drop in frequency or we cannot. In the former case, Lemma 6.2 implies that the Almgren frequency has small drop on all of $\mathcal{S}^k_{\epsilon,\eta}(v)$ (and we also get good geometric control). In the latter case, the set on which the Almgren frequency has small drop is close to a $(k-1)$-plane. In this case, even though we have no geometric control on $\mathcal{S}^k_{\epsilon,\eta}(v)$, we have very good packing control on the part with small drop in frequency. We make this formal in the following corollary.

Corollary 6.3 (key dichotomy). *Let $\gamma, \rho, \eta' \in (0,1)$ and $0 < \epsilon$ be fixed. There exist*

$$0 < \beta(n, \alpha, E_0, \epsilon, \eta', \rho) < 1 \quad \textit{and} \quad 0 < \eta_0 = \eta_0(n, \alpha, E_0, \epsilon, \eta', \gamma, \rho) \ll \rho$$

such that the following holds. For all $v \in \mathcal{A}(n, \alpha, M_0)$ with $\sup_{Q \in B_1(0)} N(Q, 2, v) \le E \in [0, E_0]$, if $\eta \le \eta_0$ and $\|\ln(h)\|_\alpha \le \eta$, then one of the following possibilities must occur:

(1) *$N(Q, \gamma\rho, v) \ge E - \eta'$ on $S^k_{\epsilon,\eta_0}(v) \cap B_1(0)$ and*

$$\mathcal{S}^k_{\epsilon,\eta_0} \cap B_1(0) \subset B_\beta(L).$$

(2) *There exists a $(k-1)$-dimensional affine plane L^{k-1} such that*

$$\{Q \in \partial\Omega^\pm : N(Q, 2\eta, v) \ge E - \eta_0\} \cap B_1(0) \subset B_\rho(L^{k-1}).$$

Remark 6.4. The former case is simply the conclusion of Lemma 6.2. In the latter case of the dichotomy, we know that all points in $\partial\Omega^\pm \cap B_1(0) \setminus B_\rho(L^{k-1})$ must have $N(Q, 2\eta, v) < E - \eta_0$. Since $N(Q, r, v)$ is almost monotonic and uniformly bounded, this can happen for each Q only finitely many times.

7. Spatial derivatives of the Almgren frequency

The main result of this section is Corollary 7.7, in which we estimate the difference between the Almgren frequency at nearby points. First, we need a preliminary estimate which extends one of the results of Lemma 4.7 to points $p \in B_1(0) \setminus \partial\Omega^\pm$.

Lemma 7.1. *Let $v \in \mathcal{A}(n, \alpha, M_0)$, and let $0 < s \le 1$, $Q \in \partial\Omega^\pm \cap B_1(0)$, and $p \in B_{s/3}(Q)$. Then we have the estimate*

$$H(p, s, v_\epsilon) \ge C(n, M_0) \frac{\omega^-(B_s(Q))^2}{s^{n-3}}.$$

Furthermore, for all $0 < s \le \frac{1}{2}$ and all $\frac{1}{2}s \le r \le 2s$,

$$H(p, r, v_\epsilon) \approx_{n,\alpha,M_0,\Gamma} H(Q, 2s, v_\epsilon) \approx_{n,\alpha,M_0,\Gamma} H\big(Q, \tfrac{1}{2}s, v_\epsilon\big).$$

Proof. Let $x_{\max}(p, s)^\pm$ denote the point in $\partial B_s(p) \cap \Omega^\pm$ which maximizes $|v|$ on $\partial B_s(p) \cap \Omega^\pm$.

If we can show that, for all p and all $0 < s \le \frac{1}{2}$,

$$|v(x_{\max}(Q, s)^-)| \sim_{M_0} \frac{\omega^-(B_s(Q))}{s^{n-2}}$$

and that $\mathrm{dist}(x_{\max}(p,s), \partial\Omega^{\pm}) \geq \delta(M_0) > 0$, then

$$\int_{\partial B_s(p)} |v|^2\, d\sigma \geq \int_{\partial B_s(p)\cap B_{\delta s}(x_{\max}(p,s))} |v|^2\, d\sigma \geq C(M_0)|v(x_{\max}(p,s))|^2(\delta s)^{n-1} \geq C(n, M_0)\frac{\omega^-(B_s(Q))^2}{s^{n-3}}.$$

If this can be shown, then recalling the doubling of harmonic measure on NTA domains, the above string of inequalities proves that $H(p,s,v_\epsilon)$, $H(Q,s,v_\epsilon)$, and $H\big(Q,\frac{1}{2}s,v_\epsilon\big)$ share a common lower bound.

The common upper bound follows from a similar argument using Remark 3.1. That is, if we can show, for all p and all $0 < s \leq \frac{1}{2}$, that $\mathrm{dist}(x_{\max}(p,s), \partial\Omega^{\pm}) \geq \delta(\alpha, M_0, \Gamma) > 0$, then by Harnack chains we know that

$$|v(x_{\max}(Q,s)^+)| \sim_{\alpha, M_0, \Gamma} \frac{\omega^+(B_s(Q))}{s^{n-2}}$$

and that

$$\begin{aligned}\int_{\partial B_s(Q)} |v|^2\, d\sigma &\leq |v(x_{\max}(Q,s))^-|^2 r^{n-1} + |v(x_{\max}(Q,s))^+|^2 r^{n-1}\\ &\leq C(M_0)(|v(A_s(Q)^-)|^2 r^{n-1} + |v(A_s(Q)^+)|^2 r^{n-1}) \leq C(n,\alpha,M_0,\Gamma)\frac{\omega^-(B_s(Q))^2}{s^{n-3}}.\end{aligned}$$

Recalling the doubling of harmonic measure on NTA domains, the above string of inequalities proves that $H(p,s,v_\epsilon)$, $H(Q,s,v_\epsilon)$, and $H\big(Q,\frac{1}{2}s,v_\epsilon\big)$ share a common lower bound. This would prove the lemma.

Let Q, p, and s be given. By the maximum principle for harmonic functions applied to v^- in Ω^-, we have $|v(x_{\max}(p,s)^{\pm})| \geq \big|v\big(x_{\max}\big(Q,\frac{1}{2}s\big)^{\pm}\big)\big|$. By NTA estimates [Engelstein 2016, Lemma 5.4], we have

$$\frac{\omega^{\pm}(B_{1/2s}(Q))}{\big(\frac{1}{2}s\big)^{n-2}} \sim_{M_0} |v(A^{\pm}_{s/2}(Q))| \leq \big|v\big(x_{\max}\big(Q,\tfrac{1}{2}s\big)^{\pm}\big)\big| \leq |v(x_{\max}(p,s)^{\pm})|.$$

Therefore, by the uniform Lipschitz estimates of Theorem 2.6 and Remark 3.1 we infer that

$$\mathrm{dist}(x_{\max}(p,s)^{\pm}, \partial\Omega^{\pm}) \gtrsim_{M_0,\Gamma,\alpha} s.$$

Therefore, we may use Harnack chains and estimate

$$|v(x_{\max}(p,s)^{\pm})| \leq |v(A^{\pm}_{2s}(Q))| \sim_{M_0} \frac{\omega^{\pm}(B_{2s}(Q))}{(2s)^{n-2}}.$$

Thus, by the doubling of harmonic measure on NTA domains (see [Jerison and Kenig 1982]), we infer that $|v(x_{\max}(p,s)^{\pm})| \sim_{M_0} |v(A_{2s}(Q)^{\pm})|$. This proves the lemma. □

Remark 7.2. As a consequence of Lemma 7.1 and Corollary 4.4, we observe that if $v \in \mathcal{A}(n,\alpha,M_0)$ then, for every $0 < r \leq \frac{1}{2}$ and every point $p \in B_1(0)$ such that $\mathrm{dist}(p, \partial\Omega^{\pm}) \leq \frac{1}{3}r$ and for any $0 < \epsilon \ll r$,

$$N(p,r,v) \leq C(n,\alpha,M_0,\Gamma) \quad \text{and} \quad N(p,r,v_\epsilon) \leq C(n,\alpha,M_0,\Gamma).$$

Lemma 7.3. *Let $v \in \mathcal{A}(n,\alpha,M_0)$ with $\|\ln(h)\|_\alpha \leq \Gamma$, and let $Q \in B_1(0)\cap\partial\Omega^{\pm}$, $0 < s \leq 1$, and $\epsilon \ll s$. For all $p \in B_{s/3}(Q)\cap\overline{\Omega}^-$ and all vectors $|\vec{v}| \leq r$,*

$$\left|\int_{B_s(p)} \langle\nabla v_\epsilon, \vec{v}\rangle \Delta v_\epsilon\, dV(x)\right| \leq C\|\ln(h)\|_\alpha s^\alpha \frac{\omega^-(B_s(Q))^2}{s^{n-2}}|\vec{v}|, \tag{7-1}$$

where $C = C(M_0)$.

Proof. Let p, Q, and s be as above. For sufficiently small $0<\epsilon$,

$$\begin{aligned}
&\left|\int_{B_s(p)} \langle \nabla v_\epsilon, \vec{v}\rangle \Delta v_\epsilon \, dV(x)\right| \\
&= \left|\int_{B_s(p)} \langle \nabla v_\epsilon, \vec{v}\rangle_\epsilon \Delta v(x)\right| \le \int_{B_s(p)} |(\langle \nabla v_\epsilon, \vec{v}\rangle)_\epsilon| \left|\frac{h(0)}{h(x)} - 1\right| d\omega^- \le \int_{B_{2s}(Q)} |(\langle \nabla v_\epsilon, \vec{v}\rangle)_\epsilon| \left|\frac{h(0)}{h(x)} - 1\right| d\omega^- \\
&\le \|\ln(h)\|_\alpha (2s)^\alpha \int_{B_{2s}(Q)} |(\langle \nabla v_\epsilon, \vec{v}\rangle)_\epsilon| \, d\omega^- \le \|\ln(h)\|_\alpha (2s)^{\alpha+1} |\vec{v}| \int_{B_{2s}(Q)} |\nabla v_\epsilon|_\epsilon \, d\omega^-.
\end{aligned}$$

Chasing through the change of variables $x = ry + Q$, we see that

$$\nabla_x v(x) = \frac{1}{r}\nabla_y v(ry+Q) = \frac{\omega^-(B_r(Q))}{r^{n-1}} \nabla_y v_r(y).$$

Thus, we calculate that for the change of variables $x = 2sy + Q$,

$$\begin{aligned}
\left|\int_{B_s(p)} \langle \nabla v_\epsilon, \vec{v}\rangle \Delta v_\epsilon \, dV(x)\right| &\le \|\ln(h)\|_\alpha (2s)^{\alpha+1} |\vec{v}| \frac{\omega^-(B_{2s}(Q))^2}{(2s)^{n-1}} \int_{B_1(0)} |\nabla v_{Q,2s} \star \phi_{\epsilon/(2s)}| \star \phi_{\epsilon/(2s)} \, d\omega^-_{Q,2s} \\
&\le \|\ln(h)\|_\alpha C s^{\alpha+1} |\vec{v}| \frac{\omega^-(B_{2s}(Q))^2}{(s)^{n-1}} \omega^-_{Q,2s}(B_2(0)) \\
&\le \|\ln(h)\|_\alpha C s^\alpha \frac{\omega^-(B_s(Q))^2}{s^{n-2}} |\vec{v}|,
\end{aligned}$$

where the last two inequalities are because the $v_{Q,r}$ are uniformly locally Lipschitz, $1+\epsilon/r<2$, the $\omega^-_{Q,r}(B_2(0))$ are uniformly bounded for $Q \in B_1(0)$ and $r<2$, and the doubling of harmonic measure on NTA domains. □

Lemma 7.4. *Let $v \in \mathcal{A}(n, \alpha, M_0)$, $Q \in \partial\Omega^\pm \cap B_1(0)$, and $0<r$. Then for $p \in B_{r/3}(Q)$ and $\vec{v} \in \mathbb{R}^n$ such that $|\vec{v}| \le r$, we calculate the spatial directional derivatives as follows*:

$$\frac{\partial}{\partial \vec{v}} H(p,r,v_\epsilon) = 2\int_{\partial B_r(p)} v_\epsilon \nabla v_\epsilon \cdot \vec{v} \, d\sigma, \tag{7-2}$$

$$\frac{\partial}{\partial \vec{v}} D(p,r,v_\epsilon) = 2\int_{\partial B_r(p)} (\nabla v_\epsilon \cdot \vec{v})(\nabla v_\epsilon \cdot \eta) \, d\sigma - \int_{B_r(p)} \frac{\partial}{\partial \vec{v}} v_\epsilon \Delta v_\epsilon \, dV, \tag{7-3}$$

$$\frac{\partial}{\partial \vec{v}} N(p,r,v_\epsilon) = \frac{2}{H(p,r,v_\epsilon)}\left(\int_{\partial B_r(p)} (r\nabla v_\epsilon \cdot \eta - N(p,r,v_\epsilon) v_\epsilon)(\nabla v_\epsilon \cdot \vec{v}) \, d\sigma\right) - \frac{r\int_{B_r(p)} \frac{\partial}{\partial \vec{v}} v_\epsilon \Delta v_\epsilon \, dV}{H(p,r,v_\epsilon)}. \tag{7-4}$$

Proof. Equation (7-4) follows immediately from the preceding equations. The spatial derivative for $H(Q,r,u)$ follows from differentiating inside the integral. To obtain the spatial derivative for $D(Q,r,v)$, we recall the divergence theorem:

$$\begin{aligned}
\frac{\partial}{\partial \vec{v}} D(p,r,v) &= \frac{\partial}{\partial \vec{v}}\left(\int_{\partial B_r(p)} v \nabla v \cdot \eta \, d\sigma(x) - \int_{B_r(p)} v_\epsilon \Delta v_\epsilon \, dV\right) \\
&= \int_{\partial B_r(p)} (\nabla v_\epsilon \cdot \vec{v})(\nabla v_\epsilon \cdot \eta) \, d\sigma + \int_{\partial B_r(p)} v_\epsilon \frac{\partial}{\partial \vec{v}} (\nabla v_\epsilon \cdot \eta) \, d\sigma.
\end{aligned}$$

Now, we focus upon the last term. Recalling Green's theorem and the fact that partial derivatives of harmonic functions are themselves harmonic,

$$\begin{aligned}\int_{\partial B_r(p)} v_\epsilon \frac{\partial}{\partial \vec{v}}(\nabla v_\epsilon \cdot \eta)\, d\sigma &= \int_{\partial B_r(p)} v_\epsilon \nabla\Big(\frac{\partial}{\partial \vec{v}} v_\epsilon\Big) \cdot \eta \, d\sigma = \int_{\partial B_r(p)} \nabla v_\epsilon \cdot \eta \frac{\partial}{\partial \vec{v}} v_\epsilon \, d\sigma - \int_{B_r(p)} \frac{\partial}{\partial \vec{v}} v_\epsilon \Delta v_\epsilon \, dV \\ &= \int_{\partial B_r(p)} (\nabla v_\epsilon \cdot \vec{v})(\nabla v_\epsilon \cdot \eta)\, d\sigma - \int_{B_r(p)} \frac{\partial}{\partial \vec{v}} v_\epsilon \Delta v_\epsilon \, dV. \qquad \square\end{aligned}$$

Definition 7.5. For the sake of concision, we define the following notation for $v \in \mathcal{A}(n, \alpha, M_0)$, $y \in \overline{\Omega}$, and radii $0 < r, R \leq 2$.

$$E_y(z) := \nabla v_\epsilon(z) \cdot (z - y) - N(y, |z-y|, v_\epsilon) v_\epsilon(z), \tag{7-5}$$

$$W_{r,R}(y) := N(y, R, v_\epsilon) - N(y, r, v_\epsilon). \tag{7-6}$$

Lemma 7.6. *Let $v \in \mathcal{A}(n, \alpha, M_0)$ with $\|\ln(h)\|_\alpha \leq \Gamma$. Let $Q \in \partial\Omega^\pm \cap B_1(0)$ and $0 < r \leq 1$. Let $p \in [Q, Q']$ with $Q' \in \partial\Omega^\pm \cap B_{r/3}(Q)$. Then, for $\vec{v} = Q' - Q$ and $0 < \epsilon \ll r$,*

$$\begin{aligned}\Big|\frac{\partial}{\partial \vec{v}} N(p, r, v_\epsilon)\Big| \lesssim_{n,\alpha,M_0,\Gamma} & \; 2(W_{r/2,2r}(Q) + W_{r/2,2r}(Q') + C\Gamma r^\alpha)\bigg(r\bigg(\frac{\int_{\partial B_r(p)} |\nabla v_\epsilon|^2 \, d\sigma}{H(p, r, v_\epsilon)}\bigg)^{\frac{1}{2}} + 1\bigg) \\ & + \frac{2}{H(p, r, v_\epsilon)}\bigg(\int_{\partial B_r(p)} |E_Q(z)|^2 + |E_{Q'}(z)|^2 \, d\sigma\bigg)^{\frac{1}{2}}\bigg(\int_{\partial B_r(p)} (\nabla v_\epsilon(z) \cdot (z-p))^2 \, d\sigma\bigg)^{\frac{1}{2}} \\ & + C(n, \Lambda)\bigg(\frac{1}{H(p, r, v_\epsilon)} \int_{\partial B_r(p)} |E_Q(z)|^2 + |E_{Q'}(z)|^2 \, d\sigma\bigg)^{\frac{1}{2}} \\ & + 2(N(Q, r, v_\epsilon) - N(Q', r, v_\epsilon))\bigg(\frac{\int_{B_r(p)} v_\epsilon \Delta v_\epsilon \, dV}{H(p, r, v_\epsilon)}\bigg) + r^\alpha \Gamma.\end{aligned}$$

Proof. We begin by noting that Lemmas 7.3 and 7.1 give

$$\bigg|\frac{r \int_{B_r(p)} \frac{\partial}{\partial \vec{v}} v_\epsilon \Delta v_\epsilon \, dV}{H(p, r, v_\epsilon)}\bigg| \leq C(n, \alpha, M_0, \Gamma) r^\alpha \Gamma.$$

Now, we write the decomposition

$$\begin{aligned}\nabla v_\epsilon \cdot (Q - Q') &= \nabla v_\epsilon \cdot (z - Q') - \nabla v_\epsilon \cdot (z - Q) \\ &= (N(Q, |z-Q|, v_\epsilon) - N(Q', |z-Q'|, v_\epsilon)) v_\epsilon + (E_Q(z) - E_{Q'}(z)).\end{aligned}$$

Therefore, plugging this into (7-4), we obtain for $v = Q' - Q$,

$$\begin{aligned}&\frac{2}{H(p, r, v_\epsilon)}\bigg(\int_{\partial B_r(p)} \Big(r \nabla v_\epsilon \cdot \eta - N(p, r, v_\epsilon) v_\epsilon\Big)(\nabla v_\epsilon \cdot \vec{v})\, d\sigma\bigg) \\ &\quad = \frac{2}{H(p, r, v_\epsilon)}\bigg(\int_{\partial B_r(p)} E_p(z)([N(Q, |z-Q|, v_\epsilon) - N(Q', |z-Q'|, v_\epsilon)] v_\epsilon + (E_Q(z) - E_{Q'}(z)))\, d\sigma\bigg) \\ &\quad = A + B - C,\end{aligned}$$

where for the purposes of this lemma

$$A := \frac{2}{H(p,r,v_\epsilon)} \int_{\partial B_r(p)} E_p(z)(N(Q,|z-Q|,v_\epsilon) - N(Q',|z-Q'|,v_\epsilon))v_\epsilon \, d\sigma,$$
$$B := \frac{2}{H(p,r,v_\epsilon)} \int_{\partial B_r(p)} \nabla v_\epsilon(z)\cdot(z-p)(E_Q(z) - E_{Q'}(z))\, d\sigma,$$
$$C := \frac{2}{H(p,r,v_\epsilon)} \int_{\partial B_r(p)} N(p,|z-p|,v_\epsilon)v_\epsilon(z)(E_Q(z) - E_{Q'}(z))\, d\sigma.$$

We begin by estimating A. We rewrite

$$N(Q,|z-Q|,v_\epsilon) - N(Q',|z-Q'|,v_\epsilon) = N(Q,r,v_\epsilon) - N(Q',r,v_\epsilon) - W_{|z-Q|,r}(Q) + W_{|z-Q'|,r}(Q').$$

Note that if $|Q-Q'| \le \frac{1}{3}r$ and $p \in [Q,Q']$, then $\frac{1}{2}r \le |z-x_i| \le 2r$ for $x_i \in \{Q,Q'\}$ and all $z \in \partial B_r(p)$. Therefore, by Lemma 4.10, for all $z \in \partial B_r(p)$,

$$|W_{|z-Q|,r}(Q)| \le W_{r/2,2r}(Q) + 2C_1\Gamma(2r)^\alpha,$$
$$|W_{|z-Q'|,r}(Q')| \le W_{r/2,2r}(Q') + 2C_1\Gamma(2r)^\alpha.$$

Furthermore, we estimate by the divergence theorem

$$\begin{aligned}\int_{\partial B_r(p)} & E_p(z)(N(Q,r,v_\epsilon) - N(Q',r,v_\epsilon))v_\epsilon \, d\sigma \\ &= (N(Q,r,v_\epsilon) - N(Q',r,v_\epsilon)) \int_{\partial B_r(p)} v_\epsilon \nabla v_\epsilon(z)\cdot(z-y) - N(p,|z-p|,v_\epsilon)v_\epsilon^2 \, d\sigma \\ &= (N(Q,r,v_\epsilon) - N(Q',r,v_\epsilon)) \cdot \left(\int_{B_r(p)} v_\epsilon \Delta v_\epsilon \, dV\right).\end{aligned}$$

Thus, we may give the following preliminary estimate on A:

$$\begin{aligned}|A| \le (W_{r/2,2r}(Q) + W_{r/2,2r}(Q') + C\Gamma r^\alpha) &\frac{2}{H(p,r,v_\epsilon)} \int_{\partial B_r(p)} |E_p(z)||v_\epsilon| \, d\sigma \\ &+ 2(N(Q,r,v_\epsilon) - N(Q',r,v_\epsilon))\left(\frac{\int_{B_r(p)} v_\epsilon \Delta v_\epsilon \, dV}{H(p,r,v_\epsilon)}\right).\end{aligned}$$

Focusing upon the term

$$\frac{2}{H(p,r,v_\epsilon)} \int_{\partial B_r(p)} |E_p(z)||v_\epsilon| \, d\sigma,$$

using Remark 7.2 and Cauchy–Schwartz we estimate

$$\begin{aligned}\frac{2}{H(p,r,v_\epsilon)} \int_{\partial B_r(p)} |E_p(z)||v_\epsilon| \, d\sigma &\le \frac{2}{H(p,r,v_\epsilon)} \int_{\partial B_r(p)} r|v_\epsilon \nabla v_\epsilon \cdot \eta| + N(p,r,v_\epsilon)|v_\epsilon|^2 \, d\sigma \\ &\le 2C\left(r\left(\frac{\int_{\partial B_r(p)} |\nabla v_\epsilon|^2 \, d\sigma}{H(p,r,v_\epsilon)}\right)^{\frac{1}{2}} + 1\right).\end{aligned}$$

Now we estimate B using Cauchy–Schwartz:

$$\left(\frac{2}{H(p,r,v_\epsilon)}\int_{\partial B_r(p)} \nabla v_\epsilon(z)\cdot(z-p)(E_Q(z)-E_{Q'}(z))\,d\sigma\right)^2$$
$$\leq \frac{4}{H(p,r,v_\epsilon)^2}\int_{\partial B_r(p)}(E_Q(z)-E_{Q'}(z))^2\,d\sigma\int_{\partial B_r(p)}(\nabla v_\epsilon(z)\cdot(z-p))^2\,d\sigma$$
$$\leq \frac{8}{H(p,r,v_\epsilon)^2}\int_{\partial B_r(p)}|E_Q(z)|^2+|E_{Q'}(z)|^2\,d\sigma\left(\int_{\partial B_r(p)}(\nabla v_\epsilon(z)\cdot(z-p))^2\,d\sigma\right).$$

For $|C|$, the same Cauchy–Schwartz argument plus Corollary 4.4 shows that

$$|C| \lesssim_{n,\alpha,M_0,\Gamma}\left(\frac{1}{H(p,r,v_\epsilon)}\int_{\partial B_r(p)}|E_Q(z)|^2+|E_{Q'}(z)|^2\,d\sigma\right)^{\frac{1}{2}}.$$

This proves the pointwise estimate

$$\frac{\partial}{\partial v}N(p,r,v_\epsilon) \lesssim_{n,\alpha,M_0,\Gamma} (W_{r/2,2r}(Q)+W_{r/2,2r}(Q')+C\Gamma r^\alpha)2\left(r\left(\frac{\int_{\partial B_r(p)}|\nabla v_\epsilon|^2\,d\sigma}{H(p,r,v_\epsilon)}\right)^{\frac{1}{2}}+1\right)$$
$$+\frac{2}{H(p,r,v_\epsilon)}\left(\int_{\partial B_r(p)}|E_Q(z)|^2+|E_{Q'}(z)|^2\,d\sigma\right)^{\frac{1}{2}}\left(\int_{\partial B_r(p)}(\nabla v_\epsilon(z)\cdot(z-p))^2\,d\sigma\right)^{\frac{1}{2}}$$
$$+\left(\frac{1}{H(p,r,v_\epsilon)}\int_{\partial B_r(p)}|E_Q(z)|^2+|E_{Q'}(z)|^2\,d\sigma\right)^{\frac{1}{2}}$$
$$+2(N(Q,r,v_\epsilon)-N(Q',r,v_\epsilon))\left(\frac{\int_{B_r(p)}v_\epsilon\Delta v_\epsilon\,dV}{H(p,r,v_\epsilon)}\right)+r^\alpha\Gamma.$$

We prove the lemma by reversing the roles of Q and Q'. □

Corollary 7.7. *Let $v\in\mathcal{A}(n,\alpha,M_0)$ with $\|\ln(h)\|_\alpha\leq\Gamma$, let $Q\in\partial\Omega^\pm\cap B_1(0)$ and $0<r\leq 1$, and let $Q'\in\partial\Omega^\pm\cap B_{r/3}(Q)$. Then*

$$|N(Q',r,v)-N(Q,r,v)| \lesssim_{n,\alpha,M_0,\Gamma} W_{r/2,2r}(Q)+W_{r/2,2r}(Q')+C\Gamma r^\alpha+1+r^{-1/2}.$$

Proof. First, since for any $0\neq c$ we have $N(Q,r,v)=N(Q,r,cv)$, we shall assume for the purposes of this lemma that $v=T_{0,1}v$. We shall show that $|N(Q',r,v_\epsilon)-N(Q,r,v_\epsilon)|$ satisfies a corresponding inequality, and let $\epsilon\to 0$. Since $N(Q',r,v_\epsilon)\to N(Q',r,v)$ as $\epsilon\to 0$, this will prove the claim.

Let $\vec{v}=Q'-Q$ and $p_t:=Q+t\vec{v}$. Then we calculate

$$|N(Q',r,v_\epsilon)-N(Q,r,v_\epsilon)|\leq\int_0^1\left|\frac{\partial}{\partial t}N(p_t,r,v_\epsilon)\right|dt \lesssim_{n,\alpha,M_0,\Gamma} A+B+C+D+E,$$

where

$$A:=\int_0^1(W_{r/2,2r}(Q)+W_{r/2,2r}(Q')+C\Gamma r^\alpha)2\left(r\left(\frac{\int_{\partial B_r(p_t)}|\nabla v_\epsilon|^2\,d\sigma}{H(p_t,r,v_\epsilon)}\right)^{\frac{1}{2}}+1\right)dt,$$

$$B:=\int_0^1\frac{2}{H(p_t,r,v_\epsilon)}\left(\int_{\partial B_r(p_t)}|E_Q(z)|^2+|E_{Q'}(z)|^2\,d\sigma\right)^{\frac{1}{2}}\left(\int_{\partial B_r(p_t)}(\nabla v_\epsilon(z)\cdot(z-p_t))^2\,d\sigma\right)^{\frac{1}{2}}dt,$$

$$C := \int_0^1 \left(\frac{1}{H(p_t, r, v_\epsilon)} \int_{\partial B_r(p_t)} |E_Q(z)|^2 + |E_{Q'}(z)|^2 \, d\sigma \right)^{\frac{1}{2}} dt,$$
$$D := 2(N(Q, r, v_\epsilon) - N(Q', r, v_\epsilon)) \int_0^1 \left(\frac{\int_{B_r(p_t)} v_\epsilon \Delta v_\epsilon \, dV}{H(p_t, r, v_\epsilon)} \right) dt,$$
$$E := C\Gamma r^\alpha.$$

We estimate each term separately.

Bounding A. We begin by rewriting A:

$$A = 2(W_{r/2,2r}(Q) + W_{r/2,2r}(Q') + C\Gamma r^\alpha) \int_0^1 \left(r \left(\frac{\int_{\partial B_r(p_t)} |\nabla v_\epsilon|^2 \, d\sigma}{H(p_t, r, v_\epsilon)} \right)^{\frac{1}{2}} + 1 \right) dt.$$

Observe that by Lemma 7.1 and Remark 4.12 we may using Hölder's inequality to estimate

$$\int_0^1 r \left(\frac{\int_{\partial B_r(p_t)} |\nabla v_\epsilon|^2 \, d\sigma}{H(p_t, r, v_\epsilon)} \right)^{\frac{1}{2}} dt \lesssim_{n,\alpha,M_0,\Gamma} \int_0^1 r \left(\frac{\int_{\partial B_r(p_t)} |\nabla v_\epsilon|^2 \, d\sigma}{H(Q, r, v_\epsilon)} \right)^{\frac{1}{2}} dt$$
$$\lesssim_{n,\alpha,M_0,\Gamma} r \frac{2}{H(Q, r, v_\epsilon)^{1/2}} \left(\int_0^1 \int_{\partial B_r(p_t)} |\nabla v_\epsilon|^2 \, d\sigma \, dt \right)^{\frac{1}{2}}.$$

Now, divide the spheres as follows: $\partial B_r(p_t) = \partial B_r(p_t)^+ \cup \partial B_r(p_t^-)$, where

$$\partial B_r(p_t)^- = \{x \in \partial B_r(p_t) : (x - p_t) \cdot \vec{v} < 0\} \quad \text{and} \quad \partial B_r(p_t)^+ = \{x \in \partial B_r(p_t) : (x - p_t) \cdot \vec{v} \geq 0\}.$$

Notice that

$$\max_{z \in \bigcup_{t\in[0,1]} \partial B_r(p_t)} \#\{t \in [0, 1] : z \in \partial B_r(p_t)^+ \text{ or } z \in \partial B_r(p_t^-)\} = 2.$$

Then, use the coarea formula for the function $\phi^\pm : \bigcup_{t\in[0,1]} \partial B_r(p_t)^\pm \to \mathbb{R}$ defined by $\phi|_{\partial B_r(p_t)^\pm} = t$. Note that if we write $L := Q + \operatorname{span}\{\vec{v}\}$ and $\operatorname{dist}(z, L) = \delta$, then

$$J\phi^\pm(z) = |\nabla \phi(z)| = \frac{r}{|Q - Q'|\sqrt{r^2 - \delta^2}} = \frac{1}{|Q - Q'| \cos(\theta(z))},$$

where

$$\theta(z) = \frac{z - p_t}{|z - p_t|} \cdot \frac{\vec{v}}{|\vec{v}|} \quad \text{for } z \in \partial B_r(p_t)^\pm.$$

Thus, we obtain

$$\int_0^1 r \left(\frac{\int_{\partial B_r(p_t)} |\nabla v_\epsilon|^2 \, d\sigma}{H(p_t, r, v_\epsilon)} \right)^{\frac{1}{2}} dt \lesssim_{n,\alpha,M_0,\Gamma} \frac{2r}{H(Q, r, v_\epsilon)^{1/2}} \left(\int_{\bigcup_{t\in[0,1]} \partial B_r(p_t)} \frac{2|\nabla v_\epsilon|^2}{|Q - Q'||\cos(\theta(z))|} \, dV \right)^{\frac{1}{2}}.$$

Note that a simple calculation gives, for any $1 \leq p < 2$,

$$\int_{\bigcup_{t\in[0,1]} \partial B_r(p_t)} |Q - Q'|^{-1} |\cos(\theta(z))|^{-p} \, dV = \int_0^r \int_{\mathscr{C}_\delta} |Q - Q'|^{-1} |\cos(\theta(z))|^{-p} \, d\mathcal{H}^{n-1} \, d\delta$$
$$\leq c(n) \int_0^r \frac{r^p \delta^{n-2}}{(r^2 - \delta^2)^{p/2}} \, d\delta \leq \frac{c(n)}{1 - \frac{1}{2}p} r^{n-1}. \tag{7-7}$$

Since $T_{Q,r}v_\epsilon$ is uniformly locally Lipschitz by Lemma 3.4, recalling Definition 2.7 and choosing $p=1$ above, we see

$$\begin{aligned}
r\frac{2}{H(Q,r,v_\epsilon)^{1/2}}&\left(2\int_{\bigcup_{t\in[0,1]}\partial B_r(p_t)}\frac{|\nabla v_\epsilon|^2}{|Q-Q'|\,|\cos(\theta(z))|}\,dV\right)^{\frac12}\\
&= 2\sqrt{2}\left(r^{1-n}\int_{\bigcup_{t\in[0,1]}\partial B_1(T_{Q,r}p_t)}\frac{|\nabla T_{Q,r}v_\epsilon|^2}{|Q-Q'|\,|\cos(\theta(z))|}\,dV\right)^{\frac12}\\
&= 2\sqrt{2}C\left(r^{1-n}\int_{\bigcup_{t\in[0,1]}\partial B_r(p_t)}|Q-Q'|^{-1}|\cos(\theta(z))|^{-1}\,dV\right)^{\frac12}\\
&\le C(n,\alpha,M_0,\Gamma).
\end{aligned}$$

Thus

$$|A|\lesssim_{n,\alpha,M_0,\Gamma}(W_{r/2,2r}(Q)+W_{r/2,2r}(Q')+C\Gamma r^\alpha).$$

Bounding C. By Hölder's inequality (or Jensen's inequality for concave functions) and Lemma 7.1, we may reduce to considering

$$\begin{aligned}
\int_0^1\frac{1}{H(p_t,r,v_\epsilon)}&\int_{\partial B_r(p_t)}|\nabla v_\epsilon(z)\cdot(z-Q)-N(Q,|z-Q|,v_\epsilon)v_\epsilon(z)|^2\,d\sigma(z)\,dt\\
&\lesssim_{n,\alpha,M_0,\Gamma}\int_0^1\frac{1}{H\big(Q,\frac12 r,v_\epsilon\big)}\int_{\partial B_r(p_t)}|\nabla v_\epsilon(z)\cdot(z-Q)-N(Q,|z-Q|,v_\epsilon)v_\epsilon(z)|^2\,d\sigma(z)\,dt.
\end{aligned}$$

Now, we change variables using Definition 2.7 and Lemma 7.1, and use Young's inequality to get

$$\begin{aligned}
\int_0^1\frac{1}{H\big(Q,\frac12 r,v_\epsilon\big)}&\int_{\partial B_r(p_t)}|\nabla v_\epsilon(z)\cdot(z-Q)-N(Q,|z-Q|,v_\epsilon)v_\epsilon(z)|^2\,d\sigma(z)\,dt\\
&\le 2\int_{\bigcup_{t\in[0,1]}\partial B_r(p_t)}\frac{|\nabla v_\epsilon(z)\cdot(z-Q)-N(Q,|z-Q|,v_\epsilon)v_\epsilon(z)|^2}{H\big(Q,\frac12 r,v_\epsilon\big)|Q-Q'|\,|\cos(\theta(z))|}\,dV\\
&\lesssim_{n,\alpha,M_0,\Gamma}2r^n\int_{\bigcup_{t\in[0,1]}\partial B_1(T_{Q,r}p_t)}\frac{|\nabla T_{Q,r}v_\epsilon(z)\cdot z-N(0,|z|,T_{Q,r}v_\epsilon)T_{Q,r}v_\epsilon(z)|^2}{r^{n-1}\left|\frac{Q-Q'}{r}\right|\,|\cos(\theta(z))|}\,dV\\
&\le 4\int_{\bigcup_{t\in[0,1]}\partial B_1(T_{Q,r}p_t)}\frac{|\nabla T_{Q,r}v_\epsilon(z)\cdot z|^2+|N(0,|z|,T_{Q,r}v_\epsilon)T_{Q,r}v_\epsilon(z)|^2}{\left|\frac{Q-Q'}{r}\right|\,|\cos(\theta(z))|}\,dV.
\end{aligned}$$

Now, by Corollary 4.4 and Lemma 3.4, the numerator is bounded by a constant. Whence, by a calculation similar to (7-7), we obtain

$$\int_0^1\frac{1}{H\big(Q,\frac12 r,v_\epsilon\big)}\int_{\partial B_r(p_t)}|\nabla v_\epsilon(z)\cdot(z-Q)-N(Q,|z-Q|,v_\epsilon)v_\epsilon(z)|^2\,d\sigma(z)\,dt\lesssim_{n,\alpha,M_0,\Gamma}1.$$

An identical argument holds for Q' in the place of Q. Thus, we have that $|C|\lesssim_{n,\alpha,M_0,\Gamma}1$.

Bounding B. Using Cauchy–Schwartz, Lemma 7.1, and the estimates of the term $|C|$ above, we obtain

$$B = \int_0^1 \frac{2}{H(p_t, r, v_\epsilon)} \left(\int_{\partial B_r(p_t)} |E_Q(z)|^2 + |E_{Q'}(z)|^2 \, d\sigma\right)^{\frac{1}{2}} \left(\int_{\partial B_r(p_t)} (\nabla v_\epsilon(z) \cdot (z - p_t))^2 \, d\sigma\right)^{\frac{1}{2}} dt$$
$$\lesssim_{n,\alpha,M_0,\Gamma} \left(\frac{r^2}{H(Q, r, v_\epsilon)} \int_{\bigcup_{t\in[0,1]} \partial B_1((p_t - Q_i)/r_i)} \frac{|\nabla v_\epsilon(z)|^2}{|Q - Q'| \, |\cos(\theta_{\vec{v}})|} \, dV\right)^{\frac{1}{2}}$$
$$\lesssim_{n,\alpha,M_0,\Gamma} r^{-1/2} \left(\int_{\bigcup_{t\in[0,1]} \partial B_1((p_t - Q_i)/r_i)} \frac{|\nabla T_{Q,r} v_\epsilon(z)|^2}{\left|\frac{Q-Q'}{r}\right| |\cos(\theta_{\vec{v}})|} \, dV\right)^{\frac{1}{2}}.$$

Thus, by Lemma 3.4 and a calculation identical to that of (7-7), we obtain $|B| \lesssim_{n,\alpha,M_0,\Gamma} r^{-1/2}$.

Bounding D. Note that

$$\left|\int_{B_r(p_t)} v_\epsilon \Delta v_\epsilon \, dV(x)\right| \le \int_{B_r(p)} |(v_\epsilon)_\epsilon| \left|\frac{h(0)}{h(x)} - 1\right| d\omega^- \le \int_{B_{2r}(Q)} |(v_\epsilon)_\epsilon| \left|\frac{h(0)}{h(x)} - 1\right| d\omega^-$$
$$\le \|\ln(h)\|_\alpha (2r)^\alpha \int_{B_{2r}(Q)} |(v_\epsilon)_\epsilon| \, d\omega^- \le \|\ln(h)\|_\alpha (2r)^\alpha C \int_{B_{2r}(Q)} |(v_\epsilon)_\epsilon| \, d\omega^-.$$

Chasing through the change of variables $x = ry + Q$, we see that

$$\nabla_x v(x) = \frac{1}{r} \nabla_y v(ry + Q) = \frac{\omega^-(B_r(Q))}{r^{n-1}} \nabla_y v_r(y).$$

Thus, we calculate that, for the change of variables $x = 2ry + Q$,

$$\left|\int_{B_r(p)} v_\epsilon \Delta v_\epsilon \, dV(x)\right| \le \|\ln(h)\|_\alpha (2r)^\alpha \frac{\omega^-(B_{2r}(0))^2}{(2r)^{n-1}} \int_{B_1(0)} |v_{Q,2r} \star \phi_{\epsilon/(2r)}| \star \phi_{\epsilon/(2r)} \, d\omega^-_{2r}$$
$$\le \|\ln(h)\|_\alpha C r^\alpha \frac{\omega^-(B_{2r}(0))^2}{r^{n-1}} C\left(\frac{\epsilon}{r}\right) \omega^-_{Q,2r}(B_2(0))$$
$$\le \|\ln(h)\|_\alpha C r^\alpha C\left(\frac{\epsilon}{r}\right) \frac{\omega^-(B_r(0))^2}{r^{n-2}},$$

where the last two inequalities are because the $v_{Q,r}$ are uniformly locally Lipschitz, $1 + \epsilon/r < 2$, the $\omega^-_{Q,r}(B_2(0))$ are uniformly bounded for $Q \in B_1(0)$ and $r < 2$, and the doubling of harmonic measure on NTA domains.

Thus, by Lemma 7.1 we have

$$2(N(Q, r, v_\epsilon) - N(Q', r, v_\epsilon)) \int_0^1 \left(\frac{\int_{B_r(p_t)} v_\epsilon \Delta v_\epsilon \, dV}{H(p_t, r, v_\epsilon)}\right) dt \le C(n, \alpha, M_0, \Gamma) \Gamma r^{\alpha-1} \left(\frac{\epsilon}{r}\right).$$

Letting $\epsilon \to 0$, we see that D vanishes.

Thus, putting together the estimates for A, B, C, D we have

$$|N(Q', r, v) - N(Q, r, v)| \le \lim_{\epsilon\to 0} |N(Q', r, v_\epsilon) - N(Q, r, v_\epsilon)|$$
$$\lesssim_{n,\alpha,M_0,\Gamma} W_{r/2,2r}(Q) + W_{r/2,2r}(Q') + \Gamma r^\alpha + 1 + r^{-1/2}.$$

This proves the lemma. □

8. Frequency pinching

In this section, we prove a "frequency pinching" result (Lemma 8.2) in the style of [De Lellis et al. 2018]. This kind of result relates Jones' beta numbers to the drop in Almgren frequency.

Definition 8.1 (Jones' beta numbers). For μ a Borel measure, we define $\beta^k_{\mu,2}(Q,r)^2$ as follows:

$$\beta^k_{\mu,2}(Q,r)^2 = \inf_{L^k} \frac{1}{r^k}\int_{B_r(p)} \frac{\operatorname{dist}(x,L)^2}{r^2}\,d\mu(x),$$

where the infimum is taken over all affine k-planes.

Taking the *infimum* here — as opposed to the *minimum* — is a convention. The space of admissible planes is compact, so a minimizing plane exists. Let $V^k_\mu(Q,r)$ denote a k-plane which minimizes the *infimum* in the definition of $\beta^k_\mu(Q,r)^2$. Note that this k-plane is not a priori unique.

Lemma 8.2 (frequency pinching). *There exists a constant* $\delta_0 = \delta_0(n,\alpha,M_0,\Gamma) > 0$ *such that, for any* $0<\delta\le\delta_0$, *if* $v\in\mathcal{A}(n,\alpha,M_0)$ *with* $\|\ln(h)\|_\alpha\le\Gamma$ *then, for any* $Q\in\partial\Omega\cap B_1(0)$ *and* $0<r\le\frac{1}{16}$, *if* v *is* $(0,\delta,8r,Q)$*-symmetric but not* $(k+1,\epsilon,8r,Q)$*-symmetric, then, for any finite Borel measure* μ *supported in* $B_r(Q)\cap\partial\Omega$,

$$\begin{aligned}\beta^k_{\mu,2}(Q,r)^2 \lesssim_{n,\alpha,M_0,\Gamma,\epsilon} \frac{r^2}{r^k}\left(\int_{B_r(Q)} W_{r/2,16r}(y)\,d\mu(y)\right) + \frac{r^2}{r^k}\int_{B_r(Q)} W_{r/2,16r}(y)^2\,d\mu(y)\\ + r^2(W_{r/2,16r}(Q)^2+\Gamma^2 r^{2\alpha}+1+r^{-1})\frac{\mu(B_r(Q))}{r^k}. \qquad (8\text{-}1)\end{aligned}$$

Before proving Lemma 8.2, we prove a few preliminary lemmas. We begin by noting that for any finite Borel measure μ and any $B_r(Q)$ we can define the μ center of mass by $X = \fint_{B_r(Q)} x\,d\mu(x)$ and define the covariance matrix of the mass distribution in $B_r(Q)$ by

$$\Sigma = \fint_{B_r(Q)} (y-X)(y-X)^\perp\,d\mu(y).$$

With this matrix, we may naturally define a symmetric, nonnegative bilinear form

$$Q(v,w) = v^\perp \Sigma w = \fint_{B_r(Q)} (v\cdot(y-X))(w\cdot(y-X))\,d\mu(y).$$

Let $\vec{v}_1,\dots,\vec{v}_n$ be an orthonormal eigenbasis and $\lambda_1\ge\cdots\ge\lambda_n\ge 0$ their associated eigenvalues. These objects enjoy the relationships

$$V^k_{\mu,2}(Q,r) = X+\operatorname{span}\{\vec{v}_1,\dots,\vec{v}_k\} \quad\text{and}\quad \beta^k_{\mu,2}(x,r)^2 = \frac{\mu(B_r(Q))}{r^k}(\lambda_{k+1}+\cdots+\lambda_n).$$

See [Hochman 2015, Section 4.2] or [Naber and Valtorta 2017, Section 7.2].

Lemma 8.3. *Let* $v\in\mathcal{A}(n,\alpha,M_0)$, *and let* $Q\in\partial\Omega\cap B_1(0)$ *and* $0<r\le\frac14$. *Let* μ, Q, λ_i, $\vec{v}_i$ *be defined as above. For any* i *and any scalar* $c\in\mathbb{R}$,

$$\lambda_i\int_{A_{3r,4r}(Q)} (\vec{v}_i\cdot\nabla v(z))^2\,dz \le \fint_{B_r(Q)}\left(\int_{A_{3r,4r}(y)} |cv(z)-\nabla v(z)\cdot(z-y)|^2\,dz\right)d\mu(y). \qquad (8\text{-}2)$$

Proof. Observe that by the definition of center of mass,

$$\fint_{B_r(Q)} \vec{w}\cdot(y-X)\,d\mu(y)=0$$

for any $\vec{w}\in\mathbb{R}^n$. Therefore, for any z for which $\nabla v(z)$ is defined,

$$\begin{aligned}
\lambda_i(\vec{v}_i\cdot\nabla v(z)) &= Q(\vec{v}_i,\nabla v(z))\\
&=\fint_{B_r(Q)}(\vec{v}_i\cdot(y-X))(\nabla v(z)\cdot(y-X))\,d\mu(y)\\
&=\fint_{B_r(Q)}(\vec{v}_i\cdot(y-X))(\nabla v(z)\cdot(y-X))\,d\mu(y)+\fint_{B_r(Q)}cv(z)(\vec{v}_i\cdot(y-X))\,d\mu(y)\\
&=\fint_{B_r(Q)}(\vec{v}_i\cdot(y-X))(cv(z)-\nabla v(z)\cdot(X-z+z-y))\,d\mu(y)\\
&=\fint_{B_r(Q)}(\vec{v}_i\cdot(y-X))(cv(z)-\nabla v(z)\cdot(z-y))\,d\mu(y)\\
&\le\lambda_i^{1/2}\left(\fint_{B_r(Q)}|cv(z)-\nabla v(z)\cdot(z-y)|^2\,d\mu(y)\right)^{\frac{1}{2}}.
\end{aligned}$$

Squaring both sides and integrating over $A_{r,R}(Q)=B_R(Q)\setminus B_r(Q)$ gives the result. □

Lemma 8.4. *Let $v\in\mathcal{A}(n,\alpha,M_0)$ with $\|\ln(h)\|_\alpha\le\Gamma$ and $0\le k\le n-2$. Let $Q\in\partial\Omega^\pm\cap B_{1/4}(0)$ and $0<r\le\frac{1}{32}$. Then, for any $Q'\in B_r(Q)\cap\partial\Omega^\pm$,*

$$\begin{aligned}
&\int_{A_{3r,4r}(Q')}\frac{|N(Q,7r,v)v(z)-\nabla v(z)\cdot(z-Q')|^2}{H(Q',|z-Q'|,v)}\,dz\\
&\qquad\lesssim_{n,\alpha,M_0,\Gamma}\int_{A_{2r,7r}(Q')}\frac{|N(Q',|z-Q'|,v)v(z)-\nabla v(z)\cdot(z-Q')|^2}{H(Q',|z-Q'|,v)}\,dz\\
&\qquad\qquad+(W_{r/2,16r}(Q)^2+W_{r/2,16r}(Q')^2+\Gamma^2r^{2\alpha}+1+r^{-1})r.
\end{aligned}$$

Proof. First, we observe that

$$N(Q,7r,v)=N(Q',|z-Q'|,v)+W_{|z-Q'|,7r}(Q)+[N(Q,|z-Q'|,v)-N(Q',|z-Q'|,v)].$$

Therefore, by the triangle inequality,

$$\begin{aligned}
&|N(Q,7r,v)v(z)-\nabla v(z)\cdot(z-Q')|^2\\
&\quad\le\big(|(W_{|z-Q'|,7r}(Q)+N(Q,|z-Q'|,v)-N(Q',|z-Q'|,v))v(z)|\\
&\qquad\qquad+|(N(Q',|z-Q'|,v))v(z)-\nabla v(z)\cdot(z-Q')|\big)^2\\
&\quad\le 2|(W_{|z-Q'|,7r}(Q)+N(Q,|z-Q'|,v)-N(Q',|z-Q'|,v))v(z)|^2\\
&\qquad\qquad+2|(N(Q',|z-Q'|,v))v(z)-\nabla v(z)\cdot(z-Q')|^2.
\end{aligned}$$

Now, using Corollary 7.7 at scale $r = |z - Q'|$ and the almost monotonicity of the Almgren frequency, we estimate

$$\int_{A_{2r,7r}(Q')} \frac{|(W_{|z-Q'|,7r}(Q) + N(Q, |z-Q'|, v) - N(Q', |z-Q'|, v))v(z)|^2}{H(Q', |z-Q'|, v)}\, dz \lesssim_{n,\alpha,M_0,\Gamma} \mathscr{C}_r(Q, Q') \int_{A_{2r,7r}(Q')} \frac{|v(z)|^2}{H(Q', |z-Q'|, v)}\, dz,$$

where the term $\mathscr{C}_r(Q, Q')$ is defined by

$$\mathscr{C}_r(Q, Q') := W_{r/2,16r}(Q)^2 + W_{r/2,16r}(Q')^2 + \Gamma^2 r^{2\alpha} + 1 + r^{-1}.$$

We finish the proof by observing that

$$\int_{A_{2r,7r}(Q')} \frac{|v(z)|^2}{H(Q', |z-Q'|, v)}\, dz \leq 7r. \qquad \square$$

Lemma 8.5. *Let $v \in \mathcal{A}(n, \alpha, M_0)$ with $\|\ln(h)\|_\alpha \leq \Gamma$ and $0 \leq k \leq n-2$. Let $Q \in B_1(0) \cap \partial\Omega^\pm$, $0 < r \leq \frac{1}{16}$. Let $0 < \epsilon$ be fixed. There exists a constant $\delta = \delta_0(n, \alpha, M_0, \Gamma, \epsilon) > 0$ and a constant $0 < C(n, \alpha, M_0, \Gamma, \epsilon)$ such that if v is $(0, \delta, 8r, Q)$-symmetric but not $(k+1, \epsilon, 8r, Q)$-symmetric, then, for any orthonormal vectors $\vec{v}_1, \ldots, \vec{v}_{k+1}$,*

$$\frac{1}{C} \leq \int_{A_{3r,4r}(Q)} \frac{r}{H(Q, r, v)} \sum_{i=1}^{k+1} (\vec{v}_i \cdot \nabla v(z))^2\, dz.$$

Proof. We argue by contradiction. Assume that there is a sequence of functions $v_i \in \mathcal{A}(n, \alpha, M_0)$, $Q_i \in B_{1/16}(0) \cap \partial\Omega_i^\pm$, and $0 < r_i \leq \frac{1}{16}$ such that v_i is $(0, 2^{-j}, 8r_i, Q_i)$-symmetric but not $(k+1, \epsilon, 8r_i, Q_i)$-symmetric. And, for each i, there exists an orthonormal collection of vectors $\{\vec{v}_{ij}\}$ such that

$$\int_{A_{3,4}(0)} \sum_{j=1}^{k+1} (\vec{v}_{ij} \cdot \nabla T_{Q,r} v_i(z))^2\, dz \leq 2^{-i}.$$

By Lemma 3.6, we may extract a subsequence $T_{Q_j,r_j} v_j$ for which $T_{Q_j,r_j} v_j$ converges to a nondegenerate function v_∞. Similarly, $\{\vec{v}_{ij}\}$ converges to an orthonormal collection $\{\vec{v}_i\}$. Given the assumptions above, v_∞ is also 0-symmetric in $B_8(0)$ and $\nabla v_\infty \cdot \vec{v}_i = 0$ for all $i = 1, \ldots, k+1$. Thus, v_∞ is $(k+1, 0)$-symmetric in $B_8(0)$. But, this is a contradiction, since the $T_{Q_j,r_j} v_j$ were supposed to stay away from $(k+1)$-symmetric functions in $L^2(B_1(0))$. $\square$

8A. *The proof of Lemma 8.2.* By Lemma 8.5 and properties of the Jones' beta numbers, we have, for $\{\vec{v}_i\}$ the orthonormal basis and λ_i the associated eigenvalues of the quadratic form in Lemma 8.3,

$$\begin{aligned}\beta^k_{\mu,2}(Q, r)^2 &\leq \frac{\mu(B_r(Q))}{r^k} n\lambda_{k+1} \leq \frac{\mu(B_r(Q))}{r^k} nC\lambda_{k+1} \frac{r}{H(Q, r, v)} \sum_{i=1}^{k+1} \int_{A_{3r,4r}(Q)} (\vec{v}_i \cdot \nabla v(z))^2\, dz \\ &\leq \frac{\mu(B_r(Q))}{r^k} nC \frac{r}{H(Q, r, v)} \sum_{i=1}^{k+1} \lambda_i \int_{A_{3r,4r}(Q)} (\vec{v}_i \cdot \nabla v(z))^2\, dz.\end{aligned}$$

By choosing $c = N(Q, 7r, v)$ in Lemma 8.3 and recalling Lemma 8.4, we have

$$\begin{aligned}
&\frac{r}{H(Q,r,v)}\lambda_i \int_{A_{3r,4r}(Q)} (\vec{v}_i \cdot \nabla v(z))^2\,dz \\
&\quad\lesssim_{n,\alpha,M_0,\Gamma} \frac{r}{H(Q,r,v)} \fint_{B_r(Q)} \biggl(\int_{A_{3r,4r}(y)} |N(Q,7r,v)v(z) - \nabla v(z)\cdot(z-y)|^2\,dz\biggr) d\mu(y) \\
&\quad\lesssim_{n,\alpha,M_0,\Gamma} r \fint_{B_r(Q)} \biggl(\int_{A_{2r,7r}(y)} \frac{|N(y,|z-y|,v)v(z) - \nabla v(z)\cdot(z-y)|^2}{H(Q,|z-y|,v)}\,dz\biggr) d\mu(y) \\
&\qquad\qquad + r^2 \fint_{B_r(Q)} (W_{r,16r}(Q)^2 + W_{r,16r}(y)^2 + \Gamma^2 r^{2\alpha} + 1 + r^{-1})\,d\mu(y).
\end{aligned}$$

Therefore, collecting constants and using Lemmas 4.9 and 7.1, we have

$$\begin{aligned}
\beta^k_{\mu,2}(Q,r)^2 &\lesssim_{n,\alpha,M_0,\Gamma} \frac{r^2}{r^k}\biggl(\int_{B_r(Q)} N(y,8r,v) - N(y,r,v)\,d\mu(y)\biggr) \\
&\quad + r^2(W_{r/2,16r}(Q)^2 + \Gamma^2 r^{2\alpha} + 1 + r^{-1})\frac{\mu(B_r(Q))}{r^k} + \frac{r^2}{r^k}\int_{B_r(Q)} W_{r/2,16r}(y)^2\,d\mu(y). \qquad \square
\end{aligned}$$

9. Packing

The following theorem of Naber and Valtorta [2017] is a powerful tool which links the sum of the $\beta^k_\mu(Q,r)^2$ over all points and scales to packing estimates.

Theorem 9.1 [Naber and Valtorta 2017, discrete Reifenberg]. *Let $\{B_{\tau_i}(x_i)\}_i$ be a collection of disjoint balls such that, for all $i = 1, 2, \dots,$ we have $\tau_i \le 1$. Let $\epsilon_k > 0$ be fixed. Define a measure*

$$\mu := \sum_i \tau_i^k \delta_{x_i},$$

and suppose that, for any $x \in B_2(0)$ and any scale $l \in \{0, 1, 2, \dots\}$, if $B_{r_l}(x) \subset B_2(0)$ and $\mu(B_{r_l}(x)) \ge \epsilon_k r_l^k$ then

$$\sum_{i \ge l} \int_{B_{2r_l(x)}} \beta^k_\mu(z, 16r_i)^2\,d\mu(z) < r_l^k \delta^2.$$

Then there exists a $\delta_0 = \delta_0(n, \epsilon_k) > 0$ such that if $\delta \le \delta_0$,

$$\mu(B_1(0)) = \sum_{i \text{ s.t. } x_i \in B_1(0)} \tau_i^k \le C(n).$$

Now we are ready to prove the crucial packing lemma.

Lemma 9.2. *Fix $0 < \epsilon$, and let $v \in \mathcal{A}(n, \alpha, M_0)$ satisfy $\|\ln(h)\|_\alpha \le \eta$ and $\sup_{Q \in B_1(0) \cap \partial\Omega^\pm} N(Q, 2, v) = E$. There is an $\eta_1(n, \alpha, M_0, \epsilon) > 0$ such that if $\eta \le \eta_1$, then for any $r > 0$ if $\{B_{2r_{Q'}}(Q')\}$ is a collection of disjoint balls satisfying*

$$N(p, \eta r_{Q'}, v) \ge E - \eta_1, \quad Q' \in S^k_{\eta_1, r}, \quad r \le r_{Q'} \le 1, \tag{9-1}$$

we have the packing estimate

$$\sum_{Q'} r^k_{Q'} \le C_2(n, \alpha, M_0, \epsilon). \tag{9-2}$$

Proof. Choose $\delta_0(n, \alpha, M_0, \epsilon)$ as in Lemma 8.2, and $\gamma(n, \alpha, M_0, \delta_0)$ as in Lemma 5.2. Note that we may assume without loss of generality that $\eta_1 \leq 1$, and so for $C_1(\alpha, M_0, 1)$ the constant in Lemma 4.9, let

$$\eta_1 \leq \frac{\min\{\delta_0, \gamma\}}{2C_1 + 1}.$$

We will employ the convention that $r_i = 2^{-i}$. For each $i \in \mathbb{N}$, define the truncated measure

$$\mu_i = \sum_{r_{Q'} \leq r_i} r_{Q'}^k \delta_{Q'}.$$

We will write $\beta_i(x, r) = \beta^k_{\mu_i,2}(x, r)$. Observe that the β_i enjoy the following properties. First, because the balls are disjoint, for all $j \geq i$,

$$\beta_i(x, r_j) = \begin{cases} \beta_j(x, r_j) & \text{if } x \in \operatorname{supp}(\mu_j), \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, for $r_i \leq 2^{-4}$, recalling Lemma 4.10 our assumption of the Almgren frequency gives that $N(16r_i, Q, v) - N(r_Q, Q, v) \leq (2C_1 + 1)\eta \leq \max\{\delta_0, \gamma\} \leq 1$ and

$$|W_{r_j/2,16r_j}(Q')| \leq \eta + C(\alpha, M_0, \eta)\eta(16r_j)^\alpha.$$

Thus, for $0 < \eta$ small enough depending only upon α and M_0, we have $|W_{r_j/2,16r_j}(Q')| \leq 1$. Therefore

$$W_{r_j/2,16r_j}(Q')^2 \leq |W_{r_j/2,16r_j}(Q')|.$$

In particular, by Lemmas 5.2 and 8.2 and our choice of $\eta \leq \eta_1$,

$$\beta^k_{\mu_i,2}(Q, r_i)^2 \lesssim_{n,\alpha,M_0,\Gamma,\epsilon} \frac{1}{r^k}\left(\int_{B_r(Q)} |W_{r/2,16r}(y)|\, d\mu(y)\right) + (|W_{r/2,16r}(Q)| + \Gamma^2 r^{2\alpha+2} + r^2 + r)\frac{\mu(B_r(Q))}{r^k}.$$

The claim of the lemma is that $\mu_0(B_1(0)) \leq C(n, \alpha, M_0, \epsilon)$. We prove the claim inductively. That is, we shall argue that there is a fixed scale $0 < R = 2^{-\ell}$ (depending only upon n, α, M_0, ϵ) such that, for $r_i \leq R$ and all $x \in B_1(0)$,

$$\mu_i(B_{r_i}(x)) \leq C_{DR}(n) r_i^k.$$

Observe that since $r_{Q'} \geq r > 0$, for $r_i < r$ the claim is trivially satisfied because $\mu_i = 0$. Assume, then, that the inductive hypothesis holds for all $j \geq i + 1$. Let $x \in B_1(0)$. We consider $\mu_i(B_{4r_i}(x))$. Observe that we can get a course bound

$$\mu_j(B_{4r_j}(x)) \leq \Gamma(n) r_j^k \quad \text{for all } j \geq i - 2 \text{ for all } x \in B_1(0)$$

by writing $\mu_j(B_{4r_j}(x)) = \mu_{j+2}(B_{4r_j}(x)) + \sum r^k_{Q'}$, where the sum is taken over all $Q' \in B_{4r_j}(x)$ with $r_{j+2} < r_{Q'} \leq r_j$. Since the balls $B_{r_{Q'}}(Q')$ are disjoint, there is a dimensional constant $c(n)$ which bounds the number of such points. Thus, we may take $\Gamma(n) = c(n) C_{DR}$.

Now, we calculate

$$
\begin{aligned}
&\sum_{r_j<2r_i} \int_{B_{2r_i}(x)} \beta_i(z, r_j)^2 \, d\mu_i(z) \\
&= \sum_{r_j<2r_i} \int_{B_{2r_i}(x)} \beta_j(z, r_j)^2 \, d\mu_j(z) \\
&\leq C \sum_{r_j<2r_i} \frac{1}{r_j^k} \int_{B_{2r_i}(x)} \left(\int_{B_{r_j}(z)} |W_{r_j/2,16r_j}(y)| d\mu_j(y) \right) d\mu_j(z) \\
&\qquad + C \sum_{r_j<2r_i} \int_{B_{2r_i}(x)} \left((|W_{r_j/2,16r_j}(z)| + \eta r_j^{2\alpha+2} + r_j^2 + r_j) \frac{\mu(B_{r_j}(z))}{r_j^k} \right) d\mu_j(z) \\
&\leq C \sum_{r_j<2r_i} \int_{B_{2r_i+r_j}(x)} \frac{\mu_j(B_{r_j}(y))}{r_j^k} |W_{r_j/2,16r_j}(y)| \, d\mu_j(y) \\
&\qquad + C \sum_{r_j<2r_i} \int_{B_{2r_i}(x)} \left((|W_{r_j/2,16r_j}(z)| + \eta r_j^{2\alpha+2} + r_j^2 + r_j) \frac{\mu(B_{r_j}(z))}{r_j^k} \right) d\mu_j(z) \\
&\leq 2C\Gamma(n) \int_{B_{4r_i}(x)} \left(\sum_{r_j<2r_i} |W_{r_j/2,16r_j}(y)| \right) d\mu_j(y) + C\Gamma(n) \sum_{r_j<2r_i} (\eta r_j^{2\alpha+2} + r_j^2 + r_j) \mu_i(B_{4r_i}(x)).
\end{aligned}
$$

Therefore, recalling $r_i = 2^{-i}$ we see that

$$
\sum_{j=i-1}^{N} |W_{r_j/2,16r_j}(Q')| \leq 6 \operatorname{var}_{r \in [r_{Q'}, r_{i-1}]} N(r, Q', v) \leq 12C(\alpha, M_0)\eta(r_{i-1}^\alpha - r_{Q'}^\alpha) + 6\eta.
$$

Therefore

$$
\begin{aligned}
\sum_{r_j<2r_i} \int_{B_{2r_i}(x)} \beta_i(z, r_j)^2 \, d\mu_i(z) &\leq C\Gamma(n)\mu_i(B_{4r_i}(x)) \left(6\eta + 12C_1 r_{i-1}^\alpha \eta \right) + C\Gamma^2(n) \left(\sum_{r_j<2r_i} \eta r_j^{2\alpha} + r_j^2 + r_j \right) r_i^k \\
&\leq C\Gamma^2(n)(1 + C(\alpha))\eta r_i^k + C\Gamma^2(n) r_i^k \sum_{r_j<2r_i} r_j^2 + r_j.
\end{aligned}
$$

Thus, for $\eta \leq \eta_1(n, \alpha, M_0, \epsilon)$ sufficiently small and $r_i \leq R(n, \alpha, M_0, \epsilon) = 2^{-\ell}$ sufficiently small,

$$
C\Gamma(n)^2(1 + C(\alpha))\eta < \tfrac{1}{2}\delta_{DR} \quad \text{and} \quad C\Gamma^2(n) \sum_{r_j<2r_i} r_j^2 + r_j < \tfrac{1}{2}\delta_{DR}.
$$

For such i and μ_i satisfying the hypotheses of Theorem 9.1,

$$
\sum_{r_j<2r_i} \int_{B_{2r_i}(x)} \beta_i(z, r_j)^2 \, d\mu_i(z) \leq \delta_{DR} r_i^k.
$$

The discreet Reifenberg theorem therefore implies that $\mu_i(B_{r_i}(x)) \leq C_{DR} r_i^k$. Thus, by induction, the claim holds for $r_i \leq R = 2^{-\ell}$. We may use a packing argument using balls of radius $2^{-\ell}$ to obtain estimates at larger scales. That is, $\mu_0(B_1(0)) \leq C_{DR} C(n, \ell)$. □

10. Tree construction

In this section, we detail two procedures for inductively refined covering schemes. We will use these covering schemes in the next section to generate the actual cover which proves Theorem 2.15. First, we fix our constants.

10A. ***Fixing constants and a definition.*** In this section, we fix our constants as follows. Fix $0 < \epsilon$, and let $v \in \mathcal{A}(n, \alpha, M_0)$. Let $E = \sup_{Q \in B_1(0) \cap \partial\Omega^{\pm}} N(Q, 2, v)$, and fix the scale of the covering we wish to construct as $R \in (0, 1]$.

We will let ρ denote the inductive scale at which we will refine our cover. For convenience, we will use the convention $r_i = \rho^{-i}$. Let $\rho < \frac{1}{10}$ be small enough that

$$2C_2(n, \alpha, M_0, \epsilon)c_2(n)\rho < \tfrac{1}{2},$$

where $C_2(n, \alpha, M_0, \epsilon)$ is as in Lemma 9.2 and $c_2(n)$ is a dimensional constant which will be given in the following lemmas.

Let $\delta(n, \alpha, M_0, \epsilon)$ be as in Lemma 8.2 and $\gamma(n, \alpha, M_0, \delta)$ as in Lemma 5.2. We also let $\eta_1(n, \alpha, M_0, \epsilon)$ be as in Lemma 9.2, and let

$$\gamma_0 = \eta' = \tfrac{1}{20}\eta_1.$$

Note that while $\gamma_0 \leq \gamma$, Lemma 5.2 still holds with γ_0 in place of γ. As in Corollary 6.3, we then let $\eta = \eta_0(n, \alpha, E + 1, \epsilon, \eta', \gamma_0, \rho)$. We shall assume that v satisfies

$$\|\ln(h)\|_\alpha \leq \frac{1}{2C_1 + 1}\eta.$$

The sorting principle for our covering comes from Corollary 6.3. To formalize this, we make the following definition.

Definition 10.1. For $Q' \in B_2(0) \cap \partial\Omega^{\pm}$ and $0 < R < r < 2$, the ball $B_r(Q)$ will be called "good" if

$$N(Q, \gamma\rho r, v) \geq E - \eta' \quad \text{for all } Q \in S^k_{\epsilon,\eta R}(v) \cap B_r(Q').$$

We will say that $B_r(Q')$ is "bad" if it is not good.

Remark 10.2. By Corollary 6.3, with $E + \frac{1}{2}\eta_0$ in place of E — which is admissible by monotonicity and our choice of $\|\ln(h)\|_\alpha \leq \eta/(2C_1 + 1)$ — in any bad ball $B_r(Q')$ there exists a $(k-1)$-dimensional affine plane L^{k-1} such that

$$\{N(Q, \gamma\rho r, v) \geq E - \tfrac{1}{2}\eta_0\} \cap B_r(Q') \subset B_{\rho r}(L^{k-1}).$$

10B. ***Good trees.*** Let $x \in B_1(0) \cap \partial\Omega^{\pm}$ and $B_{r_A}(x)$ be a good ball for $A \geq 0$. We will detail the inductive construction of a good tree based at $B_{r_A}(x)$. The induction will build a successively refined covering $B_{r_A}(x) \cap S^k_{\epsilon,\eta R}(v)$. We will terminate the process and have a cover which consists of a collection of bad balls with packing estimates and a collection of stop balls whose radii are comparable to R. We shall use the notation $\mathcal{G}_i$ to denote the collection of centers of good balls of scale r_i, and $\mathcal{B}_i$ shall denote the collection of centers of bad balls of scale r_i.

Because $B_{r_A}(x)$ is a good ball, at scale $i = A$, we set $\mathcal{G}_A = x$. We let $\mathcal{B}_A = \varnothing$. Now the inductive step. Suppose that we have constructed our collections of good and bad balls down to scale $j-1 \geq A$. Let $\{z\}_{J_i}$ be a maximal $\frac{2}{5}r_j$-net in

$$B_{r_A}(x) \cap S^k_{\epsilon,\eta R}(v) \cap B_{r_{j-1}}(\mathcal{G}_{j-1}) \setminus \bigcup_{i=A}^{j-1} B_{r_i}(\mathcal{B}_i).$$

We then sort these points into $\mathcal{G}_j$ and $\mathcal{B}_j$ depending on whether $B_{r_j}(z)$ is a good ball or a bad ball. If $r_j > R$, we proceed inductively. If $r_j \leq R$, then we stop the procedure. In this case, we let $\mathcal{S} = \mathcal{G}_j \cup \mathcal{B}_j$ and we call this the collection of "stop" balls.

The covering at which we arrive at the end of this process shall be called the "good tree at $B_{r_A}(x)$". We shall follow [Edelen and Engelstein 2019] and denote this by $\mathcal{T}_\mathcal{G} = \mathcal{T}_\mathcal{G}(B_{r_A}(x))$. We shall call the collection of "bad" ball centers $\bigcup_i \mathcal{B}_i$ the "leaves of the tree" and denote this collection by $\mathcal{F}(\mathcal{T}_\mathcal{G})$. We shall denote the collection of "stop" ball centers by $\mathcal{S}(\mathcal{T}_\mathcal{G}) = \mathcal{S}$.

For $b \in \mathcal{F}(\mathcal{T}_\mathcal{G})$ we let $r_b = r_i$ for i such that $b \in \mathcal{B}_i$. Similarly, if $s \in \mathcal{S}(\mathcal{T}_\mathcal{G})$, we let $r_s = r_j$ for the terminal j.

Theorem 10.3. *A good tree $\mathcal{T}_\mathcal{G}(B_{r_A}(x))$ enjoys the following properties*:

(A) *Tree-leaf packing*:

$$\sum_{b\in\mathcal{F}(\mathcal{T}_\mathcal{G})} r_b^k \leq C_2(n, \alpha, M_0, \epsilon) r_A^k.$$

(B) *Stop ball packing*:

$$\sum_{s\in\mathcal{S}(\mathcal{T}_\mathcal{G})} r_s^k \leq C_2(n, \alpha, M_0, \epsilon) r_A^k.$$

(C) *Covering control*:

$$\mathcal{S}^k_{\epsilon,\eta R}(v) \cap B_{r_A}(x) \subset \bigcup_{s\in\mathcal{S}(\mathcal{T}_\mathcal{G})} B_{r_s}(s) \cup \bigcup_{b\in\mathcal{F}(\mathcal{T}_\mathcal{G})} B_{r_b}(b).$$

(D) *Size control*: *for any $s \in \mathcal{S}(\mathcal{T}_\mathcal{G})$, we have $\rho R \leq r_s \leq R$.*

Proof. First, observe that by construction

$$\{B_{r_b/5}(b) : b \in \mathcal{F}(\mathcal{T}_\mathcal{G})\} \cup \{B_{r_s/5}(s) : s \in \mathcal{S}(\mathcal{T}_\mathcal{G})\}$$

is pairwise disjoint and centered in the set $\mathcal{S}^k_{\epsilon,\eta R}(v)$. Next, all bad balls and stop balls are centered in a good ball of the previous scale. By our definition of good balls, then, we have for all i

$$N(b, \gamma r_i, v) = N(b, \gamma\rho r_{i-1}, v) \geq E - \eta' \quad \text{for all } \in \mathcal{B}_i$$

and

$$N(s, \gamma r_s, v) \geq E - \eta' \quad \text{for all } s \in \mathcal{S}(\mathcal{T}_\mathcal{G}).$$

Since by monotonicity we have that $\sup_{p\in B_{r_A}(x)} N(Q, 2r_A, v) \leq E + \eta'$, we can apply Lemma 9.2 to $B_{r_A}(x)$ and get the packing estimates (A) and (B).

Covering control (C) follows from our choice of a maximal $\frac{2}{5}r_i$-net at each scale i. If i is the first scale at which a point $x \in \mathcal{S}^k_{\epsilon,\eta R}(v)$ was not contained in our inductively refined cover, it would violate the maximality assumption.

The last condition (D) follows because we stop only if j is the first scale for which $r_j \le R$. Since we decrease by a factor of ρ at each scale, (D) follows. □

10C. ***Bad trees.*** Let $B_{r_A}(x)$ be a bad ball. Note that for every bad ball, there is a $(k-1)$-dimensional affine plane L^{k-1} associated to it which satisfies the properties elaborated in Corollary 6.3. Our construction of bad trees will differ in several respects from our construction of good trees. The idea is still to define an inductively refined cover at decreasing scales of $B_{r_A}(x) \cap \mathcal{S}^k_{\epsilon,\eta R}(v)$. We shall again sort balls at each step into "good", "bad", and "stop" balls. But these balls will play slightly different roles and the "stop" balls will have different radii.

We reuse the notation $\mathcal{G}_i$ to denote the collection of centers of good balls of scale r_i, $\mathcal{B}_i$ to denote the collection of centers of bad balls of scale r_i, and $\mathcal{S}_i$ to denote the collection of centers of stop balls of scale r_i.

At scale $i = A$, we set $\mathcal{B}_A = x$, since $B_{r_A}(x)$ is a bad ball, and set $\mathcal{S}_A = \mathcal{G}_A = \varnothing$. Suppose, now that we have constructed good, bad, and stop balls for scale $i-1 \ge A$. If $r_i > R$, then define $\mathcal{S}_i$ to be a maximal $\frac{2}{5}\eta r_{i-1}$-net in

$$B_{r_A}(x) \cap \mathcal{S}^k_{\epsilon,\eta R}(v) \cap \bigcup_{b \in \mathcal{B}_{i-1}} B_{r_{i-1}}(b) \setminus B_{2\rho r_{i-1}}(L^{k-1}_b).$$

Note that $\eta \ll \rho$, so $\eta r_{i-1} < r_i$. We then let $\{z\}$ be a maximal $\frac{2}{5}r_i$-net in

$$B_{r_A}(x) \cap \mathcal{S}^k_{\epsilon,\eta R}(v) \cap \bigcup_{b \in \mathcal{B}_{i-1}} B_{r_{i-1}}(b) \cap B_{2\rho r_{i-1}}(L^{k-1}_b).$$

We then sort $\{z\}$ into the disjoint union $\mathcal{G}_i \cup \mathcal{B}_i$ depending on whether $B_{r_i}(z)$ is a good ball or a bad ball.

If $r_i \le R$, we terminate the process by defining $\mathcal{G}_i = \mathcal{B}_i = \varnothing$ and letting $\mathcal{S}_i$ be a maximal $\frac{2}{5}\eta r_{i-1}$-net in

$$B_{r_A}(x) \cap \mathcal{S}^k_{\epsilon,\eta R}(v) \cap B_{r_i}(\mathcal{B}_{i-1}).$$

The covering at which we arrive at the end of this process shall be called the "bad tree at $B_{r_A}(x)$". We shall follow [Edelen and Engelstein 2019] and denote this by $\mathcal{T}_\mathcal{B} = \mathcal{T}_\mathcal{B}(B_{r_A}(x))$. We shall call the collection of "good" ball centers, $\bigcup_i \mathcal{G}_i$, the "leaves of the tree" and denote this collection by $\mathcal{F}(\mathcal{T}_\mathcal{B})$. We shall denote the collection of "stop" ball centers by $\mathcal{S}(\mathcal{T}_\mathcal{B}) = \bigcup_i \mathcal{S}_i$.

As before, we shall use the convention that for $g \in \mathcal{F}(\mathcal{T}_\mathcal{B})$ we let $r_g = r_i$ for i such that $g \in \mathcal{G}_i$. However, note that now, if $s \in \mathcal{S}_i \subset \mathcal{S}(\mathcal{T}_\mathcal{B})$, we let $r_s = \eta r_{i-1}$.

Theorem 10.4. *A bad tree $\mathcal{T}_\mathcal{B}(B_{r_A}(x))$ enjoys the following properties*:

(A) *Tree-leaf packing*:

$$\sum_{g \in \mathcal{F}(\mathcal{T}_\mathcal{B})} r_g^k \le 2c_2(n)\rho r_A^k.$$

(B) *Stop ball packing*:

$$\sum_{s \in \mathcal{S}(\mathcal{T}_\mathcal{B})} r_s^k \le c(n,\eta) r_A^k.$$

(C) *Covering control*:

$$\mathcal{S}^k_{\epsilon,\eta R}(v) \cap B_{r_A}(x) \subset \bigcup_{s\in\mathcal{S}(\mathcal{T}_\mathcal{B})} B_{r_s}(s) \cup \bigcup_{g\in\mathcal{F}(\mathcal{T}_\mathcal{B})} B_{r_g}(g).$$

(D) *Size control*: *for any* $s \in \mathcal{S}(\mathcal{T}_\mathcal{B})$, *at least one of the following holds*:

$$\eta R \le r_s \le R \quad \text{or} \quad \sup_{Q\in B_{2r_s}(s)\cap\partial\Omega^\pm} N(Q, 2r_s, v) \le E - \tfrac{1}{2}\eta.$$

Proof. Conclusion (C) follows identically as in Theorem 10.3. Next we consider the packing estimates. Let $r_i > R$. Then, by construction, for any $b \in \mathcal{B}_{i-1}$, we have

$$\mathcal{G}_i \cup \mathcal{B}_i \cup B_{r_{i-1}}(b) \subset B_{2\rho r_{i-1}}(L^{k-1}_b).$$

Thus, since the points $\mathcal{G}_i \cup \mathcal{B}_i$ are $\frac{2}{5}r_i$ disjoint, we calculate

$$|\mathcal{G}_i \cup \mathcal{B}_i \cup B_{r_{i-1}}(b)| \le \omega_{k-1}\omega_{n-k+1}(3\rho)^{n-k+1}\frac{1}{\omega_n\left(\frac{1}{5}\rho\right)^n} \le c_2(n)\rho^{1-k}.$$

We can push this estimate up the scales as follows:

$$\begin{aligned}|\mathcal{G}_i \cup \mathcal{B}_i|r_i^k &\le c_2(n)\rho^1|\mathcal{B}_{i-1}|r^k_{i-1}\\ &\le c_2(n)\rho^1|\mathcal{B}_{i-1}\cup\mathcal{G}_{i-1}|r^k_{i-1}\\ &\vdots\\ &\le (c_2\rho)^{i-A}r^k_A.\end{aligned}$$

Summing over all $i \ge A$, then, we have that

$$\sum_{i=A+1}^{\infty} |\mathcal{B}_{i-1}\cup\mathcal{G}_{i-1}|r_i^k \le \sum_{i=A+1}^{\infty}(c_2\rho)^{i-A}r^k_A.$$

Since we chose $c_2\rho \le \frac{1}{2}$, we have that the sum converges and

$$\sum_{i=A+1}^{\infty} |\mathcal{B}_{i-1}\cup\mathcal{G}_{i-1}|r_i^k \le 2c_2\rho r^k_A.$$

This proves (A).

To see (B), we observe that for any given scale $i \ge A+1$, the collection of stop balls $\{B_{\eta r_{i-1}}(s)\}_{s\in\mathcal{S}_i}$ form a Vitali collection centered in $B_{r_{i-1}}(\mathcal{B}_{i-1})$. Thus, we have

$$|\{\mathcal{S}_i\}| \le \frac{10^n}{\eta^n}|\{\mathcal{B}_{i-1}\}|.$$

Since by construction there are no stop balls at the initial scale A, we compute that

$$\sum_{i=A+1}^{\infty} |\{\mathcal{S}_i\}|(\eta r_{i-1})^k \le 10^k\eta^{k-n}\sum_{i=A}^{\infty}|\{\mathcal{B}_i\}|r_i^k \le c(n,\eta)r^k_A.$$

This is (B).

We now argue (D). For $s \in \mathcal{S}_i$ where $r_i > R$, by construction $s \in B_{r_{i-1}}(b) \setminus B_{2\rho r_{i-1}}(L^{k-1})$ for some $b \in \mathcal{B}_{i-1}$. By Corollary 6.3, the construction, and our choice of $\eta \le \frac{1}{2}\rho$, we have

$$\sup_{p \in B_{2r_s}(s)} N(Q, 2r_s, v) \le \sup_{p \in B_{2\eta r_{i-1}}(s)} N(Q, 2\eta r_{i-1}, v) \le E - \tfrac{1}{2}\eta.$$

On the other hand, if $r_i \le R$, then $r_{i-1} > R$. Thus

$$R \ge \rho r_{i-1} \ge \eta r_{i-1} = r_s \ge \eta R.$$

This proves (D). □

11. The covering

Assuming that $\|\ln(h)\|_\alpha \le \eta/(2C_1+1)$, for $0 < \eta \le \eta_0(n, \alpha, E+1, \epsilon, \eta', \gamma_0, \rho)$ as in Section 10, we now wish to build the covering of $\mathcal{S}^k_{\epsilon,\eta R} \cap B_1(0)$ using the tree constructions above. Note that $B_1(0)$ is either a good ball or a bad ball. Therefore, we can construct a tree with $B_1(0)$ as the root. Then in each of the leaves, we construct either good trees or bad trees, depending upon the type of the leaf. Since in each construction we decrease the size of the leaves by a factor of $\rho < \frac{1}{10}$, we can continue alternating tree types until the process terminates in finite time.

Explicitly, we let $\mathcal{F}_0 = \{0\}$ and let $B_1(0)$ be the only leaf. We set $\mathcal{S}_0 = \varnothing$. Now, assume that we have defined the leaves and stop balls up to stage $i-1$. Since by hypothesis, the leaves in $\mathcal{F}_i$ are all good balls or bad balls, if they are good, we define for each $f \in \mathcal{F}_{i-1}$ the good tree $\mathcal{T}_{\mathcal{G}}(B_{r_f}(f))$. We then set

$$\mathcal{F}_i = \bigcup_{f \in \mathcal{F}_{i-1}} \mathcal{F}(\mathcal{T}_{\mathcal{G}}(B_{r_f}(f))) \quad \text{and} \quad \mathcal{S}_i = \mathcal{S}_{i-1} \cup \bigcup_{f \in \mathcal{F}_{i-1}} \mathcal{S}(\mathcal{T}_{\mathcal{G}}(B_{r_f}(f))).$$

Since all the leaves of good trees are bad balls, all the leaves of $\mathcal{F}_i$ are bad.

If, on the other hand, leaves of $\mathcal{F}_{i-1}$ are bad, then for each $f \in \mathcal{F}_{i-1}$ we construct a bad tree $\mathcal{T}_{\mathcal{B}}(B_{r_f}(f))$. In this case, we set

$$\mathcal{F}_i = \bigcup_{f \in \mathcal{F}_{i-1}} \mathcal{F}(\mathcal{T}_{\mathcal{B}}(B_{r_f}(f))) \quad \text{and} \quad \mathcal{S}_i = \mathcal{S}_{i-1} \cup \bigcup_{f \in \mathcal{F}_{i-1}} \mathcal{S}(\mathcal{T}_{\mathcal{B}}(B_{r_f}(f))).$$

Since all the leaves of bad trees are good balls, all the leaves of $\mathcal{F}_i$ are good.

This construction gives the following estimates.

Lemma 11.1. *For the construction described above, there is an $N \in \mathbb{N}$ such that $\mathcal{F}_N = \varnothing$ with the following properties*:

(A) *Leaf packing*:

$$\sum_{i=0}^{N-1} \sum_{f \in \mathcal{F}_i} r_f^k \le c(n).$$

(B) *Stop ball packing*:

$$\sum_{s \in \mathcal{S}_N} r_s^k \le c(n, \alpha, M_0, \epsilon).$$

(C) *Covering control*:

$$\mathcal{S}^k_{\epsilon,\eta R}(v) \cap B_1(0) \subset \bigcup_{s\in\mathcal{S}_N} B_{r_s}(s).$$

(D) *Size control*: *for any* $s \in \mathcal{S}_N$, *at least one of the following holds*:

$$\eta R \le r_s \le R \quad \text{or} \quad \sup_{Q\in B_{2r_s}(s)\cap\partial\Omega^\pm} N(Q, 2r_s, v) \le E - \tfrac{1}{2}\eta.$$

Proof. By construction, each of the leaves of a good or bad tree satisfy $r_f \le r_i$. Thus, there is an i sufficiently large such that $r_i < R$. Thus, N is finite.

To see (A), we use the previous theorems. That is, if the leaves $\mathcal{F}_i$ are good, then they are the leaves of bad trees rooted in $\mathcal{F}_{i-1}$. Thus, we calculate by Theorem 10.4

$$\sum_{f\in\mathcal{F}_i} r_f^k \le 2c_2(n)\rho \sum_{f'\in\mathcal{F}_{i-1}} r_{f'}^k.$$

On the other hand, if the leaves $\mathcal{F}_i$ are bad, then they are the leaves of good trees rooted in $\mathcal{F}_{i-1}$. Thus, we calculate by Theorem 10.3

$$\sum_{f\in\mathcal{F}_i} r_f^k \le C_2(n,\alpha,M_0,\epsilon) \sum_{f'\in\mathcal{F}_{i-1}} r_{f'}^k.$$

Concatenating the estimates, since we alternate between good and bad leaves, we have

$$\sum_{f\in\mathcal{F}_i} r_f^k \le c(n)(2C_2(n,\alpha,M_0,\epsilon)c_2(n)\rho)^{i/2}.$$

By our choice of ρ,

$$\sum_{f\in\mathcal{F}_i} r_f^k \le c(n)2^{-i/2}.$$

The estimate (A) follows immediately.

We now turn our attention to (B). Each stop ball $s \in \mathcal{S}_N$ is a stop ball coming from a good or a bad tree rooted in one of the leaves of a bad tree or good tree. We have the estimates from Theorems 10.3 and 10.4, which give bounds packing both leaves and stop balls. Combining these, we get

$$\sum_{s\in\mathcal{S}_N} r_s^k = \sum_{i=0}^{N}\sum_{s\in\mathcal{S}_i} r_s^k \le \sum_{i=0}^{N-1}\sum_{f\in\mathcal{F}_i} c(n,\eta) r_f^k \le C(n,\eta).$$

Recalling the dependencies of η gives the desired result.

Property (C) follows inductively from the analogous covering control in Theorems 10.3 and 10.4 applied to each tree constructed. Property (D) is immediate from these theorems as well. □

Corollary 11.2. *Fix* $0 < \epsilon$. *Let* $v \in \mathcal{A}(n,\alpha,M_0)$ *satisfy* $\sup_{p\in B_2(0)} N(Q,2,v) \le E$. *Fix* $0 < \epsilon$. *There is an* $\eta_0(n,\alpha,M_0,\epsilon,E) > 0$ *such that if* $0 < \eta \le \eta_0$ *and* $\|\ln(h)\|_\alpha \le \eta/(2C_1+1)$ *then given any* $0 < R \le 1$ *there is a collection of balls* $\{B_{r_x}(x)\}_{x\in\mathcal{U}}$ *with centers* $x \in S^k_{\epsilon,\eta R}(v) \cap B_1(0)$. *Further,* $R \le r_x \le \frac{1}{10}$ *and the collection has the following properties*:

(A) *Packing*:

$$\sum_{x\in\mathcal{U}} r_x^k \le c(n,\alpha,M_0,E,\epsilon).$$

(B) *Covering control*:

$$\mathcal{S}^k_{\epsilon,\eta R}(v)\cap B_1(0)\subset \bigcup_{x\in\mathcal{U}} B_{r_x}(x).$$

(C) *Energy drop*: *for every* $x\in\mathcal{U}$, *either*

$$r_x = R \quad \text{or} \quad \sup_{Q\in B_{2r_s}(s)\cap\partial\Omega^\pm} N(Q,2r_s,v)\le E-\tfrac{1}{2}\eta_0.$$

This follows immediately from Lemma 11.1 with $\eta\le\eta_1$, $\mathcal{S}_N=\mathcal{U}$, and setting $r_x=\max\{R,r_s\}$.

11A. *Proof of Theorem 2.15.*

Lemma 11.3. *Let* $v\in\mathcal{A}(n,\alpha,M_0)$ *with* $\|\ln(h)\|_\alpha\le\Gamma$. *There exists a scale* $\kappa(n,\alpha,M_0,\Gamma,\epsilon)>0$ *such that, for all balls* $B_r(Q)$ *with* $0<r<\kappa$ *and* $Q\in B_{1/4}(0)\cap\partial\Omega^\pm$, *the function* $\tilde v(x)=v(rx+Q)$ *on* $B_1(0)$ *satisfies the following properties*:

$$\sup_{Q'\in B_1(0)\cap T_{Q,r}\partial\Omega^\pm} N(Q',2,\tilde v)\le C(\alpha,M_0,\Gamma) \quad \textit{and} \quad \|\ln(\tilde h)\|_{C^{0,\alpha}(B_1(0))}\le\frac{\eta_0}{2C_1+1},$$

where

$$\eta_0=\eta_0(n,\alpha,C(n,\alpha,M_0,\Gamma)+1,\eta',\epsilon,\gamma_0,\rho)=\eta_0(n,\alpha,M_0,\Gamma,\epsilon)$$

is as in Corollary 6.3 and $C(n,\alpha,M_0,\Gamma)$ *is as in Corollary 4.4.*

Proof. First, note that if $\ln(h)\in C^{0,\alpha}(B_1(0))$, then $\ln(\tilde h(x))=\ln(h(rx+Q))$ satisfies

$$|\ln(\tilde h(x))-\ln(\tilde h(z))|=|\ln(h(rx+Q))-\ln(h(rz+Q))|\le\Gamma|rx-rz|^\alpha=\Gamma r^\alpha|x-z|^\alpha.$$

Since $r^\alpha\to 0$ as $r\to 0$, there exists a $\kappa(n,\alpha,M_0,\Gamma,\epsilon)>0$ such that $\|\ln(\tilde h)\|_{C^{0,\alpha}}\le\Gamma\kappa^\alpha<\eta_0/(2C_1+1)$. By a similar calculation, we see that $\mathrm{Lip}(\tilde v)\le r\,\mathrm{Lip}(v)$. Thus, the fact that $H(Q,R,v)=H(0,R/r,\tilde v)$ for any $Q\in B_{1/4}(0)\cap\partial\Omega^\pm$ and $0<r\le 2\kappa$, Lemma 3.4, and $Q'\in B_r(Q)\cap\partial\Omega^\pm$ yields

$$N(T_{Q,r}Q',2,\tilde v)=r^2N(Q,2r,v)\le r^2C(\alpha,M_0,\Gamma). \qquad \square$$

Theorem 11.4. *Let* $v\in\mathcal{A}(n,\alpha,M_0)$ *with* $\|\ln(h)\|_\alpha\le\Gamma$. *For all* $\epsilon>0$ *there exists an* $\eta_0(n,\alpha,M_0,\Gamma,\epsilon)>0$ *such that, for all* $0<R<1$ *and* $k=1,2,\dots,n-1$, *we can find a collection of balls* $\{B_R(x_i)\}_i$ *with the following properties*:

(1) $\mathcal{S}^k_{\epsilon,\eta_0 R}(v)\cap B_{1/4}(0)\subset\bigcup_i B_R(x_i)$.

(2) $|\{x_i\}_i|\le c(n,\alpha,M_0,\Gamma,\epsilon)R^{-k}$.

Proof. Cover $\mathcal{S}^k_{\epsilon,\eta r}(v)\cap B_{1/4}(0)$ by balls $B_\kappa(Q_j)$, with $Q_j\in B_{1/4}(0)\cap\partial\Omega^\pm$, such that

$$B_{1/4}(0)\cap\partial\Omega^\pm\subset\bigcup_j B_\kappa(Q_j)$$

for $0 < \kappa(n, \alpha, M_0, \Gamma, \epsilon)$ the constant in Lemma 11.3. Note that we need at most $c(n, \alpha, M_0, \Gamma, \epsilon)$ such balls.

We now wish to apply Corollary 11.2 to the rescaled functions $\tilde{v}_i(x) = v(\kappa x + Q_i)$ in $B_1(0)$. However, a careful reader may object that $\tilde{v}_i$ is not in $\mathcal{A}(n, \alpha, M_0)$, since it is possible that $\tilde{h}(0) \neq 1$. However, $\tilde{v}_i(x) = c\tilde{h}(0)u^+(x\kappa + Q) - u^-(x\kappa + Q)$, where by Remark 3.1 we can control $0 < c < \infty$ by constants that only depend upon κ and α. Thus, by multiplying the positive part by a constant controlled by Γ, α, and M_0, we obtain a new function (which we also label $\tilde{v}_i$) which is in $\mathcal{A}(n, \alpha, M_0)$.

We now construct the desired covering in $B_1(0)$ for each $\tilde{v}_i$. Ensuring that $c(n, \alpha, M_0, \Gamma, \epsilon)$ is sufficiently large, we may reduce to arguing for $r < \eta$. We now use Corollary 11.2 to build a covering $\mathcal{U}_1$. If every r_x equals R, then the packing and covering estimates give the claim directly, since

$$R^{k-n} \operatorname{Vol}(B_R(\mathcal{S}^k_{\epsilon,\eta_0 R}(\tilde{v}_i) \cap B_1(0))) \leq \omega_n R^{k-n} \sum_{\mathcal{U}_1} (2R)^n = \omega_n 2^n \sum_{\mathcal{U}_1} r_x^k \leq c(n, \alpha, M_0, \Gamma, \epsilon).$$

If there exists an $r_x \neq R$, we use Corollary 11.2 to build a finite sequence of refined covers $\mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3, \ldots$ such that, for each for each i, the covering satisfies the following properties:

(A_i) Packing:

$$\sum_{x \in \mathcal{U}_i} r_x^k \leq c(n, \alpha, M_0, \Gamma, \epsilon)\Big(1 + \sum_{x \in \mathcal{U}_{i-1}} r_x^k\Big).$$

(B_i) Covering control:

$$\mathcal{S}^k_{\epsilon,\eta_0 R}(\tilde{v}_i) \cap B_1(0) \subset \bigcup_{x \in \mathcal{U}_i} B_{r_x}(x).$$

(C_i) Energy drop: for every $x \in \mathcal{U}_i$, either

$$r_x = R \quad \text{or} \quad \sup_{Q \in B_{2r_s}(s) \cap \partial\Omega^{\pm}} N(Q, 2r_s, \tilde{v}_i) \leq C(n, \alpha, M_0, \Gamma) - i\left(\tfrac{1}{2}\eta_0\right).$$

(D_i) Radius control:

$$\sup_{x \in \mathcal{U}_i} r_x \leq 10^{-i}.$$

If we can construct such a sequence of covers, then we claim that this process will terminate in finite time, *independent of* R. Recall that blow-ups of $\tilde{v}_i$ are homogeneous harmonic polynomials. Therefore

$$N(Q, 0, \tilde{v}_i) = \lim_{r \to \infty} N(Q, r, \tilde{v}_i) \geq 1$$

for all $Q \in \partial\Omega^{\pm}$. By Remark 4.8 we have that, for all $0 < r \leq 1$,

$$N(Q, r, \tilde{v}_i) \geq 1 - C(n, \alpha, M_0, \Gamma, \epsilon)$$

for all $p \in B_1(0)$. Therefore, we know that, for i large enough that

$$i > (C(n, \alpha, M_0, \Gamma, \epsilon) + C(n, \alpha, M_0, \Gamma, \epsilon) - 1)\frac{2}{\eta_0},$$

it must be the case that $r_x = R$ for all $x \in \mathcal{U}_i$. In this case, we will have the claim with a bound of the form

$$R^{k-n} \operatorname{Vol}(B_R(\mathcal{S}^k_{\epsilon,\eta_0 R}(\tilde{v}_i) \cap B_1(0))) \leq c(n, \alpha, M_0, \Gamma, \epsilon)^{C(n,\alpha,M_0,\Gamma,\epsilon)}.$$

Thus, we reduce to inductively constructing the required covers. Suppose we have already constructed $\mathcal{U}_{i-1}$ as desired. For each $x \in \mathcal{U}_{i-1}$ with $r_x > R$, we apply Corollary 11.2 at scale $B_{r_x}(x)$ to obtain a new collection of balls $\mathcal{U}_{i,x}$. From the assumption that $r_x \le \frac{1}{10}$ and the way that Hölder norms scale, it is clear that $\tilde{v}_i$ satisfies the hypotheses of Corollary 11.2 in $B_{r_x}(x)$ with the same constants. To check packing control, we have that

$$\sum_{y \in \mathcal{U}_{i,x}} r_y^k \le c(n, \alpha, M_0, \Gamma, \epsilon) r_x^k.$$

Covering control follows immediately from the statement of Corollary 11.2. Similarly, from hypothesis (C_{i-1}), we have that $\sup_{p \in B_{2r_x}(x)} N(Q, 2r_x, \tilde{v}_i) \le C(n, \alpha, M_0, \Gamma, \epsilon) - \frac{1}{2}(i-1)\eta_0$. Thus, the statement of Corollary 11.2 at scale $B_{r_x}(x)$ gives $\sup_{p \in B_{2r_y}(y)} N(Q, 2r_y, \tilde{v}_i) \le C(n, \alpha, M_0, \Gamma, \epsilon) - \frac{1}{2} i \eta_0$ for all $y \in \mathcal{U}_{i,x}$ with $r_y > R$. Radius control follows immediately from the fact that $\sup_{y \in \mathcal{U}_{i,x}} r_y \le \frac{1}{10} r_x \le 10^{-i}$.

Thus, if we let

$$\mathcal{U}_i = \{x \in \mathcal{U}_{i-1} \mid r_x = R\} \cup \bigcup_{x \in \mathcal{U}_{i-1},\, r_x > R} \mathcal{U}_{i,x}$$

then $\mathcal{U}_i$ satisfies the inductive claim.

To obtain the cover which proves the theorem, we simply scale each covering of $\mathcal{S}^k_{\epsilon, \eta_0 R/\kappa}(\tilde{v}_i) \cap B_1(0)$ to a covering of $\mathcal{S}^k_{\epsilon,\eta_0 R}(v) \cap B_\kappa(y_i)$ and sum over the $c(n, \alpha, M_0, \Gamma, \epsilon)$ such balls which cover $\mathcal{S}^k_{\epsilon,\eta_0 R}(v) \cap B_{1/4}(0)$. This completes the proof. ☐

Proof of Theorem 2.15. By Theorem 11.4, we have

$$\mathrm{Vol}(B_R(\mathcal{S}^k_{\epsilon,\eta_0 R}(v) \cap B_{1/4}(0))) \le C(n, \alpha, M_0, \Gamma, \epsilon) R^{n-k}.$$

Thus, let $r_0 = \eta_0$ and $r = \eta_0 R'$ for $0 < R' \le 1$. For any $r \le R \le R'$, by containment, we have

$$B_R(\mathcal{S}^k_{\epsilon,r}(v) \cap B_{1/4}(0)) \subset B_{R'}(\mathcal{S}^k_{\epsilon,r}(v) \cap B_{1/4}(0)) \subset \bigcup_i B_{2R'}(x_i),$$

where $\{x_i\}$ are the centers of the balls in the covering constructed in Theorem 11.4. Therefore, the estimates in Theorem 11.4 give

$$\begin{aligned}\mathrm{Vol}(B_R(\mathcal{S}^k_{\epsilon,r}(v) \cap B_{1/4}(0))) &\le C(n, \alpha, M_0, \Gamma, \epsilon) 2^n (R')^{n-k} \\ &\le C(n, \alpha, M_0, \Gamma, \epsilon) 2^n \left(\frac{R}{\eta_0}\right)^{n-k} \le C(n, \alpha, M_0, \Gamma, \epsilon) R^{n-k}\end{aligned}$$

by increasing our constant $C(n, \alpha, M_0, \Gamma, \epsilon)$.

For any $R' \le R$, by containment, we have

$$B_R(\mathcal{S}^k_{\epsilon,r}(v) \cap B_{1/4}(0)) \subset \bigcup_i B_{2R}(x_i),$$

where $\{x_i\}$ are the centers of the balls in the covering constructed in Theorem 11.4. In this case

$$\mathrm{Vol}(B_R(\mathcal{S}^k_{\epsilon,r}(v) \cap B_{1/4}(0))) \le C(n, \alpha, M_0, \Gamma, \epsilon) 2^n (R)^{n-k} \le C(n, \alpha, M_0, \Gamma, \epsilon) R^{n-k}$$

by increasing our constant $C(n, \alpha, M_0, \Gamma, \epsilon)$. This concludes the proof of Theorem 2.15. ☐

12. Proof of Corollary 2.17

In this section, we prove that $\mathrm{sing}(\partial\Omega^{\pm}) \subset \mathcal{S}^{k-3}_{\epsilon}(v)$ for ϵ small enough.

Lemma 12.1. *Let $v \in \mathcal{A}(n, \alpha, M_0)$ with $\|\ln(h)\|_{\alpha} \le \Gamma$. Then there exists $0 < \epsilon = \epsilon(M_0, \alpha, \Gamma)$ such that* $\mathrm{sing}(\partial\Omega^{\pm}) \cap B_{1/4}(0) \subset \mathcal{S}^{n-3}_{\epsilon}(v)$.

Proof. We must argue that there is an $\epsilon > 0$ such that, for all $Q \in \mathrm{sing}(\partial\Omega^{\pm}) \cap B_1(0)$ and all radii $0 < r$,

$$\int_{B_1(0)} |T_{Q,r}v - P|^2 \, dV \ge \epsilon$$

for all $(n-2)$-symmetric functions P.

If P is $(n-2)$-symmetric, P only depends upon two variables. By complex analysis all homogeneous harmonic polynomials in two dimensions are of the form $q(z) = c(x + iy)^k$. By Theorem 2.14 (2), we need only consider $k \ge 2$. Hence, the zero set Σ_q of any $\mathrm{Re}(q)$ is the union of an even number of infinite rays equidistributed in angle. If we label the connected components of $\mathbb{R}^2 \setminus \Sigma_q$ as $\{U_i\}$, we see that by the maximum principle, the sign of q must change from one U_i to another contiguous U_j.

Thus, the zero set of P is $\Sigma_P = \Sigma_q \times \mathbb{R}^{n-2}$ for some homogeneous harmonic polynomial $\mathrm{Re}(q) : \mathbb{R}^2 \to \mathbb{R}$ of degree ≥ 2. We label the connected components of $\mathbb{R}^n \setminus \Sigma_q \times \mathbb{R}^{n-2}$ as $\{W_i\}$.

Now, we claim that there is a constant, $0 < c(M_0, \Gamma, \alpha) \le 1$, such that one of the following estimates must hold:

$$\mathcal{H}^n\Big(T_{Q,r}\Omega^- \cap \bigcup_i \{W_i : P > 0 \text{ on } W_i\} \cap B_1(0)\Big) \ge c, \qquad \text{(estimate 1)}$$

$$\mathcal{H}^n\Big(T_{Q,r}\Omega^+ \cap \bigcup_i \{W_i : P < 0 \text{ on } W_i\} \cap B_1(0)\Big) \ge c. \qquad \text{(estimate 2)}$$

Note that by Theorem 2.6 (2), we need only consider P with degree $\le d(M_0) < \infty$. Reducing to $\mathbb{R}^2$, since the rays of Σ_q are equidistributed, for q of degree k, the connected components occupy a sector of aperture π/k. Thus, if $B_{1/M_0}(A_1^{\pm}(0)) \subset T_{Q,r}\Omega^{\pm}$ is the ball guaranteed by the corkscrew condition, then, for $c = (4M_0^n)^{-1}$, there exists an integer $k(M_0)$ such that

$$\mathcal{H}^n(B_{1/M_0}(A_1^{\pm}(0)) \cap \{P \cdot T_{Q,r}v < 0\}) \ge c$$

for all P with degree $\ge k(M_0)$.

For P with degree $\le k(M_0)$, we argue by contradiction. Suppose that no such constant exists. Then there would be a sequence of functions $v_j \in \mathcal{A}(n, \alpha, M_0)$ with points $Q_j \in B_{1/4}(0)$ and radii $0 < r_j \le \frac{1}{2}$ and zero sets Σ_{P_j} for P_j satisfying $2 \le \mathrm{degree}(P_j) \le k(M_0)$ such that the scaled and translated mutual boundaries $T_{Q_j,r_j}\partial\Omega_j^{\pm}$ satisfy the properties

$$\mathcal{H}^n\Big(T_{Q_j,r_j}\Omega_j^- \cap \bigcup_i \{W_{i,j} : P_j > 0 \text{ on } W_{i,j}\} \cap B_1(0)\Big) \to 0,$$

$$\mathcal{H}^n\Big(T_{Q_j,r_j}\Omega_j^+ \cap \bigcup_i \{W_{i,j} : P_j < 0 \text{ on } W_{i,j}\} \cap B_1(0)\Big) \to 0.$$

By Lemma 3.6 there exists a subsequence for which $T_{Q_j,r_j}\partial\Omega_j^{\pm}$ converge locally in the Hausdorff metric to a limit set $A \subset \mathbb{R}^n$. By Theorem 3.8, A must be the mutual boundary of a pair of two-sided NTA domains $\Omega_\infty^{\pm}$ with constant $2M_0$. Furthermore, up to scaling and rotation, the space of homogeneous harmonic functions of two variables in $\mathbb{R}^n$ with $2 \le \text{degree}(P) \le k(M_0)$ is finite-dimensional. Since the space of rotations is compact, we may find a subsequence Σ_{P_j} which converges to Σ_{P_∞} locally in the Hausdorff metric for some $(n-2)$-symmetric P_∞. This implies that

$$\mathcal{H}^n\Big(\Omega_\infty^- \cap \bigcup_i \{W_{i,\infty} : P_\infty > 0 \text{ on } W_{i,\infty}\} \cap B_1(0)\Big) = 0, \tag{12-1}$$

$$\mathcal{H}^n\Big(\Omega_\infty^+ \cap \bigcup_i \{W_{i,\infty} : P_\infty < 0 \text{ on } W_{i,\infty}\} \cap B_1(0)\Big) = 0. \tag{12-2}$$

Indeed, if there were $p \in \bigcup_i \{W_{i,\infty} : P_\infty > 0 \text{ on } W_{i,\infty}\} \cap B_1(0)$ such that $p \in \Omega_\infty^-$, since $W_{i,\infty}$ and Ω^- are open, there would exist a ball $B_\delta(p) \subset \Omega^- \cap W_{i,\infty}$. Therefore, since $\Sigma_{P_j} \to \Sigma_{P_\infty}$ and $T_{Q_j,r_j}\partial\Omega_j^{\pm} \to A$ locally in the Hausdorff metric, for all j sufficiently large, $B_{\delta/2}(p) \subset W_{i,j} \cap T_{Q_j,r_j}\partial\Omega_j^-$. This is a contradiction. The other equation follows identically.

Now we claim that $A \cap B_1(0) = \Sigma_{P_\infty} \cap B_1(0)$. Suppose not, then there exists a point $p \in \Sigma_{P_\infty}$ with $p \notin A$ or there exists a point $Q \in A$ such that $Q \notin \Sigma_{P_\infty}$. In the former case, suppose $\text{dist}(Q, A) > \delta$. Then $B_\delta(p)$ must intersect at least two contiguous connected components, $W_{i,\infty}$ and $W_{j,\infty}$. Since they are contiguous, the sign of P_∞ must be positive on one and negative on the other. This contradicts (12-1). Similarly, if there exists a point $Q \in A$ such that $Q \notin \Sigma_{P_\infty}$ then there exists a ball $B_\delta(Q)$ which intersects both $\Omega_\infty^{\pm}$ but which is contained in a single $W_{i,\infty}$. This also contradicts (12-1).

However, if P_∞ is $(n-2)$-symmetric with degree ≥ 2, then Σ_{P_∞} does not divide $\mathbb{R}^n$ into two connected components. This contradicts our assumption that $A = \Sigma_{P_\infty}$ was the mutual boundary of a pair of two-sided NTA domains with constant $2M_0$. Therefore, such a constant $0 < c = c(M_0, \Gamma, \alpha)$ must exist.

Without loss of generality, we assume (estimate 1) holds. By Lemma A.2 we may find a radius $0 < r = r(M_0, \Gamma, \alpha)$ such that $\mathcal{H}^n(B_r(T_{Q,r}\partial\Omega^{\pm})) < \frac{1}{20}c(\alpha, M_0, \Gamma)$. Now, consider

$$p \in \bigcup \{W_i : P > 0 \text{ on } W_i\} \cap B_1(0) \setminus B_r(T_{Q,r}\partial\Omega^{\pm}).$$

By Lemma 3.5, $|T_{Q,r}v(p)| \ge c'$ for a constant $c' = c'(M_0, \Gamma, \alpha)$. Thus

$$\begin{aligned}\int_{B_1(0)} |T_{Q,r}v - P|^2\, dV &\ge \int_{B_1(0)\cap T_{Q,r}\partial\Omega^- \cap \bigcup_i \{W_i : P>0 \text{ on } W_i\}} |T_{Q,r}v - P|^2\, dV \\ &\ge \tfrac{19}{20} c(\alpha, M_0, \Gamma) c'(\alpha, M_0, \Gamma)^2.\end{aligned}$$

If (estimate 2) holds, an identical argument with signs switched proves the claim. □

Remark 12.2. The argument above can be modified to show that there is an $\epsilon' > 0$ such that if $Q \in \partial\Omega$ but $Q \notin \mathcal{S}^{n-3}_{\epsilon',r_0}$, then $Q \notin \mathcal{S}^{n-2}_{\epsilon',r_0}$. Indeed, if $Q \notin \mathcal{S}^{n-3}_{\epsilon',r_0}$, then there exists a radius $r_0 \le r$ and an $(n-2)$-symmetric function P such that $\|T_{Q,r}v - P\|^2_{L^2(B_1(0))} \le \epsilon'$. However, by taking $\epsilon' < \epsilon(\alpha, M_0, \Gamma)$ in Lemma 12.1, we see that P must be $(n-1)$-symmetric.

Appendix

The purpose of this section is to justify Lemma 3.9. We use the language of porous sets. For a nonempty set $E \subset \mathbb{R}^n,\ x \in E$, and radius $0 < r$, we write

$$P(E, x, r) = \sup\{0, h : h > 0, B_h(y) \subset B_r(x) \setminus E \text{ for some } y \in B_r(x)\}. \tag{A-1}$$

For $\alpha > 0$, we say that E is *α-porous* if

$$\liminf_{r \to 0} \frac{P(E, x, r)}{r} > \alpha \tag{A-2}$$

for all $x \in E$.

We shall say that E is *α-porous down to scale r_0* if

$$\frac{P(E, x, r)}{r} > \alpha \tag{A-3}$$

for all $x \in E$ and for all $r_0 \le r$.

Remark A.1. By definition, for $\Omega^\pm \in \mathcal{D}(n, \alpha, M_0)$, the boundary $\partial\Omega^\pm$ is $1/M_0$-porous. Similarly, $B_r(\partial\Omega^\pm)$ is $1/(2M_0)$-porous down to scale $r_0 = 2rM_0$.

Lemma A.2. *Let $E \subset \mathbb{R}^n$ be a nonempty, bounded set, $E \subset [0, 1]^n$ with $\mathbf{0} \in E$. If E is α-porous down to scale $r_0 \ll 1$, then there are $k = k(\alpha),\ k' = k'(n)$, and $N \le -\log_2(r_0)/(k + k')$ such that*

$$\mathrm{Vol}(E) \le \left(1 - \frac{1}{2^{k+k'(n)}}\right)^N.$$

Moreover, there exists $0 < \epsilon = \epsilon(\alpha, n)$ and a constant $c(n, \alpha)$ such that

$$\mathcal{M}^{n-\epsilon}_{r_0}(E) \le (1 - c)^N.$$

Proof. Let $\{Q^i_j\}_j$ be the collection of dyadic subcubes $Q^i_j \subset [0, 1]^n$ with $\ell(Q^i_j) = 2^{-i}$. Let $k \in \mathbb{N}$ be the smallest number such that $2^{-k} \le \alpha$. Note that, for any $y \in [0, 1]^n$ with $B_{\alpha/2}(y) \subset [0, 1]^n$, there exists a dyadic cube $Q^{k+k'(n)}_j \subset B_{\alpha/4}(y)$ where $k'(x)$ is the smallest integer such that $k'(n) \ge 2 + \frac{1}{2}\log_2(n)$. Let $\frac{1}{2}Q^i_j$ denote an axis-parallel cube with the same center as Q^i_j but side length half that of Q^i_j.

Now we apply the standard argument. Tile $[0, 1]^n$ by $Q^{k+k'(n)}_j$. By our porosity assumption, there exists a $Q^{k+k'(n)}_{j'}$ which does not intersect E. Thus

$$\mathrm{Vol}(E) \le \sum_{j \ne j'} \mathrm{Vol}(Q^{k+k'(n)}_j) \le (2^{(k+k'(n))n} - 1)2^{(-k-k'(n))n} \le \left(1 - \frac{1}{2^{k+k'(n)}}\right).$$

Now, within each of the $Q^{k+k'(n)}_j$ which intersects E, either E intersects $\frac{1}{2}Q^{k+k'(n)}_j$ or it doesn't. If $E \cap \frac{1}{2}Q^{k+k'(n)}_j = \varnothing$, then we tile $Q^{k+k'(n)}_j$ by cubes $\{Q^{2(k+k'(n))}_\ell\}_\ell$ and overestimate

$$\begin{aligned}\mathrm{Vol}(E \cap Q^{k+k'(n)}_j) &\le \sum_{\ell : Q^{2(k+k'(n))}_\ell \cap (E \cap Q^{k+k'(n)}_j) \ne \varnothing} \mathrm{Vol}(Q^{2(k+k'(n))}_\ell) \\ &\le (2^{2(k+k'(n))n} - 1)2^{-2(k+k'(n))n}\,\mathrm{Vol}(Q^{k+k'(n)}_j) \le \left(1 - \frac{1}{2^{k+k'(n)}}\right)\mathrm{Vol}(Q^{k+k'(n)}_j).\end{aligned}$$

If $E \cap \frac{1}{2} Q_j^{k+k'(n)} \neq \varnothing$, then there exists a ball $B_{2^{-k-k'(n)-1}}(x) \subset Q_j^{k+k'(n)}$ centered on $x \in E$. By our porosity assumption and choice of $k'(n)$, we can still tile $Q_j^{k+k'(n)}$ by $Q_\ell^{2(k+k'(n))}$ and be guaranteed that at least one such subcube does not intersect $E \cap Q_j^{k+k'(n)}$. Thus, we overestimate in the same manner as above.

We can continue, inductively, only stopping at the first N such that $2^{-(N+1)(k+k'(n))} < r_0$. This gives the desired bound

$$\mathrm{Vol}(E) \leq \left(1 - \frac{1}{2^{k+k'(n)}}\right)^N.$$

Taking a bit more care, we can actually improve these estimates. Let $0 < \epsilon = \epsilon(\alpha, n)$ be such that

$$\left(1 - \frac{1}{2^{k+k'(n)}}\right) < \left(2^{\epsilon(k+k'(n))} - \frac{1}{2^{(k+k'(n))(n-\epsilon)}}\right) < 1.$$

Then we bound $\mathcal{M}_{r_0}^{n-\epsilon}(E)$ as follows:

$$\begin{aligned}\mathcal{M}_{r_0}^{n-\epsilon}(E) &= \inf\left\{\sum_i r^{n-\epsilon} : x_i \in E,\ r_0 \leq r, E \subset \bigcup_i B_r(x_i)\right\} \\ &\leq \sum_j \ell(Q^{N(k+k'(n))})^{n-\epsilon} \leq \left(2^{\epsilon(k+k'(n))} - \frac{1}{2^{(k+k'(n))(n-\epsilon)}}\right)^N. \qquad \square\end{aligned}$$

As immediate corollaries, we have the following statements.

Corollary A.3. *If $E \subset \mathbb{R}^n$ is α-porous, then there exists $0 < \epsilon = \epsilon(\alpha, n)$ such that $\overline{\dim}_{\mathcal{M}}(E) \leq n - \epsilon$.*

Proof. Recall that $\overline{\dim}_{\mathcal{M}}(E) = \inf\{s : \mathcal{M}^{*,s}(E) = 0\}$ and that $\mathcal{M}^{*,s}(E) = \limsup_{r_0 \to 0} \mathcal{M}_{r_0}^{n-\epsilon}(E)$.

By taking $0 < \epsilon$ to be as in Lemma A.2, we have

$$\mathcal{M}_{r_0}^{n-\epsilon}(E) \leq \left(2^{\epsilon(k+k'(n))} - \frac{1}{2^{(k+k'(n))(n-\epsilon)}}\right)^N \leq (1-c)^N,$$

where $c = c(\alpha, n, \epsilon)$ and $N = N(\alpha, n, r_0)$, as in the previous lemma. Thus, letting $r_0 \to 0$ and $N \to \infty$ we have that $\mathcal{M}^{n-\epsilon}(E) = 0$. $\square$

Recalling Remark A.1, Corollary A.3 gives Lemma 3.9.

Corollary A.4. *Let $\Sigma \subset \mathbb{R}^n$ be the mutual boundary of a pair of unbounded two-sided NTA domains with NTA constant $1 < M_0$. Then, there exists $0 < \epsilon = \epsilon(M_0, n)$ such that $\overline{\dim}_{\mathcal{M}}(E) \leq n - \epsilon$.*

Acknowledgements

The author would like to thank T. Toro, whose generous insight, suggestions, and support were essential to bring this project to fruition. The author is also grateful to the reviewers for their many excellent comments which dramatically improved this paper. This research was partially supported by NSF grant DMS-1361823.

References

[Almgren 1979] F. J. Almgren, Jr., "Dirichlet's problem for multiple valued functions and the regularity of mass minimizing integral currents", pp. 1–6 in *Minimal submanifolds and geodesics* (Tokyo, 1977), edited by M. Obata, North-Holland, Amsterdam, 1979. MR Zbl

[Azzam et al. 2019] J. Azzam, M. Mourgoglou, X. Tolsa, and A. Volberg, "On a two-phase problem for harmonic measure in general domains", *Amer. J. Math.* **141**:5 (2019), 1259–1279. MR Zbl

[Badger 2011] M. Badger, "Harmonic polynomials and tangent measures of harmonic measure", *Rev. Mat. Iberoam.* **27**:3 (2011), 841–870. MR Zbl

[Badger 2013] M. Badger, "Flat points in zero sets of harmonic polynomials and harmonic measure from two sides", *J. Lond. Math. Soc.* (2) **87**:1 (2013), 111–137. MR Zbl

[Badger et al. 2017] M. Badger, M. Engelstein, and T. Toro, "Structure of sets which are well approximated by zero sets of harmonic polynomials", *Anal. PDE* **10**:6 (2017), 1455–1495. MR Zbl

[Badger et al. 2020] M. Badger, M. Engelstein, and T. Toro, "Regularity of the singular set in a two-phase problem for harmonic measure with Hölder data", *Rev. Mat. Iberoam.* **36**:5 (2020), 1375–1408. MR Zbl

[Cheeger and Naber 2013] J. Cheeger and A. Naber, "Quantitative stratification and the regularity of harmonic maps and minimal currents", *Comm. Pure Appl. Math.* **66**:6 (2013), 965–990. MR Zbl

[Cheeger et al. 2015] J. Cheeger, A. Naber, and D. Valtorta, "Critical sets of elliptic equations", *Comm. Pure Appl. Math.* **68**:2 (2015), 173–209. MR Zbl

[De Lellis et al. 2018] C. De Lellis, A. Marchese, E. Spadaro, and D. Valtorta, "Rectifiability and upper Minkowski bounds for singularities of harmonic Q-valued maps", *Comment. Math. Helv.* **93**:4 (2018), 737–779. MR Zbl

[Edelen and Engelstein 2019] N. Edelen and M. Engelstein, "Quantitative stratification for some free-boundary problems", *Trans. Amer. Math. Soc.* **371**:3 (2019), 2043–2072. MR Zbl

[Engelstein 2016] M. Engelstein, "A two-phase free boundary problem for harmonic measure", *Ann. Sci. École Norm. Sup.* (4) **49**:4 (2016), 859–905. MR Zbl

[Garnett and Marshall 2005] J. B. Garnett and D. E. Marshall, *Harmonic measure*, New Math. Monogr. **2**, Cambridge Univ. Press, 2005. MR Zbl

[Han and Lin 1994] Q. Han and F.-H. Lin, "Nodal sets of solutions of elliptic differential equations", 1994. Unpublished manuscript.

[Hochman 2015] M. Hochman, "On self-similar sets with overlaps and inverse theorems for entropy in $\mathbb{R}^d$", 2015. To appear in Mem. Amer. Math. Soc. arXiv 1503.09043

[Jerison and Kenig 1982] D. S. Jerison and C. E. Kenig, "Boundary behavior of harmonic functions in nontangentially accessible domains", *Adv. Math.* **46**:1 (1982), 80–147. MR Zbl

[Kenig and Toro 2006] C. Kenig and T. Toro, "Free boundary regularity below the continuous threshold: 2-phase problems", *J. Reine Angew. Math.* **596** (2006), 1–44. MR Zbl

[Kenig et al. 2009] C. Kenig, D. Preiss, and T. Toro, "Boundary structure and size in terms of interior and exterior harmonic measures in higher dimensions", *J. Amer. Math. Soc.* **22**:3 (2009), 771–796. MR Zbl

[Mattila 1995] P. Mattila, *Geometry of sets and measures in Euclidean spaces: fractals and rectifiability*, Cambridge Stud. Adv. Math. **44**, Cambridge Univ. Press, 1995. MR Zbl

[Naber and Valtorta 2017] A. Naber and D. Valtorta, "Rectifiable-Reifenberg and the regularity of stationary and minimizing harmonic maps", *Ann. of Math.* (2) **185**:1 (2017), 131–227. MR Zbl

Received 24 Apr 2019. Revised 19 Nov 2021. Accepted 23 Sep 2022.

SEAN MCCURDY: smccurdy@ntnu.edu.tw
National Taiwan Normal University, Taipei City, Taiwan

ANALYSIS AND PDE
Vol. 17 (2024), No. 4, pp. 1175–1236
DOI: 10.2140/apde.2024.17.1175

ON COMPLETE EMBEDDED TRANSLATING SOLITONS OF THE MEAN CURVATURE FLOW THAT ARE OF FINITE GENUS

GRAHAM SMITH

We desingularise the union of three Grim paraboloids along Costa–Hoffman–Meeks surfaces in order to obtain complete embedded translating solitons of the mean curvature flow with three ends and arbitrary finite genus.

1. Introduction

1.1. ***Main result.*** For the purposes of this paper, a *mean curvature flow* (*MCF*) *soliton* is a complete surface in $\mathbb{R}^3$ whose evolution under the mean curvature flow is given by translation. In other words, up to rescalings and rigid motions of the ambient spacetime, it is a solution of what we will call the *MCFS equation*

$$H + \langle N, e_z \rangle = 0, \tag{1-1}$$

where H here denotes the mean curvature of the surface, N its unit normal vector field, and e_z the unit vector in the direction of the z-axis. We refer the reader to the review of Martín, Savas-Halilaj and Smoczyk [Martín et al. 2015] for a good overview of the theory of MCF solitons at the time of writing.

We use surgery to construct embedded MCF solitons with three ends and arbitrary finite genus. Before stating our result, we describe the two components of our construction. First, given a positive integer g, the *Costa–Hoffman–Meeks* (*CHM*) *surface* of genus g, denoted by C_g, is a properly embedded minimal surface in $\mathbb{R}^3$ with three ends, each of which may be taken to be a graph over an unbounded annulus in $\mathbb{R}^2$; see [Hoffman and Meeks 1990; Weber 2005]. For $0 \leq k \leq g$, this surface is invariant under reflection

MSC2010: 53C44.
Keywords: mean curvature flow, soliton, singular perturbation, Costa–Hoffman–Meeks surface, Grim paraboloid, bowl soliton.

in the vertical plane making an angle of $k\pi/(g+1)$ with the x-axis at the origin. We call the group of symmetries of $\mathbb{R}^3$ generated by these reflections the group of *horizontal symmetries* of C_g.[1] Next, the *Grim paraboloid* (also known as the *bowl soliton*) is defined to be the unique simply connected MCF soliton which is symmetric under revolution about the z-axis. It is known (see [Clutterbuck et al. 2007]) that this surface is asymptotic at infinity to a vertical translate of the graph of

$$\tfrac{1}{2}r^2 - \log(r),$$

where r here denotes the distance in $\mathbb{R}^2$ to the origin.

Theorem A. *For all $g \in \mathbb{N}$ and for all sufficiently small ϵ, there exists a complete, embedded MCF soliton $\Sigma_{g,\epsilon}$ of genus g with three ends. Furthermore, letting $R := \epsilon^{-1/\lambda}$ for some $\lambda \in \,]4,5[$, we may suppose*:

(1) *$\Sigma_{g,\epsilon} \setminus (B(\epsilon R) \times \mathbb{R})$ consists of three connected components, each of which converges towards the same Grim paraboloid as ϵ tends to* 0.

(2) *Upon rescaling by a factor of $1/\epsilon$, $\Sigma_{g,\epsilon} \cap (B(\epsilon R) \times \mathbb{R})$ converges to C_g as ϵ tends to* 0.

(3) *$\Sigma_{g,\epsilon}$ is invariant under the group of horizontal symmetries of C_g.*

Remark. Theorem A follows immediately from Theorems 7.1.3 and 7.1.4, below.

Remark. All notation and terminology used in this paper is explained in detail in Appendix A. Recall, in particular, that, by elliptic regularity, all standard modes of convergence of smooth, embedded solutions of parametric elliptic functionals to smooth, embedded solutions of the same functionals are equivalent over any compact region.

Remark. The constants that appear in Theorem A have the following geometric significance. The quantity ϵ determines the scaling factor of the CHM surface. Roughly speaking, it is the "neck radius" of $\Sigma_{g,\epsilon}$. The quantity R determines how far along the end of the CHM surface the surgery is carried out. For the construction to work, ϵ and R must converge in tandem to 0 and infinity respectively, hence the condition $R^\lambda \epsilon = 1$. Distinct values of ϵ ought to yield distinct surfaces. Indeed, a refinement of our result ought to yield a continuous family $(\Sigma_{g,\epsilon})_{\epsilon < r_0}$ of distinct embedded MCF solitons with neck radii converging to 0. However, our current argument, which uses the Schauder fixed-point theorem, does not yield such fine control over the surfaces constructed.

1.2. ***Techniques.*** The proof of Theorem A follows the standard desingularisation construction for minimal surfaces originally described in [Kapouleas 1990; 1995; 1997]. In simple terms, we first use surgery to construct an approximate MCF soliton $\widehat{\Sigma}_{g,\epsilon}$ and then apply a fixed-point argument to prove the existence of an actual MCF soliton lying nearby in some suitable function space. As in all singular perturbation constructions, this is much easier said than done, and the main challenge lies in deriving the many nontrivial analytic estimates required to make the perturbation argument work.

[1] Hoffman and Meeks showed that the complete symmetry group of C_g is the dihedral group generated by the elements A and B, where A is reflection in the $(x-z)$-plane and B is rotation by an angle of $k\pi/(g+1)$ about the z-axis followed by reflection in the $(x-y)$-plane.

The use of CHM surfaces in Kapouleas' construction presents particular difficulties on account of their low orders of rotational symmetry. Indeed, rotational symmetries often serve in such constructions to improve decay rates and thus in turn provide stronger estimates. This phenomenon is well illustrated by the case of bounded solutions $u : S^1 \times [0, \infty[\to \mathbb{R}$ of Laplace's equation $\Delta u = 0$. By separation of variables, all such solutions have the form

$$u(\theta, t) = \sum_{m \in \mathbb{Z}} a_m e^{im\theta} e^{-|m|t}.$$

In particular, when u has n-th order rotational symmetry, all of its coefficients of order $0 < |m| < n$ vanish, so that the difference $(u - a_0)$ decays like e^{-nt}. Since this argument does not apply when CHM surfaces are used, we obtain our estimates by introducing instead, in Section 4.2, what we call the *hybrid norm*. This functional norm, which is a combination of Hölder and Sobolev norms, encapsulates the singular nature of our construction as ϵ tends to 0. Its main properties, established in Lemma 4.3.1, follow readily from the Sobolev embedding theorem and classical interpolation inequalities, and play a key role in the derivation of various estimates throughout the rest of the paper.

Finally, before reviewing our argument, it is worth highlighting an ingenious variant of the desingularisation construction for CHM surfaces used in [Hauswirth and Pacard 2007; Mazzeo and Pacard 2001; Morabito 2009]. In each of these papers, it is observed that the Jacobi operator $\hat{J}_{g,\epsilon}$ of the approximate minimal surface $\widehat{\Sigma}_{g,\epsilon}$ is, modulo a conformal transformation when necessary, *intrinsically* close to the Jacobi operator of the original CHM surface. A direct perturbation argument then yields a priori estimates for the norm of its Green's operator, thereby bypassing one of the main technical challenges of the perturbation part of the construction. In addition, in these works, the initial surgery is carried out in a different manner than in [Kapouleas 1990; 1995; 1997], more pleasing to the geometric eye, though it is not clear to us whether this actually leads to simpler estimates. Regardless, their argument cannot be applied in the present case where the Jacobi operator of the joined surface is not intrinsically of the correct type.

The proof is organised as follows.

1.2.1. *Rotationally symmetric Grim surfaces.* We will desingularise the join of a CHM surface with three rotationally symmetric Grim ends, that is, unbounded, rotationally symmetric MCF solitons in $\mathbb{R}^3$. The geometry of CHM surfaces is well understood (see, for example, [Hoffman and Meeks 1990; Weber 2005]) and the large-scale geometry of rotationally symmetric Grim ends has been studied by Clutterbuck, Schnürer and Schulze [Clutterbuck et al. 2007]. In Section 2, we study the small-scale geometry of rotationally symmetric Grim ends, which has not previously been addressed in the literature.

Rotationally symmetric Grim ends exhibit a dual nature over the region of interest to us. Indeed, they are roughly catenoidal towards the lower end of this region, and roughly parabolic towards its upper end. This presents us with our first main challenge, which we address via the following algebraic trick. We introduce two abstract variables, representing respectively the catenoidal part and the parabolic part of the Grim end. We then construct formal solutions to the MCFS equation in terms of these variables, and obtain the desired formulae upon applying the contraction mapping theorem to their partial sums.

The main results of this section are Theorems 2.1.1 and 2.2.1, which provide asymptotic formulae for the profiles of rotationally symmetric Grim ends over the large and small scales respectively.

1.2.2. *Green's operators.* Our perturbation argument requires estimates for the norm of a Green's operator of the MCFS Jacobi operator of the approximate MCF soliton. These are in turn derived from estimates of the norms of the corresponding operators of CHM surfaces and rotationally symmetric Grim ends. Green's operators of Jacobi operators of CHM surfaces are well understood; see, for example, [Hauswirth and Pacard 2007; Morabito 2009; Nayatani 1993; Pacard 2008]. In Sections 3 and 4, we study the Green's operators of the MCFS Jacobi operators, first of Grim paraboloids, then of rotationally symmetric Grim ends. The former are relatively straightforward, but the latter present us with our second main challenge, namely, to address the singularities that catenoids produce as their neck radii tend to 0. This simple phenomenon, which we call the *vanishing neck problem*, will be responsible for the introduction of a number of technical constructions, as we now proceed to explain.

To begin with, in Section 4.1, we modify the Jacobi operator in two ways. First, we introduce the *modified MCFS Jacobi operator*, which measures the first-order variation of mean curvature arising from first-order perturbations of the surface in the direction of a suitable modification of the unit normal vector field. At this stage, this modification serves to reduce the divergence rates of the coefficients of the Green's operators as the neck radii vanish. We underline that, since different modifications are made on different scales, the precise definition of this operator varies according to context (the general framework, unifying these definitions, is described in Section 5.4). Next, we introduce *canonical extensions* of operators across the region within the neck, which allow the modified MCFS Jacobi operators of distinct rotationally symmetric Grim ends to be compared as if they were all defined over $\mathbb{R}^2$.

Notably, even with these modifications, the vanishing neck problem still imposes restrictions on the way in which ϵ and R tend respectively to 0 and infinity. Indeed, it is precisely at this stage that we require that ϵR^5 tend to infinity in the statement of Theorem A, for otherwise we could not guarantee decay in Lemmas 4.2.1 and 4.2.2.

The main result of these two sections is Theorem 4.1.1, which provides the required uniform estimates for the Green's operators of the modified MCFS Jacobi operators of rotationally symmetric Grim ends. We prove this result using a perturbation argument. To this end, we examine the differences between the modified MCFS Jacobi operators of Grim paraboloids and those of rotationally symmetric Grim ends. We decompose these differences into regular and singular components. In Section 4.2, we show that the operator norms of the regular components tend to 0 as ϵ tends to 0, and in Section 4.3, making use of the hybrid norm described above, we prove the same result for the singular components. In particular, we see that an adequate treatment of the vanishing neck problem already calls for the hybrid norm, which will play a larger role later on in the construction.

1.2.3. *Surgery and the deformation family.* In Section 5, we describe the surgery operation used to construct the approximate MCF soliton $\widehat{\Sigma}_{g,\epsilon}$ as well as the deformation family around this surface within which the actual MCF soliton $\Sigma_{g,\epsilon}$ will be found. The surgery operation is straightforward and is described in Section 5.1. Simply put, the ends of the CHM surface are amputated, suitably chosen rescaled rotationally symmetric Grim ends are grafted in their place, and the join is smoothed out using cut-off functions. The construction of the deformation family about $\widehat{\Sigma}_{g,\epsilon}$ is more technical and is carried out in Section 5.2.

The challenge in understanding (and explaining!) this construction arises from the fact that four different families of deformations must be considered. The first concerns deformations in the direction of a suitable modification of the unit normal vector field. We refer to the resulting first-order perturbations of the surface as *microscopic perturbations*, since they decay at infinity. The remaining three families involve variations of the logarithmic parameters of the ends, starting far inside the locus of surgery, and vertical translations of the ends, starting far inside and far outside the locus of surgery respectively. We refer to the resulting first-order perturbations as *macroscopic perturbations*, since they remain large at infinity.

We associate to each of the four classes of perturbation described above the operator of first-order variation of the MCFS functional about $\widehat{\Sigma}_{g,\epsilon}$. We denote these operators by $\hat{J}_\epsilon$, X_ϵ, Y_ϵ and Z_ϵ respectively. Understanding their analytic properties is key to estimating the norm of the Green's operator of the modified MCFS Jacobi operator of $\widehat{\Sigma}_{g,\epsilon}$, and we conclude this section by deriving preliminary estimates in Sections 5.3, 5.4 and 5.5.

1.2.4. *Constructing the Green's operator.* In Section 6, we construct a Green's operator of the modified MCFS Jacobi operator of $\widehat{\Sigma}_{g,\epsilon}$, together with estimates of its operator norm. This section constitutes the hardest technical part of the paper. The determination of sufficiently strong estimates is made possible, on the one hand, by the correct choice of functional norms over the different components of $\widehat{\Sigma}_{g,\epsilon}$ and, on the other, by the use of the hybrid norm described above.

The estimates for the norm of the Green's operator are obtained in Sections 6.3, 6.4 and 6.5 via a classical iteration process which we call the "ping-pong" argument. This process, which is common to all singular perturbation constructions, involves passing successive error terms back and forth over the join region. From a conformal perspective, the join region consists of cylinders which become very long as ϵ tends to 0. More explicitly, if $R = \epsilon^{-1/\lambda}$, then these cylinders are roughly of length $(\lambda - 1)\,\mathrm{Ln}(R)$. The estimates we require to ensure the convergence of the iteration process then follow from the fact that bounded harmonic functions decay exponentially over long cylinders. In particular, we maximise decay by choosing λ as large as possible. We have already seen in Section 1.2.2, above, that λ must be less than 5. It turns out that $\lambda \in \,]4, 5[$ is sufficient for our purposes, thus explaining the condition imposed in the statement of Theorem A. We believe that the ideas underlying this technique are best illustrated by the simplest version of this construction, used in the theory of Morse homology, and described in detail in Section 2.5 of [Schwarz 1993].

The first main results of this section are Theorems 6.3.1 and 6.4.1, which provide estimates for the norms of the operators used in the two stages of the iteration process. In addition, Theorems 6.5.2, 6.5.3 and 6.5.4 provide estimates for the norms of the different components of the Green's operator that we construct.

1.2.5. *Existence and embeddedness.* Finally, in Section 7 we prove Theorem A by applying the Schauder fixed-point theorem to the MCFS functional about the approximate MCF soliton $\widehat{\Sigma}_{g,\epsilon}$. First, we determine estimates for the MCFS functional up to second order about $\widehat{\Sigma}_{g,\epsilon}$. Then, using the estimates obtained in Section 6, we prove existence in Theorem 7.1.3, and we prove embeddedness in Theorem 7.1.4 using a straightforward geometric argument.

1.3. ***Notation.*** In order not to be overwhelmed by a deluge of constants, throughout the paper we use the following notation, which we have found to be of great help. First, given two variable quantities a and b, we will write

$$a \lesssim b \tag{1-2}$$

to mean that there exists a constant C, which for our purposes we consider universal, such that

$$a \leq Cb.$$

Next, given a function f and a sequence of functions (g_m), we will write

$$f = \mathrm{O}(g_m) \tag{1-3}$$

to mean that there exists a sequence (C_m) of constants, which for our purposes we again consider universal, such that the relation

$$|D^m f| \leq C_m g_m$$

holds pointwise for all m. The indexing variable of the sequence (g_m) should be clear from the context. In certain cases, every element of this sequence may be the same. It should also be clear from the context when this occurs. All other notation and terminology is explained in detail in Appendix A.

2. Rotationally symmetric Grim surfaces

2.1. ***The large scale.*** We define a *Grim surface* to be any unit-speed MCF soliton which is a graph over some domain in $\mathbb{R}^2$. We define a *Grim end* to be a Grim surface which is defined over some unbounded annulus $A(a, \infty)$. These will be studied in more detail in Section 4. In this section, we study rotationally symmetric Grim surfaces defined over some annulus $A(a, b)$. We first recall the general formula for such surfaces. Let u be a twice differentiable function defined over some closed interval $[a, b]$ and let Σ be the surface of revolution generated by rotating its graph about the z-axis. The principal curvatures of Σ in the radial and angular directions are respectively

$$c_r = \frac{-u_{rr}}{\sqrt{1+u_r^2}^3}, \quad c_\theta = \frac{-u_r}{r\sqrt{1+u_r^2}}, \tag{2-1}$$

where r here denotes the radial distance in $A(a, b)$ from the origin, and the subscript r denotes differentiation with respect to this variable. The vertical component of the upward-pointing, unit normal vector over Σ is

$$\langle N_\Sigma, e_z \rangle = \frac{1}{\sqrt{1+u_r^2}}, \tag{2-2}$$

so that, by (1-1), Σ is a rotationally symmetric Grim surface whenever

$$r u_{rr} + (u_r - r)(1 + u_r^2) = 0. \tag{2-3}$$

Solutions of this equation have no straightforward closed form. However, it will be sufficient for our purposes to obtain approximations by semiconvergent, that is, asymptotic, series. We first derive an asymptotic formula which is valid as r tends to infinity.

Theorem 2.1.1. *If $u :]a, \infty[\to \mathbb{R}$ is a solution of (2-3) then, as $r \to +\infty$,*

$$u = \tfrac{1}{2}r^2 - \log(r) + a + \mathrm{O}(r^{-(k+2)}) \tag{2-4}$$

for some real constant a.[2]

Theorem 2.1.1 follows immediately upon integrating (2-13), below. A similar result has already been obtained in [Clutterbuck et al. 2007]. However, we consider it worth deriving (2-4) in full, not only because we use different techniques, but also because we believe it serves as good preparation for the more subtle small-scale asymptotic estimates that will be studied in the following sections.

Define the nonlinear operator $\mathcal{G}$ by

$$\mathcal{G}v := rv_r + (v-r)(1+v^2), \tag{2-5}$$

and observe that v solves $\mathcal{G}v = 0$ if and only if its integral is the profile of a rotationally symmetric Grim surface. We first derive formal solutions to (2-5). To this end, we define a *Laurent series* in the formal variable R to be a formal power series of the form

$$V := \sum_{m=-\infty}^{k} V_m R^m, \tag{2-6}$$

where, for all m, V_m is a real number and k is some finite integer, which we henceforth call the *order* of V. Since the operations of formal multiplication and formal differentiation are well-defined over the space of Laurent series, the operator $\mathcal{G}$ also has a well-defined action over this space.

Lemma 2.1.2. *There exists a unique Laurent series V such that $\mathcal{G}V = 0$. Furthermore:*

(1) *V has order* 1.

(2) $V_1 = 1,\ V_{-1} = -1.$

(3) *If m is even, then $V_m = 0$.*

(4) *If $\widehat{V}_n := \sum_{m=1-2n}^{1} V_m R^m$ denotes the n-th partial sum of V, then $\mathcal{G}\widehat{V}_n$ is a finite Laurent series of order* $(1-2n)$.

Proof. Consider the ansatz (2-6). If $k \leq -1$, then the highest-order term in $\mathcal{G}V$ is $(-R)$, if $k = 0$, then it is $(-R)(1+V_0^2)$, and if $k \geq 2$, then it is $V_k^3 R^{3k}$. Since none of these vanish, it follows that V must be of order 1. In this case, the highest-order term in $\mathcal{G}V$ is $V_1^2(V_1 - 1)R^3$ so that, in order to have nontrivial solutions, we require $V_1 = 1$. We now have

$$R\frac{dV}{dR} + (V-R)(1+V^2) = R + \sum_{m=-\infty}^{0}(m+1)V_m R^m + \sum_{m=-\infty}^{2}\Bigg(\sum_{\substack{p+q+r=m\\ p\leq 0,\ q,r\leq 1}} V_p V_q V_r\Bigg)R^m.$$

In particular, setting the respective coefficients of R^2 and R equal to 0 yields

$$V_0 = 0, \quad V_{-1} = -1.$$

[2]We refer the reader to Section 1.3 and Appendix A for a detailed review of the notation used here and throughout the sequel.

For all $m \leq -2$, setting the coefficient of R^{m+2} equal to 0 now yields

$$V_m + \Bigg(\sum_{\substack{p+q+r=m+2\\ m+1\leq p\leq -1\\ m+2\leq q,r\leq 1}} V_p V_q V_r\Bigg) + (m+3)V_{m+2} = 0. \tag{2-7}$$

The existence and uniqueness of V now follow from this recurrence relation. Furthermore, if $p+q+r = m+2$, and if m is even, then at least one of p, q and r is also even, and since $V_0 = 0$, it follows by induction that $V_m = 0$ for all even m. In addition, by (2-7), for all n, and for all $m \geq (3-2n)$, the coefficient of R^m in $\mathcal{G}\widehat{V}_n$ is equal to 0. However, since $V_{-2n} = 0$, by (2-7) again, the coefficient of R^{2-2n} in $\mathcal{G}\widehat{V}_n$ is also equal to 0, so that $\mathcal{G}\widehat{V}_n$ is a finite Laurent series of order $(1-2n)$, as desired. □

For all n, define the n-th partial sum $v_n :]0,\infty[\to \mathbb{R}$ by

$$v_n(r) := \sum_{m=1-2n}^{1} V_m r^m. \tag{2-8}$$

We now show that the sequence (v_n) yields successively better approximations over the large scale of the exact solutions of $\mathcal{G}v = 0$. We first derive zeroth order bounds.

Lemma 2.1.3. *If $v : [a,\infty[\to \mathbb{R}$ solves (2-5) then, for large r,*

$$|v_0 - v| \lesssim \frac{1}{r}. \tag{2-9}$$

Proof. Consider the family of polynomials $p_t(y) = (y-1)(t^2+y^2)$. For all $t > 0$, $y = 1$ is the unique real root of p_t. Since $y = 0$ is the unique local maximum of p_0, for sufficiently small t, the unique local maximum of p_t is also near 0, and the value of p_t at this point is less than $-t^2/2$. Since p_0 is convex over the interval $\left[\frac{1}{3},\infty\right[$, for $\frac{1}{3} < y < 1$ we have $p_0(y) \leqslant \frac{3}{2}(1-y)p_0\left(\frac{1}{3}\right) = \frac{1}{9}(y-1)$ and so, for sufficiently small t, over the smaller interval $\left[\frac{1}{2},1\right]$, $p_t(y) \leqslant \frac{1}{18}(y-1)$.

Now let v be a solution of $\mathcal{G}v = 0$. In particular, using a dot to denote differentiation with respect to r, we have $\dot{v} = -r^2 p_{1/r}(v/r)$. Suppose, furthermore, that $r \gg 1$ so that the estimates of the preceding paragraph hold for $p_{1/r}$. When $v \geq r$, we have $\dot{v} - \dot{r} = \dot{v} - 1 \leq -1$, so that, for sufficiently large r, $v(r) \leqslant r$. If $v \leqslant \frac{1}{2}r$, then $\dot{v} - \frac{1}{2}\dot{r} \geqslant \frac{1}{2}r - \frac{1}{2}$, so that, for sufficiently large r, $v(r) \geqslant \frac{1}{2}r$. Finally, if $\frac{1}{2}r \leqslant v \leqslant r$ then, by the preceding discussion, $\dot{v} \geqslant \frac{1}{18}r(r-v)$. It follows that if $w := r(v_0 - v) = r(r-v)$, then $w > 0$ and $\dot{w} = 2r - v - r\dot{v} \leqslant r + w/r - \frac{1}{18}rw$. Since this is negative for $w \geqslant 36$ and $r > 6$, the function w is bounded, and the result follows. □

Lemma 2.1.4. *If $v : [a,\infty[\to \mathbb{R}$ solves (2-5) then, for all n, and for large r,*

$$|v_n - v| \lesssim r^{-(2n+1)}. \tag{2-10}$$

Proof. For all n, let $w_n := r^{2n-1}(v_n - v)$ be the rescaled error. We prove by induction that $|w_n| \lesssim r^{-2}$ for all n. Indeed, the case $n = 0$ follows from (2-9). We suppose therefore that $n \geq 1$. Since $w_n = r^2 w_{n-1} + V_{1-2n}$, it follows by the inductive hypothesis that w_n is bounded. Now let $P(a,b)$ denote any polynomial in the variables a and b. Since $\mathcal{G}v_n$ is a finite Laurent polynomial of order $(1-2n)$, using a dot to denote differentiation with respect to r, we have

$$\begin{aligned}\dot{w}_n &= \frac{(2n-1)}{r}w_n + r^{2n-2}(r\dot{v}_n - r\dot{v})\\ &= \frac{1}{r}P\Big(\frac{1}{r}, w_n\Big) - r^{2n-2}\big((v_n - r)(1+v_n^2) - (v-r)(1+v^2)\big)\\ &= \frac{1}{r}P\Big(\frac{1}{r}, w_n\Big) - \frac{1}{r}w_n(1 - r(v_n+v) + (v_n^2 + v_n v + v^2)).\end{aligned}$$

Since $v = v_n - r^{-(2n-1)}w_n$ and since $(v_n - r)$ is also a polynomial in r^{-1} with no constant term, this yields

$$\dot{w}_n = \frac{1}{r}P\Big(\frac{1}{r}, w_n\Big) - rw_n.$$

Since w_n is bounded, there therefore exists a constant $B > 0$ such that, for all $r \geqslant 1$,

$$|\dot{w}_n + rw_n| \leqslant Br^{-1}. \tag{2-11}$$

In particular, for $r \geq 2$ and $r^2 w_n \geq 2B$,

$$\frac{d}{dr}r^2 w_n = r^2(\dot{w}_n + rw_n) + (2r - r^3)w_n \leq Br - \tfrac{1}{2}r^3 w_n \leq 0,$$

so that $r^2 w_n$ is bounded from above. Since $(-w_n)$ also satisfies (2-11), we see that $r^2 w_n$ is bounded from below, and this completes the proof. □

Lemma 2.1.5. *If* $v : [a, \infty[\to \mathbb{R}$ *solves* (2-5) *then, for all* n,

$$v_n - v = \mathrm{O}(r^{-(k+2n+1)}). \tag{2-12}$$

In particular,

$$v = r - \frac{1}{r} + \mathrm{O}(r^{-(k+3)}). \tag{2-13}$$

Proof. For all n, define $w_n := (v_n - v)$. As in the proof of Lemma 2.1.4, we obtain

$$\dot{w}_n = P_1\Big(\frac{1}{r}, w_n\Big)rw_n + \frac{1}{r}\mathcal{G}v_n,$$

where P_1 is some polynomial. Since $\mathcal{G}v_n$ is a finite Laurent polynomial of order $(1-2n)$, it follows by induction that, for all k,

$$\frac{d^k w_n}{dr^k} = P_k\Big(\frac{1}{r}, w_n\Big)r^k w_n + Q_k\Big(\frac{1}{r}, w_n\Big)r^{k-(2n+1)},$$

where P_k and Q_k are polynomials. It follows by (2-10) that, for all k,

$$\left|\frac{d^k}{dr^k}(v_n - v)\right| = \left|\frac{d^k w_n}{dr^k}\right| \lesssim r^{k-(2n+1)}.$$

However, since $(v_{n+k} - v_n)$ is a finite Laurent series of order $-(2n+1)$, for all k,

$$\left|\frac{d^k}{dr^k}(v_n - v)\right| \leq \left|\frac{d^k}{dr^k}(v_n - v_{n+k})\right| + \left|\frac{d^k}{dr^k}(v_{n+k} - v)\right| \lesssim r^{-(k+2n+1)},$$

and the result follows. □

2.2. ***The small scale: formal solutions.*** We now study solutions to (2-5) over the small scale. We fix positive constants $K \gg 1$ and $\eta \ll 1$ which we henceforth consider to be universal. Let Λ be a large, positive real number, and let ϵ, $R > 0$ and $c \in \mathbb{R}$ be such that

$$\left(\epsilon R^{4+\eta} + \frac{1}{R^{1-\eta}}\right) \leq \frac{1}{\Lambda}, \quad \epsilon R^{5-\eta} \geq \Lambda, \quad |c| \leq K. \tag{2-14}$$

These conditions will be used repeatedly throughout the paper. Observe, in particular, that (2-14) implies that ϵ becomes small and R becomes large as Λ tends to infinity. We will prove:

Theorem 2.2.1. *For all sufficiently large Λ, and for all R, ϵ satisfying (2-14), there exists a smooth function $\sigma[\epsilon, R] : \mathbb{R} \to \mathbb{R}$ such that, for all $c \in [-K, K]$, if $v : [\epsilon R, \epsilon R^4] \to \mathbb{R}$ solves $\mathcal{G}v = 0$ with initial value*

$$v(\epsilon R) = \frac{1}{R}\sigma[\epsilon, R](c) + \frac{\epsilon R}{2}, \tag{2-15}$$

then

$$v(r) = \frac{1}{2}r + \frac{c\epsilon}{r} + \mathrm{O}\left(\left[1 + \log\left(\frac{r}{\epsilon R}\right)\right]\frac{1}{r^k}\left(r + \frac{\epsilon}{r}\right)^3\right). \tag{2-16}$$

Furthermore, the function $\sigma[\epsilon, R]$ converges to the identity in the C^∞_{loc} sense as Λ tends to $+\infty$.

Remark. We leave the reader to verify that the same conclusion also holds over the interval $[\epsilon R, C\epsilon R^4]$ for any constant C not depending on ϵ or R.

The function $\sigma[\epsilon, R]$ will be defined explicitly in Section 2.3, below, and Theorem 2.2.1 will follow immediately from Lemma 2.4.2, below. The constant c will henceforth be referred to as the *logarithmic parameter* of the function v. Observe that, up to a small perturbation, it is related to the initial value of v by a linear function. This perturbation is required in order to guarantee good estimates over the whole interval. Indeed, replacing $\sigma[\epsilon, R](c)$ by c in (2-15) would increase the error in (2-16), making it then of order (ϵ/r).

In order to appreciate Theorem 2.2.1 and the argument that follows, we find it helpful to first recall the geometric properties of the function v over the interval $[\epsilon R, \epsilon R^4]$. Indeed, by definition, its integral u is the profile of some rotationally symmetric Grim surface. However, it is known (see [Clutterbuck et al. 2007]) that, near the lower end of this interval, the first term in the MCFS equation (1-1) dominates, so that the graph of u is close to some minimal catenoid in $\mathbb{R}^3$ and the function u is itself approximately logarithmic. On the other hand, near the upper end of this interval, it is the second term in the MCFS equation which dominates, and the function u is approximately quadratic, in accordance with the asymptotic formula obtained in the preceding section. These two contrasting behaviours are reflected in (2-16) by the ϵ/r terms and the r terms respectively.

In order to derive an asymptotic formula for u that simultaneously describes these two behaviours, we introduce two abstract variables M and N, where M measures its quadratic behaviour, and N measures its logarithmic behaviour. By expressing the equation $\mathcal{G}v = 0$ in terms of these new variables, the asymptotic formula for v is then obtained in the same manner as in Section 2.1 namely, by first determining formal solutions which then serve as approximations for exact solutions.

Upon applying the change of variables $r := \epsilon R e^x$ we obtain

$$\mathcal{G}v = \mathcal{D}v - \epsilon R e^x + (v - \epsilon R e^x)v^2, \tag{2-17}$$

where the operator $\mathcal{D}$ is defined by

$$\mathcal{D}v := v_x + v, \tag{2-18}$$

and the subscript x here denotes differentiation with respect to this variable. Now let $\mathbb{R}[X, M, N]$ be the ring of polynomials with real coefficients in the variables X, M and N. We consider a general element V of $\mathbb{R}[X, M, N]$ as a sum of the form

$$V = \sum_{p,q\leq k} V_{p,q}(X)M^pN^q, \tag{2-19}$$

where, for all p, q, $V_{p,q}$ is a polynomial in the variable X and k is some finite, nonnegative integer which we henceforth refer to as the *order* of V. There is a natural correspondence sending $\mathbb{R}[X, M, N]$ into the space of continuous functions over $[0, 3\log(R)]$ given by

$$V \mapsto v(x) := \sum_{p,q\leq k} V_{p,q}(x)(\epsilon Re^x)^p\Big(\frac{c}{R}e^{-x}\Big)^q. \tag{2-20}$$

In other words, this correspondence is the unique $\mathbb{R}[X]$-ring homomorphism which sends M to ϵRe^x and N to $(c/R)e^{-x}$. Although this homomorphism is not injective, it keeps track of the parameters ϵ, R and c, which is the reason why it serves our purposes. Operators $\mathcal{G}$ and $\mathcal{D}$ are also defined over $\mathbb{R}[X, M, N]$ by

$$\begin{aligned} \mathcal{G}V &:= \mathcal{D}V - M + (V - M)V^2, \\ (\mathcal{D}V)_{p,q} &:= \Big(\frac{d}{dX} + 1 + (p-q)\Big)V_{p,q}, \end{aligned} \tag{2-21}$$

where $d/(dX)$ here denotes the operator of formal differentiation with respect to the variable X. In particular, $\mathcal{G}$ and $\mathcal{D}$ both map through the above correspondence to the operators given in (2-17) and (2-18) respectively, thereby justifying this notation. Observe, furthermore, that $\mathcal{D}$ defines a surjective linear map from $\mathbb{R}[X, M, N]$ to itself and that its kernel consists of finite sums of the form

$$V = \sum_{p\leq k} a_pM^pN^{p+1},$$

where $a_0, \dots, a_k$ are real constants.

Let $\mathbb{R}[X][\![M, N]\!]$ be the ring of formal power series over the variables M and N with coefficients that are polynomials in the variable X. Observe that the operators $\mathcal{G}$ and $\mathcal{D}$ naturally extend again to well-defined operators over this space.

Lemma 2.2.2. *There exists a unique formal power series V in $\mathbb{R}[X][\![M, N]\!]$ such that*

(1) $V_{0,1} = 1$,

(2) $V_{p,p+1}(0) = 0$ *for all* $p \geq 1$, *and*

(3) $\mathcal{G}V = 0$.

Furthermore,

(4) $V_{1,0} = \frac{1}{2}$,

(5) *if* $p+q$ *is even, then* $V_{p,q} = 0$, *and*

(6) *if* $p+q = 2k+1$ *is odd, then* $V_{p,q}$ *has order at most* k *in* X.

Finally, if we define

$$\widehat{V}_k := \sum_{p+q\leq 2k+1} V_{p,q}(X)M^pN^q,$$

then,

(7) *if* $(p+q)\leq(2k+1)$, *then the coefficient of* M^pN^q *in* $\mathcal{G}\widehat{V}_k$ *vanishes,*

(8) *if* $(p+q)>(2k+1)$ *is even, then the coefficient of* M^pN^q *in* $\mathcal{G}\widehat{V}_k$ *vanishes, and*

(9) *if* $(p+q)>(2k+1)$ *is odd, then the coefficient of* M^pN^q *in* $\mathcal{G}\widehat{V}_k$ *has order at most* $\frac{1}{2}(p+q-3)$ *in* X.

Proof. Let $V=\sum_{p,q}V_{p,q}(X)M^pN^q$ be an element of $\mathbb{R}[X][\![M,N]\!]$ which solves $\mathcal{G}V=0$. For all (p,q), equating the coefficient of M^pN^q in $\mathcal{G}V$ to 0, we obtain

$$\left(\frac{d}{dX}+(1+(p-q))\right)V_{p,q}=\delta_{p1}\delta_{q0}-\sum_{\substack{p_1+p_2+p_3=p\\ q_1+q_2+q_3=q}}V_{p_1,q_1}V_{p_2,q_2}V_{p_3,q_3}+\sum_{\substack{p_1+p_2=p-1\\ q_1+q_2=q}}V_{p_1,q_1}V_{p_2,q_2}. \tag{2-22}$$

In particular,

$$\frac{dV_{0,0}}{dX}+V_{0,0}(1+V_{0,0}^2)=0,$$

and since there exists no nontrivial polynomial solution to this equation, it follows that $V_{0,0}=0$. From this it follows that the two summations on the right-hand side of (2-22) only involve terms of order at most $p+q-2$ in (M,N). In particular, $V_{0,1}$ satisfies

$$\frac{dV_{0,1}}{dX}=0.$$

It is thus constant, and we henceforth set it equal to 1. It now follows by induction that there exists a unique sequence of polynomials $(V_{p,q})$ satisfying (2-22) such that $V_{0,1}=1$ and $V_{p,p+1}(0)=0$ for all $p\geq 1$.

To prove (4), observe that $V_{1,0}$ satisfies $dV_{1,0}/(dX)+2V_{1,0}=1$ so that, since it is a polynomial, $V_{1,0}=\frac{1}{2}$, as desired. To prove (5), observe that if $p+q$ is even, then every summand on the right-hand side in (2-22) involves at least one term of the form $V_{p',q'}$, where $p'+q'$ is an even number no greater than $p+q-2$. Since $V_{0,0}=0$, it follows by induction that $V_{p,q}=0$ whenever $p+q$ is even, as desired. To prove (6), suppose that for all $l<k$, and for $p+q=2l+1$, the polynomial $V_{p,q}$ has order at most l in X. By (2-22), for all $p+q=2k+1$, the polynomial $V_{p,q}$ is obtained by integrating terms of order at most $(k-1)$ in X, and it follows by induction that $V_{p,q}$ has order at most k in X, as desired.

Finally, observe that, by (2-22), the term $V_{p,q}$ is defined by setting the coefficient of M^pN^q equal to 0 in $\mathcal{G}V$, and (7) follows. Furthermore, for $p+q>(2k+1)$, the coefficient of M^pN^q in $\mathcal{G}V$ is equal to the right-hand side of (2-22). Items (8) and (9) now follow by similar arguments used to prove (5) and (6), above, and this completes the proof. ☐

2.3. ***The small scale: exact solutions, I.*** Let V be the formal power series constructed in Lemma 2.2.2. For ϵ, R satisfying (2-14), for $c\in\mathbb{R}$, and for nonnegative, integer k, let $v_{k,c}$ be the k-th partial sum of V with logarithmic parameter c, that is,

$$v_{k,c}(x):=\sum_{p+q\leq 2k+1}V_{p,q}(x)(\epsilon Re^x)^p\left(\frac{c}{R}e^{-x}\right)^q. \tag{2-23}$$

Define the function $\sigma[\epsilon, R, k] : \mathbb{R} \to \mathbb{R}$ by

$$\sigma[\epsilon, R, k](c) := R v_{k,c}(0) - \frac{\epsilon R^2}{2}. \tag{2-24}$$

Trivially, if $v : [0, 3\log(R)] \to \mathbb{R}$ satisfies

$$v(0) = \frac{1}{R}\sigma[\epsilon, R, k](c) + \frac{\epsilon R}{2},$$

then v has the same initial value as $v_{k,c}$. Observe that $\sigma[\epsilon, R, k]$ is a polynomial in c with coefficients that depend on ϵ, R and k and, for all k, $\sigma[\epsilon, R, k]$ converges to the identity in the C^∞_{loc} sense as Λ tends to infinity. We will see presently that the estimates we require follow when k is at least 9, and we therefore define

$$\sigma[\epsilon, R](c) := \sigma[\epsilon, R, 9](c). \tag{2-25}$$

This is the function that appears in the statement of Theorem 2.2.1.

As in Section 2.1, we now determine zeroth order bounds for the difference between $v_{k,c}$ and an exact solution with the same initial value. We achieve this via the contraction mapping theorem. We first introduce the required analytic framework. For $T \in [0, 3\log(R)]$, let $C^0([0, T])$ be the Banach space of continuous functions over the interval $[0, T]$ furnished with the uniform norm and let $C^1_0([0, T])$ be the Banach space of continuously differentiable functions over this interval with initial value 0, furnished with the norm

$$\|w\|_{C^1_0} := \|w_x\|_{C^0}, \tag{2-26}$$

where the subscript x here denotes differentiation with respect to this variable. Observe that, for all $w \in C^1_0([0, T])$,

$$\|w\|_{C^0} \le T\|w\|_{C^1_0}. \tag{2-27}$$

Lemma 2.3.1. *The operator $\mathcal{D}$ defines a linear isomorphism from $C^1_0([0, T])$ into $C^0([0, T])$. Furthermore, the operator norms of $\mathcal{D}$ and its inverse satisfy*

$$\|\mathcal{D}\| \le 1 + T, \quad \|\mathcal{D}^{-1}\| \le 2. \tag{2-28}$$

Proof. First, bearing in mind (2-27),

$$\|\mathcal{D}w\|_{C^0} \le \|w_x\|_{C^0} + \|w\|_{C^0} \le (1+T)\|w\|_{C^1_0},$$

so that $\|\mathcal{D}\| \le 1 + T$. By inspection, for all w,

$$(\mathcal{D}^{-1}w)(x) = e^{-x}\int_0^x e^y w(y)\,dy.$$

In particular,

$$\|\mathcal{D}^{-1}w\|_{C^0} \le \|w\|_{C^0}.$$

Thus,

$$\|\mathcal{D}^{-1}w\|_{C^1_0} = \|(\mathcal{D}^{-1}w)_x\|_{C^0} \le \|\mathcal{D}\mathcal{D}^{-1}w\|_{C^0} + \|\mathcal{D}^{-1}w\|_{C^0} \le 2\|w\|_{C^0},$$

so that $\|\mathcal{D}^{-1}\| \le 2$. □

Consider now the functional $\mathcal{H} : C^1_0([0,T]) \to C^0([0,T])$ given by

$$\mathcal{H}(w) := \mathcal{G}(v_k + w). \tag{2-29}$$

Its Fréchet derivative at w is

$$D\mathcal{H}(w)f := \mathcal{D}f + \mathcal{E}(w)f, \tag{2-30}$$

where

$$\mathcal{E}(w)f := 3(v_{k,c} + w)^2 f - 2\epsilon Re^x (v_{k,c} + w)f. \tag{2-31}$$

Lemma 2.3.2. *For all $w \in C^1_0([0,T])$, the operator norm of $\mathcal{E}(w)$, considered as a linear map from $C^1_0([0,T])$ into $C^0([0,T])$, satisfies*

$$\|\mathcal{E}(w)\| \lesssim T\biggl((\epsilon Re^T)^2 + \frac{1}{R^2} + T^2\|w\|^2_{C^1_0}\biggr). \tag{2-32}$$

Proof. Indeed, over $[0,T]$,

$$\|\epsilon Re^x\|_{C^0} \le \epsilon Re^T, \qquad \biggl\|\frac{c}{R}e^{-x}\biggr\|_{C^0} \le \frac{c}{R}.$$

Thus, by Lemma 2.2.2 and (2-14),

$$\|v_k\|_{C^0} \lesssim \sum_{i=0}^{k}(1+T^i)\Bigl(\epsilon Re^T + \frac{1}{R}\Bigr)^{2i+1} \lesssim \epsilon Re^T + \frac{1}{R},$$

so that, by (2-27) and (2-31),

$$\begin{aligned}\|\mathcal{E}(w)f\|_{C^0} &\lesssim \Bigl((\epsilon Re^T)^2 + \frac{1}{R^2} + \|w\|^2_{C^0}\Bigr)\|f\|_{C^0}\\ &\lesssim T\Bigl((\epsilon Re^T)^2 + \frac{1}{R^2} + T^2\|w\|^2_{C^1_0}\Bigr)\|f\|_{C^1_0},\end{aligned}$$

as desired. □

Define the map $\Phi : C^1_0([0,T]) \to C^1_0([0,T])$ by

$$\Phi(w) := w - \mathcal{D}^{-1}\mathcal{H}(w). \tag{2-33}$$

Lemma 2.3.3. *For $w, \bar{w} \in C^1_0([0,T])$,*

$$\|\Phi(w) - \Phi(\bar{w})\|_{C^1_0} \lesssim T\Bigl((\epsilon Re^T)^2 + \frac{1}{R^2} + T^2\|w\|^2_{C^1_0} + T^2\|\bar{w}\|^2_{C^1_0}\Bigr)\|w - \bar{w}\|_{C^1_0}. \tag{2-34}$$

Proof. Indeed, for $w, \bar{w} \in C^1_0([0,T])$, using (2-30),

$$\begin{aligned}\Phi(w) - \Phi(\bar{w}) &= w - \bar{w} - \mathcal{D}^{-1}(\mathcal{H}(w) - \mathcal{H}(\bar{w}))\\ &= -\mathcal{D}^{-1}(\mathcal{H}(w) - \mathcal{H}(\bar{w}) - \mathcal{D}(w - \bar{w}))\\ &= -\mathcal{D}^{-1}\biggl(\int_0^1 \mathcal{E}(tw + (1-t)\bar{w})\,dt\biggr)(w - \bar{w}).\end{aligned}$$

Thus, by (2-28) and (2-32),

$$\|\Phi(w) - \Phi(\bar{w})\|_{C^1_0} \lesssim T\Bigl((\epsilon Re^T)^2 + \frac{1}{R^2} + T^2\|w\|^2_{C^1_0} + T^2\|\bar{w}\|^2_{C^1_0}\Bigr)\|w - \bar{w}\|_{C^1_0},$$

as desired. □

Applying the contraction mapping theorem now yields:

Lemma 2.3.4. *For sufficiently large Λ, if $v_{k,c}$ is the k-th partial sum of V with logarithmic parameter c, and if $v : [0, 3\log(R)] \to \mathbb{R}$ solves $\mathcal{G}v = 0$ with initial value*

$$v(0) = \frac{1}{R}\sigma[\epsilon, R, k](c) + \frac{\epsilon R}{2}, \tag{2-35}$$

then

$$\|v - v_{k,c}\|_{C^0} \lesssim (1 + T^{k+1})\Big(\epsilon Re^T + \frac{1}{R}\Big)^{2k+3}. \tag{2-36}$$

Proof. By Lemma 2.2.2,

$$\|\mathcal{G}v_k\|_{C^0} \lesssim (1 + T^k)\Big(\epsilon Re^T + \frac{1}{R}\Big)^{2k+3}.$$

By (2-28), there therefore exists $B > 0$, which we may consider to be universal, such that

$$\|\Phi(0)\|_{C^1_0} = \|\mathcal{D}^{-1}\mathcal{G}v_k\|_{C^1_0} \le B(1 + T^k)\Big(\epsilon Re^T + \frac{1}{R}\Big)^{2k+3}.$$

Let X be the closed ball of radius $2B(1 + T^k)(\epsilon Re^T + R^{-1})^{2k+3}$ about 0 in $C^1_0([0, T])$. By (2-14), if $w, \bar{w} \in X$ then, in particular,

$$T\|w\|_{C^1_0}, T\|\bar{w}\|_{C^1_0} \lesssim \Big(\epsilon Re^T + \frac{1}{R}\Big),$$

so that, by (2-34) and (2-14) again,

$$\|\Phi(w) - \Phi(\bar{w})\|_{C^1_0} \lesssim \frac{1}{\Lambda}\|w - \bar{w}\|_{C^1_0}.$$

The map Φ thus defines a contraction from X to itself, and there therefore exists $w \in X$ such that $\Phi(w) = w$. In particular $\mathcal{H}w = 0$, and

$$\|w\|_{C^0} \le T\|w\|_{C^1_0} \lesssim (1 + T^{k+1})\Big(\epsilon Re^T + \frac{1}{R}\Big)^{2k+3}.$$

Finally, by the definition of the function $\sigma[\epsilon, R, k]$, we have $v(0) = v_{k,c}(0)$ so that, by uniqueness of solutions to ODEs with prescribed initial values, $v - v_{k,c} = w$, and the result follows. □

2.4. *The small scale: exact solutions, II.* The final step in proving Theorem 2.2.1 involves extending the estimates obtained in Lemma 2.3.4 to derivatives of all orders.

Lemma 2.4.1. *If $v_{k,c}$ and v are as in Lemma 2.3.4, then*

$$v = v_{k,c} + \mathrm{O}\Big((1 + T^{k+1})\Big(\epsilon Re^T + \frac{1}{R}\Big)^{2k+3}\Big). \tag{2-37}$$

Proof. Define $w := v - v_{k,c}$. Since $v_{k,c}$ is a polynomial in ϵRe^x and $(c/R)e^{-x}$ with coefficients in $\mathbb{R}[X]$, as in the proof of Lemma 2.1.4,

$$w_x = P_1\Big(w, \epsilon Re^x, \frac{c}{R}e^{-x}\Big)w + \mathcal{G}v_{k,c}$$

for some polynomial P_1 with coefficients in $\mathbb{R}[X]$. Since $\mathcal{G}v_{k,c}$ is also a polynomial in ϵRe^x and $(c/R)e^{-x}$ with coefficients in $\mathbb{R}[X]$, it follows by induction that, for all l,

$$\frac{d^l}{dx^l}w = P_l\Big(w, \epsilon Re^x, \frac{c}{R}e^{-x}\Big)w + \sum_{p=0}^{l-1} Q_{p,l}\Big(w, \epsilon Re^x, \frac{c}{R}\Big)\frac{d^p}{dx^p}\mathcal{G}v_{k,c} \tag{2-38}$$

for suitable polynomials P_l and $(Q_{p,l})_{0\leq p\leq l-1}$ also with coefficients in $\mathbb{R}[X]$. However, by (2-36),

$$\|w\|_{C^0} \lesssim (1+T^{k+1})\Big(\epsilon Re^T + \frac{1}{R}\Big)^{2k+3}.$$

Thus, by (2-14),

$$\Big\|P_l\Big(w, \epsilon Re^x, \frac{c}{R}e^{-x}\Big)\Big\|_{C^0}, \Big\|Q_{p,l}\Big(\epsilon Re^x, \frac{c}{R}e^{-x}\Big)\Big\|_{C^0} \lesssim 1.$$

Finally, Lemma 2.2.2 and (2-14) again,

$$\Big\|\frac{d^{l-1}}{dx^{l-1}}\mathcal{G}v_k\Big\|_{C^0} \lesssim (1+T^k)\Big(\epsilon Re^T + \frac{1}{R}\Big)^{2k+3},$$

and the result follows upon combining these relations. □

Lemma 2.4.2. *If $v_{k,c}$ is the k-th partial sum of V with logarithmic parameter c, and if $v : [0, 3\log(R)] \to \mathbb{R}$ solves $\mathcal{G}v = 0$ with initial value*

$$v(0) = \frac{1}{R}\sigma[\epsilon, R, 4k+9](c) + \frac{\epsilon R}{2}, \tag{2-39}$$

then, for sufficiently large Λ,

$$v = v_{k,c} + \mathrm{O}\Big((1+x^{k+1})\Big(\epsilon Re^x + \frac{1}{R}e^{-x}\Big)^{2k+3}\Big). \tag{2-40}$$

Remark. Since $r = \epsilon Re^x$, by the chain rule,

$$\frac{d}{dr} = \frac{1}{r}\frac{d}{dx},$$

so that Theorem 2.2.1 follows immediately from (2-40) upon setting $k = 0$.

Proof. For nonnegative, integer l, if $v : [0, 3\log(R)] \to \mathbb{R}$ solves $\mathcal{G}v = 0$ with initial value as in (2-35) then, since (2-37) holds for all $T \in [0, 3\log(R)]$,

$$v = v_{l,c} + \mathrm{O}\Big((1+x^{l+1})\Big(\epsilon Re^x + \frac{1}{R}\Big)^{2l+3}\Big).$$

In particular, if $v : [0, 3\log(R)] \to \mathbb{R}$ now solves $\mathcal{G}v = 0$ with initial value given by (2-39), then, bearing in mind (2-14),

$$v = v_{4k+9,c} + \mathrm{O}\Big((1+x^{k+1})\Big(\epsilon Re^x + \frac{1}{R}e^{-x}\Big)^{2k+3}\Big).$$

However, by Lemma 2.2.2 and (2-14) again,

$$v_{4k+9,c} = v_{k,c} + \mathrm{O}\Big((1+x^{k+1})\Big(\epsilon Re^x + \frac{1}{R}e^{-x}\Big)^{2k+3}\Big),$$

and the result follows. □

2.5. *The small scale: solutions of the linearised equation.* We conclude this section by studying how solutions of the equation $\mathcal{G}v = 0$ vary with the logarithmic parameter c.

Theorem 2.5.1. *For sufficiently large* Λ *and for all* R, ϵ *satisfying* (2-14), *if, for all* $c \in [-K, K]$, *the function* $v_c : [\epsilon R, \epsilon R^4] \to \mathbb{R}$ *solves* $\mathcal{G}v_c = 0$ *with initial value*

$$v_c(\epsilon R) = \frac{1}{R}\sigma[\epsilon, R](c) + \frac{\epsilon R}{2},$$

then

$$\frac{dv_c}{dc}(r) = \frac{\epsilon}{r} + \mathrm{O}\Big(\Big[1 + \log\Big(\frac{r}{\epsilon R}\Big)\Big]\frac{1}{r^k}\Big(r + \frac{\epsilon}{r}\Big)^3\Big). \tag{2-41}$$

Theorem 2.5.1 follows from (2-49), below, via reasoning similar to that used in Section 2.4. It suffices to study solutions of the linearisation of $\mathcal{G}$ about v, the asymptotic properties of which are readily derived from the analysis of the previous sections. Indeed, let $\mathbb{R}[X][\![M, N]\!]$ be as in Section 2.2 and define the operator ∂_N over this space by

$$(\partial_N V)_{p,q} := (q+1)V_{p,q+1}. \tag{2-42}$$

In other words, ∂_N is simply the operator of formal differentiation with respect to N. By explicit calculation, $N\partial_N$ commutes with $\mathcal{D}$. Now let V be the formal power series constructed in Lemma 2.2.2 and define

$$W := N\partial_N V. \tag{2-43}$$

Applying $N\partial_N$ to the relation $\mathcal{G}V = 0$ yields

$$\mathcal{D}W + 3V^2 W - 2MVW = 0, \tag{2-44}$$

so that W is a formal solution to the linearisation of $\mathcal{G}$ about the formal series V.

