

ANALYSIS & PDE

Volume 8

No. 4

2015

Analysis & PDE

msp.org/apde

EDITORS

EDITOR-IN-CHIEF

Maciej Zworski
zworski@math.berkeley.edu
University of California
Berkeley, USA

BOARD OF EDITORS

Nicolas Burq	Université Paris-Sud 11, France nicolas.burq@math.u-psud.fr	Yuval Peres	University of California, Berkeley, USA peres@stat.berkeley.edu
Sun-Yung Alice Chang	Princeton University, USA chang@math.princeton.edu	Gilles Pisier	Texas A&M University, and Paris 6 pisier@math.tamu.edu
Michael Christ	University of California, Berkeley, USA mchrist@math.berkeley.edu	Tristan Rivière	ETH, Switzerland riviere@math.ethz.ch
Charles Fefferman	Princeton University, USA cf@math.princeton.edu	Igor Rodnianski	Princeton University, USA irod@math.princeton.edu
Ursula Hamenstaedt	Universität Bonn, Germany ursula@math.uni-bonn.de	Wilhelm Schlag	University of Chicago, USA schlag@math.uchicago.edu
Vaughan Jones	U.C. Berkeley & Vanderbilt University vaughan.f.jones@vanderbilt.edu	Sylvia Serfaty	New York University, USA serfaty@cims.nyu.edu
Herbert Koch	Universität Bonn, Germany koch@math.uni-bonn.de	Yum-Tong Siu	Harvard University, USA siu@math.harvard.edu
Izabella Laba	University of British Columbia, Canada ilaba@math.ubc.ca	Terence Tao	University of California, Los Angeles, USA tao@math.ucla.edu
Gilles Lebeau	Université de Nice Sophia Antipolis, France lebeau@unice.fr	Michael E. Taylor	Univ. of North Carolina, Chapel Hill, USA met@math.unc.edu
László Lempert	Purdue University, USA lempert@math.purdue.edu	Gunther Uhlmann	University of Washington, USA gunther@math.washington.edu
Richard B. Melrose	Massachusetts Institute of Technology, USA rbm@math.mit.edu	András Vasy	Stanford University, USA andras@math.stanford.edu
Frank Merle	Université de Cergy-Pontoise, France Frank.Merle@u-cergy.fr	Dan Virgil Voiculescu	University of California, Berkeley, USA dvv@math.berkeley.edu
William Minicozzi II	Johns Hopkins University, USA minicozz@math.jhu.edu	Steven Zelditch	Northwestern University, USA zelditch@math.northwestern.edu
Werner Müller	Universität Bonn, Germany mueller@math.uni-bonn.de		

PRODUCTION

production@msp.org
Silvio Levy, Scientific Editor

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

The subscription price for 2015 is US \$205/year for the electronic version, and \$390/year (+\$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to MSP.

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

APDE peer review and production are managed by EditFLOW[®] from MSP.

PUBLISHED BY

 **mathematical sciences publishers**
nonprofit scientific publishing

<http://msp.org/>

© 2015 Mathematical Sciences Publishers

INEQUALITY FOR BURKHOLDER’S MARTINGALE TRANSFORM

PAATA IVANISVILI

We find the sharp constant $C = C(\tau, p, \mathbb{E}G, \mathbb{E}F)$ of the inequality $\|(G^2 + \tau^2 F^2)^{1/2}\|_p \leq C \|F\|_p$, where G is the transform of a martingale F under a predictable sequence ε with absolute value 1, $1 < p < 2$, and τ is any real number.

1. Introduction	765
2. Definitions and known results	768
3. Homogeneous Monge–Ampère equation and minimal concave functions	769
4. Construction of the Bellman function	783
5. Sharp constants via foliation	793
6. Extremizers via foliation	798
Acknowledgements	806
References	806

1. Introduction

Let I be an interval of the real line \mathbb{R} , and let $|I|$ be its Lebesgue length. We write \mathcal{B} for the σ -algebra of Borel subsets of I . Let $\{F_n\}_{n=0}^\infty$ be a martingale on the probability space $(I, \mathcal{B}, dx/|I|)$ with a filtration $\{I, \emptyset\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}$. Consider any sequence of functions $\{\varepsilon_n\}_{n=1}^\infty$ such that, for each $n \geq 1$, ε_n is \mathcal{F}_{n-1} measurable and $|\varepsilon_n| \leq 1$. Let G_0 be a constant function on I ; for any $n \geq 1$, let G_n denote

$$G_0 + \sum_{k=1}^n \varepsilon_k (F_k - F_{k-1}).$$

The sequence $\{G_n\}_{n=0}^\infty$ is called the *martingale transform* of $\{F_n\}$. Obviously $\{G_n\}_{n=0}^\infty$ is a martingale with the same filtration $\{\mathcal{F}_n\}_{n=0}^\infty$. Note that, since $\{F_n\}$ and $\{G_n\}$ are martingales, we have $F_0 = \mathbb{E}F_n$ and $G_0 = \mathbb{E}G_n$ for any $n \geq 0$.

Burkholder [1984] proved that if $|G_0| \leq |F_0|$, $1 < p < \infty$, then we have the sharp estimate

$$\|G_n\|_{L^p} \leq (p^* - 1) \|F_n\|_{L^p} \quad \text{for all } n \geq 0, \tag{1}$$

where $p^* - 1 = \max\{p - 1, 1/(p - 1)\}$. Burkholder showed that it is sufficient to prove inequality (1) for the sequences of numbers $\{\varepsilon_n\}$ such that $\varepsilon_n = \pm 1$ for all $n \geq 1$. It was also noted that such an estimate

MSC2010: 42B20, 42B35, 47A30.

Keywords: martingale transform, martingale inequalities, Monge–Ampère equation, torsion, least concave function, concave envelopes, Bellman function, developable surface.

as (1) does not depend on the choice of filtration $\{\mathcal{F}_n\}$. For example, one can consider only the dyadic filtration. For more information on the estimate (1) we refer the reader to [Burkholder 1984; Choi 1992].

Vasyunin and Volberg [2010] slightly generalized the result by the Bellman function technique and Monge–Ampère equation, i.e., the estimate (1) holds if and only if

$$|G_0| \leq (p^* - 1)|F_0|. \quad (2)$$

In what follows we assume that $\{\varepsilon_n\}$ is a predictable sequence of functions such that $|\varepsilon_n| = 1$.

In [Boros et al. 2012], a perturbation of the martingale transform was investigated. Namely, under the same assumptions as (2) it was proved that, for $2 \leq p < \infty$, $\tau \in \mathbb{R}$, we have the sharp estimate

$$\|(G_n^2 + \tau^2 F_n^2)^{1/2}\|_{L^p} \leq ((p^* - 1)^2 + \tau^2)^{1/2} \|F_n\|_{L^p} \quad \text{for all } n \geq 0. \quad (3)$$

It was also claimed to be proven that the same sharp estimate holds for $1 < p < 2$, $|\tau| \leq 0.5$, and the case $1 < p < 2$, $|\tau| > 0.5$ was left open.

The inequality (3) stems from important questions concerning the L^p bounds for the perturbation of the Beurling–Ahlfors operator and hence it is of interest. We refer the reader to recent works regarding martingale inequalities and estimates of the Beurling–Ahlfors operator [Bañuelos and Janakiraman 2008; Bañuelos and Méndez-Hernández 2003; Bañuelos and Osękowski 2013; Bañuelos and Wang 1995; Boros et al. 2012] and references therein.

We should mention that Burkholder’s [1984] method and the Bellman function approach [Vasyunin and Volberg 2010; Boros et al. 2012] have similar traces in the sense that both of them reduce the required estimate to finding a certain minimal diagonally concave function with prescribed boundary conditions. However, the methods of construction of such a function are different. Unlike Burkholder’s method, in [Vasyunin and Volberg 2010; Boros et al. 2012] the construction of the function is based on the Monge–Ampère equation.

1.1. Our main results. Firstly, we should mention that the proof of (3) presented in [Boros et al. 2012] has a gap in the case $1 < p < 2$, $0 < |\tau| \leq 0.5$ (the constructed function does not satisfy a necessary concavity condition).

In the present paper we obtain the sharp L^p estimate of the perturbed martingale transform for the remaining case $1 < p < 2$ and for all $\tau \in \mathbb{R}$. Moreover, we do not require condition (2).

We define

$$u(z) \stackrel{\text{def}}{=} \tau^p (p-1)(\tau^2 + z^2)^{(2-p)/2} - \tau^2 (p-1) + (1+z)^{2-p} - z(2-p) - 1.$$

Theorem 1. *Let $1 < p < 2$, and let $\{G_n\}_{n=0}^\infty$ be a martingale transform of $\{F_n\}_{n=0}^\infty$. Set $\beta = \frac{|G_0| - |F_0|}{|G_0| + |F_0|}$. The following estimates are sharp:*

(1) *If $u(1/(p-1)) \leq 0$, then*

$$\|(\tau^2 F_n^2 + G_n^2)^{1/2}\|_{L^p} \leq \left(\tau^2 + \max \left\{ \left| \frac{G_0}{F_0} \right|, \frac{1}{p-1} \right\}^2 \right)^{\frac{1}{2}} \|F_n\|_{L^p} \quad \text{for all } n \geq 0.$$

(2) If $u(1/(p - 1)) > 0$, then

$$\|(\tau^2 F_n^2 + G_n^2)^{1/2}\|_{L^p}^p \leq C(\beta) \|F_n\|_{L^p}^p \quad \text{for all } n \geq 0,$$

where $C(\beta)$ is continuous, nondecreasing, and defined as follows:

$$C(\beta) \stackrel{\text{def}}{=} \begin{cases} (\tau^2 + |G_0|^2/|F_0|^2)^{p/2} & \text{if } \beta \geq s_0, \\ \tau^p \left(1 - \frac{2^{2-p}(1-s_0)^{p-1}}{(\tau^2 + 1)(p-1)(1-s_0) + 2(2-p)} \right)^{-1} & \text{if } \beta \leq -1 + 2/p, \\ C(\beta) & \text{if } \beta \in (-1 + 2/p, s_0), \end{cases}$$

where $s_0 \in (-1 + 2/p, 1)$ is the solution of the equation $u((1 + s_0)/(1 - s_0)) = 0$.

Explicit expression for the function $C(\beta)$ on the interval $(-1 + 2/p, s_0)$ was hard to present in a simple way. The reader can find the value of the function $C(\beta)$ in Theorem 39(ii).

Remark 2. The condition $u(1/(p - 1)) \leq 0$ holds when $|\tau| \leq 0.822$. So we also obtain Burkholder's result in the limit case when $\tau = 0$. It is worth mentioning that although the proof of the estimate (3) has a gap in [Boros et al. 2012], the claimed result in the case $1 < p < 2$, $|\tau| < 0.5$ remains true as a result of Theorem 1.

One of the important results of the current paper is that we find the function (5), and the above estimates are corollaries of this result. The argument we exploit is different from [Vasyunin and Volberg 2010; Boros et al. 2012]. Instead of writing a lot of technical computations and checking which case is valid, we present some pure geometrical facts regarding minimal concave functions with prescribed boundary conditions, and in this way we avoid computations. Moreover, we explain to the reader how we construct our Bellman function (5) based on these geometrical facts, derived in Section 3.

1.2. Plan of the paper. In Section 2 we formulate results about how to reduce the estimate (3) to finding a certain function with required properties. These results are well known and can be found in [Boros et al. 2012]. A slightly different function was investigated in [Vasyunin and Volberg 2010]; however, it possesses almost the same properties and the proof works exactly in the same way. We only mention these results and the fact that we look for a minimal continuous diagonally concave function $H(x_1, x_2, x_3)$ (see Definition 7) in the domain $\Omega = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : |x_1|^p \leq x_3\}$ with the boundary condition $H(x_1, x_2, |x_1|^p) = (x_2^2 + \tau^2 x_1^2)^{p/2}$.

Section 3 is devoted to the investigation of the minimal concave functions in two variables. It is worth mentioning that the first crucial steps in this direction for some special cases were made in [Ivanishvili et al. 2012a] (see also [Ivanishvili et al. 2012b; ≥ 2015]). In Section 3 we develop this theory for a slightly more general case. We investigate a special foliation called the *cup* and another useful object, called *force functions*.

We should note that the theory of minimal concave functions in two variables does not include the minimal diagonally concave functions in three variables. Nevertheless, this knowledge allows us to construct the candidate for H in Section 4, but with some additional technical work not mentioned in Section 3.

In Section 5 we find the good estimates for the perturbed martingale transform. In Section 6 we prove that the candidate for H constructed in Section 4 coincides with H , and as a corollary we show the sharpness of the estimates found for the perturbed martingale transform in Section 5.

In conclusion, the reader can note that the hard technical part of the current paper lies in the construction of the minimal diagonally concave function in three variables with the given boundary condition.

2. Definitions and known results

Let $\mathbb{E}F \stackrel{\text{def}}{=} \langle F \rangle_I$, where

$$\langle F \rangle_J \stackrel{\text{def}}{=} \frac{1}{|J|} \int_J F(t) dt$$

for any interval J of the real line. Let F and G be real valued integrable functions. Let $G_n = \mathbb{E}(G|\mathcal{M}_n)$ and $F_n = \mathbb{E}(F|\mathcal{M}_n)$ for $n \geq 0$, where $\{\mathcal{M}_n\}$ is a dyadic filtration (see [Boros et al. 2012]).

Definition 3. If the martingale $\{G_n\}$ satisfies $|G_{n+1} - G_n| = |F_{n+1} - F_n|$ for each $n \geq 0$, then G is called the martingale transform of F .

Recall that we are interested in the estimate

$$\|(G^2 + \tau^2 F^2)^{1/2}\|_{L^p} \leq C \|F\|_{L^p}. \quad (4)$$

We introduce the Bellman function

$$H(\mathbf{x}) \stackrel{\text{def}}{=} \sup_{F,G} \{ \mathbb{E} \mathbf{B}(\varphi(F, G)) : \mathbb{E} \varphi(F, G) = \mathbf{x}, |G_{n+1} - G_n| = |F_{n+1} - F_n|, n \geq 0 \}, \quad (5)$$

where $\varphi(x_1, x_2) = (x_1, x_2, |x_1|^p)$, $\mathbf{B}(\varphi(x_1, x_2)) = (x_2^2 + \tau^2 x_1^2)^{p/2}$ and $\mathbf{x} = (x_1, x_2, x_3)$.

Remark 4. In what follows, bold lowercase letters denote points in \mathbb{R}^3 .

Then we see that the estimate (4) can be rewritten as follows:

$$H(x_1, x_2, x_3) \leq C^p x_3.$$

We mention that the Bellman function H does not depend on the choice of the interval I . Without loss of generality, we may assume that $I = [0, 1]$.

Definition 5. Given a point $\mathbf{x} \in \mathbb{R}^3$, a pair (F, G) is said to be admissible for \mathbf{x} if G is the martingale transform of F and $\mathbb{E}(F, G, |F|^p) = \mathbf{x}$.

Proposition 6. The domain of $H(\mathbf{x})$ is $\Omega = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : |x_1|^p \leq x_3\}$, and H satisfies the boundary condition

$$H(x_1, x_2, |x_1|^p) = (x_2^2 + \tau^2 x_1^2)^{p/2}. \quad (6)$$

Definition 7. A function U is said to be diagonally concave in Ω if it is concave in both

$$\Omega \cap \{(x_1, x_2, x_3) : x_1 + x_2 = A\} \quad \text{and} \quad \Omega \cap \{(x_1, x_2, x_3) : x_1 - x_2 = A\}$$

for every constant $A \in \mathbb{R}$.

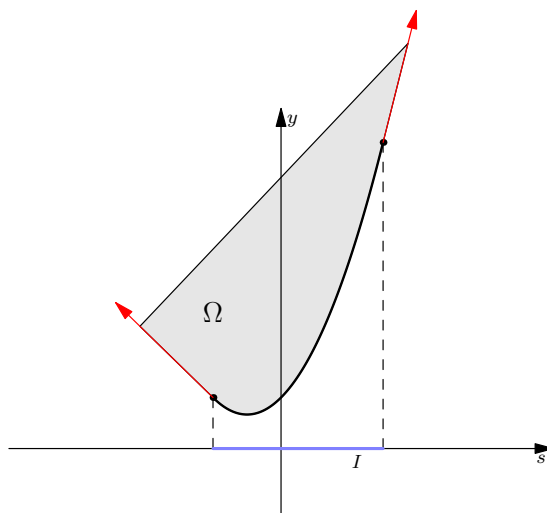


Figure 1. A domain Ω .

Proposition 8. $H(x)$ is a diagonally concave function in Ω .

Proposition 9. If U is a continuous, diagonally concave function in Ω with the boundary condition $U(x_1, x_2, |x_1|^p) \geq (x_2^2 + \tau^2 x_1^2)^{p/2}$, then $U \geq H$ in Ω .

We explain our strategy of finding the Bellman function H . We are going to find a minimal candidate \mathbf{B} that is continuous and diagonally concave, with the fixed boundary condition $\mathbf{B}|_{\partial\Omega} = (y^2 + \tau^2 x^2)^{p/2}$. We warn the reader that the symbol \mathbf{B} denoted boundary data previously, however, in Section 6 we are going to use the symbol \mathbf{B} as the candidate for the minimal diagonally concave function. Obviously, $\mathbf{B} \geq H$ by Proposition 9. We will also see that, given $x \in \Omega$ and any $\varepsilon > 0$, we can construct an admissible pair (F, G) such that $\mathbf{B}(x) < \mathbb{E}(F^2 + \tau^2 G^2)^{p/2} + \varepsilon$. This will show that $\mathbf{B} \leq H$ and hence $\mathbf{B} = H$.

In order to construct the minimal candidate \mathbf{B} , we have to elaborate a few preliminary concepts from differential geometry. We introduce the notions of *foliation* and *force* functions.

3. Homogeneous Monge–Ampère equation and minimal concave functions

3.1. Foliation. Let $g(s) \in C^3(I)$ be such that $g'' > 0$, and let Ω be a convex domain which is bounded by the curve $(s, g(s))$ and the tangents that pass through the endpoints of the curve (see Figure 1). Fix some function $f(s) \in C^3(I)$. The first question we ask is the following: how the minimal concave function $B(x_1, x_2)$ with boundary data $B(s, g(s)) = f(s)$ looks *locally* in a subdomain of Ω . In other words, take a convex hull of the curve $(s, g(s), f(s))$, $s \in I$; then the question is how the boundary of this convex hull looks.

We recall that the concavity is equivalent to the following inequalities:

$$\det(d^2 B) \geq 0, \tag{7}$$

$$B''_{x_1 x_1} + B''_{x_2 x_2} \leq 0. \tag{8}$$

The expression (7) is the Gaussian curvature of the surface $(x_1, x_2, B(x_1, x_2))$ up to a positive factor $(1 + (B'_{x_1})^2 + (B'_{x_2})^2)^2$. So, in order to minimize the function $B(x_1, x_2)$, it is reasonable to minimize the Gaussian curvature. Therefore, we will look for a surface with zero Gaussian curvature. Here the homogeneous Monge–Ampère equation arises. These surfaces are known as *developable surfaces*, that is, such a surface can be constructed by bending a plane region. The important property of such surfaces is that they consist of line segments, i.e., the function B satisfying the homogeneous Monge–Ampère equation $\det(d^2 B) = 0$ is linear along some *family of segments*. These considerations lead us to investigate such functions B . Firstly, we define a *foliation*. For any segment ℓ in the Euclidean space, by ℓ° we denote its open segment, ℓ without endpoints.

Fix any subinterval $J \subseteq I$. By $\Theta(J, g)$ we denote an arbitrary set of nontrivial segments (i.e., single points are excluded) in \mathbb{R}^2 with the following requirements:

- (1) For any $\ell \in \Theta(J, g)$ we have $\ell^\circ \in \Omega$.
- (2) For any $\ell_1, \ell_2 \in \Theta(J, g)$ we have $\ell_1 \cap \ell_2 = \emptyset$.
- (3) For any $\ell \in \Theta(J, g)$ there exists only one point $s \in J$ such that $(s, g(s))$ is one of the endpoints of the segment ℓ and, vice versa, for any point $s \in J$ there exists $\ell \in \Theta(J, g)$ such that $(s, g(s))$ is one of the endpoints of the segment ℓ .
- (4) There exists a C^1 smooth argument function $\theta(s)$.

We explain the meaning of the requirement (4). To each point $s \in J$ there corresponds only one segment $\ell \in \Theta(J, g)$ with an endpoint $(s, g(s))$. Take a nonzero vector with initial point $(s, g(s))$, parallel to the segment ℓ and having an endpoint in Ω . We define the value of $\theta(s)$ to be an argument of this vector. Since argument is defined up to addition by $2\pi k$, where $k \in \mathbb{Z}$, we take any representative from these angles. We do the same for all other points $s \in I$. In this way we get a family of functions $\theta(s)$. If there exists a $C^1(J)$ smooth function $\theta(s)$ from this family then requirement (4) is satisfied.

Remark 10. It is clear that, if $\theta(s)$ is a $C^1(J)$ smooth argument function, then, for any $k \in \mathbb{Z}$, $\theta(s) + 2\pi k$ is also a $C^1(J)$ smooth argument function. Any two $C^1(J)$ smooth argument functions differ by a constant $2\pi n$ for some $n \in \mathbb{Z}$.

This remark is a consequence of the fact that the quantity $\theta'(s)$ is well defined regardless of the choices of $\theta(s)$. Next, we define $\Omega(\Theta(J, g)) = \bigcup_{\ell \in \Theta(J, g)} \ell^\circ$. Given a point $x \in \Omega(\Theta(J, g))$, we denote by $\ell(x)$ a segment in $\Theta(J, g)$ which passes through the point x . If $x = (s, g(s))$ then, instead of $\ell((s, g(s)))$, we just write $\ell(s)$. Surely such a segment exists, and it is unique. We denote by $s(x)$ a point $s(x) \in J$ such that $(s(x), g(s(x)))$ is one of the endpoints of the segment $\ell(x)$. Moreover, in a natural way we set $s(x) = s$ if $x = (s, g(s))$. It is clear that such $s(x)$ exists, and it is unique. We introduce a function

$$K(s) = g'(s) \cos \theta(s) - \sin \theta(s), \quad s \in J. \quad (9)$$

Note that $K < 0$. This inequality becomes obvious if we rewrite

$$g'(s) \cos \theta(s) - \sin \theta(s) = \langle (1, g'), (-\sin \theta, \cos \theta) \rangle$$

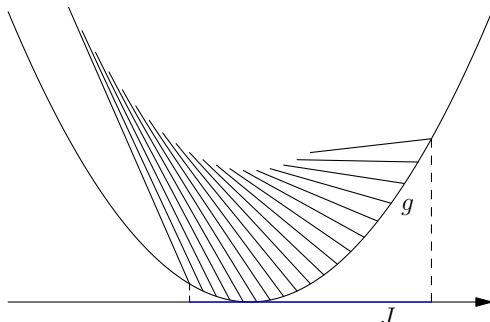


Figure 2. A foliation $\Theta(J, g)$.

and take into account requirement (1) of $\Theta(J, g)$. Note that $\langle \cdot, \cdot \rangle$ means scalar product in Euclidean space.

We need two more requirements on $\Theta(J, g)$.

(5) For any $x = (x_1, x_2) \in \Omega(\Theta(J, g))$, we have $K(s(x)) + \theta'(s(x)) \|(x_1 - s(x), x_2 - g(s(x)))\| < 0$.

(6) The function $s(x)$ is continuous in $\Omega(\Theta(J, g)) \cup \Gamma(J)$, where $\Gamma(J) = \{(s, g(s)) : s \in J\}$.

Note that if $\theta'(s) \leq 0$ (which happens in most of the cases) then requirement (5) holds. If we know the endpoints of the segments $\Theta(J, g)$, then in order to verify (5) it is enough to check it at those points $x = (x_1, x_2)$, where x is an endpoint of the segment other than $(s, g(s))$. Roughly speaking, requirement (5) means the segments of $\Theta(J, g)$ do not rotate rapidly counterclockwise.

Definition 11. A set of segments $\Theta(J, g)$ with the requirements mentioned above is called a *foliation*. The set $\Omega(\Theta(J, g))$ is called the *domain of foliation*.

A typical example of a foliation is given in Figure 2.

Lemma 12. The function $s(x)$ belongs to $C^1(\Omega(\Theta(J, g)))$. Moreover,

$$(s'_{x_1}, s'_{x_2}) = \frac{(\sin \theta, -\cos \theta)}{-K(s) - \theta' \cdot \|(x_1 - s, x_2 - g(s))\|}. \tag{10}$$

Proof. The definition of the function $s(x)$ implies that

$$-(x_1 - s) \sin \theta(s) + (x_2 - g(s)) \cos \theta(s) = 0.$$

Therefore the lemma is an immediate consequence of the implicit function theorem. □

Let $J = [s_1, s_2] \subseteq I$, and let $(s, g(s), f(s)) \in C^3(I)$ be such that $g'' > 0$ on I . Consider an arbitrary foliation $\Theta(J, g)$ with an arbitrary $C^1([s_1, s_2])$ smooth argument function $\theta(s)$. We need the following technical lemma:

Lemma 13. The solutions of the system of equations

$$t'_1(s) \cos \theta(s) + t'_2(s) \sin \theta(s) = 0, \tag{11}$$

$$t_1(s) + t_2(s)g'(s) = f'(s), \quad s \in J \tag{12}$$

are the functions

$$t_1(s) = \int_{s_1}^s \left(\frac{g''(r)}{K(r)} \sin \theta(r) \cdot t_2(r) - \frac{f''(r)}{K(r)} \sin \theta(r) \right) dr + f'(s_1) - t_2(s_1)g'(s_1),$$

$$t_2(s) = t_2(s_1) \exp\left(-\int_{s_1}^s \frac{g''(r)}{K(r)} \cos \theta(r) dr\right) + \int_{s_1}^s \frac{f''(y)}{K(y)} \exp\left(-\int_y^s \frac{g''(r)}{K(r)} \cos \theta(r) dr\right) \cos \theta(y) dy$$

for $s \in J$, where $t_2(s_1)$ is an arbitrary real number.

Proof. We differentiate (12) and combine it with (11) to obtain the system

$$\begin{pmatrix} \cos \theta & \sin \theta \\ 1 & g' \end{pmatrix} \begin{pmatrix} t_1' \\ t_2' \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & -g'' \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} + \begin{pmatrix} 0 \\ f'' \end{pmatrix}.$$

This implies that

$$\begin{pmatrix} t_1' \\ t_2' \end{pmatrix} = \frac{g''}{K} \begin{pmatrix} 0 & \sin \theta \\ 0 & -\cos \theta \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} + \frac{f''}{K} \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix}. \quad (13)$$

By solving this system of differential equations and using the fact that $t_1(s_1) + g'(s_1)t_2(s_1) = f'(s_1)$, we get the desired result. \square

Remark 14. Integration by parts allows us to rewrite the expression for $t_2(s)$ as

$$t_2(s) = \exp\left(-\int_{s_1}^s \frac{g''(r)}{K(r)} \cos \theta(r) dr\right) \left(t_2(s_1) - \frac{f''(s_1)}{g''(s_1)} \right) + \frac{f''(s)}{g''(s)} - \int_{s_1}^s \left[\frac{f''(y)}{g''(y)} \right]' \exp\left(-\int_y^s \frac{g''(r)}{K(r)} \cos \theta(r) dr\right) dy.$$

Definition 15. We say that a function B has a foliation $\Theta(J, g)$ if it is continuous on $\Omega(\Theta(J, g))$ and it is linear on each segment of $\Theta(J, g)$.

The following lemma describes how to construct a function B with a given foliation $\Theta(J, g)$ and boundary condition $B(s, g(s)) = f(s)$ such that B satisfies the homogeneous Monge–Ampère equation.

Consider the function B defined by

$$B(x) = f(s) + \langle t(s), x - (s, g(s)) \rangle, \quad x = (x_1, x_2) \in \Omega(\Theta(J, g)), \quad (14)$$

where $s = s(x)$, and $t(s) = (t_1(s), t_2(s))$ satisfies the system of equations (11), (12) with an arbitrary $t_2(s_1)$.

Lemma 16. *The function B defined by (14) satisfies the following properties:*

- (1) $B \in C^2(\Omega(\Theta(J, g))) \cap C^1(\Omega(\Theta(J, g)) \cup \Gamma)$, B has the foliation $\Theta(J, g)$ and

$$B(s, g(s)) = f(s) \quad \text{for all } s \in [s_1, s_2]. \quad (15)$$

- (2) $\nabla B(x) = t(s)$, where $s = s(x)$; moreover, B satisfies the homogeneous Monge–Ampère equation.

Proof. The fact that B has the foliation $\Theta(J, g)$ and that it satisfies the equality (15) immediately follows from the definition of the function B . We check the condition of smoothness. By Lemma 12 and Lemma 13 we have $s(x) \in C^2(\Omega(\Theta(J, g)))$ and $t_1, t_2 \in C^1(J)$, therefore the right-hand side of (14) is differentiable with respect to x . So, after differentiation of (14), we get

$$\nabla B(x) = [f'(s) - \langle t(s), (1, g'(s)) \rangle](s'_{x_1}, s'_{x_2}) + t(s) + \langle t'(s), x - (s, g(s)) \rangle (s'_{x_1}, s'_{x_2}). \tag{16}$$

Using (11) and (12) we obtain $\nabla B(x) = t(s)$. Taking the derivative with respect to x a second time we get

$$\frac{\partial^2 B}{\partial x_1^2} = t'_1(s)s'_{x_1}, \quad \frac{\partial^2 B}{\partial x_2 \partial x_1} = t'_1(s)s'_{x_2}, \quad \frac{\partial^2 B}{\partial x_1 \partial x_2} = t'_2(s)s'_{x_1}, \quad \frac{\partial^2 B}{\partial x_2^2} = t'_2(s)s'_{x_2}.$$

Using (11) we get that $t'_1(s)s'_{x_2} = t'_2(s)s'_{x_1}$, therefore $B \in C^2(\Omega(\Theta(J, g)))$. Finally, we check that B satisfies the homogeneous Monge–Ampère equation. Indeed,

$$\det(d^2 B) = \frac{\partial^2 B}{\partial x_1^2} \cdot \frac{\partial^2 B}{\partial x_2^2} - \frac{\partial^2 B}{\partial x_2 \partial x_1} \cdot \frac{\partial^2 B}{\partial x_1 \partial x_2} = t'_1(s)s'_{x_1} \cdot t'_2(s)s'_{x_1} - t'_1(s)s'_{x_2} \cdot t'_2(s)s'_{x_1} = 0. \quad \square$$

Definition 17. The function $t(s) = (t_1(s), t_2(s)) = \nabla B(x)$, $s = s(x)$, is called the *gradient function* corresponding to B .

The following lemma investigates the concavity of the function B defined by (14). Let $\|\tilde{\ell}(x)\| = \|(s(x) - x_1, g(s(x)) - x_2)\|$, where $x = (x_1, x_2) \in \Omega(\Theta(J, g))$.

Lemma 18. *The following equalities hold:*

$$\begin{aligned} \frac{\partial^2 B}{\partial x_1^2} + \frac{\partial^2 B}{\partial x_2^2} &= \frac{g''}{K(K + \theta' \|\tilde{\ell}(x)\|)} \left(-t_2 + \frac{f''}{g''} \right) \\ &= \frac{g''}{K(K + \theta' \|\tilde{\ell}(x)\|)} \left[-\exp\left(-\int_{s_1}^s \frac{g''(r)}{K(r)} \cos \theta(r) dr\right) \left(t_2(s_1) - \frac{f''(s_1)}{g''(s_1)} \right) \right. \\ &\quad \left. + \int_{s_1}^s \left[\frac{f''(y)}{g''(y)} \right]' \exp\left(-\int_y^s \frac{g''(r)}{K(r)} \cos \theta(r) dr\right) dy \right]. \end{aligned}$$

Proof. Note that

$$\frac{\partial^2 B}{\partial x_1^2} + \frac{\partial^2 B}{\partial x_2^2} = t'_1(s)s'_1 + t'_2(s)s'_2.$$

Therefore the lemma is a direct computation and application of (10), (11), (12) and Remark 14. □

Finally, we get the following important statement:

Corollary 19. *The function B is concave in $\Omega(\Theta(J, g))$ if and only if $\mathcal{F}(s) \leq 0$, where*

$$\begin{aligned} \mathcal{F}(s) &= -\exp\left(-\int_{s_1}^s \frac{g''(r)}{K(r)} \cos \theta(r) dr\right) \left(t_2(s_1) - \frac{f''(s_1)}{g''(s_1)}\right) \\ &\quad + \int_{s_1}^s \left[\frac{f''(y)}{g''(y)}\right]' \exp\left(-\int_y^s \frac{g''(r)}{K(r)} \cos \theta(r) dr\right) dy \\ &= \frac{f''(s)}{g''(s)} - t_2(s). \end{aligned} \tag{17}$$

Proof. B satisfies the homogeneous Monge–Ampère equation. Therefore, B is concave if and only if

$$\frac{\partial^2 B}{\partial x_1^2} + \frac{\partial^2 B}{\partial x_2^2} \leq 0. \tag{18}$$

Note that

$$\frac{g''}{K(K + \theta' \|\tilde{\ell}(x)\|)} > 0.$$

Hence, according to Lemma 18, the inequality (18) holds if and only if $\mathcal{F}(s) \leq 0$. □

Furthermore, the function \mathcal{F} will be called a *force function*.

Remark 20. The fact that $t_2(s) = f''/g'' - \mathcal{F}$ together with (13) implies that the force function \mathcal{F} satisfies the differential equation

$$\begin{aligned} \mathcal{F}' + \mathcal{F} \cdot \frac{\cos \theta}{K} g'' - \left[\frac{f''}{g''}\right]' &= 0, \quad s \in J, \\ \mathcal{F}(s_1) &= \frac{f''(s_1)}{g''(s_1)} - t_2(s_1). \end{aligned} \tag{19}$$

We remind the reader that, for an arbitrary smooth curve $\gamma = (s, g(s), f(s))$, the torsion has the expression

$$\frac{\det(\gamma', \gamma'', \gamma''')}{\|\gamma' \times \gamma''\|^2} = \frac{f'''g'' - g'''f''}{\|\gamma' \times \gamma''\|^2} = \frac{(g'')^2}{\|\gamma' \times \gamma''\|^2} \cdot \left[\frac{f''}{g''}\right]'$$

Corollary 21. *If $\mathcal{F}(s_1) \leq 0$ and the torsion of a curve $(s, g(s), f(s))$, $s \in J$ is negative, then the function B defined by (14) is concave.*

Proof. The corollary is an immediate consequence of (17). □

Thus, we see that the torsion of the boundary data plays a crucial role in the concavity of a surface with zero Gaussian curvature. More detailed investigations about how we choose the constant $t_2(s_1)$ will be given in Section 3.2.

Let $\Theta(J, g)$ and $\tilde{\Theta}(J, g)$ be foliations with some argument functions $\theta(s)$ and $\tilde{\theta}(s)$, respectively. Let B and \tilde{B} be the corresponding functions defined by (14), and let $\mathcal{F}, \tilde{\mathcal{F}}$ be the corresponding force functions. Note that $\mathcal{F}(s) = \tilde{\mathcal{F}}(s)$ is equivalent to the equality $t(s) = \tilde{t}(s)$, where $t(s) = (t_1(s), t_2(s))$ and $\tilde{t}(s) = (\tilde{t}_1(s), \tilde{t}_2(s))$ are the corresponding gradients of B and \tilde{B} (see (12) and Corollary 19).

Assume that the functions B and \tilde{B} are concave functions.

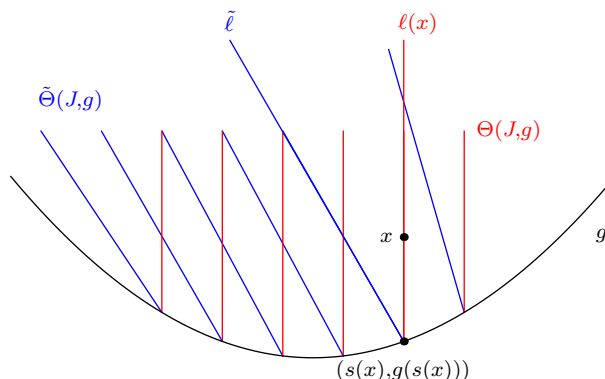


Figure 3. Foliations $\Theta(J, g)$ and $\tilde{\Theta}(J, g)$.

Lemma 22. *If $\sin(\tilde{\theta} - \theta) \geq 0$ for all $s \in J$, and $\tilde{\mathcal{F}}(s_1) = \tilde{\mathcal{F}}(s_1)$, then $\tilde{B} \leq B$ on $\Omega(\Theta(J, g)) \cap \tilde{\Omega}(\Theta(J, g))$.*

In other words, the lemma says that if, at the initial point $(s_1, g(s_1))$, gradients of the functions \tilde{B} and B coincide and the foliation $\tilde{\Theta}(J, g)$ is “to the left of” the foliation $\Theta(J, g)$ (see Figure 3), then $\tilde{B} \leq B$ provided B and \tilde{B} are concave.

Proof. Let K and \tilde{K} be the corresponding functions of B and \tilde{B} defined by (9). The condition $K, \tilde{K} < 0$ implies that the inequality $\sin(\tilde{\theta} - \theta) \geq 0$ is equivalent to the inequality

$$\frac{\cos \tilde{\theta}}{\tilde{K}} \geq \frac{\cos \theta}{K} \quad \text{for } s \in J. \tag{20}$$

Indeed, if we rewrite (20) as $K \cos \tilde{\theta} \geq \tilde{K} \cos \theta$ then this simplifies to $-\sin \theta \cos \tilde{\theta} \geq -\sin \tilde{\theta} \cos \theta$, so the result follows.

The force functions $\mathcal{F}, \tilde{\mathcal{F}}$ satisfy the differential equation (19) with the same boundary condition $\mathcal{F}(s_1) = \tilde{\mathcal{F}}(s_1)$. Then, by (20) and by comparison theorems, we get $\tilde{\mathcal{F}} \geq \mathcal{F}$ on J . This and (17) imply that $\tilde{t}_2 \leq t_2$ on J . Pick any point $x \in \Omega(\Theta(J, g)) \cap \tilde{\Omega}(\Theta(J, g))$. Then there exists a segment $\ell(x) \in \Theta(J, g)$. Let $(s(x), g(s(x)))$ be the corresponding endpoint of this segment. There exists a segment $\tilde{\ell} \in \tilde{\Theta}(J, g)$ which has $(s(x), g(s(x)))$ as an endpoint (see Figure 3).

Consider a tangent plane $L(x)$ to (x_1, x_2, \tilde{B}) at the point $(s(x), g(s(x)))$. The fact that the gradient of \tilde{B} is constant on $\tilde{\ell}$ implies that L is tangent to (x_1, x_2, \tilde{B}) on $\tilde{\ell}$. Therefore,

$$L(x) = f(s) + \langle (\tilde{t}_1(s), \tilde{t}_2(s)), (x_1 - s, x_2 - g(s)) \rangle,$$

where $x = (x_1, x_2)$ and $s = s(x)$. The concavity of \tilde{B} implies that a value of the function \tilde{B} at a point y seen from the point $(s(x), g(s(x)))$ is less than $L(y)$. In particular, $\tilde{B}(x) \leq L(x)$. Now it is enough to prove that $L(x) \leq B(x)$. By (14) we have

$$B(x) = f(s) + \langle (t_1(s), t_2(s)), (x_1 - s(x), x_2 - g(s)) \rangle.$$

Therefore, using (12), the fact that $\langle (-g', 1), (x_1 - s, x_2 - g(s)) \rangle \geq 0$ and $\tilde{t}_2 \leq t_2$, we get the desired result. □

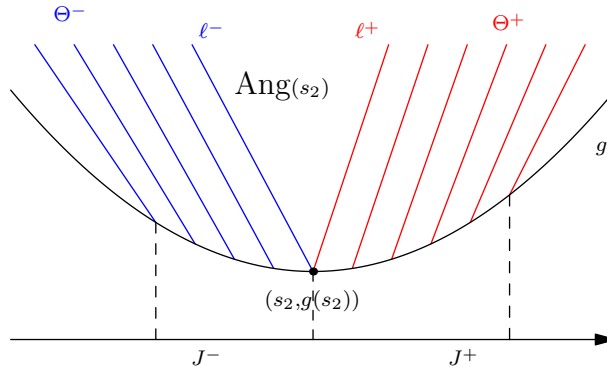


Figure 4. Gluing of B^- and B^+ .

Let $J^- = [s_1, s_2]$ and $J^+ = [s_2, s_3]$, where $J^-, J^+ \subset I$. Consider arbitrary foliations $\Theta^- = \Theta^-(J^-, g)$ and $\Theta^+ = \Theta^+(J^+, g)$ such that $\Omega(\Theta^-) \cap \Omega(\Theta^+) = \emptyset$, and let θ^- and θ^+ be the corresponding argument functions. Let B^- and B^+ be the corresponding functions defined $\ell^+(s_2)$, where $\ell^-(s_2) \in \Theta^-$ by (14), and let $t^- = (t_1^-, t_2^-)$ and $t^+ = (t_1^+, t_2^+)$ be the corresponding gradient functions. Set $\text{Ang}(s_2)$ to be a convex hull of $\ell^-(s_2)$ and $\ell^+(s_2) \in \Theta^+$ are the segments with the endpoint $(s_2, g(s_2))$ (see Figure 4). We require that $\text{Ang}(s_2) \cap \Omega(\Theta^-) = \ell^-$ and $\text{Ang}(s_2) \cap \Omega(\Theta^+) = \ell^+$.

Let $\mathcal{F}^-, \mathcal{F}^+$ be the corresponding forces, and let B_{Ang} be the function defined linearly on $\text{Ang}(s_2)$ via the values of B^- and B^+ on ℓ^-, ℓ^+ respectively.

Lemma 23. *If $t_2^-(s_2) = t_2^+(s_2)$, then the function B defined by*

$$B(x) = \begin{cases} B^-(x) & \text{if } x \in \Omega(\Theta(J^-, g)), \\ B_{\text{Ang}}(x) & \text{if } x \in \text{Ang}(s_2), \\ B^+(x) & \text{if } x \in \Omega(\Theta(J^+, g)), \end{cases}$$

belongs to the class $C^1(\Omega(\Theta^-) \cup \text{Ang}(s_2) \cup \Omega(\Theta^+) \cup \Gamma(J^- \cup J^+))$.

Proof. By (12) the condition $t_2^-(s_2) = t_2^+(s_2)$ is equivalent to the condition $t^-(s_2) = t^+(s_2)$. We recall that the gradient of B^- is constant on $\ell^-(s_2)$, and the gradient of B^+ is constant on $\ell^+(s_2)$, therefore the lemma follows immediately from the fact that $B^-(s_2, g(s_2)) = B^+(s_2, g(s_2))$. \square

Remark 24. The fact $B \in C^1$ implies that its gradient function $t(s) = \nabla B$ is well defined and is continuous. Unfortunately, it is not necessarily true that $t(s) \in C^1([s_1, s_3])$. However, it is clear that $t(s) \in C^1([s_1, s_2])$ and $t(s) \in C^1([s_2, s_3])$.

We finish this section with the following important corollary about *concave extension* of the functions with zero Gaussian curvature:

Let B^- and B^+ be defined as above (see Figure 4). Assume that $t_2^-(s_2) = t_2^+(s_2)$.

Corollary 25. *If B^- is concave in $\Omega(\Theta^-)$ and the torsion of the curve $(s, g(s), f(s))$ is nonnegative on $J^+ = [s_2, s_3]$ then the function B defined in Lemma 23 is concave in the domain $\Omega(\Theta^-) \cup \text{Ang}(s_2) \cup \Omega(\Theta^+)$.*

In other words, the corollary tells us that, if we have constructed a concave function B^- which satisfies the homogeneous Monge–Ampère equation, and we glue B^- smoothly with B^+ (which also satisfies the homogeneous Monge–Ampère equation), then the result, B , is a concave function provided that the space curve $(s, g(s), f(s))$ has nonnegative torsion on the interval J^+ .

Proof. By Corollary 19, concavity of B^- implies $\mathcal{F}^-(s_2) \leq 0$. By (17) the condition $t_2^-(s_2) = t_2^+(s_2)$ is equivalent to $\mathcal{F}^-(s_2) = \mathcal{F}^+(s_2)$. By Corollary 21 we get that B^+ is concave. Thus, concavity of B follows from Lemma 23. □

3.2. Cup. In this subsection we are going to consider a special type of foliation, which is called a *cup*. Fix an interval I and consider an arbitrary curve $(s, g(s), f(s)) \in C^3(I)$. We suppose that $g'' > 0$ on I . Let $a(s) \in C^1(J)$ be a function such that $a'(s) < 0$ on J , where $J = [s_0, s_1]$ is a subinterval of I . Assume that $a(s_0) < s_0$ and $[a(s_1), a(s_0)] \subset I$. Consider a set of open segments $\Theta_{\text{cup}}(J, g)$ consisting of those segments $\ell(s, g(s))$, $s \in J$ such that $\ell(s, g(s))$ is a segment in the plane joining the points $(s, g(s))$ and $(a(s), g(a(s)))$ (see Figure 5).

Lemma 26. *The set of segments $\Theta_{\text{cup}}(J, g)$ described above forms a foliation.*

Proof. We need to check the six requirements for a set to be the foliation. Most of them are trivial except for (4) and (5). We know the endpoints of each segment, therefore we can consider the argument function

$$\theta(s) = \pi + \arctan\left(\frac{g(s) - g(a(s))}{s - a(s)}\right).$$

Surely $\theta(s) \in C^1(J)$, so requirement (4) is satisfied. We check requirement (5). It is clear that it is enough to check this requirement for $x = (a(s), g(a(s)))$. Let $s = s(x)$; then

$$\begin{aligned} K(s) + \theta'(s)\|(a(s) - s, g(a(s)) - g(s))\| &= \frac{\langle (1, g'), (g - g(a), a - s) \rangle}{\|(g(a) - g, s - a)\|} + \frac{(g' - a'g'(a))(s - a) - (1 - a')(g - g(a))}{\|(g(a) - g, s - a)\|} \\ &= \frac{a' \cdot \langle (1, g'(a)), (g - g(a), a - s) \rangle}{\|(g(a) - g, s - a)\|}, \end{aligned}$$

which is strictly negative. □

Let $\gamma(t) = (t, g(t), f(t)) \in C^3([a_0, b_0])$ be an arbitrary curve such that $g'' > 0$ on $[a_0, b_0]$. Assume that the torsion of γ is positive on $I^- = (a_0, c)$, and it is negative on $I^+ = (c, b_0)$ for some $c \in (a_0, b_0)$.

Lemma 27. *For all P such that $0 < P < \min\{c - a_0, b_0 - c\}$, there exist $a \in I^-$, $b \in I^+$ such that $b - a = P$ and*

$$\begin{vmatrix} 1 & 1 & a - b \\ g'(a) & g'(b) & g(a) - g(b) \\ f'(a) & f'(b) & f(a) - f(b) \end{vmatrix} = 0. \tag{21}$$

Proof. Pick a number $a \in (a_0, b_0)$ such that $b = a + P \in (a_0, b_0)$. We denote

$$\mathcal{M}(a, b) = (a - b)(g'(b) - g'(a))\left(\frac{g(a) - g(b)}{a - b} - g'(a)\right).$$

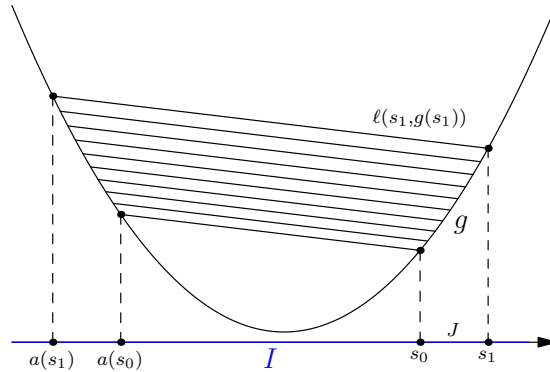


Figure 5. The foliation $\Theta_{\text{cup}}(J, g)$.

Note that the conditions $a \neq b$ and $g'' > 0$ imply $\mathcal{M}(a, b) \neq 0$. Then

$$\begin{vmatrix} 1 & 1 & a - b \\ g'(a) & g'(b) & g(a) - g(b) \\ f'(a) & f'(b) & f(a) - f(b) \end{vmatrix} = \mathcal{M}(a, b) \left[\frac{f(a) - f(b) - f'(a)(a - b)}{g(a) - g(b) - g'(a)(a - b)} - \frac{f'(b) - f'(a)}{g'(b) - g'(a)} \right].$$

Thus our equation (21) turns into

$$\frac{f(a) - f(b) - f'(a)(a - b)}{g(a) - g(b) - g'(a)(a - b)} - \frac{f'(b) - f'(a)}{g'(b) - g'(a)} = 0. \tag{22}$$

We consider the functions $V(x) = f(x) - f'(a)x$ and $U(x) = g(x) - g'(a)x$. Note that $U(a) \neq U(b)$ and $U' \neq 0$ on (a, b) . Therefore, by Cauchy's mean value theorem there exists a point $\xi = \xi(a, b) \in (a, b)$ such that

$$\frac{f(a) - f(b) - f'(a)(a - b)}{g(a) - g(b) - g'(a)(a - b)} = \frac{V(a) - V(b)}{U(a) - U(b)} = \frac{V'(\xi)}{U'(\xi)} = \frac{f'(\xi) - f'(a)}{g'(\xi) - g'(a)}.$$

Now we define

$$W_a(z) \stackrel{\text{def}}{=} \frac{f'(z) - f'(a)}{g'(z) - g'(a)}, \quad z \in (a, b).$$

So the left-hand side of (22) takes the form $W_a(\xi) - W_a(b) = 0$ for some $\xi(a, P) \in (a, b)$. We consider the curve $v(s) = (g'(s), f'(s))$, which is a graph on $[a_0, b_0]$. The fact that the torsion of the curve $\gamma(s) = (s, g(s), f(s))$ changes sign from $+$ to $-$ at the point $c \in (a_0, b_0)$ means that the curve $v(s)$ is strictly convex on the interval (a_0, c) and it is strictly concave on the interval (c, b_0) . We consider a function obtained from (22),

$$D(z) \stackrel{\text{def}}{=} \frac{f(z) - f(z + P) + f'(z)P}{g(z) - g(z + P) + g'(z)P} - \frac{f'(z + P) - f'(z)}{g'(z + P) - g'(z)}, \quad z \in [a_0, c]. \tag{23}$$

Note that $D(a_0) = W_{a_0}(\zeta) - W_{a_0}(a_0 + P)$ for some $\zeta = \zeta(a_0, P) \in (a_0, a_0 + P)$. We know that $v(s)$ is strictly convex on the interval $(a_0, a_0 + P)$. This implies that $W_{a_0}(z) - W_{a_0}(a_0 + P) < 0$ for all $z \in (a_0, a_0 + P)$. In particular, $D(a_0) < 0$. Similarly, concavity of $v(s)$ on $(c, c + P)$ implies that $D(c) > 0$. Hence, there exists $a \in (a_0, c)$ such that $D(a) = 0$. □

Let a_1 and b_1 be some solutions of (21) obtained by Lemma 27.

Lemma 28. *There exists a function $a(s) \in C^1((c, b_1]) \cap C([c, b_1])$ such that $a(b_1) = a_1$, $a(c) = c$, $a'(s) < 0$, and the pair $(a(s), s)$ solves (21) for all $s \in [c, b_1]$.*

Proof. The proof of the lemma is a consequence of the implicit function theorem. Let $a < b$, and consider the function

$$\Phi(a, b) \stackrel{\text{def}}{=} \begin{vmatrix} 1 & 1 & a - b \\ g'(a) & g'(b) & g(a) - g(b) \\ f'(a) & f'(b) & f(a) - f(b) \end{vmatrix}.$$

We are going to find the signs of the partial derivatives of $\Phi(a, b)$ at the point $(a, b) = (a_1, b_1)$. We present the calculation only for $\partial\Phi/\partial b$. The case for $\partial\Phi/\partial a$ is similar.

$$\begin{aligned} \frac{\partial\Phi(a, b)}{\partial b} &= \begin{vmatrix} 1 & 0 & a - b \\ g'(a) & g''(b) & g(a) - g(b) \\ f'(a) & f''(b) & f(a) - f(b) \end{vmatrix} \\ &= (a - b)g''(b) \left(\frac{g(a) - g(b)}{a - b} - g'(a) \right) \left[\frac{f(a) - f(b) - f'(a)(a - b)}{g(a) - g(b) - g'(a)(a - b)} - \frac{f''(b)}{g''(b)} \right]. \end{aligned}$$

Note that

$$(a - b)g''(b) \left(\frac{g(a) - g(b)}{a - b} - g'(a) \right) < 0,$$

therefore we see that the sign of $\partial\Phi/\partial b$ depends only on the sign of the expression

$$\frac{f(a) - f(b) - f'(a)(a - b)}{g(a) - g(b) - g'(a)(a - b)} - \frac{f''(b)}{g''(b)}. \tag{24}$$

We use the *cup equation* (22), and we obtain that the expression (24) at the point $(a, b) = (a_1, b_1)$ takes the form

$$\frac{f'(b) - f'(a)}{g'(b) - g'(a)} - \frac{f''(b)}{g''(b)}. \tag{25}$$

The above expression has the following geometric meaning. We consider the curve $v(s) = (g'(s), f'(s))$, and we draw a segment which connects the points $v(a)$ and $v(b)$. The above expression is the difference between the slope of the line which passes through the segment $[v(a), v(b)]$ and the slope of the tangent line of the curve $v(s)$ at the point b . In the case shown in Figure 6, this difference is positive. Recall that $v(s)$ is strictly convex on (a_1, c) and it is strictly concave on (c, b_1) . Therefore, one can easily note that this expression (25) is always positive if the segment $[v(a), v(b)]$ also intersects the curve $v(s)$ at a point ξ such that $a < \xi < b$. This always happens in our case because (22) means that the points $v(a), v(\xi), v(b)$ lie on the same line, where ξ was determined from Cauchy's mean value theorem. Thus,

$$\frac{f'(b) - f'(a)}{g'(b) - g'(a)} - \frac{f''(b)}{g''(b)} > 0. \tag{26}$$

Similarly, we can obtain that $\partial\Phi/\partial a < 0$, because this is the same as to show that

$$\frac{f'(b) - f'(a)}{g'(b) - g'(a)} - \frac{f''(a)}{g''(a)} > 0. \tag{27}$$

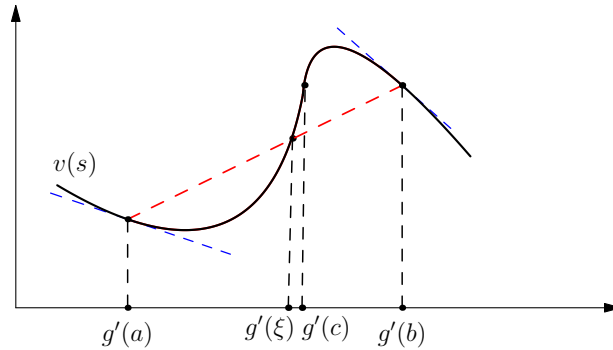


Figure 6. Graph of $v(s)$.

Thus, by the implicit function theorem there exists a C^1 function $a(s)$ in some neighborhood of b_1 such that $a'(s) = -\Phi'_b/\Phi'_a < 0$, and the pair $(a(s), s)$ solves (21).

Now we will explain that the function $a(s)$ can be defined on $(c, b_1]$ and, moreover, $\lim_{s \rightarrow c+0} a(s) = c$. Indeed, whenever $a(s) \in (a_1, c)$ and $s \in (c, b_1)$ we can use the implicit function theorem, and we can extend the function $a(s)$. It is clear that for each s we have $a(s) \in [a_1, c)$ and $s \in (c, b_1)$. Indeed, if $a(s), s \in (a_1, c]$, or $a(s), s \in [c, b_1)$, then (21) has a definite sign (see (23)). It follows that $\alpha(s) \in C^1((c, b_1])$, and the condition $a'(s) < 0$ implies $\lim_{s \rightarrow c+0} a(s) = c$. Hence $a(s) \in C([c, b_1])$. \square

It is worth mentioning that we did not use the fact that the torsion of $(s, g(s), f(s))$ changes sign from $+$ to $-$. The only thing we needed was that the torsion changes sign.

Let a_1 and b_1 be any solutions of (21) from Lemma 27, and let $a(s)$ be any function from Lemma 28. Fix an arbitrary $s_1 \in (c, b_1)$ and consider the foliation $\Theta_{\text{cup}}([s_1, b_1], g)$ constructed by $a(s)$ (see Lemma 26). Let B be the function defined by (14), where

$$t_2(s_1) = \frac{f'(s_1) - f'(a(s_1))}{g'(s_1) - g'(a(s_1))}. \tag{28}$$

Set $\Omega_{\text{cup}} = \Omega(\Theta_{\text{cup}}([s_1, b_1], g))$, and let $\bar{\Omega}_{\text{cup}}$ be the closure of Ω_{cup} .

Lemma 29. *The function B satisfies the following properties*

- (1) $B \in C^2(\Omega_{\text{cup}}) \cap C^1(\bar{\Omega}_{\text{cup}})$.
- (2) $B(a(s), g(a(s))) = f(a(s))$ for all $s \in [s_1, b_1]$.
- (3) B is a concave function in $\bar{\Omega}_{\text{cup}}$.

Proof. The first property follows from Lemma 16 and the fact that $\nabla B(x) = t(s)$ for $s = s(x)$, where $s(x)$ is a continuous function in $\bar{\Omega}_{\text{cup}}$.

We are going to check the second property. We recall (see (12)) that $t_1(s) = f'(s) - t_2(s)g'(s)$. Condition (28) implies that

$$t_1(s_1) + t_2(s_1)g'(a(s_1)) = f'(a(s_1)). \tag{29}$$

Let $B(a(s), g(a(s))) = \tilde{f}(a(s))$. After differentiation of this equality we get $t_1(s_1) + t_2(s_1)g'(a(s_1)) = \tilde{f}'(a(s_1))$. Hence, (29) implies that $f'(a(s_1)) = \tilde{f}'(a(s_1))$. It is clear that

$$\begin{aligned} t_1(s) + t_2(s)g'(s) &= f'(s), \\ t_1(s) + t_2(s)g'(a(s)) &= \tilde{f}'(a(s)), \\ t_1(s)(s - a(s)) + t_2(s)(g(s) - g(a(s))) &= f(s) - \tilde{f}(a(s)), \end{aligned}$$

which implies

$$\begin{vmatrix} 1 & 1 & s - a(s) \\ g'(s) & g'(a(s)) & g(s) - g(a(s)) \\ f'(s) & \tilde{f}'(a(s)) & f(s) - \tilde{f}(a(s)) \end{vmatrix} = 0.$$

This equality can be rewritten as follows:

$$f' \cdot \begin{vmatrix} 1 & s - a(s) \\ g'(a(s)) & g(s) - g(a(s)) \end{vmatrix} - \tilde{f}'(a) \begin{vmatrix} 1 & s - a(s) \\ g' & g(s) - g(a(s)) \end{vmatrix} + (f - \tilde{f}(a))(g'(a(s)) - g'(s)) = 0.$$

By virtue of Lemma 28 we have the same equality as above except \tilde{f} is replaced by f . We subtract one from the other:

$$[f(a(s)) - \tilde{f}(a(s))] + [f'(a(s)) - \tilde{f}'(a(s))] \cdot \frac{1}{g'(a(s)) - g'(s)} \begin{vmatrix} 1 & s - a(s) \\ g' & g(s) - g(a(s)) \end{vmatrix} = 0.$$

Note that

$$\frac{1}{g'(a(s)) - g'(s)} \begin{vmatrix} 1 & s - a(s) \\ g' & g(s) - g(a(s)) \end{vmatrix} < 0$$

and $a(s)$ is invertible. Therefore, we get the differential equation $z(u)\mathbf{C}(u) + z'(u) = 0$, where \mathbf{C} is in $C^1([a(b_1), a(s_1)])$, $z(u) = f(u) - \tilde{f}(u)$ and $\mathbf{C} < 0$. The condition $z'(a(s_1)) = 0$ implies $z(a(s_1)) = 0$. Note that $z = 0$ is a trivial solution. Therefore, by uniqueness of solutions to ODEs, we get $z = 0$.

We are going to check the concavity of B . Let \mathcal{F} be the force function corresponding to B . By Corollary 21 we only need to check that $\mathcal{F}(s_1) \leq 0$. Note that (17) and (28) imply

$$\mathcal{F}(s_1) = \frac{f''(s_1)}{g''(s_1)} - t_2(s_1) = \frac{f''(s_1)}{g''(s_1)} - \frac{f'(s_1) - f'(a(s_1))}{g'(s_1) - g'(a(s_1))},$$

which is negative by (26). □

Remark 30. The above lemma is true for all choices $s_1 \in (c, b_1)$. If we send s_1 to c then one can easily see that $\lim_{s_1 \rightarrow c^+} t_2(s_1) = 0$, therefore the force function \mathcal{F} takes the form

$$\mathcal{F}(s) = \int_c^s \left[\frac{f''(y)}{g''(y)} \right]' \exp\left(-\int_y^s \frac{g''(r)}{K(r)} \cos \theta(r) dr\right) dy.$$

This is another way to show that the force function is nonpositive.

The next lemma shows that, regardless of the choices of initial solution (a_1, b_1) of (21), the function $a(s)$ constructed by Lemma 28 is unique (i.e., it does not depend on the pair (a_1, b_1)).

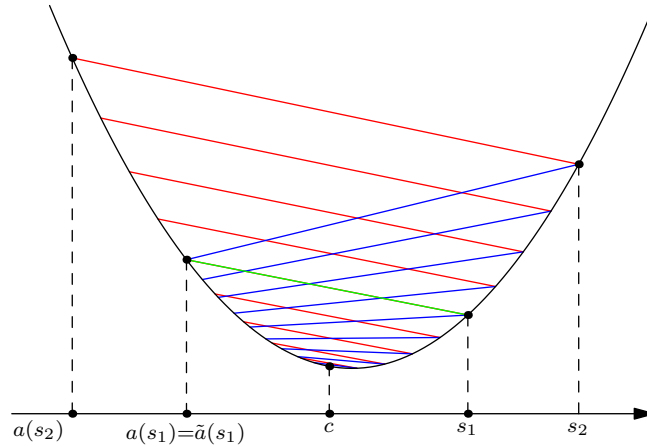


Figure 7. Uniqueness of the cup.

Lemma 31. *Let pairs $(a_1, b_1), (\tilde{a}_1, \tilde{b}_1)$ solve (21), and let $a(s), \tilde{a}(s)$ be the corresponding functions obtained by Lemma 28. Then $a(s) = \tilde{a}(s)$ on $[c, \min\{b_1, \tilde{b}_1\}]$.*

Proof. By the uniqueness result of the implicit function theorem we only need to show existence of $s_1 \in (c, \min\{b_1, \tilde{b}_1\})$ such that $a(s_1) = \tilde{a}(s_1)$. Without loss of generality, assume that $\tilde{b}_1 = b_1 = s_2$. We can also assume that $\tilde{a}(s_2) > a(s_2)$, because other cases can be solved in a similar way.

Let $\Theta = \Theta_{\text{cup}}([c, s_2], g)$ and $\tilde{\Theta} = \tilde{\Theta}_{\text{cup}}([c, s_2], g)$ be the foliations corresponding to the functions $a(s)$ and $\tilde{a}(s)$. Let B and \tilde{B} be the functions corresponding to these foliations from Lemma 29. We consider a chord T in \mathbb{R}^3 joining the points $(a(s_1), g(a(s_1)), f(a(s_1)))$ and $(s_1, g(s_1), f(s_1))$ (see Figure 7). We want to show that the chord T belongs to the graph of \tilde{B} . Indeed, concavity of \tilde{B} (see Lemma 29) implies that the chord T lies below the graph of $\tilde{B}(x_1, x_2)$, where $(x_1, x_2) \in \Omega(\tilde{\Theta})$. Moreover, concavity of B , $\Omega(\tilde{\Theta}) \subset \Omega(\Theta)$ and the fact that the graph \tilde{B} consists of chords joining the points of the curve $(t, g(t), f(t))$ imply that the graph B lies above the graph \tilde{B} . In particular, the chord T , belonging to the graph B , lies above the graph \tilde{B} . This can happen if and only if T belongs to the graph \tilde{B} . Now we show that, if $s_1 < s_2$, then the torsion of the curve $(s, g(s), f(s))$ is zero for $s \in [s_1, s_2]$. Indeed, let \tilde{T} be a chord in \mathbb{R}^3 which joins the points $(a(s_1), g(a(s_1)), f(a(s_1)))$ and $(s_2, g(s_2), f(s_2))$. We consider the tangent plane $L(x)$ to the graph \tilde{B} at the point $(x_1, x_2) = (a(s_1), g(a(s_1)))$. This tangent plane must contain both chords T and \tilde{T} , and it must be tangent to the surface at these chords. Concavity of \tilde{B} implies that the tangent plane L coincides with \tilde{B} at points belonging to the triangle, which is the convex hull of the points $(a(s_1), g(a(s_1)))$, $(s_1, g(s_1))$ and $(s_2, g(s_2))$. Therefore, it is clear that the tangent plane L coincides with \tilde{B} on the segments $\ell \in \tilde{\Theta}$ with the endpoint at $(s, g(s))$ for $s \in [s_1, s_2]$. Thus $L((s, g(s))) = \tilde{B}((s, g(s)))$ for any $s \in [s_1, s_2]$. This means that the torsion of the curve $(s, g(s), f(s))$ is zero on $s \in [s_1, s_2]$, which contradicts our assumption about the torsion. Therefore $s_1 = s_2$. \square

Corollary 32. *In the conditions of Lemma 27, for all $0 < P < \min\{c - a_0, b_0 - c\}$ there exists a unique pair (a_1, b_1) which solves (21) such that $b_1 - a_1 = P$.*

The above corollary implies that, if the pairs (a_1, b_1) and $(\tilde{a}_1, \tilde{b}_1)$ solve (21), then $a_1 \neq \tilde{a}_1$ and $b_1 \neq \tilde{b}_1$, and one of the following conditions holds: $(a_1, b_1) \subset (\tilde{a}_1, \tilde{b}_1)$ or $(\tilde{a}_1, \tilde{b}_1) \subset (a_1, b_1)$.

Remark 33. The function $a(s)$ is defined on the right of the point c . We extend naturally its definition on the left of the interval by $a(s) \stackrel{\text{def}}{=} a^{-1}(s)$.

4. Construction of the Bellman function

4.1. Reduction to the two-dimensional case. We are going to construct the Bellman function for the case $p < 2$. The case $p = 2$ is trivial, and the case $p > 2$ was solved in [Boros et al. 2012]. From the definition of H it follows that

$$H(x_1, x_2, x_3) = H(|x_1|, |x_2|, x_3) \quad \text{for all } (x_1, x_2, x_3) \in \Omega. \tag{30}$$

Also note the homogeneity condition

$$H(\lambda x_1, \lambda x_2, \lambda^p x_3) = \lambda^p H(x_1, x_2, x_3) \quad \text{for all } \lambda \geq 0. \tag{31}$$

These two conditions (30), (31), which follow from the nature of the boundary data $(x^2 + \tau^2 y^2)^{2/p}$, make the construction of H easier. However, in order to construct the function H , this information is not necessary. Further, we assume that H is $C^1(\Omega)$ smooth. Then, from the symmetry (30), it follows that

$$\frac{\partial H}{\partial x_j} = 0 \quad \text{on } x_j = 0 \text{ for } j = 1, 2. \tag{32}$$

For convenience, as in [Boros et al. 2012], we rotate the system of coordinates (x_1, x_2, x_3) . Namely, let

$$y_1 \stackrel{\text{def}}{=} \frac{x_1 + x_2}{2}, \quad y_2 \stackrel{\text{def}}{=} \frac{x_2 - x_1}{2}, \quad y_3 \stackrel{\text{def}}{=} x_3. \tag{33}$$

We define

$$N(y_1, y_2, y_3) \stackrel{\text{def}}{=} H(y_1 - y_2, y_1 + y_2, y_3) \quad \text{on } \Omega_1,$$

where $\Omega_1 = \{(y_1, y_2, y_3) : y_3 \geq 0, |y_1 - y_2|^p \leq y_3\}$. It is clear that, for fixed y_1 , the function N is concave in the variables y_2 and y_3 ; moreover, for fixed y_2 , the function N is concave with respect to the other variables. The symmetry (30) for N turns into the condition

$$N(y_1, y_2, y_3) = N(y_2, y_1, y_3) = N(-y_1, -y_2, y_3). \tag{34}$$

Thus it is sufficient to construct the function N on the domain

$$\Omega_2 \stackrel{\text{def}}{=} \{(y_1, y_2, y_3) : y_1 \geq 0, -y_1 \leq y_2 \leq y_1, (y_1 - y_2)^p \leq y_3\}.$$

Condition (32) turns into

$$\frac{\partial N}{\partial y_1} = \frac{\partial N}{\partial y_2} \quad \text{on the hyperplane } y_2 = y_1, \tag{35}$$

$$\frac{\partial N}{\partial y_1} = -\frac{\partial N}{\partial y_2} \quad \text{on the hyperplane } y_2 = -y_1. \tag{36}$$

The boundary condition (6) becomes

$$N(y_1, y_2, |y_1 - y_2|^p) = ((y_1 + y_2)^2 + \tau^2(y_1 - y_2)^2)^{p/2}. \quad (37)$$

The homogeneity condition (31) implies that $N(\lambda y_1, \lambda y_2, \lambda^p y_3) = \lambda^p N(y_1, y_2, y_3)$ for $\lambda \geq 0$. We choose $\lambda = 1/y_1$, and we obtain that

$$N(y_1, y_2, y_3) = y_1^p N\left(1, \frac{y_2}{y_1}, \frac{y_3}{y_1^p}\right) \quad (38)$$

Suppose we are able to construct the function $M(y_2, y_3) \stackrel{\text{def}}{=} N(1, y_2, y_3)$ on

$$\Omega_3 \stackrel{\text{def}}{=} \{(y_2, y_3) : -1 \leq y_2 \leq 1, (1 - y_2)^p \leq y_3\}$$

with the following conditions:

- (1) M is concave in Ω_3 .
- (2) M satisfies (37) for $y_1 = 1$.
- (3) The extension of M onto Ω_1 via formulas (38) and (34) is a function with the properties of N (see (35), (36), and concavity of N).
- (4) M is minimal among those who satisfy the conditions (1)–(3).

Then the extended function M should be N . So we are going to construct M on Ω_3 . We denote

$$g(t) \stackrel{\text{def}}{=} (1 - t)^p, \quad t \in [-1, 1], \quad (39)$$

$$f(t) \stackrel{\text{def}}{=} ((1 + t)^2 + \tau^2(1 - t)^2)^{p/2}, \quad t \in [-1, 1]. \quad (40)$$

Then we have the boundary condition

$$M(t, g(t)) = f(t), \quad t \in [-1, 1]. \quad (41)$$

We differentiate the condition (38) with respect to y_1 at the point $(y_1, y_2, y_3) = (1, -1, y_3)$ and we obtain that

$$\frac{\partial N}{\partial y_1}(1, -1, y_3) = pN(1, -1, y_3) + \frac{\partial N}{\partial y_2}(1, -1, y_3) - py_3 \frac{\partial N}{\partial y_3}, \quad y_3 \geq 0.$$

Now we use (36), so we obtain another requirement for $M(y_2, y_3)$:

$$0 = pM(-1, y_3) + 2 \frac{\partial M}{\partial y_2}(-1, y_3) - py_3 \frac{\partial M}{\partial y_3}(-1, y_3) \quad \text{for } y_3 \geq 0. \quad (42)$$

Similarly, we differentiate (38) with respect to y_1 at the point $(y_1, y_2, y_3) = (1, 1, y_3)$ and use (35), so we obtain

$$0 = pM(1, y_3) - 2 \frac{\partial M}{\partial y_2}(1, y_3) - py_3 \frac{\partial M}{\partial y_3}(1, y_3) \quad \text{for } y_3 \geq 0. \quad (43)$$

So, in order to satisfy conditions (35) and (36), the requirements (42) and (43) are necessary. It is easy to see that these requirements are also sufficient in order to satisfy these conditions.

The minimum between two concave functions with fixed boundary data is a concave function with the same boundary data. Note also that the conditions (42) and (43) are still fulfilled after taking the minimum. Thus it is quite reasonable to construct a candidate for $M(y_2, y_3)$ as a minimal concave function on Ω_3 with the boundary conditions (41), (42) and (43). We recall that we should also have the concavity of the extended function $N(y_1, y_2, y_3)$ with respect to the variables y_1, y_3 for each fixed y_2 . This condition can be verified after the construction of the function $M(y_2, y_3)$.

4.2. Construction of a candidate for M . We are going to construct a candidate B for M . Firstly, we show that, for $\tau > 0$, the torsion τ_γ of the boundary curve $\gamma(t) \stackrel{\text{def}}{=} (t, g(t), f(t))$ on $t \in (-1, 1)$, where f, g are defined by (39) and (40), changes sign once from $+$ to $-$. We call this point the root of a cup. We construct the cup around this point. Note that $g' < 0, g'' > 0$ on $[-1, 1)$. Therefore,

$$\text{sign } \tau_\gamma = \text{sign} \left(f''' - \frac{g'''}{g''} f'' \right) = \text{sign} \left(f''' - \frac{2-p}{1-t} f'' \right) = \text{sign}(v(t)),$$

where

$$v(t) \stackrel{\text{def}}{=} -(1 + \tau^2)^2(p - 1)t^3 + (1 + \tau^2)(3\tau^2 + \tau^2 p + 3 - 3p)t^2 + (2\tau^2 p - 9\tau^4 + \tau^4 p + 3 - 3p - 6\tau^2)t - p + 5\tau^4 + 2\tau^2 p - \tau^4 p - 10\tau^2 + 1.$$

Note that $v(-1) = 16\tau^4 > 0$ and $v(1) = -8((p - 1) + \tau^2) < 0$. So the function $v(t)$ changes sign from $+$ to $-$ at least once. Now, we show that $v(t)$ has only one root. For $\tau^2 < 3(p - 1)/(3 - p)$, note that the linear function $v''(t)$ is nonnegative, i.e., $v''(-1) = 8\tau^2 p(1 + \tau^2) > 0, v''(1) = -4(1 + \tau^2)(\tau^2 p - 3\tau^2 + 3p - 3) \geq 0$. Therefore, the convexity of $v(t)$ implies the uniqueness of the root $v(t)$ on $[-1, 1]$.

Suppose $\tau^2 < 3(p - 1)/(3 - p)$; we will show that $v' \leq 0$ on $[-1, 1]$. Indeed, the discriminant of the quadratic function $v'(x)$ has the expression

$$D = 16\tau^2(\tau^2 + 1)^2((3 - p)^2\tau^2 - 9(p - 1)),$$

which is negative for $0 < \tau^2 < 3(p - 1)/(3 - p)$. Moreover, $v'(-1) = -4\tau^2(\tau^2 p + 3\tau^2 + 3) < 0$. Thus we obtain that v' is negative.

We denote the root of v by c . It is an appropriate time to make the following remark:

Remark 34. Note that $v(-1 + 2/p) < 0$. Indeed,

$$v\left(-1 + \frac{2}{p}\right) = \frac{(3p - 2)(p^2 - 2p - 4)\tau^4 + (16 + 5p^3 - 8p^2 - 16p)\tau^2 + 8(1 - p)}{p^3},$$

which is negative because the coefficients of τ^4, τ^2, τ^0 are negative. Therefore, this inequality implies that $c < -1 + 2/p$.

Consider $a = -1$ and $b = 1$; the left side of (21) takes the positive value $-2^{2p-1}p(1 - p)$. However, if we consider $a = -1$ and $b = c$, then the proof of Lemma 27 (see (23)) implies that the left side of (21) is negative. Therefore, there exists a unique $s_0 \in (c, 1)$ such that the pair $(-1, s_0)$ solves (21). Uniqueness follows from Corollary 32. The equation (21) for the pair $(-1, s_0)$ is equivalent to the

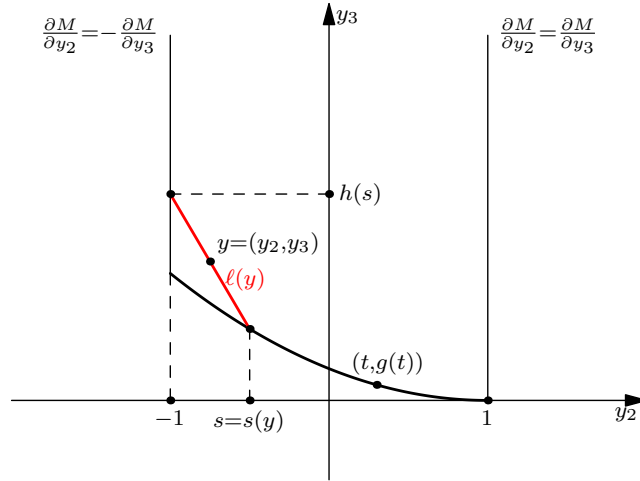


Figure 8. The segment $\ell(y)$.

equation $u((1 + s_0)/(1 - s_0)) = 0$, where

$$u(z) \stackrel{\text{def}}{=} \tau^p(p - 1)(\tau^2 + z^2)^{(2-p)/2} - \tau^2(p - 1) + (1 + z)^{2-p} - z(2 - p) - 1. \tag{44}$$

Lemma 28 gives the function $a(s)$, and Lemma 29 gives the concave function $B(y_2, y_3)$ for $s_1 = c$ with the foliation $\Theta_{\text{cup}}((c, s_0], g)$ in the domain $\Omega(\Theta_{\text{cup}}((c, s_0], g))$.

The above explanation implies the following corollary:

Corollary 35. *Pick any point $\tilde{y}_2 \in (-1, 1)$. The inequalities $s_0 < \tilde{y}_2$, $s_0 = \tilde{y}_2$ and $\tilde{y}_2 > s_0$ are equivalent to the following inequalities, respectively: $u((1 + \tilde{y}_2)/(1 - \tilde{y}_2)) < 0$, $u((1 + \tilde{y}_2)/(1 - \tilde{y}_2)) = 0$ and $u((1 + \tilde{y}_2)/(1 - \tilde{y}_2)) > 0$.*

Now we are going to extend C^1 smoothly the function B on the upper part of the cup. Recall that we are looking for a minimal concave function. If we construct a function with a foliation $\Theta([s_0, \tilde{y}_2], g)$, where $\tilde{y}_2 \in (s_0, 1)$, then the best thing we can do according to Lemma 23 and Lemma 22 is to minimize $\sin(\theta_{\text{cup}}(s_0) - \theta(s_0))$, where $\theta_{\text{cup}}(s)$ is an argument function of $\Theta_{\text{cup}}((c, s_0], g)$ and $\theta(s)$ is an argument function of $\Theta([s_0, \tilde{y}_2], g)$. In other words, we need to choose segments from $\Theta([s_0, \tilde{y}_2], g)$ close enough to the segments of $\Theta_{\text{cup}}((c, s_0], g)$.

Thus, we are going to construct the set of segments $\Theta([s_0, \tilde{y}_2])$ so that they start from $(s, g(s), f(s))$, $s \in [s_0, \tilde{y}_2]$, and they go to the boundary $y_2 = -1$ of Ω_3 .

We explain how the conditions (42) and (43) allow us to construct such a type of foliation $\Theta([s_0, \tilde{y}_2], g)$ in a unique way. Let $\ell(y)$ be the segment with the endpoints $(s, g(s))$, where $s \in (s_0, \tilde{y}_2)$ and $(-1, h(s))$ (see Figure 8).

Let $t(s) = (t_1(s), t_2(s)) = \nabla B(y)$, where $s = s(y)$ is the corresponding gradient function. Then (42) takes the form

$$0 = pB(-1, h(s)) + 2t_1(s) - ph(s)t_2(s). \tag{45}$$

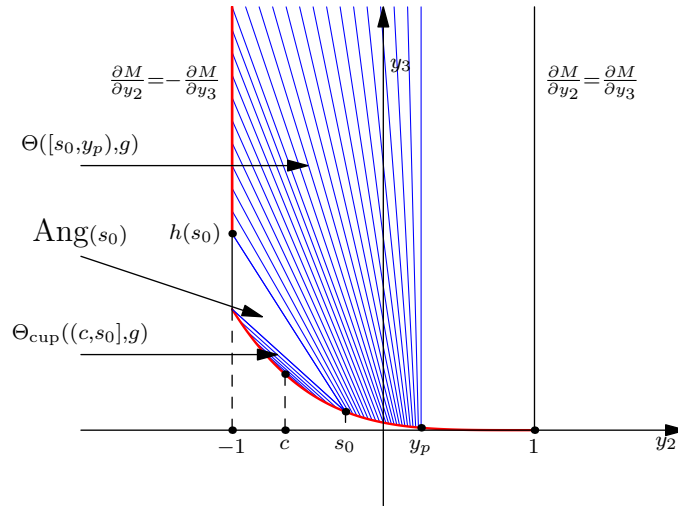


Figure 9. Foliations $\Theta_{\text{cup}}((c, s_0], g)$ and $\Theta([s_0, y_p), g)$.

We differentiate this expression with respect to s , and we obtain

$$2t'_1(s) - ph(s)t'_2(s) = 0. \tag{46}$$

Then, according to (11), we find the function $\tan \theta(s)$, and, hence, we find the quantity $h(s)$:

$$\tan \theta(s) = -\frac{ph(s)}{2} \iff \frac{h(s) - g(s)}{s + 1} = \frac{ph(s)}{2}.$$

Therefore,

$$h(s) = \frac{2g(s)}{p} \left(\frac{1}{y_p - s} \right), \quad \text{where } y_p \stackrel{\text{def}}{=} -1 + \frac{2}{p}. \tag{47}$$

We see that the function $h(s)$ is well defined, it increases, and it is differentiable on $-1 \leq s < y_p$. So we conclude that if $s_0 < y_p$ then we are able to construct the set of segments $\Theta([s_0, y_p), g)$ that pass through the points $(s, g(s))$, where $s \in [s_0, y_p)$, and through the boundary $y_2 = -1$ (see Figure 9).

It is easy to check that $\Theta([s_0, y_p), g)$ is a foliation, so, taking the value $t_2(s_0)$ of B on $\Omega(\Theta([s_0, y_p), g))$ according to Lemma 23, by Corollary 25 we have constructed a concave function B in the domain $\Omega(\Theta_{\text{cup}}((c, s_0], g)) \cup \text{Ang}(s_0) \cup \Omega(\Theta([s_0, y_p], g))$.

It is clear that the foliation $\Theta([s_0, y_p), g)$ exists as long as $s_0 < y_p$. Note that $(1 + y_p)/(1 - y_p) = 1/(p - 1)$. Therefore, Corollary 35 implies the following remark:

Remark 36. The inequalities $s_0 < y_p$, $s_0 = y_p$ and $s_0 > y_p$ are equivalent to the following inequalities respectively: $u(1/(p - 1)) < 0$, $u(1/(p - 1)) = 0$ and $u(1/(p - 1)) > 0$.

At the point y_p , the segments from $\Theta([s_0, y_p), g)$ become vertical. After the point $(y_p, g(y_p))$, we should consider vertical segments $\Theta([y_p, 1], g)$ (see Figure 10), because by Lemma 22 this corresponds to the minimal function. Surely $\Theta([y_p, 1], g)$ is the foliation. Again, choosing the value $t_2(y_p)$ of B on $\Omega(\Theta([y_p, 1], g))$ according to Lemma 23, by Corollary 25 we have constructed the concave function B

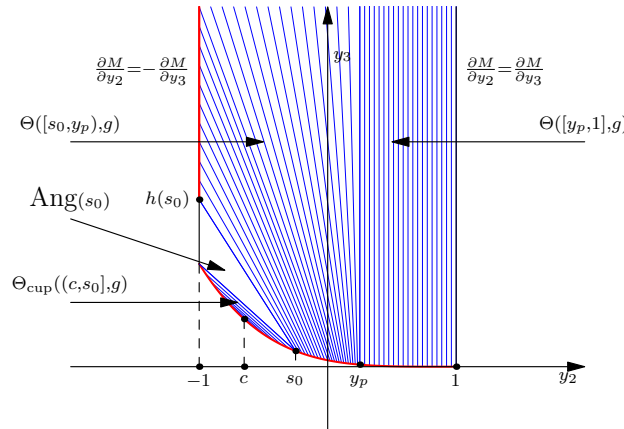


Figure 10. The case $u(1/(p - 1)) < 0$.

on Ω_3 . Note that if $s_0 \geq y_p$ (which corresponds to the inequality $u(1/(p - 1)) > 0$) then we do not have the foliation $\Theta([s_0, y_p], g)$. In this case we consider only vertical segments $\Theta([s_0, 1], g)$ (see Figure 11), and again, choosing the value $t_2(s_0)$ of B on $\Omega(\Theta([s_0, 1], g))$ according to Lemma 23, by Corollary 25 we construct a concave function B on Ω_3 . We believe that $B = M$.

We still have to check the requirements (42) and (43). A crucial role is played by symmetry of the boundary data of N . Further, the given proofs work for both of the cases $y_p < s_0$ and $y_p \geq s_0$, so we do not consider them separately.

The requirement (43) follows immediately. Indeed, the condition (14) at the point $y = (1, y_3)$ (note that in (14) instead of $x = (x_1, x_2)$ we consider $y = (y_2, y_3)$) implies that $B(1, y_3) = f(1) + t_2(1)(y_3 - g(1))$. Therefore, (43) takes the form $0 = pf(1) - 2t_1(1)$. Using (12), we obtain that $t_1(1) = f'(1)$. Therefore, we see that $pf(1) - 2t_1(1) = pf(1) - 2f'(1) = 0$.

Now, we are going to obtain the requirement (42) which is the same as (45). The quantities t_1, t_2 of B with the foliation $\Theta([s_0, y_p], g)$ satisfy the condition (46) which was obtained by differentiation of (45). So we only need to check the condition (45) at the initial point $s = s_0$. If we substitute the expression of B from (14) into (45), then (45) turns into the following equivalent condition:

$$t_1(s)(s - y_p) + t_2(s)g(s) = f(s). \tag{48}$$

Note that (12) allows us to rewrite (48) into the equivalent condition

$$t_2(s) = \frac{f(s) - (s - y_p)f'(s)}{g(s) - (s - y_p)g'(s)}. \tag{49}$$

And, as was mentioned above we only need to check condition (49) at the point $s = s_0$, i.e.,

$$t_2(s_0) = \frac{f(s_0) - (s_0 - y_p)f'(s_0)}{g(s_0) - (s_0 - y_p)g'(s_0)}. \tag{50}$$

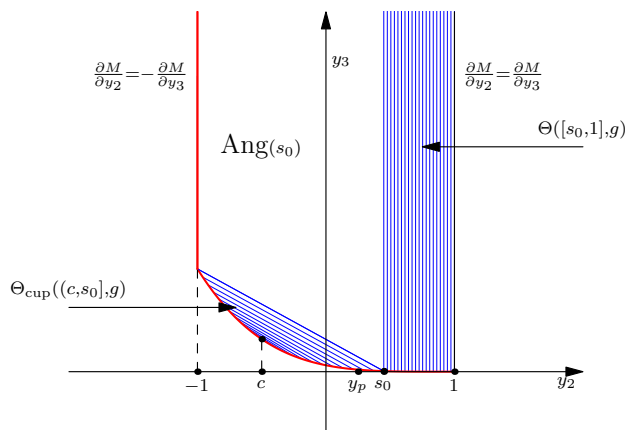


Figure 11. The case $u(1/(p - 1)) \geq 0$.

On the other hand, if we differentiate the boundary condition $B(s, g(s)) = f(s)$ at the points $s = s_0, -1$, then we obtain

$$\begin{aligned} t_1(s_0) + t_2(s_0)g'(-1) &= f'(-1), \\ t_1(s_0) + t_2(s_0)g'(s_0) &= f'(s_0). \end{aligned}$$

Thus we can find the value of $t_2(s_0)$:

$$t_2(s_0) = \frac{f'(-1) - f'(s_0)}{g'(-1) - g'(s_0)}. \tag{51}$$

So these two values (51) and (50) must coincide. In other words, we need to show

$$\frac{f(s_0) - (s_0 - y_p)f'(s_0)}{g(s_0) - (s_0 - y_p)g'(s_0)} = \frac{f'(-1) - f'(s_0)}{g'(-1) - g'(s_0)}. \tag{52}$$

It will be convenient for us to work with the following notations for the rest of the current subsection. We denote $g(-1) = g_-$, $g'(-1) = g'_-$, $f(-1) = f_-$, $f'(-1) = f'_-$, $g(s_0) = g$, $g'(s_0) = g'$, $f(s_0) = f$ and $f'(s_0) = f'$. The condition (52) is equivalent to

$$s_0 = \frac{fg'_- + f'g - fg' - gf'_-}{f'g'_- - g'f'_-} + y_p = \left(\frac{fg'_- + f'g - fg' - gf'_-}{f'g'_- - g'f'_-} - 1 \right) + \frac{2}{p}. \tag{53}$$

On the other hand, from (21) for the pair $(-1, s_0)$, we obtain that

$$s_0 = \left(\frac{fg'_- + f'g - fg' - gf'_-}{f'g'_- - g'f'_-} - 1 \right) + \frac{f'g_- + g'_-f_- - g'f'_- - f'_-g_-}{g'f'_- - f'g'_-}.$$

So, from (53) we see that it suffices to show that

$$\frac{f'g_- + g'_-f_- - g'f'_- - f'_-g_-}{g'f'_- - f'g'_-} = \frac{2}{p}.$$

We note that $g'_- = -(p/2)g_-$, $f'_- = -(p/2)f_-$, hence $g'_-f_- = f'_-g_-$. Therefore, we have

$$\frac{f'g_- + g'_f_- - g'f_- - f'_g_-}{g'f'_- - f'g'_-} = \frac{f'g_- - g'f_-}{g'f'_- - f'g'_-} = \frac{2}{p}.$$

4.3. Concavity in another direction. We are going to check the concavity of the extended function N via B in another direction. It is worth mentioning that both of the cases $y_p < s_0$, $y_p \geq s_0$ do not play any role in the following computations, therefore we consider them together. We define a candidate for N as

$$N(y_1, y_2, y_3) \stackrel{\text{def}}{=} y_1^p B\left(1, \frac{y_2}{y_1}, \frac{y_3}{y_1^p}\right) \quad \text{for } \left(\frac{y_2}{y_1}, \frac{y_3}{y_1^p}\right) \in \Omega_3, \tag{54}$$

and we extend N to Ω_1 by (34). Then, as was already discussed, $N \in C^1(\Omega_1)$. We need the following technical lemma:

Lemma 37. $N''_{y_1y_1} N''_{y_3y_3} - (N''_{y_1y_3})^2 = -t'_2 s'_{y_3} p(p-1) y_1^{p-2} \left(s t_1 + g t_2 - f + \frac{y_2}{y_1} t_1 \cdot \left(\frac{2}{p} - 1 \right) \right),$

where $s = s(y_2/y_1, y_3/y_1^p)$ and $(y_2/y_1, y_3/y_1^p) \in \text{int}(\Omega_3) \setminus \text{Ang}(s_0)$.

As was mentioned in Remark 24, the gradient function $t(s)$ is not necessarily differentiable at the point s_0 ; this is the reason for the requirement $(y_2/y_1, y_3/y_1^p) \in \text{int}(\Omega_3) \setminus \text{Ang}(s_0)$ in the lemma. However, from the proof of the lemma, the reader can easily see that $N''_{y_1y_1} N''_{y_3y_3} - (N''_{y_1y_3})^2 = 0$ whenever the points $(y_2/y_1, y_3/y_1^p)$ belong to the interior of the domain $\text{Ang}(s_0)$.

Proof. The definition of the candidate N (see (54)) implies $N''_{y_3y_3} = t'_2(s) s'_{y_3}$, $N''_{y_3y_1} = t'_2 s'_{y_1}$ and

$$N'_{y_1} = y_1^{p-1} \left(p B\left(\frac{y_2}{y_1}, \frac{y_3}{y_1^p}\right) - t_1 \frac{y_2}{y_1} - p t_2 \frac{y_3}{y_1^p} \right). \tag{55}$$

Condition (14) implies

$$B\left(\frac{y_2}{y_1}, \frac{y_3}{y_1^p}\right) = f(s) + t_1 \cdot \left(\frac{y_2}{y_1} - s\right) + t_2 \cdot \left(\frac{y_3}{y_1^p} - g(s)\right).$$

We substitute this expression for $B(y_2/y_1, y_3/y_1^p)$ into (55), and we obtain

$$N'_{y_1} = y_1^{p-1} \left(p f + \frac{y_2}{y_1} t_1 (p-1) - p s t_1 - p g t_2 \right). \tag{56}$$

The condition $(y_2/y_1, y_3/y_1^p) \in \text{int}(\Omega_3) \setminus \text{Ang}(s_0)$ implies the equality $N''_{y_1y_3} = N''_{y_3y_1}$, which in turn gives

$$t'_2 s'_{y_1} = y_1^{p-1} \left(p f' + \frac{y_2}{y_1} t'_1 (p-1) - (p s t_1 + p g t_2)'_s \right) s'_{y_3}.$$

Hence,

$$t'_2 \cdot (s'_{y_1})^2 = y_1^{p-1} \left(p f' + \frac{y_2}{y_1} t'_1 (p-1) - (p s t_1 + p g t_2)'_s \right) s'_{y_3} s'_{y_1}. \tag{57}$$

We keep in mind this identity, and continue our calculations:

$$N''_{y_1y_1} = (p-1) y_1^{p-2} \left(p f + \frac{y_2}{y_1} t_1 (p-2) - p s t_1 - p g t_2 \right) + y_1^{p-1} \left(p f' + \frac{y_2}{y_1} t'_1 (p-1) - (p s t_1 + p g t_2)'_s \right) s'_{y_1}.$$

So, finally we obtain

$$N''_{y_1 y_1} N''_{y_3 y_3} - (N''_{y_1 y_3})^2 = t'_2(N''_{y_1 y_1} s'_{y_3} - t'_2(s'_{y_1})^2).$$

Now we use the identity (57), and we substitute the expression $t'_2(s'_{y_1})^2$:

$$\begin{aligned} N''_{y_1 y_1} N''_{y_3 y_3} - (N''_{y_1 y_3})^2 &= t'_2 s'_{y_3} \left(N''_{y_1 y_1} - y_1^{p-1} \left(p f' + \frac{y_2}{y_1} t'_1 (p-1) - (p s t_1 + p g t_2)'_s \right) s'_{y_1} \right) \\ &= t'_2 s'_{y_3} (p-1) y_1^{p-2} \left(p f + \frac{y_2}{y_1} t_1 (p-2) - p s t_1 - p g t_2 \right) \\ &= -t'_2 s'_{y_3} p (p-1) y_1^{p-2} \left(s t_1 + g t_2 - f + \frac{y_2}{y_1} t_1 \cdot \left(\frac{2}{p} - 1 \right) \right). \quad \square \end{aligned}$$

Now we are going to consider several cases, when the points $(y_2/y_1, y_3/y_1^p)$ belong to the different subdomains in Ω_3 . Note that we always have $N''_{y_3 y_3} \leq 0$, because of the fact that B is concave in Ω_3 and (54). So we only have to check that the determinant of the Hessian of N is negative. If the determinant of the Hessian is zero, then it is sufficient to ensure that $N''_{y_3 y_3}$ is strictly negative, and, if $N''_{y_3 y_3}$ is also zero, then we need to ensure that N''_{y_1, y_1} is nonpositive.

The domain $\Omega(\Theta[s_0, y_p])$. In this case we can use the equality (48), and we obtain that

$$s t_1 + g t_2 - f = y_p t_1.$$

Therefore,

$$N''_{y_1 y_1} N''_{y_3 y_3} - (N''_{y_1 y_3})^2 = -t'_2 s'_{y_3} p (p-1) y_1^{p-2} t_1 y_p \left(1 + \frac{y_2}{y_1} \right) \geq 0$$

because $t_1 \geq 0$. Indeed, $t_1(s)$ is continuous on $[c, 1]$, where c is the root of the cup and $B''_{y_2 y_2} = t'_1 s'_{y_2} \leq 0$; therefore, because of the fact $s'_{y_2} > 0$, it suffices to check that $t_1(1) \geq 0$, which follows from the inequality

$$t_1(1) = f'(1) - t_2(1)g'(1) = f'(1) > 0.$$

Domain of linearity $\text{Ang}(s_0)$. This is the domain that consists of the triangle ABC with $A = (-1, g(-1))$, $B = (s_0, g(s_0))$ and $C = (-1, h(s_0))$ if $s_0 < y_p$, and the infinite domain of linearity, which is of rectangular type and which lies between the chords AB, BC' , where $C' = (s_0, +\infty)$, and AC'' , where $C'' = (-1, +\infty)$ (see Figure 11).

Suppose the points $(y_2/y_1, y_3/y_1^p)$ belong to the interior of $\text{Ang}(s_0)$. Then the gradient function $t(s)$ of B is constant, and, moreover, $s(y_2/y_1, y_3/y_1^p)$ is constant. The fact that the determinant of the Hessian is zero in the domain of linearity (note that $s'_{y_3} = 0$) implies that we only need to check $N''_{y_1 y_1} < 0$. The equality (56) implies

$$N''_{y_1 y_1} = (p-1) y_1^{p-2} \left(p f + \frac{y_2}{y_1} t_1 (p-2) - p s_0 t_1 - p g t_2 \right) \leq (p-1) y_1^{p-2} (p f - p s_0 t_1 - p g t_2 - t_1 (p-2)) = 0.$$

The last equality follows from (48). The above inequality turns into an equality if and only if $y_2/y_1 = s_0$; this is the boundary point of $\text{Ang}(s_0)$.