Fix a nonnegative integer k, let $\widehat{V}_k$ be as in Lemma 2.2.2 and define

$$\widehat{W}_k := \sum_{p+q \le 2k+1} W_{p,q}(X)M^p N^q. \tag{2-45}$$

By (2-44),

$$\mathcal{D}\widehat{W}_k + 3\widehat{V}_k^2\widehat{W}_k - 2M\widehat{V}_k\widehat{W}_k = \mathrm{O}((M+N)^{2k+3}). \tag{2-46}$$

Consider now $\Lambda, K > 0$, let $\epsilon, R > 0$ and $c \in \mathbb{R}$ satisfy (2-14), and let $v_{k,c}$ and $w_{k,c}$ be the functions corresponding to $\widehat{V}_k$ and $\widehat{W}_k$ respectively. By (2-46), for all k,

$$\mathcal{D}w_{k,c} + 3v_{k,c}^2 w_{k,c} - 2(\epsilon R e^x)v_{k,c}w_{k,c} = \mathrm{O}\Big(x^{k+1}\Big(\epsilon R e^x + \frac{1}{R}e^{-x}\Big)^{2k+3}\Big). \tag{2-47}$$

Lemma 2.5.2. *For sufficiently large* Λ *and for all* $T \in [0, 3\log(R)]$, *if* $v : [0, T] \to \mathbb{R}$ *solves* $\mathcal{G}v = 0$ *with initial value*

$$v(0) = \frac{1}{R}\sigma[\epsilon, R, k](c) + \frac{\epsilon R}{2},$$

and if $w : [0, T] \to \mathbb{R}$ *solves*

$$\mathcal{D}w + 3v^2 w - 2\epsilon R e^x v w = 0 \tag{2-48}$$

with initial value $w(0) = w_{k,c}(0)$, *then*

$$\|w - w_{k,c}\|_{C^1_0} \lesssim (1+T)^{k+1}\Big(\epsilon Re^T + \frac{1}{R}\Big)^{2k+3}. \tag{2-49}$$

Proof. Indeed, by (2-47),

$$\|\mathcal{D}w_{k,c} + 3v_{k,c}^2 w_{k,c} - 2(\epsilon Re^x)v_{k,c}w_{k,c}\| \lesssim (1+T)^{k+1}\Big(\epsilon Re^T + \frac{1}{R}\Big)^{2k+3}.$$

Observe that

$$\|v\|_{C^0},\ \|v_{k,c}\|_{C^0},\ \|w_{k,c}\|_{C^0} \lesssim 1.$$

Thus, by (2-36),

$$\begin{aligned}\|(3v_{k,c}^2 - 3v^2)w_{k,c}\|_{C^0} &= 3\|(v_{k,c}-v)(v_{k,c}+v)w_{k,c}\|_{C^0}\\ &\lesssim (1+T)^{k+1}\Big(\epsilon Re^T + \frac{1}{R}\Big)^{2k+3}.\end{aligned}$$

Likewise

$$\|(2(\epsilon Re^x)v_{k,c} - 2(\epsilon Re^x)v)w_{k,c}\|_{C^0} \lesssim (1+T)^{k+1}\Big(\epsilon Re^T + \frac{1}{R}\Big)^{2k+3}.$$

Thus

$$\begin{aligned}\|\mathcal{D}(w_{k,c}-w)+3v^2(w_{k,c}-w)-2(\epsilon Re^x)v(w_{k,c}-w)\|_{C^0} &= \|\mathcal{D}w_{k,c}+3v^2w_{k,c}-2(\epsilon Re^x)vw_{k,c}\|_{C^0}\\ &\lesssim (1+T)^{k+1}\Big(\epsilon Re^T+\frac{1}{R}\Big)^{2k+3}.\end{aligned} \tag{2-50}$$

Observe now that, for all $\phi : [0, T] \to \mathbb{R}$,

$$3v^2\phi - 2\epsilon Re^x\phi = \mathcal{E}(v - v_{k,c})\phi,$$

where $\mathcal{E}$ is given by (2-31). In particular, by (2-14), (2-32) and (2-36), the operator norm of $\mathcal{E}(v - v_{k,c})$ considered as a map from $C^1_0([0, T])$ into $C^0([0, T])$ satisfies

$$\|\mathcal{E}(v - v_{k,c})\| \lesssim T\Big((\epsilon Re^T)^2 + \frac{1}{R^2}\Big).$$

Thus, by (2-28), for sufficiently large Λ, the operator $\mathcal{D} + \mathcal{E}(v - v_{k,c})$ defines an invertible map from $C^1_0([0, T])$ into $C^0([0, T])$ and the result now follows by (2-50). □

Theorem 2.5.1 now follows as indicated above. In addition, a further iteration of this process also yields:

Theorem 2.5.3. *With the same hypotheses as in Theorem 2.5.1,*

$$\frac{d^2v_c}{dc^2}(r) = \mathrm{O}\Big(\Big[1 + \log\Big(\frac{r}{\epsilon R}\Big)\Big]\frac{1}{r^k}\Big(r + \frac{\epsilon}{r}\Big)^3\Big). \tag{2-51}$$

3. The Grim paraboloid

3.1. ***The MCFS Jacobi operator.*** The *Grim paraboloid*, which we henceforth denote by G_0, is defined to be the unique rotationally symmetric MCF soliton which is a graph over the whole of $\mathbb{R}^2$. Put differently, using the notation of Section 2, there is a unique solution v to the ODE $\mathcal{G}v = 0$ which is defined over the whole interval $]0, \infty[$. This solution tends to 0 as x tends to 0, and the Grim paraboloid is the surface of revolution generated by rotating the graph of its integral about the z-axis.

Let J be the MCFS Jacobi operator of the Grim paraboloid as defined in Section A2. In this section, we show that this operator defines a linear isomorphism over suitably weighted Sobolev and Hölder spaces. We first describe the spaces of interest to us (see Section A4 for details). Let g denote the metric induced over $\mathbb{R}^2$ by the graph G_0, that is,

$$g := (1+v^2)\,dr^2 + r^2\,d\theta^2. \tag{3-1}$$

For all nonnegative, integer m, let $\|\cdot\|_{H^m(G)}$ denote the *Sobolev norm* of order m of functions over $\mathbb{R}^2$ with respect to this metric. Likewise, for all nonnegative, integer m, and, for all $\alpha \in [0, 1]$, let $\|\cdot\|_{C^{m,\alpha}(G)}$ denote the *Hölder norm* of order (m, α) of functions over $\mathbb{R}^2$ with respect to this metric. Observe that, by (2-4), these Sobolev and Hölder norms are uniformly equivalent to the Sobolev and Hölder norms defined with respect to the more straightforward metric

$$g' := (1+r^2)\,dr^2 + r^2\,d\theta^2. \tag{3-2}$$

For all nonnegative, integer m, let $H^m(G)$ denote the *Sobolev space* of measurable functions f over $\mathbb{R}^2$ whose distributional derivatives up to and including order m are locally square integrable and which satisfy $\|f\|_{H^m(G)} < \infty$. Likewise, for all nonnegative, integer m, and, for all $\alpha \in [0, 1]$, let $C^{m,\alpha}(G)$ denote the *Hölder space* of m-times differentiable functions f over $\mathbb{R}^2$ which satisfy $\|f\|_{C^{m,\alpha}(G)} < \infty$. Recall that both $H^m(G)$ and $C^{m,\alpha}(G)$, furnished with the above norms, are Banach spaces.

For all real γ, define $\phi_\gamma : \mathbb{R}^2 \to \mathbb{R}$ by

$$\phi_\gamma := e^{(1+\gamma)u/2}. \tag{3-3}$$

where u here denotes the integral of v with initial value 0. For all nonnegative, integer m, for all $\alpha \in [0, 1]$ and for all real γ, define the *weighted Sobolev and Hölder norms* of weight γ over $\mathbb{R}^2$ by

$$\begin{aligned} \|f\|_{H^m_\gamma(G)} &:= \|\phi_\gamma f\|_{H^m(G)}, \\ \|f\|_{C^{m,\alpha}_\gamma(G)} &:= \|\phi_\gamma f\|_{C^{m,\alpha}(G)}. \end{aligned} \tag{3-4}$$

Observe that, by (2-4) again, these *weighted Sobolev and Hölder norms* are uniformly equivalent to the weighted norms defined using instead of ϕ_γ the more straightforward weight function

$$\phi'_\gamma := e^{(1+\gamma)r^2/4}. \tag{3-5}$$

For all nonnegative, integer m, for all $\alpha \in [0, 1]$, and for all real γ, define the *weighted Sobolev* and *Hölder spaces* of *weight* γ over $\mathbb{R}^2$ by

$$\begin{aligned} H^m_\gamma(G) &:= \{f \mid \phi_\gamma f \in H^m(G)\}, \\ C^{m,\alpha}_\gamma(G) &:= \{f \mid \phi_\gamma f \in C^{m,\alpha}(G)\}. \end{aligned} \tag{3-6}$$

These spaces, furnished with the weighted Sobolev and Hölder norms are trivially also Banach spaces.

Since G_0 is a graph over $\mathbb{R}^2$, its MCFS Jacobi operator may be thought of as an operator acting on functions over $\mathbb{R}^2$. In particular, as we will see presently, for all $\alpha \in [0, 1]$, and for all real γ, J defines bounded linear maps from $H^2_\gamma(G)$ into $H^0_\gamma(G)$ and from $C^{2,\alpha}_\gamma(G)$ into $C^{0,\alpha}_\gamma(G)$. We show:

Theorem 3.1.1. (1) *For all sufficiently small γ, J defines a linear isomorphism from $H^2_\gamma(G)$ into $H^0_\gamma(G)$.*

(2) *For all $\alpha \in]0, 1[$ and for all sufficiently small γ, J defines a linear isomorphism from $C^{2,\alpha}_\gamma(G)$ into $C^{0,\alpha}_\gamma(G)$.*

Theorem 3.1.1 will follow from Lemmas 3.2.6 and 3.3.4 below. Before proceeding, we first observe that, for all γ, the function ϕ_γ is strictly positive so that, for all nonnegative, integer m, and for all $\alpha \in [0, 1]$, the operator of multiplication by this function, which we denote by M_γ, defines linear isomorphisms from $H^m_\gamma(G)$ into $H^m(G)$ and from $C^{m,\alpha}_\gamma(G)$ into $C^{m,\alpha}(G)$. For all real γ, we therefore define

$$J_\gamma := M_\gamma J M_\gamma^{-1}. \tag{3-7}$$

This operator is none other than the ϕ_γ-Jacobi operator of the Grim paraboloid, which has been studied in detail in [Cheng and Zhou 2015; Cheng et al. 2014; 2015a; 2015b]. Trivially, J defines linear isomorphisms from $H^2_\gamma(G)$ into $H^0_\gamma(G)$ and from $C^{2,\alpha}_\gamma(G)$ into $C^{0,\alpha}_\gamma(G)$ if and only if J_γ defines linear isomorphisms from $H^2(G)$ into $H^0(G)$ and from $C^{2,\alpha}(G)$ into $C^{0,\alpha}(G)$ respectively.

Lemma 3.1.2. *For all real γ,*

$$J_\gamma f = \Delta^{G_0} f - \gamma\langle e_z, \nabla^{G_0} f\rangle + \frac{(\gamma^2-1)}{4} f - \frac{(1+\gamma)^2}{4}\langle e_z, N_{G_0}\rangle^2 f + \mathrm{Tr}(A^2_{G_0}) f. \tag{3-8}$$

Proof. By (A-3),

$$\nabla^{G_0}\phi_\gamma^{-1} = -\frac{(1+\gamma)}{2\phi_\gamma}\pi^{G_0}(e_z),$$
$$\mathrm{Hess}^{G_0}\phi_\gamma^{-1} = \frac{(1+\gamma)^2}{4\phi_\gamma} dz\otimes dz + \frac{(1+\gamma)}{2\phi_\gamma}\langle e_z, N_{G_0}\rangle \mathrm{II}^{G_0}.$$

However, since G_0 is a mean curvature flow soliton, $H_{G_0} = -\langle e_z, N_{G_0}\rangle$, and taking the trace therefore yields

$$\Delta^{G_0}\phi_\gamma^{-1} = \frac{(1+\gamma)^2}{4\phi_\gamma} - \frac{(1+\gamma)(3+\gamma)}{4\phi_\gamma}\langle e_z, N_{G_0}\rangle^2.$$

Thus, by (A-2),

$$\phi_\gamma J_0 \phi_\gamma^{-1} = \frac{(\gamma^2-1)}{4} - \frac{(1+\gamma)^2}{4}\langle e_z, N_{G_0}\rangle^2 + \mathrm{Tr}(A_{G_0})^2.$$

The result now follows by (A-4). □

By (A-6) and (2-13),

$$\begin{aligned}\langle e_z, N_{G_0}\rangle^2 &= \mathrm{O}(r^{-(2+k)}),\\ \mathrm{Tr}(A^2_{G_0}) &= \mathrm{O}(r^{-(2+k)}).\end{aligned} \tag{3-9}$$

It follows that, as γ tends to 0, the family (J_γ) converges to J_0 in every operator norm of relevance to us. Since invertibility is stable under small perturbations, it is therefore sufficient to consider only the case $\gamma = 0$ where, in particular, J_0 is self-adjoint.

We now derive a formula for J_0 which is better adapted to our purposes. First, let $c :]0, \infty[\to \mathbb{R}$ be such that, for all r, $c(r)$ is the geodesic curvature of the circle $C(r)$ with respect to the metric induced by the graph G_0 over $\mathbb{R}^2$.

Lemma 3.1.3. *The function c is given by*

$$c = \frac{1}{r}\langle e_z, N_{G_0}\rangle. \tag{3-10}$$

In particular, for large values of r,

$$c = \mathrm{O}(r^{-(2+k)}). \tag{3-11}$$

Proof. Let D denote the Levi-Civita covariant derivative of the Euclidean metric over $\mathbb{R}^3$. Think of $C(r)$ as a horizontal circle in $\mathbb{R}^3$ at height $u(r)$, where u here denotes the integral of v with initial value 0. In particular, $D_{e_\theta}e_\theta = (1/r)e_r$, where e_θ and e_r denote respectively the unit, horizontal vector fields in the angular and radial directions about the z-axis. Since the geodesic curvature of $C(r)$ with respect to the induced metric over G_0 is equal to the length of the tangential component of this vector, the function c is given by

$$c = \frac{1}{r}\sqrt{1 - \langle e_r, N_{G_0}\rangle^2} = \frac{1}{r}\langle e_z, N_{G_0}\rangle,$$

as desired. Equation (3-11) now follows from (3-9), and this completes the proof. □

Let $\rho :]0, \infty[\to \mathbb{R}$ be such that, for all r, $\rho(r)$ is the intrinsic distance along G_0 of any point on the circle $C(r)$ from the origin. Since ρ is obtained by integrating $\sqrt{1+v^2}$, by (2-4) again, for large values of r,

$$\begin{aligned} \rho_r &= r + \mathrm{O}(r^{-(k+1)}), \\ r_\rho &= \frac{1}{r} + \mathrm{O}(r^{-(k+3)}), \end{aligned} \tag{3-12}$$

where the subscripts r and ρ here denote differentiation with respect to the variables r and ρ respectively.

Lemma 3.1.4. *Away from the z-axis,*

$$J_0 f = f_{\rho\rho} + f_{\theta\theta} + c f_\rho - \tfrac{1}{4} f + \psi f, \tag{3-13}$$

where the subscripts ρ and θ denote differentiation along the unit radial and unit angular directions in G_0 and, for large values of ρ,

$$|\psi| \lesssim \rho^{-1}. \tag{3-14}$$

Proof. Indeed, away from the z-axis,

$$\Delta^{G_0} f = f_{\rho\rho} + f_{\theta\theta} + c f_\rho,$$

so that (3-13) follows by (3-8) and (3-9) with

$$|\psi| \lesssim r^{-2}.$$

Finally, integrating (3-12), yields $\rho \lesssim r^2$, so that $r^{-2} \lesssim \rho^{-1}$ and the result follows. □

3.2. ***Invertibility over Sobolev spaces.*** We now obtain the invertibility of J_0 for Sobolev spaces. The main technical difficulty here arises from the noncompactness of the ambient space. This is compensated for by the following estimate.

Lemma 3.2.1. *There exist $B, R > 0$ such that, for all f in $H^2(G)$,*

$$\|f|_{A(R,\infty)}\|_{L^2(G)} \leq B\big(\|f|_{A(R-1,R+1)}\|_{L^2(G)} + \|J_0 f|_{A(R-1,\infty)}\|_{L^2(G)}\big). \tag{3-15}$$

Proof. Since $C_0^\infty(G)$ is dense in $H^2(G)$, it suffices to prove the result when f is smooth and has compact support. Set $g := J_0 f$ and define $\alpha, \beta :]0, \infty[\to \mathbb{R}$ by

$$\alpha(\rho) := \int_{C(\rho)} f^2\, dl, \quad \beta(\rho) := \int_{C(\rho)} g^2\, dl,$$

where $C(\rho)$ here denotes the circle of points lying at intrinsic distance ρ along G_0 from the origin. Twice differentiating α yields

$$\alpha_\rho = \int_{C(\rho)} 2ff_\rho + f^2 c\, dl,$$
$$\alpha_{\rho\rho} = \int_{C(\rho)} 2f_\rho^2 + 2ff_{\rho\rho} + 4ff_\rho c + f^2 c_\rho + f^2 c^2\, dl,$$

where the subscript ρ here denotes differentiation with respect to this variable. By (3-13),

$$\alpha_{\rho\rho} = \int_{C(\rho)} 2f_\rho^2 - 2ff_{\theta\theta} + \tfrac{1}{2}f^2 - 2\psi f^2 + 2fg + 2ff_\rho c + f^2 c_\rho + f^2 c^2\, dl.$$

Integrating the term $2ff_{\theta\theta}$ by parts and applying the algebraic-geometric mean inequality now yields

$$\alpha_{\rho\rho} \geq \int_{C(\rho)} \big(\tfrac{1}{4} - 2\psi + c_\rho - c^2\big) f^2 - 4g^2\, dl.$$

However, by (3-11), (3-12) and (3-14), c, $c_\rho = c_r r_\rho$ and ψ all tend to 0 as ρ tends to $+\infty$ so that, for sufficiently large ρ,

$$\alpha_{\rho\rho} \geq \tfrac{1}{8}\alpha - 4\beta.$$

Since f has compact support, upon integrating this relation we obtain, for sufficiently large R,

$$\|f|_{A(R,\infty)}\|_{L^2(G)}^2 = \int_R^\infty \alpha\, d\rho \leq 32 \int_R^\infty \beta\, d\rho - 8\alpha_\rho(R) = 32\|\hat{J}_0 f|_{A(R,\infty)}\|_{L^2(G)}^2 - 8\alpha_\rho(R).$$

However, by the Sobolev trace formula and classical elliptic estimates,

$$\begin{aligned}\alpha_\rho(R) &\leq B_1\|f|_{A(R-1/2,R+1/2)}\|_{H^2(G)}^2\\ &\leq B_2(\|f|_{A(R-1,R+1)}\|_{L^2(G)}^2 + \|J_0 f|_{A(R-1,R+1)}\|_{L^2(G)}^2)\end{aligned}$$

for suitable constants B_1 and B_2. The result now follows upon combining the last two relations. □

Combining Lemma 3.2.1 with classical elliptic estimates yields:

Lemma 3.2.2. *There exist B, $R > 0$ such that, for all f in $H^2(G)$,*

$$\|f\|_{H^2(G)} \leq B(\|f|_{B(R)}\|_{L^2(G)} + \|J_0 f\|_{L^2(G)}). \tag{3-16}$$

Proof. Observe that G_0 is of bounded geometry in the sense that, as x tends to infinity in G_0, the geodesic ball of unit radius about x in this surface converges in the pointed Cheeger–Gromov sense to the unit ball about the origin in $\mathbb{R}^2$. It thus follows by classical elliptic theory (see [Gilbarg and Trudinger 1983]) that there exists $B > 0$ such that

$$\|f\|_{H^2(G)} \leq B(\|f\|_{L^2(G)} + \|J_0 f\|_{L^2(G)}).$$

The result follows upon combining this relation with (3-15). □

Since, J_0 is self-adjoint, standard arguments of the theory of elliptic operators now yield

Lemma 3.2.3. *J_0 defines a Fredholm map from $H^2(G)$ into $L^2(G)$ of Fredholm index equal to* 0.

Proof. Since $B(R)$ is relatively compact, it follows by Rellich's compactness theorem that the restriction map sending $H^2(G)$ into $L^2(B(R))$ is also compact. Thus, by (3-16), J_0 satisfies an elliptic estimate, as defined in Section A5, so that, by Theorem A5.1, J_0 has finite-dimensional kernel and closed image. Observe now that J_0 is self adjoint, so that $\mathrm{Ker}(J_0)$ is contained within the orthogonal complement of $\mathrm{Im}(J_0)$ in $L^2(G)$. We claim that these two spaces coincide. Indeed, let u be an element of the orthogonal complement of $\mathrm{Im}(J_0)$. In particular, $J_0 u = 0$ in the distributional sense. Thus, bearing in mind that G_0 is of bounded geometry, it follows by classical elliptic regularity that u is an element of $H^2(G)$. In particular, u is therefore an element of $\mathrm{Ker}(J_0)$, so that $\mathrm{Ker}(J_0)$ coincides with the orthogonal complement of $\mathrm{Im}(J_0)$ in $L^2(G)$, as asserted. It immediately follows that J_0 is a Fredholm map of Fredholm index equal to 0, and this completes the proof. □

It remains only to prove that J_0 has trivial kernel in $H^2(G)$. We obtain a slightly more general result which will serve also for the Hölder space case of the following section.

Lemma 3.2.4. *There exists no nontrivial, bounded function $f : G_0 \to \mathbb{R}$ such that $J_0 f = 0$.*

Proof. Indeed, suppose that there exists a nontrivial bounded function $f : G_0 \to \mathbb{R}$ such that $J_0 f = 0$. Upon multiplying by (-1), we may suppose that f is positive at some point. Now, since all vertical translates of G_0 are also mean curvature flow solitons, the function $\mu = \langle e_z, N_G\rangle$ is a Jacobi field over this surface, that is,

$$J_0 \phi_0 \mu = \phi_0 J \mu = 0.$$

Since G_0 is a graph, the function μ is everywhere strictly positive. It follows that $\phi_0\mu$ is also positive, so that the quotient $f/\phi_0\mu$ is smooth. Since $\phi_0 \gtrsim e^{r^2/4}$ and $\mu = \mathrm{O}(r^{-1})$, the function $\phi_0\mu$ tends to infinity as r tends to infinity, and so $f/\phi_0\mu$ attains its maximum value at some point x, say, of G_0. In particular, upon rescaling, we may suppose that $f/\phi_0\mu \leq 1$ and that $f(x)/\phi_0(x)\mu(x) = 1$.

Bearing in mind that μ is positive, we define the operator $J_\mu := M_\mu^{-1} J M_\mu$, where M_μ here denotes the operator of multiplication by μ. Since $J\mu = 0$, by (A-4), this operator has no zeroth order term. Thus, since $J_\mu(f/\mu\phi_0)=(1/\mu\phi_0)J_0f=0$, it follows by the strong maximum principle that $f/\phi_0\mu$ is constant and equal to 1. However, since $\phi_0\mu$ is unbounded, this is absurd, and the result follows. □

Corollary 3.2.5. *J_0 has trivial kernel in $H^2(G)$.*

Proof. Indeed, by the Sobolev embedding theorem, every element of $H^2(G)$ is bounded, and the result now follows by Lemma 3.2.4. □

The above results together with a perturbation argument now yield

Lemma 3.2.6. *For sufficiently small γ, J defines a linear isomorphism from $H^2_\gamma(G)$ into $H^0_\gamma(G)$.*

3.3. ***Invertibility over Hölder spaces.*** We prove the invertibility of J_0 over $C^{2,\alpha}(G)$ in essentially the same manner. We first require the following preliminary result.

Lemma 3.3.1. *Let α and β be positive constants. If $\phi:[0,\infty[\to]0,\infty[$ is a bounded, positive function such that $\phi''\geq\alpha^2\phi-\beta$ in the viscosity sense, then, for all t,*

$$\phi(t)\leq\mathrm{Max}\left(\phi(0)-\frac{\beta}{\alpha^2},0\right)e^{-\alpha t}+\frac{\beta}{\alpha^2}. \tag{3-17}$$

Proof. Let $A=\mathrm{Max}(\phi(0)-\beta/\alpha^2,0)$ and let $B=\mathrm{Sup}_{t\in[0,\infty[}\phi(t)$. Fix $T>0$ and define

$$f=\frac{Be^{\alpha T}-A}{e^{2\alpha T}-1}e^{\alpha t}+\frac{A-Be^{-\alpha T}}{1-e^{-2\alpha T}}e^{-\alpha t}+\frac{\beta}{\alpha^2}.$$

In other words, f is the unique solution of the ODE problem $f_{tt}=\alpha^2f-\beta$ with boundary values $f(0)=A+\beta/\alpha^2\geq\phi(0)$ and $f(T)=B+\beta/\alpha^2\geq\phi(T)$. Let C be the minimum value of $f-\phi$ over $[0,T]$ and let $t\in[0,T]$ be the point at which this minimum is attained. If t is a boundary point of this interval, then $C\geq 0$. Otherwise, $f-C\geq\phi$ and $f(t)-C=\phi(t)$. Thus, since ϕ is a viscosity solution of $\phi''\geq\alpha^2\phi-\beta$, at this point, we have

$$\alpha^2f-\beta=(f-C)_{tt}\geq\alpha^2(f-C)-\beta$$

so that, once again, $C\geq 0$. In each case, we therefore obtain

$$\phi\leq f=\frac{Be^{\alpha T}-A}{e^{2\alpha T}-1}e^{\alpha t}+\frac{A-Be^{-\alpha T}}{1-e^{-2\alpha T}}e^{-\alpha t}+\frac{\beta}{\alpha^2},$$

and the result follows upon taking the limit as T tends to $+\infty$. □

As in the Sobolev case, the noncompactness of the ambient space is compensated for by the following estimate.

Lemma 3.3.2. *There exist $B,R>0$ such that, for all f in $C^{2,\alpha}(G)$,*

$$\|f|_{A(R,\infty)}\|_{C^0(G)}\leq B(\|f|_{C(R)}\|_{C^0(G)}+\|J_0f|_{A(R-1,\infty)}\|_{C^0(G)}). \tag{3-18}$$

Proof. Define $\alpha:]0,\infty[\to\mathbb{R}$ by

$$\alpha(\rho):=\mathrm{Sup}_{x\in C(\rho)}f(x)^2,$$

where $C(\rho)$ here denotes the circle of points lying at intrinsic distance ρ along G_0 from the origin. Denote $g := J_0 f$, and define $B \geqslant 0$ by

$$B := \|g^2|_{A(R,\infty)}\|_{C^0(G)}.$$

Choose $x \in C(\rho)$ maximising f^2, and observe that $f f_{\theta\theta}$ is nonpositive at this point. Thus, bearing in mind (3-13),

$$\begin{aligned}(f^2)_{\rho\rho} &= 2f_\rho^2 + 2ff_{\rho\rho},\\ &\geq 2f_\rho^2 + 2fg - 2cf_\rho + \tfrac{1}{2}f^2 - 2\psi f^2,\\ &\geq \left(\tfrac{1}{4} - \tfrac{1}{2}c^2 - 2\psi\right)f^2 - 4g^2.\end{aligned}$$

By (3-11) and (3-14), for sufficiently large ρ

$$(f^2)_{\rho\rho} \geq \tfrac{1}{8}f^2 - 4g^2.$$

Since α is the envelope of the restriction of $f(x)^2$ to each radial line, it follows that over $[R,\infty[$,

$$\alpha_{\rho\rho} \geq \tfrac{1}{8}\alpha - 4B,$$

in the viscosity sense. Thus, by Lemma 3.3.1,

$$\mathrm{Sup}_{x\in A(R,\infty)} f^2(x) = \mathrm{Sup}_{\rho\geq R}\,\alpha(\rho) \leq \mathrm{Max}(\|f^2|_{C(R)}\|_{C^0} - 32B, 0) + 32B,$$

and the result follows. □

Using classical elliptic estimates again, this yields

Lemma 3.3.3. *There exist B, $R > 0$ such that for all f in $C^{2,\alpha}(G)$,*

$$\|f\|_{C^{2,\alpha}(G)} \leq B(\|f|_{B(R)}\|_{C^0(G)} + \|J_0 f\|_{C^{0,\alpha}(G)}). \tag{3-19}$$

Proof. Recall that G_0 is of bounded geometry in the sense that, as x tends to infinity in G_0, the geodesic ball of unit radius about x in this surface converges in the pointed Cheeger–Gromov sense to the unit ball about the origin in $\mathbb{R}^2$. It thus follows by classical elliptic theory (see [Gilbarg and Trudinger 1983]) that there exists $B > 0$ such that

$$\|f\|_{C^{2,\alpha}(G)} \leq B(\|f\|_{C^0(G)} + \|J_0 f\|_{C^{0,\alpha}(G)}),$$

and the result now follows upon combining this relation with (3-18). □

As before, this yields the desired invertibility result.

Lemma 3.3.4. *For all α and for all sufficiently small γ, J defines a linear isomorphism from $C^{2,\alpha}_\gamma(G)$ into $C^{0,\alpha}_\gamma(G)$.*

Proof. Recall that this is equivalent to showing that, for sufficiently small γ, J_γ defines a linear isomorphism from $C^{2,\alpha}(G)$ into $C^{0,\alpha}(G)$. Furthermore, by (3-8) and (3-9), J_γ converges to J_0 in the operator norm as γ tends to 0, so that it suffices to prove the result for J_0.

Since $B(R)$ is a relatively compact subset of G_0, it follows by the Arzelà–Ascoli theorem that the restriction map of $C^{2,\alpha}(G)$ into $C^0(B(R))$ is compact. Thus, by (3-19), J_0 satisfies an elliptic estimate, as defined in Section A5. By Theorem A5.1, the image of J_0 is closed and, in particular, is a Banach subspace of $C^{0,\alpha}(G)$. Furthermore, by Lemma 3.2.4, the kernel of J_0 in $C^{2,\alpha}(G)$ is trivial, so that, by the closed graph theorem, J_0 defines a linear isomorphism from $C^{2,\alpha}(G)$ into its image. In particular, there exists a constant $B > 0$ such that, for all $u \in C^{2,\alpha}(G)$,

$$\|u\|_{2,\alpha} \leq B\|J_0 u\|_{0,\alpha}. \tag{3-20}$$

It remains only to prove surjectivity. Choose $v \in C^{0,\alpha}(G)$ and let (v_m) be a sequence of smooth functions of compact support in $\mathbb{R}^2$ which is bounded in $C^{0,\alpha}(G)$ and which converges to v in the $C^{0,\beta}_{\text{loc}}$ sense for all $\beta < \alpha$. For all m, since v_m is a smooth function with compact support, it is an element of $L^2(G)$ so that, by Lemma 3.2.6, there exists an element u_m of $H^2(G)$ such that $J_0 u_m = v_m$. Since G_0 is of bounded geometry, it follows by classical elliptic regularity that, for all m, u_m is in fact an element of $C^{2,\alpha}(G)$. In particular, by (3-20), for all m,

$$\|u_m\|_{C^{2,\alpha}(G)} \leq B\|v_m\|_{C^{0,\alpha}(G)}.$$

Since the sequence (u_m) is uniformly bounded in $C^{2,\alpha}(G)$, it follows by the Arzelà–Ascoli theorem there exists $u \in C^{2,\alpha}(G)$ towards which (u_m) subconverges in the $C^{2,\beta}_{\text{loc}}$-topology for all $\beta < \alpha$. By continuity, $J_0 u = v$ and surjectivity follows. □

4. Rotationally symmetric Grim ends

4.1. ***The modified MCFS Jacobi operator.*** We now consider the case of rotationally symmetric Grim ends. Let Λ be a large, positive real number, let $K > 0$ be fixed, and let $\epsilon, R > 0$ and $c \in \mathbb{R}$ satisfy (2-14). Let $v : [\epsilon R, \infty[\to \mathbb{R}$ solve (2-5) with logarithmic parameter c so that, by (2-16), over the interval $[\epsilon R, \epsilon R^4]$,

$$v = \frac{1}{2}r + \frac{c\epsilon}{r} + O\left(\left[1 + \log\left(\frac{r}{\epsilon R}\right)\right]\frac{1}{r^k}\left(r + \frac{\epsilon}{r}\right)^3\right). \tag{4-1}$$

Let $u : [\epsilon R, \infty[\to \mathbb{R}$ be a primitive of v, let G be the Grim end generated by rotating the graph of u about the z-axis, and let J be its MCFS Jacobi operator, as defined in Section A2.

Since G is a graph over $A(\epsilon R, \infty)$, J may again be thought of as an operator acting on functions over this annulus. For all nonnegative, integer m, for all $\alpha \in [0, 1]$, and for all real γ, we define the norms $\|\cdot\|_{H^m_\gamma(G)}$ and $\|\cdot\|_{C^{m,\alpha}_\gamma(G)}$ as in Section 3. For all nonnegative integer m, for all $\alpha \in [0, 1]$ and for all real γ, we define the *hybrid norm* with weight γ of functions over $\mathbb{R}^m$ by

$$\|f\|_{m,\alpha,\gamma} := \|f\|_{C^{m,\alpha}_\gamma(G)} + \frac{1}{(\epsilon R)}\|f\|_{H^m_\gamma(G)}. \tag{4-2}$$

As we will see in Section 6, this norm encapsulates the asymptotic behaviour of J as Λ tends to infinity. Let $\mathcal{L}^{m,\alpha}_\gamma(G)$ denote the Banach space of m-times differentiable functions f over $\mathbb{R}^2$ with finite hybrid norm. In this section, we show that, for sufficiently small γ, and for sufficiently large Λ, the operator J *more or less* defines linear isomorphisms from $\mathcal{L}^{2,\alpha}_\gamma(G)$ into $\mathcal{L}^{0,\alpha}_\gamma(G)$ and, furthermore, that the norms of this isomorphism and its inverse are uniformly bounded as Λ tends to infinity. In order to properly formalise these assertions, we now apply the following two modifications.

First, on account of the vanishing neck problem, discussed in the Introduction, the zeroth-order coefficient of J diverges rapidly over the annulus $A(\epsilon R, \epsilon R^4)$ as Λ tends to infinity. We address this by introducing what we call the modified MCFS Jacobi operator. Recall that different modifications are applied at different scales, so that the definition of this operator varies according to context, and the general framework will be discussed in Section 5.4, below. In the present case, the modified MCFS Jacobi operator is defined as follows. Let χ_1 be the cut-off function of the transition region $A(1,2)$ as defined in Section A1 and define $\psi : A(\epsilon R, \infty) \to \mathbb{R}$ by

$$\psi(r) = \chi_1 \langle e_z, N_G \rangle + (1 - \chi_1), \tag{4-3}$$

where N_G here denotes the upward-pointing unit normal vector field over G. Bearing in mind that ψ is always positive, the *modified MCFS Jacobi operator* of G is now defined by

$$\hat{J} := M_\psi^{-1} J M_\psi, \tag{4-4}$$

where M_ψ here denotes the operator of multiplication by ψ.

Next, observe that $\hat{J}$ is in fact only defined over the annulus $A(\epsilon R, \infty)$. We thus extend it to an operator defined over the whole of $\mathbb{R}^2$ as follows. Given a function $\phi : A(\epsilon R, \infty) \to \mathbb{R}$, we define its *canonical extension* $\tilde{\phi} : \mathbb{R}^2 \to \mathbb{R}$ such that $\tilde{\phi}(x) = \phi(x)$ over $A(\epsilon R, \infty)$, $\tilde{\phi}(0)$ is equal to the mean value of ϕ over the circle $C(\epsilon R)$, and $\tilde{\phi}$ restricts to a linear function over every radial line in $B(\epsilon R)$. In particular, if ϕ is Lipschitz, then so too is $\tilde{\phi}$, and

$$\|\tilde{\phi}\|_{C^{0,1}} \leq \tfrac{\pi}{2} \|\phi\|_{C^{0,1}}.$$

Now, given a linear operator L over $A(\epsilon R, \infty)$, we define its *canonical extension* $\widetilde{L}$ to be the operator over $\mathbb{R}^2$ whose coefficients are the canonical extensions of each of the coefficients of L. We henceforth identify all operators with their canonical extensions over $\mathbb{R}^2$. Observe, in particular, that if L has any rotational symmetries, then so too does its canonical extension.

Theorem 4.1.1. *For all sufficiently small $\alpha \in]0,1[$ and for all sufficiently large Λ, $\hat{J}$ defines a linear isomorphism from $\mathcal{L}^{2,\alpha}_\gamma(G)$ into $\mathcal{L}^{0,\alpha}_\gamma(G)$. Furthermore, the operator norms of $\hat{J}$ and its inverse are uniformly bounded independent of Λ.*

Theorem 4.1.1 follows from Theorem 3.1.1 by a perturbation argument and Lemmas 4.2.7 and 4.3.4.

We conclude this section by deriving formulae for $\hat{J}$ over different regions.

Lemma 4.1.2. *Over $A(\epsilon R, 1)$, the modified MCFS Jacobi operator of G is given by*

$$\hat{J} f = g^{ij} f_{ij} - 2\mu g^{ip} g^{jq} u_{pq} u_j f_i. \tag{4-5}$$

Proof. First observe that, for every tangent vector X over G,

$$\langle \nabla^G \psi, X \rangle = X\psi = X \langle N_G, e_z \rangle = \langle D_X N_G, e_z \rangle = \langle A_G X, e_z \rangle = \langle X, A_G \pi^G(e_z) \rangle,$$

and so,

$$\nabla^G \psi = A_G \pi^G(e_z).$$

Since every vertical translate of G is also a rotationally symmetric Grim end, $J\langle e_z, N_G\rangle = 0$, and so, by (A-4),

$$\hat{J}f = \Delta^G f + \langle e_z, \nabla^G f\rangle + 2\psi^{-1}\langle A_G\nabla^G f, e_z\rangle.$$

By (A-3),

$$\mathrm{Hess}^G f = \mathrm{Hess}(f)\circ\pi - \langle D(f\circ\pi), N\rangle \mathrm{II}_G.$$

Furthermore, since $D(f\circ\pi)$ is horizontal

$$\langle D(f\circ\pi), N_G\rangle = -\frac{1}{\langle N_G, e_z\rangle}\langle D(f\circ\pi), e_z - \langle N_G, e_z\rangle N_G\rangle = -\frac{1}{\langle N_G, e_z\rangle}\langle\nabla^G f, e_z\rangle.$$

Taking the trace therefore yields

$$\Delta^G f = g^{ij}f_{ij} + \frac{1}{\langle N_G, e_z\rangle}\langle\nabla^G f, e_z\rangle H_G.$$

However, since G is a mean curvature flow soliton, $H_G = -\langle N, e_z\rangle$, and so

$$\Delta^G f = g^{ij}f_{ij} - \langle\nabla^G f, e_z\rangle.$$

We conclude that

$$\hat{J}f = g^{ij}f_{ij} + 2\psi^{-1}\langle A_G\nabla^G f, e_z\rangle,$$

and the result now follows by (A-6). □

Lemma 4.1.3. *Over $A(\epsilon R, 2\epsilon R^4)$, the modified MCFS Jacobi operator of G satisfies*

$$\hat{J}f = \Delta f - \left(\frac{1}{2} + \frac{c\epsilon}{r^2}\right)^2 x^i x^j f_{ij} - \left(\frac{1}{2} - \frac{2c^2\epsilon^2}{r^4}\right)x^i f_i + \mathcal{E}_G f, \tag{4-6}$$

where $\mathcal{E}_G f := a^{ij}f_{ij} + b^i f_i$, and a and b satisfy

$$\begin{aligned} a &= \mathrm{O}\Big(\Big[1 + \log\Big(\frac{r}{\epsilon R}\Big)\Big]\frac{1}{r^k}\Big(r + \frac{\epsilon}{r}\Big)^4\Big),\\ b &= \mathrm{O}\Big(\Big[1 + \log\Big(\frac{r}{\epsilon R}\Big)\Big]\frac{1}{r^{k+1}}\Big(r + \frac{\epsilon}{r}\Big)^4\Big). \end{aligned} \tag{4-7}$$

Proof. Indeed, by (4-1),

$$u_i = \frac{1}{2}x_i + \frac{c\epsilon}{r^2}x_i + \mathrm{O}\Big(\Big[1 + \log\Big(\frac{r}{\epsilon R}\Big)\Big]\frac{1}{r^k}\Big(r + \frac{\epsilon}{r}\Big)^3\Big).$$

Thus, by (A-6),

$$\begin{aligned} \mu^2 &= 1 - \Big(\frac{r}{2} + \frac{c\epsilon}{r}\Big)^2 + \mathrm{O}\Big(\Big[1 + \log\Big(\frac{r}{\epsilon R}\Big)\Big]\frac{1}{r^k}\Big(r + \frac{\epsilon}{r}\Big)^4\Big),\\ g^{ij} &= \delta_{ij} - \Big(\frac{1}{2} + \frac{c\epsilon}{r^2}\Big)^2 x^i x^j + \mathrm{O}\Big(\Big[1 + \log\Big(\frac{r}{\epsilon R}\Big)\Big]\frac{1}{r^{k+1}}\Big(r + \frac{\epsilon}{r}\Big)^4\Big). \end{aligned}$$

It follows that

$$g^{ij}f_{ij} = \Delta f - \Big(\frac{1}{2} + \frac{c\epsilon}{r^2}\Big)^2 x^i x^j f_{ij} + a^{ij}f_{ij},$$

where $a = \mathrm{O}([1+\log(r/\epsilon R)]r^{-k}(r+\epsilon/r)^4)$, and since $r^{-1}(r+\epsilon/r)^4$ bounds $(r+\epsilon/r)^3$,

$$-2\mu g^{ip}g^{jq}u_{pq}u_i f_j = -\left(\frac{1}{2} - \frac{2\epsilon^2 c^2}{r^4}\right)x^i f_i + b^i f_i,$$

where $b = \mathrm{O}([1+\log(r/\epsilon R)]r^{-(k+1)}(r+\epsilon/r)^4)$. The result follows. □

4.2. ***The regular component.*** Theorem 4.1.1 is derived from Theorem 3.1.1 by a perturbation argument. First, let $v_p :]0,\infty[\to \mathbb{R}$ denote the unique solution of (2-5) which is defined over the whole positive half-line, as in Section 3. Let u_p denote its primitive with initial value 0 so that its graph is a Grim paraboloid. Let $\hat{J}_p$ denote its modified MCFS Jacobi operator, as defined in Section 4.1. Over the ball $B(2\epsilon R^4)$,

$$v_p(r) = \tfrac{1}{2}r + O(r^{3-k}), \tag{4-8}$$

so that, as in Lemma 4.1.3, over $B(0, 2\epsilon R^4)$,

$$\hat{J}_p f = \Delta f - \tfrac{1}{2}x^i x^j f_{ij} - \tfrac{1}{2}x^i f_i + \mathcal{E}_p f, \tag{4-9}$$

where $\mathcal{E}_p f := a^{ij} f_{ij} + b^i f_i$ and

$$a = \mathrm{O}(r^{4-k}), \quad b = \mathrm{O}(r^{3-k}). \tag{4-10}$$

Define

$$\hat{J}_\gamma := M_\gamma^{-1}\hat{J}M_\gamma, \tag{4-11}$$

where M_γ here denotes the operator of multiplication by $\chi_2 + (1-\chi_2)\phi_\gamma$, ϕ_γ is given by (3-3), and χ_2 is the cut-off function of the transition region $A(2,4)$ as defined in Section A1. Observe that, since ϕ_γ and ψ only depend on v and its integral u, it follows by (A-2) that the coefficients of $\hat{J}_\gamma$ are functions of u, v and v_r only. Finally, define

$$\hat{J}_{p,\gamma} := M_\gamma^{-1}\hat{J}_p M_\gamma. \tag{4-12}$$

A straightforward modification of Theorem 3.1.1 shows that, for all $\alpha \in]0,1[$, and for all sufficiently small γ, $\hat{J}_{p,\gamma}$ defines a linear isomorphism from $\mathcal{L}^{2,\alpha}_\gamma(G)$ into $\mathcal{L}^{0,\alpha}_\gamma(G)$ whose Green's operator has norm uniformly bounded independent of Λ.

It will suffice to show that the difference $\hat{J}_{p,0} - \hat{J}_0$ converges to 0 with respect to the hybrid norm as Λ tends to $+\infty$. This is, in fact, a nontrivial result, since the coefficients of this operator diverge. However, the region over which they diverge itself converges to a point; the relative rates of convergence are such that the coefficients converge in the mean, which will be sufficient for us to conclude. Formally, we define the operators D and E over $A(\epsilon R, \infty)$ by

$$\begin{aligned} Df &:= (\hat{J}_0 - E)f - \hat{J}_{p,0}f, \\ Ef &:= \chi\frac{2c^2\epsilon^2}{r^4}x^i f_i, \end{aligned} \tag{4-13}$$

where χ here denotes the cut-off function of the transition region $A(\epsilon R^4, 2\epsilon R^4)$. We then extend these operators canonically to operators over the whole of $\mathbb{R}^2$, as in Section 4.1. By definition,

$$\hat{J}_0 := \hat{J}_{p,0} + D + E. \tag{4-14}$$

We call D and E respectively the *regular component* and the *singular component* of the difference. We now show that the coefficients of the regular component tend to 0 in all norms that concern us as Λ tends to infinity. We will study the singular component in the next section.

By (4-6), (4-7), (4-9), (4-10) and (4-13),

$$Df = a^{ij} f_{ij} + b^i f_i,$$

where, over $A(\epsilon R, 2\epsilon R^4)$,

$$\begin{aligned} a^{ij} &= -\frac{c\epsilon}{r^2} x^i x^j - \frac{c^2\epsilon^2}{r^4} x^i x^j + \mathrm{O}\biggl(\biggl[1+\log\Bigl(\frac{r}{\epsilon R}\Bigr)\biggr]\frac{1}{r^k}\Bigl(r+\frac{\epsilon}{r}\Bigr)^4\biggr), \\ b^i &= (1-\chi)\frac{2c^2\epsilon^2}{r^4} x^i + \mathrm{O}\biggl(\biggl[1+\log\Bigl(\frac{r}{\epsilon R}\Bigr)\biggr]\frac{1}{r^{k+1}}\Bigl(r+\frac{\epsilon}{r}\Bigr)^4\biggr). \end{aligned} \tag{4-15}$$

Lemma 4.2.1. *For sufficiently small* α,

$$\|a|_{B(\epsilon R)}\|_{C^{0,\alpha}},\ \|b|_{B(\epsilon R)}\|_{C^{0,\alpha}} \to 0, \tag{4-16}$$

as Λ tends to infinity.

Proof. Indeed, by (4-15), since χ equals 1 near $C(\epsilon R)$, over this circle,

$$\begin{aligned} a &= \mathrm{O}\biggl(\frac{1}{(\epsilon R)^k}\Bigl(\epsilon + \frac{1}{R^2} + (\epsilon R)^4 + \frac{1}{R^4}\Bigr)\biggr), \\ b &= \mathrm{O}\biggl(\frac{1}{(\epsilon R)^{k+1}}\Bigl((\epsilon R)^4 + \frac{1}{R^4}\Bigr)\biggr). \end{aligned}$$

Since the Lipschitz seminorms of the canonical extensions of a and b over $B(\epsilon R)$ are controlled by their Lipschitz seminorms over $C(\epsilon R)$, by (A-10), for all $\alpha \in [0, 1]$,

$$\begin{aligned} \|a|_{B(\epsilon R)}\|_{C^{0,\alpha}} &\lesssim \frac{\epsilon^{1-\alpha}}{R^\alpha} + \frac{1}{\epsilon^\alpha R^{2+\alpha}} + (\epsilon R)^{4-\alpha} + \frac{1}{\epsilon^\alpha R^{4+\alpha}}, \\ \|b|_{B(\epsilon R)}\|_{C^{0,\alpha}} &\lesssim (\epsilon R)^{3-\alpha} + \frac{1}{\epsilon^{1+\alpha} R^{5+\alpha}}. \end{aligned}$$

By (2-14), for sufficiently small α, these both tend to 0 as Λ tends to infinity, as desired. □

Lemma 4.2.2. *For sufficiently small* α,

$$\|a|_{A(\epsilon R, 2\epsilon R^4)}\|_{C^{0,\alpha}},\ \ \|b|_{A(\epsilon R, 2\epsilon R^4)}\|_{C^{0,\alpha}} \to 0, \tag{4-17}$$

as Λ tends to infinity.

Proof. Indeed, by (4-15), over $A(\epsilon R, 2\epsilon R^4)$,

$$a = \mathrm{O}\biggl(\frac{1}{r^k}\Bigl(\epsilon + \frac{\epsilon^2}{r^2}\Bigr)\biggr) + \mathrm{O}\biggl(\biggl[1+\log\Bigl(\frac{r}{\epsilon R}\Bigr)\biggr]\frac{1}{r^k}\Bigl(r+\frac{\epsilon}{r}\Bigr)^4\biggr),$$

and $b = b_1 + b_2$, where

$$\begin{aligned} b_1 &= \mathrm{O}\biggl(\biggl[1+\log\Bigl(\frac{r}{\epsilon R}\Bigr)\biggr]\frac{1}{r^{k+1}}\Bigl(r^4 + \frac{\epsilon^4}{r^4}\Bigr)\biggr), \\ b_2 &= (1-\chi)\frac{2c^2\epsilon^2}{r^4} x^i. \end{aligned}$$

Thus, by (A-10) and (A-20), for all $\alpha \in [0, 1]$,

$$\|a|_{A(\epsilon R, 2\epsilon R^4)}\|_{C^{0,\alpha}} \lesssim \frac{\epsilon^{1-\alpha}}{R^\alpha} + \frac{1}{\epsilon^\alpha R^{2+\alpha}} + \log(R)(\epsilon R^4)^{4-\alpha} + \frac{1}{\epsilon^\alpha R^{4+\alpha}},$$

$$\|b_1|_{A(\epsilon R, 2\epsilon R^4)}\|_{C^{0,\alpha}} \lesssim \log(R)(\epsilon R^4)^{3-\alpha} + \frac{1}{\epsilon^{1+\alpha} R^{5+\alpha}}.$$

By (2-14), for sufficiently small α, these both tend to 0 as Λ tends to infinity. Finally, over $A(\epsilon R^4, 2\epsilon R^4)$,

$$b_2 = O(\epsilon^2 r^{-(k+3)}),$$

so that, by (A-10),

$$\|b_2|_{A(\epsilon R^4, 2\epsilon R^4)}\|_{C^{0,\alpha}} \lesssim \frac{1}{\epsilon^{1+\alpha} R^{12+4\alpha}}.$$

By (2-14), for sufficiently small α, this also tends to 0 as Λ tends to infinity, and the result follows. □

Lemma 4.2.3. *If $\epsilon R < s < t < \sqrt{2}$, then*

$$|v(t) - v_p(t)| \leq |v(s) - v_p(s)|. \tag{4-18}$$

Proof. Indeed, by (2-5), using a dot to denote differentiation with respect to r, we have

$$r(\dot{v} - \dot{v}_p) = -(v - v_p)(1 - r(v + v_p) + (v^2 + v v_p + v_p^2)).$$

However,

$$1 - r(v + v_p) + (v^2 + v v_p + v_p^2) \geq 1 - \frac{r^2}{2}.$$

Thus, for $r \leq \sqrt{2}$, $|v - v_p|$ is decreasing, as desired. □

Lemma 4.2.4. *For all $\alpha \in]0, 1]$,*

$$\|a|_{A(\epsilon R^4, 1)}\|_{C^1}, \|b|_{A(\epsilon R^4, 1)}\|_{C^1} \to 0, \tag{4-19}$$

as Λ tends to infinity.

Proof. By (4-1) and (4-8), over $C(2\epsilon R^4)$,

$$|v - v_p| \lesssim \frac{1}{R^4} + \log(R)(\epsilon R^4)^3 + \log(R)\frac{1}{R^{12}}.$$

By Lemma 4.2.3, this inequality continues to hold over the whole of $A(2\epsilon R^4, 1)$. Since v and v_p both solve (2-5), it follows that, over this annulus,

$$v - v_p = \mathrm{O}\left(\frac{1}{(\epsilon R^4)^k}\left(\frac{1}{R^4} + \log(R)(\epsilon R^4)^3 + \log(R)\frac{1}{R^{12}}\right)\right).$$

Thus,

$$\|(v - v_p)|_{[2\epsilon R^4, 1]}\|_{C^2} \lesssim \frac{1}{\epsilon^2 R^{12}} + \log(R)\epsilon R^4 + \log(R)\frac{1}{\epsilon^2 R^{20}},$$

so that, by (2-14),

$$\|(v - v_p)|_{[2\epsilon R^4, 1]}\|_{C^2} \to 0,$$

as Λ tends to infinity. However, by (4-5), over $A(\epsilon R^4, 1)$, the coefficients a and b only depend on the first derivatives of v and v_p, so that

$$\|a|_{A(2\epsilon R^4,1)}\|_{C^1},\ \ \|b|_{A(2\epsilon R^4,1)}\|_{C^1} \to 0,$$

as Λ tends to infinity, as desired. □

Lemma 4.2.5. *For all* $R_0 > 1$,

$$\|a|_{A(1,R_0)}\|_{C^1}, \|b|_{A(1,R_0)}\|_{C^1} \to 0, \tag{4-20}$$

as Λ *tends to infinity.*

Proof. By (4-1), (4-8) and (4-18), over $C(1)$,

$$|v - v_p| \lesssim \frac{1}{R^4} + \log(R)(\epsilon R^4)^3 + \log(R)\frac{1}{R^{12}}.$$

Since solutions of first-order ODEs vary smoothly with their parameters,

$$\|(v - v_p)|_{[1,R_0]}\|_{C^2} \to 0,$$

as Λ tends to ∞. However, over $A(1, R_0)$, a and b only depend on v and v_p and their derivatives up to order 2, and the result follows. □

Lemma 4.2.6. *For all* $\epsilon > 0$, *there exists* $R_0 > 0$ *such that if* $|v(1) - v_p(1)| \leq 1$, *then*

$$\|a|_{A(R_0,\infty)}\|_{C^1(G)}, \|b|_{A(R_0,\infty)}\|_{C^1(G)} \leq \epsilon. \tag{4-21}$$

Proof. Indeed, over $A(4, \infty)$, both $\hat{J}_0$ and $\hat{J}_{p,0}$ are given by (3-8). The result now follows by local uniform dependence of the estimates in (3-9) on the initial value. □

Combining these results yields:

Lemma 4.2.7. (1) *The operator norm of* D, *considered as a map from* $H^2(G)$ *into* $L^2(G)$ *converges to* 0 *as* Λ *tends to infinity.*

(2) *For sufficiently small* α, *the operator norm of* D, *considered as a map from* $C^{2,\alpha}(G)$ *into* $C^{0,\alpha}(G)$ *converges to* 0 *as* Λ *tends to infinity.*

Proof. Indeed, by (4-16), (4-17), (4-19), (4-21) and (4-20), for sufficiently small α, both $\|a\|_{C^{0,\alpha}(G)}$ and $\|b\|_{C^{0,\alpha}(G)}$ converge to 0 as Λ tends to infinity, and the result follows. □

4.3. ***The singular component.*** We now write

$$Ef =: a^i f_i. \tag{4-22}$$

Since E is defined by canonical extension, over the ball $B(\epsilon R)$,

$$a^i = \frac{2c^2}{\epsilon^2 R^4} x^i. \tag{4-23}$$

At this stage we require the following key estimate, which reveals the significance of the hybrid norm.

Lemma 4.3.1. *For sufficiently small α and for sufficiently small γ,*

$$\|f\|_{C^{1,\alpha}_\gamma(G)} \lesssim (\epsilon R)^{1-2\alpha}\|f\|_{2,\alpha,\gamma}. \tag{4-24}$$

Remark. It will be useful to observe that this relation is also valid for spaces of functions defined over an unbounded annulus.

Proof. Indeed, by the Sobolev embedding theorem, for all $\beta < 1$,

$$\|f\|_{C^{0,\beta}_\gamma(G)} \lesssim \|f\|_{H^2_\gamma(G)} \lesssim (\epsilon R)\|f\|_{2,\alpha,\gamma}.$$

Setting $\beta = (1-\alpha)$ and using (A-10) and (A-11), we obtain

$$\|f\|_{C^{1,\alpha}_\gamma(G)} \lesssim (\epsilon R)^{1/(1+2\alpha)}\|f\|_{2,\alpha,\gamma} \lesssim (\epsilon R)^{1-2\alpha}\|f\|_{2,\alpha,\gamma},$$

as desired. □

Lemma 4.3.2. *For sufficiently small $\alpha \in [0,1]$ and for sufficiently small γ, the operator norm of E, considered as a map from $\mathcal{L}^{2,\alpha}_\gamma(G)$ into $C^{0,\alpha}_\gamma(G)$ tends to 0 as Λ tends to infinity.*

Proof. Indeed, over $A(\epsilon R, 2\epsilon R^4)$,

$$a^i = \mathrm{O}\left(\frac{\epsilon^2}{r^{3+k}}\right),$$

so that

$$\|a^i|_{A(\epsilon R,2\epsilon R^4)}\|_{C^0} \lesssim \frac{1}{\epsilon R^3} \quad \text{and} \quad [a^i|_{A(\epsilon R,2\epsilon R^4)}]_1 \lesssim \frac{1}{\epsilon^2 R^4}.$$

Since a^i is extended canonically over $B(\epsilon R)$, these inequalities also hold over the whole of $B(2\epsilon R^4)$ so that, by (A-10), for all $\alpha \in [0,1]$,

$$[a^i]_\alpha \lesssim \frac{1}{(\epsilon R)^\alpha \epsilon R^3}.$$

It follows by (4-24) and (A-12) that

$$\|Ef\|_{C^{0,\alpha}_\gamma(G)} \lesssim \frac{1}{(\epsilon R)^\alpha \epsilon R^3}\|f\|_{C^{1,\alpha}_\gamma(G)},$$

and the result follows by (2-14) and (4-24). □

Lemma 4.3.3. *For sufficiently small $\alpha \in [0,1]$ and for sufficiently small γ, the operator norm of $(\epsilon R)^{-1}E$ considered as a map from $\mathcal{L}^{2,\alpha}_\gamma(G)$ into $H^0_\gamma(G)$ tends to 0 as Λ tends to infinity.*

Proof. Indeed, a direct calculation yields

$$\|a^i\|_{L^2_\gamma(G)} \lesssim \frac{1}{R^2}.$$

Thus, bearing in mind (4-24),

$$\begin{aligned}\|(\epsilon R)^{-1}Ef\|_{L^2_\gamma(G)} &\lesssim (\epsilon R)^{-1}\|a^i\|_{L^2_\gamma(G)}\|Df\|_{L^\infty(G)} \\ &\lesssim (\epsilon R)^{-1}\|a^i\|_{L^2_\gamma(G)}\|f\|_{C^{1,\alpha}_\gamma(G)} \lesssim \frac{1}{(\epsilon R)^{2\alpha}R^2}\|f\|_{2,\alpha,\gamma},\end{aligned}$$

and the result follows by (2-14). □

Combining these results yields:

Lemma 4.3.4. *For sufficiently small $\alpha \in [0, 1]$ and for sufficiently small γ, the operator norm of E considered as a map from $\mathcal{L}^{2,\alpha}_\gamma(G)$ into $\mathcal{L}^{0,\alpha}_\gamma(G)$ tends to 0 as Λ tends to infinity.*

5. Surgery and the perturbation family

5.1. ***The basic surgery operation.*** Recall that our strategy for proving Theorem A consists of two stages. The first involves a surgery operation in which approximate MCF solitons are constructed out of properly embedded minimal surfaces and rotationally symmetric Grim ends. The second involves a fixed-point argument in which these approximate MCF solitons are perturbed into actual MCF solitons. In this section, we describe the surgery operation and in Section 5.2, we describe the family of deformations of the approximate MCF soliton in which the actual MCF soliton will be found. Though conceptually simple, our construction is inevitably rather technical. However, we believe that a careful reading of the following two sections will be rewarded by a clear understanding of the essence of this paper.

Consider first a properly embedded surface C in $\mathbb{R}^3$, minimal outside of some compact set, and with finitely many ends, all of which are horizontal. Let $R_0 > 0$ be such that every component of $C \cap (A(R_0, \infty) \times \mathbb{R})$ is a minimal graph over $A(R_0, \infty)$. Let $F : A(R_0, \infty) \to \mathbb{R}$ be the profile of one of these minimal ends. In Appendix B, we show how the Weierstrass representation yields

$$F = a + c\log(r) + \mathrm{O}(r^{-(1+k)})$$

for some real constants a and c, which will henceforth be referred to respectively as the *constant term* and the *logarithmic parameter* of the minimal end. In particular, planar ends are simply catenoidal ends with vanishing logarithmic parameters. We will only be concerned with minimal ends invariant under reflection in at least two distinct vertical planes. In this case, the above asymptotic series contains no terms of order (-1), so that

$$F = a + c\log(r) + \mathrm{O}(r^{-(2+k)}). \tag{5-1}$$

This asymptotic formula will be used repeatedly throughout the sequel.

Let Λ be a large, positive number, let $K > 0$ be a fixed constant, and choose ϵ, $R > 0$ and $|c| < K$ as in (2-14). Let $G : A(R/4, \infty) \to \mathbb{R}$ be the profile of a rotationally symmetric Grim end with constant term a, logarithmic parameter c and speed ϵ. Rescaling and integrating (2-16) we obtain, over the annulus $A(R/4, 2R^4)$,

$$G = a + c\log(r) + \frac{1}{4}\epsilon r^2 + \mathrm{O}\Big(\Big[1 + \log\Big(\frac{r}{R}\Big)\Big] r^{1-k}\Big(\epsilon r + \frac{1}{r}\Big)^3\Big). \tag{5-2}$$

Let χ_c be the cut-off function of the central transition region $A(R, 2R)$, as defined in Section A1, and define the function H over $A(R_0, \infty)$ by

$$H := \chi_c F + (1 - \chi_c)G. \tag{5-3}$$

Its graph will be called the *joined end*. Observe that H is entirely determined by F and the parameters ϵ and R. Furthermore, over the annuli $A(R_0, R)$ and $A(2R, \infty)$, H simply coincides with F and G

respectively whilst, over the annulus $A(R, 2R)$, by (5-1), (5-2) and the fact that $\chi_c = O(r^{-k})$,

$$H = a + c\log(r) + \tfrac{1}{4}\epsilon(1-\chi_c)r^2 + O(r^{-(2+k)}). \tag{5-4}$$

5.2. ***The deformation family.*** Continuing to use the notation of Section 5.1, let S denote the surface obtained by replacing each of the ends of C with their respective joined ends. We now construct a family of deformations of S out of which the actual MCF soliton will be selected when Λ is large. We first describe how the logarithmic parameters of C and S are varied. Let n denote the number of ends of C, and, for each $1 \leq i \leq n$, let $a_{0,i}$ and $c_{0,i}$ denote respectively the constant term and the logarithmic parameter of the i-th end. Let U be a neighbourhood of $(c_{0,1}, \dots, c_{0,n})$ in $\mathbb{R}^n$ and let $(C_c)_{c\in U}$ be a smoothly varying family of immersed surfaces in $\mathbb{R}^3$ such that $C_{c_0} = C$ and, for all $c \in U$ and for all $1 \leq i \leq n$, the i-th component of $C_c \cap (A(R_0, \infty) \times \mathbb{R})$ is a horizontal, minimal end with constant term $a_{0,i}$ and logarithmic parameter c_i. Finally, for all $c \in U$, let S_c denote the surface obtained by replacing each end of C_c with its corresponding joined end, as described in Section 5.1.

Let $E : U \times S \to \mathbb{R}^3$ be a smooth function such that

(1) for all $c \in U$, $E_c := E(c, \cdot)$ parametrises S_c, and

(2) for all $c \in U$, and for all $p \in S \cap (A(R_0, +\infty) \times \mathbb{R})$, the point $E_c(p)$ lies vertically above or below the point p.

Let χ_0, χ_0', χ_ϵ' and χ_ϵ be the cut-off functions of the transition regions $A(R_0, 2R_0)$, $A(2R_0, 4R_0)$, $A(1/(2\epsilon), 1/\epsilon)$ and $A(1/\epsilon, 2/\epsilon)$ respectively, as defined in Section A1. By composing with vertical projections onto $\mathbb{R}^2$, we think of these functions also as functions defined over S. For all $c \in U$, let N_c denote the unit normal vector field over S_c. For all $1 \leq i \leq n$, let $\mathbb{1}_i : S \to \{0, 1\}$ denote the indicator function of the i-th component of $S_c \cap (A(R_0, \infty) \times \mathbb{R})$. Observe that, since this intersection is a union of graphs, every component is transverse to the unit vertical vector e_z. For all $1 \leq i \leq n$, let $\epsilon_i \in \{\pm 1\}$ be such that $\epsilon_i e_z$ lies on the same side of the i-th component as N_c. For all $c \in U$, define the *modified normal vector field* over S_c by

$$\widehat{N}_c := (\chi_\epsilon - \chi_0)\epsilon_i e_z + (1 - (\chi_\epsilon - \chi_0))N_c. \tag{5-5}$$

Observe that, over the regions $S_c \cap (B(R_0) \times \mathbb{R})$ and $S_c \cap (A(2/\epsilon, \infty) \times \mathbb{R})$, this vector field coincides with N_c whilst, over the region $S_c \cap (A(2R_0, 1/\epsilon) \times \mathbb{R})$, it coincides with $\pm e_z$. Now let V and W be neighbourhoods of 0 in $\mathbb{R}^n$ and define $\widetilde{E} : U \times V \times W \times C^\infty(S) \to C^\infty(S, \mathbb{R}^3)$ by

$$\widetilde{E}_{c,a,b,f}(p) := E_c(p) + f(p)\widehat{N}_c(p) + \sum_{i=1}^{n} \epsilon_i \mathbb{1}_i(p)\big(a_i(1 - \chi_0'(p)) + b_i(1 - \chi_\epsilon'(p))\big)e_z. \tag{5-6}$$

Upon reducing U, V and W if necessary, there trivially exists $\delta > 0$, which is independent of Λ, ϵ and R, such that, for all $(c, a, b) \in U \times V \times W$, and, for all $\|f\|_{C^0} < \delta$, the function $\widetilde{E}_{c,a,b,f}$ defines an immersion of S into $\mathbb{R}^3$. This concludes the description of the deformation family in which the actual MCF soliton will be found.

5.3. ***Microscopic and macroscopic perturbations.*** Continuing to use the notation of Sections 5.1 and 5.2, we consider now the first-order perturbations of S defined by the above deformation family. We classify these perturbations into two main types. Those in the direction of $C^\infty(S)$ will be called *microscopic perturbations*, and those in the directions of U, V and W will be called *macroscopic perturbations*. We now describe the first-order variations of the MCFS functional resulting from macroscopic perturbations. The first-order variations resulting from microscopic perturbations will be studied in the next section.

Recall that, as in Section A2, the MCFS functional with speed ϵ of an immersion $E : S \to \mathbb{R}^3$ is given by

$$M_E := H_E + \epsilon\langle N_E, e_z\rangle, \tag{5-7}$$

where H_E here denotes the mean curvature function of E, and N_E here denotes its unit normal vector field. We define $M_\epsilon : U \times V \times W \to C_0^\infty(S)$ such that, for all $(c, a, b) \in U \times V \times W$, and, for all $p \in S$, $M_{\epsilon,c,a,b}(p)$ is the value of this functional for the immersion $E_{c,a,b}$ at the point p. We define the operators $X_\epsilon, Y_\epsilon, Z_\epsilon : \mathbb{R}^n \to C_0^\infty(S)$ by

$$\begin{aligned}(X_\epsilon u)(p) &:= \frac{1}{\langle \widehat{N}_S, N_S\rangle}\frac{d}{dt}M_{\epsilon,c_0+tu,0,0}(p)\bigg|_{t=0},\\ (Y_\epsilon v)(p) &:= \frac{1}{\langle \widehat{N}_S, N_S\rangle}\frac{d}{dt}M_{\epsilon,c_0,tv,0}(p)\bigg|_{t=0},\\ (Z_\epsilon w)(p) &:= \frac{1}{\langle \widehat{N}_S, N_S\rangle}\frac{d}{dt}M_{\epsilon,c_0,0,tw}(p)\bigg|_{t=0}.\end{aligned} \tag{5-8}$$

These are the first-order variations of the MCFS functional arising from the three types of macroscopic perturbation. In particular, since $M_{\epsilon,c,0,0}$ vanishes over $S\cap(A(2R, +\infty) \times \mathbb{R})$ for all $c \in V$, for all $u \in \mathbb{R}^d$, Xu is supported over $S\cap(B(2R) \times \mathbb{R})$. Likewise, for all $v, w \in \mathbb{R}^n$, Yv and Zw are supported over $S\cap(A(2R_0, 4R_0) \times \mathbb{R})$ and $S\cap(A(1/(2\epsilon), 1/\epsilon) \times \mathbb{R})$ respectively. In later sections, when no ambiguity arises, the subscript ϵ will be suppressed, and these operators will be denoted simply by X, Y and Z respectively.

5.4. ***Modified Jacobi operators.*** The operator of first-order variation of the MCFS functional resulting from microscopic perturbations is none other than the modified MCFS Jacobi operator. In this section, we determine asymptotic formulae for its coefficients over different regions. We recall that, since different modifications are made on different scales, the precise definition of the modified MCFS Jacobi operator varies with context. We now describe the framework which unifies these different definitions. We will then study three different cases corresponding to, in order, CHM surfaces, rotationally symmetric Grim ends, and joined surfaces.

Consider first a general immersed surface Σ in $\mathbb{R}^3$ such that, for some $R_0 > 0$, every component of $\Sigma\cap(A(R_0, \infty) \times \mathbb{R})$ is a graph over $A(R_0, \infty)$. Let $\Lambda > 0$ be a large, positive number, let $\epsilon, R > 0$ be as in (2-14), and let $\widehat{N}_\Sigma$ be the modified normal vector field over Σ as defined in (5-5). We define $E : C_0^\infty(\Sigma) \to C^\infty(\Sigma, \mathbb{R}^3)$ by

$$E_f(p) := p + f(p)\widehat{N}_\Sigma(p).$$

Observe that if f is sufficiently small, then E_f is an immersion. Define $M : C_0^\infty(\Sigma) \to C^\infty(\Sigma)$ such that, for all such f, and for all $p \in \Sigma$, $M_f(p)$ is the value of the MCFS functional (5-7) with speed ϵ for the immersion E_f at the point p. The *modified MCFS Jacobi operator* of Σ with speed ϵ is now defined by

$$(\hat{J}_{\Sigma,\epsilon} f)(p) := \frac{1}{\langle \widehat{N}_\Sigma, N_\Sigma\rangle} \frac{d}{dt} M_{tf}(p)\bigg|_{t=0}. \tag{5-9}$$

In later sections, when no ambiguity arises, the subscript ϵ will be suppressed, and this operator will be denoted simply by $\hat{J}_\Sigma$.

Over the annulus $A(R/4, 1/\epsilon)$, since $\widehat{N}_\Sigma$ here coincides with e_z, the operator $\hat{J}_{\Sigma,\epsilon}$ is simply $\langle N_\Sigma, e_z\rangle^{-1}$ times the linearisation of the MCFS functional for graphs. Consequently, if $F : A(R/4, 1/\epsilon) \to \mathbb{R}$ is the profile of a component of $\Sigma \cap (A(R/4, 1/\epsilon) \times \mathbb{R})$ then, upon differentiating (A-7) we obtain, over this annulus,

$$\hat{J}_{\Sigma,\epsilon} f = g^{ij} f_{ij} - \mu^2 g^{ij} F_{ij} F_k f_k + 2\mu^4 F_i F_j F_k F_{ij} f_k - 2\mu^2 F_{ij} F_i f_j - \epsilon\mu^2 F_i f_i. \tag{5-10}$$

In particular, for all $v, w \in \mathbb{R}^n$, and for all $p \in S$,

$$\begin{aligned}(Yv)(p) &= -\sum_{i=1}^n \mathbb{I}_i(p) v_i (\hat{J}_{\Sigma,\epsilon}\chi_0')(p),\\ (Zw)(p) &= -\sum_{i=1}^n \mathbb{I}_i(p) w_i (\hat{J}_{\Sigma,\epsilon}\chi_\epsilon')(p).\end{aligned} \tag{5-11}$$

Now let C be a minimal end over the annulus $A(R_0, \infty)$ satisfying (5-1) and let $\hat{J}_{C,\epsilon}$ be its modified MCFS Jacobi operator with speed ϵ.

Lemma 5.4.1. *Over* $A(R/4, 2R^4)$,

$$\hat{J}_{C,\epsilon} f = \Delta f - \frac{c^2}{r^4} x^i x^j f_{ij} - \frac{\epsilon c}{r^2} x^i f_i + \frac{2c^2}{r^4} x^i f_i + \mathcal{E}_{C,\epsilon} f, \tag{5-12}$$

where $\mathcal{E}_{C,\epsilon} f := a^{ij} f_{ij} + b^i f_i$ *and* a *and* b *satisfy*

$$a = \mathrm{O}(r^{-(k+4)}), \quad b = \mathrm{O}\Big(r^{-(k+4)}\Big(\epsilon r + \frac{1}{r}\Big)\Big). \tag{5-13}$$

Proof. By (5-1),

$$F_i = \frac{c}{r^2} x^i + \mathrm{O}(r^{-(k+3)}).$$

Thus, by (A-6),

$$\begin{aligned}\mu^2 &= 1 - \frac{c^2}{r^2} + \mathrm{O}(r^{-(k+4)}),\\ g^{ij} &= \delta_{ij} - \frac{c^2}{r^4} x^i x^j + \mathrm{O}(r^{-(k+4)}).\end{aligned}$$

Therefore,

$$g^{ij} f_{ij} = \Delta f - \frac{c^2}{r^4} x^i x^j f_{ij} + a^{ij} f_{ij},$$

where $a = \mathrm{O}(r^{-(k+4)})$. Likewise,

$$\begin{aligned}\mu^2 g^{ij} F_{ij} F_k f_k &= b_1^i f_i,\\ 2\mu^4 F_i F_j F_k F_{ij} f_k &= b_2^i f_i,\\ -2\mu^2 F_{ij} F_i f_j &= \frac{2c^2}{r^4} x^i f_i + b_3^i f_i,\end{aligned}$$

where $b_1^i, b_2^i, b_3^i = \mathrm{O}(r^{-(k+5)})$. Finally,

$$\epsilon\mu^2 F_i f_i = \frac{\epsilon c}{r^2} x^i f_i + b_4^i,$$

where $b_4^i = \mathrm{O}(\epsilon r^{-(k+3)})$. The result follows. □

Next let G be a rotationally symmetric Grim end of speed ϵ over the annulus $A(R/4, \infty)$ and let $\hat{J}_{G,\epsilon}$ be its modified MCFS Jacobi operator with speed ϵ. Define $\psi : G \to \mathbb{R}$ by

$$\psi := \langle \widehat{N}_G, N_G\rangle = \chi_\epsilon \langle e_z, N_G\rangle + (1 - \chi_\epsilon), \tag{5-14}$$

and denote by M_ψ the operator of multiplication by ψ.

Lemma 5.4.2. *Over $A(R/4, \infty)$,*

$$\hat{J}_{G,\epsilon} := M_\psi^{-1} J_{G,\epsilon} M_\psi, \tag{5-15}$$

where $J_{G,\epsilon}$ denotes the MCFS Jacobi operator with speed ϵ of G, as defined in Section A2.

Remark. In particular, in the case of rotationally symmetric Grim ends, the modified MCFS Jacobi operator as defined above coincides, up to rescaling, with the modified MCFS Jacobi operator as defined in Section 4.1.

Proof. Indeed, more generally, with $M := M_0$ defined as at the beginning of this section, for all $f \in C_0^\infty(\Sigma)$,

$$\hat{J}_{\Sigma,\epsilon} f = M_\psi^{-1} J_{\Sigma,\epsilon} M_\psi f + M_\psi^{-1} \langle X, \nabla M\rangle f,$$

where X here denotes the tangential component of the vector field $\widehat{N}_\Sigma$. The result now follows since M vanishes identically over G. □

In particular, rescaling (4-6) and (4-7) immediately yields:

Lemma 5.4.3. *Over $A(R/4, \infty)$,*

$$\hat{J}_{G,\epsilon} f = \Delta f - \left(\frac{\epsilon}{2} + \frac{c}{r^2}\right)^2 x^i x^j f_{ij} - \left(\frac{\epsilon^2}{2} - \frac{2c^2}{r^4}\right) x^i f_i + \mathcal{E}_G f. \tag{5-16}$$

where $\mathcal{E}_{G,\epsilon} f := a^{ij} f_{ij} + b^i f_i$, and a and b satisfy

$$\begin{aligned} a &= \mathrm{O}\left(\left[1 + \log\left(\frac{r}{R}\right)\right] \frac{1}{r^k} \left(\epsilon r + \frac{1}{r}\right)^4\right),\\ b &= \mathrm{O}\left(\left[1 + \log\left(\frac{r}{R}\right)\right] \frac{1}{r^{k+1}} \left(\epsilon r + \frac{1}{r}\right)^4\right).\end{aligned} \tag{5-17}$$

Finally, let S be a joined end, as constructed in Section 5.1, and let $\hat{J}_{S,\epsilon}$ denote its modified MCFS Jacobi operator with speed ϵ.

Lemma 5.4.4. *Over $A(R,2R)$,*

$$(\hat{J}_{S,\epsilon}-\hat{J}_{C,\epsilon})f=a_1^{ij}f_{ij}+b_1^i f_i,$$
$$(\hat{J}_{S,\epsilon}-\hat{J}_{G,\epsilon})f=a_2^{ij}f_{ij}+b_2^i f_i,$$

where a_1, a_2, b_1 and b_2 satisfy

$$a_1,a_2=\mathrm{O}(r^{-(4+k)}),\qquad b_1,b_2=\mathrm{O}(r^{-(5+k)}). \tag{5-18}$$

Proof. By (5-1), (5-4) and (2-14), over $A(R,2R)$,

$$H_i-F_i=\mathrm{O}(r^{-(3+k)}),$$
$$F_i,H_i=\mathrm{O}(r^{-(1+k)}).$$

Thus, by (A-6),

$$\mu_H-\mu_F=\mathrm{O}(r^{-(4+k)}),$$
$$g_H^{ij}-g_F^{ij}=\mathrm{O}(r^{-(4+k)}).$$

The result follows for $(\hat{J}_{S,\epsilon}-\hat{J}_{C,\epsilon})$ by (5-10). The result for $(\hat{J}_{S,\epsilon}-\hat{J}_{G,\epsilon})$ follows in a similar manner, and this completes the proof. □

We conclude this section by studying commutators of modified Jacobi operators with certain multiplication operators. Indeed, let $[\hat{J}_{C,\epsilon},\chi_l]$ denote the commutator of $\hat{J}_{C,\epsilon}$ with the operator of multiplication by the cut-off function χ_l of the lower transition region $A(R/4,R/2)$. Likewise, let $[\hat{J}_{G,\epsilon},\chi_u]$ denote the commutator of $\hat{J}_{G,\epsilon}$ with the operator of multiplication by the cut-off function χ_u of the upper transition region $A(R^4,2R^4)$. Observe that these operators are supported over the annuli $A(R/4,R/2)$ and $A(R^4,2R^4)$ respectively.

Lemma 5.4.5. $$[\hat{J}_{C,\epsilon},\chi_l]f=a_1^i f_i+b_1 f,\qquad [\hat{J}_{G,\epsilon},\chi_u]f=a_2^i f_i+b_2 f,$$

where a_1, a_2, b_1 and b_2 satisfy,

$$a_1,a_2=\mathrm{O}(r^{-(k+1)}),\qquad b_1,b_2=\mathrm{O}(r^{-(k+2)}). \tag{5-19}$$

Proof. Indeed, since $\chi_l,\chi_u=\mathrm{O}(r^{-k})$, the result follows by (5-12), (5-13), (5-16) and (5-17). □

5.5. *Controlling macroscopic perturbations.* We conclude this section by studying the first-order variation of the MCFS functional resulting from the first macroscopic perturbation. Recall that, for all $u\in\mathbb{R}^n$, Xu vanishes outside $B(2R)$. Inside this ball, we have:

Lemma 5.5.1. *For $u\in\mathbb{R}^d$ such that $\|u\|=1$, over $A(2R_0,R)$,*

$$Xu=\mathrm{O}(\epsilon r^{-(2+k)}), \tag{5-20}$$

and over $A(R,2R)$,

$$Xu=\mathrm{O}(r^{-(4+k)}). \tag{5-21}$$

Proof. For notational convenience, we suppose that C and S each only have one end and, in particular, that $u=1$. Let C_c and S_c be smooth families of immersed surfaces as in Section 5.1. For all t, let $F_t : A(2R_0,\infty)\to\mathbb{R}$ and $H_t : A(2R_0,\infty)\to\mathbb{R}$ denote the profiles of $C_{c_0+t}\cap(A(2R_0,\infty)\times\mathbb{R})$ and $S_{c_0+t}\cap(A(2R_0,\infty)\times\mathbb{R})$ respectively. Define

$$Z := \frac{d}{dt}F_t|_{t=0}, \qquad W := \frac{d}{dt}H_t|_{t=0},$$

and observe that, over $A(2R_0, 2R)$,

$$Xu = \hat{J}_{S,\epsilon}W.$$

Now, by (5-1),

$$Z = \log(r) + \mathrm{O}(r^{-(2+k)}).$$

Next, by (2-41) and (5-3), and bearing in mind that $\chi_c = \mathrm{O}(r^{-k})$, over $A(R, 2R)$, we have

$$W = \log(r) + \mathrm{O}(r^{-(2+k)}) = Z + \mathrm{O}(r^{-(2+k)}), \tag{5-22}$$

and since $Z = W$ over $A(2R_0, R)$, (5-22) in fact holds over the whole of $A(2R_0, 2R)$. We now write

$$Xu = \hat{J}_{C,\epsilon}Z + (\hat{J}_{S,\epsilon} - \hat{J}_{C,\epsilon})Z + \hat{J}_{S,\epsilon}(W - Z).$$

The second and third terms are supported over $A(R, 2R)$, and by (5-12) and (5-18),

$$(\hat{J}_{S,\epsilon} - \hat{J}_{C,\epsilon})Z = \mathrm{O}(r^{-(6+k)}),$$
$$\hat{J}_{S,\epsilon}(W - Z) = \mathrm{O}(r^{-(4+k)}).$$

Finally, since the graph of F_t is minimal for all t, by (A-6) and (A-7),

$$\hat{J}_{C,\epsilon}Z = -\epsilon\mu^2 F_{0,i}Z_i = \mathrm{O}(\epsilon r^{-(2+k)}),$$

and the result follows by (2-14). □

6. Constructing the Green's operator

6.1. ***The cylindrical, Grim and hybrid norms.*** We now prepare the ground for the perturbation argument that will be used to construct actual MCF solitons out of the approximate MCF solitons constructed in Section 5.1. In this section, we construct the Green's operator of the modified MCFS Jacobi operator of the approximate MCF soliton together with estimates of its operator norm. It is the determination of suitable estimates, requiring a careful and lengthy analysis, which constitutes the hardest part of this paper. We will see presently that sufficiently strong estimates are made possible by the correct choice of functional norms over the different components of the approximate MCF soliton, as well as the use of the hybrid norm, already mentioned in the Introduction and Section 4. Throughout this section, we will make use of (2-14) without comment.

We first study the analytic properties of Green's operators over CHM surfaces. Thus, for g a positive integer, let $C := C_g$ be the CHM surface of genus g. Observe that functions over $C\cap(A(R_0,\infty)\times\mathbb{R})$ may be considered as functions over three copies of $A(R_0,\infty)$. In defining norms over spaces of functions, we will pass between these two perspectives without comment. Consider now the triplet $(X, Y, \hat{J}_C)$, where X and Y are the operators constructed in Section 5.1 and $\hat{J}_C$ is the modified MCFS Jacobi operator of C as constructed in Section 5.4. We now construct a right inverse for this operator when Λ is large. We first gather various basic results that will be of use to us. Let D denote the total differentiation operator over $\mathbb{R}^2$ and define

$$D_{\mathrm{SF}} := r D, \tag{6-1}$$

where r here denotes the radial distance from the origin. Likewise, for $\alpha \in [0, 1]$ and for $f : \mathbb{R}^2 \to \mathbb{R}$, define

$$\delta^{\alpha}_{\mathrm{SF}} f(r) := r^{\alpha}[f|_{A(r/2,2r)}]_{\alpha}. \tag{6-2}$$

For all nonnegative integer m, for all $\alpha \in [0, 1]$ and for all real δ, define the *scale-free weighted Hölder norm* of any m-times differentiable function $f : A(R_0, \infty) \to \mathbb{R}$ by

$$\|f\|_{C^{m,\alpha}_{\delta,\mathrm{SF}}(A(R_0,\infty))} := \sum_{i=0}^{m} \|r^{\delta} D^{i}_{\mathrm{SF}} f\|_{C^0(A(R_0,\infty))} + \|r^{\delta}\delta^{\alpha}_{\mathrm{SF}} D^{m}_{\mathrm{SF}} f\|_{C^0(]2R_0,\infty[)}. \tag{6-3}$$

For nonnegative, integer m, for all $\alpha \in [0, 1]$, for all real δ and for any m-times differentiable function $f : C \to \mathbb{R}$, define

$$\|f\|_{C^{m,\alpha}_{\delta,\mathrm{SF}}(C)} := \|f|_{C\cap(B(2R_0)\times\mathbb{R})}\|_{C^{m,\alpha}} + \|f|_{C\cap(A(R_0,\infty)\times\mathbb{R})}\|_{C^{m,\alpha}_{\delta,\mathrm{SF}}(A(R_0,\infty))}. \tag{6-4}$$

For all such m, α and δ, let $C^{m,\alpha}_{\delta,\mathrm{SF},g}(C)$ denote the space of m-times differentiable functions f over C which satisfy $\|f\|_{C^{m,\alpha}_{\delta,\mathrm{SF}}(C)} < \infty$ and which also satisfy $f \circ \sigma = f$ for every horizontal symmetry σ of C. Observe in particular that, since each of X and Y has compact support, we may also think of them as taking values in $C^{0,\alpha}_{\delta+2,\mathrm{SF},g}(C)$.

Recall that, with the above symmetries imposed, for all $\delta \in]1, 2[$, and for all $\alpha \in]0, 1[$, the *Jacobi operator* J_C of C defines an injective Fredholm map of Fredholm index (-3) from $C^{2,\alpha}_{\delta,\mathrm{SF},g}(C)$ into $C^{0,\alpha}_{\delta+2,\mathrm{SF},g}(C)$; see [Hauswirth and Pacard 2007; Morabito 2009; Nayatani 1993; Pacard 2008]).[3]

Lemma 6.1.1. *For all $\alpha \in]0, 1[$, for all $\delta \in]1, 2[$, for all $R_0 > 0$ sufficiently large, and for all $\Lambda > 0$ sufficiently large, the triplet $(X, Y, \hat{J}_C)$ defines a surjective Fredholm map from $\mathbb{R}^3 \oplus \mathbb{R}^3 \oplus C^{2,\alpha}_{\delta,\mathrm{SF},g}(C)$ into $C^{0,\alpha}_{2+\delta,\mathrm{SF},g}(C)$ of Fredholm index 3. Furthermore, the right inverse (U, V, Φ) can be chosen in such a manner that its norm is uniformly bounded, independent of Λ.*

Remark. In the sequel, R_0 will be chosen large enough for Lemma 6.1.1 to hold for all large values of Λ. It will then be fixed once and for all, and Λ will be made to tend to $+\infty$.

Proof. For all $c \in U$, where U is a suitable open subset of $\mathbb{R}^3$, let C_c be as in Section 5.1 and suppose in addition that C_c is also invariant under all the horizontal symmetries of C. Let $E : U \times C \to \mathbb{R}^3$ be a smooth function such that

(1) for all $c \in U$, E_c parametrises C_c,

[3] We aim to include an overview of the perturbation theory of the Costa–Hoffman–Meeks surfaces in forthcoming work, as we are not aware of any readily accessible account in the literature.

(2) for all $c \in U$ and for all $p \in C \cap (A(R_0, +\infty) \times \mathbb{R})$, the point $E_c(p)$ lies vertically above or below the point p, and

(3) for all $c \in U$, $E_c := E(c, \cdot)$ is equivariant under all the horizontal symmetries of C.

Let V be a neighbourhood of 0 in $\mathbb{R}^3$ and define $\widetilde{E} : U \times V \times C \to \mathbb{R}^3$ such that, for all $(c, a) \in U \times V$, and, for all $p \in C$,

$$\widetilde{E}_{c,a}(p) = E_c(p) + \sum_{i=1}^{3} \epsilon_i \mathbb{I}_i(p) a_i (1 - \chi_0'(p)) e_z,$$

where $(\epsilon_i)_{1 \leq i \leq 3}$, $(\mathbb{I}_i)_{1 \leq i \leq 3}$ and χ_0' are defined as in Section 5.1. Define $H : U \times V \times C \to \mathbb{R}^3$ such that, for all $(c, a) \in U \times V$, and for all $p \in C$, $H_{c,a}(p)$ is the mean curvature of the immersion $\widetilde{E}_{c,a}$ at the point p. Define the operators $X_0, Y_0 : \mathbb{R}^3 \to C_0^\infty(C)$ by

$$(X_0 u)(p) := \frac{d}{dt} H_{c_0 + tu, 0}(p)|_{t=0},$$
$$(Y_0 v)(p) := \frac{d}{dt} H_{c_0, tv}(p)|_{t=0}.$$

By the perturbation theory of CHM surfaces (see [Hauswirth and Pacard 2007]), (X_0, Y_0, J_C) defines a surjective Fredholm map of Fredholm index 3 from $\mathbb{R}^3 \oplus \mathbb{R}^3 \oplus C^{2,\alpha}_{\delta,\mathrm{SF},g}(C)$ into $C^{0,\alpha}_{\delta+2,\mathrm{SF},g}(C)$.

Let N and $\widehat{N}$ be respectively the unit normal vector field and the modified normal vector field over C. Observe that, as Λ and R_0 tend to $+\infty$, the difference $(\widehat{N} - N)$ tends to 0 in the C^k sense for all k so that the difference $(\hat{J}_C - J_C)$ tends to 0 in the operator norm. Next, it is straightforward to show that, considered as an operator from $\mathbb{R}^3$ into $C^{0,\alpha}_{\delta+2,\mathrm{SF},g}(C)$, $\|Y - Y_0\| \lesssim \epsilon$. Finally, by (2-14), (5-20) and (5-21), considered as another operator between these two spaces, $\|X - X_0\| \lesssim R^{\delta-2}$. Since these both tend to 0 as Λ tends to $+\infty$, the result follows by the stability of surjectivity of Fredholm maps under small perturbations. □

We now review the analytic properties of rotationally symmetric Grim ends. Let G be a rotationally symmetric Grim end of speed ϵ over the annulus $A(R/4, +\infty)$. For all nonnegative, integer m, for all $\alpha \in [0, 1]$, for all $\gamma \in \mathbb{R}$ and for all $\epsilon > 0$, define the following weighted Hölder and Sobolev norms for functions over $\mathbb{R}^2$,

$$\|f\|_{C^{m,\alpha}_{\gamma,\epsilon}(G)} := \|f(\cdot/\epsilon)\|_{C^{m,\alpha}_{\gamma}(G)},$$
$$\|f\|_{H^m_{\gamma,\epsilon}(G)} := \|f(\cdot/\epsilon)\|_{H^m_\gamma(G)}, \tag{6-5}$$

and define the *hybrid norm* by

$$\|f\|_{m,\alpha,\gamma,\epsilon} := \|f\|_{C^{m,\alpha}_{\gamma,\epsilon}(G)} + \frac{1}{\epsilon R}\|f\|_{H^m_{\gamma,\epsilon}(G)}. \tag{6-6}$$

For all such m, α, γ, let $\mathcal{L}^{m,\alpha}_{\gamma,\epsilon,g}(G)$ denote the space of m-times differentiable functions with finite hybrid norm. Let $\hat{J}_G$ denote the modified MCFS Jacobi operator of G, as defined in Sections 4.1 and 5.4. Upon rescaling, Theorem 4.1.1 immediately yields

Lemma 6.1.2. *For all $\alpha \in]0, 1[$, for all sufficiently small γ, and for sufficiently large Λ, the operator $\epsilon^2 \hat{J}_G$ defines a linear isomorphism from $\mathcal{L}^{2,\alpha}_{\gamma,\epsilon,g}(G)$ into $\mathcal{L}^{0,\alpha}_{\gamma,\epsilon,g}(G)$. Furthermore, we may suppose that the operator norm of its inverse is uniformly bounded independent of Λ.*

We conclude this section by describing an alternative form of (6-5), more amenable to calculations. We define operators D_G and δ_G^α by

$$D_G := \frac{1}{\epsilon} D, \tag{6-7}$$

$$\delta_G^\alpha f(x) := \frac{1}{\epsilon^\alpha} [f|_{B(x,1/\epsilon)}]_\alpha. \tag{6-8}$$

Up to uniform equivalence, for any function f supported in $A(R/4, 2R^4)$,

$$\|f\|_{C^{m,\alpha}_{\gamma,\epsilon}(G)} = \sum_{i=0}^{m} \|D_G^i f\|_{C^0} + \|\delta_G^\alpha D_G^m f\|_{C^0}. \tag{6-9}$$

Likewise, let dVol denote the canonical volume form of $\mathbb{R}^2$ and, in analogy to (6-1), (6-2), (6-7) and (6-8), define

$$\mathrm{dVol}_{\mathrm{SF}} := \frac{1}{r^2}\,\mathrm{dVol}, \quad \mathrm{dVol}_G := \epsilon^2\,\mathrm{dVol}. \tag{6-10}$$

In particular, a formula similar to (6-9) also holds for $\|f\|_{H^m_{\gamma,\epsilon}(G)}$ when f is supported over the annulus $A(R/4, 2R^4)$. It is these forms of the norms introduced in (6-5) that we will use in the sequel.

Comparing (6-1) and (6-7) reveals a key phenomenon that must be addressed in order to obtain good estimates. Indeed, over the transition region $A(R/4, 2R)$, the respective differentiation operators of the CHM surface and the Grim ends are approximately related to one another by

$$D_G \simeq \frac{1}{\epsilon R} D_{\mathrm{SF}}, \tag{6-11}$$

so that, whenever a function is transferred from the CHM surface to one of the Grim ends, each order of differentiation introduces a factor of roughly $1/(\epsilon R)$ into the norm. This factor, which is inevitably large, would be ruinous for our estimates unless correctly addressed, and it is in order to do so that we adopt the following two measures. Firstly, we use norms of the least possible order, and likewise take α to be arbitrarily small (see Theorems 6.4.1, 6.5.2 and 6.5.3). In particular, any term involving an exponent of α may be considered heuristically to be close to 1 (see, for example, (6-15), (6-16), (6-19), and so on). Secondly, and more significantly, it is precisely in order to tame this phenomenon that the hybrid norm is introduced. To see how this works, recall that the Sobolev embedding theorem states that, for all m, the Sobolev norm of order m is roughly comparable to the Hölder norm of order $(m-1)$. That is, although the second-order Sobolev norm depends on the second derivative, from a scaling perspective, it behaves more like a first derivative. It is precisely for this reason that the introduction of the factor of $1/(\epsilon R)$ in (6-6) yields a norm which scales, roughly, like a second derivative whilst furnishing, via the Sobolev embedding theorem, stronger information about the first derivative than we would have obtained by working with the Hölder norm alone.

6.2. ***Ping-pong*: *overview.*** We now describe the iteration process used to construct the Green's operator of the approximate MCF soliton. As before, for g a positive integer, let $C := C_g$ denote the CHM surface of genus g and let $S := S_g$ denote the surface obtained by replacing each of its ends with their respective joined ends, as described in Section 5.1. Since there is a natural diffeomorphism from C to S which maps points in the ends of C vertically upwards or downwards, functions over C may equally well be considered as functions over S and vice versa. As before, we will pass between these two perspectives without comment.

Before proceeding, it is worth reviewing the role played by each component within the iteration process that we will apply. We first recall from the previous section that a CHM surface C has been joined to the union $G := G_1 \cup G_2 \cup G_3$ of three Grim ends to yield an approximate soliton S. The surgery is carried out above the annulus $A(R, 2R)$, which we call the *central transition region*. However, these surfaces also all overlap over the larger annulus $A(R/4, 2R^4)$. Consequently, functions supported above $B(2R^4)$ are viewed as functions over C, functions supported over $A(R/4, \infty)$ are viewed as functions over G, and functions supported over $A(R/4, 2R^4)$ are viewed alternately, at different stages of the process, as functions over C and G.

Our aim is to construct a right inverse of the modified Jacobi operator $\hat{J}_S$ of S, using the right inverses of the respective modified Jacobi operators $\hat{J}_C$ and $\hat{J}_G$ of C and G. Ignoring for the moment the finite-dimensional components X, Y, Z and W, we proceed as follows. First let $e : S \to \mathbb{R}$ be a function supported above $B(2R)$. Let χ_u denote the cut-off function of the annulus $A(R^4, 2R^4)$, which we call the *upper transition region*. Viewing e as a function over C, we view $\chi_u(\hat{J}_C^{-1})e$ as an approximator for $(\hat{J}_S^{-1})e$, the cut-off function being here necessary to yield a function supported over $B(2R^4)$, which we may view as a function over S. The error of this approximation is measured by the function $f := \hat{J}_S \chi_u(\hat{J}_C^{-1})e$. Since $\hat{J}_S$ coincides with $\hat{J}_C$ above $B(R)$, this function is supported above $A(R, \infty)$, and we may thus view it as a function over G. In this manner, we have concluded the "upward" stage of the process. Repeating the process in the "downward" direction then yields a function e' supported above $B(2R)$, and the process may then be iterated indefinitely.

Proceeding in this manner, we obtain two sequences $(e_n)_{n\in\mathbb{N}}$ and $(f_n)_{n\in\mathbb{N}}$ of successive errors which should ideally both converge to 0. In this and the next section, estimates for these functions will be obtained in a pointwise manner via the definitions of the norms. In this process, we will encounter some phenomena driving growth and others driving decay. Convergence is ensured upon choosing parameters in such a manner that the latter dominate. The main contributor to growth is the large norm (6-23) of $\hat{J}_G^{-1}$ resulting from the rescaling of the Grim ends. The main contributor to decay is the tendency of bounded harmonic functions to decay over long cylinders, already outlined in Section 1.2, and here encoded implicitly in the weighted Hölder norm introduced in Section 6.1. Roughly speaking, if the radii of the lower and upper transition region are respectively proportional to $R/2$ and R^λ, then the two will be separated by an annulus conformally equivalent to a cylinder roughly of length $(\lambda - 1)\,\mathrm{Ln}(R)$. We thus choose λ as large as possible in order to maximise decay. We have already seen in Section 4 that the strict upper bound $\lambda < 5$ is required in order to obtain uniform estimates for the norms of the Green's operators of the Grim ends (see the proofs of Lemmas 4.2.1 and 4.2.2), and it turns out that $\lambda \in\,]4, 5[$ is sufficient for our purposes.

It remains only to explain the finite-dimensional components in (6-13) and (6-24). It is common in singular perturbation constructions for the Green's operators used to have singular subspaces over which divergence occurs more rapidly than over the rest of the space. This can be understood as a consequence of the existence of a "kernel at infinity", itself often associated to symmetries of the construction, such as vertical translations and dilatations (or, equivalently, variations of the logarithmic parameter). It is thus common to introduce "geometric" terms which, by eliminating the kernel at infinity, allow us to focus on the essential asymptotic behaviours of the Green's operators used, and this is the role played by these finite-dimensional components. Finally, we observe that infinitesimal vertical translations can in fact be introduced in two different ways. Indeed, they can be introduced either in the "upward" stage, as infinitesimal vertical translations of the ends of the CHM surface, or in the "downward" stage as infinitesimal vertical translations of the Grim ends. The former addresses the kernel at infinity of the Green's operator of the CHM surface, whilst the latter addresses the kernel at infinity of the Grim ends. Thus, despite their superficial equivalence, they play distinct roles in the construction, and are both required for it to work.

6.3. *Ping-pong*: *batting up*. For notational convenience, we will henceforth work as if C and S had only one end. Consider now the following seminorms for functions over S:

$$\begin{aligned} \|f\|_{m,C} &:= \|f|_{B(0,4R)}\|_{C^{m,\alpha}_{(2-m)+\delta,\mathrm{SF}}(C)}, & \|f\|_{m,G,S} &:= \|f|_{A(R,\infty)}\|_{H^m_{\gamma,\epsilon}(G)}, \\ \|f\|_{m,G,H} &:= \|f|_{A(R,\infty)}\|_{C^{m,\alpha}_{\gamma,\epsilon}(G)}, & \|f\|_{m,G} &:= \|f\|_{m,G,H}+\frac{1}{\epsilon R}\|f\|_{m,G,S}. \end{aligned} \tag{6-12}$$

Let $\mathcal{E}$ denote the closure with respect to $\|\cdot\|_{0,C}$ of the space of functions supported over $S\cap(B(4R)\times\mathbb{R})$ which are invariant under every horizontal symmetry of the CHM surface C. Likewise, let $\mathcal{F}$ denote the closure with respect to $\|\cdot\|_{0,G}$ of the space of functions supported over $S\cap(A(R,\infty)\times\mathbb{R})$ that are also invariant under these symmetries.

We define the operator $A:\mathcal{E}\to\mathcal{F}$ by

$$Ae := \hat{J}_S\chi_u\Phi e + XUe + YVe - e, \tag{6-13}$$

where χ_u is the cut-off function of the upper transition region $A(R^4, 2R^4)$, and (U, V, Φ) is defined as in Lemma 6.1.1. This operator measures the extent to which $(U, V, \chi_u\Phi)$ fails to be a Green's operator of $(X, Y, \hat{J}_S)$ for functions in $\mathcal{E}$. In particular, since $\hat{J}_S$ coincides with $\hat{J}_C$ over $B(0, R)$, Ae is supported in the interior of $A(R,\infty)$ making it indeed an element of $\mathcal{F}$. In addition, by the definition of $\hat{J}_S$, and bearing in mind that X and Y are both supported in $B(2R)$,

$$Ae = [\hat{J}_G, \chi_u]\Phi e + \chi_u(\hat{J}_S - \hat{J}_C)\Phi e. \tag{6-14}$$

In this section, we prove:

Theorem 6.3.1. *For all* $\delta > 1$,

$$\|Ae\|_{0,G} \lesssim \frac{1}{(\epsilon R)^{2\alpha}}\frac{1}{R^{6+\delta}}\|e\|_{0,C}. \tag{6-15}$$

Theorem 6.3.1 follows immediately from (6-14) together with (6-16), (6-18), (6-19) and (6-21), below, and the fact that

$$\|\chi_u\|_{C^{0,\alpha}_{\gamma,\epsilon}(G)} \lesssim \frac{1}{(\epsilon R^4)^\alpha} \lesssim \frac{1}{(\epsilon R)^\alpha}.$$

For convenience, we now define $\phi := \Phi e$.

Lemma 6.3.2.
$$\|(\hat{J}_S-\hat{J}_C)\Phi e|_{A(R,2R^4)}\|_{C^{0,\alpha}_{\gamma,\epsilon}(G)}\lesssim\frac{1}{(\epsilon R)^\alpha}\frac{1}{R^{6+\delta}}\|e\|_{0,C}. \tag{6-16}$$

Proof. Indeed, by (6-1), for $k\in\{0,1,2\}$, over $A(R,2R^4)$,
$$|D^k\phi|\lesssim\frac{1}{r^{k+\delta}}\|\phi\|_{C^{2,\alpha}_{\delta,\mathrm{SF}}(C)}\lesssim\frac{1}{r^{k+\delta}}\|e\|_{0,C}.$$
Likewise, by (6-2), for all $r\in[2R,R^4]$,
$$|\delta^\alpha(D^2\phi|_{A(r/2,2r)})|\lesssim\frac{1}{r^{k+\alpha+\delta}}\|e\|_{0,C}.$$
Thus, by (5-12), (5-13), (5-16), (5-17) and (5-18), over $A(R,2R^4)$,
$$|(\hat{J}_S-\hat{J}_C)\phi|\lesssim\left(\frac{\epsilon}{r^{2+\delta}}+\frac{\epsilon^2}{r^\delta}+\left[1+\log\left(\frac{r}{R}\right)\right]\epsilon^4r^{2-\delta}+\left[1+\log\left(\frac{r}{R}\right)\right]\frac{1}{r^{6+\delta}}\right)\|e\|_{0,C}, \tag{6-17}$$
so that, by (A-20),
$$|(\hat{J}_S-\hat{J}_C)\phi|_{A(R,2R^4)}|\lesssim\frac{1}{R^{6+\delta}}\|e\|_{0,C}.$$
Likewise, using also (A-10) and (A-12), for $r\in[2R,R^4]$,
$$|\delta^\alpha((\hat{J}_S-\hat{J}_C)\phi|_{A(r/2,2r)})|\lesssim\frac{1}{r^\alpha}\left(\frac{\epsilon}{r^{2+\delta}}+\frac{\epsilon^2}{r^\delta}+\left[1+\log\left(\frac{r}{R}\right)\right]\epsilon^4r^{2-\delta}+\left[1+\log\left(\frac{r}{R}\right)\right]\frac{1}{r^{6+\delta}}\right)\|e\|_{0,C},$$
so that, by (6-8), for $r\in[2R,R^4]$,
$$|\delta^\alpha_G((\hat{J}_S-\hat{J}_C)\phi|_{A(r/2,2r)})|\lesssim\frac{1}{(\epsilon r)^\alpha}\left(\frac{\epsilon}{r^{2+\delta}}+\frac{\epsilon^2}{r^\delta}+\left[1+\log\left(\frac{r}{R}\right)\right]\epsilon^4r^{2-\delta}+\left[1+\log\left(\frac{r}{R}\right)\right]\frac{1}{r^{6+\delta}}\right)\|e\|_{0,C}.$$
Thus, by (A-14) and (A-20),
$$|\delta^\alpha_G((\hat{J}_S-\hat{J}_C)\phi|_{A(R,2R^4)})|\lesssim\frac{1}{(\epsilon R)^\alpha}\frac{1}{R^{6+\delta}}\|e\|_{0,C}.$$
The result follows upon combining the above relations. □

Lemma 6.3.3. *For all $\delta>1$,*
$$\|(\hat{J}_S-\hat{J}_C)\Phi e|_{A(R,2R^4)}\|_{H^0_{\gamma,\epsilon}(G)}\lesssim\frac{(\epsilon R)}{R^{6+\delta}}\|e\|_{0,C}. \tag{6-18}$$

Proof. By (6-10) and (6-17), over $A(R,2R^4)$,
$$|(\hat{J}_S-\hat{J}_C)\phi|^2\,d\mathrm{Vol}_G$$
$$\lesssim\left(\frac{\epsilon^4}{r^{2+2\delta}}+\epsilon^6r^{2-2\delta}+\left[1+\log\left(\frac{r}{R}\right)^2\right]\epsilon^{10}r^{6-2\delta}+\left[1+\log\left(\frac{r}{R}\right)^2\right]\frac{\epsilon^2}{r^{10+2\delta}}\right)\|e\|^2_{0,C}\,d\mathrm{Vol}_{\mathrm{SF}},$$
so that, by (A-21),
$$\int_{A(R,2R^4)}|(\hat{J}_S-\hat{J}_C)\phi|^2\,d\mathrm{Vol}_G\lesssim\frac{(\epsilon R)^2}{R^{12+2\delta}}\|e\|^2_{0,C},$$
and the result follows. □

Lemma 6.3.4. *For all* $\delta > 1$,

$$\|[\hat{J}_G, \chi_u]\Phi e\|_{C^{0,\alpha}_{\gamma,\epsilon}(G)} \lesssim \frac{1}{(\epsilon R^4)^\alpha}\frac{1}{R^{8+4\delta}}\|e\|_{0,C}. \tag{6-19}$$

Proof. By (6-1) and (6-3) for $k \in \{0, 1, 2\}$, over $A(R^4, 2R^4)$,

$$|D^k\phi| \lesssim \frac{1}{R^{4k+4\delta}}\|\phi\|_{C^{2,\alpha}_{\delta,\mathrm{SF}}(C)} \lesssim \frac{1}{R^{4k+4\delta}}\|e\|_{0,C}.$$

It follows by (5-19) that, for $k \in \{0, 1\}$, over this annulus,

$$|D^k[\hat{J}_G, \chi_u]\phi| \lesssim \frac{1}{R^{8+4k+4\delta}}\|e\|_{0,C}. \tag{6-20}$$

Thus, by (6-7), for $k \in \{0, 1\}$, over this annulus,

$$|D^k_G[\hat{J}_G, \chi_u]\phi| \lesssim \frac{1}{(\epsilon R^4)^k}\frac{1}{R^{8+4\delta}}\|e\|_{0,C},$$

and the result follows by (A-10). □

Lemma 6.3.5.

$$\|[\hat{J}_G, \chi_u]\Phi e\|_{H^0_{\gamma,\epsilon}(G)} \lesssim \frac{(\epsilon R)}{R^{5+4\delta}}\|e\|_{0,C}. \tag{6-21}$$

Proof. By (6-20) and (6-10), over $A(R^4, 2R^4)$,

$$|[\hat{J}_G, \chi_u]\phi|^2 \,\mathrm{dVol}_G \lesssim \frac{\epsilon^2}{R^{8+8\delta}}\|e\|^2_{0,C}\,\mathrm{dVol}_{\mathrm{SF}},$$

so that, by (A-21),

$$\int_{A(R^4,2R^4)} |[\hat{J}_G, \chi_u]\phi|^2 \,\mathrm{dVol}_G \lesssim \frac{\epsilon^2}{R^{8+8\delta}}\|e\|_{0,C},$$

and the result follows. □

These estimates prove Theorem 6.3.1. In addition, the following estimate will also be of use later.

Lemma 6.3.6. *For all* $\delta > 1$,

$$\|\chi_u\Phi e\|_{2,G} \lesssim \frac{1}{(\epsilon R)^\alpha}\frac{1}{\epsilon^2 R^{2+\delta}}\|e\|_{0,C}. \tag{6-22}$$

Proof. Indeed, since $\chi_u = O(r^{-k})$, we have $\|\chi_u\|_{C^{2,\alpha}_{0,\mathrm{SF}}(C)} \lesssim 1$. Thus

$$\|\chi_u\phi\|_{C^{2,\alpha}_{\delta,\mathrm{SF}}(C)} \lesssim \|\phi\|_{C^{2,\alpha}_{\delta,\mathrm{SF}}(C)} \lesssim \|e\|_{C^{0,\alpha}_{2+\delta,\mathrm{SF}}(C)} = \|e\|_{0,C}.$$

Thus, by (6-1), (6-3) and (6-7), for $k \in \{0, 1, 2\}$, over $A(R, 2R^4)$,

$$|D^k_G\chi_u\phi| \lesssim \frac{1}{(\epsilon r)^k}\frac{1}{r^\delta}\|e\|_{0,C}.$$

Likewise, by (6-2), (6-3) and (6-8), for all $r \in [2R, R^4]$,

$$|\delta^\alpha_G(D^2_G\chi_u\phi|_{A(r/2,2r)})| \lesssim \frac{1}{(\epsilon r)^{2+\alpha}}\frac{1}{r^\delta}\|e\|_{0,C},$$

so that, by (A-14),

$$|\delta^{\alpha}_{G}(D^2_G\chi_u\phi|_{A(R,2R^4)})|\lesssim\frac{1}{(\epsilon R)^{2+\alpha}}\frac{1}{R^{\delta}}\|e\|_{0,C}.$$

Combining the above relations yields

$$\|\chi_u\phi\|_{2,G,H}\lesssim\frac{1}{(\epsilon R)^{\alpha}}\frac{1}{\epsilon^2 R^{2+\delta}}\|e\|_{0,C}.$$

Likewise, by (6-10), for all k, over $A(R,2R^4)$,

$$|D^k_G\chi_u\phi|^2\,\mathrm{dVol}_G\lesssim\frac{1}{(\epsilon r)^{2k}}\frac{1}{r^{2\delta}}\|e\|^2_{0,C}(\epsilon r)^2\,\mathrm{dVol}_{\mathrm{SF}},$$

and since $\delta>1$, it follows by (A-21) that

$$\|\chi_u\phi\|_{2,G,S}\lesssim\frac{1}{R^{\delta}}\|e\|_{0,C}\lesssim\frac{1}{\epsilon R^{1+\delta}}\|e\|_{0,C}.$$

The result follows. □

6.4. *Ping-pong: batting down.* By Lemma 6.1.2, there exists a linear map $\Psi:C^{0,\alpha}_{\gamma,\epsilon,g}(G)\cap H^0_{\gamma,\epsilon,g}(G)\to C^{2,\alpha,g}_{\gamma,\epsilon,g}(G)\cap H^2_{\gamma,\epsilon,g}(G)$ such that, for all $f\in\mathcal{F}$,

$$f=\hat{J}_G\Psi f,$$

and

$$\|\Psi f\|_{2,\alpha,\gamma,\epsilon}\lesssim\frac{1}{\epsilon^2}\|f\|_{0,\alpha,\gamma,\epsilon}. \tag{6-23}$$

Define the operators $B:\mathcal{F}\to\mathcal{E}$ and $W:\mathcal{F}\to\mathbb{R}^3$ by

$$\begin{aligned}Bf&:=\hat{J}_S(1-\chi_l)(\Psi f-\chi'_{\epsilon}(Wf))-ZWf-f,\\ Wf&:=(\Psi f)(0),\end{aligned} \tag{6-24}$$

where χ_l is the cut-off function of the lower transition region $A(R/4,R/2)$, and χ'_{ϵ} is the cut-off function of the transition region $A(1/2\epsilon,1/\epsilon)$, as in Section 5.1. As before, B measures the extent to which $(-W,(1-\chi_l)(\Psi-\chi'_{\epsilon}W))$ fails to be a Green's operator of $(Z,\hat{J}_S)$ for functions in $\mathcal{F}$. In particular, by (5-11) together with the fact that $\hat{J}_S$ coincides with $\hat{J}_G$ over $A(2R,\infty)$, Bf is supported in $B(4R)$, and is thus indeed an element of $\mathcal{E}$. In addition, since $\chi'_{\epsilon}=1$ over $B(4R)$, over this ball, we have

$$Bf=-[\hat{J}_C,\chi_l](\Psi f-(\Psi f)(0))+(1-\chi_l)(\hat{J}_S-\hat{J}_G)\Psi f. \tag{6-25}$$

In this section, we prove:

Theorem 6.4.1. *For sufficiently small α,*

$$\|Bf\|_{0,C}\lesssim\frac{R^2}{(\epsilon R)}\|f\|_{0,G}. \tag{6-26}$$

Theorem 6.4.1 follows immediately from (6-25) together with (6-28) and (6-30), below, and the fact that

$$\|(1-\chi_l)\|_{C^{0,\alpha}_{0,\mathrm{SF}}(C)}\lesssim 1.$$

For convenience, we now define $\psi:=\Psi f$.

Lemma 6.4.2.
$$\|Wf\| \lesssim \frac{R^2}{(\epsilon R)}\|f\|_{0,G}. \tag{6-27}$$

Proof. Indeed, by the Sobolev embedding theorem,

$$\|Wf\| \lesssim \|\Psi f\|_{H^2_{\gamma,\epsilon}(G)} \lesssim (\epsilon R)\|\Psi f\|_{2,\alpha,\gamma,\epsilon}.$$

Thus, by (6-23),

$$\|Wf\| \lesssim \frac{R}{\epsilon}\|f\|_{0,\alpha,\gamma,\epsilon} \lesssim \frac{R^2}{(\epsilon R)}\|f\|_{0,G},$$

as desired. □

Lemma 6.4.3.
$$\|(\hat{J}_S - \hat{J}_G)\Psi f|_{A(R/4,2R)}\|_{C^{0,\alpha}_{2+\delta,\mathrm{SF}}(C)} \lesssim \frac{1}{(\epsilon R)}\frac{1}{R^{2-\delta}}\|f\|_{0,G}. \tag{6-28}$$

Proof. Indeed, by (6-7), for $k \in \{0,1,2\}$, over $A(R/4, 2R)$,

$$|D^k\psi| \lesssim \epsilon^k\|\psi\|_{C^{2,\alpha}_{\gamma,\epsilon}(G)} \lesssim \frac{1}{\epsilon^{2-k}}\|f\|_{0,G},$$

and so, by (5-12), (5-13), (5-16), (5-17) and (5-18), over $A(R/4, 2R)$,

$$|(\hat{J}_S - \hat{J}_G)\psi| \lesssim \frac{1}{(\epsilon R)}\left(\epsilon + \epsilon^2R^2 + \frac{1}{R^4}\right)\|f\|_{0,G} \lesssim \frac{1}{(\epsilon R)}\frac{1}{R^4}\|f\|_{0,G}.$$

Likewise, by (6-8),

$$|\delta^\alpha(D^2\psi|_{A(R/4,2R)})| \lesssim \epsilon^\alpha\|f\|_{0,G} \lesssim \frac{1}{R^\alpha}\|f\|_{0,G}.$$

Thus, by (6-2), using also (A-10) and (A-12),

$$|\delta^\alpha_{\mathrm{SF}}((\hat{J}_S - \hat{J}_G)\psi|_{A(R/4,2R)})| \lesssim \frac{1}{(\epsilon R)}\frac{1}{R^4}\|f\|_{0,G},$$

and the result follows. □

Lemma 6.4.4.
$$\|\Psi f - (\Psi f)(0)|_{A(R/4,2R)}\|_{2,C} \lesssim \frac{R^{2+\delta}}{(\epsilon R)^{2\alpha}}\|f\|_{0,G}. \tag{6-29}$$

Proof. Bearing in mind (6-8) and the Sobolev embedding theorem, over $A(R/4, 2R)$,

$$[\psi_0] \lesssim (\epsilon R)^{1-\alpha}\|\psi\|_{C^{0,1-\alpha}_{\gamma,\epsilon}(G)} \lesssim (\epsilon R)^{1-\alpha}\|\psi\|_{H^2_{\gamma,\epsilon}(G)} \lesssim (\epsilon R)^{2-\alpha}\|\psi\|_{2,G}.$$

Consequently, by (6-23),

$$[\psi_0] \lesssim \frac{R^2}{(\epsilon R)^\alpha}\|f\|_{0,G}.$$

Likewise, by (4-24) and the subsequent remark, over this annulus,

$$|D_G\psi| \lesssim (\epsilon R)^{1-2\alpha}\|\psi\|_{2,G} \lesssim \frac{1}{(\epsilon R)^{2\alpha}}\frac{R}{\epsilon}\|f\|_{0,G}.$$

Finally, over this annulus,

$$|D^2_G\psi| \lesssim \|\psi\|_{C^{2,\alpha}_{\gamma,\epsilon}(G)} \lesssim \frac{1}{\epsilon^2}\|f\|_{0,G},$$

and

$$|\delta^\alpha_G(D^2_G\psi|_{A(R/4,2R)})| \lesssim \|\phi\|_{C^{2,\alpha}_{\gamma,\epsilon}(G)} \lesssim \frac{1}{\epsilon^2}\|f\|_{0,G}.$$

The result now follows by (6-1), (6-2), (6-3), (6-7), (6-8) and (2-14). □

Lemma 6.4.5. *For sufficiently small* α,

$$\|[\hat{J}_C,\chi_l](\Psi f-(\Psi f)(0))\|_{C^{0,\alpha}_{2+\delta,\mathrm{SF}}(C)} \lesssim \frac{1}{(\epsilon R)^{2\alpha}}R^{2+\delta}\|f\|_{0,G}. \tag{6-30}$$

Proof. This follows from (5-19) and (6-29). □

6.5. ***Ping-pong*: *iteration*.** By (6-15) and (6-26), for $\delta\in\,]1,2[$ and for sufficiently small α, the operator norms of the products AB and BA satisfy

$$\|AB\|,\|BA\| \lesssim \frac{1}{(\epsilon R)^{2\alpha}}\frac{1}{\epsilon R^{5+\delta}} \lesssim \frac{1}{\Lambda}.$$

We therefore define $Q_E:\mathcal{E}\to\mathcal{E}$ and $Q_F:\mathcal{F}\to\mathcal{F}$ by

$$Q_E := \sum_{m=0}^{\infty}(BA)^m, \quad Q_F := \sum_{m=0}^{\infty}(AB)^m. \tag{6-31}$$

In particular, the operator norms of both Q_E and Q_F are uniformly bounded for large values of Λ. We now define

$$\begin{aligned} U_Ce &:= UQ_Ee, & U_Gf &:= -UBQ_Ff,\\ V_Ce &:= VQ_Ee, & V_Gf &:= -VBQ_Ff,\\ W_Ce &:= WAQ_Ee, & W_Gf &:= -WQ_Ff,\end{aligned}$$

$$\begin{aligned} P_Ce &:= \chi_u\Phi Q_Ee-(1-\chi_l)(\Psi AQ_Ee-\chi'_\epsilon(WAQ_Ee)),\\ P_Gf &:= -\chi_u\Phi BQ_Ff+(1-\chi_l)(\Psi Q_Ff-\chi'_\epsilon(WQ_Ff)).\end{aligned} \tag{6-32}$$

Lemma 6.5.1. *For all* $e\in\mathcal{E}$ *and for all* $f\in\mathcal{F}$,

$$\begin{aligned} \hat{J}_SP_Ce+XU_Ce+YV_Ce+ZW_Ce &= e,\\ \hat{J}_SP_Gf+XU_Gf+YV_Gf+ZW_Gf &= f.\end{aligned} \tag{6-33}$$

Proof. Indeed, bearing in mind (6-13) and (6-24),

$$\begin{aligned} &\hat{J}_SP_Ce+XU_Ce+YV_Ce+ZW_Ce\\ &\quad= \hat{J}_S\chi_u\Phi Q_Ee+XUQ_Ee+YVQ_Ee\ -\hat{J}_S(1-\chi_l)(\Psi AQ_Ee-\chi'_\epsilon(WAQ_Ee))+ZWAQ_Ee\\ &\quad= AQ_Ee+Q_Ee-BAQ_Ee-AQ_Ee=e.\end{aligned}$$

The second relation follows in a similar manner, and this completes the proof. □

Now let χ be the cut-off function of the transition region $A(2R,4R)$. Since $\chi=O(r^{-k})$, for all f,

$$\|\chi f\|_{0,C}\lesssim\|f\|_{0,C}, \quad \|(1-\chi)f\|_{0,G}\lesssim\frac{1}{(\epsilon R)^\alpha}\|f\|_{0,G}. \tag{6-34}$$

Define

$$\begin{aligned}\widehat{U}f &:= U_C\chi f + U_G(1-\chi)f, \quad \widehat{W}f := W_C\chi f + W_G(1-\chi)f,\\ \widehat{V}f &:= V_C\chi f + V_G(1-\chi)f, \quad \widehat{P}f := P_C\chi f + P_G(1-\chi)f.\end{aligned} \tag{6-35}$$

In particular, by (6-33),

$$\hat{J}_S\widehat{P}f + X\widehat{U}f + Y\widehat{V}f + Z\widehat{W}f = f, \tag{6-36}$$

so that $(\widehat{U}, \widehat{V}, \widehat{W}, \widehat{P})$ defines a Green's operator for $(X, Y, Z, \hat{J}_S)$. We conclude this section by determining the norms of its different components. First, since the operator norms of U and V are uniformly bounded, by (6-26), (6-32) and (6-34)

$$\begin{aligned}\|\widehat{U}f\| &\lesssim \|f\|_{0,C} + \frac{R^2}{(\epsilon R)^{1+\alpha}}\|f\|_{0,G},\\ \|\widehat{V}f\| &\lesssim \|f\|_{0,C} + \frac{R^2}{(\epsilon R)^{1+\alpha}}\|f\|_{0,G}.\end{aligned} \tag{6-37}$$

Theorem 6.5.2. *For sufficiently small α, for all $\delta \in]1, 2[$, and for all f,*

$$\|\widehat{W}f\| \lesssim \|f\|_{0,C} + \frac{R^2}{(\epsilon R)^{1+\alpha}}\|f\|_{0,G}. \tag{6-38}$$

Proof. For $e \in \mathcal{E}$, by (6-15) and (6-27),

$$\|W_C e\| = \|WAQ_E e\| \lesssim \frac{R^2}{(\epsilon R)}\|AQ_E e\|_{0,G} \lesssim \frac{1}{(\epsilon R)^{2\alpha}}\frac{1}{\epsilon R^{5+\delta}}\|e\|_{0,C} \lesssim \|e\|_{0,C}.$$

For $f \in \mathcal{F}$, by (6-27),

$$\|W_G f\| = \|WQ_F f\| \lesssim \frac{R^2}{(\epsilon R)}\|f\|_{0,G}.$$

The result now follows by (6-34). □

Theorem 6.5.3. *For sufficiently small α, for all $\delta \in]1, 2[$, and for all f,*

$$\|\widehat{P}f\|_{2,C} \lesssim \|f\|_{0,C} + \frac{R^2}{(\epsilon R)^{1+\alpha}}\|f\|_{0,G}. \tag{6-39}$$

Proof. Consider $e \in \mathcal{E}$. Observe that, over $B(4R)$,

$$P_C e = \Phi Q_E e - (1-\chi_l)(\Psi AQ_E e - \Psi AQ_E e(0)).$$

Now,

$$\|\Phi Q_E e\|_{2,C} \lesssim \|e\|_{0,C},$$

and by (6-15) and (6-29),

$$\|(1-\chi_l)(\Psi AQ_E e - (\Psi AQ_E e)(0))\|_{2,C} \lesssim \frac{1}{(\epsilon R)^{4\alpha}R^4}\|e\|_{0,C} \lesssim \|e\|_{0,C},$$

so that

$$\|P_C e\|_{2,C} \lesssim \|e\|_{0,C}.$$

Now consider $f\in\mathcal{F}$. Over $B(4R)$,

$$P_Gf=-\Phi BQ_Ff-(1-\chi_l)(\Psi Q_Ff-\Psi Q_Ff(0)).$$

By (6-26),

$$\|\Phi BQ_Ff\|_{2,C}\lesssim\frac{R^2}{(\epsilon R)}\|f\|_{0,G},$$

and, by (6-29),

$$\|(1-\chi_l)(\Psi Q_Ff-(\Psi Q_Ff)(0))\|_{2,C}\lesssim\frac{R^{2+\delta}}{(\epsilon R)^{2\alpha}}\|f\|_{0,G},$$

so that,

$$\|P_Gf\|_{2,C}\lesssim\frac{R^2}{(\epsilon R)}\|f\|_{0,G}.$$

The result now follows by (6-34) and (6-35). □

Theorem 6.5.4. *For sufficiently small α, for all $\delta\in\,]1,2[$, and for all f,*

$$\|\widehat{P}f\|_{2,G}\lesssim\frac{1}{(\epsilon R)^{\alpha}}\frac{1}{\epsilon^2R^{2+\delta}}\bigg(\|f\|_{0,C}+\frac{R^2}{(\epsilon R)^{1+\alpha}}\|f\|_{0,G}\bigg). \tag{6-40}$$

Proof. Consider $e\in\mathcal{E}$. Observe that, over $S\cap(A(R,\infty)\times\mathbb{R})$,

$$P_Ce=\chi_u\Phi Q_Ee-\Psi AQ_Ee+(WAQ_Ee)\chi_\epsilon'.$$

By (6-22),

$$\|\chi_u\Phi Q_Ee\|_{2,G}\lesssim\frac{1}{(\epsilon R)^{\alpha}}\frac{1}{\epsilon^2R^{2+\delta}}\|e\|_{0,C}.$$

By (6-15) and (6-23),

$$\|\Psi AQ_Ee\|_{2,G}\lesssim\frac{1}{(\epsilon R)^{2\alpha}}\frac{1}{\epsilon^2R^{6+\delta}}\|e\|_{0,C}.$$

By (6-15) and (6-27)

$$\|WAQ_Ee\|\lesssim\frac{R^2}{(\epsilon R)}\|AQ_Ee\|_{0,G}\lesssim\frac{1}{(\epsilon R)^{2\alpha}}\frac{1}{\epsilon R^{5+\delta}}\|e\|_{0,C}.$$

However,

$$\|\chi_\epsilon'\|_{2,G}\lesssim\frac{1}{(\epsilon R)},$$

and so

$$\|(WAQ_Ee)\chi_\epsilon'\|_{2,G}\lesssim\frac{1}{(\epsilon R)^{2\alpha}}\frac{1}{\epsilon^2R^{6+\delta}}\|e\|_{0,C}.$$

Combining these relations yields

$$\|P_Ce\|_{2,G}\lesssim\frac{1}{(\epsilon R)^{\alpha}}\frac{1}{\epsilon^2R^{2+\delta}}\|e\|_{0,C}.$$

Consider now $f\in\mathcal{E}$. Over $S\cap(A(R,\infty)\times\mathbb{R})$,

$$P_Gf=-\chi_u\Phi BQ_Ff+\Psi Q_Ff-(WQ_Ff)\chi_\epsilon'.$$

By (6-22) and (6-26),

$$\|\chi_u \Phi B Q_F f\|_{2,G} \lesssim \frac{1}{(\epsilon R)^\alpha}\frac{1}{\epsilon^2 R^\delta(\epsilon R)}\|f\|_{0,G}.$$

By (6-23),

$$\|\Psi Q_F f\|_{2,G} \lesssim \frac{1}{\epsilon^2}\|f\|_{0,G}.$$

By (6-27),

$$\|W Q_F f\| \lesssim \frac{R^2}{(\epsilon R)}\|f\|_{0,G},$$

so that

$$\|(W Q_F f)\chi'_\epsilon\|_{2,G} \lesssim \frac{1}{\epsilon^2}\|f\|_{0,G}.$$

Combining these relations yields

$$\|P_G f\|_{2,G} \lesssim \frac{1}{(\epsilon R)^\alpha}\frac{1}{\epsilon^2 R^\delta(\epsilon R)}\|f\|_{0,G}.$$

The result now follows by (6-34). □

7. Existence and embeddedness

7.1. ***The Schauder fixed-point theorem.*** It remains only to perturb the approximate MCF solitons constructed in Section 5 into actual MCF solitons. This perturbation will be carried out using the Schauder fixed-point theorem. It will first be convenient to modify slightly the norms introduced in (6-12). We thus define

$$\begin{aligned}\|f\|'_{m,G,H} &:= \|f|_{A(2R,\infty)}\|_{C^{m,\alpha}_{\gamma,\epsilon}(G)},\\ \|f\|'_{m,G,S} &:= \|f|_{A(2R,\infty)}\|_{H^{m}_{\gamma,\epsilon}(G)},\\ \|f\|'_{m,G} &:= \|f\|'_{m,G,H} + \frac{1}{(\epsilon R)}\|f\|'_{m,G,S}.\end{aligned} \tag{7-1}$$

By (6-35), this does not affect (6-37), (6-38), (6-39) and (6-40). In addition, we will also ignore the factor $\langle \widehat{N}_S, N_S\rangle^{-1}$ used in the definitions (5-8) and (5-9) of $(X, Y, Z, \hat{J}_S)$. Indeed, we readily show that the operator of multiplication by this function is uniformly bounded, independent of Λ, with respect to the norms $\|\cdot\|_{0,C}$ and $\|\cdot\|_{0,G}$, for which reason it also does not affect the above estimates.

For all nonnegative, integer m, for all $\alpha \in [0, 1]$ and for all real γ, let $E_{m,\alpha,\gamma}$ be the space of m-times differentiable functions $f : S \to \mathbb{R}$ which are invariant under all horizontal symmetries of C and which satisfy

$$\|f\|_{m,C}, \|f\|'_{m,G} < \infty.$$

Observe that $E_{m,\alpha,\gamma}$ furnished with these norms is a Fréchet space. Now let

$$M : U \oplus V \oplus W \oplus E_{2,\alpha,\gamma} \to E_{0,\alpha,\gamma}$$

be the MCFS functional about S, as defined in Sections 5.1 and 5.4. It only remains to study how M varies up to second order about S. As before, throughout this section, we apply (2-14) without comment.

Lemma 7.1.1.
$$\|M(0,0,0,0)\|_{0,C} \lesssim R^{\delta-2}, \quad \|M(0,0,0,0)\|'_{0,G} = 0. \tag{7-2}$$

Proof. Define $\psi := M(0,0,0,0)$. Since C is minimal, over $B(R)$,

$$\psi = \epsilon\mu.$$

Thus, by (5-1) and (A-6),

$$\|\psi|_{B(R)}\|_{C^{0,\alpha}_{2+\delta,\mathrm{SF}}(C)} \lesssim \epsilon R^{2+\delta} \lesssim R^{\delta-2}.$$

By (5-4), over $A(R,2R)$,

$$H_i = \frac{cx^i}{r^2} + \mathrm{O}(R^{-(3+k)}),$$
$$H_{ij} = \frac{c}{r^2}\left(\delta_{ij} - \frac{x^i x^j}{2r^2}\right) + \mathrm{O}(R^{-(4+k)}).$$

Thus, by (A-6), over this annulus,

$$\mu = 1 + \mathrm{O}(R^{-2+k}), \quad g^{ij} = \delta_{ij} + \mathrm{O}(R^{-2+k}),$$

so that, by (A-7),

$$\psi = \mathrm{O}(R^{-(4+k)}).$$

Consequently,

$$\|\psi|_{A(R,2R)}\|_{C^{0,\alpha}_{2+\delta,\mathrm{cyl}}(C)} \lesssim R^{\delta-2},$$

and the first estimate follows upon combining these relations. Finally, by construction, ψ vanishes over $A(2R,\infty)$, so that $\|\psi\|'_{0,G} = 0$, and this completes the proof. □

It is straightforward to show that for $\|u\|$, $\|v\|$, $\|w\|$ and $\|f\|_{2,C}$ sufficiently small, independent of Λ,

$$\|M(u,v,w,f) - M(0,0,0,0) - \hat{J}_S f - \widehat{X}u - \widehat{Y}v\|_{0,C} \lesssim \|f\|^2_{2,C} + \|u\|^2 + \|v\|^2. \tag{7-3}$$

The corresponding estimate over rotationally symmetric Grim ends is more subtle.

Lemma 7.1.2. *There exists $\eta > 0$ such that, for sufficiently large Λ, if $\epsilon(\epsilon R)^{1-2\alpha}\|f\|'_{2,G} < \eta$, then*

$$\|M(u,v,w,f) - M(0,0,0,0) - \hat{J}_S f - \widehat{Z}w\|'_{0,G} \lesssim \frac{\epsilon^2}{R}\|u\|^2 + \frac{\epsilon^2}{R}\|v\|^2 + \frac{\epsilon^2}{R}\|w\|^2 + \epsilon^3(\epsilon R)^{1-2\alpha}(\|f\|'_{2,G})^2. \tag{7-4}$$

Remark. Before continuing, it is worth reflecting on the terms that will appear in the following proof. First, on the scale of the rotationally symmetric Grim end, the perturbation that we make is of order ϵ so that, since this perturbation is quadratic, it introduces a factor of ϵ^2. Second, returning to the scale of the joined surface introduces a further factor of ϵ, thus explaining the factor of ϵ^3 in the formulae below.

Proof. Since M is a second-order quasilinear functional, upon rescaling, we obtain, for all u, for all v, and for all g with $\|\epsilon g\|'_{1,G,H}$ sufficiently small,

$$\|M(u,v,0,g) - M(u,v,0,0) - \hat{J}_{S,u,v}g\|'_{0,G} \lesssim \epsilon^3\|g\|'_{1,G,H}\|g\|'_{2,G} \lesssim \frac{\epsilon^2}{R}(\|g\|'_{1,G,H})^2 + \epsilon^3(\epsilon R)(\|g\|'_{2,G})^2.$$

Next, for all sufficiently small u and v, and for all g,

$$\|(\hat{J}_{S,u,v} - \hat{J}_S)g\|'_{0,G} \lesssim \epsilon^3(\|u\| + \|v\|)\|g\|'_{2,G} \lesssim \frac{\epsilon^2}{R}\|u\|^2 + \frac{\epsilon^2}{R}\|v\|^2 + \epsilon^3(\epsilon R)(\|g\|'_{2,G})^2.$$

Now, bearing in mind the definition of the macroscopic perturbation in the direction of w,

$$\begin{aligned}
&\|M(u,v,w,f)-M(0,0,0,0)-\hat{J}_S f-\widehat{Z}w\|'_{0,G}\\
&\qquad\lesssim \|M(u,v,w,f)-M(u,v,0,0)-\hat{J}_S f-\widehat{Z}w\|'_{0,G}\\
&\qquad\lesssim \|M(u,v,0,f+w(1-\chi'_\epsilon))-M(u,v,0,0)-\hat{J}_S(f+w(1-\chi_{\epsilon'}))\|'_{0,G}\\
&\qquad\lesssim \|M(u,v,0,f+w(1-\chi'_\epsilon))-M(u,v,0,0)-\hat{J}_{S,u,v}(f+w(1-\chi_{\epsilon'}))\|'_{0,G}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad+\|(\hat{J}_{S,u,v}-\hat{J}_S)(f+w(1-\chi_{\epsilon'}))\|'_{0,G}\\
&\qquad\lesssim \frac{\epsilon^2}{R}\|u\|^2+\frac{\epsilon^2}{R}\|v\|^2+\frac{\epsilon^2}{R}(\|f+w(1-\chi'_\epsilon)\|'_{1,G,H})^2+\epsilon^3(\epsilon R)(\|f+w(1-\chi'_\epsilon)\|'_{2,G})^2.
\end{aligned}$$

Finally,

$$\|(1-\chi'_\epsilon)|_{A(1/(2\epsilon),1/\epsilon)}\|'_{1,G,H}\lesssim 1,$$
$$\|(1-\chi'_\epsilon)|_{A(1/(2\epsilon),1/\epsilon)}\|'_{2,G}\lesssim \frac{1}{\epsilon R},$$

and the result now follows by Lemma 4.3.1 and the subsequent remark. □

This concludes our analysis of M up to second order about S. We are now ready to prove existence.

Theorem 7.1.3. *For γ sufficiently small, for all $\delta\in]1,2[$, for $\alpha\in]0,1[$ sufficiently small, and for Λ sufficiently large, there exist u, v, w and f such that*

$$M(u,v,w,f)=0.$$

Furthermore,

$$\|u\|,\ \|v\|,\ \|w\|,\ \|f\|_{2,C}\lesssim R^{\delta-2},\quad \|f\|_{2,G}\lesssim\frac{1}{(\epsilon R)^\alpha\epsilon^2R^4}. \tag{7-5}$$

Proof. Fix $\gamma\ll 1$, $\delta\in]1,2[$ and $\alpha\in]0,1[$ small. Set $\psi_0:=M(0,0,0,0)$ and define

$$(u_0,v_0,w_0,f_0):=\phi_0:=-(\widehat{U}\psi_0,\widehat{V}\psi_0,\widehat{W}\psi_0,\widehat{P}\psi_0).$$

By (6-37), (6-38), (6-39), (6-40) and (7-2), there exists a constant $B>0$, such that, for all large Λ,

$$\|u_0\|,\ \|v_0\|,\ \|w_0\|,\ \|f_0\|_{2,C}\leq BR^{\delta-2},\qquad \|f_0\|'_{2,G}\leq\frac{B}{(\epsilon R)^\alpha\epsilon^2R^4}.$$

Define $\Omega\subseteq\mathbb{R}^3\oplus\mathbb{R}^3\oplus\mathbb{R}^3\oplus E_{2,\alpha,\gamma}$ to be the set of all quadruplets (u,v,w,f) such that

$$\|u\|,\ \|v\|,\ \|w\|,\ \|f\|_{2,C}\leq 2BR^{\delta-2},\qquad \|f\|'_{2,G}\leq\frac{2B}{(\epsilon R)^\alpha\epsilon^2R^4}.$$

Observe that Ω is convex and, by the Arzelà–Ascoli theorem, for all $\alpha'<\alpha$ and $\gamma'<\gamma$, Ω is a compact subset of $\mathbb{R}^3\oplus\mathbb{R}^3\oplus\mathbb{R}^3\oplus E_{2,\alpha',\gamma'}$. For $\phi:=(u,v,w,f)$ in Ω, define

$$\Phi(\phi):=\phi_0-(\widehat{U}\psi,\widehat{V}\psi,\widehat{W}\psi,\widehat{P}\psi),$$

where

$$\psi:=M(u,v,w,f)-M(0,0,0,0)-\hat{J}_Sf-\widehat{X}u-\widehat{Y}v-\widehat{Z}w.$$

By (7-3), (7-4) and (2-14),

$$\|\psi\|_{0,C} \lesssim R^{2\delta-4}, \quad \|\psi\|'_{0,G} \lesssim \frac{1}{(\epsilon R)^{6\alpha} R^7},$$

so that, by (6-37), (6-38), (6-39) and (6-40), for sufficiently large Λ, Φ maps Ω to itself. Furthermore, for all $\alpha' < \alpha$ and $\gamma' < \gamma$, Φ is continuous with respect to the topology of $E_{2,\alpha',\gamma'}$. It follows by the Schauder fixed-point theorem (see [Gilbarg and Trudinger 1983]) that there exists a fixed point ϕ of Φ in Ω. We readily verify that $M(\phi) = 0$, and this completes the proof. □

Theorem 7.1.4. *Let (u, v, w, f) be as in Theorem 7.1.3. For sufficiently large Λ, the surface $\widetilde{E}(u, v, w, f)$ is embedded.*

Proof. We denote the joined surface by S, we denote the image of $\widetilde{E}(u, v, w, f)$ by S', and we rescale both S and S' by ϵ. Observe that the intersection of S with $A(2\epsilon R, \infty) \times \mathbb{R}$ consists of three distinct rotationally symmetric Grim ends, which we denote by G_+, G_0 and G_- respectively. Let u_+, u_0 and u_- be the respective profiles of these ends, and let v_+, v_0 and v_- be the respective derivatives of these functions in the radial direction. Observe that

$$u_+(\epsilon R) > u_0(\epsilon R) > u_-(\epsilon R),$$
$$v_+(\epsilon R) > v_0(\epsilon R) > v_-(\epsilon R).$$

Since v_+, v_0 and v_- are all solutions of the same first-order ODE, it follows that $v_+(r) > v_0(r) > v_-(r)$ for all r. In particular, the ends G_+, G_0 and G_+ are separated vertically by a distance of no less than η, where $\eta \sim \epsilon \log(R)$. Let Ω_+, Ω_0 and Ω_- denote the open sets of points lying at a vertical distance of no more than $\eta/2$ from G_+, G_0 and G_- respectively. Observe, in particular, that these three sets are disjoint.

Now let G'_+, G'_0 and G'_- be the three ends of S'. Over the annulus $A(\epsilon R, 2\epsilon R)$, by (7-5),

$$\|\epsilon f|_{A(R,2R)}\|_{C^0} \lesssim \epsilon R^{-\delta} \|f|_{(A(R,2R)}\|_{2,C} \lesssim \epsilon R^{-2},$$

so that, over this annulus, G'_+ lies strictly above G'_0, and G'_0 lies strictly above G'_-. However, by Lemma 4.3.1 and the subsequent remark and (7-5) again,

$$\|\epsilon f\|'_{1,G,H} \lesssim \frac{1}{(\epsilon R)^{3\alpha} R^3}.$$

Bearing in mind the definition of the norm $\|\cdot\|_{1,G,H}$, it follows that for sufficiently large Λ, G'_+, G'_0 and G'_- are all graphs over $A(\epsilon R, \infty)$. Furthermore, for some large R', the intersections of G'_+, G'_0 and G'_- with $A(R', \infty) \times \mathbb{R}$ are contained in Ω_+, Ω_0 and $\Omega-$ respectively. In particular, outside $B(R') \times \mathbb{R}$, G'_+ lies strictly above G'_0 and G'_0 lies strictly above G'_-. Since vertical translates of mean curvature flow solitons are also mean curvature flow solitons, it now follows by the strong maximum principle that, over the whole of $A(\epsilon R, \infty)$, G'_+ lies strictly above G'_0 and G'_0 lies strictly above G_-. □

Appendix A: Terminology, conventions and standard results

A1. *General definitions.* Let $\mathbb{R}^2$ and $\mathbb{R}^3$ denote respectively 2- and 3-dimensional Euclidean space. We consider $\mathbb{R}^2$ as the $(x-y)$ plane in $\mathbb{R}^3$. Let $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ denote the canonical projection. Let r denote a smooth positive function over $\mathbb{R}^2$ which is equal to the distance to the origin outside some (suitably large) compact set. We denote the composition of r with π also by r. Let e_x, e_y and e_z denote the vectors of the canonical basis of $\mathbb{R}^3$. Let e_r, e_θ denote respectively the unit radial and unit angular vector fields about the origin over $\mathbb{R}^2$ and about the z-axis over $\mathbb{R}^3$. Let D denote the canonical differentiation operator over $\mathbb{R}^2$ and $\mathbb{R}^3$. Let Δ denote the canonical Laplacian over $\mathbb{R}^2$ (not to be confused with Δ^Σ, defined below). Let $C(a)$ denote the circle of radius a about the origin in $\mathbb{R}^2$. Let $B(a)$ denotes the closed disk of radius a about the origin in $\mathbb{R}^2$. Let $A(a, b)$ denote the closed annulus of inner radius a and outer radius b about the origin in $\mathbb{R}^2$. Let $\chi : [0, \infty[\to \mathbb{R}$ be a nonnegative, nonincreasing function such that $\chi = 1$ over $[0, 1]$ and $\chi = 0$ over $[2, \infty[$. For all a, define $\chi_a : \mathbb{R}^2 \to \mathbb{R}$ by $\chi_a(x) = \chi(\|x\|/a)$. We call χ_a the *cut-off function* of the *transition region* $A(a, 2a)$. Composing with π, we likewise consider χ_a as a function over $\mathbb{R}^3$.

A2. ***Surface geometry.*** Let Σ be an embedded surface in $\mathbb{R}^3$. Let N_Σ denote the unit normal vector field over Σ. Let π^Σ denote the orthogonal projection onto the tangent space of Σ. Let ∇^Σ denote the gradient operator as well as the Levi-Civita covariant derivative of Σ. Let Hess^Σ denote the intrinsic Hessian operator of Σ. Let Δ^Σ denote the intrinsic Laplacian of Σ. Let II^Σ denote the second fundamental form of Σ. Let A_Σ denote the shape operator of Σ. Let H_Σ denote the mean curvature of Σ (taken to be the *sum* of the principle curvatures, or the trace of the shape operator). Let M_Σ denote the MCFS operator of Σ (with speed ϵ). It is given by

$$M_\Sigma := H_\Sigma + \epsilon\langle N_\Sigma, e_z\rangle. \tag{A-1}$$

Let J_Σ denote the MCFS Jacobi operator (with speed ϵ) of Σ. That is, J_Σ is the linearisation of the MCFS operator of Σ. It is given by

$$J_\Sigma f = \Delta^\Sigma f + \mathrm{Tr}(A_\Sigma^2) f + \epsilon\langle \nabla^G f, e_z\rangle. \tag{A-2}$$

Finally, we recall the following elementary relations. For any function f defined over a neighbourhood of Σ,

$$\begin{aligned} \nabla^\Sigma f &= Df - \langle Df, N_\Sigma\rangle N_\Sigma, \\ \mathrm{Hess}^\Sigma(f) &= \mathrm{Hess}(f) - \langle Df, N_\Sigma\rangle \mathrm{II}^\Sigma. \end{aligned} \tag{A-3}$$

Given any positive function ϕ defined over Σ, if $\hat{J}_\Sigma := M_\phi^{-1} J_\Sigma M_\phi$ denotes the conjugate of J_Σ with the operator of multiplication by ϕ, then

$$\hat{J}_\Sigma f = \Delta^\Sigma f + 2\phi^{-1}\langle \nabla^\Sigma \phi, \nabla^\Sigma f\rangle + \epsilon\langle \nabla^\Sigma f, e_z\rangle + (\phi^{-1} J_\Sigma \phi) f. \tag{A-4}$$

A3. ***Surface geometry of graphs.*** If Σ is the graph of a function u over a subset of $\mathbb{R}^2$, then we call u the *profile* of Σ. In this case, π defines a coordinate chart of Σ in $\mathbb{R}^2$. It will be more convenient to work, sometimes over Σ, and sometimes over $\mathbb{R}^2$, and we will move freely between these two perspectives. Let g_{ij} denote the intrinsic metric of Σ. Its inverse is denoted by g^{ij}. Let Γ_{ij}^k denote the Christoffel symbols of the Levi-Civita covariant derivative of g_{ij}. Setting

$$\mu := \langle e_z, N_\Sigma\rangle, \tag{A-5}$$

we readily verify the following relations:

$$\begin{aligned} \mu &= \frac{1}{\sqrt{1+\|Du\|^2}}, & \Delta^\Sigma(f) &= g^{ij}f_{ij} - g^{ij}g^{kp}u_{ij}u_pf_k, \\ g_{ij} &= \delta_{ij} + u_iu_j, & \mathrm{II}^\Sigma_{ij} &= -\mu u_{ij}, \\ g^{ij} &= \delta_{ij} - \mu^2u^iu^j, & (A^\Sigma)^i_j &= -\mu g^{ip}u_{pj}, \\ \Gamma^k_{ij} &= g^{kp}u_{ij}u_p, & H^\Sigma &= -\mu g^{ij}u_{ij}, \\ \mathrm{Hess}^\Sigma(f)_{ij} &= f_{ij} - g^{kp}u_{ij}u_pf_k, & \pi^T(e_z)_i &= \mu^2u_i. \end{aligned} \tag{A-6}$$

Finally, when Σ is a graph, the MCFS functional is given by

$$M_\Sigma = -\mu g^{ij}u_{ij} + \epsilon\mu. \tag{A-7}$$

A4. ***Function spaces.*** Let X be a metric space. For all $\alpha \in [0, 1]$, we define the *Hölder seminorm* of order α over X by

$$[f]_\alpha := \mathrm{Sup}_{x\neq y\in X}\frac{|f(x)-f(y)|}{d(x,y)^\alpha}. \tag{A-8}$$

Observe that $[f]_0$ measures the *total oscillation* of f. In particular,

$$[f]_0 \leq 2\|f\|_{C^0}. \tag{A-9}$$

For all $\alpha \in [0, 1]$,

$$[f]_\alpha \leq [f]_0^{1-\alpha}[f]_1^\alpha \leq 2^{1-\alpha}\|f\|_{C^0}^{1-\alpha}[f]_1^\alpha. \tag{A-10}$$

If X is a complete manifold, and if f is differentiable over X, then, for all $\alpha \in [0, 1[$ and for all $\beta \in]0, 1]$,

$$\|Df\|_{C^0} \leq 2[f]_\alpha^{\beta/(1+(\beta-\alpha))}[Df]_\beta^{(1-\alpha)/(1+(\beta-\alpha))}. \tag{A-11}$$

For all α,

$$[fg]_\alpha \leq \|f\|_{C^0}[g]_\alpha + [f]_\alpha\|g\|_{C^0}. \tag{A-12}$$

Finally, if $X = X_1 \cup \cdots \cup X_m$, then, for all α,

$$[f]_\alpha \leq m^{1-\alpha}\,\mathrm{Sup}_{1\leq k\leq m}[f|_{X_i}]_\alpha. \tag{A-13}$$

If, in particular, $X = [0, m+1] \times S^1$ is a cylinder and $X_i = [i, i+1] \times S^1$ for all i, then (A-13) refines to

$$[f]_\alpha \leq \sum_{i=1}^m [f|_{X_i}]_\alpha. \tag{A-14}$$

For a continuous function f over X, for all α, we define

$$\delta^\alpha f(x) := [f|_{B_1(x)}]_\alpha. \tag{A-15}$$

Now suppose that X is a smooth Riemannian manifold. For all k, α, we define the $C^{k,\alpha}$-Hölder norm over $C^\infty(M)$ by

$$\|f\|_{C^{k,\alpha}} := \sum_{i=0}^k \|D^if\|_{C^0} + \|\delta^\alpha D^kf\|_{C^0}. \tag{A-16}$$

We define the space $C^{k,\alpha}(X)$ to be the closure of $C^\infty(X)$ with respect to this norm. For all p, we define the L^p-norm over $C_0^\infty(M)$ by

$$\|f\|_{L^p}^p := \int_X |f|^p \,\mathrm{dVol}. \tag{A-17}$$

We define the space $L^p(X)$ to be the closure of $C_0^\infty(X)$ with respect to this norm. For all k, we define the H^k-Sobolev norm over $C_0^\infty(M)$ by

$$\|f\|_{H^k} := \sum_{i=0}^{k} \|D^i f\|_{L^2}. \tag{A-18}$$

The reader may verify that all surfaces studied in this paper are sufficiently regular at infinity for the Sobolev embedding theorem to hold. That is for all l, and for all $k+\alpha < l-1$,

$$\|f\|_{C^{k,\alpha}} \lesssim \|f\|_{H^l}. \tag{A-19}$$

The following formulae are readily verified:

$$\mathrm{Sup}_{t\in[1,T]} \log(t)t^\alpha \lesssim \begin{cases} \log(T)T^\alpha & \text{if } \alpha > 0, \\ 1 & \text{if } \alpha < 0 \end{cases} \tag{A-20}$$

and

$$\int_{A(1,T)} \log(r)^m r^\alpha \,\mathrm{dVol}_{\mathrm{SF}} \lesssim \begin{cases} \log(T)^m T^\alpha & \text{if } \alpha > 0, \\ 1 & \text{if } \alpha < 0. \end{cases} \tag{A-21}$$

A5. *Elliptic estimates.* Let E and F be Banach spaces and let $A : E \to F$ be a bounded linear map. We say that A satisfies an *elliptic estimate* whenever there exists a normed vector space G, a compact map $K : E \to G$, and a constant C such that, for all e in E,

$$\|e\| \le C(\|Ke\| + \|Ae\|). \tag{A-22}$$

The following straightforward result plays an important role in Fredholm theory.

Theorem A5.1. *If A satisfies an elliptic estimate, then the kernel of A is finite-dimensional and its image is a closed subset of F.*

Appendix B: Catenoidal minimal ends

In this appendix, we use the Weierstrass representation to determine the asymptotics of horizontal, catenoidal minimal ends. This is used in Sections 5 and 6 to model the asymptotics of CHM surfaces.

Let C be a horizontal, catenoidal minimal end. Its intrinsic metric is biholomorphic to the punctured disk which, for the purposes of this appendix, it is useful to view as the complement of the closed unit disk in $\mathbb{C}$, that is,

$$\Delta^* := \{\zeta \in \mathbb{C} \mid |\zeta| > 1\}. \tag{B-1}$$

The Weierstrass representation (see [Weber 2005]) is a parametrisation of C by a function $\Phi : \Delta^* \to \mathbb{R}^3$ of the form

$$\Phi(\zeta) := \mathrm{Re}\Big(\int^\zeta \Big(\frac{1}{2}\Big(G - \frac{1}{G}\Big), \frac{1}{2i}\Big(G + \frac{1}{G}\Big), 1\Big) h\, d\zeta\Big) \tag{B-2}$$

for some holomorphic functions $G, h : \Delta^* \to \mathbb{C}$. These functions are interpreted geometrically as follows. Setting $\Phi := (\Phi_1, \Phi_2, \Phi_3)$, we readily show that

$$h = 2\partial_\zeta \Phi_3. \tag{B-3}$$

That is, $hd\zeta$ is twice the holomorphic part of the derivative of the height function of C. The geometric significance of G is more subtle, but with some work we can show that it is the image under the stereographic projection of the unit normal vector field over C.

Define $\rho := |\zeta|$. Since C is a horizontal catenoidal end, Φ_3 is asymptotic to $a + c\log(\rho)$ for some constants a and c, and it follows that

$$h = \sum_{k=-\infty}^{-1} h_k \zeta^k. \tag{B-4}$$

Meanwhile, since the normal of C is asymptotically vertical, G may be chosen to vanish at infinity, so that

$$G = \sum_{k=-\infty}^{-1} G_k \zeta^k. \tag{B-5}$$

In addition, since C is a single-valued graph over some neighbourhood of infinity in $\mathbb{R}^2$, the functions h and G together satisfy a vanishing holonomy condition around the puncture at infinity. In terms of their Laurent coefficients, this holonomy condition is

$$h_{-1}G_{-2} - h_{-2}G_{-1} = 0. \tag{B-6}$$

This condition ensures, in particular, that the first two components of Φ contain no logarithmic terms. Thus, defining $\zeta =: \xi + i\nu$ and rotating and rescaling if necessary, we obtain, near infinity

$$\Phi(\xi, \nu) = \left(\xi + \alpha\left(\frac{\xi}{\rho^2}, \frac{\nu}{\rho^2}\right), \nu + \beta\left(\frac{\xi}{\rho^2}, \frac{\nu}{\rho^2}\right), a + b\log(\rho) + \gamma\left(\frac{\xi}{\rho^2}, \frac{\nu}{\rho^2}\right)\right), \tag{B-7}$$

where α, β and γ are analytic functions of their arguments defined in a neighbourhood of the origin which, furthermore, vanish at this point.

We now define

$$(x, y) := (\Phi_1(\xi, \nu), \Phi_2(\xi, \nu)) \quad \text{and} \quad r^2 := x^2 + y^2. \tag{B-8}$$

That is, (x, y) is the composition of the parametrisation Φ with the projection onto the horizontal plane. Trivially, near infinity, C is the graph of some function F defined over the (x, y)-plane. We now use (B-7) to determine the asymptotic structure of this function. First, upon observing that

$$\frac{1}{\rho^2} = \frac{\xi^2}{\rho^4} + \frac{\nu^2}{\rho^4}, \tag{B-9}$$

we find that

$$\left(\frac{x}{r^2}, \frac{y}{r^2}\right) = \Psi\left(\frac{\xi}{\rho^2}, \frac{\nu}{\rho^2}\right) \tag{B-10}$$

for some analytic function Ψ, defined in a neighbourhood of the origin, such that $\Psi(0,0)=0$ and $D\Psi(0,0)=\mathrm{Id}$. Upon applying the implicit function theorem for analytic functions, we deduce that

$$F(x,y)=a+b\log(r)+\delta\left(\frac{x}{r^2},\frac{y}{r^2}\right) \tag{B-11}$$

for some analytic function δ, vanishing at the origin. In particular, with the notation of Section 1.3,

$$F(x,y)=a+c\log(r)+\mathrm{O}(r^{-(1+k)}), \tag{B-12}$$

thus confirming the first formula of Section 5.1.

It remains only to verify (5-1). However, by (B-11),

$$F(x,y)=a+c\log(r)+\frac{\phi(x,y)}{r^2}+\mathrm{O}(r^{-(2+k)}) \tag{B-13}$$

for some linear form ϕ. Let e_z denote the unit vector in the direction of the positive z-axis. Let u be any nonzero, horizontal vector. If C is symmetric under reflection in the plane spanned by e_z and u, then ϕ annihilates the line orthogonal to u. Consequently, if C is symmetric under reflection in two distinct planes of this type, then ϕ vanishes, so that

$$F(x,y)=a+c\log(r)+\mathrm{O}(r^{-(2+k)}), \tag{B-14}$$

thus confirming (5-1).

Acknowledgements

We are grateful to Knut Smoczyk for drawing our attention to this interesting problem. We are likewise grateful to Detang Zhou and Andrew Clarke for helpful conversations and their invaluable insights.

The first version of this paper was completed in January 2015 and was circulated amongst experts in the field before publication on arXiv. At roughly the same time, [Dávila et al. 2017] was also published on arXiv. That work proves a weaker result using techniques that do not extend to the case studied in this paper.

References

[Cheng and Zhou 2015] X. Cheng and D. Zhou, "Stability properties and gap theorem for complete f-minimal hypersurfaces", *Bull. Braz. Math. Soc.* (*N.S.*) **46**:2 (2015), 251–274. MR Zbl

[Cheng et al. 2014] X. Cheng, T. Mejia, and D. Zhou, "Eigenvalue estimate and compactness for closed f-minimal surfaces", *Pacific J. Math.* **271**:2 (2014), 347–367. MR Zbl

[Cheng et al. 2015a] X. Cheng, T. Mejia, and D. Zhou, "Simons-type equation for f-minimal hypersurfaces and applications", *J. Geom. Anal.* **25**:4 (2015), 2667–2686. MR Zbl

[Cheng et al. 2015b] X. Cheng, T. Mejia, and D. Zhou, "Stability and compactness for complete f-minimal surfaces", *Trans. Amer. Math. Soc.* **367**:6 (2015), 4041–4059. MR Zbl

[Clutterbuck et al. 2007] J. Clutterbuck, O. C. Schnürer, and F. Schulze, "Stability of translating solutions to mean curvature flow", *Calc. Var. Partial Differential Equations* **29**:3 (2007), 281–293. MR Zbl

[Dávila et al. 2017] J. Dávila, M. del Pino, and X. H. Nguyen, "Finite topology self-translating surfaces for the mean curvature flow in $\mathbb{R}^3$", *Adv. Math.* **320** (2017), 674–729. MR Zbl

[Gilbarg and Trudinger 1983] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, 2nd ed., Grundl. Math. Wissen. **224**, Springer, 1983. MR Zbl

[Hauswirth and Pacard 2007] L. Hauswirth and F. Pacard, "Higher genus Riemann minimal surfaces", *Invent. Math.* **169**:3 (2007), 569–620. MR Zbl

[Hoffman and Meeks 1990] D. Hoffman and W. H. Meeks, III, "Embedded minimal surfaces of finite topology", *Ann. of Math.* (2) **131**:1 (1990), 1–34. MR Zbl

[Kapouleas 1990] N. Kapouleas, "Complete constant mean curvature surfaces in Euclidean three-space", *Ann. of Math.* (2) **131**:2 (1990), 239–330. MR Zbl

[Kapouleas 1995] N. Kapouleas, "Constant mean curvature surfaces constructed by fusing Wente tori", *Invent. Math.* **119**:3 (1995), 443–518. MR Zbl

[Kapouleas 1997] N. Kapouleas, "Complete embedded minimal surfaces of finite total curvature", *J. Differential Geom.* **47**:1 (1997), 95–169. MR Zbl

[Martín et al. 2015] F. Martín, A. Savas-Halilaj, and K. Smoczyk, "On the topology of translating solitons of the mean curvature flow", *Calc. Var. Partial Differential Equations* **54**:3 (2015), 2853–2882. MR Zbl

[Mazzeo and Pacard 2001] R. Mazzeo and F. Pacard, "Constant mean curvature surfaces with Delaunay ends", *Comm. Anal. Geom.* **9**:1 (2001), 169–237. MR Zbl

[Morabito 2009] F. Morabito, "Index and nullity of the Gauss map of the Costa–Hoffman–Meeks surfaces", *Indiana Univ. Math. J.* **58**:2 (2009), 677–707. MR Zbl

[Nayatani 1993] S. Nayatani, "Morse index and Gauss maps of complete minimal surfaces in Euclidean 3-space", *Comment. Math. Helv.* **68**:4 (1993), 511–537. MR Zbl

[Pacard 2008] F. Pacard, "Connected sum constructions in geometry and nonlinear analysis", lecture notes, 2008, available at https://indico.ictp.it/event/a07155/session/25/contribution/14/material/0/0.pdf.

[Schwarz 1993] M. Schwarz, *Morse homology*, Progress in Mathematics **111**, Birkhäuser, Basel, 1993. MR Zbl

[Weber 2005] M. Weber, "Classical minimal surfaces in Euclidean space by examples: geometric and computational aspects of the Weierstrass representation", pp. 19–63 in *Global theory of minimal surfaces*, edited by D. Hoffman, Clay Math. Proc. **2**, Amer. Math. Soc., Providence, RI, 2005. MR Zbl

Received 17 Dec 2019. Revised 13 Aug 2022. Accepted 23 Sep 2022.

GRAHAM SMITH: grahamandrewsmith@gmail.com
Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil

ANALYSIS AND PDE
Vol. 17 (2024), No. 4, pp. 1237–1259
DOI: 10.2140/apde.2024.17.1237

HAUSDORFF MEASURE BOUNDS FOR NODAL SETS OF STEKLOV EIGENFUNCTIONS

STEFANO DECIO

We study nodal sets of Steklov eigenfunctions in a bounded domain with $\mathcal{C}^2$ boundary. Our first result is a lower bound for the Hausdorff measure of the nodal set: we show that, for u_λ a Steklov eigenfunction with eigenvalue $\lambda \neq 0$, we have $\mathcal{H}^{d-1}(\{u_\lambda = 0\}) \geq c_\Omega$, where c_Ω is independent of λ. We also prove an almost sharp upper bound, namely, $\mathcal{H}^{d-1}(\{u_\lambda = 0\}) \leq C_\Omega \lambda \log(\lambda + e)$.

1. Introduction

Let Ω a bounded domain in $\mathbb{R}^d$, where $d \geq 2$. A Steklov eigenfunction $u_\lambda \in H^1(\Omega)$ is a solution of

$$\begin{cases} \Delta u_\lambda = 0 & \text{in } \Omega, \\ \partial_\nu u_\lambda = \lambda u_\lambda & \text{on } \partial\Omega. \end{cases} \tag{1}$$

Here and throughout the paper we denote by ∂_ν the outward normal derivative. A number λ for which a solution to (1) exists is called a Steklov eigenvalue, and it is well known that Steklov eigenvalues form a discrete sequence accumulating to infinity. Moreover, Steklov eigenvalues coincide with the eigenvalues of the Dirichlet-to-Neumann operator, which is the operator that maps a function on $\partial\Omega$ to the normal derivative of its harmonic extension in Ω, and a Steklov eigenfunction restricted to $\partial\Omega$ is an eigenfunction of the Dirichlet-to-Neumann operator. For a survey on the Steklov problem outlining many results and open questions see [Girouard and Polterovich 2017].

Inspired by a famous conjecture of Yau on the Hausdorff measure of nodal sets of Laplace eigenfunctions, an analogous question has been asked for nodal sets of Steklov eigenfunctions (it is stated explicitly in [Girouard and Polterovich 2017], for example); the conjecture can be formulated both for interior and boundary nodal sets. For the interior nodal set, the question is as follows:

- Is it true that there exist positive constants c and C, depending only on Ω, such that

$$c\lambda \leq \mathcal{H}^{d-1}(\{u_\lambda = 0\}) \leq C\lambda? \tag{2}$$

Similarly, for the boundary nodal set (which is the nodal set of an eigenfunction of the Dirichlet-to-Neumann operator) one can ask:

- Is it true that there exist positive constants c' and C', depending only on Ω, such that

$$c'\lambda \leq \mathcal{H}^{d-2}(\{u_\lambda = 0\} \cap \partial\Omega) \leq C'\lambda? \tag{3}$$

MSC2020: 35J15, 58J50.
Keywords: Steklov eigenfunctions, nodal set, frequency function.

Here we do not deal with question (3) and just note that the upper bound was proved in [Zelditch 2015] when $\partial\Omega$ is real-analytic. About question (2), a polynomial upper bound was proved in [Georgiev and Roy-Fortin 2019], following the corresponding polynomial upper bound in the Laplace–Beltrami eigenfunction case proved in [Logunov 2018a]. On real-analytic surfaces (that is, real-analytic metric in the interior and real-analytic boundary), the full conjecture (2) was established in [Polterovich et al. 2019]. Again in the real-analytic category, the upper bound was recently obtained in any dimension in [Zhu 2020]. Concerning lower bounds, as far as we know, the best result was contained in [Sogge et al. 2016], where the bound $\mathcal{H}^{d-1}(\{u_\lambda = 0\}) \geq c\lambda^{(2-d)/2}$ is obtained for Ω a domain with $\mathcal{C}^\infty$ boundary (actually, a smooth Riemannian manifold with smooth boundary). The first contribution of the present article is an improvement on the lower bound; we show that the Hausdorff measure of the interior nodal set is bounded below by a constant independent of λ (so the result is really an improvement over [Sogge et al. 2016] if $d \geq 3$).

Theorem 1. *Let Ω be a bounded domain in $\mathbb{R}^d$ with $\mathcal{C}^2$-smooth boundary, and let u_λ be a solution of (1) in Ω, $\lambda \neq 0$. Then there exists a constant $c_\Omega > 0$ independent of λ such that*

$$\mathcal{H}^{d-1}(\{u_\lambda = 0\}) \geq c_\Omega. \tag{4}$$

In the previous work [Decio 2022] we established a density property of the zero set near the boundary, under weaker hypothesis on the boundary regularity: we transcribe the result below.

Theorem A [Decio 2022]. *Let Ω be a Lipschitz domain in $\mathbb{R}^d$, $d \geq 2$, and let u_λ be a solution of (1), where we assume $\lambda \neq 0$. There exists a constant $C = C(\Omega)$ such that*

$$\{u_\lambda = 0\} \cap B \neq \varnothing \tag{5}$$

for any ball B in $\mathbb{R}^d$ of radius C/λ centered at a point in $\partial\Omega$.

The proof of Theorem 1 involves a combination of Theorem A and the recent breakthrough by Logunov [2018b] on Yau's conjecture. We cannot apply the results of [Logunov 2018b] directly and have to do some work to modify the necessary arguments. The fact that we are one power of λ away from the optimal result is a consequence of the deficiency of the density result, which we can only prove very close to the boundary, and not of the arguments in [Logunov 2018b].

Remark. It will be apparent from the proof that Theorem 1 extends without much difficulty to the case of manifolds equipped with a $\mathcal{C}^2$-smooth Riemannian metric and $\mathcal{C}^2$ boundary.

The conjectured upper bound in (2) would be sharp, as the example of a ball shows; the second main contribution of this article is an almost sharp upper bound for Euclidean domains with $\mathcal{C}^2$ boundary.

Theorem 2. *Let Ω be a bounded domain in $\mathbb{R}^d$ with $\mathcal{C}^2$-smooth boundary, and let u_λ be a solution of (1) in Ω. Then there is a constant $C_\Omega > 0$ independent of λ such that*

$$\mathcal{H}^{d-1}(\{u_\lambda = 0\}) \leq C_\Omega \lambda \log(\lambda + e). \tag{6}$$

Remark. The proof of Theorem 2 uses the sharp bounds of Donnelly and Fefferman [1988] in the interior of the domain and a multiscale induction argument at the boundary, which is based on a version of the hyperplane lemma of [Logunov 2018a; Logunov et al. 2021]. While, as remarked above, the proof of the lower bound can be extended almost *verbatim* to smooth Riemannian manifolds with boundary, for Theorem 2 we rely heavily on the fact that Ω is a Euclidean domain, or at least we have to require that the metric inside Ω is real analytic; this is because the results of [Donnelly and Fefferman 1988] require real analyticity. Our theorem lies in between previous results on the upper bound: the multiscale argument at the boundary allows for $\mathcal{C}^2$-regularity of the boundary only, as opposed to real analyticity as in the aforementioned paper [Zhu 2020]; on the other hand, if the metric inside is assumed to be only $\mathcal{C}^2$ (or $\mathcal{C}^\infty$), the best result attainable with these methods is still the polynomial upper bound of [Georgiev and Roy-Fortin 2019].

Plan of the paper. We prove Theorem 1 in Sections 2 and 3; in Section 2 we discuss a procedure for extending a Steklov eigenfunction across the boundary, which gives rise to an auxiliary equation for which a statement very similar to Logunov's theorem [2018b] holds (see Theorem 3), and we use this together with Theorem A to prove the lower bound. Section 3 is quite long and contains the proof of Theorem 3, which requires us to review Logunov's argument carefully and use a combination of classical elliptic estimates and frequency function techniques. Section 4 is dedicated to the proof of Theorem 2.

2. Lower bound on nodal sets

Here we deduce Theorem 1 using Theorem A and ideas stemming from Logunov's solution [2018b] of a conjecture of Nadirashvili on nodal sets of harmonic functions. In order to do this, we transform a solution to (1) into a solution of an elliptic equation in the interior of a domain. To the best of our knowledge, this idea was introduced first in [Bellová and Lin 2015] and then also applied successfully in [Georgiev and Roy-Fortin 2019; Zhu 2015].

We now describe this extension procedure, which requires $\partial\Omega$ to be of class $\mathcal{C}^2$; we follow [Bellová and Lin 2015] very closely. There is a $\delta > 0$ such that the map $\partial\Omega \times (-\delta, \delta) \ni (y, t) \to y + t\nu(y)$ is one-to-one onto a neighborhood of $\partial\Omega$ in $\mathbb{R}^d$. We set $d(x) = \operatorname{dist}(x, \partial\Omega)$, and for $\rho \le \delta$ we define $\Omega_\rho = \{x \in \Omega : d(x) < \rho\}$ and $\Omega'_\rho = \{x \in \mathbb{R}^d : d(x) < \rho\} \setminus \overline{\Omega}$. Let now u_λ be a solution of (1), and for $x \in \Omega_\delta \cup \partial\Omega$ define

$$v(x) = u_\lambda(x)\exp(\lambda d(x)); \tag{7}$$

an easy computation shows that v satisfies

$$\begin{cases} \operatorname{div}(A\nabla v) + b(x)\cdot\nabla v + c(x)v = 0 & \text{in } \Omega_\delta, \\ \partial_\nu v = 0 & \text{on } \partial\Omega, \end{cases}$$

where $A = I$, $b = -2\lambda\nabla d$ and $c = \lambda^2 - \lambda\Delta d$. Consider now the reflection map $\Psi : \Omega_\delta \to \Omega'_\delta$ given by $\Psi(y + t\nu(y)) = y - t\nu(y)$, where $y \in \partial\Omega$; since v satisfies a Neumann boundary condition on $\partial\Omega$, we can extend it "evenly" across the boundary, i.e., set $v(\Psi(x)) = v(x)$ for $x \in \Omega_\delta$. Write $\Psi(x) = x'$. Another easy computation shows that on Ω'_δ the extended function (which we still call v) satisfies the equation

$$\operatorname{div}(\tilde{A}\nabla v) + \tilde{b}\cdot\nabla v + \tilde{c}v = 0,$$

where

$$\tilde{A}(x') = \nabla\Psi(x)(\nabla\Psi(x))^T, \quad \tilde{b}^i(x') = -\sum_j \partial_{x'_j}\tilde{a}^{ij}(x') + \Delta\Psi^i(x) + \nabla\Psi^i(x)\cdot b(x), \quad \tilde{c}(x') = c(x).$$

Consider now $D = \Omega_\delta \cup \partial\Omega \cup \Omega'_\delta$; we abuse notation and denote by A, b and c the functions that are equal to the previous A, b and c in Ω_δ and equal to $\tilde{A}$, $\tilde{b}$ and $\tilde{c}$ in Ω'_δ. In [Bellová and Lin 2015] it is shown that A is Lipschitz across $\partial\Omega$ with Lipschitz constant depending only on Ω, and A is uniformly positive definite, again with constant depending only on Ω. Pasting together the pieces, one obtains that v is a strong solution of the uniformly elliptic equation

$$\operatorname{div}(A\nabla v) + b\cdot\nabla v + cv = 0 \tag{8}$$

in D, with A Lipschitz, $\|A\|_{L^\infty(D)} \le C$, $\|b\|_{L^\infty(D)} \le C\lambda$ and $\|c\|_{L^\infty(D)} \le C\lambda^2$.

We want to study (8) at wavelength scale. In order to deal with its zero set we use the theorem below, which is just an extension to more general equations of the aforementioned theorem of Logunov on harmonic functions [2018b]; its proof, which merely consists of a tedious but necessary verification that Logunov's argument carries over in this slightly more general setting, is relegated to the next section. We warn the reader that below and in the rest of the paper we do not explicitly indicate dependence of the constants on the dimension.

Theorem 3. *Consider a strong solution of the equation*

$$Lu = \operatorname{div}(A\nabla u) + b\cdot\nabla u + cu = 0 \tag{9}$$

in $B = B(0,1) \subset \mathbb{R}^d$, with the following assumptions on the coefficients:

(i) *A is a uniformly positive definite matrix; that is, $A(x)\xi\cdot\xi \ge \alpha|\xi|^2$ for any $\xi \in \mathbb{R}^d$.*

(ii) *A is Lipschitz; that is, $\sum_{i,j} |a^{ij}(x) - a^{ij}(y)| \le \gamma|x-y|$.*

(iii) $\sum_{i,j} \|a^{ij}\|_{L^\infty(B)} + \sum_i \|b^i\|_{L^\infty(B)} \le K$.

(iv) *$c \ge 0$ and $\|c\|_{L^\infty(B)} \le \varepsilon_0$, where ε_0 is a small enough constant depending on α, γ, K.*

Then there exist $r_0 = r_0(\alpha,\gamma,K) < 1$ and $c_0 = c_0(\alpha,\gamma,K)$ such that, for any solution u of (9) and any ball $B(x,r) \subset B(0,r_0)$ for which $u(x) = 0$, we have the lower measure bound

$$\mathcal{H}^{d-1}(\{u=0\}\cap B(x,r)) \ge c_0 r^{d-1}. \tag{10}$$

Assume now that λ is large enough depending on Ω and consider a ball $B(x_0, \varepsilon/\lambda) \subset D$, where ε is a small enough constant, with smallness depending only on Ω. We set $v_{x_0,\lambda}(x) = v(x_0 + \varepsilon x/\lambda)$ for $x \in B = B(0,1)$; note that $v_{x_0,\lambda}$ satisfies the equation

$$\operatorname{div}(A_{x_0,\lambda}\nabla v_{x_0,\lambda}) + b_{x_0,\lambda}\cdot\nabla v_{x_0,\lambda} + c_{x_0,\lambda}v_{x_0,\lambda} = 0, \tag{11}$$

where the ellipticity constant of $A_{x_0,\lambda}$ is the same as that of A and the Lipschitz constant is the same if not better, and the coefficients satisfy $\|A_{x_0,\lambda}\|_{L^\infty(B)} \le C$, $\|b_{x_0,\lambda}\|_{L^\infty(B)} \le C\varepsilon$ and $\|c_{x_0,\lambda}\|_{L^\infty(B)} \le C\varepsilon^2$. Note that if λ is large enough then $c_{x_0,\lambda} \ge 0$. If we then take ε small enough, $v_{x_0,\lambda}$ satisfies (9) and assumptions (i)–(iv) with constants α, γ, K depending only on Ω. By Theorem A, any ball centered at $\partial\Omega$ of radius C/λ contains a zero of the Steklov eigenfunction u_λ and hence of v. We can reduce the radius of the balls and take a maximal disjoint subcollection of balls $B(x_i, C_1/\lambda) \subset D,\ x_i \in \overline{\Omega}$, such that $v(x_i) = 0$ and consider the corresponding rescaled functions $v_{x_i,\lambda}$; we can assume that $C_1 < r_0$, so that by Theorem 3 we obtain

$$\mathcal{H}^{d-1}(\{v_{x_i,\lambda} = 0\} \cap B(0, C_1)) \geq cC_1^{d-1}. \tag{12}$$

Note also that

$$\mathcal{H}^{d-1}\left(\{u_\lambda = 0\} \cap B\left(x_i, \frac{C_1}{\lambda}\right) \cap \Omega\right) \sim \mathcal{H}^{d-1}\left(\{v = 0\} \cap B\left(x_i, \frac{C_1}{\lambda}\right)\right)$$
$$\sim \varepsilon^{d-1}\lambda^{1-d}\mathcal{H}^{d-1}(\{v_{x_i,\lambda} = 0\} \cap B(0, C_1)) \geq \widetilde{C}\lambda^{1-d},$$

where $\widetilde{C}$ depends on Ω only. Since there are $\sim \lambda^{d-1}$ such balls $B(x_i, C_1/\lambda)$, we obtain

$$\mathcal{H}^{d-1}(\{u_\lambda = 0\}) \geq c_\Omega,$$

and Theorem 1 is proved.

Remark. If one could improve the result of Theorem A by showing that every ball of radius C/λ centered at any point in a corona of fixed (independent of λ) size around the boundary contains a zero of u_λ, the optimal lower bound $\mathcal{H}^{d-1}(\{u_\lambda = 0\}) \gtrsim \lambda$ would follow immediately by the preceding argument (actually more easily, since one could directly apply Logunov's result without the need to go through Theorem 3).

3. Proof of Theorem 3

This entire long section is dedicated to the proof of Theorem 3. We follow essentially the arguments of [Logunov 2018b], which carry through in this setting with few changes; the difference is that we have to use more general elliptic estimates, such as a weaker form of the maximum principle, and a frequency function that takes into account the lower-order terms in the equation. In Sections 3.1 and 3.2 we introduce the main tools we need in the proof, namely, classical elliptic estimates and the monotonicity of the frequency function. Section 3.3 will serve as a break from technicalities: here we try to convey an idea of the scheme of the proof to the reader. Sections 3.4–3.8 contain the actual body of the proof with full details.

Throughout the section we consider the operator L defined by (9) satisfying conditions (i)–(iv). It will be convenient to denote by $L_1 = L - cI$ the operator without the zeroth-order term.

3.1. ***Elliptic estimates.*** We first recall some standard elliptic estimates for L, paraphrasing the results in [Gilbarg and Trudinger 1983] in our notation. Note that whenever we consider a bounded domain we can assume for our purposes that it is contained in the unit ball, so we can ignore the dependency of the constants on the diameter of Ω and on the radius of balls contained in Ω. We start with the weak maximum principle.

Theorem 4 [Gilbarg and Trudinger 1983, Theorem 9.1]. *Let $L_1 u \geq -\delta$ in a bounded domain Ω. Then*

$$\sup_{\Omega} u \leq \sup_{\partial\Omega} u^+ + C|\delta|,$$

where $C = C(\alpha, \gamma, K)$.

Corollary 5. *Let $Lu = 0$ in a bounded domain Ω, with ε_0 in (iv) small enough. Then*

$$\sup_{\Omega} u \leq 2 \sup_{\partial\Omega} u^+. \tag{13}$$

Proof. We can assume $\sup_\Omega u \geq 0$. Since $Lu = 0$, we have $L_1 u = -cu \geq -\varepsilon_0 \sup_\Omega u$ using assumption (iv). By Theorem 4, then $\sup_\Omega u \leq \sup_{\partial\Omega} u^+ + C\varepsilon_0 \sup_\Omega u$, and the corollary follows as soon as $C\varepsilon_0 \leq \frac{1}{2}$. □

The next theorem is a local pointwise estimate for subsolutions.

Theorem 6 [Gilbarg and Trudinger 1983, Theorem 9.20]. *Let $Lu \geq -\delta$ in Ω. Then for any ball $B(x, 2R) \subset \Omega$ and any $p > 0$ we have*

$$\sup_{B(x,R)} u \leq C_1 \left\{ \fint_{B(x,2R)} (u^+)^p \right\}^{1/p} + C_2|\delta|, \tag{14}$$

where C_1 and C_2 depend on α, K and p.

Remark. In Theorem 9.20 in [Gilbarg and Trudinger 1983], the constants depend on R. However, they get worse as R increases and improve as R decreases; in this work we will only be concerned with small R, so that we can ignore the dependency on it.

We now come to the weak Harnack inequality and then the full Harnack inequality.

Theorem 7 [Gilbarg and Trudinger 1983, Theorem 9.22]. *Let $Lu \leq \delta$ in Ω, and suppose that u is nonnegative in a ball $B(x, 2R) \subset \Omega$. Then*

$$\left\{ \fint_{B(x,R)} u^p \right\}^{1/p} \leq C(\inf_{B(x,R)} u + |\delta|), \tag{15}$$

where p and C are positive numbers depending on α and K.

Theorem 8 [Gilbarg and Trudinger 1983, Corollary 9.25]. *Let $Lu = 0$ in Ω, and suppose that u is nonnegative in a ball $B(x, 2R) \subset \Omega$. Then*

$$\sup_{B(x,R)} u \leq C \inf_{B(x,R)} u, \tag{16}$$

where $C = C(\alpha, K)$.

Corollary 9. *Let $Lu = 0$ in Ω. If $u(x_0) \geq 0$ and $B(x_0, R) \subset \Omega$, then the inequality*

$$\sup_{B(x_0,2R/3)} |u| \leq C \sup_{B(x_0,R)} u \tag{17}$$

holds for $C = C(\alpha, K)$.

Proof. Call $M = \sup_{B(x_0,R)} u$ and consider the function $h = M - u$, which is nonnegative in $B(x_0, R)$. Note that $Lh = cM$, so that $|Lh| \leq \varepsilon M$. By applying to h Theorem 6 and then Theorem 7 with $\delta = \varepsilon M$, one gets that

$$\sup_{B(x_0,2R/3)} (M-u) \leq C_1 \left\{ \fint_{B(x_0,3R/4)} u^p \right\}^{1/p} + C_2\varepsilon M \leq C_3 \inf_{B(x_0,3R/4)} (M-u) + C_4\varepsilon M \leq C_5 M,$$

where the last inequality holds because $u(x_0) \geq 0$. Hence we obtain $\sup_{B(x_0,2R/3)}(-u) \leq CM$. Since clearly we have that $\sup_{B(x_0,2R/3)} u \leq M$, the corollary is proved. □

3.2. *Frequency function and doubling index.* The frequency function, which as far as we know was used first by Almgren and then subsequently developed in the works of Garofalo and Lin [1986; 1987], is a powerful tool in the study of unique continuation and zero sets of elliptic PDEs. We are now going to define it for operators of the form (9) and state some of its properties, following mainly [Garofalo and Lin 1987; Han and Lin].

Let $u \in W^{1,2}_{\text{loc}}(B)$ be a solution of (9). In [Garofalo and Lin 1987] and [Han and Lin] a metric $g(x) = \sum_{i,j} g_{ij}(x)\, dx_i \otimes dx_j$ is introduced in the following way: let first

$$\bar{g}_{ij}(x) = a^{ij}(x)(\det A)^{1/(d-2)},$$

where, as customary, a^{ij} denote the entries of the matrix A^{-1}. To define $\bar{g}_{ij}$ we assume here $d \geq 3$; if $d = 2$, we can just add a "mute" variable. Next, one defines

$$r(x)^2 = \sum_{i,j} \bar{g}_{ij}(0) x_i x_j \quad \text{and} \quad \eta(x) = \sum_{k,l} \bar{g}^{kl}(x) \frac{\partial r}{\partial x_k}(x) \frac{\partial r}{\partial x_l}(x).$$

Finally, one sets

$$g_{ij}(x) = \eta(x) \bar{g}_{ij}(x).$$

Note that η is a positive Lipschitz function with Lipschitz constant depending on α, γ and K. Let G be the matrix (g_{ij}) and define $|g| = \det(G)$. We can now write (9) as

$$\operatorname{div}_g(\mu(x)\nabla_g u) + b_g(x) \cdot \nabla_g u + c_g(x) u = 0,$$

where $\mu = \eta^{-(d-2)/2}$ is a Lipschitz function in B with $C_1 \leq \mu(x) \leq C_2$, $b_g = Gb/\sqrt{|g|}$ and $c_g = c/\sqrt{|g|}$. Note that, since $|g|^{-1/2}$ is a Lipschitz function bounded above and below by constants depending on α, γ and K only, b_g and c_g satisfy analogous bounds to b and c in (9). The following quantities are then introduced, where the integrals are with respect to the measure induced by the metric g:

$$H(x,r) = \int_{\partial B(x,r)} \mu u^2, \quad D(x,r) = \int_{B(x,r)} \mu |\nabla_g u|^2, \quad I(x,r) = \int_{B(x,r)} \mu |\nabla_g u|^2 + u b_g \cdot \nabla_g u + c_g u^2.$$

The frequency function is finally defined as

$$\beta(x,r) = \frac{2r I(x,r)}{H(x,r)}. \tag{18}$$

Compared with the definition in [Garofalo and Lin 1987] and [Han and Lin] there is an extra factor of 2 for aesthetic reasons in later formulas. More often than not, we will forget about the point x and only write the dependance on the radius r. The key property of the frequency function is the following almost monotonicity:

Theorem 10. *There are constants r_0, c_1 and c_2 depending on α, γ and K such that*

$$\beta(x,r) \leq c_1 + c_2 \beta(x, r_0) \tag{19}$$

for $r \in (0, r_0)$. Moreover, c_2 can be chosen to be $1 + \varepsilon$ for any $\varepsilon > 0$ if $r_0 = r_0(\varepsilon)$ is small enough.

Remark. The statement of Theorem 10 is implicit in [Garofalo and Lin 1987], and the proof is contained there; in [Han and Lin] the theorem is stated as it is here, and the proof given is essentially the one of [Garofalo and Lin 1987]. The second assertion is not explicitly stated in [Garofalo and Lin 1987] or [Han and Lin] and needs some justification. In both papers, the strategy to prove the theorem is the following: one defines $\Omega_{r_0} = \{r \in (0, r_0) : \beta(r) > \max(1, \beta(r_0))\}$ and proves that it is an open subset of $\mathbb{R}$ and therefore it can be decomposed as $\Omega_{r_0} = \bigcup_{j=1}^{+\infty}(a_j, b_j)$ with a_j and b_j not belonging to Ω_{r_0}; it is then showed that $\beta'(r)/\beta(r) \geq -C$ for any $r \in \Omega_{r_0}$. By integration, one has that $\beta(r) \leq \beta(b_j)\exp(C(b_j - r))$ for any $r \in (a_j, b_j)$. Since $b_j \notin \Omega_{r_0}$, this implies that the constant c_2 can be chosen to be $\exp(Cr_0)$, which is close to 1 if r_0 is small.

In the course of the proof of Theorem 10 in [Garofalo and Lin 1987] and [Han and Lin] the differentiation formula

$$H'(r) = \left(\frac{d-1}{r} + O(1)\right)H(r) + 2I(r)$$

is obtained; the formula can be rewritten as

$$\frac{d}{dr}\left(\log \frac{H(r)}{r^{d-1}}\right) = O(1) + \frac{\beta(r)}{r}. \tag{20}$$

The next statement is an immediate consequence of this formula.

Proposition 11. *There is a constant C depending on α, γ and K such that the function $e^{Cr}H(r)/r^{d-1}$ is increasing for $r \in (0, r_0)$.*

From (20) and almost monotonicity (19), by integration one obtains the following:

Proposition 12. *The two-sided inequality*

$$c\left(\frac{r_2}{r_1}\right)^{c_2^{-1}\beta(r_1) - c_3} \leq \frac{H(r_2)}{H(r_1)} \leq C\left(\frac{r_2}{r_1}\right)^{c_2\beta(r_2) + c_3} \tag{21}$$

holds, where again c_2 can be chosen to be $1 + \varepsilon$ if r_0 is small enough.

From now on we denote with letters c, C, $c_1, \ldots$ constants which may vary from line to line and that depend only on α, γ and K without explicitly saying so every time. Additional dependencies will be indicated. We now define a quantity related to the frequency function: the doubling index.

Definition 13. For $B(x, 2r) \subset B$, the doubling index $\mathcal{N}(x, r)$ is defined by

$$2^{\mathcal{N}(x,r)} = \frac{\sup_{B(x,2r)} |u|}{\sup_{B(x,r)} |u|}. \tag{22}$$

The doubling index and the frequency function are comparable in the following sense:

Lemma 14. *Let $\varepsilon > 0$ be sufficiently small, and let r_0 be small enough that the constant c_2 in (21) is $1 + \varepsilon$; then, for $4r < r_0$,*

$$\beta(x, r(1+\varepsilon))(1 - 100\varepsilon) - c \leq \mathcal{N}(x, r) \leq \beta(x, 2r(1+\varepsilon))(1 + 100\varepsilon) + c.$$

The proof of Lemma 14 is an easy computation using the elliptic estimate (14), Proposition 11 and inequality (21); in fact, by (14),

$$\sup_{B(x,r)} |u|^2 \le C_\varepsilon \fint_{B(x,(1+\varepsilon)r)} |u|^2,$$

and further

$$\fint_{B(x,(1+\varepsilon)r)} |u|^2 \le C \frac{H((1+\varepsilon)r)}{r^{d-1}}$$

by integration and Proposition 11. From here on the computation is identical to the one in [Logunov 2018a, Lemma 7.1]. Using this, one can derive a scaling property for the doubling index; see [Logunov 2018a, Lemmas 7.2 and 7.3] for details of the computation.

Lemma 15. *Given any $\varepsilon \in (0,1)$, there exist $r_0(\varepsilon) > 0$ and $C(\varepsilon) > 0$ such that, for $u \in W^{1,2}(B)$ a solution of (9) and any $0 < 2r_1 \le r_2 \le r_0$, we have*

$$\left(\frac{r_2}{r_1}\right)^{\mathcal{N}(x,r_1)(1-\varepsilon)-C} \le \frac{\sup_{B(x,r_2)} |u|}{\sup_{B(x,r_1)} |u|} \le \left(\frac{r_2}{r_1}\right)^{\mathcal{N}(x,r_2)(1+\varepsilon)+C}. \tag{23}$$

As a consequence, the doubling index is also almost monotonic in the sense that

$$\mathcal{N}(x,r_1)(1-\varepsilon) - C \le \mathcal{N}(x,r_2)(1+\varepsilon) + C.$$

3.3. ***An informal outline of the proof.*** We include here a brief discussion of the scheme of the proof avoiding details and technicalities; the latter are all included in the next subsections. Let us first note that in dimension 2 Theorem 3 is an easy consequence of the weak maximum principle (Corollary 5): if u vanishes at the center of a ball, the weak maximum principle tells us that there can be no small loops of zeros containing the center and therefore the nodal component containing the center must exit the ball, implying that its length must be greater than the diameter of the ball.

In higher dimensions, this simple argument does not give any lower measure bound because a priori the nodal set could be a very thin tube crossing the ball. However, a slightly more sophisticated argument, still using essentially only the maximum principle, does give a nonoptimal lower bound: we prove in Proposition 16 that if $u(x) = 0$,

$$\mathcal{H}^{d-1}(\{u=0\} \cap B(x,r)) \ge c r^{d-1} N^{2-d},$$

where N is an upper bound for the doubling index $\mathcal{N}\big(x, \tfrac{1}{2}r\big)$. Note that when $d = 2$ this is already optimal, as it should be. If $d \ge 3$, this naive lower bound gets worse as the doubling index gets larger. This however contradicts intuition, since we are dealing with solutions of elliptic PDEs: if the doubling index is large, meaning that there is strong growth of u, then there should be many zeros. This suggests that one could use induction on N to promote the naive lower bound to the optimal one. The key to achieving this is Proposition 23, which shows that if the doubling index is comparable to $N \gg 1$ in balls of radii $\tfrac{1}{4}r$ to r (we call this "stable growth", see Definition 22), there are many zeros in the ball of radius r; more precisely, there are at least $[\sqrt{N}]^{d-1} f(N)$, with $f(N) \to \infty$ as $N \to \infty$, disjoint balls of radius $r/\sqrt{N}$

such that u vanishes at the center. The fact that $f(N)$ grows with N essentially shows that indeed there are more zeros as the doubling index increases, and it is needed to close the induction in Section 3.8.

The proof of Proposition 23 uses crucially Theorem 19, which tells us that if a cube is partitioned into some large number B^d of subcubes, the number of subcubes which have doubling indices dropping by an amount increasing with B compared to the doubling index of the original cube form the vast majority of the subcubes. The argument goes as follows: since the doubling index is comparable to N on scales $\frac{1}{4}r$ to r, we can assume that in the ball of radius $\frac{1}{4}r$, $|u| \leq 1$, while in the ball of radius $\frac{1}{2}r$, $|u| \geq 2^{cN}$. We then connect points where u is small to points where u is large by many chains of cubes (called "tunnels" later): since there is considerable growth of u from one endpoint of the tunnel to the other, the Harnack inequality tells us that there must be zeros and the growth happens in the cubes with zeros; an application of Theorem 19 gives us that most of the cubes in the tunnel have doubling index much smaller than N, so that the growth from one endpoint to the other cannot be realized in very few cubes, and hence each tunnel must have many cubes with zeros. The formal proof is a matter of quantifying what "small", "large", "few" and "many" mean.

The only issue remaining is ensuring that there are balls of stable growth: this is done in Claim 3, and the proof uses the estimates in Section 3.6 which are consequences of the almost monotonicity of the frequency function.

Let us emphasize once again that the proof scheme described above is due to Aleksandr Logunov, and it appeared first in [Logunov 2018b]. In our case we have to adapt it to elliptic equations with lower-order terms, but the more general estimates that we need are collected above in Sections 3.1 and 3.2, and using those estimates the proof runs in the same way as for harmonic functions.

3.4. ***Local asymmetry.*** We now derive a lower estimate for the relative volume of the set $\{u > 0\}$ in balls centered at zeros of u, and consequently a nonoptimal lower estimate for the measure of the zero set. The estimate and the proof are analogous to the Laplace–Beltrami eigenfunctions case, for which see, for example, [Logunov and Malinnikova 2018; Mangoubi 2008]. For the reader's convenience, we reproduce here essentially the same proof as [Logunov and Malinnikova 2018].

Proposition 16. *Let $B(x, r) \subset B$ and u be a solution of (9) such that $u(x) = 0$. Suppose that $\mathcal{N}\big(x, \frac{1}{2}r\big) \leq N$, where N is a positive integer. Then the lower measure bound*

$$\mathcal{H}^{d-1}(\{u = 0\} \cap B(x, r)) \geq c r^{d-1} N^{2-d} \tag{24}$$

holds for some $c > 0$.

Proof. For notational simplicity we assume $x = 0$ and write $B_r = B(0, r)$. We can also safely assume that $N \geq 4$, say. Note that by (17) and (13) we have $\sup_{B_{r/2}} |u| \leq C \max_{\partial B_{3r/4}} u$, so that

$$\frac{\max_{\partial B_r} u}{\max_{\partial B_{3r/4}} u} \leq C_1 \frac{\sup_{B_r} |u|}{\sup_{B_{r/2}} |u|} \leq C_1 2^N.$$

Let now $r_j = r\big(\frac{3}{4} + j/(4N)\big)$ for $j = 0, 1, \ldots, N$, and consider the concentric spheres $S_j = \{|x| = r_j\}$. Write $m_j^+ = \max_{S_j} u$ and $m_j^- = \min_{S_j} u$. From the weak maximum principle (13) (applied to u as well as $-u$), it follows that

$$m_j^+ > 0, \quad m_j^- < 0, \quad m_j^+ \leq 2m_{j+1}^+ \quad \text{and} \quad |m_j^-| \leq 2|m_{j+1}^-|.$$

For $j = 0, 1, \ldots, N-1$, define

$$\tau_j^+ = m_{j+1}^+/m_j^+ \quad \text{and} \quad \tau_j^- = |m_{j+1}^-|/|m_j^-|;$$

from the above, $\tau_j^{+/-} \geq \frac{1}{2}$. Moreover, we have that

$$\tau_0^+ \cdots \tau_{N-1}^+ = \frac{\max_{\partial B_r} u}{\max_{\partial B_{3r/4}} u} \leq C_1 2^N,$$

so at most $\frac{1}{4}N$, say, of the τ_j^+ are greater than some C independent of N. The same holds for the τ_j^-, so that for at least $\frac{1}{2}N$ indices there holds $m_{k+1}^+ \leq Cm_k^+$ and $|m_{k+1}^-| \leq C|m_k^-|$. Consider each such k and let $x_0 \in S_k$ be such that $u(x_0) = m_k^+$. Denote by b the ball centered at x_0 of radius $r/(8N)$; then by (13) and the choice of k,

$$\sup_b u \leq \sup_{\{|x| \leq r_{k+1}\}} u \leq 2m_{k+1}^+ \leq Cm_k^+.$$

Applying (17), we then get that $\sup_{b/2} |u| \leq Cm_k^+$. We now use this last inequality and the elliptic gradient estimate (see, for instance, [Gilbarg and Trudinger 1983, Theorem 8.32])

$$\sup_{B(y,s/2)} |\nabla u| \leq \left(\frac{C}{s}\right) \sup_{B(y,s)} |u|$$

for $y = x_0$ and $s = r/(16N)$ to get, for $x \in B(x_0, \theta r/N)$ with θ a sufficiently small number,

$$u(x) \geq u(x_0) - |x - x_0| \sup_{b/4} |\nabla u| \geq m_k^+ - C\theta m_k^+ \geq 0.$$

We thus found a ball centered on S_k of radius $\theta r/N$ where u is positive, call it b_+. Replace now u with $-u$, which is also a solution of (9): repeating the argument above with m_k^- and τ_k^- instead of m_k^+ and τ_k^+ gives us a ball centered on S_k of radius $\theta r/N$ where u is negative, call it b_-. Now consider the sections of the two balls with hyperplanes through the origin that contain the center of the balls: any path within the annulus $\{r_{k-1} < |x| < r_{k+1}\}$ that connects these two sections contains a zero of u, since u is positive on b_+ and negative on b_-. This implies that the measure of the zero set is greater than the measure of the section of the balls, that is to say,

$$\mathcal{H}^{d-1}(\{x : r_{k-1} < |x| < r_{k+1}, u(x) = 0\}) \geq c\left(\frac{r}{N}\right)^{d-1}.$$

The above holds for all indices k for which $m_{k+1}^+ \leq Cm_k^+$ and $|m_{k+1}^-| \leq C|m_k^-|$, and recall that there are at least $\frac{1}{2}N$ such indices. Summing the inequality above over those indices, we see that (24) holds. ☐

Remark. Note that the argument above also shows that

$$\frac{\text{Vol}(\{u > 0\} \cap B(x,r))}{\text{Vol}(B(x,r))} \geq \frac{c}{N^{d-1}}$$

if $u(x) = 0$, which is analogous to the best known lower bound (when $d \geq 3$) for the local asymmetry of Laplace eigenfunctions [Mangoubi 2008].

3.5. *Counting doubling indices.* We now recall some very useful results from [Logunov 2018a; 2018b; Logunov and Malinnikova 2018] that allow us to find many small cubes with better doubling index than the original ball (or cube). The proofs are combinatorial in nature. First we define a version of the doubling index for cubes, which are more suitable for partitioning than balls. Given a cube Q and a solution u of (9), we define the doubling index $N(Q)$ as

$$N(Q) = \sup_{\{x\in Q,\, r<\operatorname{diam}(Q)\}} \log \frac{\sup_{B(x,10dr)} |u|}{\sup_{B(x,r)} |u|}.$$

The constant $10d$ is there for technical reasons and the reader should not worry about it. It is clear that with this definition $N(Q_1) \le N(Q_2)$ if $Q_1 \subset Q_2$. Theorem 18 was proved in [Logunov 2018a], and then extended in [Georgiev and Roy-Fortin 2019] to the more general equation (9); the proof combines an accumulation of growth result ([Logunov 2018a, Lemma 2.1] and [Georgiev and Roy-Fortin 2019, Proposition 3.1], called the simplex lemma), and a propagation of smallness result ([Logunov 2018a, Lemma 4.1] and [Georgiev and Roy-Fortin 2019, Proposition 3.2], called the hyperplane lemma). The hyperplane lemma is a consequence of quantitative Cauchy uniqueness, which we state in a simple version below; it can be obtained from a very general result in [Alessandrini et al. 2009] (Theorem 1.7). See also [Lin 1991].

Proposition 17. *Let D be a bounded domain with $\mathcal{C}^2$ boundary, and let B be a ball of radius $\rho < 1$. Let u be a solution of (9) in $D \cap B$, $u \in \mathcal{C}^1(\bar{D} \cap B)$. There exist $\beta = \beta(\alpha, \gamma, K, D, \rho) \in (0, 1)$ and $C = C(\alpha, \gamma, K, D, \rho) > 0$ such that, if $|u| \le 1$ and $|\nabla u| \le \rho^{-1}$ in $D \cap B$, and $|u| \le \eta$ and $|\nabla u| \le \eta\rho^{-1}$ on $\partial D \cap B$, where η is a real number, then*

$$|u(x)| \le C\eta^{\beta}$$

for any $x \in D \cap \frac{1}{2}B$.

Remark. In [Logunov 2018a; Georgiev and Roy-Fortin 2019], Proposition 17 is applied when ∂D is flat; this is sufficient to prove the theorem below. We will use the proposition in the nonflat case later in Section 4, to prove a version of the hyperplane lemma.

Theorem 18 [Logunov 2018a, Theorem 5.1; Georgiev and Roy-Fortin 2019, Theorem 4.1]. *There exist a constant $c > 0$ and an integer $A > 1$ depending on the dimension only, and positive numbers $N_0 = N_0(\alpha, \gamma, K)$ and $R_0 = R_0(\alpha, \gamma, K)$ such that for any cube $Q \subset B(0, R_0)$ the following holds: if Q is partitioned into A^d equal subcubes, then the number of subcubes with doubling index greater than* $\max(N(Q)/(1+c), N_0)$ *is less than $\frac{1}{2}A^{d-1}$.*

Starting from Theorem 18, in [Logunov 2018b] an iterated version is proved, which is the one decisively used in the proof of the lower bound on zero sets. We state it below and refer to [Logunov 2018b] for the proof.

Theorem 19 [Logunov 2018b, Theorem 5.3]. *There exist positive constants c_1, c_2, C and an integer $B_0 > 1$ depending on the dimension only, and positive numbers $N_0 = N_0(\alpha, \gamma, K)$ and $R_0 = R_0(\alpha, \gamma, K)$ such that for any cube $Q \subset B(0, R_0)$ the following holds: if Q is partitioned into B^d equal subcubes, where $B > B_0$ is an integer, then the number of subcubes with doubling index greater than* $\max(N(Q)2^{-c_1 \log B/\log\log B}, N_0)$ *is less than CB^{d-1-c_2}.*

3.6. *Estimates in a spherical shell.* In the following we always indicate by u a solution of (9); the frequency function and doubling index are relative to u. Consider a ball $B(p,s) \subset B\big(0, \frac{1}{4}r_0\big)$; we are going to establish some estimates for the growth of u near a point of maximum. Let $x \in \partial B(p,s)$ be a point where the maximum of $|u|$ on $\overline{B(p,s)}$ is almost attained, in the sense that $\sup_{B(p,s)} |u| \le 2|u(x)|$; the existence of such an x is guaranteed by Corollary 5. Write $M = |u(x)|$. In the next two lemmas we will assume that there is a large enough number N and

$$\delta \in \left(\frac{1}{\log^{100} N}, \frac{1}{8}\right)$$

such that

$$\tfrac{1}{10}N \le \beta(p,t) \le 10^4 N \tag{25}$$

for $t \in I := (s(1-\delta), s(1+\delta))$.

Lemma 20 (variation on [Logunov 2018b, Lemma 4.1]). *Let (25) be satisfied. There exist positive constants C and c such that*

$$\sup_{B(p,s(1-\delta))} |u| \le CM2^{-c\delta N}, \tag{26}$$

$$\sup_{B(p,s(1+\delta))} |u| \le CM2^{C\delta N}. \tag{27}$$

Proof. Let us prove (26) only. By (21) and (25), we have that

$$\left(\frac{t_2}{t_1}\right)^{N/30} \le \frac{H(p,t_2)}{H(p,t_1)} \le C\left(\frac{t_2}{t_1}\right)^{10^5 N} \tag{28}$$

for $t_1 < t_2 \in I$, where we assume that r_0 is small enough to take $c_2 = 2$ in (21). We estimate

$$M^2 \ge C_1 s^{-d+1} H(p,s) \ge C_1 s^{-d+1} H\big(p, s\big(1-\tfrac{1}{2}\delta\big)\big)\big(1+\tfrac{1}{2}\delta\big)^{N/30},$$

where the first inequality is just the estimate of the L^2-norm by the L^∞-norm and the second inequality comes from (28). By integration and Proposition 11 we have

$$sH\big(p, s\big(1-\tfrac{1}{2}\delta\big)\big) = s\int_{\partial B(p,s(1-\delta/2))} |u|^2 \ge C_2 \int_{B(p,s(1-\delta/2))} |u|^2.$$

Let now $\tilde{x}$ be a point on $\partial B(p, s(1-\delta))$ where the supremum of $|u|$ on $B(p,s(1-\delta))$ is almost attained, i.e., $\sup_{B(p,s(1-\delta))}|u| \le 2|u(\tilde{x})|$, and write $\widetilde{M} = |u(\tilde{x})|$. Note now that

$$\int_{B(p,s(1-\delta/2))} |u|^2 \ge \int_{B(\tilde{x},\delta s/2)} |u|^2 \ge C_3(\delta s)^d \fint_{B(\tilde{x},\delta s/2)} |u|^2;$$

moreover, by (14) we have

$$\widetilde{M}^2 \le C_4 \fint_{B(\tilde{x},\delta s/2)} |u|^2.$$

Combining the estimates we obtain

$$M^2 \ge C_5 \delta^d \big(1+\tfrac{1}{2}\delta\big)^{N/30} \widetilde{M}^2.$$

Since $\log\big(1+\frac{1}{2}\delta\big) \ge \frac{1}{4}\delta$, it follows easily from the above and $\delta \gtrsim 1/\log^{100} N$ that $M^2 \ge C_6 \exp\big(\frac{1}{100}N\delta\big)\widetilde{M}^2$, from which one obtains (26) recalling the definitions of M and $\widetilde{M}$. $\square$

Using the properties of the doubling index, we now derive some estimates on small balls close to x; we keep on denoting by x the point on $\partial B(p,s)$ where the maximum of $|u|$ on $\overline{B(p,s)}$ is almost attained.

Lemma 21 (variation on [Logunov 2018b, Lemma 4.2]). *Let (25) be satisfied. There exists $C > 0$ such that*

$$\sup_{B(x,\delta s)} |u| \le M 2^{C\delta N + C} \tag{29}$$

and, for any $\tilde{x}$ with $d(x,\tilde{x}) \le \frac{1}{4}\delta s$,

$$\mathcal{N}\big(\tilde{x}, \tfrac{1}{4}\delta s\big) \le C\delta N + C, \tag{30}$$

$$\sup_{B(\tilde{x},\delta s/10N)} |u| \ge M 2^{-C\delta N \log N - C}. \tag{31}$$

Proof. Note that since $B(x,\delta s) \subset B(p, s(1+\delta))$, the first estimate (29) is an immediate consequence of (27). By definition of doubling index and (29) we have that

$$2^{\mathcal{N}(\tilde{x},\delta s/4)} \le \frac{\sup_{B(\tilde{x},\delta s/2)} |u|}{\sup_{B(\tilde{x},\delta s/4)} |u|} \le \frac{\sup_{B(x,\delta s)} |u|}{M} \le 2^{C\delta N + C},$$

and (30) is proved. Now recall the scaling properties (23); by those and (30) we obtain

$$\frac{\sup_{B(\tilde{x},\delta s/4)} |u|}{\sup_{B(\tilde{x},\delta s/10N)} |u|} \le (40N)^{2\mathcal{N}(\tilde{x},\delta s/4)+C_1} \le (40N)^{C_1\delta N + C_1} \le 2^{C_2\delta N \log N + C_2 \log N} \le 2^{C_3\delta N \log N + C_3},$$

where the last inequality holds because $\delta \gtrsim 1/\log^{100} N$. Since, by the distance condition, $\sup_{B(\tilde{x},\delta s/4)} |u| \ge |u(x)| = M$, (31) follows. □

3.7. *Finding many balls around the zero set.* We follow the arguments in Section 6 of [Logunov 2018b], in the reformulation contained in [Logunov and Malinnikova 2020]; the estimates in the spherical shell will be used together with the combinatorial results on doubling indices. We use the notion of "stable growth", which is taken from [Logunov and Malinnikova 2020] and was not present in [Logunov 2018b].

Definition 22. We say that u has a stable growth of order N in a ball $B(y,s)$ if $\mathcal{N}\big(y, \frac{1}{4}s\big) \ge N$ and $\mathcal{N}(y,s) \le 1000N$.

The number 1000 does not have any special meaning, it is just a large enough numerical constant. The following result is the key to the proof of the lower bound.

Proposition 23 (variation on [Logunov 2018b, Proposition 6.1]). *Let $B(p,2r) \subset B(0,r_0)$. There exists a number $N_0 > 0$ large enough such that, for $N > N_0$ and any solution u of (9) that has stable growth of order N in $B(p,r)$, the following holds: there exist at least $[\sqrt{N}]^{d-1} 2^{c_1 \log N/\log\log N}$ disjoint balls $B(x_i, r/\sqrt{N}) \subset B(p,r)$ such that $u(x_i) = 0$.*

Proof. Assume without loss of generality that $\sup_{B(p,r/4)} |u| = 1$. The stable growth assumption then implies that

$$\sup_{B(p,r/2)} |u| \ge 2^N \quad \text{and} \quad \sup_{B(p,2r)} |u| \le 2^{CN}.$$

We denote by x the point on $\partial B\big(p, \tfrac{1}{2}r\big)$ where the maximum over $\overline{B\big(p, \tfrac{1}{2}r\big)}$ is almost attained, so that by the above $|u(x)| \geq 2^{N-1}$. We now divide the ball $B(p, 2r)$ into cubes q_i of side length $cr/\sqrt{N}$ and organize these cubes into tunnels in the following way: the centers of the cubes in each tunnel lie on a line parallel to the segment that connects p and x. A tunnel contains at most $C\sqrt{N}$ cubes. Let us call a cube q_i "good" if

$$N(q_i) \leq \max\left(\frac{N}{2^{c \log N/\log\log N}}, N_0\right) \tag{32}$$

for some constant c. We will call a tunnel "good" if it contains only good cubes; by Theorem 19, most of the cubes are good and most of the tunnels are good. Another application of Theorem 19 gives the following:

Claim 1. *The number of good tunnels containing at least one cube with distance from x less than $r/\log^2 N$ is greater than $c(\sqrt{N}/\log^2 N)^{d-1}$.*

The proof of the proposition is then completed with the help of the next claim.

Claim 2. *Any good tunnel that contains at least one cube with distance from x less than $r/\log^2 N$ also contains at least $2^{c_2 \log N/\log\log N}$ cubes with zeros of u.*

Proof. Take one such tunnel T. Note that T contains at least one cube $q_a \subset B\big(p, \tfrac{1}{4}r\big)$, so that $\sup_{q_a} |u| \leq 1$. Call q_b a cube in T with distance from x less than $r/\log^2 N$; we want to show that the supremum of $|u|$ over q_b is large. To this end, we apply Lemma 21 with $\delta \sim 1/\log^2 N$ and $\tilde{x}$ being the center x_b of the cube q_b. By the stable growth assumption and the comparability of the doubling index and frequency function (Lemma 14), inequality (25) is satisfied for N large enough. Then (31) gives us

$$\sup_{B(x_b, \delta r/10N)} |u| \geq |u(x)| 2^{-CN/\log N - C},$$

and hence, recalling that $|u(x)| \geq 2^{N-1}$,

$$\sup_{q_b/2} |u| \geq 2^{cN}.$$

We now follow T from q_a to q_b and find many zeros. The proof is at this point identical to the one given in [Logunov 2018b]; for completeness we provide the details. We enumerate the cubes q_i from q_a to q_b such that q_a is the first and q_b is the last. Since T is a good tunnel, by (32) we have that for any two adjacent cubes

$$\log \frac{\sup_{q_{i+1}/2} |u|}{\sup_{q_i/2} |u|} \leq \log \frac{\sup_{4q_i} |u|}{\sup_{q_i/2} |u|} \leq \frac{N}{2^{c_3 \log N/\log\log N}}.$$

We split the set of indices S into two sets S_1 and S_2, where S_1 is the set of i such that u does not change sign in $\overline{q}_i \cup \overline{q}_{i+1}$ and $S_2 = S \setminus S_1$. The advantage of this is the possibility to use the Harnack inequality on S_1; the aim is to get a lower bound on the cardinality of S_2. In fact, for $i \in S_1$, by (16) we have that

$$\log \frac{\sup_{q_{i+1}/2} |u|}{\sup_{q_i/2} |u|} \leq C_1.$$

We then estimate

$$\log \frac{\sup_{q_b/2}|u|}{\sup_{q_a/2}|u|} = \sum_{S_1} \log \frac{\sup_{q_{i+1}/2}|u|}{\sup_{q_i/2}|u|} + \sum_{S_2} \log \frac{\sup_{q_{i+1}/2}|u|}{\sup_{q_i/2}|u|} \le |S_1|C_1 + |S_2| \frac{N}{2^{c_3 \log N/\log\log N}};$$

on the other hand, recall that

$$\log \frac{\sup_{q_b/2}|u|}{\sup_{q_a/2}|u|} \ge cN.$$

Combining the two estimates one obtains

$$|S_1|C_1 + |S_2| \frac{N}{2^{c_3 \log N/\log\log N}} \ge cN,$$

and noting that $|S_1|C_1 \le C_1\sqrt{N} \le \frac{1}{2}cN$ we conclude

$$|S_2| \ge c_3 2^{c_3 \log N/\log\log N}.$$

The last quantity is larger than $2^{c_2 \log N/\log\log N}$ if N is large enough, and the claim is proved. □

It is now a straightforward matter to finish the proof of Proposition 23: by Claim 1 there are at least $c(\sqrt{N}/\log^2 N)^{d-1}$ tunnels satisfying the hypothesis of Claim 2, and hence there are at least $c(\sqrt{N}/\log^2 N)^{d-1}2^{c_2 \log N/\log\log N}$ cubes that contain zeros of u; the last quantity can be made larger than $(\sqrt{N})^{d-1}2^{c_1 \log N/\log\log N}$, and then one replaces cubes by balls. □

3.8. *Proof of the lower bound.* We take r_0 small enough that (19), (21), Lemma 14 and (23) hold. Writing $N(0, r_0) = \sup_{\{B(x,r)\subset B(0,r_0)\}} \mathcal{N}(x, r)$, we define

$$F(N) = \inf \frac{\mathcal{H}^{d-1}(\{u = 0\} \cap B(x, \rho))}{\rho^{d-1}},$$

where the infimum is taken over all balls $B(x, \rho) \subset B(0, r_0)$ and all solutions u of (9) such that $u(x) = 0$ and $N(0, r_0) \le N$. Theorem 3 then follows immediately from the following:

Theorem 24. *$F(N) \ge c$, where c is independent of N.*

Proof. Let u be a solution of (9) in competition for the infimum in the definition of $F(N)$; let $F(N)$ be almost attained on u, in the sense that

$$\frac{\mathcal{H}^{d-1}(\{u = 0\} \cap B(x, \rho))}{r^{d-1}} \le 2F(N) \tag{33}$$

for some $B(x, r) \subset B(0, r_0)$ with $u(x) = 0$. Recall the easy bound (24):

$$\frac{\mathcal{H}^{d-1}(\{u = 0\} \cap B(x, r))}{r^{d-1}} \ge \frac{c_1}{\mathcal{N}(x, r/4)^{d-2}} \ge \frac{c_1}{N^{d-2}}. \tag{34}$$

Estimate (34) already finishes the proof if $\mathcal{N}\big(x, \frac{1}{4}r\big)$ is bounded uniformly in N; let us then argue by contradiction and assume that $\mathcal{N}\big(x, \frac{1}{4}r\big)$ is large enough. Denote $\widetilde{N} = \mathcal{N}\big(x, \frac{1}{4}r\big)$ and suppose first that u

has stable growth of order $\widetilde{N}$. We can then apply Proposition 23 and find at least $[\sqrt{\widetilde{N}}]^{d-1}2^{c\log\widetilde{N}/\log\log\widetilde{N}}$ disjoint balls $B(x_i, r/\sqrt{\widetilde{N}}) \subset B(x,r)$ with $u(x_i)=0$. By definition of $F(N)$, there holds:

$$\mathcal{H}^{d-1}\biggl(\{u=0\}\cap B\biggl(x_i, \frac{r}{\sqrt{\widetilde{N}}}\biggr)\biggr) \ge F(N)\biggl(\frac{r}{\sqrt{\widetilde{N}}}\biggr)^{d-1}.$$

Summing the inequality over all the balls, we obtain

$$\mathcal{H}^{d-1}(\{u=0\}\cap B(x,\rho)) \ge [\sqrt{\widetilde{N}}]^{d-1}2^{c\log\widetilde{N}/\log\log\widetilde{N}}F(N)\biggl(\frac{r}{\sqrt{\widetilde{N}}}\biggr)^{d-1};$$

the quantity on the right can be made larger than $2F(N)r^{d-1}$ if $\widetilde{N}$ is large enough, which is a contradiction with (33). Therefore we would be done if we knew a priori that u has stable growth of order $\widetilde{N}$ in $B(x,r)$, but this is not necessarily the case; fortunately we can find a smaller ball where u has stable growth.

Claim 3. *If $\mathcal{N}\bigl(x, \frac{1}{4}r\bigr)$ is large enough, there is a number $N_1 \gtrsim \mathcal{N}\bigl(x, \frac{1}{4}r\bigr)$ and a ball $B_1 \subset B(x,r)$ with radius $r_1 \sim r/\log^2 N_1$ such that u has stable growth of order $N_1/\log^2 N_1$ in B_1.*

Proof. Let us define a modified frequency function as

$$\tilde{\beta}(p,r) = \sup_{t\in(0,r]} \beta(p,t) + c_1,$$

so that $\tilde{\beta}(p,r)$ is a positive monotonic increasing function. Note that by (19) we have

$$\beta(p,r) \le \tilde{\beta}(p,r) \le c_3 + 2\beta(p,r),$$

and the rightmost expression is less than $3\beta(p,r)$ if $\beta(p,r) \ge c_3$. We use the following claim:

Claim 4 [Logunov 2018b, Lemma 3.1]. *Let f be a nonnegative, monotonic nondecreasing function in $[a,b]$, and assume $f \ge e$. Then there exist $x \in \bigl[a, \frac{1}{2}(a+b)\bigr)$ and a number $N_1 \ge e$ such that*

$$N_1 \le f(t) \le eN_1 \quad \textit{for any } t \in \biggl(x - \frac{b-a}{20\log^2 f(x)}, x + \frac{b-a}{20\log^2 f(x)}\biggr) \subset [a,b].$$

We apply Claim 4 to $\tilde{\beta}(p,\cdot)$ and hence identify a spherical shell of width $\sim r/\log^2 N_1$ about $s \in \bigl(\frac{2}{3}r, \frac{3}{4}r\bigr)$ where $\tilde{\beta}(p,\cdot)$ is comparable to N_1. Since $\mathcal{N}\bigl(x, \frac{1}{4}r\bigr)$ is large, by Lemma 14 and almost monotonicity $\beta(x,t)$ is large for $t > \frac{1}{2}r$ and then also $\beta(x,\cdot)$ is comparable to N_1 in the spherical shell. In other words, (25) holds with N_1 and $\delta \sim 1/\log^2 N$. Let now $y \in \partial B_s$ be a point where the maximum is almost attained, as in Lemmas 20 and 21. Take a ball B_1 of radius $\sim s/\log^2 N_1$ such that $\frac{1}{4}B_1 \subset B_{s(1-\delta)}$ and $y \in \frac{1}{2}B_1$; then (26) implies that

$$\mathcal{N}\bigl(\tfrac{1}{4}B_1\bigr) \ge c\frac{N_1}{\log^2 N_1}$$

and (27) implies that

$$\mathcal{N}(B_1) \le C\frac{N_1}{\log^2 N_1},$$

which means that u has stable growth of order $N_1/\log^2 N_1$ in B_1, and the claim is proved. □

Claim 3 gives an order of stable growth that is again large enough to get a contradiction with (33) if $\mathcal{N}\big(x, \frac{1}{4}r\big)$ and hence N_1 is large enough. This means that $\mathcal{N}\big(x, \frac{1}{4}r\big)$ is bounded from above by some N_0 independently of N, and therefore by (33) and (34) we obtain

$$F(N) \geq \frac{\mathcal{H}^{d-1}(\{u=0\} \cap B(x,r))}{2r^{d-1}} \geq \frac{c_3}{(N_0)^{d-2}} \geq c, \tag{35}$$

which concludes the proof of the theorem. □

4. Upper bound

Here we give the proof of Theorem 2. Throughout this section $\partial\Omega$ is assumed to be of class $\mathcal{C}^2$. As remarked in the introduction, the proof uses the Donnelly–Fefferman bound [1988] in the interior of the domain and a multiscale induction argument at the boundary. As will be apparent from the proof, the result with a $\mathcal{C}^\infty$-metric inside Ω would follow from an upper bound for zero sets of elliptic PDEs with smooth coefficients that is linear in the frequency; the best we have thus far is polynomial in the frequency [Logunov 2018a].

We introduce now a version of the doubling index that takes into account the boundary. Namely, for $x \in \overline{\Omega}$ and $u \in \mathcal{C}(\overline{\Omega})$ a harmonic function, we let

$$2^{\mathcal{N}_u^*(x,r)} = \frac{\sup_{B(x,2r)\cap\Omega} |u|}{\sup_{B(x,r)\cap\Omega} |u|}. \tag{36}$$

Note that if v is the extension across the boundary of the Steklov eigenfunction u_λ as in Section 2 and $\operatorname{dist}(x, \partial\Omega) \lesssim 1/\lambda$, $r \lesssim 1/\lambda$, we have that $\mathcal{N}_{u_\lambda}^*(x,r) \sim \mathcal{N}_v(x,r)$, where $\mathcal{N}_v(x,r)$ is defined as in (22); this will allow us to use the almost monotonicity property (23). It was proved in [Zhu 2015] (using the extension v) that for any $r < r_0(\Omega)$

$$\mathcal{N}_{u_\lambda}^*(x,r) \leq C\lambda, \tag{37}$$

mirroring a corresponding statement for Laplace eigenfunctions proved by Donnelly and Fefferman. It will once again be convenient to define a maximal version of the doubling index for cubes; for $Q \subset \mathbb{R}^d$ a cube such that $Q \cap \Omega \neq \varnothing$, we set

$$N_u^*(Q) = \sup_{\substack{x \in Q\cap\overline{\Omega} \\ r \leq \operatorname{diam}(Q)}} \mathcal{N}_u^*(x,r).$$

Definition 25. A Whitney cube in Ω is any cube Q such that $c_1 \operatorname{dist}(Q, \partial\Omega) \leq s(Q) \leq c_2 \operatorname{dist}(Q, \partial\Omega)$, where $s(Q)$ is the side length of Q and c_1 and c_2 are positive dimensional constants.

With this notation, we state the following important result of [Donnelly and Fefferman 1988].

Theorem 26. *Let u be a harmonic function in Ω. Then there is $C > 0$, independent of u, such that*

$$\mathcal{H}^{d-1}(\mathcal{Z}_u \cap Q) \leq C(N_u^*(Q)+1)s(Q)^{d-1} \tag{38}$$

for any Whitney cube Q.

From now on, we will denote by u a Steklov eigenfunction with eigenvalue λ. We will first use the theorem above to bound the measure of the zero set of u in the interior, up to a distance from the boundary comparable to $1/\lambda$. We will assume $\lambda > \lambda_0$. As in the previous section, write $d(x) = \mathrm{dist}(x, \partial\Omega)$; Let c_0 be a small constant depending only on Ω. We write the decomposition

$$\Omega = \mathrm{In} \cup \mathrm{Mid} \cup \mathrm{Bd},$$

where $\mathrm{In} = \{x \in \Omega : d(x) \geq c_0\}$, $\mathrm{Mid} = \{x \in \Omega : c_0/\lambda < d(x) < c_0\}$ and $\mathrm{Bd} = \{x \in \Omega : d(x) \leq c_0/\lambda\}$. It follows easily from Theorem 26 and (37) that

$$\mathcal{H}^{d-1}(\mathcal{Z}_u \cap \mathrm{In}) \leq C\lambda, \tag{39}$$

with C depending on Ω only. The next lemma estimates the contribution of the nodal set in Mid.

Lemma 27. *There is $C > 0$ depending only on Ω such that*

$$\mathcal{H}^{d-1}(\mathcal{Z}_u \cap \mathrm{Mid}) \leq C\lambda \log \lambda. \tag{40}$$

Proof. We set $M_k = \{x \in \Omega : c_0 2^{k-1}/\lambda < d(x) < c_0 2^k/\lambda\}$, and we have

$$\mathrm{Mid} = \bigcup_{k=1}^{c \log \lambda} M_k.$$

We perform a decomposition of Ω into Whitney cubes with disjoint interior (the statement that this is possible is usually called the Whitney covering lemma). Define

$$\mathcal{Q}_k = \{\text{Whitney cubes intersecting } M_k\}.$$

In the following lines we will denote by $|\cdot|$ both the cardinality of a discrete collection and the Lebesgue measure of cubes; it should cause no confusion. Note that if $Q \in \mathcal{Q}_k$, then

$$|Q| \sim \frac{2^{kd}}{\lambda^d};$$

it follows that $|\mathcal{Q}_k| \lesssim 2^{-kd}\lambda^{d-1}$. We can then estimate, using Theorem 26 and (37),

$$\mathcal{H}^{d-1}(\mathcal{Z}_u \cap \mathrm{Mid}) = \sum_{k=1}^{c \log \lambda} \mathcal{H}^{d-1}(\mathcal{Z}_u \cap M_k) \leq \sum_{k=1}^{c \log \lambda} \sum_{Q \in \mathcal{Q}_k} \mathcal{H}^{d-1}(\mathcal{Z}_u \cap Q)$$
$$\lesssim \lambda \sum_{k=1}^{c \log \lambda} \sum_{Q \in \mathcal{Q}_k} s(Q)^{d-1} \lesssim \lambda \sum_{k=1}^{c \log \lambda} |\mathcal{Q}_k| \frac{2^{kd}}{\lambda^{d-1}} \lesssim \lambda \log \lambda. \qquad \square$$

To prove Theorem 2 the only thing left is to estimate $\mathcal{H}^{d-1}(\mathcal{Z}_u \cap \mathrm{Bd})$. We cover Bd with $\sim \lambda^{d-1}$ cubes q_λ centered at $\partial\Omega$ of side length $s(q_\lambda) = 4c_0/\lambda$; then Theorem 2 follows from (37) and the following proposition:

Proposition 28. *Let q_λ be one of the cubes above, and suppose $N_u^*(4q_\lambda) \leq N$. Then*

$$\mathcal{H}^{d-1}(\mathcal{Z}_u \cap q_\lambda) \leq C(\Omega)(N+1)s(q_\lambda)^{d-1}. \tag{41}$$

Remark. In the following we will rescale

$$h(x) = u\Big(\frac{x}{\lambda}\Big), \tag{42}$$

so that q_λ becomes a cube Q of side length $s < 1$, where s is small enough depending on Ω but independent of λ, and h satisfies $\Delta h = 0$ in $10Q \cap \Omega$ and $\partial_\nu h = h$ on $\partial\Omega \cap \overline{10Q}$. Note that the doubling index is unchanged under this rescaling. Proposition 28 will follow from

$$\mathcal{H}^{d-1}(\mathcal{Z}_h \cap Q) \le C(\Omega)(N+1). \tag{43}$$

The main ingredient in the proof of Proposition 28 is a version of the hyperplane lemma of [Logunov 2018a] with cubes touching the boundary, the proof of which uses quantitative Cauchy uniqueness as stated in Proposition 17. The proof is very similar to the one contained in [Logunov et al. 2021], we reproduce it here for the reader's convenience.

Lemma 29. *Let h be as in (42) and Q a cube of side length s as in the remark above. There exist k and N_0 large enough depending on s and Ω such that if $Q \cap \partial\Omega$ is covered by $2^{k(d-1)}$ cubes q_j with disjoint interior centered at $\partial\Omega$ of side length $2^{-k}s$, and $N_h^*(Q) = N > N_0$, then there exists q_{j_0} such that $N_h^*(q_{j_0}) \le \frac{1}{2}N$.*

Proof. We note first that since $\partial\Omega$ is of class C^2, h is harmonic in $10Q \cap \Omega$ and $\partial_\nu h = h$ on $\partial\Omega \cap \overline{10Q}$, we can use the extension-across-the-boundary trick described in Section 4, namely, consider $v(x) = e^{d(x)}h(x)$; recall that the coefficients of the second-order term in the equation satisfied by v are at least Lipschitz. This gives us access to elliptic estimates that hold up to the boundary for h. In particular, we will use the gradient estimate

$$\sup_{B(y,r)\cap\bar{\Omega}} |\nabla h| \lesssim \frac{1}{r} \sup_{B(y,2r)\cap\bar{\Omega}} |h|, \tag{44}$$

where the implied constant depends on s and Ω. Denote now by $x_Q \in \partial\Omega$ the center of the cube Q. Consider a ball B centered at x_Q such that $2Q \subset B$, and let $M = \sup_{B\cap\Omega} |h|$. By contradiction, suppose that $N_h^*(q_j) > \frac{1}{2}N$ for any j; by definition, this implies that for any j there is $x_j \in q_j \cap \Omega$ and $r_j \le 2^{-k}\sqrt{d}s =: r_0$ such that $\mathcal{N}_h^*(x_j, r_j) > \frac{1}{2}N$. Assuming N large enough, we use (23) to get

$$\sup_{B(x_j,2r_0)\cap\Omega} |h| \le (C2^{-k})^{N/10} \sup_{B\cap\Omega} |h| \le Me^{-cNk}$$

if k is large enough. Using (44), we get

$$\sup_{B(x_j,r_0)\cap\Omega} |\nabla h| \lesssim \frac{1}{r_0} Me^{-cNk},$$

with the implied constant depending on s and Ω. Note that since $q_j \subset B(x_j, r_0)$ the two estimates above give bounds for the Cauchy data of h on $\partial\Omega \cap Q$. On the other hand, if B' is the ball centered at x_Q such that $4B' \subset Q$ we have $\sup_{2B'\cap\Omega} |h| \le M$ and $\sup_{2B'\cap\Omega} |\nabla h| \lesssim M/s$. Recalling that $r_0 = 2^{-k}\sqrt{d}s$, we can then apply Proposition 17 with $\eta = 2^k e^{-cNk}$ to get

$$\sup_{B'\cap\Omega} |h| \le C(s, \Omega) 2^{\beta k c_d} e^{-c\beta Nk} M.$$

But then

$$\mathcal{N}_h^*(x_Q, \sqrt{d}s) \geq C_d \log \frac{\sup_{B\cap\Omega} |h|}{\sup_{B'\cap\Omega} |h|} \geq C_d(c\beta Nk - c_d\beta k - C),$$

and the rightmost expression is larger than N if k and N are large enough depending on s and Ω; this is a contradiction with $N_h^*(Q) = N$. □

We are now ready to prove Proposition 28, or actually (43). The argument is an iteration at the boundary; it originates in [Logunov et al. 2021].

Proof of (43). First, we consider again $v(x) = e^{d(x)}h(x)$ and its even extension across the boundary (which we still call v). Recall from Section 2 that v satisfies an elliptic PDE with Lipschitz second-order coefficients and bounded lower-order coefficients. The results of [Hardt and Simon 1989] then apply to this situation. Let Q be any cube with $s(Q) < s_0$ small enough. By [Hardt and Simon 1989, Theorem 1.7], we have that

$$\mathcal{H}^{d-1}(\mathcal{Z}_v \cap B(x, \rho)) \leq C N_v(Q)\rho^{d-1}$$

for any ball $B(x, \rho) \subset Q$ where $v(x) = 0$ and $\rho < \rho_0(N_v(Q))$. Covering $\mathcal{Z}_h \cap Q$ with balls of such small radius and summing the estimate above over all those balls, it follows that there is a function $\tilde{A} : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\mathcal{H}^{d-1}(\mathcal{Z}_h \cap Q) \leq \tilde{A}(N_h^*(Q))s(Q)^{d-1}. \tag{45}$$

Let now Q be as above a cube centered at $\partial\Omega$ of side s, with s small enough depending on Ω. Fix a large number N_0; if $N_h^*(Q) < N_0$, (45) already implies the result. Otherwise, cover $Q \cap \Omega$ with smaller cubes of side length $2^{-k}s$, where $k = k(\Omega)$ is given by Lemma 29, in the following way: first $Q \cap \partial\Omega$ is covered by cubes $q \in \mathcal{B}$ centered at $\partial\Omega$ with disjoint interior, and then the rest of $Q \cap \Omega$ is covered by cubes $q \in \mathcal{I}$ with $\operatorname{dist}(q, \partial\Omega) > cs(q)$ for some constant $c > 0$ independent of k. Cubes in $\mathcal{B}$ will be called boundary cubes and cubes in $\mathcal{I}$ will be called inner cubes; inner cubes are allowed to overlap, while boundary cubes are not. Write $N_h^*(Q) = N$. By (38) and almost monotonicity,

$$\mathcal{H}^{d-1}\Bigl(\mathcal{Z}_h \cap \Bigl(\bigcup_{q\in\mathcal{I}} q\Bigr)\Bigr) \leq C(k)Ns^{d-1}.$$

By Lemma 29, there is a boundary cube, call it q_0, such that $N_h^*(q_0) < \frac{1}{2}N$. The other cubes in $\mathcal{B}$ will be enumerated from 1 to $2^{k(d-1)}-1$. We have that

$$\frac{\mathcal{H}^{d-1}(\mathcal{Z}_h \cap Q)}{s^{d-1}} \leq CN + \frac{\mathcal{H}^{d-1}(\mathcal{Z}_h \cap q_0)}{s^{d-1}} + \sum_{j=1}^{2^{k(d-1)}-1} \frac{\mathcal{H}^{d-1}(\mathcal{Z}_h \cap q_j)}{s^{d-1}}.$$

We define

$$A(N) = \sup \frac{\mathcal{H}^{d-1}(\mathcal{Z}_h \cap q)}{s(q)^{d-1}},$$

where the supremum is taken over all harmonic functions h in $2Q$ with $\partial_\nu h = h$ on $\partial\Omega \cap 2Q$, $N_h^*(Q) \leq N$ and all cubes $q \subset Q$. By (45), $A(N) < +\infty$. From the inequality above, we get

$$A(N) \leq C(k)N + A\bigl(\tfrac{1}{2}N\bigr)2^{-k(d-1)} + (2^{k(d-1)} - 1)A(N)2^{-k(d-1)},$$

from which

$$A(N) < C(k)N + A\left(\tfrac{1}{2}N\right).$$

(Beware that $C(k)$ changes value from line to line and depends also on Ω). Iterating the last inequality until $\frac{1}{2}N < N_0$, we obtain

$$A(N) < C(k)N + A(N_0) < C(k)(N+1),$$

which concludes the proof. □

Theorem 2 now follows by combining (39), (40), (41) and (37). We believe that the extra $\log\lambda$ factor is not necessary and is an artificial feature of the proof; it appears in the proof of (40) and it is due to the necessity of getting to cubes of side length $\sim \lambda^{-1}$.

Acknowledgements

This work owes a lot to the patient guidance of Eugenia Malinnikova; discussing the article with her has contributed greatly to both the contents and the presentation, and her reading of several drafts helped spot mistakes and inaccuracies. Many thanks are due to Aleksandr Logunov for his encouragement, as well as for reading this work at various stages and asking useful questions. I am also very grateful for the encouragement of Iosif Polterovich, whose comments helped improve the presentation of this article.

The work for the present article was started while a Visiting Student Researcher at the Department of Mathematics at Stanford University; it is a pleasure to thank the department for the hospitality and nice working conditions.

The author is supported by Project 275113 of the Research Council of Norway.

References

[Alessandrini et al. 2009] G. Alessandrini, L. Rondi, E. Rosset, and S. Vessella, "The stability for the Cauchy problem for elliptic equations", *Inverse Problems* **25**:12 (2009), art. id. 123004. MR Zbl

[Bellová and Lin 2015] K. Bellová and F.-H. Lin, "Nodal sets of Steklov eigenfunctions", *Calc. Var. Partial Differential Equations* **54**:2 (2015), 2239–2268. MR Zbl

[Decio 2022] S. Decio, "Nodal sets of Steklov eigenfunctions near the boundary: inner radius estimates", *Int. Math. Res. Not.* **2022**:21 (2022), 16709–16729. MR Zbl

[Donnelly and Fefferman 1988] H. Donnelly and C. Fefferman, "Nodal sets of eigenfunctions on Riemannian manifolds", *Invent. Math.* **93**:1 (1988), 161–183. MR Zbl

[Garofalo and Lin 1986] N. Garofalo and F.-H. Lin, "Monotonicity properties of variational integrals, A_p weights and unique continuation", *Indiana Univ. Math. J.* **35**:2 (1986), 245–268. MR Zbl

[Garofalo and Lin 1987] N. Garofalo and F.-H. Lin, "Unique continuation for elliptic operators: a geometric-variational approach", *Comm. Pure Appl. Math.* **40**:3 (1987), 347–366. MR Zbl

[Georgiev and Roy-Fortin 2019] B. Georgiev and G. Roy-Fortin, "Polynomial upper bound on interior Steklov nodal sets", *J. Spectr. Theory* **9**:3 (2019), 897–919. MR Zbl

[Gilbarg and Trudinger 1983] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, 2nd ed., Grundlehren der Math. Wissenschaften **224**, Springer, 1983. MR Zbl

[Girouard and Polterovich 2017] A. Girouard and I. Polterovich, "Spectral geometry of the Steklov problem", *J. Spectr. Theory* **7**:2 (2017), 321–359. MR Zbl

[Han and Lin] Q. Han and F.-H. Lin, "Nodal sets of solutions of elliptic differential equations", book in preparation.

[Hardt and Simon 1989] R. Hardt and L. Simon, "Nodal sets for solutions of elliptic equations", *J. Differential Geom.* **30**:2 (1989), 505–522. MR Zbl

[Lin 1991] F.-H. Lin, "Nodal sets of solutions of elliptic and parabolic equations", *Comm. Pure Appl. Math.* **44**:3 (1991), 287–308. MR Zbl

[Logunov 2018a] A. Logunov, "Nodal sets of Laplace eigenfunctions: polynomial upper estimates of the Hausdorff measure", *Ann. of Math.* (2) **187**:1 (2018), 221–239. MR Zbl

[Logunov 2018b] A. Logunov, "Nodal sets of Laplace eigenfunctions: proof of Nadirashvili's conjecture and of the lower bound in Yau's conjecture", *Ann. of Math.* (2) **187**:1 (2018), 241–262. MR Zbl

[Logunov and Malinnikova 2018] A. Logunov and E. Malinnikova, "Nodal sets of Laplace eigenfunctions: estimates of the Hausdorff measure in dimensions two and three", pp. 333–344 in 50 *years with Hardy spaces*, edited by A. Baranov et al., Oper. Theory Adv. Appl. **261**, Birkhäuser, Cham, 2018. MR Zbl

[Logunov and Malinnikova 2020] A. Logunov and E. Malinnikova, "Review of Yau's conjecture on zero sets of Laplace eigenfunctions", pp. 179–212 in *Current developments in mathematics*, 2018, edited by D. Jerison et al., Int. Press, Somerville, MA, 2020. MR Zbl

[Logunov et al. 2021] A. Logunov, E. Malinnikova, N. Nadirashvili, and F. Nazarov, "The sharp upper bound for the area of the nodal sets of Dirichlet Laplace eigenfunctions", *Geom. Funct. Anal.* **31**:5 (2021), 1219–1244. MR Zbl

[Mangoubi 2008] D. Mangoubi, "Local asymmetry and the inner radius of nodal domains", *Comm. Partial Differential Equations* **33**:7-9 (2008), 1611–1621. MR Zbl

[Polterovich et al. 2019] I. Polterovich, D. A. Sher, and J. A. Toth, "Nodal length of Steklov eigenfunctions on real-analytic Riemannian surfaces", *J. Reine Angew. Math.* **754** (2019), 17–47. MR Zbl

[Sogge et al. 2016] C. D. Sogge, X. Wang, and J. Zhu, "Lower bounds for interior nodal sets of Steklov eigenfunctions", *Proc. Amer. Math. Soc.* **144**:11 (2016), 4715–4722. MR Zbl

[Zelditch 2015] S. Zelditch, "Hausdorff measure of nodal sets of analytic Steklov eigenfunctions", *Math. Res. Lett.* **22**:6 (2015), 1821–1842. MR Zbl

[Zhu 2015] J. Zhu, "Doubling property and vanishing order of Steklov eigenfunctions", *Comm. Partial Differential Equations* **40**:8 (2015), 1498–1520. MR Zbl

[Zhu 2020] J. Zhu, "Geometry and interior nodal sets of Steklov eigenfunctions", *Calc. Var. Partial Differential Equations* **59**:5 (2020), art. id. 150. MR Zbl

Received 3 May 2021. Revised 13 Jul 2022. Accepted 19 Aug 2022.

STEFANO DECIO: sdecio@ias.edu
Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway
Current address: *Institute for Advanced Study, Princeton, NJ, United States*

ANALYSIS AND PDE
Vol. 17 (2024), No. 4, pp. 1261–1329
DOI: 10.2140/apde.2024.17.1261

msp

ON FULL ASYMPTOTICS OF REAL ANALYTIC TORSIONS FOR COMPACT LOCALLY SYMMETRIC ORBIFOLDS

BINGXIAO LIU

We consider a certain sequence of flat vector bundles on a compact locally symmetric orbifold, and we evaluate explicitly the associated asymptotic Ray–Singer real analytic torsion. The basic idea is to computing the heat trace via Selberg's trace formula, so that a key point in this paper is to evaluate the orbital integrals associated with nontrivial elliptic elements. For that purpose, we deduce a geometric localization formula, so that we can rewrite an elliptic orbital integral as a sum of certain identity orbital integrals associated with the centralizer of that elliptic element. The explicit geometric formula of Bismut for semisimple orbital integrals plays an essential role in these computations.

1. Introduction

Let (Z, g^{TZ}) be a closed Riemannian manifold of dimension m, and let $F \to Z$ be a complex vector bundle equipped with a Hermitian metric h^F and a flat connection $\nabla^{F,f}$. Let $(\Omega^\bullet(Z,F), d^{Z,F})$ be the associated de Rham complex valued in F. It is equipped with an L_2-metric induced by g^{TZ}, h^F. Let $\boldsymbol{D}^{Z,F,2}$ be the corresponding de Rham–Hodge Laplacian. The real analytic torsion $\mathcal{T}(Z,F)$ is a real-valued (graded) spectral invariant of $\boldsymbol{D}^{Z,F,2}$ introduced by Ray and Singer [1971; 1973]. When Z is odd-dimensional and $(F, \nabla^{F,f})$ is acyclic, this invariant does not depend on the metric data g^{TZ}, h^F. Ray and Singer also conjectured that, for a unitarily flat vector bundle F (i.e., $\nabla^{F,f} h^F = 0$), this invariant coincides with the Reidemeister torsion, a topological invariant associated with $(F, \nabla^{F,f}) \to Z$. This conjecture was later proved by Cheeger [1979] and Müller [1978]. Using the Witten deformation, Bismut and Zhang [1991; 1992] gave an extension of the Cheeger–Müller theorem for arbitrary flat vector bundles.

MSC2020: 11F72, 53C35, 57R18, 58C40.
Keywords: real analytic torsion, orbital integral, orbifold, locally symmetric space.

If Z is a compact orbifold, and if F is a flat orbifold vector bundle on Z, the Ray–Singer analytic torsion $\mathcal{T}(Z, F)$ extends naturally to this case (see Definition 2.2.3). In particular, if F is acyclic, and if Z and all the singular strata have odd dimensions, then $\mathcal{T}(Z, F)$ is independent of the metric data; see [Shen and Yu 2022, Corollary 4.9]. We refer to [Ma 2005; Shen and Yu 2022] for more details.

We consider a certain sequence of (acyclic) flat vector bundles $\{F_d\}_{d\in\mathbb{N}}$ on a compact locally symmetric space Z, and we study the asymptotic behavior of $\mathcal{T}(Z, F_d)$ as $d \to +\infty$. When Z is a manifold, such question was already studied by Müller [2012], by Bismut, Ma and Zhang [Bismut et al. 2011; 2017] and by Müller and Pfaff [2013b; 2013a]. In particular, Bismut, Ma and Zhang [Bismut et al. 2011; 2017] worked on the manifolds which are more general than locally symmetric manifolds. When Z is a compact hyperbolic orbifold, such question was studied by Fedosova [2015] using the method of harmonic analysis. Here, we consider this question for an arbitrary compact locally symmetric orbifold (of noncompact type).

Let G be a connected linear reductive Lie group equipped with a Cartan involution $\theta \in \mathrm{Aut}(G)$ and an invariant nondegenerate symmetric bilinear form B. Let $K \subset G$ be the fixed-point set of θ, which is a maximal compact subgroup of G. Put

$$X = G/K. \tag{1.0.1}$$

Then X is a Riemannian symmetric space with the Riemannian metric induced from B. For convenience, we also assume that G has a compact center; then X is of noncompact type.

Now let $\Gamma \subset G$ be a cocompact discrete subgroup. Set

$$Z = \Gamma\backslash X. \tag{1.0.2}$$

Then Z is a compact locally symmetric space. In general, Z is an orbifold. Let ΣZ denote the orbifold resolution of the singular points in Z whose connected components correspond exactly to the nontrivial elliptic conjugacy classes of Γ.

Since G has compact center, the compact form U of G exists and is a connected compact linear Lie group. If (E, ρ^E, h^E) is a unitary (analytic) representation of U, then it extends uniquely to a representation of G by a unitary trick. In this way, $F = G \times_K E$ is a vector bundle on X equipped with an invariant flat connection $\nabla^{F,f}$ (see Section 3.4 and (4.1.8)) and a unimodular Hermitian metric h^F induced by h^E. Moreover, $(F, \nabla^{F,f}, h^F)$ descends to a flat Hermitian orbifold vector bundle on Z, which is still denoted by $(F, \nabla^{F,f}, h^F)$. Let $\boldsymbol{D}^{Z,F,2}$ denote the corresponding de Rham–Hodge Laplacian.

The fundamental rank $\delta(G)$ (or $\delta(X)$) of G (or X) is the difference of the complex ranks of G and of K. As we will see in Theorem 4.1.4, if $\delta(G) \neq 1$, we always have

$$\mathcal{T}(Z, F) = 0. \tag{1.0.3}$$

If F is defined instead by a unitary representation of Γ, this result is obtained by Moscovici and Stanton [1991, Corollary 2.2]. If Γ is torsion-free, with F defined via a representation of G as above, (1.0.3) was proved in [Bismut et al. 2017, Remark 8.7] by using Bismut's formula for orbital integrals [2011, Theorem 6.1.1]; see also [Ma 2019, Theorems 5.4 and 5.5]. A new proof was given in [Müller and Pfaff 2013a, Proposition 4.2] (with a correction given in [Matz and Müller 2023, p. 44]). Note that in [Ma 2019, Remark 5.6], it is indicated that, using essentially Theorem 5.4 of that work, the identity (1.0.3) still holds if Γ is not torsion-free (i.e., Z is an orbifold), which gives us exactly Theorem 4.1.4 in this paper. Due to this vanishing result, we only need to deal with the case $\delta(G)=1$.

We now describe the sequence of flat vector bundles $\{F_d\}_{d\in\mathbb{N}}$ which is concerned here. Note that U contains K as a Lie subgroup. Let T be a maximal torus of K, and let T_U be the maximal torus of U containing T. Let $\mathfrak{u}$ be the Lie algebra of U, and let $\mathfrak{t}_U\subset\mathfrak{u}$ be the Lie algebra of T_U. Let $R(\mathfrak{u},\mathfrak{t}_U)$ be the associated real root system with a system of positive roots $R^+(\mathfrak{u},\mathfrak{t}_U)$. Then let $P_{++}(U)\subset\mathfrak{t}_U^*$ denote the set of (real) dominant weights of U with respect to the above root system. If $\lambda\in P_{++}(U)$, let $(E_\lambda,\rho^{E_\lambda})$ be the irreducible unitary representation of U with the highest weight λ. We extend it to a representation of G. We require λ to be nondegenerate, i.e., as G-representations, $(E_\lambda,\rho^{E_\lambda})$ is not isomorphic to $(E_\lambda,\rho^{E_\lambda}\circ\theta)$. We also take an arbitrary $\lambda_0\in P_{++}(U)$. If $d\in\mathbb{N}$, let (E_d,ρ^{E_d},h^{E_d}) be the unitary representation of U with highest weight $d\lambda+\lambda_0$. By Weyl's dimension formula, $\dim E_d$ is a polynomial in d. This way, we get a sequence of (unimodular) flat vector bundles $\{(F_d,\nabla^{F_d},h^{F_d})\}_{d\in\mathbb{N}}$ on X or on Z.

Note that in Section 8.1 (see also [Bergeron and Venkatesh 2013, Lemma 4.1]), the nondegeneracy of λ implies that, for d large enough,

$$H^\bullet(Z,F_d)=0. \tag{1.0.4}$$

Furthermore, $\dim Z$ is odd when $\delta(G)=1$. Then, for any sufficiently large d, $\mathcal{T}(Z,F_d)$ is independent of the different choices of h^{E_d} (or h^{F_d}).

Let $E[\Gamma]$ be the finite set of elliptic classes in Γ. Set $E^+[\Gamma]=E[\Gamma]\backslash\{1\}$. The first main result in this paper is the following theorem.

Theorem 1.0.1. *Assume that $\delta(G)=1$. There exists a (real) polynomial $P(d)$ in d, and for each $[\gamma]\in E^+[\Gamma]$ there exists a nice exponential polynomial $PE^{[\gamma]}(d)$ in d (i.e., a finite sum of the terms of the form $\alpha d^j e^{2\pi\sqrt{-1}\beta d}$, with $\alpha\in\mathbb{C}$, $j\in\mathbb{N}$, $\beta\in\mathbb{Q}$; see Definition 7.6.1) such that there exists a constant $c>0$ for d large, we have*

$$\mathcal{T}(Z,F_d)=P(d)+\sum_{[\gamma]\in E^+[\Gamma]}PE^{[\gamma]}(d)+\mathcal{O}(e^{-cd}). \tag{1.0.5}$$

Moreover, the degrees of $P(d)$, $PE^{[\gamma]}(d)$ can be determined in terms of λ, λ_0.

For a hyperbolic 3-manifold Z, Müller [2012, Theorem 1.1] computed explicitly the leading term of $\mathcal{T}(Z,F_d)$ as $d\to+\infty$. In [Bismut et al. 2011; 2017], under a more general setting for a closed manifold Z, Bismut, Ma and Zhang [Bismut et al. 2017, Remark 7.8] proved that there exists a constant $c>0$ such that

$$\mathcal{T}(Z,F_d)=\mathcal{T}_{L_2}(Z,F_d)+\mathcal{O}(e^{-cd}), \tag{1.0.6}$$

where $\mathcal{T}_{L_2}(Z,F_d)$ denotes the L_2-torsion [Lott 1992; Mathai 1992] associated with $F_d\to Z$. Moreover, they constructed universally an element $W\in\Omega^\bullet(Z,o(TZ))$ (where $o(TZ)$ denotes the orientation bundle of TZ) such that if $n_0=\deg E_d$, then

$$\mathcal{T}_{L_2}(Z,F_d)=d^{n_0+1}\int_Z W+\mathcal{O}(d^{n_0}). \tag{1.0.7}$$

The integral of W in the right-hand side of (1.0.7) is called a W-invariant. If we specialize (1.0.7) for a compact locally symmetric manifold Z, we get

$$\mathcal{T}_{L_2}(Z, F_d) = d^{n_0+1}\,\mathrm{Vol}(Z)[W]^{\max} + \mathcal{O}(d^{n_0}). \tag{1.0.8}$$

In [Bismut et al. 2017, Section 8.7], the explicit computation on $[W]^{\max}$ was carried out for $G = \mathrm{SL}_2(\mathbb{C})$ to recover [Müller 2012, Theorem 1.1].

We now compare (1.0.5) with (1.0.6). If ignoring that Γ may act on X noneffectively, we can extend the notion of L_2-torsion to the orbifold Z, so that $\mathcal{T}_{L_2}(Z, F_d)$ is still defined in terms of the Γ-trace of the heat operators on X. Then $P(d)$ in (1.0.5) is exactly $\mathcal{T}_{L_2}(Z, F_d)$. But different from (1.0.6), we still have the nontrivial terms $PE^{[\gamma]}(d)$, $[\gamma] \in E^+[\Gamma]$ in (1.0.5). We will see, in a refined version of (1.0.5) stated in Theorem 1.0.2, that $PE^{[\gamma]}(d)$ is essentially a linear combination of certain L_2-torsions for ΣZ associated with $[\gamma]$ and λ, λ_0. Therefore, we can define an L_2-torsion for ΣZ as

$$\widetilde{\mathcal{T}}_{L_2}(\Sigma Z, F_d) = \sum_{[\gamma]\in E^+[\Gamma]} PE^{[\gamma]}(d). \tag{1.0.9}$$

Then, as an analogue to (1.0.6), we restate our Theorem 1.0.1 as follows.

Theorem 1.0.1′. *Assume that Γ acts on X effectively. For $Z = \Gamma\backslash X$, as $d \to +\infty$, we have*

$$\mathcal{T}(Z, F_d) = \mathcal{T}_{L_2}(Z, F_d) + \widetilde{\mathcal{T}}_{L_2}(\Sigma Z, F_d) + \mathcal{O}(e^{-cd}). \tag{1.0.10}$$

Moreover, $\mathcal{T}_{L_2}(Z, F_d)$ is a polynomial in d, and $\widetilde{\mathcal{T}}_{L_2}(\Sigma Z, F_d)$ is a nice exponential polynomial in d. Their leading terms can be determined in terms of W-invariants as in (1.0.8) .

To understand better on $\widetilde{\mathcal{T}}_{L_2}(\Sigma Z, F_d)$, we need to recall the results in [Müller and Pfaff 2013a] (also in [Müller and Pfaff 2013b] for the hyperbolic case) for a compact locally symmetric manifold Z. They gave a proof to (1.0.6) using Selberg's trace formula, and then showed that $\mathcal{T}_{L_2}(Z, F_d)$ is a polynomial in d. Theorem 1.0.1′ here is an extension of their results, which shows a nontrivial contribution from ΣZ.

Let us give more detail on the results in [Müller and Pfaff 2013a]. Let $\boldsymbol{D}^{X,F_d,2}$ be the G-invariant Laplacian operator on X which is the lift of $\boldsymbol{D}^{Z,F_d,2}$. For $t > 0$, let $p_t^{X,F_d}(x, x')$ denote the heat kernel of $\frac{1}{2}\boldsymbol{D}^{X,F_d,2}$ with respect to the Riemannian volume element on X. For $t > 0$, the identity orbital integral $\mathcal{I}_X(E_d, t)$ of p_t^{X,F_d} is defined as

$$\mathcal{I}_X(F_d, t) = \mathrm{Tr}_s^{\Lambda^\bullet(T_x^*X)\otimes F_{d,x}}\Big[\Big(N^{\Lambda^\bullet(T_x^*X)} - \frac{m}{2}\Big)p_t^{X,F_d}(x, x)\Big], \tag{1.0.11}$$

where $N^{\Lambda^\bullet(T_x^*X)}$ is the number operator on $\Lambda^\bullet(T_x^*X)$, and the right-hand side of (1.0.11) is independent of the choice of $x \in X$. Let $\mathcal{MI}_X(F_d, s)$, $s \in \mathbb{C}$, denote the Mellin transform (see (7.2.57)) of $\mathcal{I}_X(F_d, t)$, which is holomorphic at 0. Set

$$\mathcal{PI}_X(F_d) = \frac{\partial}{\partial s}\Big|_{s=0}\mathcal{MI}_X(F_d, s). \tag{1.0.12}$$

The L_2-torsion is defined as

$$\mathcal{T}_{L^2}(Z, F_d) = \mathrm{Vol}(Z)\mathcal{PI}_X(F_d). \tag{1.0.13}$$

Using essentially Harish-Chandra's Plancherel theorem for $\mathcal{I}_X(F_d, t)$, Müller and Pfaff [2013a] managed to show that $\mathcal{PI}_X(F_d)$ is a polynomial in d (for d large enough). Moreover, if $\lambda_0 = 0$, there exists a constant $C_\lambda \neq 0$ such that

$$\mathcal{PI}_X(F_d) = C_\lambda d \dim E_d + R(d), \tag{1.0.14}$$

where $R(d)$ is a polynomial in d of degree no greater than $\deg \dim E_d$. They also gave concrete formulae for C_λ in some model cases [Müller and Pfaff 2013a, Corollaries 1.4 and 1.5].

In Section 7.4, we use instead an explicit geometric formula of [Bismut 2011, Theorem 6.1.1] for semisimple orbital integrals to give a different computation on $\mathcal{PI}_X(F_d)$. In Section 7.5, we verify that our computational results coincide with the ones of [Müller and Pfaff 2013a].

For the orbifold case, i.e., Γ contains nontrivial elliptic elements, a key ingredient to Theorem 1.0.1 is to evaluate explicitly the elliptic orbital integrals associated with $[\gamma] \in E^+[\Gamma]$. For that purpose, we make use of the full power of Bismut's formula [2011, Theorem 6.1.1]. Note that if Z is a hyperbolic orbifold, i.e., $G = \mathrm{Spin}(1, 2n+1)$, the result in Theorem 1.0.1 (or Theorem 1.0.1′) was obtained in [Fedosova 2015, Theorem 1.1], where she evaluated the elliptic orbital integrals using Harish-Chandra's Plancherel theorem.

In fact, we obtain in this paper a refined version of Theorem 1.0.1, where we give more explicit descriptions of the exponential polynomials $PE^{[\gamma]}(d)$ and $\widetilde{\mathcal{T}}_{L_2}(\Sigma Z, F_d)$. Before stating this refined result, we need to introduce some notation and facts.

Fix $k \in T$, and let $X(k)$ denote the fixed-point set of k acting on X. Then $X(k)$ is a connected symmetric space with $\delta(X(k)) = 1$. Let $Z(k)^0$ be the identity component of the centralizer $Z(k)$ of k in G. Then $X(k) = Z(k)^0/K(k)^0$, with $K(k)^0 = Z(k)^0 \cap K$. Let $U(k)$ denote the centralizer of k in U with Lie algebra $\mathfrak{u}(k) \subset \mathfrak{u}$. Then $U(k)^0$ is naturally a compact form of $Z(k)^0$, and the triplet $(X(k), Z(k)^0, U(k)^0)$ becomes a smaller version of (X, G, U), except that $Z(k)^0$ may have noncompact center. Note that T_U is also a maximal torus of $U(k)^0$. We get the splitting of roots

$$R(\mathfrak{u}, \mathfrak{t}_U) = R(\mathfrak{u}(k), \mathfrak{t}_U) \cup R(\mathfrak{u}^\perp(k), \mathfrak{t}_U), \tag{1.0.15}$$

where $\mathfrak{u}^\perp(k)$ is the orthogonal space of $\mathfrak{u}(k)$ in $\mathfrak{u}$ with respect to B. Let $R^+(\mathfrak{u}(k), \mathfrak{t}_U)$, $R^+(\mathfrak{u}^\perp(k), \mathfrak{t}_U)$ be the induced positive roots, and let $\rho_\mathfrak{u}$, $\rho_{\mathfrak{u}(k)}$ denote the half of the sum of the roots in $R^+(\mathfrak{u}, \mathfrak{t}_U)$, $R^+(\mathfrak{u}(k), \mathfrak{t}_U)$ respectively.

Let $W(\mathfrak{u}_\mathbb{C}, \mathfrak{t}_{U,\mathbb{C}})$ be the Weyl group associated with the pair $(\mathfrak{u}, \mathfrak{t}_U)$. Put

$$W_U^1(k) = \{\omega \in W(\mathfrak{u}_\mathbb{C}, \mathfrak{t}_{U,\mathbb{C}}) \mid \omega^{-1}(R^+(\mathfrak{u}(k), \mathfrak{t}_U)) \subset R^+(\mathfrak{u}, \mathfrak{t}_U)\}. \tag{1.0.16}$$

If $\sigma \in W_U^1(k)$, let $\varepsilon(\sigma)$ denote its sign. For $\mu \in P_{++}(U)$, set

$$\varphi_k^U(\sigma, \mu) = \varepsilon(\sigma) \frac{\xi_{\sigma(\mu+\rho_\mathfrak{u})+\rho_\mathfrak{u}}(k)}{\prod_{\alpha \in R^+(\mathfrak{u}^\perp(k), \mathfrak{t}_U)} (\xi_\alpha(k) - 1)} \in \mathbb{C}^*, \tag{1.0.17}$$

where ξ_α is the character of T_U with (dominant) weight $2\pi\sqrt{-1}\alpha$. It is clear that $\varphi_k^U(\sigma, d\lambda + \lambda_0)$ is an oscillating term of the form $c_1 e^{2\pi\sqrt{-1}c_2 d}$, with $c_1 \in \mathbb{C}^*$, $c_2 \in \mathbb{R}$. If k is of finite order, then $c_2 \in \mathbb{Q}$.

By an equivalent definition of nondegeneracy in Definition 7.3.1, for $\sigma \in W_U^1(k)$, $\sigma\lambda$ is a nondegenerate dominant weight of $U(k)^0$ with respect to $\theta|_{Z(k)^0}$. Let $E_{\sigma,d}^k$ denote the unitary representations of $U(k)^0$ (up to a finite central extension) with highest weight $d\sigma\lambda + \sigma(\lambda_0 + \rho_{\mathrm{u}}) - \rho_{\mathrm{u}(k)}$, $d \in \mathbb{N}$, and let $\{F_{\sigma,d}^k\}_{d\in\mathbb{N}}$ be the corresponding sequence of flat vector bundles on $X(k)$.

Now we state our second main theorem, which refines Theorem 1.0.1.

Theorem 1.0.2. *Assume that* $\delta(G) = 1$.

(1) *If $\Gamma \subset G$ is a cocompact discrete subgroup and $\gamma \in \Gamma$ is elliptic, let $S(\gamma)$ denote the finite subgroup of $\Gamma \cap Z(\gamma)$ which acts on $X(\gamma)$ trivially. Then there exists a constant $c > 0$, and, for each $[\gamma] \in E^+[\Gamma]$, there exists a nice exponential polynomial in d, denoted by $\mathcal{PE}_{X,\gamma}(F_d)$, such that, for $Z = \Gamma\backslash X$, as $d \to +\infty$, we have*

$$\mathcal{T}(Z, F_d) = \frac{\mathrm{Vol}(Z)}{|S(1)|}\mathcal{PI}_X(F_d) + \sum_{[\gamma]\in E^+[\Gamma]} \frac{\mathrm{Vol}(\Gamma \cap Z(\gamma)\backslash X(\gamma))}{|S(\gamma)|}\mathcal{PE}_{X,\gamma}(F_d) + \mathcal{O}(e^{-cd}). \tag{1.0.18}$$

(2) *Fix an elliptic $[\gamma] \in E^+[\Gamma]$. Then $\mathcal{PE}_{X,\gamma}(F_d)$ depends only on the conjugacy class of γ in G and is independent of the lattice Γ. If γ is conjugate to $k \in T$ by an element in G, then we have the identity*

$$\mathcal{PE}_{X,\gamma}(F_d) = \sum_{\sigma\in W_U^1(k)} \varphi_k^U(\sigma, d\lambda + \lambda_0)\mathcal{PI}_{X(k)}(F_{\sigma,d}^k), \tag{1.0.19}$$

Theorem 1.0.1 now is just a consequence of (1.0.18). Note that, for $[\gamma] \in E^+[\Gamma]$, the (compact) orbifold $\Gamma \cap Z(\gamma)\backslash X(\gamma)$ represents an orbifold stratum in ΣZ (see (3.4.13), Remark 3.4.3). An important observation on (1.0.18) is that the sequence $\{\mathcal{T}(Z, F_d)\}_{d\in\mathbb{N}}$ encodes the volume information on Z as well as on ΣZ. Moreover, combining (1.0.13), (1.0.18) with (1.0.19), we justify that the quantity $\widetilde{\mathcal{T}}_{L_2}(\Sigma Z, F_d)$ defined by (1.0.9) is indeed a linear combination of L_2-torsions such as $\mathcal{T}_{L^2}(\Gamma \cap Z(\gamma)\backslash X(\gamma), F_{\sigma,d}^{\gamma})$ for ΣZ.

Now we explain our approach to Theorem 1.0.2. Let us start with defining $\mathcal{PE}_{X,\gamma}(F_d)$ and (1.0.18). In fact, $\mathcal{T}(Z, F_d)$ can be rewritten as the derivative at 0 of the Mellin transform of

$$\mathrm{Tr}_s\Big[\Big(N^{\Lambda^\bullet(T^*Z)} - \frac{m}{2}\Big)\exp\Big(-\frac{t\boldsymbol{D}^{Z,F_d,2}}{2}\Big)\Big], \quad t > 0, \tag{1.0.20}$$

where $\mathrm{Tr}_s[\cdot]$ denotes the supertrace with respect to the $\mathbb{Z}_2$-grading on $\Lambda^\bullet(T^*Z)$.

If $\gamma \in G$ is semisimple, let $\mathcal{E}_{X,\gamma}(F_d, t)$ denote the orbital integral (see Section 3.3) of the Schwartz kernel of $(N^{\Lambda^\bullet(T^*X)} - m/2)\exp(-t\boldsymbol{D}^{X,F_d,2}/2)$ associated with γ. Note that in $\mathcal{E}_{X,\gamma}(F_d, t)$, we take the supertrace of the endomorphism on $\Lambda^\bullet(T^*X) \otimes F$ (see (4.1.16)). Moreover, $\mathcal{E}_{X,\gamma}(F_d, t)$ depends only on the conjugacy class of γ in G. Let $\mathcal{ME}_{X,\gamma}(F_d, s)$ denote the Mellin transform of $\mathcal{E}_{X,\gamma}(F_d, t)$, $t > 0$ with appropriate $s \in \mathbb{C}$. If $\gamma = 1$, they are just $\mathcal{I}_X(F_d, t)$, $\mathcal{MI}_X(F_d, s)$ introduced in (1.0.11)–(1.0.12).

We use the notation in Section 3.5. Let $[\Gamma]$ denote the set of the conjugacy classes in Γ. By applying Selberg's trace formula to $Z = \Gamma\backslash X$, we get

$$\mathrm{Tr}_s\Big[\Big(N^{\Lambda^\bullet(T^*Z)} - \frac{m}{2}\Big)\exp\Big(-\frac{t\boldsymbol{D}^{Z,F_d,2}}{2}\Big)\Big] = \sum_{[\gamma]\in[\Gamma]} \frac{\mathrm{Vol}(\Gamma \cap Z(\gamma)\backslash X(\gamma))}{|S(\gamma)|}\mathcal{E}_{X,\gamma}(F_d, t). \tag{1.0.21}$$

Now we compare (1.0.18) with (1.0.21). Then a proof to (1.0.18) mainly includes the following three parts:

(1) We show that if $[\gamma] \in E[\Gamma]$, then $\mathcal{ME}_{X,\gamma}(F_d, s)$ admits a meromorphic extension to $s \in \mathbb{C}$ which is holomorphic at $s = 0$. Thus we define

$$\mathcal{PE}_{X,\gamma}(F_d) = \frac{\partial}{\partial s}\Big|_{s=0} \mathcal{ME}_{X,\gamma}(F_d, s). \tag{1.0.22}$$

Such consideration also holds for an arbitrary elliptic element $\gamma \in G$.

(2) If $\gamma \in \Gamma$ is elliptic, then it is of finite order, and from (1.0.19), we get that $\mathcal{PE}_{X,\gamma}(F_d)$ is a nice exponential polynomial in d for d large enough.

(3) We prove that all the terms in the sum of (1.0.21) associated with nonelliptic $[\gamma] \in [\Gamma]$ contribute as $\mathcal{O}(e^{-cd})$ in $\mathcal{T}(Z, F_d)$.

Indeed, to handle the contribution of the nonelliptic $[\gamma] \in [\Gamma]$, we use a spectral gap of $\boldsymbol{D}^{Z,F_d,2}$ due to the nondegeneracy of λ. By [Bismut et al. 2011, Théorème 3.2], and [Bismut et al. 2017, Theorem 4.4] which holds for a more general setting (see also [Müller and Pfaff 2013a, Proposition 7.5, Corollary 7.6] for a proof by using representation theory for symmetric spaces), there exist constants $C > 0$, $c > 0$ such that, for $d \in \mathbb{N}$,

$$\boldsymbol{D}^{Z,F_d,2} \geq cd^2 - C. \tag{1.0.23}$$

That also explains (1.0.4) for large d. Part (3) follows essentially from the same arguments as in [Müller and Pfaff 2013a, Section 8] and [Bismut et al. 2017, Sections 6.6, 7.2, Remarks 7.8, 8.15] which makes good use of (1.0.23) and the fact that nonelliptic elements in Γ admit a uniform strictly positive lower bound for their displacement distances on X.

For elliptic $\gamma \in \Gamma$, we apply Bismut's formula [2011, Theorem 6.1.1] to evaluate $\mathcal{E}_{X,\gamma}(F_d, t)$. Then we can write $\mathcal{E}_{X,\gamma}(F_d, t)$ as a Gaussian-like integral with the integrand given as a product of an analytic function determined by the adjoint action of γ on Lie algebras and the character χ_{E_d} of the representation E_d. By coordinating these two factors, especially using all sorts of character formulae for χ_{E_d}, we can integrate it out. We show that $\mathcal{E}_{X,\gamma}(F_d, t)$ is a finite sum of the terms

$$t^{-j-\frac{1}{2}} e^{-t(cd+b)^2} Q(d), \tag{1.0.24}$$

where $j \in \mathbb{N}$, $c \neq 0$, b are real constants, and $Q(d)$ is a nice exponential polynomial in d. It is crucial that $c \neq 0$. Indeed, we will see in Section 7.3 that this quantity c measures the difference between the representations $(E_\lambda, \rho^{E_\lambda})$ and $(E_\lambda, \rho^{E_\lambda} \circ \theta)$.

As a consequence of (1.0.24), $\mathcal{PE}_{X,\gamma}(F_d)$ in (1.0.22) is well-defined, which is clearly a nice exponential polynomial in d (for d large enough). The details on these computations are carried out in Section 7.2, where we apply the techniques inspired by the computations in Shen's approach [2018, Section 7] to the Fried conjecture and also in its extension to orbifold case in [Shen and Yu 2022].

The formula (1.0.19) gives a new and geometric approach to the above results on $\mathcal{PE}_{X,\gamma}(F_d)$. It is nicer in the sense that each $\mathcal{PI}_{X(k)}(F^k_{\sigma,d})$ is already well understood and related to the L_2-torsions for the singular stratum of Z. For proving it, we apply a geometric localization formula for $\mathcal{E}_{X,\gamma}(F_d, t)$ as follows.

Theorem 1.0.3. *Assume that $\delta(G) = 1$. We use the same notation as in Theorem 1.0.2. Let $\gamma = k \in T$. Then, for $t > 0$, $d \in \mathbb{N}$,*

$$\mathcal{E}_{X,\gamma}(F_d, t) = \sum_{\sigma \in W_U^1(k)} \varphi_k^U(\sigma, d\lambda + \lambda_0) \mathcal{I}_{X(k)}(F_{\sigma,d}^k, t). \tag{1.0.25}$$

After taking the Mellin transform on both sides of (1.0.25), we get exactly (1.0.19). In Theorem 6.0.1, we will show a general version of the above geometric localization formula for $\mathcal{E}_{X,\gamma}(F_d, t)$ associated with any semisimple element $\gamma \in G$.

Our approach to Theorem 1.0.3 is a more delicate application of Bismut's formula [2011, Theorem 6.1.1]. As we said, $\mathcal{E}_{X,\gamma}(F_d, t)$, $\mathcal{I}_{X(k)}(F_{\sigma,d}^k, t)$ are equal to integrals of some integrands involving χ_{E_d}, $\chi_{E_{\sigma,d}^k}$ respectively. To relate the two sides of (1.0.25), we employ a generalized version of the Kirillov character formula (see Theorem 5.4.4), which gives an explicit way of decomposing $\chi_{E_d}|_{U(k)^0}$ into a sum of $\chi_{E_{\sigma,d}^k}$, $\sigma \in W_U^1(k)$. This character formula was proved by Duflo, Heckman and Vergne [Duflo et al. 1984, II.3, Theorem (7)] under a general setting, and we will recall its special case for our need in Section 5.4. Then we expand the integral formula for $\mathcal{E}_{X,\gamma}(F_d, t)$ carefully into a sum of certain integrals involving $\chi_{E_{\sigma,d}^k}$, $\sigma \in W_U^1(k)$, which correspond to $\mathcal{I}_{X(k)}(F_{\sigma,d}^k, t)$ via Bismut's formula. This way, we prove (1.0.25).

Theorem 1.0.3 can be interpreted as follows: the action of elliptic element γ on X could lead to a geometric localization onto its fixed-point set $X(k)$ when we evaluate the orbital integrals. Even though we only prove it for a very restrictive situation, we still expect such phenomenon in general due to a geometric formulation for the semisimple orbital integrals; see [Bismut 2011, Chapter 4].

Finally, we note that in [Bismut et al. 2017, Section 8], the authors explained well how to use Bismut's formula for semisimple orbital integrals to study the asymptotic analytic torsion. Here, we go one step further in that direction to get a refined evaluation on it. Bergeron and Venkatesh [2013] also studied the asymptotic analytic torsion but under a totally different setting. In [Liu 2018; 2021], the asymptotic equivariant analytic torsion for a locally symmetric space was studied, and the oscillating terms also appeared naturally in that case. Moreover, Finski [2018, Theorem 1.5] obtained the full asymptotic expansion of the holomorphic analytic torsions for the tensor powers of a given positive line bundle over a compact complex orbifold.

This paper is organized as follows. In Section 2, we recall the definition of Ray–Singer analytic torsion for compact orbifolds. We also include a brief introduction to the orbifolds at beginning.

In Section 3, we introduce the explicit geometric formula of Bismut for semisimple orbital integrals and the Selberg's trace formula for compact locally symmetric orbifolds. They are the main tools to study the analytic torsions in this paper.

In Section 4, we give a vanishing theorem for $\mathcal{T}(Z, F)$, so that we only need to focus on the case $\delta(G) = 1$.

In Section 5, we study the Lie algebra of G provided $\delta(G) = 1$. Furthermore, we introduce a generalized Kirillov formula for compact Lie groups.

In Section 6, we prove a general version of Theorem 1.0.3.

In Section 7, given the sequence $\{F_d\}_{d\in\mathbb{N}}$, we compute explicitly $\mathcal{E}_{X,\gamma}(F_d, t)$ in terms of root systems for elliptic γ; in particular, we prove (1.0.24). Then we give the formulae for $\mathcal{PI}_X(F_d)$, $\mathcal{PE}_{X,\gamma}(F_d)$.

Finally, in Section 8, we introduce the spectral gap (1.0.23) and we give a proof of Theorem 1.0.2.

In this paper, if V is a real vector space and if E is a complex vector space, we will use the symbol $V \otimes E$ to denote the complex vector space $V \otimes_{\mathbb{R}} E$. If both V and E are complex vector spaces, then $V \otimes E$ is just the usual tensor over $\mathbb{C}$.

2. Ray–Singer analytic torsion

In this section, we recall the definitions of the orbifold and the orbifold vector bundle. We also refer to [Satake 1956; 1957; Adem et al. 2007, Chapter 1] for more details. Then we recall the definition of Ray–Singer analytic torsion for compact orbifolds, where we refer to [Ma 2005; Shen and Yu 2022] for more details. In particular, Shen and Yu [2022] extended many important results on real analytic torsion from the manifold setting to the orbifold setting.

2.1. *Orbifolds and orbifold vector bundles.* Let Z be a topological space.

Definition 2.1.1. If U is a connected open subset of Z, an orbifold chart for U is a triple $(\widetilde{U}, \pi_U, G_U)$ such that

- $\widetilde{U}$ is a connected open set of some $\mathbb{R}^m$ and G_U is a finite group acting smoothly and effectively on $\widetilde{U}$ on the left;
- π_U is a continuous surjective $\widetilde{U} \to U$, which is invariant by a G_U-action;
- π_U induces a homeomorphism between $G_U \backslash \widetilde{U}$ and U.

If $V \subset U$ is a connected open subset, an embedding of orbifold chart for the inclusion $i : V \to U$ is an orbifold chart $(\widetilde{V}, \pi_V, G_V)$ for V and an orbifold chart $(\widetilde{U}, \pi_U, G_U)$ for U together with a smooth embedding $\phi_{UV} : \widetilde{V} \to \widetilde{U}$ such that the following diagram commutes:

$$\begin{array}{ccc} \widetilde{V}_{\pi_V} & \xrightarrow{\phi_{UV}} & \widetilde{U} \\ \downarrow & & \downarrow \pi_U \\ V & \xrightarrow{i} & U \end{array} \tag{2.1.1}$$

If U_1, U_2 are two connected open subsets of Z with the charts $(\widetilde{U}_1, \pi_{U_1}, G_{U_1})$, $(\widetilde{U}_2, \pi_{U_2}, G_{U_2})$ respectively, we say that these two orbifold charts are compatible if, for any point $z \in U_1 \cap U_2$, there exists an open connected neighborhood $V \subset U_1 \cap U_2$ of z with an orbifold chart $(\widetilde{V}, \pi_V, G_V)$ such that there exist two embeddings of orbifold charts $\phi_{U_1 V} : (\widetilde{V}, \pi_V, G_V) \to (\widetilde{U}_1, \pi_{U_1}, G_{U_1})$, $\phi_{U_2 V} : (\widetilde{V}, \pi_V, G_V) \to (\widetilde{U}_2, \pi_{U_2}, G_{U_2})$. In this case, the diffeomorphism $\phi_{U_2 V} \circ \phi_{U_1 V}^{-1} : \phi_{U_1 V}(\widetilde{V}) \to \phi_{U_2 V}(\widetilde{V})$ is called a coordinate transformation.

Definition 2.1.2. An orbifold atlas on Z is couple $(\mathcal{U}, \widetilde{\mathcal{U}})$ consisting of a cover $\mathcal{U}$ of open connected subsets of Z and a family of compatible orbifold charts $\widetilde{\mathcal{U}} = \{(\widetilde{U}, \pi_U, G_U)\}_{U \in \mathcal{U}}$.

An orbifold atlas $(\mathcal{V}, \widetilde{\mathcal{V}})$ is called a refinement of $(\mathcal{U}, \widetilde{\mathcal{U}})$ if $\mathcal{V}$ is a refinement of $\mathcal{U}$ and if every orbifold chart in $\widetilde{\mathcal{V}}$ has an embedding into some orbifold chart in $\widetilde{\mathcal{U}}$. Two orbifold atlas are said to be equivalent if they have a common refinement, and the equivalent class of an orbifold atlas is called an orbifold structure on Z.

An orbifold is a second countable Hausdorff space equipped with an orbifold structure. It is said to have dimension m if all the orbifold charts which define the orbifold structure are of dimension m.

If Z, Y are two orbifolds, a smooth map $f : Z \to Y$ is a continuous map from Z to Y such that it lifts locally to an equivariant smooth map from an orbifold chart of Z to any orbifold chart of Y. In this way, we can define the notion of smooth functions and the smooth action of Lie groups.

By [Shen and Yu 2022, Proposition 2.12], if Γ is discrete group acting smoothly and properly discontinuously on the left on an orbifold X, then $Z = \Gamma\backslash X$ has a canonical orbifold structure induced from X.

In the sequel, let Z be an orbifold with an orbifold structure given by $(\mathcal{U}, \widetilde{\mathcal{U}})$. If $z \in Z$, there exists an open connected neighborhood U_z of z with a compatible orbifold chart $(\widetilde{U}_z, G_z, \pi_z)$ such that $\pi_z^{-1}(z)$ contains only one point $x \in \widetilde{U}_z$. Then G_z does not depend on the choice of such open connected neighborhood (up to canonical isomorphisms compatible with the orbifold structure), and G_z is called the local group at z.

Put

$$Z_{\mathrm{reg}} = \{z \in Z \mid G_z = \{1\}\}, \quad Z_{\mathrm{sing}} = \{z \in Z \mid G_z \neq \{1\}\}. \tag{2.1.2}$$

Then Z_{reg} is naturally a smooth manifold. But Z_{sing} is not necessarily an orbifold. Kawasaki [1978, Section 2] explained two different methods to view Z_{sing} as an immersed image of a disjoint union of orbifolds. We just recall that method which appears naturally in Kawasaki's local index theorems for orbifolds [1978; 1979].

If $z \in Z_{\mathrm{sing}}$, let $1 = (h_z^0), (h_z^1), \dots, (h_z^{l_z})$ be the conjugacy classes in G_z. Put

$$\Sigma Z = \{(z, (h_z^j)) \mid z \in Z_{\mathrm{sing}}, j = 1, \dots, l_z\}. \tag{2.1.3}$$

Let $(\widetilde{U}_z, G_z, \pi_z)$ be the local orbifold chart for $z \in Z_{\mathrm{sing}}$ such that $\pi_z^{-1}(z)$ contains only one point. For $j = 1, \dots, l_z$, let $\widetilde{U}_z^{h_z^j} \subset \widetilde{U}_z$ be the fixed-point set of h_z^j, which is a submanifold of $\widetilde{U}_z$. Note that $\widetilde{U}_z^{h_z^j} \subset Z_{\mathrm{sing}}$. Let $Z_{G_z}(h_z^j)$ be the centralizer of h_z^j in G_z. Then $Z_{G_z}(h_z^j)$ acts smoothly on $\widetilde{U}_z^{h_z^j}$. Put

$$K_z^j = \ker(Z_{G_z}(h_z^j) \to \mathrm{Aut}(\widetilde{U}_z^{h_z^j})). \tag{2.1.4}$$

Then $(\widetilde{U}_z^{h_z^j}, Z_{G_z}(h_z^j)/K_z^j, \pi_z^j : \widetilde{U}_z^{h_z^j} \to \widetilde{U}_z^{h_z^j}/Z_{G_z}(h_z^j))$ defines an orbifold chart near $(z, (h_z^j)) \in \Sigma Z$. They form an orbifold structure for ΣZ. Let Z^i, $i = 1, \dots, l$, denote the connected components of the orbifold ΣZ.

The integer $m_z^j = |K_z^j|$ is called the multiplicity of ΣZ in Z at $(z, (h_z^j))$. This defines a function $m : \Sigma Z \to \mathbb{Z}_+$. As explained in [Kawasaki 1978, Section 1], m is locally constant on ΣZ, and let $m_i \in \mathbb{Z}_+$ be the value of m on Z^i for $i = 1, \dots, l$. We call m_i the multiplicity of Z^i in Z. We will put

$$Z^0 = Z, \quad m_0 = 1. \tag{2.1.5}$$

Remark 2.1.3. In Definition 2.1.1, for an orbifold chart, we require the action G_U on $\widetilde{U}$ to be effective. To emphasize this condition, the orbifold defined above is often called an effective orbifold. In fact, we can drop this effectiveness; then we get a general version of the (possibly ineffective) orbifold, for example, using the orbifold groupoid; see [Adem et al. 2007, Definition 1.38]. The point-view of orbifold groupoid provides a unified way to deal with effective and ineffective orbifolds.

As explained in [Adem et al. 2007, Example 2.5], for global quotient groupoids (including all the effective orbifolds and certain ineffective orbifolds), a natural stratification called the inertia groupoid was introduced as an extension of the one $\bigcup_{i=0}^{l} Z^i$ defined in (2.1.3)–(2.1.5). It plays a key role in the study of the geometry of orbifolds. We will go back to this point in Sections 3.4 and 3.5. Through this paper, the terminology orbifold will always refer to the effective one unless otherwise stated.

We say E is an orbifold vector bundle of rank r on Z if there exists a smooth map of orbifolds $\pi : E \to Z$ such that, for any $U \in \mathcal{U}$ and $(\widetilde{U}, G_U, \pi_U) \in \widetilde{\mathcal{U}}$, there exists an orbifold chart $(\widetilde{U}^E, G_U^E, \pi_U^E)$ of E such that $\widetilde{U}^E$ is an vector bundle on $\widetilde{U}$ of rank r equipped an effective action of G_U^E and $\pi_U^E(\widetilde{U}^E) = \pi^{-1}(U)$. Moreover, there exists a surjective group morphism $\psi_U : G_U^E \to G_U$ such that the action of G_U^E on $\widetilde{U}$ is identified via ψ_U with the action of G_U on $\widetilde{U}$. If we have an open embedding $\phi_{UV} : (\widetilde{V}, \pi_V, G_V) \to (\widetilde{U}, \pi_U, G_U)$, we require that it lifts to the open embedding $\phi_{UV}^E : (\widetilde{V}^E, \pi_V^E, G_V^E) \to (\widetilde{U}^E, \pi_U^E, G_U^E)$ of the orbifold charts of E such that $\phi_{UV}^E : \widetilde{V}^E \to \widetilde{U}^E$ is a morphism of vector bundles associated with the open embedding $\phi_{UV} : \widetilde{V} \to \widetilde{U}$. If every $\psi_U : G_U^E \to G_U$ is an isomorphism of groups, we call E a proper orbifold vector bundle on Z.

Note that if E is proper, then the rank of E can be extended to a locally constant function ρ on ΣZ. The orbifold chart of Z^i is given by the triples such as

$$(\widetilde{U}_z^{h_z^j}, Z_{G_z}(h_z^j)/K_z^j, \pi_z^j : \widetilde{U}_z^{h_z^j} \to \widetilde{U}_z^{h_z^j}/Z_{G_z}(h_z^j)).$$

By the above definition of E, we have an orbifold chart $(\widetilde{U}^E, G_U^E = G_U, \pi_U^E)$ such that $\widetilde{U}^E$ is a G_U-equivariant vector bundle on $\widetilde{U}$. Then, for $x \in \widetilde{U}_z^{h_z^j}$, h_z^j acts on the fibers $\widetilde{U}_z^E$ linearly, so that we can set $\rho(z, (h_z^j)) = \mathrm{Tr}^{\widetilde{U}_z^E}[h_z^j]$. Then ρ is really a locally constant function on ΣZ. For $i = 1, \ldots, l$, let ρ_i be the value of ρ on the component Z^i. We also put $\rho_0 = r$.

We call $s : Z \to E$ a smooth section of E over Z if it is a smooth map between orbifolds such that $\pi \circ s = \mathrm{Id}_Z$. We will use $C^\infty(Z, E)$ to denote the vector space of smooth sections of E over Z.

Take an orbifold chart $(\widetilde{U}, G_U, \pi_U) \in \widetilde{\mathcal{U}}$ of Z. Then G_U acts canonically on the tangent vector bundle $T\widetilde{U}$ of $\widetilde{U}$. The open embeddings of orbifold charts of Z also lift to the open embeddings of their tangent vector bundles. This way, we get a proper orbifold vector bundle TZ on Z, and the projection $\pi : TZ \to Z$ is just given by the obvious projection $T\widetilde{U} \to \widetilde{U}$. We call TZ the tangent vector bundle of Z. If we equipped TZ with Euclidean metric g^{TZ}, we will call Z a Riemannian orbifold and call g^{TZ} a Riemannian metric of Z.

Let $\Omega^\bullet(Z)$ denote the set of smooth differential forms of Z, which has a $\mathbb{Z}$-graded structure by degrees. The de Rham differential $d^Z : \Omega^\bullet(Z) \to \Omega^{\bullet+1}(Z)$ is given by the family of de Rham differential operators $d^{\widetilde{U}} : \Omega^\bullet(\widetilde{U}) \to \Omega^{\bullet+1}(\widetilde{U})$. Then we can define the de Rham complex $(\Omega^\bullet(Z), d^Z)$ of Z and the associated de Rham cohomology $H^\bullet(Z, \mathbb{R})$. By [Kawasaki 1978, Section 1], there is a natural isomorphism between $H^\bullet(Z, \mathbb{R})$ and the singular cohomology of the underlying topological space Z.

Now let us recall the integrals on Z. Assume that Z is compact. We may take a finite open covering $\{U_i\}_{i \in I}$ of the precompact orbifold charts for Z. Since Z is Hausdorff, there exists a partition of unity subordinate to this open cover. We can find a family of smooth functions $\{\phi_i \in C_c^\infty(Z)\}_{i\in I}$ with values in $[0,1]$ such that $\mathrm{Supp}(\phi_i) \subset U_i$, and that

$$\sum_{i\in I} \phi_i = 1. \tag{2.1.6}$$

Take $\tilde{\phi}_i = \pi_{U_i}^*(\phi_i) \in C_c^\infty(\widetilde{U}_i)^{G_{U_i}}$.

If $\alpha \in \Omega^m(Z, o(TZ))$, let $\tilde{\alpha}_{U_i}$ be its lift on the chart $(\widetilde{U}_i, \pi_{U_i}, G_{U_i})$. We define

$$\int_Z \alpha = \sum_i \frac{1}{|G_{U_i}|} \int_{\widetilde{U}_i} \tilde{\phi}_i \tilde{\alpha}_{U_i}. \tag{2.1.7}$$

By [Shen and Yu 2022, Section 3.2], if $\alpha \in \Omega^m(Z, o(TZ))$, then α is also integrable on Z_{reg}, so that

$$\int_Z \alpha = \int_{Z_{\mathrm{reg}}} \alpha. \tag{2.1.8}$$

Also if $\alpha \in \Omega^\bullet(Z, o(TZ))$, we have

$$\int_Z d^Z\alpha = 0. \tag{2.1.9}$$

If (Z, g^{TZ}) is a Riemannian orbifold, we can define the integration of functions on Z with respect to the Riemannian volume element. If we have a Hermitian orbifold vector bundle $(F, h^F) \to (Z, g^{TZ})$, one can define the L_2 scalar product for the space of continuous sections of F as usual. Then, after completion, we get the Hilbert space $L^2(Z, F)$.

Chern–Weil theory on the characteristic forms extends to orbifolds, where their constructions are parallel to the case of smooth manifolds. We refer to [Shen and Yu 2022, Section 3.4] for more details. Note that the characteristic forms are not only defined on Z but also defined on ΣZ. The part ΣZ has a nontrivial contribution in Kawasaki's local index theorems for orbifolds [1978; 1979].

Finally, we introduce the orbifold Euler characteristic number of (Z, g^{TZ}) [Satake 1957]. Let $\nabla^{TZ} = \{\nabla^{T\widetilde{U}_i}\}_{U_i\in\mathcal{U}}$ be the Levi-Civita connection on TZ associated with g^{TZ}. The Euler form $e(TZ, \nabla^{TZ}) \in \Omega^m(Z, o(TZ))$ is given by the family of closed forms

$$\{e(\widetilde{U}_i, \nabla^{T\widetilde{U}_i}) \in \Omega^m(\widetilde{U}_i, o(T\widetilde{U}_i))^{G_{U_i}}\}_{U_i\in\mathcal{U}}. \tag{2.1.10}$$

If Z is oriented, then we can view $e(TZ, \nabla^{TZ})$ as a differential form on Z.

If Z is compact, set

$$\chi_{\mathrm{orb}}(Z) = \int_Z e(TZ, \nabla^{TZ}). \tag{2.1.11}$$

By [Satake 1957, Section 3], $\chi_{\mathrm{orb}}(Z)$ is a rational number, and it vanishes when Z is odd-dimensional.

2.2. ***Flat vector bundles and analytic torsions of orbifolds.*** If (F, ∇^F) is an orbifold vector bundle over Z with a connection ∇^F, we call (F, ∇^F) a flat vector bundle if the curvature $R^F = \nabla^{F,2}$ vanishes identically on Z. A detailed discussion for the flat vector bundles on Z is given in [Shen and Yu 2022, Sections 2.3–2.5].

Let (Z, g^{TZ}) be a compact Riemannian orbifold of dimension m. Let (F, ∇^F) be a flat complex orbifold vector bundle of rank r on Z with Hermitian metric h^F. Note that we do not assume that F is proper.

Let $\Omega^\bullet(Z,F)$ be the set of smooth sections of $\Lambda^\bullet(T^*Z)\otimes F$ on Z. Let d^Z be the exterior differential acting on $\Omega^\bullet(Z,\mathbb{R})$.

Definition 2.2.1. For $i=0,1,\dots,m$, if $\alpha\in\Omega^i(Z,\mathbb{R})$, $s\in C^\infty(Z,F)$, the operator $d^{Z,F}$ acting on $\Omega^i(Z,F)$ is defined by

$$d^{Z,F}(\alpha\otimes s)=(d^Z\alpha)\otimes s+(-1)^i\alpha\wedge\nabla^F s\in\Omega^{i+1}(Z,F). \tag{2.2.1}$$

Since $\nabla^{F,2}=0$, then $(\Omega^\bullet(Z,F),d^{Z,F})$ is a complex, which is called the de Rham complex for the flat orbifold vector bundle (F,∇^F) on Z. Let $H^\bullet(Z,F)$ denote the corresponding de Rham cohomology group of Z valued in F, as in the case of closed manifolds, $H^\bullet(Z,F)$ is always finite-dimensional.

Let $\langle\cdot,\cdot\rangle_{\Lambda^\bullet(T^*Z)\otimes F,z}$ be the Hermitian metric on $\Lambda^\bullet(T^*_zZ)\otimes F_z$, $z\in Z$ induced by g^{TZ}_z and h^F_z. Let dv be the Riemannian volume element on Z induced by g^{TZ}. The L_2-scalar product on $\Omega^\bullet(Z,F)$ is given as follows: if $s,s'\in\Omega^\bullet(Z,F)$, then

$$\langle s,s'\rangle_{L^2}=\int_Z\langle s(z),s(z')\rangle_{\Lambda^\bullet(T^*Z)\otimes F,z}\,dv(z). \tag{2.2.2}$$

By (2.1.8), it will be the same if we take the integrals on Z_{reg}.

Let $d^{Z,F,*}$ be the formal adjoint of $d^{Z,F}$ with respect to the above L_2-metric on $\Omega^\bullet(Z,F)$; i.e., for $s,s'\in\Omega^\bullet(Z,F)$,

$$\langle d^{Z,F,*}s,s'\rangle_{L^2}=\langle s,d^{Z,F}s'\rangle_{L^2}. \tag{2.2.3}$$

Then $d^{Z,F,*}$ is a first-order differential operator acting $\Omega^\bullet(Z,F)$ on which decreases the degree by 1.

Definition 2.2.2. The de Rham–Hodge operator $\boldsymbol{D}^{Z,F}$ of $\Omega^\bullet(Z,F)$ is defined as

$$\boldsymbol{D}^{Z,F}=d^{Z,F}+d^{Z,F,*}. \tag{2.2.4}$$

It is a first-order self-adjoint elliptic differential operator acting on $\Omega^\bullet(Z,F)$.

The Hodge Laplacian is

$$\boldsymbol{D}^{F,Z,2}=[d^{Z,F},d^{Z,F,*}]=d^{Z,F}d^{Z,F,*}+d^{Z,F,*}d^{Z,F}. \tag{2.2.5}$$

Here, $[\cdot,\cdot]$ denotes the supercommutator. Then $\boldsymbol{D}^{Z,F,2}$ is a second-order essentially self-adjoint nonnegative elliptic operator, which preserves the degree.

The Hodge decomposition for $\Omega^\bullet(Z,F)$ still holds in this case (see [Ma 2005, Proposition 2.2; Dai and Yu 2017, Proposition 2.1]),

$$\Omega^\bullet(Z,F)=\ker(\boldsymbol{D}^{Z,F,2}|_{\Omega^\bullet(Z,F)})\oplus\mathrm{Im}(d^{Z,F}|_{\Omega^{\bullet-1}(Z,F)})\oplus\mathrm{Im}(d^{Z,F,*}|_{\Omega^{\bullet+1}(Z,F)}). \tag{2.2.6}$$

Then we have the canonical identification of vector spaces,

$$\mathcal{H}^\bullet(Z,F):=\ker\boldsymbol{D}^{Z,F,2}\simeq H^\bullet(Z,F). \tag{2.2.7}$$

Put

$$\chi(Z,F)=\sum_{j=0}^m(-1)^j\dim H^j(Z,F). \tag{2.2.8}$$

If F is proper, recall that the numbers ρ_i, $i = 0, \dots, l$, are defined in previous subsection as the extension of the rank of F. Then by [Shen and Yu 2022, Theorem 4.3], we have

$$\chi(Z, F) = \sum_{i=0}^{l} \rho_i \frac{\chi_{\mathrm{orb}}(Z_i)}{m_i}. \tag{2.2.9}$$

The right-hand side of (2.2.9) contains the nontrivial contributions from ΣZ.

Let P denote the orthogonal projection from $\Omega^\bullet(Z, F)$ to $\mathcal{H}^\bullet(Z, F)$. Let $\mathcal{H}^\perp$ denote the orthogonal subspace of $\mathcal{H}^\bullet(Z, F)$ in $\Omega^\bullet(Z, F)$, and let $(\boldsymbol{D}^{Z,F,2})^{-1}$ be the inverse of $\boldsymbol{D}^{Z,F,2}$ acting on $\mathcal{H}^\perp$. Let $N^{\Lambda^\bullet(T^*Z)}$ be the number operator on $\Lambda^\bullet(T^*Z)$ which acts on $\Lambda^j(T^*Z)$ by multiplication of j.

For $s \in \mathbb{C}$, $\Re(s)$ is large enough; set

$$\begin{aligned} \vartheta(F)(s) &= -\operatorname{Tr}_s[N^{\Lambda^\bullet(T^*Z)}(\boldsymbol{D}^{Z,F,2})^{-s}] \\ &= -\frac{1}{\Gamma(s)} \int_0^{+\infty} \operatorname{Tr}_s[N^{\Lambda^\bullet(T^*Z)} \exp(-t\boldsymbol{D}^{Z,F,2})(1-P)] t^{s-1}\, dt, \end{aligned} \tag{2.2.10}$$

where $\Gamma(s)$ is the gamma function for $s \in \mathbb{C}$. By the short time asymptotic expansions of the heat trace (see [Ma 2005, Proposition 2.1]), $\vartheta(F)(s)$ admits a unique meromorphic extension to $s \in \mathbb{C}$ which is holomorphic at $s = 0$.

Definition 2.2.3. Let $\mathcal{T}(g^{TZ}, \nabla^F, h^F) \in \mathbb{R}$ be given by

$$\mathcal{T}(g^{TZ}, \nabla^F, h^F) = \frac{d}{ds}\Big|_{s=0} \vartheta(F)(s). \tag{2.2.11}$$

The quantity $\mathcal{T}(g^{TZ}, \nabla^F, h^F)$ is called Ray–Singer analytic torsion associated with (F, ∇^F, h^F).

By [Shen and Yu 2022, Proposition 4.6, Corollary 4.9], for an orientable closed orbifold Z, if m is even and F is unitarily flat, then $\mathcal{T}(g^{TZ}, \nabla^F, h^F) = 0$; if m is odd and F is acyclic, then $\mathcal{T}(g^{TZ}, \nabla^F, h^F)$ is independent of the metrics g^{TZ} and h^F.

Now we explain how to evaluate $\mathcal{T}(g^{TZ}, \nabla^F, h^F)$ in practice when F is acyclic. Using the analogous arguments in [Bismut and Zhang 1992, Theorem 7.10, Section XI], as $t \to 0^+$, the heat supertrace $\operatorname{Tr}_s[(N^{\Lambda^\bullet(T^*Z)} - m/2)\exp(-t\boldsymbol{D}^{Z,F,2}/2)]$ either has a leading term as a multiple of $1/\sqrt{t}$ or is a small quantity as $\mathcal{O}(\sqrt{t})$; see [Shen and Yu 2022, equation (4.37)]. To deal with this possible divergent term $1/\sqrt{t}$ in the integral of (2.2.10), we proceed as in the proof of [Bismut and Lott 1995, Theorem 3.29]. For $t > 0$, put

$$b_t(g^{TZ}, F) = \Big(1 + 2t\frac{\partial}{\partial t}\Big) \operatorname{Tr}_s\Big[\Big(N^{\Lambda^\bullet(T^*Z)} - \frac{m}{2}\Big)\exp\Big(-\frac{t\boldsymbol{D}^{Z,F,2}}{2}\Big)\Big]. \tag{2.2.12}$$

By [Bismut and Zhang 1992, Theorem 7.10; Bismut and Lott 1995, Theorem 2.13; Shen and Yu 2022, Section 4.3] and since F is acyclic, as $t \to 0$,

$$b_t(g^{TZ}, F) = \mathcal{O}(\sqrt{t}); \tag{2.2.13}$$

as $t \to +\infty$,

$$b_t(g^{TZ}, F) = \mathcal{O}\Big(\frac{1}{\sqrt{t}}\Big). \tag{2.2.14}$$

By [Bismut and Lott 1995, Theorem 3.29; Shen and Yu 2022, Corollary 4.14], we have

$$\mathcal{T}(g^{TZ}, \nabla^F, h^F) = -\int_0^{+\infty} b_t(g^{TZ}, F)\, \frac{dt}{t}. \tag{2.2.15}$$

One particular case is that if, for $t > 0$, we always have

$$\operatorname{Tr}_s\Big[\Big(N^{\Lambda^\bullet(T^*Z)} - \frac{m}{2}\Big) \exp\Big(-\frac{t\boldsymbol{D}^{Z,F,2}}{2}\Big)\Big] = 0, \tag{2.2.16}$$

then $\mathcal{T}(g^{TZ}, \nabla^F, h^F) = 0$. This holds even for nonacyclic F.

3. Orbital integrals and locally symmetric spaces

In this section, we recall the geometry of the symmetric space X, and we recall an explicit geometric formula for semisimple orbital integrals obtained in [Bismut 2011, Chapter 6] . Then, given a cocompact discrete subgroup $\Gamma \subset G$, we describe the orbifold structure on $Z = \Gamma\backslash X$, and we give Selberg's trace formula for Z.

In this section, G is taken to be a connected linear real reductive Lie group; we do not require that it has a compact center. Then X is a symmetric space which may have de Rham components of both noncompact type and Euclidean type.

3.1. ***Real reductive Lie group.*** Let G be a connected linear real reductive Lie group with Lie algebra $\mathfrak{g}$, and let $\theta \in \operatorname{Aut}(G)$ be a Cartan involution. Let K be the fixed-point set of θ in G. Then K is a maximal compact subgroup of G, and let $\mathfrak{k}$ be its Lie algebra. Let $\mathfrak{p} \subset \mathfrak{g}$ be the eigenspace of θ associated with the eigenvalue -1. The Cartan decomposition of $\mathfrak{g}$ is given by

$$\mathfrak{g} = \mathfrak{p} \oplus \mathfrak{k}. \tag{3.1.1}$$

Put $m = \dim \mathfrak{p},\ n = \dim \mathfrak{k}$.

Let B be a G- and θ-invariant nondegenerate symmetric bilinear form on $\mathfrak{g}$, which is positive on $\mathfrak{p}$ and negative on $\mathfrak{k}$. It induces a symmetric bilinear form B^* on $\mathfrak{g}^*$, which extends to a symmetric bilinear form on $\Lambda^\bullet(\mathfrak{g}^*)$. The K-invariant bilinear form $\langle\cdot,\cdot\rangle = -B(\cdot, \theta\,\cdot)$ is a scalar product on $\mathfrak{g}$, which extends to a scalar product on $\Lambda^\bullet(\mathfrak{g}^*)$. We will use $|\cdot|$ to denote the norm under this scalar product.

Let $U\mathfrak{g}$ be the universal enveloping algebra of $\mathfrak{g}$. Let $C^{\mathfrak{g}} \in U\mathfrak{g}$ be the Casimir element associated with B; i.e., if $\{e_i\}_{i=1,\dots,m+n}$ is a basis of $\mathfrak{g}$, and if $\{e_i^*\}_{i=1,\dots,m+n}$ is the dual basis of $\mathfrak{g}$ with respect to B, then

$$C^{\mathfrak{g}} = -\sum e_i^* e_i. \tag{3.1.2}$$

We can identify $U\mathfrak{g}$ with the algebra of left-invariant differential operators over G; then $C^{\mathfrak{g}}$ is a second-order differential operator, which is $\operatorname{Ad}(G)$-invariant. Similarly, let $C^{\mathfrak{k}} \in U\mathfrak{k}$ denote the Casimir operator associated with $(\mathfrak{k}, B|_{\mathfrak{k}})$.

Let $\mathfrak{z}_{\mathfrak{g}} \subset \mathfrak{g}$ be the center of $\mathfrak{g}$. Put

$$\mathfrak{g}_{\mathrm{ss}} = [\mathfrak{g}, \mathfrak{g}]. \tag{3.1.3}$$

Then

$$\mathfrak{g} = \mathfrak{z}_{\mathfrak{g}} \oplus \mathfrak{g}_{\mathrm{ss}}. \tag{3.1.4}$$

They are orthogonal with respect to B.

Let Z_G be the center of G, and let G_{ss} be the closed analytic subgroup of G associated with $\mathfrak{g}_{\mathrm{ss}}$; see [Knapp 2002, Corollary 7.11]. Then G is the commutative product of Z_G and G_{ss}; in particular,

$$G = Z_G^0 G_{\mathrm{ss}}. \tag{3.1.5}$$

Let $i = \sqrt{-1}$ denote one square root of -1. Put

$$\mathfrak{u} = \sqrt{-1}\mathfrak{p} \oplus \mathfrak{k}. \tag{3.1.6}$$

For simplicity, if $a \in \mathfrak{p}$, we write ia or $\sqrt{-1}a \in \sqrt{-1}\mathfrak{p} \subset \mathfrak{u}$ to denote the corresponding vector.

Then $\mathfrak{u}$ is a (real) Lie algebra, which is called the compact form of $\mathfrak{g}$. Then

$$\mathfrak{g}_{\mathbb{C}} = \mathfrak{u}_{\mathbb{C}}. \tag{3.1.7}$$

Let $G_{\mathbb{C}}$ be the complexification of G with Lie algebra $\mathfrak{g}_{\mathbb{C}}$, which is closed and linear reductive [Knapp 1986, Proposition 5.6]. Then G is the analytic subgroup of $G_{\mathbb{C}}$ with Lie algebra $\mathfrak{g}$. Let $U \subset G_{\mathbb{C}}$ be the analytic subgroup associated with $\mathfrak{u}$. If G has compact center, i.e., $\mathfrak{z}_{\mathfrak{g}} \cap \mathfrak{p} = \{0\}$, then by [Knapp 1986, Proposition 5.3], U is compact; since $G_{\mathbb{C}}$ is closed, U is a maximal compact subgroup of $G_{\mathbb{C}}$.

Definition 3.1.1. An element $\gamma \in G$ is said to be semisimple if there exists $g \in G$ such that

$$\gamma = g(e^a k)g^{-1}, \quad a \in \mathfrak{p}, \quad k \in K, \quad \mathrm{Ad}(k)a = a. \tag{3.1.8}$$

We call $\gamma_h = ge^a g^{-1}$ and $\gamma_e = gkg^{-1}$ the hyperbolic and elliptic parts of γ. These two parts are uniquely determined by γ. If $\gamma_h = 1$, we say γ is elliptic, and if $\gamma_e = 1$ and $\gamma_h \neq 1$, we say γ is hyperbolic.

Let $Z(\gamma)$ be the centralizer of γ in G. If $v \in \mathfrak{g}$, let $Z(v) \subset G$ be the stabilizer of v in G via the adjoint action. Let $\mathfrak{z}(\gamma)$, $\mathfrak{z}(v)$ be the Lie algebras of $Z(\gamma)$, $Z(v)$ respectively. If $\gamma = \gamma_h\gamma_e$ is semisimple as above, by [Eberlein 1996, Theorem 2.19.23; Knapp 2002, Lemma 7.36],

$$Z(\gamma) = Z(\gamma_h) \cap Z(\gamma_e), \quad Z(\gamma_h) = Z(\mathrm{Ad}(g)a). \tag{3.1.9}$$

By [Knapp 2002, Proposition 7.25], $Z(\gamma)$ is reductive (possibly with several connected components). Set

$$\theta_g = C(g)\theta C(g^{-1}). \tag{3.1.10}$$

Then θ_g defines a Cartan involution on $Z(\gamma)$. Let $K(\gamma)$ be the fixed-point set of θ_g in $Z(\gamma)$; then

$$K(\gamma) = Z(\gamma) \cap gKg^{-1}. \tag{3.1.11}$$

Let $Z(\gamma)^0$, $K(\gamma)^0$ be the connected components of the identities of $Z(\gamma)$, $K(\gamma)$ respectively. By [Bismut 2011, Theorem 3.3.1],

$$\frac{Z(\gamma)}{K(\gamma)} = \frac{Z(\gamma)^0}{K(\gamma)^0}. \tag{3.1.12}$$

Moreover, $K(\gamma)$, $K(\gamma)^0$ are maximal compact subgroups of $Z(\gamma)$, $Z(\gamma)^0$ respectively.

Taking the corresponding Lie algebras in (3.1.9), we have

$$\mathfrak{z}(\gamma) = \mathfrak{z}(\gamma_h) \cap \mathfrak{z}(\gamma_e), \quad \mathfrak{z}(\gamma_h) = \mathfrak{z}(\mathrm{Ad}(g)a). \tag{3.1.13}$$

Let $\mathfrak{k}(\gamma) \subset \mathfrak{z}(\gamma)$ be the Lie algebra of $K(\gamma)$. Put

$$\mathfrak{p}(\gamma) = \mathfrak{z}(\gamma) \cap \mathrm{Ad}(g)\mathfrak{p}. \tag{3.1.14}$$

Then the Cartan decomposition of $\mathfrak{z}(\gamma)$ with respect to θ_g is given by

$$\mathfrak{z}(\gamma) = \mathfrak{k}(\gamma) \oplus \mathfrak{p}(\gamma). \tag{3.1.15}$$

Let $B_{\mathfrak{z}(\gamma)}$ denote the restriction of B on $\mathfrak{z}(\gamma) \times \mathfrak{z}(\gamma)$. Then $B_{\mathfrak{z}(\gamma)}$ is invariant under the adjoint action of θ_g on $\mathfrak{z}(\gamma)$. Moreover, $B_{\mathfrak{z}(\gamma)}$ is positive on $\mathfrak{p}(\gamma)$ and negative on $\mathfrak{k}(\gamma)$. The splitting in (3.1.15) is orthogonal with respect to $B_{\mathfrak{z}(\gamma)}$.

3.2. ***Symmetric space.*** Set

$$X = G/K. \tag{3.2.1}$$

Then X is a smooth manifold with the smooth structure induced by G. By definition, X is diffeomorphic to $\mathfrak{p}$.

Let $\omega^{\mathfrak{g}} \in \Omega^1(G, \mathfrak{g})$ be the canonical left-invariant 1-form on G. Then by (3.1.1),

$$\omega^{\mathfrak{g}} = \omega^{\mathfrak{p}} + \omega^{\mathfrak{k}}. \tag{3.2.2}$$

Let $p : G \to X$ denote the obvious projection. Then p is a K-principal bundle over X. Then $\omega^{\mathfrak{k}}$ is a connection form of this principal bundle. The associated curvature form

$$\Omega^{\mathfrak{k}} = d\omega^{\mathfrak{k}} + \tfrac{1}{2}[\omega^{\mathfrak{k}}, \omega^{\mathfrak{k}}] = -\tfrac{1}{2}[\omega^{\mathfrak{p}}, \omega^{\mathfrak{p}}]. \tag{3.2.3}$$

If (E, ρ^E, h^E) is a finite-dimensional unitary or Euclidean representation of K, then $F = G \times_K E$ defines a vector bundle over X equipped with a metric h^F induced by h^E and a unitary or a Euclidean connection ∇^F induced by $\omega^{\mathfrak{k}}$. Note that G acts on $(F, h^F, \nabla^F) \to X$ equivariantly on the left; more precisely, for $\gamma \in G$, $(g, v) \in G \times_K E$, the action of γ on F is represented by

$$\gamma(g, v) = (\gamma g, v) \in G \times_K E. \tag{3.2.4}$$

In particular, we have the identification

$$TX = G \times_K \mathfrak{p}, \tag{3.2.5}$$

where the right-hand side is defined by the adjoint action of K on $\mathfrak{p}$. The bilinear form B restricting to $\mathfrak{p}$ gives a Riemannian metric g^{TX}, and $\omega^{\mathfrak{k}}$ induces the associated Levi-Civita connection ∇^{TX}. Then G acts on (X, g^{TX}) isometrically. Let $d(\cdot, \cdot)$ denote the Riemannian distance on X.

Let $C(G, E)$ denote the set of continuous map from G into E. If $k \in K$, $s \in C(G, E)$, put

$$(k.s)(g) = \rho^E(k)s(gk). \tag{3.2.6}$$

Let $C_K(G, E)$ be the set of K-invariant maps in $C(G, E)$. Let $C(X, F)$ denote the continuous sections of F over X. Then

$$C_K(G, E) = C(X, F). \tag{3.2.7}$$

Also $C_K^{\infty}(G, E) = C^{\infty}(X, F)$.

The Casimir operator $C^{\mathfrak{g}}$ acting on $C^\infty(G,E)$ preserves $C^\infty_K(G,E)$, so it induces an operator $C^{\mathfrak{g},X}$ acting on $C^\infty(X,F)$. Let $\Delta^{H,X}$ be the Bochner Laplacian acting on $C^\infty(X,F)$ given by ∇^F, and let $C^{\mathfrak{k},E}\in \operatorname{End}(E)$ be the action of the Casimir $C^{\mathfrak{k}}$ on E via ρ^E. The element $C^{\mathfrak{k},E}$ induces a self-adjoint section of $\operatorname{End}(F)$ over X. Then

$$C^{\mathfrak{g},X}=-\Delta^{H,X}+C^{\mathfrak{k},E}. \tag{3.2.8}$$

Let $C^{\mathfrak{k},\mathfrak{p}}\in\operatorname{End}(\mathfrak{p}),\ C^{\mathfrak{k},\mathfrak{k}}\in\operatorname{End}(\mathfrak{k})$ be the actions of $C^{\mathfrak{k}}$ acting on $\mathfrak{p}$, $\mathfrak{k}$ via the adjoint actions. Moreover, we can also view $C^{\mathfrak{k},\mathfrak{p}}$ as a parallel section of $\operatorname{End}(TX)$.

If $A\in\operatorname{End}(E)$ commutes with K, then it can be viewed a parallel section of $\operatorname{End}(F)$ over X. Let dx be the Riemannian volume element of (X,g^{TX}).

Definition 3.2.1. Let $\mathcal{L}^X_A$ be the Bochner-like Laplacian acting on $C^\infty(X,F)$ given by

$$\mathcal{L}^X_A=\tfrac{1}{2}C^{\mathfrak{g},X}+\tfrac{1}{16}\operatorname{Tr}^{\mathfrak{p}}[C^{\mathfrak{k},\mathfrak{p}}]+\tfrac{1}{48}\operatorname{Tr}^{\mathfrak{k}}[C^{\mathfrak{k},\mathfrak{k}}]+A. \tag{3.2.9}$$

For $t>0,\ x,x'\in X$, let $p^X_t(x,x')$ denote its heat kernel with respect to dx'.

Since $\mathcal{L}^X_A$ is G-invariant, $p^X_t(x,x')$ lifts to a function $p^X_t(g,g')$ on $G\times G$ valued in $\operatorname{End}(E)$ such that, for $g''\in G,\ k,k'\in K$,

$$p^X_t(g''g,g''g')=p^X_t(g,g'),\quad p^X_t(gk,g'k')=\rho^E(k^{-1})p^X_t(g,g')\rho^E(k'). \tag{3.2.10}$$

We set

$$p^X_t(g)=p^X_t(1,g). \tag{3.2.11}$$

Then p^X_t is a $K\times K$-invariant smooth function on G valued in $\operatorname{End}(E)$. We will not distinguish the heat kernel $p^X_t(x,x')$ and the function $p^X_t(g)$ in the sequel.

3.3. ***Bismut's formula for semisimple orbital integrals.*** Let dg be the left-invariant Haar measure on G induced by $(\mathfrak{g},\langle\cdot,\cdot\rangle)$. Since G is unimodular, dg is also right-invariant. Let dk be the Haar measure on K induced by $-B|_{\mathfrak{k}}$; then

$$dg=dx\,dk. \tag{3.3.1}$$

Now let $\gamma\in G$ be a semisimple element given as in (3.1.8).

By [Eberlein 1996, Definition 2.19.21; Bismut 2011, Theorem 3.1.2], $\gamma\in G$ is semisimple if and only if the displacement function $X\ni x\mapsto d(x,\gamma x)$ on X associated with γ can reach its minimum $m_\gamma\geq 0$ in X. In this case, the minimizing set $X(\gamma)$ of this displacement function is a geodesically convex submanifold of X, and by [Bismut 2011, Theorem 3.3.1],

$$X(\gamma)\simeq\frac{Z(\gamma)^0}{K(\gamma)^0}=\frac{Z(\gamma)}{K(\gamma)}. \tag{3.3.2}$$

Moreover, we have

$$m_\gamma=|a|. \tag{3.3.3}$$

Let dy be the Riemannian volume element of $X(\gamma)$, and let dz be the bi-invariant Haar measure on $Z(\gamma)$ induced by $B_{\mathfrak{z}(\gamma)}$. Let $dk(\gamma)$ be the Haar measure on $K(\gamma)$ such that

$$dz = dy\, dk(\gamma). \tag{3.3.4}$$

Let $\mathrm{Vol}(K(\gamma)\backslash K)$ be the volume of $K(\gamma)\backslash K$ with respect to $dk, dk(\gamma)$. Then we have

$$\mathrm{Vol}(K(\gamma)\backslash K) = \frac{\mathrm{Vol}(K)}{\mathrm{Vol}(K(\gamma))}. \tag{3.3.5}$$

Let dv be the G-left invariant measure on $Z(\gamma)\backslash G$ such that

$$dg = dz\, dv. \tag{3.3.6}$$

By [Bismut 2011, Definition 4.2.2, Proposition 4.4.2], for $t > 0$, the orbital integral

$$\mathrm{Tr}^{[\gamma]}[\exp(-t\mathcal{L}_A^X)] = \frac{1}{\mathrm{Vol}(K(\gamma)\backslash K)} \int_{Z(\gamma)\backslash G} \mathrm{Tr}^E[p_t^X(v^{-1}\gamma v)]\, dv \tag{3.3.7}$$

is well-defined. As indicated by the notation, it only depends on the conjugacy class $[\gamma]$ of γ in G.

Using the theory of hypoelliptic Laplacian and the techniques from local index theory, Bismut obtained an explicit geometric formula for $\mathrm{Tr}^{[\gamma]}[\exp(-t\mathcal{L}_A^X)]$ in [Bismut 2011, Theorem 6.1.1] as well as its extension to the wave operators of $\mathcal{L}_A^X$ [Bismut 2011, Section 6.3]. Now we describe in detail this formula. We may and we will assume that

$$\gamma = e^a k, \quad a \in \mathfrak{p}, \quad k \in K, \quad \mathrm{Ad}(k)a = a. \tag{3.3.8}$$

Put

$$\mathfrak{z}_0 = \mathfrak{z}(a), \quad \mathfrak{p}_0 = \ker \mathrm{ad}(a) \cap \mathfrak{p}, \quad \mathfrak{k}_0 = \ker \mathrm{ad}(a) \cap \mathfrak{k}. \tag{3.3.9}$$

Let $\mathfrak{z}_0^\perp$, $\mathfrak{p}_0^\perp$, $\mathfrak{k}_0^\perp$ be the orthogonal vector spaces to $\mathfrak{z}_0$, $\mathfrak{p}_0$, $\mathfrak{k}_0$ in $\mathfrak{g}, \mathfrak{p}, \mathfrak{k}$ with respect to B. Then

$$\mathfrak{z}_0 = \mathfrak{p}_0 \oplus \mathfrak{k}_0, \quad \mathfrak{z}_0^\perp = \mathfrak{p}_0^\perp \oplus \mathfrak{k}_0^\perp. \tag{3.3.10}$$

By [Bismut 2011, equation (3.3.6)],

$$\mathfrak{z}(\gamma) = \mathfrak{z}_0 \cap \mathfrak{z}(k). \tag{3.3.11}$$

Also $\mathfrak{p}(\gamma)$, $\mathfrak{k}(\gamma)$ are subspaces of $\mathfrak{p}_0$, $\mathfrak{k}_0$ respectively. Let $\mathfrak{z}_0^\perp(\gamma)$, $\mathfrak{p}_0^\perp(\gamma)$, $\mathfrak{k}_0^\perp(\gamma)$ be the orthogonal spaces to $\mathfrak{z}(\gamma)$, $\mathfrak{p}(\gamma)$, $\mathfrak{k}(\gamma)$ in $\mathfrak{z}_0$, $\mathfrak{p}_0$, $\mathfrak{k}_0$. Then

$$\mathfrak{z}_0^\perp(\gamma) = \mathfrak{p}_0^\perp(\gamma) \oplus \mathfrak{k}_0^\perp(\gamma). \tag{3.3.12}$$

Also the action $\mathrm{ad}(a)$ gives an isomorphism between $\mathfrak{p}_0^\perp$ and $\mathfrak{k}_0^\perp$.

For $Y_0^{\mathfrak{k}} \in \mathfrak{k}(\gamma)$, $\mathrm{ad}(Y_0^{\mathfrak{k}})$ preserves $\mathfrak{p}(\gamma), \mathfrak{k}(\gamma), \mathfrak{p}_0^\perp(\gamma), \mathfrak{k}_0^\perp(\gamma)$, and it is an antisymmetric endomorphism with respect to the scalar product.

Recall that the function $\widehat{A}$ is given by

$$\widehat{A}(x) = \frac{x/2}{\sinh(x/2)}. \tag{3.3.13}$$

Let H be a finite-dimensional Hermitian vector space. If $B \in \mathrm{End}(H)$ is self-adjoint, then

$$\frac{B/2}{\sinh(B/2)}$$

is a self-adjoint positive endomorphism. Put

$$\widehat{A}(B) = \det{}^{\frac{1}{2}}\left[\frac{B/2}{\sinh(B/2)}\right]. \tag{3.3.14}$$

In (3.3.14), the square root is taken to be the positive square root.

If $Y_0^{\mathfrak{k}} \in \mathfrak{k}(\gamma)$, as explained in [Bismut 2011, p. 105], the following function $A(Y_0^{\mathfrak{k}})$ has a natural square root that is analytic in $Y_0^{\mathfrak{k}} \in \mathfrak{k}(\gamma)$:

$$A(Y_0^{\mathfrak{k}}) = \frac{1}{\det(1-\mathrm{Ad}(k))|_{\mathfrak{z}_0^{\perp}(\gamma)}} \cdot \frac{\det(1-\exp(-i\,\mathrm{ad}(Y_0^{\mathfrak{k}}))\,\mathrm{Ad}(k))|_{\mathfrak{k}_0^{\perp}(\gamma)}}{\det(1-\exp(-i\,\mathrm{ad}(Y_0^{\mathfrak{k}}))\,\mathrm{Ad}(k))|_{\mathfrak{p}_0^{\perp}(\gamma)}}. \tag{3.3.15}$$

Its square root is denoted by

$$\left[\frac{1}{\det(1-\mathrm{Ad}(k))|_{\mathfrak{z}_0^{\perp}(\gamma)}} \cdot \frac{\det(1-\exp(-i\,\mathrm{ad}(Y_0^{\mathfrak{k}}))\,\mathrm{Ad}(k))|_{\mathfrak{k}_0^{\perp}(\gamma)}}{\det(1-\exp(-i\,\mathrm{ad}(Y_0^{\mathfrak{k}}))\,\mathrm{Ad}(k))|_{\mathfrak{p}_0^{\perp}(\gamma)}}\right]^{\frac{1}{2}}. \tag{3.3.16}$$

The value of (3.3.16) at $Y_0^{\mathfrak{k}} = 0$ is taken to be such that

$$\frac{1}{\det(1-\mathrm{Ad}(k))|_{\mathfrak{p}_0^{\perp}(\gamma)}}. \tag{3.3.17}$$

We recall an important function J_γ defined in [Bismut 2011, equation (5.5.5)].

Definition 3.3.1. Let $J_\gamma(Y_0^{\mathfrak{k}})$ be the analytic function of $Y_0^{\mathfrak{k}} \in \mathfrak{k}(\gamma)$ given by

$$\begin{aligned} J_\gamma(Y_0^{\mathfrak{k}}) = {} & \frac{1}{|\det(1-\mathrm{Ad}(\gamma))|_{\mathfrak{z}_0^{\perp}}|^{\frac{1}{2}}} \frac{\widehat{A}(i\,\mathrm{ad}(Y_0^{\mathfrak{k}})|_{\mathfrak{p}(\gamma)})}{\widehat{A}(i\,\mathrm{ad}(Y_0^{\mathfrak{k}})|_{\mathfrak{k}(\gamma)})} \\ & \cdot \left[\frac{1}{\det(1-\mathrm{Ad}(k))|_{\mathfrak{z}_0^{\perp}(\gamma)}} \frac{\det(1-\exp(-i\,\mathrm{ad}(Y_0^{\mathfrak{k}}))\,\mathrm{Ad}(k))|_{\mathfrak{k}_0^{\perp}(\gamma)}}{\det(1-\exp(-i\,\mathrm{ad}(Y_0^{\mathfrak{k}}))\,\mathrm{Ad}(k))|_{\mathfrak{p}_0^{\perp}(\gamma)}}\right]^{\frac{1}{2}}. \end{aligned} \tag{3.3.18}$$

By [Bismut 2011, equation (6.1.1)], there exist $C_\gamma > 0$, $c_\gamma > 0$ such that, if $Y_0^{\mathfrak{k}} \in \mathfrak{k}(\gamma)$,

$$|J_\gamma(Y_0^{\mathfrak{k}})| \le C_\gamma e^{c_\gamma |Y_0^{\mathfrak{k}}|}. \tag{3.3.19}$$

Put $p = \dim \mathfrak{p}(\gamma)$, $q = \dim \mathfrak{k}(\gamma)$. Then $r = \dim \mathfrak{z}(\gamma) = p + q$. By [Bismut 2011, Theorem 6.1.1], for $t > 0$, we have

$$\mathrm{Tr}^{[\gamma]}[\exp(-t\mathcal{L}_A^X)] = \frac{e^{-\frac{|a|^2}{2t}}}{(2\pi t)^{\frac{p}{2}}} \int_{\mathfrak{k}(\gamma)} J_\gamma(Y_0^{\mathfrak{k}})\,\mathrm{Tr}^E[\rho^E(k)\exp(-i\rho^E(Y_0^{\mathfrak{k}}) - tA)]e^{-\frac{|Y_0^{\mathfrak{k}}|^2}{2t}} \frac{dY_0^{\mathfrak{k}}}{(2\pi t)^{\frac{q}{2}}}. \tag{3.3.20}$$

Remark 3.3.2. A generalization of Bismut's formula (3.3.20) to the twisted case is obtained in [Liu 2018; 2019]. An extension of this formula for considering arbitrary elements in the center of an enveloping algebra instead of the Casimir operator (3.2.8) was obtained in [Bismut and Shen 2022].

3.4. ***Compact locally symmetric spaces.*** Let Γ be a cocompact discrete subgroup of G. Then Γ acts on X isometrically and properly discontinuously. Then $Z = \Gamma\backslash X$ is compact second countable Hausdorff space.

If $x \in X$, put

$$\Gamma_x = \{\gamma \in \Gamma \mid \gamma x = x\}. \tag{3.4.1}$$

Then Γ_x is a finite subgroup of Γ. Put

$$r_x = \inf_{\gamma \in \Gamma - \Gamma_x} d(x, \gamma x). \tag{3.4.2}$$

Then we always have $r_x > 0$. Set

$$U_x = B\left(x, \frac{r_x}{4}\right) \subset X. \tag{3.4.3}$$

If $x \in X,\ \gamma \in \Gamma$, we have

$$r_{\gamma x} = r_x, \quad U_{\gamma x} = \gamma U_x. \tag{3.4.4}$$

It is clear that $\Gamma_x\backslash U_x$ can identified with a connected open subset of Z.

Set

$$S = \ker(\Gamma \to \mathrm{Diffeo}(X)) = \Gamma \cap \ker(K \xrightarrow{\mathrm{Ad}} \mathrm{Aut}(\mathfrak{p})). \tag{3.4.5}$$

Then S is a finite subgroup of $\Gamma \cap K$ and a normal subgroup of Γ.

Remark 3.4.1. Note that G_{ss} is a connected noncompact simple linear Lie group. Then

$$S = Z_G \cap \Gamma \cap K. \tag{3.4.6}$$

Put

$$\Gamma' = \Gamma/S. \tag{3.4.7}$$

Then Γ' acts on X effectively and we have $Z = \Gamma'\backslash X$.

If $x \in X$, we have

$$S \subset \Gamma_x, \quad \Gamma'_x = \Gamma_x/S. \tag{3.4.8}$$

Then the orbifold charts $(U_x, \Gamma'_x, \pi_x : U_x \to \Gamma'_x\backslash U_x)_{x\in X}$ together with the action of Γ' on these charts give an (effective) orbifold structure for Z, so that $Z = \Gamma\backslash X$ is a compact orbifold with a Riemannian metric g^{TZ} induced by g^{TX}.

By [Selberg 1960, Lemma 1], if $\gamma \in \Gamma$, then γ is semisimple. Let $[\Gamma]$ denote the set of the conjugacy classes of Γ. If $\gamma \in \Gamma$, we say $[\gamma] \in [\Gamma]$ is an elliptic class if γ is elliptic. Let $\mathrm{E}[\Gamma] \subset [\Gamma]$ be the set of elliptic classes. Then $\mathrm{E}[\Gamma]$ is always a finite set. If $\mathrm{E}[\Gamma]$ only contains the trivial conjugacy class $[1]$; i.e., Γ is torsion free, then Z is compact smooth manifold.

Let $[\Gamma']$ be the set of conjugacy classes in Γ', and let $\mathrm{E}[\Gamma']$ denote the set of elliptic classes in $[\Gamma']$. If $\gamma' \in \Gamma'$, let $Z_{\Gamma'}(\gamma')$ denote the centralizer of γ' in Γ', and let $[\gamma']'$ denote the conjugacy class of γ' in Γ'. If $\gamma' \in \Gamma'$ is elliptic, let $X(\gamma')$ be its fixed-point set in X on which $Z_{\Gamma'}(\gamma')$ acts isometrically and properly discontinuously; see [Selberg 1960, Lemma 2]. Note that if $\gamma \in \Gamma$ is a lift of $\gamma' \in \Gamma'$, then $X(\gamma) = X(\gamma')$, and γ is elliptic if and only if γ' is elliptic.

Proposition 3.4.2. *We have*

$$Z_{\mathrm{sing}} = \Gamma'\backslash\Gamma'\Big(\bigcup_{[\gamma']'\in E[\Gamma']\backslash\{1\}} X(\gamma')\Big) \subset Z. \tag{3.4.9}$$

Moreover, we have

$$\Sigma Z = \bigcup_{[\gamma']'\in E[\Gamma']\backslash\{1\}} Z_{\Gamma'}(\gamma')\backslash X(\gamma'). \tag{3.4.10}$$

Note that the right-hand side of (3.4.10) is a disjoint union of compact orbifolds.

If $\gamma' \in \Gamma'$, put

$$S'(\gamma') = \ker(Z_{\Gamma'}(\gamma') \to \mathrm{Diffeo}(X(\gamma'))). \tag{3.4.11}$$

Then $|S'(\gamma')|$ is the multiplicity of the connected component $Z_{\Gamma'}(\gamma')\backslash X(\gamma')$ in ΣZ.

Proof. Note that $z \in Z$ with a lift $x \in X$ belongs to Z_{sing} if and only if the stabilizer Γ'_x is nontrivial. Thus x is a fixed point of some $\gamma' \in \Gamma'$, from which (3.4.9) follows. By definition in Section 2.1, we get the rest of this proposition. □

Note that $\Gamma\backslash G$ is a compact smooth homogeneous space equipped with a right action of K. Moreover, the action of K is almost free; i.e., for each $\bar{g} \in \Gamma\backslash G$, the stabilizer $K_{\bar{g}}$ is finite. Then the quotient space $(\Gamma\backslash G)/K$ also has a natural orbifold structure, which, after examining the local charts, is equivalent to Z.

Let $d\bar{g}$ be the volume element on $\Gamma\backslash G$ induced by dg. By (3.3.1), we get

$$\mathrm{Vol}(\Gamma\backslash G) = \frac{\mathrm{Vol}(K)}{|S|}\,\mathrm{Vol}(Z). \tag{3.4.12}$$

In the context of geometry, we have many interesting cases where $S = \{1\}$. For instance, given a Riemannian symmetric space (X, g^{TX}) of noncompact type, let $G = \mathrm{Isom}(X)^0$ be the connected component of identity of the Lie group of isometries of X. By [Eberlein 1996, Proposition 2.1.1], G is a semisimple Lie group with trivial center (which might not be linear, but we do not need that linearity for the geometry of Z). We refer to [Eberlein 1996, Chapter 2; Bismut 2011, Chapter 3] for more details. This way, any subgroup of G acts on X effectively. In particular, if Γ is a cocompact discrete subgroup of G, then $Z = \Gamma\backslash X$ is a compact good orbifold with the orbifold fundamental group Γ. By (3.4.10), we have

$$\Sigma Z = \bigcup_{[\gamma]\in E[\Gamma]\backslash\{1\}} \Gamma\cap Z(\gamma)\backslash X(\gamma). \tag{3.4.13}$$

In general, by [Helgason 1978, Chapter V, §4, Theorem 4.1], $G = \mathrm{Isom}(X = G/K)^0$ if and only if K acts on $\mathfrak{p}$ effectively.

Remark 3.4.3. Note that, as mentioned in Remark 2.1.3, when $S \neq \{1\}$, we can also consider $Z = \Gamma\backslash X$ as an ineffective orbifold by taking the action of Γ instead of Γ' on the local charts. This way, the role of the above $Z \cup \Sigma Z$ is replaced by the inertia groupoid defined in [Adem et al. 2007, Example 2.5], which is exactly

$$\bigcup_{[\gamma]\in E[\Gamma]} \Gamma\cap Z(\gamma)\backslash X(\gamma). \tag{3.4.14}$$

It is a very natural object to use in the context here, for instance, for the Selberg's trace formula in the next subsection. In the problems we are concerned with, these two point-views on Z are equivalent.

If $\rho : \Gamma' \to \mathrm{GL}(\mathbb{C}^k)$ is a representation of Γ', which can be viewed as a representation of Γ via the projection $\Gamma \to \Gamma' = \Gamma/S$, then $F = \Gamma'\backslash(X \times \mathbb{C}^k)$ is a proper flat orbifold vector bundle on Z with the flat connection $\nabla^{F,f}$ induced from the exterior differential d^X on $\mathbb{C}^k$-valued functions. By [Shen and Yu 2022, Theorem 2.35], all the proper orbifold vector bundles on Z of rank k come from this.

Now let $\rho : \Gamma \to \mathrm{GL}(\mathbb{C}^k)$ be a representation of Γ; we do not assume that it comes from a representation of Γ'. We still have a flat orbifold vector bundle $(F = \Gamma\backslash(X \times \mathbb{C}^k), \nabla^{F,f})$ on Z, which may not be proper in general. Note that Γ acts on $C^\infty(X, \mathbb{C}^k)$ so that if $\varphi \in C^\infty(X, \mathbb{C}^k)$, $\gamma \in \Gamma$, then

$$(\gamma\varphi)(x) = \rho(\gamma)\varphi(\gamma^{-1}x). \tag{3.4.15}$$

Let $C^\infty(X, \mathbb{C}^k)^\Gamma$ denote the Γ-invariant sections in $C^\infty(X, \mathbb{C}^k)$. Then

$$C^\infty(Z, F) = C^\infty(X, \mathbb{C}^k)^\Gamma. \tag{3.4.16}$$

Definition 3.4.4. Let (V, ρ^V) be the isotypic component of $(\mathbb{C}^k, \rho|_S)$ corresponding to the trivial representation of S on $\mathbb{C}$, i.e., the maximal S-invariant subspace of $\mathbb{C}^k$ via ρ. Set

$$F^{\mathrm{pr}} = \Gamma\backslash(X \times V). \tag{3.4.17}$$

It is clear that F^{pr} is a proper flat orbifold vector bundle on Z.

Proposition 3.4.5. *We have*

$$C^\infty(Z, F) = C^\infty(Z, F^{\mathrm{pr}}). \tag{3.4.18}$$

In particular, if $\rho|_S : S \to \mathrm{GL}(\mathbb{C}^k)$ does not have the isotypic component of the trivial representation of S on $\mathbb{C}$, then

$$C^\infty(Z, F) = \{0\}. \tag{3.4.19}$$

Let (E, ρ^E) be a finite-dimensional complex representation of G. When restricting to Γ, K, we get the corresponding representations of Γ, K respectively, which are still denoted by ρ^E. As discussed in Section 3.2, associated with the K-representation (E, ρ^E) we define a homogeneous vector bundle $F = G \times_K E$ on X. Moreover, G acts on F equivariantly. By taking a Γ-quotient on the left, it descends to an orbifold vector bundle on Z, which we still denote by the same notation.

The map $(g, v) \in G \times_K E \to (pg, \rho^E(g)v) \in X \times E$ gives a canonical trivialization of F over X. This trivialization provides a flat connection $\nabla^{X,F,f}$ for $F \to X$, which is G-invariant. Then it descends to a flat connection $\nabla^{Z,F,f}$ on the orbifold vector bundle F over Z. Moreover, the above trivialization of $F \to X$ implies that the flat orbifold vector bundle $(F, \nabla^{Z,F,f})$ is exactly the one given by $\Gamma\backslash(X \times E)$ with the flat connection $\nabla^{F,f}$ induced by d^X. We will always use the notation $\nabla^{F,f}$ for the above flat connection. By (3.2.7), (3.4.16), we get

$$C^\infty(Z, F) = C_K^\infty(G, E)^\Gamma. \tag{3.4.20}$$

3.5. ***Selberg's trace formula.*** Let Z be the compact locally symmetric space discussed in Section 3.4, and let (F, h^F, ∇^F) be a Hermitian vector bundle on X defined by a unitary representation (E, ρ^E) of K. As said before, (F, h^F, ∇^F) descends to a Hermitian orbifold vector bundle on Z. Recall the Bochner-like Laplacian $\mathcal{L}_A^X$ is defined by (3.2.9). Since it commutes with G, it descends to a Bochner-like Laplacian $\mathcal{L}_A^Z$ acting on $C^\infty(Z, F)$.

Here the convergences of the integrals and infinite sums are already guaranteed by the results in [Bismut 2011, Chapters 2, 4; Shen 2018, Section 4D].

For $t > 0$, let $p_t^Z(z, z')$, $z, z' \in Z$, be the heat kernel of $\mathcal{L}_A^Z$ over Z with respect to dz'. If z, z' are identified with their lifts in X, then

$$p_t^Z(z,z') = \frac{1}{|S|}\sum_{\gamma\in\Gamma} \gamma p_t^X(\gamma^{-1}z, z') = \frac{1}{|S|}\sum_{\gamma\in\Gamma} p_t^X(z, \gamma z')\gamma. \tag{3.5.1}$$

Note that the action of γ on $F_{\gamma^{-1}z}$ or on the metric dual of $F_{z'}$ is given as in (3.2.4).

Since Z is compact, for $t > 0$, $\exp(-t\mathcal{L}_A^Z)$ is trace class. We have

$$\operatorname{Tr}[\exp(-t\mathcal{L}_A^Z)] = \int_Z \operatorname{Tr}^F[p_t^Z(z,z)]\,dz. \tag{3.5.2}$$

Combining (3.2.10), (3.2.11), (3.4.12) and (3.5.1), (3.5.2), and proceeding as in [Bismut 2011, equations (4.8.8)–(4.8.12)], we get

$$\begin{aligned}\operatorname{Tr}[\exp(-t\mathcal{L}_A^Z)] &= \frac{1}{\operatorname{Vol}(K)}\int_{\Gamma\backslash G}\sum_{\gamma\in\Gamma}\operatorname{Tr}^E[p_t^X(\bar g^{-1}\gamma\bar g)]\,d\bar g \\ &= \sum_{[\gamma]\in[\Gamma]}\frac{\operatorname{Vol}(\Gamma\cap Z(\gamma)\backslash Z(\gamma))}{\operatorname{Vol}(K(\gamma))}\operatorname{Tr}^{[\gamma]}[\exp(-t\mathcal{L}_A^X)].\end{aligned} \tag{3.5.3}$$

Take $\gamma \in \Gamma$. Recall that $X(\gamma) = Z(\gamma)/K(\gamma)$ defined in Section 3.3. Then $K(\gamma)$ acts on $Z(\gamma)$ on the right, which induces an action on $\Gamma \cap Z(\gamma)\backslash Z(\gamma)$ on the right. Set

$$S(\gamma) = \ker(\Gamma\cap Z(\gamma) \to \operatorname{Diffeo}(X(\gamma))). \tag{3.5.4}$$

Then $S(\gamma)$ represents the isotropy group of the principal orbit type for the right action of $K(\gamma)$ on $\Gamma \cap Z(\gamma)\backslash Z(\gamma)$. As in (3.4.12), we have

$$\operatorname{Vol}(\Gamma\cap Z(\gamma)\backslash Z(\gamma)) = \frac{\operatorname{Vol}(K(\gamma))}{|S(\gamma)|}\operatorname{Vol}(\Gamma\cap Z(\gamma)\backslash X(\gamma)). \tag{3.5.5}$$

Theorem 3.5.1. *For $t > 0$, we have the identity*

$$\operatorname{Tr}[\exp(-t\mathcal{L}_A^Z)] = \sum_{[\gamma]\in[\Gamma]}\frac{\operatorname{Vol}(\Gamma\cap Z(\gamma)\backslash X(\gamma))}{|S(\gamma)|}\operatorname{Tr}^{[\gamma]}[\exp(-t\mathcal{L}_A^X)]. \tag{3.5.6}$$

Proof. This is a direct consequence of (3.5.3) and (3.5.5). □

In the case where $S = 1$, the trace formula (3.5.6) shows clearly the different contributions from Z and from each components of ΣZ. Then combining (3.4.10), (3.5.6) with the results in [Bismut 2011, Theorem 7.8.2; Liu 2018, Theorem 7.7.1], we can recover (2.2.9) for Z. If we use the same settings as in [Bismut 2011, Sections 7.1, 7.2] and we use instead the results in Theorem 7.7.1 of that work, then we can recover the Kawasaki's local index theorem [1979] for Z. By taking account of Remarks 2.1.3 and 3.4.3, the above considerations also hold even for $S \neq \{1\}$.

4. Analytic torsions for compact locally symmetric spaces

In this section, we explain how to make use of Bismut's formula (3.3.20) and Selberg's trace formula (3.5.6) to study the analytic torsions of Z. We continue using the same settings as in Section 3. We will see that by a vanishing result on the analytic torsion, only the case $\delta(G) = 1$ remains interesting. For studying this case, more tools will be introduced in Sections 5 and 6.

4.1. ***A vanishing result on the analytic torsions.*** Recall that G is a connected linear real reductive Lie group. Recall that $\mathfrak{z}_\mathfrak{g}$ is the center of $\mathfrak{g}$. Set

$$\mathfrak{z}_\mathfrak{p} = \mathfrak{z}_\mathfrak{g} \cap \mathfrak{p}, \quad \mathfrak{z}_\mathfrak{k} = \mathfrak{z}_\mathfrak{g} \cap \mathfrak{k}. \tag{4.1.1}$$

Then

$$\mathfrak{z}_\mathfrak{g} = \mathfrak{z}_\mathfrak{p} \oplus \mathfrak{z}_\mathfrak{k}, \quad Z_G = \exp(\mathfrak{z}_\mathfrak{p})(Z_G \cap K). \tag{4.1.2}$$

Let T be a maximal torus of K with Lie algebra $\mathfrak{t}$; put

$$\mathfrak{b} = \{f \in \mathfrak{p} \mid [f, \mathfrak{t}] = 0\}. \tag{4.1.3}$$

It is clear that

$$\mathfrak{z}_\mathfrak{p} \subset \mathfrak{b}. \tag{4.1.4}$$

Put $\mathfrak{h} = \mathfrak{b} \oplus \mathfrak{t}$. Then $\mathfrak{h}$ is a Cartan subalgebra of $\mathfrak{g}$. Let H be analytic subgroup of G associated with $\mathfrak{h}$. Then it is also a Cartan subgroup of G; see [Knapp 1986, p. 129 and Theorem 5.22(b)]. Moreover, $\dim \mathfrak{t}$ is just the complex rank of K, and $\dim \mathfrak{h}$ is the complex rank of G.

Definition 4.1.1. Using the above notation, the deficiency of G, or the fundamental rank of G is defined as

$$\delta(G) = \mathrm{rk}_\mathbb{C}\, G - \mathrm{rk}_\mathbb{C}\, K = \dim_\mathbb{R} \mathfrak{b}. \tag{4.1.5}$$

The number $m - \delta(G)$ is even.

The following result was proved in [Shen 2018, Proposition 3.3].

Proposition 4.1.2. *If $\gamma \in G$ is semisimple, then*

$$\delta(G) \leq \delta(Z(\gamma)^0). \tag{4.1.6}$$

The two sides of (4.1.6) are equal if and only if γ can be conjugated into H.

Recall that $\mathfrak{u} = \sqrt{-1}\mathfrak{p} \oplus \mathfrak{k}$ is the compact form of G, and that $U \subset G_\mathbb{C}$ is the analytic subgroup with Lie algebra $\mathfrak{u}$. Let $U\mathfrak{u}$, $U\mathfrak{g}_\mathbb{C}$ be the enveloping algebras of $\mathfrak{u}$, $\mathfrak{g}_\mathbb{C}$ respectively. Then $U\mathfrak{g}_\mathbb{C}$ can be identified with the left-invariant holomorphic differential operators on $G_{\mathbb{C}}$. Let $C^{\mathfrak{u}} \in U\mathfrak{u}$ be the Casimir operator of $\mathfrak{u}$ associated with B. Then

$$C^{\mathfrak{u}} = C^{\mathfrak{g}} \in U\mathfrak{g} \cap U\mathfrak{u} \subset U\mathfrak{g}_{\mathbb{C}}. \tag{4.1.7}$$

In the sequel, we always assume that U is compact; this is the case when G has compact center.

Proposition 4.1.3 (unitary trick). *Assume that U is compact. Then any irreducible finite-dimensional (analytic) complex representation of U extends uniquely to an irreducible finite-dimensional complex representation of G such that their induced representations of Lie algebras are compatible.*

We now fix a unitary representation (E, ρ^E, h^E) of U, and we extend it to a representation of G, whose restriction to K is still unitary. Put $F = G \times_K E$, with the Hermitian metric h^F induced by h^E. Let ∇^F be the Hermitian connection induced by the connection form $\omega^{\mathfrak{k}}$.

Furthermore, as explained in the last part of Section 3.4, F is equipped with a canonical flat connection $\nabla^{F,f}$ as follows:

$$\nabla^{F,f} = \nabla^F + \rho^E(\omega^{\mathfrak{p}}). \tag{4.1.8}$$

If G has compact center, then $(F, h^F, \nabla^{F,f})$ is a unimodular flat vector bundle.

Let $(\Omega^\bullet_c(X, F), d^{X,F})$ be the (compactly supported) de Rham complex twisted by F. Let $d^{X,F,*}$ be the adjoint operator of $d^{X,F}$ with respect to the L_2 metric on $\Omega^\bullet_c(X, F)$. The de Rham–Hodge operator $\boldsymbol{D}^{X,F}$ of this de Rham complex is given by

$$\boldsymbol{D}^{X,F} = d^{X,F} + d^{X,F,*}. \tag{4.1.9}$$

The Clifford algebras $c(TX)$, $\hat{c}(TX)$ act on $\Lambda^\bullet(T^*X)$. We still use $e_1, \dots, e_m$ to denote an orthonormal basis of $\mathfrak{p}$ or TX, and let $e^1, \dots, e^m$ be the corresponding dual basis of $\mathfrak{p}^*$ or T^*X.

Let $\nabla^{\Lambda^\bullet(T^*X)\otimes F,u}$ be the unitary connection on $\Lambda^\bullet(T^*X) \otimes F$ induced by ∇^{TX} and ∇^F. Then the standard Dirac operator is given by

$$D^{X,F} = \sum_{j=1}^m c(e_j)\nabla^{\Lambda^\bullet(T^*X)\otimes F,u}_{e_j}. \tag{4.1.10}$$

By [Bismut et al. 2017, equation (8.42)], we have

$$\boldsymbol{D}^{X,F} = D^{X,F} + \sum_{j=1}^m \hat{c}(e_j)\rho^E(e_j). \tag{4.1.11}$$

At the same time, as explained in Section 3.2, $C^{\mathfrak{g}}$ descends to an elliptic differential operator $C^{\mathfrak{g},X}$ acting on $C^\infty(X, \Lambda^\bullet(T^*X) \otimes F)$. As in (3.2.9), we put

$$\mathcal{L}^{X,F} = \tfrac{1}{2}C^{\mathfrak{g},X} + \tfrac{1}{16}\operatorname{Tr}^{\mathfrak{p}}[C^{\mathfrak{k},\mathfrak{p}}] + \tfrac{1}{48}\operatorname{Tr}^{\mathfrak{k}}[C^{\mathfrak{k},\mathfrak{k}}]. \tag{4.1.12}$$

For simplicity, we will always put

$$\beta_{\mathfrak{g}} = \tfrac{1}{16}\operatorname{Tr}^{\mathfrak{p}}[C^{\mathfrak{k},\mathfrak{p}}] + \tfrac{1}{48}\operatorname{Tr}^{\mathfrak{k}}[C^{\mathfrak{k},\mathfrak{k}}] \in \mathbb{R}. \tag{4.1.13}$$

By [Bismut et al. 2017, Proposition 8.4], we have

$$\frac{1}{2}\boldsymbol{D}^{X,F,2} = \mathcal{L}^{X,F} - \frac{1}{2}C^{\mathfrak{g},E} - \beta_{\mathfrak{g}} =: \mathcal{L}_A^{X,F}, \tag{4.1.14}$$

where $A = -\frac{1}{2}C^{\mathfrak{g},E} - \beta_{\mathfrak{g}}$.

Let $\gamma \in G$ be a semisimple element. In the sequel, we may assume that

$$\gamma = e^a k, \quad a \in \mathfrak{p}, \quad k \in K, \quad \mathrm{Ad}(k)a = a. \tag{4.1.15}$$

We also use the same notation as in Section 3.3.

Recall that $p = \dim \mathfrak{p}(\gamma)$, $q = \dim \mathfrak{k}(\gamma)$. By (3.3.20) and (4.1.14), we have

$$\begin{aligned}\mathrm{Tr}_s^{[\gamma]}\Big[\Big(N^{\Lambda^\bullet(T^*X)} - \frac{m}{2}\Big)\exp\Big(-\frac{t\boldsymbol{D}^{X,F,2}}{2}\Big)\Big] \\ = \frac{e^{-\frac{|a|^2}{2t}}}{(2\pi t)^{\frac{p}{2}}}\exp(t\beta)\int_{\mathfrak{k}(\gamma)} J_\gamma(Y_0^{\mathfrak{k}})\,\mathrm{Tr}_s^{\Lambda^\bullet(\mathfrak{p}^*)}\Big[\Big(N^{\Lambda^\bullet(\mathfrak{p}^*)} - \frac{m}{2}\Big)\mathrm{Ad}(k)\exp(-i\,\mathrm{ad}(Y_0^{\mathfrak{k}}))\Big] \\ \cdot\mathrm{Tr}^E\Big[\rho^E(k)\exp\Big(-i\rho^E(Y_0^{\mathfrak{k}}) + \frac{t}{2}C^{\mathfrak{u},E}\Big)\Big]e^{-\frac{|Y_0^{\mathfrak{k}}|^2}{2t}}\frac{dY_0^{\mathfrak{k}}}{(2\pi t)^{\frac{q}{2}}}.\end{aligned} \tag{4.1.16}$$

Now we take a cocompact discrete subgroup $\Gamma \subset G$. Then $Z = \Gamma\backslash X$ is a compact locally symmetric orbifold. We use the same notation as in Sections 3.4 and 3.5. Then we get a flat orbifold vector bundle $(F, \nabla^{F,f}, h^F)$ on Z. Furthermore, $\boldsymbol{D}^{X,F}$ descends to the corresponding de Rham–Hodge operator $\boldsymbol{D}^{Z,F}$ acting on $\Omega^\bullet(Z, F)$. Let $\mathcal{T}(Z, F)$ denote the associated analytic torsion as in Definition 2.2.3, i.e.,

$$\mathcal{T}(Z, F) = \mathcal{T}(g^{TZ}, \nabla^{F,f}, h^F). \tag{4.1.17}$$

As explained in Section 2.2, for computing $\mathcal{T}(Z, F)$, it is enough to evaluate

$$\mathrm{Tr}_s\Big[\Big(N^{\Lambda^\bullet(T^*Z)} - \frac{m}{2}\Big)\exp\Big(-\frac{t\boldsymbol{D}^{Z,F,2}}{2}\Big)\Big], \quad t > 0. \tag{4.1.18}$$

Then we apply Selberg's trace formula in Theorem 3.5.1. We get

$$\begin{aligned}\mathrm{Tr}_s\Big[\Big(N^{\Lambda^\bullet(T^*Z)} - \frac{m}{2}\Big)\exp\Big(-\frac{t\boldsymbol{D}^{Z,F,2}}{2}\Big)\Big] \\ = \sum_{[\gamma]\in[\Gamma]}\frac{\mathrm{Vol}(\Gamma\cap Z(\gamma)\backslash X(\gamma))}{|S(\gamma)|}\mathrm{Tr}^{[\gamma]}\Big[\Big(N^{\Lambda^\bullet(T^*X)} - \frac{m}{2}\Big)\exp\Big(-\frac{t\boldsymbol{D}^{X,F,2}}{2}\Big)\Big].\end{aligned} \tag{4.1.19}$$

As in [Bismut et al. 2017, Remark 8.7], by [Ma 2019, Theorems 5.4, 5.5, Remark 5.6], we have the following vanishing theorem on $\mathcal{T}(Z, F)$.

Theorem 4.1.4. *If m is even, or if m is odd and $\delta(G) \geq 3$, then*

$$\mathcal{T}(Z, F) = 0. \tag{4.1.20}$$

Proof. By [Bismut 2011, Theorem 7.9.1; Ma 2019, Theorem 5.4], and using instead (4.1.19), we get that under the assumptions in this theorem, for $t > 0$,

$$\mathrm{Tr}_s\Big[\Big(N^{\Lambda^\bullet(T^*Z)} - \frac{m}{2}\Big)\exp(-t\boldsymbol{D}^{Z,F,2})\Big] = 0. \tag{4.1.21}$$

Then (4.1.20) follows from the definition of $\mathcal{T}(Z, F)$. ∎

Therefore, the only nontrivial case is that $\delta(G)=1$, so that m is odd. If $\gamma\in G$ is of the form (4.1.15), let $\mathfrak{t}(\gamma)\subset\mathfrak{k}(\gamma)$ be a Cartan subalgebra. Put

$$\mathfrak{b}(\gamma)=\{v\in\mathfrak{p}(k)\mid[v,\mathfrak{t}(\gamma)]=0\},\quad \mathfrak{h}(\gamma)_{\mathfrak{p}}=\mathfrak{b}(\gamma)\cap\mathfrak{p}(\gamma). \tag{4.1.22}$$

In particular, $a\in\mathfrak{b}(\gamma)$. Then $\mathfrak{h}(\gamma)=\mathfrak{h}(\gamma)_{\mathfrak{p}}\oplus\mathfrak{t}(\gamma)$ is a Cartan subalgebra of $\mathfrak{z}(\gamma)$.

Recall that H is a maximally compact Cartan subgroup of G. The following result is just an analogue of [Shen 2018, Theorem 4.12; Bismut 2011, Theorem 7.9.1].