Domain of vertical segments. On the vertical segments, the determinant of the Hessian is zero (for example, because the vertical segment is a vertical segment in all directions) and $B''_{y_3 y_3} = 0$; therefore, we must check that $N''_{y_1 y_1} \leq 0$. We note that $s(y_2, y_3) = y_2$, so

$$N''_{y_1 y_1} = y_1^{p-2} \times [(p-1)(pf + st_1(p-2) - pst_1 - pgt_2) - s(pf' - t_1' s - t_1 p - pg't_2)].$$

However, from (12) we have $pf' - t_1 p - pg't_2 = 0$; therefore,

$$N''_{y_1 y_1} = y_1^{p-2} \times [(p-1)(pf - 2st_1 - pgt_2) + s^2 t_1'].$$

The condition $t_1' \leq 0$ implies that it is sufficient to show $pf - 2st_1 - pgt_2 \leq 0$. We use (12), and we find $t_1 = f' - g't_2$. Hence,

$$pf - 2st_1 - pgt_2 = pf - gpt_2 - 2s(f' - g't_2) = pf - 2sf' - t_2(gp - 2sg').$$

Note that $gp - 2sg' \geq 0$ (because $s \geq 0$ and $g' \leq 0$). From (12) and the fact that on the vertical segments t_2 is constant (see the expression for t_2 in Lemma 13 and note that $\cos \theta(s) = 0$), it follows that $0 \geq t_1' = f'' - g''t_2$; therefore, we have $t_2 \geq f''/g''$. Therefore,

$$pf - 2sf' - t_2(gp - 2sg') \leq pf - 2sf' - \frac{f''}{g''}(gp - 2sg').$$

Now we recall the values (41), (40), and after direct calculations we obtain

$$pf - 2sf' - \frac{f''}{g''}(gp - 2sg') = \frac{f(1-s^2)p(p-2)(\tau^2(1+s)^2 + (1-s)^2 + 2\tau^2(1-s^2))}{(p-1)((1+s)^2 + \tau^2(1-s)^2)} \leq 0.$$

Domain of the cup $\Omega(\Theta_{\text{cup}}((c, s_0], g))$. The condition that $N''_{y_3 y_3}$ is strictly negative in the cup implies that we only need to show $st_2 + gt_3 - f + (y_2/y_1)t_1(2/p - 1) \geq 0$, where $s = s(y_2/y_1, y_3/y_1^p)$ and the points $y = (y_2/y_1, y_3/y_1^p)$ lie in the cup. Without loss of generality we can assume that $y_1 = 1$. Therefore it suffices to show that $st_2 + gt_3 - f + y_2 t_1(2/p - 1) \geq 0$, where $y = (y_2, y_3) \in \Omega(\Theta_{\text{cup}}((c, s_0], g))$. On a segment with the fixed endpoint $(s, g(s))$ the expressions $s, t_1, g(s), t_2$ and $f(s)$ are constant, so the expression $st_1 + gt_2 - f + y_2 t_1(2/p - 1)$ is linear with respect to y_2 on each segment of the cup. Therefore, the worst case appears when $y_2 = a(s)$ (it is the left end — an abscissa — of the given segment). This is true because $t_1 \geq 0$ (as was already shown) and $(2/p - 1) \geq 0$. So, as a result, we derive that it is sufficient to prove the inequality

$$st_1 + gt_2 - f + a(s)t_1 \cdot \left(\frac{2}{p} - 1\right) = t_1(s - a(s)) + gt_2 - f + \frac{2a(s)}{p}t_1 \geq 0. \quad (58)$$

We use the identity (14) at the point $y = (a(s), g(a(s)))$, and we find that

$$t_1(s - a(s)) + gt_2 - f = g(a(s))t_2 - f(a(s)).$$

We substitute this expression into (58), then we get that it suffices to prove the inequality

$$g(a(s))t_2 - f(a(s)) + \frac{2a(s)}{p}t_1 \geq 0. \quad (59)$$

We differentiate the condition $B(a(s), g(a(s))) = f(s)$ with respect to s . Then we find the expression for $t_1(s)$, namely $t_1(s) = f'(a(s)) - t_2(s)g'(a(s))$. After substituting this expression into (59) we obtain that

$$g(a(s))t_2 - f(a(s)) + \frac{2a(s)}{p}t_1 = \frac{1+z}{g'(z)} \left(\frac{(z-1)(\tau^2+1)f(z)}{((1+z)^2 + \tau^2(1-z)^2)g'(z)} - t_2(s) \right),$$

where $z = a(s)$. So it suffices to show that

$$\frac{(z-1)(\tau^2+1)f(z)}{((1+z)^2 + \tau^2(1-z)^2)g'(z)} - t_2(s) \leq 0 \tag{60}$$

because g' is negative. We are going to show that it is sufficient to check the condition (60) at the point $z = -1$. Indeed, note that $(t_2)'_z \geq 0$ on $[-1, c]$, where c is the root of the cup, and also note that

$$\left(\frac{(z-1)(\tau^2+1)f}{((1+z)^2 + \tau^2(1-z)^2)g'} \right)'_z = \frac{\tau^2+1}{p}(p-2)(1-z)^{-(p-1)}[(1+z)^2 + \tau^2(1-z)^2]^{p/2-2}2(1+z) \leq 0.$$

The condition (60) at the point $z = -1$ turns into the condition

$$t_2(s_0) - \frac{\tau^{p-2}(\tau^2+1)}{p} \geq 0.$$

Now we recall (27) and $t_2(s_0) = (f'(-1) - f'(s_0))/(g'(-1) - g'(s_0))$; therefore, we have

$$t_2(s_0) - \frac{\tau^{p-2}(\tau^2+1)}{p} \geq \frac{f''(-1)}{g''(-1)} - \frac{\tau^{p-2}(\tau^2+1)}{p} = \frac{\tau^p(p-1)^2 + \tau^{p-2}}{p(p-1)} > 0.$$

Thus we finish this section with the following remark:

Remark 38. We still have to check the cases when the points $(y_2/y_1, y_3/y_1^p)$ belong to the boundary of $\text{Ang}(s_0)$ and the vertical rays $y_2 = \pm 1$ in Ω_3 . The reader can easily see that, in this case, the concavity of N follows from the observation that $N \in C^1(\Omega_1)$. Symmetry of N covers the rest of the cases when $(y_2/y_1, y_3/y_1^p) \notin \Omega_3$.

Thus we have constructed the candidate N .

5. Sharp constants via foliation

5.1. Main theorem. We remind the reader the definition of the functions $u(z)$, $g(s)$ and $f(s)$ (see (44), (39) and (40)), the value $y_p = -1 + 2/p$ and the definition of the function $a(s)$ (see Lemma 28, Lemma 31 and Remark 33).

Theorem 39. Let $1 < p < 2$, let G be the martingale transform of F and let $|\mathbb{E}G| \leq \beta|\mathbb{E}F|$. Set $\beta' = \frac{\beta-1}{\beta+1}$.
 (i) If $u(1/(p-1)) \leq 0$ then

$$\mathbb{E}(\tau^2 F^2 + G^2)^{p/2} \leq \left(\tau^2 + \max \left\{ \beta, \frac{1}{p-1} \right\}^2 \right)^{\frac{p}{2}} \mathbb{E}|F|^p. \tag{61}$$

(ii) If $u(1/(p-1)) > 0$ then

$$\mathbb{E}(\tau^2 F^2 + G^2)^{p/2} \leq C(\beta')\mathbb{E}|F|^p,$$

where $C(\beta')$ is continuous, nondecreasing, and is defined by

$$C(\beta') \stackrel{\text{def}}{=} \begin{cases} (\tau^2 + \beta^2)^{p/2} & \text{if } \beta' \geq s^*, \\ \tau^p \left(1 - \frac{2^{2-p}(1-s_0)^{p-1}}{(\tau^2 + 1)(p-1)(1-s_0) + 2(2-p)} \right)^{-1} & \text{if } \beta' \leq -1 + 2/p, \\ \frac{f'(s_1) - f'(a(s_1))}{g'(s_1) - g'(a(s_1))} & \text{if } R(s_1, \beta') = 0 \text{ for } s_1 \in (\beta', s_0), \end{cases}$$

where $s_0 \in (-1 + 2/p, 1)$ is the solution of the equation $u((1 + s_0)/(1 - s_0)) = 0$, and the function $R(s, z)$ is defined as follows:

$$R(s, z) \stackrel{\text{def}}{=} -f(s) - \frac{f'(a(s))g'(s) - f'(s)g'(a(s))}{g'(s) - g'(a(s))}(z - s) + \frac{f'(s) - f'(a(s))}{g'(s) - g'(a(s))}g(s)$$

for $z \in [-1 + 2/p, s^*]$, $s \in [z, s_0]$. The value $s^* \in [-1 + 2/p, s_0]$ is the solution of the equation

$$\frac{f'(s^*) - f'(a(s^*))}{g'(s^*) - g'(a(s^*))} = \frac{f(s^*)}{g(s^*)}. \tag{62}$$

Proof. Before we investigate some of the cases mentioned in the theorem, we should make the following observation. The inequality (61) can be restated as follows:

$$H(x_1, x_2, x_3) \leq Cx_3, \tag{63}$$

where H is defined by (5) and $x_1 = \mathbb{E}F$, $x_2 = \mathbb{E}G$, $x_3 = \mathbb{E}|F|^p$. In order to derive the estimate (61), we have to find the sharp C in (63). Because of the property (30), we can assume that both of the values x_1, x_2 are nonnegative. So, the nonnegativity of x_1, x_2 and the condition $|\mathbb{E}G| \leq \beta|\mathbb{E}F|$ can be reformulated as

$$-\frac{x_1 + x_2}{2} \leq \frac{x_2 - x_1}{2} \leq \frac{\beta - 1}{\beta + 1} \cdot \frac{x_1 + x_2}{2}. \tag{64}$$

The condition (64) with (63) in terms of the function N and the variables y_1, y_2, y_3 means that we have to find the sharp C such that

$$N(y_1, y_2, y_3) \leq Cy_3 \quad \text{for } -y_1 \leq y_2 \leq \frac{\beta - 1}{\beta + 1}y_1, \mathbf{y} \in \Omega_2.$$

Because of (38), the above condition can be reformulated as

$$B(y_2, y_3) \leq Cy_3 \quad \text{for } -1 \leq y_2 \leq \frac{\beta - 1}{\beta + 1}, y_3 \geq g(y_2), \tag{65}$$

where $B(y_2, y_3) = N(1, y_2, y_3)$. So our aim is to find the sharp C , or in other words the value $\sup_{y_1, y_2} B/y_3$, where the supremum is taken from the domain mentioned in (65). Note that the quantity $B(y_2, y_3)/y_3$ increases with respect to the variable y_2 . Indeed, $(B(y_2, y_3)/y_3)'_{y_2} = t_1(s(y))/y_3$, where the function $t_1(s)$ is nonnegative on $[c_0, 1]$ (see the end of the proof of the concavity condition in the domain $\Omega(\Theta[s_0, y_p])$). Note that, as we increase the value y_2 , the range of y_3 also increases. This means that the supremum of the expression B/y_3 is attained on the subdomain where $y_2 = (\beta - 1)/(\beta + 1)$. It is worth mentioning

that, since the quantity $(\beta - 1)/(\beta + 1) \in [-1, 1]$ increases as β increases and because of the observation made above, we see that the value C increases as β' increases.

5.2. The case $y_p \leq s_0$. We are going to investigate the simple case (i). Remark 36 implies that $s_0 \leq y_p$; in other words, the foliation of vertical segments is $\Theta([y_p, 1], g)$, where the value $\theta(s)$ on $[y_p, 1]$ is equal to $\pi/2$. This means that $t_2(s)$ is constant on $[y_p, 1]$ (see Lemma 13), and it is equal to $f(y_p)/g(y_p) = (\tau^2 + 1/(p - 1)^2)^{p/2}$ (see (49)).

If $(\beta - 1)/(\beta + 1) \geq y_p$, or equivalently $\beta \geq 1/(p - 1)$, then the function B on the vertical segment with the endpoint $(\beta', g(\beta'))$, where $(\beta - 1)/(\beta + 1) = \beta' \in [y_p, 1)$, has the expression (see (14))

$$B(\beta', y_3) = f(\beta') + \frac{f(y_p)}{g(y_p)}(y_3 - g(\beta')), \quad y_3 \geq g(\beta').$$

Therefore,

$$\frac{B(\beta', y_3)}{y_3} = \frac{f(y_p)}{g(y_p)} + \frac{g(\beta')}{y_3} \left(\frac{f(\beta')}{g(\beta')} - \frac{f(y_p)}{g(y_p)} \right), \quad y_3 \geq g(\beta'). \tag{66}$$

The expression $f(s)/g(s)$ is strictly increasing on $(-1, 1)$; therefore, (66) attains its maximal value at the point $y_3 = g(\beta')$. Thus, we have

$$\frac{B(y_2, y_3)}{y_3} \leq \frac{B(\beta', y_3)}{y_3} \leq \frac{B(\beta', g(\beta'))}{g(\beta')} = \frac{f(\beta')}{g(\beta')} = (\tau^2 + \beta^2)^{p/2} \quad \text{for } -1 \leq y_2 \leq \beta', \quad y_3 \geq g(y_2).$$

If $(\beta - 1)/(\beta + 1) < y_p$, or equivalently $\beta < 1/(p - 1)$, then we can achieve the value for C which was achieved at the moment $\beta = 1/(p - 1)$, and, since the function $C = C(\beta')$ increases as β' increases, this value will be the best. Indeed, it suffices to look at the foliation (see Figure 10). For any fixed y_2 we send y_3 to $+\infty$, and we obtain that

$$\begin{aligned} \lim_{y_3 \rightarrow \infty} \frac{B(y_2, y_3)}{y_3} &= \lim_{y_3 \rightarrow \infty} \frac{f(s) + t_1(s)(y_2 - s) + t_2(s)(y_3 - g(s))}{y_3} \\ &= \lim_{y_3 \rightarrow \infty} t_2(s(y)) = t_2(y_p) = \left(\tau^2 + \frac{1}{(p - 1)^2} \right)^{\frac{p}{2}}. \end{aligned}$$

5.3. The case $y_p > s_0$. As was already mentioned, the condition in case (ii) is equivalent to the inequality $s_0 > y_p$ (see Remark 36). This means that the foliation of the vertical segments is $\Theta([s_0, 1], g)$ (see Figure 11). We know that $C(\beta')$ is increasing. We recall that we are going to maximize the function $B(y_2, y_3)/y_3$ on the domain in (65). It was already mentioned that we can require $y_2 = (\beta - 1)/(\beta + 1) = \beta'$. For such fixed $y_2 = \beta' \in [-1, 1]$, we are going to investigate the monotonicity of the function $B(\beta', y_3)/y_3$. We consider several cases. Let $\beta' \geq s_0$. We differentiate the function $B(\beta', y_3)/y_3$ with respect to y_3 , and we use the expression (14) for B to obtain that

$$\frac{\partial}{\partial y_3} \left(\frac{B(\beta', y_3)}{y_3} \right) = \frac{t_2(\beta')y_3 - B(\beta', y_3)}{y_3^2} = \frac{-f(\beta') + t_2(\beta')g(\beta')}{y_3^2}.$$

Recall that $t_2(s) = t_2(s_0)$ for $s \in [s_0, 1]$; therefore, direct calculations imply

$$t_2(\beta') = \frac{f(s_0) - (s_0 - y_p)f'(s_0)}{g(s_0) - (s_0 - y_p)g'(s_0)} < \frac{f(s_0)}{g(s_0)} \leq \frac{f(\beta')}{g(\beta')}, \quad \beta' \geq s_0.$$

This implies that

$$C(\beta') = \sup_{y_3 \geq g(\beta')} \frac{B(\beta', y_3)}{y_3} = \frac{B(\beta', y_3)}{y_3} \Big|_{y_3=g(\beta')} = \frac{f(\beta')}{g(\beta')} = (\tau^2 + \beta^2)^{p/2}.$$

Now we consider the case $\beta' < s_0$. For each point $y = (\beta', y_3)$ that belongs to the line $y_2 = \beta'$, there exists a segment $\ell(y) \in \Theta((c, s_0], g)$ with the endpoint $(s, g(s))$, where $s \in [\max\{\beta', a(\beta')\}, s_0]$. If the point y belongs to the domain of linearity $\text{Ang}(s_0)$, then we can choose the value s_0 and consider any segment with the endpoints y and $(s_0, g(s_0))$, which surely belongs to the domain of linearity. The reader can easily see that as we increase the value y_3 the value s increases as well. So,

$$\frac{\partial}{\partial y_3} \left(\frac{B(\beta', y_3)}{y_3} \right) = \frac{t_2(s)y_3 - B(\beta', y_3)}{y_3^2} = \frac{-f(s) - t_1(s)(\beta' - s) + t_2(s)g(s)}{y_3^2}.$$

Our aim is to investigate the sign of the expression $-f(s) - t_1(s)(\beta' - s) + t_2(s)g(s)$ as we vary the value $y_3 \in [g(\beta'), +\infty)$. Without loss of generality, we can forget about the variable y_3 , and we can vary only the value s on the interval $[\max\{\alpha(\beta'), \beta'\}, s_0]$.

We consider the function $R(s, z) \stackrel{\text{def}}{=} -f(s) - t_1(s)(z - s) + t_2(s)g(s)$ with the domain $-1 \leq z \leq s_0$ and $s \in [\max\{\alpha(z), z\}, s_0]$ (see Figure 12). As we have already seen, $R(s_0, s_0) < 0$. Note that $R(s_0, -1) > 0$. Indeed, $R(s_0, -1) = t_2(s_0)g(-1) - f(-1)$. This equality follows from the fact that

$$f(s_0) - f(-1) = t_1(s_0)(s_0 + 1) + t_2(s_0)(g(s_0) - g(-1)),$$

which is a consequence of Lemma 29. So, (51) and (27) imply

$$t_2(s_0) = \frac{f'(-1) - f'(s_0)}{g'(-1) - g'(s_0)} > \frac{f''(-1)}{g''(-1)} \geq \frac{f(-1)}{g(-1)}.$$

The function $R(z, s_0)$ is linear with respect to z . So, on the interval $[-1, s_0]$, it has the root $y_p = -1 + 2/p$. Indeed,

$$\frac{-f(s_0) + t_2(s_0)g(s_0) + t_1(s_0)s_0}{t_1(s_0)} = y_p.$$

The last equality follows from (51), (53) and (12). We need a few more properties of the function $R(s, z)$. For each fixed z , the function $R(s, z)$ is nonincreasing on $[\max\{\alpha(z), z\}, s_0]$. Indeed,

$$R'_s(s, z) = -f'(s) - t'_1(s)(z - s) + t_1(s) + t'_2(s)g(s) + t_2(s)g'(s). \tag{67}$$

We take into account the condition (12), so the expression (67) simplifies to

$$R'_s(s, z) = t'_2(s)g(s) + t'_1(s)(s - z).$$

We remind the reader of the equality (11) and the fact that $t'_2(s) \leq 0$. Therefore, we have $R'_s(s, z) = y_3 t'_2(s)$, where $y_3 = y_3(s) > 0$. Thus we see that $R(s, \beta') \geq 0$ for $\beta' \leq y_p$.

So, if the function $R(\cdot, z)$ at the right end on its domain $[\max\{\alpha(z), z\}, s_0]$ is positive, this will mean that the function B/y_3 is increasing; hence, the constant $C(\beta')$ will be equal to

$$C(\beta') = \lim_{y_3 \rightarrow \infty} \frac{B(z, y_3)}{y_3} = t_2(s_0) = \frac{f'(-1) - f'(s_0)}{g'(-1) - g'(s_0)}$$

(this follows from (51) and the structure of the foliation). Since $u((1 + s_0)/(1 - s_0)) = 0$ and given (52), direct computations show that

$$\frac{f'(-1) - f'(s_0)}{g'(-1) - g'(s_0)} = \tau^p \left(1 - \frac{2^{2-p}(1 - s_0)^{p-1}}{(\tau^2 + 1)(p - 1)(1 - s_0) + 2(2 - p)} \right)^{-1}. \tag{68}$$

So it follows that, if $\beta' \leq y_p$, then (68) is the value of $C(\beta')$.

If the function $R(\cdot, z)$ on the left end of its domain is nonpositive, this will mean that the function B/y_3 is decreasing, so the sharp constant will be the value of the function $B(z, y_3)/y_3$ at the left end of its domain:

$$C(\beta') = \frac{B(z, y_3)}{y_3} \Big|_{y_3=g(z)} = \frac{f(z)}{g(z)} = (\tau^2 + \beta^2)^{p/2}. \tag{69}$$

We recall that c is the root of the cup and $c < y_p$ (see Remark 34). We will show that the function $R(z, s)$ is decreasing on the boundary $s = z$ for $s \in (y_p, s_0]$. Indeed, (12) implies

$$(R(s, s))'_s = -f'(s) + t'_2(s)g(s) + t_2(s)g'(s) = -t_1(s) + t'_2(s)g(s) < 0.$$

The last inequality follows from the fact that $t'_2(s) \leq 0$ and $t_1(s) > 0$ on $(c, 1]$. Surely $R(y_p, y_p) > R(s_0, y_p) = 0$, and we recall that $R(s_0, s_0) < 0$, so there exists a unique $s^* \in [y_p, s_0]$ such that $R(s^*, s^*) = 0$. This is equivalent to (62). So it is clear that $R(z, z) \leq 0$ for $z \in [s^*, s_0]$. Therefore, $C(\beta')$ has the value (69) for $\beta' \geq s^*$.

The only case that remains is when $\beta' \in [y_p, s^*]$. We know that $R(z, z) \geq 0$ for $z \in [y_p, s^*]$ and $R(s_0, z) \leq 0$ for $z \in [y_p, s^*]$. The fact that, for each fixed z , the function $R(s, z)$ is decreasing implies the following: for each $z \in [y_p, s^*]$, there exists a unique $s_1(z) \in [z, s_0]$ such that $R(z, s_1(z)) = 0$. Therefore, for $\beta' \in [y_p, s^*]$ we have

$$C(\beta') = \frac{B(\beta', y_3(s_1(\beta')))}{y_3(s_1(\beta'))}, \tag{70}$$

where the value $s_1(\beta')$ is the root of the equation $R(s_1(\beta'), \beta') = 0$. Recall that

$$R(s_1(\beta'), \beta') = t_2(s_1)y_3(s_1) - B(\beta', y_3(s_1)) = -f(s_1) - t_1(s_1)(\beta' - s_1) + t_2(s_1)g(s_1). \tag{71}$$

So the expression (70) takes the form

$$C(\beta') = t_2(s_1) = \frac{f'(s_1) - f'(a(s_1))}{g'(s_1) - g'(a(s_1))}.$$

Finally, we remind the reader that

$$t_2(s) = \frac{f'(s) - f'(a(s))}{g'(s) - g'(a(s))},$$

$$t_1(s) = \frac{f'(a(s))g'(s) - f'(s)g'(a(s))}{g'(s) - g'(a(s))}$$

for $s \in (c, s_0]$, and we finish the proof of the theorem. \square

6. Extremizers via foliation

We set $\Psi(F, G) = \mathbb{E}(G^2 + \tau^2 F^2)^{2/p}$. Let N be the candidate that we have constructed in Section 4 (see (54)). We define the candidate \mathbf{B} for the Bellman function H (see (5)) as follows:

$$\mathbf{B}(x_1, x_2, x_3) = N\left(\frac{x_1 + x_2}{2}, \frac{x_2 - x_1}{2}, x_3\right), \quad (x_1, x_2, x_3) \in \Omega.$$

We want to show that $\mathbf{B} = H$. We already know that $\mathbf{B} \geq H$ (see Proposition 9). So, it remains to show that $\mathbf{B} \leq H$. We are going to do this as follows: for each point $\mathbf{x} \in \Omega$ and any $\varepsilon > 0$, we are going to find an admissible pair (F, G) such that

$$\Psi(F, G) > \mathbf{B}(\mathbf{x}) - \varepsilon. \quad (72)$$

Up to the end of the current section, we are going to work with the coordinates (y_1, y_2, y_3) (see (33)). It will be convenient for us to redefine the notion of admissibility of a pair.

Definition 40. We say that a pair (F, G) is admissible for the point $(y_1, y_2, y_3) \in \Omega_1$ if G is the martingale transform of F and $\mathbb{E}(F, G, |F|^p) = (y_1 - y_2, y_1 + y_2, y_3)$.

So, in this case, the condition (72) in terms of the function N takes the following form: for any point $\mathbf{y} \in \Omega_1$ and for any $\varepsilon > 0$, we are going to find an admissible pair (F, G) for the point \mathbf{y} such that

$$\Psi(F, G) > N(\mathbf{y}) - \varepsilon. \quad (73)$$

We formulate the following obvious observations:

Lemma 41. (1) A pair (F, G) is admissible for the point $\mathbf{y} = (y_1, y_2, y_3)$ if and only if $(\tilde{F}, \tilde{G}) = (\pm F, \mp G)$ is admissible for the point $\tilde{\mathbf{y}} = (\mp y_2, \mp y_1, y_3)$; moreover, $\Psi(\tilde{F}, \tilde{G}) = \Psi(F, G)$.

(2) A pair (F, G) is admissible for the point $\mathbf{y} = (y_1, y_2, y_3)$ if and only if $(\tilde{F}, \tilde{G}) = (\lambda F, \lambda G)$ (where $\lambda \neq 0$) is admissible for the point $\tilde{\mathbf{y}} = (\lambda y_1, \lambda y_2, |\lambda|^p y_3)$; moreover, $\Psi(\tilde{F}, \tilde{G}) = |\lambda|^p \Psi(F, G)$.

Definition 42. The pair of functions (F, G) is called an ε -extremizer for the point $\mathbf{y} \in \Omega_1$ if (F, G) is admissible for the point \mathbf{y} and $\Psi(F, G) > N(\mathbf{y}) - \varepsilon$.

Lemma 41, homogeneity, and the symmetry of N imply that we only need to check (73) for the points $\mathbf{y} \in \Omega_1$ where $y_1 = 1$ and $(y_2, y_3) \in \Omega_3$. In other words, we show that $\Psi(F, G) > \mathbf{B}(y_2, y_3) - \varepsilon$ for some admissible pair (F, G) for the point $(1, y_2, y_3)$, where $(y_2, y_3) \in \Omega_3$. Further, instead of saying that (F, G) is an admissible pair (or ε -extremizer) for the point $(1, y_2, y_3)$ we just say that it is an

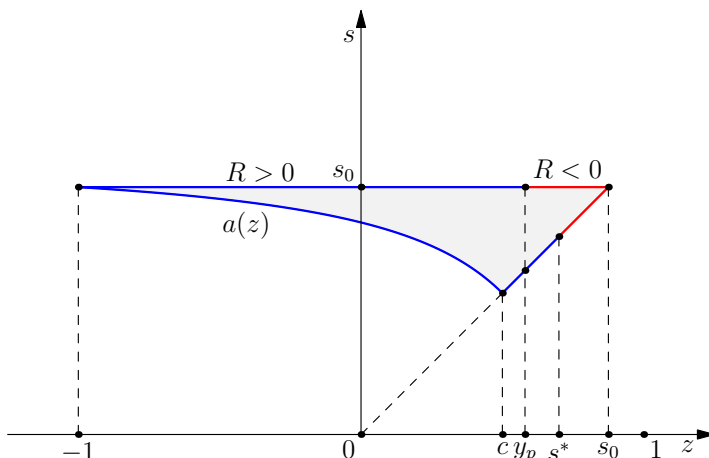


Figure 12. The domain of $R(s, z)$.

admissible pair (or an ε -extremizer) for the point (y_2, y_3) . So we only have to construct ε -extremizers in the domain Ω_3 .

It is worth mentioning that we construct ε -extremizers (F, G) such that G will be the martingale transform of F with respect to some filtration other than dyadic. The reader can find a detailed explanation on how to pass from one filtration to another in [Slavin and Vasyunin 2011].

We need a few more observations. For $\alpha \in (0, 1)$, we define the α -concatenation of the pairs (F, G) and (\tilde{F}, \tilde{G}) as follows:

$$(F \bullet \tilde{F}, G \bullet \tilde{G})_\alpha(x) = \begin{cases} (F, G)(x/\alpha) & \text{if } x \in [0, \alpha], \\ (\tilde{F}, \tilde{G})((x - \alpha)/(1 - \alpha)) & \text{if } x \in [\alpha, 1]. \end{cases}$$

Clearly, $\Psi((F \bullet \tilde{F}, G \bullet \tilde{G})_\alpha(x)) = \alpha\Psi(F, G) + (1 - \alpha)\Psi(\tilde{F}, \tilde{G})$.

Definition 43. Any domain of the type $\Omega_1 \cap \{y_1 = A\}$, where A is some real number, is said to be a *positive domain*. Any domain of the type $\Omega_1 \cap \{y_2 = B\}$, where B is some real number, is said to be a *negative domain*.

The following lemma is obvious:

Lemma 44. *If (F, G) is an admissible pair for a point \mathbf{y} and (\tilde{F}, \tilde{G}) is an admissible pair for a point $\tilde{\mathbf{y}}$ such that either of the following is true: $\mathbf{y}, \tilde{\mathbf{y}}$ belong to a positive domain, or $\mathbf{y}, \tilde{\mathbf{y}}$ belong to a negative domain, then $(F \bullet \tilde{F}, G \bullet \tilde{G})_\alpha$ is an admissible pair for the point $\alpha\mathbf{y} + (1 - \alpha)\tilde{\mathbf{y}}$.*

Let (F, G) be an admissible pair for a point \mathbf{y} , and let (\tilde{F}, \tilde{G}) be an admissible pair for a point $\tilde{\mathbf{y}}$. Let $\hat{\mathbf{y}}$ be a point which belongs to the chord joining the points \mathbf{y} and $\tilde{\mathbf{y}}$.

Remark 45. It is clear that, if (F^+, G^+) is admissible for a point (y_2^+, y_3^+) and (F^-, G^-) is admissible for a point (y_2^-, y_3^-) , then an α -concatenation of these pairs is admissible for the point $(y_2, y_3) = \alpha \cdot (y_2^+, y_3^+) + (1 - \alpha) \cdot (y_2^-, y_3^-)$.

Now we are ready to construct ε -extremizers in Ω_3 . The main idea is that these functions Ψ and \mathbf{B} are very similar: they obey almost the same properties. Moreover, foliation plays a crucial role in the contraction of ε -extremizers.

6.1. The case $s_0 \leq y_p$. We want to find ε -extremizers for the points in Ω_3 .

Extremizers in the domain $\Omega(\Theta_{\text{cup}}((c, s_0], g))$. Pick any $y = (y_2, y_3) \in \Omega(\Theta_{\text{cup}}((c, s_0], g))$. Then there exists a segment $\ell(y) \in \Theta_{\text{cup}}((c, s_0], g)$. Let $y^+ = (s, g(s))$ and $y^- = (a(s), g(a(s)))$ be the endpoints of $\ell(y)$ in Ω_3 . We know ε -extremizers at these points y^+, y^- . Indeed, we can take the ε -extremizers $(F^+, G^+) = (1 - s, 1 + s)$ and $(F^-, G^-) = (1 - a(s), 1 + a(s))$ (i.e., constant functions). Consider an α -concatenation $(F^+ \bullet F^-, G^+ \bullet G^-)_\alpha$, where α is chosen so that $y = \alpha y^+ + (1 - \alpha)y^-$. We have

$$\Psi[(F^+ \bullet F^-, G^+ \bullet G^-)_\alpha] = \alpha \Psi(F^+, G^+) + (1 - \alpha) \Psi(F^-, G^-) > \alpha \mathbf{B}(y^+) + (1 - \alpha) \mathbf{B}(y^-) - \varepsilon = \mathbf{B}(y) - \varepsilon.$$

The last equality follows from the linearity of \mathbf{B} on $\ell(y)$.

Extremizers on the vertical line $(-1, y_3)$, $y_3 \geq h(s_0)$. Now we are going to find ε -extremizers for the points $(-1, y_3)$, where $y_3 \geq h(s_0)$. We use a similar idea to one mentioned in [Vasyunin and Volberg 2010] (see the proof of Lemma 3). We define the functions (F, G) recursively:

$$G(t) = \begin{cases} -w & \text{if } 0 \leq t < \varepsilon, \\ \gamma \cdot g\left(\frac{t - \varepsilon}{1 - 2\varepsilon}\right) & \text{if } \varepsilon \leq t \leq 1 - \varepsilon, \\ w & \text{if } 1 - \varepsilon < t \leq 1, \end{cases}$$

$$F(t) = \begin{cases} d_- & \text{if } 0 \leq t < \varepsilon, \\ \gamma \cdot f\left(\frac{t - \varepsilon}{1 - 2\varepsilon}\right) & \text{if } \varepsilon \leq t \leq 1 - \varepsilon, \\ d_+ & \text{if } 1 - \varepsilon < t \leq 1, \end{cases}$$

where the nonnegative constants w, d_-, d_+ and γ will be obtained from the requirement $\mathbb{E}(F, G, |F|^p) = (2, 0, y_3)$ and the fact that G is the martingale transform of F . Surely $\langle G \rangle_{[0,1]} = 0$. The condition $\langle F \rangle_{[0,1]} = 2$ means that

$$(d_- + d_+) \varepsilon + 2\gamma(1 - 2\varepsilon) = 2. \tag{74}$$

The condition $\langle |F|^p \rangle_{[0,1]} = y_3$ implies that

$$y_3 = \frac{\varepsilon(d_+^p + d_-^p)}{1 - (1 - 2\varepsilon)\gamma^p}. \tag{75}$$

Now we use the condition $|F_0 - F_1| = |G_0 - G_1|$. In the first step we split the interval $[0, 1]$ at the point ε with the requirement $F_0 - F_1 = G_0 - G_1$, from which we obtain $w = 2 - d_-$. In the second step we split at the point $1 - \varepsilon$ with the requirement $F_1 - F_2 = G_2 - G_1$, obtaining $w = 2\gamma - d_+$. From these two conditions we obtain $d_- + d_+ = 2(1 + \gamma) - 2w$, and by substituting in (74) we find

$$\gamma = 1 + \frac{\varepsilon w}{1 - \varepsilon}.$$

Now we investigate what happens as ε tends to zero. Our aim will be to focus on the limit value $\lim_{\varepsilon \rightarrow 0} w = w_0$. We have $1 - (1 - 2\varepsilon)\gamma^p \approx \varepsilon(2 - wp)$. So (75) becomes

$$y_3 = \frac{\varepsilon(d_+^p + d_-^p)}{1 - (1 - 2\varepsilon)\gamma^p} \rightarrow \frac{2(2 - w_0)^p}{2 - w_0 p}. \tag{76}$$

Note that, for $w_0 = 1 + s$, equation (76) is the same as (47). By direct calculations we see that as $\varepsilon \rightarrow 0$ we have

$$\langle (G^2 + \tau^2 F^2)^{p/2} \rangle_{[0,1]} = \frac{\varepsilon[(w^2 + \tau^2 d_-^2)^{p/2} + (w^2 + \tau^2 d_+^2)^{p/2}]}{1 - (1 - 2\varepsilon)\gamma^p} \rightarrow \frac{2f(w_0 - 1)}{2 - w_0 p}.$$

Now we are going to calculate the value $\mathbf{B}(-1, h(s))$, where $h(s) = y_3$. From (45) we have

$$\mathbf{B}(-1, h(s)) = h(s)t_2(s) - \frac{2}{p}t_1(s).$$

By using (12) we express t_1 via t_2 ; also because of (47) and (50), we have

$$\begin{aligned} \mathbf{B}(-1, h(s)) &= h(s)t_2(s) - \frac{2}{p}t_1(s) = h(s)t_2 - \frac{2}{p}(f' - t_2g') = t_2\left(h(s) + \frac{2}{p}g'\right) - f'\frac{2}{p} \\ &= \frac{f(s) - (s - y_p)f'(s)}{g(s) - (s - y_p)g'(s)}\left(\frac{2g}{p(y_p - s)} + \frac{2}{p}g'\right) - f'\frac{2}{p} \\ &= \frac{2}{p}\left[\frac{f(s)}{y_p - s}\right] = \frac{2(2 - w_0)^p}{2 - w_0 p}. \end{aligned}$$

Thus we obtain the desired result

$$\langle (G^2 + \tau^2 F^2)^{p/2} \rangle_{[0,1]} \rightarrow \mathbf{B}(-1, y_3) \quad \text{as } \varepsilon \rightarrow 0.$$

Extremizers in the domain $\Omega(\Theta([s_0, y_p], g))$. Pick any point $y = (y_2, y_3) \in \Omega(\Theta([s_0, y_p], g))$. Then there exists a segment $\ell(y) \in \Theta([s_0, y_p], g)$. Let y^+ and y^- be the endpoints of this segment, so that $y^+ = (-1, \tilde{y}_3)$ for some $\tilde{y}_3 \geq h(s_0)$ and $y^- = (\tilde{s}, g(\tilde{s}))$ for some $\tilde{s} \in [y_p, s_0]$. We remind the reader that we know ε -extremizers for the points $(s, g(s))$, where $s \in [s_0, 1]$, and we know ε -extremizers on the vertical line $(-1, y_3)$, where $y_3 \geq h(s_0)$. Therefore, as in the case of a cup, taking the appropriate α -concatenation of these ε -extremizers and using the fact that \mathbf{B} is linear on $\ell(y)$, we find an ε -extremizer at the point y .

Extremizers in the domain $\text{Ang}(s_0)$. Pick any $y = (y_1, y_2) \in \text{Ang}(s_0)$. There exist points $y^+ \in \ell^+$, $y^- \in \ell^-$, where $\ell^+ = \ell^+(s_0, g(s_0)) \in \Theta([s_0, y_p], g)$ and $\ell^- = \ell^-(s_0, g(s_0)) \in \Theta((c, s_0], g)$, such that $y = \alpha y^+ + (1 - \alpha)y^-$ for some $\alpha \in [0, 1]$. We know ε -extremizers at the points y^+ and y^- . Then by taking an α -concatenation of these extremizers and using the linearity of \mathbf{B} on $\text{Ang}(s_0)$ we can obtain an ε -extremizer at the point y .

Extremizers in the domain $\Omega(\Theta([y_p, 1], g))$. Finally, we consider the domain of vertical segments $\Omega(\Theta([y_p, 1], g))$. Pick any point $y = (y_2, y_3) \in \Omega(\Theta([y_p, 1], g))$. Take an arbitrary point $y^+ = (-1, y_3^+)$, where y_3^+ is sufficiently large such that $y = \alpha y^+ + (1 - \alpha)y^-$ for some $\alpha \in (0, 1)$ and some $y^- = (y_2^-, y_3^-)$

with $(1, y_2^-, y_3^-) \in \partial\Omega_1$. Surely, y^+ and y^- belong to a positive domain. The condition $(1, y_2^-, y_3^-) \in \partial\Omega_1$ implies that we know an ε -extremizer (F^-, G^-) at the point y^- (these are constant functions). We also know an ε -extremizer at the point y^+ . Let $(F^+ \bullet F^-, G^+ \bullet G^-)_\alpha$ be an α -concatenation of these extremizers. Then

$$\Psi[(F^+ \bullet F^-, G^+ \bullet G^-)_\alpha] > \alpha \mathbf{B}(y^+) + (1 - \alpha) \mathbf{B}(y^-) - \varepsilon.$$

Note that the condition $y = \alpha y^+ + (1 - \alpha) y^-$ implies that

$$\alpha = \frac{y_3 - (y_2/y_2^-)y_3^-}{y_3^+ + y_3^-/y_2^-}.$$

Recall that $\mathbf{B}(y_2, g(y_2)) = f(y_2)$ and $\mathbf{B}(y^+) = f(s) + t_1(s)(-1 - s) + t_2(s)(y_3^+ - g(s))$, where $s \in [s_0, y_p]$ is such that a segment $\ell(s, g(s)) \in \Theta([s_0, y_p], g)$ has an endpoint y^+ .

Note that as $y_3^+ \rightarrow \infty$ all terms remain bounded; moreover, $\alpha \rightarrow 0$, $y^- \rightarrow (y_2, g(y_2))$ and $s \rightarrow y_p$. This means that

$$\begin{aligned} \lim_{y_3^+ \rightarrow \infty} \alpha \mathbf{B}(y^+) + (1 - \alpha) \mathbf{B}(y^-) - \varepsilon &= \lim_{y_3^+ \rightarrow \infty} t_2(s) y_3^+ \left(\frac{y_3 - (y_2/y_2^-)y_3^-}{y_3^+ + y_3^-/y_2^-} \right) + f(y_2) - \varepsilon \\ &= t_2(y_p)(y_3 - g(y_2)) + f(y_2) - \varepsilon. \end{aligned}$$

We recall that $t_2(s) = t_2(y_p)$ for $s \in [y_p, 1]$. Then

$$\mathbf{B}(y) = f(y_2) + t_2(s(y))(y_3 - g(y_2)) = f(y_2) + t_2(y_p)(y_3 - g(y_2)).$$

Thus, if we choose y_3^+ sufficiently large then we can obtain a 2ε -extremizer for the point y .

6.2. The case $s_0 > y_p$. In this case we have $s_0 \geq y_p$ (see Figure 11). This case is a little bit more complicated than the previous one. The construction of ε -extremizers (F, G) will be similar to the one presented in [Reznikov et al. 2013].

We need a few more definitions.

Definition 46. Let (F, G) be an arbitrary pair of functions. Let $(y_2, g(y_2)) \in \Omega_3$ and let J be a subinterval of $[0, 1]$. We define a new pair (\tilde{F}, \tilde{G}) as follows:

$$(\tilde{F}, \tilde{G})(x) = \begin{cases} (F, G)(x) & \text{if } x \in [0, 1] \setminus J \\ (1 - y_2, 1 + y_2) & \text{if } x \in J. \end{cases}$$

We will refer to the new pair (\tilde{F}, \tilde{G}) as *putting the constant $(y_2, g(y_2))$ on the interval J* for the pair (F, G) .

Sometimes we will denote the new pair (\tilde{F}, \tilde{G}) by the same symbol (F, G) .

Definition 47. We say that the pairs (F_α, G_α) , $(F_{1-\alpha}, G_{1-\alpha})$ are obtained from the pair (F, G) by splitting at the point $\alpha \in (0, 1)$ if

$$\begin{aligned} (F_\alpha, G_\alpha) &= (F, G)(x \cdot \alpha), & x \in [0, 1], \\ (F_{1-\alpha}, G_{1-\alpha}) &= (F, G)(x \cdot (1 - \alpha) + \alpha), & x \in [0, 1]. \end{aligned}$$

It is clear that $\Psi(F, G) = \alpha\Psi(F_\alpha, G_\alpha) + (1 - \alpha)\Psi(F_{1-\alpha}, G_{1-\alpha})$. Also note that, if (F_α, G_α) and $(F_{1-\alpha}, G_{1-\alpha})$ are obtained from the pair (F, G) by splitting at the point $\alpha \in (0, 1)$, then (F, G) is an α -concatenation of the pairs (F_α, G_α) and $(F_{1-\alpha}, G_{1-\alpha})$. Thus, splitting and concatenation are opposite operations.

Instead of explicitly presenting an admissible pair (F, G) and showing that it is an ε -extremizer, we present an algorithm which constructs the admissible pair, and we show that the result is an ε -extremizer.

By the same explanations as in the case $s_0 \leq y_p$, it is enough to construct an ε -extremizer (F, G) on the vertical line $y_2 = -1$ of the domain Ω_3 . Moreover, linearity of \mathbf{B} implies that, for any $A > 0$, it is enough to construct ε -extremizers for the points $(-1, y_3)$, where $y_3 \geq A$. Pick any point $(-1, y_3)$, where $y_3 = y_3^{(0)} > g(-1)$. Linearity of \mathbf{B} on $\text{Ang}(s_0)$ and direct calculations (see (14), (51)) show that

$$\mathbf{B}(-1, y_3) = f(-1) + t_3(s_0)(y_3 - g(-1)) = f(-1) + (y_3 - g(-1)) \frac{f'(-1) - f'(s_0)}{g'(-1) - g'(s_0)}. \tag{77}$$

We describe the first iteration. Let (F, G) be an admissible pair for the point $(-1, y_3)$, whose explicit expression will be described during the algorithm. For a pair (F, G) , we put a constant $(s_0, g(s_0))$ on an interval $[0, \varepsilon]$, where the value $\varepsilon \in (0, 1)$ will be given later. Thus we obtain a new pair (F, G) , which we denote by the same symbol. We want (F, G) to be an admissible pair for the point $(-1, y_3)$. Let the pairs $(F_\varepsilon, G_\varepsilon)$ and $(F_{1-\varepsilon}, G_{1-\varepsilon})$ be obtained from the pair (F, G) by splitting at the point ε . It is clear that $(F_\varepsilon, G_\varepsilon)$ is an admissible pair for the point $(s_0, g(s_0))$. We want $(F_{1-\varepsilon}, G_{1-\varepsilon})$ to be an admissible pair for the point $P = (\tilde{y}_2, \tilde{y}_3)$, so that

$$(-1, y_3) = \varepsilon(s_0, g(s_0)) + (1 - \varepsilon)P. \tag{78}$$

Therefore we require

$$P = \left(\frac{-1 - \varepsilon s_0}{1 - \varepsilon}, \frac{y_3 - \varepsilon g(s_0)}{1 - \varepsilon} \right). \tag{79}$$

So we make the following simple observation: if $(F_{1-\varepsilon}, G_{1-\varepsilon})$ were an admissible pair for the point P , then (F, G) (which is an ε -concatenation of the pairs $(1 - s_0, 1 + s_0)$ and $(F_{1-\varepsilon}, G_{1-\varepsilon})$) would be an admissible pair for the point $(-1, y_3)$. The explanation of this observation is simple: note that the pairs $(F_{1-\varepsilon}, G_{1-\varepsilon})$ and $(1 - s_0, 1 + s_0)$ are admissible pairs for the points P and $(s_0, g(s_0))$, which belong to a positive domain (see (78)); therefore, the rest immediately follows from Lemma 44. So we want to construct the admissible pair $(F_{1-\varepsilon}, G_{1-\varepsilon})$ for the point (79).

We recall Lemma 41, which implies that the pair $(F_{1-\varepsilon}, G_{1-\varepsilon})$ is admissible for the point

$$\left(1, \frac{-1 - \varepsilon s_0}{1 - \varepsilon}, \frac{y_3 - \varepsilon g(s_0)}{1 - \varepsilon} \right)$$

if and only if the pair (\tilde{F}, \tilde{G}) , where $(F_{1-\varepsilon}, -G_{1-\varepsilon}) = (1 + \varepsilon s_0)/(1 - \varepsilon)(\tilde{F}, \tilde{G})$, is admissible for a point

$$W = \left(1, \frac{\varepsilon - 1}{1 + \varepsilon s_0}, \frac{y_3 - \varepsilon g(s_0)}{(1 + \varepsilon s_0)^p} \cdot (1 - \varepsilon)^{p-1} \right).$$

So, if we find the admissible pair (\tilde{F}, \tilde{G}) then we automatically find the admissible pair (F, G) .

Choose ε small enough so that $((\varepsilon - 1)/(1 + \varepsilon s_0), (y_3 - \varepsilon g(s_0))/(1 + \varepsilon s_0)^p \cdot (1 - \varepsilon)^{p-1}) \in \Omega_3$ and

$$\left(\frac{\varepsilon - 1}{1 + \varepsilon s_0}, \frac{y_3 - \varepsilon g(s_0)}{(1 + \varepsilon s_0)^p} \cdot (1 - \varepsilon)^{p-1} \right) = \delta(s_0, g(s_0)) + (1 - \delta)(-1, y_3^{(1)})$$

for some $\delta \in (0, 1)$ and $y_3^{(1)} \geq g(-1)$. Then

$$\begin{aligned} \delta &= \frac{\varepsilon}{1 + \varepsilon s_0} = \varepsilon + O(\varepsilon^2), \\ y_3^{(1)} &= \frac{((y_3 - \varepsilon g(s_0))/(1 + \varepsilon s_0)^p) \cdot (1 - \varepsilon)^{p-1} - (\varepsilon/(1 + \varepsilon s_0))g(s_0)}{1 - \varepsilon/(1 + \varepsilon s_0)} \\ &= y_3(1 - \varepsilon(p + ps_0 - 2)) - 2\varepsilon g(s_0) + O(\varepsilon^2). \end{aligned} \tag{80}$$

For the pair (\tilde{F}, \tilde{G}) , we put a constant $(s_0, g(s_0))$ on the interval $[0, \delta]$. We split the new pair (\tilde{F}, \tilde{G}) at δ , so we get the pairs $(\tilde{F}_\delta, \tilde{G}_\delta)$ and $(\tilde{F}_{1-\delta}, \tilde{G}_{1-\delta})$. We make a similar observation as above. It is clear that if we know the admissible pair $(\tilde{F}_{1-\delta}, \tilde{G}_{1-\delta})$ for the point $(-1, y_3^{(1)})$, then we can obtain an admissible pair (\tilde{F}, \tilde{G}) for the point

$$\left(\frac{\varepsilon - 1}{1 + \varepsilon s_0}, \frac{y_3 - \varepsilon g(s_0)}{(1 + \varepsilon s_0)^p} \cdot (1 - \varepsilon)^{p-1} \right).$$

Surely (\tilde{F}, \tilde{G}) is a δ -concatenation of the pairs $(1 - s_0, 1 + s_0)$ and $(\tilde{F}_{1-\delta}, \tilde{G}_{1-\delta})$.

We summarize the first iteration. We took $\varepsilon \in (0, 1)$, and we started from the pair $(F^{(0)}, G^{(0)}) = (F, G)$, and after one iteration we came to the pair $(F^{(1)}, G^{(1)}) = (\tilde{F}_{1-\delta}, \tilde{G}_{1-\delta})$. We showed that, if $(F^{(1)}, G^{(1)})$ is an admissible pair for the point $(1, y_3^{(1)})$, then the pair $(F^{(0)}, G^{(0)})$ can be obtained from the pair $(F^{(1)}, G^{(1)})$; moreover, it is admissible for the point $(1, y_3^{(0)})$.

Continuing these iterations, we obtain the sequence of numbers $\{y_3^{(j)}\}_{j=0}^N$ and the sequence of pairs $\{(F^{(j)}, G^{(j)})\}_{j=0}^N$. Let N be such that $y_3^{(N)} \geq g(-1)$. It is clear that, if $(F^{(N)}, G^{(N)})$ is an admissible pair for the point $(-1, y_3^{(N)})$, then the pairs $\{(F^{(j)}, G^{(j)})\}_{j=0}^{N-1}$ can be determined uniquely, and, moreover, $(F^{(j)}, G^{(j)})$ is an admissible pair for the point $(-1, y_3^{(j)})$ for all $j = 0, \dots, N - 1$.

Note that we can choose sufficiently small $\varepsilon \in (0, 1)$, and we can find $N = N(\varepsilon)$ such that $y_3^{(N)} = g(-1)$ (see (80), and recall that $s_0 > y_p$). In this case the admissible pair $(F^{(N)}, G^{(N)})$ for the point $(-1, y_3^{(N)}) = (-1, g(-1))$ is a constant function, namely, $(F^{(N)}, G^{(N)}) = (2, 0)$. Now we try to find N in terms of ε , and we try to find the value of $\Psi(F^{(0)}, G^{(0)})$.

Condition (80) implies that $y_3^{(1)} = y_3^{(0)}(1 - \varepsilon(p + ps_0 - 2)) - 2\varepsilon g(s_0) + O(\varepsilon^2)$. We denote $\delta_0 = p + ps_0 - 2 > 0$. Therefore, after the N -th iteration we obtain

$$y_3^{(N)} = (1 - \varepsilon\delta_0)^N \left(y_3^{(0)} + \frac{2g(s_0)}{\delta_0} \right) - \frac{2g(s_0)}{\delta_0} + O(\varepsilon).$$

The requirement $y_3^{(N)} = g(-1)$ implies that

$$(1 - \varepsilon\delta_0)^{-N} = \frac{y_3^{(0)} + 2g(s_0)/\delta_0}{g(-1) + 2g(s_0)/\delta_0} + O(\varepsilon).$$

This implies that $\limsup_{\varepsilon \rightarrow 0} \varepsilon \cdot N = \limsup_{\varepsilon \rightarrow 0} \varepsilon \cdot N(\varepsilon) < \infty$. Therefore, we get

$$e^{\varepsilon \delta_0 N} = \frac{y_3^{(0)} + 2g(s_0)/\delta_0}{g(-1) + 2g(s_0)/\delta_0} + O(\varepsilon). \tag{81}$$

Also note that

$$\begin{aligned} \Psi(F^{(0)}, G^{(0)}) &= \Psi(F, G) = \varepsilon \Psi(F_\varepsilon, G_\varepsilon) + (1 - \varepsilon) \Psi(F_{1-\varepsilon}, G_{1-\varepsilon}) \\ &= \varepsilon f(s_0) + (1 - \varepsilon) \Psi(F_{1-\varepsilon}, G_{1-\varepsilon}) = \varepsilon f(s_0) + (1 - \varepsilon) \left(\frac{1 + \varepsilon s_0}{1 - \varepsilon} \right)^p \Psi(\tilde{F}, \tilde{G}) \\ &= \varepsilon f(s_0) + (1 - \varepsilon)(1 - \varepsilon) \left(\frac{1 + \varepsilon s_0}{1 - \varepsilon} \right)^p [\delta f(s_0) + (1 - \delta) \Psi(\tilde{F}_{1-\delta}, \tilde{G}_{1-\delta})] \\ &= 2\varepsilon f(s_0) + (1 + \varepsilon \delta_0) \Psi(F^{(1)}, G^{(1)}) + O(\varepsilon^2). \end{aligned}$$

Therefore, after the N -th iteration (and using the fact that $\Psi(F^{(N)}, G^{(N)}) = f(-1)$), we obtain

$$\begin{aligned} \Psi(F^{(0)}, G^{(0)}) &= (1 + \varepsilon \delta_0)^N \left(f(-1) + \frac{2f(s_0)}{\delta_0} \right) - \frac{2f(s_0)}{\delta_0} + O(\varepsilon) \\ &= e^{\varepsilon \delta_0 N} \left(f(-1) + \frac{2f(s_0)}{\delta_0} \right) - \frac{2f(s_0)}{\delta_0} + O(\varepsilon). \end{aligned} \tag{82}$$

The last equality follows from the fact that $\limsup_{\varepsilon \rightarrow 0} \varepsilon \cdot N(\varepsilon) < \infty$.

Therefore (81) and (82) imply that

$$\begin{aligned} \Psi(F^{(0)}, G^{(0)}) &= \left(\frac{y_3^{(0)} + 2g(s_0)/\delta_0}{g(-1) + 2g(s_0)/\delta_0} \right) \left(f(-1) + \frac{2f(s_0)}{\delta_0} \right) - \frac{2f(s_0)}{\delta_0} + O(\varepsilon) \\ &= f(-1) + (y_3^{(0)} - g(-1)) \left(\frac{f(-1) + 2f(s_0)/\delta_0}{g(-1) + 2g(s_0)/\delta_0} \right) + O(\varepsilon). \end{aligned}$$

Now we recall (77). So, if we show that

$$\frac{f(-1) + 2f(s_0)/\delta_0}{g(-1) + 2g(s_0)/\delta_0} = \frac{f'(-1) - f'(s_0)}{g'(-1) - g'(s_0)}, \tag{83}$$

then (83) will imply that $\Psi(F^{(0)}, G^{(0)}) = \mathbf{B}(-1, y_3^{(0)}) + O(\varepsilon)$. So, choosing ε sufficiently small, we can obtain the extremizer $(F^{(0)}, G^{(0)})$ for the point $(-1, y_3)$. Therefore, we need only to prove equality (83). It will be convenient to use the following notations: set $f_- = f(-1)$, $f'_- = f'(-1)$, $f = f(s_0)$, $f' = f'(s_0)$, $g_- = g(-1)$, $g'_- = g'(-1)$, $g = g(s_0)$ and $g' = g'(s_0)$. Then (83) turns into

$$\frac{\delta_0}{2} = \frac{fg'_- - fg' - f'_-g + f'g}{g'f_- - f'g_-}. \tag{84}$$

This simplifies into

$$s_0 - y_p = \frac{2}{p} \cdot \frac{fg'_- - fg' - f'_-g + f'g}{g'f_- - f'g_-} = \frac{fg'_- - fg' - f'_-g + f'g}{-g'f'_- + f'g'_-},$$

which is true by (53).

Acknowledgements

I would like to express my deep gratitude to Professor A. Volberg, Professor V. I. Vasyunin, and Professor S. V. Kislyakov, my research supervisors, for their patient guidance, enthusiastic encouragement, and useful critiques of this work. I would also like to thank A. Reznikov, for his assistance in finding ε -extremizers for the Bellman function. I would also like to extend my thanks to my colleagues and close friends P. Zatitskiy, N. Osipov, and D. Stolyarov for working together in Saint Petersburg, and developing a theory for minimal concave functions.

Finally, I wish to thank my parents for their support and encouragement throughout my study.

References

- [Bañuelos and Janakiraman 2008] R. Bañuelos and P. Janakiraman, “ L^p -bounds for the Beurling–Ahlfors transform”, *Trans. Amer. Math. Soc.* **360**:7 (2008), 3603–3612. MR 2009d:42032 Zbl 1220.42012
- [Bañuelos and Méndez-Hernández 2003] R. Bañuelos and P. J. Méndez-Hernández, “Space-time Brownian motion and the Beurling–Ahlfors transform”, *Indiana Univ. Math. J.* **52**:4 (2003), 981–990. MR 2004h:60067 Zbl 1080.60043
- [Bañuelos and Osękowski 2013] R. Bañuelos and A. Osękowski, “Burkholder inequalities for submartingales, Bessel processes and conformal martingales”, *Amer. J. Math.* **135**:6 (2013), 1675–1698. MR 3145007 Zbl 1286.60041
- [Bañuelos and Wang 1995] R. Bañuelos and G. Wang, “Sharp inequalities for martingales with applications to the Beurling–Ahlfors and Riesz transforms”, *Duke Math. J.* **80**:3 (1995), 575–600. MR 96k:60108 Zbl 0853.60040
- [Boros et al. 2012] N. Boros, P. Janakiraman, and A. Volberg, “Perturbation of Burkholder’s martingale transform and Monge–Ampère equation”, *Adv. Math.* **230**:4-6 (2012), 2198–2234. MR 2927369 Zbl 1263.60037
- [Burkholder 1984] D. L. Burkholder, “Boundary value problems and sharp inequalities for martingale transforms”, *Ann. Probab.* **12**:3 (1984), 647–702. MR 86b:60080 Zbl 0556.60021
- [Choi 1992] K. P. Choi, “A sharp inequality for martingale transforms and the unconditional basis constant of a monotone basis in $L^p(0, 1)$ ”, *Trans. Amer. Math. Soc.* **330**:2 (1992), 509–529. MR 92f:60073 Zbl 0747.60042
- [Ivanishvili et al. 2012a] P. Ivanishvili, N. N. Osipov, D. M. Stolyarov, V. I. Vasyunin, and P. B. Zatitskiy, “Bellman function for extremal problems in BMO”, 2012. To appear in *Trans. AMS*. arXiv 1205.7018v3
- [Ivanishvili et al. 2012b] P. Ivanishvili, N. N. Osipov, D. M. Stolyarov, V. I. Vasyunin, and P. B. Zatitskiy, “On Bellman function for extremal problems in BMO”, *C. R. Math. Acad. Sci. Paris* **350**:11-12 (2012), 561–564. MR 2956143 Zbl 1247.42018
- [Ivanishvili et al. \geq 2015] P. Ivanishvili, N. N. Osipov, D. M. Stolyarov, V. I. Vasyunin, and P. B. Zatitskiy, “Bellman function for extremal problems on BMO, II: Evolution”. In preparation.
- [Reznikov et al. 2013] A. Reznikov, V. Vasyunin, and V. Volberg, “Extremizers and Bellman function for martingale weak type inequality”, preprint, 2013. arXiv 1311.2133
- [Slavin and Vasyunin 2011] L. Slavin and V. Vasyunin, “Sharp results in the integral-form John–Nirenberg inequality”, *Trans. Amer. Math. Soc.* **363**:8 (2011), 4135–4169. MR 2012c:42005 Zbl 1223.42001
- [Vasyunin and Volberg 2010] V. Vasyunin and A. Volberg, “Burkholder’s function via Monge–Ampère equation”, *Illinois J. Math.* **54**:4 (2010), 1393–1428. MR 2981853 Zbl 1259.42016

Received 8 Mar 2014. Revised 1 Feb 2015. Accepted 25 Mar 2015.

PAATA IVANISVILI: ivanisvi@math.msu.edu

Department of Mathematics, Michigan State University, 619 Red Cedar Road, East Lansing, MI 48824, United States

CLASSIFICATION OF BLOWUP LIMITS FOR SU(3) SINGULAR TODA SYSTEMS

CHANG-SHOU LIN, JUN-CHENG WEI AND LEI ZHANG

For singular SU(3) Toda systems, we prove that the limit of energy concentration is a finite set. In addition, for fully bubbling solutions we use a Pohozaev identity to prove a uniform estimate. Our results extend previous results of Jost, Lin and Wang on regular SU(3) Toda systems.

1. Introduction

Systems of elliptic equations in two-dimensional space with exponential nonlinearity are very commonly observed in physics, geometry, chemistry and biology. In this article we consider the following general system of equations defined in \mathbb{R}^2 :

$$\Delta u_i + \sum_{j \in I} a_{ij} h_j e^{u_j} = 4\pi \gamma_i \delta_0 \quad \text{in } B_1 \subset \mathbb{R}^2 \quad \text{for } i \in I, \quad (1-1)$$

where $I = \{1, \dots, n\}$, B_1 is the unit ball in \mathbb{R}^2 , h_1, \dots, h_n are smooth functions, $A = (a_{ij})_{n \times n}$ is a constant matrix, $\gamma_i > -1$ and δ_0 is the Dirac mass at 0. If $n = 1$ and $a_{11} = 1$, the system (1-1) is reduced to a single Liouville equation, which has vast background in conformal geometry and physics. The general system (1-1) is used for many models in different disciplines of science. If the coefficient matrix A is nonnegative, symmetric and irreducible, (1-1) is called a Liouville system and is related to models in the theory of chemotaxis [Childress and Percus 1981; Keller and Segel 1971], in the physics of charged particle beams [Bennet 1934; Debye and Huckel 1923; Kiessling and Lebowitz 1994] and in the theory of semiconductors [Mock 1975]; see [Chanillo and Kiessling 1995; Chipot et al. 1997; Lin and Zhang 2010] and the references therein for more applications of Liouville systems. If A is the Cartan matrix

$$A_n = \begin{pmatrix} 2 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 2 & -1 \\ 0 & \cdots & 0 & 0 & -1 & 2 \end{pmatrix},$$

the system (1-1) is called an SU($n+1$) Toda system (which has n equations) and is related to the nonabelian gauge in Chern–Simons theory; see [Dunne et al. 1991; Dunne 1995; Ganoulis et al. 1982; Leznov 1980; Leznov and Saveliev 1992; Malchiodi and Ndiaye 2007; Malchiodi and Ruiz 2013; Mansfield

MSC2010: primary 35J60; secondary 35J55.

Keywords: SU($n+1$)-Toda system, asymptotic analysis, a priori estimate, classification theorem, topological degree, blowup solutions.

1982; Nolasco and Tarantello 1999; 2000; Yang 1997; 2001] and the references therein. There are also many works on the relationship between $SU(n+1)$ Toda systems and holomorphic curves in $\mathbb{C}\mathbb{P}^n$, the flat $SU(n+1)$ connection, complete integrability and harmonic sequences; see [Bolton and Woodward 1997; Bolton et al. 1988; Calabi 1953; Chern and Wolfson 1987; Doliwa 1997; Guest 1997; Leznov and Saveliev 1992; Lin et al. 2012a] for references.

After decades of extensive study, many important questions related to the scalar Liouville equation are answered and the behavior of blowup solutions is well understood (see [Bartolucci and Tarantello 2002a; 2002b; Bartolucci and Malchiodi 2013; Chen and Lin 2002; 2003] for related discussions). However, the understanding of blowup solutions to the more general systems (1-1) is far from complete. In recent years, much progress has been made on more general systems and we only mention a few works related to the topic of the current article. First, Lin and Zhang [2010; 2011] completed a degree-counting project for Liouville systems defined on Riemann surfaces. Second, for regular $SU(3)$ Toda systems (which have two equations), Jost, Lin and Wang [Jost et al. 2006] proved some uniform estimates for fully bubbling solutions (see Section 4 for the definition) using holonomy theory. Later, Lin, Wei and Zhao [Lin et al. 2012b] improved the estimate of [Jost et al. 2006] to the sharp form using the nondegeneracy of the global $SU(3)$ solutions, which was established by Lin, Wei and Ye [Lin et al. 2012a] among other things.

In this article we mainly focus on the asymptotic behavior of blowup solutions of (1-1) and the weak limit of energy concentration for $SU(n+1)$ Toda systems. More specifically, let $u^k = (u_1^k, \dots, u_n^k)$ be a sequence of solutions

$$\Delta u_i^k + \sum_{j=1}^n a_{ij} h_j^k e^{u_j^k} = 4\pi \gamma_i^k \delta_0 \quad \text{in } B_1, \quad i = 1, \dots, n, \quad (1-2)$$

with 0 being its only possible blowup point in B_1 :

$$\max_{K \in B_1 \setminus \{0\}} u_i^k \leq C(K). \quad (1-3)$$

Since the right-hand side of (1-2) is a Dirac mass, we define the regular part of u_i^k to be

$$\tilde{u}_i^k(x) = u_i^k(x) - 2\gamma_i^k \log|x|, \quad x \in B_1, \quad i = 1, \dots, n. \quad (1-4)$$

Then $u^k = (u_1^k, \dots, u_n^k)$ is called a *sequence of blowup solutions* if $\max_i \max_{x \in B_1} \tilde{u}_i^k \rightarrow \infty$.

We assume that $\gamma_i^k \rightarrow \gamma_i > -1$, that h_1^k, \dots, h_n^k are positive smooth functions with a uniform bound on their C^3 norm:

$$\frac{1}{C} \leq h_i^k \leq C, \quad \|h_i^k\|_{C^3(B_1)} \leq C \quad \text{in } B_1, \quad \gamma_i^k \rightarrow \gamma_i > -1 \quad \text{for all } i \in I; \quad (1-5)$$

and we suppose that there is a uniform bound on the oscillation of u_i^k on ∂B_1 and its energy, $\int_{B_1} h_i^k e^{u_i^k}$:

$$|u_i^k(x) - u_i^k(y)| \leq C \quad \text{for all } x, y \in \partial B_1, \quad \int_{B_1} h_i^k e^{u_i^k} \leq C, \quad i \in I, \quad (1-6)$$

where C is independent of k .

Note that the oscillation finiteness assumption in (1-6) is natural and generally satisfied in most applications. The energy bound in (1-6) is also natural for a system or equation defined in two-dimensional space.

If $A = A_2$, (1-2) describes SU(3) with sources. Our first main theorem is concerned with the energy limits of solutions to singular SU(3) Toda systems.

Given any $\delta > 0$, u^k has no blowup point in $B_1 \setminus B_\delta$ (in this article we use $B(x, r)$ to denote a ball centered at x with radius r and use B_r to denote $B(0, r)$). Thus we are interested in the following limit:

$$\sigma_i = \lim_{\delta \rightarrow 0} \lim_{k \rightarrow \infty} \frac{1}{2\pi} \int_{B_\delta} h_i^k e^{u_i^k}, \quad i = 1, 2. \tag{1-7}$$

Since, for each $\delta > 0$, $\int_{B_\delta} h_i^k e^{u_i^k}$ is uniformly bounded, the $\lim_{k \rightarrow \infty}$ in (1-7) is understood as the limit of a subsequence of u^k . For convenience we don't distinguish u^k and its subsequences in this article.

Let

$$\mu_i = 1 + \gamma_i, \quad i = 1, 2,$$

and let

$$\Gamma = \{(\sigma_1, \sigma_2) : \sigma_1, \sigma_2 \geq 0, \sigma_1^2 - \sigma_1\sigma_2 + \sigma_2^2 = 2\mu_1\sigma_1 + 2\mu_2\sigma_2\}$$

be a quadratic curve in the first quadrant. It is easy to see that Γ is contained in the box

$$\left[0, \frac{4}{3}\mu_1 + \frac{2}{3}\mu_2 + \frac{4}{3}\sqrt{\mu_1^2 + \mu_1\mu_2 + \mu_2^2}\right] \times \left[0, \frac{2}{3}\mu_1 + \frac{4}{3}\mu_2 + \frac{4}{3}\sqrt{\mu_1^2 + \mu_1\mu_2 + \mu_2^2}\right].$$

In Definition 1.1 below we shall define a finite set on Γ . In order to describe the relative positions of points, we say (c, d) is in the upper right part of (a, b) if $c \geq a$ and $d \geq b$.

Definition 1.1. It is easy to verify that the following six points are on Γ :

$$(0, 0) \quad (2\mu_1, 0), \quad (0, 2\mu_2), \quad (2\mu_1, 2(\mu_1 + \mu_2)), \quad (2(\mu_1 + \mu_2), 2\mu_2), \quad (2(\mu_1 + \mu_2), 2(\mu_1 + \mu_2)).$$

First we let the six points above belong to Σ , then we determine other points in Σ as follows: For $(a, b) \in \Sigma$, intersect Γ with $\sigma_1 = a + 2N$ and $\sigma_2 = b + 2N$ ($N = 0, 1, 2, \dots$) and add the point(s) of intersection to Σ that belong to the upper right part of (a, b) . For each new member $(c, d) \in \Sigma$ added by this process, we apply the same procedure based on (c, d) to obtain possible new members.

Theorem 1.2. Let $A = A_2$, h_i^k and γ_i^k satisfy (1-5). Then, for u^k satisfying (1-2), (1-3) and (1-6), we have $(\sigma_1, \sigma_2) \in \Sigma$, where σ_i is defined by (1-7) and Σ is defined as in Definition 1.1.

Remark 1.3. If $\gamma_1 = \gamma_2 = 0$, the system is a nonsingular SU(3) Toda system. One sees easily that

$$\Sigma = \{(0, 0), (2, 0), (0, 2), (2, 4), (4, 2), (4, 4)\}.$$

Indeed, when the procedure described in Definition 1.1 is applied to any of the six points in Σ , no extra point of intersection can be found. For example if we start from $(0, 0)$ and intersect Γ by lines $\sigma_1 = 2N$ (N being a nonnegative integer), then we see immediately that the intersection of Γ with $\sigma_1 = 2$ gives $(2, 0)$ and $(2, 4)$, which are already in Σ . The intersection with $\sigma_1 = 4$ gives $(4, 2)$ and $(4, 4)$, which also belong to the six types in Σ . There is no intersection between Γ and $\sigma_1 = 6$. Theorem 1.2 in this special

case was proved in [Jost et al. 2006]. Recent work of Pistoia, Musso and Wei [Musso et al. 2015] proved that all six cases for nonsingular $SU(3)$ Toda systems can occur.

Remark 1.4. It is easy to observe that the maximum value of σ_1 on Γ is

$$\frac{4}{3}\mu_1 + \frac{2}{3}\mu_2 + \frac{4}{3}\sqrt{\mu_1^2 + \mu_1\mu_2 + \mu_2^2}.$$

The maximum value of σ_2 is

$$\frac{2}{3}\mu_1 + \frac{4}{3}\mu_2 + \frac{4}{3}\sqrt{\mu_1^2 + \mu_1\mu_2 + \mu_2^2}.$$

Thus Σ is a finite set. As two special cases, we see that:

(1) If

$$\frac{4}{3}\mu_1 + \frac{2}{3}\mu_2 + \frac{4}{3}\sqrt{\mu_1^2 + \mu_1\mu_2 + \mu_2^2} < 2 \quad \text{and} \quad \frac{2}{3}\mu_1 + \frac{4}{3}\mu_2 + \frac{4}{3}\sqrt{\mu_1^2 + \mu_1\mu_2 + \mu_2^2} < 2$$

then there are only six points in Σ :

$$\Sigma = \{(0, 0), (2\mu_1, 0), (0, 2\mu_2), (2(\mu_1 + \mu_2), 2\mu_2), (2\mu_1, 2(\mu_1 + \mu_2)), (2(\mu_1 + \mu_2), 2(\mu_1 + \mu_2))\}.$$

(2) For $\gamma_1 = \gamma_2 = 1$, in addition to $(0, 0)$, $(4, 0)$, $(0, 4)$, $(4, 8)$, $(8, 4)$ and $(8, 8)$, Σ has other 14 points.

An earlier version of the current article was posted on the arXiv in March 2013. After that, some work has been done based on Theorem 1.2 (see [Battaglia and Malchiodi 2014] for example). Theorem 1.2 reflects some essential differences between Toda systems and Liouville systems. Lin et al. [2012a] proved that all the global solutions of $SU(n+1)$ Toda systems can be described by $n^2 + 2n$ parameters and the energy of global solutions is a discrete set. On the other hand, the global solutions of Liouville systems all belong to a family of three parameters but their energy forms an $(n-1)$ -dimensional hypersurface (see [Chipot et al. 1997; Lin and Zhang 2010]). These differences lead to very different approaches in their respective research. For example, [Lin et al. 2012b] obtained sharp estimates for fully bubbling solutions (see Section 4 for the definition) of $SU(3)$ Toda systems using the discreteness of energy as a key ingredient in their proof.

Here we briefly describe the strategy used to prove Theorem 1.2. First we introduce a selection process suitable for $SU(n+1)$ Toda systems. The selection process has been widely used for prescribing curvature-type equations (see [Li 1995; Chen and Lin 1998], etc) and we modify it to locate the bubbling area, which is a union of finite disks. In each of the disks, the blowup solutions have roughly the energy of a global $SU(m+1)$ Toda system on \mathbb{R}^2 (with $m \leq n$), which is the limit of the blowup solutions after scaling. If $m = n$, which means no component is lost after scaling and taking the limit, we say the sequence of solutions in the disk is fully bubbling, otherwise we call it partially bubbling. Next we introduce the “group” concept to place bubbling disks according to their relative locations. There are only finitely many bubbling disks and their relative distances may tend to 0 with very different speed. The name “group” is used to describe a few disks that are roughly closest to one another and much further from other disks. Lemma 2.4 is a Harnack-type result that plays an important role in determining the energy concentration around a group. Suppose there is a circle that surrounds a group and both components of the blowup solutions have fast decay (see Section 3 for the definition) on the circle. Then a Pohozaev

identity can be computed on this circle to determine how much energy this group carries. Because of Lemma 2.4, such a circle can always be found, so the energy within the circle can be determined. Then we consider the combination of groups by scaling. The relationship among groups is similar to that of members in a same group. For example, if the distance between two groups is scaled to be 1, the bubbling disks of one group look like a Dirac mass from afar. We can similarly find circles surrounding groups that are also suitable for computing Pohozaev identities (i.e., both components of the blowup solutions have fast decay on these circles). From these Pohozaev identities, we determine how much energy is contained in each group and all the combinations of groups. One important fact is that one component of the blowup solutions always has fast decay, even though the other component may not. It is possible for the first (fast decay) component to turn to a slow decay component as the distance to a group becomes bigger, but before that happens the second component, which used to be a slow decay component, will turn to a fast decay component first.

As another application of the Pohozaev identity we establish some uniform estimates for fully bubbling solutions. These estimates were first obtained by Li [1999] for the scalar Liouville equation without singularity (using the method of moving planes) and [Bartolucci et al. 2004] for the scalar Liouville equation with singularity (using the Pohozaev identity and potential analysis). For regular SU(3) Toda systems, [Jost et al. 2006] established similar estimates using holonomy theory. Our results (Theorem 4.1 and Theorem 4.3) apply to general SU($n + 1$) Toda systems with singularity.

This article is set out as follows. In Section 2 we introduce the selection process mentioned before and in Section 3 we prove the Pohozaev identity, which is crucial for the proof of Theorem 1.2. In Section 4 we prove a uniform estimate for fully bubbling solutions (Theorem 4.3 and Theorem 4.1). Then in Section 5 and Section 6 we finish the proof of Theorem 1.2 according to the strategy mentioned before.

2. A selection process for SU($n + 1$) Toda systems

Clearly in the proof of Theorem 1.2 we can assume 0 to be a blowup point:

$$\max_{x \in B_1, i \in I} \{u_i^k - 2\gamma_i^k \log |x|\} \rightarrow \infty, \tag{2-1}$$

because otherwise the blowup type is (0, 0). So, from now on throughout the paper, (2-1) is assumed.

Case one: $\gamma_1^k = \dots = \gamma_n^k = 0$.

Proposition 2.1. *Let $A = (a_{ij})_{n \times n}$ be the Cartan matrix A_n , h_i^k satisfy (1-5) and $u^k = (u_1^k, \dots, u_n^k)$ be a sequence of solutions to (1-2) with $\gamma_1^k = \dots = \gamma_n^k = 0$ such that (1-6) and (1-3) hold. Then there exist finite sequences of points $\Sigma_k := \{x_1^k, \dots, x_m^k\}$ (all $x_j^k \rightarrow 0, j = 1, \dots, m$) and positive numbers $l_1^k, \dots, l_m^k \rightarrow 0$ such that the following four properties hold:*

- (1) $\max_{i \in I} \{u_i^k(x_j^k)\} = \max_{B(x_j^k, l_j^k), i \in I} \{u_i^k\}$ for all $j = 1, \dots, m$.
- (2) $\exp(\frac{1}{2} \max_{i \in I} \{u_i^k(x_j^k)\}) l_j^k \rightarrow \infty, j = 1, \dots, m$.
- (3) There exists $C_1 > 0$ independent of k such that

$$u_i^k(x) + 2 \log \text{dist}(x, \Sigma_k) \leq C_1 \quad \text{for all } x \in B_1, \quad i \in I,$$

where dist stands for distance.

(4) In each $B(x_j^k, l_j^k)$ let

$$v_i^k(y) = u_i^k(\epsilon_k y + x_j^k) + 2 \log \epsilon_k, \quad \epsilon_k = e^{-M_k/2}, \quad M_k = \max_i \max_{B(x_j^k, l_j^k)} u_i^k. \quad (2-2)$$

Then one of the following two alternatives holds:

(a) The sequence is fully bubbling: along a subsequence, (v_1^k, \dots, v_n^k) converges in $C_{\text{loc}}^2(\mathbb{R}^2)$ to (v_1, \dots, v_n) which satisfies

$$\begin{aligned} \Delta v_i + \sum_{j \in I} a_{ij} h_j e^{v_j} &= 0 \quad \text{in } \mathbb{R}^2, \quad i \in I, \\ \lim_{k \rightarrow \infty} \int_{B(x_j^k, l_j^k)} \sum_{i \in I} a_{it} h_t^k e^{u_i^k} &> 4\pi, \quad i \in I. \end{aligned}$$

(b) $I = J_1 \cup J_2 \cup \dots \cup J_m \cup N$, where J_1, J_2, \dots, J_m and N are disjoint sets, $N \neq \emptyset$ and each J_t ($t = 1, \dots, m$) consists of consecutive indices. For each $i \in N$, v_j^k tends to $-\infty$ over any fixed compact subset of \mathbb{R}^2 . The components of $v^k = (v_1^k, \dots, v_n^k)$ corresponding to each J_l ($l = 1, \dots, m$) converge in $C_{\text{loc}}^2(\mathbb{R}^2)$ to an $\text{SU}(|J_l| + 1)$ Toda system, where $|J_l|$ is the number of indices in J_l . For each $i \in J_l$, we have

$$\lim_{k \rightarrow \infty} \int_{B(x_j^k, l_j^k)} \sum_{i \in J_l} a_{it} h_t^k e^{v_i^k} > 4\pi.$$

Remark 2.2. In this article we don't use different notations for sequences and subsequences.

Remark 2.3. For each $x_j^k \in \Sigma_k$, suppose $2t_j^k$ is the distance from x_j^k to $\Sigma_k \setminus \{x_j^k\}$. Then $t_j^k/l_j^k \rightarrow \infty$ as $k \rightarrow \infty$ if l_j^k is suitably chosen.

Proof of Proposition 2.1. Without loss of generality we assume

$$u_1^k(x_1^k) = \max_{i \in I, x \in B_1} u_i^k(x).$$

Clearly $x_1^k \rightarrow 0$, because $\max_i \max_{x \in B_1} u_i^k \rightarrow \infty$ and u^k is uniformly bounded from above away from the origin. Let (v_1^k, \dots, v_n^k) be defined by (2-2) with x_j^k replaced by x_1^k . Immediately we observe that $|\Delta v_i^k|$ is bounded because each $v_i^k \leq 0$. Consequently, $|v_i^k(z) - v_i^k(0)|$ is uniformly bounded in any compact subset of \mathbb{R}^2 . Thus, since $v_1^k(0) = 0$, (along a subsequence) v_1^k converges in $C_{\text{loc}}^2(\mathbb{R}^2)$ to a function v_1 . For the other components of $v^k = (v_1^k, \dots, v_n^k)$, either some of them tend to $-\infty$ over any compact subset of \mathbb{R}^2 , or all of them converge to a system of n equations. Let $J \subset I$ be the set of indices corresponding to those convergent components. That is, for all $i \in J$, v_i^k converges to v_i in $C_{\text{loc}}^2(\mathbb{R}^2)$ and, for all $j \in I \setminus J$, v_j^k tends to $-\infty$ over any fixed compact subset of \mathbb{R}^2 . For each $i \in I \setminus J$, there is $J_1 \subset J$ such that $i \in J_1$, the indices in J_1 are consecutive and the limit of the v_i^k is one component of an $\text{SU}(|J_1| + 1)$ Toda system:

$$\begin{cases} \Delta v_m + \sum_{j \in J} a_{mj} h_j e^{v_j} = 0 & \text{in } \mathbb{R}^2 \text{ for all } m \in J_1 \\ \int_{\mathbb{R}^2} h_m e^{v_m} \leq C, & m \in J_1, \end{cases} \quad (2-3)$$

where $h_m = \lim_{k \rightarrow \infty} h_m^k(x_1^k)$, $(a_{ij}) = A_{|J_1|}$, and C is the same constant as in (1-6). By the classification theorem of [Lin et al. 2012a] (if the limit is a system) or [Chen and Li 1991] (if the limit is one equation) we have

$$\sum_{j \in J_1} \int_{\mathbb{R}^2} a_{ij} h_j e^{v_j} = 8\pi \quad \text{for all } i \in J_1 \tag{2-4}$$

and

$$v_i(x) = -4 \log |x| + O(1), \quad |x| > 2, \quad \text{for all } i \in J_1. \tag{2-5}$$

Thus, for any index $i \in I$, we can find $R_k \rightarrow \infty$ such that

$$v_i^k(y) + 2 \log |y| \leq C, \quad |y| \leq R_k, \quad \text{for } i \in I. \tag{2-6}$$

Equivalently, for u^k there exist $l_1^k \rightarrow 0$ such that

$$u_i^k(x) + 2 \log |x - x_1^k| \leq C, \quad |x - x_1^k| \leq l_1^k, \quad \text{for } i \in I$$

and

$$e^{u_1^k(x_1^k)/2} l_1^k \rightarrow \infty \quad \text{as } k \rightarrow \infty, \quad i \in J.$$

Next, we let q_k be the maximum point of $\max_{|x| < 1, i \in I} u_i^k(x) + 2 \log |x - x_1^k|$. If

$$\max_{|x| \leq 1, i \in I} u_i^k(x) + 2 \log |x - x_1^k| \rightarrow \infty,$$

we let j be the index such that

$$u_j^k(q_k) + 2 \log |q_k - x_1^k| = \max_{i \in I} u_i^k(x) + 2 \log |x - x_1^k| \rightarrow \infty.$$

The following localization is to adapt the original argument of R. Schoen [1988] for the scalar curvature equation (also see [Chen and Lin 1998]). Set

$$d_k = \frac{1}{2} |q_k - x_1^k|$$

and

$$S_i^k(x) = u_i^k(x) + 2 \log(d_k - |x - q_k|) \quad \text{in } B(q_k, d_k).$$

Then clearly, for fixed k , $S_i^k \rightarrow -\infty$ as x tends to $\partial B(q_k, d_k)$. On the other hand, at least for j , we have

$$S_j^k(q_k) = u_j^k(q_k) + 2 \log d_k \rightarrow \infty.$$

Let p_k be where

$$\max_i \max_{x \in \bar{B}(q_k, d_k)} S_i^k$$

is attained and i_0 be the index corresponding to where the maximum is taken:

$$u_{i_0}^k(p_k) + 2 \log(d_k - |p_k - q_k|) \geq S_j^k(q_k) \rightarrow \infty. \tag{2-7}$$

Let

$$l_k = \frac{1}{2} (d_k - |p_k - q_k|).$$

Then for $y \in B(p_k, l_k)$, by the choice of p_k and l_k , we have

$$u_i^k(y) + 2 \log(d_k - |y - q_k|) \leq u_{i_0}^k(p_k) + 2 \log(2l_k) \quad \text{for all } i \in I.$$

On the other hand, by the definition of l_k , we have

$$d_k - |y - q_k| \geq d_k - |p_k - q_k| - |y - p_k| \geq l_k \quad \text{if } |y - p_k| < l_k,$$

and

$$u_i^k(y) \leq u_{i_0}^k(p_k) + 2 \log 2 \quad \text{for all } y \in B(p_k, l_k). \quad (2-8)$$

Next, we set

$$\mathcal{R}_k = e^{u_{i_0}^k(p_k)/2} l_k \quad (2-9)$$

and scale u_i^k by

$$\tilde{v}_i^k(y) = u_i^k(p_k + e^{-u_{i_0}^k(p_k)/2} y) - u_{i_0}^k(p_k) \quad \text{for } i \in I.$$

From (2-7) we clearly have $\mathcal{R}_k \rightarrow \infty$. By (2-8) and standard elliptic estimates for the Laplacian, \tilde{v}_i^k is bounded in $C_{\text{loc}}^2(\mathbb{R}^2)$ and there exists $\emptyset \neq J \subset I$ such that, for all $i \in J$, \tilde{v}_i^k converges to a limit system like (2-3). On the other hand, \tilde{v}_i^k converges uniformly to $-\infty$ over all compact subsets of \mathbb{R}^2 for all $i \in I \setminus J$. Clearly (2-6) holds for \tilde{v}_i^k . Going back to u^k , we have

$$u_i^k(x) + 2 \log |x - x_2^k| \leq C \quad \text{for } |x - x_2^k| \leq l_2^k,$$

where x_2^k is the point where $\max_i \max_{B(p_k, l_2^k)} u_i^k$ is attained and $l_2^k = l_k$. Here we note that x_2^k is neither q_k nor p_k and the distance between p_k and x_2^k is small: $e^{u_{i_0}^k(p_k)/2} |x_2^k - p_k| = O(1)$. If we rescale u^k around x_2^k , then v^k defined as in (2-2) satisfies (a) and (b) in Proposition 2.1. Clearly $B(x_1^k, l_1^k) \cap B(x_2^k, l_2^k) = \emptyset$.

To continue with the selection process, we let $\Sigma_{k,2} := \{x_1^k, x_2^k\}$ and consider

$$\max_{i \in I, x \in B_1} u_i^k(x) + 2 \log \text{dist}(x, \Sigma_{k,2}).$$

If, along a subsequence, the quantity above tends to infinity, we apply the same procedure to get x_3^k and l_3^k . After each selection we add a new disjoint disk, say $B(x_m^k, l_m^k)$, in which the profile of bubbling solutions is like that of a global system, so from (2-4) we see that

$$\int_{B(x_m^k, l_m^k)} \sum_i h_i^k e^{u_i^k} \geq C \quad \text{for some } C > 0 \text{ independent of } k.$$

Therefore by (1-6) the process stops after finitely many steps and we have

$$u_i^k(x) + 2 \log d(x, \Sigma_k) \leq C, \quad i \in I. \quad (2-10)$$

Proposition 2.1 is established. \square

2.1. Case two: the singular case $\exists \gamma_i \neq 0$. First, the selection process is almost the same. The difference is instead of taking the maximum of u_i^k over B_1 we require

$$0 \in \Sigma_k.$$

Clearly, in $B_1 \setminus \{0\}$, u^k satisfies the same equation as the nonsingular case. Then we consider the maximum of $u_i^k(x) + 2 \log \text{dist}(x, \Sigma_k) = u_i^k(x) + 2 \log |x|$ and the selection proceeds the same as before. Therefore, in the singular case, $\Sigma_k = \{0, x_1^k, \dots, x_m^k\}$.

Lemma 2.4. *Let Σ_k be the blowup set (thus, if $\gamma_i^k = 0$ for all i , $\Sigma_k = \{x_1^k, \dots, x_m^k\}$, and if the system is singular, $\Sigma_k = \{0, x_1^k, \dots, x_m^k\}$). In either case, for all $x_0 \in B_1 \setminus \Sigma_k$, there exists C_0 independent of x_0 and k such that*

$$|u_i^k(x_1) - u_i^k(x_2)| \leq C_0 \quad \text{for all } x_1, x_2 \in B(x_0, \frac{1}{2}d(x_0, \Sigma_k)) \quad \text{for all } i \in I.$$

Proof. We can assume $|x| < \frac{1}{10}$ because it is easy to see from Green's representation formula that the oscillation of u_i^k on $B_1 \setminus B_{1/10}$ is finite. Recall the regular part of u_i^k is defined in (1-4) and \tilde{u}_i^k satisfies

$$\Delta \tilde{u}_i^k(x) + \sum_{j \in I} a_{ij} h_j^k(x) |x|^{2\gamma_j^k} e^{\tilde{u}_j^k(x)} = 0 \quad \text{in } B_1, \quad i \in I.$$

Let σ_k be the distance between x_0 and Σ_k . Clearly, for $x_0 \in B_1 \setminus \Sigma_k$ and $x_1, x_2 \in B(x_0, \frac{1}{2}d(x_0, \Sigma_k))$,

$$u_i^k(x_1) - u_i^k(x_2) = \tilde{u}_i^k(x_1) - \tilde{u}_i^k(x_2) + O(1) = \int_{B_1} (G(x_1, \eta) - G(x_2, \eta)) \sum_{j \in I} a_{ij} h_j^k(\eta) |\eta|^{2\gamma_j^k} e^{\tilde{u}_j^k(\eta)} d\eta + O(1).$$

Here G is the Green's function on B_1 . The last term on the above is $O(1)$ because it is the difference of two points of a harmonic function that has bounded oscillation on ∂B_1 . Since both $x_1, x_2 \in B_{1/10}$, it is easy to use the uniform bound on the energy (1-6) to obtain

$$\int_{B_1} (\gamma(x_1, \eta) - \gamma(x_2, \eta)) \sum_{j \in I} a_{ij} h_j^k(\eta) |\eta|^{2\gamma_j^k} e^{\tilde{u}_j^k(\eta)} d\eta = O(1),$$

where $\gamma(\cdot, \cdot)$ is the regular part of G . Therefore, we only need to show

$$\int_{B_1} \log \frac{|x_1 - \eta|}{|x_2 - \eta|} \sum_j a_{ij} h_j^k |\eta|^{2\gamma_j} e^{\tilde{u}_j} d\eta = O(1).$$

If $\eta \in B_1 \setminus B(x_0, \frac{3}{4}\sigma_k)$, we have $\log(|x_1 - \eta|/|x_2 - \eta|) = O(1)$, then the integration over $B_1 \setminus B(x_0, \frac{3}{4}\sigma_k)$ is uniformly bounded. Therefore, we only need to show

$$\int_{B(x_0, 3\sigma_k/4)} \log \frac{|x_1 - \eta|}{|x_2 - \eta|} \sum_j a_{ij} h_j^k |\eta|^{2\gamma_j} e^{\tilde{u}_j} d\eta = \int_{B(x_0, 3\sigma_k/4)} \log \frac{|x_1 - \eta|}{|x_2 - \eta|} \sum_j a_{ij} h_j^k e^{u_j^k} d\eta = O(1).$$

To this end, let

$$v_i^k(y) = u_i^k(x_0 + \sigma_k y) + 2 \log \sigma_k, \quad y \in B_{3/4}, \quad i \in I. \tag{2-11}$$

Then we just need to show

$$\int_{B_{3/4}} \log \frac{|y_1 - \eta|}{|y_2 - \eta|} \sum_j a_{ij} h_j^k(x_0 + \sigma_k \eta) e^{v_j^k(\eta)} d\eta = O(1). \quad (2-12)$$

We assume, without loss of generality, that e_1 is the image of the closest blowup point in Σ_k . Thus, by the selection process,

$$v_i^k(\eta) \leq -2 \log |\eta - e_1| + C.$$

Therefore,

$$e^{v_i^k(\eta)} \leq C |\eta - e_1|^{-2}.$$

With this estimate, we observe that $|\eta - e_1| \geq C > 0$ for $\eta \in B_{3/4}$. Thus, for $j = 1, 2$ and any fixed $i \in I$,

$$\int_{B_{3/4}} |\log |y_j - \eta|| e^{v_i^k(\eta)} d\eta \leq C \int_{B_{3/4}} \frac{|\log |y_j - \eta||}{|\eta - e_1|^2} d\eta \leq C.$$

Lemma 2.4 is established. \square

Remark 2.5. For systems with nonnegative coefficient matrix A , the selection process can also be applied. See [Chen and Li 1993] or [Lin and Zhang 2010] for more details.

3. Pohozaev identity and related estimates on the energy

In this section we derive a Pohozaev identity for u^k satisfying (1-2), (1-3) and (1-6), h_i^k and γ_i^k satisfying (1-5), and $A = A_n$.

Proposition 3.1. *Let $A = A_n$, σ_i be defined by (1-7). Suppose $u^k = (u_1^k, \dots, u_n^k)$ satisfy (1-2), (1-6), (1-3) and (2-1), h^k and γ_i^k satisfy (1-5). Then we have*

$$\sum_{i,j \in I} a_{ij} \sigma_i \sigma_j = 4 \sum_{i=1}^n (1 + \gamma_i) \sigma_i.$$

Proof. We start with a lemma:

Lemma 3.2. *Given any $\epsilon_k \rightarrow 0$ such that $\Sigma_k \subset B(0, \frac{1}{2}\epsilon_k)$, there exist $l_k \rightarrow 0$ satisfying $l_k \geq 2\epsilon_k$ and*

$$\bar{u}_i^k(l_k) + 2 \log l_k \rightarrow -\infty \quad \text{for all } i \in I, \quad \text{where } \bar{u}_i^k(r) := \frac{1}{2\pi r} \int_{\partial B_r} u_i^k. \quad (3-1)$$

Remark 3.3. By Lemma 3.2 and Lemma 2.4,

$$u_i^k(x) + 2 \log |x| \rightarrow -\infty \quad \text{for all } i \in I \text{ and } x \in \partial B_{l_k}.$$

This is crucial for evaluating the \mathcal{R}_1 term (the first term on the right) of (3-7) below.

Proof of Lemma 3.2. Since $\Sigma_k \subset B(0, \frac{1}{2}\epsilon_k)$, we have, by Proposition 2.1(3),

$$u_i^k(x) + 2 \log |x| \leq C, \quad |x| \geq \epsilon_k. \quad (3-2)$$

The key point of the argument below is that we can always use the finite energy assumption and Lemma 2.4 to make u_1^k satisfy (3-1). Then we can adjust the radius to make the other components satisfy (3-1) as well.

First we observe that, for each fixed i , there exists $r_{k,i} \geq \epsilon_k$ such that

$$\bar{u}_i^k(r_{k,i}) + 2 \log r_{k,i} \rightarrow -\infty, \tag{3-3}$$

because otherwise we would have

$$\bar{u}_i^k(r) + 2 \log r \geq -C \quad \text{for all } r \geq \epsilon_k$$

for some $C > 0$. By Lemma 2.4, u_i^k has bounded oscillation on each ∂B_r . Thus

$$u_i^k(x) + 2 \log |x| \geq -C \quad \text{for all } x \in \partial B_r, \quad \epsilon_k < r < 1$$

for some C . Then

$$e^{u_i^k(x)} \geq C|x|^{-2}, \quad \epsilon_k \leq |x| \leq 1.$$

Integrating $e^{u_i^k}$ on $B_1 \setminus B_{\epsilon_k}$, we get a contradiction on the uniform energy bound of $\int_{B_1} h_i^k e^{u_i^k}$. Thus (3-3) is established.

Now, for u_1^k , we find $r_{k,1} \geq \epsilon_k$ so that

$$\bar{u}_1^k(r_{k,1}) + 2 \log r_{k,1} \rightarrow -\infty.$$

Here we claim that we can assume $r_{k,1} \rightarrow 0$ as well. In fact, if $r_{k,1}$ does not tend to 0, by Lemma 2.4

$$\bar{u}_1^k(r) + 2 \log r \leq -N_k + C, \quad \frac{1}{2}r_{k,1} < r < r_{k,1},$$

where $N_k \rightarrow \infty$ and satisfies

$$\bar{u}_1^k(r_{k,1}) + 2 \log r_{k,1} \leq -N_k.$$

Using Lemma 2.4 again we have

$$\bar{u}_1^k(r) + 2 \log r \leq -N_k + C, \quad \frac{1}{4}r_{k,1} < r < \frac{1}{2}r_{k,1}.$$

Obviously this process can be done \bar{N}_k times, where \bar{N}_k is chosen to tend to infinity slowly enough so that $\bar{r}_k := r_{k,1} 2^{-\bar{N}_k}$ satisfies

$$\bar{u}_1^k(\bar{r}_k) + 2 \log \bar{r}_k \leq -N_k + C\bar{N}_k \rightarrow -\infty.$$

We can use \bar{r}_k to replace $r_{k,1}$. Exactly the same argument shows the existence of $s_k \rightarrow 0, \tilde{N}_k \rightarrow \infty$ such that

$$\begin{cases} s_k/r_{k,1} \rightarrow \infty, \\ \bar{u}_1^k(r) + 2 \log r \leq -\tilde{N}_k, \quad r_{k,1} \leq r \leq s_k. \end{cases}$$

Next we claim that, between $r_{k,1}$ and s_k , there must be a $r_{k,2}$ such that

$$\bar{u}_2^k(r_{k,2}) + 2 \log r_{k,2} \leq -N_{k,2} \tag{3-4}$$

for some $N_{k,2} \rightarrow \infty$ as $k \rightarrow \infty$. The proof of (3-4) is very similar to what has been used before: If this is not the case, $e^{u_2^k} \geq Cr^{-2}$ for some $C > 0$ and $r \in (r_{k,1}, s_k)$. The fact that $s_k/r_{k,1} \rightarrow \infty$ leads to a contradiction to the uniform bound of the energy of u_2^k .

Thus, we have proved that, for $r = r_{k,2}$ both u_1^k and u_2^k decay faster than $-2 \log r$:

$$\bar{u}_i^k(r) + 2 \log r \leq -N_k, \quad r = r_{k,2}, \quad i = 1, 2,$$

for some $N_k \rightarrow \infty$. Then it is easy to see that there exist $s_k \rightarrow 0$ and $s_k/r_{k,2} \rightarrow \infty$ such that

$$\bar{u}_i^k(r) + 2 \log r \leq -N'_k, \quad r_{k,2} \leq r \leq s_k, \quad i = 1, 2,$$

for some $N'_k \rightarrow \infty$ as well. The same argument as above guarantees the existence of $l_k \in (r_{k,2}, s_k)$ and some $N''_k \rightarrow \infty$ such that

$$\bar{u}_3^k(l_k) + 2 \log l_k \leq -N''_k.$$

Clearly this argument can be applied finitely many times to exhaust all the components of the whole system. Lemma 3.2 is established. \square

Now we continue with the proof of Proposition 3.1.

Case one: $\gamma_i^k \equiv 0$. Using the definition of σ_i in (1-7), we choose $l_k \rightarrow 0$ such that $\Sigma_k \subset B(0, \frac{1}{2}l_k)$ and

$$\frac{1}{2\pi} \int_{B_{l_k}} h_i^k e^{u_i^k} = \sigma_i + o(1) \quad \text{for } i \in I. \quad (3-5)$$

Here we claim that (3-1) also holds, because otherwise we would have

$$\bar{u}_i(l_k) + 2 \log l_k \geq -C.$$

By Lemma 2.4

$$\bar{u}_i(r) + 2 \log r \geq -C_1, \quad l_k \leq r \leq 2l_k,$$

which means there is a lower bound on the energy in the annulus $B_{2l_k} \setminus B_{l_k}$. Consequently

$$\frac{1}{2\pi} \int_{B_{2l_k}} h_i^k e^{u_i^k} > \sigma_i + \epsilon$$

for some $\epsilon > 0$ independent of k , a contradiction to the definition of σ_i in (1-7).

Let

$$v_i^k(y) = u_i^k(l_k y) + 2 \log l_k, \quad i \in I.$$

Then clearly we have

$$\begin{cases} \Delta v_i^k(y) + \sum_{j=1}^n a_{ij} H_j^k(y) e^{v_j^k(y)} = 0, & |y| \leq l_k^{-1}, \quad i \in I, \\ \bar{v}_i^k(1) \rightarrow -\infty, \end{cases} \quad (3-6)$$

where

$$H_i^k(y) = h_i^k(l_k y), \quad |y| \leq l_k^{-1}, \quad i \in I.$$

The Pohozaev identity we use is

$$\begin{aligned} \sum_i \int_{B_{\sqrt{R_k}}} (x \cdot \nabla H_i^k) e^{v_i^k} + 2 \sum_i \int_{B_{\sqrt{R_k}}} H_i^k e^{v_i^k} \\ = \sqrt{R_k} \int_{\partial B_{\sqrt{R_k}}} \sum_i H_i^k e^{v_i^k} + \sqrt{R_k} \int_{\partial B_{\sqrt{R_k}}} \sum_{i,j} (a^{ij} \partial_\nu v_i^k \partial_\nu v_j^k - \frac{1}{2} a^{ij} \nabla v_i^k \nabla v_j^k), \end{aligned} \quad (3-7)$$

where $R_k \rightarrow \infty$ will be chosen later and (a^{ij}) is the inverse matrix of (a_{ij}) . The key point of the following proof is to choose R_k properly in order to estimate ∇v_i^k on $\partial B_{\sqrt{R_k}}$. In the estimate of $\partial B_{\sqrt{R_k}}$, the procedure is to get rid of unimportant parts and prove that the radial part of ∇v_i^k is the leading term. To estimate all the terms of the Pohozaev identity we first write (3-7) as

$$\mathcal{L}_1 + \mathcal{L}_2 = \mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3,$$

where \mathcal{L}_1 stands for ‘‘the first term on the left’’ and the other terms are understood similarly. First, we choose $R_k \rightarrow \infty$ such that $R_k^{3/2} = o(l_k^{-1})$, then use $l_k \rightarrow 0$ to show that $\mathcal{L}_1 = o(1)$. To evaluate \mathcal{L}_2 , we observe that, by Lemma 2.4, $v_i^k(y) \rightarrow -\infty$ over all compact subsets of $\mathbb{R}^2 \setminus B_{1/2}$. Thus we further require R_k to satisfy

$$\int_{B_{R_k} \setminus B_{3/4}} H_i^k e^{v_i^k} = o(1) \quad (3-8)$$

and, for $i \in I$, by (3-6) and Lemma 2.4,

$$v_i^k(y) + 2 \log |y| \rightarrow -\infty \quad \text{uniformly in } 1 < |y| \leq R_k. \quad (3-9)$$

By the choice of l_k we clearly have

$$\frac{1}{2\pi} \int_{B_1} H_i^k e^{v_i^k} = \frac{1}{2\pi} \int_{B_{l_k}} h_i^k e^{u_i^k} = \sigma_i + o(1), \quad i \in I.$$

By (3-8), we have

$$\mathcal{L}_2 = 4\pi \sum_{i=1}^n \sigma_i + o(1).$$

For \mathcal{R}_1 , we use (3-9) to conclude $\mathcal{R}_1 = o(1)$.