Proposition 4.1.5. *If $\delta(G)=1$, if γ is semisimple and cannot be conjugated into H by an element in G, then*

$$\operatorname{Tr}_s^{[\gamma]}\Big[\Big(N^{\Lambda^\bullet(T^*X)}-\frac{m}{2}\Big)\exp\Big(-\frac{t\boldsymbol{D}^{X,F,2}}{2}\Big)\Big]=0. \tag{4.1.23}$$

Proof. Let $\mathfrak{t}$ be a Cartan subalgebra of $\mathfrak{k}$ containing $\mathfrak{t}(\gamma)$. Then $\mathfrak{b}\subset\mathfrak{b}(\gamma)$. If $a\notin\mathfrak{b}$, then $\dim\mathfrak{b}(\gamma)\geq 2$. Therefore, by [Shen 2018, equation (4-44)], for $Y_0^{\mathfrak{k}}\in\mathfrak{k}(\gamma)$, we have

$$\operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{p}^*)}\Big[\Big(N^{\Lambda^\bullet(\mathfrak{p}^*)}-\frac{m}{2}\Big)\operatorname{Ad}(k)\exp(-i\operatorname{ad}(Y_0^{\mathfrak{k}}))\Big]=0. \tag{4.1.24}$$

This implies (4.1.23). ☐

Set

$$\mathfrak{g}'=\mathfrak{z}_{\mathfrak{k}}\oplus\mathfrak{g}_{\mathrm{ss}}. \tag{4.1.25}$$

Then $\mathfrak{g}'$ is an ideal of $\mathfrak{g}$. Let G' be the analytic subgroup of G associated with $\mathfrak{g}'$, which is closed and has a compact center; see [Knapp 2002, Proposition 7.27]. The group K is still a maximal subgroup of G'. Let $U'\subset U$ be the compact form of G' with Lie algebra $\mathfrak{u}'$. Then

$$\mathfrak{u}=\sqrt{-1}\mathfrak{z}_{\mathfrak{p}}\oplus\mathfrak{u}'. \tag{4.1.26}$$

Now we assume that $\delta(G)=1$ and that G has noncompact center, so that $\mathfrak{b}=\mathfrak{z}_{\mathfrak{p}}$ has dimension 1. Then $\delta(G')=0$. Under the hypothesis that U is compact, up to a finite cover, we may write

$$U\simeq\mathbb{S}^1\times U'. \tag{4.1.27}$$

We take $a_1\in\mathfrak{b}$ with $|a_1|=1$. If (E,ρ^E) is an irreducible unitary representation of U, then $\rho^E(a_1)$ acts on E by a real scalar operator. Let $\alpha_E\in\mathbb{R}$ be such that

$$\rho^E(a_1)=\alpha_E\operatorname{Id}_E. \tag{4.1.28}$$

Put $X'=G'/K$. Then X' is an even-dimensional symmetric space (of noncompact type). We identify $\mathfrak{z}_{\mathfrak{p}}$ with a real line $\mathbb{R}$. Then

$$G=\mathbb{R}\times G',\quad X=\mathbb{R}\times X'. \tag{4.1.29}$$

In this case, the evaluation for analytic torsions can be made more explicit. If $\gamma\in G'$, let $X'(\gamma)$ denote the minimizing set of $d_\gamma(\cdot)$ in X', so that

$$X(\gamma)=\mathbb{R}\times X'(\gamma). \tag{4.1.30}$$

Let $[\cdot]^{\max}$ denote the coefficient of a differential form (valued in $o(TX')$) on X' of the corresponding Riemannian volume form. Similarly, for $k \in T$, let $[\cdot]^{\max(k)}$ denote the analogous object on $X'(k)$. The following results are the analogues of [Shen 2018, Proposition 4.14].

Proposition 4.1.6. *Assume that G has noncompact center with $\delta(G) = 1$ and that (E, ρ^E) is irreducible. Then*

$$\mathrm{Tr}_s^{[1]}\Big[\Big(N^{\Lambda^\bullet(T^*X)} - \frac{m}{2}\Big)\exp\Big(-\frac{t\boldsymbol{D}^{X,F,2}}{2}\Big)\Big] = -\frac{e^{-\frac{1}{2}t\alpha_E^2}}{\sqrt{2\pi t}}[e(TX', \nabla^{TX'})]^{\max}\dim E. \tag{4.1.31}$$

If $\gamma = e^a k$ is such that $a \in \mathfrak{b}$, $k \in T$, then

$$\mathrm{Tr}_s^{[\gamma]}\Big[\Big(N^{\Lambda^\bullet(T^*X)} - \frac{m}{2}\Big)\exp\Big(-\frac{t\boldsymbol{D}^{X,F,2}}{2}\Big)\Big] = -\frac{1}{\sqrt{2\pi t}}e^{-\frac{|a|^2}{2t}-\frac{1}{2}t\alpha_E^2}[e(TX'(k), \nabla^{TX'(k)})]^{\max(k)}\,\mathrm{Tr}^E[\rho^E(k)]. \tag{4.1.32}$$

Proof. Let $C^{\mathfrak{u}'}$ denote the Casimir operator of $\mathfrak{u}'$ associated with $B|_{\mathfrak{u}'}$. Then we have

$$C^{\mathfrak{u}} = -a_1^2 + C^{\mathfrak{u}'}. \tag{4.1.33}$$

Since (E, ρ^E) is an irreducible representation, by (4.1.28) and (4.1.33), we get

$$C^{\mathfrak{u},E} = -\alpha_E^2 + C^{\mathfrak{u}',E}. \tag{4.1.34}$$

Then by (4.1.34) and [Bismut et al. 2017, Theorem 8.5], a modification of the proof of [Shen 2018, Proposition 4.14] proves the identities in our proposition. ☐

If we assembly the results in Proposition 4.1.6, it is enough to study the corresponding analytic torsions. We will get back to this point in Corollary 7.4.4 for asymptotic analytic torsions.

4.2. ***Symmetric spaces of noncompact type with fundamental rank 1.*** In this subsection, we focus on the case where $\delta(G) = 1$ and G has compact center (i.e., $\mathfrak{z}_\mathfrak{p} = 0$), so that X is a symmetric space of noncompact type [Shen 2018, Proposition 6.18]. For simplicity, let us also assume that G is linear semisimple in this subsection.

Note that the rank $\delta(X)$ of X (see [Eberlein 1996, Section 2.7]) is the same as $\delta(G)$. Then $\delta(X) = 1$. By the de Rham decomposition, we can write

$$X = X_1 \times X_2, \tag{4.2.1}$$

where X_1 is an irreducible symmetric space of noncompact type with $\delta(X_1) = 1$, and X_2 is a symmetric space of noncompact type with $\delta(X_2) = 0$.

As in [Bismut 2011, Remark 7.9.2], among the noncompact simple connected real linear groups such that m is odd and $\dim \mathfrak{b} = 1$, there are only $\mathrm{SL}_3(\mathbb{R})$, $\mathrm{SL}_4(\mathbb{R})$, $\mathrm{SL}_2(\mathbb{H})$, and $\mathrm{SO}^0(p,q)$ with pq odd > 1. Also, we have $\mathfrak{sl}_4(\mathbb{R}) = \mathfrak{so}(3,3)$ and $\mathfrak{sl}_2(\mathbb{H}) = \mathfrak{so}(5,1)$. Therefore, X_1 is one of the following cases (see [Shen 2018, Proposition 6.19]):

$$X_1 = \mathrm{SL}_3(\mathbb{R})/\mathrm{SO}(3) \quad \text{or} \quad \mathrm{SO}^0(p,q)/\mathrm{SO}(p+q), \quad \text{with } pq > 1 \text{ odd}. \tag{4.2.2}$$

Since $\delta(G) = 1$, we have the decomposition of Lie algebras

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2, \tag{4.2.3}$$

where

$$\mathfrak{g}_1 = \mathfrak{sl}_3(\mathbb{R}) \quad \text{or} \quad \mathfrak{so}(p, q), \tag{4.2.4}$$

with $pq > 1$ odd, and $\mathfrak{g}_2$ is semisimple with $\delta(\mathfrak{g}_2) = 0$. The Cartan involution θ preserves the splitting (4.2.3); see [Knapp 2002, VII.6, p. 471].

Let G_1 be the identity component of $Z_G(\mathfrak{g}_2)$. Then G_1 is a connected linear semisimple closed subgroup of G with Lie algebra of $\mathfrak{g}_1$. Similarly, we can find a connected linear semisimple closed subgroup G_2 of G with Lie algebra of $\mathfrak{g}_2$ such that we have canonically $G_1 \times G_2 \to G$ a finite central extension. Let θ_j be the induced Cartan involution on $G_j (j = 1, 2)$ from θ. Set $K_j = G_j \cap K$; then

$$X_j = G_j / K_j, \quad j = 1, 2. \tag{4.2.5}$$

Note that in general, G_1 is a just a finite central extension of $\mathrm{SL}_3(\mathbb{R})$ or $\mathrm{SO}^0(p, q)$ ($pq > 1$ odd). The invariant bilinear form B also splits as $B_1 \oplus B_2$ with respect to the splitting (4.2.3).

Remark 4.2.1. Let G_*, $G_{1,*}$, $G_{2,*}$ denote the identity components of the isometry groups of X, X_1, X_2 respectively. Then we have

$$G_* = G_{1,*} \times G_{2,*}. \tag{4.2.6}$$

By [Shen 2018, Proposition 6.19], $G_{1,*} = \mathrm{SL}_3(\mathbb{R})$ or $\mathrm{SO}^0(p, q)$, with $pq > 1$ odd, and $G_{2,*}$ is a semisimple Lie group with Lie algebra $\mathfrak{g}_2$ and trivial center. Also $\delta(G_{2,*}) = 0$. If we consider G_* instead of G, then the factor G_1 is exactly $\mathrm{SL}_3(\mathbb{R})$ or $\mathrm{SO}^0(p, q)$, with $pq > 1$ odd.

Let U_1, U_2 be (connected linear) compact forms of G_1, G_2. Then $U_1 \times U_2$ is a finite central extension of the compact form U of G. Let (E, ρ^E) be an irreducible unitary representation of U, and hence of $U_1 \times U_2$. Then

$$(E, \rho^E) = (E_1, \rho^{E_1}) \otimes (E_2, \rho^{E_2}), \tag{4.2.7}$$

where (E_j, ρ^{E_j}) is an irreducible unitary representation of U_j, $j = 1, 2$. Let F, F_1, F_2 be the homogeneous flat vector bundles on X, X_1, X_2 associated with these representations. Then we have

$$F = F_1 \boxtimes F_2 := \pi_1^*(F_1) \otimes \pi_2^*(F_2), \tag{4.2.8}$$

where π_i denote the projections $X \to X_i$, $i = 1, 2$.

Take $\gamma \in G$. Let $(\gamma_1, \gamma_2) \in G_1 \times G_2$ be one of its lifts. Then γ is semisimple (resp. elliptic) if and only if both γ_1, γ_2 are semisimple (resp. elliptic). Set $m_i = \dim X_i$; then m_1 is odd, and m_2 is even.

Proposition 4.2.2. *If $\gamma \in G$ is semisimple, for $t > 0$, we have*

$$\begin{aligned} &\mathrm{Tr}_s^{[\gamma]}\Big[\Big(N^{\Lambda^\bullet(T^*X)} - \frac{m}{2}\Big) \exp\Big(-\frac{t\boldsymbol{D}^{X,F,2}}{2}\Big)\Big] \\ &\qquad = \mathrm{Tr}_s^{[\gamma_1]}\Big[\Big(N^{\Lambda^\bullet(T^*X_1)} - \frac{m_1}{2}\Big) \exp\Big(-\frac{t\boldsymbol{D}^{X_1,F_1,2}}{2}\Big)\Big] \cdot \mathrm{Tr}_s^{[\gamma_2]}\Big[\exp\Big(-\frac{t\boldsymbol{D}^{X_2,F_2,2}}{2}\Big)\Big]. \end{aligned} \tag{4.2.9}$$

Then if γ_2 is nonelliptic,

$$\mathrm{Tr}_s^{[\gamma]}\Big[\Big(N^{\Lambda^\bullet(T^*X)}-\frac{m}{2}\Big)\exp\Big(-\frac{t\boldsymbol{D}^{X,F,2}}{2}\Big)\Big]=0. \tag{4.2.10}$$

If γ_2 is elliptic, then

$$\mathrm{Tr}_s^{[\gamma]}\Big[\Big(N^{\Lambda^\bullet(T^*X)}-\frac{m}{2}\Big)\exp\Big(-\frac{t\boldsymbol{D}^{X,F,2}}{2}\Big)\Big]=[e(TX_2(\gamma_2),\nabla^{TX_2(\gamma_2)})]^{\max_2(\gamma_2)}\,\mathrm{Tr}^{E_2}[\rho^{E_2}(\gamma_2)] \cdot\mathrm{Tr}_s^{[\gamma_1]}\Big[\Big(N^{\Lambda^\bullet(T^*X_1)}-\frac{m_1}{2}\Big)\exp\Big(-\frac{t\boldsymbol{D}^{X_1,F_1,2}}{2}\Big)\Big], \tag{4.2.11}$$

where $[\cdot]^{\max_2(\gamma_2)}$ is taking the coefficient of the Riemannian volume element on $X_2(\gamma_2)$.

Proof. We write

$$N^{\Lambda^\bullet(T^*X)}-\frac{m}{2}=\Big(N^{\Lambda^\bullet(T^*X_1)}-\frac{m_1}{2}\Big)+\Big(N^{\Lambda^\bullet(T^*X_2)}-\frac{m_2}{2}\Big). \tag{4.2.12}$$

Note that, since $\delta(G_1)=1$, by [Bismut 2011, Theorem 7.8.2], we always have

$$\mathrm{Tr}_s^{[\gamma_1]}\Big[\exp\Big(-\frac{t\boldsymbol{D}^{X_1,F_1,2}}{2}\Big)\Big]=0. \tag{4.2.13}$$

Combining the definition of orbital integrals (3.3.7) together with (4.2.12) and (4.2.13), we get (4.2.9).

The identities (4.2.10), (4.2.11) follow from applying the results in [Bismut 2011, Theorem 7.8.2] to $\mathrm{Tr}_s^{[\gamma_2]}[\exp(-t\boldsymbol{D}^{X_2,F_2,2}/2)]$. □

For studying $\mathcal{T}(Z,F)$, Proposition 4.2.2 helps us to reduce the computations on

$$\mathrm{Tr}_s\Big[\Big(N^{\Lambda^\bullet(T^*Z)}-\frac{m}{2}\Big)\exp\Big(-\frac{t\boldsymbol{D}^{Z,F,2}}{2}\Big)\Big]$$

to the model cases listed in (4.2.2). But it is far from enough to get an explicit evaluation. In Sections 5 and 6, we will introduce more tools, which allows us work out a proof to Theorem 1.0.2.

5. Cartan subalgebra and root system of G when $\delta(G)=1$

We use the same notation as in Section 3 and Section 4.1. In Sections 5.1–5.3, we always assume that G is a connected linear real reductive Lie group with compact center and with $\delta(G)=1$. But, as we will see in Remark 5.3.3, the constructions and results in these subsections are still true (most of them are trivial) if U is compact and if G has noncompact center with $\delta(G)=1$.

Section 5.4 is independent from other subsections, where we introduce a generalized Kirillov formula for compact Lie groups.

Recall that T is a maximal torus of K with Lie algebra $\mathfrak{t}\subset\mathfrak{k}$, and that $\mathfrak{b}\subset\mathfrak{p}$ is defined in (4.1.3). Since $\delta(G)=1$, we know $\mathfrak{b}$ is 1-dimensional. We now fix a vector $a_1\in\mathfrak{b}$, $|a_1|=1$. Recall that $\mathfrak{h}=\mathfrak{b}\oplus\mathfrak{t}$ is a Cartan subalgebra of $\mathfrak{g}$. Let $h^{\mathfrak{g}_\mathbb{C}}$ be the Hermitian product on $\mathfrak{g}_\mathbb{C}$ induced by the scalar product $-B(\cdot,\theta\cdot)$ on $\mathfrak{g}$.

5.1. ***Reductive Lie algebra with fundamental rank 1.*** Since G has compact center, $\mathfrak{b} \not\subset \mathfrak{z}_{\mathfrak{g}}$. Let $Z(\mathfrak{b})$ be the centralizer of $\mathfrak{b}$ in G, and let $Z(\mathfrak{b})^0$ be its identity component with Lie algebra $\mathfrak{z}(\mathfrak{b}) = \mathfrak{p}(\mathfrak{b}) \oplus \mathfrak{k}(\mathfrak{b}) \subset \mathfrak{g}$. Let $\mathfrak{m}$ be the orthogonal subspace of $\mathfrak{b}$ in $\mathfrak{z}(\mathfrak{b})$ (with respect to B) such that

$$\mathfrak{z}(\mathfrak{b}) = \mathfrak{b} \oplus \mathfrak{m}. \tag{5.1.1}$$

Then $\mathfrak{m}$ is a Lie subalgebra of $\mathfrak{z}(\mathfrak{b})$, which is invariant by θ.

Put

$$\mathfrak{p}_{\mathfrak{m}} = \mathfrak{m} \cap \mathfrak{p}, \quad \mathfrak{k}_{\mathfrak{m}} = \mathfrak{m} \cap \mathfrak{k}. \tag{5.1.2}$$

Then

$$\mathfrak{m} = \mathfrak{p}_{\mathfrak{m}} \oplus \mathfrak{k}_{\mathfrak{m}}, \quad \mathfrak{p}(\mathfrak{b}) = \mathfrak{b} \oplus \mathfrak{p}_{\mathfrak{m}}, \quad \mathfrak{k}(\mathfrak{b}) = \mathfrak{k}_{\mathfrak{m}}. \tag{5.1.3}$$

Let $\mathfrak{z}^{\perp}(\mathfrak{b})$, $\mathfrak{p}^{\perp}(\mathfrak{b})$, $\mathfrak{k}^{\perp}(\mathfrak{b})$ be the orthogonal subspaces of $\mathfrak{z}(\mathfrak{b})$, $\mathfrak{p}(\mathfrak{b})$, $\mathfrak{k}(\mathfrak{b})$ in $\mathfrak{g}$, $\mathfrak{p}$, $\mathfrak{k}$ respectively with respect to B. Then

$$\mathfrak{z}^{\perp}(\mathfrak{b}) = \mathfrak{p}^{\perp}(\mathfrak{b}) \oplus \mathfrak{k}^{\perp}(\mathfrak{b}). \tag{5.1.4}$$

Moreover,

$$\mathfrak{p} = \mathfrak{b} \oplus \mathfrak{p}_{\mathfrak{m}} \oplus \mathfrak{p}^{\perp}(\mathfrak{b}), \quad \mathfrak{k} = \mathfrak{k}(\mathfrak{b}) \oplus \mathfrak{k}^{\perp}(\mathfrak{b}). \tag{5.1.5}$$

Let $M \subset Z(\mathfrak{b})^0$ be the analytic subgroup associated with $\mathfrak{m}$. If we identify $\mathfrak{b}$ with $\mathbb{R}$, then

$$Z(\mathfrak{b})^0 = \mathbb{R} \times M. \tag{5.1.6}$$

Then M is a Lie subgroup of $Z(\mathfrak{b})^0$; i.e., it is closed in $Z(\mathfrak{b})^0$. Let K_M be the analytic subgroup of M associated with the Lie subalgebra $\mathfrak{k}_{\mathfrak{m}}$. Since M is reductive, K_M is a maximal compact subgroup of M. Then the splittings in (5.1.3), (5.1.4), (5.1.5) are invariant by the adjoint action of K_M.

Then $\mathfrak{t}$ is Cartan subalgebra of $\mathfrak{k}$, of $\mathfrak{k}_{\mathfrak{m}}$, and of $\mathfrak{m}$. Recall that $\mathfrak{h} = \mathfrak{b} \oplus \mathfrak{t}$ is a Cartan subalgebra of $\mathfrak{g}$. We fix $a_1 \in \mathfrak{b}$ such that $B(a_1, a_1) = 1$. The choice of a_1 fixes an orientation of $\mathfrak{b}$. Let $\mathfrak{n} \subset \mathfrak{z}^{\perp}(\mathfrak{b})$ be the direct sum of the eigenspaces of $\mathrm{ad}(a_1)$ with the positive eigenvalues. Set $\bar{\mathfrak{n}} = \theta\mathfrak{n}$. Then

$$\mathfrak{z}^{\perp}(\mathfrak{b}) = \mathfrak{n} \oplus \bar{\mathfrak{n}}. \tag{5.1.7}$$

By [Shen 2018, Section 6A], $\dim \mathfrak{n} = \dim \mathfrak{p} - \dim \mathfrak{p}_{\mathfrak{m}} - 1$. Then $\dim \mathfrak{n}$ is even under our assumption $\delta(G) = 1$. Put

$$l = \tfrac{1}{2} \dim \mathfrak{n}. \tag{5.1.8}$$

By [Shen 2018, Proposition 6.2], there exists $\beta \in \mathfrak{b}^*$ such that if $a \in \mathfrak{b}$, $f \in \mathfrak{n}$, then

$$[a, f] = \beta(a) f, \quad [a, \theta(f)] = -\beta(a)\theta(f). \tag{5.1.9}$$

The map $f \in \mathfrak{n} \mapsto f - \theta(f) \in \mathfrak{p}^{\perp}(\mathfrak{b})$ is an isomorphism of K_M-modules. Similarly, $f \in \mathfrak{n} \mapsto f + \theta(f) \in \mathfrak{k}^{\perp}(\mathfrak{b})$ is also an isomorphism of K_M-modules. Since θ fixes K_M, $\mathfrak{n} \simeq \bar{\mathfrak{n}}$ as K_M-modules via θ.

By [Shen 2018, Proposition 6.3], we have

$$[\mathfrak{n}, \bar{\mathfrak{n}}] \subset \mathfrak{z}(\mathfrak{b}), \quad [\mathfrak{n}, \mathfrak{n}] = [\bar{\mathfrak{n}}, \bar{\mathfrak{n}}] = 0. \tag{5.1.10}$$

Also

$$B|_{\mathfrak{n}\times\mathfrak{n}} = 0, \quad B|_{\bar{\mathfrak{n}}\times\bar{\mathfrak{n}}} = 0. \tag{5.1.11}$$

Then the bilinear form B induces an isomorphism of $\mathfrak{n}^*$ and $\bar{\mathfrak{n}}$ as K_M-modules. Therefore, as K_M-modules, $\mathfrak{n}$ is isomorphic to $\mathfrak{n}^*$.

As a consequence of (5.1.10), we get

$$[\mathfrak{z}(\mathfrak{b}), \mathfrak{z}(\mathfrak{b})], [\mathfrak{z}^{\perp}(\mathfrak{b}), \mathfrak{z}^{\perp}(\mathfrak{b})] \subset \mathfrak{z}(\mathfrak{b}), \quad [\mathfrak{z}(\mathfrak{b}), \mathfrak{z}^{\perp}(\mathfrak{b})] \subset \mathfrak{z}^{\perp}(\mathfrak{b}). \tag{5.1.12}$$

Then $(\mathfrak{g}, \mathfrak{z}(\mathfrak{b}))$ is a symmetric pair.

If $k \in K_M$, let $M(k)$ be the centralizer of k in M, and let $\mathfrak{m}(k)$ be its Lie algebra. Let $M(k)^0$ be the identity component of $M(k)$. The Cartan involution θ acts on $M(k)$. The associated Cartan decomposition is

$$\mathfrak{m}(k) = \mathfrak{p}_{\mathfrak{m}}(k) \oplus \mathfrak{k}_{\mathfrak{m}}(k), \tag{5.1.13}$$

where $\mathfrak{p}_{\mathfrak{m}}(k) = \mathfrak{p}_{\mathfrak{m}} \cap \mathfrak{m}(k)$, $\mathfrak{k}_{\mathfrak{m}}(k) = \mathfrak{k}_{\mathfrak{m}} \cap \mathfrak{m}(k)$.

Recall that $Z(k)$ is the centralizer of k in G and that $Z(k)^0$ is the identity component of $Z(k)$ with Lie algebra $\mathfrak{z}(k) \subset \mathfrak{g}$. Then

$$M(k) = M \cap Z(k), \quad \mathfrak{m}(k) = \mathfrak{m} \cap \mathfrak{z}(k). \tag{5.1.14}$$

Note that $Z(k)^0$ is still a reductive Lie group equipped with the Cartan involution induced by the action of θ. By the assumption that $\delta(G) = 1$, we have

$$\delta(Z(k)^0) = 1. \tag{5.1.15}$$

In particular,

$$\mathfrak{b} \subset \mathfrak{p}(k). \tag{5.1.16}$$

Set

$$\mathfrak{z}_{\mathfrak{b}}(k) = \mathfrak{z}(\mathfrak{b}) \cap \mathfrak{z}(k), \quad \mathfrak{p}_{\mathfrak{b}}(k) = \mathfrak{p}(\mathfrak{b}) \cap \mathfrak{p}(k), \quad \mathfrak{k}_{\mathfrak{b}}(k) = \mathfrak{k}(\mathfrak{b}) \cap \mathfrak{k}(k). \tag{5.1.17}$$

Then

$$\mathfrak{z}_{\mathfrak{b}}(k) = \mathfrak{b} \oplus \mathfrak{m}(k) = \mathfrak{p}_{\mathfrak{b}}(k) \oplus \mathfrak{k}_{\mathfrak{b}}(k). \tag{5.1.18}$$

We also have the identities

$$\mathfrak{p}_{\mathfrak{b}}(k) = \mathfrak{b} \oplus \mathfrak{p}_{\mathfrak{m}}(k), \quad \mathfrak{k}_{\mathfrak{b}}(k) = \mathfrak{k}_{\mathfrak{m}}(k). \tag{5.1.19}$$

Let $\mathfrak{p}_{\mathfrak{b}}^{\perp}(k)$, $\mathfrak{k}_{\mathfrak{b}}^{\perp}(k)$, $\mathfrak{z}_{\mathfrak{b}}^{\perp}(k)$ be the orthogonal spaces of $\mathfrak{p}_{\mathfrak{b}}(k)$, $\mathfrak{k}_{\mathfrak{b}}(k)$, $\mathfrak{z}_{\mathfrak{b}}(k)$ in $\mathfrak{p}(k)$, $\mathfrak{k}(k)$, $\mathfrak{z}(k)$ with respect to B, so that

$$\mathfrak{p}(k) = \mathfrak{p}_{\mathfrak{b}}(k) \oplus \mathfrak{p}_{\mathfrak{b}}^{\perp}(k), \quad \mathfrak{k}(k) = \mathfrak{k}_{\mathfrak{b}}(k) \oplus \mathfrak{k}_{\mathfrak{b}}^{\perp}(k), \quad \mathfrak{z}(k) = \mathfrak{z}_{\mathfrak{b}}(k) \oplus \mathfrak{z}_{\mathfrak{b}}^{\perp}(k). \tag{5.1.20}$$

Then

$$\mathfrak{z}_{\mathfrak{b}}^{\perp}(k) = \mathfrak{p}_{\mathfrak{b}}^{\perp}(k) \oplus \mathfrak{k}_{\mathfrak{b}}^{\perp}(k) = \mathfrak{z}^{\perp}(\mathfrak{b}) \cap \mathfrak{z}(k). \tag{5.1.21}$$

Put

$$\mathfrak{n}(k) = \mathfrak{z}(k) \cap \mathfrak{n}, \quad \bar{\mathfrak{n}}(k) = \mathfrak{z}(k) \cap \bar{\mathfrak{n}}. \tag{5.1.22}$$

Then

$$\mathfrak{z}_{\mathfrak{b}}^{\perp}(k) = \mathfrak{n}(k) \oplus \bar{\mathfrak{n}}(k). \tag{5.1.23}$$

By (5.1.17), (5.1.23), we get

$$\mathfrak{z}(k) = \mathfrak{p}_{\mathfrak{b}}(k) \oplus \mathfrak{k}_{\mathfrak{b}}(k) \oplus \mathfrak{n}(k) \oplus \bar{\mathfrak{n}}(k). \tag{5.1.24}$$

Since $\delta(\mathfrak{m}(k)) = 0$, $\dim \mathfrak{n}(k)$ is even. We set

$$l(k) = \tfrac{1}{2} \dim \mathfrak{n}(k). \tag{5.1.25}$$

Let $K_M(k)$ denote the centralizer of k in K_M. The map $f \in \mathfrak{n}(k) \mapsto f - \theta(f) \in \mathfrak{p}_{\mathfrak{b}}^{\perp}(k)$ is an isomorphism of $K_M(k)$-modules, and similarly for $\mathfrak{k}_{\mathfrak{b}}^{\perp}(k)$. Since θ fixes $K_M(k)$, we have $\mathfrak{n}(k) \simeq \bar{\mathfrak{n}}(k)$ as $K_M(k)$-modules via θ.

5.2. ***A compact Hermitian symmetric space $Y_{\mathfrak{b}}$.*** Recall that $\mathfrak{u} = \sqrt{-1}\mathfrak{p} \oplus \mathfrak{k}$ is the compact form of $\mathfrak{g}$.

Let $\mathfrak{u}(\mathfrak{b}) \subset \mathfrak{u}$, $\mathfrak{u}_{\mathfrak{m}} \subset \mathfrak{u}$ be the compact forms of $\mathfrak{z}(\mathfrak{b})$, $\mathfrak{m}$. Then

$$\mathfrak{u}(\mathfrak{b}) = \sqrt{-1}\mathfrak{b} \oplus \mathfrak{u}_{\mathfrak{m}}, \quad \mathfrak{u}_{\mathfrak{m}} = \sqrt{-1}\mathfrak{p}_{\mathfrak{m}} \oplus \mathfrak{k}_{\mathfrak{m}}. \tag{5.2.1}$$

Since M has compact center, let U_M be the analytic subgroup of U associated with $\mathfrak{u}_{\mathfrak{m}}$. Then U_M is the compact form of M. Let $U(\mathfrak{b}) \subset U$, $A_0 \subset U$ be the connected subgroups of U associated with Lie algebras $\mathfrak{u}(\mathfrak{b})$, $\sqrt{-1}\mathfrak{b}$. Then A_0 is in the center of $U(\mathfrak{b})$. By [Shen 2018, Proposition 6.6], A_0 is closed in U and is diffeomorphic to a circle $\mathbb{S}^1$. Moreover, we have

$$U(\mathfrak{b}) = A_0 U_M. \tag{5.2.2}$$

The bilinear form $-B$ induces an $\mathrm{Ad}(U)$-invariant metric on $\mathfrak{u}$. Let $\mathfrak{u}^{\perp}(\mathfrak{b}) \subset \mathfrak{u}$ be the orthogonal subspace of $\mathfrak{u}(\mathfrak{b})$. Then

$$\mathfrak{u}^{\perp}(\mathfrak{b}) = \sqrt{-1}\mathfrak{p}^{\perp}(\mathfrak{b}) \oplus \mathfrak{k}^{\perp}(\mathfrak{b}). \tag{5.2.3}$$

By (5.1.12), we get

$$[\mathfrak{u}(\mathfrak{b}), \mathfrak{u}(\mathfrak{b})]\,, [\mathfrak{u}^{\perp}(\mathfrak{b}), \mathfrak{u}^{\perp}(\mathfrak{b})] \subset \mathfrak{u}(\mathfrak{b}), \quad [\mathfrak{u}(\mathfrak{b}), \mathfrak{u}^{\perp}(\mathfrak{b})] \subset \mathfrak{u}^{\perp}(\mathfrak{b}). \tag{5.2.4}$$

Then $(\mathfrak{u}, \mathfrak{u}(\mathfrak{b}))$ is a symmetric pair.

Put $a_0 = a_1/\beta(a_1) \in \mathfrak{b}$. Set

$$J = \sqrt{-1}\,\mathrm{ad}(a_0)|_{\mathfrak{u}^{\perp}(\mathfrak{b})} \in \mathrm{End}(\mathfrak{u}^{\perp}(\mathfrak{b})). \tag{5.2.5}$$

By (5.1.9), J is an $U(\mathfrak{b})$-invariant complex structure on $\mathfrak{u}^{\perp}(\mathfrak{b})$ which preserves $B|_{\mathfrak{u}^{\perp}(\mathfrak{b})}$. The spaces $\mathfrak{n}_{\mathbb{C}} = \mathfrak{n} \otimes_{\mathbb{R}} \mathbb{C}$, $\bar{\mathfrak{n}}_{\mathbb{C}} = \bar{\mathfrak{n}} \otimes_{\mathbb{R}} \mathbb{C}$ are exactly the eigenspaces of J associated with eigenvalues $\sqrt{-1}$, $-\sqrt{-1}$.

The following proposition is just the summary of the results in [Shen 2018, Section 6B].

Proposition 5.2.1. *Set*

$$Y_{\mathfrak{b}} = U/U(\mathfrak{b}). \tag{5.2.6}$$

Then $Y_{\mathfrak{b}}$ is a compact symmetric space, and J induces an integrable complex structure on $Y_{\mathfrak{b}}$ such that

$$T^{(1,0)}Y_{\mathfrak{b}} = U \times_{U(\mathfrak{b})} \mathfrak{n}_{\mathbb{C}}, \quad T^{(0,1)}Y_{\mathfrak{b}} = U \times_{U(\mathfrak{b})} \bar{\mathfrak{n}}_{\mathbb{C}}. \tag{5.2.7}$$

The form $-B(\cdot\,, J\,\cdot)$ induces a Kähler form $\omega^{Y_{\mathfrak{b}}}$ on $Y_{\mathfrak{b}}$.

Let $\omega^{\mathfrak{u}}$ be the canonical left-invariant 1-form on U with values in $\mathfrak{u}$. Let $\omega^{\mathfrak{u}(\mathfrak{b})}$ and $\omega^{\mathfrak{u}^{\perp}(\mathfrak{b})}$ be the $\mathfrak{u}(\mathfrak{b})$ and $\mathfrak{u}^{\perp}(\mathfrak{b})$ components of $\omega^{\mathfrak{u}}$, so that

$$\omega^{\mathfrak{u}} = \omega^{\mathfrak{u}(\mathfrak{b})} + \omega^{\mathfrak{u}^{\perp}(\mathfrak{b})}. \tag{5.2.8}$$

Moreover, $\omega^{\mathfrak{u}(\mathfrak{b})}$ defines a connection form on the principal $U(\mathfrak{b})$-bundle $U \to Y_{\mathfrak{b}}$. Let $\Omega^{\mathfrak{u}(\mathfrak{b})}$ be the curvature form. Then

$$\Omega^{\mathfrak{u}(\mathfrak{b})} = -\tfrac{1}{2}[\omega^{\mathfrak{u}^{\perp}(\mathfrak{b})}, \omega^{\mathfrak{u}^{\perp}(\mathfrak{b})}]. \tag{5.2.9}$$

Note that the real tangent bundle of $Y_{\mathfrak{b}}$ is given by

$$TY_{\mathfrak{b}} = U \times_{U(\mathfrak{b})} \mathfrak{u}^{\perp}(\mathfrak{b}). \tag{5.2.10}$$

Then $-B|_{\mathfrak{u}^{\perp}(\mathfrak{b})}$ induces a Riemannian metric $g^{TY_{\mathfrak{b}}}$ on $Y_{\mathfrak{b}}$. The corresponding Levi-Civita connection is induced by $\omega^{\mathfrak{u}(\mathfrak{b})}$.

Recall that the first splitting in (5.2.1) is orthogonal with respect to $-B$. Let $\Omega^{\mathfrak{u}_{\mathfrak{m}}}$ be the $\mathfrak{u}_{\mathfrak{m}}$-component of $\Omega^{\mathfrak{u}(\mathfrak{b})}$. Since the Kähler form $\omega^{Y_{\mathfrak{b}}}$ is invariant under the left action of U on $Y_{\mathfrak{b}}$, we also can view $\omega^{Y_{\mathfrak{b}}}$ as an element in $\Lambda^2(\mathfrak{u}_{\mathfrak{b}}^{\perp})^*)$. By [Shen 2018, equation (6-48)],

$$\Omega^{\mathfrak{u}(\mathfrak{b})} = \beta(a_1)\omega^{Y_{\mathfrak{b}}} \otimes \sqrt{-1}a_1 + \Omega^{\mathfrak{u}_{\mathfrak{m}}}. \tag{5.2.11}$$

Moreover, by [Shen 2018, Proposition 6.9], we have

$$B(\Omega^{\mathfrak{u}(\mathfrak{b})}, \Omega^{\mathfrak{u}(\mathfrak{b})}) = 0, \quad B(\Omega^{\mathfrak{u}_{\mathfrak{m}}}, \Omega^{\mathfrak{u}_{\mathfrak{m}}}) = \beta(a_1)^2 \omega^{Y_{\mathfrak{b}},2}. \tag{5.2.12}$$

Remark 5.2.2. By [Shen 2018, Proposition 6.20], if G has compact center, then as symmetric spaces, the Kähler manifold $Y_{\mathfrak{b}}$ is isomorphic either to $\mathrm{SU}(3)/\mathrm{U}(2)$ or to $\mathrm{SO}(p+q)/\mathrm{SO}(p+q-2) \times \mathrm{SO}(2)$ with $pq > 1$ odd. This way, the computations on $Y_{\mathfrak{b}}$ can be made more explicit.

Now we fix $k \in K_M$. Let $U(k)$ be the centralizer of k in U, and let $U(k)^0$ be its identity component. Let $\mathfrak{u}(k)$ be the Lie algebra of $U(k)^0$. Then $\mathfrak{u}(k)$ is the compact form of $\mathfrak{z}(k)$, and $U(k)^0$ is the compact form of $Z(k)^0$.

We will use the same notation as in Section 5.1. Then the compact form of $\mathfrak{m}(k)$ is given by

$$\mathfrak{u}_{\mathfrak{m}}(k) = \sqrt{-1}\mathfrak{p}_{\mathfrak{m}}(k) \oplus \mathfrak{k}_{\mathfrak{m}}(k). \tag{5.2.13}$$

Let $\mathfrak{u}_{\mathfrak{b}}(k)$ be the compact form of $\mathfrak{z}_{\mathfrak{b}}(k)$. Then

$$\mathfrak{u}_{\mathfrak{b}}(k) = \sqrt{-1}\mathfrak{b} \oplus \mathfrak{u}_{\mathfrak{m}}(k). \tag{5.2.14}$$

Let $U_{\mathfrak{b}}(k)$ be the analytic subgroup associated with $\mathfrak{u}_{\mathfrak{b}}(k)$. Then

$$U_{\mathfrak{b}}(k) = U(\mathfrak{b}) \cap U(k)^0. \tag{5.2.15}$$

Set

$$Y_{\mathfrak{b}}(k) = U(k)^0 / U_{\mathfrak{b}}(k). \tag{5.2.16}$$

As in Proposition 5.2.1, $Y_{\mathfrak{b}}(k)$ is a connected complex manifold equipped with a Kähler form $\omega^{Y_{\mathfrak{b}}(k)}$.

Let $\mathfrak{u}_{\mathfrak{b}}^{\perp}(k)$ be the orthogonal space of $\mathfrak{u}_{\mathfrak{b}}(k)$ in $\mathfrak{u}(k)$ with respect to B. Then

$$\mathfrak{u}_{\mathfrak{b}}^{\perp}(k) = \sqrt{-1}\mathfrak{p}_{\mathfrak{b}}^{\perp}(k) \oplus \mathfrak{k}_{\mathfrak{b}}^{\perp}(k). \tag{5.2.17}$$

Then the real tangent bundle of $Y_{\mathfrak{b}}(k)$ is given by

$$TY_{\mathfrak{b}}(k) = U(k)^0 \times_{U_{\mathfrak{b}}(k)} \mathfrak{u}_{\mathfrak{b}}^{\perp}(k). \tag{5.2.18}$$

Moreover,

$$T^{(1,0)}Y_{\mathfrak{b}}(k) = U(k)^0 \times_{U_{\mathfrak{b}}(k)} \mathfrak{n}(k)_{\mathbb{C}}, \quad T^{(0,1)}Y_{\mathfrak{b}}(k) = U(k)^0 \times_{U_{\mathfrak{b}}(k)} \bar{\mathfrak{n}}(k)_{\mathbb{C}}. \tag{5.2.19}$$

Let $\Omega^{\mathfrak{u}_{\mathfrak{b}}(k)}$ be the curvature form as in (5.2.9) for the pair $(U(k)^0, U_{\mathfrak{b}}(k))$, which can be viewed as an element in $\Lambda^2(\mathfrak{u}_{\mathfrak{b}}^{\perp}(k)^*) \otimes \mathfrak{u}_{\mathfrak{b}}(k)$. Using the splitting (5.2.14), let $\Omega^{\mathfrak{u}_{\mathfrak{m}}(k)}$ be the $\mathfrak{u}_{\mathfrak{m}}(k)$-component of $\Omega^{\mathfrak{u}_{\mathfrak{b}}(k)}$. Then as in (5.2.11) and (5.2.12), we have

$$\Omega^{\mathfrak{u}_{\mathfrak{b}}(k)} = \beta(a_1)\omega^{Y_{\mathfrak{b}}(k)} \otimes \sqrt{-1}a_1 + \Omega^{\mathfrak{u}_{\mathfrak{m}}(k)}, \tag{5.2.20}$$

$$B(\Omega^{\mathfrak{u}_{\mathfrak{b}}(k)}, \Omega^{\mathfrak{u}_{\mathfrak{b}}(k)}) = 0, \quad B(\Omega^{\mathfrak{u}_{\mathfrak{m}}(k)}, \Omega^{\mathfrak{u}_{\mathfrak{m}}(k)}) = \beta(a_1)^2\omega^{Y_{\mathfrak{b}}(b),2}. \tag{5.2.21}$$

5.3. ***Positive root system and character formula.*** Recall that $\mathfrak{t}$ is Cartan subalgebra of $\mathfrak{k}$, of $\mathfrak{k}_{\mathfrak{m}}$, and of $\mathfrak{m}$. Recall that $\mathfrak{h} = \mathfrak{b} \oplus \mathfrak{t}$ is a Cartan subalgebra of $\mathfrak{g}$, and H is the associated maximally compact Cartan subgroup of G.

Put

$$\mathfrak{t}_U = \sqrt{-1}\mathfrak{b} \oplus \mathfrak{t} \subset \mathfrak{u}. \tag{5.3.1}$$

Then $\mathfrak{t}_U$ is a Cartan subalgebra of $\mathfrak{u}$. Let $T_U \subset U$ be the corresponding maximal torus. Then A_0 is a circle in T_U. Then $\mathfrak{t}$ is a Cartan subalgebra of $\mathfrak{u}_{\mathfrak{m}}$, and the corresponding maximal torus is T.

Let $R(\mathfrak{u}, \mathfrak{t}_U)$ be the real root system for the pair (U, T_U) [Bröcker and tom Dieck 1985, Chapter V]. The root system for the complexified pair $(\mathfrak{u}_{\mathbb{C}}, \mathfrak{t}_{U,\mathbb{C}}) = (\mathfrak{g}_{\mathbb{C}}, \mathfrak{h}_{\mathbb{C}})$ is given by $2\pi i R(\mathfrak{u}, \mathfrak{t}_U)$. Similarly, let $R(\mathfrak{u}(\mathfrak{b}), \mathfrak{t}_U)$, $R(\mathfrak{u}_{\mathfrak{m}}, \mathfrak{t})$ denote the real root systems for the pairs $(\mathfrak{u}(\mathfrak{b}), \mathfrak{t}_U)$, $(\mathfrak{u}_{\mathfrak{m}}, \mathfrak{t})$. When we embed $\mathfrak{t}^*$ into $\mathfrak{t}_U^*$ by the splitting in (5.3.1),

$$R(\mathfrak{u}(\mathfrak{b}), \mathfrak{t}_U) = R(\mathfrak{u}_{\mathfrak{m}}, \mathfrak{t}). \tag{5.3.2}$$

For a root $\alpha \in R(\mathfrak{u}, \mathfrak{t}_U)$, if $\alpha(\sqrt{-1}a_1) = 0$, then $\alpha \in R(\mathfrak{u}_{\mathfrak{m}}, \mathfrak{t})$. Fix a positive root system $R^+(\mathfrak{u}_{\mathfrak{m}}, \mathfrak{t})$. We get a positive root system $R^+(\mathfrak{u}, \mathfrak{t}_U)$ consisting of an element α such that $\alpha(\sqrt{-1}a_1) > 0$ and the elements in $R^+(\mathfrak{u}_{\mathfrak{m}}, \mathfrak{t})$.

Let $W(\mathfrak{u}, \mathfrak{t}_U)$ denote the algebraic Weyl group associated with $R(\mathfrak{u}, \mathfrak{t}_U)$. If $\omega \in W(\mathfrak{u}, \mathfrak{t}_U)$, let $l(\omega)$ denote the length of ω with respect to $R^+(\mathfrak{u}, \mathfrak{t}_U)$. Set

$$\varepsilon(\omega) = (-1)^{l(\omega)}. \tag{5.3.3}$$

Let $W(U, T_U)$ be the analytic Weyl group. Then $W(\mathfrak{u}, \mathfrak{t}_U) = W(U, T_U)$.

Put

$$W_u = \{\omega \in W(U, T_U) \mid \omega^{-1} \cdot \alpha > 0 \text{ for all } \alpha \in R^+(\mathfrak{u}_{\mathfrak{m}}, \mathfrak{t})\}. \tag{5.3.4}$$

Put

$$\rho_{\mathfrak{u}} = \frac{1}{2} \sum_{\alpha^0 \in R^+(\mathfrak{u}, \mathfrak{t}_U)} \alpha^0 \in \mathfrak{t}_U^*, \quad \rho_{\mathfrak{u}_\mathfrak{m}} = \frac{1}{2} \sum_{\alpha^0 \in R^+(\mathfrak{u}_\mathfrak{m}, \mathfrak{t})} \alpha^0 \in \mathfrak{t}^*. \tag{5.3.5}$$

Then $\rho_{\mathfrak{u}}|_{\mathfrak{t}} = \rho_{\mathfrak{u}_\mathfrak{m}}$.

Let $P_{++}(U) \subset \mathfrak{t}_U^*$ be the set of dominant weights of (U, T_U) with respect to $R^+(\mathfrak{u}, \mathfrak{t}_U)$. If $\lambda \in P_{++}(U)$, let $(E_\lambda, \rho^{E_\lambda})$ be the irreducible unitary representation of U with the highest weight λ, which by the unitary trick extends to an irreducible representation of G.

By [Warner 1972, Lemmas 1.1.2.15, 2.4.2.1], if $\omega \in W_u$, then $\omega(\lambda + \rho_{\mathfrak{u}}) - \rho_{\mathfrak{u}}$ is a dominant weight for $R^+(\mathfrak{u}(\mathfrak{b}), \mathfrak{t}_U)$. Let $V_{\lambda,\omega}$ denote the representation of $U(\mathfrak{b})$ with the highest weight $\omega(\lambda + \rho_{\mathfrak{u}}) - \rho_{\mathfrak{u}}$.

Recall that $U(\mathfrak{b})$ acts on $\mathfrak{n}_\mathbb{C}$. Let $H^\bullet(\mathfrak{n}_\mathbb{C}, E_\lambda)$ be the Lie algebra cohomology of $\mathfrak{n}_\mathbb{C}$ with coefficients in E_λ; see [Kostant 1961]. By [Warner 1972, Theorem 2.5.1.3], for $i = 0, \dots, 2l$, we have the identification of $U(\mathfrak{b})$-modules

$$H^i(\mathfrak{n}_\mathbb{C}, E_\lambda) \simeq \bigoplus_{\substack{\omega \in W_u \\ l(\omega) = i}} V_{\lambda,\omega}. \tag{5.3.6}$$

By (5.3.6) and the Poincaré duality, we get the following identifications as $U(\mathfrak{b})$-modules:

$$\bigoplus_{i=0}^{2l} (-1)^i \Lambda^i \mathfrak{n}_\mathbb{C}^* \otimes E_\lambda = \bigoplus_{\omega \in W_u} \varepsilon(\omega) V_{\lambda,\omega}. \tag{5.3.7}$$

Note that if we apply the unitary trick, the above identification also holds as $Z(\mathfrak{b})^0$-modules.

Definition 5.3.1. Let $P_0 : \mathfrak{t}_U^* \to \mathfrak{t}^*$ denote the orthogonal projection with respect to $B^*|_{\mathfrak{t}_U^*}$. For $\omega \in W_u$, $\lambda \in P_{++}(U)$, put

$$\eta_\omega(\lambda) = P_0(\omega(\lambda + \rho_{\mathfrak{u}}) - \rho_{\mathfrak{u}}) \in \mathfrak{t}^*. \tag{5.3.8}$$

Note that

$$P_0 \rho_{\mathfrak{u}} = \rho_{\mathfrak{u}_\mathfrak{m}}. \tag{5.3.9}$$

Then

$$\eta_\omega(\lambda) = P_0(\omega(\lambda + \rho_{\mathfrak{u}})) - \rho_{\mathfrak{u}_\mathfrak{m}}. \tag{5.3.10}$$

Proposition 5.3.2. *If $\lambda \in P_{++}(U)$, for $\omega \in W_u$, then $\eta_\omega(\lambda)$ is a dominant weight of (U_M, T) with respect to $R^+(\mathfrak{u}_\mathfrak{m}, \mathfrak{t})$. Moreover, the restriction of the $U(\mathfrak{b})$-representation $V_{\lambda,\omega}$ to the subgroup U_M is irreducible, which has the highest weight $\eta_\omega(\lambda)$.*

Proof. Since $\omega(\lambda + \rho_{\mathfrak{u}}) - \rho_{\mathfrak{u}}$ is analytically integrable, $\eta_\omega(\lambda)$ is also analytically integrable as a weight associated with (U_M, T). By (5.3.2) and the corresponding identification of positive root systems, we know that $\eta_\omega(\lambda)$ is dominant with respect to $R^+(\mathfrak{u}_\mathfrak{m}, \mathfrak{t})$.

Recall that $A_0 \simeq \mathbb{S}^1$ is defined in Section 5.2. By (5.2.2), we get that A_0 acts on $V_{\lambda,\omega}$ as scalars given by its character, and then U_M act irreducibly on $V_{\lambda,\omega}$, which clearly has the highest weight $\eta_\omega(\lambda)$. □

Remark 5.3.3. In general, U is just the analytic subgroup of $G_\mathbb{C}$ with Lie algebra $\mathfrak{u}$. If U is compact but G has noncompact center, i.e., $\mathfrak{z}_\mathfrak{p} = \mathfrak{b}$, then $\mathfrak{n} = \bar{\mathfrak{n}} = 0$, so that $l = 0$. Recall that in this case, G', U' are defined in Section 4.1. Then

$$M = G', \quad U_M = U'. \tag{5.3.11}$$

The compact symmetric space $Y_{\mathfrak{b}}$ now reduces to one point.

Moreover, in (5.3.4), $W_u = \{1\}$, so that $V_{\lambda,\omega}$ becomes just E_λ itself. The identities (5.3.6), (5.3.7) are trivially true; so is Proposition 5.3.2.

5.4. *Kirillov character formula for compact Lie groups.* In this subsection, we recall the Kirillov character formula for compact Lie groups. We only use the group U_M as an explanatory example. We fix the maximal torus T and the positive (real) root system $R^+(\mathfrak{u}_\mathfrak{m}, \mathfrak{t})$.

Let $\lambda \in \mathfrak{t}^*$ be a dominant (analytically integrable) weight of U_M with respect to the above root system. Let $(V_\lambda, \rho^{V_\lambda})$ be the irreducible unitary representation of U_M with the highest weight λ.

Put

$$\mathcal{O} = \mathrm{Ad}^*(U_M)(\lambda + \rho_{\mathfrak{u}_\mathfrak{m}}) \subset \mathfrak{u}_\mathfrak{m}^*. \tag{5.4.1}$$

Then $\mathcal{O}$ is an even-dimensional closed manifold.

Since $\lambda + \rho_{\mathfrak{u}_\mathfrak{m}}$ is regular, we have the following identifications of U_M-manifolds:

$$\mathcal{O} \simeq U_M/T. \tag{5.4.2}$$

For $u \in \mathfrak{u}_\mathfrak{m}$, an associated vector field $\tilde{u}$ on $\mathcal{O}$ is defined as follows: if $f \in \mathcal{O}$, then

$$\tilde{u}_f = -\,\mathrm{ad}^*(u) f \in T_f\mathcal{O}. \tag{5.4.3}$$

Such vector fields span the whole tangent space at each point. Let ω_L denote the real 2-form on $\mathcal{O}$ such that if $u, v \in \mathfrak{u}_\mathfrak{m}$, $f \in \mathcal{O}$,

$$\omega_L(\tilde{u}, \tilde{v})_f = -\langle f, [u, v]\rangle. \tag{5.4.4}$$

Then ω_L is a U_M-invariant symplectic form on $\mathcal{O}$. Put $r^+ = \frac{1}{2}\dim \mathfrak{u}_\mathfrak{m}/\mathfrak{t}$. In fact, if we can define an almost complex structure on $T\mathcal{O}$ such that the holomorphic tangent bundle is given by the positive root system $R^+(\mathfrak{u}_\mathfrak{m}, \mathfrak{t})$. Then $(\mathcal{O}, \omega_L)$ become a closed Kähler manifold, and r^+ is its complex dimension.

The Liouville measure on $\mathcal{O}$ is defined as

$$d\mu_L = \frac{(\omega_L)^{r^+}}{(r^+)!}. \tag{5.4.5}$$

It is invariant by the left action of U_M. Let $\mathrm{Vol}_L(\mathcal{O})$ denote the symplectic volume of $\mathcal{O}$ with respect to the Liouville measure. Then we have (see [Berline et al. 1992, Proposition 7.26])

$$\mathrm{Vol}_L(\mathcal{O}) = \Pi_{\alpha^0 \in R^+(\mathfrak{u}_\mathfrak{m}, \mathfrak{t})} \frac{\langle \alpha^0, \lambda + \rho_{\mathfrak{u}_\mathfrak{m}}\rangle}{\langle \alpha^0, \rho_{\mathfrak{u}_\mathfrak{m}}\rangle} = \dim V_\lambda. \tag{5.4.6}$$

The second identity is the Weyl dimension formula (see [Knapp 1986, Theorem 4.48]).

By the Kirillov formula (see [Berline et al. 1992, Theorem 8.4]), if $y \in \mathfrak{u}_\mathfrak{m}$, we have

$$\widehat{A}^{-1}(\mathrm{ad}(y)|_{\mathfrak{u}_\mathfrak{m}})\, \mathrm{Tr}^{V_\lambda}[\rho^{V_\lambda}(e^y)] = \int_{f \in \mathcal{O}} e^{2\pi i \langle f, y\rangle}\, d\mu_L. \tag{5.4.7}$$

To shorten the notation here, if $k \in T$, put $Y = U_M(k)^0$ with Lie algebra $\mathfrak{y} = \mathfrak{u}_\mathfrak{m}(k)$. Then $T \subset Y$, and it also a maximal torus of Y.

In the sequel, we will give a generalized version of (5.4.7) for describing the function $\mathrm{Tr}^{V_\lambda}[\rho^{V_\lambda}(ke^y)]$, with $y \in \mathfrak{y}$.

Let $\mathfrak{q}$ be the orthogonal space of $\mathfrak{y}$ in $\mathfrak{u}_\mathfrak{m}$ with respect to B, so that

$$\mathfrak{u}_\mathfrak{m} = \mathfrak{y} \oplus \mathfrak{q}. \tag{5.4.8}$$

Since the adjoint action of T preserves the splitting in (5.4.8). Then $R(\mathfrak{u}_\mathfrak{m}, \mathfrak{t})$ splits into two disjoint parts

$$R(\mathfrak{u}_\mathfrak{m}, \mathfrak{t}) = R(\mathfrak{y}, \mathfrak{t}) \cup R(\mathfrak{q}, \mathfrak{t}), \tag{5.4.9}$$

where $R(\mathfrak{q}, \mathfrak{t})$ is just the set of real roots for the adjoint action of $\mathfrak{t}$ on $\mathfrak{q}_\mathbb{C}$.

The positive root system $R^+(\mathfrak{u}_\mathfrak{m}, \mathfrak{t})$ induces a positive root system $R^+(\mathfrak{y}, \mathfrak{t})$. Set

$$R^+(\mathfrak{q}, \mathfrak{t}) = R^+(\mathfrak{u}_\mathfrak{m}, \mathfrak{t}) \cap R(\mathfrak{q}, \mathfrak{t}). \tag{5.4.10}$$

Then we have the disjoint union

$$R^+(\mathfrak{u}_\mathfrak{m}, \mathfrak{t}) = R^+(\mathfrak{y}, \mathfrak{t}) \cup R^+(\mathfrak{q}, \mathfrak{t}). \tag{5.4.11}$$

Put

$$\rho_\mathfrak{y} = \frac{1}{2} \sum_{\alpha^0 \in R^+(\mathfrak{y},\mathfrak{t})} \alpha^0, \quad \rho_\mathfrak{q} = \frac{1}{2} \sum_{\alpha^0 \in R^+(\mathfrak{q},\mathfrak{t})} \alpha^0. \tag{5.4.12}$$

Then

$$\rho_{\mathfrak{u}_\mathfrak{m}} = \rho_\mathfrak{y} + \rho_\mathfrak{q} \in \mathfrak{t}^*. \tag{5.4.13}$$

Let $\mathcal{C} \subset \mathfrak{t}^*$ denote the Weyl chamber corresponding to $R^+(\mathfrak{u}_\mathfrak{m}, \mathfrak{t})$, and let $\mathcal{C}_0 \subset \mathfrak{t}^*$ denote the Weyl chamber corresponding to $R^+(\mathfrak{y}, \mathfrak{t})$. Then $\mathcal{C} \subset \mathcal{C}_0$.

Let $W(U_M, T)$, $W(Y, T)$ be the Weyl groups associated with the pairs (U_M, T), (Y, T) respectively. Then $W(Y, T)$ is canonically a subgroup of $W(U_M, T)$. Put

$$W^1(k) = \{\omega \in W(U_M, T) \mid \omega(\mathcal{C}) \subset \mathcal{C}_0\}. \tag{5.4.14}$$

Note that the set $W^1(k)$ is similar to the set W_u defined in (5.3.4).

Lemma 5.4.1. *The inclusion $W^1(k) \hookrightarrow W(U_M, T)$ induces a bijection between $W^1(k)$ and the quotient $W(Y, T)\backslash W(U_M, T)$.*

Proof. This lemma follows from $W(Y, T)$ acting simply transitively on the Weyl chambers associated with $(\mathfrak{y}, \mathfrak{t})$. □

Let $\mathcal{O}^k$ denote the fixed-point set of the holomorphic action of k on $\mathcal{O}$. We embed $\mathfrak{y}^*$ in $\mathfrak{u}_\mathfrak{m}^*$ by the splitting (5.4.8). Then

$$\mathcal{O}^k = \mathcal{O} \cap \mathfrak{y}^*. \tag{5.4.15}$$

Lemma 5.4.2 (see [Duflo et al. 1984, I.2, Lemma (7); Bouaziz 1987, Lemmas 6.1.1, 7.2.2]). *As subsets of $\mathfrak{y}^*$, we have the identification*

$$\mathcal{O}^k = \bigcup_{\sigma \in W^1(k)} \mathrm{Ad}^*(Y)(\sigma(\lambda + \rho_{\mathfrak{u}_\mathfrak{m}})) \subset \mathfrak{y}^*, \tag{5.4.16}$$

where the union is disjoint.

For each $\sigma \in W^1(k)$, put

$$\mathcal{O}^k_{\sigma(\lambda+\rho_{\mathfrak{u}_\mathfrak{m}})} = \mathrm{Ad}^*(Y)(\sigma(\lambda+\rho_{\mathfrak{u}_\mathfrak{m}})) \subset \mathfrak{y}^*. \tag{5.4.17}$$

Let $d\mu^k_\sigma$ denote the Liouville measure on $\mathcal{O}^k_{\sigma(\lambda+\rho_{\mathfrak{u}_\mathfrak{m}})}$ as defined in (5.4.5).

If $\delta \in \mathfrak{t}^*$ is (real) analytically integrable, let ξ_δ denote the character of T with differential $2\pi i\delta$. Note that for $\sigma \in W^1(k)$, $\sigma\rho_{\mathfrak{u}_\mathfrak{m}} + \rho_{\mathfrak{u}_\mathfrak{m}}$ is analytically integrable even though $\rho_{\mathfrak{u}_\mathfrak{m}}$ may not be analytically integrable.

Definition 5.4.3. For $\sigma \in W^1(k)$, set

$$\varphi_k(\sigma,\lambda) = \varepsilon(\sigma)\frac{\xi_{\sigma(\lambda+\rho_{\mathfrak{u}_\mathfrak{m}})+\rho_{\mathfrak{u}_\mathfrak{m}}}(k)}{\Pi_{\alpha^0\in R^+(\mathfrak{q},\mathfrak{t})}(\xi_{\alpha^0}(k)-1)}. \tag{5.4.18}$$

Note that if $y \in \mathfrak{y}$, the analytic function

$$\frac{\det(1-e^{\mathrm{ad}(y)}\,\mathrm{Ad}(k))|_\mathfrak{q}}{\det(1-\mathrm{Ad}(k))|_\mathfrak{q}} \tag{5.4.19}$$

has a square root which is analytic in $y \in \mathfrak{y}$ and equal to 1 at $y = 0$. We denote this square root by

$$\left[\frac{\det(1-e^{\mathrm{ad}(y)}\,\mathrm{Ad}(k))|_\mathfrak{q}}{\det(1-\mathrm{Ad}(k))|_\mathfrak{q}}\right]^{\frac{1}{2}}. \tag{5.4.20}$$

The following theorem is a special case of a generalized Kirillov formula obtained by Duflo, Heckman and Vergne [Duflo et al. 1984, II.3, Theorem (7)]. We will also include a simpler proof for the sake of completeness.

Theorem 5.4.4 (generalized Kirillov formula). *For $y \in \mathfrak{y}$, we have the identity of analytic functions*

$$\widehat{A}^{-1}(\mathrm{ad}(y)|_\mathfrak{y})\left[\frac{\det(1-e^{\mathrm{ad}(y)}\,\mathrm{Ad}(k))|_\mathfrak{q}}{\det(1-\mathrm{Ad}(k))_\mathfrak{q}}\right]^{\frac{1}{2}}\mathrm{Tr}^{V_\lambda}[\rho^{V_\lambda}(ke^y)] = \sum_{\sigma\in W^1(k)}\varphi_k(\sigma,\lambda)\int_{f\in\mathcal{O}^k_{\sigma(\lambda+\rho_{\mathfrak{u}_\mathfrak{m}})}} e^{2\pi i\langle f,y\rangle}\,d\mu^k_\sigma. \tag{5.4.21}$$

If $k = 1$, (5.4.21) *is reduced to* (5.4.7).

Proof. Let $\mathfrak{t}'$ denote the set of regular element in $\mathfrak{t}$ associated with the root $R(\mathfrak{u}_\mathfrak{m},\mathfrak{t})$, which is an open dense subset of $\mathfrak{t}$. Since both sides of (5.4.21) are analytic and invariant by the adjoint action of Y, we only need to prove (5.4.21) for $y \in \mathfrak{t}'$.

We firstly compute the left-hand side of (5.4.21).

For $y \in \mathfrak{t}'$, we have

$$\widehat{A}^{-1}(\mathrm{ad}(y)|_\mathfrak{y}) = \Pi_{\alpha^0\in R^+(\mathfrak{y},\mathfrak{t})}\frac{e^{\pi i\langle\alpha^0,y\rangle}-e^{-\pi i\langle\alpha^0,y\rangle}}{\langle 2\pi i\alpha^0,y\rangle}. \tag{5.4.22}$$

Let $y_0 \in \mathfrak{t}$ be such that $k = \exp(y_0)$. Then

$$\left[\frac{\det(1-e^{\mathrm{ad}(y)}\,\mathrm{Ad}(k))|_\mathfrak{q}}{\det(1-\mathrm{Ad}(k))|_\mathfrak{q}}\right]^{\frac{1}{2}} = \Pi_{\alpha^0\in R^+(\mathfrak{q},\mathfrak{t})}\frac{e^{\pi i\langle\alpha^0,y+y_0\rangle}-e^{-\pi i\langle\alpha^0,y+y_0\rangle}}{e^{\pi i\langle\alpha^0,y_0\rangle}-e^{-\pi i\langle\alpha^0,y_0\rangle}}. \tag{5.4.23}$$

By the Weyl character formula for (U_M, T), we get

$$\mathrm{Tr}^{V_\lambda}[\rho^{V_\lambda}(ke^y)] = \mathrm{Tr}^{V_\lambda}[\rho^{V_\lambda}(e^{y+y_0})] = \frac{\sum_{\omega\in W(\mathfrak{u}_{\mathfrak{m},\mathbb{C}},\mathfrak{t}_{\mathbb{C}})} \varepsilon(\omega)e^{2\pi i\langle\omega(\lambda+\rho_{\mathfrak{u}_\mathfrak{m}}),y+y_0\rangle}}{\prod_{\alpha^0\in R^+(\mathfrak{u}_\mathfrak{m},\mathfrak{t})}(e^{\pi i\langle\alpha^0,y+y_0\rangle}-e^{-\pi i\langle\alpha^0,y+y_0\rangle})}. \tag{5.4.24}$$

Note that we have $\xi_{\alpha_0}(k) = 1$ for $\alpha_0 \in R^+(\mathfrak{y},\mathfrak{t})$. Then

$$\prod_{\alpha^0\in R^+(\mathfrak{y},\mathfrak{t})} \frac{e^{\pi i\langle\alpha^0,y+y_0\rangle}-e^{-\pi i\langle\alpha^0,y+y_0\rangle}}{e^{\pi i\langle\alpha^0,y\rangle}-e^{-\pi i\langle\alpha^0,y\rangle}} = e^{-2\pi i\langle\rho_\mathfrak{y},y_0\rangle}. \tag{5.4.25}$$

Combining (5.4.22)–(5.4.25), we get the left-hand side of (5.4.21) is equal the to function

$$\frac{e^{2\pi i\langle\rho_\mathfrak{y},y_0\rangle}}{\prod_{\alpha^0\in R^+(\mathfrak{y},\mathfrak{t})}\langle 2\pi i\alpha^0, y\rangle} \frac{\sum_{\omega\in W(\mathfrak{u}_{\mathfrak{m},\mathbb{C}},\mathfrak{t}_{\mathbb{C}})} \varepsilon(\omega)e^{2\pi i\langle\omega(\lambda+\rho_{\mathfrak{u}_\mathfrak{m}}),y+y_0\rangle}}{\prod_{\alpha^0\in R^+(\mathfrak{q},\mathfrak{t})}(e^{\pi i\langle\alpha^0,y\rangle}-e^{-\pi i\langle\alpha^0,y\rangle})}. \tag{5.4.26}$$

Now we show that the right-hand side of (5.4.21) is also equal to (5.4.26).

Note that, for $\omega \in W(Y,T)$, $\omega\rho_{\mathfrak{u}_\mathfrak{m}} - \rho_{\mathfrak{u}_\mathfrak{m}}$ is analytically integrable. We claim that if $\omega \in W(Y,T)$, then

$$\xi_{\omega\rho_{\mathfrak{u}_\mathfrak{m}}-\rho_{\mathfrak{u}_\mathfrak{m}}}(k) = e^{2\pi i\langle\omega\rho_{\mathfrak{u}_\mathfrak{m}}-\rho_{\mathfrak{u}_\mathfrak{m}},y_0\rangle} = 1. \tag{5.4.27}$$

Actually, we have $\xi_{2\rho_{\mathfrak{u}_\mathfrak{m}}}(k) = \xi_{2\omega\rho_{\mathfrak{u}_\mathfrak{m}}}(k) = 1$. Then, after taking the square roots, we get $\xi_{\omega\rho_{\mathfrak{u}_\mathfrak{m}}-\rho_{\mathfrak{u}_\mathfrak{m}}}(k) = \xi_{\omega\rho_{\mathfrak{u}_\mathfrak{m}}-\rho_{\mathfrak{u}_\mathfrak{m}}}(e^{y_0}) = \pm 1$. The continuity of the character implies exactly (5.4.27).

As a consequence of (5.4.27), we get that for $\sigma \in W^1(\mathfrak{k})$, if $\omega \in W(Y,T)$, then

$$e^{2\pi i\langle\omega\sigma(\lambda+\rho_{\mathfrak{u}_\mathfrak{m}}),y_0\rangle} = e^{2\pi i\langle\sigma(\lambda+\rho_{\mathfrak{u}_\mathfrak{m}}),y_0\rangle}. \tag{5.4.28}$$

For $\sigma \in W^1(\mathfrak{k})$, since $\sigma(\lambda+\rho_{\mathfrak{u}_\mathfrak{m}}) \in \mathcal{C}_0$ and y is regular, by [Berline et al. 1992, Corollary 7.25], we have

$$\int_{f\in\mathcal{O}^k_{\sigma(\lambda+\rho_{\mathfrak{u}_\mathfrak{m}})}} e^{2\pi i\langle f,y\rangle}\, d\mu^k_\sigma = \frac{1}{\prod_{\alpha^0\in R^+(\mathfrak{y},\mathfrak{t})}\langle 2\pi i\alpha^0, y\rangle} \sum_{\omega\in W(Y,T)} \varepsilon(\omega)e^{2\pi i\langle\omega\sigma(\lambda+\rho_{\mathfrak{u}_\mathfrak{m}}),y\rangle}. \tag{5.4.29}$$

We rewrite $\varphi_k(\sigma,\lambda)$ as

$$\varepsilon(\sigma)\frac{e^{2\pi i\langle\rho_\mathfrak{y},y_0\rangle}}{\prod_{\alpha^0\in R^+(\mathfrak{q},\mathfrak{t})}(e^{\pi i\langle\alpha^0,y\rangle}-e^{-\pi i\langle\alpha^0,y\rangle})}e^{2\pi i\langle\sigma(\lambda+\rho_{\mathfrak{u}_\mathfrak{m}}),y_0\rangle}. \tag{5.4.30}$$

Combining together Lemma 5.4.1 and (5.4.28)–(5.4.30), a direct computation shows that the right-hand side of (5.4.21) is given exactly by (5.4.26). □

Remark 5.4.5. Let C^0 denote the identity component of the center of Y, and let Y_{ss} be the closed analytic subgroup of Y associated with $\mathfrak{y}_{ss} = [\mathfrak{y},\mathfrak{y}]$. By Weyl's theorem [Knapp 1986, Theorem 4.26], the universal covering group of Y_{ss} is compact, which we denote by $\widetilde{Y}_{ss}$. Put

$$\widetilde{Y} = C^0 \times \widetilde{Y}_{ss}. \tag{5.4.31}$$

Then $\widetilde{Y}$ is clearly a finite central extension of Y. Let $\widetilde{T}$ be the maximal torus of $\widetilde{Y}$ associated with the Cartan subalgebra $\mathfrak{t}$, which is also a finite extension of T. By [Knapp 1986, Corollary 4.25], the weights

$\rho_{\mathfrak{u}_{\mathfrak{m}}}, \rho_{\mathfrak{y}}$ are analytically integrable with respect to $\widetilde{T}$, since they are algebraically integrable [Knapp 1986, Propositions 4.15, 4.33].

Note that, for $\sigma \in W^1(k)$, $\sigma(\lambda + \rho_{\mathfrak{u}_{\mathfrak{m}}})$ is regular and positive with respect to $R^+(\mathfrak{y}, \mathfrak{t})$; thus $\sigma(\lambda + \rho_{\mathfrak{u}_{\mathfrak{m}}}) - \rho_{\mathfrak{y}}$ is nonnegative with respect to $R^+(\mathfrak{y}, \mathfrak{t})$ by the property of $\rho_{\mathfrak{y}}$ [Knapp 1986, Proposition 4.33]. Since now $\sigma(\lambda + \rho_{\mathfrak{u}_{\mathfrak{m}}}) - \rho_{\mathfrak{y}}$ is also analytically integrable with respect to $\widetilde{T}$, it is a dominant weight for $(\widetilde{Y}, \widetilde{T})$ with respect to $R^+(\mathfrak{y}, \mathfrak{t})$. In this case, let $V^k_{\lambda,\sigma}$ be the irreducible unitary representation of $\widetilde{Y}$ with highest weight $\sigma(\lambda + \rho_{\mathfrak{u}_{\mathfrak{m}}}) - \rho_{\mathfrak{y}}$. Then by (5.4.7), (5.4.21), we get that, for $y \in \mathfrak{y}$,

$$\left[\frac{\det(1 - e^{\mathrm{ad}(y)} \operatorname{Ad}(k))|_{\mathfrak{q}}}{\det(1 - \operatorname{Ad}(k))_{\mathfrak{q}}}\right]^{\frac{1}{2}} \operatorname{Tr}^{V_\lambda}[\rho^{V_\lambda}(ke^y)] = \sum_{\sigma \in W^1(k)} \varphi_k(\sigma, \lambda) \operatorname{Tr}^{V^k_{\lambda,\sigma}}[\rho^{V^k_{\lambda,\sigma}}(e^y)]. \tag{5.4.32}$$

6. A geometric localization formula for orbital integrals

Recall that $G_{\mathbb{C}}$ is the complexification of G with Lie algebra $\mathfrak{g}_{\mathbb{C}}$, and that G, U are the analytic subgroups of $G_{\mathbb{C}}$ with Lie algebras $\mathfrak{g}$, $\mathfrak{u}$ respectively. In this section, we always assume that U is compact; we do not require that G have compact center. We need not to assume $\delta(G) = 1$ either.

Under the settings in Section 4.1, for $t > 0$ and semisimple $\gamma \in G$, we set

$$\mathcal{E}_{X,\gamma}(F, t) = \operatorname{Tr}_s^{[\gamma]}\Big[\Big(N^{\Lambda^\bullet(T^*X)} - \frac{m}{2}\Big) \exp\Big(-\frac{t\boldsymbol{D}^{X,F,2}}{2}\Big)\Big]. \tag{6.0.1}$$

The indices X, F in this notation indicate precisely the symmetric space and the flat vector bundle which are concerned with defining the orbital integrals.

If $\gamma \in G$ is semisimple, then there exists a unique elliptic element γ_e and a unique hyperbolic element γ_h in G such that $\gamma = \gamma_e\gamma_h = \gamma_h\gamma_e$. Here, we will show that $\mathcal{E}_{X,\gamma}(F, t)$ becomes a sum of the orbital integrals associated with γ_h, but defined for the centralizer of γ_e instead of G. This suggests that the elliptic part of γ should lead to a localization for the geometric orbital integrals.

We still fix a maximal torus T of K with Lie algebra $\mathfrak{t}$. For simplicity, if $\gamma \in G$ is semisimple, we may and we will assume

$$\gamma = e^a k, \quad k \in T, \quad a \in \mathfrak{p}, \quad \operatorname{Ad}(k^{-1})a = a. \tag{6.0.2}$$

In this case,

$$\gamma_e = k \in T, \quad \gamma_h = e^a. \tag{6.0.3}$$

Recall that $Z(\gamma_e)^0$ is the identity component of the centralizer of γ_e in G. Then

$$\gamma_h \in Z(\gamma_e)^0. \tag{6.0.4}$$

The Cartan involution θ preserves $Z(\gamma_e)^0$ such that $Z(\gamma_e)^0$ is a connected linear reductive Lie group. Then we have the diffeomorphism

$$Z(\gamma_e)^0 = K(\gamma_e)^0 \exp(\mathfrak{p}(\gamma_e)). \tag{6.0.5}$$

It is clear that $\delta(Z(\gamma_e)^0) = \delta(G)$.

Recall that T_U is a maximal torus of U with Lie algebra $\mathfrak{t}_U = \sqrt{-1}\mathfrak{b} \oplus \mathfrak{t} \subset \mathfrak{u}$. Let $R^+(\mathfrak{u}, \mathfrak{t}_U)$ be a positive root system for $R(\mathfrak{u}, \mathfrak{t}_U)$, which is not necessarily the same as in Section 5.3 when $\delta(G) = 1$.

Since U is the compact form of G, $U(\gamma_e)^0$ is the compact form for $Z(\gamma_e)^0$. Moreover, T_U is also a maximal torus of $U(\gamma_e)^0$. Let $R(\mathfrak{u}(\gamma_e), \mathfrak{t}_U)$ be the corresponding real root system with the positive root system $R^+(\mathfrak{u}(\gamma_e), \mathfrak{t}_U) = R(\mathfrak{u}(\gamma_e), \mathfrak{t}_U) \cap R^+(\mathfrak{u}, \mathfrak{t}_U)$. Let $\rho_{\mathfrak{u}}$, $\rho_{\mathfrak{u}(\gamma_e)}$ be the corresponding half sums of positive roots.

Let $\widetilde{U}(\gamma_e)$ be a connected finite covering group of $U(\gamma_e)^0$ such that $\rho_{\mathfrak{u}}$, $\rho_{\mathfrak{u}(\gamma_e)}$ are analytically integrable with respect to the maximal torus $\widetilde{T}_U$ of $\widetilde{U}(\gamma_e)$ associated with $\mathfrak{t}_U$. It always exists by a construction similar to that in Remark 5.4.5.

Let $\widetilde{K}(\gamma_e)$ be the analytic subgroup of $\widetilde{U}(\gamma_e)$ associated with the Lie algebra $\mathfrak{k}(\gamma_e)$. By [Knapp 2002, Proposition 7.12], $\widetilde{U}(\gamma_e)$ has a unique complexification $\widetilde{U}(\gamma_e)_{\mathbb{C}}$ which is a connected linear reductive Lie group. Let $\widetilde{Z}(\gamma_e)$ be the analytic subgroup of $\widetilde{U}(\gamma_e)_{\mathbb{C}}$ associated with $\mathfrak{z}(\gamma_e) \subset \mathfrak{u}(\gamma_e)_{\mathbb{C}} = \mathfrak{z}(\gamma_e)_{\mathbb{C}}$. Then we have the Cartan decomposition

$$\widetilde{Z}(\gamma_e) = \widetilde{K}(\gamma_e) \exp(\mathfrak{p}(\gamma_e)). \tag{6.0.6}$$

We still denote by θ the corresponding Cartan involution on $\widetilde{Z}(\gamma_e)$.

The Lie group $\widetilde{Z}(\gamma_e)$ is a finite covering group of $Z(\gamma_e)^0$. Moreover, we have the identification of symmetric spaces

$$X(\gamma_e) \simeq \widetilde{Z}(\gamma_e)/\widetilde{K}(\gamma_e). \tag{6.0.7}$$

Note that even under an additional assumption that G has compact center, $\widetilde{Z}(\gamma_e)$ may still have noncompact center.

Let λ be a dominant weight for (U, T_U) with respect to $R^+(\mathfrak{u}, \mathfrak{t}_U)$. Let $(E_\lambda, \rho^{E_\lambda})$ be the associated irreducible unitary representation of U. As before, let $(F_\lambda, h^{F_\lambda})$ be the corresponding homogeneous vector bundle on X with the G-invariant flat connection $\nabla^{F_\lambda, f}$. Let $\boldsymbol{D}^{X, F_\lambda, 2}$ denote the associated de Rham–Hodge Laplacian.

Let $W^1_U(\gamma_e) \subset W(U, T_U)$ be the set defined as in (5.4.14) but with respect to the group U and to $\gamma_e = k \in T \subset T_U$. As in Definition 5.4.3, for $\sigma \in W^1_U(\gamma_e)$, set

$$\varphi^U_{\gamma_e}(\sigma, \lambda) = \varepsilon(\sigma) \frac{\xi_{\sigma(\lambda+\rho_{\mathfrak{u}})+\rho_{\mathfrak{u}}}(\gamma_e)}{\prod_{\alpha^0 \in R^+(\mathfrak{u}^\perp(\gamma_e), \mathfrak{t}_U)} (\xi_{\alpha^0}(\gamma_e) - 1)}. \tag{6.0.8}$$

As explained in Remark 5.4.5, if $\sigma \in W^1_U(\gamma_e)$, then $\sigma(\lambda + \rho_{\mathfrak{u}}) - \rho_{\mathfrak{u}(\gamma_e)}$ is a dominant weight of $\widetilde{U}(\gamma_e)$ with respect to $R^+(\mathfrak{u}(\gamma_e), \mathfrak{t}_U)$. Let $E_{\sigma,\lambda}$ be the irreducible unitary representation of $\widetilde{U}(\gamma_e)$ with highest weight $\sigma(\lambda + \rho_{\mathfrak{u}}) - \rho_{\mathfrak{u}(\gamma_e)}$.

We extend $E_{\sigma,\lambda}$ to an irreducible representation of $\widetilde{Z}(\gamma_e)$ by the unitary trick. Then

$$F_{\sigma,\lambda} = \widetilde{Z}(\gamma_e) \times_{\widetilde{K}(\gamma_e)} E_{\sigma,\lambda}$$

is a homogeneous vector bundle on $X(\gamma_e)$ with an invariant flat connection $\nabla^{F_{\sigma,\lambda}, f}$ as explained in Section 4. Let $\boldsymbol{D}^{X(\gamma_e), F_{\sigma,\lambda}, 2}$ denote the associated de Rham–Hodge Laplacian acting on $\Omega^\bullet(X(\gamma_e), F_{\sigma,\lambda})$.

We also view $\gamma_h = e^a$ as a hyperbolic element in $\widetilde{Z}(\gamma_e)$. For $\sigma \in W^1_U(\gamma_e)$, as in (6.0.1), we set

$$\mathcal{E}_{X(\gamma_e),\gamma_h}(F_{\sigma,\lambda}, t) = \operatorname{Tr}_s^{[\gamma_h]}\left[\left(N^{\Lambda^\bullet(T^*X(\gamma_e))} - \frac{p'}{2}\right) \exp\left(-\frac{t \boldsymbol{D}^{X(\gamma_e), F_{\sigma,\lambda}, 2}}{2}\right)\right]. \tag{6.0.9}$$

Note that we use $B|_{\mathfrak{z}(\gamma_e)}$ on $\mathfrak{z}(\gamma_e)$ to define this orbital integral for $\widetilde{Z}(\gamma_e)$.

Set

$$c(\gamma)=\left|\frac{\det(1-\mathrm{Ad}(\gamma_e))|_{\mathfrak{z}^\perp(\gamma_e)}}{\det(1-\mathrm{Ad}(\gamma))|_{\mathfrak{z}^\perp(\gamma_e)}}\right|^{\frac{1}{2}}>0. \tag{6.0.10}$$

In particular, $c(\gamma_e)=1$.

The following theorem is essentially a consequence of the generalized Kirillov formula in Theorem 5.4.4.

Theorem 6.0.1. *Let $\gamma\in G$ be given as in (6.0.2). For $t>0$, we have the identity*

$$\mathcal{E}_{X,\gamma}(F_\lambda,t)=c(\gamma)\sum_{\sigma\in W_U^1(\gamma_e)}\varphi^U_{\gamma_e}(\sigma,\lambda)\mathcal{E}_{X(\gamma_e),\gamma_h}(F_{\sigma,\lambda},t). \tag{6.0.11}$$

We call (6.0.11) a localization formula for the geometric orbital integral.

Proof. Set $p'=\dim\mathfrak{p}(\gamma_e)=\dim X(\gamma_e)$. At first, if m is even, then p' is even. Then the both sides of (6.0.11) are 0 by [Bismut 2011, Theorem 7.9.1].

If m is odd, then p' is odd, and $\delta(G)=\delta(Z(\gamma_e)^0)$ is odd. If $\delta(G)\geq 3$, then the both sides of (6.0.11) are 0 by [Bismut 2011, Theorem 7.9.1].

Now we consider the case where $\delta(G)=\delta(Z(\gamma_e)^0)=1$. If γ cannot be conjugated into H by an element in G, then γ_h cannot be conjugated into H by an element in $Z(\gamma_e)^0$. Then both sides of (6.0.11) are 0 by Proposition 4.1.5.

Now we assume that $\delta(G)=1$ and $a\in\mathfrak{b}$. Note that $\mathfrak{z}(\gamma)$ is the centralizer of γ_h in $\mathfrak{z}(\gamma_e)$. We will prove (6.0.11) using (4.1.16)

For $y\in\mathfrak{k}(\gamma)$, let $J^{\sim}_{\gamma_h}(y)$ be the function defined in 3.3.1 for $\gamma_h=e^a\in\widetilde{Z}(\gamma_e)$:

$$J^{\sim}_{\gamma_h}(y)=\frac{1}{|\det(1-\mathrm{Ad}(\gamma_h))|_{\mathfrak{z}_0^\perp\cap\mathfrak{z}(\gamma_e)}|^{\frac{1}{2}}}\frac{\widehat{A}(i\,\mathrm{ad}(y)|_{\mathfrak{p}(\gamma)})}{\widehat{A}(i\,\mathrm{ad}(y)|_{\mathfrak{k}(\gamma)})}. \tag{6.0.12}$$

The Casimir operator $C^{\mathfrak{u}(\gamma_e),E_{\sigma,\lambda}}$ acts on $E_{\sigma,\lambda}$ by the scalar given by

$$-4\pi^2(|\lambda+\rho_{\mathfrak{u}}|^2-|\rho_{\mathfrak{u}(\gamma_e)}|^2). \tag{6.0.13}$$

Similar to (4.1.13), set

$$\beta_{\mathfrak{z}(\gamma_e)}=\tfrac{1}{16}\,\mathrm{Tr}^{\mathfrak{p}(\gamma_e)}[C^{\mathfrak{k}(\gamma_e),\mathfrak{p}(\gamma_e)}]+\tfrac{1}{48}\,\mathrm{Tr}^{\mathfrak{k}(\gamma_e)}[C^{\mathfrak{k}(\gamma_e),\mathfrak{k}(\gamma_e)}]. \tag{6.0.14}$$

Then by [Bismut 2011, Propositions 2.6.1, 7.5.1],

$$2\pi^2|\rho_{\mathfrak{u}(\gamma_e)}|^2=-\beta_{\mathfrak{z}(\gamma_e)}. \tag{6.0.15}$$

By (4.1.16), (6.0.13), (6.0.15), for $\sigma\in W_U^1(\gamma_e)$, we get

$$\begin{aligned}\mathcal{E}_{X(\gamma_e),\gamma_h}(F_{\sigma,\lambda},t)&=\frac{e^{-\frac{|a|^2}{2t}}}{(2\pi t)^{\frac{p}{2}}}\exp(-2\pi^2t|\lambda+\rho_{\mathfrak{u}}|^2)\\&\quad\cdot\int_{\mathfrak{k}(\gamma)}J^{\sim}_{\gamma_h}(y)\,\mathrm{Tr}_s^{\Lambda^\bullet(\mathfrak{p}(\gamma_e)^*)}\Big[\Big(N^{\Lambda^\bullet(\mathfrak{p}(\gamma_e)^*)}-\frac{p'}{2}\Big)\exp(-i\,\mathrm{ad}(y))\Big]\\&\qquad\cdot\mathrm{Tr}^{E_{\sigma,\lambda}}[\exp(-i\rho^{E_{\sigma,\lambda}}(y))]e^{-\frac{|y|^2}{2t}}\frac{dy}{(2\pi t)^{\frac{q}{2}}}.\end{aligned} \tag{6.0.16}$$

Note that $\dim \mathfrak{p}^{\perp}(\gamma_e)$ is even. We claim that if $y \in \mathfrak{k}(\gamma)$, then

$$\operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{p}^*)}\left[\left(N^{\Lambda^\bullet(\mathfrak{p}^*)} - \frac{m}{2}\right)\exp(-i\operatorname{ad}(y))\operatorname{Ad}(k^{-1})\right] = \operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{p}(\gamma_e)^*)}\left[\left(N^{\Lambda^\bullet(\mathfrak{p}(\gamma_e)^*)} - \frac{p'}{2}\right)e^{-i\operatorname{ad}(y)}\right]\det(1-e^{-i\operatorname{ad}(y)}\operatorname{Ad}(k^{-1}))|_{\mathfrak{p}^{\perp}(\gamma_e)}. \tag{6.0.17}$$

Indeed, we can verify (6.0.17) for $y \in \mathfrak{t}$. Since both sides of (6.0.17) are invariant by the adjoint action of $K(\gamma_e)^0$, (6.0.17) holds in full generality.

Also $K(\gamma)^0$ preserves the splitting

$$\mathfrak{p}^{\perp}(\gamma_e) = \mathfrak{p}_0^{\perp}(\gamma) \oplus (\mathfrak{p}^{\perp}(\gamma_e) \cap \mathfrak{p}_0^{\perp}). \tag{6.0.18}$$

The action $\operatorname{ad}(a)$ gives an isomorphism between $\mathfrak{p}^{\perp}(\gamma_e) \cap \mathfrak{p}_0^{\perp}$ and $\mathfrak{k}^{\perp}(\gamma_e) \cap \mathfrak{k}_0^{\perp}$ as $K(\gamma)$-vector spaces.

Note that

$$\mathfrak{z}^{\perp}(\gamma_e) \cap \mathfrak{z}_0^{\perp} = (\mathfrak{p}^{\perp}(\gamma_e) \cap \mathfrak{p}_0^{\perp}) \oplus (\mathfrak{k}^{\perp}(\gamma_e) \cap \mathfrak{k}_0^{\perp}). \tag{6.0.19}$$

Then

$$\det(1-e^{-i\operatorname{ad}(y)}\operatorname{Ad}(\gamma_e))|_{\mathfrak{p}^{\perp}(\gamma_e)} = \det(1-e^{-i\operatorname{ad}(y)}\operatorname{Ad}(\gamma_e))|_{\mathfrak{p}_0^{\perp}(\gamma_e)}[\det(1-e^{-i\operatorname{ad}(y)}\operatorname{Ad}(\gamma_e))|_{\mathfrak{z}^{\perp}(\gamma_e)\cap\mathfrak{z}_0^{\perp}}]^{\frac{1}{2}}. \tag{6.0.20}$$

Here the square root is taken to be positive at $y = 0$.

By Definition 3.3.1 and (6.0.12), for $y \in \mathfrak{k}(\gamma)$,

$$J_\gamma(y) = J_{\gamma_h}^{\sim}(y)\frac{1}{|\det(1-\operatorname{Ad}(\gamma))|_{\mathfrak{z}_0^{\perp}\cap\mathfrak{z}^{\perp}(\gamma_e)}|^{\frac{1}{2}}} \cdot\left[\frac{1}{\det(1-\operatorname{Ad}(\gamma_e))|_{\mathfrak{z}_0^{\perp}(\gamma)}}\frac{\det(1-\exp(-i\operatorname{ad}(y))\operatorname{Ad}(\gamma_e))|_{\mathfrak{k}_0^{\perp}(\gamma)}}{\det(1-\exp(-i\operatorname{ad}(y))\operatorname{Ad}(\gamma_e))|_{\mathfrak{p}_0^{\perp}(\gamma)}}\right]^{\frac{1}{2}}. \tag{6.0.21}$$

Combining (6.0.17), (6.0.20) and (6.0.21), we get

$$J_\gamma(y)\operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{p}^*)}\left[\left(N^{\Lambda^\bullet(\mathfrak{p}^*)} - \frac{m}{2}\right)\exp(-i\operatorname{ad}(y))\operatorname{Ad}(\gamma_e)\right] = c(\gamma)J_{\gamma_h}^{\sim}(y)\operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{p}(\gamma_e)^*)} \cdot\left[\left(N^{\Lambda^\bullet(\mathfrak{p}(\gamma_e)^*)} - \frac{p'}{2}\right)e^{-i\operatorname{ad}(y)}\right]\left[\frac{\det(1-\exp(-i\operatorname{ad}(y))\operatorname{Ad}(\gamma_e))|_{\mathfrak{z}^{\perp}(\gamma_e)}}{\det(1-\operatorname{Ad}(\gamma_e))|_{\mathfrak{z}^{\perp}(\gamma_e)}}\right]^{\frac{1}{2}}. \tag{6.0.22}$$

Note that, for $y \in \mathfrak{k}(\gamma)$,

$$\left[\frac{\det(1-\exp(-i\operatorname{ad}(y))\operatorname{Ad}(\gamma_e))|_{\mathfrak{z}^{\perp}(\gamma_e)}}{\det(1-\operatorname{Ad}(\gamma_e))|_{\mathfrak{z}^{\perp}(\gamma_e)}}\right]^{\frac{1}{2}} = \left[\frac{\det(1-\exp(-i\operatorname{ad}(y))\operatorname{Ad}(\gamma_e))|_{\mathfrak{u}^{\perp}(\gamma_e)}}{\det(1-\operatorname{Ad}(\gamma_e))|_{\mathfrak{u}^{\perp}(\gamma_e)}}\right]^{\frac{1}{2}}. \tag{6.0.23}$$

By (4.1.16), (6.0.13), (6.0.15), (6.0.22) and (6.0.23), we get

$$\begin{aligned}\mathcal{E}_{X,\lambda}(F_\lambda,t) = {} & c(\gamma)\frac{e^{-\frac{|a|^2}{2t}}}{(2\pi t)^{\frac{p}{2}}}\exp(-2\pi^2 t|\lambda+\rho_{\mathfrak{u}}|^2)\\ & \cdot\int_{\mathfrak{k}(\gamma)} J^{\sim}_{\gamma_h}(y)\,\mathrm{Tr}_s^{\Lambda^\bullet(\mathfrak{p}(\gamma_e)^*)}\Big[\Big(N^{\Lambda^\bullet(\mathfrak{p}(\gamma_e)^*)}-\frac{p'}{2}\Big)e^{-i\,\mathrm{ad}(y)}\Big]\\ & \cdot\Big[\frac{\det(1-e^{-i\,\mathrm{ad}(y)}\,\mathrm{Ad}(\gamma_e))|_{\mathfrak{u}^\perp(\gamma_e)}}{\det(1-\mathrm{Ad}(\gamma_e))|_{\mathfrak{u}^\perp(\gamma_e)}}\Big]^{\frac{1}{2}}\mathrm{Tr}^{E_\lambda}[\rho^{E_\lambda}(\gamma_e)e^{-i\rho^{E_\lambda}(y)}]e^{-\frac{|y|^2}{2t}}\frac{dy}{(2\pi t)^{\frac{q}{2}}}.\end{aligned}\tag{6.0.24}$$

Then (6.0.11) follows from (5.4.32), (6.0.16) and (6.0.24). □

Remark 6.0.2. A similar consideration can be made for $\mathrm{Tr}_s^{[\gamma]}[\exp(-t\boldsymbol{D}^{X,F_\lambda,2})]$, where (6.0.11) will become an analogue of the index theorem for orbifolds as in (2.2.9). The related computation can be found in [Bismut and Shen 2022, Section 10.4].

7. Full asymptotics of elliptic orbital integrals

In this section, we always assume that $\delta(G)=1$ and that U is compact. We also use the notation and settings as in Sections 5.1, 5.2 and 5.3.

In this section, given a irreducible unitary representation E of U with certain nondegenerate highest weight Λ, and for elliptic γ, we will compute explicitly $\mathcal{E}_{X,\gamma}(F=G\times_K E,t)$ and its Mellin transform in terms of the root systems. Note that, when $\gamma=1$, $\mathcal{E}_{X,\gamma}(F_d,t)$ is already computed in [Bergeron and Venkatesh 2013; Müller and Pfaff 2013a] using the Plancherel formula for identity orbital integral. We here give a different approach via Bismut's formula as in (4.1.16).

Then in Section 7.3, we apply these results to a sequence of flat vector bundles $\{F_d\}_{d\in\mathbb{N}}$ on X defined by a sequence of nondegenerate dominant weights $\Lambda = d\lambda+\lambda_0$. This way, we show that the Mellin transforms of the elliptic orbital integrals are exponential polynomials in d.

7.1. ***Estimates of elliptic orbital integrals for small time t.*** Recall that T is a maximal torus of K, T_U is a maximal torus of U, and $W(U,T_U)$ denotes the (analytic) Weyl group of (U,T_U). The positive root system $R^+(\mathfrak{u},\mathfrak{t}_U)$ is given in Section 5.3. Recall that $P_{++}(U)$ is the set of dominant weights of (U,T_U) with respect to $R^+(\mathfrak{u},\mathfrak{t}_U)$.

Let (E,ρ^E) be the irreducible unitary representation of U associated with the highest weight $\Lambda\in P_{++}(U)$. We will prove our main result of this subsection and next subsection for this (E,ρ^E).

Our homogeneous flat vector bundle concerned here is given by $F=G\times_K E$. Let $\boldsymbol{D}^{X,F,2}$ denote the associated de Rham–Hodge Laplacian.

For $t>0$, if $\gamma\in G$ is semisimple, as in (6.0.1), set

$$\mathcal{E}_{X,\gamma}(F,t)=\mathrm{Tr}_s^{[\gamma]}\Big[\Big(N^{\Lambda^\bullet(T^*X)}-\frac{m}{2}\Big)\exp\Big(-\frac{t\boldsymbol{D}^{X,F,2}}{2}\Big)\Big].\tag{7.1.1}$$

It is clear that $\mathcal{E}_{X,\gamma}(F_d,t)$ only depends on the conjugacy class $[\gamma]$ in G. If $\gamma=1$, we also write

$$\mathcal{I}_X(F,t)=\mathcal{E}_{X,1}(F,t).\tag{7.1.2}$$

In the sequel, we only consider the case of elliptic γ.

By (4.1.16), (6.0.13), (6.0.15), if $\gamma = k \in K$, we have

$$\mathcal{E}_{X,\gamma}(F,t) = \frac{1}{(2\pi t)^{\frac{p}{2}}} \exp(-2\pi^2 t|\Lambda + \rho_{\mathfrak{u}}|^2) \cdot \int_{\mathfrak{k}(\gamma)} J_\gamma(Y_0^{\mathfrak{k}}) \operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{p}^*)}\Big[\Big(N^{\Lambda^\bullet(\mathfrak{p}^*)} - \frac{m}{2}\Big) \operatorname{Ad}(k) \exp(-i \operatorname{ad}(Y_0^{\mathfrak{k}}))\Big] \cdot \operatorname{Tr}^E[\rho^E(k)\exp(-i\rho^E(Y_0^{\mathfrak{k}}))] e^{-\frac{|Y_0^{\mathfrak{k}}|^2}{2t}} \frac{dY_0^{\mathfrak{k}}}{(2\pi t)^{\frac{q}{2}}}. \tag{7.1.3}$$

By (3.3.18), we have the following formula for $J_\gamma(Y_0^{\mathfrak{k}})$, $Y_0^{\mathfrak{k}} \in \mathfrak{k}(\gamma)$:

$$J_\gamma(Y_0^{\mathfrak{k}}) = \frac{\widehat{A}(i \operatorname{ad}(Y_0^{\mathfrak{k}})|_{\mathfrak{p}(\gamma)})}{\widehat{A}(i \operatorname{ad}(Y_0^{\mathfrak{k}})|_{\mathfrak{k}(\gamma)})} \left[\frac{1}{\det(1-\operatorname{Ad}(k))|_{\mathfrak{z}^\perp(\gamma)}} \frac{\det(1-\exp(-i \operatorname{ad}(Y_0^{\mathfrak{k}}))\operatorname{Ad}(k))|_{\mathfrak{k}^\perp(\gamma)}}{\det(1-\exp(-i \operatorname{ad}(Y_0^{\mathfrak{k}}))\operatorname{Ad}(k))|_{\mathfrak{p}^\perp(\gamma)}}\right]^{\frac{1}{2}}. \tag{7.1.4}$$

Proposition 7.1.1. *For an elliptic element $\gamma \in G$, there exists a constant $C_\gamma > 0$ (depending on Λ) such that for $t \in]0,1]$*

$$|\sqrt{t}\mathcal{E}_{X,\gamma}(F,t)| \le C_\gamma, \qquad \Big|\Big(1+2t\frac{\partial}{\partial t}\Big)\mathcal{E}_{X,\gamma}(F,t)\Big| \le C_\gamma\sqrt{t}. \tag{7.1.5}$$

As $t \to 0$, $\mathcal{E}_{X,\gamma}(E,t)$ has the asymptotic expansion in the form of

$$\frac{1}{\sqrt{t}} \sum_{j=0}^{+\infty} a_j^\gamma t^j, \tag{7.1.6}$$

with $a_j^\gamma \in \mathbb{C}$.

Proof. If γ is elliptic, up to a conjugation, we assume that $\gamma = k \in T$. Thus the subgroup H defined in Section 4.1 is also a Cartan subgroup of $Z(\gamma)^0$. Then $\mathfrak{b}(\gamma) = \mathfrak{b}$. Let $\mathfrak{b}^\perp(\gamma)$ be the orthogonal complementary space of $\mathfrak{b}(\gamma)$ in $\mathfrak{p}(\gamma)$, whose dimension is $p-1$. Note that similar estimates have been proved in [Liu 2021, Theorem 4.4.1]; here we only sketch a proof to (7.1.5).

By (7.1.3), we have

$$\mathcal{E}_{X,\gamma}(F,t) = \frac{1}{(2\pi t)^{\frac{p}{2}}} \exp(-2\pi^2 t|\Lambda + \rho_{\mathfrak{u}}|^2) \cdot \int_{\mathfrak{k}(k)} J_k(\sqrt{t}Y_0^{\mathfrak{k}}) \operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{p}^*)}\Big[\Big(N^{\Lambda^\bullet(\mathfrak{p}^*)} - \frac{m}{2}\Big) \operatorname{Ad}(k) \exp(-i \operatorname{ad}(\sqrt{t}Y_0^{\mathfrak{k}}))\Big] \cdot \operatorname{Tr}^E[\rho^E(k)\exp(-i\rho^E(\sqrt{t}Y_0^{\mathfrak{k}}))] e^{-\frac{|Y_0^{\mathfrak{k}}|^2}{2}} \frac{dY_0^{\mathfrak{k}}}{(2\pi)^{\frac{q}{2}}}, \tag{7.1.7}$$

where the integral is rescaled by $\sqrt{t}$.

In this proof, we denote by C or c a positive constant independent of the variables t and $Y_0^{\mathfrak{k}}$. We use the symbol $\mathcal{O}_{\text{ind}}$ to denote the big-O convention which does not depend on t and $Y_0^{\mathfrak{k}}$.

The same computations as in [Liu 2021, equations (4.4.8)–(4.4.10)] show that, for $Y_0^{\mathfrak{k}} \in \mathfrak{t}$,

$$\begin{aligned} J_k(\sqrt{t}Y_0^{\mathfrak{k}}) &= \frac{1}{\det(1-\operatorname{Ad}(k))|_{\mathfrak{p}^\perp(k)}} + \mathcal{O}_{\text{ind}}(\sqrt{t}|Y_0^{\mathfrak{k}}|e^{C\sqrt{t}|Y_0^{\mathfrak{k}}|}) \\ &\quad \cdot \frac{1}{t^{(p-1)/2}} \operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{p}^*)}\Big[\Big(N^{\Lambda^\bullet(\mathfrak{p}^*)} - \frac{m}{2}\Big)\rho^{\Lambda^\bullet(\mathfrak{p}^*)}(k)\exp(-i\rho^{\Lambda^\bullet(\mathfrak{p}^*)}(\sqrt{t}Y_0^{\mathfrak{k}}))\Big] \\ &= -\det(i \operatorname{ad}(Y_0^{\mathfrak{k}}))|_{\mathfrak{b}^\perp(k)} \det(1-\operatorname{Ad}(k))|_{\mathfrak{p}^\perp(k)} + \mathcal{O}_{\text{ind}}(\sqrt{t}|Y_0^{\mathfrak{k}}|e^{C\sqrt{t}|Y_0^{\mathfrak{k}}|}). \end{aligned} \tag{7.1.8}$$

Using the adjoint invariance, the further estimates on the above quantities by a function in $|Y_0^{\mathfrak{k}}|$ hold for all $Y_0^{\mathfrak{k}} \in \mathfrak{k}(k)$.

It is clear that

$$|\mathrm{Tr}^E[\rho^E(k)\exp(-i\rho^E(\sqrt{t}Y_0^{\mathfrak{k}}))]| \leq C\exp(C\sqrt{t}|Y_0^{\mathfrak{k}}|). \tag{7.1.9}$$

Combining (7.1.8) and (7.1.9), we see that there exists a number $N \in \mathbb{N}$ big enough such that if $t \in]0,1]$,

$$|\sqrt{t}\mathcal{E}_{X,\gamma}(F,t)| \leq C_k' \int_{\mathfrak{k}(k)} (1+|Y_0^{\mathfrak{k}}|)^N \exp\left(C|Y_0^{\mathfrak{k}}| - \frac{|Y_0^{\mathfrak{k}}|^2}{2}\right) dY_0^{\mathfrak{k}}. \tag{7.1.10}$$

The second estimate in (7.1.5) can be proved using the same arguments as in [Liu 2021, equations (4.4.24)–(4.4.29)].

The asymptotic expansion in (7.1.6) is just a consequence of (7.1.5) and (7.1.7). □

7.2. ***Elliptic orbital integrals for Hodge Laplacians.*** In this subsection, we explain how to use Bismut's formula (4.1.16) to compute explicitly the expansion of $\mathcal{E}_{X,\gamma}(F,t)$ in $t>0$ when $\gamma \in G$ is elliptic. Then we study the corresponding Mellin transform. After conjugation, we may and we will assume that $\gamma = k \in T$. Then T is also a maximal torus for $K(\gamma)^0$, and $\mathfrak{b}(\gamma) = \mathfrak{b}$.

Recall that $\omega^{Y_{\mathfrak{b}}(\gamma)}$, $\Omega^{\mathfrak{u}_{\mathfrak{b}}(\gamma)}$, $\Omega^{\mathfrak{u}_{\mathfrak{m}}(\gamma)}$ are defined in Section 5.2. Note that $\dim \mathfrak{u}_{\mathfrak{b}}^{\perp}(\gamma) = 4l(\gamma)$. If $\nu \in \Lambda^{\bullet}(\mathfrak{u}_{\mathfrak{b}}^{\perp}(\gamma)^*)$, let $[\nu]^{\max(\gamma)} \in \mathbb{R}$ be such that

$$\nu - [\nu]^{\max(\gamma)} \frac{\omega^{Y_{\mathfrak{b}}(\gamma),2l(\gamma)}}{(2l(\gamma))!} \tag{7.2.1}$$

is of degree strictly smaller than $4l(\gamma)$.

Recall that $-B(\cdot,\theta\cdot)$ is a Euclidean product on $\mathfrak{g}$. Let $\mathfrak{n}^{\perp}(\gamma)$, $\bar{\mathfrak{n}}^{\perp}(\gamma)$ be the orthogonal spaces of $\mathfrak{n}(\gamma)$, $\bar{\mathfrak{n}}(\gamma)$ in $\mathfrak{n}$, $\bar{\mathfrak{n}}$ respectively. As T-modules, $\mathfrak{n}^{\perp}(\gamma) \simeq \bar{\mathfrak{n}}^{\perp}(\gamma)$.

Since $\mathfrak{t} \subset \mathfrak{k}(\gamma) \subset \mathfrak{k}$, $R(\mathfrak{k}(\gamma),\mathfrak{t})$ is a subroot system of $R(\mathfrak{k},\mathfrak{t})$. Let $R^+(\mathfrak{k}(\gamma),\mathfrak{t})$ be the positive root system for $(\mathfrak{k}(\gamma),\mathfrak{t})$ induced by $R^+(\mathfrak{k},\mathfrak{t})$. We use the notation in Sections 5.1, 5.2. Then $\mathfrak{t}$ is a Cartan subalgebra for $\mathfrak{k}_{\mathfrak{m}}(\gamma)$, $\mathfrak{u}_{\mathfrak{m}}(\gamma)$, $\mathfrak{m}(\gamma)$. Let $R(\mathfrak{k}_{\mathfrak{m}}(\gamma),\mathfrak{t})$, $R(\mathfrak{u}_{\mathfrak{m}}(\gamma),\mathfrak{t})$ be the corresponding root systems.

Similar to (5.4.10), we have the disjoint union

$$R(\mathfrak{u}_{\mathfrak{m}}(\gamma),\mathfrak{t}) = R(\sqrt{-1}\mathfrak{p}_{\mathfrak{m}}(\gamma),\mathfrak{t}) \cup R(\mathfrak{k}_{\mathfrak{m}}(\gamma),\mathfrak{t}). \tag{7.2.2}$$

Since $R(\mathfrak{u}_{\mathfrak{m}}(\gamma),\mathfrak{t}) \subset R(\mathfrak{u}_{\mathfrak{m}},\mathfrak{t})$, by intersecting with $R^+(\mathfrak{u}_{\mathfrak{m}},\mathfrak{t})$, we get a positive root system $R^+(\mathfrak{u}_{\mathfrak{m}}(\gamma),\mathfrak{t})$. Moreover,

$$R^+(\mathfrak{u}_{\mathfrak{m}}(\gamma),\mathfrak{t}) = R^+(\sqrt{-1}\mathfrak{p}_{\mathfrak{m}}(\gamma),\mathfrak{t}) \cup R^+(\mathfrak{k}_{\mathfrak{m}}(\gamma),\mathfrak{t}). \tag{7.2.3}$$

Let $\mathrm{Vol}(K/T)$, $\mathrm{Vol}(U_M/T)$ be the Riemannian volumes of K/T, U_M/T with respect to the restriction of $-B$ to $\mathfrak{k}$, $\mathfrak{u}_{\mathfrak{m}}$ respectively. We have explicit formulae for them in terms of the roots; for example,

$$\mathrm{Vol}(U_M,T) = \Pi_{\alpha^0 \in R^+(\mathfrak{u}_{\mathfrak{m}},\mathfrak{t})} \frac{1}{2\pi\langle\alpha^0,\rho_{\mathfrak{u}_{\mathfrak{m}}}\rangle}. \tag{7.2.4}$$

For $\gamma = k \in T$, set

$$c_G(\gamma) = \frac{(-1)^{\frac{p-1}{2}+1}\,\mathrm{Vol}(K(\gamma)^0/T)|W(U_M(\gamma)^0,T)|}{\mathrm{Vol}(U_M(\gamma)^0/T)|W(K(\gamma)^0,T)|}\frac{1}{\det(1-\mathrm{Ad}(\gamma))|_{\mathfrak{n}^\perp(\gamma)}}. \tag{7.2.5}$$

If $\gamma = 1$, we define

$$c_G = c_G(1) = \frac{(-1)^{\frac{m-1}{2}+1}\mathrm{Vol}(K/T)|W(U_M,T)|}{\mathrm{Vol}(U_M/T)|W(K,T)|}. \tag{7.2.6}$$

We will use the same notation as in Sections 5.3 and 5.4. In particular, W_u is defined by (5.3.4) as a subset of $W(U,T_U)$, and $W^1(\gamma)$ is defined by (5.4.14) as a subset of $W(U_M,T)$. As explained in Remark 5.4.5, for $\omega \in W_u$, $\sigma \in W^1(\gamma)$, let $E^\gamma_{\omega,\sigma}$ denote the irreducible unitary representation of $Y = U_M(\gamma)^0$ or its finite central extension with highest weight $\sigma(\eta_\omega(\Lambda)+\rho_{\mathfrak{u}_\mathfrak{m}}) - \rho_{\mathfrak{y}}$.

Definition 7.2.1. For $j = 0, 1, \ldots, l(\gamma)$, $\omega \in W_u$, $\sigma \in W^1(\gamma)$, set

$$Q^\gamma_{j,\omega,\sigma}(\Lambda) = \frac{(-1)^j \beta(a_1)^{2j}}{j!\,(2l(\gamma)-2j)!\,(8\pi^2)^j}\dim E^\gamma_{\omega,\sigma}[\omega^{Y_{\mathfrak{b}}(\gamma),2j}\langle\omega(\Lambda+\rho_{\mathfrak{u}}),\Omega^{\mathfrak{u}_\mathfrak{m}(\gamma)}\rangle^{2l(\gamma)-2j}]^{\max(\gamma)}. \tag{7.2.7}$$

In particular, if $l(\gamma) \geq 1$, we have

$$\begin{aligned} Q^\gamma_{0,\omega,\sigma}(\Lambda) &= \frac{1}{(2l)!}\dim E^\gamma_{\omega,\sigma}[\langle\omega(\Lambda+\rho_{\mathfrak{u}}),\Omega^{\mathfrak{u}_\mathfrak{m}(\gamma)}\rangle^{2l(\gamma)}]^{\max(\gamma)},\\ Q^\gamma_{l(\gamma),\omega,\sigma}(\Lambda) &= \frac{(-1)^{l(\gamma)}\beta(a_1)^{2l(\gamma)}(2l(\gamma)-1)!!}{(4\pi^2)^{l(\gamma)}}\dim E^\gamma_{\omega,\sigma}. \end{aligned} \tag{7.2.8}$$

Recall that $a_1 \in \mathfrak{b}$ is such that $B(a_1,a_1) = 1$. For $\omega \in W_u$, set

$$b_{\Lambda,\omega} = \langle \omega\cdot(\Lambda+\rho_{\mathfrak{u}}), \sqrt{-1}a_1\rangle \in \mathbb{R}. \tag{7.2.9}$$

Then we have

$$|\eta_\omega(\Lambda)+\rho_{\mathfrak{u}_\mathfrak{m}}|^2 - |\Lambda+\rho_{\mathfrak{u}}|^2 = -b^2_{\Lambda,\omega}. \tag{7.2.10}$$

Note that $\varphi_\gamma(\sigma,\eta_\omega(\Lambda))$ is defined in Definition 5.4.3.

Theorem 7.2.2. *For $t > 0$, we have the identity*

$$\mathcal{E}_{X,\gamma}(F,t) = \frac{c_G(\gamma)}{\sqrt{2\pi t}}\sum_{j=0}^{l(\gamma)} t^{-j}\sum_{\substack{\omega\in W_u\\ \sigma\in W^1(\gamma)}}\varepsilon(\omega)\varphi_\gamma(\sigma,\eta_\omega(\Lambda))e^{-2\pi^2 t b^2_{\Lambda,\omega}}Q^\gamma_{j,\omega,\sigma}(\Lambda). \tag{7.2.11}$$

Remark 7.2.3. The formula (7.2.11) is compatible with the estimate (7.1.5). For example, we take $\gamma = 1$; then $W^1(\gamma)$ reduces to $\{1\}$, the representation $E^\gamma_{\omega,\sigma}$ is just $V_{\Lambda,\omega}$ introduced in (5.3.6), and $l(\gamma) = l$, $\varphi_\gamma(\sigma,\eta_\omega(\Lambda)) = 1$. Then we take the asymptotic expansion of the right-hand side of (7.4.2) as $t \to 0$, the coefficient of $t^{-l-1/2}$ is given by

$$\frac{c_G}{\sqrt{2\pi}}\sum_{\omega\in W_u}\varepsilon(\omega)Q^{\gamma=1}_{l,\omega,1}(\Lambda). \tag{7.2.12}$$

By (5.3.7), if $l \geq 1$, we get

$$\sum_{\omega \in W_u} \varepsilon(\omega) \dim V_{\Lambda,\omega} = \mathrm{Tr}_s^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)}[1] \dim E = 0. \tag{7.2.13}$$

Then by (7.2.8) and (7.2.13), the quantity in (7.2.12) is 0 (provided $l \geq 1$).

Before proving Theorem 7.2.2, we need some preparation work.

Definition 7.2.4. For $y \in \mathfrak{t}$, put

$$\begin{aligned}
\pi_{\mathfrak{u}_\mathfrak{m}(\gamma)/\mathfrak{t}}(y) &= \prod_{\alpha^0 \in R^+(\mathfrak{u}_\mathfrak{m}(\gamma),\mathfrak{t})} \langle 2\pi\sqrt{-1}\alpha^0, y\rangle, \\
\pi_{\sqrt{-1}\mathfrak{p}_\mathfrak{m}(\gamma)/\mathfrak{t}}(y) &= \prod_{\alpha^0 \in R^+(\sqrt{-1}\mathfrak{p}_\mathfrak{m}(\gamma),\mathfrak{t})} \langle 2\pi\sqrt{-1}\alpha^0, y\rangle, \\
\pi_{\mathfrak{k}_\mathfrak{m}(\gamma)/\mathfrak{t}}(y) &= \prod_{\alpha^0 \in R^+(\mathfrak{k}_\mathfrak{m}(\gamma),\mathfrak{t})} \langle 2\pi\sqrt{-1}\alpha^0, y\rangle.
\end{aligned} \tag{7.2.14}$$

For $y \in \mathfrak{t}$, put

$$\begin{aligned}
\sigma_{\mathfrak{u}_\mathfrak{m}(\gamma)/\mathfrak{t}}(y) &= \prod_{\alpha^0 \in R^+(\mathfrak{u}_\mathfrak{m}(\gamma),\mathfrak{t})} \left(\exp(\langle \pi\sqrt{-1}\alpha^0, y\rangle) - \exp(-\langle \pi\sqrt{-1}\alpha^0, y\rangle)\right), \\
\sigma_{\sqrt{-1}\mathfrak{p}_\mathfrak{m}(\gamma)/\mathfrak{t}}(y) &= \prod_{\alpha^0 \in R^+(\sqrt{-1}\mathfrak{p}_\mathfrak{m}(\gamma),\mathfrak{t})} \left(\exp(\langle \pi\sqrt{-1}\alpha^0, y\rangle) - \exp(-\langle \pi\sqrt{-1}\alpha^0, y\rangle)\right), \\
\sigma_{\mathfrak{k}_\mathfrak{m}(\gamma)/\mathfrak{t}}(y) &= \prod_{\alpha^0 \in R^+(\mathfrak{k}_\mathfrak{m}(\gamma)t,\mathfrak{t})} \left(\exp(\langle \pi\sqrt{-1}\alpha^0, y\rangle) - \exp(-\langle \pi\sqrt{-1}\alpha^0, y\rangle)\right).
\end{aligned} \tag{7.2.15}$$

We can always extend analytically the above functions to $y \in \mathfrak{t}_\mathbb{C}$. If $\gamma = 1$, the above functions become $\pi_{\mathfrak{u}_\mathfrak{m}/\mathfrak{t}}(y)$, $\pi_{\sqrt{-1}\mathfrak{p}_\mathfrak{m}/\mathfrak{t}}(y)$, $\pi_{\mathfrak{k}_\mathfrak{m}/\mathfrak{t}}(y)$, $\sigma_{\mathfrak{u}_\mathfrak{m}/\mathfrak{t}}(y)$, $\sigma_{\sqrt{-1}\mathfrak{p}_\mathfrak{m}/\mathfrak{t}}(y)$, $\sigma_{\mathfrak{k}_\mathfrak{m}/\mathfrak{t}}(y)$.

If the adjoint action of T preserves certain orthogonal splittings of $\mathfrak{u}_\mathfrak{m}$, $\mathfrak{u}_\mathfrak{m}(\gamma)$, etc., so that we have the corresponding splitting of the root systems, then we can also define the associated π-function or σ-function as above.

It is clear that if $y \in \mathfrak{t}_\mathbb{C}$,

$$\begin{aligned}
\pi_{\mathfrak{u}_\mathfrak{m}(\gamma)/\mathfrak{t}}(y) &= \pi_{\sqrt{-1}\mathfrak{p}_\mathfrak{m}(\gamma)/\mathfrak{t}}(y)\pi_{\mathfrak{k}_\mathfrak{m}(\gamma)/\mathfrak{t}}(y), \\
\sigma_{\mathfrak{u}_\mathfrak{m}(\gamma)/\mathfrak{t}}(y) &= \sigma_{\sqrt{-1}\mathfrak{p}_\mathfrak{m}(\gamma)/\mathfrak{t}}(y)\sigma_{\mathfrak{k}_\mathfrak{m}/\mathfrak{t}}(y).
\end{aligned} \tag{7.2.16}$$

Set

$$\begin{aligned}
\mathfrak{k}'_\mathfrak{m}(\gamma) &= \mathfrak{k}^\perp(\gamma) \cap \mathfrak{k}_\mathfrak{m}, \qquad & \mathfrak{p}'_\mathfrak{m}(\gamma) &= \mathfrak{p}^\perp(\gamma) \cap \mathfrak{p}_\mathfrak{m}, \\
\mathfrak{k}''_\mathfrak{m}(\gamma) &= \mathfrak{k}^\perp(\gamma) \cap \mathfrak{k}^\perp(\mathfrak{b}), & \mathfrak{p}''_\mathfrak{m}(\gamma) &= \mathfrak{p}^\perp(\gamma) \cap \mathfrak{p}^\perp(\mathfrak{b}).
\end{aligned} \tag{7.2.17}$$

Let $\mathfrak{m}^\perp(\gamma)$ be the orthogonal space of $\mathfrak{m}(\gamma)$ in $\mathfrak{m}$ with respect to B. Then

$$\mathfrak{m}^\perp(\gamma) = \mathfrak{p}'_\mathfrak{m}(\gamma) \oplus \mathfrak{k}'_\mathfrak{m}(\gamma). \tag{7.2.18}$$

We also have

$$\mathfrak{k}_\mathfrak{m} = \mathfrak{k}_\mathfrak{m}(\gamma) \oplus \mathfrak{k}'_\mathfrak{m}(\gamma), \quad \mathfrak{p}_\mathfrak{m} = \mathfrak{p}_\mathfrak{m}(\gamma) \oplus \mathfrak{p}'_\mathfrak{m}(\gamma) \tag{7.2.19}$$

and

$$\mathfrak{k}^{\perp}(\gamma)=\mathfrak{k}'_{\mathfrak{m}}(\gamma)\oplus\mathfrak{k}''_{\mathfrak{m}}(\gamma),\quad \mathfrak{p}^{\perp}(\gamma)=\mathfrak{p}'_{\mathfrak{m}}(\gamma)\oplus\mathfrak{p}''_{\mathfrak{m}}(\gamma). \tag{7.2.20}$$

Set

$$\mathfrak{u}^{\perp}_{\mathfrak{m}}(\gamma)=\sqrt{-1}\mathfrak{p}'_{\mathfrak{m}}(\gamma)\oplus\mathfrak{k}'_{\mathfrak{m}}(\gamma). \tag{7.2.21}$$

Then it is the orthogonal space of $\mathfrak{u}_{\mathfrak{m}}(\gamma)$ in $\mathfrak{u}_{\mathfrak{m}}$ with respect to B.

Lemma 7.2.5. *The following spaces are isomorphic to each other as modules of T by the adjoint actions*:

$$\mathfrak{n}^{\perp}(\gamma)\simeq\bar{\mathfrak{n}}^{\perp}(\gamma)\simeq\mathfrak{k}''_{\mathfrak{m}}(\gamma)\simeq\mathfrak{p}''_{\mathfrak{m}}(\gamma). \tag{7.2.22}$$

Proof. Note that

$$\dim\mathfrak{n}=\dim\mathfrak{k}-\dim\mathfrak{k}_{\mathfrak{m}},\quad \dim\mathfrak{n}(\gamma)=\dim\mathfrak{k}(\gamma)-\dim\mathfrak{k}_{\mathfrak{m}}(\gamma). \tag{7.2.23}$$

Together with the splittings (7.2.19), (7.2.20), we get

$$\dim\mathfrak{k}''_{\mathfrak{m}}(\gamma)=\dim\mathfrak{n}^{\perp}(\gamma). \tag{7.2.24}$$

Similarly, $\dim\mathfrak{p}''_{\mathfrak{m}}(\gamma)=\dim\mathfrak{n}^{\perp}(\gamma)$.

If $f\in\mathfrak{n}^{\perp}(\gamma)$, then $f+\theta(f)\in\mathfrak{k}$; we can verify directly that $f+\theta(f)\in\mathfrak{k}''_{\mathfrak{m}}(\gamma)$. Then the map $f\in\mathfrak{n}^{\perp}(\gamma)\mapsto f+\theta(f)\in\mathfrak{k}''_{\mathfrak{m}}(\gamma)$ defines an isomorphisms of T-modules. Similar for $\mathfrak{n}^{\perp}(\gamma)\simeq\mathfrak{p}''_{\mathfrak{m}}(\gamma)$. ☐

Since $\gamma=k\in T$, let $y_0\in\mathfrak{t}$ be such that $\exp(y_0)=\gamma$. Note that y_0 is not unique.

Lemma 7.2.6. *If $y\in\mathfrak{t}$ is regular with respect to $R(\mathfrak{k}_{\mathfrak{m}}(\gamma),\mathfrak{t})$, then we have*

$$\begin{aligned}J_{\gamma}(y)\operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{p}^*)}\Big[\Big(N^{\Lambda^\bullet(\mathfrak{p}^*)}-\frac{m}{2}\Big)\operatorname{Ad}(k)\exp(-\mathrm{i}\operatorname{ad}(y))\Big]\\ =\frac{(-1)^{\dim\mathfrak{p}_{\mathfrak{m}}(\gamma)/2+1}}{\det(1-\operatorname{Ad}(k))|_{\mathfrak{n}^{\perp}(\gamma)}}\operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{n}^*_{\mathbb{C}})}[e^{-i\operatorname{ad}(y)}\operatorname{Ad}(k)]\\ \cdot\frac{\pi_{\sqrt{-1}\mathfrak{p}_{\mathfrak{m}}(\gamma)/\mathfrak{t}}(iy)}{\pi_{\mathfrak{k}_{\mathfrak{m}}(\gamma)/\mathfrak{t}}(iy)}\frac{\sigma_{\mathfrak{u}_{\mathfrak{m}}(\gamma)/\mathfrak{t}}(iy)\sigma_{\mathfrak{u}^{\perp}_{\mathfrak{m}}(\gamma)/\mathfrak{t}}(-iy+y_0)}{\sigma_{\mathfrak{u}^{\perp}_{\mathfrak{m}}(\gamma)/\mathfrak{t}}(y_0)}.\end{aligned} \tag{7.2.25}$$

Proof. Using (5.4.23), (7.2.20) and Lemma 7.2.5, we get that, for $y\in\mathfrak{t}$,

$$\begin{aligned}\left[\frac{1}{\det(1-\operatorname{Ad}(k))|_{\mathfrak{z}^{\perp}(\gamma)}}\frac{\det(1-e^{-\mathrm{i}\operatorname{ad}(y)}\operatorname{Ad}(k))|_{\mathfrak{k}^{\perp}(\gamma)}}{\det(1-e^{-\mathrm{i}\operatorname{ad}(y)}\operatorname{Ad}(k))|_{\mathfrak{p}^{\perp}(\gamma)}}\right]^{\frac{1}{2}}\\ =\frac{(-1)^{\frac{1}{2}\dim\mathfrak{p}'_{\mathfrak{m}}(\gamma)}}{\det(1-\operatorname{Ad}(k))|_{\mathfrak{n}^{\perp}(\gamma)}}\frac{1}{\sigma_{\mathfrak{u}^{\perp}_{\mathfrak{m}}(\gamma)/\mathfrak{t}}(y_0)}\frac{\sigma_{\mathfrak{k}'_{\mathfrak{m}}(\gamma)/\mathfrak{t}}(-iy+y_0)}{\sigma_{\sqrt{-1}\mathfrak{p}'_{\mathfrak{m}}(\gamma)/\mathfrak{t}}(-iy+y_0)}.\end{aligned} \tag{7.2.26}$$

Recall that in Section 5.1, as K_M-modules, we have the isomorphism

$$\mathfrak{p}\simeq\mathfrak{b}\oplus\mathfrak{p}_{\mathfrak{m}}\oplus\mathfrak{n}. \tag{7.2.27}$$

Note that

$$\operatorname{Ad}(k)=e^{\operatorname{ad}(y_0)}. \tag{7.2.28}$$

If $y\in\mathfrak{t}$, when acting on $\mathfrak{p}$, we have

$$\operatorname{Ad}(k)\exp(-i\operatorname{ad}(y))=\exp(\operatorname{ad}(-iy+y_0)). \tag{7.2.29}$$

Note that $\dim \mathfrak{b} = 1$. Then, for $y \in \mathfrak{t}$, we get

$$\begin{aligned}\operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{p}^*)}\Big[\Big(N^{\Lambda^\bullet(\mathfrak{p}^*)} - \frac{m}{2}\Big)\operatorname{Ad}(k)\exp(-\mathrm{i}\operatorname{ad}(y))\Big] \\ &= -\operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{p}_\mathfrak{m}^*)}[\operatorname{Ad}(k)e^{-i\operatorname{ad}(y)}]\operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)}[\operatorname{Ad}(k)e^{-i\operatorname{ad}(y)}] \\ &= -\det(1-\operatorname{Ad}(k^{-1})e^{i\operatorname{ad}(y)})|_{\mathfrak{p}_\mathfrak{m}}\operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)}[\operatorname{Ad}(k)e^{-i\operatorname{ad}(y)}], \qquad (7.2.30)\end{aligned}$$

where we have the identity

$$\det(1-\operatorname{Ad}(k^{-1})e^{i\operatorname{ad}(y)})|_{\mathfrak{p}_\mathfrak{m}} = (-1)^{\frac{1}{2}\dim\mathfrak{p}_\mathfrak{m}}\sigma_{\sqrt{-1}\mathfrak{p}'_\mathfrak{m}(\gamma)/\mathfrak{t}}(-iy+y_0)^2\sigma_{\sqrt{-1}\mathfrak{p}_\mathfrak{m}(\gamma)/\mathfrak{t}}(iy)^2. \qquad (7.2.31)$$

Note that analogous to (7.2.27), we have $\mathfrak{p}(\gamma) \simeq \mathfrak{b} \oplus \mathfrak{p}_\mathfrak{m}(\gamma) \oplus \mathfrak{n}(\gamma)$; using [Bismut 2011, equation (7.5.24)], if $y \in \mathfrak{t}$, we have

$$\begin{aligned}\widehat{A}(i\operatorname{ad}(y)|_{\sqrt{-1}\mathfrak{p}(\gamma)}) &= \frac{\pi_{\sqrt{-1}\mathfrak{p}_\mathfrak{m}(\gamma)/\mathfrak{t}}(iy)}{\sigma_{\sqrt{-1}\mathfrak{p}_\mathfrak{m}(\gamma)/\mathfrak{t}}(iy)}\widehat{A}(i\operatorname{ad}(y)|_{\mathfrak{n}(\gamma)}), \\ \widehat{A}(i\operatorname{ad}(y)|_{\mathfrak{k}(\gamma)}) &= \frac{\pi_{\mathfrak{k}(\gamma)/\mathfrak{t}}(iy)}{\sigma_{\mathfrak{k}(\gamma)/\mathfrak{t}}(iy)} = \frac{\pi_{\mathfrak{k}_\mathfrak{m}(\gamma)/\mathfrak{t}}(iy)}{\sigma_{\mathfrak{k}_\mathfrak{m}(\gamma)/\mathfrak{t}}(iy)}\widehat{A}(i\operatorname{ad}(y)|_{\mathfrak{n}(\gamma)}).\end{aligned} \qquad (7.2.32)$$

Combining (7.1.4), (7.2.26) and (7.2.30)–(7.2.32), we get (7.2.25). □

Now we prove Theorem 7.2.2.

Proof of Theorem 7.2.2. Put

$$\begin{aligned}F_\gamma(\Lambda,t) = \frac{1}{(2\pi t)^{\frac{p}{2}}}\int_{\mathfrak{k}(\gamma)} J_\gamma(Y_0^\mathfrak{k})\operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{p}^*)}\Big[\Big(N^{\Lambda^\bullet(\mathfrak{p}^*)} - \frac{m}{2}\Big)\operatorname{Ad}(k)e^{-i\operatorname{ad}(Y_0^\mathfrak{k})}\Big] \\ \cdot\operatorname{Tr}^E[\rho^E(k)e^{-i\rho^E(Y_0^\mathfrak{k})}]e^{-\frac{|Y_0^\mathfrak{k}|^2}{2t}}\frac{dY_0^\mathfrak{k}}{(2\pi t)^{\frac{q}{2}}}. \qquad (7.2.33)\end{aligned}$$

By (7.1.3), we have

$$\mathcal{E}_{X,\gamma}(F,t) = \exp(-2\pi^2 t|\Lambda+\rho_\mathfrak{u}|^2)F_\gamma(\Lambda,t). \qquad (7.2.34)$$

Recall that $r = p+q = \dim_\mathbb{R}\mathfrak{z}(\gamma)$. By the Weyl integration formula,

$$\begin{aligned}F_\gamma(\Lambda,t) = \frac{\operatorname{Vol}(K(\gamma)^0/T)}{(2\pi t)^{\frac{r}{2}}|W(K(\gamma)^0,T)|}\int_\mathfrak{t}|\pi_{\mathfrak{k}(\gamma)/\mathfrak{t}}(y)|^2 J_\gamma(y)\operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{p}^*)}\Big[\Big(N^{\Lambda^\bullet(\mathfrak{p}^*)} - \frac{m}{2}\Big)\operatorname{Ad}(k)e^{-i\operatorname{ad}(y)}\Big] \\ \cdot\operatorname{Tr}^E[\rho^E(k)\exp(-i\rho^E(y))]e^{-\frac{|y|^2}{2t}}\,dy. \qquad (7.2.35)\end{aligned}$$

Recall that $l(\gamma) = \frac{1}{2}\dim\mathfrak{n}(\gamma)$. We can verify directly that if $y \in \mathfrak{t}$,

$$\pi_{\mathfrak{k}(\gamma)/\mathfrak{t}}(iy)^2 = (-1)^{l(\gamma)}\pi_{\mathfrak{k}_\mathfrak{m}(\gamma)/\mathfrak{t}}(iy)^2\det(i\operatorname{ad}(y))|_{\mathfrak{n}(\gamma)_\mathbb{C}}. \qquad (7.2.36)$$

Moreover, if $y \in \mathfrak{t}$ is such that $\pi_{\mathfrak{u}_\mathfrak{m}(\gamma)/\mathfrak{t}}(y) \neq 0$,

$$\frac{|\pi_{\mathfrak{k}(\gamma)/\mathfrak{t}}(y)|^2}{|\pi_{\mathfrak{u}_\mathfrak{m}(\gamma)/\mathfrak{t}}(y)|^2} = \frac{\pi_{\mathfrak{k}(\gamma)/\mathfrak{t}}(iy)^2}{\pi_{\mathfrak{u}_\mathfrak{m}(\gamma)/\mathfrak{t}}(iy)^2}. \qquad (7.2.37)$$

Then by Lemma 7.2.6 and (7.2.5), (7.2.32), (7.2.36), we get

$$\frac{|\pi_{\mathfrak{k}(\gamma)/\mathfrak{t}}(y)|^2}{|\pi_{\mathfrak{u}_\mathfrak{m}(\gamma)/\mathfrak{t}}(y)|^2} J_\gamma(y)\operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{p}^*)}\Big[\Big(N^{\Lambda^\bullet(\mathfrak{p}^*)}-\frac{m}{2}\Big)\operatorname{Ad}(k)\exp(-\mathrm{i}\,\operatorname{ad}(y))\Big]$$
$$=\frac{(-1)^{l(\gamma)+\frac{1}{2}\dim\mathfrak{p}_\mathfrak{m}(\gamma)+1}}{\det(1-\operatorname{Ad}(k))|_{\mathfrak{n}^\perp(\gamma)}}\operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)}[e^{-i\,\operatorname{ad}(y)}\operatorname{Ad}(k)]$$
$$\cdot\det(i\,\operatorname{ad}(y))|_{\mathfrak{n}(\gamma)_\mathbb{C}}\widehat{A}^{-1}(i\,\operatorname{ad}(y)|_{\mathfrak{u}_\mathfrak{m}(\gamma)})\left[\frac{\det(1-e^{-i\,\operatorname{ad}(y)}\operatorname{Ad}(k))|_{\mathfrak{u}_\mathfrak{m}^\perp(\gamma)}}{\det(1-\operatorname{Ad}(k))_{\mathfrak{u}_\mathfrak{m}^\perp(\gamma)}}\right]^{\frac{1}{2}}. \tag{7.2.38}$$

Note that we have the even number

$$p-1=\dim\mathfrak{p}_\mathfrak{m}(\gamma)+2l(\gamma). \tag{7.2.39}$$

Now we can rewrite (7.2.35) as

$$F_\gamma(\Lambda,t)=\frac{(-1)^{\frac{p-1}{2}+1}\operatorname{Vol}(K(\gamma)^0/T)}{(2\pi t)^{\frac{r}{2}}|W(K(\gamma)^0,T)|}\frac{1}{\det(1-\operatorname{Ad}(k))|_{\mathfrak{n}^\perp(\gamma)}}$$
$$\cdot\int_\mathfrak{t}|\pi_{\mathfrak{u}_\mathfrak{m}(\gamma)/\mathfrak{t}}(y)|^2\det(i\,\operatorname{ad}(y))|_{\mathfrak{n}(\gamma)_\mathbb{C}}\cdot\widehat{A}^{-1}(i\,\operatorname{ad}(y)|_{\mathfrak{u}_\mathfrak{m}(\gamma)})$$
$$\cdot\left[\frac{\det(1-e^{-i\,\operatorname{ad}(y)}\operatorname{Ad}(k))|_{\mathfrak{u}_\mathfrak{m}^\perp(\gamma)}}{\det(1-\operatorname{Ad}(k))_{\mathfrak{u}_\mathfrak{m}^\perp(\gamma)}}\right]^{\frac{1}{2}}$$
$$\cdot\operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)\otimes E}[e^{-i\rho^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)\otimes E}(y)}\rho^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)\otimes E}(k)]e^{-|y|^2/2t}\,dy. \tag{7.2.40}$$

Note that the function in $y\in\mathfrak{t}$

$$\det(i\,\operatorname{ad}(y))|_{\mathfrak{n}(\gamma)_\mathbb{C}}\cdot\widehat{A}^{-1}(i\,\operatorname{ad}(y)|_{\mathfrak{u}_\mathfrak{m}(\gamma)})\left[\frac{\det(1-e^{-i\,\operatorname{ad}(y)}\operatorname{Ad}(k))|_{\mathfrak{u}_\mathfrak{m}^\perp(\gamma)}}{\det(1-\operatorname{Ad}(k))_{\mathfrak{u}_\mathfrak{m}^\perp(\gamma)}}\right]^{\frac{1}{2}}$$
$$\cdot\operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)\otimes E}[e^{-i\rho^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)\otimes E}(y)}\rho^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)\otimes E}(k)] \tag{7.2.41}$$

can be extended directly to a $U_M(\gamma)^0$-invariant function in $y\in\mathfrak{u}_\mathfrak{m}(\gamma)$. Since $\mathfrak{t}$ is a Cartan subalgebra of $\mathfrak{u}_\mathfrak{m}(\gamma)$, we can apply the Weyl integration formula for the pair $(\mathfrak{u}_\mathfrak{m}(\gamma),\mathfrak{t})$; we get

$$F_\gamma(\Lambda,t)=\frac{c_G(\gamma)}{(2\pi t)^{\frac{r}{2}}}\int_{y\in\mathfrak{u}_\mathfrak{m}(\gamma)}\det(i\,\operatorname{ad}(y))|_{\mathfrak{n}(\gamma)_\mathbb{C}}\cdot\widehat{A}^{-1}(i\,\operatorname{ad}(y)|_{\mathfrak{u}_\mathfrak{m}(\gamma)})\left[\frac{\det(1-e^{-i\,\operatorname{ad}(y)}\operatorname{Ad}(k))|_{\mathfrak{u}_\mathfrak{m}^\perp(\gamma)}}{\det(1-\operatorname{Ad}(k))_{\mathfrak{u}_\mathfrak{m}^\perp(\gamma)}}\right]^{\frac{1}{2}}$$
$$\cdot\operatorname{Tr}_s^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)\otimes E}[e^{-i\rho^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)\otimes E}(y)}\rho^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)\otimes E}(k)]e^{-\frac{|y|^2}{2t}}\,dy. \tag{7.2.42}$$

The constant $c_G(\gamma)$ is defined by (7.2.5).

Note that

$$r=\dim\mathfrak{u}_\mathfrak{m}(\gamma)+4l(\gamma)+1. \tag{7.2.43}$$

If $y\in\mathfrak{u}_\mathfrak{m}(\gamma)$, then

$$B\Big(y,\frac{\Omega^{\mathfrak{u}_\mathfrak{m}(\gamma)}}{2\pi}\Big)\in\Lambda^2(\mathfrak{u}_\mathfrak{b}^\perp(\gamma)^*). \tag{7.2.44}$$

If $y \in \mathfrak{u}_\mathfrak{m}(\gamma)$, by [Shen 2018, equation (7-27)], we have

$$\frac{\det(i\,\mathrm{ad}(y))|_{\mathfrak{n}(\gamma)_\mathbb{C}}}{(2\pi t)^{2l(\gamma)}} = \left[\exp\left(\frac{1}{t}B\left(y, \frac{\Omega^{\mathfrak{u}_\mathfrak{m}(\gamma)}}{2\pi}\right)\right)\right]^{\max(\gamma)}. \tag{7.2.45}$$

Combining (7.2.42)–(7.2.45), we get

$$F_\gamma(\Lambda,t)=\frac{c_G(\gamma)}{\sqrt{2\pi t}}\left[\int_{y\in\mathfrak{u}_\mathfrak{m}(\gamma)} \widehat{A}^{-1}(i\,\mathrm{ad}(y)|_{\mathfrak{u}_\mathfrak{m}(\gamma)})\left[\frac{\det(1-e^{-i\,\mathrm{ad}(y)}\,\mathrm{Ad}(k))|_{\mathfrak{u}_\mathfrak{m}^\perp(\gamma)}}{\det(1-\mathrm{Ad}(k))_{\mathfrak{u}_\mathfrak{m}^\perp(\gamma)}}\right]^{\frac{1}{2}}\right.$$
$$\left.\cdot\mathrm{Tr}_s^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)\otimes E}[\rho^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)\otimes E}(e^{-iy}k)]e^{\frac{1}{t}B\left(y,\frac{\Omega^{\mathfrak{u}_\mathfrak{m}(\gamma)}}{2\pi}\right)-\frac{|y|^2}{2t}}\frac{dy}{(2\pi t)^{\dim\mathfrak{u}_\mathfrak{m}(\gamma)/2}}\right]^{\max(\gamma)}. \tag{7.2.46}$$

By (5.2.21) if $y \in \mathfrak{u}_\mathfrak{m}(\gamma)$, then

$$B\left(y, \frac{\Omega^{\mathfrak{u}_\mathfrak{m}(\gamma)}}{2\pi}\right) - \frac{|y|^2}{2} = \frac{1}{2}B\left(y+\frac{\Omega^{\mathfrak{u}_\mathfrak{m}(\gamma)}}{2\pi}, y+\frac{\Omega^{\mathfrak{u}_\mathfrak{m}(\gamma)}}{2\pi}\right) - \frac{\beta(a_1)^2}{8\pi^2}\omega^{Y_\mathfrak{b}(\gamma),2}. \tag{7.2.47}$$

Let $\Delta^{\mathfrak{u}_\mathfrak{m}(\gamma)}$ be the standard negative Laplace operator on the Euclidean space $(\mathfrak{u}_\mathfrak{m}(\gamma), -B|_{\mathfrak{u}_\mathfrak{m}(\gamma)})$. Then by considering the heat kernel of $-\Delta^{\mathfrak{u}_\mathfrak{m}(\gamma)}$, we can rewrite (7.2.46) as

$$F_\gamma(\Lambda,t) = \frac{c_G(\gamma)}{\sqrt{2\pi t}}\left[\exp\left(-\frac{\beta(a_1)^2\omega^{Y_\mathfrak{b}(\gamma),2}}{8\pi^2 t}\right)\right.$$
$$\cdot\exp\left(\frac{t}{2}\Delta^{\mathfrak{u}_\mathfrak{m}(\gamma)}\right)\left\{\widehat{A}^{-1}(i\,\mathrm{ad}(y)|_{\mathfrak{u}_\mathfrak{m}(\gamma)})\left[\frac{\det(1-e^{-i\,\mathrm{ad}(y)}\,\mathrm{Ad}(k))|_{\mathfrak{u}_\mathfrak{m}^\perp(\gamma)}}{\det(1-\mathrm{Ad}(k))_{\mathfrak{u}_\mathfrak{m}^\perp(\gamma)}}\right]^{\frac{1}{2}}\right.$$
$$\left.\left.\cdot\mathrm{Tr}_s^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)\otimes E}[\rho^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)\otimes E}(e^{-iy}k)]\right\}\right|_{y=-\frac{\Omega^{\mathfrak{u}_\mathfrak{m}(\gamma)}}{2\pi}}\Bigg]^{\max(\gamma)}. \tag{7.2.48}$$

Recall that $V_{\Lambda,\omega}$ is an irreducible unitary representation of U_M with highest weight $\eta_\omega(\Lambda)$. By (5.3.7), for $y \in \mathfrak{u}_\mathfrak{m}(\gamma)$, then

$$\mathrm{Tr}_s^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)\otimes E}[\rho^{\Lambda^\bullet(\mathfrak{n}_\mathbb{C}^*)\otimes E}(e^{-iy}k)] = \sum_{\omega\in W_u}\varepsilon(\omega)\,\mathrm{Tr}^{V_{\Lambda,\omega}}[\rho^{V_{\Lambda,\omega}}(e^{-iy}k)]. \tag{7.2.49}$$

Then we apply the generalized Kirillov formula (5.4.21) to each term in the right-hand side of (7.2.49), we conclude that, for $\omega \in W_u$, the function in $y \in \mathfrak{u}_\mathfrak{m}(\gamma)$

$$\widehat{A}^{-1}(i\,\mathrm{ad}(y)|_{\mathfrak{u}_\mathfrak{m}(\gamma)})\left[\frac{\det(1-e^{-i\,\mathrm{ad}(y)}\,\mathrm{Ad}(k))|_{\mathfrak{u}_\mathfrak{m}^\perp(\gamma)}}{\det(1-\mathrm{Ad}(k))_{\mathfrak{u}_\mathfrak{m}^\perp(\gamma)}}\right]^{\frac{1}{2}}\mathrm{Tr}^{V_{\Lambda,\omega}}[\rho^{V_{\Lambda,\omega}}(e^{-iy}k)] \tag{7.2.50}$$

is an eigenfunction of $\Delta^{\mathfrak{u}_\mathfrak{m}(\gamma)}$ associated with the eigenvalue $4\pi^2|\eta_\omega(\Lambda)+\rho_{\mathfrak{u}_\mathfrak{m}}|^2$. Then the heat operator $\exp\left(\frac{t}{2}\Delta^{\mathfrak{u}_\mathfrak{m}(\gamma)}\right)$ acts on the function (7.2.50) as a scalar $e^{2\pi^2 t|\eta_\omega(\Lambda)+\rho_{\mathfrak{u}_\mathfrak{m}}|^2}$. By (5.3.8), (5.3.9), for $\omega \in W_u$, we get

$$\eta_\omega(\Lambda)+\rho_{\mathfrak{u}_\mathfrak{m}} = P_0(\omega(\Lambda+\rho_\mathfrak{u})). \tag{7.2.51}$$

Combing the above computation with the term $e^{-2\pi^2 t|\Lambda+\rho_\mathfrak{u}|^2}$ in (7.2.34), by (7.2.10), we get the factor $e^{-2\pi^2 t b_{\Lambda,\omega}^2}$ in (7.2.11).

Now we deal with the main part in (7.2.48) after removing the heat operator $\exp\left(\frac{t}{2}\Delta^{\mathfrak{u}_{\mathfrak{m}}(\gamma)}\right)$. We will use the same notation as in Section 5.4. The orbit $\mathcal{O}^{\gamma}_{\sigma(\eta_\omega(\Lambda)+\rho_{\mathfrak{u}_{\mathfrak{m}}})}$ is defined in (5.4.17) equipped with a Liouville measure $d\mu^{\gamma}_{\sigma}$. We claim the identity

$$\begin{aligned}&\Bigg[\exp\left(-\frac{\beta(a_1)^2\omega^{Y_{\mathfrak{b}}(\gamma),2}}{8\pi^2 t}\right)\\&\quad\cdot\left\{\widehat{A}^{-1}(i\,\mathrm{ad}(y)|_{\mathfrak{u}_{\mathfrak{m}}(\gamma)})\left[\frac{\det(1-e^{-i\,\mathrm{ad}(y)}\mathrm{Ad}(k))|_{\mathfrak{u}^{\perp}_{\mathfrak{m}}(\gamma)}}{\det(1-\mathrm{Ad}(k))_{\mathfrak{u}^{\perp}_{\mathfrak{m}}(\gamma)}}\right]^{\frac{1}{2}}\mathrm{Tr}_s^{V_{\Lambda,\omega}}[\rho^{V_{\Lambda,\omega}}(e^{-iy}k)]\right\}\Bigg|_{y=-\frac{\Omega^{\mathfrak{u}_{\mathfrak{m}}(\gamma)}}{2\pi}}\Bigg]^{\max(\gamma)}\\&=\sum_{\sigma\in W^1(\gamma)}\varphi_\gamma(\sigma,\eta_\omega(\Lambda))\cdot\dim E^{\gamma}_{\omega,\sigma}\left[\exp\left(-\frac{\beta(a_1)^2\omega^{Y_{\mathfrak{b}}(\gamma),2}}{8\pi^2 t}-\langle\sigma(\eta_\omega(\Lambda)+\rho_{\mathfrak{u}_{\mathfrak{m}}}),\Omega^{\mathfrak{u}_{\mathfrak{m}}(\gamma)}\rangle\right)\right]^{\max(\gamma)}.\end{aligned}\tag{7.2.52}$$

Indeed, by (5.4.21), we have the following identity as elements in $\Lambda^{\bullet}(\mathfrak{u}^{\perp}_{\mathfrak{b}}(\gamma)^*)$:

$$\begin{aligned}&\left\{\widehat{A}^{-1}(i\,\mathrm{ad}(y)|_{\mathfrak{u}_{\mathfrak{m}}(\gamma)})\left[\frac{\det(1-e^{-i\,\mathrm{ad}(y)}\mathrm{Ad}(k))|_{\mathfrak{u}^{\perp}_{\mathfrak{m}}(\gamma)}}{\det(1-\mathrm{Ad}(k))_{\mathfrak{u}^{\perp}_{\mathfrak{m}}(\gamma)}}\right]^{\frac{1}{2}}\mathrm{Tr}_s^{V_{\Lambda,\omega}}[\rho^{V_{\Lambda,\omega}}(e^{-iy}k)]\right\}\Bigg|_{y=-\frac{\Omega^{\mathfrak{u}_{\mathfrak{m}}(\gamma)}}{2\pi}}\\&\qquad=\sum_{\sigma\in W^1(\gamma)}\varphi_\gamma(\sigma,\eta_\omega(\Lambda))\int_{f\in\mathcal{O}^{\gamma}_{\sigma(\eta_\omega(\Lambda)+\rho_{\mathfrak{u}_{\mathfrak{m}}})}}e^{-\langle f,\Omega^{\mathfrak{u}_{\mathfrak{m}}(\gamma)}\rangle}\,d\mu^{\gamma}_{\sigma}.\end{aligned}\tag{7.2.53}$$

Recall that the curvature form $\Omega^{\mathfrak{u}_{\mathfrak{b}}(\gamma)}$ is invariant by the action of $U_M(\gamma)^0$ on $Y_{\mathfrak{b}}(\gamma)$. Since a_1 and $\omega^{Y_{\mathfrak{b}}(\gamma)}$ are invariant by $U_M(\gamma)^0$-action, so is $\Omega^{\mathfrak{u}_{\mathfrak{m}}(\gamma)}$. Therefore, for $f\in\mathfrak{u}_{\mathfrak{m}}(\gamma)^*$, $u\in U_M(\gamma)^0$,

$$\begin{aligned}&\left[\exp\left(-\frac{\beta(a_1)^2\omega^{Y_{\mathfrak{b}}(\gamma),2}}{8\pi^2 t}\right)\exp(-\langle\mathrm{Ad}^*(u)f,\Omega^{\mathfrak{u}_{\mathfrak{m}}(\gamma)}\rangle)\right]^{\max(\gamma)}\\&\qquad=\det\mathrm{Ad}(u)|_{\mathfrak{u}^{\perp}_{\mathfrak{b}}(\gamma)}\left[\exp\left(-\frac{\beta(a_1)^2\omega^{Y_{\mathfrak{b}}(\gamma),2}}{8\pi^2 t}\right)\exp(-\langle f,\Omega^{\mathfrak{u}_{\mathfrak{m}}(\gamma)}\rangle)\right]^{\max(\gamma)}.\end{aligned}\tag{7.2.54}$$

Since $U_M(\gamma)^0$ acts on $\mathfrak{u}^{\perp}_{\mathfrak{b}}(\gamma)$ isometrically with respect to $-B|_{\mathfrak{u}^{\perp}_{\mathfrak{b}}(\gamma)}$,

$$\det\mathrm{Ad}(u)|_{\mathfrak{u}^{\perp}_{\mathfrak{b}}(\gamma)}=1.\tag{7.2.55}$$

Then (7.2.52) follows from (5.4.6) and (7.2.53)–(7.2.55).

The right-hand side of (7.2.52) is a polynomial in t^{-1}. Recall that $\dim\mathfrak{u}^{\perp}_{\mathfrak{b}}(\gamma)=4l(\gamma)$. Then, for each $\sigma\in W^1(\gamma)$, we can rewrite the term $[\cdots]^{\max(\gamma)}$ in the right-hand side of (7.2.52) as follows:

$$\sum_{j=0}^{l(\gamma)}\frac{1}{t^j}\frac{(-1)^j\beta(a_1)^{2j}}{j!(2l(\gamma)-2j)!(8\pi^2)^j}[\omega^{Y_{\mathfrak{b}}(\gamma),2j}\langle\omega(\Lambda+\rho_{\mathfrak{u}}),\Omega^{\mathfrak{u}_{\mathfrak{m}}(\gamma)}\rangle^{2l(\gamma)-2j}]^{\max(\gamma)}.\tag{7.2.56}$$

Finally, putting together (7.2.7), (7.2.34), (7.2.48), (7.2.49), (7.2.52), and (7.2.56), we get (7.2.11). □

The Mellin transform of $\mathcal{E}_{X,\gamma}(F,t)$ (if applicable) is defined by the following formula as a function in $s\in\mathbb{C}$ with $\Re(s)\gg 0$:

$$\mathcal{M}\mathcal{E}_{X,\gamma}(F,s)=-\frac{1}{\Gamma(s)}\int_0^{+\infty}\mathcal{E}_{X,\gamma}(F,t)t^{s-1}\,dt.\tag{7.2.57}$$

If $\mathcal{ME}_{X,\gamma}(F,s)$ admits a meromorphic extension on $\mathbb{C}$ which is holomorphic at $s=0$, we will set

$$\mathcal{PE}_{X,\gamma}(F) = \frac{\partial}{\partial s}\Big|_{s=0} \mathcal{ME}_{X,\gamma}(F,s). \tag{7.2.58}$$

Theorem 7.2.7. *Suppose that the dominant weight Λ is such that, for every $\omega \in W_u$, $b_{\Lambda,\omega} \neq 0$. Then, for $s \in \mathbb{C}$ with $\Re(s) > l(\gamma)+1$, $\mathcal{ME}_{X,\gamma}(F,s)$ is well-defined and holomorphic, which admits a meromorphic extension to $s \in \mathbb{C}$.*

Moreover, we have the identity

$$\mathcal{ME}_{X,\gamma}(F,s) = -\frac{c_G(\gamma)}{\sqrt{2\pi}} \sum_{j=0}^{l(\gamma)} \frac{\Gamma\big(s-j-\frac{1}{2}\big)}{\Gamma(s)} \Bigg[\sum_{\substack{\omega\in W_u\\ \sigma\in W^1(\gamma)}} \varepsilon(\omega)\varphi_\gamma(\sigma,\eta_\omega(\Lambda)) Q^\gamma_{j,\omega,\sigma}(\Lambda)(2\pi^2 b^2_{\Lambda,\omega})^{j+\frac{1}{2}-s}\Bigg]. \tag{7.2.59}$$

Then $\mathcal{ME}_{X,\gamma}(F,s)$ is holomorphic at $s=0$. We have

$$\mathcal{PE}_{X,\gamma}(F) = -\frac{c_G(\gamma)}{\sqrt{2}} \sum_{j=0}^{l(\gamma)} \frac{(-4)^{j+1}(j+1)!}{(2j+2)!} \Bigg[\sum_{\substack{\omega\in W_u\\ \sigma\in W^1(\gamma)}} \varepsilon(\omega)\varphi_\gamma(\sigma,\eta_\omega(\Lambda)) Q^\gamma_{j,\omega,\sigma}(\Lambda)(2\pi^2 b^2_{\Lambda,\omega})^{j+\frac{1}{2}}\Bigg]. \tag{7.2.60}$$

Proof. By Theorem 7.2.2, the assumption on Λ implies that $\mathcal{E}_{X,\gamma}(F,t)$ decays exponentially as $t\to+\infty$. By (7.1.6) and (7.2.11), we get (7.2.59). This proves the first part of this theorem.

Equation (7.2.60) is a direct consequence of (7.2.59) by taking its derivative at 0. This completes the proof of our theorem. □

The formula in the right-hand side of (7.2.60) still looks complicated; we can rewrite it in a neat way as follows. We introduce the following functions.

Definition 7.2.8. Let $a^1 \in \mathfrak{b}^*$ take value -1 at a_1. Note that $\gamma \in T$. For $\omega \in W_u$, $\sigma \in W^1(\gamma)$, if $\Lambda \in P_{++}(U)$, for $z\in\mathbb{C}$, set

$$P^\gamma_{\omega,\sigma,\Lambda}(z) = \dim E^\gamma_{\omega,\sigma} \cdot \big[\exp\big(\langle \Omega^{\mathfrak{u}_\mathfrak{b}(\gamma)}, \sigma(\eta_\omega(\Lambda)+\rho_{\mathfrak{u}_\mathfrak{m}}) + z\sqrt{-1}a^1\rangle\big)\big]^{\max(\gamma)}. \tag{7.2.61}$$

Since θ fixes $\Omega^{\mathfrak{u}_\mathfrak{b}(\gamma)}$, by the fact that $\det\theta|_{\mathfrak{u}_\mathfrak{b}^\perp(\gamma)} = 1$, we have $P^\gamma_{\omega,\sigma,\Lambda}(z)$ is an even polynomial in z. Moreover, by the dimension formula (5.4.6), the coefficients of z^j, $j\in\mathbb{N}$, in $P^\gamma_{\omega,\sigma,\Lambda}(z)$ are polynomials in Λ. Such polynomials are related to the Plancherel measures in the representation theory.

Lemma 7.2.9. *We have the identity*

$$\sum_{j=0}^{l(\gamma)} \frac{(-4)^{j+1}(j+1)!}{\sqrt{2}(2j+2)!} Q^\gamma_{j,\omega,\sigma}(\Lambda)(2\pi^2(b_{\Lambda,\omega})^2)^{j+\frac{1}{2}} = -2\pi\int_0^{|b_{\Lambda,\omega}|} P^\gamma_{\omega,\sigma,\Lambda}(t)\,dt. \tag{7.2.62}$$

Proof. We have

$$\langle \eta_\omega(\Lambda)+\rho_{\mathfrak{u}_\mathfrak{m}} + z\sqrt{-1}a^1, \Omega^{\mathfrak{u}_\mathfrak{b}(\gamma)}\rangle = z\beta(a_1)\omega^{Y_\mathfrak{b}(\gamma)} + \langle \omega(\Lambda+\rho_\mathfrak{u}), \Omega^{\mathfrak{u}_\mathfrak{m}(\gamma)}\rangle. \tag{7.2.63}$$

Since $P^{\gamma}_{\omega,\sigma,\Lambda}(z)$ is an even function in z,

$$\begin{aligned} P^{\gamma}_{\omega,\sigma,\Lambda}(z) &= \dim E^{\gamma}_{\omega,\sigma}\cdot\frac{1}{(2l(\gamma))!}\Big[\big(z\beta(a_1)\omega^{Y_{\mathfrak{b}}(\gamma)}+\langle\omega(\Lambda+\rho_{\mathfrak{u}}),\Omega^{\mathfrak{u}_{\mathfrak{m}}(\gamma)}\rangle\big)^{2l(\gamma)}\Big]^{\max(\gamma)} \\ &= \dim E^{\gamma}_{\omega,\sigma}\cdot\sum_{j=0}^{l(\gamma)}\frac{\beta(a_1)^{2j}z^{2j}}{(2l(\gamma)-2j)!(2j)!}\Big[\omega^{Y_{\mathfrak{b}}(\gamma),2j}\langle\omega(\Lambda+\rho_{\mathfrak{u}}),\Omega^{\mathfrak{u}_{\mathfrak{m}}(\gamma)}\rangle^{2l(\gamma)-2j}\Big]^{\max(\gamma)}. \end{aligned} \tag{7.2.64}$$

Note that, for $j=0,1,\cdots,l(\gamma)$,

$$\int_0^{|b_{\Lambda,\omega}|} t^{2j}\,dt = \frac{1}{2j+1}|b_{\Lambda,\omega}|^{2j+1}. \tag{7.2.65}$$

Then (7.2.62) is a consequence of (7.2.7), (7.2.64) and (7.2.65). □

As a consequence, we get the following formula for $\mathcal{P}\mathcal{E}_{X,\gamma}(F)$.

Theorem 7.2.10. *Suppose that the dominant weight Λ is such that, for every $\omega\in W_u$, $b_{\Lambda,\omega}\neq 0$. Then*

$$\mathcal{P}\mathcal{E}_{X,\gamma}(F) = 2\pi c_G(\gamma)\cdot\sum_{\substack{\omega\in W_u\\ \sigma\in W^1(\gamma)}}\varepsilon(\omega)\varphi_\gamma(\sigma,\eta_\omega(\Lambda))\int_0^{|b_{\Lambda,\omega}|}P^{\gamma}_{\omega,\sigma,\Lambda}(t)\,dt. \tag{7.2.66}$$

7.3. ***A family of representations of G.*** We recall a definition of nondegeneracy of λ in [Bismut et al. 2017, Definition 1.13, Proposition 8.12].

Definition 7.3.1. A dominant weight $\Lambda\in P_{++}(U)$ is said to be nondegenerate with respect to the Cartan involution θ if

$$W(U,T_U)\cdot\Lambda\cap\mathfrak{t}^* = \varnothing. \tag{7.3.1}$$

It is equivalent to

$$\mathrm{Ad}^*(U)\Lambda\cap\mathfrak{k}^* = \varnothing. \tag{7.3.2}$$

Note that if such dominant weight exists, we must have $\delta(G)>0$.

Let (E,ρ^E) be the irreducible unitary representation of U with highest weight $\Lambda\in P_{++}(U)$. By the unitary trick, it extends to an irreducible representation of G, which we still denote by (E,ρ^E). Then Λ being nondegenerate is equivalent to saying that (E,ρ^E) is not isomorphic to $(E,\rho^E\circ\theta)$ as G-representations (as in [Müller and Pfaff 2013a]).

Definition 7.3.2. If $\lambda\in\mathfrak{t}^*_U$, for $\omega\in W(U,T_U)$, put

$$a_{\lambda,\omega} = \langle\omega\cdot\lambda,\sqrt{-1}a_1\rangle\in\mathbb{R}. \tag{7.3.3}$$

Recall the real number $b_{\lambda,\omega}$ is already defined by (7.2.9); then $b_{\lambda,\omega}=a_{\lambda,\omega}+a_{\rho_{\mathfrak{u}},\omega}$. In particular, we simply put $a_\lambda=a_{\lambda,1}$, $b_\lambda=b_{\lambda,1}$.

Lemma 7.3.3. *If $\lambda\in P_{++}(U)$ is nondegenerate, then, for $\omega\in W(U,T_U)$, $a_{\lambda,\omega}\neq 0$.*

Now we fix two dominant weights $\lambda, \lambda_0 \in P_{++}(U)$. Let $\{(E_d, \rho^{E_d})\}|_{d\in\mathbb{N}}$ be the sequence of representations of G given by the irreducible unitary representations of U with the highest weights $d\lambda + \lambda_0$, $d \in \mathbb{N}$.

Put $F_d = G \times_K E_d$. Let $\boldsymbol{D}^{X,F_d,2}$ denote the associated de Rham–Hodge Laplacian. For $t > 0$, let $\exp(-t\boldsymbol{D}^{X,F_d,2}/2)$ denote the heat operator associated with $\boldsymbol{D}^{X,F_d,2}/2$. By taking $\Lambda = d\lambda + \lambda_0$, we apply our results in previous subsection to the sequence $\mathcal{E}_{X,\gamma}(F_d, t)$, $d \in \mathbb{N}$.

7.4. ***Asymptotics for identity orbital integrals.*** In this subsection, we specialize our results in Section 7.2 for $\gamma = 1$ and $\Lambda = d\lambda + \lambda_0$. Now the set $W^1(\gamma)$ reduces to $\{1\}$, and $l(\gamma) = l$, $\varphi_\gamma(\sigma, \eta_\omega(\Lambda)) = 1$. We will drop the superscript γ and subscript σ in our notation.

Moreover, for $\omega \in W_u$, the representation $E^{\gamma=1}_{\omega,\sigma=1}$ is just $V_{\Lambda,\omega}$ introduced in (5.3.6), which is the irreducible unitary representation of U_M with highest weight $\eta_\omega(\Lambda)$ given by (5.3.8).

Definition 7.4.1. By taking $\Lambda = d\lambda + \lambda_0$ in (7.2.7), we define the following functions in d: for $j = 0, 1, \dots, l$, $\omega \in W_u$, set

$$\begin{aligned} Q^{\lambda,\lambda_0}_{j,\omega}(d) &= Q_{j,\omega}(d\lambda + \lambda_0) \\ &= \frac{(-1)^j \beta(a_1)^{2j}}{j!(2l-2j)!(8\pi^2)^j} \dim V_{d\lambda+\lambda_0,\omega}[\omega^{Y_{\mathfrak{b}},2j} \langle \omega(d\lambda + \lambda_0 + \rho_{\mathfrak{u}}), \Omega^{\mathfrak{u}_{\mathfrak{m}}}\rangle^{2l-2j}]^{\max}. \end{aligned} \tag{7.4.1}$$

By the Weyl dimension formula, $\dim V_{d\lambda+\lambda_0,\omega}$ is a polynomial in d. Then $Q^{\lambda,\lambda_0}_{j,\omega}(d)$ is a polynomial in d of degree $\leq \frac{1}{2}\dim(\mathfrak{g}/\mathfrak{h}) - 2j$.

By Theorem 7.2.2 and (7.4.1), we get directly the following results.

Theorem 7.4.2. *For $t > 0$, we have the identity*

$$\mathcal{I}_X(F_d, t) = \frac{c_G}{\sqrt{2\pi t}} \sum_{j=0}^{l} t^{-j} \sum_{\omega\in W_u} \varepsilon(\omega) e^{-2\pi^2 t (da_{\lambda,\omega} + b_{\lambda_0,\omega})^2} Q^{\lambda,\lambda_0}_{j,\omega}(d). \tag{7.4.2}$$

Theorem 7.4.3. *Suppose that λ is nondegenerate with respect to θ. For $d \in \mathbb{N}$ large enough and for $s \in \mathbb{C}$ with $\Re(s) \gg 0$, $\mathcal{M}\mathcal{I}_X(F_d, s)$ is well-defined and holomorphic, which admits a unique meromorphic extension to $s \in \mathbb{C}$ and is holomorphic at $s = 0$.*

Moreover, we have the identities

$$\mathcal{M}\mathcal{I}_X(F_d, s) = -\frac{c_G}{\sqrt{2\pi}} \sum_{j=0}^{l} \frac{\Gamma\big(s - j - \frac{1}{2}\big)}{\Gamma(s)} \Bigg[\sum_{\omega\in W_u} \varepsilon(\omega) Q^{\lambda,\lambda_0}_{j,\omega}(d) (2\pi^2 (da_{\lambda,\omega} + b_{\lambda_0,\omega})^2)^{j+\frac{1}{2}-s}\Bigg], \tag{7.4.3}$$

$$\mathcal{P}\mathcal{I}_X(F_d) = -\frac{c_G}{\sqrt{2}} \sum_{j=0}^{l} \frac{(-4)^{j+1}(j+1)!}{(2j+2)!} \Bigg[\sum_{\omega\in W_u} \varepsilon(\omega) Q^{\lambda,\lambda_0}_{j,\omega}(d) (2\pi^2 (da_{\lambda,\omega} + b_{\lambda_0,\omega})^2)^{j+\frac{1}{2}}\Bigg]. \tag{7.4.4}$$

In particular, the quantity $\mathcal{P}\mathcal{I}_X(F_d)$ is a polynomial in d for d large enough, whose coefficients depend only on the given root system and λ, λ_0, and has degree $\leq \frac{1}{2}\dim(\mathfrak{g}/\mathfrak{h}) + 1$.

Proof. Since λ is nondegenerate, by Lemma 7.3.3, $a_{\lambda,\omega} \neq 0$, $\omega \in W_u$. Then there exists $d_0 \in \mathbb{N}$ such that for $d \geq d_0$, $(da_{\lambda,\omega} + b_{\lambda_0,\omega})^2 > 0$. Then by Theorem 7.2.7, we get first part of this theorem and (7.4.3), (7.4.4).

Note that

$$[(da_{\lambda,\omega} + b_{\lambda_0,\omega})^2]^{\frac{1}{2}} = |da_{\lambda,\omega} + b_{\lambda_0,\omega}|.$$

For $d \gg d_0$,

$$|da_{\lambda,\omega} + b_{\lambda_0,\omega}| = \operatorname{sign}(a_{\lambda,\omega})(da_{\lambda,\omega} + b_{\lambda_0,\omega}).$$

Then we see that $\mathcal{PI}_X(F_d)$ is a polynomial in d for d large enough. □

As explained in Remark 5.3.3, when G has noncompact center with $\delta(G) = 1$ (but U is still assumed to be compact), most of the above computations can be reduce into very simple ones. Recall that $a_\lambda, b_{\lambda_0} \in \mathbb{R}$ are defined in Definition 7.3.2.

Corollary 7.4.4. *Assume that U is compact and that G has noncompact center with $\delta(G) = 1$, and assume that λ is nondegenerate. Then, for $t > 0$, $s \in \mathbb{C}$,*

$$\mathcal{I}_X(F_d, t) = \frac{c_G}{\sqrt{2\pi t}} e^{-2\pi^2 t (da_\lambda + b_{\lambda_0})^2} \dim E_d,$$

$$\mathcal{MI}_X(F_d, s) = -\frac{c_G}{\sqrt{2\pi}} \frac{\Gamma\left(s - \frac{1}{2}\right)}{\Gamma(s)} (2\pi^2 (da_\lambda + b_{\lambda_0})^2)^{1/2 - s} \dim E_d. \tag{7.4.5}$$

Furthermore,

$$\mathcal{PI}_X(F_d) = 2\pi c_G |da_\lambda + b_{\lambda_0}| \dim E_d. \tag{7.4.6}$$

Proof. By the hypothesis, we get that $l = 0$, $W_u = \{1\}$ and $Q_{0,1}^{\lambda,\lambda_0}(d) = \dim E_d$. Then (7.4.5), (7.4.6) are just special cases of (7.4.2), (7.4.3) and (7.4.4).

However, we can prove them more directly using a result of Proposition 4.1.6. It is enough to prove the first identity in (7.4.5). Note that by (5.3.11), we have

$$X' = M/K, \tag{7.4.7}$$

with $\delta(X') = 0$.

By [Müller and Pfaff 2013a, Proposition 5.2] or [Shen 2018, Proposition 4.1], we have

$$[e(TX', \nabla^{TX'})]^{\max} = (-1)^{\frac{m-1}{2}} \frac{|W(U_M, T)|/|W(K, T)|}{\operatorname{Vol}(U_M/K)}. \tag{7.4.8}$$

Then by (7.2.6), we have

$$[e(TX', \nabla^{TX'})]^{\max} = -c_G. \tag{7.4.9}$$

By (4.1.28) and (7.3.3), we have

$$\alpha_{E_d} = -2\pi (da_\lambda + b_{\lambda_0}). \tag{7.4.10}$$

Combing (4.1.31) and (7.4.8) - (7.4.10), we get the first identity in (7.4.5), and hence the other identities. This gives a second proof to this corollary. □

7.5. *Connection to Müller and Pfaff's results.* In this subsection, we assume that G has compact center with $\delta(G)=1$. We explain here how to connect our computations in the previous subsection to the results in [Müller and Pfaff 2013a].

For $\gamma=1$, $\omega\in W_u$, the function $P^{\gamma}_{\omega,\sigma,\Lambda}$ defined in (7.2.61) now reduces to

$$P_{\omega,\Lambda}(z)=\dim V_{\Lambda,\omega}\big[\exp\big(\langle\eta_\omega(\Lambda)+\rho_{\mathfrak{u}_\mathfrak{m}}+z\sqrt{-1}a^1,\Omega^{\mathfrak{u}(\mathfrak{b})}\rangle\big)\big]^{\max}. \tag{7.5.1}$$

We can verify directly that

$$P_{\omega,\Lambda}(z)=\frac{\mathrm{Vol}(U_M/T)}{\mathrm{Vol}(U/T_U)}\Pi_{\alpha^0\in R^+(\mathfrak{u},\mathfrak{t}_U)}\frac{\langle\alpha^0,\eta_\omega(\Lambda)+\rho_{\mathfrak{u}_\mathfrak{m}}+z\sqrt{-1}a^1\rangle}{\langle\alpha^0,\rho_\mathfrak{u}\rangle}. \tag{7.5.2}$$

The scalar product in (7.5.2) is taken with respect to $-B|_\mathfrak{u}$. Up to a universal constant, $P_{\omega,\Lambda}(z)$ is just the polynomial related to the Plancherel measure of representation $V_{\Lambda,\omega}$ as given in [Müller and Pfaff 2013a, equation (6.10)]. Note that there is no factor $(2\pi)^{2l}$ in (7.5.2) because of our normalization for $[\cdot]^{\max}$.

By Theorem 7.2.10, we have the following result for sufficiently large d.

Corollary 7.5.1. *Suppose that λ is nondegenerate with respect to θ. Then*

$$\mathcal{PI}_X(F_d)=2\pi c_G\sum_{\omega\in W_u}\varepsilon(\omega)\int_0^{|da_{\lambda,\omega}+b_{\lambda_0,\omega}|}P_{\omega,d\lambda+\lambda_0}(t)\,dt. \tag{7.5.3}$$

By [Müller and Pfaff 2013a, Lemma 6.1], we can get the identity

$$|W(K,T)|=2|W(K_M,T)|. \tag{7.5.4}$$

Combining (7.2.6), (7.5.2), (7.5.4), we see that the formula in Corollary 7.5.1, is exactly the same formula of [Müller and Pfaff 2013a, Proposition 6.6] for $\mathcal{PI}_X(F_d)$.

Recall that the U-representation E_d has highest weight $d\lambda+\lambda_0\in P_{++}(U)$. Then by Weyl dimension formula, $\dim E_d$ is a polynomial in d. If λ is regular, then the degree (in d) of $\dim E_d$ is $\frac{1}{2}\dim\mathfrak{g}/\mathfrak{h}$.

For determining the leading term of $\mathcal{PI}_X(F_d)$, as mentioned in the Introduction, we can specialize the result of [Bismut et al. 2017, Theorem 0.1] as in Section 8 of that work for the symmetric space X. Here to emphasize $\mathcal{PI}_X(F_d)$ being a polynomial in d, we state a result of [Müller and Pfaff 2013a, Proposition 1.3] as follows.

Proposition 7.5.2. *Suppose that λ is nondegenerate and that $\lambda_0=0$. Then there exists a constant $C_{X,\lambda}\neq 0$ such that*

$$\mathcal{PI}_X(F_d)=C_{X,\lambda}d\dim E_d+R(d), \tag{7.5.5}$$

where $R(d)$ is a polynomial whose degree is no greater than the degree of $\dim E_d$.

Remark 7.5.3. Note that Müller and Pfaff [2013a, Proposition 1.3] proved Proposition 7.5.2 by reducing the problems to the cases $G=\mathrm{SL}_3(\mathbb{R})$ and $\mathrm{SO}^0(p,q)$ ($pq>1$ odd). In particular, for certain examples of λ, they also worked out explicitly the constant $C_{X,\lambda}$ [Müller and Pfaff 2013a, Corollaries 1.4, 1.5].

Similarly, if we take a nonzero λ_0, we can repeat their computations for $G=\mathrm{SL}_3(\mathbb{R})$ and $\mathrm{SO}^0(p,q)$ ($pq>1$ odd) in order to get more explicit information on the leading terms of $\mathcal{PI}_X(F_d)$.

An important step in Müller and Pfaff's proof to Proposition 7.5.2 is reducing the computation of $\mathcal{PI}_X(F_d)$ to the cases where $\mathfrak{g} = \mathfrak{sl}_3(\mathbb{R})$ or $\mathfrak{so}(p,q)$ with $pq > 1$ odd. Such reduction is already explained in Section 4.2. More precisely, we have

$$X = X_1 \times X_2, \tag{7.5.6}$$

where X_1 is one case listed in (4.2.1), and X_2 is a symmetric space with rank 0.

We use the notation in Section 4.2 and assume G to be semisimple. Let λ_i, $\lambda_{0,i}$ be dominant weights of U_i, $i = 1, 2$, such that

$$\lambda = \lambda_1 + \lambda_2, \quad \lambda_0 = \lambda_{0,1} + \lambda_{0,2}. \tag{7.5.7}$$

Now we consider the sequence $d\lambda + \lambda_0$, $d \in \mathbb{N}$. Then

$$E_d = E_{d\lambda_1 + \lambda_{0,1}} \otimes E_{d\lambda_2 + \lambda_{0,2}}. \tag{7.5.8}$$

Since G_2 is equal rank, the nondegeneracy of λ with respect to θ is equivalent to the nondegeneracy of λ_1 with respect to θ_1. Then by Proposition 4.2.2, after taking the Mellin transform, we have

$$\mathcal{MI}_X(F_d, s) = [e(TX_2, \nabla^{TX_2})]^{\max 2} \dim E_{d\lambda_2 + \lambda_{0,2}} \mathcal{MI}_{X_1}(F_{d\lambda_1 + \lambda_{0,1}}, s). \tag{7.5.9}$$

Then

$$\mathcal{PI}_X(F_d) = [e(TX_2, \nabla^{TX_2})]^{\max 2} \dim E_{d\lambda_2 + \lambda_{0,2}} \mathcal{PI}_{X_1}(F_{d\lambda_1 + \lambda_{0,1}}). \tag{7.5.10}$$

Then we only need to evaluate $\mathcal{PI}_{X_1}(F_{d\lambda_1 + \lambda_{0,1}})$ explicitly, which has been dealt with in [Müller and Pfaff 2013a, Section 6].

7.6. *Asymptotic elliptic orbital integrals.*

Definition 7.6.1. A function $f(d)$ in d is called an exponential polynomial in d if it is a finite sum of the term $c_{j,s} e^{2\pi\sqrt{-1}sd} d^j$ with $j \in \mathbb{N}$, $s \in \mathbb{R}$, $c_{j,s} \in \mathbb{C}$. The largest $j \geq 0$ such that $c_{j,s} \neq 0$ in $f(d)$ is called the degree of $f(d)$.