Therefore we are left with the estimates of \mathcal{R}_2 and \mathcal{R}_3 , for which we shall estimate ∇v_i^k on ∂B_{R_k} . Let

$$G_k(y, \eta) = -\frac{1}{2\pi} \log |y - \eta| + \gamma_k(y, \eta)$$

be the Green’s function on $B_{l_k^{-1}}$ with respect to the Dirichlet boundary condition. Clearly

$$\gamma_k(y, \eta) = \frac{1}{2\pi} \log \frac{|y|}{l_k^{-1}} \left| \frac{l_k^{-2} y}{|y|^2} - \eta \right|,$$

and we have

$$\nabla_y \gamma_k(y, \eta) = O(l_k), \quad y \in \partial B_{\sqrt{R_k}}, \quad \eta \in B_{l_k^{-1}}. \quad (3-10)$$

We first estimate ∇v_i^k on $\partial B_{R_k^{1/2}}$. By Green's representation formula,

$$v_i^k(y) = \int_{B_{l_k^{-1}}} G(y, \eta) \sum_{j=1}^n a_{ij} H_i^k e^{v_j^k} d\eta + H_{ik},$$

where H_{ik} is the harmonic function satisfying $H_{ik} = v_i^k$ on $\partial B_{l_k^{-1}}$. Since $H_{ik} - c_k = O(1)$ for some c_k , $|\nabla H_{ik}(y)| = O(l_k)$, so

$$\begin{aligned} \nabla v_i^k(y) &= \int_{B_{l_k^{-1}}} \nabla_y G_k(y, \eta) \sum_{j=1}^n a_{ij} H_j^k e^{v_j^k} d\eta + \nabla H_{ik}(y) \\ &= -\frac{1}{2\pi} \int_{B_{l_k^{-1}}} \frac{y-\eta}{|y-\eta|^2} \sum_{j=1}^n a_{ij} H_j^k e^{v_j^k} d\eta + O(l_k). \end{aligned} \quad (3-11)$$

We estimate the integral in (3-11) over a few subregions. First, the integral over $B_{l_k^{-1}} \setminus B_{R_k^{2/3}}$ is $o(1)R_k^{-1/2}$ because, over this region, $1/|y-\eta| \sim 1/|\eta| \leq o(R_k^{-1/2})$. For the integral over B_1 , we use

$$\frac{y-\eta}{|y-\eta|^2} = \frac{y}{|y|^2} + O\left(\frac{1}{|y|^2}\right)$$

to obtain

$$-\frac{1}{2\pi} \int_{B_1} \frac{y-\eta}{|y-\eta|^2} \sum_{j=1}^n a_{ij} H_j^k e^{v_j^k} = \left(-\frac{y}{|y|^2} + O\left(\frac{1}{|y|^2}\right)\right) \left(\sum_{j=1}^n a_{ij} \sigma_j + o(1)\right).$$

This is the leading term. For the integral over the region $B(0, \sqrt{R_k}/2) \setminus B_1$, we use $1/|y-\eta| \sim 1/|y|$ and (3-8) to get

$$\int_{B_{R_k^{1/2}/2} \setminus B_1} \frac{y-\eta}{|y-\eta|^2} \sum_{j=1}^n a_{ij} H_j^k e^{v_j^k} = o(1)|y|^{-1}.$$

By a similar argument we also have

$$\int_{B_{R_k^{2/3}} \setminus (B_{R_k^{1/2}/2} \cup B(y, |y|/2))} \frac{y-\eta}{|y-\eta|^2} \sum_{j=1}^n a_{ij} H_j^k e^{v_j^k} = o(1)|y|^{-1}.$$

Finally, over the region $B(y, |y|/2)$ we use $e^{v_i^k(\eta)} = o(1)|\eta|^{-2}$ to get

$$\int_{B(y, |y|/2)} \frac{y-\eta}{|y-\eta|^2} \sum_{j=1}^n a_{ij} H_j^k e^{v_j^k} = o(1)|y|^{-1}.$$

Combining the estimates on all the subregions mentioned above, we have

$$\nabla v_i^k(y) = -\frac{y}{|y|^2} \left(\sum_{j=1}^n a_{ij} \sigma_j + o(1)\right) + o(|y|^{-1}), \quad |y| = R_k^{1/2}.$$

Using the above in \mathcal{R}_2 and \mathcal{R}_3 , we have

$$\sum_{i,j=1}^n a_{ij} \sigma_i \sigma_j = 4 \sum_{i=1}^n \sigma_i + o(1).$$

Proposition 3.1 is established for the nonsingular case.

Case two: the singular case $\exists \gamma_i \neq 0$.

Lemma 3.4. For $\sigma \in (0, 1)$, the following Pohozaev identity holds:

$$\begin{aligned} \sigma \int_{\partial B_\sigma} \sum_{i,j \in I} a^{ij} (\partial_\nu u_i^k \partial_\nu u_j^k - \frac{1}{2} \nabla u_i^k \cdot \nabla u_j^k) + \sum_{i \in I} \sigma \int_{\partial B_\sigma} h_i^k e^{u_i^k} \\ = 2 \sum_{i \in I} \int_{B_\sigma} h_i^k e^{u_i^k} + \sum_{i \in I} \int_{B_\sigma} (x \cdot \nabla h_i^k) e^{u_i^k} + 4\pi \sum_{i,j \in I} a^{ij} \gamma_i^k \gamma_j^k. \end{aligned}$$

Proof. First, we claim that, for each fixed k ,

$$\nabla u_i^k(x) = 2\gamma_i^k x/|x|^2 + O(1) \quad \text{near the origin.} \tag{3-12}$$

Indeed, recall the equation for the regular part \tilde{u}_i^k is

$$\Delta \tilde{u}_i^k(x) + \sum_j |x|^{2\gamma_j^k} h_j^k(x) e^{\tilde{u}_j^k(x)} = 0 \quad \text{in } B_1.$$

By the argument of Lemma 4.1 in [Lin and Zhang 2010], for fixed k , \tilde{u}_i^k is bounded above near 0, then an elliptic estimate leads to (3-12).

Let $\Omega = B_\sigma \setminus B_\epsilon$. The standard Pohozaev identity on Ω is

$$\sum_{i \in I} \left(\int_\Omega (x \cdot \nabla h_i^k) e^{u_i^k} + 2h_i^k e^{u_i^k} \right) = \int_{\partial\Omega} \left(\sum_i (x \cdot \nu) h_i^k e^{u_i^k} + \sum_{i,j} a^{ij} (\partial_\nu u_j^k (x \cdot \nabla u_i^k) - \frac{1}{2} (x \cdot \nu) (\nabla u_i^k \cdot \nabla u_j^k)) \right).$$

Let $\epsilon \rightarrow 0$, then the integration over Ω extends to B_σ by the integrability of $h_i^k e^{u_i^k}$ and (1-5). For the terms on the right-hand side, clearly $\partial\Omega = \partial B_\sigma \cup \partial B_\epsilon$. Thanks to (3-12), the integral on ∂B_ϵ is $-4\pi \sum_{i,j} a^{ij} \gamma_i^k \gamma_j^k$. Lemma 3.4 is established. \square

Let

$$\sigma_i^k(r) = \frac{1}{2\pi} \int_{B_r} h_i^k e^{u_i^k}, \quad i \in I.$$

Lemma 3.5. Let $\epsilon_k \rightarrow 0$ such that $\Sigma_k \subset B(0, \frac{1}{2}\epsilon_k)$ and

$$u_i^k(x) + 2 \log |x| \rightarrow -\infty, \quad |x| = \epsilon_k, \quad i \in I. \tag{3-13}$$

Then we have

$$\sum_{i,j \in I} a_{ij} \sigma_i^k(\epsilon_k) \sigma_j^k(\epsilon_k) = 4 \sum_{i \in I} (1 + \gamma_i^k) \sigma_i^k(\epsilon_k) + o(1). \tag{3-14}$$

Proof of Lemma 3.5. First the existence of ϵ_k that satisfies (3-13) is guaranteed by Lemma 2.4. In B_{ϵ_k} , we let $\tilde{u}_i^k(x)$ be defined as in (1-4). Then

$$v_i^k(y) = \tilde{u}_i^k(\epsilon_k y) + 2(1 + \gamma_i^k) \log \epsilon_k.$$

Using $v_i^k \rightarrow -\infty$ on ∂B_1 , we obtain, by Green’s representation formula and standard estimates,

$$\nabla v_i^k(y) = \left(\sum_{j \in I} a_{ij} \sigma_j^k(\epsilon_k) + o(1) \right) y, \quad y \in \partial B_1.$$

After translating the above to estimates of u_i^k , we have

$$\nabla u_i^k(x) = \left(\sum_{j \in I} (a_{ij} \sigma_j^k(\epsilon_k) - 2\gamma_j^k) \right) \frac{x}{|x|^2} + \frac{o(1)}{|x|}, \quad |x| = \epsilon_k. \tag{3-15}$$

As we observe the Pohozaev identity in Lemma 3.4 with $\sigma = \epsilon_k$, we see easily that the second term on the left-hand side and the second term on the right-hand side are both $o(1)$. The first term on the right-hand side is clearly $4\pi \sum_i \sigma_i^k(\epsilon_k)$. Therefore we only need to evaluate the first term on the left-hand side, for which we use (3-15). Lemma 3.5 is established by similar estimates as in the nonsingular case. \square

Thus Proposition 3.1 is established for the singular case as well. \square

Remark 3.6. The proof of Proposition 3.1 clearly indicates the following statements when it is applied to an SU(3) Toda system. Let $B(p_k, l_k)$ be a circle centered at p_k with radius l_k . Let Σ'_k be a subset of Σ_k . Suppose $\text{dist}(\Sigma'_k, \partial B(p_k, l_k)) = o(1) \text{dist}(\Sigma_k \setminus \Sigma'_k, \partial B(p_k, l_k))$, and we consider the following two situations: If $p_k = 0$, we have

$$\tilde{\sigma}_1^k(l_k)^2 - \tilde{\sigma}_1^k(l_k)\tilde{\sigma}_2^k(l_k)^2 + \tilde{\sigma}_2^k(l_k) = 2\mu_1\tilde{\sigma}_1^k(l_k) + 2\mu_2\tilde{\sigma}_2^k(l_k) + o(1).$$

If $0 \in \Sigma_k \setminus \Sigma'_k$, then

$$\tilde{\sigma}_1^k(l_k)^2 - \tilde{\sigma}_1^k(l_k)\tilde{\sigma}_2^k(l_k) + \tilde{\sigma}_2^k(l_k)^2 = 2\tilde{\sigma}_1^k(l_k) + 2\tilde{\sigma}_2^k(l_k) + o(1),$$

where $\tilde{\sigma}_i^k(l_k) = (1/2\pi) \int_{B(p_k, l_k)} h_i^k e^{u_i^k}$. This fact will be used in the final step of the proof of Theorem 1.2.

Remark 3.7. From the proof of Proposition 3.1, we see that the Pohozaev identity has to be evaluated on *fast decay* components in order to rule out the \mathcal{R}_1 term. A component is called *fast decay* if the difference between itself and the threshold harmonic function tends to $-\infty$; for example, see (3-13). A component is called a *slow decay* component if it is not a fast decay component. Later, in the remaining part of the proof of Theorem 1.2, we shall derive Pohozaev identities over different regions and all of them will have to be evaluated on fast decay components.

4. Fully bubbling systems

Next we consider a typical blowup situation for systems: fully bubbling solutions. First, let $\gamma_i^k \equiv 0$ for all $i \in I$. Let

$$\lambda^k = \max \left\{ \max_{B_1} u_1^k, \dots, \max_{B_1} u_n^k \right\} \tag{4-1}$$

and $x^k \rightarrow 0$ be where λ^k is attained. Let

$$v_i^k(y) = u_i^k(x_k + e^{-\lambda^k/2}y) - \lambda^k, \quad y \in \Omega_k, \quad i \in I, \tag{4-2}$$

where $\Omega_k = \{y : e^{-\lambda^k/2}y + x_k \in B_1\}$. The sequence is called fully bubbling if, along a subsequence,

$$\{v_1^k, \dots, v_n^k\} \text{ converge in } C_{\text{loc}}^2(\mathbb{R}^2) \text{ to } (v_1, \dots, v_n) \tag{4-3}$$

that satisfies

$$\Delta v_i + \sum_{j \in I} a_{ij} h_j e^{v_j} = 0 \quad \text{in } \mathbb{R}^2, \quad i \in I, \tag{4-4}$$

where $h_i = \lim_{k \rightarrow \infty} h_i^k(0)$. Our next theorem is concerned with the closeness between $u^k = (u_1^k, \dots, u_n^k)$ and $v = (v_1, \dots, v_n)$.

Theorem 4.1. *Let $A = A_n, u^k$ be a sequence of solutions to (1-2) with $\gamma_i^k = 0$ for all $i \in I$. Suppose u^k satisfies (1-3) and (1-6), h^k satisfies (1-5), and λ^k, x^k and v^k are described by (4-1) and (4-2), respectively. Suppose u^k is fully bubbling; then there exists $C > 0$ independent of k such that*

$$|u_i^k(e^{-\frac{1}{2}\lambda^k}y + x^k) - \lambda^k - v_i(y)| \leq C + o(1) \log(1 + |y|) \quad \text{for } x \in \Omega_k, \quad i \in I. \tag{4-5}$$

Remark 4.2. If A is nonnegative, i.e., the system is a Liouville system, Theorem 4.1 and Theorem 4.3 below are established in [Lin and Zhang 2010]. For $A = A_2$, [Jost et al. 2006] proved

$$|u_i^k(e^{-\lambda^k/2}y + x^k) - \lambda^k - v_i(y)| \leq C \quad \text{for } x \in \Omega_k, \quad i = 1, 2.$$

Clearly this estimate is slightly stronger than (4-5) for $n = 2$. The Jost–Lin–Wang proof uses holonomy theory but the proof of Theorem 4.1 is a simple application of the Pohozaev identity proved in Section 3.

If there is a $\gamma_i \neq 0$, we let

$$\tilde{\lambda}^k = \max \left\{ \frac{\max_{B_1} \tilde{u}_1^k}{1 + \gamma_1^k}, \dots, \frac{\max_{B_1} \tilde{u}_n^k}{1 + \gamma_n^k} \right\},$$

and

$$\tilde{v}_i^k(y) = \tilde{u}_i^k(e^{-\tilde{\lambda}^k/2}y) - (1 + \gamma_i^k)\tilde{\lambda}^k$$

for $i \in I$ and $y \in \Omega_k := \{y : e^{-\tilde{\lambda}^k/2}y \in B_1\}$. We assume

$$(\tilde{v}_1^k, \dots, \tilde{v}_n^k) \text{ converge in } C_{\text{loc}}^2(\mathbb{R}^2) \text{ to } (\tilde{v}_1, \dots, \tilde{v}_n) \tag{4-6}$$

that satisfies

$$\Delta \tilde{v}_i + \sum_{j=1}^n a_{ij} |x|^{2\gamma_j} h_j e^{\tilde{v}_j} = 0 \quad \text{in } \mathbb{R}^2, \quad i \in I, \tag{4-7}$$

where $h_i = \lim_{k \rightarrow \infty} h_i^k(0)$.

Theorem 4.3. *Let $A = A_n, \tilde{u}^k, \tilde{v}^k, (\tilde{v}_1, \dots, \tilde{v}_n), \tilde{\lambda}^k, \epsilon_k$ and Ω_k be as described above, and h_i^k and γ_i^k satisfy (1-5); then, under assumption (4-6), there exists $C > 0$ independent of k such that*

$$|\tilde{u}_i^k(e^{\tilde{\lambda}^k/2}y) - (1 + \gamma_i^k)\tilde{\lambda}^k - \tilde{v}_i(y)| \leq C + o(1) \log(1 + |y|) \quad \text{for } x \in \Omega_k. \tag{4-8}$$

Proof of Theorem 4.1. Recall that σ_i is defined in (1-7). By Proposition 3.1, we have

$$\sum_{i,j \in I} a_{ij} \sigma_i \sigma_j = 4 \sum_{i \in I} \sigma_i. \quad (4-9)$$

On the other hand, let

$$\sigma_{iv} := \frac{1}{2\pi} \int_{\mathbb{R}^2} h_i e^{v_i} \quad \text{for } i = 1, \dots, n,$$

where $v = (v_1, \dots, v_n)$ is the limit of the fully bubbling sequence after scaling. Clearly $\sigma_v = (\sigma_{1v}, \dots, \sigma_{nv})$ also satisfies (4-9). We claim that

$$\sigma_i = \sigma_{iv} \quad \text{for } i = 1, \dots, n. \quad (4-10)$$

Let $s_i = \sigma_i - \sigma_{iv}$; we obviously have $s_i \geq 0$. The difference between σ and σ_v on (4-9) gives

$$\sum_{i,j \in I} a_{ij} s_i s_j + 2 \sum_{i \in I} \left(\sum_{j \in I} a_{ij} \sigma_{vj} - 2 \right) s_i = 0. \quad (4-11)$$

First, by Proposition 2.1, we have $\sum_{j \in I} a_{ij} \sigma_{vj} - 2 > 0$. Next, if either A is nonnegative ($a_{ij} \geq 0$ for all $i, j = 1, \dots, n$) or A is positive definite, we have $\sum_{i,j \in I} a_{ij} s_i s_j \geq 0$. Then (4-11) and $s_i \geq 0$ imply (4-10).

From the convergence from v_i^k to v_i in $C_{\text{loc}}^2(\mathbb{R}^2)$, we can find $R_k \rightarrow \infty$ such that

$$|v_i^k(y) - v_i(y)| = o(1), \quad |y| \leq R_k.$$

For $|y| > R_k$, let

$$\bar{v}_i^k(r) = \frac{1}{2\pi r} \int_{\partial B_r} v_i^k(y) dS_y.$$

Then

$$\frac{d}{dr} \bar{v}_i^k(r) = \frac{1}{2\pi r} \int_{B_r} \Delta v_i^k = -\frac{1}{2\pi r} \int_{B_r} \sum_{j \in I} a_{ij} h_j^k e^{v_j^k} = -\frac{\sum_j a_{ij} \sigma_j + o(1)}{r}.$$

Hence

$$\bar{v}_i^k(r) = -\left(\sum_{j \in I} a_{ij} \sigma_j + o(1) \right) \log r + O(1) \quad \text{for all } r > 2.$$

Since $v_i^k(y) = \bar{v}_i^k(|y|) + O(1)$ and

$$v_i(y) = -\left(\sum_j a_{ij} \sigma_j \right) \log |y| + O(1) \quad \text{for } |y| > 1,$$

we see that (4-5) holds. Theorem 4.1 is established. \square

Proof of Theorem 4.3. By (3-14) we have

$$\sum_{i,j \in I} a_{ij} \sigma_i \sigma_j = 4 \sum_{i \in I} (1 + \gamma_i) \sigma_i. \quad (4-12)$$

Recall that $v = (v_1, \dots, v_n)$ satisfies (4-7). Let

$$\sigma_{iv} = \frac{1}{2\pi} \int_{\mathbb{R}^2} h_i |x|^{2\gamma_i} e^{v_i}.$$

On the one hand, $(\sigma_{1v}, \dots, \sigma_{iv})$ also satisfies (4-12); on the other hand, the classification theorem of [Lin et al. 2012a] gives

$$\sum_{j \in I} a_{ij} \sigma_{jv} > 2 + 2\gamma_i, \quad i \in I. \tag{4-13}$$

Let $s_i = \sigma_i - \sigma_{iv}$ ($i \in I$); then (4-12), which is satisfied by both $(\sigma_1, \dots, \sigma_n)$ and $(\sigma_{1v}, \dots, \sigma_{nv})$, gives

$$\sum_{i,j \in I} a_{ij} s_i s_j + 2 \sum_{i \in I} \left(\sum_{j \in J} a_{ij} \sigma_{jv} - 2 - 2\gamma_i \right) s_i = 0.$$

By (4-13) and the assumption on A , we have $s_i = 0$ for all $i \in I$. The remaining part of the proof is exactly like the last part of the proof of Theorem 4.1. Theorem 4.3 is established. \square

5. Asymptotic behavior of solutions in each simple blowup area

In this section, we derive some results on the energy classification around each blowup point. First we let $A = A_n$ (the Cartan matrix) and consider:

The neighborhood around 0. Since 0 is postulated to belong to Σ_k first, it means there may not be a bubbling picture in a neighborhood of 0.

Let $\tau_k = \frac{1}{2} \text{dist}(0, \Sigma_k \setminus \{0\})$; we consider the energy limits of $h_i^k e^{u_i^k}$ in B_{τ_k} . By the selection process and Lemma 2.4,

$$u_i^k(x) + 2 \log |x| \leq C, \quad u_i^k(x) = \bar{u}_i^k(|x|) + O(1), \quad |x| \leq \tau_k, \quad i \in I, \tag{5-1}$$

where $\bar{u}_i^k(|x|)$ is the average of u_i^k on $\partial B_{|x|}$. Let \tilde{u}_i^k be defined by (1-4). Then we have

$$\Delta \tilde{u}_i^k(x) + \sum_{j \in I} a_{ij} |x|^{2\gamma_j} h_j^k(x) e^{\tilde{u}_j^k(x)} = 0, \quad |x| \leq \tau_k.$$

Let

$$-2 \log \delta_k = \max_{i \in I} \max_{x \in B(0, \tau_k)} \frac{\tilde{u}_i^k}{1 + \gamma_i^k}$$

and

$$v_i^k(y) = \tilde{u}_i^k(\delta_k y) + 2(1 + \gamma_i^k) \log \delta_k, \quad |y| \leq \tau_k \delta_k^{-1}. \tag{5-2}$$

It is easy to see the equation for v_i^k is

$$\Delta v_i^k(y) + \sum_{j \in I} a_{ij} |y|^{2\gamma_j^k} h_j^k(\delta_k y) e^{v_j^k(y)} = 0, \quad |y| \leq \tau_k \delta_k^{-1}.$$

Then we consider two trivial cases, first, $\tau_k \delta_k^{-1} \leq C$. This is the case that there is no entire bubble after scaling.

Let f_i^k solve

$$\begin{cases} \Delta f_i^k + \sum_{j \in I} a_{ij} |y|^{2\gamma_j^k} h_j^k(\delta_k y) e^{f_j^k} = 0, & |y| \leq \tau_k \delta_k^{-1}, \\ f_i^k = 0 & \text{on } |y| = \tau_k \delta_k^{-1}. \end{cases}$$

Using $v_i \leq 0$ we have $|f_i^k| \leq C$ on $B(0, \tau_k \delta_k^{-1})$. Since $v_i^k - f_i^k$ is harmonic and v_i^k has bounded oscillation on $\partial B(0, \tau_k \delta_k^{-1})$, we have

$$v_i^k(x) = \bar{v}_i^k(\partial B(0, \tau_k \delta_k^{-1})) + O(1) \quad \text{for all } x \in B(0, \tau_k \delta_k^{-1}), \tag{5-3}$$

where $\bar{v}_i^k(\partial B(0, \tau_k \delta_k^{-1}))$ stands for the average of v_i^k on $\partial B(0, \tau_k \delta_k^{-1})$. Direct computation shows that

$$\int_{B(0, \tau_k)} e^{u_i^k(x)} dx = \int_{B(0, \tau_k \delta_k^{-1})} e^{v_i^k(y)} |y|^{2\gamma_i^k} dy.$$

Therefore,

$$\int_{B_{\tau_k}} h_i^k e^{u_i^k} dx = O(1) e^{\bar{v}_i^k(\partial B(0, \tau_k \delta_k^{-1}))}. \tag{5-4}$$

So, if $\bar{v}_i^k(\partial B(0, \tau_k \delta_k^{-1})) \rightarrow -\infty$, then $\int_{B_{\tau_k}} h_i^k e^{u_i^k} dx = o(1)$.

The second trivial case is when the blowup sequence is fully bubbling. We now have

$$\tau_k \delta_k^{-1} \rightarrow \infty \tag{5-5}$$

and we assume that $(v_1^k, \dots, v_n^k) \rightarrow (v_1, \dots, v_n)$ in $C_{loc}^2(\mathbb{R}^2)$. Clearly,

$$\Delta v_i + \sum_{j=1}^n a_{ij} |x|^{2\gamma_j} h_j e^{v_j} = 0 \quad \text{in } \mathbb{R}^2, \quad i \in I,$$

where $h_i = \lim_{k \rightarrow \infty} h_i^k(0)$. By the classification theorem of [Lin et al. 2012a], we have

$$\frac{1}{2\pi} \sum_{j \in I} a_{ij} \int_{\mathbb{R}^2} |y|^{2\gamma_j} e^{v_j} h_j dy = 2(2 + \gamma_i + \gamma_{n+1-i})$$

and

$$v_i(y) = -(4 + 2\gamma_{n+1-i}) \log |y| + O(1), \quad |y| > 1, \quad i \in I.$$

By the proof of Theorem 4.3, there is only one bubble.

The final case we consider is a partially blown-up picture. Note that (5-5) is assumed. For the following two propositions we assume $n = 2$, i.e., we consider SU(3) Toda systems.

Proposition 5.1. *Suppose (1-2), (1-3), (1-5) and (1-6) hold for u^k, h_i^k and γ_i . The matrix A equals A_2 , and (5-5) also holds. Suppose $s_k \in (0, \tau_k)$ satisfies*

$$u_i^k(x) \leq -2 \log |x| - N_k, \quad i = 1, 2,$$

for all $|x| = s_k$ and some $N_k \rightarrow \infty$. Then $(\sigma_1^k(s_k), \sigma_2^k(s_k))$ is an $o(1)$ perturbation of one of the following five types:

$$(2\mu_1, 0), \quad (0, 2\mu_2), \quad (2(\mu_1 + \mu_2), 2\mu_2), \quad (2\mu_1, 2(\mu_1 + \mu_2)), \quad (2\mu_1 + 2\mu_2, 2\mu_1 + 2\mu_2).$$

On $\partial B(0, \tau_k)$, for each i either

$$u_i^k(x) + 2 \log |x| \geq -C, \quad |x| = \tau_k,$$

for some $C > 0$ or

$$u_i^k(x) + 2 \log |x| < -(2 + \delta) \log |x| + \delta \log \delta_k, \quad |x| = \tau_k, \tag{5-6}$$

for some $\delta > 0$. If (5-6) holds for some i , then

$$\sigma_i^k(\tau_k) = o(1), \quad 2\mu_i + o(1) \text{ or } 2\mu_1 + 2\mu_2 + o(1).$$

Moreover, there exists at least one i_0 such that (5-6) holds for $i = i_0$.

Similarly, for bubbles away from the origin we have:

Proposition 5.2. *Suppose (1-2), (1-3), (1-5) and (1-6) hold for u^k, h_i^k and γ_i . The matrix A equals A_2 . Let $x_k \in \Sigma_k \setminus \{0\}$, $\bar{\tau}_k = \frac{1}{2} \text{dist}(x_k, \Sigma_k \setminus \{0, x_k\})$ and*

$$\bar{\delta}_k = \exp\left(-\frac{1}{2} \max_{\substack{i=1,2 \\ x \in B(x_k, \bar{\tau}_k)}} u_i^k(x)\right).$$

Then, for all $s_k \in (0, \bar{\tau}_k)$, if

$$u_i^k(x) + 2 \log |x - x_k| \leq -N_k \quad \text{for all } |x - x_k| = s_k, \quad i = 1, 2,$$

for some $N_k \rightarrow \infty$, then $((1/2\pi) \int_{B(x_k, s_k)} h_1^k e^{u_1^k}, (1/2\pi) \int_{B(x_k, s_k)} h_2^k e^{u_2^k})$ is an $o(1)$ perturbation of one of the following five types:

$$(2, 0), \quad (0, 2), \quad (2, 4), \quad (4, 2), \quad (4, 4).$$

On $\partial B(x_k, \bar{\tau}_k)$, for each i either

$$u_i^k(x) + 2 \log \bar{\tau}_k \geq -C \quad \text{for all } x \in \partial B(x_k, \bar{\tau}_k)$$

or

$$u_i^k(x) \leq -(2 + \delta) \log \bar{\tau}_k + \delta \log \bar{\delta}_k \quad \text{for all } x \in \partial B(x_k, \bar{\tau}_k). \tag{5-7}$$

If (5-7) holds for some i , then $(1/2\pi) \int_{B(x_k, \bar{\tau}_k)} h_i^k e^{u_i^k}$ is $o(1), 2 + o(1)$ or $4 + o(1)$. Moreover, there exists at least one i_0 such that (5-7) holds for i_0 .

We shall only prove Proposition 5.1, as the proof for Proposition 5.2 is similar.

Proof of Proposition 5.1. Let v_i^k be defined by (5-2). Since we only need to consider a partially blown-up situation, without loss of generality we assume v_1^k converges to v_1 in $C_{loc}^2(\mathbb{R}^2)$ and v_2^k tends to $-\infty$ over any compact subset of \mathbb{R}^2 . The equation for v_1 is

$$\Delta v_1 + 2h_1|y|^{2\gamma_1} e^{v_1} = 0 \quad \text{in } \mathbb{R}^2, \quad \int_{\mathbb{R}^2} h_1|y|^{2\gamma_1} e^{v_1} < \infty,$$

where $h_1 = \lim_{k \rightarrow \infty} h_1^k(0)$. By the classification result of [Prajapat and Tarantello 2001] we have

$$2 \int_{\mathbb{R}^2} h_1|y|^{2\gamma_1} e^{v_1} = 8\pi \mu_1$$

and

$$v_1(y) = -4\mu_1 \log |y| + O(1), \quad |y| > 1.$$

Thus we can find $R_k \rightarrow \infty$ (without loss of generality, $R_k = o(1)\tau_k\delta_k^{-1}$) such that

$$\frac{1}{2\pi} \int_{B_{R_k}} h_1^k(\delta_k y) |y|^{2\gamma_1^k} e^{v_1^k} = 2\mu_1 + o(1),$$

i.e., $\sigma_1^k(\delta_k R_k) = 2\mu_1 + o(1)$, and

$$\int_{B_{R_k}} h_2^k(\delta_k y) |y|^{2\gamma_2^k} e^{v_2^k} = o(1).$$

For $r \geq R_k$, recall that

$$\sigma_i^k(\delta_k r) = \frac{1}{2\pi} \int_{B_r} h_i^k(\delta_k y) |y|^{2\gamma_i^k} e^{v_i^k} dy;$$

then we have

$$\begin{aligned} \frac{d}{dr} \bar{v}_1^k(r) &= \frac{-2\sigma_1^k(\delta_k r) + \sigma_2^k(\delta_k r)}{r}, \\ \frac{d}{dr} \bar{v}_2^k(r) &= \frac{\sigma_1^k(\delta_k r) - 2\sigma_2^k(\delta_k r)}{r}, \quad R_k \leq r \leq \tau_k \delta_k^{-1}. \end{aligned}$$

Clearly we have

$$R_k \frac{d}{dr} \bar{v}_1^k(R_k) = -4\mu_1 + o(1), \quad R_k \frac{d}{dr} \bar{v}_2^k(R_k) = 2\mu_1 + o(1). \quad (5-8)$$

The following lemma says that as long as both components stay well below the harmonic function $-2 \log |y|$ (i.e., both of them are fast decay components), there is no essential change on the energy for either component:

Lemma 5.3. *Suppose $L_k \in (R_k, \tau_k \delta_k^{-1})$ satisfies*

$$v_i^k(y) + 2\gamma_i^k \log |y| \leq -2 \log |y| - N_k, \quad R_k \leq |y| \leq L_k, \quad i = 1, 2, \quad (5-9)$$

for some $N_k \rightarrow \infty$, then

$$\sigma_i^k(\delta_k R_k) = \sigma_i^k(\delta_k L_k) + o(1), \quad i = 1, 2.$$

Proof of Lemma 5.3. We aim to prove that σ_i^k does not change much from $\delta_k R_k$ to $\delta_k L_k$. Suppose this is not the case; then there exists i such that $\sigma_i^k(\delta_k L_k) > \sigma_i^k(\delta_k R_k) + \delta$ for some $\delta > 0$. Let $\tilde{L}_k \in (R_k, L_k)$ be such that

$$\max_{i=1,2} (\sigma_i^k(\delta_k \tilde{L}_k) - \sigma_i^k(\delta_k R_k)) = \epsilon \quad \text{for } i = 1, 2, \quad (5-10)$$

where $\epsilon > 0$ is sufficiently small. Then, for v_1^k ,

$$\frac{d}{dr} \bar{v}_1^k(r) \leq \frac{-4(1 + \gamma_1) + \epsilon}{r} \leq -\frac{2(1 + \gamma_1) + \epsilon}{r}. \quad (5-11)$$

It is easy to see from Lemma 2.4 that

$$\int_{B_{\tilde{L}_k} \setminus B_{R_k}} |y|^{2\gamma_1^k} e^{v_1^k} = o(1),$$

which is $\sigma_1^k(\delta_k \tilde{L}_k) = \sigma_1^k(\delta_k R_k) + o(1)$. Indeed, by Lemma 2.4,

$$\int_{B_{L_k} \setminus B_{R_k}} |y|^{2\gamma_1^k} e^{v_1^k} = O(1) \int_{B_{L_k} \setminus B_{R_k}} |y|^{2\gamma_1^k} e^{\bar{v}_1^k} = o(1).$$

The second equality above is because, by (5-11),

$$\bar{v}_1^k(r) + 2\gamma_1^k \log r \leq -N_k - 2 \log R_k + \left(-2 - \frac{1}{2}\epsilon\right) \log r, \quad R_k \leq r \leq L_k.$$

Thus $\sigma_2^k(\delta_k \tilde{L}_k) = \sigma_2^k(\delta_k R_k) + \epsilon$. However, since (5-9) holds, by Remark 3.6 we have

$$\lim_{k \rightarrow \infty} (\sigma_1^k(\delta_k \tilde{L}_k), \sigma_2^k(\delta_k \tilde{L}_k)) \in \Gamma.$$

The two points on Γ that have the first component equal to $2\mu_1$ are $(2\mu_1, 0)$ and $(2\mu_1, 2(\mu_1 + \mu_2))$. Thus (5-10) is impossible. Lemma 5.3 is established. \square

From Lemma 5.3 and (5-8) we see that, for $r \geq R_k$, either

$$v_i^k(y) + 2\gamma_i^k \log |y| \leq -2 \log |y| - N_k, \quad R_k \leq |y| \leq \tau_k \delta_k^{-1}, \quad i = 1, 2, \tag{5-12}$$

or there exists $L_k \in (R_k, \tau_k \delta_k^{-1})$ such that

$$v_2^k(y) + 2\gamma_2^k \log L_k \geq -2 \log L_k - C, \quad |y| = L_k, \tag{5-13}$$

for some $C > 0$, while, for $R_k \leq |y| \leq L_k$,

$$v_1^k(y) + 2\gamma_1^k \log |y| \leq -(2 + \delta) \log |y|, \quad R_k \leq |y| \leq L_k, \tag{5-14}$$

for some $\delta > 0$. Indeed, from (5-8) we see that if the energy has to change, σ_2^k has to change first. L_k can be chosen so that $\sigma_2^k(\delta_k L_k) - \sigma_2^k(\delta_k R_k) = \epsilon$ for some $\epsilon > 0$ small.

Lemma 5.4. *Suppose there exist $L_k \geq R_k$ such that (5-13) and (5-14) hold. For L_k , we assume $L_k = o(1)\tau_k \delta_k^{-1}$. Then there exist \tilde{L}_k such that $\tilde{L}_k/L_k \rightarrow \infty$ and $\tilde{L}_k = o(1)\tau_k \delta_k^{-1}$ still holds. For $|y| = \tilde{L}_k$, we have*

$$v_i^k(y) + 2(1 + \gamma_i^k) \log |y| \leq -N_k, \quad |y| = \tilde{L}_k, \quad i = 1, 2, \tag{5-15}$$

for some $N_k \rightarrow \infty$. In particular,

$$v_1^k(y) + 2\left(1 + \gamma_1^k + \frac{1}{4}\delta\right) \log |y| \leq 0, \quad |y| = \tilde{L}_k, \tag{5-16}$$

$$\sigma_1^k(\delta_k \tilde{L}_k) = 2\mu_1 + o(1), \quad \sigma_2^k(\delta_k \tilde{L}_k) = 2\mu_1 + 2\mu_2 + o(1). \tag{5-17}$$

Remark 5.5. The statement of Lemma 5.4 can be understood as follows: Suppose, starting from ∂B_{L_k} , σ_2^k starts to change because (5-13) holds. Then, from L_k to \tilde{L}_k , σ_1^k does not change much and v_1^k is still far below $-2(1 + \gamma_1^k) \log |y|$, but v_2^k has changed from decaying slowly (which is (5-13)) to a fast decay (the $i = 2$ part of (5-16)). In other words, as σ_2^k changes from L_k to \tilde{L}_k , v_2^k changes from slow decay to fast decay but v_1^k still has fast decay in the meanwhile. The change of σ_2^k has influenced the derivative of \bar{v}_1^k , but has not made σ_1^k change much because σ_2^k changes too fast from L_k to \tilde{L}_k .

Proof of Lemma 5.4. First we observe that, by Lemma 5.3, the energy does not change if both components satisfy (5-12). Thus we can assume that $\sigma_2^k(\delta_k L_k) \leq \epsilon$ for some $\epsilon > 0$ small. Consequently,

$$\frac{d}{dr} \bar{v}_1^k(r) \leq \frac{-4(1 + \gamma_1) + 2\epsilon}{r}, \quad r \geq R_k.$$

Now we claim that there exists $N > 1$ such that

$$\sigma_2^k(\delta_k(L_k N)) \geq 2 + \gamma_1 + \gamma_2 + o(1). \tag{5-18}$$

If this is not true, we would have $\epsilon_0 > 0$ and $\tilde{R}_k \rightarrow \infty$ such that

$$\sigma_2^k(\delta_k \tilde{R}_k L_k) \leq 2 + \gamma_1 + \gamma_2 - \epsilon_0. \tag{5-19}$$

On the other hand, \tilde{R}_k can be chosen to tend to infinity slowly, so that, by Lemma 2.4 and (5-14),

$$v_1^k(y) + 2(1 + \gamma_1^k) \log |y| \leq -\frac{1}{2} \delta \log |y|, \quad L_k \leq |y| \leq \tilde{R}_k L_k. \tag{5-20}$$

Clearly (5-20) implies $\sigma_1^k(\delta_k L_k) = \sigma_1^k(\delta_k \tilde{R}_k L_k) + o(1)$. Thus, by (5-19),

$$\frac{d}{dr} \bar{v}_2^k(r) \geq \frac{-2 - 2\gamma_2 + \epsilon_0/2}{r}. \tag{5-21}$$

Using (5-21) and

$$v_2^k(y) = (-2 - 2\gamma_2^k) \log |y| + O(1), \quad |y| = L_k,$$

we see easily that

$$\int_{B(0, \tilde{R}_k L_k) \setminus B(0, L_k)} |y|^{2\gamma_2^k} e^{v_2^k} \rightarrow \infty,$$

a contradiction to (1-6). Therefore (5-18) holds.

By Lemma 2.4,

$$v_i^k(y) + 2 \log(NL_k) = \bar{v}_i^k(NL_k) + 2 \log(NL_k) + O(1), \quad |y| = NL_k, \quad i = 1, 2.$$

Thus we have

$$\begin{aligned} \bar{v}_1^k(NL_k) &\leq (-2 - 2\gamma_1^k - \frac{1}{2}\delta) \log(NL_k), \\ \bar{v}_2^k(NL_k) &\geq (-2 - 2\gamma_2^k) \log(NL_k) - C. \end{aligned}$$

Consequently,

$$\bar{v}_2^k((N + 1)L_k) \geq (-2 - 2\gamma_2^k) \log L_k - C$$

leads to

$$\frac{1}{2\pi} \int_{B(0, (N+1)L_k)} h_2^k(\delta_k y) |y|^{2\gamma_2^k} e^{v_2^k(y)} dy \geq 2 + \gamma_1 + \gamma_2 + \epsilon_0$$

for some $\epsilon_0 > 0$. Going back to the equation for \bar{v}_2^k , we have

$$\frac{d}{dr} \bar{v}_2^k(r) \leq -\frac{2 + 2\gamma_2 + \epsilon_0}{r}, \quad r = (N + 1)L_k.$$

Therefore we can find $\tilde{R}_k \rightarrow \infty$ such that $\tilde{R}_k L_k = o(1)\tau_k \delta_k^{-1}$ and

$$\begin{aligned} v_2^k(y) &\leq (-2 - 2\gamma_2^k - \epsilon_0) \log |y| - N_k, & |y| &= \tilde{R}_k L_k, \\ v_1^k(y) &\leq (-2 - 2\gamma_1^k - \frac{1}{4}\delta) \log |y|, & L_k &\leq |y| \leq \tilde{R}_k L_k. \end{aligned}$$

Obviously,

$$\sigma_1^k(\delta_k \tilde{R}_k L_k) = \sigma_1^k(\delta_k L_k) + o(1) = \sigma_1^k(\delta_k R_k) + o(1) = 2(1 + \gamma_1) + o(1).$$

By computing the Pohozaev identity on $\tilde{R}_k L_k$, we have

$$\sigma_2^k(\delta_k \tilde{R}_k L_k) = 2\mu_1 + 2\mu_2 + o(1).$$

Letting $\tilde{L}_k = \tilde{R}_k L_k$, we have proved Lemma 5.4. □

To finish the proof of Proposition 5.1, we need to consider the region $\tilde{L}_k \leq |y| \leq \tau_k \delta_k^{-1}$ if $L_k = o(1)\tau_k \delta_k^{-1}$ (in which case \tilde{L}_k can be made to be $o(1)\tau_k \delta_k^{-1}$), or $L_k = O(1)\tau_k \delta_k^{-1}$. First we consider the region $\tilde{L}_k \leq |y| \leq \tau_k \delta_k^{-1}$ when $\tilde{L}_k = o(1)\tau_k \delta_k^{-1}$. It is easy to verify that

$$\begin{aligned} \frac{d}{dr} \bar{v}_1^k(r) &= -\frac{2\gamma_1 - 2\gamma_2}{r} + \frac{o(1)}{r}, & r &= \tilde{L}_k, \\ \frac{d}{dr} \bar{v}_2^k(r) &= -\frac{6 + 2\gamma_1 + 4\gamma_2 + o(1)}{r}, & r &= \tilde{L}_k. \end{aligned}$$

The second equation above implies

$$\frac{d}{dr} \bar{v}_2^k(r) \leq -\frac{2\mu_2 + \delta}{r}, \quad r = \tilde{L}_k,$$

for some $\delta > 0$. So $\sigma_2^k(r)$ does not change for $r \geq \tilde{L}_k$ unless σ_1^k changes. By the same argument as before, either v_1^k rises to $-2 \log |y| + O(1)$ on $|y| = \tau_k \delta_k^{-1}$, or there is $\hat{L}_k = o(1)\tau_k \delta_k^{-1}$ such that

$$\sigma_i^k(\delta_k \hat{L}_k) = 2\mu_1 + 2\mu_2 + o(1), \quad i = 1, 2.$$

Since this is the energy of a fully bubbling system, we have in this case both

$$v_i^k(y) \leq -(2\mu_i + \delta) \log |y|, \quad |y| = \tau_k \delta_k^{-1}, \quad i = 1, 2,$$

for some $\delta > 0$.

If $L_k = O(1)\tau_k \delta_k^{-1}$, it is easy to use Lemma 2.4 to see that one component is $-2(1 + \gamma_i^k) \log |y| + O(1)$ and the other component has the fast decay. Proposition 5.1 is established. □

6. Combination of bubbling areas

The following definition plays an important role:

Definition 6.1. Let $Q_k = \{p_1^k, \dots, p_q^k\}$ be a subset of Σ_k such that Q_k has more than one point in it and $\Sigma_k \setminus Q_k = \emptyset$. Q_k is called a group if:

$$(1) \quad \text{dist}(p_i^k, p_j^k) \sim \text{dist}(p_s^k, p_t^k),$$

where $p_i^k, p_j^k, p_s^k, p_t^k$ are any points in Q_k such that $p_i^k \neq p_j^k$ and $p_t^k \neq p_s^k$.

(2) For any $p_k \in \Sigma_k \setminus Q_k$, $\text{dist}(p_i^k, p_j^k)/\text{dist}(p_i^k, p_k) \rightarrow 0$ for all $p_i^k, p_j^k \in Q_k$ with $p_i^k \neq p_j^k$.

Proof of Theorem 1.2. Let $2\tau_k$ be the distance between 0 and $\Sigma_k \setminus \{0\}$. For each $z_k \in \Sigma_k \cap \partial B(0, 2\tau_k)$, if $\text{dist}(z_k, \Sigma_k \setminus \{z_k\}) \sim \tau_k$, let G_0 be the group that contains the origin. On the other hand, if there exists $z'_k \in \partial B(0, 2\tau_k)$ such that $\tau_k/\text{dist}(z'_k, \Sigma_k \setminus \{z'_k\}) \rightarrow \infty$, we let G_0 be 0 itself. By the definition of a group, all members of G_0 are in $B(0, N\tau_k)$ for some N independent of k . Let

$$v_i^k(y) = u_i^k(\tau_k y) + 2 \log \tau_k, \quad |y| \leq \tau_k^{-1}.$$

Then we have

$$\Delta v_i^k(y) + \sum_{j=1}^2 a_{ij} h_j^k(\tau_k y) e^{v_j^k(y)} = 4\pi \gamma_i^k \delta_0, \quad |y| \leq \tau_k^{-1}. \tag{6-1}$$

Let $0, Q_1, \dots, Q_m$ be the images of members of G_0 after the scaling from y to $\tau_k y$. Then all $Q_i \in B_N$. By Proposition 5.1 and Proposition 5.2, at least one component decays fast on ∂B_1 . Without loss of generality, we assume

$$v_1^k \leq -N_k \quad \text{on } \partial B_1$$

for some $N_k \rightarrow \infty$, and

$$\sigma_1^k(\tau_k) = o(1), 2\mu_1 + o(1) \text{ or } 2\mu_1 + 2\mu_2 + o(1).$$

Specifically, if $\tau_k \delta_k^{-1} \leq C$, then $\sigma_1^k(\tau_k) = o(1)$. Otherwise, $\sigma_1^k(\tau_k)$ is equal to one of the two other cases mentioned above. By Lemma 2.4, $v_1^k \leq -N_k + C$ on all $\partial B(Q_t, 1)$ ($t = 1, \dots, m$); therefore, by Proposition 5.2,

$$\frac{1}{2\pi} \int_{B(Q_t, 1)} h_1^k(\tau_k \cdot) e^{v_1^k} = 2m_t + o(1), \quad t = 1, \dots, m,$$

where, for each t , $m_t = 0, 1$ or 2 . Let $2\tau_k L_k$ be the distance from 0 to the nearest group other than G_0 . Then $L_k \rightarrow \infty$. By Lemma 2.4 and the proof of Lemma 3.2, we can find $\tilde{L}_k \leq L_k, \tilde{L}_k \rightarrow \infty$, such that most of the energy of v_1^k in $B(0, \tilde{L}_k)$ is contributed by bubbles and v_2^k decays faster than $-2 \log \tilde{L}_k$ on $\partial B(0, \tilde{L}_k)$:

$$\frac{1}{2\pi} \int_{B(0, \tilde{L}_k)} h_1^k(0) e^{v_1^k} = 2m + o(1), 2\mu_1 + 2m + o(1) \text{ or } 2(\mu_1 + \mu_2) + 2m + o(1) \tag{6-2}$$

for some nonnegative integer m , and

$$v_2^k(y) + 2 \log \tilde{L}_k \rightarrow -\infty, \quad |y| = \tilde{L}_k. \tag{6-3}$$

Then we evaluate the Pohozaev identity on $B(0, \tilde{L}_k)$. Since (6-3) holds, by Remark 3.6 we have

$$\lim_{k \rightarrow \infty} (\sigma_1^k(\tau_k \tilde{L}_k), \sigma_2^k(\tau_k \tilde{L}_k)) \in \Gamma.$$

Moreover, by (6-2) we see that $\lim_{k \rightarrow \infty} (\sigma_1^k(\tau_k \tilde{L}_k), \sigma_2^k(\tau_k \tilde{L}_k)) \in \Sigma$ because the limit point is the intersection between the line $\sigma_1 = \lim_{k \rightarrow \infty} \sigma_1^k(\tau_k \tilde{L}_k)$ and Γ .

The Pohozaev identity for $(\sigma_1^k(\tau_k \tilde{L}_k), \sigma_2^k(\tau_k \tilde{L}_k))$ can be written as

$$\sigma_1^k(\tau_k \tilde{L}_k)(2\sigma_1^k(\tau_k \tilde{L}_k) - \sigma_2^k(\tau_k \tilde{L}_k) - 4\mu_1) + \sigma_2^k(\tau_k \tilde{L}_k)(2\sigma_2^k(\tau_k \tilde{L}_k) - \sigma_1^k(\tau_k \tilde{L}_k) - 4\mu_2) = o(1).$$

Thus either

$$2\sigma_1^k(\tau_k \tilde{L}_k) - \sigma_2^k(\tau_k \tilde{L}_k) \geq 4\mu_1 + o(1) \tag{6-4}$$

or

$$2\sigma_2^k(\tau_k \tilde{L}_k) - \sigma_1^k(\tau_k \tilde{L}_k) \geq 4\mu_2 + o(1).$$

Moreover, if

$$2\sigma_1^k(\tau_k \tilde{L}_k) - \sigma_2^k(\tau_k \tilde{L}_k) \geq 2\mu_1 + o(1) \quad \text{and} \quad 2\sigma_2^k(\tau_k \tilde{L}_k) - \sigma_1^k(\tau_k \tilde{L}_k) \geq 2\mu_2 + o(1),$$

then, by the proof of Theorem 4.3,

$$\int_{B_{l_k} \setminus \tau_k \tilde{L}_k} h_i^k e^{u_i^k} = o(1), \quad i = 1, 2,$$

for any $l_k \rightarrow 0$. In this case we have

$$\sigma_i = \lim_{k \rightarrow \infty} \sigma_i^k(\tau_k \tilde{L}_k), \quad i = 1, 2,$$

and Theorem 1.2 is proved in this case.