We say that the oscillating term $e^{2\pi\sqrt{-1}sd}$ is nice if $s \in \mathbb{Q}$. We say that an exponential polynomial $f(d)$ in d is nice if all its oscillating terms are nice.

Remark 7.6.2. If $f(d)$ is a nice exponential polynomial in d, then there exists an $N_0 \in \mathbb{N}_{>0}$ such that the function $f(dN_0)$ is a polynomial in d.

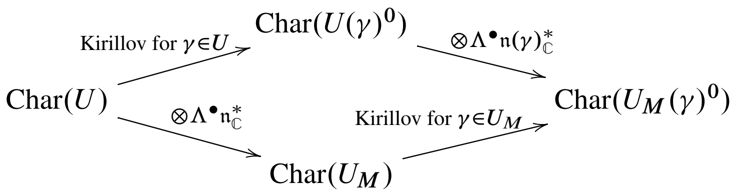
Note that by (5.4.18), $\varphi_\gamma(\sigma, \eta_\omega(d\lambda + \lambda_0))$ is an oscillating term in d, which is nice when $\gamma \in T$ is of finite order. The following theorem is a direct consequence of Theorem 7.2.10.

Theorem 7.6.3. *Suppose that λ is nondegenerate, and that $\gamma = k \in T$. Then, for sufficiently large d, $\mathcal{PE}_{X,\gamma}(F_d)$ is an exponential polynomial in d. Moreover, we have*

$$\mathcal{PE}_{X,\gamma}(F_d) = 2\pi c_G(\gamma) \cdot \sum_{\substack{\omega \in W_u \\ \sigma \in W^1(\gamma)}} \varepsilon(\omega) \varphi_\gamma(\sigma, \eta_\omega(d\lambda + \lambda_0)) \int_0^{|da_{\lambda,\omega} + b_{\lambda_0,\omega}|} P^\gamma_{\omega,\sigma,d\lambda+\lambda_0}(t)\, dt. \tag{7.6.1}$$

If we consider $G = \mathrm{Spin}(1, 2n+1)$, $n \geq 1$, as in [Fedosova 2015], then up to a constant, the exponential polynomial $\sum_{\sigma \in W^1(\gamma)} \varphi_\gamma(\sigma, \eta_\omega(d\lambda + \lambda_0)) P^\gamma_{\omega,\sigma,d\lambda+\lambda_0}(t)$ is just the one defined in [Fedosova 2015, Proposition 5.1]. This way, our results are compatible with her results in [Fedosova 2015, Theorem 1,1] for hyperbolic orbifolds.

Remark 7.6.4. Let $\mathrm{Char}(A)$ denote the character ring of the complex representations of a compact Lie group A. One key ingredient in (7.2.66) is an explicit decomposition of characters of U into characters of $U_M(\gamma)^0$. In the diagram below, we give two different ways of getting to this decomposition:



(7.6.2)

The formula in (7.2.66) is obtained by the computations along the lower path in (7.6.2). We also have the upper path, which is essentially the geometric localization formula obtained in Theorem 6.0.1.

We will use the same notation as in Section 6. The following theorem is a consequence of the geometric localization formula obtained in Theorem 6.0.1.

For $k \in T$, let $W^1_U(k) \subset W(U, T_U)$ be defined as in (5.4.14) with respect to $R^+(\mathfrak{u}, \mathfrak{t}_U)$. For $\sigma \in W^1_U(k)$, the term $\varphi^U_k(\sigma, d\lambda + \lambda_0)$ defined as in (6.0.8) is an oscillating term, which is nice if k is of finite order.

Theorem 7.6.5. *Suppose that $\gamma = k \in T$ is elliptic and that λ is nondegenerate with respect to θ. Then, for $\sigma \in W^1_U(k)$, $\sigma\lambda \in P_{++}(\widetilde{U}(k))$ is nondegenerate with respect to the Cartan involution θ on $\mathfrak{z}(k)$. For $d \in \mathbb{N}$, let $E^k_{\sigma,d}$ be the irreducible unitary representation of $\widetilde{U}(k)$ with highest weight $d\sigma\lambda + \sigma(\lambda_0 + \rho_{\mathfrak{u}}) - \rho_{\mathfrak{u}(k)}$. This way we get a sequence of flat vector bundles $\{F^k_{\sigma,d}\}_{d\in\mathbb{N}}$ on $X(k)$. Then, for sufficiently large d, we have*

$$\mathcal{P}\mathcal{E}_{X,\gamma}(F_d) = \sum_{\sigma \in W^1_U(k)} \varphi^U_k(\sigma, d\lambda + \lambda_0) \mathcal{P}\mathcal{I}_{X(k)}(F^k_{\sigma,d}). \tag{7.6.3}$$

Proof. The nondegeneracy of $\sigma\lambda$ ($\sigma \in W^1_U(k)$) follows easily from the nondegeneracy of λ and the definition of $W^1_U(k)$. For proving this theorem, we only need to prove (7.6.3). Actually, by Theorem 6.0.1, for $t > 0$, we get

$$\mathcal{E}_{X,\gamma}(F_d, t) = \sum_{\sigma \in W^1_U(k)} \varphi^U_k(\sigma, d\lambda + \lambda_0) \mathcal{I}_{X(k)}(F^k_{\sigma,d}, t), \tag{7.6.4}$$

Then (7.6.3) follows from the linearity of Mellin transform. □

8. A proof of Theorem 1.0.2

In this section, we complete the proof of Theorem 1.0.2; then Theorem 1.0.1 (and Theorem 1.0.1′) follows as a consequence. We assume that G is a connected linear real reductive Lie group with $\delta(G) = 1$ and compact center, so that U is a compact Lie group.

8.1. ***A lower bound for the Hodge Laplacian on X.*** We use the notation from Section 4. Recall that $e_1, \dots, e_m$ is an orthogonal basis of TX or $\mathfrak{p}$. Put

$$C^{\mathfrak{g},H} = -\sum_{j=1}^{m} e_j^2 \in U\mathfrak{g}. \tag{8.1.1}$$

Let $C^{\mathfrak{g},H,E}$ be its action on E via ρ^E. Then

$$C^{\mathfrak{g},E} = C^{\mathfrak{g},H,E} + C^{\mathfrak{k},E}. \tag{8.1.2}$$

Let $\Delta^{H,X}$ be the Bochner–Laplace operator on the bundle $\Lambda^\bullet(T^*X) \otimes F$ associated with the unitary connection $\nabla^{\Lambda^\bullet(T^*X)\otimes F,u}$. Put

$$\Theta(F) = \tfrac{1}{4}S^X - \tfrac{1}{8}\langle R^{TX}(e_i, e_j)e_k, e_\ell\rangle c(e_i)c(e_j)\hat{c}(e_k)\hat{c}(e_\ell) \\ - C^{\mathfrak{g},H,E} + \tfrac{1}{2}\big(c(e_i)c(e_j) - \hat{c}(e_i)\hat{c}(e_j)\big)R^F(e_i, e_j), \tag{8.1.3}$$

where R^F is the curvature of the unitary connection ∇^F on F.

Then $\Theta(F)$ is a self-adjoint section of $\mathrm{End}(\Lambda^\bullet(T^*X) \otimes F)$, which is parallel with respect to $\nabla^{\Lambda^\bullet(T^*X)\otimes F,u}$. Equivalently, $\Theta(F)$ is an element in $\mathrm{End}(\Lambda^\bullet(\mathfrak{p}^*) \otimes E)$ which commutes with the K-action. By [Bismut et al. 2017, equation (8.39)], we have

$$\boldsymbol{D}^{X,F,2} = -\Delta^{H,X} + \Theta(F). \tag{8.1.4}$$

Then, for $s \in \Omega^\bullet_c(X, F)$, we have

$$\langle \boldsymbol{D}^{X,F,2}s, s\rangle_{L_2} \geq \langle \Theta(F)s, s\rangle_{L_2}. \tag{8.1.5}$$

Let $\Delta^{H,X,i}$ denote the Bochner–Laplace operator acting on $\Omega^i(X, F)$, and let $p_t^{H,i}(x, x')$ be the kernel of $\exp(t\Delta^{H,X,i}/2)$ on X with respect to dx'. We will denote by $p_t^{H,i}(g) \in \mathrm{End}(\Lambda^i(\mathfrak{p}^*) \otimes E)$ its lift to G explained in Section 3.2. Let Δ_0^X be the scalar Laplacian on X with the heat kernel $p_t^{X,0}$.

Let $\|p_t^{H,i}(g)\|$ be the operator norm of $p_t^{H,i}(g)$ in $\mathrm{End}(\Lambda^i(\mathfrak{p}^*) \otimes E)$. By [Müller and Pfaff 2013b, Proposition 3.1], if $g \in G$; then

$$\|p_t^{H,i}(g)\| \leq p_t^{X,0}(g). \tag{8.1.6}$$

Let p_t^H be the kernel of $\exp(t\Delta^{H,X}/2)$, then

$$p_t^H = \bigoplus_{i=1}^{p} p_t^{H,i}. \tag{8.1.7}$$

Let $q_t^{X,F}$ be the heat kernel associated with $\boldsymbol{D}^{X,F,2}/2$, by (8.1.4), for $g \in X$,

$$q_t^{X,F}(g) = \exp\Big(-\frac{t\Theta(F)}{2}\Big)p_t^H(g). \tag{8.1.8}$$

Recall that $P_{++}(U)$ is the set of dominant weights of U with respect to $R^+(\mathfrak{u}, \mathfrak{t}_U)$ defined in Section 5.3. As in Section 7.3, we fix $\lambda, \lambda_0 \in P_{++}(U)$ such that λ is nondegenerate with respect to θ. Recall that, for $d \in \mathbb{N}$, (E_d, ρ^{E_d}) is the irreducible unitary representation of U with highest weight $d\lambda + \lambda_0$, which extends uniquely to a representation of G. By [Bismut et al. 2011, Théorème 3.2; 2017, Theorem 4.4, Remark 4.5; Müller and Pfaff 2013a, Proposition 7.5], there exist $c>0$, $C>0$ such that, for $d\in\mathbb{N}$,

$$\Theta(F_d)\geq cd^2-C, \tag{8.1.9}$$

where the estimate d^2 comes from the positive operator $C^{\mathfrak{g},H,E_d}$. By (8.1.4), (8.1.5), (8.1.9), we get

$$\boldsymbol{D}^{X,F_d,2}\geq cd^2-C. \tag{8.1.10}$$

Lemma 8.1.1. *There exists $d_0\in\mathbb{N}$ and $c_0>0$ such that if $d\geq d_0$, $g\in G$,*

$$\|q_t^{X,F_d}(g)\|\leq e^{-c_0d^2t}p_t^{X,0}(g). \tag{8.1.11}$$

Proof. By (8.1.9), there exist $d_0\in\mathbb{N}$, $c'>0$ such that if $d\geq d_0$,

$$\Theta(F_d)\geq c'd^2. \tag{8.1.12}$$

Then if $t>0$,

$$\left\|\exp\left(-\frac{t\Theta(F_d)}{2}\right)\right\|\leq e^{-\frac{1}{2}c'd^2t}. \tag{8.1.13}$$

By (8.1.6), (8.1.7), (8.1.8), (8.1.13), we get (8.1.11). □

The locally symmetric orbifold Z is defined as $\Gamma\backslash X$, where Γ is a cocompact discrete subgroup of G. For $\gamma\in\Gamma$, the number $m_\gamma\geq 0$ is given by (3.3.3), which only depends on the conjugacy class of γ (in G or Γ). Recall that $E[\Gamma]$ is the finite set of elliptic conjugacy classes in Γ.

For $t>0$, $x\in X$, $\gamma\in\Gamma$, set

$$v_t(F_d,\gamma,x)=\mathrm{Tr}_s^{\Lambda^\bullet(T^*X)\otimes F_d}\left[\left(N^{\Lambda^\bullet(T^*X)}-\frac{m}{2}\right)q_t^{X,F_d}(x,\gamma(x))\gamma\right]. \tag{8.1.14}$$

Then by Lemma 8.1.1, we have the following result.

Lemma 8.1.2. *There exist $C_0>0$, $c_0>0$ such that if d is large enough, for $t>0$, $x\in X$, $\gamma\in\Gamma$,*

$$|v_t(F_d,\gamma,x)|\leq C_0(\dim E_d)e^{-c_0d^2t}p_t^{X,0}(x,\gamma(x)). \tag{8.1.15}$$

Set

$$m_\Gamma=\inf_{[\gamma]\in[\Gamma]-E[\Gamma]}m_\gamma. \tag{8.1.16}$$

By [Liu 2018, Proposition 1.8.5], $m_\Gamma>0$.

Proposition 8.1.3. *There exist constants $C>0$, $c>0$ such that if $x\in X$, $t\in\,]0,1]$, then*

$$\sum_{\gamma\in\Gamma,\gamma\ \textit{nonelliptic}}p_t^{X,0}(x,\gamma(x))\leq C\exp\left(-\frac{c}{t}\right). \tag{8.1.17}$$

Proof. By [Donnelly 1979, Theorem 3.3], there exists $C_0>0$ such that when $0<t\leq 1$,

$$p_t^{X,0}(x,x')\leq C_0t^{-\frac{m}{2}}\exp\left(-\frac{d^2(x,x')}{4t}\right). \tag{8.1.18}$$

By [Liu 2018, Lemma 1.8.6], there exist $c>0$, $C>0$ such that for $R>0$, $x\in X$,

$$\#\{\gamma\in\Gamma \mid \gamma \text{ nonelliptic}, d_\gamma(x)\leq R\}\leq C\exp(cR). \tag{8.1.19}$$

By (8.1.16), (8.1.18), (8.1.19), and using the same arguments as in the proof of [Müller and Pfaff 2013b, Proposition 3.2], we get (8.1.17). □

8.2. ***A proof of Theorem 1.0.2.*** In this subsection, we complete our proof of Theorem 1.0.2. Note that every elliptic element $\gamma\in\Gamma$ is of finite order, then part (2) of Theorem 1.0.2 is an easy consequence of Theorem 7.6.5. We only need to prove part (1). We restate it as follows.

Proposition 8.2.1. *Let $\Gamma\subset G$ be a cocompact discrete subgroup and set $Z=\Gamma\backslash X$. There exists $c>0$ such that, for d large enough,*

$$\mathcal{T}(Z,F_d)=\frac{\mathrm{Vol}(Z)}{|S|}\mathcal{P}\mathcal{I}_X(F_d)+\sum_{[\gamma]\in E^+[\Gamma]}\frac{\mathrm{Vol}(\Gamma\cap Z(\gamma)\backslash X(\gamma))}{|S(\gamma)|}\mathcal{P}\mathcal{E}_{X,\gamma}(F_d)+\mathcal{O}(e^{-cd}), \tag{8.2.1}$$

where $E^+[\Gamma]=E^+[\Gamma]\backslash\{[1]\}$ is the finite set of nontrivial elliptic classes in $[\Gamma]$.

Proof. By (8.1.10), we have

$$\boldsymbol{D}^{Z,F_d,2}\geq cd^2-C. \tag{8.2.2}$$

Then if d is large enough, we have

$$H^\bullet(Z,F_d)=0. \tag{8.2.3}$$

Then $\mathcal{T}(Z,F_d)$ can be computed using (2.2.15).

As in (2.2.12), for $t>0$, set

$$b(F_d,t)=\Big(1+2t\frac{\partial}{\partial t}\Big)\mathrm{Tr_s}\Big[\Big(N^{\Lambda^\bullet(T^*Z)}-\frac{m}{2}\Big)\exp\Big(-\frac{t\boldsymbol{D}^{Z,F_d,2}}{2}\Big)\Big]. \tag{8.2.4}$$

As in [Bismut et al. 2017, Section 7.2], by (8.2.2), there exist constants $\tilde{c}>0$, $\widetilde{C}>0$ such that, for d large enough and for $t>\frac{1}{d}$,

$$|b(F_d,t)|\leq\widetilde{C}\exp(-\tilde{c}d-\tilde{c}t). \tag{8.2.5}$$

By (2.2.15), we have

$$\mathcal{T}(Z,F_d)=-\int_0^{+\infty}b(F_d,t)\,\frac{dt}{t}. \tag{8.2.6}$$

We rewrite it as

$$\mathcal{T}(Z,F_d)=-\int_{1/d}^{+\infty}b(F_d,t)\,\frac{dt}{t}-\int_0^d b\Big(\frac{F_d,t}{d^2}\Big)\,\frac{dt}{t}. \tag{8.2.7}$$

By (8.2.5), there exists $c>0$ such that, for d large enough,

$$\int_{1/d}^{+\infty}b(F_d,t)\,\frac{dt}{t}=\mathcal{O}(e^{-cd}). \tag{8.2.8}$$

By (3.5.1), (8.1.14), (8.2.4), we get

$$b(F_d,t) = \Big(1+2t\frac{\partial}{\partial t}\Big)\int_Z \frac{1}{|S|}\sum_{\gamma\in\Gamma} v_t(F_d,\gamma,z)\,dz. \tag{8.2.9}$$

We split the sum in (8.2.9) into two parts,

$$\sum_{\gamma\in\Gamma,\gamma \text{ elliptic}} + \sum_{\gamma\in\Gamma,\gamma \text{ nonelliptic}}, \tag{8.2.10}$$

so that we write

$$b(F_d,t) = b_{\text{elliptic}}(F_d,t) + b_{\text{nonelliptic}}(F_d,t). \tag{8.2.11}$$

Similar to Selberg's trace formula in Section 3.5, we get

$$b_{\text{elliptic}}(F_d,t) = \sum_{[\gamma]\in E[\Gamma]} \frac{\mathrm{Vol}(\Gamma\cap Z(\gamma)\backslash X(\gamma))}{|S(\gamma)|}\Big(1+2t\frac{\partial}{\partial t}\Big)\mathcal{E}_{X,\gamma}(F_d,t). \tag{8.2.12}$$

By (7.4.2) and (7.6.4), the terms in $\mathcal{E}_{X,\gamma}(F_d,t)$ are of the form

$$t^{-j+\frac{1}{2}}\exp(-2\pi^2 t(da'+b')^2)Q(d), \tag{8.2.13}$$

where $Q(d)$ is a nice exponential polynomial in d, and $a',b'\in\mathbb{R}$ with $a'\neq 0$ due to the nondegeneracy of λ. By (8.2.13), there exists $c>0$ such that, for d large enough,

$$\int_0^d b_{\text{elliptic}}\Big(F_d,\frac{t}{d^2}\Big)\frac{dt}{t} = \int_0^{+\infty} b_{\text{elliptic}}(F_d,t)\frac{dt}{t} + \mathcal{O}(e^{-cd}). \tag{8.2.14}$$

Using Proposition 7.1.1 and by (8.2.13), we get

$$\mathcal{P}\mathcal{E}_{X,\gamma}(F_d) = -\int_0^{+\infty}\Big(1+2t\frac{\partial}{\partial t}\Big)\mathcal{E}_{X,\gamma}(F_d,t)\frac{dt}{t}. \tag{8.2.15}$$

Now we consider the contribution from the nonelliptic elements. If $x\in X$, put

$$h_t(F_d,x) = \frac{1}{|S|}\sum_{\gamma\in\Gamma,\gamma \text{ nonelliptic}} v_t(F_d,\gamma,x). \tag{8.2.16}$$

Then

$$b_{\text{nonelliptic}}(F_d,t) = \Big(1+2t\frac{\partial}{\partial t}\Big)\int_Z h_t(F_d,z)\,dz. \tag{8.2.17}$$

Now we prove the following uniform estimates for $x\in X$:

$$\int_0^d \Big(1+2t\frac{\partial}{\partial t}\Big)h_{t/d^2}(F_d,x)\frac{dt}{t} = \mathcal{O}(e^{-cd}). \tag{8.2.18}$$

Indeed, using Lemma 8.1.2 and Proposition 8.1.3, there exist $C>0$, $c'>0$, $c''>0$ such that if d is large enough, $0<t\leq d$, then

$$|h_{t/d^2}(F_d,x)| \leq C\dim(E_d)e^{-c't}\exp\Big(-\frac{c''d^2}{t}\Big). \tag{8.2.19}$$

Recall that $\dim E_d$ is a polynomial in d. Then by (8.2.19), we have

$$\begin{aligned}\left|\int_0^1 h_{t/d^2}(F_d,x)\,\frac{dt}{t}\right| &\le Ce^{-c''d^2/2}\dim(E_d)\int_0^1 e^{-c''d^2/2t}\,\frac{dt}{t}=\mathcal{O}(e^{-cd}),\\ \left|\int_1^d h_{t/d^2}(F_d,x)\frac{dt}{t}\right| &\le Ce^{-c''d}\dim(E_d)\int_1^d e^{-c't}\,\frac{dt}{t}=\mathcal{O}(e^{-cd}).\end{aligned} \tag{8.2.20}$$

By (8.2.19)–(8.2.20), we get (8.2.18).

At last, we assembly together (8.2.7), (8.2.8), (8.2.11), (8.2.14)–(8.2.18), we get exactly (8.2.1). □

Note that since $\mathcal{T}(Z,F_d)$ is always a real number, (8.2.1) still holds if we take the real part of $\mathcal{PE}_{X,\gamma}(F_d)$ instead.

Acknowledgments

I would like to thank Prof. Jean-Michel Bismut and Prof. Werner Müller for encouraging me to work on this subject, and for many useful discussions. I also thank Dr. Taiwang Deng for educating me about the cohomology of arithmetic groups, and Dr. Ksenia Fedosova for explanations of her results on the hyperbolic case.

This work was carried out during my stay at the Max Planck Institute for Mathematics (MPIM) in Bonn. I also want to express my sincere gratitude to MPIM for providing such a nice research environment. Last but not least, I also would like to thank the anonymous referee(s) for the suggestions and comments that greatly improved the paper.

References

[Adem et al. 2007] A. Adem, J. Leida, and Y. Ruan, *Orbifolds and stringy topology*, Cambridge Tracts in Math. **171**, Cambridge Univ. Press, 2007. MR Zbl

[Bergeron and Venkatesh 2013] N. Bergeron and A. Venkatesh, "The asymptotic growth of torsion homology for arithmetic groups", *J. Inst. Math. Jussieu* **12**:2 (2013), 391–447. MR Zbl

[Berline et al. 1992] N. Berline, E. Getzler, and M. Vergne, *Heat kernels and Dirac operators*, Grundlehren der Math. Wissenschaften **298**, Springer, 1992. MR Zbl

[Bismut 2011] J.-M. Bismut, *Hypoelliptic Laplacian and orbital integrals*, Ann. of Math. Stud. **177**, Princeton Univ. Press, 2011. MR Zbl

[Bismut and Lott 1995] J.-M. Bismut and J. Lott, "Flat vector bundles, direct images and higher real analytic torsion", *J. Amer. Math. Soc.* **8**:2 (1995), 291–363. MR Zbl

[Bismut and Shen 2022] J.-M. Bismut and S. Shen, "Geometric orbital integrals and the center of the enveloping algebra", *Compos. Math.* **158**:6 (2022), 1189–1253. MR Zbl

[Bismut and Zhang 1991] J.-M. Bismut and W. Zhang, "Métriques de Reidemeister et métriques de Ray–Singer sur le déterminant de la cohomologie d'un fibré plat: une extension d'un résultat de Cheeger et Müller", *C. R. Acad. Sci. Paris Sér. I Math.* **313**:11 (1991), 775–782. MR Zbl

[Bismut and Zhang 1992] J.-M. Bismut and W. Zhang, *An extension of a theorem by Cheeger and Müller*, Astérisque **205**, Soc. Math. France, Paris, 1992. MR Zbl

[Bismut et al. 2011] J.-M. Bismut, X. Ma, and W. Zhang, "Opérateurs de Toeplitz et torsion analytique asymptotique", *C. R. Math. Acad. Sci. Paris* **349**:17-18 (2011), 977–981. MR Zbl

[Bismut et al. 2017] J.-M. Bismut, X. Ma, and W. Zhang, "Asymptotic torsion and Toeplitz operators", *J. Inst. Math. Jussieu* **16**:2 (2017), 223–349. MR Zbl

[Bouaziz 1987] A. Bouaziz, "Sur les caractères des groupes de Lie réductifs non connexes", *J. Funct. Anal.* **70**:1 (1987), 1–79. MR Zbl

[Bröcker and tom Dieck 1985] T. Bröcker and T. tom Dieck, *Representations of compact Lie groups*, Grad. Texts in Math. **98**, Springer, 1985. MR Zbl

[Cheeger 1979] J. Cheeger, "Analytic torsion and the heat equation", *Ann. of Math.* (2) **109**:2 (1979), 259–322. MR Zbl

[Dai and Yu 2017] X. Dai and J. Yu, "Comparison between two analytic torsions on orbifolds", *Math. Z.* **285**:3-4 (2017), 1269–1282. MR Zbl

[Donnelly 1979] H. Donnelly, "Asymptotic expansions for the compact quotients of properly discontinuous group actions", *Illinois J. Math.* **23**:3 (1979), 485–496. MR Zbl

[Duflo et al. 1984] M. Duflo, G. Heckman, and M. Vergne, "Projection d'orbites, formule de Kirillov et formule de Blattner", pp. 65–128 in *Harmonic analysis on Lie groups and symmetric spaces* (Kleebach, Germany, 1983), edited by M. Duflo et al., Mém. Soc. Math. France (N.S.) **15**, Soc. Math. France, Paris, 1984. MR Zbl

[Eberlein 1996] P. B. Eberlein, *Geometry of nonpositively curved manifolds*, Univ. Chicago Press, 1996. MR Zbl

[Fedosova 2015] K. Fedosova, "On the asymptotics of the analytic torsion for compact hyperbolic orbifolds", preprint, 2015. arXiv 1511.04281

[Finski 2018] S. Finski, "On the full asymptotics of analytic torsion", *J. Funct. Anal.* **275**:12 (2018), 3457–3503. MR Zbl

[Helgason 1978] S. Helgason, *Differential geometry*, *Lie groups*, *and symmetric spaces*, Pure Appl. Math. **80**, Academic Press, New York, 1978. MR Zbl

[Kawasaki 1978] T. Kawasaki, "The signature theorem for V-manifolds", *Topology* **17**:1 (1978), 75–83. MR Zbl

[Kawasaki 1979] T. Kawasaki, "The Riemann–Roch theorem for complex V-manifolds", *Osaka Math. J.* **16**:1 (1979), 151–159. MR Zbl

[Knapp 1986] A. W. Knapp, *Representation theory of semisimple groups: an overview based on examples*, Princeton Math. Ser. **36**, Princeton Univ. Press, 1986. MR Zbl

[Knapp 2002] A. W. Knapp, *Lie groups beyond an introduction*, 2nd ed., Progr. Math. **140**, Birkhäuser, Boston, 2002. MR Zbl

[Kostant 1961] B. Kostant, "Lie algebra cohomology and the generalized Borel–Weil theorem", *Ann. of Math.* (2) **74**:2 (1961), 329–387. MR Zbl

[Liu 2018] B. Liu, *Hypoelliptic Laplacian and twisted trace formula*, Ph.D. thesis, Université Paris-Saclay, 2018, available at https://theses.hal.science/tel-01841334.

[Liu 2019] B. Liu, "Hypoelliptic Laplacian and twisted trace formula", *C. R. Math. Acad. Sci. Paris* **357**:1 (2019), 74–83. MR Zbl

[Liu 2021] B. Liu, "Asymptotic equivariant real analytic torsions for compact locally symmetric spaces", *J. Funct. Anal.* **281**:7 (2021), art. id. 109117. MR Zbl

[Lott 1992] J. Lott, "Heat kernels on covering spaces and topological invariants", *J. Differential Geom.* **35**:2 (1992), 471–510. MR Zbl

[Ma 2005] X. Ma, "Orbifolds and analytic torsions", *Trans. Amer. Math. Soc.* **357**:6 (2005), 2205–2233. MR Zbl

[Ma 2019] X. Ma, "Geometric hypoelliptic Laplacian and orbital integrals (after Bismut, Lebeau and Shen)", exposé 1130, pp. 333–389 in *Séminaire Bourbaki*, 2016/2017, Astérisque **407**, Soc. Math. France, Paris, 2019. MR Zbl

[Mathai 1992] V. Mathai, "L^2-analytic torsion", *J. Funct. Anal.* **107**:2 (1992), 369–386. MR Zbl

[Matz and Müller 2023] J. Matz and W. Müller, "Analytic torsion for arithmetic locally symmetric manifolds and approximation of L^2 torsion", *J. Funct. Anal.* **284**:1 (2023), art. id. 109727. Zbl

[Moscovici and Stanton 1991] H. Moscovici and R. J. Stanton, "R-torsion and zeta functions for locally symmetric manifolds", *Invent. Math.* **105**:1 (1991), 185–216. MR Zbl

[Müller 1978] W. Müller, "Analytic torsion and R-torsion of Riemannian manifolds", *Adv. Math.* **28**:3 (1978), 233–305. MR Zbl

[Müller 2012] W. Müller, "The asymptotics of the Ray–Singer analytic torsion of hyperbolic 3-manifolds", pp. 317–352 in *Metric and differential geometry* (Beijing, 2009), edited by X. Dai and X. Rong, Progr. Math. **297**, Birkhäuser, Basel, 2012. MR Zbl

[Müller and Pfaff 2013a] W. Müller and J. Pfaff, "Analytic torsion and L^2-torsion of compact locally symmetric manifolds", *J. Differential Geom.* **95**:1 (2013), 71–119. MR Zbl

[Müller and Pfaff 2013b] W. Müller and J. Pfaff, "On the asymptotics of the Ray–Singer analytic torsion for compact hyperbolic manifolds", *Int. Math. Res. Not.* **2013**:13 (2013), 2945–2983. MR Zbl

[Ray and Singer 1971] D. B. Ray and I. M. Singer, "R-torsion and the Laplacian on Riemannian manifolds", *Adv. Math.* **7** (1971), 145–210. MR Zbl

[Ray and Singer 1973] D. B. Ray and I. M. Singer, "Analytic torsion", pp. 167–181 in *Partial differential equations* (Berkeley, CA, 1971), edited by D. C. Spencer, Proc. Sympos. Pure Math. **23**, Amer. Math. Soc., Providence, RI, 1973. MR Zbl

[Satake 1956] I. Satake, "On a generalization of the notion of manifold", *Proc. Nat. Acad. Sci. U.S.A.* **42** (1956), 359–363. MR Zbl

[Satake 1957] I. Satake, "The Gauss–Bonnet theorem for V-manifolds", *J. Math. Soc. Japan* **9**:4 (1957), 464–492. MR Zbl

[Selberg 1960] A. Selberg, "On discontinuous groups in higher-dimensional symmetric spaces", pp. 147–164 in *Contributions to function theory* (Bombay, 1960), edited by K. Chandrasekharan, Tata Inst. Fund. Res., Bombay, 1960. MR Zbl

[Shen 2018] S. Shen, "Analytic torsion, dynamical zeta functions, and the Fried conjecture", *Anal. PDE* **11**:1 (2018), 1–74. MR Zbl

[Shen and Yu 2022] S. Shen and J. Yu, "Flat vector bundles and analytic torsion on orbifolds", *Comm. Anal. Geom.* **30**:3 (2022), 575–656. MR Zbl

[Warner 1972] G. Warner, *Harmonic analysis on semi-simple Lie groups, I*, Grundlehren der Math. Wissenschaften **188**, Springer, 1972. MR Zbl

Received 25 Jul 2021. Revised 16 Aug 2022. Accepted 23 Sep 2022.

BINGXIAO LIU: bingxiao.liu@uni-koeln.de
Mathematisches Institut, Universität zu Köln, Köln, Germany

ANALYSIS AND PDE
Vol. 17 (2024), No. 4, pp. 1331–1375
DOI: 10.2140/apde.2024.17.1331

THE LANDAU EQUATION AS A GRADIENT FLOW

JOSÉ A. CARRILLO, MATIAS G. DELGADINO, LAURENT DESVILLETTES AND JEREMY S.-H. WU

We propose a gradient flow perspective to the spatially homogeneous Landau equation for soft potentials. We construct a tailored metric on the space of probability measures based on the entropy dissipation of the Landau equation. Under this metric, the Landau equation can be characterized as the gradient flow of the Boltzmann entropy. In particular, we characterize the dynamics of the PDE through a functional inequality which is usually referred as the energy dissipation inequality (EDI). Furthermore, analogous to the optimal transportation setting, we show that this interpretation can be used in a minimizing movement scheme to construct solutions to a regularized Landau equation.

1. Introduction

The Landau equation is an important partial differential equation in kinetic theory. It gives a description of colliding particles in plasma physics [Lifshitz and Pitaevskiĭ 1981], and it can be formally derived as a limit of the Boltzmann equation where grazing collisions are dominant [Degond and Lucquin-Desreux 1992; Villani 1998a]. Similar to the Boltzmann equation (see [Boblylev et al. 2013] for a consistency result and related derivation issues), the rigorous derivation of the Landau equation from particle dynamics is still a huge challenge. For a spatially homogeneous density of particles $f = f_t(v)$ for $t \in (0, \infty)$, $v \in \mathbb{R}^d$, the homogeneous Landau equation reads

$$\partial_t f(v) = \nabla_v \cdot \left(f(v) \int_{\mathbb{R}^d} |v - v_*|^{2+\gamma}\, \Pi[v - v_*](\nabla_v \log f(v) - \nabla_{v_*} \log f(v_*)) f(v_*)\, dv_* \right). \tag{1}$$

For notational convenience, we sometimes abbreviate $f = f_t(v)$ and $f_* = f_t(v_*)$. We also denote the differentiations by $\nabla = \nabla_v$ and $\nabla_* = \nabla_{v_*}$. The physically relevant parameters are usually $d = 2, 3$ and $\gamma \geq -d-1$ with $\Pi[z] = I - (z \otimes z)/|z|^2$ being the projection matrix onto $\{z\}^\perp$. In this paper, for simplicity we will focus in the case $d = 3$ and vary the weight parameter γ, although most of our results are valid in arbitrary dimension. The regime $0 < \gamma < 1$ corresponds to the so-called *hard potentials*, while $\gamma < 0$ corresponds to the *soft potentials* with a further classification of $-2 \leq \gamma < 0$ as the moderately soft potentials and $-4 \leq \gamma < -2$ as the very soft potentials. The particular instances of $\gamma = 0$ and $\gamma = -d$ are known as the Maxwellian and Coulomb cases respectively.

The purpose of this work is to propose a new perspective inspired from gradient flows for weak solutions to (1), which is in analogy with the relationship of the heat equation and the 2-Wasserstein metric; see [Jordan et al. 1998; Ambrosio et al. 2008]. Our main result is inspired by and extends [Erbar 2023]. There, he establishes the gradient flow perspective for the closely related spatially homogeneous Boltzmann

MSC2020: primary 35Q82, 49Q22; secondary 82C40.
Keywords: Landau equation, gradient flow, steepest descent.

equation with bounded collision kernels ($\gamma = 0$) which we perform in the case of Landau for $\gamma \in (-3, 0]$ (see Theorem 12). One of the fundamental steps is to symmetrize the right-hand side of (1). More specifically, if we consider a test function $\phi \in C_c^\infty(\mathbb{R}^d)$, we can formally characterize the equation by

$$\frac{d}{dt}\int_{\mathbb{R}^d} \phi f\,dv = -\frac{1}{2}\iint_{\mathbb{R}^{2d}} ff_*|v-v_*|^{2+\gamma}(\nabla\phi - \nabla_*\phi_*)\cdot\Pi[v-v_*](\nabla\log f - \nabla_*\log f_*)\,dv_*\,dv, \quad (2)$$

where the change of variables $v \leftrightarrow v_*$ has been exploited. Building our analogy with the heat equation and the 2-Wasserstein distance, we define an appropriate gradient

$$\tilde\nabla\phi := |v-v_*|^{1+\gamma/2}\Pi[v-v_*](\nabla\phi - \nabla_*\phi_*),$$

so that (2) now looks like

$$\frac{d}{dt}\int_{\mathbb{R}^d}\phi f\,dv = -\frac{1}{2}\iint_{\mathbb{R}^{2d}} ff_*\tilde\nabla\phi\cdot\tilde\nabla\log f\,dv_*\,dv,$$

noting that $\Pi^2 = \Pi$. To highlight the use of this interpretation, we notice that $\tilde\nabla\phi = 0$, when we choose as test functions $\phi = 1$, v_i, $|v|^2$ for $i = 1, \ldots, d$, which immediately shows that formally the equation conserves mass, momentum and energy. The action functional defining the Landau metric mimics the Benamou–Brenier formula [2000] for the 2-Wasserstein distance; see [Dolbeault et al. 2009; Erbar 2014; Erbar and Maas 2014] for other distances defined analogously for nonlinear and nonlocal mobilities. In fact, the Landau metric is built by considering a minimizing action principle over curves that are solutions to the appropriate continuity equation, that is,

$$d_L(f,g) := \min_{\substack{\partial_t\mu + \tilde\nabla\cdot(V\mu\mu_*)/2 = 0\\ \mu_0 = f,\ \mu_1 = g}} \left\{\frac{1}{2}\int_0^1\iint_{\mathbb{R}^{2d}} |V|^2\,d\mu(v)\,d\mu(v_*)\,dt\right\}, \quad (3)$$

where the $\tilde\nabla\cdot$ is the appropriate divergence; the formal adjoint to the appropriate gradient (see Section 2.1).

Also, we notice that analogously to the heat equation, written as the continuity equation $\partial_t f = \nabla\cdot(f\nabla\log f)$, the Landau equation can be formally rewritten as

$$\partial_t f = \tfrac{1}{2}\tilde\nabla\cdot(ff_*\tilde\nabla\log f),$$

equivalent to the continuity equation with nonlocal velocity field given by

$$\begin{cases}\partial_t f + \nabla\cdot(U(f)f) = 0,\\ U(f) := -\int_{\mathbb{R}^d}|v-v_*|^{2+\gamma}\Pi[v-v_*](\nabla\log f - \nabla_*\log f_*)f_*\,dv_*.\end{cases} \quad (4)$$

This is a direct way to write (1) in the form of a continuity equation. Considering the evolution of Boltzmann entropy we formally obtain

$$\frac{d}{dt}\int_{\mathbb{R}^d} f\log f\,dv =: -D(f_t) = -\frac{1}{2}\iint_{\mathbb{R}^{2d}}|\tilde\nabla\log f|^2 ff_*\,dv_*\,dv \le 0. \quad (5)$$

In physical terms this is referred to as the *entropy dissipation* (referred to as entropy production in the physics literature from defining $\mathcal{H}$ with a minus sign) since it formally shows that the entropy functional

$$\mathcal{H}[f] := \int_{\mathbb{R}^d} f\log f\,dv$$

is nonincreasing along the dynamics of the Landau equation. Moreover, by integrating (5) in time one formally obtains

$$\mathcal{H}[f_t]+\int_0^t D(f_s)\,ds=\mathcal{H}[f_0]. \tag{6}$$

Villani [1998a] introduced the notion of H-solution, which captures this formal property. Motivated by the physical considerations of certain conserved quantities and entropy dissipation, H-solutions provided a step towards well-posedness of the Landau equation in the soft potential case. One advantage to this approach is that it avoids assuming that the solutions belongs to $L^p(\mathbb{R}^3)$ for $p>1$. For moderately soft potentials, the propagation of L^p norms is proven and this is enough to make sense of classical weak solutions [Wu 2014]. In the very soft potential case, there is no longer a guarantee of L^p propagation due to the singularity of the weight. We refer to [Desvillettes 2015, Section 1.2] for a heuristic description of this difficulty.

Similar to H-solutions our approach will also be based on the entropy dissipation (6). Following De Giorgi's minimizing movement ideas [Ambrosio 1995; Ambrosio et al. 2008], we characterize the Landau equation by its associated energy dissipation inequality. More specifically, we show that weak solutions to (1) with initial data f_0 are completely determined by the functional inequality

$$\mathcal{H}[f_t]+\frac{1}{2}\int_0^t |\dot f|^2_{d_L}(s)\,ds+\frac{1}{2}\int_0^t D(f_s)\,ds\le\mathcal{H}[f_0]\quad\text{for a.e. every } t>0,$$

where $|\dot f|^2_{d_L}(s)$ stands for the metric derivative associated to the Landau metric defined above. Our analysis is also largely inspired by Erbar's approach [2023] in viewing the Boltzmann equation as a gradient flow and recent numerical simulations of the homogeneous Landau equation in [Carrillo et al. 2020] based on a regularized version of (4). In contrast with the classical 2-Wasserstein metric, one of the main features of the Landau equation (1) and metric (3) is that they are nonlocal. To be precise, gradient flow theory has been successfully applied to the study of many nonlocal PDEs [Carrillo et al. 2010; 2012; Blanchet et al. 2008] by viewing them as gradient flows of appropriate energy functionals with respect to the 2-Wasserstein metric. The novelty in this work is the construction of the *nonlocal* metric d_L with respect to which (1) can be viewed as the gradient flow of $\mathcal{H}$. Hence, the convergence analysis usually relying on convexity and lower-semicontinuity needs to be adapted to deal with the nonlocality of this equation. In particular, our characterization Theorem 12 is based in using (expected) a priori estimates to deal with the nonlocality through appropriate bounds.

On the other hand, the state of the art related to the uniqueness for the Landau equation depends on the range of values γ may take. In the cases of hard potentials or Maxwellian, the uniqueness theory is very well understood due to Villani and the third author [Desvillettes and Villani 2000a; 2000b; Villani 1998b]. In the soft potential case, one of the first major contributions to the general theory of the spatially inhomogeneous Landau equation ($\gamma\ge-3$) was the global existence and uniqueness result of [Guo 2002]. This result was achieved in a perturbative framework with high regularity assumptions on the initial data. Through probabilistic arguments, the next major improvement to uniqueness for $\gamma\in(-3,0)$ came from [Fournier and Guérin 2009]. Their result established uniqueness in a class of solutions that shrinks as γ decreases towards -3, as more L^p and moments assumptions are needed. In their proof, uniqueness is shown by proving stability with respect to the 2-Wasserstein metric.

Still lots of open questions for the soft potential case remain. In particular, a fundamental question like uniqueness for the Coulomb case is unresolved. To tackle this and other problems an array of novel methods have been employed. Here is an incomplete sample of the contributions made in this direction which highlight the difficulties of the soft potential case [Desvillettes and Villani 2000a; 2000b; Alexandre et al. 2015; Carrapatoso and Mischler 2017; Carrapatoso et al. 2017; Wu 2014; Gualdani and Zamponi 2017, 2018a; 2018b; Gualdani and Guillen 2016; Strain and Wang 2020; Golse et al. 2019a; 2019b; Silvestre 2017]. A brief glance at some of these references illustrates the breadth of techniques that have found partial success at answering the open questions: probability-based arguments, kinetic and parabolic theory, and many more.

The purpose of this paper is to bring in another set of techniques to help answer some of these fundamental questions. The gradient flow theory applied to PDEs has flourished in the last decades. In their seminal paper, Jordan, Kinderlehrer, and Otto [Jordan et al. 1998] proposed a variational approach (JKO scheme) extended later on to a wide class of PDEs using the optimal transportation distance of probability measures. These results and many more achievements from their contemporaries allowed for novel approaches to questions of existence, uniqueness, convergence to equilibrium, and other aspects of a large class of PDEs; we mention [Ambrosio et al. 2008; Santambrogio 2017] for a coherent exposition of these techniques and the relevant literature, even as more advances have been made since then.

The advantage of our variational characterization of the Landau equation is that it unveils new possible routes of showing convergence results for this equation. First of all, it allows for natural regularizations of the Landau equation by taking the steepest descent of regularized entropy functionals instead of the Boltzmann entropy as in [Carrillo et al. 2019]. This idea was recently developed in [Carrillo et al. 2020] leading to structure-preserving particle schemes with good accuracy. We can also consider the framework of convergence of gradient flows based on Γ-convergence introduced in [Sandier and Serfaty 2004; Serfaty 2011] to attack the convergence of these numerical methods [Carrillo et al. 2020]. Moreover, this approach is flexible enough to also study the rigorous convergence of the grazing collision limit of the Boltzmann equation to the Landau equation. The grazing collision limit was recently revisited in the gradient flow framework by three of the authors [Carrillo et al. 2022]. There, ideas from Γ-convergence were used to pass from Erbar's gradient flow description [2023] for the Boltzmann equation to the present work's description of the Landau equation. Finally, deriving uniqueness from the variational structure is classically done through convexity properties of the entropy functional with respect to the geodesics of the Landau metric. This is another important avenue of research that our work opens. Moreover, gradient flows of convex entropies typically enjoy instantaneous smoothing [Ambrosio et al. 2008]; even if the entropy at $t = 0$ is infinite, for $t > 0$, the entropy becomes finite. In the case of Landau, we are not aware if this property holds for $\mathcal{H}$.

We mention briefly the connection between (1) and the Fokker–Planck equation. For $\gamma = 0$, one can formally compute the evolution of $\int v^i v^j f(v)\, dv$ through (1). This a priori information allows one to reduce (1) to a linear Fokker–Planck equation for $\gamma = 0$. The present work proposes the alternative viewpoint that the resultant Fokker–Planck equation can be viewed as the d_L-gradient flow of $\mathcal{H}$ for $\gamma = 0$. Since many variants of the linear Fokker–Planck equation have been well-studied, this case serves as a nice benchmark to test the gradient flow theory developed here.

The plan of this paper is as follows. Section 2 introduces the prerequisites and contains the statements of the main results. We first construct and analyze in Section 3 the Landau metric based on (3). For a regularized problem, Section 4 shows the equivalence between weak solutions and gradient flows, while Section 5 shows the existence of gradient flow solutions via a minimizing movement scheme. Finally, we show in Section 6 that a gradient flow solution is equivalent to H-solutions of the Landau equation (1) under some integrability assumptions. The Appendix is devoted to some technical lemmas needed in the proof of the main theorems regarding the chain rule identity behind the definition of weak solutions for the regularized Landau equation.

2. Preliminaries and the main results

We start by introducing the necessary notation and definitions together with a quick overview of gradient flow concepts to make our main results fully self-contained.

2.1. ***Notation and definitions.*** We define

$$a \lesssim_{\dots} b \quad \iff \quad \text{there exists } C(\cdots) > 0 \text{ such that } a \le C(\cdots) b.$$

We adopt the Japanese angle bracket notation for a smooth alternative to absolute value

$$\langle v\rangle^2 = 1 + |v|^2, \quad v \in \mathbb{R}^d.$$

For $\epsilon > 0$, we define our regularization kernel to be an exponential distribution:

$$G^\epsilon(v) = \epsilon^{-d} G\left(\frac{v}{\epsilon}\right), \quad G(v) = C_d \exp(-\langle v\rangle), \quad C_d = \left(\int_{\mathbb{R}^d} \exp(-\langle v\rangle)\, dv\right)^{-1}.$$

Our results work for some general tail behavior in the kernels given by

$$G^{s,\epsilon}(v) = \epsilon^{-d} G^s\left(\frac{v}{\epsilon}\right), \quad G^s(v) = C_{s,d} \exp(-\langle v\rangle^s), \quad C_{s,d} = \left(\int_{\mathbb{R}^d} \exp(-\langle v\rangle^s)\, dv\right)^{-1}$$

for $s > 0$; we point out some of the limitations and restrictions on $s > 0$ in the later estimates. We shall refer to $G^{2,\epsilon}$ as the Maxwellian regularization. We denote the space of probability measures over $\mathbb{R}^d$ by $\mathscr{P}(\mathbb{R}^d)$, endowed with the weak topology against bounded continuous functions. We will mostly be dealing with the Lebesgue measure on $\mathbb{R}^d$ as our reference measure, which we denote by $\mathcal{L}$. The subset $\mathscr{P}^a(\mathbb{R}^d) \subset \mathscr{P}(\mathbb{R}^d)$ denotes the set of absolutely continuous probability measures with respect to Lebesgue measure. For $p > 0$, we also define the probability measures with finite p-moments $\mathscr{P}_p(\mathbb{R}^d)$ by

$$\mathscr{P}_p(\mathbb{R}^d) := \left\{\mu \in \mathscr{P}(\mathbb{R}^d) \;\middle|\; m_p(\mu) := \int_{\mathbb{R}^d} \langle v\rangle^p \, d\mu(v) < \infty\right\}.$$

Finally, for $E > 0$, we consider the subset $\mathscr{P}_{p,E}(\mathbb{R}^d) \subset \mathscr{P}_p(\mathbb{R}^d)$ of probability measures with p-moments uniformly bounded by E:

$$\mathscr{P}_{p,E}(\mathbb{R}^d) := \{\mu \in \mathscr{P}_p(\mathbb{R}^d) \mid m_p(\mu) \le E\}.$$

We denote by $\mathcal{M}$ the space of signed Radon measures on $\mathbb{R}^d \times \mathbb{R}^d$ with the standard weak* topology against the continuous and compactly supported functions of $\mathbb{R}^d \times \mathbb{R}^d$. The space $\mathcal{M}^d$ is the space of signed d-length Radon measures. For $T > 0$, we will add the time contribution of the measures by defining $\mathcal{M}_T$ to be the space of signed Radon measures on $\mathbb{R}^d \times \mathbb{R}^d \times [0, T]$ with the usual weak* topology. Similarly, $\mathcal{M}_T^d$ will be the space of signed d-length Radon measures on $\mathbb{R}^d \times \mathbb{R}^d \times [0, T]$.

For $\mu \in \mathscr{P}(\mathbb{R}^d)$, we define a family of regularized entropies $\mathcal{H}_\epsilon[\mu]$ by

$$\mathcal{H}_\epsilon[\mu] := \int_{\mathbb{R}^d} [\mu * G^\epsilon](v) \log[\mu * G^\epsilon](v)\, dv,$$

which we shall see is well-defined provided μ has a finite moment in Lemma 30. Formally, one can calculate the first variation of this functional in $\mathscr{P}_2$ as

$$\frac{\delta \mathcal{H}_\epsilon}{\delta \mu}(v) = G^\epsilon * \log[\mu * G^\epsilon](v).$$

This can be formally obtained by calculating Fréchet derivatives in the sense of identifying the limit

$$\int_{\mathbb{R}^d} \frac{\delta \mathcal{H}_\epsilon}{\delta \mu}(v)\phi(v)\, dv = \lim_{t \downarrow 0} \frac{\mathcal{H}_\epsilon[\mu + t\phi] - \mathcal{H}_\epsilon[\mu]}{t}$$

for arbitrary $\phi \in C_c^\infty(\mathbb{R}^d)$ with zero mean $\int_{\mathbb{R}^d} \phi = 0$. To be precise, the first variation (in an L^2 setting) would actually be $\delta\mathcal{H}_\epsilon/\delta\mu = 1 + G^\epsilon * \log[\mu * G^\epsilon]$. We drop the constant term since our functional space is $\mathscr{P}$ and the first variation typically appears with derivatives applied to it. For a functional $\mathcal{F} : \mathscr{P}^a(\mathbb{R}^d) \to \mathbb{R}$ with first variation $\delta\mathcal{F}/\delta f$, we refer to the $\mathcal{F}$ Landau equation as

$$\partial_t f = \nabla \cdot \left(f \int_{\mathbb{R}^d} f_* |v - v_*|^{2+\gamma} \Pi[v - v_*] \left(\nabla \frac{\delta \mathcal{F}}{\delta f} - \nabla_* \frac{\delta \mathcal{F}_*}{\delta f_*} \right) dv_* \right). \tag{7}$$

To clarify the meaning of $\tilde{\nabla} \cdot$, for a given test function $\phi = \phi(v) \in \mathbb{R}^d$ and vector-valued test function $A = A(v, v_*) \in \mathbb{R}^d$, we have

$$\iint_{\mathbb{R}^{2d}} [\tilde{\nabla}\phi](v, v_*) \cdot A(v, v_*)\, dv_*\, dv = -\int_{\mathbb{R}^d} \phi(v) [\tilde{\nabla} \cdot A](v)\, dv.$$

In this way, the $\mathcal{F}$ Landau equation (7) can be concisely written as

$$\partial_t f = \frac{1}{2} \tilde{\nabla} \cdot \left(f f_* \tilde{\nabla} \frac{\delta \mathcal{F}}{\delta f} \right).$$

Note, by formally testing (7) with $\phi = \delta\mathcal{F}/\delta f$, one obtains an analogy of Boltzmann's H-theorem with the functional $\mathcal{F}$:

$$\frac{d}{dt} \mathcal{F}[f_t] = -D_{\mathcal{F}}(f_t) := -\frac{1}{2} \iint_{\mathbb{R}^{2d}} f f_* \left| \tilde{\nabla} \frac{\delta \mathcal{F}}{\delta f} \right|^2 dv\, dv_* \le 0.$$

We will refer to $D_{\mathcal{F}}$ as the $\mathcal{F}$ dissipation. This notation induces our notion of weak solutions to the $\mathcal{F}$ Landau equation (7) closely following Villani's H-solutions [1998a].

Definition 1 (weak $\mathcal{F}$ solutions). For $T > 0$, we say that a curve $f \in C([0, T]; L^1(\mathbb{R}^d))$ is a weak solution to the $\mathcal{F}$ Landau equation (7) if the following hold:

(1) $f\mathcal{L}$ is a probability measure with uniformly bounded second moment so that

$$f_t \geq 0, \qquad \int_{\mathbb{R}^d} f_t(v)\,dv = 1 \quad \text{for all } t \in [0,T], \qquad \sup_{t\in[0,T]} \int_{\mathbb{R}^d} \langle v\rangle^2 f_t(v)\,dv < \infty.$$

(2) The functional $\mathcal{F}$ evaluated along the curve is bounded by its initial value:

$$\mathcal{F}[f_t] \leq \mathcal{F}[f_0] < +\infty \quad \text{for all } t \in [0,T].$$

(3) The $\mathcal{F}$ dissipation is time integrable:

$$\int_0^T D_{\mathcal{F}}(f_t)\,dt = \frac{1}{2}\int_0^T \iint_{\mathbb{R}^{2d}} ff_* \left|\tilde{\nabla}\frac{\delta\mathcal{F}}{\delta f}\right|^2 dv\,dv_*\,dt < \infty.$$

(4) For every test function $\phi \in C_c^\infty((0,T)\times\mathbb{R}^d)$, equation (7) is satisfied in weak form:

$$\int_0^T \int_{\mathbb{R}^d} \partial_t\phi f_t(v)\,dv\,dt = \frac{1}{2}\int_0^T \iint_{\mathbb{R}^{2d}} ff_* \tilde{\nabla}\phi\cdot\tilde{\nabla}\frac{\delta\mathcal{F}}{\delta f}\,dv\,dv_*\,dt.$$

For $\epsilon > 0$, we will refer to the weak $\mathcal{H}_\epsilon$ solutions as *ϵ-solutions* and, recalling $\mathcal{H}$ is the Boltzmann entropy, we will refer to weak $\mathcal{H}$ solutions as just *weak solutions* or *H-solutions*. We deliberately use the terminology of H-solutions since the time integrability of $D_{\mathcal{H}}(f_t)$, as for [Villani 1998a], is essential in our analysis.

2.2. *Quick review of gradient flow theory.* We recall the basic definitions of gradient flow theory that can be found in more generality in [Ambrosio et al. 2008, Chapter 1]. Throughout, (X,d) denotes a complete (pseudo-)metric space X with (pseudo-)metric d. Points $a < b \in \mathbb{R}$ will refer to endpoints of some interval. $F : X \to (-\infty,\infty]$ will denote a proper function.

Definition 2 (absolutely continuous curve). A function $\mu : t \in (a,b) \mapsto \mu_t \in X$ is said to be an *absolutely continuous curve* if there exists $m \in L^2(a,b)$ such that for every $s \leq t \in (a,b)$

$$d(\mu_t,\mu_s) \leq \int_s^t m(r)\,dr.$$

Among all possible functions m in Definition 2, one can make the following minimal selection.

Definition 3 (metric derivative). For an absolutely continuous curve $\mu : (a,b) \to X$, we define its *metric derivative* at every $t \in (a,b)$ by

$$|\dot{\mu}|(t) := \lim_{h\to 0}\frac{d(\mu_{t+h},\mu_t)}{|h|}.$$

Further properties of the metric derivative can be found in [Ambrosio et al. 2008, Theorem 1.1.2].

Definition 4 (strong upper gradient). The function $g : X \to [0,\infty]$ is a *strong upper gradient* with respect to F if for every absolutely continuous curve $\mu : t \in (a,b) \mapsto \mu_t \in X$ we have that $g\circ\mu : (a,b) \to [0,\infty]$ is Borel and the following inequality holds:

$$|F[\mu_t] - F[\mu_s]| \leq \int_s^t g(\mu_r)|\dot{\mu}|(r)\,dr \quad \text{for all } a < s \leq t < b.$$

Using Young's inequality and moving everything to one side, the inequality in Definition 4 implies

$$F[\mu_t] - F[\mu_s] + \frac{1}{2}\int_s^t g(\mu_r)^2\,dr + \frac{1}{2}\int_s^t |\dot\mu|^2(r)\,dr \geq 0 \quad \text{for all } a < s \leq t < b.$$

If the reverse inequality also holds, one obtains the stronger energy dissipation *equality*. This leads to our notion of gradient flows.

Definition 5 (curve of maximal slope). An absolutely continuous curve $\mu : (a, b) \to X$ is said to be a *curve of maximal slope* for F with respect to its strong upper gradient $g : X \to [0, \infty]$ if $F \circ \mu : (a, b) \to [0, \infty]$ is nonincreasing and the following inequality holds:

$$F[\mu_t] - F[\mu_s] + \frac{1}{2}\int_s^t g(\mu_r)^2\,dr + \frac{1}{2}\int_s^t |\dot\mu|^2(r)\,dr \leq 0 \quad \text{for all } a < s \leq t < b.$$

F has the following natural candidates for upper gradient.

Definition 6 (slopes). We define the *local slope of F* by

$$|\partial F|(\mu) := \limsup_{\nu\to\mu} \frac{(F(\nu) - F(\mu))^+}{d(\nu, \mu)}.$$

The superscript "+" refers to the positive part. The *relaxed slope of F* is given by

$$|\partial^- F|(\mu) := \inf\{\liminf_{n\to\infty} |\partial F|(\mu_n) \mid \mu_n \to \mu,\ \sup_{n\in\mathbb{N}}(d(\mu_n, \mu), F(\mu_n)) < +\infty\}.$$

2.3. *Main results.* In order to understand the Landau equation as a gradient flow, we need to clarify what type of object the corresponding metric is.

Theorem 7 (distance on $\mathscr{P}_{2,E}(\mathbb{R}^d)$). *The (pseudo-)metric d_L on $\mathscr{P}_{2,E}(\mathbb{R}^d)$ satisfies:*

- *d_L-convergent sequences are weakly convergent.*
- *d_L-bounded sets are weakly compact.*
- *The map $(\mu_0, \mu_1) \mapsto d_L(\mu_0, \mu_1)$ is weakly lower semicontinuous.*
- *For any $\tau \in \mathscr{P}_2(\mathbb{R}^d)$ the subset $\mathscr{P}_\tau(\mathbb{R}^d) := \{\mu \in \mathscr{P}_{2,m_2(\tau)}(\mathbb{R}^d) \mid d_L(\mu, \tau) < \infty\}$ is a complete geodesic space.*

The content of this theorem is essentially that our new proposed distance actually provides a meaningful topological structure on $\mathscr{P}_{2,E}(\mathbb{R}^d)$. Furthermore, the connection to ϵ-solutions of Landau is established when considering the previous notions of slope and upper gradient with respect to d_L. General conditions which guarantee $d_L(\mu_0, \mu_1) < +\infty$ are presently unknown. In Lemma 15, we will see that a necessary condition is that μ_0 and μ_1 have the same mean velocity. Moreover, for $\gamma \in [-4, -2]$, Lemma 15 asserts that they should have the same second moment. In the construction of d_L detailed in Section 3, if $\mu = \mu(t)$ for $t \in [0, T]$ is an H-solution of Landau, then it is certainly true that $d_L(\mu(t),\ \mu(s)) < +\infty$ for all $0 \leq t, s \leq T$.

Theorem 8 (epsilon equivalence). *Fix any $\epsilon, E > 0$, $\gamma \in [-4, 0]$. Assume that a curve $\mu : [0, T] \to \mathscr{P}_{2,E}(\mathbb{R}^d)$ has a density $\mu_t = f_t\mathcal{L}$. Then μ is a curve of maximal slope for $\mathcal{H}_\epsilon$ with respect to its upper gradient $\sqrt{D_{\mathcal{H}_\epsilon}}$ if and only if its density f is an ϵ-solution to the Landau equation.*

From the numerical perspective, we can also construct ϵ-solutions using the JKO scheme (see Section 5) which is the following:

Theorem 9 (existence of curves of maximal slope). *For any* $\epsilon, E > 0$, $\gamma \in [-4, 0]$, *and initial data* $\mu_0 \in \mathscr{P}_{2,E}(\mathbb{R}^d)$, *there exists a curve of maximal slope in* $\mathscr{P}_{2,E}(\mathbb{R}^d)$ *for* $\mathcal{H}_\epsilon$ *with respect to its upper gradient* $\sqrt{D_{\mathcal{H}_\epsilon}}$.

Remark 10. The curves constructed in Theorem 9 do not necessarily have a density with respect to Lebesgue measure; the regularization allows $\mathcal{H}_\epsilon[\mu] < +\infty$ without μ being absolutely continuous with respect to Lebesgue measure. Moreover, uniqueness of such curves is beyond the scope of the present work although it would be interesting to see what convexity properties are available for $\mathcal{H}_\epsilon$ with respect to d_L. This could also shed some insight into the available convexity of $\mathcal{H}$ with respect to d_L.

Remark 11. The choice of an exponential convolution kernel G^ϵ for the regularized entropy $\mathcal{H}_\epsilon$ is perhaps unnatural compared to the Maxwellian regularization $G^{2,\epsilon}$. We discuss in more detail the estimates that fail using $G^{2,\epsilon}$ in Remark 33 as it pertains to Theorem 8. With respect to Theorem 9, the general construction of *some* curve can be done even with the Maxwellian regularization. However, due to the same lack of estimates, this curve might not be a curve of maximal slope with respect to $\sqrt{D_{\mathcal{H}_\epsilon}}$. This is discussed in Remark 37.

Motivated by recent numerical experiments [Carrillo et al. 2020], Theorems 8 and 9 provide the theoretical basis to this ϵ-approximated Landau equation. In the limit $\epsilon \to 0$, more assumptions are required.

Theorem 12 (full equivalence). *We fix* $d = 3$ *and* $\gamma \in (-3, 0]$. *Suppose that, for some* $T > 0$, *a curve* $\mu : [0, T] \to \mathscr{P}(\mathbb{R}^3)$ *has a density* $\mu_t = f_t \mathcal{L}$ *that satisfies the following set of assumptions*:

(A1) (moments and L^p) *Assume that, for some* $0 < \eta \le \gamma + 3$, *we have*

$$\langle v\rangle^{2-\gamma} f_t(v) \in L_t^\infty(0, T; L_v^1 \cap L_v^{(3-\eta)/(3+\gamma-\eta)}(\mathbb{R}^3)).$$

(A2) (finite entropy) *We assume that the initial entropy is finite*

$$\mathcal{H}[f_0] = \int_{\mathbb{R}^3} f_0 \log f_0 < +\infty.$$

(A3) (finite entropy-dissipation) *We assume that the entropy-dissipation of* f *is integrable in time*:

$$\begin{aligned} D(f_t) = D_{\mathcal{H}}(f_t) &= \frac{1}{2} \iint_{\mathbb{R}^6} f f_* \left| \tilde{\nabla} \frac{\delta \mathcal{H}}{\delta f} \right|^2 dv\, dv_* \\ &= \frac{1}{2} \iint_{\mathbb{R}^6} f f_* |v - v_*|^{\gamma+2} |\Pi[v - v_*](\nabla \log f - \nabla_* \log f_*)|^2 \, dv\, dv_* \in L_t^1(0, T). \end{aligned}$$

Then μ *is a curve of maximal slope for* $\mathcal{H}$ *with respect to its upper gradient* $\sqrt{D}$ *if and only if its density* f *is a weak solution of the Landau equation.*

Remark 13. When $\gamma \in [-2, 0]$, it is known that for suitable initial data (lying in weighted L^p spaces for p large enough and for a sufficient power-like weight), weak solutions of Landau equation satisfying (A1)–(A3) are known to exist (and to be strong and unique under extra conditions). We refer to [Wu 2014], and Appendix B of [Desvillettes 2022] when $\gamma > -2$, for details.

When $\gamma \in (-3, -2)$, (A1) is not known to hold for global weak solutions with large initial data. Solutions satisfying (A1)–(A3) are nevertheless known to exist for initial data close to equilibrium (see [Guo 2002] in a much larger spatially inhomogeneous context), or in the Coulomb case $\gamma = -3$ (in that case $(3-\eta)/(3+\gamma-\eta)$ being replaced by ∞) for large initial data, but on specific intervals of times only [Desvillettes et al. 2023; Arsenev and Peskov 1977].

The focus on the Maxwellian and soft potential regime $\gamma \le 0$ here is motivated by building a gradient flow framework to address the open questions for Landau. The hard potential case $\gamma \in (0, 1)$ has already been studied in detail in [Desvillettes and Villani 2000b; 2000a]. We believe that our results also carry to the hard potentials. In particular, the exponents in (A1) should be modified to

$$\langle v\rangle^{2+\gamma} f_t(v) \in L_t^\infty(0, T;\ L_v^1(\mathbb{R}^3)), \qquad f_t(v) \in L_t^\infty(0, T;\ L_v^{(3/(3-\gamma))+}(\mathbb{R}^3)), \qquad 0 < \gamma < 1.$$

We emphasize that these conditions are guaranteed since the required moments and L^p integrability are propagated from appropriate initial data when $\gamma > 0$ [Desvillettes and Villani 2000a; 2000b]. This condition appears in [Desvillettes 2016, Corollary 2.7]. It is the hard potential version of Theorem 41, which is crucial to the proof of Theorem 12. Much of our analysis remains the same; however, the space $\mathscr{P}_2$ should be changed to $\mathscr{P}_{2+\gamma}$ cohering with the moment condition above and trivializing Lemma 43, for example.

It is an open problem to find the range of values γ under which we can show the existence of curves of maximal slope for the original Landau equation (1), or equivalently, constructing solutions of the original Landau equation passing $\epsilon \to 0$ in Theorem 9. Some of the difficulties to achieve this result are the propagation of moments for the regularized Landau equation uniformly in ϵ and the compactness of sequences with bounded in ϵ regularized entropy dissipation $D_{\mathcal{H}_\epsilon}$. The rest of this work is devoted to showing the main four theorems in the next four sections.

3. The Landau metric d_L

Our approach to defining the distance d_L mentioned in Theorem 7 closely follows the dynamic formulation of transport distances originally due to Benamou and Brenier [2000] and further extended by Dolbeault, Nazaret, and Savaré [Dolbeault et al. 2009]. We also refer the reader to [Erbar 2023] for a similar approach.

3.1. ***Grazing continuity equation.*** We consider for $\gamma \in [-4, 0]$ the *grazing continuity equation*

$$\partial_t \mu_t + \tfrac{1}{2}\tilde{\nabla} \cdot M_t = 0 \quad \text{in } (0, T) \times \mathbb{R}^d, \tag{8}$$

which is interpreted in the sense of distributions. For every $\phi \in C_c^\infty((0, T) \times \mathbb{R}^d)$, we have

$$\int_0^T \int_{\mathbb{R}^d} \partial_t \phi(t, v)\, d\mu_t(v)\, dt + \frac{1}{2} \int_0^T \iint_{\mathbb{R}^{2d}} [\tilde{\nabla}\phi](t, v, v_*)\, dM_t(v, v_*)\, dt = 0.$$

Another formulation (see Lemma 14) is the following for $\zeta \in C_c^\infty(\mathbb{R}^d)$:

$$\frac{d}{dt}\int_{\mathbb{R}^d} \zeta(v)\, d\mu_t(v) = \frac{1}{2}\iint_{\mathbb{R}^{2d}} \tilde{\nabla}\zeta(v, v_*)\, dM_t(v, v_*). \tag{9}$$

The curves $(\mu_t)_{t\in[0,T]}$, $(M_t)_{t\in[0,T]}$ are Borel families of measures belonging to $\mathcal{M}_+$ and $\mathcal{M}^d$ respectively. We will refer to μ from the pair as a *curve* and M as a *grazing rate*. For some regularity properties, we will also need to assume the moment condition

$$\int_0^T \iint_{\mathbb{R}^{2d}} (1 + |v| + |v_*|)\, d|M_t|(v, v_*)\, dt < \infty. \tag{10}$$

We first establish some a priori properties of solutions to the grazing continuity equation.

Lemma 14 (continuous representative). *For families (μ_t), (M_t) satisfying the grazing continuity equation and the finite moment condition (10), there exists a unique weakly* continuous representative curve $(\tilde{\mu}_t)_{t\in[0,T]}$ such that $\tilde{\mu}_t = \mu_t$ for a.e. $t \in [0, T]$. Furthermore, for any $\phi \in C_c^\infty((0, T) \times \mathbb{R}^d)$ and any $t_0, t_1 \in [0, T]$, we have the formula*

$$\int_{\mathbb{R}^d} \phi_{t_1}\, d\tilde{\mu}_{t_1} - \int_{\mathbb{R}^d} \phi_{t_0}\, d\tilde{\mu}_{t_0} = \int_{t_0}^{t_1}\int_{\mathbb{R}^d} \partial_t\phi\, d\mu_t\, dt + \frac{1}{2}\int_{t_0}^{t_1}\iint_{\mathbb{R}^{2d}} \tilde{\nabla}\phi\, dM_t\, dt.$$

Proof. This proof is nearly identical to [Ambrosio et al. 2008, Lemma 8.1.2]. There, it was crucial to estimate the distributional time derivative of $t \mapsto \mu_t$. We perform the analogous estimate here to highlight the difference in our context. Fix $\zeta \in C_c^\infty(\mathbb{R}^d)$ and consider the map

$$t \in (0, T) \mapsto \mu_t(\zeta) = \int_{\mathbb{R}^d} \zeta(v)\, d\mu_t(v) \in \mathbb{R}.$$

According to (9), the distributional time derivative is

$$\dot{\mu}_t(\zeta) = \frac{1}{2}\iint_{\mathbb{R}^{2d}} \tilde{\nabla}\zeta\, dM_t(v, v_*) = \frac{1}{2}\iint_{\mathbb{R}^{2d}} |v - v_*|^{1+\gamma/2}\Pi[v - v_*](\nabla\zeta - \nabla_*\zeta_*)\, dM_t(v, v_*).$$

Depending on the values of γ above or below -2, the integrand can be estimated:

$$\big||v-v_*|^{1+\gamma/2}\Pi[v-v_*](\nabla\zeta-\nabla_*\zeta_*)\big| \le \begin{cases} 2^{1+\gamma/2}\sup_{w\in\mathbb{R}^d}|\nabla\zeta(w)|(|v|^{1+\gamma/2}+|v_*|^{1+\gamma/2}), & \gamma\in[-2,0], \\ \sup_{w\in\mathbb{R}^d}|D^2\zeta(w)|\,|v-v_*|^{2+\gamma/2}, & \gamma\in[-4,-2). \end{cases}$$

Consequently, using the moment condition (10), we have the following estimates depending on $\gamma \in [-4, 0]$:

$$|\dot{\mu}_t(\zeta)| \lesssim \begin{cases} \sup_{w\in\mathbb{R}^d}|\nabla\zeta(w)| \iint_{\mathbb{R}^{2d}}(1+|v|+|v_*|)d|M_t|(v, v_*), & \gamma\in[-2,0], \\ \sup_{w\in\mathbb{R}^d}|D^2\zeta(w)| \iint_{\mathbb{R}^{2d}}(1+|v|+|v_*|)d|M_t|(v, v_*), & \gamma\in[-4,-2). \end{cases}$$

The rest of the proof proceeds as in [Ambrosio et al. 2008, Lemma 8.1.2] using the C^2-norm of ζ for the soft potentials $\gamma \in [-4, -2)$ as opposed to their C^1 control of ζ. □

Lemma 15 (conservation lemma). *Fix $\gamma \in [-4, 0]$ and let $(\mu_t)_{t\in[0,T]}$, $(M_t)_{t\in[0,T]}$ be Borel families of measures in $\mathcal{M}_+$, $\mathcal{M}^d$ respectively satisfying (8) and the moment condition (10). Assume further that*

$(\mu_t)_{t\in[0,T]}$ *is weakly* continuous with respect to* t. *We have that mass and momentum are conserved*:

$$\mu_t(\mathbb{R}^d) = \mu_0(\mathbb{R}^d), \qquad \int_{\mathbb{R}^d} v\, d\mu_t(v) = \int_{\mathbb{R}^d} v\, d\mu_0(v) \quad \textit{for all } t \in [0, T].$$

In the case $\gamma \in [-4, -2]$ *we have that the energy is conserved*:

$$\int_{\mathbb{R}^d} |v|^2\, d\mu_t(v) = \int_{\mathbb{R}^d} |v|^2\, d\mu_0(v) \quad \textit{for all } t \in [0, T].$$

Proof. To minimize clutter, we introduce $w = |v - v_*|^{1+\gamma/2}$. We show the proof of the conservation of energy for $\gamma \in [-4, -2]$. We consider a fixed $\varphi \in C_c^\infty(B_2)$ which satisfies

$$0 \le \varphi \le 1 \quad \text{and} \quad \varphi(v) = 1 \quad \text{in } B_1.$$

We define

$$\varphi_R(v) = \varphi\Big(\frac{v}{R}\Big).$$

Using the grazing continuity equation, we have

$$\begin{aligned}\int_{\mathbb{R}^d} &|v|^2 \varphi_R(v)\, d\mu_t(v) - \int_{\mathbb{R}^d} |v|^2 \varphi_R(v)\, d\mu_0(v) \\ &= \int_0^t \iint_{\mathbb{R}^{2d}} w\Pi\Big(v\varphi_R(v) + |v|^2 \frac{\nabla\varphi(v/R)}{R} - v_*\varphi_R(v_*) - |v_*|^2 \frac{\nabla\varphi(v_*/R)}{R}\Big)\, dM_s(v, v_*)\, ds. \qquad (11)\end{aligned}$$

We estimate the contribution of $v\phi_R(v) - v_*\phi_R(v_*)$ from the integral in (11) using the cancellation from the projection $\Pi[v - v_*]$ to obtain

$$\begin{aligned}\Big|\int_0^t \iint_{\mathbb{R}^{2d}} w\Pi(v\varphi_R(v) - v_*\varphi_R(v_*))\, dM_s\Big| &\le \int_0^t \iint_{(B_R\times B_R)^c} w|v\varphi_R(v) - v_*\varphi_R(v_*)|\, d|M_s| \\ &\lesssim \int_0^t \iint_{(B_R\times B_R)^c} (1 + |v| + |v_*|)\, d|M_s|,\end{aligned}$$

where we have used $\gamma \in [-4, -2]$ to bound

$$w|v\varphi_R(v) - v_*\varphi_R(v_*)| \lesssim \begin{cases} 1, & |v - v_*| \le 1, \\ |v| + |v_*|, & |v - v_*| \ge 1. \end{cases}$$

Similarly, using that $\nabla\phi_R$ is supported in $B_{2R} \setminus B_R$ and that $|\partial_{v^i}\{|v|^2 \partial_{v^j}\varphi(v/R)/R\}| \lesssim 1$ for every index $i, j \in \{1, \dots, d\}$, we obtain

$$\Big|\int_0^t \iint_{\mathbb{R}^{2d}} w\Pi\Big(|v|^2 \frac{\nabla\varphi(v/R)}{R} - |v_*|^2 \frac{\nabla\varphi(v_*/R)}{R}\Big)\, dM_s\Big| \lesssim \iint_{(B_R\times B_R)^c} 1 + |v| + |v_*|\, d|M_s|,$$

where we have controlled the difference with a mean-value-type estimate. From the previous bounds, we can use hypothesis (10) to take $R \to \infty$ in (11) and obtain the conservation of energy

$$\lim_{R\to\infty} \int_{\mathbb{R}^d} |v|^2 \varphi_R(v)\, d\mu_t(v) = \lim_{R\to\infty} \int_{\mathbb{R}^d} |v|^2 \varphi_R(v)\, d\mu_0(v).$$

The proofs for conservation of mass and momentum involve testing the grazing continuity equation against ϕ_R and $v_i\phi_R$ respectively, where v_i is the i-th component of v. For these statements, the case $\gamma \in [-4, -2]$ follows in the same way. For $\gamma \in [-2, 0]$, the estimates can be more blunt since the weight is no longer singular. □

Remark 16. Note that as γ increases into the range $(-2, 0]$, the weight function w starts adding growth so the mean-value-type argument in Lemma 15 no longer helps unless more moments of M are assumed than (10). Due to the conservation of mass, the unique weakly* continuous representative $(\tilde{\mu}_t)$ of Lemma 14 has the additional property of being weakly continuous in the context of $\mathscr{P}(\mathbb{R}^d)$.

Based on the previous results, we propose the following definition.

Definition 17 (grazing continuity equation). For some terminal time $T > 0$, we define $\mathcal{GCE}_T$ to be the set of pairs of measures $(\mu_t, M_t)_{t\in[0,T]}$ satisfying the following:

(1) $\mu_t \in \mathscr{P}(\mathbb{R}^d)$ is weakly continuous with respect to $t \in [0, T]$. $(M_t)_{t\in[0,T]}$ is a family of Borel measures belonging to $\mathcal{M}^d$.

(2) We have the moment bound

$$\int_0^T \iint_{\mathbb{R}^{2d}} (1 + |v| + |v_*|)\, d|M_t|(v, v_*)\, dt < \infty.$$

(3) The grazing continuity equation (8) is satisfied in the distributional sense. That is, for every $\phi \in C_c^\infty((0, T) \times \mathbb{R}^d)$,

$$\int_0^T \int_{\mathbb{R}^d} \partial_t \phi \, d\mu_t \, dt + \frac{1}{2} \int_0^T \iint_{\mathbb{R}^{2d}} \tilde{\nabla} \phi \, dM_t \, dt = 0,$$

or equivalently, for every $\zeta \in C_c^\infty(\mathbb{R}^d)$,

$$\frac{d}{dt} \int_{\mathbb{R}^d} \zeta(v)\, d\mu_t(v) = \frac{1}{2} \iint_{\mathbb{R}^{2d}} \tilde{\nabla} \zeta(v, v_*)\, dM_t(v, v_*).$$

For fixed probability measures λ, ν, we may also specify the subset $\mathcal{GCE}(\lambda, \nu)$ as those pairs $(\mu, M) \in \mathcal{GCE}_T$ such that $\mu_0 = \lambda$, $\mu_T = \nu$. For $E > 0$, we will speak of curves $(\mu, M) \in \mathcal{GCE}_T^{2,E}$ such that

$$\int_{\mathbb{R}^d} |v|^2 \, d\mu_t(v) \le E \quad \text{for all } t \in [0, T].$$

3.2. ***Action of a curve.*** In this section, we construct the action of a curve under the grazing continuity equation. We introduce the function $\alpha : \mathbb{R}^d \times \mathbb{R}_{\ge 0} \to [0, \infty]$ defined by

$$\alpha(u, s) := \begin{cases} |u|^2/(2s), & s \neq 0, \\ 0, & s = 0, u = 0, \\ \infty, & s = 0, u \neq 0. \end{cases}$$

Remark 18. The function α is lower semicontinuous (lsc), convex, and positively 1-homogeneous.

For fixed $\mu \in \mathscr{P}(\mathbb{R}^d)$, $M \in \mathcal{M}^d$, we consider the tensorized probability measure $\mu \otimes \mu \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d)$ given by $\mu \otimes \mu(dv, dv_*) = \mu(dv)\mu(dv_*)$. Define $\tau \in \mathcal{M}$ given by $\tau = \mu \otimes \mu + |M|$ and the decompositions $\mu \otimes \mu = f^1 \tau$ and $M = N\tau$. We define the action functional as

$$\mathcal{A}(\mu, M) := \iint_{\mathbb{R}^{2d}} \alpha(N, f^1) d\tau. \tag{12}$$

This is well-defined by the 1-homogeneity of α. The following lemma establishes a more concrete expression for the action functional.

Lemma 19. *Let $\mu \in \mathscr{P}(\mathbb{R}^d)$ be absolutely continuous with respect to $\mathcal{L}$ and $\mu = f\mathcal{L}$. Let $M \in \mathcal{M}^d$ be given such that $\mathcal{A}(\mu, M) < \infty$. Then, M is absolutely continuous with respect to $ff_* \, dv\, dv_*$ given by some density $U : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ such that $M = ff_* U \, dv\, dv_* = m\, dv\, dv_*$ and*

$$\mathcal{A}(\mu, M) = \frac{1}{2} \iint_{\mathbb{R}^{2d}} ff_* |U|^2 \, dv\, dv_* = \frac{1}{2} \iint_{\mathbb{R}^{2d}} \frac{|m|^2}{ff_*} \, dv\, dv_*.$$

Proof. The proof is identical to [Erbar 2023, Lemma 3.6] up to appropriate modifications. Define $\tau \in \mathcal{M}$ by $\tau = \mu \otimes \mu + |M|$ and label the corresponding densities (which may be infinite) $\mu \otimes \mu = g\tau$ and $M = N\tau$. It suffices to show that M is absolutely continuous with respect to $\mu \otimes \mu$, which is the goal of this proof.

Suppose $S \subset \mathbb{R}^{2d}$ is a measurable set such that $\mu \otimes \mu(S) = 0$. This is equivalent to saying $g = 0$ τ-almost everywhere in S. Since α is positive, the assumption $\mathcal{A}(\mu, M) < +\infty$ certainly implies $\alpha(N, g) < +\infty$ τ-almost everywhere in S. By the definition of α, we must also have $N = 0$ τ-almost everywhere in S, which is equivalent to saying $M(S) = 0$. □

Lemma 20 (lower semicontinuity of action functional). *The action functional $\mathcal{A}$ as defined in (12) is lower semicontinuous in both arguments. Specifically, if $\mu_n \rightharpoonup \mu$ weakly in $\mathscr{P}(\mathbb{R}^d)$ and $M_n \overset{*}{\rightharpoonup} M$ weakly* in $\mathcal{M}^d$, we have*

$$\mathcal{A}(\mu, M) \le \liminf_{n\to\infty} \mathcal{A}(\mu_n, M_n).$$

Proof. This result is an application of the general lsc result in [Buttazzo 1989, Theorem 3.4.3] since α satisfies the required convexity, lsc, and homogeneity assumptions by Remark 18. □

Another useful property of the action functional is the compactness provided by bounded action. We first state:

Lemma 21. *Let $F : \mathbb{R}^{2d} \to [0, \infty]$ be measurable and fix any $\mu \in \mathscr{P}(\mathbb{R}^d)$, $M \in \mathcal{M}^d$. We have the following bound*:

$$\iint_{\mathbb{R}^{2d}} F(v, v_*) d|M|(v, v_*) \le \sqrt{2}\mathcal{A}(\mu, M)^{1/2} \left(\iint_{\mathbb{R}^{2d}} F(v, v_*)^2 d\mu(v) d\mu(v_*) \right)^{1/2}. \tag{13}$$

Proof. This proof follows [Erbar 2023, Lemma 3.8]. We assume $\mathcal{A}(\mu, M) < +\infty$ or else (13) holds automatically. This implies that whenever $A \subset \mathbb{R}^{2d}$ is a measurable set, $\mu \otimes \mu(A) = 0$ if and only if $|M|(A) = 0$. Therefore, in the following computations we are implicitly integrating away from sets of zero $\mu \otimes \mu$-measure. We provide the simple argument by Cauchy–Schwarz for completeness. By considering $\tau = \mu \otimes \mu + |M|$, we estimate

$$\begin{aligned}\iint_{\mathbb{R}^{2d}} F d|M|(v, v_*) &\le \iint_{\mathbb{R}^{2d}} F \left| \frac{dM}{d\tau} \right| d\tau(v, v_*) = \iint_{\mathbb{R}^{2d}} F \left(\left| \frac{dM}{d\tau} \right| \Big/ \sqrt{2 \frac{d\mu \otimes \mu}{d\tau}} \right) \sqrt{2 \frac{d\mu \otimes \mu}{d\tau}} d\tau \\ &\le \left(\iint_{\mathbb{R}^{2d}} \alpha\left(\frac{dM}{d\tau}, \frac{d\mu \otimes \mu}{d\tau} \right) d\tau \right)^{1/2} \left(\iint_{\mathbb{R}^{2d}} 2F^2 d\mu \otimes \mu \right)^{1/2} \\ &= \sqrt{2}\mathcal{A}(\mu, M)^{1/2} \left(\iint_{\mathbb{R}^{2d}} F(v, v_*)^2 d\mu(v) d\mu(v_*) \right)^{1/2}.\end{aligned}$$ □

Remark 22. Suppose we have $\mu_t \in \mathscr{P}(\mathbb{R}^d)$ such that

$$\int_0^T m_2(\mu_t)\,dt = \int_0^T \int_{\mathbb{R}^d} \langle v\rangle^2\,d\mu_t(v)\,dt < \infty.$$

Then for $M \in \mathcal{M}_T^d$ the previous estimate (13) yields

$$\int_0^T \iint_{\mathbb{R}^{2d}} (1+|v|+|v_*|)\,d|M_t|(v,v_*)\,dt \lesssim \int_0^T \mathcal{A}(\mu_t, M_t)^{1/2}\Big(1+2\int_{\mathbb{R}^d}|v|^2\,d\mu_t\Big)^{1/2}\,dt. \tag{14}$$

Therefore, if the integral in time of the second moment of μ is bounded, then M satisfies the moments conditions (10) and the energy is conserved (Lemma 15). In the sequel, we will be considering curves that have bounded second moment which guarantee (14).

Proposition 23. *Let $(\mu_t^n, M_t^n)_n$ be a sequence in $\mathcal{GCE}_T$ such that $(\mu_0^n)_n$ is tight and we have the uniform bounds*

$$\sup_{n\in\mathbb{N}} \int_0^T \int_{\mathbb{R}^d} |v|^2\,d\mu_t^n\,dt < \infty \quad \textit{and} \quad \sup_{n\in\mathbb{N}} \int_0^T \mathcal{A}(\mu_t^n, M_t^n)\,dt < \infty. \tag{15}$$

Then, there exists $(\mu_t, M_t) \in \mathcal{GCE}_T$ such that, possibly after extracting a subsequence, we have the convergences

$$\begin{aligned} \mu_t^n &\rightharpoonup \mu_t && \textit{weakly in } \mathscr{P}(\mathbb{R}^d) \textit{ for all } t\in[0,T],\\ M_t^n\,dt &\overset{*}{\rightharpoonup} M_t dt && \textit{weakly* in } \mathcal{M}_T^d. \end{aligned}$$

Furthermore, along this subsequence we have the lower semicontinuity

$$\int_0^T \mathcal{A}(\mu_t, M_t)\,dt \le \liminf_{n\to\infty} \int_0^T \mathcal{A}(\mu_t^n, M_t^n)\,dt.$$

Sketch of the proof. This result follows from a proof similar to that of [Dolbeault et al. 2009, Lemma 4.5] and [Erbar 2023, Proposition 3.11], which we sketch. The second moment bound for μ^n in (15) produces a limit μ. Recalling the application of Lemma 21 in Remark 22, the bounded action in (15) and the estimate (14) produce a limit $M_t\,dt$ for a subsequence of $M_t^n\,dt$. The lower semicontinuity follows from Fatou's lemma and Lemma 20. □

3.3. ***Properties of the Landau metric.*** We define the distance, d_L induced by the action functional on $\mathscr{P}_{2,E}(\mathbb{R}^d)$. Throughout, we will be working in the grazing continuity equation space defined earlier by $\mathcal{GCE}_T^{2,E}$ for $T>0$ some terminal time and $E>0$ any second moment bound.

Definition 24. For $\lambda, \nu \in \mathscr{P}_{2,E}(\mathbb{R}^d)$ we define the (square of the) Landau distance by

$$d_L^2(\lambda,\nu) := \inf\bigg\{ T\int_0^T \mathcal{A}(\mu_t, M_t)\,dt \;\bigg|\; (\mu, M)\in\mathcal{GCE}_T^{2,E}(\lambda,\nu)\bigg\}. \tag{16}$$

Notice this definition is independent of $T>0$ considering the scaling of the grazing collision equation and the 1-homogeneity of $\mathcal{A}$. We have an equivalent characterization of d_L which can be seen in other PDE contexts such as [Erbar 2023; Dolbeault et al. 2009].

Lemma 25. *Given $\lambda, \nu \in \mathscr{P}_{2,E}(\mathbb{R}^d)$, we have*

$$d_L(\lambda, \nu) = \inf\left\{ \int_0^T \sqrt{\mathcal{A}(\mu_t, M_t)}\, dt \;\middle|\; (\mu, M) \in \mathcal{GCE}_T^{2,E}(\lambda, \nu) \right\}. \tag{17}$$

Proof. This proof uses the same reparametrization technique in [Dolbeault et al. 2009, Theorem 5.4]. □

Proposition 26 (minimizing curve). *Suppose that $\mu_0, \mu_1 \in \mathscr{P}_{2,E}(\mathbb{R}^d)$ are probability measures such that $d_L(\mu_0, \mu_1) < \infty$. Then there exists a curve $(\mu, M) \in \mathcal{GCE}_1^{2,E}(\mu_0, \mu_1)$ attaining the infimum of (16) (equivalently, also (17)) and $\mathcal{A}(\mu_t, M_t) = d_L^2(\mu_0, \mu_1)$ for almost every $t \in [0, 1]$.*

Proof. This result follows from the direct method of calculus of variations where the lower semicontinuity comes from Proposition 23. □

Proof of Theorem 7. We prove the statements in exactly the order they are presented in the theorem, starting with the properties of the proposed Landau distance as a metric. The positivity of d_L follows from the positivity of α. We now check that d_L satisfies the properties of a metric.

d_L distinguishes points: Fix $\mu_0, \mu_1 \in \mathscr{P}_{2,E}(\mathbb{R}^d)$. We check that $d_L(\mu_0, \mu_1) = 0 \iff \mu_0 = \mu_1$. Suppose that $d_L(\mu_0, \mu_1) = 0$. By Proposition 26 we can find $(\mu, M) \in \mathcal{GCE}_1^{2,E}(\mu_0, \mu_1)$, which is a minimizing curve and moreover $0 = d_L(\mu_0, \mu_1) = \mathcal{A}(\mu_t, M_t)$ implies $M = 0$. The grazing continuity equation reduces to $\partial_t \mu_t = 0$, which implies μ_t is constant in time.

The converse statement follows similarly by pairing the constant curve $\mu : t \mapsto \mu_0 = \mu_1$ with the zero measure so that $(\mu, 0) \in \mathcal{GCE}_1^{2,E}(\mu_0, \mu_1)$.

Symmetry: Symmetry follows because time can be reversed for every curve. For instance, if $(\mu, M) \in \mathcal{GCE}_T^{2,E}(\mu_0, \mu_1)$, then one can check that the pair

$$\mu^r : t \mapsto \mu(T - t), \qquad M^r : t \mapsto -M(T - t)$$

belongs to $\mathcal{GCE}_T^{2,E}(\mu_1, \mu_0)$ with the same action.

Triangle inequality: We sketch the argument using a gluing lemma as in [Dolbeault et al. 2009, Lemma 4.4]. Let $\mu^0, \mu^1, \mu^2 \in \mathscr{P}_{2,E}(\mathbb{R}^d)$ be such that $d_L(\mu^0, \mu^1) < \infty$ and $d_L(\mu^1, \mu^2) < \infty$. If not, $d_L(\mu^0, \mu^2) \le d_L(\mu^0, \mu^1) + d_L(\mu^1, \mu^2)$ holds trivially. By Proposition 26, we can find minimizing curves connecting these probability measures

$$\left\{ \begin{array}{ll} (\mu^{0\to1}, M^{0\to1}) & \in \mathcal{GCE}_1^{2,E}(\mu^0, \mu^1), \\ (\mu^{1\to2}, M^{1\to2}) & \in \mathcal{GCE}_1^{2,E}(\mu^1, \mu^2) \end{array} \right\}.$$

Their concatenation from time 0 to 1 is given by

$$\mu_t := \begin{cases} \mu_{2t}^{0\to1}, & 0 \le t \le \frac{1}{2}, \\ \mu_{2(t-1/2)}^{1\to2}, & \frac{1}{2} \le t \le 1, \end{cases} \qquad M_t := \begin{cases} 2M_{2t}^{0\to1}, & 0 \le t \le \frac{1}{2}, \\ 2M_{2t-1/2}^{1\to2}, & \frac{1}{2} < t \le 1. \end{cases}$$

One can check that $(\mu, M) \in \mathcal{GCE}_1^{2,E}(\mu^0, \mu^2)$, so it is an admissible competitor in the computation of $d_L(\mu^0, \mu^2)$. By looking at the action on the different time pieces, we obtain

$$d_L(\mu^0, \mu^2) \le \int_0^1 \sqrt{\mathcal{A}(\mu_t, M_t)}\, dt = d_L(\mu^0, \mu^1) + d_L(\mu^1, \mu^2).$$

d_L-convergence/boundedness implies weak convergence/compactness: Fix $\mu^n, \mu^\infty \in \mathscr{P}_{2,E}$ for $n \in \mathbb{N}$ such that $d_L(\mu^\infty, \mu^n) \to 0$ as $n \to \infty$. By Proposition 26, take minimizing curves $(\nu^n, M^n) \in \mathcal{GCE}_1^{2,E}(\mu^\infty, \mu^n)$ such that

$$d_L(\mu^\infty, \mu^n) = \sqrt{\mathcal{A}(\nu_t^n, M_t^n)} \quad \text{for a.e. } t \in [0, 1].$$

By compactness in Proposition 23, there are limits $(\nu, M) \in \mathcal{GCE}_1^{2,E}$ such that $\nu^n \rightharpoonup \nu$ and $M^n \overset{*}{\rightharpoonup} M$ up to a subsequence. Moreover, the lower semicontinuity in Proposition 23 gives

$$\mathcal{A}(\nu_t, M_t) \le \liminf_{n\to\infty} \mathcal{A}(\nu_t^n, M_t^n) = 0;$$

hence $M = 0$ so that ν is a constant in time. Since $\nu(0) = \mu^\infty$, this implies $\mu^\infty = \nu(1) = \lim_{n\to\infty} \mu^n$, which establishes the weak convergence.

$(\mathscr{P}_\tau, d_L)$ is a complete geodesic space: We start with the geodesic property from completely analogous arguments to [Erbar 2023]; the remaining statement that $\mathscr{P}_\tau$ equipped with d_L is a complete geodesic space follows. Fix $\tau \in \mathscr{P}_{2,E}(\mathbb{R}^d)$ with $\mu_0, \mu_1 \in \mathscr{P}_\tau$. The triangle inequality ensures $d_L(\mu_0, \mu_1) < \infty$ so Proposition 26 guarantees the existence of a minimizing curve $(\mu, M) \in \mathcal{GCE}_1^{2,E}(\mu_0, \mu_1)$. One easily sees that this also induces a minimizing curve for intermediate times. More precisely, for every $0 \le r \le s \le 1$, we have that $(t \mapsto \mu_{t+r}, t \mapsto M_{t+r}) \in \mathcal{GCE}_{s-r}^{2,E}(\mu_r, \mu_s)$ also minimizes $d_L(\mu_r, \mu_s)$.

To show completeness, let $(\mu^n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathscr{P}_\tau$. The sequence is certainly d_L-bounded so by Proposition 23, we can find, up to extraction of a weakly convergent subsequence, $\mu^\infty \in \mathscr{P}_{2,E}(\mathbb{R}^d)$ such that $\mu^n \rightharpoonup \mu^\infty$ in $\mathscr{P}_{2,E}(\mathbb{R}^d)$. Lower semicontinuity of d_L and the Cauchy property of the subsequence give

$$d_L(\mu^n, \mu^\infty) \le \liminf_{m\to\infty} d_L(\mu^n, \mu^m) \to 0 \quad \text{as } n \to \infty.$$

For any $n \in \mathbb{N}$ the triangle inequality gives

$$d_L(\mu^\infty, \tau) \le d_L(\mu^\infty, \mu^n) + d_L(\mu^n, \tau) < \infty,$$

So $\mu^\infty \in \mathscr{P}_\tau$. □

Proposition 27 (metric derivative). *A curve $(\mu_t)_{t\in[0,T]} \subset \mathscr{P}_{2,E}(\mathbb{R}^d)$ is absolutely continuous with respect to d_L if and only if there exists a Borel family $(M_t)_{t\in[0.T]}$ belonging to $\mathcal{M}_T^d$ such that $(\mu, M) \in \mathcal{GCE}_T^{2,E}$ with the property that*

$$\int_0^T \sqrt{\mathcal{A}(\mu_t, M_t)}\, dt < \infty.$$

In this equivalence, we have a bound on the metric derivative

$$\lim_{h\downarrow 0} \frac{d_L^2(\mu_{t+h}, \mu_t)}{h^2} =: |\dot{\mu}|^2(t) \le \mathcal{A}(\mu_t, M_t) \quad \textit{for a.e. } t \in (0, T).$$

Furthermore, there exists a unique Borel family $(\widetilde{M}_t)_{t\in[0,T]}$ belonging to $\mathcal{M}^d$ which is characterized by

$$M_t = U\mu_t \otimes \mu_t \quad \textit{and} \quad U \in T_\mu := \overline{\{\tilde{\nabla}\phi \mid \phi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\mu_t\otimes\mu_t)}$$

such that $(\mu, \widetilde{M}) \in \mathcal{GCE}_T^E(\mu_0, \mu_T)$ where we have equality

$$|\dot{\mu}|^2(t) = \mathcal{A}(\mu_t, \widetilde{M}_t) \quad \textit{for a.e. } t \in (0, T).$$

Proof. The argument follows exactly as in [Dolbeault et al. 2009, Theorem 5.17]. □

4. Energy dissipation equality

The goal in this section is to prove Theorem 8, which states that the notions of gradient flow solutions coincide with ϵ-solutions to the Landau equation. To fix ideas, we recall the regularized entropy functionals acting on probability measures

$$\mathcal{H}_\epsilon[\mu] = \int_{\mathbb{R}^d} (\mu * G^\epsilon)(v) \log(\mu * G^\epsilon)(v)\, dv,$$

with $G^\epsilon(v)$ given by

$$G^\epsilon(v) = \epsilon^{-d} C_d \exp\left\{-\left\langle \frac{v}{\epsilon}\right\rangle\right\}.$$

The crucial ingredient to prove Theorem 8 is the following:

Proposition 28 (chain rule ϵ). *Fix $\gamma \in [-4, 0]$ and suppose $(\mu, M) \in \mathcal{GCE}_T^{2,E}$ and*

$$\int_0^T \mathcal{A}(\mu_t, M_t)\, dt < \infty.$$

Then, $\sup_{t\in[0,T]} \mathcal{H}_\epsilon[\mu_t] < \infty$ *and the "chain rule" holds*:

$$\mathcal{H}_\epsilon[\mu_r] - \mathcal{H}_\epsilon[\mu_s] = \frac{1}{2}\int_s^r \iint_{\mathbb{R}^{2d}} \tilde{\nabla}\left[\frac{\delta \mathcal{H}_\epsilon}{\delta\mu}\right] \cdot dM_t\, dt \quad \textit{for all } 0 \le s \le r \le T. \tag{18}$$

Remark 29. Recall the expression for the dissipation

$$D_\epsilon[\mu] = \frac{1}{2}\iint_{\mathbb{R}^{2d}} \left|\tilde{\nabla}\left[\frac{\delta \mathcal{H}_\epsilon}{\delta\mu}\right]\right|^2 d\mu(v) d\mu(v_*).$$

Using a time integrated version of Lemma 21, we have the estimate

$$\frac{1}{2}\int_s^r \iint_{\mathbb{R}^{2d}} \left|\tilde{\nabla}\left[\frac{\delta \mathcal{H}_\epsilon}{\delta\mu}\right]\right| \cdot d|M_t|(v, v_*)\, dt \le \int_s^r \mathcal{A}(\mu_t, M_t)^{1/2} D_\epsilon[\mu_t]^{1/2}\, dt.$$

Therefore, under the hypothesis of Proposition 28, we have that

$$|\mathcal{H}_\epsilon(\mu_r) - \mathcal{H}_\epsilon(\mu_r)| \le \int_s^r |\dot{\mu}|(t) D_\epsilon[\mu_t]^{1/2}\, dt,$$

which implies that $D_\epsilon[\mu_t]^{1/2}$ is a strong upper gradient of $\mathcal{H}_\epsilon$; see Definition 4.

Taking Proposition 28 for granted, we can prove Theorem 8.

Proof of Theorem 8. Throughout, $\mu = f\mathcal{L}$ is a curve of probability measures with uniformly bounded second moment.

<u>Weak ϵ-solution $\Longrightarrow$ curve of maximal slope</u>: Consider f an ϵ-solution to the Landau equation. Define $m = -ff_*\tilde{\nabla}(\delta\mathcal{H}_\epsilon/\delta f)$ so that the pair of measures $(\mu = f\mathcal{L}, M = m\mathcal{L}\otimes\mathcal{L})$ therefore belong to $\mathcal{GCE}_T^E$. Indeed, the distributional grazing continuity equation from Definition 17 is precisely the weak ϵ-Landau equation. Based on the definition of M and the finite $\mathcal{H}_\epsilon$ dissipation, we have the bound

$$\int_0^T \mathcal{A}(\mu_t, M_t)\, dt = \int_0^T D_\epsilon(f_t)\, dt < \infty,$$

which implies the weak continuity of μ. By Proposition 27, we have

$$|\dot{\mu}|^2(t) = \mathcal{A}(\mu_t, M_t) = D_\epsilon(f_t) < \infty \quad \text{for a.e. } t \in [0, T].$$

Using Proposition 28, we have, for any $0 \le s \le r \le T$,

$$\mathcal{H}_\epsilon[\mu_r] - \mathcal{H}_\epsilon[\mu_s] + \frac{1}{2}\int_s^r D_\epsilon(\mu_t)\,dt + \frac{1}{2}\int_s^r |\dot{\mu}|^2(t)\,dt \le 0.$$

According to Definition 5, this is the curve of maximal slope property.

Curve of maximal slope $\Longrightarrow$ weak ϵ-solution: Assume that $\mu = f\mathcal{L}$ is a curve of maximal slope for $\mathcal{H}_\epsilon$ with respect to the upper gradient $\sqrt{D_\epsilon}$. Since μ is absolutely continuous with respect to d_L, Proposition 27 guarantees existence of a unique curve $M : t \in [0, T] \mapsto M_t \in \mathcal{M}^d$ such that $\int_0^T \sqrt{\mathcal{A}(\mu_t, M_t)}\,dt < \infty$ and $|\dot{\mu}|^2(t) = \mathcal{A}(\mu_t, M_t)$ for a.e. $t \in [0, T]$. Furthermore, $(\mu, M) \in \mathcal{GCE}_T^E$. According to Lemma 19, let $M = m\mathcal{L} \otimes \mathcal{L}$ for some measurable function m. We apply the chain rule (18) with Cauchy–Schwarz and Young's inequalities with minus signs in the following computations:

$$\begin{aligned}
\mathcal{H}_\epsilon[f_T] - \mathcal{H}_\epsilon[f_0] &= \frac{1}{2}\int_0^T \iint_{\mathbb{R}^{2d}} \tilde{\nabla}\frac{\delta \mathcal{H}_\epsilon}{\delta f} \cdot m\,dv\,dv_*\,dt \\
&\ge -\frac{1}{2}\int_0^T \left(\iint_{\mathbb{R}^{2d}} f f_* \left|\tilde{\nabla}\frac{\delta \mathcal{H}_\epsilon}{\delta f}\right|^2 dv\,dv_*\right)^{1/2} \left(\iint_{\mathbb{R}^{2d}} \frac{|m|^2}{f f_*}\,dv\,dv_*\right)^{1/2} dt \\
&\ge -\frac{1}{2}\int_0^T \left(\frac{1}{2}\iint_{\mathbb{R}^{2d}} f f_* \left|\tilde{\nabla}\frac{\delta \mathcal{H}_\epsilon}{\delta f}\right|^2 dv\,dv_*\right) dt - \frac{1}{2}\int_0^T \left(\frac{1}{2}\iint_{\mathbb{R}^{2d}} \frac{|m|^2}{f f_*}\,dv\,dv_*\right) dt \\
&= -\frac{1}{2}\int_0^T D_\epsilon(f_t)\,dt - \frac{1}{2}\int_0^T |\dot{f}|^2(t)\,dt.
\end{aligned}$$

All the inequalities in the calculations above are actually equalities owing to the fact that μ is a curve of maximal slope. In particular, since we have the equality in the Young's inequality, this implies

$$\frac{m}{\sqrt{f f_*}} = -\sqrt{f f_*}\,\tilde{\nabla}\frac{\delta \mathcal{H}_\epsilon}{\delta f}.$$

As in the previous direction, the weak ϵ Landau equation coincides with the grazing continuity equation when m is equal to $-f f_* \tilde{\nabla}(\delta \mathcal{H}_\epsilon / \delta f)$. □

The rest of this section is devoted to proving Proposition 28. We need some lemmas to establish crucial estimates. The following result is a variation of [Carlen and Carvalho 1992, Lemma 2.6].

Lemma 30 [Carlen and Carvalho 1992]. *Let μ be a probability measure on $\mathbb{R}^d$ with finite second moment/energy, $m_2(\mu) \le E$ for $E > 0$. Then, for every $\epsilon > 0$, there exists a constant $C = C(\epsilon, E) > 0$ such that*

$$|\log(\mu * G^\epsilon)(v)| \le C\left\langle \frac{v}{\epsilon} \right\rangle.$$

Proof. Starting with an upper bound, we easily see

$$\mu * G^\epsilon(v) = \int_{\mathbb{R}^d} G^\epsilon(v - v')\,d\mu(v') \lesssim_\epsilon 1.$$

Turning to the lower bound, we cut off the integration domain to $|v'| \le R$ for some $R > 0$ to be chosen later. We estimate, for $\epsilon > 0$ small enough,

$$\left\langle \frac{v-v'}{\epsilon} \right\rangle = \sqrt{1 + \left|\frac{v-v'}{\epsilon}\right|^2} \le \sqrt{1 + 2\left|\frac{v}{\epsilon}\right|^2 + 2\left(\frac{R}{\epsilon}\right)^2} \le \sqrt{2}\left(\left\langle \frac{v}{\epsilon}\right\rangle + \left\langle \frac{R}{\epsilon}\right\rangle\right).$$

This is substituted into $G^\epsilon(v-v')$ to obtain

$$\mu * G^\epsilon(v) \ge \int_{|v'|\le R} G^\epsilon(v-v')\, d\mu(v') \gtrsim_\epsilon \exp\left\{-\sqrt{2}\left(\left\langle \frac{v}{\epsilon}\right\rangle + \left\langle \frac{R}{\epsilon}\right\rangle\right)\right\} \int_{|v'|\le R} d\mu(v').$$

At this point, we appeal to Chebyshev's inequality to see

$$\int_{|v'|\le R} d\mu(v') = 1 - \int_{|v'|\ge R} d\mu(v') \ge 1 - \frac{1}{R^2}\int_{|v'|\ge R} |v'|^2\, d\mu(v').$$

We can now choose, for example, large R such that $1 - E/R^2 \ge \frac{1}{2}$ to uniformly lower bound the integral $\int_{|v'|\le R} d\mu(v')$ away from 0 and then conclude the result after applying logarithms. □

Lemma 31 (log-derivative estimates). *For fixed $\epsilon > 0$ we have the formula*

$$\nabla G^\epsilon(v) = -\frac{1}{\epsilon}\left\langle \frac{v}{\epsilon}\right\rangle^{-1} G^\epsilon(v)\frac{v}{\epsilon}. \tag{19}$$

For $\mu \in \mathscr{P}(\mathbb{R}^d)$, defining $\partial^i = \partial/\partial v^i$ and $\partial^{ij} = \partial^2/(\partial v^i \partial v^j)$, we obtain

$$|\nabla \log(\mu * G^\epsilon)(v)| \le \frac{1}{\epsilon}, \quad |\partial^{ij} \log(\mu * G^\epsilon)(v)| \le \frac{4}{\epsilon^2}. \tag{20}$$

Proof. Equation (19) is a direct computation after noticing

$$\frac{\nabla G^\epsilon}{G^\epsilon} = \nabla \log G^\epsilon = \nabla\left(-\left\langle \frac{v}{\epsilon}\right\rangle + \text{const.}\right) = -\frac{1}{\epsilon}\left\langle \frac{v}{\epsilon}\right\rangle^{-1}\frac{v}{\epsilon}.$$

The first order log-derivative estimate of (20) is calculated using formula (19) to obtain

$$\begin{aligned}|\nabla(\mu * G^\epsilon)(v)| = |\mu * \nabla G^\epsilon(v)| &\le \frac{1}{\epsilon}\int_{\mathbb{R}^d} \left\langle \frac{v-v'}{\epsilon}\right\rangle^{-1} \left|\frac{v-v'}{\epsilon}\right| G^\epsilon(v-v')\, d\mu(v') \\ &\le \frac{1}{\epsilon}\int_{\mathbb{R}^d} G^\epsilon(v-v')\, d\mu(v') = \frac{1}{\epsilon}(\mu * G^\epsilon)(v).\end{aligned}$$

For the second order, we first look at $\partial^{ij}\mu * G^\epsilon$ which can be computed with the help of (19):

$$\begin{aligned}&|\partial^{ij}\mu * G^\epsilon(v)| \\ &\quad = \left|\partial^i\left(-\frac{1}{\epsilon}\int_{\mathbb{R}^d}\left\langle \frac{v-v'}{\epsilon}\right\rangle^{-1}\frac{v^j - v'^j}{\epsilon} G^\epsilon(v-v')\, d\mu(v')\right)\right| \\ &\quad = \left|\frac{1}{\epsilon^2}\int_{\mathbb{R}^d}\left(\left\langle \frac{v-v'}{\epsilon}\right\rangle^{-3}\frac{v^i-v'^i}{\epsilon}\frac{v^j-v'^j}{\epsilon} + \delta^{ij}\left\langle \frac{v-v'}{\epsilon}\right\rangle^{-1} - \left\langle \frac{v-v'}{\epsilon}\right\rangle^{-2}\frac{v^i-v'^i}{\epsilon}\frac{v^j-v'^j}{\epsilon}\right) G^\epsilon(v-v')\, d\mu(v')\right| \\ &\quad \le \frac{3}{\epsilon^2}\mu * G^\epsilon(v).\end{aligned}$$

Combining this estimate with the previous first-order one, we have

$$|\partial^{ij}\log(\mu * G^\epsilon)(v)| = \left|\frac{\partial^{ij}\mu * G^\epsilon}{\mu * G^\epsilon} - \frac{(\partial^i\mu * G^\epsilon)(\partial^j\mu * G^\epsilon)}{(\mu * G^\epsilon)^2}\right| \le \frac{4}{\epsilon^2}. \qquad \square$$

Lemma 32. *Fix $\epsilon > 0$ and $\gamma \in [-4, 0]$, with $\mu \in \mathscr{P}_{2,E}(\mathbb{R}^d)$ for some $E > 0$. We have*

(1) *Moderately soft case $\gamma \in [-2, 0]$*:

$$\left|\tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu}\right| = \left|\tilde{\nabla}[G^\epsilon * \log(\mu * G^\epsilon)](v, v_*)\right| \lesssim_\epsilon |v|^{1+\gamma/2} + |v_*|^{1+\gamma/2}.$$

(2) *Very soft case $\gamma \in [-4, -2]$*:

$$\left|\tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu}\right| \lesssim_\epsilon 1.$$

In particular, it holds

$$\iint_{\mathbb{R}^{2d}}\left|\tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu}\right|^2 d\mu(v)d\mu(v_*) \le E.$$

Proof. We develop the expression for $\tilde{\nabla}(\delta\mathcal{H}_\epsilon/\delta\mu)$ in integral form to be used throughout this proof:

$$\begin{aligned}\tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu} &= \tilde{\nabla}G^\epsilon * \log(\mu * G^\epsilon)(v, v_*)\\ &= |v - v_*|^{1+\gamma/2}\Pi[v - v_*](\nabla_v G^\epsilon * \log(\mu * G^\epsilon)(v) - \nabla_{v_*}G^\epsilon * \log(\mu * G^\epsilon)(v_*))\\ &= |v - v_*|^{1+\gamma/2}\Pi[v - v_*]\int_{\mathbb{R}^d} G^\epsilon(v')\left(\frac{\nabla\mu * G^\epsilon}{\mu * G^\epsilon}(v - v') - \frac{\nabla\mu * G^\epsilon}{\mu * G^\epsilon}(v_* - v')\right)dv'. \end{aligned} \tag{21}$$

(1) Moderately soft case $\gamma \in [-2, 0]$: We use (a concave version of) the triangle inequality (valid since $1 + \gamma/2 \ge 0$) and the first estimate of (20) to bound the last line of (21):

$$\left|\tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu}\right| \le 2^{1+\gamma/2}(|v|^{1+\gamma/2} + |v_*|^{1+\gamma/2})\frac{2}{\epsilon}\int_{\mathbb{R}^d} G^\epsilon(v')\,dv' \lesssim_\epsilon |v|^{1+\gamma/2} + |v_*|^{1+\gamma/2}.$$

(2) Very soft case $\gamma \in [-4, -2]$: We perform estimates in two cases, the far field $|v - v_*| \ge 1$ and near field $|v - v_*| \le 1$.

$|v - v_*| \ge 1$: In the far field, we have $|v - v_*|^{1+\gamma/2} \le 1$; hence we can brutally estimate (21) using again the first estimate of (20) to obtain, similar to the moderately soft case, the estimate

$$\left|\tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu}\right| \le \frac{2}{\epsilon}.$$

$|v - v_*| \le 1$: We can remove the singularity from the weight with a mean-value estimate and the second estimate of (20):

$$\left|\frac{\nabla\mu * G^\epsilon}{\mu * G^\epsilon}(v - v') - \frac{\nabla\mu * G^\epsilon}{\mu * G^\epsilon}(v_* - v')\right| \le \sup_{i,j=1,\dots,d}\left\|\partial^i\left(\frac{\partial^j\mu * G^\epsilon}{\mu * G^\epsilon}\right)\right\|_{L^\infty}|v - v_*| \le \frac{4}{\epsilon^2}|v - v_*|.$$

Inserting this into (21), we have

$$\left|\tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu}\right| \le \frac{4}{\epsilon^2}|v-v_*|^{2+\gamma/2}\int_{\mathbb{R}^d} G^\epsilon(v')\,dv' \le \frac{4}{\epsilon^2}. \qquad \square$$

Remark 33. Originally, we considered the general family of convolution kernels $G^{s,\epsilon}$ described in Section 2.1. Besides the context of the Landau equation, Lemma 31 (excluding the second-order log-derivative estimate) can be generalized to this family of s-order tailed exponential distributions with additional moment assumptions on μ. In particular, (19) and (20) (for $s \ge 1$) become

$$\frac{\nabla G^{s,\epsilon}}{G^{s,\epsilon}}(v) = -\frac{s}{\epsilon}\left\langle\frac{v}{\epsilon}\right\rangle^{s-2}\frac{v}{\epsilon}, \qquad \frac{|\nabla(\mu * G^{s,\epsilon})|}{\mu * G^{s,\epsilon}}(v) \lesssim \frac{1}{\epsilon^s}\langle v\rangle^{s-1}.$$

Since Maxwellians are known to be stationary solutions for the Landau equation, we wanted to perform the regularization with $s=2$. However, the analogous estimates of Lemma 31 for $s=2$ are not sufficient for Lemma 32 in the $\mathscr{P}_2$ framework. For example, in the moderately soft potential case, the estimate reads

$$\left|\tilde{\nabla}\frac{\delta\mathcal{H}_{2,\epsilon}}{\delta\mu}\right| \lesssim_\epsilon \langle v\rangle^{2+\gamma/2} + \langle v_*\rangle^{2+\gamma/2} \notin L^2(\mu\otimes\mu).$$

However, there is one value of $\gamma=-2$ for which the estimates hold when using a Maxwellian regularization kernel $G^{2,\epsilon}$. A restriction to $\mathscr{P}_4$ resolves the issue mentioned above for the moderately soft potential case, but then a fourth moment propagation is needed, which we did not pursue. A similar issue is present in the very soft potential case.

Proof of Proposition 28. To prove (18), our strategy is to regularize the pair (μ, M) in time with parameter $\delta > 0$ and differentiate the regularization. Then we obtain uniform bounds in δ needed to take the limit $\delta \to 0$.

Finite regularized entropy: We have the following chain of inequalities:

$$\mathcal{H}_\epsilon[\mu_t] = \int_{\mathbb{R}^d}(\mu_t * G^\epsilon)(v)\log(\mu_t * G^\epsilon)(v)\,dv \lesssim_{\epsilon,E} \int_{\mathbb{R}^d}(\mu_t * G^\epsilon)(v)\langle v\rangle\,dv \lesssim_\epsilon 1+E.$$

The first inequality comes from Lemma 30 because $\log(\mu_t * G^\epsilon)$ has linear growth (uniform in time) while in the second inequality, one realizes that $\mu_t * G^\epsilon$ has as many moments as μ_t with computable constants.

Time regularization with $\delta > 0$: Without loss of generality, let μ be the weakly time continuous representative (Lemma 14) and M be the optimal grazing rate (Proposition 27) achieving the finite distance d_L. We first regularize the pair (μ, M) in time for a fixed parameter $\delta > 0$ as follows. Take $\eta \in C_0^\infty(\mathbb{R})$ with the following properties:

$$\operatorname{supp}\eta \subset (-1,1), \quad \eta \ge 0, \quad \eta(t) = \eta(-t), \quad \int_{-1}^1 \eta(t)\,dt = 1.$$

We define the following measures for $t \in [0, T]$, by taking convex combinations:

$$\mu_t^\delta := \int_{-1}^1 \eta(t')\mu_{t-\delta t'}\,dt', \qquad M_t^\delta := \int_{-1}^1 \eta(t')M_{t-\delta t'}\,dt'.$$

Here, we constantly extend the measures in time. That is, if $t-\delta t' \in [-\delta, 0]$, we treat $\mu_{t-\delta t'} = \mu_0$, $M_{t-\delta t'} = 0$. For the other end point, if $t-\delta t' \in [T, T+\delta]$, we set $\mu_{t-\delta t'} = \mu_T$, $M_{t-\delta t'} = 0$. This transformation is stable so that $(\mu^\delta, M^\delta) \in \mathcal{GCE}_T$ and in particular, the distributional grazing continuity equation holds:

$$\partial_t \mu_t^\delta + \tfrac{1}{2}\tilde{\nabla} \cdot M_t^\delta = 0.$$

We derive (18) using this regularized grazing continuity equation. Consider

$$\mathcal{H}_\epsilon[\mu_t^\delta] = \int_{\mathbb{R}^d} (\mu_t^\delta * G^\epsilon)(v) \log(\mu_t^\delta * G^\epsilon)(v)\, dv,$$

which we differentiate with respect to t by appealing to the dominated convergence theorem. Firstly, due to the time regularization, we have

$$\partial_t \{(\mu_t^\delta * G^\epsilon) \log(\mu_t^\delta * G^\epsilon)\} = [(\partial_t \mu_t^\delta) * G^\epsilon](\log(\mu_t^\delta * G^\epsilon) + 1).$$

The L_v^1 bound is obtained on the following difference quotient for a fixed time step $h > 0$:

$$\begin{aligned}&\left|\frac{1}{h}[(\mu_{t+h}^\delta * G^\epsilon) \log(\mu_{t+h}^\delta * G^\epsilon) - (\mu_t^\delta * G^\epsilon) \log(\mu_t^\delta * G^\epsilon)]\right| \\ &\qquad\qquad \le \frac{1}{h}|(\mu_{t+h}^\delta * G^\epsilon) - (\mu_t^\delta * G^\epsilon)| \sup_{s\in[t,t+h]} |\log(\mu_s^\delta * G^\epsilon) + 1|,\end{aligned}$$

where we have used the mean value theorem with the chain rule. Applying Lemma 30, we obtain

$$\left|\frac{1}{h}[(\mu_{t+h}^\delta * G^\epsilon) \log(\mu_{t+h}^\delta * G^\epsilon) - (\mu_t^\delta * G^\epsilon) \log(\mu_t^\delta * G^\epsilon)]\right| \lesssim_{\epsilon,E} \frac{1}{h}|(\mu_{t+h}^\delta * G^\epsilon) - (\mu_t^\delta * G^\epsilon)|\langle v\rangle.$$

We apply the mean value theorem on the difference quotient again to get

$$\left|\frac{1}{h}[(\mu_{t+h}^\delta * G^\epsilon) \log(\mu_{t+h}^\delta * G^\epsilon) - (\mu_t^\delta * G^\epsilon) \log(\mu_t^\delta * G^\epsilon)]\right| \lesssim_{\delta,\epsilon} \|\eta'\|_{L^\infty}\left(\mu_0 * G^\epsilon + \int_0^T \mu_t * G^\epsilon\, dt\right)\langle v\rangle.$$

Since μ has finite second-order moments, this last expression belongs to L_v^1. By the dominated convergence theorem,

$$\frac{d}{dt}\mathcal{H}_\epsilon[\mu_t^\delta] = \int_{\mathbb{R}^d} [(\partial_t \mu_t^\delta) * G^\epsilon](\log(\mu_t^\delta * G^\epsilon) + 1)\, dv = \int_{\mathbb{R}^d} (\partial_t \mu_t^\delta) \cdot [G^\epsilon * \log(\mu_t^\delta * G^\epsilon)]\, dv.$$

The last line is achieved by the self-adjointness of convolution with G^ϵ and eliminating the constant term due to the conserved mass of μ^δ. Integrating in t, we obtain

$$\begin{aligned}\mathcal{H}_\epsilon[\mu_r^\delta] - \mathcal{H}_\epsilon[\mu_s^\delta] &= \int_s^r \int_{\mathbb{R}^d} (\partial_t \mu_t^\delta) \cdot [G^\epsilon * \log(\mu_t^\delta * G^\epsilon)]\, dv\, dt \\ &= \frac{1}{2}\int_s^r \iint_{\mathbb{R}^{2d}} [\tilde{\nabla} G^\epsilon * \log(\mu_t^\delta * G^\epsilon)] \cdot dM_t^\delta\, dt \\ &= \frac{1}{2}\int_s^r \iint_{\mathbb{R}^{2d}} \tilde{\nabla} \frac{\delta \mathcal{H}_\epsilon}{\delta \mu_t^\delta} \cdot dM_t^\delta\, dt. \end{aligned} \tag{22}$$

We now turn to establishing estimates independent of $\delta > 0$ to pass to the limit.

Estimates on the right-hand side of (22): According to Lemma 32, we have the estimate

$$\left|\tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu^\delta}\right| \lesssim_{\epsilon,E} |v|^p + |v_*|^p,$$

where $p \le 1$. By the first moment assumption of M_t, we have

$$\int_0^T\iint_{\mathbb{R}^{2d}}\left|\tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t^\delta}\right| d|M_t|(v,v_*)\,dt \lesssim_{\epsilon,E} \int_0^T\iint_{\mathbb{R}^{2d}} |v|+|v_*|\,d|M_t|(v,v_*)\,dt < \infty.$$

This estimate also extends to M_t^δ

$$\int_0^T\iint_{\mathbb{R}^{2d}}\left|\tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t^\delta}\right| d|M_t^\delta|(v,v_*)\,dt < \infty.$$

Note that these estimates are independent of $\delta > 0$.

Convergence $\delta \to 0$: Firstly, we establish the following identity which will be useful later. For fixed functions f^1, f^2 we have

$$\begin{aligned}\tilde{\nabla}&[G^\epsilon * f^1] - \tilde{\nabla}[G^\epsilon * f^2]\\ &= |v-v_*|^{1+\gamma/2}\Pi[v-v_*](\nabla[G^\epsilon * f^1] - \nabla[G^\epsilon * f^2] - (\nabla_*[G^\epsilon * f^1]_* - \nabla_*[G^\epsilon * f^2]_*))\\ &= |v-v_*|^{1+\gamma/2}\Pi[v-v_*]\int_{\mathbb{R}^d}(\nabla G^\epsilon(v-v') - \nabla G^\epsilon(v_*-v'))(f^1(v')-f^2(v'))\,dv'. \end{aligned} \tag{23}$$

Using the weak in time continuity of μ, we can consider

$$|\mu_t^\delta * G^\epsilon(v') - \mu_t * G^\epsilon(v')| \le \int_{-1}^1 \eta(t')|\langle\mu_{t-\delta t'}, G^\epsilon(v'-\cdot)\rangle - \langle\mu_t, G^\epsilon(v'-\cdot)\rangle|\,dt'.$$

The "$\cdot$" stands for the convoluted variable. Since t belongs to a compact set, the function $t \mapsto \langle\mu_t, G^\epsilon(v'-\cdot)\rangle$ is *uniformly* continuous from the weak continuity of μ. In particular, using the continuity in v' and the lower bound from Lemma 30 we conclude that for any $R > 0$

$$|\log(\mu_t^\delta * G^\epsilon) - \log(\mu_t * G^\epsilon)| \to 0 \quad \text{uniformly on } B_R. \tag{24}$$

Therefore by Lemma 30, defining $w = |v-v_*|^{1+\gamma/2}$, and using (23) with $f^1 = \log(\mu_t^\delta * G^\epsilon)$ and $f^2 = \log(\mu_t * G^\epsilon)$, we have

$$\begin{aligned}\left|\tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t^\delta} - \tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t}\right| &= |\tilde{\nabla}G^\epsilon * \log(\mu_t^\delta * G^\epsilon)(v,v_*) - \tilde{\nabla}G^\epsilon * \log(\mu_t * G^\epsilon)(v,v_*)|\\ &\le \int_{\mathbb{R}^d} w|\nabla G^\epsilon(v-v') - \nabla G^\epsilon(v_*-v')||\log(\mu_t^\delta * G^\epsilon(v')) - \log(\mu_t * G^\epsilon(v'))|\,dv'\\ &\le \int_{B_{R_0}^c} w|\nabla G^\epsilon(v-v') - \nabla G^\epsilon(v_*-v')|C_\epsilon\langle v'\rangle\,dv'\\ &\qquad + \sup_{B_{R_0}}|\log(\mu_t^\delta * G^\epsilon) - \log(\mu_t * G^\epsilon)|\int_{B_{R_0}} w|\nabla G^\epsilon(v-v') - \nabla G^\epsilon(v_*-v')|\,dv'.\end{aligned}$$

For a fixed (v, v_*), we obtain the convergence to zero by taking $\delta \to 0$ and $R_0 \to \infty$ in the previous estimate. This holds for all $\gamma \in [-4, 0]$ by taking advantage of the regularity of G^ϵ. Using continuity, we obtain that for any $R > 0$

$$\left|\tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t^\delta}(v, v_*) - \tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t}(v, v_*)\right| \to 0 \quad \text{uniformly on } [0, T] \times B_R \times B_R. \tag{25}$$

We turn to the limit estimate for the right-hand side of (22). For any $R > 0$, we have

$$\begin{aligned}
&\left|\int_s^r \iint_{\mathbb{R}^{2d}} \tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t^\delta} \cdot dM_t^\delta \, dt - \int_s^r \iint_{\mathbb{R}^{2d}} \tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t} \cdot dM_t \, dt\right| \\
&\quad \le \left|\int_s^r \iint_{\mathbb{R}^{2d}} \left(\tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t^\delta} - \tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t}\right) \cdot dM_t^\delta \, dt\right| + \left|\int_s^r \iint_{\mathbb{R}^{2d}} \tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t} \cdot dM_t^\delta \, dt - \int_s^r \iint_{\mathbb{R}^{2d}} \tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t} \cdot dM_t \, dt\right| \\
&\quad \le \int_s^r \iint_{B_R \times B_R} \left|\tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t^\delta} - \tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t}\right| d|M_t^\delta| \, dt + \int_s^r \iint_{(B_R \times B_R)^C} \left|\tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t^\delta} - \tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t}\right| d|M_t^\delta| \, dt + o(1).
\end{aligned}$$

The last term is $o(1)$ as $\delta \to 0$ due to similar estimates from the previous step. By sending $\delta \to 0$ (the first term vanishes due to (25)) and then sending $R \to \infty$ (the second term vanishes again due to the estimate from the previous step), we obtain the convergence

$$\lim_{\delta\to 0} \frac{1}{2} \int_s^r \iint_{\mathbb{R}^{2d}} \tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t^\delta} \cdot dM_t^\delta \, dt = \frac{1}{2} \int_s^r \iint_{\mathbb{R}^{2d}} \tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t} \cdot dM_t^\delta \, dt. \tag{26}$$

<u>Convergence of the left-hand side of (22)</u>: By (24), Lemma 30 and the uniform bound on the second moment, we have

$$\begin{aligned}
|\mathcal{H}_\epsilon[\mu_t^\delta] - \mathcal{H}_\epsilon[\mu_t]| &\le \int_{\mathbb{R}^d} |(\mu_t^\delta * G^\epsilon)\log(\mu_t^\delta * G^\epsilon)(v) - (\mu_t * G^\epsilon)\log(\mu_t * G^\epsilon)(v)| \, dv \\
&\to 0 \quad \text{as } \delta \to 0.
\end{aligned}$$

Therefore, by the previous equation and (26) we can take $\delta \to 0$ in (22) to obtain

$$\mathcal{H}_\epsilon[\mu_r] - \mathcal{H}_\epsilon[\mu_s] = \frac{1}{2} \int_s^r \iint_{\mathbb{R}^{2d}} \tilde{\nabla}\frac{\delta\mathcal{H}_\epsilon}{\delta\mu_t} \cdot dM_t(v, v_*) \, dt,$$

which is the desired result. □

5. JKO scheme for ϵ-Landau equation

This section is devoted to the proof of Theorem 9 after a series of preliminary lemmas. Our construction of curves of maximal slope in Theorem 9 uses the basic minimizing movement/variational approximation scheme of [Jordan et al. 1998]. Fix a small time step $\tau > 0$ and initial datum $\mu_0 \in \mathscr{P}_{2,E}(\mathbb{R}^d)$ and consider the recursive minimization procedure for $n \in \mathbb{N}$

$$\nu_0^\tau := \mu_0, \quad \nu_n^\tau \in \operatorname{argmin}_{\lambda \in \mathscr{P}_{2,E}} \left[\mathcal{H}_\epsilon(\lambda) + \frac{1}{2\tau} d_L^2(\nu_{n-1}^\tau, \lambda)\right]. \tag{27}$$

Then, we concatenate these minimizers into a curve by setting

$$\mu_0^\tau := \mu_0, \qquad \mu_t^\tau := \nu_n^\tau \quad \text{for } t \in ((n-1)\tau, n\tau]. \tag{28}$$

The scheme given by (27) and (28) satisfies the abstract formulation in [Ambrosio et al. 2008] giving:

Proposition 34 (Landau JKO scheme). *For any $\tau > 0$ and $\mu_0 \in \mathscr{P}_{2,E}(\mathbb{R}^d)$, there exists $\nu_n^\tau \in \mathscr{P}_{2,E}(\mathbb{R}^d)$ for every $n \in \mathbb{N}$ as described in (27). Furthermore, up to a subsequence of μ_t^τ described in (28) as $\tau \to 0$, there exists a locally absolutely continuous curve $(\mu_t)_{t\geq 0}$ such that*

$$\mu_t^\tau \rightharpoonup \mu_t \quad \textit{for all } t \in [0, \infty).$$

Proof. Our metric setting is $(\mathscr{P}_{\mu_0}, d_L)$ (see Theorem 7) with the weak topology σ. This space is essentially $\mathscr{P}_{2,E}(\mathbb{R}^d)$ except we need to make sure that d_L is a proper metric; hence we remove the probability measures with infinite Landau distance. We follow the proof of [Erbar 2023], which consists in verifying [Ambrosio et al. 2008, Assumptions 2.1(a)–(c)]. These assumptions are listed and verified now.

(1) $\mathcal{H}_\epsilon$ is sequentially σ-lsc on d_L-bounded sets: Suppose $\mu_n \in \mathscr{P}_{2,E}(\mathbb{R}^d) \rightharpoonup \mu \in \mathscr{P}_{2,E}(\mathbb{R}^d)$. This implies $\mu_n * G^\epsilon \rightharpoonup \mu * G^\epsilon$ in $\mathscr{P}_2(\mathbb{R}^d)$. It is known that

$$\mathcal{H}(\mu) = \begin{cases} \int_{\mathbb{R}^d} f(v) \log f(v)\, dv, & \mu = f\mathcal{L}, \\ +\infty, & \text{else} \end{cases}$$

is σ-lsc and since $\mathcal{H}_\epsilon(\mu) = \mathcal{H}(\mu * G^\epsilon)$, we achieve the first property.

(2) $\mathcal{H}_\epsilon$ is lower bounded: By Lemma 30 for fixed $\epsilon > 0$, $\log(\mu * G^\epsilon)$ is uniformly lower bounded by a linearly growing term. For fixed $\mu \in \mathscr{P}_{2,E}(\mathbb{R}^d)$, we have, with Cauchy–Schwarz,

$$\mathcal{H}_\epsilon(\mu) \gtrsim_\epsilon -\int_{\mathbb{R}^d} \langle v\rangle \mu * G^\epsilon(v)\, dv \geq -\left(\int_{\mathbb{R}^d} \langle v\rangle^2 \mu * G^\epsilon(v)\, dv\right)^{1/2} \geq -(\mathcal{O}(\epsilon) + E)^{1/2} > -\infty.$$

(3) d_L-bounded sets are relatively sequentially σ-compact: This is one of the consequences from Theorem 7.

The existence of minimizers, ν_n^τ, to (27) and limits, μ_t, to (28) is guaranteed from [Ambrosio et al. 2008, Corollary 2.2.2 and Proposition 2.2.3], respectively. □

At the abstract level, the limit curve constructed in Proposition 34 has no relation to $\sqrt{D_\epsilon}$. The following lemmas bridge this gap.

Lemma 35. *For any $\mu_0 \in \mathscr{P}_2(\mathbb{R}^d)$, we have*

$$\sqrt{D_\epsilon(\mu_0)} \leq |\partial^- \mathcal{H}_\epsilon|(\mu_0).$$

Proof. For fixed $\epsilon, R_1, R_2 > 0$ and $\gamma \in \mathbb{R}$, take $T > 0$ from Theorem 48 in the Appendix and the unique weak solution $\mu \in C([0, T]; \mathscr{P}_2(\mathbb{R}^d))$ to

$$\begin{cases} \partial_t \mu = \nabla \cdot \{\mu \phi_{R_1} \int_{\mathbb{R}^d} \phi_{R_1 *} \psi_{R_2}(v - v_*)|v - v_*|^{\gamma+2} \Pi[v - v_*](J_0^\epsilon - J_{0*}^\epsilon)\, d\mu(v_*)\}, \\ \mu(0) = \mu_0. \end{cases}$$

The functions $0 \leq \phi_{R_1}, \psi_{R_2} \leq 1$ are smooth cut-off functions with the following properties:

$$\phi_{R_1}(v) = \begin{cases} 1, & |v| \leq R_1, \\ 0, & |v| \geq R_1 + 1, \end{cases} \qquad \psi_{R_2}(z) = \begin{cases} 0, & |z| \leq 1/R_2, \\ 1, & |z| \geq 2/R_2. \end{cases}$$

The notation J_0^ϵ from the Appendix means

$$J_0^\epsilon = \nabla G^\epsilon * \log[\mu_0 * G^\epsilon] \in C^\infty(\mathbb{R}^d; \mathbb{R}^d).$$

For this proof alone, we define the reduced ϵ-entropy-dissipation

$$D_\epsilon^{R_1,R_2}(\mu_0) := \frac{1}{2}\iint_{\mathbb{R}^{2d}} \phi_{R_1}\phi_{R_1 *}\psi_{R_2}(v-v_*)|v-v_*|^{\gamma+2}|\Pi[v-v_*](J_0^\epsilon - J_{0*}^\epsilon)|^2 d\mu_0(v)d\mu_0(v_*).$$

On the other hand, as the ϵ-entropy dissipation comes from the negative time derivative of entropy, we have

$$\begin{aligned} D_\epsilon^{R_1,R_2}(\mu_0) &= \lim_{t\downarrow 0}\frac{\mathcal{H}_\epsilon(\mu_0)-\mathcal{H}_\epsilon(\mu_t)}{t} = \lim_{t\downarrow 0}\frac{\mathcal{H}_\epsilon(\mu_0)-\mathcal{H}_\epsilon(\mu_t)}{d_L(\mu_0,\mu_t)}\frac{d_L(\mu_0,\mu_t)}{t} \\ &\le \lim_{t\downarrow 0}\Bigg\{\frac{\mathcal{H}_\epsilon(\mu_0)-\mathcal{H}_\epsilon(\mu_t)}{d_L(\mu_0,\mu_t)}\times\frac{1}{t} \\ &\qquad \times\left(\int_0^t\sqrt{\frac{1}{2}\iint_{\mathbb{R}^{2d}}\phi_{R_1}^2\phi_{R_1 *}^2\psi_{R_2}^2|v-v_*|^{\gamma+2}|\Pi[v-v_*](J_0^\epsilon-J_{0*}^\epsilon)|^2 d\mu_s(v)\,d\mu_s(v_*)\,ds}\right)\Bigg\} \\ &\le |\partial\mathcal{H}_\epsilon|(\mu_0)\sqrt{D_\epsilon^{R_1,R_2}(\mu_0)}. \end{aligned}$$

In the first inequality, we estimated $d_L(\mu_0, \mu_t)$ by considering the PDE in this lemma as the grazing collision equation with $M = -(\mu\otimes\mu)\tilde{\nabla}\log\mu_0$. In the last inequality, we have used the Lebesgue differentiation theorem with strong-weak convergence since μ is continuous in time as well as the fact that $\phi_{R_1}^2 \le \phi_{R_1}$ and $\psi_{R_2}^2 \le \psi_{R_2}$ since $0 \le \phi_{R_1}, \psi_{R_2} \le 1$. We are left with the inequality

$$\sqrt{D_\epsilon^{R_1,R_2}(\mu_0)} \le |\partial\mathcal{H}_\epsilon|(\mu_0) \quad \text{for all } R_1, R_2 > 0.$$

Owing to the many regularizations applied, the ϵ-entropy-dissipation $\mu \mapsto D_\epsilon^{R_1,R_2}(\mu)$ is continuous with respect to weak convergence of probability measures. By considering weakly convergent sequences and passing to the limit inferior, we deduce the same inequality with the relaxed slope

$$\sqrt{D_\epsilon^{R_1,R_2}(\mu_0)} \le |\partial^-\mathcal{H}_\epsilon|(\mu_0) \quad \text{for all } R_1, R_2 > 0.$$

As functions of R_1, R_2 individually, $D_\epsilon^{R_1,R_2}(\mu_0)$ is nondecreasing. Furthermore, the integrand of $D_\epsilon^{R_1,R_2}(\mu_0)$ converges to the integrand of $D_\epsilon(\mu_0)$ pointwise μ_0-almost every v, v_*. Thus, an application of the monotone convergence theorem in the limit $R_1, R_2 \to \infty$ on the above inequality completes the proof. □

Lemma 36. *$|\partial^-\mathcal{H}_\epsilon|$ is a strong upper gradient for $\mathcal{H}_\epsilon$ in $\mathscr{P}_{\mu_0}(\mathbb{R}^d)$, where $\mu_0 \in \mathscr{P}_{2,E}(\mathbb{R}^d)$.*

Proof. Fix $\lambda, \nu \in \mathscr{P}_{\mu_0}(\mathbb{R}^d)$ so that by the triangle inequality of Theorem 7, we have $d_L(\lambda,\nu) < \infty$. Now by Proposition 26, there exists a pair of curves $(\mu, M) \in \mathcal{GCE}_1^E$ connecting λ, ν and $\mathcal{A}(\mu_t, M_t) = d_L^2(\lambda,\nu)$ for almost every $t \in [0, 1]$. Using Remark 29 and Lemma 35, we have

$$|\mathcal{H}_\epsilon(\lambda) - \mathcal{H}_\epsilon(\nu)| \le \int_0^1 \sqrt{D_\epsilon(\mu_t)}|\dot{\mu}|(t)\,dt \le \int_0^1 |\partial^-\mathcal{H}_\epsilon|(\mu_t)|\dot{\mu}|(t)\,dt. \qquad \square$$

We now have all the ingredients to prove Theorem 9 so that we can relate curves of maximal slope to weak solutions of the ϵ-Landau equation.

Proof of Theorem 9. Take a limit curve μ_t constructed in Proposition 34. By the previous Lemma 36, the assumptions of [Ambrosio et al. 2008, Theorem 2.3.3] are fulfilled so the curve is of maximal slope with respect to $|\partial^- \mathcal{H}_\epsilon|$ and satisfies the associated energy dissipation inequality

$$\mathcal{H}_\epsilon(\mu_r) - \mathcal{H}_\epsilon(\mu_s) + \frac{1}{2}\int_s^r |\partial^- \mathcal{H}_\epsilon(\mu_t)|^2\, dt + \frac{1}{2}\int_s^r |\dot{\mu}|^2(t)\, dt \le 0.$$

The inequality of Lemma 35 gives

$$\mathcal{H}_\epsilon(\mu_r) - \mathcal{H}_\epsilon(\mu_s) + \frac{1}{2}\int_s^r D_\epsilon(\mu_t)\, dt + \frac{1}{2}\int_s^r |\dot{\mu}|^2(t)\, dt \le 0,$$

which is precisely the statement that the limit curve μ_t is a curve of maximal slope with respect to $\sqrt{D_\epsilon}$. □

Remark 37. The results of Proposition 34 and Lemma 35 can be generalized to other regularization kernels $G^{s,\epsilon}$, in particular, the Maxwellian regularization. However, this is not the case for Lemma 36 since the proof relies on Proposition 28; see Remark 33.

6. Recovering the full Landau equation as $\epsilon \to 0$

Theorems 8 and 9 provide the basic existence theory for the $\epsilon > 0$ approximation of the Landau equation. In this section, we prove the $\epsilon \downarrow 0$ analogue of Theorem 8, which is Theorem 12. By definition, both H-solutions and curves of maximal slope to the full Landau equation dissipate the entropy. Therefore, the assumption of finite initial entropy (A2) automatically ensures

$$\sup_{t\in[0,T]} \mathcal{H}[f_t] = \sup_{t\in[0,T]} \int_{\mathbb{R}^3} f_t \log f_t < +\infty.$$

In the sequel, every quotation of (A2) will refer to this bound.

Sketch of the proof of Theorem 12. By repeating the proof of Theorem 8, we see that the crucial ingredient is the chain rule (18) in Proposition 28. For now assume the following:

Claim 38. *Assume (A1), (A2), (A3) and let M be any grazing rate such that $(\mu, M) \in \mathcal{GCE}_T^E$ and*

$$\int_0^T \mathcal{A}(\mu_t, M_t)\, dt < \infty.$$

Then we have the chain rule

$$\mathcal{H}[\mu_r] - \mathcal{H}[\mu_s] = \frac{1}{2}\int_s^r \iint_{\mathbb{R}^6} \tilde{\nabla}\left[\frac{\delta \mathcal{H}}{\delta \mu}\right] \cdot dM_t\, dt. \tag{29}$$

By following the steps of the proof of Theorem 8 and using (29) instead of (18), one completes the proof of Theorem 12. We dedicate this section to proving Claim 38.

Equation (29) is clearly the $\epsilon \downarrow 0$ limit of (18). The left-hand side of (29) can be obtained from the left-hand side of (18) using the finite entropy (A2) and the fact that $\epsilon \mapsto \mathcal{H}_\epsilon[\mu_t]$ is nonincreasing for every t. We refer to [Erbar 2023, Proof of Proposition 4.2, Step 4(d)] for more details on a similar argument.

The difficulty remains in deducing that the right-hand side of (18) converges to the right-hand side of (29) as $\epsilon \downarrow 0$ given by

$$\int_0^T \iint_{\mathbb{R}^6} \tilde{\nabla} \frac{\delta \mathcal{H}_\epsilon}{\delta \mu} \cdot dM_t \, dt \to \int_0^T \iint_{\mathbb{R}^6} \tilde{\nabla} \frac{\delta \mathcal{H}}{\delta \mu} \cdot dM_t \, dt, \quad \epsilon \downarrow 0, \tag{30}$$

under the additional assumptions (A1), (A2), (A3) on f. The key result which we will use repeatedly in this section is the following theorem which is a specific case of the result in [Royden 1963, Chapter 4, Theorem 17].

Theorem 39 (extended dominated convergence theorem (EDCT)). *Let $(H_\epsilon)_{\epsilon>0}$ and $(I_\epsilon)_{\epsilon>0}$ be sequences of measurable functions on X satisfying $I_\epsilon \geq 0$ and suppose there exists measurable functions H, I satisfying*:

(1) *$|H_\epsilon| \leq I_\epsilon$ for every $\epsilon > 0$ and pointwise a.e.*

(2) *H_ϵ and I_ϵ converge pointwise a.e. to H and I, respectively.*

(3) $\lim_{\epsilon \downarrow 0} \int_X I_\epsilon = \int_X I < \infty$.

Then, we have the convergence

$$\lim_{\epsilon \downarrow 0} \int_X H_\epsilon = \int_X H.$$

Setting $M = m\mathcal{L} \otimes \mathcal{L}$ (valid by Lemma 19) and using Young's inequality on the right-hand side of (18), we obtain the majorants

$$\tilde{\nabla}\left[\frac{\delta \mathcal{H}_\epsilon}{\delta \mu}\right] \cdot m_t \leq \frac{1}{2} f f_* \left|\tilde{\nabla}\left[\frac{\delta \mathcal{H}_\epsilon}{\delta \mu}\right]\right|^2 + \frac{1}{2} \frac{|m_t|^2}{f f_*}.$$

Notice that the first term is precisely the integrand of D_ϵ, while the second term is the integrand of the action functional $\mathcal{A}(\mu_t, M_t)$, which has no dependence on ϵ and is henceforth ignored. We can apply the EDCT (Theorem 39) with $X = (0, T) \times \mathbb{R}^6$ to prove (30) once we show

$$\int_0^T \iint_{\mathbb{R}^6} f f_* \left|\tilde{\nabla}\left[\frac{\delta \mathcal{H}_\epsilon}{\delta \mu}\right]\right|^2 dv_* \, dv \, dt \to \int_0^T \iint_{\mathbb{R}^6} f f_* \left|\tilde{\nabla}\left[\frac{\delta \mathcal{H}}{\delta \mu}\right]\right|^2 dv_* \, dv \, dt, \quad \epsilon \downarrow 0. \tag{31}$$

The pointwise a.e. convergence hypothesis of Theorem 39 is straightforward based on the regularization of $\mathcal{H}_\epsilon$ through G^ϵ. Focusing on (31), we will use a standard dominated convergence theorem (DCT) for the integration in the t-variable, by proving

$$\begin{aligned} &\iint_{\mathbb{R}^6} \frac{1}{2} f f_* \left|\tilde{\nabla}\left[\frac{\delta \mathcal{H}_\epsilon}{\delta \mu}\right]\right|^2 dv_* \, dv \to \iint_{\mathbb{R}^6} \frac{1}{2} f f_* \left|\tilde{\nabla}\left[\frac{\delta \mathcal{H}}{\delta \mu}\right]\right|^2 dv_* \, dv \quad \text{for a.e. } t, \\ &\iint_{\mathbb{R}^6} \frac{1}{2} f f_* \left|\tilde{\nabla}\left[\frac{\delta \mathcal{H}_\epsilon}{\delta \mu}\right]\right|^2 dv_* \, dv \leq C \iint_{\mathbb{R}^6} \frac{1}{2} f f_* \left|\tilde{\nabla}\left[\frac{\delta \mathcal{H}}{\delta \mu}\right]\right|^2 dv_* \, dv \quad \text{for a.e. } t, \text{ for all } \epsilon > 0, \end{aligned} \tag{32}$$

where $C > 0$ is a constant independent of $\epsilon > 0$. The estimate of (32) guarantees the L_t^1 majorization due to the finite entropy-dissipation (A3). □

Our estimates in this section accomplish both the convergence and the estimate of (32) by nested application of Theorem 39. The significance of all three assumptions (A1), (A2), and (A3) will be apparent in proving the convergence in (32).

Remark 40. In this section, the only properties of G^ϵ we use are that it is a nonnegative radial approximate identity with sufficiently many moments. As in the construction of minimizing movement curves in Section 5, the results of this section can be achieved with other radial approximate identities.

6.1. *Outline of technical strategy to prove* (32).

The need to apply Theorem 39 instead of the more classical Lebesgue DCT is that we are unable to prove pointwise estimates in v for the function $v \to f \int_{\mathbb{R}^3} f_* |\tilde{\nabla}[\partial \mathcal{H}_\epsilon / \partial f]|^2 \, dv_*$. Instead, our estimates in this section rely on the self-adjointness of convolution against radial exponentials (SACRE) to construct a convergent majorant in ϵ.

Step 1: finding majorants and appealing to Theorem 39. We seek to find pointwise a.e. majorants in the v-variable:

$$f \int_{\mathbb{R}^3} f_* \left| \tilde{\nabla} \left[\frac{\delta \mathcal{H}_\epsilon}{\delta \mu} \right] \right|^2 dv_* \le I^1_\epsilon(v),$$

where $I^1_\epsilon(v)$ satisfies the hypothesis for the majorant in Theorem 39. We show that I^1_ϵ converges pointwise to some I^1, since I^1_ϵ depends on ϵ only through convolutions against G^ϵ, which is an approximation of the identity. Hence, we are left with showing the integral convergence of Theorem 39(3)

$$\int_{\mathbb{R}^3} I^1_\epsilon(v) \, dv \, dt \to \int_{\mathbb{R}^3} I^1(v) \, dv, \quad \epsilon \to 0.$$

Step 2: use SACRE with G^ϵ. To show the integral convergence for I^1_ϵ, we find functions A^1 and B^1 such that

$$I^1_\epsilon(v) \le A^1(v)(G^\epsilon * B^1)(v)$$

and apply Theorem 39. As in the previous step, the pointwise convergence is easily proved. Hence, we are left to show the integral convergence

$$\int_{\mathbb{R}^3} A^1(G^\epsilon * B^1) \, dv \to \int_{\mathbb{R}^3} A^1 B^1, \quad \epsilon \to 0.$$

The key observation is applying SACRE to obtain

$$\int_{\mathbb{R}^3} A^1(G^\epsilon * B^1) = \int_{\mathbb{R}^3} \overbrace{(G^\epsilon * A^1) B^1}^{=: I^2_\epsilon}.$$

Therefore, we have reduced the problem to showing integral convergence of Theorem 39(3) for I^2_ϵ (as the pointwise convergence is easily proved).

Step 3: repeat Step 2. We repeat the process outlined in Step 2 by finding functions A^2 and B^2 such that we have the pointwise bound

$$I^2_\epsilon(v) \le A^2(v)(G^\epsilon * B^2)(v).$$

Again the pointwise convergence for the majorant follows easily; hence we only need to check the integral convergence of Theorem 39(3) given by

$$\int_{\mathbb{R}^3} A^2(G^\epsilon * B^2) \to \int_{\mathbb{R}^3} A^2 B^2.$$

Using SACRE, we study instead the integral convergence of

$$I_\epsilon^3(v) = (G^\epsilon * A^2)B^2.$$

Eventually, after a finite number of times of finding majorants and applying SACRE, we will obtain a majorant I_ϵ^i for which the estimates and the convergence as $\epsilon \to 0$ follow from the standard Lebesgue DCT, using the bound of the weighted Fisher information in terms of the entropy-dissipation (see Theorem 41) and (A3).

6.2. ***Preparatory results.*** As mentioned in the previous section, for the final step of the proof we need a bound on the weighted Fisher information and a closely related variant in terms of the entropy-dissipation originally discovered by the third author in [Desvillettes and Fellner 2006].

Theorem 41. *Suppose $\gamma \in (-4, 0]$ and let $f \geq 0$ be a probability density belong to $L^1_{2-\gamma} \cap L \log L(\mathbb{R}^3)$. We have*

$$\int_{\mathbb{R}^3} f(v)\langle v\rangle^\gamma \left|\nabla \frac{\delta \mathcal{H}}{\delta f}\right|^2 dv + \int_{\mathbb{R}^3} f(v)\langle v\rangle^\gamma \left|v \times \nabla \frac{\delta \mathcal{H}}{\delta f}\right|^2 dv \leq C(1 + D_{w,\mathcal{H}}(f)),$$

where $C > 0$ is a constant depending only on the bounds of $m_{2-\gamma}(f)$ and the Boltzmann entropy, $\mathcal{H}[f]$, of f.

The estimate in this precise form can be found in [Desvillettes 2022, Proposition 4, p. 10]. We will refer to the second term on the left-hand side as a "cross Fisher information". We mention here that (A2) enters in the sequel since the constant $C > 0$ in Theorem 41 depends on bounds for $\mathcal{H}[f]$.

To decompose the entropy-dissipation in a manageable way that makes the cross Fisher term more apparent, we have the following linear algebra fact.

Lemma 42. *For $x, y \in \mathbb{R}^3$, we have*

$$|x|^2(y \cdot \Pi[x]y) = |x \times y|^2.$$

Proof. Without loss of generality, we assume neither $x, y = 0$ or else the statement holds trivially. Let θ be an oriented angle between x and y. We expand the definition of $\Pi[x]$ and observe

$$\begin{aligned} |x|^2(y \cdot \Pi[x]y) &= y \cdot (|x|^2 I - x \otimes x)y = |x|^2|y|^2 - |x \cdot y|^2 \\ &= |x|^2|y|^2(1 - \cos^2\theta) = |x|^2|y|^2 \sin^2\theta = |x \times y|^2. \end{aligned}$$

□

The following lemma shows how we use (A1) to control the singularity of the weight.

Lemma 43. *Given $\gamma \in (-3, 0]$, assume that f satisfies (A1) for some $0 < \eta \leq \gamma + 3$. Then we have for a.e. t*

$$\int_{\mathbb{R}^3} f_*(t)|v - v_*|^\gamma \, dv_* \leq C_1(t)\langle v\rangle^\gamma, \qquad \int_{\mathbb{R}^3} f_*(t)|v_*|^2|v - v_*|^\gamma \, dv_* \leq C_2(t)\langle v\rangle^\gamma, \tag{33}$$

where

$$\|C_1\|_{L^\infty(0,T)} \lesssim_{\gamma,\eta} \|\langle\cdot\rangle^{-\gamma} f(t)\|_{L^\infty(0,T;L^1\cap L^{(3-\eta)/(3+\gamma-\eta)}(\mathbb{R}^3))},$$
$$\|C_2\|_{L^\infty(0,T)} \lesssim_{\gamma,\eta} \|\langle\cdot\rangle^{2-\gamma} f(t)\|_{L^\infty(0,T;L^1\cap L^{(3-\eta)/(3+\gamma-\eta)}(\mathbb{R}^3))}.$$

Proof. We will only prove the first inequality of (33) since the second inequality uses the same procedure. We split the estimation for local $|v| \le 1$ and far-field $|v| \ge 1$.

Case 1: $|v| \le 1$. We split the integral over v_* into two regions

$$\begin{aligned}\int_{\mathbb{R}^3} f_*|v-v_*|^\gamma\,dv_* &= \int_{|v-v_*|\ge 1} f_*|v-v_*|^\gamma\,dv_* + \int_{|v-v_*|\le 1} f_*|v-v_*|^\gamma\,dv_*\\ &\le 1 + \int_{|v-v_*|\le 1} f_*|v-v_*|^\gamma\,dv_*,\end{aligned}$$

where we have used that $\int_{\mathbb{R}^3} f = 1$ and $\gamma \le 0$. For the integral with the singularity, we apply Young's convolution inequality with conjugate exponents $((3-\eta)/(3+\gamma-\eta), (-3+\eta)/\gamma)$

$$\begin{aligned}\int_{|v-v_*|\le 1} f_*|v-v_*|^\gamma\,dv_* &\le \|f*(\chi_{B_1}|\cdot|^\gamma)\|_{L^\infty}\\ &\le \|f\|_{L^{(3-\eta)/(3+\gamma-\eta)}}\|\chi_{B_1}|\cdot|^\gamma\|_{L^{(-3+\eta)/\gamma}} \le \left(\frac{\omega_2}{\eta}\right)^{(-3+\eta)/\gamma}\|f\|_{L^{(3-\eta)/(3+\gamma-\eta)}}.\end{aligned}$$

Here, ω_2 is the volume of the unit sphere in $\mathbb{R}^3$.

Case 2: $|v| \ge 1$. Once again, we split the integral into two parts

$$\begin{aligned}\int_{\mathbb{R}^3} f_*|v-v_*|^\gamma\,dv_* &= \int_{|v_*|\le|v|/2} f_*|v-v_*|^\gamma\,dv_* + \int_{|v_*|\ge|v|/2} f_*|v-v_*|^\gamma\,dv_*\\ &\le 2^{-\gamma}|v|^\gamma\int_{|v_*|\le|v|/2} f_*\,dv_* + 2^{-\gamma}|v|^\gamma\int_{|v_*|\ge|v|/2} f_*|v_*|^{-\gamma}|v-v_*|^\gamma\,dv_*.\end{aligned}$$

The first term and second term come from the following inequalities based on their respective integration regions:

$$|v-v_*| \ge |v|-|v_*| \ge \tfrac{1}{2}|v|, \quad 1 \le 2^{-\gamma}|v|^\gamma|v_*|^{-\gamma}.$$

We estimate the first integral using the unit mass of f, while the second integral is more delicate but again uses the splitting of the previous step to obtain

$$\begin{aligned}&\int_{\mathbb{R}^3} f_*|v-v_*|^\gamma\,dv_*\\ &\quad\le 2^{-\gamma}|v|^\gamma + 2^{-\gamma}|v|^\gamma\left(\int_{|v-v_*|\ge 1} f_*|v_*|^{-\gamma}|v-v_*|^\gamma\,dv_* + \int_{|v-v_*|\le 1} f_*|v_*|^{-\gamma}|v-v_*|^\gamma\,dv_*\right).\end{aligned}$$

In the large brackets, the first integral can be estimated by $m_{-\gamma}(f)$. Now we use the same Young's inequality argument for the remaining integral to obtain

$$\int_{\mathbb{R}^3} f_*|v-v_*|^\gamma\,dv_* \le 2^{-\gamma}|v|^\gamma + 2^{-\gamma}|v|^\gamma\left(m_{-\gamma}(f) + \left(\frac{\omega_2}{\eta}\right)^{(-3+\eta)/\gamma}\||\cdot|^{-\gamma} f\|_{L^{(3-\eta)/(3+\gamma-\eta)}(\mathbb{R}^3)}\right).$$

The proof is complete by combining the estimates for $|v| \le 1$ and $|v| \ge 1$. □

Lemma 44 (Peetre). *For any $p \in \mathbb{R}$ and $x, y \in \mathbb{R}^d$, we have*

$$\frac{\langle x\rangle^p}{\langle y\rangle^p} \le 2^{|p|/2}\langle x-y\rangle^{|p|}.$$

Proof. Our proof follows [Barros-Neto 1973]. Starting with the case $p=2$, for fixed vectors $a, b \in \mathbb{R}^d$ we have, with the help of Young's inequality,

$$\begin{aligned}1+|a-b|^2 &\le 1+|a|^2+2|a|\,|b|+|b|^2 \le 1+2|a|^2+2|b|^2\\ &\le 2+2|a|^2+2|a|^2|b|^2+2|b|^2 = 2(1+|a|^2)(1+|b|^2).\end{aligned}$$

Dividing by $\langle b\rangle^2$ and setting $a=x-y$, $b=-y$, we obtain the inequality for $p=2$

$$\frac{\langle x\rangle^2}{\langle y\rangle^2} \le 2\langle x-y\rangle^2.$$

By taking nonnegative powers, this proves the inequality for $p \ge 0$. On the other hand, when we divided by $\langle b\rangle^2$ we could have also set $a=x-y,\ b=x$ to obtain

$$\frac{\langle y\rangle^2}{\langle x\rangle^2} \le 2\langle x-y\rangle^2.$$

Taking strictly nonnegative powers here proves the inequality for $p<0$. □

Next, we prove an estimate for algebraic functions (growing or decaying) convoluted against G^ϵ with respect to the original function.

Lemma 45. *For any $p \in \mathbb{R}$, we have*

$$\int_{\mathbb{R}^d} \langle w\rangle^p G^\epsilon(v-w)\,dw \le C\langle v\rangle^p,$$

where $C>0$ is a constant depending only on $|p|$ and $m_{|p|}(G)$.

Proof. We use Peetre's inequality in Lemma 44 to introduce $v-w$ into the angle brackets

$$\begin{aligned}\int_{\mathbb{R}^d} \langle w\rangle^p G^\epsilon(v-w)\,dw &\le 2^{|p|/2}\langle v\rangle^p \int_{\mathbb{R}^d} \langle v-w\rangle^{|p|} G^\epsilon(v-w)\,dw\\ &= 2^{|p|/2}\langle v\rangle^p \int_{\mathbb{R}^d} (1+|w|^2)^{|p|/2}\epsilon^{-d} G\Big(\frac{w}{\epsilon}\Big)\,dw\\ &= 2^{|p|/2}\langle v\rangle^p \int_{\mathbb{R}^d} (1+\epsilon^2|w|^2)^{|p|/2} G(w)\,dw\\ &\le C_{|p|}\langle v\rangle^p\Big[1+\epsilon^{|p|}\int_{\mathbb{R}^d}|w|^{|p|}G(w)\,dw\Big] \le C_{|p|}[1+\epsilon^{|p|}m_{|p|}(G)]\langle v\rangle^p.\end{aligned}$$ □

We stress that Peetre's inequality in Lemma 44 is necessary for the estimate of Lemma 45 with *nonpositive* powers p which we apply in the sequel. Finally, the last result we will need is an integration by parts formula for the differential operator associated to the cross Fisher information.

Lemma 46 (twisted integration by parts). *Let f, g be smooth scalar functions of $\mathbb{R}^3$ which are sufficiently integrable. Then, we have the formula*

$$\int_{\mathbb{R}^3} (v \times \nabla_v g(v)) f(v)\, dv = -\int_{\mathbb{R}^3} g(v)(v \times \nabla_v f(v))\, dv.$$

Here, the meaning of $v \times \nabla_v$ is

$$v \times \nabla_v f(v) = (v^2 \partial^3 f(v) - v^3 \partial^2 f(v),\, v^3 \partial^1 f(v) - v^1 \partial^3 f(v),\, v^1 \partial^2 f(v) - v^2 \partial^1 f(v)).$$

6.3. *Proof of (32) using Theorem 39.* We start by decomposing and estimating the integrand of D_ϵ. With the help of Lemma 42, we expand the square term of the integrand to see

$$\begin{aligned}
&\left|\tilde{\nabla}\left[\frac{\delta \mathcal{H}_\epsilon}{\delta \mu}\right]\right|^2 \\
&= |v-v_*|^{2+\gamma} |\Pi[v-v_*](b^\epsilon * a^\epsilon - b^\epsilon * a_*^\epsilon)|^2 \\
&\le |v-v_*|^\gamma (4|v \times (b^\epsilon * a^\epsilon)|^2 + 4|v_* \times (b^\epsilon * a_*^\epsilon)|^2 + 4|v \times (b^\epsilon * a_*^\epsilon)|^2 + 4|v_* \times (b^\epsilon * a^\epsilon)|^2) \\
&\le 4|v-v_*|^\gamma \underbrace{|v \times (b^\epsilon * a^\epsilon)|^2}_{①} + 4|v-v_*|^\gamma \underbrace{|v_* \times (b^\epsilon * a_*^\epsilon)|^2}_{②} + 4|v|^2|v-v_*|^\gamma \underbrace{|b^\epsilon * a_*^\epsilon|^2}_{③} + 4|v_*|^2|v-v_*|^\gamma \underbrace{|b^\epsilon * a^\epsilon|^2}_{④},
\end{aligned}$$

where we use the shorthand notation

$$b^\epsilon = G^\epsilon \quad \text{and} \quad a^\epsilon = \nabla \log(G^\epsilon * f). \tag{34}$$

By using that G^ϵ is an approximation of the identity, we know that the integrand of D_ϵ converges pointwise a.e. to the integrand of D as $\epsilon \downarrow 0$. As well, each ⓘ for $i = 1, 2, 3, 4$ converges pointwise a.e. to

$$① \to \frac{|v \times \nabla f|^2}{f^2}, \quad ② \to \frac{|v_* \times \nabla_* f_*|^2}{f_*^2}, \quad ③ \to \frac{|\nabla_* f_*|^2}{f_*^2}, \quad ④ \to \frac{|\nabla f|^2}{f^2}.$$

By Theorem 39, to show the integral convergence in (32), it suffices to show, for example,

$$\iint_{\mathbb{R}^6} f f_* |v-v_*|^\gamma\, ①\, dv\, dv_* \to \iint_{\mathbb{R}^6} f f_* |v-v_*|^\gamma \frac{|v \times \nabla f|^2}{f^2}\, dv\, dv_*,$$

and similarly for each ⓘ for $i = 2, 3, 4$. By symmetry considerations when swapping the variables $v \leftrightarrow v_*$, the convergence for the terms ① and ④ controls the convergence for ② and ③, respectively. Hence we will focus on the term ④ first and then on term ①.

6.3.1. *Term* ④. We seek to show in the limit $\epsilon \downarrow 0$

$$\begin{aligned}
\iint_{\mathbb{R}^6} f f_* |v_*|^2 |v-v_*|^\gamma |b^\epsilon * a^\epsilon|^2\, dv_*\, dv &= \int_{\mathbb{R}^3} \left(\int_{\mathbb{R}^3} f_* |v_*|^2 |v-v_*|^\gamma\, dv_* \right) f |b^\epsilon * a^\epsilon|^2\, dv \\
&\to \int_{\mathbb{R}^3} \left(\int_{\mathbb{R}^3} f_* |v_*|^2 |v-v_*|^\gamma\, dv_* \right) \frac{|\nabla f|^2}{f}\, dv.
\end{aligned} \tag{35}$$

By the reordering of integrations written above, we now think of the double integral over v, v_* of $f f_* |v_*|^2 |v-v_*|^\gamma |b^\epsilon * a^\epsilon|^2$ as a single integral of the function $\left(\int_{\mathbb{R}^3} f_* |v_*|^2 |v-v_*|^\gamma\, dv_*\right) f |b^\epsilon * a^\epsilon|^2$ over v.

To be precise, we wish to apply Theorem 39 with $X = \mathbb{R}^3$ with $H_\epsilon = \left(\int_{\mathbb{R}^3} f_* |v_*|^2 |v-v_*|^\gamma \, dv_*\right) f |b^\epsilon * a^\epsilon|^2$. We can use Cauchy–Schwarz on the convolution integral to absorb the power term as follows:

$$\begin{aligned} |b^\epsilon * a^\epsilon|^2 = \left| \int_{\mathbb{R}^3} b^\epsilon(v-w) a^\epsilon(w)\, dw \right|^2 &\le \left(\int_{\mathbb{R}^3} \langle w \rangle^{-\gamma} b^\epsilon(v-w)\, dw \right) \left(\int_{\mathbb{R}^3} b^\epsilon(v-w) \langle w \rangle^\gamma |a^\epsilon(w)|^2 \, dw \right) \\ &\le C \langle v \rangle^{-\gamma} b^\epsilon * [\langle \cdot \rangle^\gamma |a^\epsilon(\cdot)|^2], \end{aligned}$$

where the last inequality comes from Lemma 45. Continuing with Lemma 43, we have

$$\left(\int_{\mathbb{R}^3} f_* |v_*|^2 |v-v_*|^\gamma \, dv_* \right) f |b^\epsilon * a^\epsilon|^2 \le C f b^\epsilon * [\langle \cdot \rangle^\gamma |a^\epsilon|^2].$$

By Theorem 39, we reduce the problem to showing in the limit $\epsilon \downarrow 0$

$$\int_{\mathbb{R}^3} f b^\epsilon * [\langle \cdot \rangle^\gamma |a^\epsilon|^2] \, dv \to \int_{\mathbb{R}^3} \langle v \rangle^\gamma \frac{|\nabla f|^2}{f} \, dv.$$

This is were we use SACRE, Step 2 of our general strategy in Section 6.1. Application of SACRE and further simplification using the specific forms of a^ϵ and b^ϵ (see (34)) yields

$$\int_{\mathbb{R}^3} f b^\epsilon * [\langle \cdot \rangle^\gamma |a^\epsilon|^2] \, dv = \int_{\mathbb{R}^3} [b^\epsilon * f] \langle v \rangle^\gamma |a^\epsilon|^2 \, dv = \int_{\mathbb{R}^3} \langle v \rangle^\gamma \frac{|b^\epsilon * \nabla f|^2}{b^\epsilon * f} \, dv. \tag{36}$$

We work with this simplified expression and note that pointwise convergence is still valid

$$\frac{|b^\epsilon * \nabla f|^2}{b^\epsilon * f} \to \frac{|\nabla f|^2}{f}.$$

Next, we notice that the function $\beta : (F, f) \mapsto |F|^2/f$ is jointly convex in $F \in \mathbb{R}^3$ and $f > 0$, so we can use Jensen's inequality with $b^\epsilon = G^\epsilon$ as the reference probability measure to obtain a further pointwise majorant for the integrand of (36)

$$\begin{aligned} \frac{|b^\epsilon * \nabla f|^2}{b^\epsilon * f}(v) &= \beta(b^\epsilon * \nabla f, b^\epsilon * f)(v) = \beta\left(\int_{\mathbb{R}^3} \nabla f(v-y) b^\epsilon(y)\, dy, \int_{\mathbb{R}^3} f(v-y) b^\epsilon(y)\, dy \right) \\ &\le \int_{\mathbb{R}^3} \beta(\nabla f(v-y), f(v-y)) b^\epsilon(y)\, dy = \int_{\mathbb{R}^3} \frac{|\nabla f(v-y)|^2}{f(v-y)} b^\epsilon(y)\, dy = b^\epsilon * \left[\frac{|\nabla f|^2}{f} \right](v). \end{aligned}$$

Using Theorem 39 again, we reduce the problem to showing in the limit $\epsilon \downarrow 0$

$$\int_{\mathbb{R}^3} \langle v \rangle^\gamma b^\epsilon * \left[\frac{|\nabla f|^2}{f} \right] dv \to \int_{\mathbb{R}^3} \langle v \rangle^\gamma \frac{|\nabla f|^2}{f} \, dv.$$

We use SACRE once more and place the convolution onto the weight term

$$\int_{\mathbb{R}^3} \langle v \rangle^\gamma b^\epsilon * \left[\frac{|\nabla f|^2}{f} \right] dv = \int_{\mathbb{R}^3} [b^\epsilon * \langle \cdot \rangle^\gamma] \frac{|\nabla f|^2}{f} \, dv.$$

Now, we are in a position to apply the classical dominated convergence theorem. We notice that we have the pointwise convergence

$$[b^\epsilon * \langle \cdot \rangle^\gamma] \to \langle v \rangle^\gamma.$$

Furthermore, using Lemma 45, we can estimate $b^\epsilon * \langle\cdot\rangle^\gamma$ uniformly in ϵ to find the domination

$$[b^\epsilon * \langle\cdot\rangle^\gamma]\frac{|\nabla f|^2}{f} \le C\langle v\rangle^\gamma \frac{|\nabla f|^2}{f}.$$

Using Theorem 41, the finite entropy-dissipation (A3), and uniformly bounded entropy (A2) (remember the constant in Theorem 41 depends also on bounds for the entropy), we know that the right-hand side belongs to L^1_v for a.e. $t \in (0, T)$. Therefore, for a.e. $t \in (0, T)$ the conditions of the dominated convergence theorem are satisfied so we have the integral convergence

$$\int_{\mathbb{R}^3} [b^\epsilon * \langle\cdot\rangle^\gamma]\frac{|\nabla f|^2}{f}\,dv \to \int_{\mathbb{R}^3} \langle v\rangle^\gamma \frac{|\nabla f|^2}{f}\,dv.$$

We have closed the argument for the convergence of (35) after retracing the previous estimates with Theorem 39.

6.3.2. *Term* ①. We seek to show in the limit $\epsilon \downarrow 0$

$$\begin{aligned}\iint_{\mathbb{R}^6} ff_*|v-v_*|^\gamma |v\times(b^\epsilon * a^\epsilon)|^2\,dv_*\,dv &= \int_{\mathbb{R}^3}\left(\int_{\mathbb{R}^3} f_*|v-v_*|^\gamma\,dv_*\right) f|v\times(b^\epsilon * a^\epsilon)|^2\,dv \\ &\to \int_{\mathbb{R}^3}\left(\int_{\mathbb{R}^3} f_*|v-v_*|^\gamma\,dv_*\right)\frac{|v\times\nabla f|^2}{f}\,dv \end{aligned}\tag{37}$$

using the same strategy of nested applications of Theorem 39 like in Section 6.3.1. We will encounter difficulty when trying to use Jensen's inequality due to the cross Fisher information term. As in Section 6.3.1, we have written this double integral over v, v_* as a single integral over v. By Theorem 39 and Lemma 43, it suffices to show the integral convergence of

$$\int_{\mathbb{R}^3} \langle v\rangle^\gamma f|v\times(b^\epsilon * a^\epsilon)|^2\,dv \to \int_{\mathbb{R}^3}\langle v\rangle^\gamma \frac{|v\times\nabla f|^2}{f}\tag{38}$$

to obtain the integral convergence of (37). Pointwise, we can make the following manipulations:

$$\begin{aligned} v\times(b^\epsilon * a^\epsilon) &= v\times\left(\int_{\mathbb{R}^3} G^\epsilon(v-w)\nabla\log(f*G^\epsilon(w))\,dw\right) = v\times\left(\int_{\mathbb{R}^3}\nabla G^\epsilon(v-w)\log(f*G^\epsilon(w))\,dw\right) \\ &= \int_{\mathbb{R}^3} w\times\nabla G^\epsilon(v-w)\log(f*G^\epsilon(w))\,dw = \int_{\mathbb{R}^3} G^\epsilon(v-w)\,w\times\nabla\log(f*G^\epsilon(w))\,dw, \end{aligned}\tag{39}$$

where we have used the radial symmetry of G^ϵ to get the cancellation $(v-w)\times\nabla G^\epsilon(v-w) = 0$ and the twisted integration by parts Lemma 46 (we note that we do not pick up any signs in the integration by parts, as the variable w appears with a minus sign in the argument of G^ϵ).

We apply Cauchy–Schwarz, multiply and divide by $\langle w\rangle^\gamma$, and use Lemma 45 to obtain

$$\begin{aligned}|v\times(b^\epsilon * a^\epsilon)|^2 &\le \left(\int_{\mathbb{R}^3} G^\epsilon(v-w)\langle w\rangle^{-\gamma}\,dw\right)\left(\int_{\mathbb{R}^3} G^\epsilon(v-w)\langle w\rangle^\gamma\left|w\times\frac{\nabla f * G^\epsilon(w)}{f*G^\epsilon(w)}\right|^2 dw\right) \\ &\lesssim_\gamma \langle v\rangle^{-\gamma}\left(\int_{\mathbb{R}^3} G^\epsilon(v-w)\langle w\rangle^\gamma\left|w\times\frac{\nabla f*G^\epsilon(w)}{f*G^\epsilon(w)}\right|^2 dw\right).\end{aligned}$$

Remembering that this majorant holds pointwise on the integrand of (38), we multiply by $\langle v\rangle^\gamma f(v)$ and obtain

$$\langle v\rangle^\gamma f(v)|v\times(b^\epsilon * a^\epsilon)|^2 \lesssim f\bigg(\int_{\mathbb{R}^3} G^\epsilon(v-w)\langle w\rangle^\gamma \bigg| w\times \frac{\nabla f * G^\epsilon(w)}{f*G^\epsilon(w)}\bigg|^2\,dw\bigg).$$

Now, we recognize a convolution inside the brackets. Hence, using SACRE we can rewrite

$$\int_{\mathbb{R}^3} f\bigg(\int_{\mathbb{R}^3} G^\epsilon(v-w)\langle w\rangle^\gamma \bigg| w\times \frac{\nabla f * G^\epsilon(w)}{f*G^\epsilon(w)}\bigg|^2\,dw\bigg)\,dv = \int_{\mathbb{R}^3} \langle v\rangle^\gamma \frac{|v\times\nabla f * G^\epsilon(v)|^2}{f*G^\epsilon(v)}\,dv.$$

Using Theorem 39, we need to show the convergence of the right-hand side. Here, it is now possible to use Jensen's inequality after some more manipulations.

Claim 47.
$$\frac{|v\times\nabla f * G^\epsilon(v)|^2}{f*G^\epsilon(v)} \le \int_{\mathbb{R}^3} G^\epsilon(v-w)\frac{|w\times\nabla f(w)|^2}{f(w)}\,dw. \tag{40}$$

Proof of Claim 47. We start by repeating an argument similar to (39). Using that G^ϵ is radially symmetric and the twisted integration by parts Lemma 46, we obtain

$$\begin{aligned} v\times\nabla f * G^\epsilon(v) &= v\times\bigg(\int_{\mathbb{R}^3}\nabla G^\epsilon(v-w)f(w)\,dw\bigg)\\ &= \int_{\mathbb{R}^3} w\times\nabla G^\epsilon(v-w)f(w)\,dw = \int_{\mathbb{R}^3} G^\epsilon(v-w)\overbrace{(w\times\nabla_w f(w))}^{=:F(w)}\,dw. \end{aligned}$$

Therefore, since $\beta:(F,f)\mapsto |F|^2/f$ is jointly convex in $F\in\mathbb{R}^3$ and $f>0$, we apply Jensen's inequality with G^ϵ as the reference probability measure to the left-hand side of (40) to see

$$\begin{aligned} \frac{|v\times\nabla f * G^\epsilon(v)|^2}{f*G^\epsilon(v)} &= \frac{|F*G^\epsilon|^2}{f*G^\epsilon}(v) = \beta(G^\epsilon * F,\ G^\epsilon * f)(v)\\ &= \beta\bigg(\int_{\mathbb{R}^3} F(v-w)G^\epsilon(w)\,dw,\ \int_{\mathbb{R}^3} f(v-w)G^\epsilon(w)\,dw\bigg)\\ &\le \int_{\mathbb{R}^3}\beta(F(v-w),\ f(v-w))G^\epsilon(w)\,dw = \int_{\mathbb{R}^3}\frac{|(v-w)\times\nabla F(v-w)|^2}{f(v-w)}G^\epsilon(w)\,dw, \end{aligned}$$

which proves the claim. □

Continuing, by Theorem 39, we seek to establish the integral convergence of

$$\int_{\mathbb{R}^3}\langle v\rangle^\gamma\bigg[\frac{|F|^2}{f} * G^\epsilon\bigg](v)\,dv = \int_{\mathbb{R}^3}[\langle\cdot\rangle^\gamma * G^\epsilon](v)\frac{|v\times\nabla f(v)|^2}{f(v)}\,dv.$$

Finally, the integrand of the right-hand side has a majorant due to Lemma 45

$$[\langle\cdot\rangle^\gamma * G^\epsilon](v)\frac{|v\times\nabla f(v)|^2}{f(v)} \lesssim \langle v\rangle^\gamma\frac{|v\times\nabla f(v)|^2}{f(v)}.$$

Once again, using Theorem 41 and Assumptions (A3) and (A2), we obtain that for a.e. $t\in(0,T)$ the right-hand side belongs to $L^1_v(\mathbb{R}^3)$. Using dominated convergence theorem, we see that the integral converges. Tracing back the estimates, this takes care of the convergence of the term ① and establishes the convergence in (38).