Thus, without loss of generality, we assume that (6-4) holds. From the equation for u_1^k , this means that, for some $\delta > 0$,

$$\bar{u}_1^k(\tau_k \tilde{L}_k) \leq -2 \log(\tau_k \tilde{L}_k) - N_k, \quad \frac{d}{dr} \bar{u}_1^k(r) < \frac{-2-\delta}{r}, \quad r = \tau_k \tilde{L}_k. \tag{6-5}$$

The property above implies, by the proof of Proposition 5.1, that, as r grows from $\tau_k \tilde{L}_k$ to $\tau_k L_k$, the following three situations may occur:

Case one. Both u_i^k satisfy, for some $N_k \rightarrow \infty$, that

$$u_i^k(x) + 2 \log |x| \leq -N_k, \quad \tau_k \tilde{L}_k \leq |x| \leq \tau_k L_k, \quad i = 1, 2.$$

In this case,

$$\sigma_i^k(\tau_k \tilde{L}_k) = \sigma_i^k(\tau_k L_k) + o(1), \quad i = 1, 2.$$

So, on $\partial B(0, \tau_k L_k)$, u_1^k is still a fast decaying component.

Case two. There exist $L_{1,k}$ and $L_{2,k} \in (\tilde{L}_k, L_k)$ such that

$$\begin{aligned} u_2^k(x) &\geq -2 \log(\tau_k L_{1,k}) - C, \quad |x| = \tau_k L_{1,k}, \\ u_i^k(x) &\leq -2 \log(\tau_k L_{2,k}) - N_k, \quad |x| = \tau_k L_{2,k}, \quad i = 1, 2, \end{aligned} \tag{6-6}$$

and

$$\sigma_1^k(\tau_k \tilde{L}_k) = \sigma_1^k(\tau_k L_{2,k}) + o(1). \tag{6-7}$$

Since (6-6) holds, by Remark 3.6 we have $(\lim_{k \rightarrow \infty} \sigma_1^k(\tau_k L_{2,k}), \lim_{k \rightarrow \infty} \sigma_2^k(\tau_k L_{2,k})) \in \Gamma$. Then we further observe that, since (6-7) holds, $\lim_{k \rightarrow \infty} (\sigma_1^k(\tau_k L_{2,k}), \sigma_2^k(\tau_k L_{2,k})) \in \Sigma$, because this point is obtained by intersecting Γ with $\sigma_1 = \lim_{k \rightarrow \infty} \sigma_1^k(\tau_k \tilde{L}_k)$. In other words, the new point $\lim_{k \rightarrow \infty} (\sigma_1^k(\tau_k L_{2,k}), \sigma_2^k(\tau_k L_{2,k}))$ is on the upper right part of the old point $\lim_{k \rightarrow \infty} (\sigma_1^k(\tau_k \tilde{L}_k), \sigma_2^k(\tau_k \tilde{L}_k))$.

Case three.
$$u_2^k(x) \geq -2 \log \tau_k L_k - C, \quad |x| = \tau_k L_k,$$

for some $C > 0$ and $\sigma_1^k(\tau_k \tilde{L}_k) = \sigma_1^k(\tau_k L_k) + o(1)$. This means that $\partial B(0, \tau_k L_k)$, u_1^k is still the fast decaying component.

If the second case above happens, the relationship between σ_1^k and σ_2^k on $B(0, \tau_k L_k) \setminus B(0, \tau_k L_{2,k})$ is the same as discussed before. In any case, on $\partial B(0, \tau_k L_k)$ at least one of the two components has fast decay and has its energy equal to a corresponding component of a point in Σ . For any group not equal to G_0 , it is easy to see that the fast decay component has its energy equal to 0, 2 or 4. The combination of bubbles for groups is very similar to the combination of bubbling disks as we have done before. For example, let G_0, G_1, \dots, G_t be groups in $B(0, \epsilon_k)$ for some $\epsilon_k \rightarrow 0$. Suppose the distances between any two of G_0, \dots, G_t are comparable and

$$\text{dist}(G_i, G_j) = o(1)\epsilon_k \quad \text{for all } i, j = 0, \dots, t, \quad i \neq j.$$

Also we require $(\Sigma_k \setminus (\bigcup_{i=0}^t G_i)) \cap B(0, 2\epsilon_k) = \emptyset$. Let $\epsilon_{1,k} = \text{dist}(G_0, G_1)$; then all G_0, \dots, G_t are in $B(0, N\epsilon_{1,k})$ for some $N > 0$. Without loss of generality let u_1^k be a fast decaying component on $\partial B(0, N\epsilon_{1,k})$. Then we have

$$\sigma_1^k(N\epsilon_{1,k}) = \sigma_1^k(\tau_k L_k) + 2m + o(1),$$

where m is a nonnegative integer because, by Lemma 2.4, u_1^k is also a fast decaying component for G_1, \dots, G_t . Moreover, by Proposition 5.2, the energy of u_1^k in G_s ($s = 1, \dots, t$) is $o(1), 2 + o(1)$ or $4 + o(1)$. If u_2^k also has fast decay on $\partial B(0, N\epsilon_{1,k})$, then $\lim_{k \rightarrow \infty} (\sigma_1^k(N\epsilon_{1,k}), \sigma_1^k(N\epsilon_{1,k})) \in \Sigma$ because this is a point of intersection between Γ and $\sigma_1 = \lim_{k \rightarrow \infty} \sigma_1^k(\tau_k L_k) + 2m$. If

$$u_2^k(x) \geq -2 \log N\epsilon_{1,k} - C, \quad |x| = N\epsilon_{1,k},$$

then, as before, we can find $\epsilon_{3,k}$ in $(N\epsilon_{1,k}, \epsilon_k)$ such that, for some $N_k \rightarrow \infty$,

$$u_i^k(x) + 2 \log \epsilon_{3,k} \leq -N_k, \quad |x| = \epsilon_{3,k}, \quad i = 1, 2,$$

and

$$\sigma_1^k(N\epsilon_{1,k}) = \sigma_1^k(\epsilon_{3,k}).$$

Thus we have

$$\lim_{k \rightarrow \infty} (\sigma_1^k(\epsilon_{3,k}), \sigma_2^k(\epsilon_{3,k})) \in \Sigma,$$

because this point is the intersection between Γ and $\sigma_1 = \lim_{k \rightarrow \infty} \sigma_1^k(N\epsilon_{1,k})$.

The last possibility on $B(0, \epsilon_k) \setminus B(0, \epsilon_{1,k})$ is

$$\sigma_1^k(\epsilon_k) = \sigma_1^k(N\epsilon_{1,k}) + o(1)$$

and

$$u_2^k(x) + 2 \log \epsilon_k \geq -C, \quad |x| = \epsilon_k.$$

In this case, u_1^k is the fast decaying component on $\partial B(0, \epsilon_k)$.

Such a procedure can be applied to include groups further away from G_0 . Since we have only finitely many blowup disks this procedure only needs to be applied finitely many times. Finally, let $s_k \rightarrow 0$ be such that

$$\sigma_i = \lim_{k \rightarrow \infty} \lim_{s_k \rightarrow 0} \sigma_i^k(s_k), \quad i = 1, 2,$$

and, for some $N_k \rightarrow \infty$,

$$u_i^k(x) + 2 \log s_k \leq -N_k, \quad |x| = s_k, \quad i = 1, 2.$$

Then we see that $(\sigma_1, \sigma_2) \in \Sigma$. Theorem 1.2 is established. \square

Acknowledgements

Part of the paper was finished when Zhang was visiting Chinese University of Hong Kong in April, May and December in 2012, and Taida Institute of Mathematical Sciences (TIMS) in June 2012. He would like to thank both institutes for their warm hospitality and generous financial support. The research of Wei is partially supported by NSERC of Canada. The authors are grateful to the anonymous referee for scrutinizing the whole paper.

References

- [Bartolucci and Malchiodi 2013] D. Bartolucci and A. Malchiodi, “An improved geometric inequality via vanishing moments, with applications to singular Liouville equations”, *Comm. Math. Phys.* **322**:2 (2013), 415–452. MR 3077921 Zbl 1276.58005
- [Bartolucci and Tarantello 2002a] D. Bartolucci and G. Tarantello, “The Liouville equation with singular data: a concentration-compactness principle via a local representation formula”, *J. Differential Equations* **185**:1 (2002), 161–180. MR 2003h:35063 Zbl 1247.35032
- [Bartolucci and Tarantello 2002b] D. Bartolucci and G. Tarantello, “Liouville type equations with singular data and their applications to periodic multivortices for the electroweak theory”, *Comm. Math. Phys.* **229**:1 (2002), 3–47. MR 2003e:58026 Zbl 1009.58011
- [Bartolucci et al. 2004] D. Bartolucci, C.-C. Chen, C.-S. Lin, and G. Tarantello, “Profile of blow-up solutions to mean field equations with singular data”, *Comm. Partial Differential Equations* **29**:7-8 (2004), 1241–1265. MR 2005i:58022 Zbl 1062.35146
- [Battaglia and Malchiodi 2014] L. Battaglia and A. Malchiodi, “A Moser–Trudinger inequality for the singular Toda system”, *Bull. Inst. Math., Acad. Sin. (N.S.)* **9**:1 (2014), 1–23. MR 3234966 Zbl 1295.35372
- [Bennet 1934] W. H. Bennet, “Magnetically self-focusing streams”, *Phys. Rev.* **45** (1934), 890–897.
- [Bolton and Woodward 1997] J. Bolton and L. M. Woodward, “Some geometrical aspects of the 2-dimensional Toda equations”, pp. 69–81 in *Geometry, topology and physics* (Campinas, 1996), edited by B. N. Apanasov et al., de Gruyter, Berlin, 1997. MR 98m:58049 Zbl 0911.58019
- [Bolton et al. 1988] J. Bolton, G. R. Jensen, M. Rigoli, and L. M. Woodward, “On conformal minimal immersions of S^2 into $\mathbb{C}P^n$ ”, *Math. Ann.* **279**:4 (1988), 599–620. MR 88m:53110 Zbl 0642.53063
- [Calabi 1953] E. Calabi, “Isometric imbedding of complex manifolds”, *Ann. of Math. (2)* **58** (1953), 1–23. MR 15,160c Zbl 0051.13103

- [Chanillo and Kiessling 1995] S. Chanillo and M. K.-H. Kiessling, “Conformally invariant systems of nonlinear PDE of Liouville type”, *Geom. Funct. Anal.* **5**:6 (1995), 924–947. MR 96j:35052 Zbl 0858.35035
- [Chen and Li 1991] W. X. Chen and C. Li, “Classification of solutions of some nonlinear elliptic equations”, *Duke Math. J.* **63**:3 (1991), 615–622. MR 93e:35009 Zbl 0768.35025
- [Chen and Li 1993] W. X. Chen and C. Li, “Qualitative properties of solutions to some nonlinear elliptic equations in \mathbb{R}^2 ”, *Duke Math. J.* **71**:2 (1993), 427–439. MR 94f:35046 Zbl 0923.35055
- [Chen and Lin 1998] C.-C. Chen and C.-S. Lin, “Estimate of the conformal scalar curvature equation via the method of moving planes, II”, *J. Differential Geom.* **49**:1 (1998), 115–178. MR 2000h:35045 Zbl 0961.35047
- [Chen and Lin 2002] C.-C. Chen and C.-S. Lin, “Sharp estimates for solutions of multi-bubbles in compact Riemann surfaces”, *Comm. Pure Appl. Math.* **55**:6 (2002), 728–771. MR 2003d:53056 Zbl 1040.53046
- [Chen and Lin 2003] C.-C. Chen and C.-S. Lin, “Topological degree for a mean field equation on Riemann surfaces”, *Comm. Pure Appl. Math.* **56**:12 (2003), 1667–1727. MR 2004h:35065 Zbl 1032.58010
- [Chern and Wolfson 1987] S. S. Chern and J. G. Wolfson, “Harmonic maps of the two-sphere into a complex Grassmann manifold, II”, *Ann. of Math. (2)* **125**:2 (1987), 301–335. MR 88g:58038 Zbl 0627.58017
- [Childress and Percus 1981] S. Childress and J. K. Percus, “Nonlinear aspects of chemotaxis”, *Math. Biosci.* **56**:3-4 (1981), 217–237. MR 83d:92019 Zbl 0481.92010
- [Chipot et al. 1997] M. Chipot, I. Shafrir, and G. Wolansky, “On the solutions of Liouville systems”, *J. Differential Equations* **140**:1 (1997), 59–105. MR 98j:35053 Zbl 0902.35039
- [Debye and Huckel 1923] P. Debye and E. Huckel, “Zur Theorie der Electrolyte, II: Das Grenzgesetz für die elektrische Leitfähigkeit”, *Phys. Zft.* **24**:10 (1923), 305–325.
- [Doliwa 1997] A. Doliwa, “Holomorphic curves and Toda systems”, *Lett. Math. Phys.* **39**:1 (1997), 21–32. MR 98a:58047 Zbl 0869.35089
- [Dunne 1995] G. Dunne, *Self-dual Chern–Simons theories*, Lecture Notes in Physics Monographs **36**, Springer, Berlin, 1995. Zbl 0834.58001
- [Dunne et al. 1991] G. V. Dunne, R. Jackiw, S.-Y. Pi, and C. A. Trugenberger, “Self-dual Chern–Simons solitons and two-dimensional nonlinear equations”, *Phys. Rev. D* (3) **43**:4 (1991), 1332–1345. MR 92g:81120
- [Ganoulis et al. 1982] N. Ganoulis, P. Goddard, and D. Olive, “Self-dual monopoles and Toda molecules”, *Nuclear Phys. B* **205**:4 (1982), 601–636. MR 84c:81045
- [Guest 1997] M. A. Guest, *Harmonic maps, loop groups, and integrable systems*, London Mathematical Society Student Texts **38**, Cambridge University Press, Cambridge, 1997. MR 99g:58036 Zbl 0898.58010
- [Jost et al. 2006] J. Jost, C. Lin, and G. Wang, “Analytic aspects of the Toda system, II: Bubbling behavior and existence of solutions”, *Comm. Pure Appl. Math.* **59**:4 (2006), 526–558. MR 2007h:35099 Zbl 1207.35140
- [Keller and Segel 1971] E. F. Keller and L. A. Segel, “Traveling bands of Chemotactic Bacteria: A theoretical analysis”, *J. Theor. Biol.* **30**:2 (1971), 235–248. Zbl 1170.92308
- [Kiessling and Lebowitz 1994] M. K.-H. Kiessling and J. L. Lebowitz, “Dissipative stationary plasmas: kinetic modeling, Bennett’s pinch and generalizations”, *Phys. Plasmas* **1**:6 (1994), 1841–1849. MR 95a:82111
- [Leznov 1980] A. N. Leznov, “On complete integrability of a nonlinear system of partial differential equations in two-dimensional space”, *Teoret. Mat. Fiz.* **42**:3 (1980), 343–349. In Russian; translated in *Theor. Math. Phys.* **42** (1980), 224–229. MR 82e:58047 Zbl 0426.35028
- [Leznov and Saveliev 1992] A. N. Leznov and M. V. Saveliev, *Group-theoretical methods for integration of nonlinear dynamical systems*, Progress in Physics **15**, Birkhäuser, Basel, 1992. MR 93d:58194 Zbl 0746.58004
- [Li 1995] Y. Y. Li, “Prescribing scalar curvature on S^n and related problems, I”, *J. Differential Equations* **120**:2 (1995), 319–410. MR 98b:53031 Zbl 0827.53039
- [Li 1999] Y. Y. Li, “Harnack type inequality: the method of moving planes”, *Comm. Math. Phys.* **200**:2 (1999), 421–444. MR 2000c:58024 Zbl 0928.35057
- [Lin and Zhang 2010] C.-S. Lin and L. Zhang, “Profile of bubbling solutions to a Liouville system”, *Ann. Inst. H. Poincaré Anal. Non Linéaire* **27**:1 (2010), 117–143. MR 2011a:35118 Zbl 1182.35107

- [Lin and Zhang 2011] C.-S. Lin and L. Zhang, “A topological degree counting for some Liouville systems of mean field type”, *Comm. Pure Appl. Math.* **64**:4 (2011), 556–590. MR 2012d:35388 Zbl 1214.58008
- [Lin et al. 2012a] C.-S. Lin, J. Wei, and D. Ye, “Classification and nondegeneracy of $SU(n+1)$ Toda system with singular sources”, *Invent. Math.* **190**:1 (2012), 169–207. MR 2969276 Zbl 1258.35089
- [Lin et al. 2012b] C.-S. Lin, J. Wei, and C. Zhao, “Sharp estimates for fully bubbling solutions of a $SU(3)$ Toda system”, *Geom. Funct. Anal.* **22**:6 (2012), 1591–1635. MR 3000499 Zbl 1271.30017
- [Malchiodi and Ndiaye 2007] A. Malchiodi and C. B. Ndiaye, “Some existence results for the Toda system on closed surfaces”, *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.* **18**:4 (2007), 391–412. MR 2008h:35064 Zbl 1148.35021
- [Malchiodi and Ruiz 2013] A. Malchiodi and D. Ruiz, “A variational analysis of the Toda system on compact surfaces”, *Comm. Pure Appl. Math.* **66**:3 (2013), 332–371. MR 3008227 Zbl 1275.35095
- [Mansfield 1982] P. Mansfield, “Solution of Toda systems”, *Nuclear Phys. B* **208**:2 (1982), 277–300. MR 85e:58145
- [Mock 1975] M. S. Mock, “Asymptotic behavior of solutions of transport equations for semiconductor devices”, *J. Math. Anal. Appl.* **49** (1975), 215–225. MR 50 #7857 Zbl 0298.35033
- [Musso et al. 2015] M. Musso, A. Pistoia, and J. Wei, “New concentration phenomena for $SU(3)$ Toda system”, preprint, 2015.
- [Nolasco and Tarantello 1999] M. Nolasco and G. Tarantello, “Double vortex condensates in the Chern–Simons–Higgs theory”, *Calc. Var. Partial Differential Equations* **9**:1 (1999), 31–94. MR 2001e:58022 Zbl 0951.58030
- [Nolasco and Tarantello 2000] M. Nolasco and G. Tarantello, “Vortex condensates for the $SU(3)$ Chern–Simons theory”, *Comm. Math. Phys.* **213**:3 (2000), 599–639. MR 2001j:58025 Zbl 0998.81047
- [Prajapat and Tarantello 2001] J. Prajapat and G. Tarantello, “On a class of elliptic problems in \mathbb{R}^2 : symmetry and uniqueness results”, *Proc. Roy. Soc. Edinburgh Sect. A* **131**:4 (2001), 967–985. MR 2002j:35108 Zbl 1009.35018
- [Schoen 1988] R. Schoen, “Topics in differential geometry”, lecture notes, Stanford University, 1988.
- [Yang 1997] Y. Yang, “The relativistic non-abelian Chern–Simons equations”, *Comm. Math. Phys.* **186**:1 (1997), 199–218. MR 98h:58047 Zbl 0874.58093
- [Yang 2001] Y. Yang, *Solitons in field theory and nonlinear analysis*, Springer, New York, 2001. MR 2002m:58001 Zbl 0982.35003

Received 24 Apr 2014. Revised 21 Jan 2015. Accepted 6 Mar 2015.

CHANG-SHOU LIN: cs1in@math.ntu.edu.tw

Department of Mathematics, Taida Institute of Mathematical Sciences, National Taiwan University, Taipei 106, Taiwan

JUN-CHENG WEI: jcwei@math.ubc.ca

Department of Mathematics, University of British Columbia, Vancouver B.C, Canada V6T 1Z2,

and

Department of Mathematics, Chinese University of Hong Kong, Shatin, Hong Kong

LEI ZHANG: leizhang@ufl.edu

Department of Mathematics, University of Florida, 358 Little Hall P.O. Box 118105, Gainesville, FL 32611-8105, United States

RICCI FLOW ON SURFACES WITH CONIC SINGULARITIES

RAFE MAZZEO, YANIR A. RUBINSTEIN AND NATASA SESUM

We establish short-time existence of the Ricci flow on surfaces with a finite number of conic points, all with cone angle between 0 and 2π , with cone angles remaining fixed or changing in some smooth prescribed way. For the angle-preserving flow we prove long-time existence; if the angles satisfy the Troyanov condition, this flow converges exponentially to the unique constant-curvature metric with these cone angles; if this condition fails, the conformal factor blows up at precisely one point. These geometric results rely on a new refined regularity theorem for solutions of linear parabolic equations on manifolds with conic singularities. This is proved using methods from geometric microlocal analysis, which is the main novelty of this article.

1. Introduction	839
2. Preliminaries on Ricci flow	842
3. Linear estimates and existence of the flow	848
4. Convergence of the flow in the Troyanov case	869
5. Convergence in the non-Troyanov case	872
Acknowledgements	880
References	880

1. Introduction

This article studies the local and global properties of Ricci flow on compact surfaces with conic singularities. This is a natural continuation of various efforts, including recent work of Mazzeo and Sesum, to develop a comprehensive understanding of Ricci flow in two dimensions in various natural geometries. This work is also partly motivated by extensive recent efforts in higher-dimensional complex geometry toward finding Kähler–Einstein edge metrics with prescribed cone angle along a divisor, as approached by Mazzeo and Rubinstein using a stationary (continuity) method with features suggested by the Ricci flow, together with geometric microlocal techniques. A final motivation is the Hamilton–Tian conjecture, stipulating that Kähler–Ricci flow on Fano manifolds should converge in a suitable sense to a Kähler–Ricci soliton with mild singularities; we make some progress toward the analogue of this conjecture in our setting.

We investigate here the dynamical problem of Ricci flow on a Riemann surface (M, J) , with conic singularities at a specified k -tuple of points \vec{p} , where the cone angle at p_j is $2\pi\beta_j$. Our main theorems provide optimal regularity for flow in this setting for cone angle smaller than 2π . We state these results, deferring explanation of the notations and terminology until later in the introduction and the next section.

MSC2010: 53C44, 58J35.

Keywords: Ricci flow, conic singularities, heat kernels.

Theorem 1.1. *Consider a set of conic data $(M, J, \vec{p}, \vec{\beta})$ with all $\beta_j \in (0, 1)$, and let g_0 be a $C_b^{2,\gamma}$ conic metric compatible with this data (this regularity class is defined in Section 3B) with curvature $K_{g_0} \in C_b^{0,\gamma}$. If $\beta_j(t) : [0, t_0] \rightarrow (0, 1)$ is a k -tuple of C^∞ functions with $\beta_j(0) = \beta_j$, then there exists a solution $g(t)$ of (2-1) defined on some interval $0 \leq t < T \leq t_0$ with conic singularities at the points p_j with cone angle $2\pi\beta_j(t)$ at time t . For $t > 0$, $g(t)$ is smooth away from the p_j and polyhomogeneous at these conic points, and satisfies $\lim_{t \searrow 0} g(t) = g_0$.*

The special case of this theorem when $\beta_j(t) \equiv \beta_j(0)$ is called the angle-preserving flow and is the two-dimensional case of a recent short-time existence result for the Yamabe flow with edge singularities by [Bahuaud and Vertman 2014]; the methods developed here to obtain the necessary bounds for the linear parabolic problem are somewhat more flexible than theirs and yield stronger estimates.

The key step in the proof of this short-time existence result is a new regularity statement for the linearized parabolic equation. This regularity is one of the main new technical contributions of this article.

Theorem 1.2. *Let $\beta \in (0, 1)$. Suppose that $(\partial_t - \Delta_g - V)u = f$, $u(0, \cdot) = \phi$, where $g, V, \phi \in C_b^{0,\gamma}(M)$ and f lies in the parabolic regularity space $C_b^{\gamma,\gamma/2}(M \times [0, T])$. Then, near each conic point,*

$$u = a_0(t) + r^{1/\beta}(a_{11}(t) \cos y + a_{12}(t) \sin y) + \mathcal{O}(r^2), \tag{1-1}$$

where $a_0, a_{ij}(t) \in C^{1+\gamma/2}([0, T])$. When g, V, f and ϕ are all polyhomogeneous, then the solution u is polyhomogeneous on $[0, T] \times M$. If $\beta > 1$, a similar expansion holds but there exist additional terms of order $r^\delta (\log r)^k$ with $\delta \in (1/\beta, 2)$.

This refined regularity for solutions of singular parabolic equations seems to be new and requires some delicate analysis that is mostly contained in Propositions 3.6 and 3.9. We expect this type of estimate should be a standard tool in problems where such equations arise; see [Gell-Redman 2011] for a recent application.

We go beyond this short-time existence result only for the angle-preserving flow. Theorem 1.2 allows us to directly adapt Hamilton’s method to get long-time existence of the normalized flow. Convergence, however, is more subtle. As we explain below, there is a set of linear inequalities (2-14), discovered by Troyanov, which is known to be necessary and sufficient for the existence of constant-curvature metrics with this prescribed conic data (for cone angles less than 2π).

Theorem 1.3. *Let $g(t)$ be the angle-preserving solution of the normalized Ricci flow from Theorem 1.1. Then $g(t)$ exists for all $t > 0$. If $\chi(M) \leq 0$, or if $\chi(M) > 0$ and (2-14) holds, then $g(t)$ converges exponentially to the unique constant-curvature metric compatible with this conic data.*

In the remaining cases we have two parallel results.

Theorem 1.4. *Let $g(t)$ be the angle-preserving solution of the normalized Ricci flow, as above. Suppose that $\chi(M) > 0$ and (2-14) fails.*

- Define $\psi(t)$ to be the t -dependent diffeomorphism generated by the vector field $\nabla f(t)$, where $\Delta f(t) = R_{g(t)} - \rho$ (where ρ is the average of R). Then $\hat{g}(t) := \psi^*g(t)$ satisfies $\partial \hat{g}(t) / \partial t = 2\hat{\mu}(t)$,

where $\hat{\mu}$ is the tensor defined by (5-1) with respect to the metric $\hat{g}(t)$, and we prove that

$$\lim_{t \rightarrow \infty} \int_M |\hat{\mu}(t)|_{\hat{g}(t)}^2 d\hat{A}(t) = \lim_{t \rightarrow \infty} \int_M |\mu(t)|_{g(t)}^2 dA(t) = 0.$$

Furthermore, the vector field $X = \nabla R + R\nabla f$ satisfies

$$\lim_{t \rightarrow \infty} \int_M |X(t)|_{g(t)}^2 dA(t) = 0.$$

- Returning to the unmodified normalized Ricci flow, and writing $g(t) = u(t, \cdot)g_0$, the conformal factor u blows up at precisely one point $q \in M$.

The significance of the tensor μ and the vector field X , is that they both vanish on a Ricci soliton. It would be very interesting to connect the two different conclusions of this theorem.

Remark 1.5. It should be possible to show that there is a t -dependent family of conformal dilations $F(t)$ fixing the point of blowup of $u(t)$, and such that $F(t)^*g(t)$ converges (on every compact set $K \subset S^2 \setminus \vec{p}$) to an eternal solution of normalized Ricci flow. One would hope to prove that the family of metrics $F(t)^*g(t)$ converges to a soliton metric, but, unfortunately, this does not seem to be possible with the present methods. It would also be quite interesting to identify the unique point of blowup of $u(t)$; the natural conjecture is that this blowup occurs at the unique conic point $p_j \in \vec{p}$ where the Troyanov condition fails.

We learned only in November 2014 of [Phong et al. 2014], where this conjecture is verified. The proof uses the machinery developed in the recent proof of the Yau–Tian–Donaldson conjecture.

Our goals are, first, to provide a clear and direct analytic treatment of the short-time existence for this problem, thus circumventing the approximation methods of [Yin 2010], and, second, to establish convergence to a constant-curvature metric when the Troyanov condition holds. This generalizes [Yin 2013], where only the negative case is handled. We assume throughout that all cone angles lie between 0 and 2π . As explained below, this restriction has significant geometric and analytical ramifications. The regularity theorem accounts for a substantial amount of the analysis here, and is one of our new innovations. Our methods provide a new approach for obtaining estimates for heat operators on conic spaces on the naturally associated Hölder spaces.

This article is organized as follows. In Section 2 we review some basic facts regarding the two-dimensional Ricci flow. The heart of the article, Section 3, develops the linear parabolic edge theory on Riemann surfaces. In particular, Sections 3A–3E review the relevant elliptic theory, based on the methods of [Mazzeo 1991; Jeffres et al. 2014], but emphasizing the simplifications that occur in this dimension compared to [Jeffres et al. 2014]. Building on this, Section 3F develops the corresponding parabolic regularity theory. Short-time existence, Theorem 1.1, is proved in Section 3G, while Section 3H contains Theorem 1.2 on the asymptotic expansion for solutions and the further results on higher regularity. The long-time existence of the flow is a fairly easy consequence of all of this and appears in Section 3J. The convergence result in the Troyanov regime is the subject of Section 4, while in Section 5 we study the complementary regime.

2. Preliminaries on Ricci flow

The normalized Ricci flow equation on surfaces is

$$\partial_t g(t) = (\rho - R(t, \cdot))g(t), \quad (2-1)$$

where R is the scalar curvature function of the metric $g(t)$ and ρ is the (time-independent!) average of the scalar curvature. For this choice of ρ , the area $A(M, g(t))$ remains constant in time. This flow preserves the conformal class of $g(t)$, so (2-1) can be written as a scalar equation for the conformal factor: if g_0 is the metric at $t = 0$ and $g(t) = u(t, \cdot)g_0$, then (2-1) is equivalent to

$$\partial_t u = \Delta_{g_0} \log u - R_{g_0} + \rho u, \quad u(0) \equiv 1. \quad (2-2)$$

This is the fundamental equation studied in this article.

2A. Miscellaneous formulae. In two dimensions, $\text{Ric}(g) = \frac{1}{2}Rg$, where R is the scalar curvature, so Ricci flow coincides with the Yamabe flow, and both are given by (2-1). This flow preserves the conformal class of the metric, and so can be written as a scalar parabolic equation. Indeed, if $g = e^\phi g_0$, then the scalar curvatures of these two metrics satisfy

$$\Delta_0 \phi - R_0 + R e^\phi = 0, \quad (2-3)$$

so, with $u = e^\phi$, (2-1) is equivalent to (2-2). (The reader should note that the conformal factor is often written $e^{2\phi}$ elsewhere, but this is compensated for here by the fact that $R = 2K$.)

If g_0 is any metric with finite Hölder regularity and isolated conic points, then its conformal class $[g_0]$ admits a representative \bar{g}_0 which is smooth on all of M . We can even assume that \bar{g}_0 is exactly Euclidean in a ball around each p_j . Fix any such metric, then choose a conformal factor $\phi_0 \in C^\infty(M \setminus \{p_1, \dots, p_k\})$ which equals $(\beta_j - 1) \log r$ in Euclidean coordinates near each p_j . The metric $\tilde{g}_0 = e^{2\phi_0} \bar{g}_0$ is then smooth away from each p_j and has the exact conic form $dr^2 + \beta_j^2 r^2 dy^2$ near p_j . Finally, write the metric g_0 , the initial condition for the Ricci flow, as $u_0 \tilde{g}_0$. This encodes the finite regularity entirely in the conformal factor. Using this regular background conic metric \tilde{g}_0 allows for some technical simplifications in the presentation below. Henceforth we relabel \tilde{g}_0 as g_0 , and then consider the initial value problem (2-2) with $u(0) = u_0$, assuming that g_0 is C^∞ on $M \setminus \{p_1, \dots, p_k\}$ and exactly conic near each p_j .

We record some other useful formulae. First, using (2-3) in (2-2) with $\phi = \frac{1}{2} \log u$ gives

$$\partial_t u = (\rho - R)u \iff \partial_t \log u = \rho - R. \quad (2-4)$$

Another formulation of the equations for the angle-fixing flow includes a distributional contribution from the cone points:

$$\partial_t \log u = \rho - R + 2\pi \sum (1 - \beta_i) \delta_{p_i}.$$

This conforms with a standard presentation in higher dimensions, but we primarily work with the equations (2-2) without the extra delta summands. Denoting the area form for $g(t)$ by dA , then

$$\frac{d}{dt} dA = (\rho - R) dA. \quad (2-5)$$

Consequently, the area $\mathcal{A}(t) := \int_M dA$ satisfies

$$\frac{d}{dt}\mathcal{A}(t) = \int_M (\rho - R) dA = \rho\mathcal{A}(t) - 4\pi\chi(M, \vec{\beta}),$$

so, if we now fix

$$\rho = \frac{4\pi\chi(M, \vec{\beta})}{\mathcal{A}(0)}, \tag{2-6}$$

then $\mathcal{A}(t) \equiv \mathcal{A}(0)$ for all t .

Note that, by (2-4), uniform bounds on $R(t)$ imply bounds and (at least subsequential) convergence for $\log u(t)$ as $t \nearrow \infty$. This means we can focus on the curvature function rather than the conformal factor. Differentiating (2-3), assuming that $g(t)$ is a solution of (2-1) on some interval $0 \leq t < T$, we obtain

$$\partial_t R = \Delta_{g(t)} R + R(R - \rho). \tag{2-7}$$

When M is compact and smooth, (2-7) implies that the minimum of R is nondecreasing in t . Indeed, $R_{\min}(t) := \inf_M R(q, t)$ satisfies

$$\frac{d}{dt}R_{\min} \geq R_{\min}(R_{\min} - \rho).$$

Since $R_{\min}(t)$ is only Lipschitz, the term on the left is defined as the limit infimum of the forward difference quotient of $R_{\min}(t)$. Since ρ is the average of R , $R_{\min} \leq \rho$; hence, if $\rho \leq 0$, then the right-hand side is nonnegative, and the claim about R_{\min} being nondecreasing holds. If $\rho > 0$, then, choosing $r(t)$ so that $dr(t)/dt = r(t)(r(t) - \rho)$, $r(0) = R_{\min}(0)$, a similar argument applied to the difference $R_{\min} - r(t)$ leads to the same conclusion.

Estimating R_{\max} is more difficult, especially when $R > 0$, and we discuss this later.

2B. Conic singularities. In two dimensions, there are two equivalent ways to describe conic singularities. The first is conformal: using a local holomorphic coordinate, we can write

$$g = e^{2\phi} |z|^{2\beta-2} |dz|^2, \tag{2-8}$$

where $\beta > 0$ and ϕ is a bounded function (with regularity to be specified later); the second is the polar coordinate model

$$g = dr^2 + r^2 h(r, y)^2 dy^2, \quad y \in S_{2\pi}^1, \tag{2-9}$$

where h is a strictly positive function with $h(0, y) = \beta$, again with regularity to be specified later. The equivalence between these two representations, at least in the model case where $\phi \equiv 0$ and $h \equiv 1$, is exhibited by writing $|dz|^2 = d\rho^2 + \rho^2 dy^2$, $y \in S_{2\pi}^1 = \mathbb{R}/2\pi\mathbb{Z}$, and setting $r = \rho^\beta / \beta$, since then

$$dr = \rho^{\beta-1} d\rho \implies g = e^{2\phi} (dr^2 + \beta^2 r^2 dy^2).$$

The fact that more general conic metrics can be written in either of these two forms is considered in [Trojanov 1991]. We refer also to [Jeffres et al. 2014, §2.1] for a thorough discussion of this correspondence. Consequently, if g has a conic singularity at p , then the underlying conformal class $[g]$ extends smoothly across p , or, in other words, the conformal class $[g]$ determined by a conic metric contains

a representative which is smooth across the conic points. (This holds for isolated conic singularities only in two dimensions, or, more generally, for nonisolated “edge” singularities in complex codimension one.)

It is also convenient to use

$$\alpha = \beta - 1,$$

and we refer to either α or β as the cone angle parameter, hopefully without causing confusion. We focus in this article exclusively on surfaces with conic singularities for which the equivalent conditions

$$2\pi\beta \in (0, 2\pi), \quad \beta \in (0, 1), \quad \alpha \in (-1, 0) \tag{2-10}$$

hold. There are good reasons for this restriction: for such cone angles, the uniformization results are definitive, and, in addition, conic surfaces with cone angles in this range have certain favourable geometric and analytic properties which are very helpful, and perhaps crucial, in certain parts of the analysis below. Related issues appear in [Jeffres et al. 2014].

2C. Uniformization of conical Riemann surfaces. Fix a smooth compact surface M , along with a conformal, or, equivalently, a complex structure J . Denote by $\vec{p} = \{p_1, \dots, p_k\} \subset M$ a collection of k distinct points, and let $\vec{\beta} = \{\beta_1, \dots, \beta_k\} \in (0, 1)^k$ be a corresponding set of cone angle parameters. As above, write $\alpha_j = \beta_j - 1$. The *conic Euler characteristic* associated to this data is the number

$$\chi(M, \vec{\beta}) = \chi(M) + \sum_{j=1}^k \alpha_j = \chi(M) + \sum_{j=1}^k \beta_j - k. \tag{2-11}$$

In the higher-dimensional language of [Jeffres et al. 2014], this is the twisted anticanonical class of the pair $(M, \sum(1 - \beta_i)p_i)$, i.e., $-K_M - \sum(1 - \beta_i)p_i$, where $K_M = T^{1,0,*}M$ denotes the class of the canonical divisor of M .

The uniformization problem asks for the existence of a conic metric g compatible with the complex structure J with cone parameters β_j at p_j and with constant curvature away from these conic points. This can also be phrased in terms of the distributional equation

$$R_g - 2\pi \sum (1 - \beta_i)\delta_{p_i} = \text{const}. \tag{2-12}$$

Indeed, in conformal coordinates, $R_g = -\Delta_g \log \gamma$ up to a constant factor, where $g = \sqrt{-1}\gamma dz \otimes \bar{d}z = \sqrt{-1}\gamma |dz|^2$, and the Poincaré–Lelong formula asserts that $-\Delta_g \log |z|$ is a multiple of the delta function at $\{z = 0\}$ (this can be seen by excising a small neighbourhood near the cone point and using Stokes’ formula). Then (2-12) follows, since, for a conic metric, $\gamma = |z|^{2\beta-2}F$ near a cone point, with F bounded.

A consequence of this formulation is the Gauss–Bonnet theorem in this setting: if g is any metric with this conic data, then

$$2\pi \chi(M, \vec{\beta}) = \int_M K_g dA_g. \tag{2-13}$$

Therefore, if a constant-curvature metric with this conic data exists, then the sign of its curvature K_g agrees with the sign of $\chi(M) + \sum \alpha_i$. Note that, because of (2-10), this sign can be positive only when $M = S^2$ (or $\mathbb{R}P^2$, but for simplicity we always work in the oriented case).

The next theorem combines results of McOwen [1988; 1993] and Troyanov [Troyanov 1991] (the existence), Luo and Tian [1992] (uniqueness and nonexistence), and Jeffres, Mazzeo and Rubinstein [Jeffres et al. 2014] (higher regularity).

Theorem 2.1. *Let $(M, J, \vec{p}, \vec{\beta})$ be as above. Then there exists a conic metric with constant curvature associated to the data $(J, \vec{p}, \vec{\beta})$ if and only if either $\chi(M, \vec{\beta}) \leq 0$, in which case $\{\beta_j\} \in (0, 1)^k$ can be arbitrary, or else $\chi(M, \vec{\beta}) > 0$ and, for each $j = 1, \dots, k$,*

$$\alpha_j > \sum_{i \neq j} \alpha_i \quad \text{or, equivalently,} \quad 2\alpha_j > \sum_{i=1}^k \alpha_i. \tag{2-14}$$

This metric, when it exists, is unique, except when $\chi(M, \vec{\beta}) = 0$, in which case it is unique up to a constant positive multiple, or when $M = S^2$ and there are no more than two conic singularities, in which case it is unique up to Möbius transformations which fix the cone points. Finally, the metric is polyhomogeneous with a complete asymptotic expansion of the form

$$g \sim \left(\sum_{j,k \geq 0} \sum_{\ell=0}^{N_{j,k}} a_{jk\ell}(y) r^{j+k/\beta} (\log r)^\ell \right) |z|^{2\beta-2} |dz|^2$$

The existence and regularity statements here were recently generalized to any dimension in [Jeffres et al. 2014, Theorems 1 and 2]; in that setting, the Troyanov condition is replaced by the coercivity of the twisted Mabuchi K-energy functional. Following [Ross and Thomas 2011], these conditions can also be reinterpreted as saying that the twisted Futaki invariant of the pair $(M, \sum(1 - \beta_i)p_i)$ is nonnegative, or, equivalently, that this pair is logarithmically K-stable. The generalization of the uniqueness part of this result to higher dimensions has been accomplished by Berndtsson [2015]. Nonexistence when coercivity fails can be easily deduced from [Jeffres et al. 2014] together with work of Berman [2013]. We also remark that Berman’s work gave a new proof of Troyanov’s original results. We refer to [Rubinstein 2014] for a survey of the results mentioned in this paragraph and further references.

The rather curious linear inequalities (2-14) were discovered by Troyanov [1991, Theorem 5], and we refer to them henceforth as the *Troyanov conditions*. As just noted, they guarantee coercivity in the variational approach to this problem, which is key to proving existence, and which plays a key role in our considerations about the flow below. This coercivity is automatic when $\chi(M) \leq 0$, where simpler barrier methods suffice [McOwen 1988].

We also remark that, if $k > 2$, then (2-14) can fail for no more than one value of j . Indeed, if these inequalities fail for two distinct index values j, j' , which we may as well take as $j = 1$ and $j' = 2$, then

$$\alpha_1 \leq \alpha_2 + \sum_{j=3}^k \alpha_j, \quad \alpha_2 \leq \alpha_1 + \sum_{j=3}^k \alpha_j \quad \implies \quad 0 \leq \sum_{j=3}^k \alpha_j,$$

which is impossible since all the α_i are negative.

We discuss the cases $k = 1, 2$ separately. Using that a constant-curvature metric is rotationally symmetric near each conic point, we see that there can be no constant-curvature metric with only one conic point,

while, if there are precisely two conic points, then the surface is globally rotationally symmetric, the cone angles are equal and the metric is the standard suspension $dr^2 + \beta^2 \sin^2 r dy^2$, $0 \leq r \leq \pi$. When $k \leq 2$ and no constant-curvature metrics exist, there are well-known soliton metrics: the teardrop ($k = 1$ and any $\beta \in (0, 1)$) and the (American) football ($k = 2$ and any pair $0 < \beta_1 < \beta_2 < 1$). These can be obtained by ODEs methods; see [Hamilton 1988; Yin 2010; Ramos 2013]; Ramos's paper gives a particularly complete and incisive analysis.

The variational approach has recently been extended considerably through the work of Malchiodi et al. to allow angles bigger than 2π , even when coercivity fails; see, e.g., [Bartolucci et al. 2011; Carlotto and Malchiodi 2012]. Our regularity result, Theorem 1.2, holds for such angles, but our proofs of long-time existence and convergence do not carry over to that angle regime.

2D. Optimal regularity. We have already identified the central role of the refined regularity in Theorem 1.2. This result considerably sharpens the linear estimates proved by Jeffres and Loya [2003]. At the technical level, that paper establishes control on two “ b -derivatives”, i.e., with respect to the vector fields $r\partial_r$ and ∂_y , which vanish at the cone points, which imply only that $\partial_r u = O(r^{-1})$, for example. Our Theorem 1.2 shows that both $\partial_r u$ and $r^{1-1/\beta}\partial_r u$ are bounded. It also parallels the recent result [Jeffres et al. 2014, Proposition 3.3], which concerns the corresponding elliptic Poisson equation $\Delta_g u = f$ for the Laplacian of a Kähler edge metric g (generalizing the conic metrics considered here). This result in the elliptic case for smooth (or polyhomogeneous) edge metrics and with data lying in Sobolev spaces appears in [Mazzeo 1991].

These refined regularity statements represent basic phenomena associated to elliptic and parabolic edge operators. The fact that “singular” terms with noninteger exponents appear in solutions goes back to the work of Kondratiev and his school in the 1960s. However, since the methods and the particular choice of function spaces used here are less well known to geometric analysts, we pause to make some additional remarks. One key fact is that, even for the model (exact conic) case, if $\Delta_g u = f$ is Hölder continuous with respect to the metric g (i.e., defining Hölder seminorms using the distance determined by g), then it is *not* the case — unlike in the smooth setting — that all second derivatives of u are even bounded, let alone Hölder continuous. A basic example of this is the harmonic function $u = \operatorname{Re} z = r^{1/\beta} \cos y$, since, if $\frac{1}{2} < \beta < 1$, then $\partial_r^2 u \sim r^{1/\beta-2}$ blows up as $r \rightarrow 0$. The optimal regularity is that $[\partial_r u]_{g;0,1/\beta-1} < \infty$, where

$$[v]_{g;0,\gamma} = \sup \frac{|v(z) - v(z')|}{d_g(z, z')^\gamma}.$$

The results described above show that the phenomena in these examples provide the only mechanism through which control of second derivatives is lost. They also show that, if $\beta \in (0, \frac{1}{2}]$ (the easier “orbifold regime”), one has full control on the Hessian, since $1/\beta \geq 2$. One can obtain a slightly weaker statement using classical methods; see [Donaldson 2012]. As shown here, and in line with [Jeffres et al. 2014], one can go further by taking advantage of a detailed description of the structure of the Green function and heat kernel. Thus, we use here the so-called b -Hölder spaces $C_b^{k,\gamma}$, which are defined using the slightly different seminorms

$$[v]_{b;0,\gamma} = \sup \frac{|v(z) - v(z')|(r + r')^\gamma}{d_g(z, z')^\gamma},$$

where $r = r(z)$ and $r' = r(z')$ are the g -distances of these respective points to the nearest conic points.

As already noted, [Bahuaud and Vertman 2014] contains a result similar to Theorem 1.1 for the higher-dimensional Yamabe flow for metrics with edges, while, as announced in [Mazzeo and Rubinstein 2012], direct analogues of Theorems 1.1 and 1.2 for the higher-dimensional Kähler–Ricci edge flow will appear in [Mazzeo and Rubinstein \geq 2015].

2E. Historical remarks. The survey [Isenberg et al. 2011] provides a fairly recent account of what is known about Ricci flow on various classes of smooth surfaces, both compact and noncompact. The survey [Rubinstein 2014] reviews results on geometry and analysis related to Kähler edge metrics, including the special case of conic metrics on Riemann surfaces. The Ricci flow on conic surfaces presents several new challenges, some geometric and some analytic. For example, the uniformization problem in this setting is obstructed, in the sense that it is not always possible to find metrics of constant curvature in a given conformal class with certain prescribed cone angles. In addition, the flow starting at an initial singular surface is not uniquely defined: there are solutions which immediately smooth out the cone points [Simon 2002; Ramos 2011], and others which immediately become complete and send the cone points to infinity [Giesen and Topping 2010; 2011]. The solutions studied here, by contrast, either preserve the cone angles or allow them to change in some prescribed, smoothly varying manner. Our methods are drawn from geometric microlocal analysis, and are continuations of the elliptic methods used in [Jeffres et al. 2014; Mazzeo and Rubinstein 2012; \geq 2015] to study the existence problem for Kähler–Einstein edge metrics. These provide very detailed information about the asymptotic behaviour of solutions near the conic points. Indeed, we have already noted that Theorem 1.2, concerning a regularity and asymptotics theorem for solutions of linear heat equations on manifolds with conic singularities, is a key ingredient, and should be useful elsewhere too.

The angle-preserving flow for Riemann surfaces with conic singularities was previously studied by Yin [2010]; his approach provides few details about the geometric nature of the solution and does not yield precise analytic or geometric control of the solution for positive time. More recently, in [Yin 2013], he establishes long-time existence of the normalized Ricci flow for conic surfaces, and proves convergence to a constant-curvature metric when the conic Euler characteristic (see Section 2C for the definition) is negative. However, he only establishes smooth convergence away from the conic points, and does not describe the precise limiting behaviour near these conic points. There is other work on this problem by Ramos, contained in his thesis but not yet released (see, however, [Ramos 2011; 2013]). Another related paper is [Bahuaud and Vertman 2014], which proves a short-time existence result for the Yamabe flow on higher-dimensional manifolds with edge singularities. Their methods are not far from the ones here, but our approach to regularity theory developed is simpler in many regards. Recently, Chen and Wang [2013] use quite different ideas to study the Kähler–Ricci flow on Kähler manifolds with edges.

We also mention the work of Rochon [2014], where a “propagation of polyhomogeneity” result is proved in the spirit of Theorem 1.2 but in the *complete* asymptotically hyperbolic setting; see also Albin, Aldana and Rochon [Albin et al. 2013], and also [Rochon and Zhang 2012] concerning a similar result in higher dimensions.

Finally, we make some remarks about the history of these results and of this particular work. The initial draft of this paper was completed in the Fall of 2011, though the work on it had started a few years before, and this material has been presented at conferences since then and announced in the survey article [Isenberg et al. 2011]. The appearance of this final draft was held up by other commitments of the authors, as well as our efforts to obtain the most incisive results possible. We now comment on the relationship between this work and other recent papers. These recent works include Yin’s original [2010] paper and his very recent follow-up [2013]; these certainly have substantial overlap with the present work, although our more detailed treatment of the linear and nonlinear regularity theory should be useful in further and more refined investigations of this problem. In addition, some time ago we were informed that D. Ramos had obtained results on this problem, relying on the short-time existence results in [Yin 2010]. His work was done independently of ours and has many points of overlap as well, though we have not seen details beyond what is contained in [Ramos 2011; 2013]. We acknowledge some very interesting and helpful conversations with him, clarifying his work, shortly before this paper was initially posted. Finally, we mention the very recent paper by Chen and Wang [2013], which has made substantial inroads into the higher-dimensional Kähler–Ricci flow in the presence of edge singularities using rather different methods that do not give higher regularity, and the announcement of Tian and Zhang [2013] concerning the Hamilton–Tian conjecture in the smooth setting in dimension three.

3. Linear estimates and existence of the flow

We now review some of the basic theory of the Laplacian and its associated heat operator on manifolds with conic singularities. For brevity, we focus entirely on the two-dimensional case. The main part of this section is an extension of standard parabolic regularity estimates to this conic setting; the main goal is a refined regularity result which is necessary for understanding our particular geometric problem. These estimates also lead directly to a proof of short-time existence.

3A. Elliptic operators on conic manifolds. Let g be a metric on a compact two-dimensional surface M with a finite number of conic singularities; in fact, to simplify the discussion below, assume that there is only one conic point, p . Write $g = e^\phi g_0$, where g_0 is smooth and exactly conic near p . We now study some analytic properties of the operator $\Delta_g + V$, where g and V have some specified Hölder regularity. Since

$$(\Delta_g + V)u = (e^{-\phi}(\Delta_{g_0} + e^\phi V)u = f \implies (\Delta_{g_0} + e^\phi V)u = e^\phi f,$$

we may as well replace g by g_0 and the potential V by $e^\phi V$, and hence it suffices to study operators of the form $\Delta_g + V$, where g is smooth and exactly conic, and V satisfies an appropriate Hölder condition.