We note that the estimates in the previous subsections not only establish the a.e. pointwise convergence of (32), but also the majorization

$$\iint_{\mathbb{R}^6} \frac{1}{2} f f_* \left| \tilde{\nabla} \left[\frac{\delta \mathcal{H}_\epsilon}{\delta \mu} \right] \right|^2 dv_* \, dv \leq C \iint_{\mathbb{R}^6} \frac{1}{2} f f_* \left| \tilde{\nabla} \left[\frac{\delta \mathcal{H}}{\delta \mu} \right] \right|^2 dv_* \, dv \quad \text{for a.e. } t, \text{ for all } \epsilon > 0,$$

where

$$C \lesssim \| \langle \cdot \rangle^{-\gamma} f(t) \|_{L^\infty(0,T; L^1 \cap L^{(3-\eta)/(3+\gamma-\eta)}(\mathbb{R}^3))} + \| \langle \cdot \rangle^{2-\gamma} f(t) \|_{L^\infty(0,T; L^1 \cap L^{(3-\eta)/(3+\gamma-\eta)}(\mathbb{R}^3))}$$

by Lemma 43. Hence, using (A3) and (32) we can apply Lebesgue DCT to pass to the limit in the time integral and show the desired chain rule Claim 38.

Appendix: An auxiliary PDE for Lemma 35

In this section, we fix $\epsilon > 0$ throughout and study weak solutions to the PDE

$$\begin{cases} \partial_t \mu = \nabla \cdot \{ \mu \phi_{R_1} \int_{\mathbb{R}^d} \phi_{R_1 *} \psi_{R_2}(v - v_*) |v - v_*|^{\gamma+2} \Pi[v - v_*] (J_0^\epsilon - J_{0*}^\epsilon) \, d\mu(v_*) \}, \\ \mu(0) = \mu_0. \end{cases} \tag{41}$$

We assume the initial data μ_0 belongs to $\mathscr{P}_2(\mathbb{R}^d)$. For $R_1, R_2 > 0$, the functions $0 \leq \phi_{R_1}, \psi_{R_2} \leq 1$ are smooth cut-off functions used to approximate the identity function in different ways:

$$\phi_{R_1}(v) = \begin{cases} 1, & |v| \leq R_1, \\ 0, & |v| \geq R_1 + 1, \end{cases} \qquad \psi_{R_2}(z) = \begin{cases} 0, & |z| \leq 1/R_2, \\ 1, & |z| \geq 2/R_2. \end{cases}$$

For $\epsilon > 0$, J_0^ϵ is the gradient of first variation of $\mathcal{H}_\epsilon$ *applied to* μ_0, meaning

$$J_0^\epsilon = \nabla G^\epsilon * \log[\mu_0 * G^\epsilon] \in C^\infty(\mathbb{R}^d; \mathbb{R}^d).$$

The main result of this section is:

Theorem 48. *Fix $\epsilon, R_1, R_2 > 0$, $\gamma \in \mathbb{R}$, and $\mu_0 \in \mathscr{P}_2(\mathbb{R}^d)$. Then, there is a global unique weak solution $\mu \in C([0, +\infty); \mathscr{P}_2(\mathbb{R}^d))$ to (41).*

By Lemma 31, we know that J_0^ϵ is uniformly bounded (with constant depending on ϵ and μ_0 only through bounds on its second moment). The purpose of $\phi_{R_1}, \phi_{R_1 *}$ is to cut off the growth of $J_0^\epsilon, J_{0*}^\epsilon$ to ensure that the "velocity field" in the right-hand side of (41) is globally Lipschitz (it is, in fact, smooth and compactly supported). The $\psi_{R_2}(v - v_*)$-term avoids the possible singularities coming from the weight $|v - v_*|^{\gamma+2}$ for soft potentials $\gamma < 0$.

The construction of the solution in Theorem 48 is given in two steps. Firstly, a local well-posedness theory established to some finite time interval $T > 0$ which depends on ϵ, γ, R_1, R_2 and μ_0. Secondly, the time of existence (and uniqueness) is extended to $+\infty$ since T depends on μ_0 only through its second moment, which is conserved by the evolution of (41).

We fix $T > 0$ to be determined explicitly later. Our strategy is to employ a fixed-point argument in the space $C([0, T]; \mathscr{P}_2(\mathbb{R}^d))$, which we will equip with the metric

$$d(\mu, \nu) := \sup_{t \in [0,T]} W_2(\mu(t), \nu(t)), \qquad \mu, \nu \in C([0, T]; \mathscr{P}_2(\mathbb{R}^d)),$$

where W_2 is the 2-Wasserstein distance on $\mathscr{P}_2(\mathbb{R}^d)$. We have closely followed the procedure in [Cañizo et al. 2011] with appropriate modifications for this setting.

Remark 49. Since we are cutting off the "velocity" field at radius R_1, R_2, the growth of J_0^ϵ is inconsequential. Hence the results of this section can be applied when replacing the convolution kernel of J_0^ϵ with general tailed exponential distributions $G^{s,\epsilon}(v)$ for $s > 0$.

For $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, we will denote by $U[\mu](v)$ the function

$$U[\mu](v) := -\phi_{R_1} \int_{\mathbb{R}^d} \phi_{R_1 *} \psi_{R_2}(v - v_*) |v - v_*|^{\gamma+2} \Pi[v - v_*](J_0^\epsilon - J_{0*}^\epsilon)\, d\mu(v_*),$$

so that the PDE in (41) can be written as a nonlinear transport/continuity equation:

$$\partial_t \mu(t) = -\nabla \cdot \{\mu(t) U[\mu(t)]\}.$$

To fix ideas, the weak formulation of (41) is such that the following equality holds for all test functions $\tau \in C_c^\infty(\mathbb{R}^d)$ and times $t \in [0, T]$

$$\int_{\mathbb{R}^d} \tau(v)\, d\mu_r(v) - \int_{\mathbb{R}^d} \tau(v)\, d\mu_0(v) \\ = \int_0^t \int_{\mathbb{R}^d} \phi_{R_1} \nabla \tau(v) \cdot \int_{\mathbb{R}^d} \phi_{R_1 *} \psi_{R_2}(v - v_*) |v - v_*|^{\gamma+2} \Pi[v - v_*](J_0^\epsilon - J_{0*}^\epsilon) d\mu_s(v_*)\, d\mu_s(v)\, ds.$$

Thanks to all the smooth cutoffs from ϕ_{R_1}, $\phi_{R_1 *}$, and ψ_{R_2} and $\mu_0 \in \mathscr{P}_2(\mathbb{R}^d)$, we can enlarge the class of test functions to smooth functions with quadratic growth. In particular, by choosing $\tau(v) = |v|^2$ and symmetrizing the right-hand side by swapping $v \leftrightarrow v_*$, we see that the second moment of μ_0 is conserved along the evolution of (41).

Our first step is to look at the level of the characteristic equation associated to (41).

Lemma 50 (characteristic equation). *For any $T > 0$, $\mu \in C([0, T]; \mathscr{P}_2(\mathbb{R}^d))$ and $v_0 \in \mathbb{R}^d$, there exists a unique solution $v \in C^1((0, T); \mathbb{R}^d) \cap C([0, T]; \mathbb{R}^d)$ to the ODE*

$$\frac{dv}{dt} = U[\mu(t)](v), \qquad v(0) = v_0.$$

Furthermore, the growth rate satisfies

$$|v(t)| \le \max\{|v_0|, R_1 + 1\} \quad \textit{for all } t \in [0, T].$$

Proof. $U[\mu(t)](\cdot)$ is smooth and compactly supported uniformly in t, so classical Cauchy–Lipschitz theory gives existence and uniqueness of solution v with the promised regularity.

For the estimate on the growth rate, note that $U[\mu]$ has support contained in B_{R_1+1}. Points outside this ball do not change in time according to this ODE. □

We will denote by Φ_μ^t the flow map associated to this ODE, so that

$$\frac{d}{dt} \Phi_\mu^t(v_0) = U[\mu(t)](\Phi_\mu^t(v_0)), \qquad \Phi_\mu^0(v_0) = v_0.$$

It is known that, given $\nu \in C([0,T];\mathscr{P}_2(\mathbb{R}^d))$, the curve of probability measures $\mu(t) = \Phi_\nu^t \# \mu_0$ is a weak solution to

$$\partial_t \mu(t) = -\nabla \cdot \{\mu(t) U[\nu(t)]\}, \qquad \mu(0) = \mu_0.$$

Here, $\Phi_\nu^t \# \mu_0$ is the push-forward measure of μ_0 defined in duality with $\tau \in C_b(\mathbb{R}^d)$ by

$$\int_{\mathbb{R}^d} \tau(v)\, d(\Phi_\nu^t \# \mu_0)(v) = \int_{\mathbb{R}^d} \tau(\Phi_\nu^t(v))\, d\mu_0(v).$$

We seek to find a fixed point to the map $\mu \mapsto \Phi_\mu^t \# \mu_0$ as it would weakly solve (41). To better understand the properties of this map, we need to establish estimates on the flow map through U as a function of time and measures.

Lemma 51 (L^∞ estimate for velocity field). *There exists a constant $C = C(\epsilon, \gamma, R_1, R_2, \mu_0) > 0$ such that, for every $T > 0$ and $\nu \in C([0,T];\mathscr{P}_2(\mathbb{R}^d))$, we have*

$$|U[\nu(t)](v)| \le C \quad \textit{for all } t \in [0,T],\, v \in \mathbb{R}^d.$$

Proof. Estimate for $\gamma \ge -2$: We have the three inequalities

$$|v - v_*|^{\gamma+2} \lesssim_\gamma |v|^{\gamma+2} + |v_*|^{\gamma+2}, \qquad \|\Pi[v - v_*]\| \le 1, \qquad J_0^\epsilon \lesssim_{\epsilon,\mu_0} 1$$

due to the range of γ, boundedness of Π, and Lemma 31, respectively. These three inequalities provide the estimate

$$|U[\nu(t)](v)| \lesssim_{\gamma,\epsilon,\mu_0} \phi_{R_1}(v) \int_{\mathbb{R}^d} \phi_{R_1}(v_*)(|v|^{\gamma+2} + |v_*|^{\gamma+2})\, d\nu_t(v_*),$$

where we have dropped ψ_{R_2} altogether. For the integral term, we apply Hölder's inequality taking advantage of the compact support of ϕ_{R_1} and the unit mass of ν_t to further obtain

$$|U[\nu(t)](v)| \lesssim_{\gamma,\epsilon,\mu_0} \phi_{R_1}(v)(R_1^{2+\gamma} + \langle v\rangle^{2+\gamma}) \int_{\mathbb{R}^d} d\nu_t(v_*) \lesssim_{R_1} \phi_{R_1}(v)\langle v\rangle^{2+\gamma}.$$

Again, since ϕ_{R_1} has compact support, we can brutally estimate the polynomial to conclude.

Estimate for $\gamma < -2$: Unlike the previous case, we change one of the inequalities due to the unavailability of a triangle inequality and use

$$\psi_{R_2}(v - v_*)|v - v_*|^{\gamma+2} \lesssim 1/R_2^{\gamma+2}, \qquad \|\Pi[v - v_*]\| \le 1, \qquad J_0^\epsilon \lesssim_{\epsilon,\mu_0} 1.$$

From these inequalities and the compact support of ϕ_{R_1}, we have

$$|U[\nu(t)](v)| \lesssim_{\gamma,\epsilon,\mu_0,R_2} \phi_{R_1}(v) \int_{\mathbb{R}^d} \phi_{R_1}(v_*)\, d\nu_t(v_*) \le 1. \qquad \square$$

The next result follows exactly as in [Cañizo et al. 2011].

Lemma 52 (time continuity of flow map). *Let $C = C(\epsilon, \gamma, R_1, R_2, \mu_0) > 0$ be the same constant from Lemma 51. Then, for any $T > 0$, and $\nu \in C([0,T];\mathscr{P}_2(\mathbb{R}^d))$ we have*

$$\|\Phi_\nu^t - \Phi_\nu^s\|_{L^\infty(\mathbb{R}^d)} \le C|t - s|.$$

Our next objective is to establish the regularity of the flow map with respect to the measures in the subscript. To simplify the subsequent lemmas, let us use the notation in the following:

Lemma 53. *Define*

$$F : (v, w) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto \phi_{R_1}(v)\phi_{R_1}(w)\psi_{R_2}(v-w)|v-w|^{\gamma+2}\Pi[v-w](J_0^\epsilon(v) - J_0^\epsilon(w)).$$

The function F is smooth and compactly supported. In particular, for every $k, l \in \mathbb{N}$, there is a constant $C = C(\epsilon, \gamma, R_1, R_2, \mu_0, k, l) > 0$ such that

$$\| D_v^k D_w^l F \|_{L^\infty(\mathbb{R}^d \times \mathbb{R}^d)} \leq C.$$

More precisely, the constant C depends on μ_0 only through bounds on its second moment as in Lemma 31.

Proof. The compact support property comes from the factor of $\phi_{R_1}(v)\phi_{R_1}(w)$ in the definition. The regularity comes from the avoidance of $v = w$ due to the factor $\psi_{R_2}(v-w)$. □

Corollary 54 (pointwise and measurewise regularity of U). *Consider the constant C from Lemma 53 above. We have the following*:

(1) *Take $C_1 = C(\epsilon, \gamma, R_1, R_2, \mu_0, 0, 1) > 0$. For every $T > 0$, $\nu^1, \nu^2 \in C([0, T]; \mathscr{P}_2(\mathbb{R}^d))$, $t \in [0, T]$, $v \in \mathbb{R}^d$ we have the estimate*

$$|U[\nu^1(t)](v) - U[\nu^2(t)](v)| \leq C_1 W_2(\nu_t^1, \nu_t^2).$$

(2) *Take $C_2 = C(\epsilon, \gamma, R_1, R_2, \mu_0, 1, 0) > 0$. For every $T > 0$, $\nu \in C([0, T]; \mathscr{P}_2(\mathbb{R}^d))$, $t \in [0, T]$, $v_1, v_2 \in \mathbb{R}^d$ we have the estimate*

$$|U[\nu(t)](v_1) - U[\nu(t)](v_2)| \leq C_2 |v_1 - v_2|.$$

Remark 55. By considering the antisymmetric property of F when swapping variables $v \leftrightarrow w$, one really obtains $C_1 = C_2$. Their distinction in this corollary is artificial.

Proof. (1) Firstly, for every $t \in [0, T]$ take $\pi(t) \in \mathscr{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$, the 2-Wasserstein optimal transportation plan connecting $\nu^1(t)$ and $\nu^2(t)$ which exists; see [Villani 2009]. We estimate the difference with notation from Lemma 53:

$$\begin{aligned} |U[\nu^1(t)](v) - U[\nu^2(t)](v)| &= \left| \int_{\mathbb{R}^d} F(v, w)\, d\nu_t^1(w) - \int_{\mathbb{R}^d} F(v, \bar{w})\, d\nu_t^2(\bar{w}) \right| \\ &= \left| \iint_{\mathbb{R}^{2d}} F(v, w) - F(v, \bar{w}) d\pi_t(w, \bar{w}) \right| \\ &\leq C_1 \iint_{\mathbb{R}^{2d}} |w - \bar{w}| d\pi_t(w, \bar{w}) \leq C_1 W_2(\nu_t^1, \nu_t^2). \end{aligned}$$

The first inequality uses a mean-value-type estimate (in the second variable of F) and the second inequality uses Cauchy–Schwarz, or equivalently, that W_2 is stronger than W_1.

(2) As with item (1), we estimate the difference using F to find

$$|U[\nu(t)](v_1) - U[\nu(t)](v_2)| = \left| \int_{\mathbb{R}^d} F(v_1, w) - F(v_2, w)\, d\nu_t(w) \right| \\ \leq \int_{\mathbb{R}^d} |F(v_1, w) - F(v_2, w)|\, d\nu_t(w) \leq C_2 |v_1 - v_2|.$$

Once more, a mean-value-type estimate is applied (in the first variable of F) and we recall ν_t is a probability measure. □

The next result combines both items of Corollary 54 to estimate the regularity of the flow map with respect to measures and follows exactly as in [Cañizo et al. 2011].

Lemma 56 (continuity of flow map with respect to measures). *For $T > 0$ fix any $\nu^1, \nu^2 \in C([0, T]; \mathscr{P}_2(\mathbb{R}^d))$ and $t \in [0, T]$. With $C := C_1 = C_2$ the same constants in Corollary 54, we have the estimate*

$$\| \Phi^t_{\nu^1} - \Phi^t_{\nu^2} \|_{L^\infty(\mathbb{R}^d)} \leq (e^{Ct} - 1) d(\nu^1, \nu^2),$$

recalling that $d(\nu^1, \nu^2) = \sup_{t \in [0,T]} W_2(\nu^1_t, \nu^2_t)$.

It is by now classical how to obtain Theorem 48 from Corollary 54 and Lemma 56; see [Cañizo et al. 2011; Carrillo et al. 2014; Golse 2016] for instance. The time of existence can be given by any $0 < T < (1/C) \log 2$, where $C > 0$ is chosen as in Lemma 56 and the result follows by a fixed-point argument. The extension to all times is owed to the fact that $C > 0$ depends on the initial data μ_0 only through its second moment. This quantity is conserved through by the evolution of (41) and so the maximal time of existence is $+\infty$.

Acknowledgements

Carrillo was supported the Advanced Grant Nonlocal-CPD (Nonlocal PDEs for Complex Particle Dynamics: Phase Transitions, Patterns and Synchronization) of the European Research Council Executive Agency (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 883363). Carrillo and Delgadino were partially supported by EPSRC grant no. EP/P031587/1. Delgadino was partially supported by CNPq-Brazil (#308800/2019-2) and Instituto Serrapilheira. Wu was funded by the President's PhD Scholarship program of Imperial College London. Carrillo and Wu were also partially supported by the Royal Society through the International Exchange Scheme 2016 CNRS France. Carrillo and Delgadino would like to thank the American Institute of Mathematics since our attendance to the AIM workshop "Nonlocal differential equations in collective behavior" in June 2018 triggered this research.

References

[Alexandre et al. 2015] R. Alexandre, J. Liao, and C. Lin, "Some a priori estimates for the homogeneous Landau equation with soft potentials", *Kinet. Relat. Models* **8**:4 (2015), 617–650. MR Zbl

[Ambrosio 1995] L. Ambrosio, "Movimenti minimizzanti", *Rend. Accad. Naz. Sci. XL Mem. Mat. Appl.* (5) **19** (1995), 191–246. MR Zbl

[Ambrosio et al. 2008] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, 2nd ed., Birkhäuser, Basel, 2008. MR Zbl

[Arsenev and Peskov 1977] A. A. Arsenev and N. V. Peskov, "The existence of a generalized solution of Landau's equation", *Zh. Vychisl. Mat. Mat. Fiz.* **17**:4 (1977), 1063–1068. In Russian; translated in *USSR Comput. Math. and Math. Phys.* **17**:4 (1977), 241–246. MR Zbl

[Barros-Neto 1973] J. Barros-Neto, *An introduction to the theory of distributions*, Pure Appl. Math. **14**, Dekker, New York, 1973. MR Zbl

[Benamou and Brenier 2000] J.-D. Benamou and Y. Brenier, "A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem", *Numer. Math.* **84**:3 (2000), 375–393. MR Zbl

[Blanchet et al. 2008] A. Blanchet, V. Calvez, and J. A. Carrillo, "Convergence of the mass-transport steepest descent scheme for the subcritical Patlak–Keller–Segel model", *SIAM J. Numer. Anal.* **46**:2 (2008), 691–721. MR Zbl

[Boblylev et al. 2013] A. V. Boblylev, M. Pulvirenti, and C. Saffirio, "From particle systems to the Landau equation: a consistency result", *Comm. Math. Phys.* **319**:3 (2013), 683–702. MR Zbl

[Buttazzo 1989] G. Buttazzo, *Semicontinuity, relaxation and integral representation in the calculus of variations*, Pitman Res. Notes Math. Ser. **207**, Longman Sci. Tech., Harlow, England, 1989. MR Zbl

[Cañizo et al. 2011] J. A. Cañizo, J. A. Carrillo, and J. Rosado, "A well-posedness theory in measures for some kinetic models of collective motion", *Math. Models Methods Appl. Sci.* **21**:3 (2011), 515–539. MR Zbl

[Carlen and Carvalho 1992] E. A. Carlen and M. C. Carvalho, "Strict entropy production bounds and stability of the rate of convergence to equilibrium for the Boltzmann equation", *J. Statist. Phys.* **67**:3-4 (1992), 575–608. MR Zbl

[Carrapatoso and Mischler 2017] K. Carrapatoso and S. Mischler, "Landau equation for very soft and Coulomb potentials near Maxwellians", *Ann. PDE* **3**:1 (2017), art. id. 1. MR Zbl

[Carrapatoso et al. 2017] K. Carrapatoso, L. Desvillettes, and L. He, "Estimates for the large time behavior of the Landau equation in the Coulomb case", *Arch. Ration. Mech. Anal.* **224**:2 (2017), 381–420. MR Zbl

[Carrillo et al. 2010] J. A. Carrillo, S. Lisini, G. Savaré, and D. Slepčev, "Nonlinear mobility continuity equations and generalized displacement convexity", *J. Funct. Anal.* **258**:4 (2010), 1273–1309. MR Zbl

[Carrillo et al. 2012] J. A. Carrillo, M. Di Francesco, A. Figalli, T. Laurent, and D. Slepčev, "Confinement in nonlocal interaction equations", *Nonlinear Anal.* **75**:2 (2012), 550–558. MR Zbl

[Carrillo et al. 2014] J. A. Carrillo, Y.-P. Choi, and M. Hauray, "The derivation of swarming models: mean-field limit and Wasserstein distances", pp. 1–46 in *Collective dynamics from bacteria to crowds*, edited by A. Muntean and F. Toschi, CISM Courses and Lect. **553**, Springer, 2014. MR

[Carrillo et al. 2019] J. A. Carrillo, K. Craig, and F. S. Patacchini, "A blob method for diffusion", *Calc. Var. Partial Differential Equations* **58**:2 (2019), art. id. 53. MR Zbl

[Carrillo et al. 2020] J. A. Carrillo, J. Hu, L. Wang, and J. Wu, "A particle method for the homogeneous Landau equation", *J. Comput. Phys. X* **7** (2020), art. id. 100066. MR

[Carrillo et al. 2022] J. A. Carrillo, M. G. Delgadino, and J. Wu, "Boltzmann to Landau from the gradient flow perspective", *Nonlinear Anal.* **219** (2022), art. id. 112824. MR Zbl

[Degond and Lucquin-Desreux 1992] P. Degond and B. Lucquin-Desreux, "The Fokker–Planck asymptotics of the Boltzmann collision operator in the Coulomb case", *Math. Models Methods Appl. Sci.* **2**:2 (1992), 167–182. MR Zbl

[Desvillettes 2015] L. Desvillettes, "Entropy dissipation estimates for the Landau equation in the Coulomb case and applications", *J. Funct. Anal.* **269**:5 (2015), 1359–1403. MR Zbl

[Desvillettes 2016] L. Desvillettes, "Entropy dissipation estimates for the Landau equation: general cross sections", pp. 121–143 in *From particle systems to partial differential equations, III* (Braga, Portugal, 2014), edited by P. Gonçalves and A. J. Soares, Springer Proc. Math. Stat. **162**, Springer, 2016. MR Zbl

[Desvillettes 2022] L. Desvillettes, "Autour du thèorème *H* de Boltzmann", art. id. 9 in *Séminaire Laurent Schwartz: EDP et applications* 2019/2020, Ècole Polytech., Palaiseau, France, 2022.

[Desvillettes and Fellner 2006] L. Desvillettes and K. Fellner, "Exponential decay toward equilibrium via entropy methods for reaction-diffusion equations", *J. Math. Anal. Appl.* **319**:1 (2006), 157–176. MR Zbl

[Desvillettes and Villani 2000a] L. Desvillettes and C. Villani, "On the spatially homogeneous Landau equation for hard potentials, I: Existence, uniqueness and smoothness", *Comm. Partial Differential Equations* **25**:1-2 (2000), 179–259. MR Zbl

[Desvillettes and Villani 2000b] L. Desvillettes and C. Villani, "On the spatially homogeneous Landau equation for hard potentials, II: H-theorem and applications", *Comm. Partial Differential Equations* **25**:1-2 (2000), 261–298. MR Zbl

[Desvillettes et al. 2023] L. Desvillettes, L.-B. He, and J.-C. Jiang, "A new monotonicity formula for the spatially homogeneous Landau equation with Coulomb potential and its applications", *J. Eur. Math. Soc.* (online publication January 2023).

[Dolbeault et al. 2009] J. Dolbeault, B. Nazaret, and G. Savaré, "A new class of transport distances between measures", *Calc. Var. Partial Differential Equations* **34**:2 (2009), 193–231. MR Zbl

[Erbar 2014] M. Erbar, "Gradient flows of the entropy for jump processes", *Ann. Inst. Henri Poincaré Probab. Stat.* **50**:3 (2014), 920–945. MR Zbl

[Erbar 2023] M. Erbar, "A gradient flow approach to the Boltzmann equation", *J. Eur. Math. Soc.* (online publication June 2023).

[Erbar and Maas 2014] M. Erbar and J. Maas, "Gradient flow structures for discrete porous medium equations", *Discrete Contin. Dyn. Syst.* **34**:4 (2014), 1355–1374. MR Zbl

[Fournier and Guérin 2009] N. Fournier and H. Guérin, "Well-posedness of the spatially homogeneous Landau equation for soft potentials", *J. Funct. Anal.* **256**:8 (2009), 2542–2560. MR Zbl

[Golse 2016] F. Golse, "On the dynamics of large particle systems in the mean field limit", pp. 1–144 in *Macroscopic and large scale phenomena*: *coarse graining*, *mean field limits and ergodicity* (Enschede, Netherlands, 2012), edited by A. Muntean et al., Lect. Notes Appl. Math. Mech. **3**, Springer, 2016. MR

[Golse et al. 2019a] F. Golse, M. P. Gualdani, C. Imbert, and A. Vasseur, "Partial regularity in time for the space homogeneous Landau equation with Coulomb potential", preprint, 2019. arXiv 1906.02841

[Golse et al. 2019b] F. Golse, C. Imbert, C. Mouhot, and A. F. Vasseur, "Harnack inequality for kinetic Fokker–Planck equations with rough coefficients and application to the Landau equation", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **19**:1 (2019), 253–295. MR Zbl

[Gualdani and Guillen 2016] M. P. Gualdani and N. Guillen, "Estimates for radial solutions of the homogeneous Landau equation with Coulomb potential", *Anal. PDE* **9**:8 (2016), 1772–1809. MR Zbl

[Gualdani and Zamponi 2017] M. P. Gualdani and N. Zamponi, "Spectral gap and exponential convergence to equilibrium for a multi-species Landau system", *Bull. Sci. Math.* **141**:6 (2017), 509–538. MR Zbl

[Gualdani and Zamponi 2018a] M. Gualdani and N. Zamponi, "A review for an isotropic Landau model", pp. 115–144 in *PDE models for multi-agent phenomena* (Rome, 2016), edited by P. Cardaliaguet and F. Salvarani, Springer INdAM Ser. **28**, Springer, 2018. MR Zbl

[Gualdani and Zamponi 2018b] M. P. Gualdani and N. Zamponi, "Global existence of weak even solutions for an isotropic Landau equation with Coulomb potential", *SIAM J. Math. Anal.* **50**:4 (2018), 3676–3714. MR Zbl

[Guo 2002] Y. Guo, "The Landau equation in a periodic box", *Comm. Math. Phys.* **231**:3 (2002), 391–434. MR Zbl

[Jordan et al. 1998] R. Jordan, D. Kinderlehrer, and F. Otto, "The variational formulation of the Fokker–Planck equation", *SIAM J. Math. Anal.* **29**:1 (1998), 1–17. MR Zbl

[Lifshitz and Pitaevskiĭ 1981] E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical kinetics*, Course of Theoret. Phys. **10**, Pergamon, Oxford, 1981. MR Zbl

[Royden 1963] H. L. Royden, *Real analysis*, Macmillan, New York, 1963. MR Zbl

[Sandier and Serfaty 2004] E. Sandier and S. Serfaty, "Gamma-convergence of gradient flows with applications to Ginzburg–Landau", *Comm. Pure Appl. Math.* **57**:12 (2004), 1627–1672. MR Zbl

[Santambrogio 2017] F. Santambrogio, "(Euclidean, metric, and Wasserstein) gradient flows: an overview", *Bull. Math. Sci.* **7**:1 (2017), 87–154. MR Zbl

[Serfaty 2011] S. Serfaty, "Gamma-convergence of gradient flows on Hilbert and metric spaces and applications", *Discrete Contin. Dyn. Syst.* **31**:4 (2011), 1427–1451. MR Zbl

[Silvestre 2017] L. Silvestre, "Upper bounds for parabolic equations and the Landau equation", *J. Differential Equations* **262**:3 (2017), 3034–3055. MR Zbl

[Strain and Wang 2020] R. M. Strain and Z. Wang, "Uniqueness of bounded solutions for the homogeneous relativistic Landau equation with Coulomb interactions", *Quart. Appl. Math.* **78**:1 (2020), 107–145. MR Zbl

[Villani 1998a] C. Villani, "On a new class of weak solutions to the spatially homogeneous Boltzmann and Landau equations", *Arch. Ration. Mech. Anal.* **143**:3 (1998), 273–307. MR Zbl

[Villani 1998b] C. Villani, "On the spatially homogeneous Landau equation for Maxwellian molecules", *Math. Models Methods Appl. Sci.* **8**:6 (1998), 957–983. MR Zbl

[Villani 2009] C. Villani, *Optimal transport: old and new*, Grundlehren der Math. Wissenschaften **338**, Springer, 2009. MR Zbl

[Wu 2014] K.-C. Wu, "Global in time estimates for the spatially homogeneous Landau equation with soft potentials", *J. Funct. Anal.* **266**:5 (2014), 3134–3155. MR Zbl

Received 23 Oct 2021. Revised 1 Jun 2022. Accepted 27 Oct 2022.

JOSÉ A. CARRILLO: carrillo@maths.ox.ac.uk
Mathematical Institute, University of Oxford, Oxford, United Kingdom

MATIAS G. DELGADINO: matias.delgadino@math.utexas.edu
Department of Mathematics, University of Texas, Austin, TX, United States

LAURENT DESVILLETTES: desvillettes@math.univ-paris-diderot.fr
Université Paris Cité and Sorbonne Université, CNRS, IUF, IMJ-PRG, Paris, France

JEREMY S.-H. WU: jeremywu@math.ucla.edu
Department of Mathematics, University of California, Los Angeles, CA, United States

ANALYSIS AND PDE
Vol. 17 (2024), No. 4, pp. 1377–1395
DOI: 10.2140/apde.2024.17.1377

DEGENERATING HYPERBOLIC SURFACES AND SPECTRAL GAPS FOR LARGE GENUS

YUNHUI WU, HAOHAO ZHANG AND XUWEN ZHU

We study the differences of two consecutive eigenvalues $\lambda_i - \lambda_{i-1}$, i up to $2g-2$, for the Laplacian on hyperbolic surfaces of genus g, and show that the supremum of such spectral gaps over the moduli space has infimum limit at least $\frac{1}{4}$ as the genus goes to infinity. A min-max principle for eigenvalues on degenerating hyperbolic surfaces is also established.

1. Introduction

For a closed Riemann surface X_g of genus $g \geq 2$, consider the hyperbolic metric uniquely determined by its complex structure. We study the spectrum of the Laplacian on X_g, which is a discrete subset in $\mathbb{R}^{\geq 0}$ and consists of eigenvalues with finite multiplicities. The eigenvalues, counted with multiplicities, are listed in the following increasing order:

$$0 = \lambda_0(X_g) < \lambda_1(X_g) \leq \lambda_2(X_g) \leq \cdots \to \infty.$$

Let $\mathcal{M}_g$ be the moduli space of Riemann surfaces of genus g, which is an open orbifold of dimension equal to $6g-6$. For each index i, the i-th eigenvalue $\lambda_i(\cdot)$ is a bounded continuous function on $\mathcal{M}_g$. In this paper we study the differences of two consecutive eigenvalues and will focus on the behavior of such spectral gaps when $g \to \infty$.

Definition. For all $i \geq 1$, the *i-th spectral gap* $\mathrm{SpG}_i(\cdot)$ is a bounded continuous function over the moduli space $\mathcal{M}_g$ defined as

$$\mathrm{SpG}_i : \mathcal{M}_g \to \mathbb{R}^{\geq 0}, \quad X_g \mapsto \lambda_i(X_g) - \lambda_{i-1}(X_g). \tag{1}$$

By definition, $\mathrm{SpG}_1(X_g) = \lambda_1(X_g)$. For all $i \geq 1$, the i-th spectral gap $\mathrm{SpG}_i(\cdot)$ can be arbitrarily close to zero (e.g., see Proposition 3.7). In this paper we mainly study the quantity $\sup_{X_g \in \mathcal{M}_g} \mathrm{SpG}_i(X_g)$ for large g and a family of indices i.

The main result of this article is the limiting behavior of the lower bound of the spectral gaps.

Theorem 4.1. *Let $\{\eta(g)\}_{g=2}^{\infty}$ be any sequence of integers with $\eta(g) \in [1, 2g-2]$. Then*

$$\liminf_{g\to\infty} \sup_{X_g \in \mathcal{M}_g} \mathrm{SpG}_{\eta(g)}(X_g) \geq \tfrac{1}{4}.$$

MSC2020: primary 32G15; secondary 58C40.
Keywords: spectral gaps, min-max principle, large genus.

Remark. The sequence $\{\eta(g)\}$ is arbitrary as long as it satisfies the bounds: examples include $\eta(g) \equiv 2$, $\eta(g) = \{2, 3, 2, 3, \dots\}$, and $\eta(g) = 2g - 2$.

On the other hand, by [Cheng 1975, Corollary 2.3], we know that

$$\lambda_i(X_g) \le \frac{1}{4} + i^2 \cdot \frac{16\pi^2}{\mathrm{Diam}^2(X_g)}.$$

By Gauss–Bonnet, $\mathrm{Area}(X_g) = 4\pi(g-1)$. A simple area argument implies that the diameter satisfies $\mathrm{Diam}(X_g) \ge C \ln(g)$ for some universal constant $C > 0$. So if $\eta(g)$ satisfies

$$\lim_{g\to\infty} \frac{\eta(g)}{\ln(g)} = 0,$$

we have

$$\limsup_{g\to\infty} \sup_{X_g \in \mathcal{M}_g} \mathrm{SpG}_{\eta(g)}(X_g) \le \tfrac{1}{4}.$$

Together with Theorem 4.1 this yields the following direct consequence.

Corollary 1.1. *If $\eta(g) = o(\ln(g))$, then*

$$\lim_{g\to\infty} \sup_{X_g \in \mathcal{M}_g} \mathrm{SpG}_{\eta(g)}(X_g) = \tfrac{1}{4}.$$

For $\eta(g) = 1$, both Theorem 4.1 and Corollary 1.1 are due to Hide and Magee [2023, Corollary 1.3], who used a probabilistic method to solve the conjecture (e.g., see [Buser 1984; Buser et al. 1988]) that there exists a sequence of closed hyperbolic surfaces with first eigenvalues tending to $\frac{1}{4}$ as the genus goes to infinity.

The following result is important in the proof of Theorem 4.1, which we include for independent interest. The proof is highly motivated by the work of Burger, Buser and Dodziuk [Buser et al. 1988], where they studied the case when the limiting surface is connected (e.g., see Theorem 2.6).

Proposition 3.1 (min-max principle). *Let $X_g(0) \in \partial\mathcal{M}_g$ be the limit of a family of Riemann surfaces $\{X_g(t)\}$ obtained by pinching certain simple closed geodesics such that $X_g(0)$ has k connected components, i.e., $X_g(0) = Y_1 \sqcup Y_2 \sqcup \cdots \sqcup Y_k$, where $k \ge 2$. Let $\lambda_1(Y_1), \dots, \lambda_1(Y_k)$ be the first nonzero eigenvalue of $Y_1, \dots, Y_k$ (if Y_i has no discrete eigenvalues then write $\lambda_1(Y_i) = \infty$) and write $\bar{\lambda}_1(*) = \min\{\lambda_1(*), \frac{1}{4}\}$ for $* = Y_1, \dots, Y_k$. Then*

$$\liminf_{t\to 0} \lambda_k(X_g(t)) \ge \min_{1\le i\le k} \{\bar{\lambda}_1(Y_i)\}.$$

Remark. Each component Y_i in the proposition above is a complete open hyperbolic surface of finite volume, whose spectrum consists of possibly discrete eigenvalues and the continuous spectrum $\left[\frac{1}{4}, \infty\right)$. Therefore, in the statement above, $\bar{\lambda}_1(Y_i)$ is the nonzero minimum of the spectrum of Y_i.

Proof sketch of Theorem 4.1. In the proof of Theorem 4.1, we will apply Proposition 3.1 to the case when all the $\bar{\lambda}_1(Y_i)$ are close to $\frac{1}{4}$. The main idea is the following: for each $\eta(g)$ we construct a sequence of genus g surfaces that degenerate into $\eta(g)$ components using only pieces that are known to have the first eigenvalue close to $\frac{1}{4}$. Then by the min-max principle, the $\eta(g)$-th eigenvalue of these surfaces will be close to $\frac{1}{4}$. On the other hand, by a result of Schoen, Wolpert and Yau (see Theorem 2.5), the $(\eta(g)-1)$-th eigenvalue is close to zero. This way we find sequences of surfaces that achieve the spectral gap of $\frac{1}{4}$. For the regime $\eta(g) > g$, the components used in the construction only include the thrice-punctured sphere and a twice-punctured torus. On the other hand, for $\eta(g) \leq g$, the essential components also include a large genus piece that relies on the work of Hide and Magee [2023].

Plan of the paper. Section 2 will first discuss properties of the boundary degeneration of the Riemann moduli spaces; then we will provide a review of the background and recent developments on spectral gaps on hyperbolic surfaces, including a list of punctured surface components with eigenvalue bounds which will be used in the degeneration limits. In Section 3 we will provide a proof for Proposition 3.1 regarding the min-max principle for eigenvalues of degenerating hyperbolic surfaces and a few immediate applications. In Section 4 we will complete the proof of Theorem 4.1.

2. Preliminaries

Boundary of the Riemann moduli spaces. Denote by $\mathcal{M}_{g,n}$ the moduli space of hyperbolic surfaces of genus g with n punctures, and by $\mathcal{M}_g := \mathcal{M}_{g,0}$ the moduli space of compact hyperbolic surfaces with genus g. It is well known that $\dim_{\mathbb{R}}(\mathcal{M}_{g,n}) = 6g + 2n - 6$. In particular, $\mathcal{M}_{0,3}$ contains only one point represented by the hyperbolic thrice-punctured sphere. The Deligne–Mumford compactification of $\mathcal{M}_{g,n}$ is obtained by adding nodal surfaces into $\mathcal{M}_{g,n}$, which is homeomorphic to the completion of $\mathcal{M}_{g,n}$ endowed with the Weil–Petersson metric. Let $\partial\mathcal{M}_{g,n}$ be the boundary of the Deligne–Mumford compactification of $\mathcal{M}_{g,n}$. Recall that $\partial\mathcal{M}_{g,n}$ is stratified, and each stratum of $\partial\mathcal{M}_{g,n}$ is a product of lower-dimensional moduli spaces. Points in $\partial\mathcal{M}_{g,n}$ are represented by hyperbolic nodal surfaces in $\mathcal{M}_{g,n}$ (see for example [Masur 1976] for more details on the completion of $\mathcal{M}_{g,n}$). Locally the process of pinching a simple closed geodesic into a pair of cusp points can be written with respect to hyperbolic collar coordinates (ρ, θ) with ℓ the length of the central geodesic circle. As $\ell \to 0$, the hyperbolic neck degenerates into a pair of cusps, which can be seen with the choice of the correct coordinates (see for example [Ji 1993; Masur 1976]). Another way to see this would be using the complex "plumbing" coordinates, which we will not discuss. Hyperbolic nodal surfaces are obtained by pinching certain disjoint geodesic circles, and we call such a family of hyperbolic metrics approaching nodal surfaces a degenerating family (see, e.g., [Wolpert 1990], and see Figure 1 for an example).

We also recall the collar lemma on structures of disjoint hyperbolic collars around short geodesics, which will be useful later in decomposing the surfaces.

Lemma 2.1 (collar lemma [Buser 1992, Theorem 4.1.1]). *Let $\gamma_1, \gamma_2, \ldots, \gamma_m$ be disjoint simple closed geodesics on a closed hyperbolic Riemann surface X_g, and let $\ell(\gamma_i)$ be the length of γ_i. Then $m \leq 3g - 3$, and we can define the collar of γ_i by*

$$T(\gamma_i) = \{x \in X_g : \mathrm{dist}(x, \gamma_i) \leq w(\gamma_i)\},$$

where

$$w(\gamma_i) = \operatorname{arcsinh} \frac{1}{\sinh\left(\frac{1}{2}\ell(\gamma_i)\right)} \tag{2}$$

is the width of the collar.

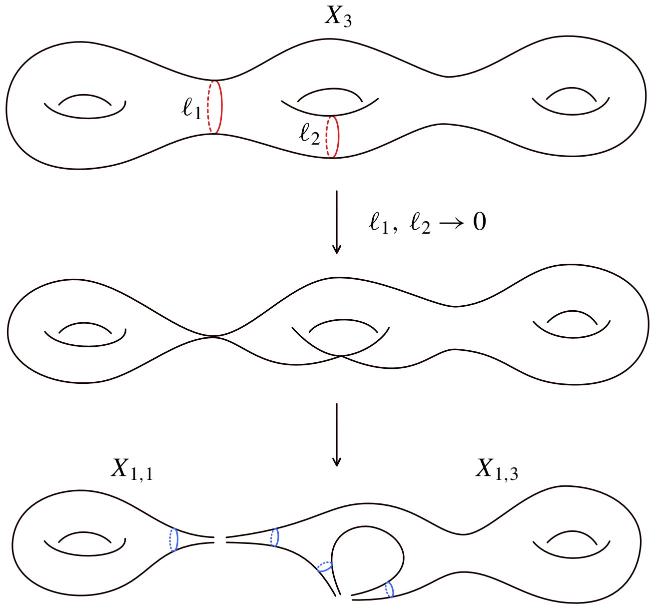


Figure 1. An example of a degenerating family in $\mathcal{M}_3$ whose limit is $X_{1,1} \sqcup X_{1,3}$, which is disconnected.

Then the collars are pairwise disjoint for $i = 1, \ldots, m$. *Each* $T(\gamma_i)$ *is isomorphic to a cylinder* $(\rho, \theta) \in [-w(\gamma_i), w(\gamma_i)] \times \mathbb{S}^1$, *where* $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$, *with the metric*

$$ds^2 = d\rho^2 + \ell(\gamma_i)^2 \cosh^2 \rho \, d\theta^2. \tag{3}$$

For a point (ρ, θ), *the point* $(0, \theta)$ *is its projection on the geodesic* γ_i, $|\rho|$ *is the distance to* γ_i, *and* θ *is the coordinate on* $\gamma_i \cong \mathbb{S}^1$.

As the length $\ell(\gamma)$ of the central closed geodesic goes to zero, the width $w(\gamma)$ is approximately $\ln(1/\ell(\gamma))$, which tends to infinity. We have the following as an easy corollary.

Corollary 2.2. *For a degenerating family of hyperbolic surfaces* $\{X_g(t)\}$, *the diameter satisfies*

$$\mathrm{Diam}(X_g(t)) \to \infty.$$

The following two lemmas will be useful in the proof of Theorem 4.1.

Lemma 2.3. *For each integer* $\eta(g) \in [g-1, 2g-2]$ *with* $g \geq 2$, *there exist two nonnegative integers* i *and* j *such that*

(1) $i + j = \eta(g)$,

(2) $\underbrace{\mathcal{M}_{0,3} \times \cdots \times \mathcal{M}_{0,3}}_{i \text{ copies}} \times \underbrace{\mathcal{M}_{1,2} \times \cdots \times \mathcal{M}_{1,2}}_{j \text{ copies}} \subset \partial \mathcal{M}_g$.

Remark. Here i and j depend on g and satisfy $i + 2j = 2g - 2$ by the additivity of the Euler characteristic.

Proof. If $\eta(g) = 2g - 2$, the conclusion is obvious by choosing $i = 2g - 2$ and $j = 0$, which is obtained by pinching $3g - 3$ disjoint simple closed curves in a closed surface X_g of genus g.

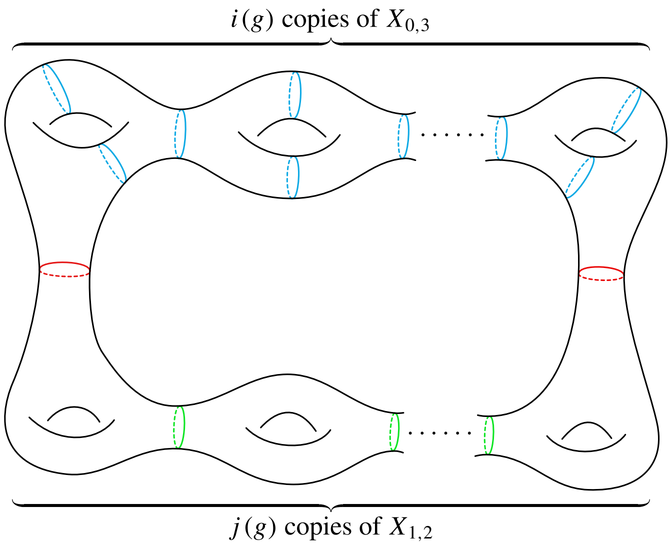


Figure 2. An example of the degeneration of a genus g surface into $i(g)$ copies of $X_{0,3}$ and $j(g)$ copies of $X_{1,2}$ by pinching all the simple geodesics marked in the picture.

Now we assume $g \leq \eta(g) \leq 2g-3$. Given a closed surface X_g of genus g, first one may pinch X_g along two disjoint simple closed curves σ_1 and σ_2 such that $X_g \setminus (\sigma_1 \cup \sigma_2)$ has two connected components $X_{g_1,2} \sqcup X_{g_2,2}$, where g_1 and g_2 are two nonnegative integers satisfying $g_1 + g_2 = g-1$. Here we choose

$$g_1 = (2g-2) - \eta(g) \quad \text{and} \quad g_2 = \eta(g) - (g-1).$$

For the second step, we pinch $X_{g_1,2}$ along $g_1 - 1$ disjoint simple closed curves $\{\gamma_l\}_{1 \leq l \leq g_1 - 1}$ such that the complement decomposes further into g_1 components:

$$X_{g_1,2} \setminus \bigcup_{1 \leq l \leq g_1 - 1} \gamma_l = \underbrace{X_{1,2} \sqcup \cdots \sqcup X_{1,2}}_{g_1 \text{ copies}}.$$

For $X_{g_2,2}$, one may pinch along $3g_2 - 1$ disjoint simple closed curves $\{\gamma'_m\}_{1 \leq m \leq 3g_2 - 1}$ such that the complement decomposes further into $2g_2$ components:

$$X_{g_2,2} \setminus \bigcup_{1 \leq m \leq 3g_2 - 1} \gamma'_m = \underbrace{X_{0,3} \sqcup \cdots \sqcup X_{0,3}}_{2g_2 \text{ copies}}.$$

Pinching all these simple closed curves during cutting above to zero, the conclusion follows since

$$i = 2g_2 = 2\eta(g) - (2g-2) \quad \text{and} \quad j = g_1 = (2g-2) - \eta(g). \tag{4}$$

For an illustration, see Figure 2.

If $\eta(g) = g-1$, we first pinch X_g along a nonseparating simple closed curve to get a surface $X_{g-1,2}$. Then in the same way as with $X_{g_1,2}$ in the previous case, we pinch $X_{g-1,2}$ along $g-2$ disjoint simple closed curves to get $g-1$ copies of $X_{1,2}$. Then the conclusion follows with $i = 0$ and $j = g-1$.

Combining the three cases above, the proof is complete. □

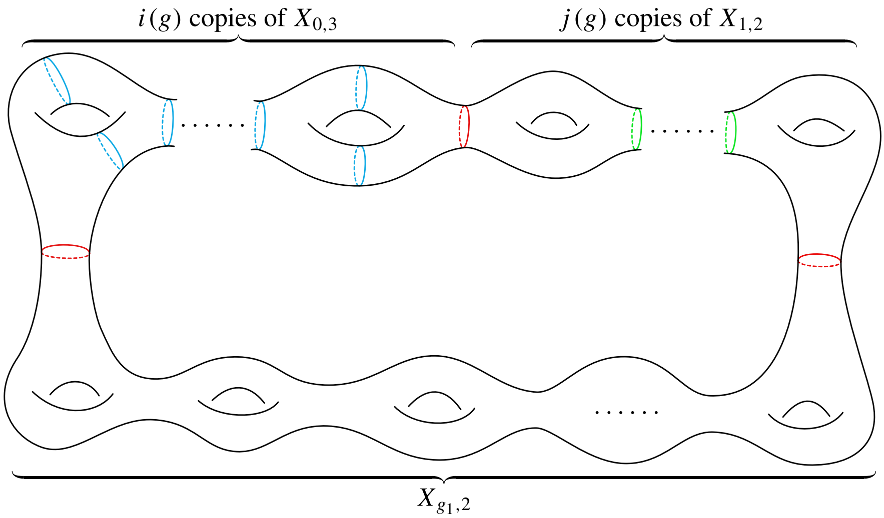


Figure 3. An example of decomposing a surface of genus g into i copies of $X_{0,3}$, j copies of $X_{1,2}$ and a copy of $X_{g_1,2}$, where i, j, and g_1 are given in the proof of Lemma 2.4.

Lemma 2.4. *For each integer $\eta(g) \in [2, g]$ with $g \geq 3$, there exist three nonnegative integers g_1, i and j such that*

(1) $2g_1 \geq g - 2$,

(2) $i + j + 1 = \eta(g)$,

(3) $\underbrace{\mathcal{M}_{0,3} \times \cdots \times \mathcal{M}_{0,3}}_{i(g) \text{ copies}} \times \underbrace{\mathcal{M}_{1,2} \times \cdots \times \mathcal{M}_{1,2}}_{j(g) \text{ copies}} \times \mathcal{M}_{g_1,2} \subset \partial \mathcal{M}_g$.

Remark. Similar to the previous lemma, i, j and g_1 depend on g. By calculating the Euler characteristics, these numbers should satisfy $i + 2j + 2g_1 = 2g - 2$.

Proof. Similar to the proof of Lemma 2.3 above, we first decompose X_g as $X_g \setminus (\sigma_1 \cup \sigma_2) = X_{g_1,2} \sqcup X_{g_2,2}$ for two disjoint simple closed curves σ_1 and σ_2, where g_1 and $g_2 := g - 1 - g_1$ will be determined in different cases below. Next we decompose $X_{g_2,2}$ into the disjoint union of i copies of $X_{0,3}$ and j copies of $X_{1,2}$ to obtain the desired properties. For an illustration, see Figure 3.

The proof contains the following three cases.

Case 1: $2 \leq \eta(g) \leq \frac{1}{2}g + 1$. The conclusion follows by choosing

$$i = 0, \quad j = \eta(g) - 1 \quad \text{and} \quad g_1 = g - \eta(g).$$

Case 2: $\frac{1}{2}g + 1 < \eta(g) \leq g$ and $\eta(g)$ is odd. The conclusion follows by choosing

$$i = \eta(g) - 1, \quad j = 0 \quad \text{and} \quad g_1 = g - \tfrac{1}{2}(1 + \eta(g)).$$

Case 3: $\frac{1}{2}g + 1 < \eta(g) \leq g$ and $\eta(g)$ is even. The conclusion follows by choosing

$$i = \eta(g) - 2, \quad j = 1 \quad \text{and} \quad g_1 = g - 1 - \tfrac{1}{2}\eta(g).$$

□

Eigenvalues of hyperbolic surfaces. The study of eigenvalues of the Laplacian on hyperbolic surfaces has a long history and has recently seen much progress. For a compact hyperbolic surface, the eigenvalues are discrete. On the other hand, when the hyperbolic surface degenerates to one with cusps, by [Lax and Phillips 1982] it is known that the spectrum is no longer discrete, rather it consists of a continuous spectrum $\left[\frac{1}{4}, \infty\right)$ and (possibly) additional discrete eigenvalues. The study of spectral degeneration has seen many developments; see [Hejhal 1990; Ji 1993; Ji and Zworski 1993; Wolpert 1987; 1992a; 1992b] for some of the earlier works.

An eigenvalue of a hyperbolic surface is said to be "small" if it is less than $\frac{1}{4}$, where the number $\frac{1}{4}$ shows up as the bottom of the continuous spectrum of a hyperbolic surface with cusps. The questions of existence of eigenvalues less than $\frac{1}{4}$ for both noncompact and compact hyperbolic surfaces not only arise in the field of spectral geometry, but also have deep relations to number theory regarding arithmetic hyperbolic surfaces, dating back to Selberg's famous $\frac{3}{16}$ theorem [1965]. We refer to [Gelbart and Jacquet 1978; Kim 2003; Luo et al. 1995] for more recent developments. Regarding the estimates and multiplicity counting of small eigenvalues, the history goes back to McKean [1972], Randol [1974], and Buser [1982; 1984]. Recently there have been many developments; see [Ballmann et al. 2016; 2017; 2018; Brooks and Makover 2001; Buser 1992; Buser et al. 1988; Mondal 2015; Otal and Rosas 2009; Schoen et al. 1980]. Among these are two classical results of particular relevance to our current work. The first regards bounds of eigenvalues on degenerating hyperbolic surfaces by Schoen, Wolpert and Yau [Schoen et al. 1980]:

Theorem 2.5 [Schoen et al. 1980]. *For any compact hyperbolic surface X_g of genus g and integer $i \in (0, 2g-2)$, the i-th eigenvalue satisfies*

$$\alpha_i(g) \cdot \ell_i \le \lambda_i \le \beta_i(g) \cdot \ell_i$$

and

$$\alpha(g) \le \lambda_{2g-2},$$

where $\alpha_i(g) > 0$ and $\beta_i(g) > 0$ depend only on i and g, $\alpha(g) > 0$ depends only on g, and ℓ_i is the minimal possible sum of the lengths of simple closed geodesics in X_g which cut X_g into $i+1$ connected components.

Dodziuk and Randol [1986] gave an alternative proof of Theorem 2.5, and one may also see Dodziuk, Pignataro, Randol and Sullivan [Dodziuk et al. 1987] on similar results for Riemann surfaces with punctures. It was proved by Otal and Rosas [2009] that the constant $\alpha(g)$ can be optimally chosen to be $\frac{1}{4}$. For large genus g, it was recently proved by the first-named author and Xue [Wu and Xue 2022a; 2022c] that up to multiplication by a universal constant, $\alpha_1(g)$ can be optimally chosen to be $1/g^2$.

The other result that is relevant is [Buser et al. 1988, Theorem 2.1] regarding the first eigenvalue when the limiting degenerating surface is connected:

Theorem 2.6 [Buser et al. 1988]. *Let $\{X_g(t)\} \subset \mathcal{M}_g$ such that $Y = \lim_{t\to 0} X_g(t) \in \partial\mathcal{M}_g$ is connected. Denote by $\lambda_1(Y)$ the first nonzero eigenvalue of Y (if Y has no discrete eigenvalues we write $\lambda_1(Y) = \infty$). Then*

$$\limsup_{t\to 0} \lambda_1(X_g(t)) \ge \bar{\lambda}_1(Y) = \min\left\{\lambda_1(Y), \tfrac{1}{4}\right\}.$$

In Section 3 we will give a similar description of $\lambda_k(X_g(t))$ when the limiting surface has k connected components.

Another related direction in this topic is to understand how the genus of the hyperbolic surface, in particular when $g \to \infty$, affects the eigenvalues via different models of random hyperbolic surfaces. Brooks and Makover [2004] gave a uniform lower bound on the first spectral gap for their combinatorial model of random surfaces by gluing hyperbolic ideal triangles. In terms of Weil–Petersson random closed hyperbolic surfaces, Mirzakhani [2013] showed that the first eigenvalue is greater than 0.0024 with probability one as $g \to \infty$. Recently, the first-named author and Xue [Wu and Xue 2022b] improved this lower bound 0.0024 to be $\frac{3}{16} - \epsilon$, which was also independently obtained by Lipnowski and Wright [2024]. One may also see [Hide 2022] for similar results on Weil–Petersson random punctured hyperbolic surfaces and [Monk 2021] for related results. Recently there have also been many exciting developments in the case of random covers of both compact and noncompact hyperbolic surfaces; see [Magee and Naud 2020; 2021, Magee and Puder 2023,Magee et al. 2022]. For example, Magee, Naud and Puder [Magee et al. 2022] showed that a generic covering of a hyperbolic surface has relative spectral gap of size $\frac{3}{16} - \epsilon$, which was improved to $\frac{1}{4} - \epsilon$ by Hide and Magee [2023] for random covers of punctured hyperbolic surfaces. As an important application, [Hide and Magee 2023] proved that

$$\lim_{g\to\infty} \sup_{X_g \in \mathcal{M}_g} \lambda_1(X_g) = \tfrac{1}{4}.$$

This result provides major inspiration for our current paper.

One major ingredient of our proof is the existence of punctured surfaces with first eigenvalue close to $\frac{1}{4}$. We summarize those components in the two theorems below.

Theorem 2.7. (1) $\lambda_1(X_{0,3}) \geq \frac{1}{4}$;

(2) [Mondal 2015] *There exists a surface* $X_{1,2} \in \mathcal{M}_{1,2}$ *such that* $\lambda_1(X_{1,2}) \geq \frac{1}{4}$.

Proof. The first item is well known; see for example [Otal and Rosas 2009] or [Ballmann et al. 2016]. The existence of the second item was proved by Mondal [2015, Theorem 1.3]. □

The third component is from the recent breakthrough by Hide and Magee [2023]. They use probabilistic methods to show that for any $\epsilon > 0$, there exists an integer $\delta(\epsilon) > 0$ only depending on ϵ such that for all $g > \delta(\epsilon)$ there exists a $2g$-cover $\mathcal{X}$ of $X_{0,3}$ such that

$$\bar{\lambda}_1(\mathcal{X}) = \min\{\lambda_1(\mathcal{X}), \tfrac{1}{4}\} > \tfrac{1}{4} - \epsilon.$$

It is not hard to see that $\mathcal{X}$ must have an even number of punctures because the Euler characteristic of $\mathcal{X}$ is equal to $-2g$, which is even. Then one may apply the handle lemma of [Buser et al. 1988] (or see [Brooks and Makover 2001, Lemma 1.1]) to get the following.

Theorem 2.8. *For any* $\epsilon > 0$ *and large enough* $g > 0$, *there exists a hyperbolic surface* $\mathcal{X}_{g,2} \in \mathcal{M}_{g,2}$ *such that*

$$\bar{\lambda}_1(\mathcal{X}_{g,2}) = \min\{\lambda_1(\mathcal{X}_{g,2}), \tfrac{1}{4}\} > \tfrac{1}{4} - \epsilon.$$

Proof. For completeness we sketch an outline of the proof. Suppose by contradiction there exists a constant $\epsilon_0 > 0$ such that

$$\liminf_{g\to\infty} \sup_{X\in\mathcal{M}_{g,2}} \lambda_1(X) \le \tfrac{1}{4} - \epsilon_0. \tag{5}$$

It follows by [Hide and Magee 2023] that, for any $\epsilon > 0$ and large enough g, there exists a $2g$-cover $\mathcal{X}$ of $X_{0,3}$ such that

$$\bar{\lambda}_1(\mathcal{X}) = \min\{\lambda_1(\mathcal{X}), \tfrac{1}{4}\} > \tfrac{1}{4} - \epsilon.$$

Since the Euler characteristic $\chi(\mathcal{X}) = -2g$ is even, one may assume that $\mathcal{X}$ has an even number of cusps. As in [Buser et al. 1988] we can construct a family of hyperbolic surfaces $\{X_{g,2}(t)\} \subset \mathcal{M}_{g,2}$ such that

$$\lim_{t\to 0} X_{g,2}(t) = \mathcal{X} \in \partial\mathcal{M}_{g,2}.$$

By [Lax and Phillips 1982] we know that, for a hyperbolic surface with cusps, the spectrum below $\frac{1}{4}$ is discrete and only contains eigenvalues. By (5), for some large g one may assume that ϕ_t is the first eigenfunction on $X_{g,2}(t)$ with $\Delta\phi_t = \lambda_1(X_{g,2}(t)) \cdot \phi_t$ on $X_{g,2}(t)$. Then one may apply the handle lemma of [Buser et al. 1988] (or see [Brooks and Makover 2001, Lemma 1.1]) to obtain

$$\limsup_{t\to 0} \lambda_1(X_{g,2}(t)) \ge \bar{\lambda}_1(\mathcal{X}) = \min\{\lambda_1(\mathcal{X}), \tfrac{1}{4}\} > \tfrac{1}{4} - \epsilon,$$

which is a contradiction to (5) since $\epsilon > 0$ can be chosen to be arbitrarily small. □

3. Eigenvalues on a family of degenerating Riemann surfaces

In this section we will prove the following min-max principle, which was stated earlier.

Proposition 3.1 (min-max principle). *Let $X_g(0) \in \partial\mathcal{M}_g$ be the limit of a family of Riemann surfaces $\{X_g(t)\}$ obtained by pinching certain simple closed geodesics such that $X_g(0)$ has k connected components, i.e., $X_g(0) = Y_1 \sqcup Y_2 \sqcup \cdots \sqcup Y_k$, where $k \ge 2$. Let $\lambda_1(Y_1), \ldots, \lambda_1(Y_k)$ be the first nonzero eigenvalue of $Y_1, \ldots, Y_k$ (if Y_i has no discrete eigenvalues then write $\lambda_1(Y_i) = \infty$) and write $\bar{\lambda}_1(*) = \min\{\lambda_1(*), \frac{1}{4}\}$ for $* = Y_1, \ldots, Y_k$. Then*

$$\liminf_{t\to 0} \lambda_k(X_g(t)) \ge \min_{1\le i\le k} \{\bar{\lambda}_1(Y_i)\}.$$

To prove the theorem, we will start by discussing the subsequence limits of eigenfunctions. Denote by $\phi_t \in C^\infty(X_g(t))$ (one of) the normalized eigenfunctions corresponding to $\lambda_k(X_g(t))$, i.e.,

$$\Delta_{X_g(t)}\phi_t = \lambda_k(X_g(t)) \cdot \phi_t \quad \text{and} \quad \int_{X_g(t)} |\phi_t|^2 \,\mathrm{dVol}_{X_g(t)} = 1.$$

By [Cheng 1975, Corollary 2.3] we know that for any compact hyperbolic surface X there is an upper bound

$$\lambda_k(X) \le \frac{1}{4} + k^2 \cdot \frac{16\pi^2}{\mathrm{Diam}^2(X)}.$$

Note that $\mathrm{Diam}(X_g(t)) \to \infty$ as $t \to 0$ by Corollary 2.2 for any family of degenerating hyperbolic surfaces $\{X_g(t)\}$ as described in the proposition above. This gives that, for any fixed $k \geq 1$,

$$\liminf_{t\to 0} \lambda_k(X_g(t)) \leq \limsup_{t\to 0} \lambda_k(X_g(t)) \leq \tfrac{1}{4}. \tag{6}$$

On the other hand, by Theorem 2.5 we know that the lowest $k-1$ eigenvalues of $X_g(t)$ go to zero when the degenerating limit has k components, while the k-th eigenvalue $\lambda_k(X_g(t))$ stays bounded away from zero. Therefore

$$\liminf_{t\to 0} \lambda_k(X_g(t)) > 0. \tag{7}$$

Now consider

$$\lambda_k(0) := \liminf_{t\to 0} \lambda_k(X_g(t)). \tag{8}$$

By the discussion above we know that

$$0 < \lambda_k(0) \leq \tfrac{1}{4}. \tag{9}$$

By the collar lemma, Lemma 2.1, each $X_g(t)$ can be decomposed into a number of disjoint degenerating hyperbolic necks and a compact part (which has possibly several connected components). The width of each hyperbolic neck is determined by the central shrinking geodesic γ and can be chosen to be $w(\gamma)-1$, for example, where $w(\gamma)$ is given in (2). For the degenerating family $\{X_g(t)\}$ with N shrinking geodesics $\{\gamma_m(t)\}_{m=1}^N$, we denote the width of each hyperbolic neck by the following N-tuple:

$$\vec{w} := (w(\gamma_1(t)) - 1, w(\gamma_2(t)) - 1, \dots, w(\gamma_N(t)) - 1).$$

Note that $\vec{w}$ depends on t, and each entry in $\vec{w}$ goes to ∞ as t goes to zero. Geometrically each hyperbolic neck degenerates into a pair of cusps. We remark here that in the definition of $\vec{w}$, the choice $w(\gamma)-1$ is for convenience and can be replaced by $w(\gamma)-c$ for any $c > 0$.

For any $X_g(t)$, we denote the union of all N hyperbolic necks as $C_{\vec{w}}(t)$. In local hyperbolic geodesic coordinates given by $d\rho^2 + \ell^2 \cosh^2 \rho\, d\theta^2$ where ℓ is the length of the central geodesic circle γ_i,

$$C_{\vec{w}}(t) = \bigcup_{m=1}^{N} \{(\rho,\theta) : 0 \leq |\rho| \leq w(\gamma_m(t)) - 1\}. \tag{10}$$

In addition, we also denote the union of all "shells" near the collars by

$$S_{\vec{w}}(t) = \bigcup_{m=1}^{N} \{(\rho,\theta) : w(\gamma_m(t)) - 1 \leq |\rho| \leq w(\gamma_m(t))\}. \tag{11}$$

Then it follows by the collar lemma that all such collar neighborhoods (and shells) are disjoint; see Figure 4 for an illustration of collars and shells.

Denote the compact part by $F_{\vec{w}}(t) = X_g(t) \setminus C_{\vec{w}}(t)$. The compact area and nodal degeneration area are grafted together [Melrose and Zhu 2018; 2019; Wolpert 1990]. For small t, the $F_{\vec{w}}(t)$ are all diffeomorphic. In particular, the metric on $F_{\vec{w}}(t)$ can be written as $e^{2u_t} g_0$, where g_0 is the metric on $F_{\vec{w}}(0)$ and u_t is polyhomogeneous and uniformly bounded in all derivatives [Melrose and Zhu 2019]. That is, we can write the diffeomorphism $D_t : F_{\vec{w}}(t) \to F_{\vec{w}}(0)$ such that $g_t = D_t^* g_0$ and D_t are uniformly bounded. From now on, when we consider the convergence of eigenfunctions $\phi(t)$ on $X_g(t)$, the functions are all defined

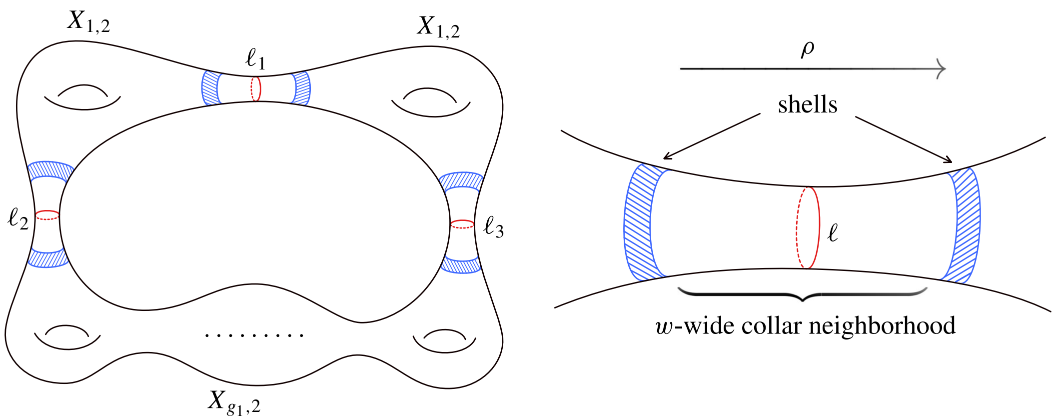


Figure 4. An example of collar neighborhoods and shells.

on $X_g(0)$ via the pullback $(D_t^{-1})^*\phi(t)$; see [Wolpert 1992a; 1992b] for similar approaches. See also another related approach via universal covers in [Buser et al. 1988].

Now take a sequence of metrics such that the corresponding sequence of eigenvalues approaches $\lambda_k(0)$, which is defined in (8). Denote the sequence by $\{X_g(t_i)\}_{i=1}^{\infty}$. By definition,

$$\lim_{i\to\infty} t_i = 0 \quad \text{and} \quad \lim_{i\to\infty} \lambda_k(X_g(t_i)) = \lambda_k(0).$$

Denote the corresponding eigenfunction on $X_g(t_i)$ by ϕ_{t_i}; we discuss the convergence of the sequence of functions $\{\phi_{t_i}\}_{i=1}^{\infty}$ below. One key ingredient is the following Sobolev–Gårding Inequality on the compact part $F_{\vec{w}}(t)$. Denote by $\operatorname{inj}(\cdot)$ the injectivity radius function. Denote by $\nabla^j\phi_{t_i}$ the j-th covariant derivative of ϕ_{t_i}, where $j \in \mathbb{N}$. Then we have the following.

Lemma 3.2. *For any $x \in F_{\vec{w}}(t)$, $j \in \mathbb{N}$ and $r < \operatorname{inj}(F_{\vec{w}}(t))$, there exist a constant $c_{r,j} > 0$ and an integer $N_j > 0$ independent of x such that we have the following pointwise bound for any j-th derivative*:

$$|\nabla^j \phi_t(x)| \le c_{r,j} \sum_{\ell=0}^{N_j} \|\Delta^{\ell}_{X_g(t)} \phi_t\|_{L^2(B_r(x))}. \tag{12}$$

Proof. This equality was shown in [Buser et al. 1988, Theorem 2.1]. The inequality is from the combination of the Sobolev and Gårding inequalities, for example, see [Bers et al. 1964]. □

With the above inequality we have the following uniform bound on $\{\phi_{t_i}\}_{i=1}^{\infty}$ and their derivatives.

Lemma 3.3. *For any $j \in \mathbb{N}$, we have that $\{\nabla^j\phi_{t_i}\}_i$ is uniformly bounded on any compact set of $X_g(0)$.*

Proof. Using (12) in the previous lemma, $\Delta\phi_t = \lambda_k(t)\phi_t$ and $0 < \lambda_k(t) < \frac{1}{3}$, we have

$$|\nabla^j \phi_t(x)| \le c_{r,j} \sum_{\ell=0}^{\infty} \left(\frac{1}{3}\right)^{\ell} \|\phi_t\|_{L^2(X_g(t))} \le 2c_{r,j},$$

where the bound is independent of x. Hence all derivatives of ϕ_t (in particular the sequence $\{\phi_{t_i}\}$) are uniformly bounded. □

Lemma 3.4. *There exists a subsequence of ϕ_{t_i} (denoted by ϕ_i) and $\phi_0 \in H^1(X_g(0))$ such that any derivatives satisfy*

$$\nabla^j \phi_i \to \nabla^j \phi_0$$

uniformly on connected compact set of $X_g(0)$.

Proof. Viewing $\{\phi_t\}$ as functions on F_0 where F_0 is any connected compact set of $X_g(0)$, by the previous lemma we have uniform boundedness of ϕ_t and all their derivatives. Hence by the Arzelà–Ascoli diagonal argument there exists a subsequence ϕ_i such that the function and its derivative converge uniformly on any compact set. □

By the convergence above we have

$$\int_{X_g(0)} |\phi_0|^2 \le 1, \quad \int_{X_g(0)} |\nabla\phi_0|^2 \le 1$$

and

$$\Delta_{X_g(0)}\phi_0 = \lambda_k(0)\cdot\phi_0.$$

Now we show the following statement regarding the limit $(\lambda_k(0), \phi_0)$. The argument is similar to [Wu and Xue 2022a, Lemma 9] and [Dodziuk et al. 1987, Lemma 3.3].

Proposition 3.5. *The limit $(\lambda_k(0), \phi_0)$ must satisfy one of the following conditions*:

(1) *ϕ_0 is an eigenfunction of $\Delta_{X_g(0)}$ and also restricts to at least one of the components Y_k as an eigenfunction*; *or*

(2) *$\phi_0 = 0$ everywhere on $X_g(0)$ and $\lambda_k(0) = \frac{1}{4}$.*

Proof. If ϕ_0 is not zero everywhere, then ϕ_0 belongs to $H^1(X_g(0))$ and is an eigenfunction. In particular, it must restrict to a nonzero function on at least one component of $X_g(0)$.

Otherwise suppose $\phi_0 = 0$ everywhere on $X_g(0)$, that is, $\phi_i \to 0$ pointwise everywhere. Then following a similar argument as in [Wu and Xue 2022a, Lemma 9] or [Dodziuk et al. 1987, Lemma 3.3], we can show that $\lambda_k(0) \ge \frac{1}{4}$. For completeness we write out the proof in detail here.

Recall the definitions of collars and shells on hyperbolic necks in (10) and (11). Similar to the definition above, we denote by $C_{\vec{w}}(i)$ the union of $\vec{w}$-wide collar neighborhoods near all degenerating geodesic circles on $X_g(t_i)$, and by $S_{\vec{w}}(i)$ the union of the "shells". To simplify the argument below, we also denote by $C_{i,m}$ and $S_{i,m}$ the individual hyperbolic neck and shell, respectively, with central geodesic circle $\gamma_m(i)$, where $1 \le m \le N$, and denote the corresponding width by $w_{i,m} := w(\gamma_m(i)) - 1$. Hence

$$C_{\vec{w}}(i) = \bigcup_{m=1}^{N} C_{i,m} \quad \text{and} \quad S_{\vec{w}}(i) = \bigcup_{m=1}^{N} S_{i,m}.$$

Fix any $\epsilon \in (0, 1)$ and $\delta \in \left(0, \frac{1}{16}\right)$. We write $c = 1 - \epsilon$. Since ϕ_i converges to zero uniformly on any compact set, there exists $N_0 \in \mathbb{N}$ such that for any $i > N_0$ we have

$$\int_{C_{\vec{w}}(i)} |\phi_i|^2 \ge c > 0, \quad \int_{S_{\vec{w}}(i)} |\phi_i|^2 < \delta c \quad \text{and} \quad \int_{S_{\vec{w}}(i)} |\nabla\phi_i|^2 < \delta c.$$

Define a new function on $C_{\vec{w}}(i) \cup S_{\vec{w}}(i)$ as follows:

$$\Phi_i := \begin{cases} \phi_i, & |\rho| \leq w_{i,m}, \\ (w_{i,m}+1-|\rho|)\phi_i, & w_{i,m} \leq |\rho| \leq w_{i,m}+1. \end{cases}$$

Then Φ_i gives a function in $H_0^1(C_{\vec{w}}(i) \cup S_{\vec{w}}(i))$ with $\Phi_i|_{\partial(C_{\vec{w}}(i)\cup S_{\vec{w}}(i))} = 0$. Therefore by applying [Wu and Xue 2022a, Lemma 7] to a union of hyperbolic collars we have

$$\int_{C_{\vec{w}}(i)\cup S_{\vec{w}}(i)} |\nabla \Phi_i|^2 > \frac{1}{4} \int_{C_{\vec{w}}(i)\cup S_{\vec{w}}(i)} |\Phi_i|^2.$$

On the other hand we have

$$\begin{aligned}
\int_{S_{\vec{w}}(i)} |\nabla \Phi_i|^2 &= \sum_{m=1}^{N} \int_{S_{i,m}} |\nabla((w_{i,m}+1-|\rho|)\phi_i)|^2 \\
&= \sum_{m=1}^{N} \int_{S_{i,m}} |\nabla(w_{i,m}+1-|\rho|)\cdot\phi_i + (w_{i,m}+1-|\rho|)\cdot\nabla\phi_i|^2 \\
&\leq \sum_{m=1}^{N} \int_{S_{i,m}} (|\phi_i| + (w_{i,m}+1-|\rho|)\cdot|\nabla\phi_i|)^2 \leq 2\sum_{m=1}^{N}\int_{S_{i,m}} |\phi_i|^2 + 2\sum_{m=1}^{N}\int_{S_{i,m}} |\nabla\phi_i|^2 \leq 4\delta c.
\end{aligned}$$

Therefore for any $i > N_0$ we have

$$\begin{aligned}
\int_{C_{\vec{w}}(i)} |\nabla\phi_i|^2 = \int_{C_{\vec{w}}(i)} |\nabla\Phi_i|^2 &= \int_{C_{\vec{w}}(i)\cup S_{\vec{w}}(i)} |\nabla\Phi_i|^2 - \int_{S_{\vec{w}}(i)} |\nabla\Phi_i|^2 \\
&\geq \frac{1}{4}\int_{C_{\vec{w}}(i)\cup S_{\vec{w}}(i)} |\Phi_i|^2 - \int_{S_{\vec{w}}(i)} |\nabla\Phi_i|^2 \\
&\geq \frac{1}{4}\int_{C_{\vec{w}}(i)} |\phi_i|^2 - \int_{S_{\vec{w}}(i)} |\nabla\Phi_i|^2 \geq \frac{1}{4}c - 4\delta c = \frac{1-16\delta}{4}(1-\epsilon),
\end{aligned}$$

which implies

$$\lambda_k(X_g(t_i)) = \frac{\int_{X_g(t_i)} |\nabla\phi_{t_i}|^2}{\int_{X_g(t_i)} |\phi_i|^2} \geq \frac{\int_{C_{\vec{w}}(i)} |\nabla\phi_i|^2}{\int_{X_g(t_i)} |\phi_i|^2} \geq \frac{1-16\delta}{4}(1-\epsilon).$$

Since this argument holds for any $\epsilon \in (0, 1)$ and $\delta \in \left(0, \frac{1}{16}\right)$, we have

$$\lambda_k(0) = \liminf_{i\to\infty} \lambda_k(X_g(t_i)) \geq \tfrac{1}{4}.$$

On the other hand $\lambda_k(0) \leq \frac{1}{4}$ by (9), therefore we have $\lambda_k(0) = \frac{1}{4}$. □

Now we are ready to prove Proposition 3.1.

Proof of Proposition 3.1. By the previous proposition, either $\lambda_k(0) = \lambda_1(Y_i)$ for at least one of the components Y_i, or $\lambda_k(0) = \frac{1}{4}$, therefore we obtain

$$\lambda_k(0) \geq \min_{1\leq i\leq k}\left\{\min\left\{\lambda_1(Y_i), \tfrac{1}{4}\right\}\right\}$$

as desired. □

We enclose in this section the following result, which is an easy application of Proposition 3.1.

Proposition 3.6. *Let $X_g(0) \in \partial\mathcal{M}_g$ be the limit of a family of Riemann surfaces $\{X_g(t)\} \subset \mathcal{M}_g$ by pinching certain simple closed geodesics such that $X_g(0)$ has k connected components, i.e., $X_g(0) = Y_1 \sqcup Y_2 \sqcup \cdots \sqcup Y_k$ for some $k \geq 2$. Assume in addition that $\bar{\lambda}_1(Y_i) = \min\{\lambda_1(Y_i), \frac{1}{4}\} \geq \frac{1}{4}$ for all $1 \leq i \leq k$, where $\lambda_1(Y_i)$ is the first nonzero eigenvalue of Y_i. Then*

$$\lim_{t\to 0} \lambda_k(X_g(t)) = \tfrac{1}{4}.$$

Proof. From (6) we have that

$$\limsup_{t\to 0} \lambda_k(X_g(t)) \leq \tfrac{1}{4}.$$

On the other hand, it follows by Proposition 3.1 that

$$\liminf_{t\to 0} \lambda_k(X_g(t)) \geq \min_{1\leq i\leq k} \left\{\min\left\{\lambda_1(Y_i), \tfrac{1}{4}\right\}\right\} = \tfrac{1}{4}.$$

The conclusion immediately follows. □

We now prove spectral gaps can be arbitrarily close to zero by using this result. Recall that, for all $i \geq 1$ and $X_g \in \mathcal{M}_g$, the i-th spectral gap $\mathrm{SpG}_i(X_g)$ of X is defined as

$$\mathrm{SpG}_i(X_g) := \lambda_i(X_g) - \lambda_{i-1}(X_g).$$

We prove the following.

Proposition 3.7. *For all $i \geq 1$,*

$$\inf_{X_g \in \mathcal{M}_g} \mathrm{SpG}_i(X_g) = 0.$$

Proof. We split the proof into three cases.

Case 1: $1 \leq i \leq 2g-3$. One may choose a closed hyperbolic surface $\mathcal{X}_g \in \mathcal{M}_g$ which is close enough to the maximal nodal surface

$$\underbrace{X_{0,3} \sqcup \cdots \sqcup X_{0,3}}_{2g-2 \text{ copies}} \in \partial\mathcal{M}_g,$$

then $\lambda_i(\mathcal{X}_g)$ is close to zero by Theorem 2.5. So the conclusion follows for this case.

Case 2: $i = 2g-2$. Let $Z_{1,2} \in \mathcal{M}_{1,2}$ such that $\bar{\lambda}_1(Z_{1,2}) = \min\left\{\frac{1}{4}, \lambda_1(Z_{1,2})\right\} \geq \frac{1}{4}$ by Theorem 2.7. Recall that $\lambda_1(X_{0,3}) \geq \frac{1}{4}$ from the same theorem. Let $\{X_g(t)\} \subset \mathcal{M}_g$ be a family of hyperbolic surfaces such that

$$\lim_{t\to 0} X_g(t) = \underbrace{X_{0,3} \sqcup \cdots \sqcup X_{0,3}}_{2g-4 \text{ copies}} \sqcup Z_{1,2} \in \partial\mathcal{M}_g.$$

Then it follows from Proposition 3.6 that

$$\lim_{t\to 0} \lambda_{2g-3}(X_g(t)) = \tfrac{1}{4}.$$

Meanwhile, by [Otal and Rosas 2009, Theorem 2], we know that

$$\lambda_{2g-2}(X_g(t)) \geq \tfrac{1}{4}.$$

Since $\mathrm{Diam}(X_g(t)) \to \infty$ as $t \to 0$, by [Cheng 1975, Corollary 2.3] we have that

$$\limsup_{t\to 0} \lambda_{2g-2}(X_g(t)) \le \tfrac{1}{4}.$$

Thus, we have

$$\lim_{t\to 0} \lambda_{2g-2}(X_g(t)) = \tfrac{1}{4}.$$

Then the conclusion also follows for this case because

$$\inf_{X_g\in\mathcal{M}_g} \mathrm{SpG}_{2g-2}(X_g) \le \lim_{t\to 0} \mathrm{SpG}_{2g-2}(X_g(t)) = 0.$$

Case 3: $i > 2g-2$. Let $\{Y_g(t)\} \subset \mathcal{M}_g$ be a family of hyperbolic surfaces such that

$$\lim_{t\to 0} Y_g(t) \in \partial\mathcal{M}_g.$$

Similar to Case 2, by [Otal and Rosas 2009, Theorem 2] and [Cheng 1975, Corollary 2.3], we have

$$\lim_{t\to 0} \lambda_i(Y_g(t)) = \tfrac{1}{4} \quad \text{and} \quad \lim_{t\to 0} \lambda_{i-1}(Y_g(t)) = \tfrac{1}{4}.$$

This implies $\inf_{X_g\in\mathcal{M}_g} \mathrm{SpG}_i(X_g) = 0$ for all $i > 2g-2$. □

4. Proof of Theorem 4.1

Now we are ready to prove Theorem 4.1.

Theorem 4.1. *Let $\{\eta(g)\}_{g=2}^{\infty}$ be any sequence of integers with $\eta(g) \in [1, 2g-2]$. Then*

$$\liminf_{g\to\infty} \sup_{X_g\in\mathcal{M}_g} \mathrm{SpG}_{\eta(g)}(X_g) \ge \tfrac{1}{4}.$$

Proof. We will show that for any $\eta(g)$ with sufficiently large g, one can find a genus g surface X_g with $\mathrm{SpG}_{\eta(g)}(X_g)$ close to $\frac{1}{4}$. To see this, we split the proof into the following four cases.

Case 1: $\eta(g) = 2g-2$. Let $X_g(t) : (0,1) \to \mathcal{M}_g$ be a family of closed hyperbolic surfaces such that

$$\lim_{t\to 0} X_g(t) = \underbrace{X_{0,3} \sqcup \cdots \sqcup X_{0,3}}_{2g-2 \text{ copies}} \in \partial\mathcal{M}_g.$$

First by [Otal and Rosas 2009, Theorem 2], $\lambda_{2g-2}(X_g(t)) \ge \frac{1}{4}$ for all $t \in (0,1)$. Secondly by Theorem 2.5 we know that $\lambda_{2g-3}(X_g(t)) \to 0$ as $t \to 0$. Thus,

$$\sup_{X_g\in\mathcal{M}_g} \mathrm{SpG}_{2g-2}(X_g) \ge \liminf_{t\to 0} \mathrm{SpG}_{2g-2}(X_g(t)) \ge \tfrac{1}{4}.$$

Case 2: $\eta(g) \in [g+1, 2g-3]$. First we choose a hyperbolic surface $Z_{1,2} \in \mathcal{M}_{1,2}$ such that $\bar{\lambda}_1(Z_{1,2}) \ge \frac{1}{4}$ by Theorem 2.7. Recall also that $\lambda_1(X_{0,3}) \ge \frac{1}{4}$. By Lemma 2.3 we can construct $X_g(t) : (0,1) \to \mathcal{M}_g$ as a family of closed hyperbolic surfaces such that

$$\lim_{t\to 0} X_g(t) = \underbrace{X_{0,3} \sqcup \cdots \sqcup X_{0,3}}_{i \text{ copies}} \sqcup \underbrace{Z_{1,2} \sqcup \cdots \sqcup Z_{1,2}}_{j \text{ copies}} \in \partial\mathcal{M}_g,$$

where i and j are two nonnegative integers satisfying $i+j=\eta(g)$. By Theorem 2.5 we know that $\lim_{t\to 0}\lambda_{\eta(g)-1}(X_g(t))=0$. By Proposition 3.6 we have

$$\lim_{t\to 0}\lambda_{\eta(g)}(X_g(t))=\tfrac{1}{4},$$

which implies

$$\sup_{X_g\in\mathcal{M}_g}\mathrm{SpG}_{\eta(g)}(X_g)\geq\lim_{t\to 0}\mathrm{SpG}_{\eta(g)}(X_g(t))=\tfrac{1}{4}.$$

<u>Case 3</u>: $\eta(g)\in[2,g]$. As in Case 2, we choose a hyperbolic surface $Z_{1,2}\in\mathcal{M}_{1,2}$ such that $\bar{\lambda}_1(Z_{1,2})\geq\frac{1}{4}$. Let $g_1>0$ be the integer determined in Lemma 2.4. Note that g_1 tends to ∞ as $g\to\infty$ because $2g_1\geq g-2$. Then by Theorem 2.8 we know that, for any $\epsilon>0$ and large enough $g>0$, one may choose a hyperbolic surface $\mathcal{X}_{g_1,2}\in\mathcal{M}_{g_1,2}$ such that

$$\bar{\lambda}_1(\mathcal{X}_{g_1,2})>\tfrac{1}{4}-\epsilon.$$

Fix any such large g. Then by Lemma 2.4 we construct $X_g(t):(0,1)\to\mathcal{M}_g$ as a family of closed hyperbolic surfaces such that

$$\lim_{t\to 0}X_g(t)=\underbrace{X_{0,3}\sqcup\cdots\sqcup X_{0,3}}_{i\text{ copies}}\sqcup\underbrace{Z_{1,2}\sqcup\cdots\times Z_{1,2}}_{j\text{ copies}}\sqcup\mathcal{X}_{g_1,2}\in\partial\mathcal{M}_g,$$

where i and j are two nonnegative integers satisfying $i+j=\eta(g)-1$. By Theorem 2.5 we know that $\lim_{t\to 0}\lambda_{\eta(g)-1}(X_g(t))=0$. Applying the min-max principle in Proposition 3.1 to this sequence with $k=\eta(g)$ (note that g is a fixed large integer hence $\eta(g)$ is also fixed), we have

$$\liminf_{t\to 0}\lambda_{\eta(g)}(X_g(t))\geq\min\{\bar{\lambda}_1(\mathcal{M}_{0,3}),\ \bar{\lambda}_1(Z_{1,2}),\ \bar{\lambda}_1(\mathcal{X}_{g_1,2})\}\geq\tfrac{1}{4}-\epsilon,$$

which implies

$$\liminf_{g\to\infty}\sup_{X_g\in\mathcal{M}_g}\mathrm{SpG}_{\eta(g)}(X_g)\geq\tfrac{1}{4}-\epsilon$$

because

$$\sup_{X_g\in\mathcal{M}_g}\mathrm{SpG}_{\eta(g)}(X_g)\geq\liminf_{t\to 0}\mathrm{SpG}_{\eta(g)}(X_g(t)).$$

Since $\epsilon>0$ can be arbitrarily small, we have

$$\liminf_{g\to\infty}\sup_{X_g\in\mathcal{M}_g}\mathrm{SpG}_{\eta(g)}(X_g)\geq\tfrac{1}{4}.$$

<u>Case 4</u>: $\eta(g)=1$. This is due to [Hide and Magee 2023, Corollary 1.3] because $\mathrm{SpG}_1(X_g)=\lambda_1(X_g)$.

The four cases above cover all possible $\eta(g)$ and hence complete the proof. □

Remark. The method in this paper works for indices in the range of $[1,2g-2]$ in Theorem 4.1. The restriction comes from the lack of suitable components with λ_1 close to $\frac{1}{4}$ when constructing the degenerating family. It would be interesting to know whether the assumption $\eta(g)\in[1,2g-2]$ can be dropped.

We also note that, together with [Cheng 1975, Corollary 2.3], the proof of Theorem 4.1 above actually gives the following result.

Theorem 4.2. *For any* $0 \le j < i$ *with* $i = o(\ln(g))$,

$$\lim_{g\to\infty} \sup_{X_g \in \mathcal{M}_g} (\lambda_i(X_g) - \lambda_j(X_g)) = \tfrac{1}{4}.$$

Acknowledgements

The authors would like to thank the referees for their careful reading and valuable comments which improved the quality of this paper. Wu and Zhang would like to thank Yuxin He, Yang Shen and Yuhao Xue for helpful discussions on this project. Wu is partially supported by NSFC grants 12171263 and 12361141813. Zhu is supported by NSF grants DMS-2041823 and DMS-2305363.

References

[Ballmann et al. 2016] W. Ballmann, H. Matthiesen, and S. Mondal, "Small eigenvalues of closed surfaces", *J. Differential Geom.* **103**:1 (2016), 1–13. MR Zbl

[Ballmann et al. 2017] W. Ballmann, H. Matthiesen, and S. Mondal, "On the analytic systole of Riemannian surfaces of finite type", *Geom. Funct. Anal.* **27**:5 (2017), 1070–1105. MR Zbl

[Ballmann et al. 2018] W. Ballmann, H. Matthiesen, and S. Mondal, "Small eigenvalues of surfaces: old and new", *ICCM Not.* **6**:2 (2018), 9–24. MR Zbl

[Bers et al. 1964] L. Bers, F. John, and M. Schechter, *Partial differential equations*, Lectures in Appl. Math. **3**, Interscience, New York, 1964. MR Zbl

[Brooks and Makover 2001] R. Brooks and E. Makover, "Riemann surfaces with large first eigenvalue", *J. Anal. Math.* **83** (2001), 243–258. MR Zbl

[Brooks and Makover 2004] R. Brooks and E. Makover, "Random construction of Riemann surfaces", *J. Differential Geom.* **68**:1 (2004), 121–157. MR Zbl

[Buser 1982] P. Buser, "A note on the isoperimetric constant", *Ann. Sci. École Norm. Sup.* (4) **15**:2 (1982), 213–230. MR Zbl

[Buser 1984] P. Buser, "On the bipartition of graphs", *Discrete Appl. Math.* **9**:1 (1984), 105–109. MR Zbl

[Buser 1992] P. Buser, *Geometry and spectra of compact Riemann surfaces*, Progr. Math. **106**, Birkhäuser, Boston, 1992. MR Zbl

[Buser et al. 1988] P. Buser, M. Burger, and J. Dodziuk, "Riemann surfaces of large genus and large λ_1", pp. 54–63 in *Geometry and analysis on manifolds* (Katata/Kyoto, 1987), edited by T. Sunada, Lecture Notes in Math. **1339**, Springer, 1988. MR Zbl

[Cheng 1975] S. Y. Cheng, "Eigenvalue comparison theorems and its geometric applications", *Math. Z.* **143**:3 (1975), 289–297. MR Zbl

[Dodziuk and Randol 1986] J. Dodziuk and B. Randol, "Lower bounds for λ_1 on a finite-volume hyperbolic manifold", *J. Differential Geom.* **24**:1 (1986), 133–139. MR Zbl

[Dodziuk et al. 1987] J. Dodziuk, T. Pignataro, B. Randol, and D. Sullivan, "Estimating small eigenvalues of Riemann surfaces", pp. 93–121 in *The legacy of Sonya Kovalevskaya* (Cambridge/Amherst, MA, 1985), edited by L. Keen, Contemp. Math. **64**, Amer. Math. Soc., Providence, RI, 1987. MR Zbl

[Gelbart and Jacquet 1978] S. Gelbart and H. Jacquet, "A relation between automorphic representations of GL(2) and GL(3)", *Ann. Sci. École Norm. Sup.* (4) **11**:4 (1978), 471–542. MR Zbl

[Hejhal 1990] D. A. Hejhal, *Regular b-groups, degenerating Riemann surfaces, and spectral theory*, Mem. Amer. Math. Soc. **437**, Amer. Math. Soc., Providence, RI, 1990. MR Zbl

[Hide 2022] W. Hide, "Spectral gap for Weil–Petersson random surfaces with cusps", *Int. Math. Res. Not.* **2022** (online publication October 2022). Zbl

[Hide and Magee 2023] W. Hide and M. Magee, "Near optimal spectral gaps for hyperbolic surfaces", *Ann. of Math.* (2) **198**:2 (2023), 791–824. MR Zbl

[Ji 1993] L. Ji, "Spectral degeneration of hyperbolic Riemann surfaces", *J. Differential Geom.* **38**:2 (1993), 263–313. MR Zbl

[Ji and Zworski 1993] L. Ji and M. Zworski, "The remainder estimate in spectral accumulation for degenerating hyperbolic surfaces", *J. Funct. Anal.* **114**:2 (1993), 412–420. MR Zbl

[Kim 2003] H. H. Kim, "Functoriality for the exterior square of GL_4 and the symmetric fourth of GL_2", *J. Amer. Math. Soc.* **16**:1 (2003), 139–183. MR Zbl

[Lax and Phillips 1982] P. D. Lax and R. S. Phillips, "The asymptotic distribution of lattice points in Euclidean and non-Euclidean spaces", *J. Funct. Anal.* **46**:3 (1982), 280–350. MR Zbl

[Lipnowski and Wright 2024] M. Lipnowski and A. Wright, "Towards optimal spectral gaps in large genus", *Ann. Probab.* **52**:2 (2024), 545–575. MR Zbl

[Luo et al. 1995] W. Luo, Z. Rudnick, and P. Sarnak, "On Selberg's eigenvalue conjecture", *Geom. Funct. Anal.* **5**:2 (1995), 387–401. MR Zbl

[Magee and Naud 2020] M. Magee and F. Naud, "Explicit spectral gaps for random covers of Riemann surfaces", *Publ. Math. Inst. Hautes Études Sci.* **132** (2020), 137–179. MR Zbl

[Magee and Naud 2021] M. Magee and F. Naud, "Extension of Alon's and Friedman's conjectures to Schottky surfaces", preprint, 2021. Zbl arXiv 2106.02555

[Magee and Puder 2023] M. Magee and D. Puder, "The asymptotic statistics of random covering surfaces", *Forum Math. Pi* **11** (2023), art. id. E15. Zbl

[Magee et al. 2022] M. Magee, F. Naud, and D. Puder, "A random cover of a compact hyperbolic surface has relative spectral gap $\frac{3}{16} - \varepsilon$", *Geom. Funct. Anal.* **32**:3 (2022), 595–661. MR Zbl

[Masur 1976] H. Masur, "Extension of the Weil–Petersson metric to the boundary of Teichmuller space", *Duke Math. J.* **43**:3 (1976), 623–635. MR Zbl

[McKean 1972] H. P. McKean, "Selberg's trace formula as applied to a compact Riemann surface", *Comm. Pure Appl. Math.* **25** (1972), 225–246. Correction in **27**:1 (1974), 134. MR Zbl

[Melrose and Zhu 2018] R. Melrose and X. Zhu, "Resolution of the canonical fiber metrics for a Lefschetz fibration", *J. Differential Geom.* **108**:2 (2018), 295–317. MR Zbl

[Melrose and Zhu 2019] R. Melrose and X. Zhu, "Boundary behaviour of Weil–Petersson and fibre metrics for Riemann moduli spaces", *Int. Math. Res. Not.* **2019**:16 (2019), 5012–5065. MR Zbl

[Mirzakhani 2013] M. Mirzakhani, "Growth of Weil–Petersson volumes and random hyperbolic surfaces of large genus", *J. Differential Geom.* **94**:2 (2013), 267–300. MR Zbl

[Mondal 2015] S. Mondal, "On largeness and multiplicity of the first eigenvalue of finite area hyperbolic surfaces", *Math. Z.* **281**:1-2 (2015), 333–348. MR Zbl

[Monk 2021] L. Monk, *Geometry and spectrum of typical hyperbolic surfaces*, Ph.D. thesis, Université de Strasbourg, 2021, available at https://lauramonk.github.io/thesis.pdf. Zbl

[Otal and Rosas 2009] J.-P. Otal and E. Rosas, "Pour toute surface hyperbolique de genre g, $\lambda_{2g-2} > 1/4$", *Duke Math. J.* **150**:1 (2009), 101–115. MR Zbl

[Randol 1974] B. Randol, "Small eigenvalues of the Laplace operator on compact Riemann surfaces", *Bull. Amer. Math. Soc.* **80** (1974), 996–1000. MR Zbl

[Schoen et al. 1980] R. Schoen, S. Wolpert, and S. T. Yau, "Geometric bounds on the low eigenvalues of a compact surface", pp. 279–285 in *Geometry of the Laplace operator* (Honolulu, HI, 1979), edited by R. Osserman and A. Weinstein, Proc. Sympos. Pure Math. **36**, Amer. Math. Soc., Providence, RI, 1980. MR Zbl

[Selberg 1965] A. Selberg, "On the estimation of Fourier coefficients of modular forms", pp. 1–15 in *Theory of numbers* (Pasadena, CA, 1963), edited by A. L. Whiteman, Proc. Sympos. Pure Math. **8**, Amer. Math. Soc., Providence, RI, 1965. MR Zbl

[Wolpert 1987] S. A. Wolpert, "Asymptotics of the spectrum and the Selberg zeta function on the space of Riemann surfaces", *Comm. Math. Phys.* **112**:2 (1987), 283–315. MR Zbl

[Wolpert 1990] S. A. Wolpert, "The hyperbolic metric and the geometry of the universal curve", *J. Differential Geom.* **31**:2 (1990), 417–472. MR Zbl

[Wolpert 1992a] S. A. Wolpert, "Spectral limits for hyperbolic surfaces, I", *Invent. Math.* **108**:1 (1992), 67–89. MR Zbl

[Wolpert 1992b] S. A. Wolpert, "Spectral limits for hyperbolic surfaces, II", *Invent. Math.* **108**:1 (1992), 91–129. MR Zbl

[Wu and Xue 2022a] Y. Wu and Y. Xue, "Optimal lower bounds for first eigenvalues of Riemann surfaces for large genus", *Amer. J. Math.* **144**:4 (2022), 1087–1114. MR Zbl

[Wu and Xue 2022b] Y. Wu and Y. Xue, "Random hyperbolic surfaces of large genus have first eigenvalues greater than $\frac{3}{16}-\epsilon$", *Geom. Funct. Anal.* **32**:2 (2022), 340–410. MR Zbl

[Wu and Xue 2022c] Y. Wu and Y. Xue, "Small eigenvalues of closed Riemann surfaces for large genus", *Trans. Amer. Math. Soc.* **375**:5 (2022), 3641–3663. MR Zbl

Received 25 Jan 2022. Revised 29 Jul 2022. Accepted 23 Sep 2022.

YUNHUI WU: yunhui_wu@tsinghua.edu.cn
Yau Mathematical Sciences Center, Tsinghua University, Beijing, China

HAOHAO ZHANG: zhh21@mails.tsinghua.edu.cn
Tsinghua University, Beijing, China

XUWEN ZHU: x.zhu@northeastern.edu
Department of Mathematics, Northeastern University, Boston, MA, United States

ANALYSIS AND PDE
Vol. 17 (2024), No. 4, pp. 1397–1438
DOI: 10.2140/apde.2024.17.1397

PLATEAU FLOW OR THE HEAT FLOW FOR HALF-HARMONIC MAPS

MICHAEL STRUWE

Using the interpretation of the half-Laplacian on S^1 as the Dirichlet-to-Neumann operator for the Laplace equation on the ball B, we devise a classical approach to the heat flow for half-harmonic maps from S^1 to a closed target manifold $N \subset \mathbb{R}^n$, recently studied by Wettstein, and for arbitrary finite-energy data we obtain a result fully analogous to the author's 1985 results for the harmonic map heat flow of surfaces and in similar generality. When N is a smoothly embedded, oriented closed curve $\Gamma \subset \mathbb{R}^n$, the half-harmonic map heat flow may be viewed as an alternative gradient flow for a variant of the Plateau problem of disc-type minimal surfaces.

1. Background and results

Half-harmonic maps and their heat flow. Let $N \subset \mathbb{R}^n$ be a closed submanifold, that is, compact and without boundary. The concept of a half-harmonic map $u\colon S^1 \to N \subset \mathbb{R}^n$ was introduced by Da Lio and Rivière [2011], who together with Martinazzi [Da Lio et al. 2015, Theorem 2.9] also made the interesting observation that the harmonic extension of a half-harmonic map yields a free boundary minimal surface supported by N, a fact which also was noticed by Millot and Sire [2015, Remark 4.28].

In his Ph.D. thesis, Wettstein [2021; 2022; 2023], studied the corresponding heat flow given by

$$d\pi_N(u)(u_t + (-\Delta)^{1/2}u) = 0 \quad \text{on } S^1 \times [0, \infty[, \tag{1-1}$$

where $u_t = \partial_t u$ and where $\pi_N\colon N_\rho \to N$ is the smooth nearest-neighbor projection on a ρ-neighborhood N_ρ of the given target manifold to N, and, with the help of a fine analysis of the fractional differential operators involved, he showed global existence for initial data of small energy.

Moser [2011] and Millot and Sire [2015] contributed important results to the study of half-harmonic maps by exploiting the fact that for any smooth $u\colon S^1 \to \mathbb{R}^n$ we can represent the half-Laplacian classically in the form

$$(-\Delta)^{1/2}u = \partial_r U, \tag{1-2}$$

where $U\colon B \to \mathbb{R}^n$ is the harmonic extension of u to the unit disc B.[1] Here, using the identity (1-2), we are able to remove the smallness assumption in Wettstein's work and show the existence of a "global"

I thank Amélie Loher and the anonymous referee for careful reading of the manuscript and useful suggestions.
MSC2020: primary 35K55, 35K65, 53E99; secondary 53A10.
Keywords: half-harmonic maps, harmonic map heat flow, Plateau problem.

[1]The classical formula (1-2) is a special case of a much more general result due to Caffarelli and Silvestre [2007], who pointed out that many nonlocal problems involving fractional powers of the Laplacian can be related to a local, possibly degenerate, elliptic equation via a suitable extension of the solution to a half-space.

weak solution to the heat flow (1-1) for data of arbitrarily large (but finite) energy, which is defined for all times and smooth away from finitely many "blow-up points" where energy concentrates, and whose energy is nonincreasing. The solution is unique in this class in exact analogy with the classical result by the author [Struwe 1985] on the harmonic map heat flow for maps from a closed surface to a closed target manifold $N \subset \mathbb{R}^n$; see Theorem 1.2 below.

In order to describe our work in more detail, let

$$H^{1/2}(S^1; N) = \{u \in H^{1/2}(S^1; \mathbb{R}^n) : u(z) \in N \text{ for almost every } z \in S^1\}.$$

Interpreting $S^1 = \partial B$, where $B = B_1(0; \mathbb{R}^2)$, and tacitly identifying a map $u \in H^{1/2}(S^1; N)$ with its harmonic extension $U \in H^1(B; \mathbb{R}^n)$, for a given function $u_0 \in H^{1/2}(S^1; N)$ we then seek to find a family of harmonic functions $u(t) \in H^1(B; \mathbb{R}^n)$ with traces $u(t) \in H^{1/2}(S^1; N)$ for $t > 0$, solving the equation

$$d\pi_N(u)(u_t + \partial_r u) = u_t + d\pi_N(u)\partial_r u = 0 \quad \text{on } S^1 \times [0, \infty[, \tag{1-3}$$

with initial data

$$u|_{t=0} = u_0 \in H^{1/2}(S^1; N). \tag{1-4}$$

Energy. The half-harmonic heat flow may be regarded as the heat flow for the half-energy

$$E_{1/2}(u) = \frac{1}{2}\int_{S^1} |(-\Delta)^{1/4}u|^2 \, d\phi$$

of a map $u \in H^{1/2}(S^1; N)$. Note that the half-energy of u equals the standard Dirichlet energy

$$E(u) = \frac{1}{2}\int_B |\nabla u|^2 \, dz$$

of its harmonic extension $u \in H^1(B; \mathbb{R}^n)$. Indeed, integrating by parts we have

$$\int_B |\nabla u|^2 \, dz = \int_{S^1} u\partial_r u \, d\phi = \int_{S^1} u(-\Delta)^{1/2}u \, d\phi = \int_{S^1} |(-\Delta)^{1/4}u|^2 \, d\phi, \tag{1-5}$$

where we use the identity (1-2) and where the last identity easily follows from the representation of the operators $(-\Delta)^{1/2}$ and $(-\Delta)^{1/4}$ in Fourier space with symbols $|\xi|$ and $\sqrt{|\xi|}$, respectively, and Parceval's identity.[2] Therefore, in the following for convenience we may always work with the classically defined Dirichlet energy. Moreover, we may interpret the half-harmonic heat flow as the heat flow for the Dirichlet energy in the class of harmonic functions with trace in $H^{1/2}(S^1; N)$; see Section 2 below for details.

Results. Identifying $\mathbb{R}^2$ with $\mathbb{C}$, we denote by M the three dimensional Möbius group of conformal transformations of the unit disc, given by

$$M = \left\{\Phi(z) = e^{i\theta}\frac{z+a}{1+\bar{a}z} \in C^\infty(\bar{B}; \bar{B}) : |a| < 1,\ \theta \in \mathbb{R}\right\}.$$

Observe that the Dirichlet energy is invariant under conformal transformations, and we have $E(u) = E(u \circ \Phi)$ for any $u \in H^1(B; \mathbb{R}^n)$ and any $\Phi \in M$.

[2]Conversely, via Fourier expansion we also can prove (1-5) directly. Computing the first variations of E and $E_{1/2}$, respectively, we then obtain (1-2).

For smooth data we then have the following result.

Theorem 1.1. *Let $N \subset \mathbb{R}^n$ be a closed, smooth submanifold of $\mathbb{R}^n$, and suppose that the normal bundle $T^{\perp}N$ is parallelizable. Then the following holds*:

(i) *For any smooth $u_0 \in H^{1/2}(S^1; N)$ there exists a time $T_0 \le \infty$ and a unique smooth solution $u = u(t)$ of (1-3), and hence of (1-1), with data (1-4) for $0 < t < T_0$.*

(ii) *If $T_0 < \infty$, we have concentration in the sense that, for some $\delta > 0$ and any $R > 0$,*

$$\sup_{\substack{z_0 \in B \\ 0<t<T_0}} \int_{B_R(z_0)\cap B} |\nabla u(t)|^2\, dz \ge \delta,$$

and for suitable $t_k \uparrow T_0$ there exist finitely many points $z_k^{(1)}, \dots, z_k^{(i_0)}$ and conformal maps $\Phi_k^{(i)} \in M$ with $z_k^{(i)} \to z^{(i)} \in \bar{B}$ and $\Phi_k^{(i)} \to \Phi_\infty^{(i)} \equiv z^{(i)}$ weakly in $H^1(B)$ such that $u(t_k)\circ\Phi_k^{(i)} \to \bar{u}^{(i)}$ weakly in $H^1(B)$ as $k \to \infty$, where $\bar{u}^{(i)}$ is nonconstant and conformal and satisfies

$$d\pi_N(\bar{u}^{(i)})\partial_r \bar{u}^{(i)} = 0, \quad 1 \le i \le i_0. \tag{1-6}$$

Moreover, there exists $\delta = \delta(N) > 0$ such that $E(\bar{u}^{(i)}) \ge \delta$, and $i_0 \le E(u_0)/\delta$. Finally, $u(t_k)$ smoothly converges to a limit $u_1 \in H^{1/2}(S^1; N)$ on $\bar{B} \setminus \{z^{(1)}, \dots, z^{(i_0)}\}$.

(iii) *If $T_0 = \infty$, then, as $t \to \infty$ suitably, $u(t)$ smoothly converges to a half-harmonic limit map away from at most finitely many concentration points where nonconstant half-harmonic maps "bubble off" as in (ii).*

By the Da Lio–Rivière interpretation of (1-6), the "bubbles" $\bar{u}^{(i)}$ as well as the limit u_∞ of the flow conformally parametrize minimal surfaces with free boundary on N, meeting N orthogonally along their free boundaries.

The hypothesis regarding the target manifold N in particular is fulfilled if N is a closed, orientable hypersurface of codimension 1 in $\mathbb{R}^n$, or if N is a smoothly embedded, closed curve $\Gamma \subset \mathbb{R}^3$.

It would be interesting to find examples of initial data for which the flow blows up in finite time, as in the work of Chang, Ding, and Ye [Chang et al. 1992] on the harmonic map heat flow.

For data in $H^{1/2}(S^1; N)$ the following global existence result holds, which is our main result.

Theorem 1.2. *For $N \subset \mathbb{R}^n$ as in Theorem 1.1 the following holds*:

(i) *For any $u_0 \in H^{1/2}(S^1; N)$ there exists a unique global weak solution of (1-3) with data (1-4) as in Definition 6.3, whose energy is nonincreasing and which is smooth for positive time away from finitely many points in space-time where nontrivial half-harmonic maps "bubble off" in the sense of Theorem 1.1(ii).*

(ii) *As $t \to \infty$ suitably, $u(t)$ smoothly converges to a half-harmonic limit map away from at most finitely many concentration points where nonconstant half-harmonic maps "bubble off" as in Theorem 1.1(iii).*

Note that uniqueness is only asserted within the class of partially regular weak solutions with nonincreasing energy, as in the case of the harmonic map heat flow. It would be interesting to find out if the latter condition suffices, as in the work of Freire [1995], and, conversely, to explore the possibility of "backward bubbling" in (1-3), as in the examples of Topping [2002] for the latter flow.

Key features of the proof and related flow equations. In our approach, in a similar vein as [Lenzmann and Schikorra 2020], we uncover and exploit surprising regularity properties of the normal component $d\pi_N^\perp(u)\partial_r u$ for the harmonic extension of u, likely related to the fractional commutator estimates for the normal projection in the work of Da Lio and Rivière [2011] or the regularity estimates of Da Lio and Pigati [2020], Mazowiecka and Schikorra [2018], and others.

The use of the Dirichlet-to-Neumann map for the harmonic extension $u\colon B \to \mathbb{R}^n$ of u instead of the half-Laplacian, and the simple identity (3-2) as well as (3-5) allow us to perform the analysis using only local, classically defined operators, avoiding fractional calculus almost entirely.

Note that (1-3) is similar to the equation governing the (scalar) evolution problem for conformal metrics $e^{2u} g_{\mathbb{R}^2}$ of prescribed geodesic boundary curvature and vanishing Gauss curvature on the unit disc B, studied for instance by Brendle [2002] or Gehrig [2020]. In contrast to the latter flows, due to the presence of the projection operator mapping u_r to its tangent component, the flow (1-3) at first sight appears to be degenerate. However, surprisingly, within our framework we are able to obtain similar smoothing properties as in the case of the harmonic map heat flow of surfaces.

A different heat flow associated with half-harmonic maps, using the half-heat operator $(\partial_t - \Delta)^{1/2}$ instead of (1-1), was suggested by Hyder et al. [2022], and they obtained global existence of partially regular, but possibly nonunique, weak solutions for their flow, with a possibly large singular set of measure zero.

Applications to the Plateau problem. In the case when N is a smoothly embedded, oriented closed curve $\Gamma \subset \mathbb{R}^3$, the half-harmonic heat flow (1-3) may furnish an alternative gradient flow for the Plateau problem of minimal surfaces of the type of the disc, which has a long and famous tradition in geometric analysis.

Posed in the 1890's, Plateau's problem was finally solved independently by Douglas [1931] and Radó [1930]. In order to analyze the set of *all* minimal surfaces solving the Plateau problem, including saddle points of the Dirichlet integral, thereby building on Douglas' ideas, Morse and Tompkins [1939] proposed a critical point theory for Plateau's problem in the sense of [Morse 1937], attempting to characterize nonminimizing solutions as "homotopy-critical" points of Dirichlet's integral. However, Tromba [1984; 1985] pointed out that it was not even clear that all smooth, nondegenerate minimal surfaces would be "homotopy-critical" in the sense of [Morse and Tompkins 1939]. To overcome this problem, Tromba developed a version of degree theory that could be applied in this case and which yielded at least a proof of the "last" Morse inequality, which is an identity for the total degree.

Finally, this author [Struwe 1984] recast the Plateau problem as a variational problem on a closed convex set and was able to develop a version of the Palais–Smale type critical point theory for the problem within this frame-work, which allowed him to obtain all Morse inequalities in a rigorous fashion; see [Struwe 1988] and [Imbusch and Struwe 1999] for further details. In [Struwe 1986] and [Jost and Struwe 1990], the approach was extended to the case of multiple boundaries and/or higher genus.

A key element of critical point theory for a variational problem is the construction of a pseudogradient flow for the problem at hand. In [Struwe 1984] this was achieved in an ad-hoc way. However, starting with the work of Eells and Sampson [1964] on the harmonic map heat flow, it is now an established approach in geometric analysis to study the (negative) (L^2-)gradient flow related to a variational problem, similar to the standard heat equation. For Plateau's problem, such a flow was obtained by Chang and Liu [2005] within the frame-work laid out by Struwe [1984] in the form of a parabolic variational inequality, for which Chang and Liu obtained a solution of class H^2 by means of a time-discrete minimization scheme. Rupflin [2017] and Rupflin and Schrecker [2018] studied the analogous parabolic variational inequality in the case of an annulus, which again had previously been studied in [Struwe 1986] by means of an ad-hoc pseudogradient flow.

In view of the much better regularity properties of the flow equation (1-3) it would be tempting to regard this as the correct definition of the canonical gradient flow for the Plateau problem, but an important issue still needs to be addressed.

Monotonicity. Recall that in the classical Plateau problem $u(t)$ is required to induce a (weakly) monotone parametrization of Γ for each $t > 0$. Even though it may seem likely that — at least for curves Γ on the boundary of a convex body in $\mathbb{R}^3$ — this Plateau boundary condition will be preserved along the flow (1-3) whenever it is satisfied initially, at this moment even for a strictly convex planar curve $\Gamma \subset \mathbb{R}^2$ it is not clear whether this actually happens. However, the results that we obtain also seem to be of interest if we drop the Plateau condition. In particular, our results motivate the study of smooth minimal surfaces with continuous trace covering only a part of the given boundary curve Γ; dropping the monotonicity condition also brings the parametric approach to the Plateau problem closer to the approach via geometric measure theory or level sets.

Plateau flow. It should be straightforward to extend our results to the case when the disc B is replaced by a surface Σ of higher genus with boundary $\partial\Sigma \cong S^1$, if for given initial data $u_0 \in H^{1/2}(S^1; N)$ we consider a family $u = u(t)$ in $H^{1/2}(S^1; N)$ solving (1-3), that is,

$$u_t + d\pi_N(u)\partial_\nu u = 0$$

instead of (1-1), where for each time we harmonically extend $u(t)$ to Σ and denote by $\partial_\nu u$ the outward normal derivative of u along $\partial\Sigma$, as was proposed and analyzed by Da Lio and Pigati [2020] in the time-independent case. Similarly, one might study the flow (1-3) on a domain Σ with multiple boundaries. Of course, in order for the flow to converge to a minimal surface in the case of higher genus or higher connectivity it will be necessary to couple the flow (1-3) with a corresponding evolution equation for the conformal structure on Σ, as in the work of Rupflin and Topping [2019] on minimal immersions. Note that on a general domain Σ the flow equations (1-1) and (1-3) no longer agree. In order to clearly distinguish the flow equation (1-3) from the equation (1-1) defining the half-harmonic map heat flow, we therefore propose to say that (1-3) defines the "Plateau flow".

Outline. After a brief discussion of energy estimates in Section 2, in Section 3 we present the analytic core of the argument for higher regularity in Section 4 and for the blow-up analysis, later presented in Section 8. These tools are also instrumental in proving uniqueness of partially regular weak solutions in Section 7. The L^2-bounds for higher and higher derivatives which we establish in Section 4, assuming that energy does not concentrate, may be of particular interest. These bounds either concern estimates for $\nabla\partial_\phi^k u$ on B or on ∂B, and we view the latter bounds as stronger by an order of $\frac{1}{2}$. These bounds may be iterated interlaced, as we later do in Section 6, to prove uniform smooth estimates, locally in time, for smooth flows with smooth initial data converging in $H^{1/2}(N; S^1)$. Since the latter data are dense in $H^{1/2}(N; S^1)$, we thus not only obtain existence of weak solutions for arbitrary data $u_0 \in H^{1/2}(N; S^1)$ but also can show their smoothness for positive time and hence are able to derive Theorem 1.2 from Theorem 1.1. A peculiar feature is that one set of regularity estimates can only be obtained globally, that is on all of B, whereas the other set of estimates may be localized using cut-off functions. Similar estimates for a regularized version of (1-3) are employed in Section 5 to prove local existence of smooth solutions of (1-3) for smooth data (1-4). Finally, in Section 9 the large-time behavior of smooth solutions to (1-3) is discussed, finishing the proof of Theorem 1.1.

Notation. The letter C is used throughout to denote a generic constant, possibly depending on the "target" N and the initial energy $E(u_0)$.

Moreover, since $T^\perp N$ by assumption is parallelizable and compact, there exists $\rho > 0$ such that the representation

$$T \colon N \times B_\rho(0; \mathbb{R}^m) \ni (p, y) \ \to \ p + \sum_{i=1}^m y^i \nu_i(p) \in N_\rho$$

of the tubular neighborhood $N_\rho = \bigcup_{p\in N} B_\rho(p)$ of N is a diffeomorphism, where $\nu_1, \dots, \nu_m$ is a suitable smooth orthonormal frame along N and where we let $y = (y^1, \dots, y^m) \in \mathbb{R}^m$. For $q \in N_\rho$, then $T^{-1}(q) = (p, h)$ with $p = \pi_N(q)$ defines a (vector-valued) signed distance function $h = h(q) = (h^1(q), \dots, h^m(q))$ with $h^i(q) = \nu_i(p) \cdot (q - \pi_N(q))$ for each $1 \le i \le m$. Fixing a smooth function $\eta \colon \mathbb{R} \to \mathbb{R}$ such that $\eta(s) = s$ for $|s| < \frac{1}{2}\rho$, and with $\eta(s) = 0$ for $|s| \ge \frac{3}{4}\rho$, we then let

$$\mathrm{dist}_N(q) = (\mathrm{dist}_N^1(q), \dots, \mathrm{dist}_N^m(q)),$$

with

$$\mathrm{dist}_N^i(q) = \eta(h^i(q)) \ \text{ for } q \in N_\rho, \quad \text{otherwise} \quad \mathrm{dist}_N^i(q) = 0, \qquad 1 \le i \le m.$$

Then for any smooth $u \in H^{1/2}(S^1; N)$ with harmonic extension $u \in H^1(B; \mathbb{R}^n)$ we have

$$\sum_{i=1}^m \nu_i(u)\partial_r \,\mathrm{dist}_N^i(u) = \sum_{i=1}^m \nu_i(u)\nu_i(u) \cdot u_r = d\pi_N^\perp(u)u_r \quad \text{on } \partial B = S^1, \tag{1-7}$$

where for each $p \in N$ we denote by $d\pi_N^\perp(p) = 1 - d\pi_N(p) \colon \mathbb{R}^n \to T_p^\perp N$ the orthogonal projection. In the sequel, we abbreviate

$$\sum_{i=1}^m \nu_i(u)\nu_i(u) \cdot u_r =: \nu(u)\nu(u) \cdot u_r = \nu(u)\partial_r \,\mathrm{dist}_N(u);$$

moreover, we extend the vector fields ν_i to the whole ambient space by letting $\nu_i(q) = \nabla \,\mathrm{dist}_N^i(q)$ for $q \in \mathbb{R}^n$, $1 \le i \le m$.

Finally, we fix a smooth cut-off function $\varphi \in C_c^\infty(B)$ satisfying $0 \le \varphi \le 1$ with $\varphi \equiv 1$ on $B_{1/2}(0)$, and for any $z_0 \in B$ and any $0 < R < 1$ we scale

$$\varphi_{z_0,R}(z) = \varphi\left(\frac{(z - z_0)}{R}\right) \in C_c^\infty(B_R(z_0)).$$

2. Energy inequality and first consequences

The half-harmonic heat flow may be regarded as the heat flow for the Dirichlet energy in the class $H^{1/2}(S^1; N)$. Indeed, let $u(t)$ be a smooth solution of (1-3) and (1-4) for $0 < t < T_0$. Then we have the following result.

Lemma 2.1. *For any* $0 \le S < T < T_0$,

$$E(u(T)) + \int_S^T \int_{\partial B} |u_t|^2 \, d\phi \, dt \le E(u(S)).$$

Proof. Integrating by parts and using (1-3) we compute

$$\frac{d}{dt} E(u) = \int_B \nabla u \nabla u_t \, dz = \int_{\partial B} u_r \cdot u_t \, d\phi = -\int_{\partial B} |d\pi_N(u) u_r|^2 \, d\phi = -\int_{\partial B} |u_t|^2 \, d\phi$$

for any $0 < t < T_0$. The claim follows by integration. □

Moreover, there holds a localized version of this energy inequality.

Lemma 2.2. *There exists a constant* $C > 0$ *such that, for any* $z_0 \in B$, *any* $0 < R < 1$, *any* $\varepsilon > 0$, *and any* $0 < t_0 < t_1 \le t_0 + \varepsilon R < T_0$,

$$\int_B |\nabla u(t_1)|^2 \varphi_{z_0,R}^2 \, dz + 4 \int_{t_0}^{t_1} \int_{\partial B} |u_t|^2 \varphi_{z_0,R}^2 \, d\phi \, dt \le 4 \int_B |\nabla u(t_0)|^2 \varphi_{z_0,R}^2 \, dz + C\varepsilon E(u_0).$$

Proof. Writing $\varphi = \varphi_{z_0,R}$ for brevity, integrating by parts, and using Young's inequality, similar to the proof of Lemma 2.1 for any $0 < t < T_0$ we have

$$\begin{aligned} \frac{d}{dt}\left(\frac{1}{2}\int_B |\nabla u|^2 \varphi^2 \, dz\right) &= \int_{\partial B} u_t \cdot u_r \varphi^2 \, d\phi - \int_B u_t \operatorname{div}(\nabla u \varphi^2) \, dz \\ &= -\int_{\partial B} |d\pi_N(u) u_r|^2 \varphi^2 \, d\phi - 2\int_B u_t \nabla u \varphi \nabla \varphi \, dz \\ &\le -\int_{\partial B} |u_t|^2 \varphi^2 \, d\phi + (8\varepsilon R)^{-1} \int_B |\nabla u|^2 \varphi^2 \, dz + 8\varepsilon R \int_B |u_t|^2 |\nabla \varphi|^2 \, dz. \end{aligned} \tag{2-1}$$

Letting

$$A = \sup_{t_0 < t < t_1} \left(\frac{1}{2}\int_B |\nabla u(t)|^2 \varphi^2 \, dz\right),$$

then upon integration we find

$$A + \int_{t_0}^{t_1} \int_{\partial B} |u_t|^2 \varphi^2 \, d\phi \, dt \le \int_B |\nabla u(t_0)|^2 \varphi^2 \, dz + \frac{t_1 - t_0}{2\varepsilon R} A + C\varepsilon R^{-1} \int_{t_0}^{t_1} \int_{B_R(z_0) \cap B} |u_t|^2 \, dz \, dt.$$

But with $u = u(t)$, then also $u_t = u_t(t)$ is harmonic for each t. Expanding

$$u_t(re^{i\phi}) = \sum_{k \ge 0} a_k r^k e^{ik\phi}$$

in a Fourier series, we see that the map

$$r \mapsto \int_{\partial B_r(0)} |u_t|^2 \, ds = 2\pi \sum_{k \ge 0} |a_k|^2 r^{2k+1},$$

with ds denoting the element of length along $\partial B_r(0)$, is nondecreasing. Thus, for any $z_0 \in B$, any $0 < R < 1$, and any $t_0 < t < t_1$,

$$\int_{B_R(z_0) \cap B} |u_t|^2 \, dz \le 2R \int_{\partial B} |u_t|^2 \, d\phi, \tag{2-2}$$

and we may use Lemma 2.1 to conclude. □

3. A regularity estimate

To illustrate the key ideas that later will allow us to prove higher regularity and analyze blow-up of solutions of (1-3), we first consider smooth solutions $u \in H^{1/2}(S^1; N)$ of the equation

$$d\pi_N(u)\partial_r u + f = 0 \quad \text{on } \partial B = S^1, \tag{3-1}$$

where $f \in L^2(S^1)$. We prove the following a-priori estimate, where we use classical estimates similar to [Wettstein 2022, Lemma 3.4], which in turn is a fractional version of a result by Rivière [1993, Chapter 4, pp. 96-104]. Note that with the truncated signed distance function $\text{dist}_N : \mathbb{R}^n \to \mathbb{R}^m$ we have the orthogonal decomposition

$$\partial_r u = d\pi_N(u)\partial_r u + d\pi_N^{\perp}(u)\partial_r u = d\pi_N(u)\partial_r u + \nu(u)\partial_r(\text{dist}_N(u)) \tag{3-2}$$

on $\partial B = S^1$, where we recall that we use the shorthand notation

$$\nu(u)\partial_r(\text{dist}_N(u)) = \sum_{i=1}^m \nu_i(u)\partial_r(\text{dist}_N^i(u)) = \sum_{i=1}^m \nu_i(u)\nu_i(u) \cdot \partial_r u$$

and extend $\nu_i(p) = \nabla \, \text{dist}_N^i(p), \ p \in \mathbb{R}^n$.

Proposition 3.1. *There exist constants C and $\delta_0 = \delta_0(N) > 0$ such that, for any smooth solution $u \in H^{1/2}(S^1; N)$ of (3-1) with $E(u) \le \delta^2 < \delta_0^2$,*

$$\int_{S^1} |\partial_\phi u|^2 \, d\phi \le C \|f\|^2_{L^2(S^1)}. \tag{3-3}$$

Proof. Multiplying (3-2) with $\partial_r u$, we find the Pythagorean identity

$$|\partial_r u|^2 = |d\pi_N(u)\partial_r u|^2 + |d\pi_N^{\perp}(u)\partial_r u|^2 = |d\pi_N(u)\partial_r u|^2 + |\partial_r(\text{dist}_N(u))|^2. \tag{3-4}$$

Note that $\text{dist}_N(u) \in H_0^1(B)$; moreover, for each $1 \le i \le m$ we have $\nabla(\text{dist}_N^i(u)) = \nu_i(u) \cdot \nabla u$, and there holds the equation

$$\Delta(\text{dist}_N^i(u)) = \text{div}(\nu_i(u) \cdot \nabla u) = \nabla u \cdot d\nu_i(u)\nabla u \quad \text{in } B. \tag{3-5}$$

The divergence theorem now gives

$$\begin{aligned}\|\partial_r(\mathrm{dist}_N(u))\|^2_{L^2(S^1)} &= (\nabla(\mathrm{dist}_N(u)), \nabla(\mathrm{dist}_N(u))_r)_{L^2(B)} + (\Delta(\mathrm{dist}_N(u)), (\mathrm{dist}_N(u))_r)_{L^2(B)} \\ &\le C\|\nabla u\|_{L^2(B)}\|\nabla^2(\mathrm{dist}_N(u))\|_{L^2(B)} \le C\delta\|\nabla^2(\mathrm{dist}_N(u))\|_{L^2(B)},\end{aligned}$$

where the basic L^2-theory for the Laplace equation (3-5) yields the bound

$$\|\nabla^2(\mathrm{dist}_N(u))\|_{L^2(B)} \le C\|\Delta(\mathrm{dist}^i_N(u))\|_{L^2(B)} \le C\|\nabla u\|^2_{L^4(B)}.$$

With Sobolev's embedding $H^{1/2}(B) \hookrightarrow L^4(B)$ we then conclude

$$\|\partial_r(\mathrm{dist}_N(u))\|^2_{L^2(S^1)} \le C\delta\|\nabla u\|^2_{H^{1/2}(B)}.$$

Thus from (3-4) and (3-1) we have

$$\|\partial_r u\|^2_{L^2(S^1)} \le \|f\|^2_{L^2(S^1)} + \|\partial_r(\mathrm{dist}_N(u))\|^2_{L^2(S^1)} \le \|f\|^2_{L^2(S^1)} + C\delta\|\nabla u\|^2_{H^{1/2}(B)}. \tag{3-6}$$

But Fourier expansion of the harmonic function u gives

$$\|\partial_\phi u\|^2_{L^2(S^1)} = \|\partial_r u\|^2_{L^2(S^1)} = \tfrac{1}{2}\|\nabla u\|^2_{L^2(S^1)} \tag{3-7}$$

as well as the bound

$$\|\nabla u\|^2_{H^{1/2}(B)} \le C\|\nabla u\|^2_{L^2(S^1)},$$

and from (3-6) we obtain

$$\|\partial_r u\|^2_{L^2(S^1)} \le \|f\|^2_{L^2(S^1)} + C\delta\|\nabla u\|^2_{H^{1/2}(B)} \le \|f\|^2_{L^2(S^1)} + C\delta\|\partial_r u\|^2_{L^2(S^1)},$$

which for sufficiently small $\delta > 0$ by (3-7) yields the claim. □

In particular, from Proposition 3.1 we obtain a positive energy threshold for nonconstant solutions of (1-6).

Corollary 3.2. *Suppose $u \in H^{1/2}(S^1; N)$ smoothly solves (1-6). Then, either u is constant or $E(u) \ge \delta_0^2$, with $\delta_0 = \delta_0(N) > 0$ given by Proposition 3.1.*

Combining the ideas in the proof of the previous result with ideas from the classical proof of the Courant–Lebesgue lemma in minimal surface theory, we can obtain the following local version of Proposition 3.1.

Proposition 3.3. *There exists a constant $\delta > 0$ with the following property. Given any smooth solution $u \in H^{1/2}(S^1; N)$ of (3-1) with harmonic extension $u \in H^1(B)$, any $z_0 \in \partial B$, and any $0 < R \le \frac{1}{2}$ such that*

$$\int_{B_R(z_0)\cap B} |\nabla u|^2\, dz < \delta^2, \tag{3-8}$$

with a constant $C = C(R) > 0$ there holds

$$\int_{B_{R^2}(z_0)\cap S^1} |\partial_\phi u|^2\, d\phi \le C\|f\|^2_{L^2(B_R(z_0)\cap S^1)} + CE(u).$$

Proof. Fix any $z_0 \in \partial B$ and $0 < R \le \frac{1}{2}$ such that (3-8) holds. For suitable $\rho \in [R^2, R]$, with s denoting arc-length along the curve $C_\rho = \{z_0 + \rho e^{i\theta} \in B : \theta \in \mathbb{R}\}$ with end-points $z_j = z_0 + \rho e^{i\theta_j} = e^{i\phi_j} \in \partial B$, $j = 1, 2$, we have

$$\rho \int_{C_\rho} |\nabla u|^2 \, ds \le 2 \inf_{R^2 < \rho' < R} \left(\rho' \int_{C_{\rho'}} |\nabla u|^2 \, ds \right).$$

We can bound the latter infimum by the average over $\rho \in [R^2, R]$ with respect to the measure with density ρ^{-1} to obtain the bound

$$\rho \int_{C_\rho} |\nabla u|^2 \, ds \le \frac{2 \int_{R^2}^{R} \int_{C_\rho} |\nabla u|^2 \, ds \, d\rho}{\int_{R^2}^{R} \rho^{-1} \, d\rho} \le \frac{2 \int_B |\nabla u|^2 \, dz}{|\log(R)|} = \frac{4E(u)}{|\log(R)|}. \tag{3-9}$$

Let $\Phi_0 \colon B \to B$ be the conformal map fixing the circular arc C_ρ and mapping the point z_0 to the point $-z_0$, obtained as the composition $\Phi_0 = \pi_0^{-1} \circ \Psi_0 \circ \pi_0$ of a conformal diffeomorphism $\pi_0 \colon B \to \mathbb{R}^2_+$ mapping the points z_0 and $-z_0$ to the origin and infinity, respectively, and the reflection $\Psi_0 \colon \mathbb{R}^2_+ \to \mathbb{R}^2_+$ of the upper half-plane $\mathbb{R}^2_+$ in the half-circle $\pi_0(C_\rho)$. Replacing u by the map $u \circ \Phi_0$ in $B \setminus B_\rho(z_0)$ we obtain a piecewise smooth map $v_1 \colon B \to \mathbb{R}^n$ which is harmonic on $B \setminus C_\rho$ and continuous on all of B. Let $v_0 \in H^1(B)$ be harmonic with $w := v_1 - v_0 \in H^1_0(B)$. Note that by the variational characterization of harmonic functions and conformal invariance of the Dirichlet integral we have

$$E(v_0) \le E(v_1) \le \int_{B_R(z_0) \cap B} |\nabla u|^2 \, dz \le \delta^2. \tag{3-10}$$

Moreover, for any smooth $\varphi \in H^1_0(B)$, by (3-9) we can estimate

$$\begin{aligned} \left| \int_B \nabla w \nabla \varphi \, dz \right| = \left| \int_B \nabla v_1 \nabla \varphi \, dz \right| = \left| \int_{C_\rho} [\partial_\nu v_1] \varphi \, ds \right| &\le \left(\int_{C_\rho} |\nabla u|^2 \, ds \right)^{1/2} \left(\int_{C_\rho} |\varphi|^2 \, ds \right)^{1/2} \\ &\le C(R) E(u)^{1/2} \|\varphi\|_{H^{1/2}(B)}, \end{aligned}$$

where $[\partial_\nu v_1]$ denotes the difference of the outer and inner normal derivatives of v_1 along C_ρ. Thus we have $\Delta w \in H^{-1/2}(B)$, and the basic L^2-theory for the Laplace equation gives $w \in H^{3/2} \cap H^1_0(B)$ with

$$\|w\|_{H^{3/2}(B)} \le \sup_{\substack{\varphi \in H^1_0(B) \\ \|\varphi\|_{H^{1/2}(B)} \le 1}} \left(\int_B \nabla w \nabla \varphi \, dz \right) \le C(R) E(u)^{1/2},$$

and then also

$$\|\partial_r w\|^2_{L^2(S^1)} \le C \|w\|^2_{H^{3/2}(B)} \le C(R) E(u). \tag{3-11}$$

In view of (3-10), for sufficiently small $\delta > 0$, from Proposition 3.1 we obtain the estimate

$$\|\partial_\phi v_0\|^2_{L^2(S^1)} \le C \|d\pi_N(v_0) \partial_r v_0\|^2_{L^2(S^1)}. \tag{3-12}$$

Observe that since $v_0 = v_1$ on $\partial B = S^1$ and since we also have $v_1 = u$ on $B \cap B_\rho(z_0)$ and $v_1 = u \circ \Phi_0$ on $B \setminus B_\rho(z_0)$, respectively, we can bound

$$\|d\pi_N(v_0) \partial_r v_0\|^2_{L^2(S^1)} = \|d\pi_N(v_1) \partial_r v_0\|^2_{L^2(S^1)} \le 2 \|d\pi_N(v_1) \partial_r v_1\|^2_{L^2(S^1)} + 2 \|\partial_r w\|^2_{L^2(S^1)}$$

and

$$\|d\pi_N(v_1)\partial_r v_1\|^2_{L^2(S^1)} \le C(R)\|d\pi_N(u)\partial_r u\|^2_{L^2(S^1\cap B_\rho(z_0))}.$$

Thus from (3-11) we obtain

$$\begin{aligned}\|d\pi_N(v_0)\partial_r v_0\|^2_{L^2(S^1)} &\le C(R)\|d\pi_N(u)\partial_r u\|^2_{L^2(S^1\cap B_\rho(z_0)} + C\|\partial_r w\|^2_{L^2(S^1)}\\ &\le C(R)\|f\|^2_{L^2(S^1\cap B_\rho(z_0))} + C(R)E(u),\end{aligned}$$

and from (3-12) there results the bound

$$\begin{aligned}\|\partial_\phi u\|^2_{L^2(S^1\cap B_\rho(z_0))} &= \|\partial_\phi v_0\|^2_{L^2(S^1\cap B_\rho(z_0))} \le \|\partial_\phi v_0\|^2_{L^2(S^1)}\\ &\le C\|d\pi_N(v_0)\partial_r v_0\|^2_{L^2(S^1)} \le C(R)\|f\|^2_{L^2(S^1\cap B_R(z_0)))} + C(R)E(u),\end{aligned}$$

as claimed. □

The local estimate Proposition 3.3 also implies the following global bound.

Proposition 3.4. *There exists a constant* $\delta > 0$ *with the following property. Given any smooth solution* $u \in H^{1/2}(S^1; N)$ *of* (3-1) *and any* $0 < R \le \frac{1}{2}$ *with*

$$\sup_{z_0\in B}\int_{B_R(z_0)\cap B} |\nabla u|^2\,dz < \delta^2, \tag{3-13}$$

there holds

$$\int_{S^1} |\partial_\phi u|^2\,d\phi \le C(R)\|f\|^2_{L^2(S^1)} + C(R)E(u).$$

Proof. Covering ∂B with balls $B_{R^2}(z_i)$, $1 \le i \le i_0$, from Proposition 3.3 we obtain the claim. □

Remark 3.5. The proofs of the above propositions only require $u \in H^1(S^1; N)$ with harmonic extension $u \in H^{3/2}(B)$.

4. Higher regularity

Again let $u(t)$ be a smooth solution of the half-harmonic heat flow (1-3) for $0 < t < T_0$ with smooth initial data (1-4). We show that as long as the flow does not concentrate energy in the sense of Theorem 1.1(ii) the solution remains smooth and can be a-priori bounded in any H^k-norm in terms of the data.

H^2-bound. In a first step we show an L^2-bound in space-time for the second derivatives of our solution to the flow (1-3). Recall that by harmonicity, writing $u = u(t)$, $\partial_\phi u = u_\phi$, and so on, for any $0 < t < T_0$ we have (3-7), that is,

$$\int_{\partial B} |u_\phi|^2\,d\phi = \int_{\partial B} |u_r|^2\,d\phi,$$

as Fourier expansion shows, with similar identities for partial derivatives of u of higher order. Indeed, writing

$$\Delta u = \frac{1}{r}(ru_r)_r + \frac{1}{r^2}u_{\phi\phi} \tag{4-1}$$

we see that $\partial_\phi^j u$ and then also $\nabla^{k-j}\partial_\phi^j u$ is harmonic for any $j \le k$ in $\mathbb{N}_0$, where $\nabla u = (u_x, u_y)$ in Euclidean coordinates $z = x + iy$. Thus by induction we obtain

$$\int_{\partial B} |\nabla^k u|^2 \, d\phi = 2 \int_{\partial B} |\nabla^{k-1} u_\phi|^2 \, d\phi = \cdots = 2^k \int_{\partial B} |\partial_\phi^k u|^2 \, d\phi \tag{4-2}$$

for any $k \in \mathbb{N}$. Similarly, for any $\frac{1}{4} < r < 1$ with uniform constants $C > 0$ we have

$$\int_{\partial B_r(0)} |\nabla^k u|^2 \, dz \le C \int_{\partial B_r(0)} |\nabla^{k-1} u_\phi|^2 \, dz \le \cdots \le C \int_{\partial B_r(0)} |\partial_\phi^k u|^2 \, dz.$$

Integrating and using the mean value property of harmonic functions together with (4-2) to bound

$$\sup_{B_{1/4}(0)} |\nabla^k u|^2 \le C \int_{B \setminus B_{1/4}(0)} |\nabla^k u|^2 \, dz \le C \int_B |\nabla \partial_\phi^{k-1} u|^2 \, dz,$$

in particular, for any $k \in \mathbb{N}$, we have the bound

$$\int_B |\nabla^k u|^2 \, dz \le C \int_B |\nabla \partial_\phi^{k-1} u|^2 \, dz \tag{4-3}$$

with an absolute constant $C > 0$.

The following lemma is strongly reminiscent of analogous results for the harmonic map heat flow in two space dimensions.

Lemma 4.1. *With a constant $C > 0$ depending only on N,*

$$\frac{d}{dt}\left(\int_{\partial B} |u_\phi|^2 \, d\phi\right) + \int_B |\nabla u_\phi|^2 \, dz \le C \int_B |\nabla u|^2 |u_\phi|^2 \, dz.$$

Proof. Writing $d\pi_N(u) = 1 - d\pi_N^\perp(u)$ with

$$d\pi_N^\perp(u)X = \nu(u)\nu(u) \cdot X = \sum_{i=1}^m \nu_i(u)\nu_i(u) \cdot X$$

for any $X \in \mathbb{R}^n$, we compute

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}\left(\int_{\partial B} |u_\phi|^2 \, d\phi\right) &= \int_{\partial B} u_\phi \cdot u_{\phi,t} \, d\phi = -\int_{\partial B} u_{\phi\phi} \cdot u_t \, d\phi \\
&= \int_{\partial B} u_{\phi\phi} \cdot d\pi_N(u) u_r \, d\phi = -\int_{\partial B} (u_\phi \cdot u_{r\phi} - u_\phi \cdot \partial_\phi(\nu(u)\nu(u) \cdot u_r)) \, d\phi \\
&= -\frac{1}{2}\int_{\partial B} \partial_r(|u_\phi|^2) \, d\phi - \int_{\partial B} u_\phi \cdot d\nu(u) u_\phi \nu(u) \cdot u_r \, d\phi,
\end{aligned}$$

where we use orthogonality $u_\phi \cdot \nu_i(u) = 0$ on ∂B, $1 \le i \le m$, in the last step. But u_ϕ is harmonic. So with $\Delta |u_\phi|^2 = 2|\nabla u_\phi|^2$, from Gauss' theorem we obtain

$$\frac{1}{2}\int_{\partial B} \partial_r(|u_\phi|^2) \, d\phi = \int_B |\nabla u_\phi|^2 \, dz.$$

On the other hand, by Young's inequality we can estimate

$$\int_{\partial B} u_r \cdot \nu(u)u_\phi \cdot d\nu(u)u_\phi \, d\phi = \int_B \nabla u \cdot \nabla(\nu(u)u_\phi \cdot d\nu(u)u_\phi)\, dz$$
$$\le C \int_B |\nabla u_\phi||\nabla u|\,|u_\phi|\, dz + C\int_B |\nabla u|^2|u_\phi|^2\, dz$$
$$\le \frac{1}{2}\int_B |\nabla u_\phi|^2\, dz + C\int_B |\nabla u|^2|u_\phi|^2\, dz,$$

and our claim follows. □

Combining the previous result with a quantitative bound for the concentration of energy, we obtain a space-time bound for the second derivatives of u. Note that since u is smooth by assumption, for any $\delta > 0$ and any $T < T_0$, there exists a number $R = R(T, u) > 0$ such that

$$\sup_{\substack{z_0\in B\\ 0<t<T}} \int_{B_R(z_0)\cap B} |\nabla u(t)|^2\, dz < \delta. \tag{4-4}$$

Proposition 4.2. *There exist constants $\delta = \delta(N) > 0$ and $C > 0$ such that, for any $T < T_0$ with $R > 0$ as in (4-4),*

$$\sup_{0<t<T} \int_{\partial B} |u_\phi(t)|^2\, d\phi + \int_0^T\int_B |\nabla u_\phi|^2\, dx\, dt \le C\int_{\partial B} |u_{0,\phi}|^2\, d\phi + CTR^{-2}E(u_0). \tag{4-5}$$

Proof. For given $T < T_0$ and $\delta > 0$ to be determined, we fix $R > 0$ such that (4-4) holds. Let $B_{R/2}(z_i)$, $1 \le i \le i_0$, be a cover of B such that any point $z_0 \in B$ belongs to at most L of the balls $B_R(z_i)$, where $L \in \mathbb{N}$ is independent of $R > 0$. We then use the decomposition

$$\int_B |\nabla u|^2|u_\phi|^2\, dz \le \sum_{i=1}^{i_0}\int_{B_{R/2}(z_i)} |\nabla u|^4\, dz \le \sum_{i=1}^{i_0}\int_B |\nabla u\varphi_{z_i,R}|^4\, dz.$$

Using the multiplicative inequality (A-2) in the Appendix, for each i we can bound

$$\int_B |\nabla u\varphi_{z_i,R}|^4\, dz \le C\delta\int_{B_R(z_i)} (|\nabla^2 u|^2 + R^{-2}|\nabla u|^2)\, dz.$$

Summing over $1 \le i \le i_0$, we thus obtain the bound

$$\int_B |\nabla u|^2|u_\phi|^2\, dz \le CL\delta\int_B |\nabla^2 u|^2\, dz + CL\delta R^{-2}E(u) \le CL\delta\int_B |\nabla u_\phi|^2\, dz + CL\delta R^{-2}E(u_0),$$

and for sufficiently small $\delta > 0$ we obtain the claim from Lemma 4.1. □

With the help of Proposition 4.2 we can now bound u in $H^2(B)$ also uniformly in time. For this, we first note the following estimate, which also will be useful later for bounding higher-order derivatives.

Lemma 4.3. *For any $k \in \mathbb{N}$, with a constant $C > 0$ depending only on k and N, for the solution $u = u(t)$ to (1-3) and (1-4) for any $0 < t < T_0$*

$$\frac{d}{dt}(\|\nabla\partial_\phi^k u\|^2_{L^2(B)}) + \|\partial_\phi^k u_r\|^2_{L^2(S^1)} \le C\sum_{\substack{1\le j_i\le k+1\\ \Sigma_i j_i\le k+2}} \|\nabla\partial_\phi^k u\|_{L^2(B)}\Big\|\prod_i \nabla^{j_i}u\Big\|_{L^2(B)}.$$

Proof. For any $k \in \mathbb{N}$ we use harmonicity of $\partial_\phi^{2k} u$ to compute

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}(\|\nabla\partial_\phi^k u\|^2_{L^2(B)}) &= (-1)^k \int_B \nabla\partial_\phi^{2k} u \nabla u_t\, dx \\ &= (-1)^k(\partial_\phi^{2k} u_r, u_t)_{L^2(S^1)} = (-1)^{k+1}(\partial_\phi^{2k} u_r, d\pi_N(u)u_r)_{L^2(S^1)} \\ &= -(\partial_\phi^k u_r, \partial_\phi^k u_r)_{L^2(S^1)} + (\partial_\phi^k u_r, \partial_\phi^k(\nu(u)\nu(u)\cdot u_r))_{L^2(S^1)} = -\|\partial_\phi^k u_r\|^2_{L^2(S^1)} + I, \quad \text{(4-6)}\end{aligned}$$

where we use the decomposition $I = \sum_{j=0}^k \binom{k}{j} I_j$ with

$$I_j = (\partial_\phi^k u_r, \partial_\phi^j(\nu(u)\nu(u))\partial_\phi^{k-j} u_r)_{L^2(S^1)} = (\nabla\partial_\phi^k u, \nabla(\partial_\phi^j(\nu(u)\nu(u))\cdot\partial_\phi^{k-j} u_r))_{L^2(B)}.$$

Hence for any $1 \le j \le k$ we can bound

$$\begin{aligned}|I_j| &\le C \sum_{0\le i\le j} \|\nabla\partial_\phi^k u\|_{L^2(B)} \|\nabla\partial_\phi^{j-i}\nu(u)\partial_\phi^i \nu(u)\partial_\phi^{k-j} u_r\|_{L^2(B)} \\ &\qquad + C \sum_{0\le i\le j} \|\nabla\partial_\phi^k u\|_{L^2(B)} \|\partial_\phi^{j-i}\nu(u)\partial_\phi^i\nu(u)\nabla\partial_\phi^{k-j} u_r\|_{L^2(B)} \\ &\le C \sum_{\substack{1\le j_i\le k+1 \\ \Sigma_i j_i = k+2}} \|\nabla\partial_\phi^k u\|_{L^2(B)} \Big\|\prod_i \nabla^{j_i} u\Big\|_{L^2(B)},\end{aligned}$$

as claimed. It remains to bound the term $I_0 = \|\partial_\phi^k u_r\cdot\nu(u)\|^2_{L^2(S^1)}$. With the signed distance function we can express

$$\nu(u)\cdot u_{\phi r} = (\nu(u)\cdot u_r)_\phi - u_r\cdot d\nu(u)u_\phi = (\mathrm{dist}_N(u))_{\phi r} - u_r\cdot d\nu(u)u_\phi,$$

so that

$$\begin{aligned}I_0 = \|\partial_\phi^k u_r\cdot\nu(u)\|^2_{L^2(S^1)} &= (\partial_\phi^k u_r\cdot\nu(u), \partial_\phi^k(\mathrm{dist}_N(u))_r)_{L^2(S^1)} + II \\ &= (\nabla\partial_\phi^k u, \nabla(\nu(u)\partial_\phi^k(\mathrm{dist}_N(u))_r))_{L^2(B)} + II,\end{aligned}$$

where all terms in II can be dealt with as in the case $1 \le j \le k$. Finally, we have

$$\begin{aligned}&(\nabla\partial_\phi^k u, \nabla(\nu(u)\partial_\phi^k(\mathrm{dist}_N(u))_r))_{L^2(B)} \\ &\qquad \le \|\nabla\partial_\phi^k u\|_{L^2(B)}(\|\nabla^2\partial_\phi^k(\mathrm{dist}_N(u))\|_{L^2(B)} + \|\nabla\nu(u)\partial_\phi^k(\mathrm{dist}_N(u))_r\|_{L^2(B)}).\end{aligned}$$

But by the chain rule we can bound

$$\|\nabla\nu(u)\partial_\phi^k(\mathrm{dist}_N(u))_r\|_{L^2(B)} \le C\|\nabla u\nabla^{k+1}(\mathrm{dist}_N(u))\|_{L^2(B)} \le C \sum_{\substack{1\le j_i\le k+1 \\ \Sigma_i j_i = k+2}} \Big\|\prod_i \nabla^{j_i} u\Big\|_{L^2(B)}.$$

Moreover, by (3-5) and elliptic regularity theory,

$$\begin{aligned}\|\nabla^{k+2}(\mathrm{dist}_N(u))\|^2_{L^2(B)} &\le C\|\Delta(\mathrm{dist}_N(u))\|^2_{H^k(B)} \\ &\le C\|\nabla u\cdot d\nu_i(u)\nabla u\|^2_{H^k((B)} \le C \sum_{\substack{1\le j_i\le k+1 \\ \Sigma_i j_i \le k+2}} \Big\|\prod_i \nabla^{j_i} u\Big\|_{L^2(B)},\end{aligned}$$

which gives the claim. □

For $k = 1$, from Proposition 4.2 we now easily derive a uniform L^2-bound for the second derivatives of the flow.

Proposition 4.4. *For any smooth $u_0 \in H^{1/2}(S^1; N)$ and any $T < T_0$ with $R > 0$ as in Proposition 4.2 with a constant $C_1 = C_1(T, R, u_0) > 0$ depending on the right-hand side of (4-5),*

$$\sup_{0<t<T} \int_B |\nabla u_\phi(t)|^2 \, dz + \int_0^T \int_{\partial B} |u_{\phi r}|^2 \, d\phi \, dt \le C_1 \int_B |\nabla u_{0,\phi}|^2 \, dz + C_1.$$

Proof. For $k = 1$, by Lemma 4.3 we need to bound the term

$$J = \sum_{\substack{1\le j_i \le 2\\ \Sigma_i j_i \le 3}} \Big\| \prod_i \nabla^{j_i} u \Big\|_{L^2(B)} \le C \| |\nabla^2 u| |\nabla u| + |\nabla u|^3 \|_{L^2(B)} + J_1,$$

where J_1 contains all terms of lower order. By the maximum principle and Sobolev's embedding $H^1(\partial B) \hookrightarrow L^\infty(\partial B)$ we can estimate

$$\|\nabla u\|^2_{L^\infty(B)} \le \|\nabla u\|^2_{L^\infty(\partial B)} \le C \|\nabla u\|^2_{H^1(\partial B)} \le C \|u_{\phi r}\|^2_{L^2(\partial B)} + C_1,$$

where we have also used (3-7) and Proposition 4.2. Also bounding

$$\|\nabla u\|^3_{L^6(B)} \le \|\nabla u\|^2_{L^4(B)} \|\nabla u\|_{L^\infty(B)} \le C(\|\nabla^2 u\|_{L^2(B)} \|\nabla u\|_{L^2(B)} + E(u)) \|\nabla u\|_{L^\infty(B)}$$

via (A-2), and again using (3-7) (and with similar, but simpler bounds for J_1), we arrive at the estimate

$$\begin{aligned} J \le C \| |\nabla^2 u| |\nabla u| + |\nabla u|^3 \|_{L^2(B)} + C_1 &\le C(\|\nabla^2 u\|_{L^2(B)} + E(u)) \|\nabla u\|_{L^\infty(B)} + C_1 \\ &\le C(1 + \|\nabla u_\phi\|_{L^2(B)} + E(u_0))(\|u_{\phi r}\|_{L^2(\partial B)} + C_1). \end{aligned}$$

With Lemma 4.3 and Young's inequality we then have

$$\begin{aligned} \frac{d}{dt}(1 + \|\nabla u_\phi\|^2_{L^2(B)}) + \|u_{\phi r}\|^2_{L^2(S^1)} &\le C \|\nabla u_\phi\|_{L^2(B)} (\|\nabla u_\phi\|_{L^2(B)} + E(u_0))(\|u_{\phi r}\|_{L^2(\partial B)} + C_1) \\ &\le \tfrac{1}{2} \|u_{\phi r}\|^2_{L^2(\partial B)} + C(1 + \|\nabla u_\phi\|^2_{L^2(B)})(\|\nabla u_\phi\|^2_{L^2(B)} + C_1). \end{aligned} \tag{4-7}$$

Absorbing the term $\frac{1}{2}\|u_{\phi r}\|^2_{L^2(\partial B)}$ into the left-hand side of this inequality and dividing by $1 + \|\nabla u_\phi\|^2_{L^2(B)}$ we obtain

$$\frac{d}{dt}(\log(1 + \|\nabla u_\phi\|^2_{L^2(B)})) \le C \|\nabla u_\phi\|^2_{L^2(B)} + C_1,$$

and from Proposition 4.2 we obtain the bound

$$\sup_{0<t<T} \|\nabla u_\phi(t)\|^2_{L^2(B)} \le C_1 (1 + \|\nabla u_{0,\phi}\|^2_{L^2(B)}).$$

The claim then follows from (4-7). □

H^3-bounds. The derivation of a-priori L^2-bounds for third derivatives of the solution u to the flow (1-3), (1-4) requires special care, which is why we highlight this case.

Proposition 4.5. *For any smooth $u_0 \in H^{1/2}(S^1; N)$ and any $T < T_0$,*

$$\sup_{0<t<T} \int_B |\nabla u_{\phi\phi}(t)|^2 \, dz + \int_0^T \int_{\partial B} |u_{\phi\phi r}|^2 \, d\phi \, dt \le C_2 \int_B |\nabla u_{0,\phi\phi}|^2 \, dz + C_2,$$

where we denote by $C_2 = C_2(T, R, u_0) > 0$ a constant bounded by the terms on the right-hand side in the statements of Propositions 4.2 and 4.4.

Proof. For $k = 2$, by Lemma 4.3 we need to bound the term

$$J = \sum_{\substack{1\le j_i\le 3\\ \Sigma_i j_i=4}} \Big\| \prod_i \nabla^{j_i} u \Big\|_{L^2(B)} \le C \| |\nabla u|^4 + |\nabla u|^2 |\nabla^2 u| + |\nabla^2 u|^2 + |\nabla u| |\nabla^3 u| \|_{L^2(B)}$$

and corresponding terms involving at most three derivatives in total, which we will omit.

For the first term, by the multiplicative inequality (A-2) and Sobolev's embedding $H^2(B) \hookrightarrow L^\infty(B)$, we can estimate

$$\begin{aligned}\|\nabla u\|^4_{L^8(B)} &\le \|\nabla u\|^2_{L^4(B)} \|\nabla u\|^2_{L^\infty(B)} \le C\|\nabla u\|_{H^1(B)} \|\nabla u\|_{L^2(B)} \|\nabla u\|^2_{L^\infty(B)} \\ &\le C(\|\nabla^2 u\|^2_{L^2(B)} + E(u))\|\nabla u\|^2_{L^\infty(B)} \le C_2 \|\nabla u\|^2_{L^\infty(B)} \\ &\le C_2(\|\nabla^3 u\|_{L^2(B)} + \|\nabla u\|_{L^2(B)})\|\nabla u\|_{L^\infty(B)}\end{aligned}$$

with a constant $C_2 = C_2(T, R, u_0) > 0$ as in the statement of the proposition. Similarly,

$$\|\nabla^2 u\|^2_{L^4(B)} \le C\|\nabla^2 u\|_{H^1(B)} \|\nabla^2 u\|_{L^2(B)} \le \|\nabla^3 u\|_{L^2(B)} \|\nabla^2 u\|_{L^2(B)} + \|\nabla^2 u\|^2_{L^2(B)} \le C_2(1 + \|\nabla^3 u\|_{L^2(B)}).$$

Hence we can also bound

$$\| |\nabla u|^2 |\nabla^2 u| \|_{L^2(B)} \le \|\nabla u\|^4_{L^8(B)} + \|\nabla^2 u\|^2_{L^4(B)} \le C_2(1 + \|\nabla^3 u\|_{L^2(B)})(1 + \|\nabla u\|_{L^\infty(B)}).$$

Finally, we estimate

$$\| |\nabla u| |\nabla^3 u| \|_{L^2(B)} \le \|\nabla^3 u\|_{L^2(B)} \|\nabla u\|_{L^\infty(B)}$$

to obtain

$$J \le C_2(1 + \|\nabla^3 u\|_{L^2(B)})(1 + \|\nabla u\|_{L^\infty(B)}).$$

But with the inequality

$$\|f\|_{L^\infty(B)} \le C\|f\|_{H^1(B)} \left(1 + \log^{1/2}\left(1 + \frac{\|f\|_{H^2(B)}}{\|f\|_{H^1(B)}}\right)\right)$$

for $f \in H^2(B)$ due to Brezis and Gallouet [1980] (see also [Brézis and Wainger 1980] for a more general version), we have

$$\|\nabla u\|^2_{L^\infty(B)} \le C\|\nabla u\|^2_{H^1(B)} \left(1 + \log\left(1 + \frac{\|\nabla u\|_{H^2(B)}}{\|\nabla u\|_{H^1(B)}}\right)\right) \le C_2(1 + \log(1 + \|\nabla^3 u\|_{L^2(B)})),$$

and Lemma 4.3 yields the differential inequality

$$\frac{d}{dt}(\|\nabla \partial_\phi^2 u\|^2_{L^2(B)}) + \|u_{\phi\phi r}\|^2_{L^2(\partial B)} \le C_2 \|\nabla \partial_\phi^2 u\|_{L^2(B)} (1 + \|\nabla^3 u\|_{L^2(B)})(1 + \log(1 + \|\nabla^3 u\|_{L^2(B)})).$$

Simplifying, and recalling that $\|\nabla^3 u\|^2_{L^2(B)} \le C\|\nabla \partial_\phi^2 u\|^2_{L^2(B)}$ by (4-3), we then find

$$\frac{d}{dt}(1 + \|\nabla \partial_\phi^2 u\|_{L^2(B)}) \le C_2(1 + \|\nabla \partial_\phi^2 u\|_{L^2(B)})(1 + \log(1 + \|\nabla \partial_\phi^2 u\|_{L^2(B)}));$$

that is, we have

$$\frac{d}{dt}(1 + \log(1 + \|\nabla \partial_\phi^2 u\|_{L^2(B)})) \le C_2(1 + \log(1 + \|\nabla \partial_\phi^2 u\|_{L^2(B)})).$$

Arguing as in the proof of Proposition 4.4 we then obtain the claim. □

H^m-bounds, $m \geq 4$. In view of Proposition 4.5 we can now use induction to prove the following result.

Proposition 4.6. *For any $k \geq 3$, any smooth $u_0 \in H^{1/2}(S^1; N)$, and any $T < T_0$,*

$$\sup_{0<t<T} \int_B |\nabla \partial_\phi^k(t)|^2 \, dz + \int_0^T \int_{\partial B} |\partial_\phi^k u_r|^2 \, d\phi \, dt \leq C_k \int_B |\nabla \partial_\phi^k u_0|^2 \, dz + C_k,$$

where we denote by $C_k = C_k(T, R, u_0) > 0$ a constant bounded by the terms on the right-hand side in the statement of the proposition for $k-1$.

Proof. By Proposition 4.5 the claimed result holds true for $k = 2$. Suppose the claim holds true for some $k_0 \geq 2$ and let $k = k_0 + 1$. Note that by Sobolev's embedding $H^2(B) \hookrightarrow W^{1,4} \cap C^0(\bar{B})$ and (4-3) for $0 \leq t < T$ we then have the uniform bounds

$$\|\nabla^{k_0+1} u\|^2_{L^2(B)} + \|\nabla^{k_0} u\|^2_{L^4(B)} + \sum_{1 \leq j \leq k_0-1} \|\nabla^j u\|^2_{L^\infty(B)} \leq C_{k_0} \|\nabla^{k_0+1} u_0\|^2_{L^2(B)} + C_{k_0} \leq C_k < \infty \quad (4\text{-}8)$$

with a constant of the type C_k, as defined above.

By Lemma 4.3 again we only need to bound the term

$$J = \sum_{\substack{1 \leq j_i \leq k+1 \\ \Sigma_i j_i \leq k+2}} \Big\| \prod_i \nabla^{j_i} u \Big\|_{L^2(B)}.$$

Clearly we have

$$\begin{aligned} J &\leq \|\nabla^{k+1} u\|_{L^2(B)} \|\nabla u\|_{L^\infty(B)} + \|\nabla^k u\|_{L^2(B)} \|\nabla u\|^2_{L^\infty(B)} + \|\nabla^k u \nabla^2 u\|_{L^2(B)} \\ &\qquad + \|\nabla^{k-1} u \nabla^3 u\|_{L^2(B)} + \|\nabla^{k-1} u \nabla^2 u\|_{L^2(B)} \|\nabla u\|_{L^\infty(B)} + C_k \\ &\leq C_k \|\nabla^{k+1} u\|_{L^2(B)} + \|\nabla^k u \nabla^2 u\|_{L^2(B)} + \|\nabla^{k-1} u \nabla^3 u\|_{L^2(B)} + C_k. \end{aligned}$$

We now distinguish the following cases: If $k - 1 = k_0 \geq 3$, by (4-8) we can bound

$$\|\nabla^k u \nabla^2 u\|_{L^2(B)} \leq \|\nabla^k u\|_{L^2(B)} \|\nabla^2 u\|_{L^\infty(B)} \leq C_{k_0} \|\nabla^{k_0+1} u\|^2_{L^2(B)} + C_{k_0} \leq C_k$$

as well as

$$\|\nabla^{k-1} u \nabla^3 u\|_{L^2(B)} \leq \|\nabla^{k-1} u\|_{L^4(B)} \|\nabla^3 u\|_{L^4(B)} \leq C_{k_0} \|\nabla^{k_0} u\|^2_{L^4(B)} + C_{k_0} \leq C_k$$

to obtain the estimate

$$J \leq C_k \|\nabla^{k+1} u\|_{L^2(B)} + C_k.$$

If, on the other hand, $k_0 = k - 1 = 2$, by our induction hypothesis (4-8) we have

$$\begin{aligned} \|\nabla^{k-1} u \nabla^3 u\|_{L^2(B)} = \|\nabla^2 u \nabla^k u\|^2_{L^2(B)} &\leq \|\nabla^k u\|_{L^4(B)} \|\nabla^2 u\|_{L^4(B)} \\ &\leq C_k \|\nabla^k u\|_{H^1(B)} + C_k \leq C_k \|\nabla^{k+1} u\|_{L^2(B)} + C_k, \end{aligned}$$

and we find

$$J \leq C_k \|\nabla^{k+1} u\|_{L^2(B)} + C_k$$

as before.

In any case, inequality (4-3) and Lemma 4.3 now may be invoked to obtain

$$\frac{d}{dt}(\|\nabla\partial_\phi^k u\|_{L^2(B)}) \le C_k\|\nabla\partial_\phi^k u\|_{L^2(B)} + C_k,$$

and our claim follows. □

Local H^k-bounds. The bounds established so far all require the initial data to be sufficiently smooth for the estimate at hand and do not yet allow to show smoothing of the flow. For the latter purpose we next prove a second set of "intermediate" estimates that in combination with the first set of estimates later will allow boot-strapping. Moreover, in contrast to the estimates in Lemma 4.3, the following estimates may be localized. This will be important for showing regularity of the flow at blow-up times away from concentration points of the energy on ∂B.

For the localized estimates, fix a point $z_0 \in \partial B$ and some radius $0 < R_0 < \frac{1}{4}$ and for $k \in \mathbb{N}$ set $R_k = 2^{-k}R_0$ and $\varphi_k = \varphi_{z_0,R_k}$. Set $\varphi_k = 1$ for each $k \in \mathbb{N}$ for the analogous global bounds.

We first establish the following localized version of Lemma 4.1.

Lemma 4.7. *With a constant $C > 0$ depending only on N,*

$$\frac{d}{dt}\left(\int_{\partial B}|u_\phi|^2\varphi_1^2\,d\phi\right) + \int_B|\nabla u_\phi|^2\varphi_1^2\,dz \le C\int_B|\nabla u|^2|u_\phi|^2\varphi_1^2\,dz + CR_0^{-2}E(u_0).$$

Proof. Similar to the proof of Lemma 4.1, we compute

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\left(\int_{\partial B}|u_\phi|^2\varphi_1^2\,d\phi\right) &= \int_{\partial B}u_\phi\cdot u_{\phi,t}\varphi_1^2\,d\phi = -\int_{\partial B}\partial_\phi(u_\phi\varphi_1^2)\cdot u_t\,d\phi\\ &= \int_{\partial B}\partial_\phi(u_\phi\varphi_1^2)\cdot d\pi_N(u)u_r\,d\phi = -\int_{\partial B}(u_\phi\cdot u_{r\phi} - u_\phi\cdot\partial_\phi(\nu(u)\nu(u)\cdot u_r))\varphi_1^2\,d\phi\\ &= -\frac{1}{2}\int_{\partial B}\partial_r(|u_\phi|^2)\varphi_1^2\,d\phi - \int_{\partial B}u_\phi\cdot d\nu(u)u_\phi\nu(u)\cdot u_r\varphi_1^2\,d\phi.\end{aligned}$$

With $\Delta|u_\phi|^2 = 2|\nabla u_\phi|^2$ we obtain

$$\frac{1}{2}\int_{\partial B}\partial_r(|u_\phi|^2)\varphi_1^2\,d\phi = \int_B|\nabla u_\phi|^2\varphi_1^2\,dz + \int_B\nabla|u_\phi|^2\varphi_1\nabla\varphi_1\,dz,$$

where

$$\left|\int_B\nabla|u_\phi|^2\varphi_1\nabla\varphi_1\,dz\right| \le \frac{1}{4}\int_B|\nabla u_\phi|^2\varphi_1^2\,dz + C\int_B|u_\phi|^2|\nabla\varphi_1|^2\,dz$$

by Young's inequality. Finally, we can bound

$$\begin{aligned}\int_{\partial B}u_r\cdot\nu(u)u_\phi\cdot d\nu(u)u_\phi\varphi_1^2\,d\phi &= \int_B\nabla u\cdot\nabla(\nu(u)u_\phi\cdot d\nu(u)u_\phi\varphi_1^2)\,dz\\ &\le C\int_B(|\nabla u_\phi||\nabla u||u_\phi| + |\nabla u|^2|u_\phi|^2)\varphi_1^2\,dz + C\int_B|\nabla u||\nabla\varphi_1||u_\phi|^2\varphi_1\,dz\\ &\le \frac{1}{4}\int_B|\nabla u_\phi|^2\varphi_1^2\,dz + C\int_B|\nabla u|^2|u_\phi|^2\varphi_1^2\,dz + C\int_B|\nabla u|^2|\nabla\varphi_1|^2\,dz,\end{aligned}$$

and our claim follows. □

We need a substitute for the global bound (4-3). For this, we note that (4-1) also implies the pointwise bound

$$|u_{rr}|^2 \le \frac{2|u_{\phi\phi}|^2}{r^4} + \frac{2|u_r|^2}{r^2};$$

hence we have

$$|\nabla^2 u|^2 \le C(|\nabla u_\phi|^2 + |\nabla u|^2) \quad \text{in } B_{R_0}(z_0)$$

with an absolute constant $C > 0$, uniformly in $z_0 \in \partial B$ and $0 < R_0 < \frac{1}{4}$. By induction then, similarly, we have

$$|\nabla^{k+1} u|^2 \le C(|\nabla^k \partial_\phi u|^2 + |\nabla^k u|^2) \le C \sum_{j=0}^{k} |\nabla \partial_\phi^j u|^2 \quad \text{in } B_{R_0}(z_0) \tag{4-9}$$

with an absolute constant $C = C(k) > 0$, uniformly in $z_0 \in \partial B$ and $0 < R_0 < \frac{1}{4}$ for any $k \in \mathbb{N}$.

Likewise, as a substitute for the global nonconcentration condition (4-4) we now suppose that $z_0 \in \partial B$ is not a concentration point in the sense that for suitably chosen $\delta > 0$ to be determined in the sequel and some $0 < R_0 < \frac{1}{4}$ as above,

$$\sup_{0<t<T_0} \int_{B_{R_0}(z_0)\cap B} |\nabla u(t)|^2 \, dz < \delta. \tag{4-10}$$

We then obtain the following localized version of Proposition 4.2.

Proposition 4.8. *There exist constants $\delta > 0$ and $C > 0$ independent of $R_0 > 0$ such that whenever (4-10) holds then for any $T \le T_0$ we have*

$$\sup_{0<t<T} \int_{\partial B} |u_\phi(t)|^2 \varphi_1^2 \, d\phi + \int_0^T \int_B |\nabla u_\phi|^2 \varphi_1^2 \, dz \, dt \le 2 \int_{\partial B} |u_{0,\phi}|^2 \varphi_1^2 \, d\phi + CT R_0^{-2} E(u_0).$$

Proof. With the help of inequality (A-1) in the Appendix we can bound

$$\int_B |\nabla u|^4 \varphi_1^2 \, dz \le C\delta \int_{B_R(z_i)} |\nabla^2 u|^2 \varphi_1^2 \, dz + C\delta R_0^{-2} \int_{B_R(z_i)} |\nabla u|^2 \, dz.$$

Thus, for sufficiently small $\delta > 0$ our claim follows from Lemma 4.7. □

The next lemma again prepares for a proposition that later will allow us to obtain higher-derivative bounds by induction. Note the differences to Lemma 4.3.

Lemma 4.9. *For any $k \ge 2$, with a constant $C > 0$ depending only on k and N, for the solution $u = u(t)$ to (1-3) and (1-4) for any $0 < t < T_0$,*

$$\frac{d}{dt}(\|\partial_\phi^k u \varphi_k\|^2_{L^2(\partial B)}) + \|\nabla \partial_\phi^k u \varphi_k\|^2_{L^2(B)} \le C \sum_{\substack{1\le j_i \le k \\ \Sigma_i j_i \le 2k+2}} \Big\| \prod_i \nabla^{j_i} u \varphi_k^2 \Big\|_{L^1(B)} + C \sum_{\substack{1\le j_0, j_i \le k \\ \Sigma_{i\ge 0} j_i \le k+1}} \Big\| \prod_{i>0} \nabla^{j_i} u \nabla^{j_0} \varphi_k \Big\|^2_{L^2(B)} + C R_0^{-2k} E(u_0).$$

Proof. Fix $k \geq 2$. With $\Delta|\partial_\phi^k u|^2 = 2|\nabla\partial_\phi^k u|^2$ we compute

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}(\|\partial_\phi^k u\varphi_k\|^2_{L^2(\partial B)}) &= (-1)^k\int_{\partial B}\partial_\phi^k(\partial_\phi^k u\varphi_k^2)\cdot u_t\,d\phi = (-1)^{k+1}\int_{\partial B}\partial_\phi^k(\partial_\phi^k u\varphi_k^2)\cdot(u_r-\nu(u)\nu(u)\cdot u_r)\,d\phi\\ &= -\frac{1}{2}\int_{\partial B}\partial_r(|\partial_\phi^k u|^2)\varphi_k^2\,d\phi + \int_{\partial B}\partial_\phi^k u\cdot\partial_\phi^k(\nu(u)\nu(u)\cdot u_r)\varphi_k^2\,d\phi\\ &= -\int_B|\nabla\partial_\phi^k u|^2\varphi_k^2\,dz - \int_B\nabla(|\partial_\phi^k u|^2)\varphi_k\nabla\varphi_k\,dz + I,\end{aligned}$$

where the term $\int_B\nabla(|\partial_\phi^k u|^2)\varphi_k\nabla\varphi_k\,dz$ can be bounded as claimed. We use the decomposition

$$I = \int_{\partial B}\partial_\phi^k u\cdot\partial_\phi^k(\nu(u)\nu(u)\cdot u_r)\varphi_k^2\,d\phi = \sum_{j=0}^k\binom{k}{j}I_j$$

with

$$I_j = (\partial_\phi^k u\cdot\partial_\phi^j(\nu(u)\nu(u))\varphi_k^2,\,\partial_\phi^{k-j}u_r)_{L^2(\partial B)} = (\nabla(\partial_\phi^k u\cdot\partial_\phi^j(\nu(u)\nu(u))\varphi_k^2),\,\nabla\partial_\phi^{k-j}u)_{L^2(B)},\quad 0\leq j\leq k.$$

For $1\leq j\leq k$ we bound

$$\begin{aligned}|I_j| \leq C\sum_{0\leq i\leq j}&\|\nabla\partial_\phi^k u\varphi_k\|_{L^2(B)}\|\partial_\phi^{j-i}\nu(u)\partial_\phi^i\nu(u)\nabla\partial_\phi^{k-j}u\varphi_k\|_{L^2(B)}\\ &+C\sum_{0\leq i\leq j}\|\partial_\phi^k u\cdot\nabla(\partial_\phi^{j-i}\nu(u)\partial_\phi^i\nu(u)\varphi_k^2)\cdot\nabla\partial_\phi^{k-j}u\|_{L^1(B)}.\end{aligned}$$

By the chain rule then for $1\leq j\leq k$ we have

$$\begin{aligned}|I_j| \leq C\sum_{\substack{1\leq j_i\leq k\\ \Sigma_i j_i=k+1}}&\|\nabla\partial_\phi^k u\varphi_k\|_{L^2(B)}\Big\|\prod_i\nabla^{j_i}u\varphi_k\Big\|_{L^2(B)}\\ &+C\sum_{\substack{1\leq j_i\leq k\\ \Sigma_i j_i=k+2}}\Big\|\partial_\phi^k u\cdot\prod_i\nabla^{j_i}u\varphi_k^2\Big\|_{L^1(B)} + C\sum_{\substack{1\leq j_i\leq k\\ \Sigma_i j_i=k+1}}\Big\|\partial_\phi^k u\cdot\prod_i\nabla^{j_i}u\varphi_k\nabla\varphi_k\Big\|_{L^1(B)}.\end{aligned}$$

By Cauchy–Schwarz and Young's inequalities then we can bound

$$\begin{aligned}\sum_{1\leq j\leq k}|I_j| &\leq \tfrac{1}{4}\|\nabla\partial_\phi^k u\varphi_k\|^2_{L^2(B)} + C\sum_{\substack{1\leq j_i\leq k\\ \Sigma_i j_i=k+1}}\Big\|\prod_i\nabla^{j_i}u\varphi_k\Big\|^2_{L^2(B)}\\ &\qquad + C\sum_{\substack{1\leq j_i\leq k\\ \Sigma_i j_i=2k+2}}\Big\|\prod_i\nabla^{j_i}u\varphi_k^2\Big\|_{L^1(B)} + C\|\partial_\phi^k u\nabla\varphi_k\|^2_{L^2(B)}\\ &\leq \tfrac{1}{4}\|\nabla\partial_\phi^k u\varphi_k\|^2_{L^2(B)} + C\sum_{\substack{1\leq j_i\leq k\\ \Sigma_i j_i=2k+2}}\Big\|\prod_i\nabla^{j_i}u\varphi_k^2\Big\|_{L^1(B)} + C\|\partial_\phi^k u\nabla\varphi_k\|^2_{L^2(B)},\end{aligned}$$

as claimed. Finally, with

$$\nu(u)\cdot u_{\phi r} = (\mathrm{dist}_N(u))_{\phi r} - u_r\cdot d\nu(u)u_\phi$$

as in the proof of Lemma 4.3, for $j=0$ we can write

$$\nu(u)\cdot\partial_\phi^k u_r = \partial_\phi^{k-1}(\nu(u)\cdot u_{\phi r}) + II = \partial_\phi^k(\mathrm{dist}_N(u))_r + III,$$

where the terms in II and III involve products of at least two derivatives of orders between 1 and k of u. Thus we have

$$I_0 = (\partial_\phi^k u \cdot \nu(u)\varphi_k^2, \nu(u) \cdot \partial_\phi^k u_r)_{L^2(\partial B)} = (\partial_\phi^k u \cdot \nu(u)\varphi_k^2, \partial_\phi^k (\mathrm{dist}_N(u))_r)_{L^2(\partial B)} + II_0$$

with a term II_0 that can be dealt with in the same way as the terms $I_j,\ 1 \le j \le k$.