We use tools from geometric microlocal analysis to study elliptic operators on surfaces with cone points. As references for these results, see the monograph by Melrose [1993] and the articles of Mazzeo [1991], Gil, Krainer and Mendoza [Gil et al. 2006], and [Jeffres et al. 2014, §3] for a more extended expository review. This approach takes advantage of the approximate homogeneity of the Laplacian of a conic metric of the cone point, as well as the resulting approximate homogeneity of the Schwartz kernels

of the corresponding Green functions. The strategy is to use these to obtain refined mapping properties of the operator, as well as regularity properties of its solutions.

In much of the following, it is convenient to replace the conic manifold M with a manifold with boundary \tilde{M} which is obtained by blowing up the conic point. This blowup procedure (which is described in more generality below) corresponds to introducing polar coordinates (r, y) around the conic point p and then replacing p by the circle $\{(0, y)\} = \{0\} \times S^1$ at $r = 0$. The space \tilde{M} is then given the smallest smooth structure for which these polar coordinate functions give a smooth chart.

3B. Function spaces. We first introduce various function spaces used later. The key to all these definitions is that it is advantageous to base them on differentiations with respect to the elements of $\mathcal{V}_b(\tilde{M})$, the space of all smooth vector fields on \tilde{M} which are unconstrained in the interior but tangent to the boundary. In local coordinates, any element of this space is a linear combination, with $C^\infty(\tilde{M})$ coefficients, of the vector fields $r\partial_r$ and ∂_y . Natural differential operators are built out of these; for example, the Laplacian of an exactly conic metric with cone angle $2\pi\beta$ takes the form

$$\Delta_\beta = r^{-2}((r\partial_r)^2 + \beta^{-2}\partial_y^2)$$

near p , where $y \in S^1_{2\pi}$. In other words, up to the factor r^{-2} , this is an elliptic combination (sum of squares) of the basis elements of \mathcal{V}_b .

Now define

$$C^k_b(\tilde{M}) = \{u : V_1 \cdots V_\ell u \in C^0(\tilde{M}) \text{ for all } \ell \leq k \text{ and } V_j \in \mathcal{V}_b(M)\}.$$

Because these spaces are based on differentiating by elements of \mathcal{V}_b , observe that C^k_b contains functions like $r^\zeta \psi(y)$, where $\psi \in C^k(S^1)$ and $\text{Re } \zeta > 0$. We also use the corresponding family of b -Hölder spaces $C^{k+\delta}_b(\tilde{M})$. The space $C^\delta_b(\tilde{M})$ consists of functions ϕ such that $\|\phi\|_{b,\delta} := \sup |\phi| + [\phi]_{b,\delta} < \infty$, where this Hölder seminorm is the ordinary one away from $\partial\tilde{M}$, while, in a neighbourhood $\mathcal{U} = \{r < 2\}$,

$$[\phi]_{b,\delta,\mathcal{U}} = \sup_{(r,y) \neq (r',y')} \frac{|\phi(r,y) - \phi(r',y')|(r+r')^\delta}{|r-r'|^\delta + (r+r')^\delta|y-y'|^\delta}.$$

Observe that, if we decompose \mathcal{U} into a union of overlapping dyadic annuli, $\bigcup_{\ell \geq 0} A_\ell$, where each $A_\ell = \{(r,y) : 2^{-\ell-1} \leq r \leq 2^{-\ell+1}\}$, then this seminorm (for functions supported in \mathcal{U}) is equivalent to the supremum over ℓ of the Hölder seminorm on each annulus,

$$[\phi]_{\delta,\mathcal{U}} \approx \sup_{\ell \geq 0} [\phi]_{\delta,A_\ell}. \tag{3-1}$$

Said differently, the seminorm can be computed assuming $\frac{1}{2} \leq r/r' \leq 2$. To verify this, simply note that, if $(r,y) \in A_\ell$ and $(r',y') \in A_{\ell'}$ with $|\ell - \ell'| \geq 2$, then

$$\frac{|r-r'|}{|r+r'|} \approx 1,$$

so that

$$\frac{|\phi(r,y) - \phi(r',y')|(r+r')^\delta}{|r-r'|^\delta + (r+r')^\delta|y-y'|^\delta} \leq C \sup |\phi|$$

with C independent of ℓ and ℓ' .

We also let $\mathcal{C}_b^{k+\delta}(\tilde{M})$ consist of the space of ϕ such that $V_1 \cdots V_\ell \phi \in \mathcal{C}_b^\delta(\tilde{M})$ for all $\ell \leq k$, and where $V_j \in \mathcal{V}_b(\tilde{M})$ for every j ; finally, define $r^\gamma \mathcal{C}_b^{k+\delta}(\tilde{M}) = \{\phi = r^\gamma \psi : \psi \in \mathcal{C}_b^{k+\delta}(\tilde{M})\}$.

The intersection of all these spaces, $\bigcap_k \mathcal{C}_b^k(\tilde{M})$, is the space of conormal functions, denoted by $\mathcal{A}(\tilde{M})$. It contains the very useful subspace of polyhomogeneous functions, \mathcal{A}_{phg} . By definition, \mathcal{A}_{phg} consists of all conormal functions which admit asymptotic expansions of the form

$$\phi \sim \sum_{\text{Re } \gamma_j \nearrow \infty} \sum_{\ell=0}^{N_j} \phi_{j,\ell}(y) r^{\gamma_j} (\log r)^\ell.$$

Note that the conormality condition requires that each coefficient $\phi_{j,\ell}$ lies in $\mathcal{C}^\infty(S^1)$. As an important special case, $\mathcal{C}^\infty(\tilde{M}) \subset \mathcal{A}_{\text{phg}}(\tilde{M})$, since smoothness corresponds to demanding that the exponents in the expansion above are all nonnegative integers, i.e., $\gamma_j = j$ and $N_j = 0$ for all $j \geq 0$. Finally, define $\mathcal{A}^0(\tilde{M}) = \mathcal{A}(\tilde{M}) \cap L^\infty$ and $\mathcal{A}_{\text{phg}}^0(\tilde{M}) = \mathcal{A}_{\text{phg}}(\tilde{M}) \cap L^\infty(M)$.

A metric g is $\mathcal{C}_b^{k+\delta}$, conormal, polyhomogeneous or smooth if $g = u g_0$, where the background metric g_0 is smooth and exactly conic, and where the function u satisfies any one of these regularity conditions.

3C. Mapping properties. Suppose that $L = \Delta_g + V$, where both g and V are polyhomogeneous (and V is real-valued). There is a canonical self-adjoint realization of this operator, which we still denote by L , defined via the Friedrichs construction associated to the quadratic form $\int |\nabla u|^2 - V|u|^2 dA_g$ and core domain $\mathcal{C}_0^\infty(M \setminus \{p\})$. It is well known that the Friedrichs domain of L obtained from this construction is compactly contained in L^2 , so this operator has discrete spectrum. We let G denote its generalized inverse. As an operator on $L^2(\tilde{M})$, this satisfies

$$\Delta_g \circ G = G \circ \Delta_g = \text{Id} - \Pi, \tag{3-2}$$

where Π is the orthogonal projector onto the nullspace of L . Thus Π has finite rank, and a basic regularity theorem in the subject (see the references cited earlier) states that, if g and V are polyhomogeneous, then the range of Π , which is the nullspace of L , lies in \mathcal{A}_{phg} . When $V \equiv 0$, we have $\text{rank}(\Pi) = 1$ and Π projects onto the constant functions. We regard each of these integral operators as corresponding to a Schwartz kernel, which is an element of $\mathcal{D}'(\tilde{M} \times \tilde{M})$. The “integration”, or distributional pairing, is taken with respect to the density dA_g . In local coordinates this equals $r dr dy$; the reader should note that this is *not* the standard b -density $r^{-1} dr dy$ that is commonly used in setting up the b -calculus. The differences are minor and notational only.

In this subsection we apply the theory of b -pseudodifferential operators to describe the fine structure of the Schwartz kernel of G . There are many reasons for wanting to know this structure, beyond the simplest statement that G is bounded on L^2 . One example is that, once we know the pointwise structure of this Schwartz kernel, we can show that G and Π are bounded operators acting between certain weighted b -Hölder spaces. Since the equality of operators (3-2) remains true on these spaces as well, we deduce that the operator L is Fredholm between these weighted Hölder spaces as well as just on L^2 or Sobolev spaces. This is very helpful when studying nonlinear problems.

We are primarily interested in the mapping

$$L : \mathcal{C}_b^{2+\delta}(\tilde{M}) \longrightarrow \mathcal{C}_b^\delta(\tilde{M}). \tag{3-3}$$

This is unbounded because, for a general $u \in \mathcal{C}_b^{2+\delta}$, it need only be true that $\Delta_g u \in r^{-2}\mathcal{C}_b^\delta$. Thus the domain of (3-3) is

$$\mathcal{D}_b^\gamma(L) := \{u \in \mathcal{C}_b^{2+\delta}(\tilde{M}) : Lu = f \in \mathcal{C}_b^\delta(\tilde{M})\}, \tag{3-4}$$

which we call the *Friedrichs–Hölder domain* of L . This space is independent of the potential V . Indeed, if $u \in \mathcal{D}_b^\delta(L)$, then $\Delta_g u = f - Vu \in \mathcal{C}_b^\delta$, so $u \in \mathcal{D}_b^\delta(\Delta_g)$. Note finally that $\mathcal{D}_b^\delta(\Delta_g)$ is complete with respect to the Banach norm

$$\|u\|_{\mathcal{D}_b^\delta} := \|u\|_{\mathcal{C}_b^\delta} + \|\Delta_g u\|_{\mathcal{C}_b^\delta}.$$

An essentially tautological characterization of this space is that

$$\mathcal{D}_b^\delta(L) = \{u = Gf + w : f \in \mathcal{C}_b^\delta \text{ and } w \in \ker L \cap \mathcal{C}_b^{2+\delta}\}. \tag{3-5}$$

However, there is an even more explicit characterization of this space:

Proposition 3.1. *Suppose that $L = \Delta_g + V$ with $g, V \in \mathcal{C}_b^\delta$, and $u \in \mathcal{D}_b^\delta(L)$ satisfies $Lu = f \in \mathcal{C}_b^\delta(\tilde{M})$. Then*

$$u = a_0 + (a_{11} \cos y + a_{12} \sin y)r^{1/\beta} + \tilde{u},$$

where a_0, a_{11}, a_{12} are constants and $\tilde{u} \in r^2\mathcal{C}_b^{2+\delta}$. (Note that the middle term on the right can be absorbed into \tilde{u} if $\beta \leq \frac{1}{2}$.)

To explain the relevance of the terms in this expansion, recall that, using the exactly conic structure of g near the conic points, we have that, if $\gamma \in \mathbb{R}$ and $\phi \in C^\infty(S^1)$, then

$$\Delta_g r^\gamma \phi(y) = (\beta^{-2}\phi''(y) + \gamma^2\phi(y))r^{\gamma-2} \quad \text{and} \quad Vr^\gamma \phi(y) = \mathcal{O}(r^\gamma).$$

Thus, in terms of its formal action on Taylor series, Δ_g is the principal part. The operator Δ_g has special locally defined solutions $r^{j/\beta}(a_{j1} \cos(jy) + a_{j2} \sin(jy))$, and the terms in the statement of this result are simply those special solutions with exponent less than 2.

The L^2 version of this proposition is a special case of Theorem 7.14 in [Mazzeo 1991], and it is not hard to deduce the corresponding statement in these b -Hölder spaces from that. We sketch a direct proof below in Section 3E.

Remark 3.2. The higher-dimensional version of this decomposition for solutions of Schrödinger-type equations on manifolds with edges plays a crucial role in [Jeffres et al. 2014].

3D. Structure of the generalized inverse. We now describe the detailed structure of G . First recall the definition of conormal and polyhomogeneous distributions. We say that u is conormal of order γ on \tilde{M} , written $u \in \mathcal{A}^\gamma(\tilde{M})$, if $V_1 \cdots V_\ell u \in r^\gamma L^\infty$ for every $\ell \geq 0$ and all $V_j \in \mathcal{V}_b(M)$. Such a u is smooth away from the conic points. Next, let E be an *index set*, i.e., a discrete subset $\{(\gamma_j, p_j)\} \subset \mathbb{C} \times \mathbb{N}_0$ such that there are only finitely many pairs with γ_j lying in any half-plane $\text{Re } z < C$. We also assume that

$(\gamma_j, p_j) \in E$ implies that $(\gamma_j + \ell, p_j) \in E$ for every $\ell \in \mathbb{N}$. We then say that u is polyhomogeneous with index set E , written $u \in \mathcal{A}_{\text{phg}}^E(M)$, if $u \in \mathcal{A}'(\tilde{M})$ and

$$u \sim \sum_{(\gamma_j, p_j) \in E} \sum_{\ell \leq p_j} a_{j\ell}(y)r^{\gamma_j}(\log r)^\ell,$$

where each $a_{j\ell} \in C^\infty(S^1)$. Similarly, if X is any manifold with corners, then we can define the space of polyhomogeneous functions on X ; these have the same type of asymptotic expansion at all boundary faces and product-type expansions at the corners of X .

The reason for introducing polyhomogeneity is that the Schwartz kernel G is polyhomogeneous, not on $(\tilde{M})^2$, but rather on a certain manifold with corners $(\tilde{M})_b^2$ which is obtained by blowing up $(\tilde{M})^2$ along the codimension two corner $(\partial\tilde{M})^2$. This new space has three boundary hypersurfaces: two are lifts of the faces $\partial\tilde{M} \times \tilde{M}$ and $\tilde{M} \times \partial\tilde{M}$ and called the left and right faces, lf and rf, respectively, and the third is the *front face*, ff, which is the one produced by the blowup. There is a natural blowdown map $\mathfrak{b} : (\tilde{M})_b^2 \rightarrow (\tilde{M})^2$, and the precise statement is that $G = (\mathfrak{b})_* K_G$, where K_G is polyhomogeneous on $(\tilde{M})_b^2$, with an additional conormal singularity along the lifted diagonal in $(\tilde{M})_b^2$.

There are several useful coordinate systems on $(\tilde{M})_b^2$. Using coordinates (r, y) near the boundary on the first copy of \tilde{M} and an identical set (r', y') on the second copy, this blowup is tantamount to introducing the polar coordinates $r = R \cos \theta, r' = R \sin \theta$ and replacing the corner $\{r = r' = 0\}$ by the hypersurface $\{R = 0, \theta \in [0, \pi/2]\}$. Thus lf corresponds to $\theta = \pi/2$, rf corresponds to $\theta = 0$, and the front face ff corresponds to $R = 0$. The lifted diagonal is the submanifold $\{\theta = \pi/4, y = y'\}$. If $\mathcal{E} = (E_{\text{lf}}, E_{\text{rf}})$ is a pair of index sets, the first for lf and the second for rf, then we say that a pseudodifferential operator A lies in the space $\Psi_b^{-\infty, r, \mathcal{E}}(\tilde{M})$ if the lift K_A of its Schwartz kernel to $(\tilde{M})_b^2$ lies in $\mathcal{A}_{\text{phg}}^{r, \mathcal{E}}((\tilde{M})_b^2)$, where the initial superscript r indicates that $K_A = R^{r-2}K'_A$, where K'_A is C^∞ up to the front face and is polyhomogeneous at the side faces with index sets E_{lf} and E_{rf} , respectively. Finally, $A \in \Psi_b^{m, r, \mathcal{E}}(\tilde{M})$ if $K_A = R^{r-2}(K'_A + K''_A)$, where the first term lies in $\Psi_b^{-\infty, r, \mathcal{E}}$ and K''_A is supported in a small neighbourhood of the lifted diagonal, and in particular vanishes near $\text{lf} \cup \text{rf}$, has a conormal singularity of pseudodifferential order m along the lifted diagonal (so its Fourier transform on the fibres of the normal bundle to the lifted diagonal is a symbol of order $-2 + m$), and is smoothly extendible across the front face. The reason for the slightly odd normalization of the singularity along ff is to make the identity operator an element of $\Psi_b^{0, 0, \emptyset, \emptyset}(M)$. Indeed, relative to the measure $r'dr'dy'$, the Schwartz kernel of Id is $r^{-1}\delta(r - r')\delta(y - y')$, and this lifts to $(\tilde{M})_b^2$ as $R^{-2}\delta(\theta - \pi/4)\delta(y - y')$.

If g is a smooth conic metric and $\beta \notin \mathbb{Q}$, then the index set for the expansion of K_G at lb and rb is

$$E = \left\{ \left(\frac{j}{\beta} + \ell, 0 \right) : j, \ell \in \mathbb{N}_0, (j, \ell) \neq (0, 1) \right\}.$$

This excluded element $(0, 1)$ corresponds to requiring that the expansion not include the term $\log r$. If β is rational, or if g is only polyhomogeneous, then we are able to state that the generalized inverse G lies in $\Psi_b^{-2, 2, E', E'}(M)$ for some index set E' , which may contain extra terms, including log terms, high up in the index set; however, the initial part of this index set (and hence the exponents in the initial part of the expansion of any solution) up to order 2 remains the same. The fact that the index r in the general definition

equals 2 for the particular kernel K_G turns out to be very helpful. This correspond to precisely the order of approximate homogeneity needed to compensate for the fact that the identity operator behaves like R^{-2} at the front face, and Δ_g is approximately homogeneous of order 2. The index sets of G at the left and right faces are equal to one another because G is a symmetric operator. The fact that E does not contain the term $(0, 1)$ is because G is the generalized inverse for the Friedrichs extension. It can also be verified by direct calculation that, in fact, E does not contain the element $(1, 0)$, for, if it did, then we could produce a poly-homogeneous element $u = Gf$ in the Friedrichs domain which contains a term $u_1(y)r$; this holds because $\Delta_g r = \mathcal{O}(r^{-1})$. We refer to [Jeffres et al. 2014, §3] for a more careful description of all of these facts.

Let us say that $A \in \Psi_b^{m,r,\mathcal{E}}$ is of nonnegative type if $m \leq 0, r \geq 0$, all the terms (γ, s) in the index sets E_{lf} and E_{rf} are nonnegative and, if $(0, s)$ lies in either index set, then $s = 0$. Proposition 3.27 in [Mazzeo 1991] implies that, if A is of nonnegative type, then $A : \mathcal{C}_b^{0,\delta} \rightarrow \mathcal{C}_b^{0,\delta}$ is a bounded mapping.

3E. Mapping properties, revisited. We are now ready for:

Proof of Proposition 3.1. Rewrite $Lu = f$ as $\Delta_g u = f - Vu := \tilde{f} \in \mathcal{C}_b^\delta$. Let G denote the generalized inverse of the Friedrichs extension of Δ_g , so that $u = G\tilde{f} - \Pi u$; since Πu is a constant, we can concentrate on the first term.

Decompose the Schwartz kernel of G into a sum $G' + G''$, where G' is supported in a small neighbourhood of the lifted diagonal of \tilde{M}_b^2 (and hence vanishes near $\text{lf} \cup \text{rf}$), and $G'' \in \mathcal{A}_{\text{phg}}(\tilde{M}_b^2)$; see Section 3F3, where the parabolic version of this decomposition is described more carefully. Since $G' \in \Psi_b^{-2,2,\emptyset,\emptyset}$, we can write $G' = r^2 \hat{G}'$, where $\hat{G}' \in \Psi_b^{-2,0,\emptyset,\emptyset}$ and hence is nonnegative. Since $\hat{G}' \tilde{f} \in \mathcal{C}_b^{2+\delta}$, we obtain that $u' \in r^2 \mathcal{C}_b^{2+\delta}$.

Turning now to u'' , first observe that $r\partial_r$ and ∂_y lift to the left factor of $(\tilde{M})_b^2$ as smooth vector fields on \tilde{M}_b^2 that are tangent to all boundaries. It follows that $(r\partial_r)^j \partial_y^\ell G'' \in \Psi_b^{-\infty,2,0,0}$ for all $j, \ell \geq 0$, from which it follows that $u'' \in \mathcal{A}^0(\tilde{M})$. Moreover, the initial part of the expansion G — and hence of G'' — at rf takes the form $A_0 r^0 + (A_{11} \cos y + A_{12} \sin y) r^{1/\beta} + \mathcal{O}(r^2)$, which means that the kernel $(r\partial_r - \beta^{-1})(r\partial_r) \circ G$ is not only of nonnegative type (and of pseudodifferential order $-\infty$), but in fact vanishes to order 2 at rf . Since G'' already vanishes to this order at ff , we can remove a factor of r^2 , i.e., write $(r\partial_r - \beta^{-1})r\partial_r \circ G'' = r^2 \hat{G}''$, where \hat{G}'' is of nonnegative type and smoothing. This means that

$$(r\partial_r)(r\partial_r - \beta^{-1})u'' \in r^2 \mathcal{A}^0(\tilde{M}).$$

Integrating in r gives that $u'' = a_0(y) + a_1(y)r^{1/\beta} + r^2 \mathcal{A}^0$. Finally, since $\Delta_g u''$ is bounded, we conclude that a_0 is constant and $a_1(y) = a_{11} \cos y + a_{12} \sin y$, as claimed. □

We conclude this discussion with the following application of Proposition 3.1 to our geometric problem.

Proposition 3.3. *Let g_0 be a conic metric and suppose that its scalar curvature R_{g_0} lies in \mathcal{C}_b^δ and, in particular, is bounded near the conic points. If $g = e^\phi g_0$ is another conformally related metric, with $\phi \in \mathcal{C}_b^{2+\delta}$, then $R_g \in \mathcal{C}_b^\delta$ if and only if*

$$\phi = c_0 + r^{1/\beta}(c_{11} \cos y + c_{12} \sin y) + \tilde{\phi}, \quad \tilde{\phi} \in r^2 \mathcal{C}_b^{2+\delta},$$

or, more succinctly, $\phi \in \mathcal{D}_b^\delta(\tilde{M})$.

Proof. Apply the generalized inverse G for the Friedrichs extension of Δ_{g_0} to the curvature transformation equation

$$\Delta_{g_0}\phi = R_{g_0} - \frac{1}{2}R_g e^\phi$$

to get

$$\phi = \Pi\phi + G\left(R_{g_0} - \frac{1}{2}R_g e^{2\phi}\right).$$

Suppose now that $R_g \in \mathcal{C}_b^\delta$. The first term $\Pi\phi$ is just a constant, while, by Proposition 3.1, $G\left(R_{g_0} - \frac{1}{2}R_g e^{2\phi}\right)$ has an expansion up to order r^2 .

On the other hand, if ϕ has an expansion as in the statement of this proposition, then $R_g \in \mathcal{C}_b^\delta$. □

Remark 3.4. The results in 3A–3E are special cases of the ones in [Jeffres et al. 2014, §3], which are proved for Kähler manifolds of arbitrary dimension. We have presented this material in some detail since the statements and proofs in the Riemann surface case are simpler than in higher dimensions, and also because the discussion above sets the stage for the derivation of the parabolic estimates, which occupies the remainder of this section.

3F. Parabolic Schauder estimates. We now turn to the parabolic problem, and in particular to the analogue of Proposition 3.1.

Let (M, g) be a smooth exactly conic metric with cone angle $2\pi\beta < 2\pi$, and set $L = \Delta_g + V$, where V is polyhomogeneous; later we relax this to assume that $V \in \mathcal{C}_b^\delta$. We are interested in the homogeneous and inhomogeneous problems

$$\begin{cases} (\partial_t - L)v = 0, \\ v(0, z) = \phi(z), \end{cases} \quad \text{and} \quad \begin{cases} (\partial_t - L)u = f, \\ u(0, z) = 0, \end{cases} \tag{3-6}$$

for which the solutions can be represented as

$$v(t, z) = \int_M H(t, z, z')\phi(z') dA_g(z'), \tag{3-7}$$

$$u(t, z) = \int_0^t \int_M H(t - t', z, z')f(t', z') dt' dA_g; \tag{3-8}$$

here H is the heat kernel associated to L . In order to study the regularity properties of the solution u , we describe a fine structure theorem for H , similar to the one for the Green function G above. This leads to a definition of parabolic weighted Hölder spaces, and finally a derivation of the estimates for solutions in these spaces. As in the previous section, we work exclusively with the Friedrichs extension of the Laplacian.

3F1. Structure of the heat kernel. Denote by g_β the complete flat conic metric $dr^2 + \beta^2 r^2 dy^2$ and by Δ_β its Laplacian. The first observation is that the model heat operator $\partial_t - \Delta_\beta$ is homogeneous with respect to the dilation $(t, r, y) \mapsto (\lambda^2 t, \lambda r, y)$, $\lambda > 0$, and hence, if H_β is the heat kernel associated to (the Friedrichs realization of) Δ_β , then

$$H_\beta(\lambda^2 t, \lambda r, y, \lambda r', y') = \lambda^{-2} H_\beta(t, r, y, r', y'). \tag{3-9}$$

In fact, there are explicit expressions:

$$\begin{aligned} H_\beta(t, r, y, r', y') &= \frac{1}{\pi} \sum_{\ell=0}^{\infty} \left(\int_0^\infty e^{-\lambda^2 t} J_{\ell/\alpha}(\lambda r) J_{\ell/\alpha}(\lambda r') \lambda \, d\lambda \right) \cos \ell(y - y') \\ &= \sum_{\ell=0}^{\infty} \frac{1}{t} \exp\left(\frac{-(r^2 + (r')^2)}{2t}\right) I_{\ell/\alpha}\left(\frac{rr'}{2t}\right) \cos \ell(y - y'). \end{aligned}$$

These expressions are better suited for studying the action of H_β on L^2 Sobolev spaces than weighted Hölder spaces, so, just as for the operator G earlier, we describe this model heat kernel, and then the true heat kernel, using the language of blowups and polyhomogeneous distributions. This structure theory for the Laplacian on a conic space appears in [Mooers 1999], with basic mapping properties later determined by Jeffres and Loya [2003].

The function $H(t, z, z')$ is a distribution on $\mathbb{R}^+ \times (\tilde{M})^2$, but the key point is that its lift to the “conic heat space” $(\tilde{M})_h^2$ is polyhomogeneous. This will be obvious for the model heat kernel H_β once we define $(\tilde{M})_h^2$ and, conversely, starting from the ansatz that this lift is polyhomogeneous, we can construct (the lift of) H as a polyhomogeneous object by standard heat operator parametrix methods.

The conic heat space is defined, starting from $\mathbb{R}^+ \times (\tilde{M})^2$, through a sequence of blowups. The first step is to blow up the corner $r = r' = t = 0$, with a parabolic homogeneity in the variable t , and, following that, to blow up the diagonal in $(\tilde{M})^2$ at $t = 0$. The first blowup is tantamount to introducing the parabolic spherical coordinates $\rho \geq 0$ and $\omega = (\omega_0, \omega_1, \omega_2) \in S^2_+ = S^2 \cap (\mathbb{R}^+)^3$, where

$$\rho = \sqrt{t + r^2 + (r')^2} \quad \text{and} \quad \omega = \left(\frac{t}{\rho^2}, \frac{r}{\rho}, \frac{r'}{\rho} \right).$$

Thus ρ, ω, y, y' are nondegenerate local coordinates near the new face created by this first step. For the second blowup we use the coordinates

$$R = \sqrt{t + |z - z'|^2}, \quad \theta = \frac{z - z'}{R}, \quad z',$$

where z is any interior coordinate system and z' an identical chart on the second copy of \tilde{M} . This sequence of blowups is summarized by the notation

$$M_h^2 := [\mathbb{R}^+ \times \tilde{M}^2; \{0\} \times (\partial \tilde{M})^2, \{dt\}; \{0\} \times \text{diag}_{\tilde{M}}, \{dt\}].$$

This manifold with corners has five boundary faces (see Figure 1): the left and right faces $\text{lf} = \{\omega_2 = 0\}$ and $\text{rf} = \{\omega_1 = 0\}$, which are the lifts of the faces $r' = 0$ and $r = 0$, respectively; the front face $\text{ff} = \{\rho = 0\}$; the temporal diagonal $\text{td} = \{R = 0\}$, which covers the diagonal at $t = 0$, and bf , the original bottom face at $t = 0$ away from the diagonal.

The construction in [Mooers 1999] shows that H is polyhomogeneous on $(\tilde{M})_h^2$ with index set $E = \{(j/\beta, 0) : j \in \mathbb{N}_0\}$ at the left and right faces; note that these are exactly the same as the index sets for the Green function G at the corresponding faces. The kernel H vanishes to infinite order at bf , while at td it has an expansion in powers of R , starting with R^{-2} (in general, this is $R^{-\dim M}$). Finally, at ff it

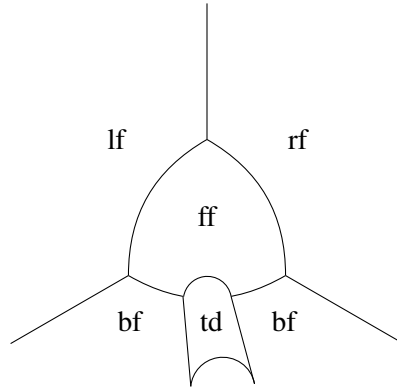


Figure 1. The faces of the manifold with corners.

has an expansion in integer powers of ρ , beginning with ρ^{-1} . The leading coefficient of the expansion at this face is precisely the model heat kernel H_β .

3F2. Function spaces. We now describe a family of function spaces commonly used in parabolic problems. We refer to [Lunardi 1995, Chapter 5] for a careful description of these (in the setting of interior and standard boundary problems). In the definitions and discussion below, we first introduce a scale of fully dilation-invariant spaces (jointly in the variables (t, r)), where the parabolic estimates are obtained by using scaling arguments to reduce to standard interior parabolic estimates. After that, we refine the estimates to obtain the maximal expected regularity in t .

First, for $0 < \delta < 2$, define $\mathcal{C}_{b0}^{0,\delta/2}([0, T] \times \tilde{M})$ to consist of all $u \in \mathcal{C}^0([0, T] \times \tilde{M})$ for which $u(\cdot, z) \in \mathcal{C}^{\delta/2}([0, T])$ for all $z \in \tilde{M} \setminus \partial\tilde{M}$ and

$$[u]_{b0;0,\delta/2} := \sup_z r^\delta [u(\cdot, z)]_{\delta/2, [0, T]} < \infty; \quad (3-10)$$

by contrast, the standard Hölder space in t , $\mathcal{C}^{0,\delta/2}([0, T] \times \tilde{M})$ is defined using the usual seminorm

$$[u]_{0,\delta/2} := \sup_z [u(\cdot, z)]_{\delta/2, [0, T]}$$

(without the extra weight factor r^δ). Next, spatial regularity is measured using the spaces

$$\mathcal{C}_b^{\delta,0}([0, T] \times \tilde{M}) = \{u \in \mathcal{C}^0([0, T] \times \tilde{M}) : u(t, \cdot) \in \mathcal{C}_b^\delta(\tilde{M}) \text{ for all } t \in [0, T]\},$$

where the norm is $\|u\|_{b;\delta,0} = \sup_t \|u(t, \cdot)\|_{b;\delta}$. We still let $0 < \delta < 2$, with the understanding that if $\delta = 1$ then this is the Zygmund space (so that interpolation arguments can be used). For simplicity below we omit discussion of this special case. Taking intersections yields the two natural parabolic Hölder spaces

$$\mathcal{C}_{b0}^{\delta,\delta/2}([0, T] \times \tilde{M}) = \mathcal{C}_{b0}^{0,\delta/2}([0, T] \times \tilde{M}) \cap \mathcal{C}_b^{\delta,0}([0, T] \times \tilde{M}), \quad (3-11a)$$

$$\mathcal{C}_b^{\delta,\delta/2}([0, T] \times \tilde{M}) = \mathcal{C}^{0,\delta/2}([0, T] \times \tilde{M}) \cap \mathcal{C}_b^{\delta,0}([0, T] \times \tilde{M}). \quad (3-11b)$$

Thus, functions in $C_{b0}^{\delta, \delta/2}$ have no regularity in t at $r = 0$, while functions in $C_b^{\delta, \delta/2}$ satisfy the ordinary Hölder regularity in t even at $r = 0$. The seminorms on these two spaces agree away from p , while, in a neighbourhood \mathcal{U} of this conic point, these seminorms are described as follows. Decomposing \mathcal{U} into a countable union of dyadic annuli, $\bigcup_{\ell \geq 0} A_\ell$, we have

$$[u]_{b0; \delta, \delta/2, \mathcal{U}} = \sup_{\ell \in \mathbb{N}_0} \sup_{|t-t'| < 2^{-2\ell}} \sup_{z, z' \in A_\ell} \frac{|u(t, r, y) - u(t', r', y')|(r+r')^\delta}{|r-r'|^\delta + |t-t'|^{\delta/2} + (r+r')^\delta |y-y'|^\delta}$$

and

$$[u]_{b; \delta, \delta/2, \mathcal{U}} = \sup_{t, t'} \sup_{\ell \in \mathbb{N}_0} \sup_{z, z' \in A_\ell} \frac{|u(t, r, y) - u(t', r', y')|(r+r')^\delta}{|r-r'|^\delta + (r+r')^\delta (|t-t'|^{\delta/2} + |y-y'|^\delta)}.$$

These seminorms are equivalent to

$$\sup_{(t, z) \neq (t', z')} \frac{|u(t, z) - u(t', z')| \max\{r(z)^\delta, r'(z')^\delta\}}{|t-t'|^{\delta/2} + \text{dist}_g(z, z')^\delta}$$

and

$$\sup_{(t, z) \neq (t', z')} \frac{|u(t, z) - u(t', z')| \max\{r(z)^\delta, r'(z')^\delta\}}{|t-t'|^{\delta/2} \max\{r(z)^\delta, r'(z')^\delta\} + \text{dist}_g(z, z')^\delta},$$

respectively, where the radial function r has been extended from \mathcal{U} to the rest of \tilde{M} to be smooth and strictly positive.

We also define higher regularity versions of these spaces,

$$C_{b0}^{k+\delta, (k+\delta)/2}([0, T] \times \tilde{M}) \quad \text{and} \quad C_b^{k+\delta, (k+\delta)/2}([0, T] \times \tilde{M}),$$

where k is an even positive integer and $0 < \delta < 2$. The former space consists of functions u such that $V_1 \cdots V_i (r^2 \partial_t)^j u \in C_{b0}^{\delta, \delta/2}$ for $i + 2j \leq k$, where every $V_\ell \in \mathcal{V}_b(\tilde{M})$, while the latter consists of all u such that $V_1 \cdots V_i \partial_t^j u \in C_b^{\delta, \delta/2}$ for $i + 2j \leq k$ and every $V_\ell \in \mathcal{V}_b(\tilde{M})$. As before, these are Zygmund spaces when $\delta = 1$. We also introduce weighted versions of these spaces, $r^\gamma C_*^{k+\delta, (k+\delta)/2}$, $*$ = $b0$ or b . For later reference, for the same ranges of k and δ , $C_*^{0, (k+\delta)/2}([0, T] \times \tilde{M})$ is the space of functions u with $\partial_t^j u \in C_*^{0, \delta/2}([0, T] \times \tilde{M})$ for $2j \leq k$.

Finally, we define the analogues of the Friedrichs–Hölder domain:

$$\mathcal{D}_*^{\delta, \delta/2}([0, T] \times \tilde{M}) = \{u \in C_*^{\delta, \delta/2} : \Delta u \in C_*^{\delta, \delta/2}([0, T] \times \tilde{M})\}, \quad * = b0 \text{ or } * = b,$$

again with the higher regularity analogues.

If $h(t, r, y) \in C_{b0}^{k+\delta, (k+\delta)/2}$ is supported in $\mathbb{R}^+ \times \mathcal{U}$, then the rescaled function $h_\lambda(t, r, y) = h(\lambda^2 t, \lambda r, y)$ satisfies

$$\|h_\lambda\|_{b0; k+\delta, (k+\delta)/2, \gamma} = \lambda^\gamma \|h\|_{b0; k+\delta, (k+\delta)/2, \gamma}$$

(the final subscript in the norms indicates the weight factor). In other words, these spaces are compatible with the approximate dilation invariance of the heat operator $\partial_t - L$, which means that we will be able to prove the basic a priori estimates on them by exploiting this scaling. On the other hand, it is clearly important to obtain better regularity of solutions in t near $r = 0$. We obtain estimates on the b -spaces

starting from the estimates on the b_0 -spaces and using induction and interpolation. Note that the analogue of (3-11b) is not true when $k > 0$; namely, there is a proper inclusion

$$\mathcal{C}_b^{k+\delta, (k+\delta)/2} \subsetneq \mathcal{C}_{b_0}^{k+\delta, (k+\delta)/2} \cap \mathcal{C}^{0, (k+\delta)/2}, \quad k > 0.$$

3F3. Estimates. The basic Hölder estimates for the homogeneous problem were already determined by Jeffres and Loya [2003].

Proposition 3.5. *Suppose that $\phi \in \mathcal{C}_b^{k+\delta}(\tilde{M})$ and*

$$(\partial_t - L)v = 0, \quad v|_{t=0} = \phi.$$

Then $v \in \mathcal{C}_b^{k+\delta, (k+\delta)/2}([0, T] \times \tilde{M})$ and, furthermore, $v(t, \cdot) \in \mathcal{A}_{\text{phg}}(\tilde{M}) \cap \mathcal{D}_b^{0, \delta}(\tilde{M})$ for all $t > 0$.

The proof in [Jeffres and Loya 2003] of the first assertion here proceeds by direct and rather intricate estimates in various local coordinate systems, but they do not consider the issue of membership in $\mathcal{D}_b^{0, \delta}$. The polyhomogeneity of v when $t > 0$ is immediate from the polyhomogeneous structure of H on M_h^2 ; also, $v \in \mathcal{D}_b^{0, \delta}$ implies that $v(t, \cdot) \sim c_0(t) + (c_{11}(t) \cos y + c_{12}(t) \sin y)r^{1/\beta}$ as $r \rightarrow 0$; using polyhomogeneity again, these coefficients are smooth when $t > 0$.

There are a couple of variants of the inhomogeneous problem, depending on the regularity assumptions placed on f . We start with the version in dilation-invariant spaces.

Proposition 3.6. *Let $f \in \mathcal{C}_{b_0}^{k+\delta, (k+\delta)/2}([0, T] \times \tilde{M})$ and suppose that u is the unique solution in the Friedrichs domain to $(\partial_t - L)u = f$, $u|_{t=0} = 0$. Then $u \in \mathcal{C}_{b_0}^{k+2+\delta, (k+2+\delta)/2}([0, T] \times \tilde{M})$ and*

$$\|u\|_{b_0; k+2+\delta, (k+2+\delta)/2} \leq C \|f\|_{b_0; k+\delta, (k+\delta)/2}, \tag{3-12}$$

where C is a constant independent of u and f . In addition,

$$u(t, z) = \hat{u}(t, z) + \tilde{u}(t, z), \quad \text{where } \tilde{u} \in r^2 \mathcal{C}_{b_0}^{k+2+\delta, (k+2+\delta)/2}(\tilde{M}) \text{ and } \hat{u}(t, z) \in \bigcap_{\ell \geq 0} \mathcal{C}_{b_0}^{2\ell, \ell}.$$

The proof of this, which relies on the approximate homogeneity structure of H , adapts readily to the homogeneous case too, and gives a new proof of Proposition 3.5 which is conceptually simpler than the one in [Jeffres and Loya 2003].

Proof. Write u as in (3-8). We analyze this integral by decomposing H into a sum of two terms, as follows. Choose a smooth nonnegative cutoff function $\chi = \chi^{(1)}(\rho)\chi^{(2)}(\omega)$ on M_h^2 , where $\chi^{(1)}(\rho)$ equals 1 for $\rho \leq 1$ and vanishes for $\rho \geq 2$, and $\chi^{(2)}(\omega)$ has support in $\{\frac{1}{2} \leq \omega_1/\omega_2 \leq 2, \omega_0 \leq \frac{1}{2}\}$ and equals 1 near $(0, 1/\sqrt{2}, 1/\sqrt{2})$ (which is where the diagonal $\{t = 0, r = r'\}$ intersects ff). Note that χ is (locally) invariant under the parabolic dilations $(t, r, y, r', y') \mapsto (\lambda^2 t, \lambda r, y, \lambda r', y')$. Then set

$$H = H_0 + H_1, \quad H_0 = (1 - \chi(\rho, \omega))H, \quad H_1 = \chi(\rho, \omega)H,$$

and

$$u = u_0 + u_1, \quad u_j = H_j \star f, \quad j = 0, 1.$$

We study u_1 first. Introduce a partition of unity $\{\psi_\ell\}$ relative to the covering $\mathcal{U} = \bigcup A_\ell$; for example, take $\psi_\ell(r) = \psi(2^\ell r)$, where $\psi(r) \in C_0^\infty((\frac{1}{4}, 4)) \geq 0$ equals 1 for $\frac{1}{2} \leq r \leq 1$ and is chosen so that $\sum_{\ell \geq 0} \psi(2^\ell r) = 1$ for $0 < r \leq 1$. Now write

$$f = \sum f_\ell(t, r, y), \quad f_\ell = \psi_\ell f, \quad \text{and} \quad u_{1\ell} = H_1 \star f_\ell.$$

Thus f_ℓ has support in $\mathbb{R}^+ \times A_\ell$, while the support of $u_{1\ell}$ lies in $\mathbb{R}^+ \times (A_{\ell-1} \cup A_\ell \cup A_{\ell+1})$. We can also assume that f_ℓ is supported in some time interval $[\tau, \tau + 2^{2-2\ell}]$, since if $|t - t'| > (r + r')^2$ then the b -Hölder seminorm can be estimated by $C \sup |f_\ell|$. By the support properties of H_1 , $u_{1\ell}$ is supported in a time interval of at most twice this length. We replace t by $t - \tau$ without further comment.

Fix $\ell \in \mathbb{N}_0$ and let $\lambda = 2^{\ell-1}$; for any function h , define $(D_\lambda h)(\bar{t}, \bar{r}, y) = h(\lambda^{-2}\bar{t}, \lambda^{-1}\bar{r}, y)$. Thus, if h is supported in A_ℓ , then $D_\lambda h$ is supported in $A_1 := \{(\bar{t}, \bar{r}, y) : \frac{1}{4} \leq \bar{r} \leq 1\}$. In particular, $D_\lambda f_\ell$ and $D_\lambda u_{1\ell}$ are supported in $[0, 1] \times A_1$ and $[0, 1] \times (A_0 \cup A_1 \cup A_2)$, respectively. We shall use that $\|D_\lambda u_{1\ell}\|_{b0; k+2+\delta, (k+2+\delta)/2} = \|u_{1\ell}\|_{k+2+\delta, (k+2+\delta)/2}$, and similarly for $D_\lambda f_\ell$.

For convenience in the next few paragraphs, drop the indices ℓ and 1, and simply write $D_\lambda u = u_\lambda$, $D_\lambda f = f_\lambda$. Since it also just complicates the notation, we also assume that $k = 0$. Using these conventions, change variables in $u = H_1 \star f$ by setting

$$\bar{t} = \lambda^2 t, \quad \hat{t} = \lambda^2 t', \quad \bar{r} = \lambda r, \quad \hat{r} = \lambda r'.$$

This yields

$$u_\lambda(\bar{t}, \bar{r}, y) = \int_0^{\bar{t}} \int \lambda^{-4} H_1(\lambda^{-2}(\bar{t} - \hat{t}), \lambda^{-1}\bar{r}, y, \lambda^{-1}\hat{r}, y') f_\lambda(\hat{t}, \hat{r}, y') \hat{r} d\hat{r} dy' d\hat{t}.$$

For simplicity we have replaced the measure $dA_g dt'$ in the initial integral by $r' dr' dy' dt'$.

The key point is that the polyhomogeneous structure of H_1 on M_h^2 implies that the family of dilated kernels

$$(H_1)_\lambda(\bar{t} - \hat{t}, \bar{r}, y, \hat{r}, y') := \lambda^{-2} H_1(\lambda^{-2}(\bar{t} - \hat{t}), \lambda^{-1}\bar{r}, y, \lambda^{-1}\hat{r}, y')$$

converges in \mathcal{A}_{phg} on the portion of the heat space with $\bar{r}, \hat{r} \in [\frac{1}{4}, 4]$ as $\lambda \rightarrow \infty$. In fact, its limit is simply the heat kernel for the model operator Δ_β on the complete warped product cone restricted to this range of radial variables. Since this region remains away from the vertex, we invoke the classical parabolic Schauder estimates to deduce that, as an operator between ordinary parabolic Hölder spaces, the norm of $(H_1)_\lambda$ restricted to functions supported in $[0, 1] \times (A_0 \cup A_1 \cup A_2)$ is uniformly bounded in λ . Hence, comparing the last two displayed formulae, we see that

$$\|u_\lambda\|_{b0; 2+\delta, 1+\delta/2} \leq C \lambda^{-2} \|f_\lambda\|_{b0; \delta, \delta/2} \implies \|r^{-2} u_\lambda\|_{b0; 2+\delta, 1+\delta/2} \leq C \|f_\lambda\|_{b0; \delta, \delta/2}$$

with C independent of λ . Restoring the indices, and using the fact that, analogous to (3-1),

$$\|h\|_{b0; k\delta, (k+\delta)/2} \approx \sup_\ell \|h\|_{b0; k+\delta, (k+\delta)/2}$$

for any function h and any $k \in \mathbb{N}_0$, we conclude finally that

$$\|r^{-2} u_1\|_{b0; 2+\delta, 1+\delta/2} \leq C \|f\|_{b0; \delta, \delta/2}, \tag{3-13}$$

so $u_1 \in r^2 C_{b_0}^{2+\delta, (1+\delta)/2}$.

We now turn to the estimate for $u_0 = H_0 \star f$, which is the same as the function \hat{u} in the statement of the theorem. The polyhomogeneous structure of H_0 is slightly simpler than that for H ; indeed, H_0 vanishes to infinite order not only along bf but along td as well. This means that H_0 is polyhomogeneous on the space obtained from M_h^2 by blowing down td . We first claim that $\|H_0 \star f\|_{C^0} \leq C \|f\|_{C^0}$. The proof reduces immediately to verifying that $\int_0^t \int_M H_0(t-s, r, y, r', y') r' dr' dy' ds \leq C$ independently of t , and this can be done by changing to polar coordinates in M_h^2 near ff to see that the integrand is actually bounded. Details are left to the reader. Since the vector fields $r^2 \partial_t$, $r \partial_r$ and ∂_y lift to M_h^2 to be tangent to the side and front faces, and because of the infinite order vanishing along $t = 0$, the differentiated kernel $(r^2 \partial_t)^i (r \partial_r)^j \partial_y^s H_0$ has the same polyhomogeneous structure as H_0 for any $i, j, s \in \mathbb{N}_0$. This means that $(r^2 \partial_t)^i (r \partial_r)^j \partial_y^s u_0$ satisfies precisely the same estimates as u_0 does, whence $u_0 = \hat{u} \in C_{b_0}^{2\ell, \ell}$ for all $\ell \geq 0$, as claimed.

This discussion has focussed entirely on the behaviour of H near ff . This is because if we localize H by multiplying by a cutoff function which vanishes near ff and the side faces, then the estimates reduce to those for a standard local interior problem with no conic degeneracy. \square

Remark 3.7. There is one other dilation-invariant vector field, namely $t \partial_t$, and it is natural to ask about the regularity of $t \partial_t u$ when $f \in C_{b_0}^{k+\delta, (k+\delta)/2}$. Write $t \partial_t = (t/r^2) r^2 \partial_t$, and note that, in the support of H_1 , t/r^2 is a smooth bounded function; in addition, $t \partial_t$ is tangent to the front face of the heat space, and hence preserves the expansion of H_0 . Taking these two facts together, we see that

$$(r \partial_r)^i (\partial_y)^j (r^2 \partial_t)^\ell (t \partial_t)^m u \in C_{b_0}^{\delta, \delta/2}$$

provided $i + j + 2\ell + 2m \leq k + 2$. In particular, we see that u obtains more regularity in t than was initially apparent near $r = 0$ when $t > 0$.

The next estimate is for the Friedrichs–Hölder domain norm.

Proposition 3.8. *Suppose that $f \in C_{b_0}^{k+\delta, (k+\delta)/2}([0, T] \times \tilde{M})$ and u is the unique solution to $(\partial_t - L)u = f$, $u|_{t=0} = 0$. Then u lies in the Friedrichs–Hölder domain $\mathcal{D}_{b_0}^{k+\delta, (k+\delta)/2}$ and satisfies*

$$\|u\|_{\mathcal{D}_{b_0}^{k+\delta, (k+\delta)/2}} := \|u\|_{b_0; k+\delta, (k+\delta)/2} + \|\Delta_g u\|_{b_0; k+\delta, (k+\delta)/2} \leq C \|f\|_{b_0; k+\delta, (k+\delta)/2}. \tag{3-14}$$

Proof. We must estimate

$$\Delta_g u = \int_0^t \int_M \Delta_g H(t-t', z, z') f(s, z') dA_g dt'$$

in $C_{b_0}^{\delta, \delta/2}$. The key observation is that the Schwartz kernel K of $\Delta_g \circ H$ is an operator of heat type which we say is of “nonnegative type” (by analogy with the stationary case), and which therefore gives a bounded map of the spaces $C_{b_0}^{\delta, \delta/2}$. To be more specific, K is polyhomogeneous at all the faces of M_h^2 , and the terms in its expansions at the left and right faces are nonnegative, while the leading terms at ff and td are $\rho^{-4} \cong t^{-2}$ and R^{-4} , respectively. To see this, note that Δ_g differentiates tangentially to the left face (where $r' \rightarrow 0$) so K has the same leading order as H there; at the right face ($r \rightarrow 0$), Δ_g annihilates

the initial terms r^0 and $r^{1/\beta} \cos y$ and $r^{1/\beta} \sin y$ in the expansion of H , so the leading order of K is nonnegative here too; the leading orders exhibit the maximal drop in order to ρ^{-4} and R^{-2} at the other two faces because Δ_g is not tangent to these faces and acts as a second-order conic operator in (r, y) , and the leading coefficients in the expansion of H there are not annihilated by this operator.

We now proceed as in the preceding proof, decomposing K into $K_0 + K_1$ and estimating the integrals corresponding to each. The details are almost exactly the same, except for two facts. First, the extra factor of $\lambda^{-2} = 2^{-2\ell}$ no longer appears when rescaling the terms $K_1 \star f_\ell$ because of the drop in leading order homogeneity (from ρ^{-2} to ρ^{-4}) at the front face. In addition, we appeal to the standard interior estimate $\|\Delta u\|_{\delta, \delta/2} \leq C \|f\|_{\delta, \delta/2}$, where u and f are defined on the product of $[0, 1]$ with a ball of radius 1, Δ is a nondegenerate Laplacian on that ball, and, as usual, the norm on the left is only computed over a ball of radius $\frac{1}{2}$. A generalization of this interior estimate is that, if J is a kernel on the double heat space of \mathbb{R}^2 with compact support in all variables, and which vanishes to infinite order at $t = 0$ but blows up like t^{-2} at the new face, td , of the blowup, then $\|Jf\|_{\delta, \delta/2} \leq C \|f\|_{\delta, \delta/2}$. The simpler integral estimate for $K_0 \star f$ is again essentially the same, since $\int K_0(t, z, z') dt dz'$ is still bounded as a function of z . This proves that $\|\Delta_g u\|_{b_0; k+\delta, (k+\delta)/2} \leq C \|f\|_{b_0; k+\delta, (k+\delta)/2}$. \square

We can now turn to the estimates in the b -spaces.