Using the divergence theorem and integrating by parts we can write the leading term as

$$\begin{aligned}\hat{I}_0 &:= (\partial_\phi^k u \cdot \nu(u)\varphi_k^2, \partial_\phi^k (\mathrm{dist}_N(u))_r)_{L^2(\partial B)} \\ &= (\nabla(\partial_\phi^k u \cdot \nu(u)\varphi_k^2), \nabla\partial_\phi^k (\mathrm{dist}_N(u)))_{L^2(B)} + (\partial_\phi^k u \cdot \nu(u)\varphi_k^2, \Delta\partial_\phi^k (\mathrm{dist}_N(u)))_{L^2(B)} \\ &= (\nabla(\partial_\phi^k u \cdot \nu(u)\varphi_k^2), \nabla\partial_\phi^k (\mathrm{dist}_N(u)))_{L^2(B)} - (\partial_\phi(\partial_\phi^k u \cdot \nu(u)\varphi_k^2), \Delta\partial_\phi^{k-1} (\mathrm{dist}_N(u)))_{L^2(B)}\end{aligned}$$

to see that this term may be bounded:

$$|\hat{I}_0| \le C\|(|\nabla\partial_\phi^k u| + |\partial_\phi^k u \nabla u|)\varphi_k + |\partial_\phi^k u \nabla\varphi_k|\|_{L^2(B)} \|\nabla^{k+1}(\mathrm{dist}_N(u))\varphi_k\|_{L^2(B)}.$$

But by elliptic regularity we again have

$$\begin{aligned}\|\nabla^{k+1}(\mathrm{dist}_N(u))\varphi_k\|_{L^2(B)} &\le \|\nabla^{k+1}(\mathrm{dist}_N(u)\varphi_k)\|_{L^2(B)} + C\sum_{1\le j\le k+1} \|\nabla^{k+1-j}(\mathrm{dist}_N(u))\nabla^j\varphi_k\|_{L^2(B)} \\ &\le C\|\Delta(\mathrm{dist}_N(u))\varphi_k\|_{H^{k-1}(B)} + C\sum_{1\le j\le k+1} \|\nabla^{k+1-j}(\mathrm{dist}_N(u))\nabla^j\varphi_k\|_{L^2(B)},\end{aligned}$$

where from (3-5) we can bound the first term on the right:

$$\|\Delta(\mathrm{dist}_N(u))\varphi_k\|_{H^{k-1}(B)} \le \sum_{0\le j<k} \|\nabla^j(\nabla u \cdot d\nu(u)\nabla u\varphi_k)\|_{L^2(B)} \le C \sum_{\substack{0\le j_0<k \\ 1\le j_i\le k \\ \Sigma_{i\ge 0} j_i\le k+1}} \Big\|\prod_i \nabla^{j_i} u \nabla^{j_0}\varphi_k\Big\|_{L^2(B)}.$$

Moreover, using that $\mathrm{dist}_N(u) = 0$ on ∂B, with the help of Poincaré's inequality we find the bound

$$\|\mathrm{dist}_N(u)\nabla^{k+1}\varphi_k\|^2_{L^2(B)} \le CR_k^{-2k}\|\nabla(\mathrm{dist}_N(u))\|^2_{L^2(B_{R_k}(z_0))} \le CR_0^{-2k}E(u).$$

The remaining terms for $1 \le j \le k$ can be estimated as

$$\|\nabla^{k+1-j}(\mathrm{dist}_N(u))\nabla^j\varphi_k\|_{L^2(B)} \le C \sum_{\substack{1\le j_i\le k \\ \Sigma_i j_i = k+1-j}} \Big\|\prod_i \nabla^{j_i} u \nabla^j\varphi_k\Big\|_{L^2(B)}$$

via the chain rule. Thus, finally, we obtain the bound

$$\|\nabla^{k+1}(\mathrm{dist}_N(u))\varphi_k\|^2_{L^2(B)} \le C \sum_{\substack{1\le j_0, j_i\le k \\ \Sigma_{i\ge 0} j_i\le k+1}} \Big\|\prod_{i>0} \nabla^{j_i} u \nabla^{j_0}\varphi_k\Big\|^2_{L^2(B)} + CR_0^{-2k}E(u_0).$$

By Cauchy–Schwarz and Young's inequalities thus we can bound

$$|\hat{I}_0| \le \tfrac{1}{4}\|\nabla\partial_\phi^k u\varphi_k\|^2_{L^2(B)} + C\|\partial_\phi^k u\nabla u\varphi_k\|^2_{L^2(B)} + C \sum_{\substack{1\le j_0, j_i\le k \\ \Sigma_{i\ge 0} j_i\le k+1}} \Big\|\prod_{i>0} \nabla^{j_i} u \nabla^{j_0}\varphi_k\Big\|^2_{L^2(B)} + CR_0^{-2k}E(u_0),$$

and together with our above estimate for the terms $I_j,\ j \ge 1$, our claim follows. □

Proposition 4.10. *There exists a constant $\delta > 0$ independent of $R_0 > 0$ such that whenever (4-10) holds then for any $T \le T_0$ with a constant $C_2 = C_2(T, R, u_0) > 0$ bounded by the terms on the right-hand side in the statement of Proposition 4.8 there holds the estimate*

$$\sup_{0<t<T} \int_{\partial B} |u_{\phi\phi}(t)|^2 \varphi_2^2 \, d\phi + \int_0^T \int_B |\nabla u_{\phi\phi}|^2 \varphi_2^2 \, dz \, dt \le C_2 \int_{\partial B} |u_{0,\phi\phi}|^2 \varphi_2^2 \, d\phi + C_2.$$

Proof. For $k = 2$, with the help of Young's inequality we can bound

$$\begin{aligned} J_1 = \sum_{\substack{1 \le j_i \le k \\ \Sigma_i j_i \le 2k+2}} \Big\| \prod_i \nabla^{j_i} u \varphi_k^2 \Big\|_{L^1(B)} &\le C \| (|\nabla^2 u|^3 + |\nabla^2 u|^2 |\nabla u|^2 + |\nabla^2 u| \, |\nabla u|^4 + |\nabla u|^6 + 1) \varphi_2^2 \|_{L^1(B)} \\ &\le C \| (|\nabla^2 u|^3 + |\nabla u|^6 + 1) \varphi_2^2 \|_{L^1(B)} \end{aligned}$$

and

$$J_2 = \sum_{\substack{1 \le j_0, j_i \le k \\ \Sigma_{i \ge 0} j_i \le k+1}} \Big\| \prod_{i>0} \nabla^{j_i} u \nabla^{j_0} \varphi_2 \Big\|_{L^2(B)}^2 \le C \| (|\nabla^2 u|^2 + |\nabla u|^4 + 1) |\nabla \varphi_2|^2 + (|\nabla u|^2 + 1) |\nabla^2 \varphi_2|^2 \|_{L^1(B)}.$$

Observing that $\varphi_1 = 1$ on the support of φ_2, by (A-2) for the first term in J_1 we have

$$\begin{aligned} \| |\nabla^2 u|^3 \varphi_2^2 \|_{L^1(B)} &\le \| \nabla^2 u \varphi_2 \|_{L^4(B)}^2 \| \nabla^2 u \varphi_1 \|_{L^2(B)} \le C \| \nabla^2 u \varphi_2 \|_{H^1(B)} \| \nabla^2 u \varphi_2 \|_{L^2(B)} \| \nabla^2 u \varphi_1 \|_{L^2(B)} \\ &\le C (\| \nabla^3 u \varphi_2 \|_{L^2(B)} + \| \nabla^2 u \varphi_1 \|_{L^2(B)}) \| \nabla^2 u \varphi_2 \|_{L^2(B)} \| \nabla^2 u \varphi_1 \|_{L^2(B)}. \end{aligned}$$

Moreover, arguing as in (A-1) for the function $|\nabla u|^6 \varphi_2^2$ in place of $|v|^4 \varphi^2$, we can bound

$$\begin{aligned} \int_B |\nabla u|^6 \varphi_2^2 \, dz &\le C \Big(\int_B (|\nabla^2 u| \, |\nabla u|^2 \varphi_2 + |\nabla u|^3 |\nabla \varphi_2|) \, dz \Big)^2 \\ &\le C \Big(\int_B |\nabla^2 u|^3 \varphi_2^2 \, dz \Big)^{2/3} \Big(\int_B |\nabla u|^3 \varphi_2^{1/2} \, dz \Big)^{4/3} + C \Big(\int_B |\nabla u|^3 |\nabla \varphi_2| \, dz \Big)^2, \end{aligned}$$

where by Hölder's inequality we have

$$\int_B |\nabla u|^3 \varphi_2^{1/2} \, dz \le \Big(\int_B |\nabla u|^6 \varphi_2^2 \, dz \Big)^{1/4} \Big(\int_B |\nabla u|^2 \varphi_1^2 \, dz \Big)^{3/4},$$

so that with Young's inequality we obtain

$$\begin{aligned} \int_B |\nabla u|^6 \varphi_2^2 \, dz &\le C\delta \Big(\int_B |\nabla^2 u|^3 \varphi_2^2 \, dz \Big)^{2/3} \Big(\int_B |\nabla u|^6 \varphi_2^2 \, dz \Big)^{1/3} + C \Big(\int_B |\nabla u|^3 |\nabla \varphi_2| \, dz \Big)^2 \\ &\le \frac{1}{2} \int_B |\nabla u|^6 \varphi_2^2 \, dz + C \int_B |\nabla^2 u|^3 \varphi_2^2 \, dz + C \Big(\int_B |\nabla u|^3 |\nabla \varphi_2| \, dz \Big)^2. \end{aligned}$$

With Young's inequality for suitable $\varepsilon > 0$, and using (4-9), we then can bound

$$\begin{aligned} J_1 &\le C \| (|\nabla^2 u|^3 + 1) \varphi_2^2 \|_{L^1(B)} + C \| |\nabla u|^3 |\nabla \varphi_2| \|_{L^1(B)}^2 \\ &\le \varepsilon \| \nabla^3 u \varphi_2 \|_{L^2(B)}^2 + C (1 + \| \nabla^2 u \varphi_2 \|_{L^2(B)}^2) \| \nabla^2 u \varphi_1 \|_{L^2(B)}^2 + C \| |\nabla u|^3 |\nabla \varphi_2| \|_{L^1(B)}^2 \\ &\le \tfrac{1}{2} \| \nabla \partial_\phi^2 u \varphi_2 \|_{L^2(B)}^2 + C (1 + \| \nabla \partial_\phi u \varphi_2 \|_{L^2(B)}^2) \| \nabla \partial_\phi u \varphi_1 \|_{L^2(B)}^2 + C, \end{aligned}$$

where we also have estimated

$$\begin{aligned}\||\nabla u|^3|\nabla\varphi_2|\|^2_{L^1(B)} &\le C\|\nabla u\varphi_1\|^4_{L^4(B)}\|\nabla u\varphi_1\|^2_{L^2(B)}\\ &\le C(\|\nabla^2 u\varphi_1\|^2_{L^2(B)}+E(u))\|\nabla u\varphi_1\|^4_{L^2(B)} \le C\|\nabla\partial_\phi u\varphi_1\|^2_{L^2(B)}+C.\end{aligned}$$

Similarly, with (A-2) we have

$$J_2 \le C\|\nabla^2 u\varphi_1\|^2_{L^2(B)}+C.$$

Thus, from Lemma 4.9 we obtain

$$\frac{d}{dt}(\|\partial^2_\phi u\varphi_2\|^2_{L^2(\partial B)})+\tfrac{1}{2}\|\nabla\partial^2_\phi u\varphi_2\|^2_{L^2(B)} \le C(1+\|\nabla\partial_\phi u\varphi_2\|^2_{L^2(B)})\|\nabla\partial_\phi u\varphi_1\|^2_{L^2(B)}+C. \tag{4-11}$$

Denote by $C_1 = C_1(T, R, u_0) > 0$ a constant bounded by the terms on the right-hand side in the statement of Proposition 4.8. By elliptic regularity, using that $|\Delta(u\varphi_2)| \le 2|\nabla u\nabla\varphi_2|+C$ we can bound

$$\begin{aligned}\|\nabla^2 u\varphi_2\|^2_{L^2(B)} &\le \|u\varphi_2\|^2_{H^2(B)}+C\|\nabla u\nabla\varphi_2\|^2_{L^2(B)}+C\\ &\le C\|u\varphi_2\|^2_{H^2(\partial B)}+\|\Delta(u\varphi_2)\|^2_{L^2(B)}+C\|\nabla u\nabla\varphi_2\|^2_{L^2(B)}+C\\ &\le C\|\partial^2_\phi u\varphi_2\|^2_{L^2(\partial B)}+CE(u)+C_1.\end{aligned}$$

From (4-11) we then obtain the differential inequality

$$\frac{d}{dt}(1+\|\partial^2_\phi u\varphi_2\|^2_{L^2(\partial B)}) \le C_1(1+\|\partial^2_\phi u\varphi_2\|^2_{L^2(\partial B)})\|\nabla\partial_\phi u\varphi_1\|^2_{L^2(B)}+C_1;$$

that is,

$$\frac{d}{dt}(\log(1+\|\partial^2_\phi u\varphi_2\|^2_{L^2(\partial B)})) \le C_1\|\nabla\partial_\phi u\varphi_1\|^2_{L^2(B)}+C_1,$$

and the right-hand side is integrable in time by Proposition 4.8. The claim follows. □

We continue by induction.

Proposition 4.11. *There exists a constant $\delta > 0$ independent of $R_0 > 0$ with the following property. Whenever (4-10) holds, then, for any $k \ge 3$, any smooth $u_0 \in H^{1/2}(S^1; N)$, and any $T \le T_0$,*

$$\sup_{0<t<T}\int_{\partial B}|\partial^k_\phi u(t)|^2\varphi_k^2\,d\phi+\int_0^T\int_B|\nabla\partial^k_\phi u|^2\varphi_k^2\,dz\,dt \le C_k\int_{\partial B}|\partial^k_\phi u_0|^2\varphi_k^2\,d\phi+C_k,$$

where we denote by $C_k = C_k(T, R, u_0) > 0$ a constant bounded by the terms on the right-hand side in the statement of the proposition for $k-1$.

Proof. By Proposition 4.10 the claimed result holds true for $k = 2$. Suppose the claim holds true for some $k_0 \ge 2$ and let $k = k_0+1$. Note that by elliptic regularity, as in the proof of Proposition 4.10, we can bound

$$\begin{aligned}\|\nabla^k u\varphi_k\|^2_{L^2(B)} &\le \|u\varphi_k\|^2_{H^k(B)}+C\sum_{j<k}\|\nabla^j u\nabla^{k-j}\varphi_k\|^2_{L^2(B)}\\ &\le C\|u\varphi_k\|^2_{H^k(\partial B)}+C\|\Delta(u\varphi_k)\|^2_{H^{k-2}(B)}+C\sum_{j<k}\|\nabla^j u\nabla^{k-j}\varphi_k\|^2_{L^2(B)}\\ &\le C\|\partial^k_\phi u\varphi_k\|^2_{L^2(\partial B)}+C\sum_{j<k}\|\nabla^j u\nabla^{k-j}\varphi_k\|^2_{L^2(B)}+C_k.\end{aligned}$$

By the induction hypothesis and Sobolev's embedding $H^2(B) \hookrightarrow W^{1,4} \cap C^0(\bar{B})$ for $0 \le t < T$, we then have the uniform bounds

$$\|\nabla^{k_0} u\varphi_{k_0}\|^2_{L^2(B)} + \|\nabla^{k_0-1} u\varphi_{k_0}\|^2_{L^4(B)} + \sum_{j=1}^{k_0-2} \|\nabla^j u\varphi_{k_0}\|^2_{L^\infty(B)} \le C_k,$$

and it follows that

$$\|\nabla^k u\varphi_k\|^2_{L^2(B)} + \|\nabla^{k_0} u\varphi_k\|^2_{L^4(B)} + \|\nabla^{k_0-1} u\varphi_k\|^2_{L^\infty(B)} \le C\|\partial^k_\phi u\varphi_k\|^2_{L^2(\partial B)} + C_k.$$

Again let

$$\begin{aligned} J_1 &:= \sum_{\substack{1\le j_i\le k \\ \Sigma_i j_i = 2k+2}} \Big\|\prod_i \nabla^{j_i} u\varphi_k^2\Big\|_{L^1(B)} \\ &\le \|(|\nabla^k u|^2(|\nabla^2 u| + |\nabla u|^2) + |\nabla^k u|\,|\nabla^{k_0} u|\,|\nabla^3 u| + \cdots + |\nabla u|^{2k+2})\varphi_k^2\|_{L^1(B)} \end{aligned}$$

and set

$$J_2 = \sum_{\substack{1\le j_0, j_i\le k \\ \Sigma_{i\ge 0} j_i \le k+1}} \Big\|\prod_{i>0} \nabla^{j_i} u\nabla^{j_0}\varphi_k\Big\|^2_{L^2(B)}.$$

Suppose $k_0 = 2$. Recalling that $\varphi_k = \varphi_k\varphi_{k_0}$, we can bound the terms

$$\begin{aligned} \||\nabla^3 u|^2(|\nabla^2 u| + |\nabla u|^2)\varphi_3^2\|_{L^1(B)} &\le \|\nabla^3 u\varphi_3\|^2_{L^4(B)}(\|\nabla^2 u\varphi_2\|_{L^2(B)} + \|\nabla u\varphi_2\|^2_{L^4(B)}) \\ &\le C_3\|\nabla\partial^3_\phi u\varphi_3\|_{L^2(B)}\|\nabla^3 u\varphi_3\|_{L^2(B)} + C_3\|\nabla^3 u\varphi_2\|^2_{L^2(B)} + C_3 \\ &\le C_3\|\nabla\partial^3_\phi u\varphi_3\|_{L^2(B)}(\|\partial^3_\phi u\varphi_3\|_{L^2(\partial B)} + 1) + C_3\|\nabla\partial^2_\phi u\varphi_2\|^2_{L^2(B)} + C_3 \\ &\le \varepsilon\|\nabla\partial^3_\phi u\varphi_3\|^2_{L^2(B)} + C_3\|\partial^3_\phi u\varphi_3\|^2_{L^2(\partial B)} + C_3\|\nabla\partial^2_\phi u\varphi_2\|^2_{L^2(B)} + C_3 \end{aligned}$$

and

$$\||\nabla u|^8\varphi_3^2\|_{L^1(B)} \le \|\nabla u\varphi_3\|^2_{L^\infty(B)}\|\nabla u\varphi_2\|^6_{L^6(B)} \le C_3\|\partial^3_\phi u\varphi_3\|^2_{L^2(\partial B)} + C_3$$

from the estimate of J_1. Here we also have used (A-1) and (A-2) to bound

$$\begin{aligned} \|\nabla u\varphi_2\|^3_{L^6(B)} \le \|\nabla(|\nabla u|^3\varphi_2^3)\|_{L^1(B)} &\le C\|(|\nabla^2 u|\varphi_2 + |\nabla u|\,|\nabla\varphi_2|)|\nabla u|^2\varphi_2^2\|_{L^1(B)} \\ &\le C(\|\nabla^2 u\varphi_2\|_{L^2(B)} + \|\nabla u\nabla\varphi_2\|_{L^2(B)})\|\nabla u\varphi_2\|^2_{L^4(B)} \\ &\le C(\|\nabla^2 u\varphi_2\|_{L^2(B)} + \|\nabla u\nabla\varphi_2\|_{L^2(B)})^2\|\nabla u\varphi_2\|_{L^2(B)} \le C_3. \end{aligned}$$

Similarly, we can bound the remaining terms and the terms in J_2 to obtain

$$\frac{d}{dt}(\|\partial^3_\phi u\varphi_3\|^2_{L^2(\partial B)}) + \tfrac{1}{2}\|\nabla\partial^3_\phi u\varphi_3\|^2_{L^2(B)} \le C_3(1 + \|\partial^3_\phi u\varphi_3\|^2_{L^2(\partial B)})(1 + \|\nabla\partial^2_\phi u\varphi_2\|^2_{L^2(B)}) + C_3$$

from Lemma 4.9 and then

$$\frac{d}{dt}(\log(1 + \|\partial^3_\phi u\varphi_2\|^2_{L^2(\partial B)})) \le C_3(1 + \|\nabla\partial^2_\phi u\varphi_2\|^2_{L^2(B)}),$$

where the right-hand side is integrable in time by Proposition 4.10. The claim for $k = 3$ thus follows.

For $k \ge 4$ the analysis is similar (but simpler) and may be left to the reader. □

5. Local existence

In order to show local existence we approximate the flow equation (1-3) by the equation

$$u_t = -(\varepsilon + d\pi_N(u))u_r \quad \text{on } \partial B, \tag{5-1}$$

where $\varepsilon > 0$ and where we smoothly extend the nearest-neighbor projection π_N, originally defined only in the ρ-neighborhood N_ρ of N, to the whole ambient $\mathbb{R}^n$. Our aim then is to show that for given smooth initial data u_0 the evolution problem (5-1), (1-4) admits a smooth solution u_ε which remains uniformly smoothly bounded on a uniform time interval as $\varepsilon \downarrow 0$. Fixing some $0 < \varepsilon < \frac{1}{2}$, we show existence for the problem (5-1) with data (1-4) by means of a fixed-point argument.

To set up the argument, fix smooth initial data $u_0 \colon S^1 \to N$ with harmonic extension $u_0 \in C^\infty(\bar{B}; \mathbb{R}^n)$ and some $k \geq 2$. For suitable $T > 0$ to be determined let

$$X = L^\infty([0, T]; H^{k+1}(B; \mathbb{R}^n)) \cap H^1(S^1 \times [0, T]; \mathbb{R}^n)$$

and set

$$V = \Big\{ v \in X : v(0) = u_0,\ \Delta v(t) = 0 \text{ in } B \text{ for } 0 \leq t \leq T, \\ \|v\|_X^2 = \sup_{0\leq t\leq T} \|v(t)\|^2_{H^{k+1}(B)} + \int_0^T \int_{S^1} |v_t|^2 \, d\phi \, dt \leq 4R_0^2 \Big\},$$

where $R_0 = \|u_0\|_{H^{k+1}(B)}$. We endow the space V with the metric derived from the seminorm

$$|v|_X^2 = \sup_{0\leq t\leq T} \|\nabla v(t)\|^2_{L^2(B)} + \int_0^T \int_{S^1} |v_t|^2 \, d\phi \, dt.$$

Note that this metric is positive definite on V in view of the initial condition that we impose.

Lemma 5.1. *V is a complete metric space.*

Proof. Let $(v_m)_{m\in\mathbb{N}} \subset V$ with $|v_l - v_m|_X \to 0$ $(l, m \to \infty)$. By the theorem of Banach–Alaoglu a subsequence $v_m \rightharpoonup v$ weakly-$*$ in $L^\infty([0, T]; H^{k+1}(B))$ with $v_{m,t} \to v_t$ weakly in $L^2([0, T] \times S^1)$, and by weak lower semicontinuity of the norm

$$\|v\|_X^2 \leq \limsup_{m\to\infty} \|v_m\|_X^2 \leq 4R_0^2.$$

Moreover, we have $\Delta v(t) = 0$ for all $0 \leq t \leq T$ and $v(0) = u_0$ by compactness of the trace operator $H^1(S^1 \times [0, T]) \ni u \mapsto u(0) \in L^2(S^1)$. Hence $v \in V$.

Moreover, we have

$$|v_l - v|_X \leq \limsup_{m\to\infty} |v_l - v_m|_X \to 0 \quad \text{as } l \to \infty. \qquad \square$$

Lemma 5.2. *There is $T_2 > 0$ such that for any $T \leq T_2$ and any $v \in V$ there is a solution $u = \Phi(v) \in V$ of the equation*

$$u_t = -(\varepsilon + d\pi_N(v))u_r \quad \text{on } \partial B \times [0, T_2[\tag{5-2}$$

satisfying (1-4).

Proof. For $v \in V$ we construct a solution $u = \Phi(v) \in X$ of (5-2) via Galerkin approximation. For this let $(\varphi_l)_{l\in\mathbb{N}_0}$ be Steklov eigenfunctions of the Laplacian, satisfying

$$\Delta\varphi_l = 0 \quad \text{in } B$$

with boundary condition

$$\partial_r\varphi_l = \lambda_l\varphi_l \quad \text{on } \partial B, \qquad l \in \mathbb{N}_0.$$

Note that the Steklov eigenvalues are given by $\lambda_0 = 0$ and $\lambda_{2l-1} = \lambda_{2l} = l$, $l \in \mathbb{N}$. In fact, we may choose $\varphi_0 \equiv 1/\sqrt{2\pi}$ and

$$\varphi_{2l-1}(re^{i\theta}) = \frac{1}{\sqrt{\pi}} r^l \sin(l\theta), \quad \varphi_{2l}(re^{i\theta}) = \frac{1}{\sqrt{\pi}} r^l \cos(l\theta), \qquad l \in \mathbb{N} \tag{5-3}$$

to obtain an orthonormal basis for $L^2(S^1)$ consisting of these functions. Given $m \in \mathbb{N}$ then let

$$u^{(m)}(t,z) = \sum_{l=0}^{m} a_l^{(m)}(t)\varphi_l(z)$$

solve the system of equations

$$\begin{aligned}\partial_t a_l^{(m)} = (\varphi_l, u_t^{(m)})_{L^2(S^1)} &= -(\varphi_l, (\varepsilon + d\pi_N(v))u_r^{(m)})_{L^2(S^1)} \\ &= -\sum_{j=0}^{m} a_j^{(m)}\lambda_j(\varphi_l, (\varepsilon + d\pi_N(v))\varphi_j)_{L^2(S^1)}, \quad 0 \le l \le m.\end{aligned} \tag{5-4}$$

Since for any $m \in \mathbb{N}$ the coefficients $\lambda_j(\varphi_l, (\varepsilon + d\pi_N(v))\varphi_j)_{L^2(S^1)}$ of this system are uniformly bounded for any $v \in V$, for any $m \in \mathbb{N}$ there exists a unique solution $a^{(m)} = (a_l^{(m)})_{0\le l\le m}$ of (5-4) on $[0, T]$ with initial data $a_l^{(m)}(0) = a_{l0} = (u_0, \varphi_l)_{L^2(S^1)}$, $0 \le l \le m$.

Note that for any $m \in \mathbb{N}$ and any $j \in \mathbb{N}_0$ we have

$$\partial_\phi^{2j}(ru_r^{(m)}) \in \operatorname{span}\{\varphi_l : 0 \le l \le m\},$$

and the function $\partial_\phi^{2j}u^{(m)}$ is harmonic. In particular, for $j = 0$ we obtain

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}(\|\nabla u^{(m)}\|^2_{L^2(B)}) &= \int_B \nabla u^{(m)}\nabla u_t^{(m)}\,dz = (u_r^{(m)}, u_t^{(m)})_{L^2(S^1)} = -(u_r^{(m)}, (\varepsilon + d\pi_N(v))u_r^{(m)})_{L^2(S^1)} \\ &= -\varepsilon\|u_r^{(m)}\|^2_{L^2(S^1)} - \|d\pi_N(v)u_r^{(m)}\|^2_{L^2(S^1)} \le -\tfrac{1}{2}\|u_t^{(m)}\|^2_{L^2(S^1)} \le 0,\end{aligned} \tag{5-5}$$

and we find the uniform H^1-bound

$$\begin{aligned}\sup_{t\ge 0}\|\nabla u^{(m)}(t)\|^2_{L^2(B)} + \varepsilon\|u_r^{(m)}\|^2_{L^2([0,\infty[\times S^1)} + \|u_t^{(m)}\|^2_{L^2([0,\infty[\times S^1)} &\le 2\|\nabla u^{(m)}(0)\|^2_{L^2(B)} \\ &\le 2\|\nabla u_0\|^2_{L^2(B)} \le 2R_0^2.\end{aligned} \tag{5-6}$$

Moreover, for $j = k \in \mathbb{N}$ as in the definition of X, upon integrating by parts we find

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}(\|\nabla\partial_\phi^k u^{(m)}\|^2_{L^2(B)}) &= (-1)^k\int_B \nabla\partial_\phi^{2k}u^{(m)}\nabla u_t^{(m)}\,dz = (-1)^k(\partial_\phi^{2k}u_r^{(m)}, u_t^{(m)})_{L^2(S^1)} \\ &= (-1)^{k+1}(\partial_\phi^{2k}u_r^{(m)}, (\varepsilon + d\pi_N(v))u_r^{(m)})_{L^2(S^1)} \\ &= -\varepsilon\|\partial_\phi^k u_r^{(m)}\|^2_{L^2(S^1)} - \|d\pi_N(v)\partial_\phi^k u_r^{(m)}\|^2_{L^2(S^1)} + I,\end{aligned} \tag{5-7}$$

where $I = \sum_{j=1}^{k} \binom{k}{j} I_j$ with

$$I_j = -(\partial_\phi^k u_r^{(m)}, \partial_\phi^j (d\pi_N(v))\partial_\phi^{k-j} u_r^{(m)})_{L^2(S^1)}$$

similar to the proof of Lemma 4.3. However, now we simply bound

$$|I_j| \le C \sum_{\Sigma_i j_i = j} \|\partial_\phi^k u_r^{(m)}\|_{L^2(S^1)} \Big\| \prod_i \partial_\phi^{j_i} v \partial_\phi^{k-j} u_r^{(m)} \Big\|_{L^2(S^1)}, \quad 1 \le j \le k.$$

Note that by compactness of Sobolev's embedding $H^1(S^1) \hookrightarrow L^\infty(S^1)$ and Ehrlich's lemma for any number $1 \le j \le k$, any $\delta > 0$ we can bound

$$\begin{aligned}\|\partial_\phi^{k-j} u_r^{(m)}\|_{L^\infty(S^1)} &\le \delta \|\partial_\phi^{k-j+1} u_r^{(m)}\|_{L^2(S^1)} + C(\delta)\|\partial_\phi^{k-j} u_r^{(m)}\|_{L^2(S^1)} \\ &\le 2\delta \|\partial_\phi^k u_r^{(m)}\|_{L^2(S^1)} + C(\delta)\|u_r^{(m)}\|_{L^2(S^1)}.\end{aligned}$$

On the other hand, for any $v \in V$ by the trace theorem we have

$$\|\partial_\phi^k v\|_{L^2(S^1)} \le C\|\partial_\phi^k v\|_{H^1(B)} \le C\|v\|_{H^{k+1}(B)} \le CR_0,$$

and we therefore also can bound

$$\|\partial_\phi^j v\|_{L^\infty(S^1)} \le C\|\partial_\phi^k v\|_{L^2(S^1)} + \|\partial_\phi^j v\|_{L^2(S^1)} \le C\|v\|_{H^{k+1}(B)} \le CR_0$$

for any $1 \le j < k$.

Thus, for sufficiently small $\delta > 0$ with a constant $C > 0$ depending on $\varepsilon > 0$ and R_0,

$$|I| \le \tfrac{1}{2}\varepsilon \|\partial_\phi^k u_r^{(m)}\|_{L^2(S^1)}^2 + C\|u_r^{(m)}\|_{L^2(S^1)}^2,$$

and from (5-7) with the help of (3-7) we obtain the inequality

$$\begin{aligned}\frac{d}{dt}(\|\nabla \partial_\phi^k u^{(m)}\|_{L^2(B)}^2) &\le C\|u_r^{(m)}\|_{L^2(S^1)}^2 = C\|u_\phi^{(m)}\|_{L^2(S^1)}^2 \le C\|u_\phi^{(m)}\|_{H^1(B)}^2 \\ &\le C\|\nabla \partial_\phi^k u^{(m)}\|_{L^2(B)}^2 + C\|\nabla u^{(m)}\|_{L^2(B)}^2 \le C(1 + \|\nabla \partial_\phi^k u^{(m)}\|_{L^2(B)}^2),\end{aligned}$$

where we recall (5-6) for the last conclusion.

It follows that for suitably small $T > 0$ there holds $\|u^{(m)}\|_X^2 \le 4R_0^2$ for all $m \in \mathbb{N}$. Thus, there is a sequence $m \to \infty$ such that $u^{(m)} \rightharpoonup u$ weakly-$*$ in $L^\infty([0, T]; H^{k+1}(B))$ with $u_t^{(m)} \rightharpoonup u_t$ weakly in $L^2([0, T] \times S^1)$, where $u =: \Phi(v) \in V$ solves (5-2). □

Lemma 5.3. *There is $T > 0$ such that, for $v_1, v_2 \in V$,*

$$|\Phi(v_1) - \Phi(v_2)|_X \le \tfrac{1}{2}|v_1 - v_2|_X.$$

Proof. Let $T_2 > 0$ be as determined in Lemma 5.2 and fix some $0 < T \le T_2$. For $v_1, v_2 \in V$ then we have $u_i =: \Phi(v_i) \in V,\ i = 1, 2$. Set $w = u_1 - u_2$ and $v = v_1 - v_2$, and compute

$$w_t = -(\varepsilon + d\pi_N(v_1))w_r - (d\pi_N(v_1) - d\pi_N(v_2))u_{2,r} \quad \text{on } \partial B = S^1. \tag{5-8}$$

Multiplying with w_r and integrating we obtain

$$\frac{1}{2}\frac{d}{dt}(\|\nabla w\|^2_{L^2(B)}) = \int_B \nabla w \nabla w_t\,dx = (w_r, w_t)_{L^2(S^1)}$$
$$= -\varepsilon\|w_r\|^2_{L^2(S^1)} - \|d\pi_N(v_1)w_r\|^2_{L^2(S^1)} - (w_r, (d\pi_N(v_1) - d\pi_N(v_2))u_{2,r})_{L^2(S^1)},$$

where with $\|u_{2,r}\|_{L^\infty(S^1)} \le C\|u_2\|_{H^3(B)} \le CR_0$ we can bound

$$|(w_r, (d\pi_N(v_1) - d\pi_N(v_2))u_{2,r})_{L^2(S^1)}| \le C\|w_r\|_{L^2(S^1)}\|v\|_{L^2(S^1)}\|u_{2,r}\|_{L^\infty(S^1)}$$
$$\le C\|w_r\|_{L^2(S^1)}\|v\|_{L^2(S^1)} \le \tfrac{1}{2}\varepsilon\|w_r\|^2_{L^2(S^1)} + C(\varepsilon)\|v\|^2_{L^2(S^1)}.$$

Thus, with a constant $C = C(\varepsilon) > 0$ we find

$$\frac{d}{dt}\|\nabla w\|^2_{L^2(B)} + \varepsilon\|w_r\|^2_{L^2(S^1)} \le C\|v\|^2_{L^2(S^1)}. \tag{5-9}$$

Similarly, from (5-8) we can bound

$$\|w_t\|^2_{L^2(S^1)} \le C\|w_r\|^2_{L^2(S^1)} + C\|v\|^2_{L^2(S^1)}. \tag{5-10}$$

Integrating over $0 \le t \le T$ and observing that we have

$$\sup_{0\le t\le T}\|v(t)\|^2_{L^2(S^1)} \le \left(\int_0^T \|v_t(t)\|_{L^2(S^1)}\,dt\right)^2 \le T\int_0^T \|v_t(t)\|^2_{L^2(S^1)}\,dt,$$

from (5-9) we first obtain

$$\sup_{0\le t\le T}\|\nabla w(t)\|^2_{L^2(B)} + \varepsilon\|w_r\|^2_{L^2([0,T]\times S^1)} \le CT\sup_{0\le t\le T}\|v(t)\|^2_{L^2(S^1)} \le CT^2|v|^2_X,$$

which we may use together with (5-10) to bound

$$|w|^2_X = \sup_{0\le t\le T}\|\nabla w(t)\|^2_{L^2(B)} + \|w_t\|^2_{L^2([0,T]\times S^1)} \le CT^2|v|^2_X.$$

For sufficiently small $T > 0$ then our claim follows. □

Thus, by Banach's fixed-point theorem, for any $\varepsilon > 0$ and any smooth $u_0 \in H^{1/2}(S^1; N)$, there exists $T > 0$ and a solution $u = u(t) \in V$ of the initial value problem (5-1), (1-4). We now show that the number $T > 0$ may be chosen uniformly as $\varepsilon \downarrow 0$. Indeed, we have the following result.

Lemma 5.4. *There exists a constant $C > 0$ such that, for any $k \ge 2$, any smooth $u_0 \in H^{1/2}(S^1; N)$, and any $0 < \varepsilon \le \frac{1}{2}$ for the solution u to (5-1) with $u(0) = u_0$,*

$$\frac{d}{dt}(\|\nabla\partial^k_\phi u\|^2_{L^2(B)}) \le C(1 + \|\nabla u\|^2_{L^2(B)} + \|\nabla\partial^k_\phi u\|_{L^2(B)})^{k+3}.$$

Proof. Similar to the proof of Lemma 5.2, for given $2 \le k \in \mathbb{N}$ we compute

$$\frac{1}{2}\frac{d}{dt}(\|\nabla\partial^k_\phi u\|^2_{L^2(B)}) = (-1)^k\int_B \nabla\partial^{2k}_\phi u\nabla u_t\,dx = (-1)^k(\partial^{2k}_\phi u_r, u_t)_{L^2(S^1)}$$
$$= (-1)^{k+1}(\partial^{2k}_\phi u_r, (\varepsilon + d\pi_N(u))u_r)_{L^2(S^1)}$$
$$\le -\|d\pi_N(u)\partial^k_\phi u_r\|^2_{L^2(S^1)} - I, \tag{5-11}$$

where we now drop the term $\varepsilon\|\partial_\phi^k u_r\|^2_{L^2(S^1)}$ from (5-7). Again we use the decomposition $I = \sum_{j=1}^k \binom{k}{j} I_j$ with

$$I_j = (\partial_\phi^k u_r, \partial_\phi^j (d\pi_N(u))\partial_\phi^{k-j} u_r)_{L^2(S^1)} = (\nabla\partial_\phi^k u, \nabla(\partial_\phi^j (d\pi_N(u))\partial_\phi^{k-j} u_r))_{L^2(B)},$$

but now we bound these terms as in the proof of Lemma 4.3 via

$$\begin{aligned} |I_j| &\le C\|\nabla\partial_\phi^k u\|_{L^2(B)}(\|\nabla\partial_\phi^j (d\pi_N(u))\partial_\phi^{k-j} u_r\|_{L^2(B)} + \|\partial_\phi^j (d\pi_N(u))\nabla\partial_\phi^{k-j} u_r\|_{L^2(B)}) \\ &\le C \sum_{\substack{1\le j_i\le k+1 \\ \Sigma_i j_i = k+2}} \|\nabla\partial_\phi^k u\|_{L^2(B)} \Big\|\prod_i \nabla^{j_i} u\Big\|_{L^2(B)}. \end{aligned}$$

Using that for any $k \ge 2$ by Sobolev's embedding $H^2(B) \hookrightarrow W^{1,4} \cap C^0(\bar{B})$ we can bound

$$\sum_{\substack{1\le j_i\le k+1 \\ \Sigma_i j_i = k+2}} \Big\|\prod_i \nabla^{j_i} u\Big\|_{L^2(B)} \le C(1 + \|\nabla u\|_{L^2(B)} + \|\nabla^{k+1} u\|_{L^2(B)})^{k+2}$$

and also using (4-3), we obtain the claim. □

We now are able to conclude.

Proposition 5.5. *For any $k \ge 2$ and any smooth $u_0 \in H^{1/2}(S^1; N)$ there exist $T > 0$ and a solution $u \in V$ to (1-3) on $[0, T]$ with initial data $u(0) = u_0$.*

Proof. In view of Lemma 5.4, there exists a uniform number $T > 0$ such that, with V as defined above, for any $0 < \varepsilon \le \frac{1}{2}$ there exists a solution $u_\varepsilon \in V$ to (5-1) on $[0, T]$. By definition of V, as $\varepsilon \downarrow 0$ suitably, we have $u_\varepsilon \to u$ weakly-$*$ in $L^\infty([0, T]; H^{k+1}(B)) \cap H^1(S^1 \times [0, T])$. But this suffices to pass to the limit $\varepsilon \downarrow 0$ in (5-1), and $u \in V$ solves (1-3) with $u(0) = u_0$. □

Proof of Theorem 1.1(i). By Proposition 5.5 for any smooth $u_0 \in H^{1/2}(S^1; N)$ and any $k \ge 2$ there exists $T > 0$ and a solution $u \in V$ of (1-3), (1-4) for $0 < t < T$. Alternatingly employing Propositions 4.11 and 4.6, we then obtain smoothness of u for $0 < t \le T$, including the final time T. (This argument later appears in more detail in Section 6 after Lemma 6.2.) Iterating, the solution u may be extended smoothly until some maximal time T_0 where condition (4-4) ceases to hold. Uniqueness (even within a much larger class of competing functions) is established in Section 7. □

6. Weak solutions

Given $u_0 \in H^{1/2}(S^1; N)$, there are smooth functions $u_{0k} \in H^{1/2}(S^1; N)$ with $u_{0k} \to u_0$ in $H^1(B)$ as $k \to \infty$. Indeed, similar to an argument of Schoen and Uhlenbeck [1982, Theorem 3.1], with a standard mollifying sequence $(\rho_k)_{k\in\mathbb{N}}$ for the mollified functions $v_{0k} := u_0 * \rho_k$ we have $\mathrm{dist}_N(v_{0k}) \to 0$ uniformly, and $u_{0k} := \pi_N(v_{0k}) \to u_0 \in H^{1/2}(S^1; N)$ as $k \to \infty$.

Let u_k be the corresponding solutions of (1-4) with initial data $u_k(0) = u_{0k}$, defined on a maximal time interval $[0, T_k[$, $k \in \mathbb{N}$. We claim that each function u_k can be smoothly extended to a uniform time interval $[0, T[$ for some $T > 0$. To see this, we first establish the following nonconcentration result.

Lemma 6.1. *For any $\delta > 0$ there exists a number $R > 0$ and a time $T_0 > 0$ such that*

$$\sup_{\substack{z_0 \in B\\ 0<t<T_0}} \int_{B_R(z_0)\cap B} |\nabla u_k(t)|^2\,dz < \delta \quad \textit{for all } k \in \mathbb{N}.$$

Proof. Given $\delta > 0$, by absolute continuity of the Lebesgue integral and H^1-convergence $u_{0k} \to u_0$ $(k \to \infty)$ we can find $R > 0$ such that

$$\sup_{z_0 \in B} \int_{B_{2R}(z_0)\cap B} |\nabla u_{0k}|^2\,dz < \delta \quad \text{for all } k \in \mathbb{N}.$$

Choosing $T_0 = \delta R$, by Lemma 2.2 then we have

$$\sup_{\substack{z_0 \in B\\ 0<t<T_0}} \int_{B_R(z_0)\cap B} |\nabla u_k(t)|^2\,dz < 4\delta + C\delta E(u_{k0}) < L\delta$$

with a uniform constant $L > 0$ for all $k \in \mathbb{N}$. The claim follows if we replace δ with δ/L. □

In view of Proposition 3.4, from Lemmas 6.1 and 2.1 we obtain the following bound for u_k in $H^1(S^1)$.

Lemma 6.2. *There exist a time $T_0 > 0$ and constants $C > 0$, $C_0 = C_0(E(u_0)) > 0$ such that*

$$\int_0^{T_0}\int_{S^1} |\partial_\phi u_k(t)|^2\,d\phi\,dt \le CE(u_{k0}) \le C_0 \quad \textit{for all } k \in \mathbb{N}.$$

From Lemma 6.2 we obtain locally in time uniform smooth bounds for (u_k) for $t > 0$ by iteratively applying our previous regularity results. More precisely, Fatou's lemma and Lemma 6.2 first yield the bound

$$\int_0^{T_0} \liminf_{k\to\infty}\left(\int_{S^1} |\partial_\phi u_k(t)|^2\,d\phi\right)dt \le C_0.$$

Thus, for almost every $0 < t_0 < T_0$,

$$\liminf_{k\to\infty} \int_{S^1} |\partial_\phi u_k(t_0)|^2\,d\phi < \infty.$$

For any such $0 < t_0 < T_0$, if $\delta > 0$ is sufficiently small, from Proposition 4.2 with another appeal to Fatou's lemma we may conclude

$$\int_{t_0}^{T_0} \liminf_{k\to\infty} \int_B |\nabla\partial_\phi u_k|^2\,dz\,dt \le \liminf_{k\to\infty} \int_{t_0}^{T_0}\int_B |\nabla\partial_\phi u_k|^2\,dz\,dt \le C_1$$

for some $C_1 > 0$, so that now we even have

$$\liminf_{k\to\infty} \int_B |\nabla\partial_\phi u_k(t_1)|^2\,dz < \infty$$

for almost every $t_0 < t_1 < T_0$. Hence we may next invoke Proposition 4.4 and (4-2) to obtain the bound

$$\liminf_{k\to\infty} \int_{t_1}^{T_0}\int_{\partial B} |\nabla\partial_\phi u_k|^2\,dz\,dt < \infty$$

for any such $t_0 < t_1 < T_0$, and Fatou's lemma gives

$$\liminf_{k\to\infty} \int_{\partial B} |\nabla \partial_\phi u_k(t_2)|^2 \, d\phi < \infty$$

for almost every $t_1 < t_2 < T_0$. Now Proposition 4.10 may be applied with $\varphi_0 = 1$, and we obtain

$$\liminf_{k\to\infty} \int_{t_2}^{T_0} \int_B |\nabla \partial_\phi^2 u_k|^2 \, dz \, dt < \infty$$

for any such $t_1 < t_2 < T_0$. Another application of Fatou's lemma gives

$$\liminf_{k\to\infty} \int_B |\nabla \partial_\phi^2 u_k(t_3)|^2 \, dz < \infty$$

for almost every $t_2 < t_3 < T_0$, and Proposition 4.5 yields

$$\liminf_{k\to\infty} \int_{t_3}^{T_0} \int_{\partial B} |\nabla \partial_\phi^2 u_k|^2 \, d\phi \, dt < \infty$$

for any such $t_2 < t_3 < T_0$. We may then iterate, using (3-7) and alternatingly employing Propositions 4.11 and 4.6 for $3 \le k \in \mathbb{N}$, to find a subsequence (u_k) satisfying uniform smooth bounds on $]t_0, T_0]$ for any $t_0 > 0$. Passing to the limit as $k \to \infty$ for this subsequence we obtain a weak solution to (1-3), (1-4) of energy-class in the following sense.

Definition 6.3. A function $u \in H^1([0, T_0] \times S^1; N) \cap L^\infty([0, T_0]; H^{1/2}(S^1; N))$ with harmonic extension $u = u(t)$ for each t is a weak solution of (1-3), (1-4) of energy-class, if (1-3) is satisfied in the weak sense, that is, if

$$\int_0^{T_0} \int_{\partial B} (u_t + d\pi_N(u) u_r) \cdot \varphi \, d\phi \, dt = \int_0^{T_0} \int_{\partial B} u_t \cdot \varphi \, d\phi \, dt + \int_0^{T_0} \int_B \nabla u \cdot \nabla (d\pi_N(u)\varphi) \, dz \, dt = 0 \quad (6\text{-}1)$$

for all $\varphi \in C_c^\infty(\bar{B} \times]0, T_0[)$, and if there holds the energy inequality

$$E(u(T)) + \int_S^T \int_{\partial B} |u_t|^2 \, d\phi \, dt \le E(u(S)) \quad (6\text{-}2)$$

for any $0 \le S < T < T_0$, with the initial data $u_0 \in H^{1/2}(S^1; N)$ being attained in the sense of traces.

We then may summarize our results as follows.

Proposition 6.4. *For any $u_0 \in H^{1/2}(S^1; N)$ there exists $T_0 > 0$ and a weak solution u to (1-3), (1-4) on $[0, T_0]$ of energy-class, which is smooth for $t > 0$.*

Proof. For any open $U \subset]0, T_0[$ we have uniform smooth bounds for u_k on U; thus a suitable subsequence u_k approaches u smoothly locally as $k \to \infty$. Equation (6-1) follows from the corresponding identities for u_k.

Moreover, (6-2) follows from the energy identity, Lemma 2.1, for u_k, where we also use H^1-convergence $u_{0k} \to u_0$ as well as weak lower semicontinuity of the energy and of the L^2-norm.

Finally, with error $o(1) \to 0$ as $k \to \infty$ for $0 < t < T_0$ we can estimate

$$\|u(t) - u_0\|^2_{L^2(\partial B)} \le \|u_k(t) - u_{0k}\|^2_{L^2(\partial B)} + o(1) \le \left(\int_0^t \|\partial_t u_k(t')\|_{L^2(\partial B)}\, dt'\right)^2 + o(1)$$
$$\le t \int_0^t \|\partial_t u_k(t')\|^2_{L^2(\partial B)}\, dt' + o(1) \le t E(u_0) + o(1),$$

and $u(t) \to u_0$ weakly in $H^{1/2}(S^1; N) \cap H^1(B; \mathbb{R}^n)$ as $t \downarrow 0$. In fact, by (6-2) we then even have strong convergence. □

7. Uniqueness

With the help of the tools developed in Section 3 we can show uniqueness of partially regular weak energy-class solutions as in Proposition 6.4.

Theorem 7.1. *Let $u_0 \in H^{1/2}(S^1; N)$. Suppose u and v both are weak energy-class solutions of (1-3), (1-4) on $[0, T_0]$ for some $T_0 > 0$ with initial data u_0, and suppose u and v are smooth for $t > 0$. Then $u = v$.*

Proof. Using (3-2) for u and v, for the function $w = u - v$ for almost every $0 < t < T_0$ we have

$$\partial_t w + \partial_r w = \nu(u)\partial_r(\operatorname{dist}_N(u)) - \nu(v)\partial_r(\operatorname{dist}_N(v))$$
$$= (\nu(u) - \nu(v))\partial_r(\operatorname{dist}_N(u)) + \nu(v)\partial_r(\operatorname{dist}_N(u) - \operatorname{dist}_N(v)) \tag{7-1}$$

on $\partial B = S^1$. From (3-5), moreover, we obtain

$$|\Delta(\operatorname{dist}_N(u) - \operatorname{dist}_N(v))| = |\nabla u \cdot d\nu(u)\nabla u - \nabla v \cdot d\nu(v)\nabla v|$$
$$\le C(|w||\nabla u|^2 + (|\nabla u| + |\nabla v|)|\nabla w|) \quad \text{in } B. \tag{7-2}$$

Observing that

$$|\operatorname{dist}_N(u) - \operatorname{dist}_N(v)| \le C|w|,$$

upon multiplying (7-2) with the function $(\operatorname{dist}_N(u) - \operatorname{dist}_N(v)) \in H^1_0(B)$, integrating by parts, and using Young's inequality, for any $\varepsilon > 0$ we obtain

$$\|\nabla(\operatorname{dist}_N(u) - \operatorname{dist}_N(v))\|^2_{L^2(B)} \le C \int_B (|w|^2|\nabla u|^2 + (|\nabla u| + |\nabla v|)|\nabla w||w|)\, dz$$
$$\le \varepsilon \|\nabla w\|^2_{L^2(B)} + C(\varepsilon)\|w\|^2_{L^4(B)}(\|\nabla u\|^2_{L^4(B)} + \|\nabla v\|^2_{L^4(B)}). \tag{7-3}$$

On the other hand, for any $0 < t_0 < T \le T_0$, multiplying (7-1) with w and integrating by parts on $S^1 \times [t_0, T]$, upon letting $t_0 \downarrow 0$ we find

$$\sup_{0<t<T} \|w(t)\|^2_{L^2(\partial B)} + \int_0^T \int_B |\nabla w|^2\, dz\, dt$$
$$\le C \int_0^T \int_{\partial B} (\partial_t w + \partial_r w) w\, d\phi\, dt$$
$$= C \int_0^T \int_{\partial B} w(\nu(u) - \nu(v))\partial_r(\operatorname{dist}_N(u))\, d\phi\, dt + C \int_0^T \int_{\partial B} w\nu(v)\partial_r(\operatorname{dist}_N(u) - \operatorname{dist}_N(v))\, d\phi\, dt$$
$$=: C \int_0^T (I + II)\, dt.$$

We first estimate the term

$$I = I(t) = \int_{\partial B} w(\nu(u)-\nu(v))\partial_r(\mathrm{dist}_N(u))\,d\phi$$
$$= \int_B \nabla(w(\nu(u)-\nu(v)))\nabla(\mathrm{dist}_N(u))\,dz + \int_B w(\nu(u)-\nu(v))\Delta(\mathrm{dist}_N(u))\,dz.$$

Using

$$|\nabla(w(\nu(u)-\nu(v)))| \le C|\nabla w||w| + |w((d\nu(u)-d\nu(v))\nabla u + d\nu(v)\nabla w)|$$
$$\le C(|\nabla w||w| + |w|^2|\nabla u|)$$

we can bound

$$\left|\int_B \nabla(w(\nu(u)-\nu(v)))\nabla(\mathrm{dist}_N(u))\,dz\right| \le C\int_B (|\nabla w||w|+|w|^2|\nabla u|)|\nabla u|\,dz$$
$$\le \varepsilon\|\nabla w\|^2_{L^2(B)} + C(\varepsilon)\|w\|^2_{L^4(B)}\|\nabla u\|^2_{L^4(B)}$$

for each t. Also using (3-5), we can moreover estimate

$$\left|\int_B w(\nu(u)-\nu(v))\Delta(\mathrm{dist}_N(u))\,dz\right| \le C\|w\|^2_{L^4(B)}\|\nabla u\|^2_{L^4(B)}$$

for almost every $0 < t < T$ to obtain

$$|I| \le \varepsilon\|\nabla w\|^2_{L^2(B)} + C(\varepsilon)\|w\|^2_{L^4(B)}\|\nabla u\|^2_{L^4(B)}.$$

Similarly, we estimate the term

$$II = II(t) = \int_{\partial B} w\nu(v)\partial_r(\mathrm{dist}_N(u)-\mathrm{dist}_N(v))\,d\phi$$
$$= \int_B \nabla(w\nu(v))\nabla(\mathrm{dist}_N(u)-\mathrm{dist}_N(v))\,dz + \int_B w\nu(v)\Delta(\mathrm{dist}_N(u)-\mathrm{dist}_N(v))\,dz.$$

Noting that with (7-3) we can bound

$$\left|\int_B \nabla(w\nu(v))\nabla(\mathrm{dist}_N(u)-\mathrm{dist}_N(v))\,dz\right| \le C(\|\nabla w\|_{L^2(B)}+\|w\nabla v\|_{L^2(B)})\|\nabla(\mathrm{dist}_N(u)-\mathrm{dist}_N(v))\|_{L^2(B)}$$
$$\le \varepsilon\|\nabla w\|^2_{L^2(B)} + C(\varepsilon)\|w\|^2_{L^4(B)}(\|\nabla u\|^2_{L^4(B)}+\|\nabla v\|^2_{L^4(B)})$$

and that with (7-2) we have

$$\left|\int_B w\nu(v)\Delta(\mathrm{dist}_N(u)-\mathrm{dist}_N(v))\,dz\right| \le C\int_B (|w|^2|\nabla u|^2 + |w||\nabla w|(|\nabla u|+|\nabla v|))\,dz$$
$$\le \varepsilon\|\nabla w\|^2_{L^2(B)} + C(\varepsilon)\|w\|^2_{L^4(B)}(\|\nabla u\|^2_{L^4(B)}+\|\nabla v\|^2_{L^4(B)}),$$

we find the estimate

$$|II| \le \varepsilon\|\nabla w\|^2_{L^2(B)} + C(\varepsilon)\|w\|^2_{L^4(B)}(\|\nabla u\|^2_{L^4(B)}+\|\nabla v\|^2_{L^4(B)})$$

for almost every $0 < t < T$.

But Sobolev's embedding $H^{1/2}(B) \hookrightarrow L^4(B)$ and Fourier expansion give the bound

$$\|w\|^2_{L^4(B)} \le C\|w\|^2_{H^{1/2}(B)} \le C\|w\|^2_{L^2(\partial B)}$$

and similar bounds for ∇u as well as ∇v. Moreover, since by the energy inequality (6-2) we have $u(t), v(t) \to u_0$ strongly in $H^1(B)$ as $t \downarrow 0$, there exist a radius $0 < R \le \frac{1}{2}$ and a time $0 < T < T_0$ such that condition (3-13) in Proposition 3.4 holds true on $[0, T]$ for both u and v, allowing us to bound

$$\begin{aligned}\int_0^T \|\nabla u(t)\|^2_{L^4(B)}\,dt &\le C\int_0^T \|\nabla u(t)\|^2_{L^2(\partial B)}\,dt \le C\int_0^T \|\partial_\phi u(t)\|^2_{L^2(\partial B)}\,dt \\ &\le C\int_0^T\int_{\partial B} |u_t|^2\,d\phi\,dt + C(R)TE(u_0) \le C(R)(1+T_0)E(u_0)\end{aligned}$$

with the help of (3-7), and similarly for $|\nabla v|$. Choosing $\varepsilon = \frac{1}{4}$, for sufficiently small $0 < T < T_0$ by absolute continuity of the integral we thus can estimate

$$\begin{aligned}&\sup_{0<t<T} \|w(t)\|^2_{L^2(\partial B)} + \int_0^T\int_B |\nabla w|^2\,dz\,dt \\ &\qquad\le \tfrac{1}{2}\|\nabla w\|^2_{L^2(B\times[0,T])} + C\sup_{0<t<T}\|w(t)\|^2_{L^2(\partial B)}\int_0^T (\|\nabla u\|^2_{L^4(B)} + \|\nabla v\|^2_{L^4(B)})\,dt \\ &\qquad\le \frac{1}{2}\Big(\sup_{0<t<T}\|w(t)\|^2_{L^2(\partial B)} + \int_0^T\int_B |\nabla w|^2\,dz\,dt\Big),\end{aligned}$$

and it follows that $w = 0$, as claimed. □

Proof of Theorem 1.2. Existence for short time and uniqueness of a partially regular weak solution to (1-3), (1-4) for given data $u_0 \in H^{1/2}(S^1; N)$ follow from Proposition 6.4 and Theorem 7.1, respectively. Since by Proposition 6.4 our weak solution is smooth for $t > 0$, the remaining assertions follow from Theorem 1.1.

Note that at any blow-up time T_{i-1}, $i \ge 1$, of the flow as in Theorem 1.1(ii) there exists a unique weak limit $u_i = \lim_{t\uparrow T_{i-1}} u(t) \in H^{1/2}(S^1; N)$, and we may uniquely continue the flow using Proposition 6.4. □

8. Blow-up

Preparing for the proof of part (ii) of Theorem 1.1, suppose now that for the solution constructed in part (i) of that theorem there holds $T_0 < \infty$. Then, as we shall see in more detail below, by the results in Section 4 condition (4-4) must be violated for $T = T_0$ and there exist $\delta > 0$ and points $z_k \in B$ as well as radii $r_k \downarrow 0$ as $k \to \infty$ such that, for suitable $t_k \uparrow T_0$,

$$\int_{B_{r_k}(z_k)\cap B} |\nabla u(t_k)|^2\,dz = \sup_{\substack{z_0\in B\\ t\le t_k}} \int_{B_{r_k}(z_0)\cap B} |\nabla u(t)|^2\,dz = \delta.$$

We may later choose a smaller constant $\delta > 0$, if necessary. Moreover, for later use from now on we consider local concentrations in the sense that, for some $z_0 \in B$ and some fixed radius $r_0 > 0$ for a sequence of points $z_k \in B$ with $z_k \to z_0$ and radii $r_k \downarrow 0$ for suitable $t_k \uparrow T_0$ as $k \to \infty$,

$$\int_{B_{r_k}(z_k)\cap B} |\nabla u(t_k)|^2\,dz = \sup_{\substack{z'\in B_{r_0}(z_0)\\ t\le t_k}} \int_{B_{r_k}(z')\cap B} |\nabla u(t)|^2\,dz = \delta.$$

Scale

$$u_k(z,t) = u(z_k + r_k z, t_k + r_k t)$$

for

$$z \in \Omega_k = \{z : z_k + r_k z \in B\}, \quad t \in I_k = \{t : 0 \le t_k + t r_k < T_0\}.$$

Note that then

$$\int_{B_1(0)\cap\Omega_k} |\nabla u_k(0)|^2\,dz = \sup_{\substack{z_k + r_k z' \in B_{r_0}(z_0)\\ -t_k/r_k \le t < 0}} \int_{B_1(z')\cap\Omega_k} |\nabla u_k(t)|^2\,dz = \delta. \tag{8-1}$$

Passing to a subsequence we may assume that the domains Ω_k exhaust a limit domain $\Omega_\infty \subset \mathbb{R}^2$, which either is the whole space $\mathbb{R}^2$ or a half-space H.

By the energy inequality Lemma 2.1 for $t \in I_k$,

$$\int_{\Omega_k} |\nabla u_k(t)|^2\,dz = \int_B |\nabla u(t_k + r_k t)|^2\,dz \le 2E(u_0), \tag{8-2}$$

and for any $t_0 < 0$ and sufficiently large $k \in \mathbb{N}$ we have

$$\begin{aligned}\int_{t_0}^0 \int_{\partial\Omega_k} |\partial_t u_k|^2\,ds\,dt &= \int_{t_0}^0 \int_{\partial\Omega_k} |d\pi_N(u_k)\partial_{\nu_k} u_k|^2\,ds\,dt \\ &= \int_{t_k + r_k t_0}^{t_k} \int_{\partial B} |u_t|^2\,d\phi\,dt \le \int_{t_k + r_k t_0}^{T_0} \int_{\partial B} |u_t|^2\,d\phi\,dt \to 0 \end{aligned} \tag{8-3}$$

as $k \to \infty$, where ds is the element of length and where ν_k is the outward unit normal along $\partial\Omega_k$. Expressing the harmonic functions $\partial_t u_k(t)$ in Fourier series for each $t < 0$, it then also follows that $\partial_t u_k \to 0$ locally in L^2 on $\Omega_\infty \times]-\infty, 0[$. Finally, again using the fact that $u_k(t)$ for each t is harmonic, by the maximum principle we have the uniform bound $|u_k| \le \sup_{p\in N} |p|$ as well as uniform smooth bounds locally away from the boundary of Ω_∞.

Hence we may assume that as $k \to \infty$ we have $u_k \to u_\infty$ weakly locally in H^1 on $\Omega_\infty \times]-\infty, 0[$, where $u_\infty(z,t) = u_\infty(z)$ is independent of time, harmonic, and bounded. Moreover, we have smooth convergence away from $\partial\Omega_\infty$. Thus, if we assume that $\Omega_\infty = \mathbb{R}^2$, by (8-1) it follows that

$$\int_{B_1(0)} |\nabla u_\infty|^2\,dz = \delta.$$

But any function $v : \mathbb{R}^2 \to \mathbb{R}$ which is bounded and harmonic must be constant, which rules out this possibility. Hence Ω_∞ can only be a half-space.

After a suitable rotation of the domain B and shift of coordinates in $\mathbb{R}^2 \cong \mathbb{C}$ we may then assume that $z_k = (0, -y_k)$ with $1 - y_k \le M r_k$ for some $M \in \mathbb{N}$ and that $\Omega_\infty = \{(x,y) : y > y_0\}$ for some y_0. Finally, replacing $r_k > 0$ with $(M+1)r_k$ and z_k with $z_k = (0, -1)$, if necessary, we may assume that

$$\Omega_k \subset \mathbb{R}^2_+ = \{(x, y) : y > 0\}$$

is the ball of radius $1/r_k$ around the point $(0, 1/r_k)$ with $0 \in \partial\Omega_k$, while from (8-1) with a uniform number $L \in \mathbb{N}$ we have

$$L \int_{B_1(0)\cap\Omega_k} |\nabla u_k(0)|^2 \, dz \geq L\delta \geq \sup_{\substack{|z'|\leq r_0/r_k \\ -t_k/r_k \leq t < 0}} \int_{B_1(z')\cap\Omega_k} |\nabla u_k(t)|^2 \, dz \tag{8-4}$$

for any $k \in \mathbb{N}$. Let $\Phi_k : \mathbb{R}^2_+ \to \Omega_k$ be the conformal maps given by

$$\Phi_k(z) = \frac{2z}{1 - ir_k z}, \quad z \in \mathbb{R}^2_+, \ \ k \in \mathbb{N},$$

with $\Phi_k \to 2 \cdot \mathrm{id}$ locally uniformly on $\mathbb{R}^2 \cong \mathbb{C}$ as $k \to \infty$.

Let $v_k = u_k \circ \Phi_k$ with $k \in \mathbb{N}$. By conformal invariance of the Dirichlet energy, from (8-2) for any t we have

$$\int_{\mathbb{R}^2_+} |\nabla v_k(t)|^2 \, dz = \int_{\Omega_k} |\nabla u_k(t)|^2 \, dz \leq 2E(u_0), \tag{8-5}$$

and by (8-4) with a uniform number $L_1 \in \mathbb{N}$ there holds

$$L_1 \int_{B_2^+(0)} |\nabla v_k(0)|^2 \, dz \geq L_1\delta \geq \sup_{\substack{|z'|\leq r_0/r_k \\ -t_k/r_k \leq t < 0}} \int_{B_1^+(z')} |\nabla v_k(t)|^2 \, dz, \tag{8-6}$$

where

$$B_r^+(z) = B_r(z) \cap \mathbb{R}^2_+$$

for any $r > 0$ and any $z = (x, y) \in \mathbb{R}^2$. Moreover, from (8-3) for any $t_0 < 0$ and any $R > 0$ for the integral over $]-R, R[\times \{0\} \subset \partial\mathbb{R}^2_+$ we obtain

$$\int_{t_0}^0 \int_{-R}^R |\partial_t v_k|^2 \, dx \, dt \leq C \int_{t_0}^0 \int_{-R}^R |d\pi_N(v_k)\partial_y v_k|^2 \, dx \, dt \to 0 \quad \text{as } k \to \infty, \tag{8-7}$$

and $\partial_t v_k \to 0$ locally in L^2 on $\overline{\mathbb{R}}^2_+ \times]-\infty, 0[$. In addition, from our choice of (u_k) it follows that $v_k \to v_\infty$ weakly locally in H^1 on $\overline{\mathbb{R}}^2_+ \times]-\infty, 0[$ as $k \to \infty$, where $v_\infty(z, t) =: w_\infty(z)$ is harmonic and bounded.

For a suitable sequence of times $t_0 < s_k < 0$, we also have locally weak convergence $w_k := v_k(s_k) \to w_\infty$ in H^1 on $\overline{\mathbb{R}}^2_+$ and, in addition,

$$d\pi_N(w_k)\partial_y w_k \to 0 \quad \text{in } L^2_{\mathrm{loc}}(\partial\mathbb{R}^2_+) \quad \text{as } k \to \infty. \tag{8-8}$$

Thus, for sufficiently small $\delta > 0$, by Proposition 3.3 applied to the functions $w_k \circ \Psi$, where $\Psi : B \to \mathbb{R}^2_+$ is a suitable conformal map, we also have uniform local L^2-bounds for $\partial_x w_k$ on $\partial\mathbb{R}^2_+$, and we may assume that $w_k \to w_\infty$ locally uniformly and weakly locally in H^1 on $\partial\mathbb{R}^2_+$ as $k \to \infty$. Since w_k is harmonic, we then also have locally strong H^1-convergence $w_k \to w_\infty$ on $\overline{\mathbb{R}}^2_+$.

To see that w_∞ is nonconstant, let $\varphi_k = \varphi_{z_0,4r_k}$, $k \in \mathbb{N}$. Integrating the identity (2-1) from the proof of Lemma 2.2 in time, with error $o(1) \to 0$ and suitable numbers $\varepsilon_k \downarrow 0$ as $k \to \infty$ in view of (8-3), we find

$$\begin{aligned}\frac{1}{2}\left|\int_B |\nabla u(t_k)|^2 \varphi_k^2\,dz - \int_B |\nabla u(t_k + r_k s_k)|^2 \varphi_k^2\,dz\right| \\ \le \int_{t_k+r_k s_k}^{t_k} \int_{\partial B} |u_t|^2 \varphi_k^2\,d\phi\,dt + 2\int_{t_k+r_k s_k}^{t_k}\int_B |u_t \nabla u \varphi_k \nabla\varphi_k|\,dz\,dt \\ \le o(1) + 8\varepsilon_k r_k \int_{t_k+r_k s_k}^{t_k}\int_B |\nabla u|^2 |\nabla\varphi_k|^2\,dz\,dt + (8\varepsilon_k r_k)^{-1}\int_{t_k+r_k s_k}^{t_k}\int_B |u_t|^2\varphi_k^2\,dz\,dt. \quad (8\text{-}9)\end{aligned}$$

With the help of (2-2) and (8-3) for suitable $\varepsilon_k \downarrow 0$ we can bound

$$(8\varepsilon_k r_k)^{-1}\int_{t_k+r_k s_k}^{t_k}\int_B |u_t|^2\varphi_k^2\,dz\,dt \le C\varepsilon_k^{-1}\int_{t_k+r_k s_k}^{t_k}\int_{\partial B}|u_t|^2\,dz\,dt \to 0.$$

Since for any choice $t_0 < s_k < 0$ we also can estimate

$$8\varepsilon_k r_k \int_{t_k+r_k s_k}^{t_k}\int_B |\nabla u|^2|\nabla\varphi_k|^2\,dz\,dt \le C\varepsilon_k |t_0| E(u_0) \to 0,$$

from (8-9) and (8-6) it follows that with error $o(1) \to 0$ as $k \to \infty$ we have

$$\begin{aligned}\int_{B_4^+(0)} |\nabla w_k|^2\,dz + o(1) = \int_{B_4^+(0)} |\nabla v_k(s_k)|^2\,dz + o(1) \ge \int_B |\nabla u(t_k + r_k s_k)|^2\varphi_k^2\,dz + o(1) \\ \ge \int_B |\nabla u(t_k)|^2 \varphi_k^2\,dz \ge \int_{B_2^+(0)} |\nabla v_k(0)|^2\,dz \ge \delta. \quad (8\text{-}10)\end{aligned}$$

Finally, in view of locally uniform convergence $w_k \to w_\infty$ and weak local L^2-convergence of the traces $\nabla w_k \to \nabla w_\infty$ on $\partial\mathbb{R}^2_+$, we may pass to the limit $k \to \infty$ in (8-8) to conclude that $w_\infty \colon \partial\mathbb{R}^2_+ \to N$ with

$$d\pi_N(w_\infty)\partial_y w_\infty = 0 \quad \text{on } \partial\mathbb{R}^2_+. \quad (8\text{-}11)$$

Since w_∞ is harmonic, the Hopf differential

$$f = |\partial_x w_\infty|^2 - |\partial_y w_\infty|^2 - 2i\,\partial_x w_\infty \cdot \partial_y w_\infty$$

defines a holomorphic function $f \in L^1(\mathbb{R}^2_+, \mathbb{C})$. Moreover, $w_\infty \in H^{3/2}_{\mathrm{loc}}(\overline{\mathbb{R}}^2_+)$ with trace $\nabla w_\infty \in L^2_{\mathrm{loc}}(\partial\mathbb{R}^2_+)$; thus also the trace of f is well-defined on $\partial\mathbb{R}^2_+$. By (8-11) now the trace of f is real-valued; thus $f \equiv c$ for some constant $c \in \mathbb{R}$. But $\nabla w_\infty \in L^2(\mathbb{R}^2_+)$; hence $f \in L^1(\mathbb{R}^2_+)$. It follows that $c = 0$, and w_∞ is conformal.

With a conformal diffeomorphism $\Phi \colon B \to \mathbb{R}^2_+$ mapping a point $z_0 \in \partial B$ to infinity, define the map $\bar u = w_\infty \circ \Phi \in H^{1/2}(S^1; N)$. By conformal invariance, $\bar u$ again is harmonic with finite Dirichlet integral and satisfies (1-6) on $\partial B \setminus \{z_0\}$; since the point $\{z_0\}$ has vanishing H^1-capacity, $\bar u$ then is stationary in the sense of [Grüter et al. 1981]. Moreover, $\bar u$ is conformal. For such mappings, smooth regularity on $\overline{B}$ was shown by Grüter, Hildebrandt, and Nitsche [Grüter et al. 1981]; thus condition (1-6) holds everywhere on

∂B in the pointwise sense, and $\bar{u}$ parametrizes a minimal surface of finite area supported by N which meets N orthogonally along its boundary.

Proof of Theorem 1.1(ii). For given smooth data $u_0 \in H^{1/2}(S^1; N)$ let u be the unique solution to (1-3), (1-4) guaranteed by part (i) of the theorem, and suppose that the maximal time of existence T_0 is less than ∞. Then condition (4-4) must fail as $t \uparrow T_0$; else from Propositions 4.11 and 4.6 we obtain smooth bounds for $u(t)$ as $t \uparrow T_0$ and there exists a smooth trace $u_1 = \lim_{t\uparrow T_0} u(t)$. But by the first part of the theorem there is a smooth solution to the initial value problem for (1-3) with initial data u_1 at time T_0, and this solution extends the original solution u to an interval $[0, T_1[$ for some $T_1 > T_0$, contradicting the maximality of T_0.

Let $z^{(i)} \in B,\ 1 \le i \le i_0$, such that, for some number $\delta > 0$ and suitable $t_k^{(i)} \uparrow T_0,\ z_k^{(i)} \to z^{(i)},\ r_k^{(i)} \to 0$ as $k \to \infty$,

$$\liminf_{k\to\infty} \int_{B_{r_k^{(i)}}(z_k^{(i)})\cap B} |\nabla u(t_k^{(i)})|^2\, dz \ge \delta.$$

By the argument following (8-9), for a suitable sequence of radii $0 < r_k^{(0)} \to 0$ such that $r_k^{(i)}/r_k^{(0)} \to 0$ as well as $(T_0 - t_k^{(i)})/r_k^{(0)} \to 0$, then with error $o(1) \to 0$ as $k \to \infty$,

$$\int_{B_{r_k^{(0)}}(z^{(i)})\cap B} |\nabla u(t)|^2\, dz + o(1) \ge \int_{B_{r_k^{(i)}}(z_k^{(i)})\cap B} |\nabla u(t_k^{(i)})|^2\, dz \ge \delta$$

for all $T_0 - r_k^{(0)} < t < T_0$, uniformly in $1 \le i \le i_0$. For sufficiently large $k \in \mathbb{N}$ such that

$$r_k^{(0)} < \inf_{i<j} \tfrac{1}{4}|z^{(i)} - z^{(j)}|,$$

it follows that $i_0 \le E(u_0)/\delta$, and we may fix $r_0 > 0$ and redefine $t_k^{(i)},\ r_k^{(i)}$, and $z_k^{(i)}$, if necessary, such that, for each $1 \le i \le i_0$,

$$\int_{B_{r_k^{(i)}}(z_k^{(i)})\cap B} |\nabla u(t_k^{(i)})|^2\, dz = \sup_{\substack{z'\in B_{r_0}(z^{(i)}) \\ 0<t\le t_k^{(i)}}} \int_{B_{r_k^{(i)}}(z')\cap B} |\nabla u(t)|^2\, dz = \delta.$$

Moreover, we may assume that $\delta < \delta_0$ as defined in Proposition 3.1. The characterization of the concentration points as in Theorem 1.2(ii) via solutions $\bar{u}^{(i)}$ of (1-6) then follows from our above analysis.

In addition, Corollary 3.2 yields the uniform lower bound

$$\lim_{r_0\downarrow 0} \liminf_{t\uparrow T} \int_{B_{r_0}(z^{(i)})\cap B} |\nabla u(t)|^2\, dz \ge 2E(\bar{u}^{(i)}) \ge 2\delta_0^2$$

for the concentration energy quanta, which gives the claimed upper bound for the total number of concentration points.

Finally, with the help of Proposition 4.11 we can smoothly extend the solution u to $B \setminus \{z^{(1)}, \dots, z^{(i_0)}\}$ at time $t = T_0$. □

9. Asymptotics

Suppose next that the solution u to (1-3), (1-4) exists for all time $0 < t < \infty$. Then u either concentrates for suitable $t_k \uparrow \infty$ in the sense that condition (4-4) does not hold true uniformly in time, or u satisfies uniform smooth bounds, as shown in Section 4.

In the latter case, the claim made in Theorem 1.1(iii) easily follows.

Proposition 9.1. *Suppose that for any $\delta > 0$ there exists $R > 0$ such that condition (4-4) holds true for all $0 < t < \infty$. Then there exists a smooth solution $u_\infty \in H^{1/2}(S^1; N)$ of (1-6) such that $u(t) \to u_\infty$ smoothly as $t \to \infty$ suitably, and u_∞ parametrizes a minimal surface of finite area supported by N which meets N orthogonally along its boundary.*

Proof. For sufficiently small $\delta > 0$ and for any $j \in \mathbb{N}$, by iterative reference to Propositions 4.2, 4.4–4.6, and 4.10, 4.11 as in Section 6 we can find constants $C_j > 0$ such that $\|u(t)\|_{H^j(B)} \le C_j$ for all $t > 1$. Moreover, by the energy inequality Lemma 2.1 for a suitable sequence $t_k \to \infty$ there holds $u_t(t_k) \to 0$ in $L^2(\partial B)$ as $k \to \infty$. Then for any $j \in \mathbb{N}$ a subsequence $u(t_k)$ approaches u_∞ in $H^j(B)$, and a diagonal subsequence converges smoothly, where u_∞ solves (1-6). By the argument after (8-11) in Section 8, u_∞ is conformal and u_∞ parametrizes a minimal surface with free boundary on N which meets N orthogonally along its boundary. □

In the remaining case that for some $\delta > 0$ condition (4-4) fails to hold, there exists a sequence $t_k \uparrow \infty$ and points $z^{(1)}, \dots, z^{(i_0)}$ such that, for sequences $z_k^{(i)} \to z^{(i)}$ and radii $r_k^{(i)} \to 0$ as $k \to \infty$,

$$\liminf_{k\to\infty} \int_{B_{r_k^{(i)}}(z_k^{(i)})\cap B} |\nabla u(t_k)|^2\, dz \ge \delta, \quad 1 \le i \le i_0.$$

By Lemma 2.1 there holds the a-priori bound $i_0 \le E(u_0)/\delta$ for the number of concentration points. By the argument leading to (8-10) then, for a suitable number

$$0 < r_0 \le \inf_{i<j} \tfrac{1}{4}|z^{(i)} - z^{(j)}|$$

with error $o(1) \to 0$ as $k \to \infty$ and with some constant $L \in \mathbb{N}$ for all $1 \le i \le i_0$,

$$L \int_{B_{2r_k^{(i)}}(z_k^{(i)})\cap B} |\nabla u(t_k)|^2\, dz + o(1) \ge \sup_{\substack{z_0 \in B_{r_0}(z_k^{(i)}) \\ t_k - r_0 \le t \le t_k}} \int_{B_{r_k^{(i)}}(z_0)\cap B} |\nabla u(t)|^2\, dz \ge \delta.$$

Fixing any index $1 \le i \le i_0$ and renaming $z_k^{(i)} =: z_k$ and $r_k^{(i)} =: r_k$, we then scale

$$u_k(z, t) = u(z_k + r_k z, t_k + r_k t), \quad z \in \Omega_k = \{z : z_k + r_k z \in B\}, \; -t_k/r_k \le t \le 0,$$

as before and observe that, for any $t_0 < 0$,

$$\begin{aligned}\int_{t_0}^{0}\int_{\partial\Omega_k} |\partial_t u_k|^2\, ds\, dt &= \int_{t_0}^{0}\int_{\partial\Omega_k} |d\pi_N(u_k)\partial_{\nu_k} u_k|^2\, ds\, dt \\ &= \int_{t_k + r_k t_0}^{t_k}\int_{\partial B} |u_t|^2\, d\phi\, dt \le \int_{t_k + r_k t_0}^{\infty}\int_{\partial B} |u_t|^2\, d\phi\, dt \to 0 \end{aligned} \tag{9-1}$$

as $k \to \infty$, where ν_k is the outward unit normal along $\partial\Omega_k$. Just as in Section 8 for suitable $t_0 < s_k < 0$ we then obtain local uniform and H^1-convergence of a subsequence of the conformally rescaled maps $w_k = u_k(s_k) \circ \Phi_k \in H^1_{\mathrm{loc}}(\mathbb{R}^2_+)$ to a smooth, nonconstant, harmonic and conformal limit w_∞ with finite energy and continuously mapping $\partial\mathbb{R}^2_+$ to N, inducing a solution $\bar{u}_\infty = w_\infty \circ \Phi \in H^{1/2}(S^1; N)$ of (1-6) corresponding to a minimal surface with free boundary on N. This ends the proof of Theorem 1.1(iii).

Appendix

In this section, for the convenience of the reader we derive two interpolation inequalities that play a crucial role in our arguments.

Let $v \in H^1(B)$, and for $r > 0$ let $(z_i)_{1\le i\le i_0}$ be such that the collection of balls $B_{r/2}(z_i)$, $1 \le i \le i_0$, covers $\bar{B}$ with at most L balls $B_r(z_i)$ overlapping at any $z \in B$, with $L \in \mathbb{N}$ independent of $r > 0$. We may assume $r < \frac{1}{8}$ so that for any $1 \le i \le i_0$ there is a pair of orthogonal vectors $e_{1,i}$ and $e_{2,i}$ such that for any $z \in B_r(z_i)$ there holds $z + se_{1,i} + te_{2,i} \in B$ for any $0 \le s, t \le 2r$. After a rotation of coordinates, we may assume that $e_{1,i} = (1, 0)$ and $e_{2,i} = (0, 1)$ are the standard basis vectors. Writing φ for $\varphi_{z_i,r}$, by arguing as Ladyzhenskaya [1963] for any $z = (x, y) \in B_r(z_i)$ and using that

$$(v^2\varphi)(x + 2r, y) = 0 = (v^2\varphi)(x, y + 2r)$$

then we can estimate

$$\begin{aligned} v^4(z) = |(v^2\varphi)(z)|^2 &\le \int_0^{2r} |\partial_x(v^2\varphi)(x+s, y)|\, ds \cdot \int_0^{2r} |\partial_y(v^2\varphi)(x, y+t)|\, dt \\ &\le \int_{\{s:(s,y)\in B\}} |\partial_x(v^2\varphi)(s, y)|\, ds \cdot \int_{\{t:(x,t)\in B\}} |\partial_y(v^2\varphi)(x, t)|\, dt, \end{aligned} \tag{A-1}$$

and with the help of Fubini's theorem we find

$$\begin{aligned} \int_{B_{r/2}(z_i)} |v|^4\, dz &\le \int_B |v|^4\varphi^2\, dz \le \int_{-\infty}^{\infty} \left(\int_{\{x:(x,y)\in B\}} |(v^2\varphi)(x, y)|^2\, dx \right) dy \\ &\le \int_{-\infty}^{\infty} \int_{\{s:(s,y)\in B\}} |\partial_x(v^2\varphi)(s, y)|\, ds\, dy \cdot \int_{-\infty}^{\infty} \int_{\{t:(x,t)\in B\}} |\partial_y(v^2\varphi)(x, t)|\, dt\, dx \\ &\le \left(\int_B |\nabla(v^2\varphi)|\, dz \right)^2 \\ &\le \left(\int_B (2|\nabla v|\,|v\varphi| + v^2|\nabla\varphi|)\, dz \right)^2 \\ &\le C \left(\int_{B_r(z_i)} |\nabla v|^2\, dz + r^{-2} \int_{B_r(z_i)} v^2\, dz \right) \int_{B_r(z_i)} v^2\, dz. \end{aligned}$$

Fixing $r = \frac{1}{10}$ and summing over $1 \le i \le i_0$, with an absolute constant $C > 0$ we obtain the bound

$$\|v\|^4_{L^4(B)} \le C\|v\|^2_{H^1(B)}\|v\|^2_{L^2(B)} \tag{A-2}$$

for any $v \in H^1(B)$.

References

[Brendle 2002] S. Brendle, "Curvature flows on surfaces with boundary", *Math. Ann.* **324**:3 (2002), 491–519. MR Zbl

[Brézis and Gallouet 1980] H. Brézis and T. Gallouet, "Nonlinear Schrödinger evolution equations", *Nonlinear Anal.* **4**:4 (1980), 677–681. MR Zbl

[Brézis and Wainger 1980] H. Brézis and S. Wainger, "A note on limiting cases of Sobolev embeddings and convolution inequalities", *Comm. Partial Differential Equations* **5**:7 (1980), 773–789. MR Zbl

[Caffarelli and Silvestre 2007] L. Caffarelli and L. Silvestre, "An extension problem related to the fractional Laplacian", *Comm. Partial Differential Equations* **32**:7-9 (2007), 1245–1260. MR Zbl

[Chang and Liu 2005] K.-C. Chang and J.-Q. Liu, "Boundary flow for the minimal surfaces in $\mathbb{R}^n$ with Plateau boundary condition", *Proc. Roy. Soc. Edinburgh Sect. A* **135**:3 (2005), 537–562. MR Zbl

[Chang et al. 1992] K.-C. Chang, W. Y. Ding, and R. Ye, "Finite-time blow-up of the heat flow of harmonic maps from surfaces", *J. Differential Geom.* **36**:2 (1992), 507–515. MR Zbl

[Da Lio and Pigati 2020] F. Da Lio and A. Pigati, "Free boundary minimal surfaces: a nonlocal approach", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **20**:2 (2020), 437–489. MR Zbl

[Da Lio and Rivière 2011] F. Da Lio and T. Rivière, "Three-term commutator estimates and the regularity of $\frac{1}{2}$-harmonic maps into spheres", *Anal. PDE* **4**:1 (2011), 149–190. MR Zbl

[Da Lio et al. 2015] F. Da Lio, L. Martinazzi, and T. Rivière, "Blow-up analysis of a nonlocal Liouville-type equation", *Anal. PDE* **8**:7 (2015), 1757–1805. MR Zbl

[Douglas 1931] J. Douglas, "Solution of the problem of Plateau", *Trans. Amer. Math. Soc.* **33**:1 (1931), 263–321. MR Zbl

[Eells and Sampson 1964] J. Eells, Jr. and J. H. Sampson, "Harmonic mappings of Riemannian manifolds", *Amer. J. Math.* **86** (1964), 109–160. MR Zbl

[Freire 1995] A. Freire, "Uniqueness for the harmonic map flow from surfaces to general targets", *Comment. Math. Helv.* **70**:2 (1995), 310–338. Correction in **71**:2 (1996), 330-337. MR Zbl

[Gehrig 2020] M. I. Gehrig, *Prescribed curvature on the boundary of the disc*, Ph.D. thesis, ETH Zurich, 2020, available at https://doi.org/10.3929/ethz-b-000445412.

[Grüter et al. 1981] M. Grüter, S. Hildebrandt, and J. C. C. Nitsche, "On the boundary behavior of minimal surfaces with a free boundary which are not minima of the area", *Manuscripta Math.* **35**:3 (1981), 387–410. MR Zbl

[Hyder et al. 2022] A. Hyder, A. Segatti, Y. Sire, and C. Wang, "Partial regularity of the heat flow of half-harmonic maps and applications to harmonic maps with free boundary", *Comm. Partial Differential Equations* **47**:9 (2022), 1845–1882. MR Zbl

[Imbusch and Struwe 1999] C. Imbusch and M. Struwe, "Variational principles for minimal surfaces", pp. 477–498 in *Topics in nonlinear analysis*, edited by J. Escher and G. Simonett, Progr. Nonlinear Differential Equations Appl. **35**, Birkhäuser, Basel, 1999. MR Zbl

[Jost and Struwe 1990] J. Jost and M. Struwe, "Morse–Conley theory for minimal surfaces of varying topological type", *Invent. Math.* **102**:3 (1990), 465–499. MR Zbl

[Ladyzhenskaya 1963] O. A. Ladyzhenskaya, *The mathematical theory of viscous incompressible flow*, Gordon and Breach, New York, 1963. MR Zbl

[Lenzmann and Schikorra 2020] E. Lenzmann and A. Schikorra, "Sharp commutator estimates via harmonic extensions", *Nonlinear Anal.* **193** (2020), art. id. 111375. MR Zbl

[Mazowiecka and Schikorra 2018] K. Mazowiecka and A. Schikorra, "Fractional div-curl quantities and applications to nonlocal geometric equations", *J. Funct. Anal.* **275**:1 (2018), 1–44. MR Zbl

[Millot and Sire 2015] V. Millot and Y. Sire, "On a fractional Ginzburg–Landau equation and 1/2-harmonic maps into spheres", *Arch. Ration. Mech. Anal.* **215**:1 (2015), 125–210. MR Zbl

[Morse 1937] M. Morse, "Functional topology and abstract variational theory", *Ann. of Math.* (2) **38**:2 (1937), 386–449. MR Zbl

[Morse and Tompkins 1939] M. Morse and C. Tompkins, "The existence of minimal surfaces of general critical types", *Ann. of Math.* (2) **40**:2 (1939), 443–472. MR Zbl

[Moser 2011] R. Moser, "Intrinsic semiharmonic maps", *J. Geom. Anal.* **21**:3 (2011), 588–598. MR Zbl

[Radó 1930] T. Radó, "On Plateau's problem", *Ann. of Math.* (2) **31**:3 (1930), 457–469. MR

[Rivière 1993] T. Rivière, *Applications harmoniques entre variétés*, Ph.D. thesis, Université Paris 6, 1993, available at https://people.math.ethz.ch/~triviere/pdf/pub/flot-harmonique-93.pdf.

[Rupflin 2017] M. Rupflin, "Teichmüller harmonic map flow from cylinders", *Math. Ann.* **368**:3-4 (2017), 1227–1276. MR Zbl

[Rupflin and Schrecker 2018] M. Rupflin and M. R. I. Schrecker, "Analysis of boundary bubbles for almost minimal cylinders", *Calc. Var. Partial Differential Equations* **57**:5 (2018), art. id. 121. MR Zbl

[Rupflin and Topping 2019] M. Rupflin and P. M. Topping, "Global weak solutions of the Teichmüller harmonic map flow into general targets", *Anal. PDE* **12**:3 (2019), 815–842. MR Zbl

[Schoen and Uhlenbeck 1982] R. Schoen and K. Uhlenbeck, "A regularity theory for harmonic maps", *J. Differential Geom.* **17**:2 (1982), 307–335. MR Zbl

[Struwe 1984] M. Struwe, "On a critical point theory for minimal surfaces spanning a wire in $\mathbb{R}^n$", *J. Reine Angew. Math.* **349** (1984), 1–23. MR Zbl

[Struwe 1985] M. Struwe, "On the evolution of harmonic mappings of Riemannian surfaces", *Comment. Math. Helv.* **60**:4 (1985), 558–581. MR Zbl

[Struwe 1986] M. Struwe, "A Morse theory for annulus-type minimal surfaces", *J. Reine Angew. Math.* **368** (1986), 1–27. MR Zbl

[Struwe 1988] M. Struwe, *Plateau's problem and the calculus of variations*, Math. Notes **35**, Princeton Univ. Press, 1988. MR Zbl

[Topping 2002] P. Topping, "Reverse bubbling and nonuniqueness in the harmonic map flow", *Int. Math. Res. Not.* **2002**:10 (2002), 505–520. MR Zbl

[Tromba 1984] A. J. Tromba, "Degree theory on oriented infinite-dimensional varieties and the Morse number of minimal surfaces spanning a curve in $\mathbb{R}^n$", *Manuscripta Math.* **48**:1-3 (1984), 139–161. MR Zbl

[Tromba 1985] A. J. Tromba, "Degree theory on oriented infinite-dimensional varieties and the Morse number of minimal surfaces spanning a curve in $\mathbb{R}^n$, I: $n \geq 4$", *Trans. Amer. Math. Soc.* **290**:1 (1985), 385–413. MR Zbl

[Wettstein 2021] J. Wettstein, "Existence, uniqueness and regularity of the fractional harmonic gradient flow in general target manifolds", preprint, 2021. arXiv 2109.11458

[Wettstein 2022] J. Wettstein, "Uniqueness and regularity of the fractional harmonic gradient flow in S^{n-1}", *Nonlinear Anal.* **214** (2022), art. id. 112592. MR Zbl

[Wettstein 2023] J. D. Wettstein, "Half-harmonic gradient flow: aspects of a non-local geometric PDE", *Math. Eng.* **5**:3 (2023), art. id. 058. MR

Received 3 Feb 2022. Revised 14 Jul 2022. Accepted 19 Aug 2022.

MICHAEL STRUWE: michael.struwe@math.ethz.ch
Departement Mathematik, ETH Zürich, Zürich, Switzerland

ANALYSIS AND PDE
Vol. 17 (2024), No. 4, pp. 1439–1471
DOI: 10.2140/apde.2024.17.1439

NONCOMMUTATIVE MAXIMAL OPERATORS WITH ROUGH KERNELS

XUDONG LAI

This paper is devoted to the study of noncommutative maximal operators with rough kernels. More precisely, we prove the weak-type (1, 1) boundedness for noncommutative maximal operators with rough kernels. The proof of the weak-type (1, 1) estimate is based on the noncommutative Calderón–Zygmund decomposition. To deal with the rough kernel, we use the microlocal decomposition in the proofs of both the bad and good functions.

1. Introduction and state of main result

In recent years, there has been extensive research on noncommutative harmonic analysis, especially on noncommutative Calderón–Zygmund theory; see, e.g., [Parcet 2009; Mei and Parcet 2009; Cadilhac 2018; Chen et al. 2013]. The main content of this topic is focused on investigating the boundedness property of various operators in harmonic analysis on the noncommutative L_p space. Due to the lack of commutativity (i.e., $ab=ba$ may not hold in general case), many problems in the study of noncommutative Calderón–Zygmund theory seem to be more difficult, for instance the weak-type (1, 1) bound of integral operators.

It is well known that the real-variable theory of classical harmonic analysis was initiated by A. P. Calderón and A. Zygmund [1952]. One of the remarkable techniques in [Calderón and Zygmund 1952] is the so-called Calderón–Zygmund decomposition, which is now a widely used method in harmonic analysis. This technique not only gives a real-variable method to show weak-type (1, 1) bounds of singular integrals, but also provides a basic idea of stopping-time arguments for many topics in harmonic analysis, such as the theory of Hardy and BMO spaces; see, e.g., [Grafakos 2014a; 2014b; Stein 1993]. The noncommutative Calderón–Zygmund decomposition was recently established in [Parcet 2009] via the theory of noncommutative martingales. With this tool, the weak-type (1, 1) bound theory of the standard Calderón–Zygmund operator was developed there. It was pointed out in [Parcet 2009] that the noncommutative Calderón–Zygmund decomposition and the related method should open a door to work for a more general class of operators. For the subsequent works related to weak-type (1, 1) problem and noncommutative Calderón–Zygmund decomposition, see [Mei and Parcet 2009; Cadilhac 2018; Caspers et al. 2018; 2019; Hong and Xu 2021; Hong et al. 2023; Cadilhac et al. 2022].

This work was supported by the National Natural Science Foundation of China (nos. 12322107 and 12271124), the Heilongjiang Provincial Natural Science Foundation of China (YQ2022A005), Fundamental Research Funds for the Central Universities (no. FRFCU5710050121).

MSC2020: primary 46L52, 42B25; secondary 46L51, 42B20.

Keywords: noncommutative L_p space, weak (1,1) bound, maximal operator, singular integral operator, rough kernel.

On the other hand, the classical theory of singular integral operators tells us that to ensure the weak-type (1, 1) bound of the Calderón–Zygmund operator, the regularity condition of kernel can be relaxed to the so-called Hörmander condition; see, e.g., [Hörmander 1960; Grafakos 2014a]. Moreover, Calderón and Zygmund [1956] further studied the singular integral operator with a rough homogeneous kernel defined by

$$\text{p.v.} \int_{\mathbb{R}^d} \frac{\Omega(x-y)}{|x-y|^d} f(y)\,dy \tag{1-1}$$

and established its L_p boundedness for all $1 < p < \infty$. For its weak-type (1, 1) boundedness, it was quite later showed by S. Hofmann [1988] (and independently by M. Christ and Rubio de Francia [1988]) in two dimensions and by A. Seeger [1996] in higher dimensions (see further results in [Tao 1999]). Therefore a natural question inspired by [Parcet 2009] is whether can we weaken the Lipschitz regularity of kernel to the Hörmander condition or even rough homogeneous kernel. This problem has been open since then. The purpose of this paper is to develop some theory in this aspect for a class of rough operators. We consider the most fundamental operator: the maximal operator with a rough kernel which is defined by (in the sense of classical harmonic analysis)

$$M_\Omega f(x) = \sup_{r>0} |M_r f(x)|, \qquad M_r f(x) = \frac{1}{|B(x,r)|} \int_{B(x,r)} \Omega(x-y) f(y)\,dy, \tag{1-2}$$

where $B(x,r)$ is a ball in $\mathbb{R}^d$ with center x and radius r, the kernel Ω is a homogeneous function defined on $\mathbb{R}^d \setminus \{0\}$ with degree zero, that is,

$$\Omega(rx') = \Omega(x') \quad \text{for any } r > 0 \text{ and } x' \in \boldsymbol{S}^{d-1}. \tag{1-3}$$

Notice that the maximal operator M_Ω is a generalization of the Hardy–Littlewood maximal operator (by setting Ω as a constant, M_Ω is exactly the Hardy–Littlewood maximal operator). M_Ω is very important in the theory of rough singular integrals since it could be used to control many operators with rough kernels, just like the Hardy–Littlewood maximal operator plays an important role in analysis. By the method of rotation, it is easy to see that M_Ω is bounded on $L_p(\mathbb{R}^d)$ for all $1 < p \le \infty$ if $\Omega \in L_1(\boldsymbol{S}^{d-1})$; see, e.g., [Grafakos 2014a]. However, the weak-type (1, 1) boundedness of M_Ω is quite challenging. It was proved by Christ [1988] that M_Ω is of weak-type (1, 1) if $\Omega \in L_q(\boldsymbol{S}^1)$ with $1 < q \le \infty$ in two dimensions. Later Christ and Rubio de Francia [1988] showed in higher dimensions M_Ω is weak-type (1, 1) bounded if $\Omega \in L \log^+ L(\boldsymbol{S}^{d-1})$ by a depth investigation of the geometry in Euclidean space. For more topics, including open problems related to the maximal operator M_Ω, we refer to the reader to [Stein 1998; Grafakos and Stefanov 1999; Grafakos et al. 2017].

The noncommutative version of M_Ω should be important in the theory of noncommutative rough singular integral operators as expected. For instance, the noncommutative M_Ω will play a crucial role in the study of the noncommutative maximal operator of truncated operator in (1-1). In this paper, we will study the boundedness of M_Ω on the noncommutative L_p space for $1 \le p \le \infty$. In a special case Ω is a constant (i.e., M_Ω is the Hardy–Littlewood maximal operator), T. Mei [2007] investigated its noncommutative L_p, $1 < p \le \infty$, and weak-type (1, 1) boundedness. For general kernel Ω, there is no proper theory for the noncommutative M_Ω. To illustrate our noncommutative result of M_Ω, we should first give some basic notation.

Let us first introduce the noncommutative L_p space. Let $\mathcal{M}$ be a semifinite von Neumann algebra equipped with a normal semifinite faithful (n.s.f. for short) trace τ. We consider the algebra $\mathcal{A}_B$ of essentially bounded $\mathcal{M}$-valued functions

$$\mathcal{A}_B = \Big\{ f : \mathbb{R}^d \to \mathcal{M} \mid f \text{ is strong measurable such that } \operatorname*{ess\,sup}_{x\in\mathbb{R}^d} \|f(x)\|_{\mathcal{M}} < \infty \Big\}.$$

equipped with the n.s.f. trace $\varphi(f) = \int_{\mathbb{R}^d} \tau(f(x))\,dx$. Let $\mathcal{A}$ be the weak-operator closure of $\mathcal{A}_B$. Then $\mathcal{A}$ is a von Neumann algebra. For $1 \le p < \infty$, define $L_p(\mathcal{M})$ as the noncommutative L_p space associated to the pairs $(\mathcal{M}, \tau)$ with the L_p norm given by $\|x\|_{L_p(\mathcal{M})} = (\tau(|x|^p))^{1/p}$. The space $L_p(\mathcal{A})$ is defined as the closure of $\mathcal{A}_B$ with respect to the norm

$$\|f\|_{L_p(\mathcal{A})} = \left(\int_{\mathbb{R}^d} \tau(|f(x)|^p)\,dx \right)^{1/p}, \tag{1-4}$$

which means that $L_p(\mathcal{A})$ is the noncommutative L_p space associated to the pairs $(\mathcal{A}, \varphi)$. On the other hand, from (1-4) we see that $L_p(\mathcal{A})$ is isometric to the Bochner L_p space with values in $L_p(\mathcal{M})$. For convenience, we set $L_\infty(\mathcal{M}) = \mathcal{M}$ and $L_\infty(\mathcal{A}) = \mathcal{A}$ equipped with the operator norm. The lattices of projections are written as $\mathcal{M}_\pi$ and $\mathcal{A}_\pi$, while $1_{\mathcal{M}}$ and $1_{\mathcal{A}}$ stand for the unit elements. Let $L_p^+(\mathcal{A})$ be the positive part of $L_p(\mathcal{A})$. A lot of basic properties of classical L_p spaces, such as Minkowski's inequality, Hölder's inequality, the dual property, real and complex interpolation, etc., have been transferred to this noncommutative setting. We refer to the very detailed introduction in [Parcet 2009] or the survey article [Pisier and Xu 2003] for more about the noncommutative L_p space, the noncommutative $L_{1,\infty}$ space and related topics.

We next define a noncommutative analogue of M_Ω. For two general elements belong to a von Neumann algebra, they may not be comparable (i.e., neither $a < b$ nor $a \ge b$ holds for $a, b \in \mathcal{A}$). Hence it is difficult to define the noncommutative maximal function directly. This obstacle could be overcome by straightforwardly defining the maximal weak-type (1, 1) norm or L_p norm. We adopt the definition of the noncommutative maximal norm introduced by G. Pisier [1998] and M. Junge [2002].

Definition 1.1. For any index set I, we define $L_p(\mathcal{M}; \ell_\infty(I))$, the space of all sequences $x = \{x_n\}_{n\in I}$ in $L_p(\mathcal{M})$ which admit a factorization of the following form: there exist $a, b \in L_{2p}(\mathcal{M})$ and a bounded sequence $y = \{y_n\}_{n\in I}$ in $L_\infty(\mathcal{M})$ such that $x_n = a y_n b$ for all $n \in I$. The norm of x in $L_p(\mathcal{M}; \ell_\infty(I))$ is given by

$$\|\{x_k\}_{k\in I}\|_{L_p(\mathcal{M};\ell_\infty(I))} = \inf\Big\{ \|a\|_{L_{2p}(\mathcal{M})} \sup_{n\in I} \|y_n\|_{L_\infty(\mathcal{M})} \|b\|_{L_{2p}(\mathcal{M})} \Big\},$$

where the infimum is taken over all factorizations of x as above. We define a sequence $x = \{x_k\}_{k\in I}$ in $L_{1,\infty}(\mathcal{M})$ with quasinorm given by

$$\|\{x_k\}_{k\in I}\|_{\Lambda_{1,\infty}(\mathcal{M};\ell_\infty(I))} = \sup_{\lambda>0} \lambda \inf_{e\in\mathcal{M}_\pi} \{\tau(e^\perp) : \|e x_k e\|_\infty \le \lambda \text{ for all } k \in I\}.$$

If $x = \{x_n\}_{n\in I}$ is a sequence of positive elements, then $x \in L_p(\mathcal{M}; \ell_\infty(I))$ if and only if there exists a positive element $a \in L_p(\mathcal{M})$ such that $0 < x_n \le a$, and

$$\|x\|_{L_p(\mathcal{M};\ell_\infty(I))} = \inf\{\|a\|_{L_p(\mathcal{M})} : 0 < x_n \le a \text{ for all } n \in I\}, \tag{1-5}$$

$$\|(x_k)_{k\in I}\|_{\Lambda_{1,\infty}(\mathcal{M};\ell_\infty(I))} = \sup_{\lambda>0} \lambda \inf_{e\in\mathcal{M}_\pi} \{\tau(e^\perp) : e x_n e \le \lambda \text{ for all } n\}. \tag{1-6}$$

Now we can state our main result as follows.

Theorem 1.2. *Suppose that* Ω *satisfies* (1-3) *and* $\Omega \in L(\log^+ L)^2(S^{d-1})$. *Then the operator sequence* $\{M_r\}_{r>0}$ *is of maximal weak-type* $(1, 1)$, *i.e.*,

$$\|\{M_r f\}_{r>0}\|_{\Lambda_{1,\infty}(\mathcal{A},\ell_\infty(0,\infty))} \lesssim \mathcal{C}_\Omega \|f\|_{L_1(\mathcal{A})},$$

where $\mathcal{C}_\Omega$ *is a constant depending only on the dimension* d *and* Ω. *Equivalently, for every* $f \in L_1(\mathcal{A})$ *and* $\lambda > 0$, *there exists a projection* $e \in \mathcal{A}_\pi$ *such that*

$$\sup_{r>0} \|e M_r f e\|_{L_\infty(\mathcal{A})} \le \lambda \quad \text{and} \quad \lambda\varphi(1_{\mathcal{A}} - e) \lesssim \mathcal{C}_\Omega \|f\|_{L_1(\mathcal{A})}.$$

It is very easy to show that $\{M_r\}_{r>0}$ is of maximal strong type (p, p) for $1 < p \le \infty$ by the method of rotation. For completeness, we give a proof in the Appendix for this result.

The strategy in the proof of Theorem 1.2 is as follows. Firstly we convert the study of the maximal operator to a linearized singular integral operator T with a rough kernel (see Section 2). Secondly, to prove the weak-type $(1, 1)$ bound of this singular integral operator T, we use the noncommutative Calderón–Zygmund decomposition to split the function f into two parts: good functions and bad functions (see Section 3A). Roughly speaking, the proof is reduced to obtain some decay estimates for the good and bad functions separately. For the proof related to the bad functions, since the kernel Ω is rough, we will use a further decomposition, the so-called microlocal decomposition, to the operator T_j. Then we apply the L_2 norm and the L_1 norm to control the weak-type estimate (see Lemmas 3.4 and 3.5), where vector-valued Plancherel's theorem and an orthogonal geometric argument are involved in the proof of the L_2 estimate and the stationary phase method is used in the L_1 estimate (see Section 3B and the proofs in Section 4). For the proof of good functions we use the so-called pseudolocalization arguments to obtain some decay estimate for the L_2 norm of the singular integral operator T outside the support of functions on which it acts. To get such decay estimates, we adopt a similar method (the microlocal decomposition) from the proof of bad functions (see Section 3C).

In the classical Calderón–Zygmund decomposition, one can easily deal with the good function by the L_2 estimate. However, the proof of good functions from the noncommutative Calderón–Zygmund is much elaborated as showed in the case of smooth kernel by J. Parcet [2009]. In this paper, to overcome the nonsmoothness of kernel, we use the microlocal decomposition in the proofs of both bad and good functions. To the best knowledge of the author, this method seems to be new in the noncommutative Calderón–Zygmund theory. We should point out that the proof of bad functions is quite different from that in the classical case of[Christ and Rubio de Francia 1988], where they used the TT^* argument to obtain some regularity of the kernel $T_j T_j^*$ by some depth geometry but without using the Fourier transform. However, our method presented in this paper heavily depends on the Fourier transform where Plancherel's theorem and the stationary phase method are involved. These ideas are mainly inspired by [Seeger 1996]. Recall the following important pointwise property is crucial in the classical TT^* argument: $|Q|^{-1} \int_Q |b_Q(y)|\, dy \lesssim \lambda$, where b_Q is a bad function from the Calderón–Zygmund decomposition which is supported in a cube Q. Since in the noncommutative setting such kind of inequality may not hold for the off-diagonal terms of bad functions, our noncommutative TT^* argument is more complicated than that of the classical case. In fact only one pointwise property holds in the noncommutative Calderón–Zygmund decomposition: $q_k f_k q_k \lesssim \mathcal{C}_\Omega^{-1} \lambda q_k$ (see Lemma 3.1) and all pointwise estimates in the proof should finally be transferred to this property (see Section 4B for the details in the proof of the L_2 estimate).

This paper is organized as follows. First the study for maximal operator of M_r is reduced to a linearized singular integral operator in Section 2. In Section 3, by the noncommutative Calderón–Zygmund decomposition and microlocal decomposition, we finish the proof of our main theorem based on the estimates of bad and good functions. The proofs of lemmas related to the bad functions are all presented in Section 4. In Section 5, we give all proofs of lemmas related to the good functions. Finally in the Appendix, we give a proof of strong type (p, p), $1 < p \le \infty$, for $\{M_r\}_{r>0}$.

Further remark. After we finished this manuscript, L. Cadilhac [2022] found a more efficient noncommutative Calderón–Zygmund decomposition (see also [Hong et al. 2023]) so that the off-diagonal terms of the good functions vanish and the argument for the pseudolocalization can be avoided. Of course using this new Calderón–Zygmund decomposition, we only need to apply the L_2 estimate to deal with the good function and the proof related to the good functions in this paper can be greatly shortened. However, we point out that using this new method, the proof for the bad functions will be significantly more complicated than our arguments presented in this paper. So our proof in this paper still has its own interest. Nevertheless, we hope to show this in the study of weak-type (1, 1) boundedness for singular integral operators with rough kernels (1-1) which is our ongoing work.

Notation. Throughout this paper, we only consider the dimension $d \ge 2$ and the letter C stands for a positive finite constant which is independent of the essential variables, not necessarily the same one in each occurrence. $A \lesssim B$ means $A \le CB$ for some constant C. By the notation C_ε we mean that the constant depends on the parameter ε, $A \approx B$ means that $A \lesssim B$ and $B \lesssim A$, $\mathbb{Z}_+$ denotes the set of all nonnegative integers and

$$\mathbb{Z}_+^d = \underbrace{\mathbb{Z}_+ \times \cdots \times \mathbb{Z}_+}_{d}.$$

For $\alpha \in \mathbb{Z}_+^d$ and $x \in \mathbb{R}^d$, we define $x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}$ and $|x|$ denotes the ℓ_2 norm. For all $s \in \mathbb{R}_+$, $[s]$ denotes the integer part of s. For any set A with finite elements, we define $\operatorname{card}(A)$ or $\#(A)$ as the number of elements in A. Let $s \ge 0$, we define

$$\|\Omega\|_{L(\log^+ L)^s} := \int_{\boldsymbol{S}^{d-1}} |\Omega(\theta)| [\log(2 + |\Omega(\theta)|)]^s \, d\sigma(\theta),$$

where $d\sigma(\theta)$ denotes the sphere measure of $\boldsymbol{S}^{d-1}$. When $s = 0$, we use the standard notation $\|\Omega\|_1 := \|\Omega\|_{L(\log^+ L)^0}$.

Define $\mathcal{F} f$ (or $\hat{f}$) and $\mathcal{F}^{-1} f$ (or $\check{f}$) the Fourier transform and the inverse Fourier transform of f by

$$\mathcal{F} f(\xi) = \int_{\mathbb{R}^d} e^{-i\langle x, \xi\rangle} f(x)\, dx, \quad \mathcal{F}^{-1} f(\xi) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{i\langle x, \xi\rangle} f(x)\, dx.$$

Let $\mathcal{Q}$ be the set of all dyadic cubes in $\mathbb{R}^d$. For any $Q \in \mathcal{Q}$, denote by $\ell(Q)$ the side length of the cube Q. Let sQ be the cube with the same center of Q such that $\ell(sQ) = s\ell(Q)$. Given an integer $k \in \mathbb{Z}$,

$\mathcal{Q}_k$ will be defined as the set of dyadic cubes of side length 2^{-k}. Let $|Q|$ be the volume of the cube Q. If $Q \in \mathcal{Q}$ and $f : \mathbb{R}^d \to \mathcal{M}$ is integrable on Q, we define its average as $f_Q = |Q|^{-1} \int_Q f(y)\,dy$.

For $k \in \mathbb{Z}$, set σ_k as the k-th dyadic σ-algebra, i.e., σ_k is generated by the dyadic cubes with side lengths equal to 2^{-k}. Let E_k be the conditional expectation which is associated to the classical dyadic filtration σ_k on $\mathbb{R}^d$. We also use E_k for the tensor product $\mathsf{E}_k \otimes \mathrm{id}_{\mathcal{M}}$ acting on $\mathcal{A}$. Then for $1 \le p < \infty$ and $f \in L_p(\mathcal{A})$, we get

$$\mathsf{E}_k(f) = \sum_{Q \in \mathcal{Q}_k} f_Q \chi_Q,$$

where χ_Q is the characteristic function of Q. Similarly, $\{\mathcal{A}_k\}_{k\in\mathbb{Z}}$ will stand for the corresponding filtration, i.e., $\mathcal{A}_k = \mathsf{E}_k(\mathcal{A})$. For simplicity, we will define the conditional expectation $f_k := \mathsf{E}_k(f)$ and the martingale difference $\Delta_k(f) := f_k - f_{k-1} =: df_k$.

2. Reduction to singular integral operator

In this section, we reduce the study of maximal operator of M_r to a singular integral operator with a rough kernel. This will be done by several steps as follows.

Step 1: By decomposing the functions Ω and f into four parts (i.e., real positive part, real negative part, imaginary positive part, imaginary negative part), together with the quasitriangle inequality for the quasi-norm $\|\cdot\|_{\Lambda_{1,\infty}(\mathcal{A},\ell_\infty(0,\infty))}$, we only consider the case that Ω is a positive function and f is positive in $\mathcal{A}$. Then by (1-6), it is enough to show that for any $f \in L_1^+(\mathcal{A})$ and $\lambda > 0$ there exists a projection $e \in \mathcal{A}_\pi$ such that

$$eM_r f e \le \lambda \quad \text{for all } r > 0 \qquad \text{and} \qquad \lambda\varphi(1_{\mathcal{A}} - e) \lesssim C_\Omega \|f\|_{L_1(\mathcal{A})}. \tag{2-1}$$

Step 2: Next we show that the study of M_r can be reduced to a dyadic smooth operator. More precisely, let ϕ be a $C_c^\infty(\mathbb{R}^d)$, radial, positive function which is supported in $\{x \in \mathbb{R}^d : \frac{1}{2} \le |x| \le 2\}$ and $\sum_{i\in\mathbb{Z}} \phi_j(x) = 1$ for all $x \in \mathbb{R}^d \setminus \{0\}$, where $\phi_j(x) = \phi(2^{-j}x)$. Define an operator $\mathfrak{M}_j$ by

$$\mathfrak{M}_j f(x) = \int_{\mathbb{R}^d} \frac{\Omega(x-y)}{|x-y|^d} \phi_j(x-y) f(y)\,dy.$$

We will prove that the maximal operator of $\mathfrak{M}_j$ is of weak-type $(1,1)$ below and (2-1) follows from it.

Theorem 2.1. *Let Ω be a positive function satisfying (1-3) and $\Omega \in L(\log^+ L)^2(S^{d-1})$. For any $f \in L_1^+(\mathcal{A})$, $\lambda > 0$, there exists a projection $e \in \mathcal{A}_\pi$ such that*

$$\sup_{j\in\mathbb{Z}} \|e\mathfrak{M}_j f e\|_{L_\infty(\mathcal{A})} \lesssim \lambda, \quad \lambda\varphi(1_{\mathcal{A}} - e) \lesssim C_\Omega \|f\|_{L_1(\mathcal{A})},$$

where the constant C_Ω only depends on Ω and the dimension.

The proof of Theorem 2.1 will be given later. We apply Theorem 2.1 to show (2-1). Let Ω be a positive function and f be positive in $L_1^+(\mathcal{A})$. Then by our choice of ϕ_j, for any $r > 0$, we have

$$\begin{aligned} M_r f(x) &= \frac{1}{|B(x,r)|} \int_{B(x,r)} \Omega(x-y) f(y)\,dy \\ &= \frac{C_d}{r^d} \sum_{j \le [\log r]+1} \int_{|x-y|\le r} \phi_j(x-y)\Omega(x-y) f(y)\,dy \lesssim \frac{1}{r^d} \sum_{j\le[\log r]+1} 2^{jd} \mathfrak{M}_j f(x). \end{aligned}$$

Notice that Ω is positive and $f \in L_1^+(\mathcal{A})$; thus the inequality $e\mathfrak{M}_j f e \le \lambda$ is equivalent to $\|e\mathfrak{M}_j f e\|_{\mathcal{A}} = \|e\mathfrak{M}_j f e\|_{L_\infty(\mathcal{A})} \le \lambda$. By Theorem 2.1, there exists a projection $e \in \mathcal{A}_\pi$ such that

$$e\mathfrak{M}_j f e \lesssim \lambda \quad \text{for all } j \in \mathbb{Z}, \qquad \lambda\varphi(1_{\mathcal{A}} - e) \lesssim C_\Omega \|f\|_{L_1(\mathcal{A})}.$$

Then it is easy to see that, for any $r > 0$,

$$eM_r f e \lesssim \frac{1}{r^d} \sum_{j \le [\log r]+1} 2^{jd} e\mathfrak{M}_j f e \lesssim \lambda.$$

Step 3: We will reduce the study of the maximal operator of $\mathfrak{M}_j$ to a class of square functions. Notice that the kernel Ω of $\mathfrak{M}_j$ has no cancellation. Formally we cannot study the operator $\big(\sum_j |\mathfrak{M}_j|^2\big)^{1/2}$ directly since it may not even be L_2 bounded. To avoid such case, we define a new operator T_j which is a modified version of the operator $\mathfrak{M}_j$

$$T_j f(x) = \int_{\mathbb{R}^d} \phi_j(x-y) \frac{\widetilde{\Omega}(x-y)}{|x-y|^d} f(y)\, dy, \tag{2-2}$$

where

$$\widetilde{\Omega}(x) = \Omega(x) - \frac{1}{\sigma_{d-1}} \int_{S^{d-1}} \Omega(\theta)\, d\sigma(\theta)$$

and σ_{d-1} is measure of the unit sphere. Then it is easy to see that $\widetilde{\Omega}$ has mean value zero over $\boldsymbol{S}^{d-1}$. Then formally the study of the maximal operator of $\mathfrak{M}_j$ may follow from that of the square function $\big(\sum_j |T_j|^2\big)^{1/2}$ and the maximal operator. In the following we use rigorous noncommutative language to explain how to do it. To define a noncommutative square function, we should first introduce the so-called column and row function space. Let $\{f_j\}_j$ be a finite sequence in $L_p(\mathcal{A})$, $1 \le p \le \infty$. Define

$$\|\{f_j\}_j\|_{L_p(\mathcal{A};\ell_2^r)} = \bigg\|\Big(\sum |f_j^*|^2\Big)^{1/2}\bigg\|_{L_p(\mathcal{A})}, \qquad \|(f_j)\|_{L_p(\mathcal{A};\ell_2^c)} = \bigg\|\Big(\sum |f_j|^2\Big)^{1/2}\bigg\|_{L_p(\mathcal{A})}.$$

This procedure is also used to define the spaces $L_{1,\infty}(\mathcal{A};\, \ell_2^r)$ and $L_{1,\infty}(\mathcal{A};\, \ell_2^c)$; i.e.,

$$\|\{f_j\}_j\|_{L_{1,\infty}(\mathcal{A};\ell_2^r)} = \bigg\|\Big(\sum |f_j^*|^2\Big)^{1/2}\bigg\|_{L_{1,\infty}(\mathcal{A})}, \qquad \|\{f_j\}\|_{L_{1,\infty}(\mathcal{A};\ell_2^c)} = \bigg\|\Big(\sum |f_j|^2\Big)^{1/2}\bigg\|_{L_{1,\infty}(\mathcal{A})}.$$

Let $L_{1,\infty}(\mathcal{A},\, \ell_2^{rc})$ space be the weak-type square function of $\{T_j\}_j$ defined as

$$\|\{T_j\}_j\|_{L_{1,\infty}(\mathcal{A},\ell_2^{rc})} = \inf_{T_j f = g_j + h_j} \{\|\{g_j\}_j\|_{L_{1,\infty}(\mathcal{A},\ell_2^c)} + \|\{h_j\}_j\|_{L_{1,\infty}(\mathcal{A};\ell_2^r)}\}.$$

We have the following weak-type (1, 1) estimate of square function of $\{T_j\}_j$.

Theorem 2.2. *Suppose that Ω satisfies (1-3) and $\Omega \in L(\log^+ L)^2(\boldsymbol{S}^{d-1})$. Let T_j be defined in (2-2). Then we have*

$$\|\{T_j\}_j\|_{L_{1,\infty}(\mathcal{A},\ell_2^{rc})} \lesssim C_\Omega \|f\|_{L_1(\mathcal{A})},$$

where the constant C_Ω only depends on Ω and the dimension.

In the following we use Theorem 2.2 to prove Theorem 2.1. Our goal is to find a projection $e \in \mathcal{A}_\pi$ such that

$$\sup_{j\in\mathbb{Z}} \|e\mathfrak{M}_j f e\|_{L_\infty(\mathcal{A})} \lesssim \lambda, \quad \lambda\varphi(1_{\mathcal{A}} - e) \lesssim C_\Omega \|f\|_{L_1(\mathcal{A})}.$$

We first decompose $\mathfrak{M}_j f$ into two parts

$$T_j f(x) + \frac{1}{\sigma_{d-1}} \int_{S^{d-1}} \Omega(\theta)\, d\sigma(\theta) \int_{\mathbb{R}^d} \frac{\phi_j(x-y)}{|x-y|^d} f(y)\, dy =: T_j f(x) + \widetilde{M}_j f(x).$$

Notice that $(1/\sigma_{d-1}) \int_{S^{d-1}} \Omega(\theta)\, d\sigma(\theta)$ is a harmless constant which is bounded by $\|\Omega\|_1$. By using the fact that the noncommutative Hardy–Littlewood maximal operator is of weak-type (1, 1) (see, e.g., [Mei 2007]), it is not difficult to see that the maximal operator of $\widetilde{M}_j$ is of weak-type (1, 1). Thus we can find a projection $e_1 \in \mathcal{A}_\pi$ such that

$$\sup_{j\in\mathbb{Z}} \|e_1 \widetilde{M}_j f e_1\|_{L_\infty(\mathcal{A})} \le \lambda \quad \text{and} \quad \lambda\varphi(1_{\mathcal{A}} - e_1) \lesssim \|\Omega\|_1 \|f\|_{L_1(\mathcal{A})}.$$

Next we utilize Theorem 2.2 to construct other projection. By the definition of infimum, there exists a decomposition $T_j f = g_j + h_j$ satisfying

$$\|\{g_j\}\|_{L_{1,\infty}(\mathcal{A};\ell_2^c)} + \|\{h_j\}\|_{L_{1,\infty}(\mathcal{A};\ell_2^r)} \le \tfrac{1}{2}\mathcal{C}_\Omega \|f\|_{L_1(\mathcal{A})}.$$

We now take $e_2 = \chi_{(0,\lambda]}\big(\big(\sum_{j\in\mathbb{Z}} |g_j|^2\big)^{1/2}\big)$ and $e_3 = \chi_{(0,\lambda]}\big(\big(\sum_{j\in\mathbb{Z}} |h_j^*|^2\big)^{1/2}\big)$. Then

$$\bigg\| \bigg(\bigg(\sum_{j\in\mathbb{Z}} |g_j|^2\bigg)^{1/2} \bigg) e_2 \bigg\|_{L_\infty(\mathcal{A})} \le \lambda \quad \text{and} \quad \lambda\varphi(1_{\mathcal{A}} - e_2) \lesssim \mathcal{C}_\Omega \|f\|_{L_1(\mathcal{A})}.$$

Also for e_3, we have

$$\bigg\| \bigg(\bigg(\sum_{j\in\mathbb{Z}} |h_j^*|^2\bigg)^{1/2} \bigg) e_3 \bigg\|_{L_\infty(\mathcal{A})} \le \lambda \quad \text{and} \quad \lambda\varphi(1_{\mathcal{A}} - e_3) \lesssim \mathcal{C}_\Omega \|f\|_{L_1(\mathcal{A})}.$$

Let $e = e_1 \wedge e_2 \wedge e_3$. Then it is easy to see that

$$\sup_{j\in\mathbb{Z}} \|e \widetilde{M}_j f e\|_{L_\infty(\mathcal{A})} \le \lambda, \quad \lambda\varphi(1_{\mathcal{A}} - e) \lesssim \mathcal{C}_\Omega \|f\|_{L_1(\mathcal{A})}.$$

Hence to finish the proof of Theorem 2.1, it is sufficient to show

$$\sup_{j\in\mathbb{Z}} \|e T_j f e\|_{L_\infty(\mathcal{A})} \lesssim \lambda.$$

Recall the definition of $L_\infty(\mathcal{A})$, $\|f\|_{L_\infty(\mathcal{A})} = \|f\|_{\mathcal{A}}$. Then we get

$$\|e T_j f e\|_{L_\infty(\mathcal{A})} \le \|e g_j e\|_{\mathcal{A}} + \|e h_j e\|_{\mathcal{A}} = \|e g_j e\|_{\mathcal{A}} + \|e h_j^* e\|_{\mathcal{A}}.$$

Now using polar decomposition $g_j = u_j |g_j|$ and $h_j^* = v_j |h_j^*|$, we continue to estimate the above term as follows:

$$\begin{aligned}
\|e u_j |g_j| e\|_{\mathcal{A}} + \|e v_j |h_j^*| e\|_{\mathcal{A}} &\le \||g_j| e\|_{\mathcal{A}} + \||h_j^*| e\|_{\mathcal{A}} = \|e |g_j|^2 e\|_{\mathcal{A}}^{1/2} + \|e |h_j^*|^2 e\|_{\mathcal{A}}^{1/2} \\
&\le \bigg\| e \sum_{j\in\mathbb{Z}} |g_j|^2 e \bigg\|_{\mathcal{A}}^{1/2} + \bigg\| e \sum_{j\in\mathbb{Z}} |h_j^*|^2 e \bigg\|_{\mathcal{A}}^{1/2} \\
&= \bigg\| \bigg(\sum_{j\in\mathbb{Z}} |g_j|^2\bigg)^{1/2} e_2 e \bigg\|_{\mathcal{A}} + \bigg\| \bigg(\sum_{j\in\mathbb{Z}} |h_j^*|^2\bigg)^{1/2} e_3 e \bigg\|_{\mathcal{A}} \lesssim \lambda.
\end{aligned}$$

Hence we finish the proof of Theorem 2.1.

Step 4: We reduce the study of the square function to a linear operator. To simplify the notation, we still use Ω in (2-2), i.e.,

$$T_j f(x) = \int_{\mathbb{R}^d} K_j(x-y)\Omega(x-y) f(y)\, dy, \quad \text{with } K_j(x) = \phi_j(x)|x|^{-d}, \tag{2-3}$$

but we suppose that Ω satisfies the cancellation property $\int_{S^{d-1}} \Omega(\theta)\, d\sigma(\theta) = 0$. To linearize the square function, we use the following noncommutative Khintchine's inequality in $L_{1,\infty}(\mathcal{A}, \ell_2^{rc})$ which was recently established in [Cadilhac 2019].

Lemma 2.3. *Let $\{\varepsilon_j\}_j$ be a Rademacher sequence on a probability space $(\mathfrak{m}, P)$. Suppose that $f = \{f_j\}_j$ is a finite sequence in $L_{1,\infty}(\mathcal{A})$. Then we have*

$$\Big\| \sum_{j\in\mathbb{Z}} f_j \varepsilon_j \Big\|_{L_{1,\infty}(L_\infty(\mathfrak{m})\overline{\otimes}\mathcal{A})} \approx \|\{f_j\}_j\|_{L_{1,\infty}(\mathcal{A};\ell_2^{rc})}.$$

Now by the preceding lemma, Theorem 2.2 immediately follows from the result below.

Theorem 2.4. *Suppose that Ω satisfies (1-3), $\Omega \in L(\log^+ L)^2(S^{d-1})$ and the cancellation property $\int_{S^{d-1}} \Omega(\theta)\, d\sigma(\theta) = 0$. Let T_j be defined in (2-3). Assume $\{\varepsilon_j\}_j$ is the Rademacher sequence on a probability space $(\mathfrak{m}, P)$. Define $Tf(x,z) = \sum_j T_j f(x)\varepsilon_j(z)$ and the tensor trace $\tilde{\varphi} = \int_{\mathfrak{m}} \otimes \varphi$. Then T maps $L_1(\mathcal{A})$ to $L_{1,\infty}(L_\infty(\mathfrak{m}) \overline{\otimes} \mathcal{A})$, i.e., for any $\lambda > 0$, $f \in L_1(\mathcal{A})$,*

$$\lambda \tilde{\varphi}\{|Tf| > \lambda\} \lesssim C_\Omega \|f\|_{L_1(\mathcal{A})},$$

where the constant C_Ω only depends on Ω and the dimension.

At present our main result Theorem 1.2 is reduced to Theorem 2.4. In the rest of this paper, we give effort to the proof of Theorem 2.4.

3. Proof of Theorem 2.4

In this section we give the proof of Theorem 2.4 based on some lemmas; their proofs will be given in Sections 4 and 5. We first introduce the noncommutative Calderón–Zygmund decomposition.

3A. ***Noncommutative Calderón–Zygmund decomposition.*** By the standard density argument, we only need to consider the following dense class of $L_1(\mathcal{A})$:

$$\mathcal{A}_{c,+} = \{f : \mathbb{R}^d \to \mathcal{M} \mid f \in \mathcal{A}_+,\ \overrightarrow{\operatorname{supp}} f \text{ is compact}\}.$$

Here $\overrightarrow{\operatorname{supp}} f$ represents the support of f as an operator-valued function in $\mathbb{R}^d$, which means that $\overrightarrow{\operatorname{supp}} f = \{x \in \mathbb{R}^d : \|f(x)\|_{\mathcal{M}} \neq 0\}$. Let $\Omega \in L(\log^+ L)^2(S^{d-1})$. Set a constant

$$C_\Omega = \|\Omega\|_{L(\log^+ L)^2} + \int_{S^{d-1}} |\Omega(\theta)| \bigg(1 + \bigg[\log^+\bigg(\frac{|\Omega(\theta)|}{\|\Omega\|_1}\bigg)\bigg]^2\bigg)\, d\sigma(\theta), \tag{3-1}$$

where $\log^+ a = 0$ if $0 < a < 1$ and $\log^+ a = \log a$ if $a \geq 1$. Since $\|\Omega\|_{L(\log^+ L)^2} < +\infty$, one can easily check that C_Ω is a finite constant. Now we fix $f \in \mathcal{A}_{c,+}$, and set $f_k = \mathsf{E}_k f$ for all $k \in \mathbb{Z}$. Then the sequence $\{f_k\}_{k\in\mathbb{Z}}$ is a positive dyadic martingale in $L_1(\mathcal{A})$. Applying the so-called Cuculescu construction introduced in [Parcet 2009, Lemma 3.1] at level $\lambda\mathcal{C}_\Omega{}^{-1}$, we get the following result.

Lemma 3.1. *There exists a decreasing sequence* $\{q_k\}_{k\in\mathbb{Z}}$ *depending on* f *and* $\lambda\mathcal{C}_\Omega{}^{-1}$, *where* q_k *is a projection in* $\mathcal{A}_\pi$ *satisfying the following conditions*:

(i) q_k *commutes with* $q_{k-1}f_kq_{k-1}$ *for every* $k\in\mathbb{Z}$.

(ii) q_k *belongs to* $\mathcal{A}_k$ *for every* $k\in\mathbb{Z}$ *and* $q_kf_kq_k\le\lambda\mathcal{C}_\Omega{}^{-1}q_k$.

(iii) *Set* $q=\bigwedge_{k\in\mathbb{Z}}q_k$. *We have the inequality*

$$\varphi(1_{\mathcal{A}}-q)\le\lambda^{-1}\mathcal{C}_\Omega\|f\|_{L_1(\mathcal{A})}.$$

(iv) *The expression of* q_k *can be written as follows*: *for some negative integer* $m\in\mathbb{Z}$

$$q_k=\begin{cases}1_{\mathcal{A}} & \text{if } k<m,\\ \chi_{(0,\lambda\mathcal{C}_\Omega^{-1}]}(f_k) & \text{if } k=m,\\ \chi_{(0,\lambda\mathcal{C}_\Omega^{-1}]}(q_{k-1}f_kq_{k-1}) & \text{if } k>m.\end{cases}$$

Below we introduce another expression of the projection q_k given in the previous lemma as done in [Parcet 2009]. We point out that such kind of expression will be quite helpful when we give some estimates to the terms related to q_k. In fact we can write $q_k=\sum_{Q\in\mathcal{Q}_k}\xi_Q\chi_Q$ for all $k\in\mathbb{Z}$, where ξ_Q is a projection in $\mathcal{M}$ which satisfies the following conditions:

(i) ξ_Q has the following explicit expression: $\widehat{Q}$ below is the father dyadic cube of Q,

$$\xi_Q=\begin{cases}1_{\mathcal{M}} & \text{if } k<m,\\ \chi_{(0,\lambda\mathcal{C}_\Omega^{-1}]}(f_Q) & \text{if } k=m,\\ \chi_{(0,\lambda\mathcal{C}_\Omega^{-1}]}(\xi_{\widehat{Q}}f_Q\xi_{\widehat{Q}}) & \text{if } k>m.\end{cases}$$

(ii) $\xi_Q\in\mathcal{M}_\pi$ and $\xi_Q\le\xi_{\widehat{Q}}$.

(iii) ξ_Q commutes with $\xi_{\widehat{Q}}f_Q\xi_{\widehat{Q}}$ and $\xi_Qf_Q\xi_Q\le\mathcal{C}_\Omega^{-1}\lambda\xi_Q$.

Define the projection $p_k=q_{k-1}-q_k$. By applying the above more explicit expression, we see that p_k equals $\sum_{Q\in\mathcal{Q}_k}(\xi_{\widehat{Q}}-\xi_Q)\chi_Q=:\sum_{Q\in\mathcal{Q}_k}\pi_Q\chi_Q$, where $\pi_Q=\xi_{\widehat{Q}}-\xi_Q$. Then it is easy to see that all p_k's are pairwise disjoint and $\sum_{k\in\mathbb{Z}}p_k=1_{\mathcal{A}}-q$.

Now we define the associated good functions and bad functions related to f as follows:

$$f=g+b,\quad g=\sum_{i,j\in\widehat{\mathbb{Z}}}p_if_{i\vee j}p_j,\quad b=\sum_{i,j\in\widehat{\mathbb{Z}}}p_i(f-f_{i\vee j})p_j,$$

where we set $p_\infty=q$, $\widehat{\mathbb{Z}}=\mathbb{Z}\cup\{\infty\}$ and $i\vee j=\max(i,j)$. If i or j is infinite, $i\vee j$ is just ∞ and $f_\infty=f$ by definition. We further decompose g as the diagonal terms and the off-diagonal terms:

$$g_d=qfq+\sum_{k\in\mathbb{Z}}p_kf_kp_k,\quad g_{\rm off}=\sum_{i\ne j}p_if_{i\vee j}p_j+qf(1_{\mathcal{A}}-q)+(1_{\mathcal{A}}-q)fq.$$

The proofs for diagonal terms g_d and off-diagonal terms $g_{\rm off}$ will be different as we shall see below. For the bad function b, we can deal with the diagonal and off-diagonal terms uniformly. So it is unnecessary for us to decompose it as we did for the good functions. By the linearity of T, we get

$$\tilde{\varphi}(|Tf|>\lambda)\le\tilde{\varphi}\Big(|Tg|>\frac{\lambda}{2}\Big)+\tilde{\varphi}\Big(|Tb|>\frac{\lambda}{2}\Big).$$

In the following we give estimates for the good and bad functions, respectively. Before that we state a lemma to construct a projection in $\mathcal{A}$ such that the proof can be reduced to the case that the operators are restricted on this projection.

Lemma 3.2. *There exists a projection $\zeta\in\mathcal{A}_\pi$ which satisfies the following conditions*:

(i) $\lambda\varphi(1_{\mathcal{A}}-\zeta)\lesssim\mathcal{C}_\Omega\|f\|_{L_1(\mathcal{A})}$.

(ii) *If $Q_0\in\mathcal{Q}$ and $x\in(2^{101}+1)Q_0$, then $\zeta(x)\le 1_{\mathcal{M}}-\xi_{\widehat{Q}_0}+\xi_{Q_0}$ and $\zeta(x)\le\xi_{Q_0}$.*

The proof of this lemma can be easily modified from that of [Parcet 2009, Lemma 4.2]. Here the exact value of $2^{101}+1$ above is not essential and the reason we choose this value is just for convenience in a later calculation (see (3-3)). Now let us consider the bad functions first since our method presented here is also needed for the good functions.

3B. ***Estimates for the bad functions.*** We first use Lemma 3.2 to reduce the study of the operator T to that of $\zeta T\zeta$. Split Tb into four terms as follows:

$$(1_{\mathcal{A}}-\zeta)Tb(1_{\mathcal{A}}-\zeta)+\zeta Tb(1-\zeta)+(1-\zeta)Tb\zeta+\zeta Tb\zeta.$$

By the property (i) in Lemma 3.2, we get

$$\tilde{\varphi}\Big(|Tb|>\frac{\lambda}{2}\Big)\lesssim\varphi(1_{\mathcal{A}}-\zeta)+\tilde{\varphi}\Big(|\zeta Tb\zeta|>\frac{\lambda}{4}\Big)\lesssim\lambda^{-1}\mathcal{C}_\Omega\|f\|_{L_1(\mathcal{A})}+\tilde{\varphi}\Big(|\zeta Tb\zeta|>\frac{\lambda}{4}\Big).$$

Therefore it is enough to show that the term $\tilde{\varphi}(|\zeta Tb\zeta|>\lambda/4)$ satisfies our desired estimate. Recall the bad function

$$b=\sum_{k\in\mathbb{Z}}p_k(f-f_k)p_k+\sum_{s\ge1}\sum_{k\in\mathbb{Z}}p_k(f-f_{k+s})p_{k+s}+p_{k+s}(f-f_{k+s})p_k=:\sum_{s=0}\sum_{k\in\mathbb{Z}}b_{k,s},$$

where

$$b_{k,0}=p_k(f-f_k)p_k,\quad b_{k,s}=p_k(f-f_{k+s})p_{k+s}+p_{k+s}(f-f_{k+s})p_k. \tag{3-2}$$

By the definition of T, we further rewrite Tb as follows: for any $x\in\mathbb{R}^d$ and $z\in\mathfrak{m}$,

$$Tb(x,z)=\sum_{j\in\mathbb{Z}}T_j\Big[\sum_{s\ge0}\sum_{n\in\mathbb{Z}}b_{n-j,s}\Big](x)\varepsilon_j(z)=\sum_{s\ge0}\sum_{n\in\mathbb{Z}}\sum_{j\in\mathbb{Z}}T_jb_{n-j,s}(x)\varepsilon_j(z).$$

For any $Q\in\mathcal{Q}_{n-j+s}$, set $Q_{n-j}\in\mathcal{Q}_{n-j}$ as the s-th ancestor of Q. Consider x in the support of ζ (i.e., $\zeta(x)\neq0$) and let $n<100$. Then we get that, for all s, $\zeta(x)T_jb_{n-j,s}(x)\zeta(x)$ equals

$$\begin{aligned}&\sum_{Q\in\mathcal{Q}_{n-j+s}}\zeta(x)\int_Q K_j(x-y)b_{n-j,s}(y)\,dy\,\zeta(x)\\&\quad=\sum_{Q\in\mathcal{Q}_{n-j+s}}\zeta(x)\chi_{((2^{101}+1)Q_{n-j})^c}(x)\int_Q K_j(x-y)[\pi_{Q_{n-j}}(f(y)-f_Q)\pi_Q+\pi_Q(f(y)-f_Q)\pi_{Q_{n-j}}]\,dy\,\zeta(x)\\&\quad=0,\end{aligned}\tag{3-3}$$

where in the first equality we apply $\zeta(x)\pi_{Q_{n-j}} = 0$ if $x \in (2^{101}+1)Q_{n-j}$ by the property (ii) of ζ in Lemma 3.2 and the second inequality follows from the fact $x \in ((2^{101}+1)Q_{n-j})^c$ and $y \in Q$ implies that $|x-y| \geq 2^{100+j-n}$, which is a contradiction with the support of K_j and $n < 100$. Therefore we get

$$\zeta T b \zeta = \zeta \sum_{n\geq 100} \sum_{s\geq 0} \sum_{j\in\mathbb{Z}} T_j b_{n-j,s} \varepsilon_j \zeta.$$

Hence, to finish the proof related to the bad functions, it suffices to verify the following estimate:

$$\tilde{\varphi}\left(\left|\zeta \sum_{n\geq 100} \sum_{s\geq 0} \sum_{j\in\mathbb{Z}} T_j b_{n-j,s} \varepsilon_j \zeta\right| > \frac{\lambda}{4}\right) \lesssim \lambda^{-1} \mathcal{C}_\Omega \|f\|_{L_1(\mathcal{A})}. \tag{3-4}$$

Some important decompositions play key roles in the proof of (3-4). We present them by some lemmas, which will be proved in Section 4. It should be pointed out that the methods used here also work for the good functions, which will be clear in the next subsection.

The first lemma shows that (3-4) holds if Ω is restricted in some subset of $\boldsymbol{S}^{d-1}$. More precisely, for fixed $n \geq 100$ and $s \geq 0$, define

$$D^\iota = \{\theta \in \boldsymbol{S}^{d-1} : |\Omega(\theta)| \geq 2^{\iota(n+s)} \|\Omega\|_1\},$$

where $\iota > 0$ will be chosen later. Let $T^{n,s}_{j,\iota}$ be defined by

$$T^{n,s}_{j,\iota} h(x) = \int_{\mathbb{R}^d} \Omega \chi_{D^\iota}\left(\frac{x-y}{|x-y|}\right) K_j(x-y) \cdot h(y)\, dy. \tag{3-5}$$

Lemma 3.3. *Suppose $\Omega \in L(\log^+ L)^2(\boldsymbol{S}^{d-1})$. With all the notation above, we get*

$$\tilde{\varphi}\left(\left|\zeta \sum_{n\geq 100} \sum_{s\geq 0} \sum_{j\in\mathbb{Z}} T^{n,s}_{j,\iota} b_{n-j,s} \varepsilon_j \zeta\right| > \frac{\lambda}{8}\right) \lesssim \lambda^{-1} \mathcal{C}_\Omega \|f\|_{L_1(\mathcal{A})}.$$

Thus, by Lemma 3.3, to finish the proof for bad functions, it suffices to verify (3-4) under the condition that for fixed $n \geq 100$ and $s \geq 0$ the kernel function Ω satisfies $\|\Omega\|_\infty \leq 2^{\iota(n+s)}\|\Omega\|_1$ in each T_j.

In the following, we introduce the *microlocal decomposition* of kernel. To do this, we give a partition of unity on the unit surface $\boldsymbol{S}^{d-1}$. Let $k \geq 100$. Choose $\{e^k_v\}_{v\in\Theta_k}$ be a collection of unit vectors on $\boldsymbol{S}^{d-1}$ which satisfies the following two conditions:

(a) $|e^k_v - e^k_{v'}| \geq 2^{-k\gamma-4}$ if $v \neq v'$.

(b) If $\theta \in \boldsymbol{S}^{d-1}$, there exists an e^k_v such that $|e^k_v - \theta| \leq 2^{-k\gamma-3}$.

The constant $0 < \gamma < 1$ in (a) and (b) will be chosen later. To choose such an $\{e^k_v\}_{v\in\Theta_k}$, we simply take a maximal collection $\{e^k_v\}_v$ for which (a) holds and then (b) holds automatically by the maximality. Notice that there are $C2^{k\gamma(d-1)}$ elements in the collection $\{e^k_v\}_v$. For every $\theta \in \boldsymbol{S}^{d-1}$, there only exist finite e^k_v such that $|e^k_v - \theta| \leq 2^{-k\gamma-4}$. Now we can construct an associated partition of unity on the unit surface $\boldsymbol{S}^{d-1}$. Let η be a smooth, nonnegative, radial function with $\eta(u) = 1$ for $|u| \leq \frac{1}{2}$ and $\eta(u) = 0$ for $|u| > 1$. Define

$$\widetilde{\Gamma}^k_v(u) = \eta\left(2^{k\gamma}\left(\frac{u}{|u|} - e^k_v\right)\right), \quad \Gamma^k_v(u) = \widetilde{\Gamma}^k_v(u)\left(\sum_{v\in\Theta_k} \widetilde{\Gamma}^k_v(u)\right)^{-1}.$$

Then it is easy to see that Γ_v^k is homogeneous of degree 0 with $\sum_{v\in\Theta_k}\Gamma_v^k(u)=1$ for all $u\neq 0$ and all k. Now we define the operator $T_j^{n,s,v}$ by

$$T_j^{n,s,v}h(x)=\int_{\mathbb{R}^d}\Omega(x-y)\Gamma_v^{n+s}(x-y)\cdot K_j(x-y)\cdot h(y)\,dy. \tag{3-6}$$

Then it is easy to see that $T_j=\sum_{v\in\Theta_{n+s}}T_j^{n,s,v}$.

In the sequel, we will use the Fourier transform since we need to separate the phase in frequency space into different directions. Hence we define a Fourier multiplier operator by

$$\widehat{G_{k,v}h}(\xi)=\Phi\Big(2^{k\gamma}\Big\langle e_v^k,\frac{\xi}{|\xi|}\Big\rangle\Big)\hat{h}(\xi),$$

where $\hat{h}$ is the Fourier transform of h and Φ is a smooth, nonnegative, radial function such that $0\le\Phi(x)\le 1$ and $\Phi(x)=1$ on $|x|\le 2$, $\Phi(x)=0$ on $|x|>4$. Now we can split $T_j^{n,s,v}$ into two parts: $T_j^{n,s,v}=G_{n+s,v}T_j^{n,s,v}+(I-G_{n+s,v})T_j^{n,s,v}$.

The following lemma gives the L^2 estimate involving $G_{n+s,v}T_j^{n,s,v}$, which will be proved in Section 4.

Lemma 3.4. *Let $n\ge 100$ and $s\ge 0$. Suppose $\|\Omega\|_\infty\le 2^{\iota(n+s)}\|\Omega\|_1$ in each T_j. With all the notation above, we get the estimate*

$$\sum_j\Big\|\sum_{v\in\Theta_{n+s}}G_{n+s,v}T_j^{n,s,v}b_{n-j,s}\Big\|_{L_2(\mathcal{A})}^2\lesssim 2^{-(n+s)\gamma+2(n+s)\iota}\lambda\mathcal{C}_\Omega\|f\|_{L_1(\mathcal{A})}.$$

The terms involving $(I-G_{n+s,v})T_j^{n,s,v}$ are more complicated. For convenience, we set $L_j^{n,s,v}=(I-G_{n+s,v})T_j^{n,s,v}$. In Section 4, we shall prove the following lemma.

Lemma 3.5. *Let $n\ge 100$ and $s\ge 0$. Suppose $\|\Omega\|_\infty\le 2^{\iota(n+s)}\|\Omega\|_1$ in each T_j. With all the notation above, then there exists a positive constant α such that*

$$\sum_j\sum_{v\in\Theta_{n+s}}\|L_j^{n,s,v}b_{n-j,s}\|_{L_1(\mathcal{A})}\lesssim 2^{-(n+s)\alpha}\mathcal{C}_\Omega\|f\|_{L_1(\mathcal{A})}.$$

We now complete the proof of (3-4). It is sufficient to prove (3-4) under the condition that for all fixed $n\ge 100$ and $s\ge 0$ we have $\|\Omega\|_\infty\le 2^{\iota(n+s)}\|\Omega\|_1$ in T_j. By Chebyshev's inequality and the triangle inequality, we get

$$\begin{aligned}\tilde{\varphi}\Big(\Big|\zeta\sum_{n\ge 100}\sum_{s\ge 0}\sum_{j\in\mathbb{Z}}T_jb_{n-j,s}\varepsilon_j\zeta\Big|>\frac{\lambda}{8}\Big)&\\ \lesssim\lambda^{-2}\Big\|\zeta\sum_{n\ge 100}\sum_{s\ge 0}\sum_j\sum_{v\in\Theta_{n+s}}G_{n+s,v}T_j^{n,s,v}b_{n-j,s}\varepsilon_j\zeta\Big\|_{L_2(L_\infty\mathfrak{m}\bar{\otimes}\mathcal{A})}^2&\\ +\lambda^{-1}\sum_{n\ge 100}\sum_{s\ge 0}\sum_j\sum_{v\in\Theta_{n+s}}\|\zeta L_j^{n,s,v}b_{n-j,s}\varepsilon_j\zeta\|_{L_1(L_\infty(\mathfrak{m})\bar{\otimes}\mathcal{A})}&\\ =: I+II.&\end{aligned}$$

First we consider the term I. Recall that $\{\varepsilon_j\}_j$ is a Rademacher sequence on a probability space $(\mathfrak{m}, P)$. So we have the orthogonal equality

$$\Big\|\sum_{j\in\mathbb{Z}}\varepsilon_ja_j\Big\|_{L_2(L_\infty(\mathfrak{m})\bar{\otimes}\mathcal{A})}^2=\sum_j\|a_j\|_{L_2(\mathcal{A})}^2. \tag{3-7}$$

Choose $0<\iota<\frac{\gamma}{2}<\frac{1}{2}$. By the triangle inequality, the above orthogonal equality, using Hölder's inequality to remove ζ since ζ is a projection in $\mathcal{A}$, and finally by Lemma 3.4, we get

$$\begin{aligned}
I &\lesssim \lambda^{-2}\bigg(\sum_{n\geq 100}\sum_{s\geq 0}\Big\|\sum_{j}\varepsilon_j\zeta\sum_{v\in\Theta_{n+s}}G_{n+s,v}T_j^{n,s,v}b_{n-j,s}\zeta\Big\|_{L_2(L_\infty(\mathfrak{m})\bar{\otimes}\mathcal{A})}\bigg)^2\\
&\lesssim \lambda^{-2}\bigg(\sum_{n\geq 100}\sum_{s\geq 0}\Big(\sum_{j}\Big\|\zeta\sum_{v\in\Theta_{n+s}}G_{n+s,v}T_j^{n,s,v}b_{n-j,s}\zeta\Big\|^2_{L_2(\mathcal{A})}\Big)^{1/2}\bigg)^2\\
&\lesssim \lambda^{-2}\bigg(\sum_{n\geq 100}\sum_{s\geq 0}(2^{-(n+s)\gamma+2(n+s)\iota}\mathcal{C}_\Omega\lambda\|f\|_{L_1()})^{\frac{1}{2}}\bigg)^2\lesssim \lambda^{-1}\mathcal{C}_\Omega\|f\|_{L_1(\mathcal{A})}.
\end{aligned}$$

For the term II, by the fact that $\{\varepsilon_j\}_{j\in\mathbb{Z}}$ is a bounded sequence, using Hölder's inequality to remove ζ and by Lemma 3.5, we get

$$\begin{aligned}
II &\lesssim \lambda^{-1}\sum_{n\geq 100}\sum_{s\geq 0}\sum_{j}\sum_{v\in\Theta_{n+s}}\|L_j^{n,s,v}b_{n-j,s}\|_{L_1(\mathcal{A})}\\
&\lesssim \lambda^{-1}\sum_{n\geq 100}\sum_{s\geq 0}2^{-(n+s)\alpha}\mathcal{C}_\Omega\|f\|_{L_1(\mathcal{A})}\lesssim \mathcal{C}_\Omega\lambda^{-1}\|f\|_{L_1(\mathcal{A})}.
\end{aligned}$$

Hence we complete the proof of (3-4) based on Lemmas 3.3–3.5. Their proofs will be given in Section 4.

3C. *Estimates for the good functions.* Now we turn to the estimates for good functions. The proofs of diagonal terms and off-diagonal terms will be quite different. We first consider the diagonal terms, which are simpler since they behave similar to those in the classical Calderón–Zygmund decomposition. Following the classical strategy, we should first establish the L_2 boundedness of T. In this situation, the condition for the kernel Ω in fact can be relaxed to $\Omega\in L(\log^+L)^{1/2}(\boldsymbol{S}^{d-1})$.

Lemma 3.6. *Suppose that Ω satisfy (1-3), $\Omega\in L(\log^+L)^{1/2}(\boldsymbol{S}^{d-1})$ and the cancellation property $\int_{\boldsymbol{S}^{d-1}}\Omega(\theta)\,d\sigma(\theta)=0$. Then we have*

$$\|Tf\|_{L_2(L_\infty(\mathfrak{m})\bar{\otimes}\mathcal{A})}\lesssim\|f\|_{L_2(\mathcal{A})},$$

where the implicit constant above depends only on the dimension and Ω.

Remark. It should be pointed out that the cancellation condition $\int_{\boldsymbol{S}^{d-1}}\Omega(\theta)\,d\theta=0$ in Theorem 2.4 is only used in this lemma to guarantee the L_2 boundedness of T.

The proof of Lemma 3.6 will be given in Section 5. Based on this lemma, we could prove required bound for the diagonal term g_d of good functions as follows. By using the property of q_k's in Lemma 3.1, Parcet [2009] obtained the following basic property of g_d:

$$\|g_d\|_{L_1(\mathcal{A})}\lesssim\|f\|_{L_1(\mathcal{A})},\quad \|g_d\|_{L_\infty(\mathcal{A})}\lesssim\lambda\mathcal{C}_\Omega^{-1}. \tag{3-8}$$

By Lemma 3.6, it is not difficult to see that the L_2 norm of T is bounded by $\mathcal{C}_\Omega$ (see the details in Section 5A for its proof). Therefore we get the estimate for g_d as follows:

$$\begin{aligned}
\tilde{\varphi}\Big(|Tg_d|>\frac{\lambda}{4}\Big)&\lesssim\lambda^{-2}\|Tg_d\|^2_{L_2(L_\infty(\mathfrak{m})\bar{\otimes}\mathcal{A})}\lesssim\lambda^{-2}\mathcal{C}_\Omega^2\|g_d\|^2_{L_2(\mathcal{A})}\\
&\lesssim\lambda^{-2}\mathcal{C}_\Omega^2\|g_d\|_{L_1(\mathcal{A})}\|g_d\|_{L_\infty(\mathcal{A})}\lesssim\lambda^{-1}\mathcal{C}_\Omega\|f\|_{L_1(\mathcal{A})},
\end{aligned}$$

where in the first inequality we use Chebyshev's inequality, the second inequality follows from Lemma 3.6, the third and fourth inequalities just follow from (3-8).

In the remaining parts of this subsection, we give effort to the estimate of g_{off}. We first use Lemma 3.2 to reduce the proof to the case $\zeta T g_{\mathrm{off}}\zeta$. In fact

$$Tg_{\mathrm{off}} = (1_{\mathcal{A}}-\zeta)Tg_{\mathrm{off}}(1_{\mathcal{A}}-\zeta)+\zeta Tg_{\mathrm{off}}(1-\zeta)+(1-\zeta)Tg_{\mathrm{off}}\zeta+\zeta Tg_{\mathrm{off}}\zeta.$$

By Lemma 3.2 and the same argument as used for the bad functions, it is sufficient to consider the last term $\zeta T g_{\mathrm{off}}\zeta$ above. Thus our goal is to prove

$$\tilde{\varphi}\Big(|\zeta Tg_{\mathrm{off}}\zeta|>\frac{\lambda}{8}\Big)\lesssim \lambda^{-1}\mathcal{C}_\Omega\|f\|_{L_1(\mathcal{A})}. \tag{3-9}$$

Next we introduce another expression of the off-diagonal terms g_{off} and related estimates which were proved in [Parcet 2009].

Lemma 3.7. *Let df_s be martingale difference. We can rewrite g_{off} as*

$$g_{\mathrm{off}}=\sum_{s\ge 1}\sum_{k\in\mathbb{Z}} p_k df_{k+s}q_{k+s-1}+q_{k+s-1}df_{k+s}p_k =: \sum_{s\ge 1}\sum_{k\in\mathbb{Z}} g_{k,s} =: \sum_{s\ge 1} g_{(s)}.$$

The martingale difference sequence of $g_{(s)}$ satisfies $dg_{(s)_{k+s}}=g_{k,s}$ and $\mathrm{supp}^*\, g_{k,s}\le p_k\le 1_{\mathcal{A}}-q_k$, where supp^* *is weak support projection defined by* $\mathrm{supp}^*\, a=1_{\mathcal{A}}-\boldsymbol{q}$, *with $\boldsymbol{q}$ is the greatest projection satisfying $\boldsymbol{q}a\boldsymbol{q}=0$. Meanwhile, we have the estimates*

$$\sup_{s\ge 1}\|g_{(s)}\|^2_{L_2(\mathcal{A})}=\sup_{s\ge 1}\sum_{k\in\mathbb{Z}}\|g_{k,s}\|^2_{L_2(\mathcal{A})}\lesssim \lambda\mathcal{C}_\Omega^{-1}\|f\|_{L_1(\mathcal{A})}.$$

The strategy to deal with the off-diagonal terms g_{off} is similar to that we use in the proof for the bad functions, although the technical proofs may be different. By the expression of g_{off} in Lemma 3.7 and the formula $f=\sum_{n\in\mathbb{Z}} df_n$, we can write

$$\begin{aligned}\zeta Tg_{\mathrm{off}}\zeta &= \zeta\sum_{s\ge 1}\sum_{j\in\mathbb{Z}}\varepsilon_j T_j g_{(s)}\zeta=\zeta\sum_{s\ge 1}\sum_{j\in\mathbb{Z}}\sum_{n\in\mathbb{Z}}\varepsilon_j T_j d(g_{(s)})_{n-j+s}\zeta\\ &=\zeta\sum_{s\ge 1}\sum_{j\in\mathbb{Z}}\sum_{n\in\mathbb{Z}}\varepsilon_j T_j g_{n-j,s}\zeta=\zeta\sum_{s\ge 1}\sum_{n\ge 100}\sum_{j\in\mathbb{Z}}\varepsilon_j T_j g_{n-j,s}\zeta,\end{aligned}$$

where the last equality follows from the fact if $\zeta(x)\neq 0$ and $n<100$, we get $T_j g_{n-j,s}(x)=0$ for all $s\ge 1$ by property in (ii) of Lemma 3.2, $\mathrm{supp}^*\, g_{k,s}\le p_k\le 1_{\mathcal{A}}-q_k$ in Lemma 3.7 and the similar arguments in (3-3).

By Chebyshev's inequality, the triangle inequality, ζ is a projection in $\mathcal{A}$ and the orthogonal equality (3-7), we then get

$$\begin{aligned}\tilde{\varphi}(|\zeta Tg_{\mathrm{off}}\zeta|>\lambda)&\lesssim \lambda^{-2}\|\zeta Tg_{\mathrm{off}}\zeta\|^2_{L_2(L_\infty(\mathrm{m})\overline{\otimes}\mathcal{A})}\\ &\lesssim \lambda^{-2}\bigg(\sum_{s\ge 1}\sum_{n\ge 100}\Big(\sum_j\|T_j g_{n-j,s}\|^2_{L_2(\mathcal{A})}\Big)^{1/2}\bigg)^2.\end{aligned}$$

Hence to finish the proof for the off-diagonal terms g_{off}, it is sufficient to show that

$$\sum_{s\ge 1}\sum_{n\ge 100}\Big(\sum_j\|T_j g_{n-j,s}\|^2_{L_2(\mathcal{A})}\Big)^{1/2}\lesssim (\mathcal{C}_\Omega\lambda\|f\|_{L_1(\mathcal{A})})^{1/2}. \tag{3-10}$$

As done in the proof for the bad functions, we first show that (3-10) holds if Ω is restricted in $D^\iota = \{\theta \in \boldsymbol{S}^{d-1} : |\Omega(\theta)| \geq 2^{\iota(n+s)}\|\Omega\|_1\}$, where $\iota \in (0,1)$. Recall the definition of $T_{j,\iota}^{n,s}$ in (3-5). Then we have the following lemma.

Lemma 3.8. *Suppose $\Omega \in L(\log^+ L)^2(\boldsymbol{S}^{d-1})$. With all the notation above, we get*

$$\sum_{s\geq 1}\sum_{n\geq 100}\Big(\sum_j \|T_{j,\iota}^{n,s} g_{n-j,s}\|_{L_2(\mathcal{A})}^2\Big)^{1/2} \lesssim (\mathcal{C}_\Omega \lambda \|f\|_{L_1(\mathcal{A})})^{1/2}.$$

The proof of Lemma 3.8 will be given in Section 5. By Lemma 3.8, to prove (3-10), we only need to show (3-10) under the condition that the kernel function Ω satisfies $\|\Omega\|_\infty \leq 2^{\iota(n+s)}\|\Omega\|_1$ in each T_j. For each fixed $s \geq 1$ and $n \geq 100$, we make a microlocal decomposition of T_j as follows:

$$T_j = \sum_{v\in\Theta_{n+s}} T_j^{n,s,v}, \qquad T_j^{n,s,v} = G_{n+s,v} T_j^{n,s,v} + (I - G_{n+s,v}) T_j^{n,s,v}.$$

Here the notation $T_j^{n,s,v}$, $G_{n+s,v}$ is the same as those in the proof of the bad functions.

Lemma 3.9. *Let $n \geq 100$ and $s \geq 1$. Suppose that $\|\Omega\|_\infty \leq 2^{\iota(n+s)}\|\Omega\|_1$ in each T_j. Then we get the estimate*

$$\Big\|\sum_{v\in\Theta_{n+s}} G_{n+s,v} T_j^{n,s,v} g_{n-j,s}\Big\|_{L_2(\mathcal{A})}^2 \lesssim 2^{-(n+s)(\gamma-2\iota)}\|\Omega\|_1^2 \|g_{n-j,s}\|_{L_2(\mathcal{A})}^2.$$

Lemma 3.10. *Let $n \geq 100$ and $s \geq 1$. Suppose $\|\Omega\|_\infty \leq 2^{\iota(n+s)}\|\Omega\|_1$ in each T_j. There exists a constant $\kappa > 0$ such that*

$$\sum_{v\in\Theta_{n+s}} \|(I - G_{n+s,v}) T_j^{n,s,v} g_{n-j,s}\|_{L_2(\mathcal{A})} \lesssim 2^{-(n+s)\kappa}\|\Omega\|_1 \|g_{n-j,s}\|_{L_2(\mathcal{A})}.$$

The proofs of Lemmas 3.9 and 3.10 will be given in Section 5. Now we use Lemmas 3.9 and 3.10 to prove (3-10) as follows:

$$\begin{aligned}\sum_{s\geq 1}\sum_{n\geq 100}\Big(\sum_j \|T_j g_{n-j,s}\|_{L_2(\mathcal{A})}^2\Big)^{1/2} &\leq \sum_{s\geq 1}\sum_{n\geq 100}\Big(\sum_j\Big(\sum_{v\in\Theta_{n+s}} \|(I-G_{n+s,v})T_j^{n,s,v} g_{n-j,s}\|_{L_2(\mathcal{A})}\Big)^2\Big)^{1/2} \\ &\qquad + \sum_{s\geq 1}\sum_{n\geq 100}\Big(\sum_j\Big\|\sum_{v\in\Theta_{n+s}} G_{n+s,v}T_j^{n,s,v} g_{n-j,s}\Big\|_{L_2(\mathcal{A})}^2\Big)^{1/2} \\ &\lesssim (\mathcal{C}_\Omega \lambda \|f\|_{L_1(\mathcal{A})})^{1/2},\end{aligned}$$

where in the second inequality we use Lemmas 3.9 and 3.10, with the fact that $\sum_j \|g_{n-j,s}\|_{L_2(\mathcal{A})}^2 \lesssim \lambda \mathcal{C}_\Omega^{-1}\|f\|_{L_1(\mathcal{A})}$ for all $s \geq 1$ in Lemma 3.7. Thus to finish the proof for good functions, it remains to show Lemmas 3.6, 3.8, 3.9 and 3.10, which are all given in Section 5.

Remark 3.11. At present, it is easy to see that the proofs for off-diagonal terms of good functions are similar to that of bad functions. Notice that for the bad functions, we can deal with the diagonal terms (i.e., $s = 0$) and the off-diagonals terms (i.e., $s > 1$) in a unified way. However this cannot be done for the good functions; thus we prove the diagonal terms g_d and the off-diagonal terms g_{off} using different methods. The main reason comes from the fact that g_{off} has the following property: for all $Q \in \mathcal{Q}_{n-j+s-1}$,

$\int_Q g_{n-j,s}(y)\,dy = 0$. Such kind of cancellation property is crucial in the proof of Lemma 3.10. But the diagonal terms g_d do not have this cancellation property.

4. Proofs of lemmas related to the bad functions

In this section, we begin to prove all the lemmas for the bad functions in Section 3B. Before that we introduce some lemmas needed in our proof. We first state Schur's Lemma which will be used later.

Lemma 4.1 (Schur's lemma). *Suppose that T is an operator with the kernel $K(x, y)$. Thus*

$$Tf(x) = \int_{\mathbb{R}^d} K(x, y) f(y)\,dy.$$

Then T is bounded on $L_2(\mathcal{A})$ with bound $\sqrt{c_1 c_2}$, where

$$c_1 = \sup_{x\in\mathbb{R}^d} \int_{\mathbb{R}^d} |K(x, y)|\,dy, \qquad c_2 = \sup_{y\in\mathbb{R}^d} \int_{\mathbb{R}^d} |K(x, y)|\,dx.$$

The proof of this lemma could be found in [Parcet 2009; Grafakos 2014a]. We also need the following convexity inequality (or the Cauchy–Schwarz-type inequality) for the operator-valued function; see [Mei 2007, page 9]. Let $(\mathfrak{m}, \mu)$ be a measure space. Suppose that $f : \mathfrak{m} \to \mathcal{M}$ is a weak-* integrable function and $g : \mathfrak{m} \to \mathbb{C}$ is an integrable function. Then

$$\left|\int_{\mathfrak{m}} f(x)g(x)\,d\mu(x)\right|^2 \le \int_{\mathfrak{m}} |f(x)|^2\,d\mu(x) \int_{\mathfrak{m}} |g(x)|^2\,d\mu(x). \tag{4-1}$$

Below we introduce some basic properties of the bad functions that we will use in our proof.

Lemma 4.2. *Let $b_{k,s}$ be defined in (3-2). Fix any $s \ge 0$. Then we have the following properties for the bad functions $b_{k,s}$:*

(i) *The L_1 estimate $\sum_{k\in\mathbb{Z}} \|b_{k,s}\|_{L_1(\mathcal{A})} \lesssim \|f\|_{L_1(\mathcal{A})}$ holds.*

(ii) *For all $k \in \mathbb{Z}$ and $Q \in \mathcal{Q}_{k+s}$, the cancellation property $\int_Q b_{k,s}(y)\,dy = 0$ holds.*

The proof of Lemma 4.2 can be found in [Cadilhac 2018; Parcet 2009]. Now we start to prove Lemmas 3.3, 3.4 and 3.5.

4A. *Proof of Lemma 3.3.* Denote the kernel of the operator $T^{n,s}_{j,\iota}$ by

$$K^{n,s}_{j,\iota}(x-y) := \Omega\chi_{D^\iota}\left(\frac{x-y}{|x-y|}\right) K_j(x-y). \tag{4-2}$$

By the support of K_j, it is easy to see that

$$\|K^{n,s}_{j,\iota}\|_{L_1(\mathbb{R}^d)} \lesssim \int_{D^\iota} \int_{2^{j-1}}^{2^{j+1}} |\Omega(\theta)| r^{d-1} 2^{-jd}\,dr\,d\sigma(\theta) \lesssim \int_{D^\iota} |\Omega(\theta)|\,d\sigma(\theta).$$

Therefore by Chebyshev's inequality and the triangle inequality, using Hölder's inequality to remove the projection ζ, we get

$$\tilde{\varphi}\left(\left|\zeta \sum_{n\geq 100}\sum_{s\geq 0}\sum_{j\in\mathbb{Z}} T_{j,\iota}^{n,s} b_{n-j,s}\varepsilon_j\zeta\right| > \lambda\right) \leq \lambda^{-1}\sum_{n\geq 100}\sum_{s\geq 0}\sum_{j\in\mathbb{Z}} \|T_{j,\iota}^{n,s} b_{n-j,s}\varepsilon_j\|_{L_1(L_\infty(\mathfrak{m})\overline{\otimes}\mathcal{A})}.$$

Since $\{\varepsilon_j\}_j$ is the Rademacher sequence, $\{\varepsilon_j\}_j$ is a bounded sequence. Then from above we have

$$\begin{aligned}\lambda^{-1}\sum_{n\geq 100}\sum_{s\geq 0}\sum_{j\in\mathbb{Z}} \|T_{j,\iota}^{n,s} b_{n-j,s}\|_{L_1(\mathcal{A})} &\lesssim \lambda^{-1}\sum_{n\geq 100}\sum_{s\geq 0}\sum_{j\in\mathbb{Z}} \|K_{j,\iota}^{n,s}\|_{L_1(\mathbb{R}^d)}\|b_{n-j,s}\|_{L_1(\mathcal{A})} \\ &\lesssim \lambda^{-1}\sum_{n\geq 100}\sum_{s\geq 0}\int_{D^\iota}|\Omega(\theta)|\,d\sigma(\theta)\sum_j \|b_{n-j,s}\|_{L_1(\mathcal{A})}.\end{aligned}$$

Now applying the property (i) in Lemma 4.2, the above estimate is bounded by

$$\begin{aligned}\lambda^{-1}\|f\|_{L_1(\mathcal{A})}&\int_{S^{d-1}}\#\left\{(n,s): n\geq 100,\ s\geq 0,\ 2^{\iota(n+s)}\leq \frac{|\Omega(\theta)|}{\|\Omega\|_1}\right\}|\Omega(\theta)|\,d\sigma(\theta) \\ &\lesssim \lambda^{-1}\|f\|_{L_1(\mathcal{A})}\int_{S^{d-1}}|\Omega(\theta)|\left(\left(\log^+\frac{|\Omega(\theta)|}{\|\Omega\|_1}\right)^2\right)d\sigma(\theta) \lesssim \lambda^{-1}\mathcal{C}_\Omega\|f\|_{L_1(\mathcal{A})}. \qquad \square\end{aligned}$$

4B. *Proof of Lemma 3.4.* The proof of Lemma 3.4 is based on the following observation of some orthogonality of the support of $\mathcal{F}(G_{k,v})$: For a fixed $k\geq 100$, we have

$$\sup_{\xi\neq 0}\sum_{v\in\Theta_k}\left|\Phi^2\left(2^{k\gamma}\left\langle e_v^k,\frac{\xi}{|\xi|}\right\rangle\right)\right| \lesssim 2^{k\gamma(d-2)}. \tag{4-3}$$

In fact, by the homogeneity of $\Phi^2(2^{k\gamma}\langle e_v^k,\xi/|\xi|\rangle)$, it suffices to take the supremum over the surface S^{d-1}. For $|\xi|=1$ and $\xi\in\operatorname{supp}\Phi^2(2^{k\gamma}\langle e_v^k,\xi/|\xi|\rangle)$, denote by $\xi^\perp$ the hyperplane perpendicular to ξ. Then it is easy to see that

$$\operatorname{dist}(e_v^k,\xi^\perp)\lesssim 2^{-k\gamma}. \tag{4-4}$$

Since the mutual distance of e_v^k's is bounded by $2^{-k\gamma-4}$, there are at most $2^{k\gamma(d-2)}$ vectors satisfying (4-4). We hence get (4-3).

Notice that $L_2(\mathcal{M})$ is a Hilbert space; then the following vector-valued Plancherel's theorem holds:

$$\|\mathcal{F}f\|_{L^2(\mathcal{A})} = (2\pi)^{d/2}\|f\|_{L^2(\mathcal{A})} = (2\pi)^d\|\mathcal{F}^{-1}f\|_{L^2(\mathcal{A})}.$$

By applying this Plancherel's theorem, the convex inequality for the operator-valued function (4-1), the fact (4-3) and finally Plancherel's theorem again, we get

$$\begin{aligned}\Big\|\sum_{v\in\Theta_{n+s}} G_{n+s,v}T_j^{n,s,v}b_{n-j,s}\Big\|_{L_2(\mathcal{A})}^2 &= (2\pi)^{-d/2}\int_{\mathbb{R}^d}\tau\left(\Big|\sum_{v\in\Theta_{n+s}}\Phi\left(2^{(n+s)\gamma}\left\langle e_v^{n+s},\frac{\xi}{|\xi|}\right\rangle\right)\mathcal{F}(T_j^{n,s,v}b_{n-j,s})(\xi)\Big|^2\right)d\xi \\ &\lesssim \int_{\mathbb{R}^d}\sum_{v\in\Theta_{n+s}}\Phi^2\left(2^{(n+s)\gamma}\left\langle e_v^{n+s},\frac{\xi}{|\xi|}\right\rangle\right)\tau\left(\sum_{v\in\Theta_{n+s}}|\mathcal{F}(T_j^{n,s,v}b_{n-j,s})(\xi)|^2\right)d\xi \\ &\lesssim 2^{(n+s)\gamma(d-2)}\sum_{v\in\Theta_{n+s}}\|T_j^{n,s,v}b_{n-j,s}\|_{L_2(\mathcal{A})}^2.\end{aligned} \tag{4-5}$$

Once it is showed that, for a fixed e_v^{n+s},

$$\sum_j \|T_j^{n,s,v} b_{n-j,s}\|_{L_2(\mathcal{A})}^2 \lesssim 2^{-2(n+s)\gamma(d-1)+2(n+s)\iota} \lambda \|\Omega\|_1 \|f\|_{L_1(\mathcal{A})}, \tag{4-6}$$

by $\mathrm{card}(\Theta_{n+s}) \lesssim 2^{(n+s)\gamma(d-1)}$, and applying (4-5) and (4-6) we get

$$\sum_j \Big\| \sum_{v \in \Theta_{n+s}} G_{n+s,v} T_j^{n,s,v} b_{n-j,s} \Big\|_{L_2(\mathcal{A})}^2 \lesssim 2^{-(n+s)\gamma+2(n+s)\iota} \lambda \mathcal{C}_\Omega \|f\|_{L_1(\mathcal{A})},$$

which is the asserted bound of Lemma 3.4. Thus, to finish the proof of Lemma 3.4, it is sufficient to show (4-6).

Recall the definition of $b_{n-j,s}$ in (3-2). By using triangle inequality, to prove (4-6), it is enough to prove the four terms

$$\sum_j \|T_j^{n,s,v} p_{n-j} f p_{n-j+s}\|_{L_2(\mathcal{A})}^2, \quad \sum_j \|T_j^{n,s,v} p_{n-j} f_{n-j+s} p_{n-j+s}\|_{L_2(\mathcal{A})}^2,$$
$$\sum_j \|T_j^{n,s,v} p_{n-j+s} f p_{n-j}\|_{L_2(\mathcal{A})}^2, \quad \sum_j \|T_j^{n,s,v} p_{n-j+s} f_{n-j+s} p_{n-j}\|_{L_2(\mathcal{A})}^2$$

satisfy the desired bound in (4-6). In the following we will only give the detailed proofs of the first and the second terms above, since the proofs of the third and the fourth terms are similar.

We first consider the second term, which involves $p_{n-j} f_{n-j+s} p_{n-j+s}$. Set the kernel of $T_j^{n,s,v}$ as

$$K_j^{n,s,v}(x) = \Gamma_v^{n+s}(x)\Omega(x)\phi_j(x)|x|^{-d}.$$

By Young's inequality, we get

$$\|T_j^{n,s,v} p_{n-j} f_{n-j+s} p_{n-j+s}\|_{L_2(\mathcal{A})}^2 \lesssim \|K_j^{n,s,v}\|_{L_1(\mathbb{R}^d)}^2 \|p_{n-j} f_{n-j+s} p_{n-j+s}\|_{L_2(\mathcal{A})}^2. \tag{4-7}$$

Below we give some estimates for the bound in (4-7). Recall that $|\Omega(\theta)| \le 2^{(n+s)\iota}\|\Omega\|_1$ and the definition of Γ_v^{n+s} in Section 3B. Then by some elementary calculation, we get

$$\|K_j^{n,s,v}\|_{L_1(\mathbb{R}^d)} \lesssim 2^{-(n+s)\gamma(d-1)+(n+s)\iota}\|\Omega\|_1. \tag{4-8}$$

Notice that f is positive in $\mathcal{A}$. By some basic properties of trace φ, we write

$$\begin{aligned}\|p_{n-j} f_{n-j+s} p_{n-j+s}\|_{L_2(\mathcal{A})}^2 &= \varphi(|p_{n-j} f_{n-j+s} p_{n-j+s}|^2)\\ &= \varphi(|p_{n-j+s} f_{n-j+s} p_{n-j}|^2) = \varphi(p_{n-j} f_{n-j+s} p_{n-j+s} f_{n-j+s} p_{n-j})\\ &\le \varphi(p_{n-j} f_{n-j+s}^{1/2} f_{n-j+s}^{1/2} p_{n-j}) \cdot \|f_{n-j+s}^{1/2} p_{n-j+s} f_{n-j+s}^{1/2}\|_{\mathcal{A}}.\end{aligned} \tag{4-9}$$

By the trace invariance and modularity of conditional expectations, the first term in the last line above has the trace-preserving property

$$\varphi(p_{n-j} f_{n-j+s} p_{n-j}) = \varphi(p_{n-j} f p_{n-j}) = \varphi(p_{n-j} f). \tag{4-10}$$

Applying the basic property of C^* algebra, $\|aa^*\|_{\mathcal{A}} = \|a^*a\|_{\mathcal{A}}$, we get

$$\begin{aligned}\|f_{n-j+s}^{1/2} p_{n-j+s} f_{n-j+s}^{1/2}\|_{\mathcal{A}} &= \|p_{n-j+s} f_{n-j+s} p_{n-j+s}\|_{\mathcal{A}} \\ &= \|p_{n-j+s} q_{n-j+s-1} f_{n-j+s} q_{n-j+s-1} p_{n-j+s}\|_{\mathcal{A}} \\ &\le 2^d \|p_{n-j+s} q_{n-j+s-1} f_{n-j+s-1} q_{n-j+s-1} p_{n-j+s}\|_{\mathcal{A}} \\ &\lesssim \lambda \mathcal{C}_\Omega^{-1}, \end{aligned} \tag{4-11}$$

where the second equality follows from the identity $p_k = p_k q_{k-1}$ by the definition of p_k and the last inequality follows from $q_k f_k q_k \le \lambda \mathcal{C}_\Omega^{-1} q_k$, property (ii) in Lemma 3.1. Now combining (4-7)–(4-11), we get

$$\begin{aligned}\sum_j \|T_j^{n,s,v} p_{n-j} f_{n-j+s} p_{n-j+s}\|_{L_2(\mathcal{A})}^2 &\lesssim \mathcal{C}_\Omega^{-1} \lambda 2^{-2(n+s)\gamma(d-1)+2(n+s)\iota} \|\Omega\|_1^2 \sum_j \varphi(p_{n-j} f) \\ &\lesssim \lambda 2^{-2(n+s)\gamma(d-1)+2(n+s)\iota} \|\Omega\|_1 \|f\|_{L_1(\mathcal{A})},\end{aligned}$$

which is the required estimate in (4-6).

Next we give an estimate of the term corresponding to $p_{n-j} f p_{n-j+s}$. Notice that there is no average of f in this case and the crucial property $q_k f_k q_k \le \lambda \mathcal{C}_\Omega^{-1} q_k$ cannot be applied in the estimate (4-11). Our strategy here is to *add* an average of f. In the following we first reduce the proof to the case that the kernel is positive. To do that, we first take the decomposition

$$K_j^{n,s,v} = (K_j^{n,s,v})^+ - (K_j^{n,s,v})^-,$$

where $(K_j^{n,s,v})^+$ and $(K_j^{n,s,v})^-$ are positive functions. Then by using triangle inequality, we get

$$\begin{aligned}\sum_j \|T_j^{n,s,v} p_{n-j} f p_{n-j+s}\|_{L_2(\mathcal{A})}^2 &\lesssim \sum_j \Big\| \int (K_j^{n,s,v}(\cdot - y))^+ p_{n-j} f p_{n-j+s}(y)\, dy \Big\|_{L_2(\mathcal{A})}^2 \\ &\quad + \sum_j \Big\| \int (K_j^{n,s,v}(\cdot - y))^- p_{n-j} f p_{n-j+s}(y)\, dy \Big\|_{L_2(\mathcal{A})}^2 .\end{aligned}$$

Therefore we need to consider the terms related to $(K_j^{n,s,v})^+$ and $(K_j^{n,s,v})^-$, respectively. We only consider the term related to $(K_j^{n,s,v})^+$ since the proof of the other term is similar. For convenience, in the remaining part of this section we still use the abused notation $K_j^{n,s,v}$ to represent $(K_j^{n,s,v})^+$.

Denote the support of $K_j^{n,s,v}$ by $E_j^{n,s,v}$. Then it is not difficult to see

$$\begin{aligned}E_j^{n,s,v} &\subset \Big\{ x \in \mathbb{R}^d : \Big| \frac{x}{|x|} - e_v^{n+s} \Big| \le 2^{-(n+s)\gamma},\ 2^{j-1} \le |x| \le 2^{j+1} \Big\} \\ &\subset \{ x \in \mathbb{R}^d : |\langle x, e_v^{n+s}\rangle| \le 2^{j+1},\ |x - \langle x, e_v^{n+s}\rangle e_v^{n+s}| \le 2^{j+1-(n+s)\gamma} \}.\end{aligned}$$

For any $Q \in \mathcal{Q}_{n-j+s}$, let $Q_{n-j} \in \mathcal{Q}_{n-j}$ be the s-th ancestor of Q. By the definition of p_k, we may write

$$\begin{aligned}T_j^{n,s,v}(p_{n-j} f p_{n-j+s})(x) &= \int_{\mathbb{R}^d} K_j^{n,s,v}(x-y)(p_{n-j} f p_{n-j+s})(y)\, dy \\ &= \sum_{\substack{Q \in \mathcal{Q}_{n-j+s} \\ Q \cap \{x - E_j^{n,s,v}\} \ne \varnothing}} \pi_{Q_{n-j}} \Big(\int_Q K_j^{n,s,v}(x-y) f(y)\, dy \Big) \pi_Q\end{aligned}$$

$$= \sum_{\substack{Q \in \mathcal{Q}_{n-j+s} \\ Q \cap \{x - E_j^{n,s,v}\} \neq \varnothing}} \int_Q [p_{n-j}(K_j^{n,s,v}(x - \cdot) f(\cdot))_{n-j+s} p_{n-j+s}](z)\, dz$$
$$= \int_{E_j^{n,s,v}(x)} [p_{n-j}(K_j^{n,s,v}(x - \cdot) f(\cdot))_{n-j+s} p_{n-j+s}](z)\, dz,$$

where we use the notation

$$E_j^{n,s,v}(x) = \bigcup_{\substack{Q \in \mathcal{Q}_{n-j+s} \\ Q \cap \{x - E_j^{n,s,v}\} \neq \varnothing}} Q.$$

By the support of $E_j^{n,s,v}$, we see that $E_j^{n,s,v}$ is contained in a rectangle with one sidelength at most 2^{j+1} and $d-1$ sidelength at most $2^{j+1-(n+s)\gamma}$. Since for any $Q \in \mathcal{Q}_{n-j+s}$, the sidelength satisfies $l(Q) = 2^{j-(n+s)} \leq 2^{j+1-(n+s)\gamma}$. So we get $E_j^{n,s,v}(x)$ is contained in a rectangle with one sidelength at most 2^{j+2} and $d-1$ sidelength at most $2^{j+2-(n+s)\gamma}$. Therefore we have the estimate

$$|E_j^{n,s,v}(x)| \lesssim 2^{jd-(n+s)\gamma(d-1)}.$$

Next by using the convexity inequality for the operator-valued function (4-1) and the preceding inequality, we get

$$|T_j^{n,s,v}(p_{n-j} f p_{n-j+s})(x)|^2 \lesssim 2^{jd-(n+s)\gamma(d-1)} \int_{E_j^{n,s,v}(x)} |p_{n-j}(K_j^{n,s,v}(x - \cdot) f(\cdot))_{n-j+s} p_{n-j+s}(z)|^2\, dz.$$

Combining the above estimates, we get

$$\|T_j^{n,s,v} p_{n-j} f p_{n-j+s}\|^2_{L_2(\mathcal{A})} \lesssim 2^{jd-(n+s)\gamma(d-1)} \int_{\mathbb{R}^d} \int_{E_j^{n,s,v}(x)} \tau\big(|p_{n-j}(K_j^{n,s,v}(x - \cdot) f(\cdot))_{n-j+s} p_{n-j+s}(z)|^2\big)\, dz\, dx. \tag{4-12}$$

Since $K_j^{n,s,v}$ is a positive function and f is a positive operator-valued function in $\mathcal{A}$, we see that $K(x - \cdot) f(\cdot)$ is positive in $\mathcal{A}$. Therefore

$$(K_j^{n,s,v}(x - \cdot) f(\cdot))_{n-j+s} = \sum_{Q \in \mathcal{Q}_{n-j+s}} \frac{1}{|Q|} \int_Q K_j^{n,s,v}(x - y) f(y)\, dy \chi_Q$$
$$\lesssim \sum_{Q \in \mathcal{Q}_{n-j+s}} \frac{1}{|Q|} \int_Q f(y)\, dy \chi_Q 2^{-jd+(n+s)\iota} \|\Omega\|_1$$
$$= 2^{-jd+(n+s)\iota} \|\Omega\|_1 f_{n-j+s}.$$

Now applying the above estimate and using the same idea in the estimates of (4-9) and (4-11), we get

$$\begin{aligned}
&\tau\big(|p_{n-j}(K_j^{n,s,v}(x - \cdot) f(\cdot))_{n-j+s} p_{n-j+s}(z)|^2\big) \\
&\quad = \tau\big(p_{n-j}(K_j^{n,s,v}(x - \cdot) f(\cdot))_{n-j+s} p_{n-j+s} (K_j^{n,s,v}(x - \cdot) f(\cdot))_{n-j+s} p_{n-j}(z)\big) \\
&\quad \leq \tau\big(p_{n-j}(K_j^{n,s,v}(x - \cdot) f(\cdot))_{n-j+s} p_{n-j}(z)\big) \|p_{n-j+s}(K_j^{n,s,v}(x - \cdot) f(\cdot))_{n-j+s} p_{n-j+s}(z)\|_{\mathcal{M}} \\
&\quad \lesssim 2^{-2jd+2(n+s)\iota} \|\Omega\|_1^2 \tau(p_{n-j} f_{n-j+s} p_{n-j}(z)) \|p_{n-j+s} f_{n-j+s} p_{n-j+s}\|_{\mathcal{A}} \\
&\quad \lesssim 2^{-2jd+2(n+s)\iota} \|\Omega\|_1 \lambda \tau(p_{n-j} f_{n-j+s} p_{n-j}(z)).
\end{aligned} \tag{4-13}$$

By the definition of $E_j^{n,s,v}(x)$, for any fixed $z \in \mathbb{R}^d$, we have the estimate

$$\left|\int_{\{x:\, E_j^{n,s,v}(x)\ni z\}} dx\right| \lesssim 2^{jd-(n+s)\gamma(d-1)}. \tag{4-14}$$

Plugging (4-13) into (4-12), then applying Fubini's theorem with (4-14), and finally using the trace-preserving property (4-10), we get

$$\begin{aligned}\sum_j \|T_j^{n,s,v} p_{n-j} f p_{n-j+s}\|_{L_2(\mathcal{A})}^2 &\lesssim 2^{-2(n+s)\gamma(d-1)+2(n+s)\iota}\|\Omega\|_1\lambda \sum_{j\in\mathbb{Z}} \varphi(p_{n-j} f_{n-j+s} p_{n-j}) \\ &\lesssim 2^{-2(n+s)\gamma(d-1)+2(n+s)\iota}\|\Omega\|_1\lambda\|f\|_{L_1(\mathcal{A})}.\end{aligned}$$

Hence, we complete the proof of Lemma 3.4. □

4C. *Proof of Lemma 3.5.* To prove Lemma 3.5, we have to face some oscillatory integrals which come from $L_j^{n,s,v}$. Before stating the proof of Lemma 3.5, let us first give some notation. We introduce the Littlewood–Paley decomposition. Let ψ be a radial C^∞ function such that $\psi(\xi)=1$ for $|\xi|\le 1$, $\psi(\xi)=0$ for $|\xi|\ge 2$ and $0\le\psi(\xi)\le 1$ for all $\xi\in\mathbb{R}^d$. Define $\beta_k(\xi)=\psi(2^k\xi)-\psi(2^{k+1}\xi)$. Then β_k is supported in $\{\xi : 2^{-k-1}\le|\xi|\le 2^{-k+1}\}$. Choose $\tilde\beta$ be a radial C^∞ function such that $\tilde\beta(\xi)=1$ for $\frac{1}{2}\le|\xi|\le 2$, $\tilde\beta$ is supported in $\left\{\xi : \frac{1}{4}\le|\xi|\le 4\right\}$ and $0\le\tilde\beta(\xi)\le 1$ for all $\xi\in\mathbb{R}^d$. Set $\tilde\beta_k(\xi)=\tilde\beta(2^k\xi)$. Then it is easy to see $\beta_k=\tilde\beta_k\beta_k$. Define the convolution operators Λ_k and $\tilde\Lambda_k$ with the Fourier multipliers β_k and $\tilde\beta_k$, respectively. That is,

$$\widehat{\Lambda_k f}(\xi)=\beta_k(\xi)\hat f(\xi), \qquad \widehat{\tilde\Lambda_k f}(\xi)=\tilde\beta_k(\xi)\hat f(\xi).$$

Then by the construction of β_k and $\tilde\beta_k$, we have $\Lambda_k=\tilde\Lambda_k\Lambda_k$, $I=\sum_{k\in\mathbb{Z}}\Lambda_k$.

Write

$$L_j^{n,s,v}=\sum_k (I-G_{n+s,v})\Lambda_k T_j^{n,s,v}.$$

Then triangle inequality gives us

$$\|L_j^{n,s,v} b_{n-j,s}\|_{L_1(\mathcal{A})}\le\sum_{k\in\mathbb{Z}}\|(I-G_{n+s,v})\Lambda_k T_j^{n,s,v} b_{n-j,s}\|_{L_1(\mathcal{A})}.$$

In the remaining part of this subsection, we show that two different estimates can be established for $\|(I-G_{n+s,v})\Lambda_k T_j^{n,s,v} b_{n-j,s}\|_{L_1(\mathcal{A})}$, which will give Lemma 3.5 by taking a sum over $k\in\mathbb{Z}$ with these two different estimates.

Lemma 4.3. *With all the notation above. Then there exists $N>0$ such that the following estimate holds*:

$$\|(I-G_{n+s,v})\Lambda_k T_j^{n,s,v} b_{n-j,s}\|_{L_1(\mathcal{A})}\lesssim 2^{-(n+s)\gamma(d-1)+(n+s)\iota+(k-j)+(n+s)\gamma(1+2N)}\|\Omega\|_1\|b_{n-j,s}\|_{L_1(\mathcal{A})}. \tag{4-15}$$

Proof. Applying Fubini's theorem, we may write

$$(I-G_{n+s,v})\Lambda_k T_j^{n,s,v} b_{n-j,s}(x) =: \int_{\mathbb{R}^d} D_k^{n,s,v}(x-y) b_{n-j,s}(y)\,dy, \tag{4-16}$$

where $D_k^{n,s,v}(x)$ is defined as the kernel of the operator $(I-G_{n+s,v})\Lambda_k T_j^{n,s,v}$. More precisely, $D_k^{n,s,v}$ can be written as

$$D_k^{n,s,v}(x)=\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}e^{ix\cdot\xi}h_{k,n,s,v}(\xi)\int_{\mathbb{R}^d}e^{-i\xi\cdot\omega}\Omega(\omega)\Gamma_v^{n+s}(\omega)K_j(\omega)\,d\omega\,d\xi, \tag{4-17}$$

where $h_{k,n,s,v}(\xi)=(1-\Phi(2^{(n+s)\gamma}\langle e_v^{n+s},\xi/|\xi|\rangle))\beta_k(\xi)$. Using Young's inequality, we get

$$\|(I-G_{n+s,v})\Lambda_k T_j^{n,s,v}b_{n-j,s}\|_{L_1(\mathcal{A})}\le\|D_k^{n,s,v}\|_{L_1(\mathbb{R}^d)}\|b_{n-j,s}\|_{L_1(\mathcal{A})}.$$

Hence in the following we only need to give an L_1 estimate of $D_k^{n,s,v}$. In order to separate the rough kernel, we make a change of variable $\omega=r\theta$. By Fubini's theorem, $D_k^{n,s,v}(x)$ can be written as

$$\frac{1}{(2\pi)^d}\int_{S^{d-1}}\Omega(\theta)\Gamma_v^{n+s}(\theta)\left\{\int_{\mathbb{R}^d}\int_0^\infty e^{i\langle x-r\theta,\xi\rangle}h_{k,n,s,v}(\xi)K_j(r)r^{d-1}\,dr\,d\xi\right\}d\sigma(\theta). \tag{4-18}$$

Concerning the support of K_j, we have $2^{j-1}\le r\le 2^{j+1}$. Integrating by parts with r, the integral involving r can be rewritten as

$$\int_0^\infty e^{-i\langle r\theta,\xi\rangle}(i\langle\theta,\xi\rangle)^{-1}\partial_r[K_j(r)r^{d-1}]\,dr.$$

Since $\theta\in\operatorname{supp}\Gamma_v^{n+s}$, we have $|\theta-e_v^{n+s}|\le 2^{-(n+s)\gamma}$. By the support of Φ, we see $|\langle e_v^{n+s},\xi/|\xi|\rangle|\ge 2^{1-(n+s)\gamma}$. Thus,

$$\left|\left\langle\theta,\frac{\xi}{|\xi|}\right\rangle\right|\ge\left|\left\langle e_v^{n+s},\frac{\xi}{|\xi|}\right\rangle\right|-\left|\left\langle e_v^{n+s}-\theta,\frac{\xi}{|\xi|}\right\rangle\right|\ge 2^{-(n+s)\gamma}. \tag{4-19}$$

After integrating by parts with r, integrating by parts N times with ξ, the integral in (4-18) can be rewritten as

$$\begin{aligned}\frac{1}{(2\pi)^d}\int_{S^{d-1}}\Omega(\theta)\Gamma_v^{n+s}(\theta)\int_{\mathbb{R}^d}\int_0^\infty e^{i\langle x-r\theta,\xi\rangle}\partial_r[K_j(r)r^{d-1}]&\\ \times\frac{(I-2^{-2k}\Delta_\xi)^N}{(1+2^{-2k}|x-r\theta|^2)^N}(h_{k,n,s,v}(\xi)(i\langle\theta,\xi\rangle)^{-1})\,dr\,d\xi\,d\sigma(\theta).&\end{aligned} \tag{4-20}$$

In the following, we give explicit estimates of all terms in (4-20). We show that the following estimate holds:

$$|(I-2^{-2k}\Delta_\xi)^N[\langle\theta,\xi\rangle^{-1}h_{k,n,s,v}(\xi)]|\lesssim 2^{(n+s)\gamma+k+2(n+s)\gamma N}. \tag{4-21}$$

Firstly we prove (4-21) when $N=0$. By (4-19), we have

$$|(-i\langle\theta,\xi\rangle)^{-1}\cdot h_{k,n,s,v}(\xi)|\lesssim|\langle\theta,\xi\rangle|^{-1}\lesssim 2^{(n+s)\gamma+k}.$$

Next we consider $N=1$ in (4-21). By using product rule and some elementary calculation, we get

$$\begin{aligned}|\partial_{\xi_i}h_{k,n,s,v}(\xi)|&\le\left|-\partial_{\xi_i}\left[\Phi\left(2^{(n+s)\gamma}\left\langle e_v^{n+s},\frac{\xi}{|\xi|}\right\rangle\right)\right]\cdot\beta_k(\xi)\right|+\left|\partial_{\xi_i}\beta_k(\xi)\cdot\left(1-\Phi\left(2^{(n+s)\gamma}\left\langle e_v^{n+s},\frac{\xi}{|\xi|}\right\rangle\right)\right)\right|\\&\lesssim 2^{(n+s)\gamma+k}.\end{aligned}$$

Therefore by induction, we have $|\partial_\xi^\alpha h_{k,n,s,v}(\xi)| \lesssim 2^{((n+s)\gamma+k)|\alpha|}$ for any multi-indices $\alpha \in \mathbb{Z}_+^n$. By using product rule again and (4-19), we have

$$|\partial_{\xi_i}^2((\langle\theta,\xi\rangle)^{-1}h_{k,n,s,v}(\xi))| \le |2\langle\theta,\xi\rangle^{-3}\cdot\theta_i^2\cdot h_{k,n,s,v}(\xi)|+|2\langle\theta,\xi\rangle^{-2}\cdot\theta_i\partial_{\xi_i}h_{k,n,s,v}(\xi)|+|\langle\theta,\xi\rangle^{-1}\partial_{\xi_i}^2 h_{k,n,s,v}(\xi)| \lesssim 2^{3((n+s)\gamma+k)}.$$

Hence we conclude that $2^{-2k}|\Delta_\xi[(\langle\theta,\xi\rangle)^{-1}h_{k,n,s,v}(\xi)]| \lesssim 2^{(n+s)\gamma+k+2(n+s)\gamma}$. Proceeding by induction, we get (4-21).

By the definition of K_j and using product rule, it is not difficult to get

$$|\partial_r(K_j(r)r^{d-1})| \lesssim 2^{-2j}. \tag{4-22}$$

Now we choose $N=[d/2]+1$. Since we need to get the L^1 estimate of (4-20), by the support of $h_{k,n,s,v}$, $|\xi|\approx 2^{-k}$,

$$\int_{|\xi|\approx 2^{-k}}\int_{\mathbb{R}^d}(1+2^{-2k}|x-r\theta|^2)^{-N}\,dx\,d\xi \le C.$$

Now combine (4-22), (4-21) and above estimates. Next integrating with r, we get a bound 2^j. Note that we suppose that $\|\Omega\|_\infty \le 2^{(n+s)\iota}\|\Omega\|_1$. Then integrating with θ, we get a bound $2^{-(n+s)\gamma(d-1)+(n+s)\iota}\|\Omega\|_1$. So we finally get

$$\begin{aligned}\|D_k^{n,s,v}\|_{L_1(\mathbb{R}^d)} &\lesssim 2^{-2j+(n+s)\gamma+k+2(n+s)\gamma N+j-(n+s)\gamma(d-1)+(n+s)\iota}\|\Omega\|_1\\ &= 2^{-(n+s)\gamma(d-1)+(n+s)\iota-j+k+(n+s)\gamma(1+2N)}\|\Omega\|_1.\end{aligned} \tag{4-23}$$

Hence we complete the proof of Lemma 4.3 with $N=[d/2]+1$. □

Lemma 4.4. *With all the notation above, the following estimate holds*:

$$\|(I-G_{n+s,v})\Lambda_k T_j^{n,s,v}b_{n-j,s}\|_{L_1(\mathcal{A})} \lesssim 2^{-(n+s)\gamma(d-1)-(n+s)+j-k+(n+s)\iota}\|\Omega\|_1\|b_{n-j,s}\|_{L_1(\mathcal{A})}.$$

Proof. Using $\Lambda_k=\Lambda_k\tilde{\Lambda}_k$, we write

$$\begin{aligned}\|(I-G_{n+s,v})\Lambda_k T_j^{n,s,v}b_{n-j,s}\|_{L_1(\mathcal{A})} &= \|(I-G_{n+s,v})\tilde{\Lambda}_k\Lambda_k T_j^{n,s,v}b_{n-j,s}\|_{L_1(\mathcal{A})}\\ &\lesssim \|(I-G_{n+s,v})\tilde{\Lambda}_k\|_{L_1(\mathcal{A})\to L_1(\mathcal{A})}\|\Lambda_k T_j^{n,s,v}b_{n-j,s}\|_{L_1(\mathcal{A})}.\end{aligned}$$

Then it is easy to see that the proof of this lemma follows from the two estimates

$$\|(I-G_{n+s,v})\tilde{\Lambda}_k\|_{L_1(\mathcal{A})\to L_1(\mathcal{A})} \lesssim 1 \tag{4-24}$$

and

$$\|\Lambda_k T_j^{n,s,v}b_{n-j,s}\|_{L_1(\mathcal{A})} \lesssim 2^{-(n+s)\gamma(d-1)-(n+s)+j-k+(n+s)\iota}\|\Omega\|_1\|b_{n-j,s}\|_{L_1(\mathcal{A})}. \tag{4-25}$$

We first consider the estimate (4-24). The kernel of $(I-G_{n+s,v})\tilde{\Lambda}_k$ is the inverse Fourier transform of $\tilde{h}_{k,n,s,v}(\xi)=[1-\Phi(2^{(n+s)\gamma}\langle e_v^{n+s},\xi/|\xi|\rangle)]\tilde{\beta}_k(\xi)$. So

$$\|(I-G_{n+s,v})\tilde{\Lambda}_k\|_{L_1(\mathcal{A})\to L_1(\mathcal{A})} \lesssim \|\mathcal{F}(\tilde{h}_{k,n,s,v})\|_{L^1(\mathbb{R}^d)} = \|\mathcal{F}[\tilde{h}_{k,n,s,v}(A_k^{n,s,v}\,\cdot)]\|_{L_1(\mathbb{R}^d)},$$

where $A_k^{n,s,v}$ is an invertible linear transform such that $A_k^{n,s,v} e_v^{n+s} = 2^{-(n+s)\gamma-k} e_v^{n+s}$ and $A_k^{n,s,v} y = 2^{-k} y$ if $\langle y, e_v^{n+s}\rangle = 0$. For all $\alpha \in \mathbb{Z}_+^d$, it is straightforward to check that

$$\|\partial^\alpha[\tilde{h}_{k,n,s,v}(A_k^{n,s,v}\,\cdot)]\|_{L_2(\mathbb{R}^d)} \lesssim C_\alpha$$

uniformly with k, n, s, v; see [Seeger 1996, page 100]. Therefore splitting the following integral into two parts and using Plancherel's theorem, we get

$$\begin{aligned}
&\|\mathcal{F}[\tilde{h}_{k,n,s,v}(A_k^{n,s,v}\,\cdot)]\|_{L^1(\mathbb{R}^d)} \\
&= \biggl(\int_{|\xi|\geq 1} + \int_{|\xi|<1}\biggr) |\mathcal{F}[\tilde{h}_{k,n,s,v}(A_k^{n,s,v}\,\cdot)](\xi)|\,d\xi \\
&\lesssim \biggl(\int_{|\xi|\geq 1} \frac{d\xi}{|\xi|^{2([d/2]+1)}}\biggr)^{1/2} \sum_{|\alpha|=[d/2]+1} \biggl(\int_{\mathbb{R}^d} |\xi^\alpha \mathcal{F}[\tilde{h}_{k,n,s,v}(A_k^{n,s,v}\,\cdot)](\xi)|^2\,d\xi\biggr)^{1/2} + \|\mathcal{F}[\tilde{h}_{k,n,s,v}(A_k^{n,s,v})]\|_{L_2(\mathbb{R}^d)} \\
&\lesssim \sum_{|\alpha|=[d/2]+1} \|\partial^\alpha[\tilde{h}_{k,n,s,v}(A_k^{n,s,v}\,\cdot)]\|_{L_2(\mathbb{R}^d)} + \|\tilde{h}_{k,n,s,v}(A_k^{n,s,v}\,\cdot)\|_{L_2(\mathbb{R}^d)} \lesssim 1,
\end{aligned}$$

which completes the proof of (4-24).

Now we turn to another estimate (4-25). Write

$$\Lambda_k T_j^{n,s,v} b_{n-j,s} = \check{\beta}_k * K_j^{n,s,v} * b_{n-j,s} = K_j^{n,s,v} * \check{\beta}_k * b_{n-j,s}.$$

Then by the estimate (4-8) of $K_j^{n,s,v}$, we get

$$\begin{aligned}
\|\Lambda_k T_j^{n,s,v} b_{n-j,s}\|_{L_1(\mathcal{A})} &\leq \|K_j^{n,s,v}\|_{L_1(\mathbb{R}^d)} \|\check{\beta}_k * b_{n-j,s}\|_{L_1(\mathcal{A})} \\
&\lesssim 2^{-(n+s)(\gamma(d-1)-\iota)} \|\Omega\|_1 \|\check{\beta}_k * b_{n-j,s}\|_{L_1(\mathcal{A})}.
\end{aligned} \tag{4-26}$$

Note that $\beta_k(\xi) = \beta(2^k\xi)$; we get $\check{\beta}_k(x) = 2^{-kd}\check{\beta}(2^{-k}x)$. Therefore we see

$$\int_{\mathbb{R}^d} |\nabla[\check{\beta}_k](x)|\,dx = 2^{-k(d+1)} \int_{\mathbb{R}^d} |\nabla(\check{\beta})(2^{-k}x)|\,dx = 2^{-k} \int_{\mathbb{R}^d} |\nabla(\check{\beta})(x)|\,dx. \tag{4-27}$$

Using the cancellation property (ii) in Lemma 4.2, we see that, for all $Q \in \mathcal{Q}_{n-j+s}$, $\int_Q b_{n-j,s}(y)\,dy = 0$. Let y_Q be the center of Q. Notice that, for all $y \in Q$, $|y - y_Q| \lesssim 2^{j-n-s}$. Using this cancellation property, we then get

$$\begin{aligned}
\|\check{\beta}_k * b_{n-j,s}\|_{L_1(\mathcal{A})} &= \int_{\mathbb{R}^d} \tau\biggl(\biggl|\sum_{Q\in\mathcal{Q}_{n-j+s}} \int_Q [\check{\beta}_k(x-y) - \check{\beta}_k(x-y_Q)] b_{n-j,s}(y)\,dy\biggr|\biggr)\,dx \\
&\leq \int_{\mathbb{R}^d} \sum_{Q\in\mathcal{Q}_{n-j+s}} \int_Q \biggl|\int_0^1 \langle y-y_Q, \nabla[\check{\beta}_k](x-\rho y-(1-\rho)y_Q)\rangle\,d\rho\biggr| \tau(|b_{n-j,s}(y)|)\,dy\,dx \\
&\lesssim 2^{j-n-s-k} \|b_{n-j,s}\|_{L_1(\mathcal{A})},
\end{aligned}$$

where in the second inequality we just use the mean value formula. Combining this inequality with (4-26) yields the estimate (4-25). Hence we finish the proof of this lemma. □

Now we conclude the proof of Lemma 3.5 as follows. Let $\varepsilon_0 \in (0, 1)$ be a constant which will be chosen later. Notice that $\operatorname{card}(\Theta_{n+s}) \lesssim 2^{(n+s)\gamma(d-1)}$. Then by Lemma 4.3 with $N = [d/2]+1$, Lemma 4.4 and the property $\sum_j \|b_{n-j,s}\|_{L_1(\mathcal{A})} \lesssim \|f\|_{L_1(\mathcal{A})}$ in Lemma 4.2, we get

$$\begin{aligned}\sum_j \sum_{v\in\Theta_{n+s}} \|L_j^{n,s,v} b_{n-j,s}\|_{L_1(\mathcal{A})} &\le \sum_j \sum_{v\in\Theta_{n+s}} \Big(\sum_{k\le j-[(n+s)\varepsilon_0]} + \sum_{k\ge j-[(n+s)\varepsilon_0]} \Big) \|(I-G_{n+s,v})\Lambda_k T_j^{n,s,v} b_{n-j,s}\|_{L_1(\mathcal{A})} \\ &\lesssim \sum_j (2^{-(n+s)(\varepsilon_0-\gamma(3+2[d/2])-\iota)} + 2^{-(n+s)(1-\varepsilon_0-\iota)}) \|b_{n-j,s}\|_{L_1(\mathcal{A})} \lesssim 2^{-(n+s)\alpha} \|f\|_{L_1(\mathcal{A})},\end{aligned}$$

where we choose the constants $0<\iota \ll \gamma \ll \varepsilon_0 \ll 1$ such that the constant α is defined by

$$\alpha = \min\Big\{\varepsilon_0 - \gamma\Big(3+2\Big[\frac{d}{2}\Big]\Big) - \iota,\ 1-\varepsilon_0-\iota\Big\} > 0. \qquad \square$$

5. Proofs of lemmas related to the good functions

In this section, we begin to prove all lemmas for the good functions in Section 3C. The proofs for off-diagonal terms are similar to those for bad functions in Section 4, so we shall be brief and only indicate necessary changes in the proofs of off-diagonal terms. We first consider the proofs of diagonal terms.

5A. ***Proof of Lemma 3.6.*** Recall the definition of T. Let $\boldsymbol{K}_j$ be the kernel of the operator T_j, i.e., $\boldsymbol{K}_j(x) = \Omega(x)\phi_j(x)|x|^{-d}$. Notice that $\{\varepsilon_j\}_j$ is a Rademacher sequence on a probability space $(\mathfrak{m}, P)$; then applying the equality (3-7), we can write

$$\|Tf\|^2_{L_2(L_\infty(\mathfrak{m}\bar{\otimes}\mathcal{A}))} = \sum_{j\in\mathbb{Z}} \|T_j f\|^2_{L_2(\mathcal{A})} = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \sum_{j\in\mathbb{Z}} |\widehat{\boldsymbol{K}}_j(\xi)|^2 \tau(|\hat{f}(\xi)|^2)\, d\xi,$$

where the second equality follows from Plancherel's theorem since $L_2(\mathcal{M})$ is a Hilbert space. In the following we show that

$$\sum_{j\in\mathbb{Z}} |\widehat{\boldsymbol{K}}_j(\xi)|^2 < \infty \tag{5-1}$$

holds for almost every $\xi \in \mathbb{R}^d$. Once we prove the inequality (5-1), Lemma 3.6 follows from Plancherel's theorem. Now we fix $\xi \ne 0$. By the cancellation property of Ω, $\int_{\boldsymbol{S}^{d-1}} \Omega(\theta)\, d\sigma(\theta) = 0$, we get

$$|\widehat{\boldsymbol{K}}_j(\xi)| = \Big| \int_{\mathbb{R}^d} K_j(x)(e^{-i\xi x}-1)\, dx \Big| \lesssim 2^j |\xi| \|\Omega\|_1.$$

Therefore the sum over all j's satisfying $2^j|\xi| \le 1$ is convergent.

Now we turn to the case $2^j|\xi| > 1$. We split the kernel $\Omega(\theta)$ into two parts:

$$\Omega_1(\theta) = \Omega(\theta)\chi_{\{\theta\in\boldsymbol{S}^{d-1}:|\Omega(\theta)|\le 2^{j\nu}|\xi|^\nu\|\Omega\|_1\}} \quad \text{and} \quad 1-\Omega_1(\theta)$$

for some constant $\nu \in \big(0, \frac{1}{2}\big)$. We first consider Ω_1. By making a change of variable $x = r\theta$, we get

$$|\widehat{\boldsymbol{K}}_j(\xi)| \le \int_{\boldsymbol{S}^{d-1}} |\Omega_1(\theta)| \Big| \int_{\mathbb{R}} e^{-ir\langle\theta,\xi\rangle} \phi_j(r) r^{-1}\, dr \Big|\, d\sigma(\theta). \tag{5-2}$$

It is easy to see that $\left|\int_{\mathbb{R}} e^{-ir\langle\theta,\xi\rangle}\phi_j(r)r^{-1}\,dr\right|$ is finite. By integrating by parts with the variable r, we get

$$\left|\int_{\mathbb{R}} e^{-ir\langle\theta,\xi\rangle}\phi_j(r)r^{-1}\,dr\right| = \left|\int_{\mathbb{R}} e^{-ir\langle\theta,\xi\rangle}\langle\theta,\xi\rangle^{-1}\partial_r[\phi_j(r)r^{-1}]\,dr\right| \lesssim (2^j|\xi|)^{-1}|\langle\theta,\xi'\rangle|^{-1},$$

where $\xi' = \xi/|\xi|$. Interpolating these two estimates we get, that, for any $\delta \in \left(\frac{1}{2}, 1\right)$,

$$\left|\int_{\mathbb{R}} e^{-ir\langle\theta,\xi\rangle}\phi_j(r)r^{-1}\,dr\right| \lesssim (2^j|\xi|)^{-\delta}|\langle\theta,\xi'\rangle|^{-\delta}.$$

Plugging the above estimate into (5-2) with the fact $\int_{\boldsymbol{S}^{d-1}} |\langle\theta,\xi'\rangle|^{-\delta}\,d\sigma(\theta) < \infty$, we hence get

$$|\widehat{\boldsymbol{K}}_j(\xi)| \lesssim (2^j|\xi|)^{-\delta+\nu}\|\Omega\|_1,$$

which is sufficient for us taking a sum over all j's satisfying $2^j|\xi| > 1$. Consider the other term $1-\Omega_1$. Then we get

$$\begin{aligned}\sum_{j:2^j|\xi|>1} &|\widehat{\boldsymbol{K}}_j(\xi)|^2 \\ &\lesssim \sum_{j:2^j|\xi|>1}\left(\int_{\{\theta\in\boldsymbol{S}^{d-1}:|\Omega(\theta)|\geq(2^j|\xi|)^\nu\|\Omega\|_1\}} |\Omega(\theta)|\,d\sigma(\theta)\right)^2 \\ &= \int_{\boldsymbol{S}^{d-1}\times\boldsymbol{S}^{d-1}} \#\left\{j : 1 < 2^j|\xi| \leq \min\left\{\left(\frac{|\Omega(\theta)|}{\|\Omega\|_1}\right)^{1/\nu}, \left(\frac{|\Omega(\alpha)|}{\|\Omega\|_1}\right)^{1/\nu}\right\}\right\}|\Omega(\theta)|\,|\Omega(\alpha)|\,d\sigma(\theta)\,d\sigma(\alpha) \\ &\lesssim \left(\int_{\boldsymbol{S}^{d-1}} |\Omega(\theta)|\left(1+\left[\log^+\frac{|\Omega(\theta)|}{\|\Omega\|_1}\right]^{1/2}\right)d\sigma(\theta)\right)^2 < \infty,\end{aligned}$$

where the last inequality just follows from $\Omega \in L(\log^+ L)^{1/2}(\boldsymbol{S}^{d-1})$. Hence we complete the proof. □

5B. *Proof of Lemma 3.8.* Recall the definition of the kernel $K^{n,s}_{j,\iota}$ in (4-2). By Young's inequality, it is easy to see that

$$\|T^{n,s}_{j,\iota} g_{n-j,s}\|_{L_2(\mathcal{A})} \leq \|K^{n,s}_{j,\iota}\|_{L_1(\mathbb{R}^d)}\|g_{n-j,s}\|_{L_2(\mathcal{A})} \lesssim \int_{D^\iota} |\Omega(\theta)|\,d\sigma(\theta)\|g_{n-j,s}\|_{L_2(\mathcal{A})}.$$

Now applying $\sum_j \|g_{n-j,s}\|^2_{L_2(\mathcal{A})} \lesssim \lambda\mathcal{C}_\Omega^{-1}\|f\|_{L_1(\mathcal{A})}$ in Lemma 3.7 and the above estimate, we get

$$\begin{aligned}\sum_{s\geq1}\sum_{n\geq100}&\left(\sum_j \|T^{n,s}_{j,\iota} g_{n-j,s}\|^2_{L_2(\mathcal{A})}\right)^{1/2} \\ &\lesssim \sum_{s\geq1}\sum_{n\geq100}\int_{D^\iota} |\Omega(\theta)|\,d\sigma(\theta)(\lambda\mathcal{C}_\Omega^{-1}\|f\|_{L_1(\mathcal{A})})^{1/2} \\ &\lesssim \int_{\boldsymbol{S}^{d-1}} \#\{(s,n) : s\geq1, n\geq100, |\Omega(\theta)| \geq 2^{(n+s)\iota}\|\Omega\|_1\}|\Omega(\theta)|\,d\sigma(\theta)(\lambda\mathcal{C}_\Omega^{-1}\|f\|_{L_1(\mathcal{A})})^{1/2} \\ &\lesssim (\mathcal{C}_\Omega\lambda\|f\|_{L_1(\mathcal{A})})^{1/2},\end{aligned}$$

which is our desired estimate. Hence we complete the proof. □

5C. *Proof of Lemma 3.9.* By applying Plancherel's theorem, the convex inequality for the operator-valued function (4-1), the fact (4-3) and finally Plancherel's theorem again, we get

$$\begin{aligned}\Big\|\sum_{v\in\Theta_{n+s}} G_{n+s,v}T_j^{n,s,v}g_{n-j,s}\Big\|^2_{L_2(\mathcal{A})} &= (2\pi)^{-d/2}\int_{\mathbb{R}^d}\tau\bigg(\Big|\sum_{v\in\Theta_{n+s}}\Phi\Big(2^{(n+s)\gamma}\Big\langle e_v^{n+s},\frac{\xi}{|\xi|}\Big\rangle\Big)\mathcal{F}(T_j^{n,s,v}g_{n-j,s})(\xi)\Big|^2\bigg)\,d\xi\\ &\lesssim \int_{\mathbb{R}^d}\sum_{v\in\Theta_{n+s}}\Phi^2\Big(2^{(n+s)\gamma}\Big\langle e_v^{n+s},\frac{\xi}{|\xi|}\Big\rangle\Big)\sum_{v\in\Theta_{n+s}}\tau(|\mathcal{F}(T_j^{n,s,v}g_{n-j,s})(\xi)|^2)\,d\xi\\ &\lesssim 2^{(n+s)\gamma(d-2)}\sum_{v\in\Theta_{n+s}}\|T_j^{n,s,v}g_{n-j,s}\|^2_{L_2(\mathcal{A})}. \end{aligned}\tag{5-3}$$

Using Young's inequality and (4-8), we get that $\|T_j^{n,s,v}g_{n-j,s}\|^2_{L_2(\mathcal{A})}$ is bounded by

$$\|K_j^{n,s,v}\|^2_{L_1(\mathbb{R}^d)}\|g_{n-j,s}\|^2_{L_2(\mathcal{A})}\lesssim 2^{2(n+s)(-\gamma(d-1)+\iota)}\|\Omega\|_1^2\|g_{n-j,s}\|^2_{L_2(\mathcal{A})}.\tag{5-4}$$

Now plugging (5-4) into (5-3) and using the fact $\operatorname{card}(\Theta_{n+s})\lesssim 2^{(n+s)\gamma(d-1)}$, we get

$$\Big\|\sum_{v\in\Theta_{n+s}} G_{n+s,v}T_j^{n,s,v}g_{n-j,s}\Big\|^2_{L_2(\mathcal{A})}\lesssim 2^{(n+s)(-\gamma+2\iota)}\|\Omega\|_1^2\|g_{n-j,s}\|^2_{L_2(\mathcal{A})},$$

which is just our desired estimate. □

5D. *Proof of Lemma 3.10.* Using $I=\sum_k\Lambda_k$ and the triangle inequality, we get

$$\|(I-G_{n+s,v})T_j^{n,s,v}g_{n-j,s}\|_{L_2(\mathcal{A})}\le\sum_k\|(I-G_{n+s,v})\Lambda_kT_j^{n,s,v}g_{n-j,s}\|_{L_2(\mathcal{A})}.$$

Let $\varepsilon_0\in(0,1)$ be a constant which will be chosen later. Separating the above sum into two parts, we will prove that

$$\begin{aligned}\sum_{k\le j-[(n+s)\varepsilon_0]}&\|(I-G_{n+s,v})\Lambda_kT_j^{n,s,v}g_{n-j,s}\|_{L_2(\mathcal{A})}\\ &\lesssim 2^{-(n+s)(\gamma(d-1)+\varepsilon_0-\gamma(3+2[d/2])-\iota)}\|\Omega\|_1\|g_{n-j,s}\|_{L_2(\mathcal{A})}\end{aligned}\tag{5-5}$$

and

$$\sum_{k>j-[(n+s)\varepsilon_0]}\|(I-G_{n+s,v})\Lambda_kT_j^{n,s,v}g_{n-j,s}\|_{L_2(\mathcal{A})}\lesssim 2^{-(n+s)(\gamma(d-1)+1-\varepsilon_0-\iota)}\|\Omega\|_1\|g_{n-j,s}\|_{L_2(\mathcal{A})}.\tag{5-6}$$

Based on (5-5), (5-6) and the fact $\operatorname{card}(\Theta_{n+s})\lesssim 2^{(n+s)\gamma(d-1)}$, we finish the proof of this lemma by choosing the constants $0<\iota\ll\gamma\ll\varepsilon_0\ll 1$ such that the constant κ is defined by

$$\kappa=\min\Big\{\varepsilon_0-\gamma\Big(3+2\Big[\frac{d}{2}\Big]\Big)-\iota,\,1-\varepsilon_0-\iota\Big\}>0.$$

Now we give the proof of (5-5) and (5-6). Consider (5-5) first. Recall that $D_k^{n,s,v}(x)$ is defined as the kernel of the operator $(I-G_{n+s,v})\Lambda_kT_j^{n,s,v}$ in (4-17). Applying Young's inequality and the estimate of $D_k^{n,s,v}$ in (4-23), we get

$$\begin{aligned}\|(I-G_{n+s,v})\Lambda_k T_j^{n,s,v} g_{n-j,s}\|_{L_2(\mathcal{A})} &\le \|D_k^{n,s,v}\|_{L_1(\mathbb{R}^d)} \|g_{n-j,s}\|_{L_2(\mathcal{A})} \\ &\lesssim 2^{-(n+s)(\gamma(d-1)-\iota-(3+2[d/2])\gamma)-j+k} \|\Omega\|_1 \|g_{n-j,s}\|_{L_2(\mathcal{A})}.\end{aligned}$$

Taking a sum over $k \le j-[(n+s)\varepsilon_0]$ yields (5-5).

Next we turn to the proof of (5-6). By Plancherel's theorem, we see that

$$\|(I-G_{n+s,v})\Lambda_k T_j^{n,s,v} g_{n-j,s}\|_{L_2(\mathcal{A})} \lesssim \|\Lambda_k T_j^{n,s,v} g_{n-j,s}\|_{L_2(\mathcal{A})}. \tag{5-7}$$

Write

$$\Lambda_k T_j^{n,s,v} g_{n-j,s} = \check{\beta}_k * K_j^{n,s,v} * g_{n-j,s} = K_j^{n,s,v} * \check{\beta}_k * g_{n-j,s}.$$

Then by Young's inequality and (4-8), we get

$$\begin{aligned}\|\Lambda_k T_j^{n,s,v} g_{n-j,s}\|_{L_2(\mathcal{A})} &\le \|K_j^{n,s,v}\|_{L_1(\mathbb{R}^d)} \|\check{\beta}_k * g_{n-j,s}\|_{L_2(\mathcal{A})} \\ &\lesssim 2^{-(n+s)(\gamma(d-1)-\iota)} \|\Omega\|_1 \|\check{\beta}_k * g_{n-j,s}\|_{L_2(\mathcal{A})}.\end{aligned} \tag{5-8}$$

Recall the definition of $g_{n-j,s}$; we have the following cancellation property: for all $s \ge 1$ and $Q \in \mathcal{Q}_{n-j+s-1}$, we have $\int_Q g_{n-j,s}(y)\,dy = 0$. Let y_Q be the center of Q. Using this cancellation property, we get

$$\begin{aligned}\check{\beta}_k * g_{n-j,s}(x) &= \int_{\mathbb{R}^d} \sum_{Q\in\mathcal{Q}_{n-j+s-1}} [\check{\beta}_k(x-y)-\check{\beta}_k(x-y_Q)]\chi_Q(y) g_{n-j,s}(y)\,dy \\ &=: \int_{\mathbb{R}^d} K_k(x,y) g_{n-j,s}(y)\,dy,\end{aligned}$$

with $K_k(x,y)=\sum_{Q\in\mathcal{Q}_{n-j+s-1}} [\check{\beta}_k(x-y)-\check{\beta}_k(x-y_Q)]\chi_Q(y)$. Below we will apply Schur's lemma to give an estimate of $\|\check{\beta}_k * g_{n-j,s}\|_{L_2(\mathcal{A})}$. We first consider $K_k(x,y)$ as follows: For any y, there exists a unique cube $Q \in \mathcal{Q}_{n-j+s-1}$ such that $y\in Q$. Then by (4-27),

$$\int_{\mathbb{R}^d} |K_k(x,y)|\,dx \le \int_{\mathbb{R}^d} |y-y_Q| \int_0^1 |\nabla[\check{\beta}_k](x-\rho y-(1-\rho)y_Q)|\,d\rho\,dx \lesssim 2^{j-n-s-k}. \tag{5-9}$$

For any $x\in\mathbb{R}^d$, we have the estimate

$$\begin{aligned}\int_{\mathbb{R}^d} |K_k(x,y)|\,dy &\le \sum_{Q\in\mathcal{Q}_{n-j+s-1}} \int_Q |y-y_Q| \int_0^1 |\nabla[\check{\beta}_k](x-\rho y-(1-\rho)y_Q)|\,d\rho\,dy \\ &\lesssim 2^{j-n-s-k} \int_0^1 \sum_{Q\in\mathcal{Q}_{n-j+s-1}} 2^{-kd} \int_Q |\nabla[\check{\beta}](2^{-k}(x-\rho y-(1-\rho)y_Q))|\,dy\,d\rho \\ &\lesssim 2^{j-n-s-k}\end{aligned} \tag{5-10}$$

once we can show that the estimate below holds uniformly in x, ρ, k

$$\sum_{Q\in\mathcal{Q}_{n-j+s-1}} 2^{-kd} \int_Q |\nabla[\check{\beta}](2^{-k}(x-\rho y-(1-\rho)y_Q))|\,dy \lesssim 1. \tag{5-11}$$

In the following we prove (5-11). Making a change of variables $\tilde{y} = 2^{-k}y$, the integral now integrates over all cubes $Q \in \mathcal{Q}_{n-j+s-1+k}$ with $\tilde{y}_Q = 2^{-k}y_Q$ the center of this cube Q, which is rewritten as

$$\begin{aligned}\sum_{Q\in\mathcal{Q}_{n-j+s-1+k}} &\int_Q |\nabla[\check{\beta}](2^{-k}x-\rho\tilde{y}-(1-\rho)\tilde{y}_Q)|\,d\tilde{y}\\ &=\Bigg(\sum_{\mathrm{dist}(Q,2^{-k}x)\le 2}+\sum_{l=1}^{\infty}\sum_{2^l<\mathrm{dist}(Q,2^{-k}x)\le 2^{l+1}}\Bigg)\int_Q |\nabla[\check{\beta}](2^{-k}x-\rho\tilde{y}-(1-\rho)\tilde{y}_Q)|\,d\tilde{y}\\ &=: I+II,\end{aligned}$$

where in the second line we split the sum $\sum_{Q\in\mathcal{Q}_{n-j+s-1+k}}$ into two parts. Notice that the sidelength of $Q\in\mathcal{Q}_{n-j+s-1+k}$ is $2^{-n+j-s+1-k}$, which is less than 1 since we only consider the sum over $k > j-[(n+s)\varepsilon_0]$ and $0<\varepsilon_0\ll 1$. For I, note that the cubes belonging in $\mathcal{Q}_{n-j+s-1+k}$ are disjoint with interior; therefore the sum $\sum_{\mathrm{dist}(Q,2^{-k}x)\le 2}$ over these cubes is supported in $B(2^{-k}x, 2+\sqrt{d})$, a ball with center $2^{-k}x$ and radius $2+\sqrt{d}$. Thus we get

$$|I| \lesssim \sum_{\mathrm{dist}(Q,2^{-k}x)\le 2} |Q| \le |B(2^{-k}x, 2+\sqrt{d})| \le C.$$

Consider II. Since $\tilde{y}$ lies in a cube $Q\in\mathcal{Q}_{n-j+s-1+k}$ and $\tilde{y}_Q$ is the center of this cube, we get $\rho\tilde{y}+(1-\rho)\tilde{y}_Q$ lies in a line segment which is started at $\tilde{y}_Q$ and ended at $\tilde{y}$. So we have $\rho\tilde{y}+(1-\rho)\tilde{y}_Q\in Q$ for any $\rho\in[0,1]$. Because of $2^l<\mathrm{dist}(Q,2^{-k}x)\le 2^{l+1}$ and $l(Q)\le 1$, we get $|2^{-k}x-\rho\tilde{y}-(1-\rho)\tilde{y}_Q|\approx 2^l$. Combining the above estimates, we get

$$|II| \lesssim \sum_{l=1}^{\infty}\sum_{2^l<\mathrm{dist}(Q,2^{-k}x)\le 2^{l+1}} |Q|2^{-(d+1)l} \lesssim \sum_{l=1}^{\infty} 2^{-l} \le C,$$

where in the first inequality we also use the fact $\nabla[\check{\beta}]$ is a Schwartz function which decays fast away from the origin, while the second inequality follows from the fact that the sum over all cubes $2^l < \mathrm{dist}(Q, 2^{-k}x) \le 2^{l+1}$ is supported in a ball with center $2^{-k}x$ and approximate radius 2^l. Hence we finish the proof of (5-11).

Now utilizing Schur's lemma in Lemma 4.1 with (5-9) and (5-10), we get

$$\|\check{\beta}_k * g_{n-j,s}\|_{L_2(\mathcal{A})} \lesssim 2^{j-n-s-k}\|g_{n-j,s}\|_{L_2(\mathcal{A})}.$$

Plugging this inequality into (5-8) and later (5-7) , we get

$$\|(I-G_{n+s,v})\Lambda_k T_j^{n,s,v} g_{n-j,s}\|_{L_2(\mathcal{A})} \lesssim 2^{j-k-(n+s)(\gamma(d-1)+1-\iota)}\|\Omega\|_1\|g_{n-j,s}\|_{L_2(\mathcal{A})}.$$

Taking a sum of the above estimate over $k > j-[(n+s)\varepsilon_0]$ yields (5-6). Hence we complete the proof. □

Appendix: Strong (p, p) bound for $\{M_r\}_{r>0}$

Theorem A.1. *Suppose that Ω satisfies (1-3) and $\Omega\in L_1(S^{d-1})$. Then the operator sequence $\{M_r\}_{r>0}$ is of maximal strong type (p, p) for $1 < p \le \infty$, i.e.,*

$$\|\{M_r f\}_{r>0}\|_{L_p(\mathcal{A},\ell_\infty(0,\infty))} \lesssim \|\Omega\|_1\|f\|_{L_p(\mathcal{A})}.$$

Proof. By decomposing the functions Ω and f into four parts (i.e., real positive part, real negative part, imaginary positive part, imaginary negative part), together with triangle inequality for the norm $\|\cdot\|_{L_p(\mathcal{A},\ell_\infty(0,\infty))}$, we only consider the case that Ω is a positive function and f is positive in $\mathcal{A}$. Then by (1-5), it is enough to show that for any $f \in L_p^+(\mathcal{A})$ there exists a positive function $F \in L_p^+(\mathcal{A})$ such that

$$M_r f \le F \quad \text{for all } r > 0 \qquad \text{and} \qquad \|F\|_{L_p(\mathcal{A})} \lesssim \|\Omega\|_1 \|f\|_{L_p(\mathcal{A})}. \tag{A-1}$$

We will use the method of rotation. Let $f \in L_p^+(\mathcal{A})$, by making a change of variables $x - y = r\theta$, we get

$$\begin{aligned} M_r f(x) &= \frac{1}{|B(x,r)|} \int_{B(x,r)} \Omega(x-y) f(y)\, dy \\ &= \frac{1}{v_n} \int_{\boldsymbol{S}^{d-1}} \Omega(\theta) \frac{1}{r^d} \int_0^r f(x - s\theta) s^{d-1}\, ds\, d\sigma(\theta) \\ &\lesssim \int_{\boldsymbol{S}^{d-1}} \Omega(\theta) \biggl(\frac{1}{r} \int_0^r f(x - s\theta)\, ds \biggr) d\sigma(\theta). \end{aligned}$$

For a fixed $\theta \in \boldsymbol{S}^{d-1}$, we define the directional Hardy–Littlewood average operator as

$$\mathfrak{M}_r^\theta f(x) = \frac{1}{r} \int_0^r f(x - s\theta)\, ds.$$

We will prove at the end of this section the result

$$\|\{\mathfrak{M}_r^\theta f\}_{r>0}\|_{L_p(\mathcal{A},\ell_\infty(0,\infty))} \lesssim \|f\|_{L_p(\mathcal{A})}. \tag{A-2}$$

Assuming (A-2) and using (1-5), there exists a positive function $F_\theta \in L_p^+(\mathcal{A})$ such that

$$\mathfrak{M}_r^\theta f \le F_\theta \quad \text{for all } r > 0 \qquad \text{and} \qquad \|F_\theta\|_{L_p(\mathcal{A})} \lesssim \|f\|_{L_p(\mathcal{A})}.$$

Now if set $F(x) = \int_{\boldsymbol{S}^{d-1}} \Omega(\theta) F_\theta(x)\, d\sigma(\theta)$, then $M_r f(x) \lesssim F(x)$ and

$$\|F\|_{L_p(\mathcal{A})} \lesssim \int_{\boldsymbol{S}^{d-1}} \Omega(\theta) \|F_\theta\|_{L_p(\mathcal{A})}\, d\sigma(\theta) \lesssim \|\Omega\|_1 \|f\|_{L_p(\mathcal{A})}.$$

Thus F is the desired function satisfying (A-1).

It remains to show (A-2). Let $e_1 = (1, 0, \dots, 0)$ be the unit vector. Now, for any orthogonal matrix A, we have

$$\mathfrak{M}_r^{A(e_1)} f(x) = \mathfrak{M}_r^{e_1} (f \circ A)(A^{-1}x), \tag{A-3}$$

which implies that the L_p boundedness of $\{\mathfrak{M}_r^\theta\}_{r>0}$ can be reduced to that of $\{\mathfrak{M}_r^{e_1}\}_{r>0}$. Let $f \in L_p(L_\infty(\mathbb{R}^d) \overline{\otimes} \mathcal{M})$. Without loss of generality, we may assume that f is positive. Fixing $x_2, \dots, x_d \in \mathbb{R}$, we consider $f(\cdot, x_2, \dots, x_d)$ as a function in $L_p(L_\infty(\mathbb{R}) \overline{\otimes} \mathcal{M})_+$. By the strong-type (p, p) boundedness of noncommutative Hardy–Littlewood maximal operator (see [Mei 2007]), we know that, for $1 < p \le \infty$,

$$\|\{\mathfrak{M}_r^{e_1} f(\cdot, x_2, \dots, x_d)\}_{r>0}\|_{L_p(L_\infty(\mathbb{R}) \overline{\otimes} \mathcal{M}, \ell_\infty(0,\infty))} \lesssim \|f(\cdot, x_2, \dots, x_d)\|_{L_p(L_\infty(\mathbb{R}) \overline{\otimes} \mathcal{M})}.$$

By (1-5), there exists a positive function $F(\cdot, x_2, \dots, x_d) \in L_p(L_\infty(\mathbb{R}) \overline{\otimes} \mathcal{M})$ such that, for any $r > 0$, $\mathfrak{M}_r^{e_1} f(x) \le F(x)$ and

$$\|F(\cdot, x_2, \dots, x_d)\|_{L_p(L_\infty(\mathbb{R})\overline{\otimes}\mathcal{M})} \lesssim \|f(\cdot, x_2, \dots, x_d)\|_{L_p(L_\infty(\mathbb{R})\overline{\otimes}\mathcal{M})}.$$

Then it is easy to see that

$$\|F\|_{L_p(L_\infty(\mathbb{R}^d)\overline{\otimes}\mathcal{M})} \lesssim \|f\|_{L_p(L_\infty(\mathbb{R}^d)\overline{\otimes}\mathcal{M})}.$$

Therefore, we conclude that $\{\mathfrak{M}_r^{e_1}\}_{r>0}$ is of strong-type (p, p). □

Acknowledgement

The author would like to thank the referee for very careful reading and many valuable suggestions.

References

[Cadilhac 2018] L. Cadilhac, "Weak boundedness of Calderón–Zygmund operators on noncommutative L_1-spaces", *J. Funct. Anal.* **274**:3 (2018), 769–796. MR Zbl

[Cadilhac 2019] L. Cadilhac, "Noncommutative Khintchine inequalities in interpolation spaces of L_p-spaces", *Adv. Math.* **352** (2019), 265–296. MR Zbl

[Cadilhac et al. 2022] L. Cadilhac, J. M. Conde-Alonso, and J. Parcet, "Spectral multipliers in group algebras and noncommutative Calderón–Zygmund theory", *J. Math. Pures Appl.* (9) **163** (2022), 450–472. MR Zbl

[Calderón and Zygmund 1952] A. P. Calderón and A. Zygmund, "On the existence of certain singular integrals", *Acta Math.* **88** (1952), 85–139. MR Zbl

[Calderón and Zygmund 1956] A. P. Calderón and A. Zygmund, "On singular integrals", *Amer. J. Math.* **78** (1956), 289–309. MR Zbl

[Caspers et al. 2018] M. Caspers, F. Sukochev, and D. Zanin, "Weak type operator Lipschitz and commutator estimates for commuting tuples", *Ann. Inst. Fourier* (*Grenoble*) **68**:4 (2018), 1643–1669. MR Zbl

[Caspers et al. 2019] M. Caspers, D. Potapov, F. Sukochev, and D. Zanin, "Weak type commutator and Lipschitz estimates: resolution of the Nazarov–Peller conjecture", *Amer. J. Math.* **141**:3 (2019), 593–610. MR Zbl

[Chen et al. 2013] Z. Chen, Q. Xu, and Z. Yin, "Harmonic analysis on quantum tori", *Comm. Math. Phys.* **322**:3 (2013), 755–805. MR Zbl

[Christ 1988] M. Christ, "Weak type $(1, 1)$ bounds for rough operators", *Ann. of Math.* (2) **128**:1 (1988), 19–42. MR Zbl

[Christ and Rubio de Francia 1988] M. Christ and J. L. Rubio de Francia, "Weak type $(1, 1)$ bounds for rough operators, II", *Invent. Math.* **93**:1 (1988), 225–237. MR Zbl

[Grafakos 2014a] L. Grafakos, *Classical Fourier analysis*, 3rd ed., Grad. Texts in Math. **249**, Springer, 2014. MR Zbl

[Grafakos 2014b] L. Grafakos, *Modern Fourier analysis*, 3rd ed., Grad. Texts in Math. **250**, Springer, 2014. MR Zbl

[Grafakos and Stefanov 1999] L. Grafakos and A. Stefanov, "Convolution Calderón–Zygmund singular integral operators with rough kernels", pp. 119–143 in *Analysis of divergence* (Orono, ME, 1997), edited by W. O. Bray, Birkhäuser, Boston, 1999. MR Zbl

[Grafakos et al. 2017] L. Grafakos, D. Oliveira e Silva, M. Pramanik, A. Seeger, and B. Stovall, "Some problems in harmonic analysis", preprint, 2017. arXiv 1701.06637

[Hofmann 1988] S. Hofmann, "Weak $(1, 1)$ boundedness of singular integrals with nonsmooth kernel", *Proc. Amer. Math. Soc.* **103**:1 (1988), 260–264. MR Zbl

[Hong and Xu 2021] G. Hong and B. Xu, "A noncommutative weak type $(1, 1)$ estimate for a square function from ergodic theory", *J. Funct. Anal.* **280**:9 (2021), art. id. 108959. MR Zbl

[Hong et al. 2023] G. Hong, X. Lai, and B. Xu, "Maximal singular integral operators acting on noncommutative L_p-spaces", *Math. Ann.* **386**:1-2 (2023), 375–414. MR Zbl

[Hörmander 1960] L. Hörmander, "Estimates for translation invariant operators in L^p spaces", *Acta Math.* **104** (1960), 93–140. MR Zbl

[Junge 2002] M. Junge, "Doob's inequality for non-commutative martingales", *J. Reine Angew. Math.* **549** (2002), 149–190. MR Zbl

[Mei 2007] T. Mei, *Operator valued Hardy spaces*, Mem. Amer. Math. Soc. **881**, Amer. Math. Soc., Providence, RI, 2007. MR Zbl

[Mei and Parcet 2009] T. Mei and J. Parcet, "Pseudo-localization of singular integrals and noncommutative Littlewood–Paley inequalities", *Int. Math. Res. Not.* **2009**:8 (2009), 1433–1487. MR Zbl

[Parcet 2009] J. Parcet, "Pseudo-localization of singular integrals and noncommutative Calderón–Zygmund theory", *J. Funct. Anal.* **256**:2 (2009), 509–593. MR Zbl

[Pisier 1998] G. Pisier, *Non-commutative vector valued L_p-spaces and completely p-summing maps*, Astérisque **247**, Math. Soc. France, Paris, 1998. MR Zbl

[Pisier and Xu 2003] G. Pisier and Q. Xu, "Non-commutative L^p-spaces", pp. 1459–1517 in *Handbook of the geometry of Banach spaces*, *II*, edited by W. B. Johnson and J. Lindenstrauss, North-Holland, Amsterdam, 2003. MR Zbl

[Seeger 1996] A. Seeger, "Singular integral operators with rough convolution kernels", *J. Amer. Math. Soc.* **9**:1 (1996), 95–105. MR Zbl

[Stein 1993] E. M. Stein, *Harmonic analysis*: *real-variable methods, orthogonality, and oscillatory integrals*, Princeton Math. Series **43**, Princeton Univ. Press, 1993. MR Zbl

[Stein 1998] E. M. Stein, "Singular integrals: the roles of Calderón and Zygmund", *Not. Amer. Math. Soc.* **45**:9 (1998), 1130–1140. MR Zbl

[Tao 1999] T. Tao, "The weak-type (1, 1) of $L \log L$ homogeneous convolution operator", *Indiana Univ. Math. J.* **48**:4 (1999), 1547–1584. MR Zbl

Received 8 Mar 2022. Revised 9 Jul 2022. Accepted 30 Aug 2022.

XUDONG LAI: xudonglai@hit.edu.cn
Institute for Advanced Study in Mathematics, Harbin Institute of Technology, Harbin, China

ANALYSIS AND PDE
Vol. 17 (2024), No. 4, pp. 1473–1500
DOI: 10.2140/apde.2024.17.1473

STRUCTURE OF SETS WITH NEARLY MAXIMAL FAVARD LENGTH

ALAN CHANG, DAMIAN DĄBROWSKI, TUOMAS ORPONEN AND MICHELE VILLA

Let $E \subset B(1) \subset \mathbb{R}^2$ be an $\mathcal{H}^1$ measurable set with $\mathcal{H}^1(E) < \infty$, and let $L \subset \mathbb{R}^2$ be a line segment with $\mathcal{H}^1(L) = \mathcal{H}^1(E)$. It is not hard to see that $\text{Fav}(E) \leq \text{Fav}(L)$. We prove that in the case of near equality, that is,

$$\text{Fav}(E) \geq \text{Fav}(L) - \delta,$$

the set E can be covered by an ϵ-Lipschitz graph, up to a set of length ϵ. The dependence between ϵ and δ is polynomial: in fact, the conclusions hold with $\epsilon = C\delta^{1/70}$ for an absolute constant $C > 0$.

1. Introduction

Let $E \subset \mathbb{R}^2$ be $\mathcal{H}^1$ measurable with $\mathcal{H}^1(E) < \infty$. We recall the definition of Favard length:

$$\text{Fav}(E) = \int_0^\pi \mathcal{H}^1(\pi_\theta(E))\, d\theta.$$

Here $\pi_\theta : \mathbb{R}^2 \to \mathbb{R}$ is the orthogonal projection $\pi_\theta(x) = x \cdot (\cos\theta, \sin\theta)$. The definition of $\text{Fav}(E)$ can be posed without the assumption $\mathcal{H}^1(E) < \infty$, but this hypothesis will be crucial for most of the statements below, and it will be assumed unless otherwise stated. A fundamental result in geometric measure theory is the Besicovitch projection theorem [1939] which relates Favard length and rectifiability: $\text{Fav}(E) > 0$ if and only if $\mathcal{H}^1(E \cap \Gamma) > 0$ for some Lipschitz graph $\Gamma \subset \mathbb{R}^2$ — in other words, E is not purely 1-unrectifiable.

Dąbrowski and Orponen are supported by the Academy of Finland via the project *Incidences on Fractals*, grant 321896. Orponen is also supported by the Academy of Finland via the project *Quantitative rectifiability in Euclidean and non-Euclidean spaces*, grants 309365, 314172. Villa was supported by a starting grant of the University of Oulu.
MSC2020: primary 28A75; secondary 28A78.
Keywords: Favard length, Besicovitch projection theorem, Lipschitz graph.

The proof of the Besicovitch projection theorem is famous for being difficult to quantify, partly because of its reliance on the Lebesgue differentiation theorem: it is hard to decipher from the argument just how large the intersection $E \cap \Gamma$ is, and what the Lipschitz constant of Γ is. In fact, it is nontrivial to even find the right question: for example, if $E \subset B(1)$, $\mathcal{H}^1(E) = 1$, and $\text{Fav}(E) \geq \delta$ for some small but fixed constant $\delta > 0$, then it is not true that $\mathcal{H}^1(E \cap \Gamma) \geq \epsilon$ for some ϵ^{-1}-Lipschitz graph $\Gamma \subset \mathbb{R}^2$, where $\epsilon = \epsilon(\delta) > 0$. We construct a relevant counterexample in Section 6.

In Theorem 1.1, we show that similar counterexamples are no longer possible if the assumption "$\text{Fav}(E) \geq \delta$" is upgraded to "$\text{Fav}(E) \geq 2\,\mathcal{H}^1(E) - \delta$" for a sufficiently small constant $\delta > 0$. The number 2 comes from the fact that $\text{Fav}([0, 1] \times \{0\}) = 2$ and that $[0, 1] \times \{1\}$ has the maximal Favard length among sets of length unity (see (2.4)).

Theorem 1.1. *For every $\epsilon > 0$ there exists $\delta > 0$ such that the following holds: Let $E \subset B(1)$ be an $\mathcal{H}^1$ measurable set with $\mathcal{H}^1(E) < \infty$, and assume that*

$$\text{Fav}(E) \geq \text{Fav}(L) - \delta, \tag{1.2}$$

where $L \subset \mathbb{R}^2$ is a line segment with $\mathcal{H}^1(L) = \mathcal{H}^1(E)$. Then, there exists an ϵ-Lipschitz graph $\Gamma \subset \mathbb{R}^2$ such that $\mathcal{H}^1(E \cap \Gamma) \geq \mathcal{H}^1(E) - \epsilon$. One can take $\delta = \epsilon^{70}/C$ for an absolute constant $C > 1$.

By an *ϵ-Lipschitz graph* we mean a set of the form $R(\text{Graph}_f)$, where $R\colon \mathbb{R}^2 \to \mathbb{R}^2$ is a rotation, and $\text{Graph}_f = \{(x, f(x)) : x \in \mathbb{R}\}$ is the graph of an ϵ-Lipschitz function $f\colon \mathbb{R} \to \mathbb{R}$. This means that

$$|f(x) - f(y)| \leq \epsilon |x - y|$$

for all $x, y \in \mathbb{R}$. It is easy to check that the intersection of an ϵ-Lipschitz graph with $B(1)$ is contained in the 2ϵ-neighborhood of some line $\ell \subset \mathbb{R}^2$, so in particular the same is true of $E \cap \Gamma$ (as in Theorem 1.1).

Theorem 1.1 shows that if $\text{Fav}(E)$ is nearly maximal, the Besicovitch projection theorem can be quantified in a very strong way, whereas the example constructed in Section 6 shows that any similar conclusion fails completely if we make the weaker assumption $\text{Fav}(E) \geq \delta$. However, it remains plausible that the assumption $\text{Fav}(E) \geq \delta$ is sufficient to guarantee a quantitative version of Besicovitch's theorem under the additional assumption that E is 1-Ahlfors regular, or satisfies other *multiscale 1-dimensionality* hypotheses. For recent partial results, and more discussion on this question; see [Davey and Taylor 2022; Martikainen and Orponen 2018; Orponen 2021; Tao 2009]. The problem is closely related to Vitushkin's conjecture [1967] on the connection between analytic capacity and Favard length; see [Chang and Tolsa 2020; Dąbrowski and Villa 2022].

We briefly mention another closely related topic: if $E \subset \mathbb{R}^2$ is self-similar and purely 1-unrectifiable, then $\text{Fav}(E) = 0$ by the Besicovitch projection theorem. It is an interesting and very popular question to attempt quantifying the (sharp) rate of decay at which $\text{Fav}(E_n) \to 0$, where E_n is the *n-th iteration* of the self-similar set. For recent developments; see [Bateman and Volberg 2010; Bond et al. 2014; Bond and Volberg 2010; 2012; Bongers and Taylor 2023; Cladek et al. 2022; Łaba and Zhai 2010; Łaba 2015; Łaba and Marshall 2022; Nazarov et al. 2010; Peres and Solomyak 2002].

It is tempting to consider the following scale-invariant version of Theorem 1.1: *for any* $\epsilon_1, \epsilon_2 > 0$ *there exists* $\delta > 0$ *such that if* $E \subset B(1)$ *satisfies* $\mathcal{H}^1(E) < \infty$ *and*

$$\operatorname{Fav}(E) \geq (1-\delta)\operatorname{Fav}(L),$$

then there exists an ϵ_1*-Lipschitz graph* $\Gamma \subset \mathbb{R}^2$ *such that* $\mathcal{H}^1(E \setminus \Gamma) \leq \epsilon_2 \mathcal{H}^1(E)$. Note that for sets E with $\mathcal{H}^1(E) \sim 1$ this statement is equivalent to Theorem 1.1; however, in general, the statement is false. Consider a set E_n consisting of four horizontal segments of length $1/n$ placed in the corners of $[0, 1]^2$. Clearly, one may cover at most half of E_n using a single 1-Lipschitz graph. At the same time, $\operatorname{Fav}(E_n)/\operatorname{Fav}(L_n) \to 1$, where $L_n = [0, 4/n] \times \{0\}$. To see this, let $\mathcal{B}_n := \{\theta \in [0, \pi) : \pi_\theta \text{ is not injective on } E_n\}$. Note that $\mathcal{H}^1(\mathcal{B}_n) \to 0$, and at the same time for $\theta \notin \mathcal{B}_n$ we have $\mathcal{H}^1(\pi_\theta(E_n)) = \mathcal{H}^1(\pi_\theta(L_n))$. It follows easily that $\operatorname{Fav}(E_n)/\operatorname{Fav}(L_n) \to 1$.

1A. ***Outline of the paper.*** A quick outline of the article is as follows: In Section 2 we introduce Crofton's formula and prove that line segments maximize Favard length. In Section 3 we prove Theorem 1.1 using two main propositions, Proposition 3.3 and Proposition 3.11. The moral of these propositions is discussed at the beginning of Section 3. These two propositions are then proven in Section 4 and Section 5, respectively. Section 6 contains the counterexample mentioned above to the scale-invariant version of Theorem 1.1. Finally, in the Appendix we give an exact formula for the measure of lines spanned by two rectifiable curves — this is used in Section 5 but it might be of independent interest.

2. Measure-theoretic preliminaries

2A. ***Notation.*** For $x \in \mathbb{R}^d$ and $r > 0$, the notation $B(x, r)$ stands for a closed ball of radius r centered at x. For $A \subset \mathbb{R}^d$, we denote the cardinality of A by $\#A$, and we write $A(r) := \{x \in \mathbb{R}^d : \operatorname{dist}(x, A) \leq r\}$, where "dist" is Euclidean distance. For $f, g \geq 0$, we write $f \lesssim g$ if there exists an absolute constant $C > 0$ such that $f \leq Cg$. The notation $f \gtrsim g$ means the same as $g \lesssim f$, and $f \sim g$ is shorthand for $f \lesssim g \lesssim f$. If the constant $C > 0$ is allowed to depend on some parameter p, we signify this by writing $f \lesssim_p g$.

2B. ***Integralgeometry and Crofton's formula.*** One of the main tools is Crofton's formula for rectifiable sets, which states the following: if $E \subset \mathbb{R}^2$ is an $\mathcal{H}^1$ measurable 1-rectifiable set with $\mathcal{H}^1(E) < \infty$, then

$$\mathcal{H}^1(E) = \frac{1}{2}\int_0^\pi \int_{\mathbb{R}} \#(E \cap \pi_\theta^{-1}\{t\})\, dt\, d\theta. \tag{2.1}$$

Equation (2.1) is false without the rectifiability assumption, but the inequality "$\geq$" remains valid in this case. This formula (and the inequality) is a special case of a more general relation between Hausdorff measure and integralgeometric measure for n-rectifiable sets in $\mathbb{R}^d$; see [Federer 1947, Theorem 9.7; 1969, Theorem 3.2.26]. We next rephrase the formula (2.1) in slightly more abstract terms. We define the following measure η on the family $\mathcal{A} := \mathcal{A}(2, 1)$ of all affine lines in $\mathbb{R}^2$:

$$\eta(\mathcal{L}) = \int_0^\pi \mathcal{H}^1(\{t \in \mathbb{R} : \pi_\theta^{-1}\{t\} \in \mathcal{L}\})\, d\theta, \quad \mathcal{L} \subset \mathcal{A}.$$

With this notation, the Crofton formula (2.1) can be rewritten as

$$\mathcal{H}^1(E) = \frac{1}{2} \int_{\mathcal{L}(E)} \#(E \cap \ell)\, d\eta(\ell), \tag{2.2}$$

where

$$\mathcal{L}(E) := \{\ell \in \mathcal{A} : E \cap \ell \neq \varnothing\}.$$

Lemma 2.3 (the line segment maximizes Favard length). *If $E \subset \mathbb{R}^2$ is $\mathcal{H}^1$ measurable, $\mathcal{H}^1(E) < \infty$, and $L \subset \mathbb{R}^2$ is a line segment with $\mathcal{H}^1(E) = \mathcal{H}^1(L)$, then*

$$\operatorname{Fav}(E) \le \operatorname{Fav}(L) \tag{2.4}$$

and

$$\operatorname{Fav}(L) - \operatorname{Fav}(E) \ge \int_{\mathcal{L}(E)} (\#(E \cap \ell) - 1)\, d\eta(\ell). \tag{2.5}$$

If E is rectifiable, then equality holds in (2.5).

Proof. Suppose $E \subset \mathbb{R}^2$ is $\mathcal{H}^1$ measurable, $\mathcal{H}^1(E) < \infty$, and $L \subset \mathbb{R}^2$ is a line segment with $\mathcal{H}^1(E) = \mathcal{H}^1(L)$. Then

$$\operatorname{Fav}(E) = \eta(\mathcal{L}(E)) = \int_{\mathcal{L}(E)} 1\, d\eta(\ell) \le \int_{\mathcal{L}(E)} \#(E \cap \ell)\, d\eta(\ell) \le 2\mathcal{H}^1(E). \tag{2.6}$$

If we replace E with the line segment L, then equality holds in both inequalities above. Thus, $\operatorname{Fav}(L) = 2\mathcal{H}^1(L) = 2\mathcal{H}^1(E)$, which combined with (2.6) (for E) proves (2.5).

Next, (2.4) follows from the fact that the right-hand side of (2.5) is nonnegative. Finally, if E is rectifiable, then the second inequality in (2.6) becomes an equality, which implies that equality holds in (2.5). □

2C. *Coarea formula.* We now record another tool in the proof of Theorem 1.1. It is closely related to Crofton's formula, but only considers the intersections with lines in a fixed direction. The price to pay is that the tangent of the rectifiable set enters the formula. It is a generalization of the following standard fact: if $f : [a, b] \to \mathbb{R}$ is α-Lipschitz, then

$$\mathcal{H}^1\big(\{(t, f(t)) : t \in [a, b]\}\big) = \int_a^b \sqrt{1 + f'(t)^2}\, dt \le \sqrt{1 + \alpha^2}\,(b - a).$$

Lemma 2.7 (coarea formula). *Let $\alpha > 0$. Let $E \subset \mathbb{R}^2$ be a countable union of α-Lipschitz graphs over the x-axis. Then,*

$$\mathcal{H}^1(A) \le \sqrt{1 + \alpha^2} \int_{\mathbb{R}} \#(A \cap \pi_0^{-1}\{t\})\, dt \tag{2.8}$$

for all $\mathcal{H}^1$ measurable subsets $A \subset E$. (Recall that $\pi_0 : \mathbb{R}^2 \to \mathbb{R}$ is the projection onto the x-axis.)

Proof. This follows from the coarea formula for rectifiable sets. (See, e.g., [Federer 1969, Theorem 3.2.22] or [Krantz and Parks 2008, Theorem 5.4.9].) □

3. Proof of Theorem 1.1 in two main steps

In this section we prove Theorem 1.1 using Propositions 3.3 and 3.11 introduced below. Proposition 3.3 says roughly the following: Assume a priori that E is a union of line segments (we reduce matters to something like this in Section 3A), fix a small *angle* $\alpha > 0$, and let $E_{\ell,\alpha}$ be the union of those segments which make an angle $\leq \alpha$ with some given line $\ell \subset \mathbb{R}^2$. Evidently E can be expressed as the union of $\sim 1/\alpha$ sets of the form $E_{\ell,\alpha}$. Proposition 3.3 says that if the parameter δ in our hypothesis $\mathrm{Fav}(E) \geq \mathrm{Fav}(L) - \delta$ is sufficiently small, then each of the sets $E_{\ell,\alpha}$ can be (almost) covered by a single ($\sim\alpha$)-Lipschitz graph over ℓ. After this step, we know that E can be (almost) covered by a union of $\sim 1/\alpha$ Lipschitz graphs with constant $\sim\alpha$. Thereafter, to complete the proof of Theorem 1.1, it remains to show that only one of these graphs can have a nontrivial intersection with E. This uses the hypothesis $\mathrm{Fav}(E) \geq \mathrm{Fav}(L) - \delta$ once more, and is accomplished in Proposition 3.11 (and the discussion right below).

3A. ***Step 1: first reductions.*** Let $E \subset \mathbb{R}^2$ be a Borel set with $\mathcal{H}^1(E) < \infty$. We start with the following simple lemma:

Lemma 3.1. *It suffices to prove Theorem 1.1 under the additional assumption that E is a finite union of disjoint C^1 curves.*

Proof. We may assume that $E \subset B(1)$ is rectifiable, because by the Besicovitch projection theorem, the rectifiable part of E continues to satisfy all the assumptions of Theorem 1.1 (with the same constant $\delta > 0$). By this assumption, $\mathcal{H}^1$ almost all of E can be covered by a countable union of C^1-curves. Decomposing the curves further, we may assume that they are disjoint, and for any given $\eta > 0$ we may write

$$E = \bigcup_{j=1}^{M_1} (\gamma_j \cap E) \cup S,$$

where $\mathcal{H}^1(S) \leq \eta$, and $\mathcal{H}^1(E \cap \gamma_j) \geq (1-\eta)\mathcal{H}^1(\gamma_j)$. Now, the set $\bar{E} := \bigcup_{j=1}^{M_1} \gamma_j$ satisfies

$$\mathcal{H}^1(\bar{E}) \leq (1-\eta)^{-1}\mathcal{H}^1(E) \quad \text{and} \quad \mathrm{Fav}(\bar{E}) \geq \mathrm{Fav}(E) - \eta$$

and is additionally a finite union of disjoint C^1-curves. If Theorem 1.1 is already known under this additional assumption, we may now infer that $\mathcal{H}^1(\bar{E} \setminus \Gamma) \leq \epsilon$, where Γ is an ϵ-Lipschitz graph. But then also $\mathcal{H}^1(E \setminus \Gamma) \leq \mathcal{H}^1(E \setminus \bar{E}) + \mathcal{H}^1(\bar{E} \setminus \Gamma) \leq \eta + \epsilon$, and Theorem 1.1 follows for E by choosing the parameters ϵ, η appropriately. □

3B. ***Step 2: minigraphs and how to merge them.*** By Lemma 3.1, we may assume that E is a finite union of disjoint C^1-curves $\gamma_1, \ldots, \gamma_{M_1}$. We further chop up each curve γ_j into connected pieces whose tangent varies by less than α, where α is a small constant depending on ϵ fixed later on (see (3.5)). At this point, we have managed to write E as a finite union of disjoint α-Lipschitz graphs $\gamma_1, \ldots, \gamma_{M_1'}$, where $M_1 \leq M_1' < +\infty$. At this point we have no quantitative control on the constant M_1'. Each of the graphs γ_j will be called a *minigraph*, and their collection is denoted $\mathcal{E}$. The main tasks in Theorem 1.1 are to combine the minigraphs into roughly $1/\alpha$ bigger graphs, and to show that nearly all of E lies on just one of these bigger graphs.

To begin with, let $M_2 = \lceil \pi\alpha^{-1} \rceil \sim \alpha^{-1}$. We would like to divide the collection of minigraphs $\mathcal{E}$ into M_2 subcollections $\mathcal{E}_1, \dots \mathcal{E}_{M_2}$, each of them containing the minigraphs with roughly the same direction. To do this, we consider M_2 vectors of the form

$$v_k := (\cos(k\pi/M_2), \sin(k\pi/M_2)) \quad \text{for } 1 \le k \le M_2 \sim \alpha^{-1}.$$

Observe that for each minigraph $\gamma \in \mathcal{E}$ there exists $k \in \{1, \dots, M_2\}$ such that γ is a 2α-Lipschitz graph over the line $\operatorname{span}(v_k)$. The vector v_k will be called the *direction* of the minigraph (if there are several suitable vectors for one minigraph, fix any one of them; we will only need to know that each minigraph is a 2α-Lipschitz graph over the line spanned by its direction). Statements about the (relative) angles of minigraphs should always be interpreted as statements about the relative angles of the direction vectors v_k.

For $k \in \{1, \dots, M_2\}$ fixed, we define $\mathcal{E}_k \subset \mathcal{E}$ as the collection of minigraphs with direction v_k. We suggest that the reader visualize the minigraphs as line segments I with $\angle(I, \operatorname{span}(v_k)) \le \alpha$. It seems likely that Theorem 1.1 could be reduced to the case where E is a finite union of line segments, but employing the minigraphs seems to spare us some unnecessary steps.

We write $E_k := \bigcup \mathcal{E}_k$. Thus

$$E = E_1 \cup \dots \cup E_{M_2}. \tag{3.2}$$

It turns out that, except for a small error, each set E_k is covered by a single Lipschitz graph with constant $\sim\alpha$ over $\operatorname{span}(v_k)$. Indeed, note that Lemma 2.3 and (1.2) together imply $\int_{\mathcal{L}(E)} \#(E \cap \ell) - 1 \, d\eta(\ell) \le \delta$. Since for each $k \in \{1, \dots, M_2\}$ we have $E_k \subset E$, one sees immediately that $\mathcal{L}(E_k) \subset \mathcal{L}(E)$ and $\#(E_k \cap \ell) \le \#(E \cap \ell)$, so that we also get $\int_{\mathcal{L}(E_k)} \#(E_k \cap \ell) - 1 \, d\eta(\ell) \le \delta$. Then, the desired Lipschitz graph Γ covering most of E_k is constructed in the following proposition, whose proof will be carried out in Section 4:

Proposition 3.3. *There exist absolute constants* $C_0, \alpha_0 \in (0, 1)$ *and* $\boldsymbol{C}_{\text{lip}} > 1$ *such that the following holds: Let* $\delta, \epsilon \in (0, 1)$ *and* $\alpha \in (0, \alpha_0)$ *be such that* $\delta \le C_0\alpha^3\epsilon^2$. *Let* $E \subset B(1)$ *be a set with* $\mathcal{H}^1(E) < \infty$ *of the form*

$$E = \bigcup_{\gamma \in \mathcal{E}} \gamma,$$

where $\mathcal{E}$ *is a finite collection of disjoint* α*-Lipschitz graphs over a fixed line* $L \subset \mathbb{R}^2$. *Assume further that* E *satisfies*

$$\int_{\mathcal{L}(E)} (\#(E \cap \ell) - 1) \, d\eta(\ell) \le \delta. \tag{3.4}$$

Then, there exists a Lipschitz graph Γ *over* L, *with Lipschitz constant at most* $\boldsymbol{C}_{\text{lip}} \cdot \alpha$, *such that*

$$\mathcal{H}^1(E \setminus \Gamma) \le \epsilon.$$

We remark that the absolute constants α_0 and $\boldsymbol{C}_{\text{lip}}$ are such that $\alpha_0 \le \boldsymbol{C}_{\text{lip}}^{-1}$. In particular, the $\boldsymbol{C}_{\text{lip}}\alpha$-Lipschitz graph Γ from above has a Lipschitz constant bounded by 1.

The proof of Proposition 3.3 recycles most of the ideas from Besicovitch's original proof of the Besicovitch projection theorem [1939]. Indeed, we first use (in Lemma 4.1) the assumption (3.4) to show that E must have arbitrarily low conical density in arbitrarily wide cones centered at most points $x \in E$, whose axis is perpendicular to the line L. The quantifications of *arbitrarily low* and *arbitrarily wide* can be made stronger by reducing the value of the constants α and δ. After this step, we use Besicovitch's *two cones* argument (quantified in Lemma 4.18) to show that most of E can be contained on a Lipschitz graph over L.

3C. ***Step 3*: *there can only be one graph.*** In Proposition 3.3 we managed to pack a majority of each set E_j (as defined in (3.2)) to a Lipschitz graph of constant $\sim\alpha$, up to errors which tend to zero as $\delta \to 0$ in the main assumption (1.2). However, at this point there might be up to $\sim\alpha^{-1}$ distinct Lipschitz graphs, and to prove Theorem 1.1, we would (roughly speaking) like to reduce their number to one. That this should be possible is not hard to believe: if E consists of several distinct Lipschitz graphs of substantial measure, which nevertheless cannot be fit into a single Lipschitz graph, then $\mathrm{Fav}(E)$ cannot possibly be maximal.

We turn to the details. We recall the *given* constant $\epsilon > 0$ from the statement of Theorem 1.1, and we set

$$\delta := \frac{\epsilon^{70}}{\boldsymbol{C}_{\mathrm{thm}}}$$

for a sufficiently large absolute constant $\boldsymbol{C}_{\mathrm{thm}} > 1$. We define also

$$\alpha := \left(\frac{\epsilon}{\boldsymbol{C}_{\mathrm{alp}}}\right)^{10} \tag{3.5}$$

for some universal $\boldsymbol{C}_{\mathrm{alp}} > 1$. The universal constant $\boldsymbol{C}_{\mathrm{thm}}$ will depend on $\boldsymbol{C}_{\mathrm{alp}}$, whereas $\boldsymbol{C}_{\mathrm{alp}}$ depends only on $\boldsymbol{C}_{\mathrm{lip}}$ and another constant $\boldsymbol{C}_{\mathrm{sep}}$, which is introduced below. The additional constant $\boldsymbol{C}_{\mathrm{alp}}$ will make it easier for us to ensure that the Lipschitz graph Γ obtained from the application of Proposition 3.3 has Lipschitz constant smaller than ϵ; see the discussion around (3.8). We record that

$$\alpha^7 = \boldsymbol{C}_{\mathrm{alp}}^{-70}\epsilon^{70} = \boldsymbol{C}_{\mathrm{thm}}\boldsymbol{C}_{\mathrm{alp}}^{-70}\cdot\delta. \tag{3.6}$$

Recall, once more, the decompositions $\mathcal{E} = \mathcal{E}_0 \cup \cdots \cup \mathcal{E}_{M_2}$ and $E = E_0 \cup \cdots \cup E_{M_2}$ from the previous subsection: this decomposition depends on the parameter α fixed above. In addition to the decomposition $E = E_0 \cup \cdots \cup E_{M_2}$, we will also need another, coarser, decomposition of E in this section. Write $\kappa := \frac{1}{10}$, fix $M_3 \sim \alpha^{-\kappa}$, and decompose $\mathcal{E} = \mathcal{F}_0 \cup \cdots \cup \mathcal{F}_{M_3}$ in such a way that

- each $\mathcal{F}_k$ is a union of finitely many consecutive families $\mathcal{E}_j$, and
- $\mathcal{F}_k$ contains those minigraphs whose direction makes an angle no larger than α^κ with $w_k = (\cos(k\pi/M_3), \sin(k\pi/M_3))$ for $0 \le k \le M_3$.

We write

$$F_k := \bigcup \mathcal{F}_k, \quad 0 \le k \le M_3 \sim \alpha^{-\kappa}.$$

At this point, we consider two distinct cases. Let $\boldsymbol{C}_{\mathrm{sep}}$ be a large constant depending only on the absolute constant $\boldsymbol{C}_{\mathrm{lip}}$ appearing in Proposition 3.3 (the letters sep stand for *separation*). Thus, the constant $\boldsymbol{C}_{\mathrm{sep}}$ is also absolute, and we may (and will) assume that $\boldsymbol{C}_{\mathrm{alp}}$ is large relative to $\boldsymbol{C}_{\mathrm{sep}}$.

Case 1. Given the constant $\epsilon > 0$ from Theorem 1.1, the first case is that we can find consecutive sets $F_k, F_{k+1}, \dots, F_{k+\boldsymbol{C}_{\text{sep}}}$ with the property

$$\mathcal{H}^1(E \setminus (F_k \cup \dots \cup F_{k+\boldsymbol{C}_{\text{sep}}})) \le \epsilon. \tag{3.7}$$

In this case we note that $F := F_k \cup \dots \cup F_{k+\boldsymbol{C}_{\text{sep}}}$ is a union of minigraphs whose directions are within $\lesssim \boldsymbol{C}_{\text{sep}}\alpha^{\kappa}$ of the fixed vector w_k. In particular, F can be expressed as a union of finitely many disjoint $\bar{\alpha}$-Lipschitz graphs over the line $\text{span}(w_k)$, with $\bar{\alpha} \sim \boldsymbol{C}_{\text{sep}}\alpha^{\kappa}$. This will place us in a position to use Proposition 3.3 (with E replaced by F and α replaced by $\bar{\alpha}$). Of course also

$$\int_{\mathcal{L}(F)} (\#(F \cap \ell) - 1)\, d\eta(\ell) \le \int_{\mathcal{L}(E)} (\#(E \cap \ell) - 1)\, d\eta(\ell) \le \delta,$$

so the analogue of the assumption (3.4) is valid for F in place of E. We also note that

$$\delta = \boldsymbol{C}_{\text{thm}}^{-1}\epsilon^{70} \le \boldsymbol{C}_{\text{thm}}^{-1}\boldsymbol{C}_{\text{alp}}^{3} \cdot (\epsilon/\boldsymbol{C}_{\text{alp}})^3 \cdot \epsilon^2 = (\boldsymbol{C}_{\text{thm}}^{-1}\boldsymbol{C}_{\text{alp}}^{3}) \cdot \alpha^{3\kappa}\epsilon^2 \sim (\boldsymbol{C}_{\text{thm}}^{-1}\boldsymbol{C}_{\text{alp}}^{3}\boldsymbol{C}_{\text{sep}}^{-3}) \cdot \bar{\alpha}^3\epsilon^2,$$

so if $\boldsymbol{C}_{\text{thm}}$ is sufficiently large relative to $\boldsymbol{C}_{\text{alp}}$, then the hypothesis in Proposition 3.3 on the relation between $\delta, \bar{\alpha}$, and ϵ is satisfied (the constant $\boldsymbol{C}_{\text{sep}}$ is large, so it can be safely ignored here). Consequently, there exists a Lipschitz graph $\Gamma \subset \mathbb{R}^2$ of constant $\lesssim \boldsymbol{C}_{\text{lip}}\boldsymbol{C}_{\text{sep}} \cdot \alpha^{\kappa} = \boldsymbol{C}_{\text{lip}}\boldsymbol{C}_{\text{sep}} \cdot \epsilon/\boldsymbol{C}_{\text{alp}}$ with the property

$$\mathcal{H}^1(F \setminus \Gamma) \le \epsilon, \tag{3.8}$$

and consequently $\mathcal{H}^1(E \setminus \Gamma) \le 2\epsilon$. By choosing $\boldsymbol{C}_{\text{alp}}$ sufficiently large relative to $\boldsymbol{C}_{\text{sep}}$ and $\boldsymbol{C}_{\text{lip}}$, we may ensure that Γ is an ϵ-Lipschitz graph, as desired.

Case 2. We then move to consider the other option, where E cannot be exhausted, up to measure ϵ, by a constant number of consecutive sets $F_k, F_{k+1}, \dots, F_{k+\boldsymbol{C}_{\text{sep}}}$. Since (3.7) fails for every k, we may find an index pair $k, l \in \{0, \dots, M_3\}$ with $|k - l| \ge \boldsymbol{C}_{\text{sep}}$ such that

$$\mathcal{H}^1(F_k) \ge \alpha^{2\kappa} \quad \text{and} \quad \mathcal{H}^1(F_l) \ge \alpha^{2\kappa}. \tag{3.9}$$

This follows immediately from the pigeonhole principle, recalling that the cardinality of the pieces F_k is $\lesssim \alpha^{-\kappa}$, and also that α^{κ} is much smaller than ϵ by (3.5).

Remark 3.10. Recall that the *separation* constant $\boldsymbol{C}_{\text{sep}}$ above has been chosen to be large relative to the constant $\boldsymbol{C}_{\text{lip}}$ in Proposition 3.3: morally, if Γ_1, Γ_2 are two $\boldsymbol{C}_{\text{lip}}\alpha^{\kappa}$-Lipschitz graphs over lines L_1, L_2 with $\angle(L_1, L_2) \ge \boldsymbol{C}_{\text{sep}}\alpha^{\kappa}$, we need to know that Γ_1 and Γ_2 are still *transversal* (their tangents form angles $\ge \frac{1}{2}\boldsymbol{C}_{\text{sep}}\alpha^{\kappa}$ with each other).

The next key proposition will imply that Case 2 cannot happen:

Proposition 3.11. *Suppose that $\boldsymbol{C}_{\text{sep}} > 0$ is sufficiently large, and suppose that there are $k, l \in \{0, \dots, M_3\}$ with $|k - l| \ge \boldsymbol{C}_{\text{sep}}$ such that*

$$\mathcal{H}^1(F_k) \ge \alpha^{2\kappa} \quad \textit{and} \quad \mathcal{H}^1(F_l) \ge \alpha^{2\kappa}.$$

Then

$$\int_{\mathcal{L}(E)} (\#(E \cap \ell) - 1)\, d\eta(\ell) \gtrsim \alpha^7. \tag{3.12}$$

As we recorded in (3.6), we have $\alpha^7 = \boldsymbol{C}_{\mathrm{thm}}\boldsymbol{C}_{\mathrm{alp}}^{-70} \cdot \delta$. Thus, if $\boldsymbol{C}_{\mathrm{thm}}$ is chosen sufficiently large relative to $\boldsymbol{C}_{\mathrm{alp}}$ and the implicit absolute constants in (3.12), then (3.12) would lead to the contradiction

$$\delta \geq \int_{\mathcal{L}(E)} (\#(E \cap \ell) - 1)\, d\eta(\ell) > \delta.$$

(For the first inequality, recall (2.5) and our main assumption (1.2).) Thus, with the choices of constants specified in this section, Case 2 cannot occur. This concludes the proof of Theorem 1.1.

In the next two sections we prove the two key results used above, Propositions 3.3 and 3.11.

4. Proof of Proposition 3.3

Let $E \subset \mathbb{R}^2$ be as in the proposition. With no loss of generality, we may assume that L is the x-axis, so the minigraphs in $\mathcal{E}$ are roughly horizontal. We introduce further notation. We write

$$\mathcal{C}_\beta := \{(x, y) \in \mathbb{R}^2 : |y| \geq \beta|x|\}, \quad \beta > 0.$$

Thus, the smaller the β, the wider the cone. We also write

$$\mathcal{C}_\beta(x) := x + \mathcal{C}_\beta \quad \text{and} \quad \mathcal{C}_\beta(x, r) := \mathcal{C}_\beta(x) \cap B(x, r).$$

With this notation, if a set $\Gamma \subset \mathbb{R}^2$ satisfies $\Gamma \cap \mathcal{C}_\beta(x) = \{x\}$ for all $x \in \Gamma$, then Γ is (a subset of) a β-Lipschitz graph. Thus, in view of Proposition 3.3, it would be desirable to show that $E \cap \mathcal{C}_{\boldsymbol{C}_{\mathrm{lip}}\alpha}(x) = \{x\}$ for all $x \in E$. In reality, we will prove a similar statement about a subset of E (of nearly full length). It is worth noting that a toy version of these statements is already present in our hypotheses: each minigraph $\gamma \in \mathcal{E}$ is an α-Lipschitz graph over the x-axis.

Define the maximal conical density

$$\Theta^*_{E,\beta}(x) = \sup_{r>0} \frac{\mathcal{H}^1(\mathcal{C}_\beta(x, r) \cap E)}{r}.$$

Lemma 4.1 says that points of high conical density are negligible, whereas Lemma 4.18 says that points of low conical density can be mostly contained in a Lipschitz graph.

Lemma 4.1 (high conical density points are negligible). *Let $E \subset B(1)$, $\alpha \in (0, \alpha_0)$ and $\delta \in (0, 1)$ be as in Proposition 3.3, so that in particular (3.4) holds. Let $\varepsilon > 0$. If the absolute constant $\boldsymbol{C}_{\mathrm{lip}} > 0$ is chosen sufficiently large, then*

$$\mathcal{H}^1(\{x \in E : \Theta^*_{E,\alpha'}(x) \geq \varepsilon\}) \lesssim \frac{\delta}{\varepsilon\alpha^2}, \tag{4.2}$$

where $\alpha' := \boldsymbol{C}_{\mathrm{lip}}\alpha/2$.

Write $\ell_{x,\theta} := \pi_\theta^{-1}\{\pi_\theta(x)\}$ for $\theta \in [0, \pi)$, so that $\ell_{0,\theta} = \mathrm{span}(\cos\theta, \sin\theta)^\perp$. Let $J(\beta) \subset [0, \pi)$ be the set of directions in the cone $\mathcal{C}_\beta$, i.e.,

$$J(\beta) = \{\theta \in [0, \pi) : \ell_{0,\theta} \subset \mathcal{C}_\beta\} = \{\theta \in [0, \pi) : \mathrm{span}(\cos\theta, \sin\theta)^\perp \subset \mathcal{C}_\beta\}.$$

If ℓ is a line, we let $\ell(w)$ denote the tube that is the w-neighborhood of ℓ. For a tube $T = \ell(w)$, we write $w(T) = w$.

To prove Lemma 4.1, we rely on the following lemma:

Lemma 4.3 (the Besicovitch alternative). *Let $E \subset \mathbb{R}^2$ and $\beta \leq 1$. Then for all $x \in E$ and $H \geq 1$, at least one of the following two alternatives holds*:

(A1) *There exists a set $I_x \subset J(\beta)$ of measure $\mathcal{H}^1(I_x) \geq H^{-1}$ such that*

$$\#(E \cap \ell_{x,\theta}) \geq 2, \quad \theta \in I_x.$$

(A2) *There exists a set $J_x \subset J(\beta)$ of measure $\mathcal{H}^1(J_x) \gtrsim H^{-1}$ with the following property*: *for every $\theta \in J_x$, there is a tube $T = T_{x,\theta} = \ell_{x,\theta}(w(T))$ centered around $\ell_{x,\theta}$ such that*

$$\mathcal{H}^1(E \cap T) \gtrsim \Theta^*_{E,\beta}(x) \cdot H \cdot w(T).$$

We call this lemma the *Besicovitch alternative*, because its proof is part of Besicovitch's original argument [1939] for his projection theorem. For a more recent presentation; see [Falconer 1986, Lemma 6.11] or [Mattila 1995, Lemma 18.7]. Neither the hypotheses nor the conclusion of Falconer's lemma are exactly the same as ours, but the reader can easily convince himself that the proof of Lemma 4.3 heavily draws inspiration from his proof.

Proof of Lemma 4.3. Let E, x, β, H be as in the statement of the lemma. Let $\varepsilon := \frac{1}{2}\Theta^*_{E,\beta}(x)$, so that there exists an $r > 0$ such that $\mathcal{H}^1(\mathcal{C}_\beta(x, r) \cap E) \geq \varepsilon r$. We set also $J := J(\beta)$.

If the alternative (A1) fails, then

$$\mathcal{H}^1\big(\{\theta \in J : \#(\mathcal{C}_\beta(x, r) \cap E \cap \ell_{x,\theta}) \geq 2\}\big) \leq \mathcal{H}^1(\{\theta \in J : \#(E \cap \ell_{x,\theta}) \geq 2\}) \leq H^{-1}.$$

Since evidently $x \in \mathcal{C}_\beta(x, r) \cap E \cap \ell_{x,\theta}$, this implies that most of the lines $\ell_{x,\theta}$ do not intersect the set $\mathcal{C}_\beta(x, r) \cap E$ outside x. Consequently, $\mathcal{C}_\beta(x, r) \cap E$ is contained in a union of narrow cones $\mathcal{C}_1, \mathcal{C}_2, \ldots$ which are centered around certain lines ℓ_{x,θ_j} with $\theta_j \in J$, and whose opening angles $\beta_1, \beta_2, \ldots$ satisfy $\sum \beta_j \leq 2H^{-1}$. We may arrange that the cones have the form

$$\mathcal{C}_j := \mathcal{C}(I_j) := \bigcup\{\ell_{x,\theta} : \theta \in I_j\},$$

where $I_j \subset J$ is a dyadic interval, $|I_j| = \beta_j$, and $\theta_j \in J$ is the midpoint of I_j. We may also assume that the dyadic intervals I_j are disjoint, so the sets $\mathcal{C}_j \setminus \{x\}$ are disjoint.

To use these cones to arrive at alternative (A2), recall that $\mathcal{H}^1(\mathcal{C}_\beta(x, r) \cap E) \geq \varepsilon r$, where $\varepsilon = \frac{1}{2}\Theta^*_{E,\beta}(x)$. Now, we throw away cones which are not *heavy*: we call a cone *heavy* if it satisfies

$$\mathcal{H}^1(\mathcal{C}_j \cap B(x, r) \cap E) \geq \tfrac{1}{4} \cdot \varepsilon H |I_j| \cdot r. \tag{4.4}$$

The total length of $\mathcal{C}_\beta(x, r) \cap E$ contained in the nonheavy cones is bounded from above by

$$\tfrac{1}{4}\varepsilon H r \sum_{j \in \mathbb{N}} |I_j| \leq \tfrac{1}{2}\varepsilon r \leq \tfrac{1}{2}\mathcal{H}^1(\mathcal{C}_\beta(x, r) \cap E),$$

so at least half of the length in $\mathcal{C}_\beta(x, r) \cap E$ is contained in the union of the heavy cones. In the sequel, we assume that all the cones $\mathcal{C}_j$ are heavy.

Next, we would like to prove that $\sum \beta_j = \sum |I_j| \gtrsim H^{-1}$. This would be easy if the heavy cones also satisfied an upper bound roughly matching the lower bound in (4.4). If we knew this, then we could estimate

$$\sum_{j\in\mathbb{N}} |I_j| \gtrsim (\varepsilon H r)^{-1} \sum_{j\in\mathbb{N}} \mathcal{H}^1(\mathcal{C}_j \cap B(x,r)\cap E) \gtrsim H^{-1}. \tag{4.5}$$

This desired upper bound in (4.4) need not be true to begin with, but can be easily arranged. Fix a heavy cone $\mathcal{C}(I_j)$, and perform the following stopping time argument: the dyadic interval I_j is successively replaced by its parent $\hat{I}_j$ until either the upper bound

$$\mathcal{H}^1(\mathcal{C}(\hat{I}_j)\cap B(x,r)\cap E) \le \varepsilon H|\hat{I}_j|\cdot r \tag{4.6}$$

holds, or then $\hat{I}_j = J$. This procedure gives rise to a new collection of cones $\mathcal{C}(\hat{I}_j)$ which are evidently still heavy, and whose union covers the union of the initial heavy cones. Since the intervals $\hat{I}_j$ are dyadic, we may arrange that the new heavy cones are disjoint outside $\{x\}$ without violating the previous two properties.

At this point, either $\hat{I}_j = J$ for some index j, in which case (4.5) is trivially true (using $|J| \sim 1$), or then the upper bound (4.6) holds for all the heavy cones. In this case the lower bound (4.5) holds by the very calculation shown in (4.5).

We are now fully equipped to establish alternative (A2). Consider a line $\ell_{x,\theta}$ contained in the union of the heavy cones. According to (4.5), the set of angles $\theta \in J$ of such lines has length $\gtrsim H^{-1}$. This set of angles is the set $J_x \subset J$ whose existence is claimed in (A2). It remains to associate the tube $T_{x,\theta}$ to each line $\ell_{x,\theta}$ with $\theta \in J_x$. Let $\mathcal{C}(I_j) = \mathcal{C}_j \supset \ell_{x,\theta}$ be the (unique) heavy cone containing $\ell_{x,\theta}$. The opening angle of $\mathcal{C}_j$ is $\beta_j = |I_j| \in (0,|J|]$, and it follows by elementary geometry that

$$\mathcal{C}_j \cap B(x,r) \subset \ell_{x,\theta}(2\beta_j r) =: T_{x,\theta}.$$

Finally,

$$\mathcal{H}^1(E\cap T_{x,\theta}) \ge \mathcal{H}^1(\mathcal{C}_j \cap B(x,r)\cap E) \gtrsim \varepsilon H \beta_j \cdot r \sim \varepsilon H \cdot w(T),$$

as claimed in alternative (A2). □

Proof of Lemma 4.1. Recall that E is a union of finitely many disjoint α-Lipschitz minigraphs $\gamma \in \mathcal{E}$, all defined over the x-axis. The main geometric observation is the following: every minigraph in $\mathcal{E}$ is an α^{-1}-Lipschitz graph over every line $L_\theta := \operatorname{span}(\cos\theta, \sin\theta) = \ell_{0,\theta}^{\perp}$ with $\theta \in J(\alpha')$ (recall that $\alpha' = \boldsymbol{C}_{\mathrm{lip}}\alpha/2$). This is simply because the minigraphs in $\mathcal{E}$ are α-Lipschitz graphs over the x-axis, but for all $\theta \in J(\alpha')$, the lines L_θ form an angle $\gtrsim \alpha$ with the y-axis. See Figure 1. Thus, E is a union of finitely many α^{-1}-Lipschitz graphs over L_θ, for every $\theta \in J(\alpha')$. This places us in a position to use the coarea formula (2.8): for every $\theta \in J(\alpha')$ and every $\mathcal{H}^1$ measurable subset $E' \subset E$ we have

$$\int_{\pi_\theta(E')} \#(E' \cap \pi_\theta^{-1}\{t\})\,dt \gtrsim \alpha \mathcal{H}^1(E'). \tag{4.7}$$

Let

$$R = \{x \in E : \Theta^*_{E,\alpha'}(x) \ge \varepsilon\}.$$

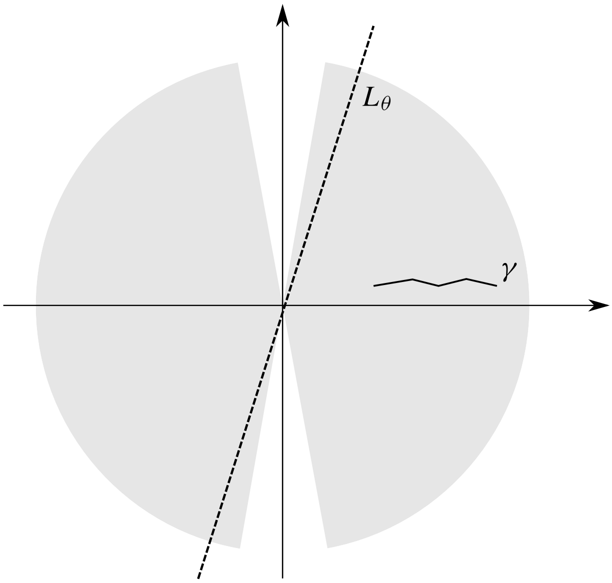


Figure 1. Every minigraph $\gamma \in \mathcal{E}$ is an α^{-1}-Lipschitz graph over every line L_θ with $\theta \in J(\alpha')$.

Fix $H \geq 1$. (We will eventually choose $H \sim 1/(\alpha\varepsilon)$; see (4.16) below.) By Lemma 4.3 (with $\beta = \alpha'$), we can write $R = R_1 \cup R_2$, where alternative (A1) holds on R_1 and (A2) holds on R_2. To prove (4.2), it suffices to show

$$\mathcal{H}^1(R_i) \lesssim \frac{\delta}{\varepsilon\alpha^2} \quad \text{for } i = 1, 2. \tag{4.8}$$

We first consider R_1. Recall the sets $I_x \subset J(\alpha')$ defined in (A1). Since E is a union of finitely many compact Lipschitz graphs, there are no measurability issues, and we may freely use Fubini's theorem:

$$H^{-1}\mathcal{H}^1(R_1) \leq \int_{R_1} \mathcal{H}^1(I_x)\, d\mathcal{H}^1(x) = \int_{J(\alpha')} \mathcal{H}^1(\{x \in R_1 : \theta \in I_x\})\, d\theta. \tag{4.9}$$

For $\theta \in J(\alpha')$ fixed, abbreviate $R'_\theta := \{x \in R_1 : \theta \in I_x\}$. Write also

$$E'_\theta := \bigcup_{t \in \pi_\theta(R'_\theta)} (E \cap \pi_\theta^{-1}\{t\}),$$

so certainly $R'_\theta \subset E'_\theta$. Note that if $t \in \pi_\theta(E'_\theta)$, then $t = \pi_\theta(x)$ for some $x \in R'_\theta$. Thus $\theta \in I_x$ by definition, so

$$\#(E'_\theta \cap \pi_\theta^{-1}\{t\}) = \#(E \cap \ell_{x,\theta}) \geq 2.$$

Therefore

$$\#(E'_\theta \cap \pi_\theta^{-1}\{t\}) - 1 \sim \#(E'_\theta \cap \pi_\theta^{-1}\{t\}), \quad t \in \pi_\theta(E'_\theta). \tag{4.10}$$

We may now deduce from (4.7) applied to $E' := E'_\theta$, and (4.10), that

$$\int_{\pi_\theta(E'_\theta)} (\#(E'_\theta \cap \pi_\theta^{-1}\{t\}) - 1)\, dt \sim \int_{\pi_\theta(E'_\theta)} \#(E'_\theta \cap \pi_\theta^{-1}\{t\})\, dt \gtrsim \alpha\mathcal{H}^1(E'_\theta) \geq \alpha\mathcal{H}^1(R'_\theta),$$

and finally

$$\int_{\mathcal{L}(E)} (\#(E \cap \ell) - 1)\, d\eta(\ell) \geq \int_{J(\alpha')} \int \#(E'_\theta \cap \pi_\theta^{-1}\{t\}) - 1\, dt\, d\theta \overset{(4.9)}{\geq} \alpha H^{-1} \mathcal{H}^1(R_1).$$

By (3.4) the left-hand side is bounded from above by δ, so

$$\mathcal{H}^1(R_1) \lesssim \frac{\delta H}{\alpha}. \tag{4.11}$$

Recalling that we promised to choose $H \sim 1/(\alpha\varepsilon)$ in the end, the bound above implies (4.8) for R_1.

Next, we tackle R_2. This time we define $R'_\theta := \{x \in R_2 : \theta \in J_x\} \subset E$, and we deduce exactly as in (4.9) that

$$H^{-1}\mathcal{H}^1(R_2) \lesssim \int_{J(\alpha')} \mathcal{H}^1(R'_\theta)\, d\theta. \tag{4.12}$$

Fix $\theta \in J(\alpha')$ with $R'_\theta \neq \varnothing$. For each $x \in R'_\theta$, by definition, there exists a tube $T = T_{x,\theta}$ centered around $\ell_{x,\theta}$ with the property

$$\mathcal{H}^1(E \cap T) \gtrsim \varepsilon H \cdot w(T). \tag{4.13}$$

The tubes $\{T_{x,\theta} : x \in R'_\theta\}$ may overlap, but they are all parallel. By the Besicovitch covering theorem (e.g., [Mattila 1995, Theorem 2.7]) applied to the projections $\pi_\theta(T_{x,\theta}) \subset \mathbb{R}$, there exists a countable subcollection $\mathcal{T}_\theta \subset \{T_{x,\theta} : x \in R'_\theta\}$, with the properties

$$R'_\theta \subset \bigcup_{x \in R'_\theta} T_{x,\theta} \subset \bigcup_{T \in \mathcal{T}_\theta} T \quad \text{and} \quad \sum_{T \in \mathcal{T}_\theta} \mathbf{1}_T \lesssim 1. \tag{4.14}$$

Fix $T \in \mathcal{T}_\theta$, and let $\mathrm{Bad}(E \cap T) \subset E \cap T$ consist of those points $x \in E \cap T$ with $\#(\ell_{x,\theta} \cap E) = 1$. We apply the coarea formula (2.8) to the set $A := \mathrm{Bad}(E \cap T) \subset E$. Recalling that for every $\theta \in J(\alpha')$ the set E is a union of finitely many α^{-1}-Lipschitz graphs over L_θ (see the remark above (4.7)) we get that

$$\mathcal{H}^1(\mathrm{Bad}(E \cap T)) \lesssim \frac{1}{\alpha} \int_{\pi_\theta(T)} 1\, dt = \frac{w(T)}{\alpha}. \tag{4.15}$$

Now, for a suitable choice $H \sim 1/(\alpha\varepsilon)$, a combination of (4.13) and (4.15) shows that

$$\mathcal{H}^1((E \cap T) \setminus \mathrm{Bad}(E \cap T)) \geq \tfrac{1}{2}\mathcal{H}^1(E \cap T). \tag{4.16}$$

At this point, we simplify notation by setting

$$E_\theta := \bigcup_{T \in \mathcal{T}_\theta} (E \cap T) \setminus \mathrm{Bad}(E \cap T) \subset E.$$

By the definition of the sets $\mathrm{Bad}(E \cap T)$, if $x \in E_\theta$, then $\#(E \cap \ell_{x,\theta}) \geq 2$, and therefore

$$\#(E \cap \pi_\theta^{-1}\{t\}) - 1 \sim \#(E \cap \pi_\theta^{-1}\{t\}) \geq \#(E_\theta \cap \pi_\theta^{-1}\{t\}), \quad t \in \pi_\theta(E_\theta). \tag{4.17}$$

It follows that

$$\begin{aligned}\int_{\mathcal{L}(E)} \#(E\cap\ell)-1\,d\eta(\ell) &\geq \int_{J(\alpha')}\int \#(E\cap\pi_\theta^{-1}\{t\})-1\,dt\,d\theta\\ &\overset{(4.17)}{\gtrsim} \int_{J(\alpha')}\int_{\pi_\theta(E_\theta)} \#(E_\theta\cap\pi_\theta^{-1}\{t\})\,dt\,d\theta\\ &\overset{(4.14)}{\gtrsim} \int_{J(\alpha')}\sum_{T\in\mathcal{T}_\theta}\int_{\pi_\theta(E_\theta\cap T)} \#(E_\theta\cap\pi_\theta^{-1}\{t\})\,dt\,d\theta\\ &\overset{(4.7)}{\gtrsim} \alpha\int_{J(\alpha')}\sum_{T\in\mathcal{T}_\theta}\mathcal{H}^1(E_\theta\cap T)\,d\theta\\ &\overset{(4.16)}{\geq} \frac{\alpha}{2}\int_{J(\alpha')}\sum_{T\in\mathcal{T}_\theta}\mathcal{H}^1(E\cap T)\,d\theta\\ &\overset{(4.14)}{\geq} \alpha\int_{J(\alpha')}\mathcal{H}^1(R'_\theta)\,d\theta\\ &\overset{(4.12)}{\geq} \frac{\alpha}{H}\cdot\mathcal{H}^1(R_2).\end{aligned}$$

Recalling once again from (3.4) that the left-hand side above is $\leq\delta$, we deduce that

$$\mathcal{H}^1(R_2)\lesssim\frac{\delta H}{\alpha}\sim\frac{\delta}{\varepsilon\alpha^2},$$

which is (4.8) for R_2. The proof of Lemma 4.1 is complete. □

Next, repeating the classical *two cones* argument of Besicovitch (e.g., [Mattila 1995, Lemma 15.14]), we show that we can pack most of points of low conical density into a single Lipschitz graph:

Lemma 4.18 (most low conical density points fit into a Lipschitz graph). *Let $E\subset B(1)\subset\mathbb{R}^2$ and let $\varepsilon\in(0,1)$, $\beta\in\left(0,\frac{1}{2}\right)$. Then, there exists a 2β-Lipschitz graph $\Gamma\subset\mathbb{R}^2$ over the x-axis such that*

$$\mathcal{H}^1(\{x\in E:\Theta^*_{E,\beta}(x)\leq\varepsilon\}\setminus\Gamma)\lesssim\varepsilon/\beta.$$

Proof. Let $G=\{x\in E:\Theta^*_{E,\beta}(x)\leq\varepsilon\}$. Our task is to find a subset $\Gamma\subset G$ with $\mathcal{H}^1(G\setminus\Gamma)\lesssim\varepsilon/\beta$ and the property $\mathcal{C}_{2\beta}(x)\cap\Gamma=\{x\}$ for all $x\in\Gamma$. Then Γ extends to a 2β-Lipschitz graph, as desired.

Let B be the set of points $x\in G$ with the "bad" property that there exists a point $y\in G\cap\mathcal{C}_{2\beta}(x)$ with $y\neq x$. The goal is to show that $\mathcal{H}^1(B)\lesssim\varepsilon/\beta$. For each $x\in B$, let $r(x)=\sup\{|x-y|:y\in G\cap\mathcal{C}_{2\beta}(x)\}$, so

$$B\cap\mathcal{C}_{2\beta}(x)\subset B(x,r(x)),\qquad x\in B. \tag{4.19}$$

See Figure 2 for an illustration.

Let T_x be the tube around the vertical line passing through x with $w(T_x):=\frac{1}{10}\beta r(x)$. Then

$$T_x\setminus B\left(x,\tfrac{1}{2}\beta r(x)\right)\subset\mathcal{C}_1(x)\subset\mathcal{C}_{2\beta}(x)\subset\mathcal{C}_\beta(x). \tag{4.20}$$

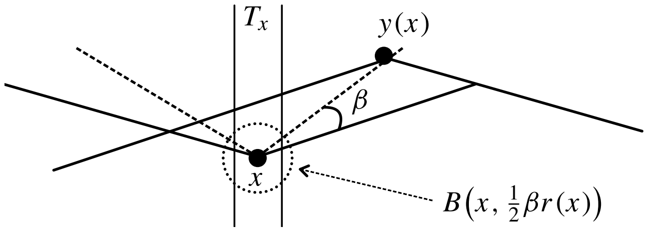


Figure 2. Containing the tube T_x in the union of the cones $\mathcal{C}_\beta(x)$ and $\mathcal{C}_\beta(y(x))$. The dotted cone illustrates $\mathcal{C}_{2\beta}(x) \ni y(x)$.

(Recall that $2\beta \leq 1$.) In particular, (4.20) implies $T_x \setminus B(x, r(x)) \subset \mathcal{C}_{2\beta}(x)$. Using this, we observe that

$$\begin{aligned} B \cap T_x &\subset B(x, r(x)) \cup [(B \cap T_x) \setminus B(x, r(x))] \\ &= B(x, r(x)) \cup [B \cap (T_x \setminus B(x, r(x)))] \subset B(x, r(x)) \cup [B \cap \mathcal{C}_{2\beta}(x)] \overset{(4.19)}{\subset} B(x, r(x)). \end{aligned} \tag{4.21}$$

Choose a point $y(x) \in G \cap \mathcal{C}_{2\beta}(x)$ such that $|x - y(x)| \geq \frac{9}{10} r(x)$. A slightly more delicate geometric fact is that

$$T_x \subset \mathcal{C}_\beta(x) \cup \mathcal{C}_\beta(y(x)).$$

This is an exercise in elementary geometry; see Figure 2 (or the proof in [Mattila 1995, Lemma 15.14] for a more formal argument): the disc $B\big(x, \frac{1}{2}\beta r(x)\big)$, and in particular the intersection $T_x \cap B\big(x, \frac{1}{2}\beta r(x)\big)$, is contained in the cone $\mathcal{C}_\beta(y(x))$, whereas the rest of T_x is contained in $\mathcal{C}_\beta(x)$, as already noted in (4.20). Consequently, using (4.21), the trivial inclusion $B(x, r(x)) \subset B(y(x), 2r(x))$, and $x, y(x) \in G$, we have

$$\mathcal{H}^1(B \cap T_x) \leq \mathcal{H}^1(\mathcal{C}_\beta(y(x), 2r(x)) \cap E) + \mathcal{H}^1(\mathcal{C}_\beta(x, r(x)) \cap E) \leq 2\varepsilon r(x) + \varepsilon r(x) \leq 30(\varepsilon/\beta) \cdot w(T_x).$$

We have now shown that every point $x \in B$ is contained on the central line of a vertical tube T_x satisfying the estimate above. By the Besicovitch covering theorem, as in the proof of Lemma 4.1, we may then find a countable, boundedly overlapping subfamily $\mathcal{T}$ of these tubes which still cover B. All the tubes intersect $B(1) \supset B$, so $\sum_{T \in \mathcal{T}} w(T) \lesssim 1$. It follows that

$$\mathcal{H}^1(B) \leq \sum_{T \in \mathcal{T}} \mathcal{H}^1(B \cap T) \leq \frac{30\varepsilon}{\beta} \sum_{T \in \mathcal{T}} w(T) \lesssim \frac{\varepsilon}{\beta}.$$

This completes the proof of Lemma 4.18. □

We are then ready to prove Proposition 3.3:

Proof of Proposition 3.3. Fix $\epsilon > 0$ as in the statement of the proposition, and set $\alpha' = \boldsymbol{C}_{\text{lip}}\alpha/2$. Define $\epsilon_1 := \alpha\epsilon/C$ for a suitable absolute constant $C > 0$. By Lemma 4.1 applied to $\varepsilon = \epsilon_1$, we know that the set $R \subset E$ of bad points $x \in E$ with

$$\Theta^*_{E_\alpha, \alpha'}(x) \geq \epsilon_1$$

satisfies

$$\mathcal{H}^1(R) \lesssim \delta \cdot \epsilon_1^{-1} \alpha^{-2} = C\delta \cdot \epsilon^{-1} \alpha^{-3}.$$

Since $\delta \le C_0\epsilon^2\alpha^3$, taking $C_0 = C^{-2}$ gives $\mathcal{H}^1(R) \le \epsilon/2$ (assuming that $C > 0$ was large enough).

The set $G := E \setminus R$ satisfies the hypotheses of Lemma 4.18 (with $\beta = \alpha' = \boldsymbol{C}_{\text{lip}}\alpha/2$ and $\varepsilon = \epsilon_1$), so there exists a $\boldsymbol{C}_{\text{lip}}\alpha$-Lipschitz graph $\Gamma \subset \mathbb{R}$ over the x-axis such that $\mathcal{H}^1(G \setminus \Gamma) \lesssim \epsilon_1/\alpha = \epsilon/C$. If the constant $C > 0$ was chosen large enough, we see that

$$\mathcal{H}^1(E \setminus \Gamma) \le \mathcal{H}^1(R) + \mathcal{H}^1(G \setminus \Gamma) \le \tfrac{1}{2}\epsilon + \tfrac{1}{2}\epsilon = \epsilon.$$

This concludes the proof of Proposition 3.3. □

5. Proof of Proposition 3.11

In this section we prove Proposition 3.11. Recall that we are assuming to be in Case 2; that is, E cannot be exhausted, up to measure ϵ, by a constant number of consecutive sets $F_k, F_{k+1}, \dots, F_{k+\boldsymbol{C}_{\text{sep}}}$ (recall this notation from Section 3C). More precisely, this means that

$$\mathcal{H}^1(E \setminus (F_k \cup \dots \cup F_{k+\boldsymbol{C}_{\text{sep}}})) \le \epsilon \tag{5.1}$$

fails for every k; thus we find an index pair $k, l \in \{0, \dots, M_3\}$ with $|k - l| \ge \boldsymbol{C}_{\text{sep}}$ such that

$$\mathcal{H}^1(F_k) \ge \alpha^{2\kappa} \quad \text{and} \quad \mathcal{H}^1(F_l) \ge \alpha^{2\kappa}. \tag{5.2}$$

Recall that all the minigraphs in $\mathcal{F}_k$ make an angle $\le \alpha^\kappa$ with

$$L_k := \text{span}(w_k) = \text{span}(\cos(k\pi/M_3), \sin(k\pi/M_3)),$$

and similarly all the minigraphs in $\mathcal{F}_l$ make an angle $\le \alpha^\kappa$ with $L_l = \text{span}(w_l)$.

The existence of F_k and F_l will imply a configuration such as the one depicted in Figure 3. A more precise definition is given in the lemma below.

Lemma 5.3. *If the inequalities in (5.2) hold, then there exists an absolute constant $C \sim \boldsymbol{C}_{\text{lip}}$ (the constant from Proposition 3.3) such that the following objects exist:*

(1) *affine lines ℓ_k and ℓ_l with $\angle(\ell_k, L_k) \le \alpha^\kappa$ and $\angle(\ell_l, L_l) \le \alpha^\kappa$,*

(2) *tubes $T_k' := \ell_k(C\alpha)$ and $T_k := \ell_k(\alpha^{1/2})$,*

(3) *tubes $T_l' := \ell_l(C\alpha)$ and $T_l := \ell_l(\alpha^{1/2})$,*

(4) *$\boldsymbol{C}_{\text{lip}}\alpha$-Lipschitz graphs γ_k, γ_l over the lines ℓ_k, ℓ_l, respectively such that*

$$\gamma_k \cap B(1) \subset T_k' \quad \textit{and} \quad \gamma_l \cap B(1) \subset T_l',$$

(5) *compact subsets*

$$G_k \subset (E \cap \gamma_k) \setminus T_l \subset B(1) \quad \textit{and} \quad G_l \subset (E \cap \gamma_l) \setminus T_k \subset B(1) \tag{5.4}$$

of measure $\mathcal{H}^1(G_k) \ge \alpha^3/C$ and $\mathcal{H}^1(G_l) \ge \alpha^3/C$.

Once the objects in Lemma 5.3 are found, it follows from a relatively simple geometric argument, presented below, that positively many lines intersect E twice (the lines in question are depicted in red in Figure 3):

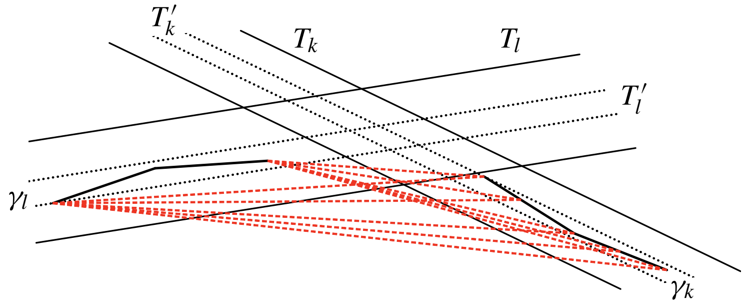


Figure 3. A configuration where positively many lines hit E twice.

Lemma 5.5. *There exists a set of lines $\mathcal{L}(G_k, G_l)$ of measure $\eta(\mathcal{L}(G_k, G_l)) \gtrsim \alpha^7$ such that $\ell \cap G_k \neq \varnothing$ and $\ell \cap G_l \neq \varnothing$ for all $\ell \in \mathcal{L}(G_k, G_l)$. In particular, since $G_k, G_l \subset E$ are disjoint,*

$$\int_{\mathcal{L}(E)} (\#(E \cap \ell) - 1)\, d\eta(\ell) \gtrsim \eta(\mathcal{L}(G_k, G_l)) \gtrsim \alpha^7. \tag{5.6}$$

Proposition 3.11 follows immediately by Lemma 5.5. We will next derive Lemma 5.5 from Lemma 5.3. (See Remark 5.10 and the Appendix for an alternative proof of Lemma 5.5.)

Proof. The key geometric observation is the following: if $\ell \subset \mathbb{R}^2$ is any line with

$$G_k \cap \ell \neq \varnothing \neq G_l \cap \ell,$$

then ℓ must make an angle $\gtrsim \alpha^{1/2}$ with both ℓ_k and ℓ_l; see Figure 3: indeed, if for example $\angle(\ell, \ell_l) \ll \alpha^{1/2}$ and $\ell \cap G_l \neq \varnothing$, then $\ell \cap B(1) \subset T_l$, and hence $\ell \cap G_k = \varnothing$ by (5.4). It follows that both ℓ_k, ℓ_l are $C\alpha^{-1/2}$-graphs over $\ell^\perp$, for any line ℓ connecting G_k and G_l. But since γ_k, γ_l were by definition $\boldsymbol{C}_{\text{lip}}\alpha$-Lipschitz graphs over ℓ_k, ℓ_l, it follows that also γ_k, γ_l are $C\alpha^{-1/2}$-Lipschitz graphs over $\ell^\perp$ (assuming that $\alpha > 0$ is small enough).

To prove the lower bound (5.6), start by fixing $x \in G_l \subset \gamma_l$, recall that $\ell_{x,\theta} := \pi_\theta^{-1}\{\pi_\theta(x)\}$, and consider the set of directions

$$\Theta(x, G_k) := \{\theta \in [0, \pi) : \ell_{x,\theta} \cap G_k \neq \varnothing\}.$$

With this notation, we claim that

$$\mathcal{H}^1(\Theta(x, G_k)) \gtrsim \alpha^{1/2} \mathcal{H}^1(G_k), \qquad x \in G_l. \tag{5.7}$$

Indeed, if $\{B(\theta_j, r_j)\}_{j \in \mathbb{N}}$ is an arbitrary cover of $\Theta(x, G_k)$, then the tubes $\ell_{x,\theta_j}(Cr_j)$ cover G_k, where $C > 0$ is an absolute constant. This is because G_k is covered by the cones $C_j := \bigcup\{\ell_{x,\theta} : \theta \in B(\theta_j, r_j)\}$ by definition, and each intersection $G_k \cap C_j \subset B(1) \cap C_j$ is further covered by a tube of the form $\ell_{x,\theta_j}(Cr_j)$. Now recall that $\gamma_k \supset G_k$ is an $\alpha^{-1/2}$-Lipschitz graph over each line $\ell_{x,\theta_j}^\perp$: this gives

$$\alpha^{-1/2} \sum_{j \in \mathbb{N}} r_j \gtrsim \sum_{j \in \mathbb{N}} \mathcal{H}^1(G_k \cap \ell_{x,\theta_j}(r_j)) \geq \mathcal{H}^1(G_k),$$

which implies (5.7).

We now infer from (5.7) and Fubini's theorem that

$$\int_0^{\pi} \mathcal{H}^1(\{x \in G_l : \theta \in \Theta(x, G_k)\})\, d\theta = \int_{G_l} \mathcal{H}^1(\Theta(x, G_k))\, d\mathcal{H}^1(x) \gtrsim \alpha^{1/2}\mathcal{H}^1(G_k)\mathcal{H}^1(G_l). \tag{5.8}$$

To proceed, write $G_l(\theta) := \{x \in G_l : \theta \in \Theta(x, G_k)\}$. We claim that

$$\mathcal{H}^1(G_l(\theta)) \neq 0 \quad \Longrightarrow \quad \mathcal{H}^1\big(\pi_\theta(G_l(\theta))\big) \gtrsim \alpha^{1/2}\mathcal{H}^1(G_l(\theta)), \quad \theta \in [0, \pi). \tag{5.9}$$

This will complete the proof of the corollary, because (5.8) then implies

$$\int_0^{\pi} \mathcal{H}^1(\pi_\theta(G_l(\theta))\, d\theta \overset{(5.8)}{\gtrsim} \alpha\mathcal{H}^1(G_k)\mathcal{H}^1(G_l) \overset{\text{Lem. 5.3}}{\gtrsim} \alpha^7,$$

and the left-hand side above is a lower bound for $\eta(\mathcal{L}(G_k, G_l))$.

Finally, let us prove (5.9). If $\mathcal{H}^1(G_l(\theta)) \neq 0$, then $\theta \in \Theta(x, \gamma_k)$ for at least one $x \in G_l$, which means that $\ell_{x,\theta} = \pi_\theta^{-1}\{\pi_\theta(x)\}$ intersects both G_k and G_l. Thus, γ_l is a $C\alpha^{-1/2}$-Lipschitz graph over the line $\ell_{x,\theta}^{\perp}$. Consequently, the relation $\mathcal{H}^1(\pi_\theta(H)) \gtrsim \alpha^{1/2}\mathcal{H}^1(H)$ holds for all $\mathcal{H}^1$ measurable subsets $H \subset \gamma_l$, in particular for $H := G_l(\theta)$. □

Remark 5.10. In fact, we have an exact expression for $\eta(\mathcal{L}(G_k, G_l))$:

$$\eta(\mathcal{L}(G_k, G_l)) = \iint_{G_k \times G_l} \frac{|\pi_{\theta(x_k,x_l)}(\tau_k(x_k))|\,|\pi_{\theta(x_k,x_l)}(\tau_l(x_l))|}{|x_k - x_l|}\, d(\mathcal{H}^1 \times \mathcal{H}^1)(x_k, x_l). \tag{5.11}$$

In (5.11), $\tau_k(x)$ denotes the unit tangent vector to γ_k at $x \in \gamma_k$, and $\tau_l(x)$ is defined similarly. For distinct $x, x' \in \mathbb{R}^2$, $\theta(x, x')$ denotes the angle θ such that $\pi_\theta(x) = \pi_\theta(x')$.

Now we show how (5.11) implies Lemma 5.5. By the key geometric observation in the first paragraph of the proof of Lemma 5.5 and the fact that $G_k, G_l \subset B(1)$, the integrand in (5.11) is $\gtrsim \alpha^{1/2}\alpha^{1/2}/1 = \alpha$. Thus, $\eta(\mathcal{L}(G_k, G_l)) \gtrsim \alpha\mathcal{H}^1(G_k)\mathcal{H}^1(G_l) \gtrsim \alpha^7$.

We state and prove a more general form of (5.11) in the Appendix.

The remainder of this section is devoted to constructing the objects listed in Lemma 5.3. This is based on the assumption (3.9), that is, $\mathcal{H}^1(F_k) \geq \alpha^{2\kappa}$ and $\mathcal{H}^1(F_l) \geq \alpha^{2\kappa}$. Recall also that F_k, F_l were the unions of the minigraphs in $\mathcal{F}_k$ and $\mathcal{F}_l$. The minigraphs in $\mathcal{F}_k$ make an angle $\leq \alpha^\kappa$ with L_k, while the minigraphs in $\mathcal{F}_l$ make an angle $\leq \alpha^\kappa$ with L_l. Furthermore, $\angle(L_k, L_l) \geq \boldsymbol{C}_{\text{sep}}\alpha^\kappa$, so the minigraphs from $\mathcal{F}_k$ and $\mathcal{F}_l$ point in quantitatively different directions. We also recall that $\mathcal{F}_k$ (respectively $\mathcal{F}_l$) can be expressed as a union of certain consecutive families $\mathcal{E}_i$:

$$\mathcal{F}_k = \mathcal{E}_s \cup \mathcal{E}_{s+1} \cup \cdots \cup \mathcal{E}_{s+m} \quad \text{and} \quad \mathcal{F}_l = \mathcal{E}_t \cup \cdots \cup \mathcal{E}_{t+m}. \tag{5.12}$$

Some of these families may be empty, but not all, according to (5.2). Of course

$$m \lesssim \alpha^{-1}, \tag{5.13}$$

since there were no more than α^{-1} of the families $\mathcal{E}_j$ altogether.

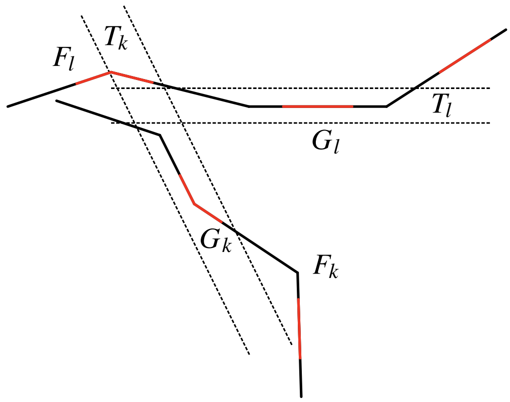


Figure 4. Finding the graphs and tubes claimed by Lemma 5.3.

5A. ***Sketch of the proof.*** We now explain the proof strategy with a picture. In Figure 4, we have depicted the sets F_k and F_l, which are roughly speaking α^κ-Lipschitz graphs over the lines L_k, L_l by Proposition 3.3 (details will follow). Both F_k and F_l are, moreover, tiled by $\lesssim \alpha^{-1}$ of the sets E_j. Most of sets E_j are (individually) contained on α-Lipschitz graphs γ_j, by another application of Proposition 3.3. The red sets shown in Figure 4 illustrate sets of the form

$$G_j = E_j \cap \gamma_j \cap B_j,$$

where B_j is some ball of radius α with the property that $\mathcal{H}^1(G_j) \sim_\alpha \mathcal{H}^1(E_j)$. Each G_j is contained in a tube T_j of width $\alpha^{1/2}$ (or even a tube of width α, which was also required in Lemma 5.3). So, picking $G_k \subset F_k$ and $G_l \subset F_l$ arbitrarily, we would satisfy all the points (1)-(5) in Lemma 5.3, except for the inclusions (5.4).

The problem is that if we pick $G_k \subset F_k$ and $G_l \subset F_l$ arbitrarily, the tube T_k associated with G_k might intersect G_l, or vice versa, violating (5.4). To satisfy (5.4), we need to pick G_k, G_l in such a way that the G_k-tube avoids G_l and the G_l-tube avoids G_k. To achieve this, we roughly choose three well-separated sets $G_1^l, G_2^l, G_3^l \subset F_l$, and two further well-separated sets $G_1^k, G_2^k \subset F_k$.

Then, we use the *transversality* of the graphs F_k, F_l to deduce the following: each G_i^k-tube can intersect at most one of the sets G_j^l, and vice versa. At this point, we may deduce from the pigeonhole principle that there must exists a pair (G_i^k, G_j^l) such that the G_i^k-tube does not intersect G_j^l, and the G_j^l-tube does not intersect G_i^k. Indeed, there are six pairs (G_i^k, G_j^l), but only five tubes. This will complete the proof.

5B. ***Proof.*** We turn to the details. First, we apply Proposition 3.3 to the sets F_k, F_l, each of which can be written as a finite union of α^κ-Lipschitz minigraphs over the lines L_k, L_l, respectively. It follows from the choice of constants $\delta = \epsilon^{70}/\boldsymbol{C}_{\mathrm{thm}}$ and $\alpha = (\epsilon/\boldsymbol{C}_{\mathrm{alp}})^{10}$ made in Section 3C that $\delta \ll \alpha^{5\kappa}$, assuming that $\boldsymbol{C}_{\mathrm{thm}}$ is chosen sufficiently small compared to the absolute constant $\boldsymbol{C}_{\mathrm{alp}}$. Writing $\alpha^{5\kappa} = (\alpha^\kappa)^3\alpha^{2\kappa}$, this means that the main hypothesis of Proposition 3.3 is valid with constants α^κ and $\frac{1}{2}\alpha^{2\kappa}$ in place of α and ϵ. It follows that there exist $\boldsymbol{C}_{\mathrm{lip}}\alpha^\kappa$-Lipschitz graphs Γ_k, Γ_l over L_k, L_l, respectively, which cover most of F_k and F_l in the sense

$$\mathcal{H}^1(F_k \setminus \Gamma_k) \le \tfrac{1}{2}\alpha^{2\kappa} \overset{(3.9)}{\le} \tfrac{1}{2}\mathcal{H}^1(F_k) \quad \text{and} \quad \mathcal{H}^1(F_l \setminus \Gamma_l) \le \tfrac{1}{2}\mathcal{H}^1(F_l).$$

We write $F_k' := F_k \cap \Gamma_k$ and $F_l' := F_l \cap \Gamma_l$. Next, recall from (5.12) that

$$F_k = E_s \cup \cdots \cup E_{s+m} \quad \text{and} \quad F_l = E_t \cup \cdots \cup E_{t+m},$$

and each E_j is a finite union of α-Lipschitz minigraphs $\mathcal{E}_j$ over a certain line (which makes an angle $\le \alpha^\kappa$ with L_k). Applying Proposition 3.3 again, for each E_j with either $j \in \{s, \dots, s+m\}$ or $j \in \{t, \dots, t+m\}$, we find Lipschitz graphs γ_j with constant $\le \boldsymbol{C}_{\text{lip}}\alpha$ and the property

$$\mathcal{H}^1(E_j \setminus \gamma_j) \lesssim \alpha^2, \quad s \le j \le s+m \text{ or } t \le j \le t+m.$$

For this application of Proposition 3.3 to be legitimate, we need $\delta \ll \alpha^3(\alpha^2)^2 = \alpha^7$, which also follows from our choice of constants recalled above, taking $\boldsymbol{C}_{\text{thm}} \gg \boldsymbol{C}_{\text{alp}}^{70}$. We write $E_j' := E_j \cap \gamma_j$. With these choices, a major part of F_k' is covered by the union of the graphs γ_j: indeed since $F_k' \subset F_k \subset (E_s \cup \cdots \cup E_{s+m})$, we have

$$\mathcal{H}^1\Big(F_k' \setminus \bigcup_{j=1}^{m} E_{s+j}'\Big) \le \sum_{j=1}^{m} \mathcal{H}^1(E_{s+j} \setminus \gamma_{s+j}) \lesssim \sum_{j=1}^{m} \alpha^2 \overset{(5.13)}{\lesssim} \alpha.$$

Since $\mathcal{H}^1(F_k') \gtrsim \mathcal{H}^1(F_k) \ge \alpha^{2\kappa}$, and $\kappa = \frac{1}{10}$, we infer that at least half of F_k' is covered by the (subsets of) α-Lipschitz graphs E_j' with $s \le j \le s+m$. The same conclusion *mutatis mutandis* holds for F_l' and the sets E_j' with $t \le j \le t+m$. We finally redefine

$$F_k := F_k' \cap \bigcup_{j=1}^{m} E_{s+j}' \quad \text{and} \quad F_l := F_l' \cap \bigcup_{j=1}^{m} E_{t+j}'.$$

This should cause no confusion, since the original sets F_k, F_l will no longer be used. We list all the properties of F_k, F_l we will need in the sequel:

- $F_k, F_l \subset E$ and $\mathcal{H}^1(F_k) \gtrsim \alpha^{2\kappa}$ and $\mathcal{H}^1(F_l) \gtrsim \alpha^{2\kappa}$ (compare with (3.9)).
- F_k is covered by the Lipschitz graph Γ_k over L_k with constant $\le \boldsymbol{C}_{\text{lip}}\alpha^\kappa$.
- F_l is covered by the Lipschitz graph Γ_l over L_l with constant $\le \boldsymbol{C}_{\text{lip}}\alpha^\kappa$.
- F_k is covered by the union of $\lesssim \alpha^{-1}$ Lipschitz graphs $\gamma_s, \dots, \gamma_{s+m}$ with constant $\le \boldsymbol{C}_{\text{lip}}\alpha$ over certain lines ℓ_{s+j} making an angle $\le \alpha^\kappa$ with L_k.
- F_l is covered by the union of $\lesssim \alpha^{-1}$ Lipschitz graphs $\gamma_t, \dots, \gamma_{t+m}$ with constant $\le \boldsymbol{C}_{\text{lip}}\alpha$ over certain lines ℓ_{t+j} making an angle $\le \alpha^\kappa$ with L_l.

We have now defined carefully the objects F_k and F_l in Figure 4. In defining the objects E_k and E_l in the same picture, there is the technical problem that the *initial* sets E_j need not be localized, as the picture suggests. This will be easily fixed by intersecting the initial sets E_j with balls. First, using that $\mathcal{H}^1(F_k) \gtrsim \alpha^{2\kappa}$, we choose two special points $x_1, x_2 \in \overline{F}_k$ with the properties

$$|x_1 - x_2| \gtrsim \alpha^{2\kappa} \quad \text{and} \quad \mathcal{H}^1(F_k \cap B(x_j, \alpha)) \ge \alpha^2 \quad \text{for } j \in \{1, 2\}. \tag{5.14}$$

This can be arranged, because the set of points $x \in F_k$ with $\mathcal{H}^1(F_k \cap B(x,\alpha)) \le \alpha^2$ has total length at most $\lesssim \alpha \ll \mathcal{H}^1(F_k)$. Thus, the admissible points for the second condition in (5.14) have total length $\ge \frac{1}{2}\mathcal{H}^1(F_k) \gtrsim \alpha^{2\kappa}$. Then, to finish the selection, it remains to pick two of these points with separation $\alpha^{2\kappa}$: this is possible because F_k lies on a Lipschitz graph with constant ≤ 1, so in particular $\mathcal{H}^1(F_k \cap B(x,r)) \lesssim r$ for all $r > 0$.

Next, we move attention from F_k to F_l. This time we pick three special points $y_1, y_2, y_3 \in F_l$ with properties similar to those in (5.14):

$$|y_i - y_j| \gtrsim \alpha^{2\kappa} \text{ for } i \neq j \qquad \text{and} \qquad \mathcal{H}^1(F_l \cap B(y_j,\alpha)) \ge \alpha^2 \quad \text{for } j \in \{1,2,3\}. \tag{5.15}$$

The details of the selection are the same as we have seen above.

Next, recall that both F_k and F_l can be written as a finite union of (subsets of) $\boldsymbol{C}_{\mathrm{lip}}\alpha$-Lipschitz graphs: the covering graphs for F_k were denoted $\gamma_s, \dots, \gamma_{s+m}$ and the covering graphs for F_l were denoted $\gamma_t, \dots, \gamma_{t+m}$, where $m \lesssim \alpha^{-1}$. Since $\mathcal{H}^1(F_k \cap B(x_1,\alpha)) \ge \alpha^2$, at least one of the graphs $\gamma_s, \dots, \gamma_{s+m}$ must have large intersection with $F_k \cap B(x_1,\alpha)$. We denote this graph by γ_1^k; then we have

$$\mathcal{H}^1(F_k \cap \gamma_1^k \cap B(x_1,\alpha)) \gtrsim \alpha^3. \tag{5.16}$$

We find similarly a graph $\gamma_2^k \in \{\gamma_s, \dots, \gamma_{s+m}\}$ such that $\mathcal{H}^1(F_k \cap \gamma_2^k \cap B(x_2,\alpha)) \gtrsim \alpha^3$. Then, we also repeat the argument for the three balls $B(y_j,\alpha)$: we find three graphs $\gamma_1^l, \gamma_2^l, \gamma_3^l \in \{\gamma_t, \dots, \gamma_{t+m}\}$ with the property

$$\mathcal{H}^1(F_l \cap B(y_j,\alpha) \cap \gamma_j^l) \gtrsim \alpha^3, \quad 1 \le j \le 3. \tag{5.17}$$

The sets

$$G_i^k := F_k \cap \gamma_i^k \cap B(x_i,\alpha), \quad i = 1,2, \qquad \text{and} \qquad G_j^l := F_l \cap \gamma_j^l \cap B(y_j,\alpha), \quad j = 1,2,3, \tag{5.18}$$

are the ones we informally discussed below Figure 4.

Next, we associate the lines and tubes (required by Lemma 5.3) to the sets G_i^k, G_j^l. We associate to each graph γ_i^k or γ_j^l an affine line ℓ_i^k or ℓ_j^l with the following properties:

- γ_i^k is a $\boldsymbol{C}_{\mathrm{lip}}\alpha$-Lipschitz graph over ℓ_i^k for $i \in \{1,2\}$.
- γ_j^l is a $\boldsymbol{C}_{\mathrm{lip}}\alpha$-Lipschitz graph over ℓ_j^l for $j \in \{1,2,3\}$.
- The lines are chosen so that

  $$G_j^k \subset \ell_i^k(C\alpha) \text{ for } i \in \{1,2\} \quad \text{and} \quad G_j^l \subset \ell_j^l(C\alpha) \quad \text{for } j \in \{1,2,3\},$$

  where $C \sim \boldsymbol{C}_{\mathrm{lip}}$.

We now define

$$(T_i^k)' := \ell_i^k(C\alpha) \quad \text{and} \quad T_i^k := \ell_i^k(\alpha^{1/2})$$

for $i \in \{1,2\}$, and similarly

$$(T_j^l)' := \ell_j^l(C\alpha) \quad \text{and} \quad T_j^l := \ell_j^l(\alpha^{1/2})$$

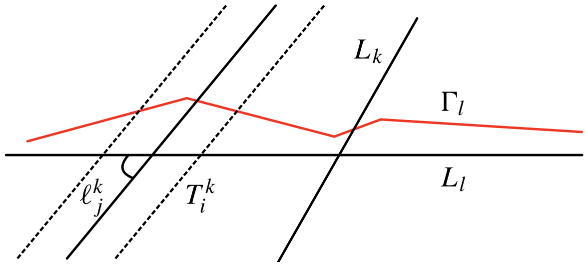


Figure 5. Transversality of T_i^k and Γ_l. The angle between ℓ_j^k and L_l is $\gtrsim C\alpha^\kappa$.

for $j \in \{1,2,3\}$. Thus, $G_i^k \subset (T_i^k)' \subset T_i^k$ and $G_j^l \subset (T_j^l)' \subset T_j^l$. Since moreover $\mathcal{H}^1(G_i^k) \gtrsim \alpha^3$ and $\mathcal{H}^1(G_j^l) \gtrsim \alpha^3$ by (5.16)–(5.17), *any* pair (G_i^k, G_j^l) (with associated lines and tubes) would now satisfy all the requirements of Lemma 5.3, except perhaps the inclusions (5.4).

We will now use the pigeonhole principle to show that at least one of the pairs (G_i^k, G_j^l) also satisfies the inclusions (5.4). The main geometric observation is

$$\operatorname{diam}(T_i^k \cap \Gamma_l) \lesssim \alpha^{1/2-\kappa} \quad \text{and} \quad \operatorname{diam}(T_j^l \cap \Gamma_k) \lesssim \alpha^{1/2-\kappa}. \tag{5.19}$$

The first inequality holds for $i \in \{1,2\}$, the second for $j \in \{1,2,3\}$. The proof of (5.19) is contained in Figure 5. Recall that T_i^k is an $\alpha^{1/2}$-tube around a certain line ℓ_i^k with $\angle(\ell_i^k, L_k) \le \alpha^\kappa$. On the other hand, $\angle(L_k, L_l) \ge \boldsymbol{C}_{\text{sep}}\alpha^\kappa$, so also $\angle(\ell_i^k, L_l) \ge (\boldsymbol{C}_{\text{sep}} - 1)\alpha^\kappa$. Finally, Γ_l is a $\boldsymbol{C}_{\text{lip}}\alpha^\kappa$-Lipschitz graph over L_l, so every tangent of Γ_l makes an angle $\gtrsim \boldsymbol{C}_{\text{sep}}\alpha^\kappa$ with ℓ_i^k, since we chose $\boldsymbol{C}_{\text{sep}}$ much larger than $\boldsymbol{C}_{\text{lip}}$ in Section 3C. Thus Γ_l is an $\alpha^{-\kappa}$-Lipschitz graph over $(\ell_j^k)^\perp$. It follows that

$$\operatorname{diam}(T_i^k \cap \Gamma_l) \le \mathcal{H}^1(T_i^k \cap \Gamma_l) \lesssim \alpha^{1/2-\kappa}.$$

Now that we have proved (5.19), recall from (5.15) the three balls $B(y_j, \alpha)$, all of which were centered at $y_j \in F_l \subset \Gamma_l$, and whose centers y_j had pairwise separation $\gtrsim \alpha^{2\kappa}$. Since $\kappa = \frac{1}{10}$, we have $\alpha^{1/2-\kappa} \ll \alpha^{2\kappa}$ for $\alpha > 0$ small enough (or in other words assuming that the constant $\boldsymbol{C}_{\text{alp}} > 0$ is chosen large enough), and therefore (5.19) implies that

$$\#\{j \in \{1,2,3\} : T_i^k \cap B(y_j, \alpha) \ne \varnothing\} \le 1, \quad i \in \{1,2\}. \tag{5.20}$$

By a similar argument,

$$\#\{i \in \{1,2\} : T_j^l \cap B(x_i, \alpha) \ne \varnothing\} \le 1, \quad j \in \{1,2,3\}. \tag{5.21}$$

We finally claim, as a consequence of (5.20)- (5.21) and the pigeonhole principle, that there exists a pair of balls $(B(x_{i_0}, \alpha), B(y_{j_0}, \alpha))$, for some $i_0 \in \{1,2\}$ and $j_0 \in \{1,2,3\}$ with the property

$$T_{i_0}^k \cap B(y_{j_0}, \alpha) = \varnothing \quad \text{and} \quad T_{j_0}^l \cap B(x_{i_0}, \alpha) = \varnothing. \tag{5.22}$$

This, by definition, yields

$$G_{i_0}^k \overset{(5.18)}{\subset} B(x_{i_0}, \alpha) \setminus T_{j_0}^l \quad \text{and} \quad G_{j_0}^l \overset{(5.18)}{\subset} B(y_{j_0}, \alpha) \setminus T_{i_0}^k,$$

which (combined with (5.18)) completes the proof of the inclusions (5.4), and Lemma 5.3.

To prove (5.22), consider the bipartite graph with 5 vertices $\{v_1, v_2\} \cup \{w_1, w_2, w_3\}$ and the following edge set:

- For $i \in \{1, 2\}$ and $j \in \{1, 2, 3\}$, the edge (v_i, w_j) is included if $T_i^k \cap B(y_j, \alpha) \neq \varnothing$.
- For $j \in \{1, 2, 3\}$ and $i \in \{1, 2\}$, the edge (w_j, v_i) is included if $T_j^l \cap B(x_i, \alpha) \neq \varnothing$.

Now, (5.20)–(5.21) can be restated as follows: for v_i fixed, there can be at most one edge (v_i, w_j), and for w_i fixed, there can be at most one edge (w_i, v_j). Thus, the edge set contains at most five edges. On the other hand, the product set $\{v_1, v_2\} \times \{w_1, w_2, w_3\}$ contains six elements, so there must be a pair $\{v_i, w_j\}$ so that neither (v_i, w_j) nor (w_j, v_i) lies in the edge set. This is equivalent to (5.22). This completes the proof of Lemma 5.3.

6. The grid example

In this section we provide an example showing that Theorem 1.1 is optimal in the sense that the assumption $\mathrm{Fav}(E) \geq \mathrm{Fav}(L) - \delta$ cannot be relaxed to $\mathrm{Fav}(E) \geq \delta$.

Proposition 6.1. *There exists an absolute constant $\delta > 0$ and a sequence of compact rectifiable sets $E_n \subset [0, 1]^2 \subset \mathbb{R}^2$ such that*

(1) $\mathcal{H}^1(E_n) = 1$,

(2) $\mathrm{Fav}(E_n) \geq \delta$,

(3) *for any $\alpha \in [2n^{-2}, 1)$ and any curve Γ with $\mathcal{H}^1(\Gamma \cap E_n) \geq \alpha$ we have $\mathcal{H}^1(\Gamma) \gtrsim \alpha n$.*

In particular, property (3) *implies that if $M \geq 1$, then for any M-Lipschitz graph Γ, $\mathcal{H}^1(\Gamma \cap E_n) \lesssim M n^{-1}$.*

We begin the construction. Fix an integer $n \geq 2$, and let $[n] := \{1, \dots, n\}$. For any $j = (k, l) \in [n]^2$ set

$$x_j = \left(\frac{k}{n+1}, \frac{l}{n+1}\right) \tag{6.2}$$

and

$$B_j = B\left(x_j, \frac{1}{2\pi n^2}\right).$$

Note that $B_j \subset [0, 1]^2$ and if $i, j \in [n]^2$, $i \neq j$, then

$$\mathrm{dist}(B_i, B_j) \geq \frac{1}{n+1} - \frac{2}{2\pi n^2} \geq \frac{1}{2n}. \tag{6.3}$$

Define $S_j = \partial B_j$, and observe that $\mathcal{H}^1(S_j) = n^{-2}$.

We define the set E_n as

$$E_n := \bigcup_{j \in [n]^2} S_j.$$

Since $\mathcal{H}^1(S_j) = n^{-2}$, we have $\mathcal{H}^1(E_n) = 1$. This verifies property (1) for E_n. It is also clear that E_n is compact and rectifiable.

Now we check property (3). We will use the following result:

Lemma 6.4 [Schul 2007, Lemma 3.7]. *Any compact connected set $\Gamma \subset \mathbb{R}^2$ with $\mathcal{H}^1(\Gamma) < \infty$ can be parametrized with $\gamma : [0,1] \to \mathbb{R}^2$ such that $\gamma([0,1]) = \Gamma$ and* $\mathrm{Lip}(\gamma) \le 32\,\mathcal{H}^1(\Gamma)$.

Lemma 6.5. *For any $\alpha \in [2n^{-2}, 1)$ and any curve Γ with $\mathcal{H}^1(\Gamma \cap E_n) \ge \alpha$ we have $\mathcal{H}^1(\Gamma) \gtrsim \alpha n$.*

Proof. Suppose that $\alpha \in [2n^{-2}, 1)$ and let Γ be a curve with $\mathcal{H}^1(\Gamma \cap E_n) \ge \alpha$. Since each circle S_j comprising E_n has length n^{-2}, we get that Γ intersects at least αn^2 different circles. Let $J_0 \subset [n]^2$ be the set of indices such that for $j \in J_0$ we have $\Gamma \cap S_j \ne \varnothing$, so that

$$N := \# J_0 \ge \alpha n^2. \tag{6.6}$$

To estimate $\mathcal{H}^1(\Gamma)$, we are going to use (6.6) together with the fact that the circles S_j are centered on a well-separated grid (6.2), (6.3). We provide the details below:

Let γ be the parametrization of the curve Γ given by Lemma 6.4. Without loss of generality, we may assume that the curve Γ begins and ends on E_n, i.e., $\gamma(0), \gamma(1) \in \Gamma \cap E_n$. For all $j \in J_0$ we choose a point $y_j \in \Gamma \cap S_j$, and let $t_j \in [0,1]$ be such that $\gamma(t_j) = y_j$ (γ might be noninjective, in which case t_j is nonunique, but in this case we pick t_j arbitrarily among the admissible options). The only constraint we make on our choice of $\{y_j\}_{j \in J_0}$ is so that $\gamma(0), \gamma(1) \in \{y_j\}_{j \in J_0}$. For convenience, we relabel the points t_j in “ascending order”: for all $i \in \{1, \dots, N\}$ we set $t_i := t_j$ for some $j \in J_0$, in such a way that $t_1 < t_2 < \cdots < t_N$. We relabel in a similar way y_j and S_j.

Recalling that the circles S_j are centered on a grid (6.2), it follows from the separation property (6.3) that, for any $i \in \{1, \dots, N\}$,

$$\frac{1}{2n} \le |y_{i+1} - y_i| = |\gamma(t_{i+1}) - \gamma(t_i)| \le \mathrm{Lip}(\gamma) \cdot |t_{i+1} - t_i| = \mathrm{Lip}(\gamma) \cdot (t_{i+1} - t_i).$$

Summing over $i \in \{1, \dots, N-1\}$ we get

$$\frac{N-1}{2n} \le \mathrm{Lip}(\gamma) \cdot (t_N - t_1) \le 32\,\mathcal{H}^1(\Gamma) \cdot (t_N - t_1).$$

Since we assumed $\gamma(0), \gamma(1) \in \{y_j\}_{j \in J_0}$, we get that $t_N = 1$ and $t_1 = 0$. Thus,

$$32\,\mathcal{H}^1(\Gamma) \ge \frac{N-1}{2n} \overset{(6.6)}{\ge} \frac{\alpha n^2 - 1}{2n} \ge \frac{\alpha n}{4}.$$

This completes the proof of the lemma. □

It remains to prove the property (2), that is, $\mathrm{Fav}(E_n) \ge \delta$. Let

$$G_n = \bigcup_{j \in [n]^2} B_j,$$

so that $E_n = \partial G_n$. Note that $\mathrm{Fav}(E_n) = \mathrm{Fav}(G_n)$. We define an auxiliary measure

$$\mu = \mu_n = \frac{1}{\mathcal{L}^2(G_n)} \mathcal{L}^2|_{G_n}.$$

Recall that the 1-energy of μ is defined as

$$I_1(\mu) = \iint \frac{1}{|x-y|}\, d\mu(x)\, d\mu(y).$$

Lemma 6.7. *We have*

$$I_1(\mu) \lesssim 1.$$

As a consequence,

$$\mathrm{Fav}(E_n) = \mathrm{Fav}(G_n) \gtrsim 1. \tag{6.8}$$

Proof. We write

$$\begin{aligned}
I_1(\mu) &= \iint \frac{1}{|x-y|}\, d\mu(x) d\mu(y) \\
&= \sum_{i,j\in[n]^2} \int_{B_i}\int_{B_j} \frac{1}{|x-y|}\, d\mu(x) d\mu(y) \\
&= \sum_{i\in[n]^2} \int_{B_i}\int_{B_i} \frac{1}{|x-y|}\, d\mu(x) d\mu(y) + \sum_{\substack{i,j\in[n]^2 \\ i\neq j}} \int_{B_i}\int_{B_j} \frac{1}{|x-y|}\, d\mu(x) d\mu(y) \\
&= A_1 + A_2.
\end{aligned}$$

To estimate A_1 we note that for any $i \in [n]^2$ and any fixed $x \in B_i$

$$\begin{aligned}
\int_{B_i} \frac{1}{|x-y|}\, d\mu(y) &\leq \sum_{k=\lfloor \log_2 n^2 \rfloor}^{\infty} \int_{B(x,2^{-k})\setminus B(x,2^{-k-1})} \frac{1}{|x-y|}\, d\mu(y) \\
&\sim \sum_{k=\lfloor \log_2 n^2 \rfloor}^{\infty} 2^k \mu(B(x,2^{-k}) \setminus B(x,2^{-k-1})) \\
&\lesssim \frac{1}{\mathcal{L}^2(G_n)} \sum_{k=\lfloor \log_2 n^2 \rfloor}^{\infty} 2^k \mathcal{L}^2(B(x,2^{-k})) \sim n^2 \sum_{k=\lfloor \log_2 n^2 \rfloor}^{\infty} 2^k \cdot 2^{-2k} \sim 1.
\end{aligned}$$

Hence,

$$A_1 = \sum_{i\in[n]^2} \int_{B_i}\int_{B_i} \frac{1}{|x-y|}\, d\mu(x)\, d\mu(y) \lesssim \sum_{i\in[n]^2} \mu(B_i) = 1.$$

We move on to estimating A_2. Let Q_j denote the square centered at x_j with sidelength $1/(n+1)$. Note that $B_j \subset Q_j$, and the squares $Q_j,\ j \in [n]^2$ are pairwise disjoint. If $x \in B_i$ and $y \in B_j$, with $i \neq j$, then $|x-y| \sim \mathrm{dist}(B_i, B_j) \sim |x-z|$ for any $z \in Q_j$. It follows that for a fixed $x \in B_i$

$$\int_{B_j} \frac{1}{|x-y|}\, d\mu(y) \sim \mathrm{dist}(B_i,B_j)^{-1}\, \mu(B_j) \sim \mathrm{dist}(B_i,B_j)^{-1}\, \mathcal{L}^2(Q_j) \sim \int_{Q_j} \frac{1}{|x-z|}\, d\mathcal{L}^2(z).$$

Summing over $j \in [n]^2 \setminus \{i\}$ yields

$$\begin{aligned}
\sum_{j\in[n]^2\setminus\{i\}} \int_{B_j} \frac{1}{|x-y|}\, d\mu(y) \sim \sum_{j\in[n]^2\setminus\{i\}} \int_{Q_j} \frac{1}{|x-z|}\, d\mathcal{L}^2(z) &\leq \int_{[-1,2]^2} \frac{1}{|x-z|}\, d\mathcal{L}^2(z) \\
&\lesssim \sum_{k=-1}^{\infty} \int_{B(x,2^{-k})\setminus B(x,2^{-k-1})} 2^k\, d\mathcal{L}^2(z) \lesssim 1.
\end{aligned}$$

Thus,

$$A_2 = \sum_{i\in[n]^2} \int_{B_i} \Big(\sum_{j\in[n]^2\setminus\{i\}} \int_{B_j} \frac{1}{|x-y|}\, d\mu(y) \Big)\, d\mu(x) \lesssim \sum_{i\in[n]^2} \mu(B_i) = 1.$$

It follows that $I_1(\mu) \lesssim 1$.

To see (6.8), we use Theorem 4.3 from [Mattila 2015] to conclude that

$$\mathrm{Fav}(E_n) = \mathrm{Fav}(G_n) \gtrsim \frac{1}{I_1(\mu)} \gtrsim 1.$$

This concludes the proof of Proposition 6.1. □

Appendix: Lines spanned by rectifiable curves

We state and prove a generalization of (5.11), which was mentioned in Remark 5.10:

Lemma A.1. *Let $\gamma_1, \gamma_2 \subset \mathbb{R}^2$ be rectifiable curves. For $\mathcal{H}^1$ almost every $x \in \gamma_i$, let $\tau_i(x)$ denote the unit tangent vector to γ_i at x. (The choice of direction is irrelevant.) Then for any $G_1 \subset \gamma_1$ and $G_2 \subset \gamma_2$, we have*

$$\int_{\mathcal{A}} \#\{(x_1, x_2) \in G_1 \times G_2 : x_1 \neq x_2 \text{ and } x_1, x_2 \in \ell\}\, d\eta(\ell) = \iint_{G_1 \times G_2} \frac{|\pi_{\theta(x_1,x_2)}(\tau_1(x_1))|\,|\pi_{\theta(x_1,x_2)}(\tau_2(x_2))|}{|x_1 - x_2|}\, d(\mathcal{H}^1 \times \mathcal{H}^1)(x_1, x_2),$$

where $\theta(x_1, x_2)$ denotes the angle θ such that $\pi_\theta(x_1) = \pi_\theta(x_2)$.

Proof. Let $\phi_i(s)$ be a parametrization of γ_i by arclength. Consider the map $\Psi : (s_1, s_2) \mapsto (\theta, t)$ defined implicitly by

$$\pi_\theta(\phi_1(s_1)) = \pi_\theta(\phi_2(s_2)) = t. \tag{A.2}$$

By the change of variables formula,

$$\begin{aligned}\int_{\mathcal{A}} \#\{(x_1, x_2) \in G_1 \times G_2 : x_1 \neq x_2 \text{ and } x_1, x_2 \in \ell\}\, d\eta(\ell) &= \int_{[0,\pi]\times\mathbb{R}} \#\{(x_1, x_2) \in G_1 \times G_2 : x_1 \neq x_2 \text{ and } x_1, x_2 \in \pi_\theta^{-1}(t)\}\, d\mathcal{H}^2(\theta, t) \\ &= \iint_{s_1 \in \phi_1^{-1}(G_1),\, s_2 \in \phi_2^{-1}(G_2)} J\Psi(s_1, s_2)\, ds_1\, ds_2,\end{aligned}$$

where $J\Psi$ denotes the Jacobian determinant of Ψ. (Note that the set $\{(s_1, s_2) : \phi_1(s_1) = \phi_2(s_2)\}$ has $\mathcal{H}^2$-measure zero.)

We now prove that

$$J\Psi(s_1, s_2) := \mathrm{abs}\begin{vmatrix} \partial_{s_1}\theta & \partial_{s_2}\theta \\ \partial_{s_1} t & \partial_{s_2} t \end{vmatrix} = \frac{|\pi_{\theta(s_1,s_2)}(\gamma_1'(s_1))|\,|\pi_{\theta(s_1,s_2)}(\gamma_2'(s_2))|}{|\gamma_1(s_1) - \gamma_2(s_2)|}. \tag{A.3}$$

Note that this would finish the proof of the lemma. To show (A.3), define $e_\theta = (\cos\theta, \sin\theta)$ and $e_\theta^\perp = d/d\theta\, e_\theta = (-\sin\theta, \cos\theta)$. By differentiating (A.2) with respect to s_1 and s_2, we obtain

$$\begin{aligned} e_\theta \cdot \phi_1'(s_1) + e_\theta^\perp \cdot \phi_1(s_1)\partial_{s_1}\theta &= e_\theta^\perp \cdot \phi_2(s_2)\partial_{s_1}\theta = \partial_{s_1} t, \\ e_\theta \cdot \phi_2'(s_2) + e_\theta^\perp \cdot \phi_2(s_2)\partial_{s_2}\theta &= e_\theta^\perp \cdot \phi_1(s_1)\partial_{s_2}\theta = \partial_{s_2} t. \end{aligned}$$

The two equalities on the left give

$$|\partial_{s_i}\theta| = \frac{|e_\theta \cdot \phi_i'(s_i)|}{|e_\theta^\perp \cdot (\phi_1(s_1) - \phi_2(s_2))|} \quad \text{for } i = 1, 2,$$

which, when combined with the two equalities on the right, give

$$J\Psi(s_1, s_2) = |\partial_{s_1}\theta|\,|\partial_{s_2}\theta|\,|e_\theta^\perp \cdot (\phi_1(s_1) - \phi_2(s_2))| = \frac{|e_\theta \cdot \phi_1'(s_1)|\,|e_\theta \cdot \phi_2'(s_2)|}{|e_\theta^\perp \cdot (\phi_1(s_1) - \phi_2(s_2))|}.$$

Finally, observe that $e_\theta \cdot (\phi_1(s_1) - \phi_2(s_2)) = 0$ by the definition of Ψ, which implies $|e_\theta^\perp \cdot (\phi_1(s_1) - \phi_2(s_2))| = |\phi_1(s_1) - \phi_2(s_2)|$. This completes the proof of (A.3). □

By using the coarea formula for rectifiable sets (e.g., [Krantz and Parks 2008, Theorem 5.4.9]), it is not hard to show that Lemma A.1 can be generalized to Lemma A.4, below. We omit the details.

Lemma A.4. *Let $E \subset \mathbb{R}^2$ be a 1-rectifiable set. For $\mathcal{H}^1$ almost every $x \in E$, let $\tau(x)$ denote the unit tangent vector to E at x. (The choice of direction is irrelevant.) Then for any $G \subset (E \times E) \setminus \{(x, x) : x \in E\}$, we have*

$$\int_{\mathcal{A}} \#\{(x_1, x_2) \in G : x_1, x_2 \in \ell\}\, d\eta(\ell) = \iint_G \frac{|\pi_{\theta(x_1,x_2)}(\tau(x_1))|\,|\pi_{\theta(x_1,x_2)}(\tau(x_2))|}{|x_1 - x_2|}\, d(\mathcal{H}^1 \times \mathcal{H}^1)(x_1, x_2), \quad \text{(A.5)}$$

where $\theta(x_1, x_2)$ denotes the angle θ such that $\pi_\theta(x_1) = \pi_\theta(x_2)$.

A version of Lemma A.4 was discovered independently by Steinerberger [2024]; see the sixth displayed equation in Section 1.2.

Acknowledgements

The research was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC-2047/1 - 390685813.

The paper was written while the authors were visiting the Hausdorff Research Institute for Mathematics in Bonn during the research trimester *Interactions between Geometric measure theory, Singular integrals, and PDE*. We would like to thank the institute and its staff for creating this opportunity for collaboration. Finally, we are grateful to the referee for a thorough reading of the manuscript, and for making numerous thoughtful suggestions.

References

[Bateman and Volberg 2010] M. Bateman and A. Volberg, "An estimate from below for the Buffon needle probability of the four-corner Cantor set", *Math. Res. Lett.* **17**:5 (2010), 959–967. MR Zbl

[Besicovitch 1939] A. S. Besicovitch, "On the fundamental geometrical properties of linearly measurable plane sets of points, III", *Math. Ann.* **116**:1 (1939), 349–357. MR Zbl

[Bond and Volberg 2010] M. Bond and A. Volberg, "Buffon needle lands in ϵ-neighborhood of a 1-dimensional Sierpinski gasket with probability at most $|\log \epsilon|^{-c}$", *C. R. Math. Acad. Sci. Paris* **348**:11-12 (2010), 653–656. MR Zbl

[Bond and Volberg 2012] M. Bond and A. Volberg, "Buffon's needle landing near Besicovitch irregular self-similar sets", *Indiana Univ. Math. J.* **61**:6 (2012), 2085–2109. MR Zbl

[Bond et al. 2014] M. Bond, I. Łaba, and A. Volberg, "Buffon's needle estimates for rational product Cantor sets", *Amer. J. Math.* **136**:2 (2014), 357–391. MR Zbl

[Bongers and Taylor 2023] R. Bongers and K. Taylor, "Transversal families of nonlinear projections and generalizations of Favard length", *Anal. PDE* **16**:1 (2023), 279–308. MR Zbl

[Chang and Tolsa 2020] A. Chang and X. Tolsa, "Analytic capacity and projections", *J. Eur. Math. Soc.* **22**:12 (2020), 4121–4159. MR Zbl

[Cladek et al. 2022] L. Cladek, B. Davey, and K. Taylor, "Upper and lower bounds on the rate of decay of the Favard curve length for the four-corner Cantor set", *Indiana Univ. Math. J.* **71**:3 (2022), 1003–1025. MR Zbl

[Davey and Taylor 2022] B. Davey and K. Taylor, "A quantification of a Besicovitch non-linear projection theorem via multiscale analysis", *J. Geom. Anal.* **32**:4 (2022), art. id. 138. MR Zbl

[Dąbrowski and Villa 2022] D. Dąbrowski and M. Villa, "Analytic capacity and dimension of sets with plenty of big projections", preprint, 2022. arXiv 2204.05804

[Falconer 1986] K. J. Falconer, *The geometry of fractal sets*, Cambridge Tracts in Math. **85**, Cambridge Univ. Press, 1986. MR Zbl

[Federer 1947] H. Federer, "The (φ, k) rectifiable subsets of n-space", *Trans. Amer. Math. Soc.* **62** (1947), 114–192. MR Zbl

[Federer 1969] H. Federer, *Geometric measure theory*, Grundlehren der Math. Wissenschaften **153**, Springer, 1969. MR Zbl

[Krantz and Parks 2008] S. G. Krantz and H. R. Parks, *Geometric integration theory*, Birkhäuser, Boston, 2008. MR Zbl

[Łaba 2015] I. Łaba, "Recent progress on Favard length estimates for planar Cantor sets", pp. 117–145 in *Operator-related function theory and time-frequency analysis* (Oslo, 2012), edited by K. Gröchenig et al., Abel Symp. **9**, Springer, 2015. MR Zbl

[Łaba and Marshall 2022] I. Łaba and C. Marshall, "Vanishing sums of roots of unity and the Favard length of self-similar product sets", *Discrete Anal.* **2022** (2022), art. id. 19. MR

[Łaba and Zhai 2010] I. Łaba and K. Zhai, "The Favard length of product Cantor sets", *Bull. Lond. Math. Soc.* **42**:6 (2010), 997–1009. MR Zbl

[Martikainen and Orponen 2018] H. Martikainen and T. Orponen, "Characterising the big pieces of Lipschitz graphs property using projections", *J. Eur. Math. Soc.* **20**:5 (2018), 1055–1073. MR Zbl

[Mattila 1995] P. Mattila, *Geometry of sets and measures in Euclidean spaces: fractals and rectifiability*, Cambridge Stud. Adv. Math. **44**, Cambridge Univ. Press, 1995. MR Zbl

[Mattila 2015] P. Mattila, *Fourier analysis and Hausdorff dimension*, Cambridge Stud. Adv. Math. **150**, Cambridge Univ. Press, 2015. MR Zbl

[Nazarov et al. 2010] F. Nazarov, Y. Peres, and A. Volberg, "The power law for the Buffon needle probability of the four-corner Cantor set", *Algebra i Analiz* **22**:1 (2010), 82–97. MR Zbl

[Orponen 2021] T. Orponen, "Plenty of big projections imply big pieces of Lipschitz graphs", *Invent. Math.* **226**:2 (2021), 653–709. MR Zbl

[Peres and Solomyak 2002] Y. Peres and B. Solomyak, "How likely is Buffon's needle to fall near a planar Cantor set?", *Pacific J. Math.* **204**:2 (2002), 473–496. MR Zbl

[Schul 2007] R. Schul, "Subsets of rectifiable curves in Hilbert space: the analyst's TSP", *J. Anal. Math.* **103** (2007), 331–375. MR Zbl

[Steinerberger 2024] S. Steinerberger, "Quadratic Crofton and sets that see themselves as little as possible", *Monatshefte für Mathematik* (2024), 1–13.

[Tao 2009] T. Tao, "A quantitative version of the Besicovitch projection theorem via multiscale analysis", *Proc. Lond. Math. Soc.* (3) **98**:3 (2009), 559–584. MR Zbl

[Vitushkin 1967] A. G. Vitushkin, "Analytic capacity of sets in problems of approximation theory", *Uspekhi Mat. Nauk* **22**:6(138) (1967), 141–199. In Russian; translated in *Russian Math. Surv.* **22** (1967), 139–200. MR Zbl

Received 7 Apr 2022. Revised 6 Aug 2022. Accepted 27 Oct 2022.

ALAN CHANG: alanchang@math.wustl.edu
Department of Mathematics, Washington University in St. Louis, MO, United States

DAMIAN DĄBROWSKI: damian.m.dabrowski@jyu.fi
Department of Mathematics and Statistics, University of Jyväskylä, Finland

TUOMAS ORPONEN: tuomas.t.orponen@jyu.fi
Department of Mathematics and Statistics, University of Jyväskylä, Finland

MICHELE VILLA: michele.villa@oulu.fi
Research Unit of Mathematical Sciences, University of Oulu, Finland

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. 4 2024