Proposition 3.9. *Suppose that $f \in C_b^{k+\delta, (k+\delta)/2}([0, T] \times \tilde{M})$ and u is the unique Friedrichs solution to $(\partial_t - L)u = f$, $u|_{t=0} = 0$. Then u lies in the Friedrichs–Hölder domain $\mathcal{D}_b^{k+\delta, (k+\delta)/2}$ and satisfies*

$$\|u\|_{b; k+2+\delta, (k+2+\delta)/2} \leq C \|f\|_{b; k+\delta, (k+\delta)/2} \tag{3-15}$$

and

$$\|u\|_{\mathcal{D}_b^{k+\delta, (k+\delta)/2}} := \|u\|_{b; k+\delta, (k+\delta)/2} + \|\Delta_g u\|_{b; k+\delta, (k+\delta)/2} \leq C \|f\|_{b; k+\delta, (k+\delta)/2}. \tag{3-16}$$

Moreover, $u = \hat{u} + \tilde{u}$, where $\tilde{u} \in r^2 C_b^{k+2+\delta, (k+2+\delta)/2}$ and

$$\hat{u}(t, z) = a_0(t) + (a_{11}(t) \cos y + a_{12}(t) \sin y)r^{1/\beta} \tag{3-17}$$

with $a_0, a_{11}, a_{12} \in C^{1+\delta/2}([0, T])$.

Proof. First suppose that $k = 0$. We prove (3-16) using (3-11b). By Proposition 3.8, we already know that $u \in C_{b_0}^{2+\delta, 1+\delta/2} \cap \mathcal{D}_{b_0}^{\delta, \delta/2}$. Thus it suffices to show that u and $\Delta_g u$ lie in $C^{0, \delta/2}$ as well. Defining $K = \Delta_g \circ H$, we first prove that

$$K \star : C_{b_0}^{\delta, \delta/2} \cap C^{0, \ell} \longrightarrow C_{b_0}^{\delta, \delta/2} \cap C^{0, \ell}$$

is bounded for $\ell = 0, 1$. For $\ell = 0$, observe first that if $f = C$ is constant then $K \star f \equiv 0$, since $H \star 1 = t$. This means that we may reduce to considering functions which vanish at $t = r = 0$. Next, if f vanishes near $t = r = 0$, then direct inspection of the integral defining $K \star f$ shows that this function also vanishes near $t = r = 0$; taking the closure in the C^0 norm (or rather, the $C^0 \cap C_{b_0}^{\delta, \delta/2}$) norm preserves the property of vanishing at $t = r = 0$. The case $\ell = 1$ follows by noting that ∂_t commutes with H and hence with K . By interpolation, we conclude the boundedness of

$$K \star : C_{b_0}^{\delta, \delta/2} \cap C^{0, \delta/2} \longrightarrow C_{b_0}^{\delta, \delta/2} \cap C^{0, \delta/2}.$$

This finishes the proof of (3-16).

To obtain (3-15) when $k = 0$, we must show that $u \in C_b^{2+\delta, 1+\delta/2}$, or equivalently (in a neighbourhood of the conic point), that $(r\partial_r)^i \partial_y^j \partial_t^\ell u \in C_b^{\delta, \delta/2}$ if $i + j + 2\ell \leq 2$. If $\ell = 1$ (so $i = j = 0$), we use that $\partial_t u = \Delta_g u + f \in C_b^{\delta, \delta/2}$, as per the last paragraph. If $\ell = 0$, we observe, as before, that $(r\partial_r)^i \partial_y^j \circ H$ is bounded on $C_{b0}^{\delta, \delta/2} \cap C^{0, \ell}$ for $\ell = 0, 1$, and hence, by interpolation, is bounded on $C_b^{\delta, \delta/2}$.

Now suppose that k is a strictly positive even integer. We use induction, supposing that (3-16) and (3-15) have been proved for $0, 2, \dots, k - 2$. To prove that $K = \Delta_g \circ H \star$ is bounded on $C_b^{k+\delta, (k+\delta)/2}$, we must show that $K_{i,j,\ell} := (r\partial_r)^i \partial_y^j \partial_t^\ell \circ K \star : C_b^{k+\delta, (k+\delta)/2} \rightarrow C_b^{\delta, \delta/2}$ is bounded whenever $i + j + 2\ell \leq k$. There are three cases. First, if $1 \leq \ell \leq k/2 - 1$, then $K_{i,j,\ell} : C_b^{k+\delta, (k+\delta)/2} \rightarrow C_b^{\delta, \delta/2}$ is bounded provided $K_{i,j,0} : C_b^{k+\delta-2\ell, (k+\delta-2\ell)/2} \rightarrow C_b^{\delta, \delta/2}$ is, and, since $i + j \leq k - 2\ell \leq k - 2$, this is known by induction. Next, if $\ell = k/2$ then, since $\partial_t^{k/2} \circ K = K \circ \partial_t^{k/2}$, we reduce directly to the boundedness of K on $C_b^{\delta, \delta/2}$. Finally, when $\ell = 0$, a bit more work is needed. If V is any b -vector field, we consider either the commutator $[V, H \star]$ or, more or less equivalently, the commutator $[V, \partial_t - \Delta]$. The latter is slightly more elementary, so we follow that route. Writing $g = e^\phi(dr^2 + (1 + \beta)^2 r^2 dy^2)$ near the conic point, it is easy to check that

$$[V, \Delta] = p\Delta + q + W,$$

where W is a second-order operator with coefficients supported away from $r = 0$. Since the estimates we seek are standard in the support of W , we shall systematically neglect this term in the calculations below. For this part of the estimate we induct in integer steps, so, to unify the notation, assume that $k \in \mathbb{N}$ and $0 < \delta < 1$. Now, suppose that $f \in C_b^{k+\delta, (k+\delta)/2}$ and that we have proved by induction that $u \in C_b^{k+1+\delta, (k+1+\delta)/2}$ and $\Delta u \in C_b^{k-1+\delta, (k-1+\delta)/2}$. We then compute that

$$(\partial_t - \Delta)Vu = Vf + (p\Delta + q)u \in C_b^{k-1+\delta, (k-1+\delta)/2},$$

which implies that $Vu \in C_b^{k+1+\delta, (k+1+\delta)/2}$ and $\Delta Vu \in C_b^{k-1+\delta, (k-1+\delta)/2}$. Finally, $V\Delta u = \Delta Vu + (p\Delta + q)u$ is in $C_b^{k-1+\delta, (k-1+\delta)/2}$. Since this is true for every b -vector field V , we conclude that $u \in C_b^{k+2+\delta, (k+2+\delta)/2}$ and $\Delta u \in C_b^{k+\delta, (k+\delta)/2}$, as required. This proves (3-16) and (3-15) in general.

It remains to study the expansion as $r \rightarrow 0$. We explain the case $k = 0$ and leave the extension to spaces with higher regularity to the reader. Recalling the decomposition $H = H_0 + H_1$ from the proof of Proposition 3.6, the same interpolation argument as earlier implies that

$$H_1 \star : C_b^{\delta, \delta/2} \longrightarrow r^2 C_b^{2+\delta, 1+\delta/2}.$$

Next, similarly to what we did in the stationary (elliptic) case, note that $r\partial_r(r\partial_r - \beta^{-1}) \circ H_0 = r^2 H'_0$, where H'_0 has nonnegative index sets at $\text{ff} \cup \text{lf} \cup \text{rf}$ (and vanishes to infinite order at td), which means that $r\partial_r(r\partial_r - \beta^{-1})u_0 \in r^2 C_{b0}^{k, k/2}$ for all $k \geq 0$. Applying interpolation once more, this time for the mappings

$$(r\partial_r)^i \partial_y^j \partial_t^\ell r\partial_r(r\partial_r - \beta^{-1})H_0 \star : C_{b0}^{\delta, \delta/2} \cap C^{0, m} \longrightarrow r^2 C_{b0}^{\delta, \delta/2} \cap C^{0, m},$$

gives that $r\partial_r(r\partial_r - \beta^{-1})u_0 \in r^2 C_b^{k+\delta, (k+\delta)/2}$ for every $k \geq 0$. Both this and the previous interpolation involving H_1 are complicated slightly by the fact that $[\partial_t, H_j \star]$ is no longer zero, but the extra terms can still be handled.

Finally, integrating in r gives that $u_0 = a_0(t, y) + a_1(t, y)r^{1/\beta} + \tilde{u}'$, where $\tilde{u}' \in r^2\mathcal{C}_b^{2+\delta, 1+\delta/2}$. Applying $(\partial_t - \Delta_g)$ to $u = u_0 + u_1$ shows first that $a_0 = a_0(t)$ and $a_1 = a_{11}(t) \cos y + a_{12}(t) \sin y$, and then that $a_0, a_{11}, a_{12} \in C^{1+\delta/2}([0, T])$. \square

Corollary 3.10. *Let u and f be as in Proposition 3.9. Then*

$$\|u\|_{b, k+\delta, (k+\delta)/2} \leq CT \|f\|_{b, k+\delta, (k+\delta)/2}. \tag{3-18}$$

Proof. The inequality (3-18) is actually a formal consequence of (3-12) and (3-16). Indeed, since $u(0, z) = 0$,

$$u(t, z) = \int_0^t \partial_\tau u(\tau, z) d\tau \implies \|u\|_{b; \delta, 0} \leq \int_0^T \|\partial_\tau u(\tau, \cdot)\|_{b; \delta, 0} d\tau \leq T \|u\|_{b; 2+\delta, 1+\delta/2} \leq CT \|f\|_{b; \delta, \delta/2}.$$

Similarly, since $\partial_\tau u(0, z) = \Delta_g u(0, z) = 0$,

$$\begin{aligned} |u(t, z) - u(t', z)| &\leq \int_{t'}^t |\partial_\tau u(\tau, z)| d\tau = \int_{t'}^t |\partial_\tau u(\tau, z) - \partial_\tau u(0, z)| d\tau \\ &\leq \|u\|_{b; 2+\delta, 1+\delta/2} \int_{t'}^t \tau^{\delta/2} d\tau \leq C|t - t'| \cdot (|t + t'|^{\delta/2} + 1) \|u\|_{b; 2+\delta, 1+\delta/2} \end{aligned}$$

for some constant $C = C(\delta) > 0$, whence

$$[u]_{b; 0, \delta/2} \leq CT \|f\|_{b; \delta, \delta/2}.$$

Combining these two inequalities yields (3-18). \square

We make a special note of the fact that the estimate (3-16) is the main one here, since both (3-15) and (3-18) follow from it.

Corollary 3.11. *Let g_0 be any smooth conic metric, and suppose that $g_1 = e^\phi g_0$ with $\phi \in \mathcal{C}_b^{k+\delta}(\tilde{M})$, where $\phi = 0$ at $\partial\tilde{M}$. For any $R_1 \in \mathcal{C}_b^{k+\delta}(\tilde{M})$, i.e., not necessarily the scalar curvature of g_1 , set $L_1 = \Delta_{g_1} + R_1$. Then the solution operator H_1 to $(\partial_t - L_1)u = f$, $u|_{t=0} = 0$, satisfies the same set of bounds (3-12), (3-14), (3-15), (3-16) and (3-18) (for that particular value of k , with constants depending only on g_0 and the norms $\|\phi\|_{b; k+\delta}$, $\|R_1\|_{b; k+\delta}$).*

Proof. We may as well absorb the term $R_1 u$ into f . Choose a function $\tilde{a} \in \mathcal{C}_b^{k+\delta}$ which agrees with e^ϕ in a small neighbourhood of $\partial\tilde{M}$ and which is chosen uniformly close to 1 on the rest of \tilde{M} , so that $\|(\tilde{a} - 1)\Delta_0 H_0 \star\|_{b; k+\delta} < \epsilon$, where H_0 is the heat kernel for $\partial_t - \Delta_0$. Writing $\tilde{\Delta}_1 = \tilde{a} \Delta_0$,

$$(\partial_t - \tilde{\Delta}_1)H_0 \star = \text{Id} - (\tilde{a} - 1)\Delta_0 H_0 \star;$$

by our choice of \tilde{a} , the right-hand side is invertible by Neumann series, so we may represent the heat kernel \tilde{H}_1 for $\tilde{\Delta}_1$ as

$$\tilde{H}_1 = H_0 \star (\text{Id} - (\tilde{a} - 1)\Delta_0 H_0 \star)^{-1}.$$

This shows that the solution \tilde{u} to $(\partial_t - \tilde{\Delta}_1)\tilde{u} = f$ satisfies all the same estimates as the solution u to $(\partial_t - \Delta_0)u = f$, with constants depending only on the norm of ϕ .

Taking as given that the solution u exists, but may not satisfy the correct estimates near \tilde{M} , observe that

$$(\partial_t - \tilde{\Delta}_1)(\tilde{u} - u) = b\Delta_0 u$$

for some function $b \in C_b^{k+\delta}$ which vanishes in a fixed neighbourhood of the conic points. Noting that, by standard local parabolic regularity theory, u certainly satisfies the correct estimates on the support of b , we observe finally that

$$u = \tilde{u} - \tilde{H}_1 \star (b\Delta_0 u) = \tilde{H}_1 \star (f - b\Delta_0 u),$$

from which we again obtain all necessary estimates. It is clear that the constants depend on ϕ only through its norm $\|\phi\|_{b;k+\delta}$. □

3G. Short-time existence. We can now apply the mapping properties of the last section to establish the short-time existence for the angle-preserving solution of the flow (2-2). For this short-time result, we may as well assume that $\rho = 0$, and we consider the flow starting at any $\mathcal{D}_b^{k,\delta}$ metric g_0 . Recall that this means that $g_0 = e^{w_0} \bar{g}_0$, where \bar{g}_0 is a smooth and exact conic and $w_0 \in \mathcal{D}_b^{k,\delta}$. Now let $g(t) = e^{\phi(t)} g_0$, so that (2-2) becomes

$$\begin{aligned} \partial_t \phi &= e^{-\phi} \Delta_{g_0} \phi - R_0 e^{-\phi} = (\Delta_{g_0} + R_0) \phi - R_0 + (e^{-\phi} - 1) \Delta_0 \phi - R_0 (e^{-\phi} - 1 + \phi) \\ &:= L\phi - R_0 + Q(\phi, \Delta_0 \phi) \end{aligned} \tag{3-19}$$

with $\phi(0, \cdot) = 0$. By Corollary 3.11, the heat kernel H for $\partial_t - L$, $L = \Delta_{g_0} + R_{g_0}$, satisfies the same estimates as before.

Proposition 3.12. *Let g_0 be a $\mathcal{D}_b^{k,\delta}$ metric. Then there exists a unique solution $\phi \in \mathcal{D}_b^{k+\delta, (k+\delta)/2}([0, T] \times \tilde{M})$ to (3-19) with $\phi|_{t=0} = 0$ for some $T > 0$ depending on the $\mathcal{D}_b^{k,\delta}$ norm of g_0 .*

Proof. We suppose that $k = 0$, leaving the case of general k to the reader. The equation (3-19) is equivalent to the integral equation

$$\phi(t, z) = \int_0^t \int_M H(t-s, z, z') (Q(\phi, \Delta_0 \phi)(s, z') - R_0(s, z')) ds dA_{z'}. \tag{3-20}$$

Denote the operator on the right by $\mathcal{T}(\phi)$. We claim that there are constants η and T so that the convex, closed set

$$\mathcal{J} = \{ \phi \in \mathcal{D}_b^{\delta, \delta/2}([0, T] \times \tilde{M}) : \|\phi\|_{b;\delta, \delta/2} + \|\Delta_0 \phi\|_{b;\delta, \delta/2} \leq \eta \}$$

is mapped to itself by \mathcal{T} and, moreover, $\mathcal{T} : \mathcal{J} \rightarrow \mathcal{J}$ is a contraction.

For notational simplicity below, write

$$\|\phi\|_{b;\delta, \delta/2} + \|\Delta_0 \phi\|_{b;\delta, \delta/2} := \|\phi\|_{\mathcal{D}}.$$

Denote by B the norm of $H \star : \mathcal{C}_b^{\delta, \delta/2} \rightarrow \mathcal{D}_b^{\delta, \delta/2}$; cf. Proposition 3.9. Writing $\Phi = H \star (-R_0)$, we then take $\eta = 2\|\Phi\|_{\mathcal{D}}$.

To proceed, recall first that, if $\chi \in \mathcal{C}_b^{\delta, \delta/2}$ vanishes at $t = 0$, then, for $0 \leq t \leq T$,

$$|\chi(t, z)| = |\chi(t, z) - \chi(0, z)| \leq T^{\delta/2} \|\chi\|_{b;\delta, \delta/2},$$

and hence

$$[\chi_1 \chi_2]_{b;\delta,\delta/2} \leq \|\chi_1\|_\infty [\chi_2]_{b;\delta,\delta/2} + [\chi_1]_{b;\delta,\delta/2} \|\chi_2\|_\infty \leq T^{\delta/2} \|\chi_1\|_{b;\delta,\delta/2} \|\chi_2\|_{b;\delta,\delta/2}.$$

Therefore,

$$\|(e^{-\phi} - 1)\Delta_0\phi\|_{b;\delta,\delta/2} \leq CT^{\delta/2} \|\phi\|_{b;\delta,\delta/2} \|\Delta_0\phi\|_{b;\delta,\delta/2},$$

where the constant C depends on η ; hence,

$$\|Q(\phi, \Delta_0\phi)\|_{b;\delta,\delta/2} \leq C_1 T^{\delta/2} \eta^2.$$

Thus, if $\phi \in \mathcal{J}$, then

$$\|\mathcal{T}(\phi)\|_{\mathcal{D}} \leq BC_1 T^{\delta/2} \eta^2 + \|\Phi\|_{\mathcal{D}}.$$

By taking T sufficiently small, we can make this less than η again, so \mathcal{T} maps \mathcal{J} to itself.

By the same reasoning, adding and subtracting $(e^{-\phi_2} - 1)\Delta_0\phi_1$ shows that

$$\|(e^{-\phi_1} - 1)\Delta_0\phi_1 - (e^{-\phi_2} - 1)\Delta_0\phi_2\|_{b;\delta,\delta/2} \leq CT^{\delta/2} (\|\phi_1\|_{\mathcal{D}} + \|\phi_2\|_{\mathcal{D}}) \|\phi_1 - \phi_2\|_{\mathcal{D}}.$$

The identical estimate for the other term in $Q(\phi, \Delta_0\phi)$, which does not involve derivatives of the ϕ_j , is easier. We deduce that

$$\|\mathcal{T}(\phi_1) - \mathcal{T}(\phi_2)\|_{\mathcal{D}} \leq BCT^{\delta/2} (2\eta) \|\phi_1 - \phi_2\|_{\mathcal{D}},$$

so, by taking T still smaller, we can make the coefficient less than $\frac{1}{2}$. This proves that \mathcal{T} is a contraction on \mathcal{J} , and hence that there exists a unique solution $\phi \in \mathcal{D}_b^{\delta,\delta/2}$ to (3-20) in \mathcal{J} . \square

We now prove the short-time existence result for the angle-changing flow. Since this is a side note of the paper, we make some simplifying assumptions about the initial metric to remove some irrelevant details from the proof. We assume that the prescribed angle functions $\beta_i(t)$ are smooth functions of t , although the optimal result should allow these to have only finite Hölder regularity. Assume too that there is only one conic point, and that the initial metric g_0 is the exact conic metric $dr^2 + \beta^2 r^2 dy^2$ near $r = 0$. Reverting back to the conformal form of the metric, define

$$\hat{g}_0(t) = |z|^{2\beta(t)-2} |dz|^2.$$

We have $\hat{g}'_0(t) = 2\beta'(t) \log |z| \hat{g}_0(t)$, or, in terms of the (r, y) coordinates,

$$\hat{g}'_0(t) = \kappa \beta'(t) \log r \hat{g}_0(t), \quad \kappa = \frac{2}{\beta}.$$

Setting $g(t) = u(t, \cdot) \hat{g}_0(t)$, the Ricci flow equation (with $\rho = 0$) thus becomes

$$(\partial_t u + u C \kappa \beta' \log r) = \Delta_{\hat{g}_0(t)} \log u - R_{\hat{g}_0(t)},$$

or, finally, in terms of $\phi = \log u$,

$$\partial_t \phi = e^{-\phi} \Delta_{\hat{g}_0(t)} \phi - R_{\hat{g}_0(t)} e^{-\phi} - \kappa \beta' \log r. \tag{3-21}$$

We seek a local-in- t solution to this equation with initial value $\phi(0, \cdot) \equiv 0$.

Unlike the case considered before, the reference metric $\hat{g}_0(t)$ now depends on t , and there is an extra inhomogeneous term $-\kappa\beta'(t) \log r$. For the first issue we say nothing, because short-time existence for the heat operators associated to time-dependent metrics is standard; see [Chow et al. 2006]. Regarding the second issue, since this additional term is polyhomogeneous, we may choose a polyhomogeneous function $\hat{\phi}(t, \cdot)$ with leading term $C\kappa r^2 \log r$ that satisfies

$$(\partial_t - e^{-\hat{\phi}} \Delta_{\hat{g}_0(t)})\hat{\phi} + R_{\hat{g}_0(t)}e^{-\hat{\phi}} = -\kappa\beta'(t) \log r + \chi,$$

where χ is smooth and vanishes to infinite order at $r = 0$. Now set $\phi = \hat{\phi} + \psi$ and rewrite (3-21) as an equation for the unknown function ψ . It is straightforward to check that this equation is different from the one for the angle-fixing flow in only a few minor ways. There are additional terms in the coefficients of the nonlinear terms; these, however, are polyhomogeneous in (r, y, t) and vanish at least like $r^2 \log r$. Next, there is an additional inhomogeneous term coming from the “error term” χ . The general structure of the equation is very similar to the one considered earlier in this section, and it is a straightforward exercise to check that this equation has a solution $\psi(t, \cdot)$ for $0 \leq t < T$ for T sufficiently small.

It is important to note that, unlike in the angle-changing flow, the fact that the conformal factor now includes a term $r^2 \log r$ means that the curvature $R_{g(t)}$ is unbounded for $t > 0$ near $r = 0$. This is in accord with the results in the thesis of Ramos.

3H. Higher regularity. It will be very helpful for us later to be able to appeal to some higher regularity properties of the solution, so we prove these now.

Proposition 3.13. *Suppose that $g(t)$ is the solution to the Ricci flow equation with $g(t) = u(t)g_0$, where g_0 is smooth and exactly conic, $u(0) \in \mathcal{C}_b^{0,\delta}$, and $u \in \mathcal{D}_b^{\delta,\delta/2}$ is given by Proposition 3.12. Then u is polyhomogeneous on $(0, T) \times \tilde{M}$.*

Proof. Write $u = e^\phi$ with ϕ satisfying (3-19) and $\phi(0) = \phi_0 \in \mathcal{C}_b^{0,\delta}$. Since the initial condition is no longer zero, we have

$$\phi(t, z) = \int_M H_0(t, z, z')\phi_0(z') dA_{z'} + H_0 \star (Q(\phi, \Delta_0\phi) - R_0).$$

The first term is polyhomogeneous when $t > 0$ because of the polyhomogeneous structure of H_0 . The second term lies in $\mathcal{C}_b^{2+\delta, 1+\delta/2}$, so its restriction to any $t = \epsilon > 0$ lies in $\mathcal{C}_b^{2,\delta}$. Consider the equation starting at $t = \epsilon$, i.e., replace t by $t + \epsilon$. Then Proposition 3.12 and the uniqueness of solutions shows that $u \in \mathcal{D}_b^{2+\delta, 1+\delta/2}$ for $t \geq \epsilon$ and, since ϵ is arbitrary, this holds for all $t > 0$. Bootstrapping in the obvious way gives that $u \in \mathcal{D}_b^{k+\delta, (k+\delta)/2}$ for every k , all in the same interval of existence $(0, T)$. In other words, $(r\partial_r)^j \partial_\theta^\ell \partial_t^s \Delta_0 u \in \mathcal{C}_b^{\delta,\delta/2}$ for all $j, \ell, s \geq 0$, which means that u is conormal when $t > 0$.

From Proposition 3.6, $\phi = a_0(t) + r^{1/\beta}(a_{11}(t) \cos y + a_{12}(t) \sin y) + \tilde{\phi}$; by what we have just shown, $\tilde{\phi} \in r^2 \mathcal{A}((0, T) \times \tilde{M})$ and $a_0, a_{11}, a_{12} \in \mathcal{C}^\infty((0, T))$. In order to extend this expansion to all higher orders, assume g_0 is exactly conic (so $R_0 \equiv 0$) in some neighbourhood of $r = 0$ and write (3-19) there as

$$r^2 \partial_t e^\phi = ((r\partial_r)^2 + \beta^{-2} \partial_y^2) \phi.$$

Since ϕ is conormal, we may study this formally. Taking advantage of information we have already obtained, inserting the expansion of ϕ to order 2 shows that the expression on the left has a finite expansion $r^2 a'_0(t) + r^{2+1/\beta}(a'_{11}(t) \cos y + a'_{12}(t) \sin y)$ and a conormal error term of order r^4 . Using the operator on the right shows that ϕ must have an expansion up to order 4, with new terms of orders r^2 and $r^{2+1/\beta}$ as well as $r^{2/\beta}$ if $\beta > \frac{1}{2}$, with a conormal error term of order 4. Continuing in this way, we see that ϕ has an expansion to all orders, as claimed. \square

Corollary 3.14. *Let $R(t)$ denote the curvature function of the solution metric $g(t)$. Then $R(t)$ is also polyhomogeneous on $(0, T) \times \tilde{M}$, and the initial terms in its expansion have the form*

$$R(t) \sim b_0(t) + r^{1/\beta}(b_{11}(t) \cos y + b_{12}(t) \sin y) + \mathcal{O}(r^2).$$

In particular, $\Delta_0 R$ is bounded and polyhomogeneous for all $t > 0$.

Proof. This follows directly from the polyhomogeneity of ϕ and equation (2-4). \square

3I. Maximum principles. Before embarking on the remainder of the proof of long-time existence and convergence, we present some results which show how the maximum principle may be extended to this conic setting. We adapt the trick of [Jeffres 2005].

The possible difficulty in applying the maximum principle directly is if the maximum of the solution were to occur at a conic point, so the idea is to perturb the solution slightly to ensure that the maximum cannot occur at the singular locus.

Lemma 3.15. *Suppose that $(M, g(t))$ is a family of metrics which is in $\mathcal{D}_b^{\delta, \delta/2}([0, T) \times \tilde{M})$, polyhomogeneous on $(0, T) \times \tilde{M}$, and that w satisfies*

$$\partial_t w \geq \Delta w + X \cdot \nabla w + a(w^2 - A^2),$$

where X and a are a given vector field and function, respectively, with the same regularity as $g(t)$ and with $a > 0$; here $A \geq 0$ is a constant. Suppose too that $w(0, \cdot) \geq -A$ and that $\sup(|w(t, \cdot)| + r^\sigma |\nabla w(t, \cdot)|) < \infty$ for every $t > 0$, where $0 < \sigma < 1$. Then $w \geq -A$ for all $t < T$.

Proof. Define $w_{\min}(t) = \inf_{q \in \tilde{M}} w(t, q)$. By hypothesis, $w_{\min}(0) \geq -A$. Suppose that, at some time $t > 0$, this minimum is achieved at some point q . If q is not one of the conic points, then $\Delta w(t, q) \geq 0$ and $\nabla w(t, q) = 0$; hence

$$\frac{d}{dt} w_{\min}(t) \geq a(w_{\min}(t)^2 - A^2). \tag{3-22}$$

Suppose for the moment that we have established this differential inequality regardless of the location of the minimum. But then, if $w_{\min}(t)$ were ever to achieve a value less than $-A$ at some $t_0 > 0$, (3-22) would give that $w'_{\min}(t_0) > 0$, which is impossible (if $t_0 > 0$ is the smallest time at which $w_{\min}(t_0) < -A$).

Thus it suffices to show that (3-22) is always true. Fix γ with $0 < \gamma < 1 - \sigma$. Then, for any $k \geq 1$, define $w_k(q, t) = w(q, t) - r^\gamma/k$ (where r is a fixed radial function near each conic point such that r is smooth and strictly positive in the interior and $r = 0$ at a conic point). Suppose that $w_{\min}(t)$ is achieved at some conic point p . We first observe that, for q sufficiently near p , using the hypothesis on $|\nabla w|$,

$$w(t, q) \leq w_{\min}(t) + Cr^{1-\sigma} = w(t, p) + Cr^{1-\sigma},$$

where $r = r(q)$, and hence

$$w_k(t, q) \leq w(t, p) + Cr^{1-\sigma} - \frac{1}{k}r^\gamma < w(t, p) = w_k(t, p)$$

for r sufficiently small. In other words, $(w_k)_{\min}(t)$ cannot occur at p . Now, the differential inequality satisfied by w_k is

$$\partial_t w_k \geq \Delta \left(w_k + \frac{1}{k}r^\gamma \right) + X \cdot \nabla \left(w_k + \frac{1}{k}r^\gamma \right) + a \left(\left(w_k + \frac{1}{k}r^\gamma \right)^2 - A^2 \right).$$

At a spatial minimum (away from the conic point), $\Delta w_k \geq 0$ and $\nabla w_k = 0$. On the other hand, $\Delta r^\gamma \geq Cr^{\gamma-2}$ and $|X \cdot \nabla r^\gamma| \leq Cr^{\gamma-1}$ near $r = 0$, and, since the first of these terms is positive, these two terms together satisfy

$$\frac{1}{k}(\Delta r^\gamma + X \cdot \nabla r^\gamma) \geq \frac{C}{k}.$$

Thus altogether, applying the same reasoning as before (and using that $(w_k)_{\min}$ does not occur at a conic point), we deduce that

$$\frac{d}{dt}(w_k)_{\min} \geq \frac{C}{k} + a \left(\left((w_k)_{\min} + \frac{1}{k}r(q_k(t))^\gamma \right)^2 - A^2 \right),$$

where the minimum of w_k is achieved at $q_k(t)$. The same arguments as above give $(w_k)_{\min} \geq -A - C'/k$, and hence $w_{\min} \geq -A - C''/k$. Letting $k \nearrow \infty$ proves the result. \square

Essentially the same proof gives the following version of the maximum principle:

Lemma 3.16. *Suppose that the setup is exactly the same as in the previous lemma, and that*

$$\partial_t w = \Delta w + aw^2 + bw.$$

Then

$$\frac{d}{dt} w_{\max} \leq aw_{\max}^2 + bw_{\max} \quad \text{and} \quad \frac{d}{dt} w_{\min} \geq aw_{\min}^2 + bw_{\min}. \tag{3-23}$$

3J. Long-time existence. We are finally able to complete the proof of long-time existence of the solution of the Ricci flow with prescribed conic singularities. In fact, the proof is a straightforward adaptation of the original proof of this same fact for the Ricci flow on smooth compact surfaces in [Hamilton 1988]. We refer to that article as well as [Isenberg et al. 2011] for all the details of the proof. We supply here only the key results which then allow the proofs in those articles to be applied verbatim.

The strategy is to consider the “potential function” f for the metric $g(t)$. (In the language of [Jeffres et al. 2014], f is the Ricci potential for g .) By definition, this is a solution to the equation

$$\Delta_{g(t)} f = R_{g(t)} - \rho, \tag{3-24}$$

where ρ is the average scalar curvature. The crucial property that it must satisfy is that $|\nabla f| \leq C$. Observe that f is only defined up to an arbitrary additive constant, which may depend on t , but that the proof in [Hamilton 1988] shows how to choose this constant using the evolution equation satisfied by f .

In any case, we now show that a potential function with bounded gradient exists. Interestingly, this is one place where the assumption that the cone angles are less than 2π plays a crucial role.

Proposition 3.17. *Suppose that g is a conic metric with all cone angles less than 2π ; suppose too that $g = ug_0$, where g_0 is smooth (or polyhomogeneous) on \tilde{M} , $u \in C_b^{2,\delta}$ and, furthermore, $R_g \in C_b^{0,\delta}$. Then the solution f to $\Delta_g f = R_g - \rho$ which lies in the Friedrichs domain and satisfies $\int f dA_g = 0$ has $|\nabla f| \leq C$.*

Proof. By Proposition 3.1 (as well as the fact that the integral of $R - \rho$ is zero), there exists a unique solution f which has integral zero, and this function has a partial expansion

$$f \sim a_0 + (a_{11} \cos y + a_{12} \sin y)r^{1/\beta} + \tilde{u}, \quad \tilde{u} \in r^2 C_b^{2,\delta}.$$

Since $\beta < 1$, it follows immediately that $|\nabla f| \leq C$. □

We recall very briefly that the rest of the proof of long-time existence involves getting a priori uniform bound on $R_{g(t)}$ where $g(t)$ is the family of solution metrics, and then using (2-4) to find bounds for $\log u$. The bounds on R_{\min} follow easily from the maximum principle, while the bound for R_{\max} is derived by considering the evolution equation satisfied by $h := \Delta f + |\nabla f|^2$. For both of these steps, one needs the maximum principle from the previous subsection, which is permissible since R and h both satisfy the conditions of Lemma 3.15.

4. Convergence of the flow in the Troyanov case

We are now in a position to be able to prove that the solution $g(\cdot, t)$ converges exponentially as $t \rightarrow \infty$ to a constant-curvature metric with the same cone angles, provided the Troyanov condition (2-14) holds.

Let $W^{1,2}$ denote the usual Sobolev space of L^2 functions whose gradient is in L^2 (with respect to g_0). Following [Troyanov 1991; Struwe 2002], consider the energy functional $\mathcal{F} : W^{1,2} \rightarrow \mathbb{R}$,

$$\mathcal{F}(\phi) := \int_M (|\nabla_0 \phi|^2 + 2R_0 \phi) dA_0,$$

where the conformal factor has been rewritten as $u = e^\phi$. (The function spaces $W^{1,2}$ and $W^{2,2}$ used below are taken with respect to any fixed conic metric that is smooth in the (r, y) coordinates.) The next lemma says that the Ricci flow is the gradient flow of \mathcal{F} with respect to the Calabi L^2 metric (see, e.g., [Clarke and Rubinstein 2013, §2]).

Lemma 4.1. *If u is a solution of (2-2), then*

$$\frac{d}{dt} \mathcal{F}(\phi) = -2 \int_M (R - \rho)^2 dA_g. \tag{4-1}$$

Proof. On smooth, closed surfaces the formula is well known [Struwe 2002, Equation (49)]. Indeed, recall that, using (2-1) and (2-2),

$$\partial_t \phi = e^{-\phi} (\Delta_0 u - R_0) + \rho = \rho - R;$$

from this we get

$$\begin{aligned} \frac{d}{dt} \mathcal{F}(\phi) &= 2 \int_M (\nabla \phi \cdot \nabla \phi_t + R_0 \phi_t) dA_0 = 2 \int_M \phi_t (R_0 - \Delta_0 \phi) dA_0 \\ &= 2 \int_M R e^\phi \phi_t dA_0 = -2 \int_M (R - \rho) R dA_g, \end{aligned}$$

and the result follows since $\int (R - \rho) dA_g = 0$. Concealed here is the fact that these integrations by parts remain valid in this conic setting. This sort of computation will be used repeatedly in the remainder of this paper. The key point is that the functions involved enjoy sufficient regularity near the conic points that one may integrate by parts on the complement of an ϵ -neighbourhood of these points and show that the boundary term tends to 0 with ϵ . □

Troyanov [1991] proves that the conditions (2-14) ensure that there exists a constant C such that

$$\mathcal{F}(\phi(t)) \geq -C \quad \text{for all } t \geq 0.$$

(In fact, Troyanov considers the stationary problem from a variational point of view and proves that \mathcal{F} is bounded below on $W^{1,2}$ if (2-14) holds.)

We now prove that $\phi(\cdot, t)$ is uniformly bounded in $W^{2,2}$. This too follows arguments in [Troyanov 1991; Struwe 2002].

Proposition 4.2. *With all notation as above, if the conditions (2-14) hold and ϕ is a solution to the flow, then*

$$\|\phi(\cdot, t)\|_{W^{2,2}} \leq C.$$

Proof. We sketch the argument and refer to [Troyanov 1991; Struwe 2002] for more details. The starting point is the uniform lower bound $\mathcal{F}(\phi(\cdot, t)) \geq -C$. We first claim that

$$\|\phi(\cdot, t)\|_{W^{1,2}} \leq C, \quad t \geq 0. \tag{4-2}$$

There are three cases to consider. We only give details for the case when $\chi(M, \vec{\beta}) > 0$, since the cases where $\chi(M, \vec{\beta}) \leq 0$ are similar but simpler. The Troyanov condition (2-14) is equivalent to $0 < 2\pi\gamma := 2\pi\chi(M, \vec{\beta}) < 4\pi \min_i \{\beta_i\}$. Choose b such that $\pi\gamma = \pi\chi(M, \vec{\beta}) < b < 2\pi \min_i \{\beta_i\}$ and set

$$I(\phi) := \frac{1}{2b} \int_M |\nabla\phi|^2 dA_0 + \frac{1}{\pi\gamma} \int_M R_0\phi dA_0.$$

As in the proof of Theorem 5 in [Troyanov 1991], we have $I(\phi) \geq -C$ for all $\phi \in W^{1,2}$. But

$$\frac{1}{2\pi\gamma} \mathcal{F}(\phi) = I(\phi) + \frac{1}{2} \left(\frac{1}{\pi\gamma} - \frac{1}{b} \right) \int_M |\nabla\phi|^2 dA_0 > I(\phi) \geq -C.$$

Since $\mathcal{F}(\phi) \leq m$,

$$\int_M |\nabla\phi|^2 dA_0 \leq C, \quad t \geq 0,$$

and Troyanov’s argument then shows that also the L^2 is uniformly bounded [Troyanov 1991, p. 817], whence $\|\phi(\cdot, t)\|_{W^{1,2}} \leq C$ for all $t \geq 0$.

It is proved in [Troyanov 1991] that, if $0 < b < 2\pi \min_i \{2 + 2\alpha_i\}$, then there exists a constant C such that

$$\int_M e^{bu^2} dA_0 \leq C$$

for all $u \in W^{1,2}$ such that $\int_M u dA_0 = 0$ and $\int_M |\nabla u|^2 dA_0 \leq 1$. This is the Moser–Trudinger–Cherrier inequality for surfaces with conic singularities.

We now prove that

$$\int_M |\nabla^2 \phi|^2 dA_0 \leq C \quad \text{for all } t \in [0, \infty).$$

Carrying out a standard integration by parts argument over the complement of the ϵ -balls around the conic points, we obtain

$$\begin{aligned} & \int_{M \setminus B(\vec{p}, \epsilon)} |\nabla^2 \phi|^2 dA_0 \\ &= \int_{M \setminus B(\vec{p}, \epsilon)} |\Delta_0 \phi|^2 dA_0 - \frac{1}{2} \int_{M \setminus B(\vec{p}, \epsilon)} R |\nabla \phi|^2 dA_0 + \int_{\partial B(\vec{p}, \epsilon)} \partial_\nu \nabla \phi \cdot \nabla \phi d\sigma_0 - \int_{\partial B(\vec{p}, \epsilon)} \Delta_0 \phi \partial_\nu \phi d\sigma_0. \end{aligned}$$

Using Proposition 3.3 and letting $\epsilon \rightarrow 0$ gives

$$\int_M |\nabla^2 \phi|^2 dA_0 = \int_M |\Delta_0 \phi|^2 dA_0 - \frac{1}{2} \int_M R |\nabla \phi|^2 dA_0. \tag{4-3}$$

By (2-3),

$$\int_M |\Delta_0 \phi|^2 dA_0 \leq 2 \left(\int_M R_0^2 dA_0 + \int_M R^2 e^{2\phi} dA_0 \right) \leq C \left(1 + \int_M e^{2|\phi|} dA_0 \right) \leq C \left(1 + \int_M e^{b^2 |\phi|^2} dA_0 \right),$$

since, by Corollary 5.7 (proved later), the scalar curvature is uniformly bounded in time, where b is any real number such that $0 < b^2 < 2\pi \min_i \{2 + 2\alpha_i\}$ and C may depend on the choice of b . Now, by [Troyanov 1991, Proposition 11], the map $\phi \mapsto e^\phi$ is a compact embedding of $W^{1,2}$ in L^2 , which thus yields

$$\int_M |\Delta_0 \phi|^2 dA_0 \leq C,$$

and hence, finally,

$$\int_M |\nabla^2 \phi|^2 dA_0 \leq C \quad \text{for all } t \geq 0. \quad \square$$

Proposition 4.3. *Let $g(t)$ be the angle-preserving solution of (2-1) provided by Theorem 1.1. If (2-14) holds, then $g(t)$ converges exponentially to the unique constant-curvature metric in the conformal class of g_0 with specified conic data.*

Proof. We have already shown that $\phi(\cdot, t)$ exists and $\|\phi(\cdot, t)\|_{W^{2,2}} \leq C$ for all $t \geq 0$. We now invoke the arguments of [Struwe 2002] verbatim to deduce that $g(t)$ converges exponentially to a constant-curvature metric g_∞ in the conformal class of g_0 .

It remains to show that g_∞ has the same conic data $\{\vec{p}, \vec{\beta}\}$ as g_0 . The $W^{2,2}$ bound and the Sobolev embedding theorem give a uniform C^0 bound $|\phi(\cdot, t)| \leq C$. This implies that the conic points do not merge in the limit. Indeed, if $i \neq j$ and γ_{ij}^t is the geodesic for $g(t)$ joining these two conic points, then

$$\text{dist}_{g(t)}(p_i, p_j) = \int_{\gamma_{ij}^t} e^{\phi/2} \geq \bar{c} \int_{\gamma_{ij}^t} \geq \bar{c} \text{dist}_{g(0)}(p_i, p_j).$$

Next, suppose that g_∞ has cone angle parameter $\tilde{\beta}_i$ at p_i . Thus, in local conformal coordinates,

$$g_0 = e^{2\phi_0} |z|^{2\beta_i - 2} |dz|^2 \quad \text{and} \quad g_\infty = e^{2\phi_\infty} |z|^{2\tilde{\beta}_i - 2} |dz|^2,$$

so by the uniform C^0 bound it is clear that $\tilde{\beta}_i \geq \beta_i$ for all i . Since

$$\chi(M) + \sum \beta_i = \chi(M) + \sum \tilde{\beta}_i,$$

we see that $\beta_i = \tilde{\beta}_i$ for all i . □

5. Convergence in the non-Troyanov case

In this final section we consider the case where the Troyanov condition (2-14) fails. As remarked earlier, the angle inequality fails at just one of the points p_j , say p_1 , and necessarily $M = S^2$. Then, $(M, J, \vec{p}, \vec{\beta})$ does not admit a constant-curvature metric, and hence, even if $g(\cdot, t)$ converges, its limit must either not be of constant curvature or else some of the conic data is destroyed in the limit. More precisely, the limit might be a surface with fewer conic points and different cone angles, and hence might conceivably still admit a constant-curvature metric. The existence of nonconstant-curvature, soliton metrics with one or two conic points (the teardrop or American football) on S^2 can be ascertained using ODEs arguments, [Yin 2010], and these are the reasonable candidates for limiting metrics in the non-Troyanov case. To this end, we first show that every compact two-dimensional shrinking Ricci soliton which does not have constant curvature has at most two conic points. Furthermore, if (2-14) holds, then any shrinking Ricci soliton must have constant curvature. The next lemma also appears in [Ramos 2013].

Lemma 5.1. *If g is a shrinking Ricci soliton metric on M with conic data $(\vec{p}, \vec{\beta})$ and there are at least three conic points, then g has constant curvature.*

Proof. View g as a Kähler–Ricci soliton; then

$$(R - 1)g_{i\bar{j}} = \nabla_i \nabla_{\bar{j}} f,$$

where the vector field $X^i := \nabla^i f$ is a holomorphic vector field on $S^2 \setminus \vec{p}$. The trace of the soliton equation gives $\Delta f = R - 1$, and hence, using the static case of Theorem 1.2—see also [Jeffres et al. 2014, Propositions 3.3 and 3.8]—it follows that $\nabla f = \mathcal{O}(r^{1/\beta-1})$, so must vanish at each of the points p_i . This may also be deduced as in [Luo and Tian 1992, Lemma 3]. Using this same regularity, we can integrate by parts to get

$$\int_{S^2 \setminus \vec{p}} |X|^2 dA = \int_{S^2 \setminus \vec{p}} |\nabla f|^2 dA = \int_{S^2 \setminus \vec{p}} (1 - R)f dA < \infty.$$

However, there is no nontrivial holomorphic vector field on S^2 which vanishes at more than two points, so $X = 0$ and hence $\nabla_{\bar{j}} \nabla_i f = 0$. Finally, using the soliton equation again, $R \equiv 1$. □

Lemma 5.2. *If $(M, J, \vec{\beta}, \vec{p})$ satisfies (2-14) and g is a shrinking Ricci soliton metric, then g has constant curvature, i.e., $f = \text{const}$.*

Proof. The argument carries over from the smooth setting, by virtue of Theorem 1.2. We already know that there exists a constant-curvature metric \bar{g} with this prescribed data. By rescaling, assume $R_{\bar{g}} = 1$. Write $g = e^\phi \bar{g}$. Since g is a shrinking soliton, it moves under Ricci flow by a 1-parameter family of

diffeomorphisms $\psi(t)$, so $g(t) = \psi(t)^*g$. Hence, $\phi(\cdot, t) = \psi^*\phi$ solves

$$\partial_t \phi = \langle \nabla f, \nabla \phi \rangle_g = e^{-\phi} (\Delta_{\bar{g}} \phi - R_{\bar{g}}) + 1.$$

However, $R_g = e^{-\phi}(1 - \Delta_{\bar{g}} \phi)$ and $R_{\bar{g}} = 1$, so $\langle \nabla f, \nabla \phi \rangle_g = -R_g + 1$, which implies that

$$\langle \nabla f, \nabla \phi \rangle_g = R_g - 1 = -\Delta_g f,$$

or, equivalently, $\operatorname{div}(e^\phi \nabla f) = 0$. Multiplying by f and integrating by parts on $M \setminus B_\epsilon(\vec{p})$ gives

$$\int_{M \setminus B_\epsilon(\vec{p})} |\nabla f|^2 e^\phi dA_g = \int_{\partial B_\epsilon(\vec{p})} f \partial_\nu f e^\phi d\sigma,$$

and this converges to 0 as $\epsilon \rightarrow 0$. Hence, $\int_M |\nabla f|^2 dA_g = 0$, so $f = \text{const}$. Thus $R_g \equiv 1$ and, by the uniqueness of constant-curvature metrics with given conic data [Luo and Tian 1992], $g = \bar{g}$. \square

Our goal in the remainder of this section is to prove:

Proposition 5.3. *Let $g(t)$ be the angle-preserving flow on $(M, J, \vec{p}, \vec{\beta})$ and assume that (2-14) fails. Define $\psi(t)$ to be the t -dependent diffeomorphism generated by the vector field $\nabla f(t)$, where $\Delta f(t) = R_{g(t)} - \rho$. Then $\hat{g}(t) := \psi^*g(t)$ satisfies $\partial \hat{g}(t) / \partial t = 2\hat{\mu}(t)$, where $\hat{\mu}$ is the tensor defined by (5-1) with respect to the metric $\hat{g}(t)$. We prove that*

$$\lim_{t \rightarrow \infty} \int_M |\hat{\mu}(t)|_{\hat{g}(t)}^2 d\hat{A} = \lim_{t \rightarrow \infty} \int_M |\mu(t)|_{g(t)}^2 dA = 0$$

and, moreover,

$$\lim_{t \rightarrow \infty} \int_M |X(t)|_{g(t)}^2 dA = 0,$$

where $X = \nabla R + R \nabla f$.

In the next subsections we assemble various facts which lead to the proof of this proposition. These were all initially developed in the smooth case, and the main work here consists mainly in verifying that they remain true in this conic setting.

The outline of this proof is as follows: In Section 5A we adapt Perelman’s arguments for volume noncollapsing for the Kähler–Ricci flow; see [Sesum and Tian 2008]. We then follow the arguments in [Hamilton 1988], making use of the entropy functional $N(g) = \int_M R \log R dV_g$, and showing that $N(g(t)) \leq C$ here too. In Section 5C we explain how to apply the maximum principle in the proof of the Harnack inequality, and hence obtain that $R_{\text{sup}} \leq C R_{\text{inf}}$. Area noncollapsing, entropy monotonicity and the Harnack estimate then show that $R \leq C$ for all $t \in [0, \infty)$. We also show $R \geq c > 0$ for $t \geq t_0$.

5A. Area noncollapsing via Perelman’s monotonicity formula. Our first goal is to prove an estimate on the area of small geodesic balls.

Lemma 5.4. *Let $(M, g(t))$ be a compact conic surface evolving by the angle-preserving area-normalized Ricci flow. Define $R_{\text{max}}(t) = \sup_{q \in M} R_{g(t)}$. Then there exists $C > 0$ so that, for all $p \in M$ and $t > 0$, we have*

$$\operatorname{Area}_{g(t)} B(p, R_{\text{max}}(t)^{-1/2}) \geq \frac{C}{R_{\text{max}}(t)}.$$

Proof. The proof relies on monotonicity properties with respect to the unnormalized Ricci flow of Perelman’s \mathcal{W} functional,

$$\mathcal{W}(g, f, \tau) = (4\pi\tau)^{-1} \int_M [\tau(|\nabla f|^2 + R) + f - 2]e^{-f} dA_g,$$

where g is a metric, f is a function and $\tau \in \mathbb{R}^+$. For us, g is a polyhomogeneous conic metric and $f(z, t) = a_0(t) + (a_{11}(t) \cos y + a_{12}(t) \sin y)r^{1/\beta} + r^2 \tilde{f}(z, t)$, where both f and \tilde{f} lie in $\mathcal{C}^1([0, \infty); \mathcal{C}_b^{2,\delta})$. This integral is convergent, since $|\nabla f|$ is bounded.

We review the proof of monotonicity of this functional to check that the singularities of g and f do not cause difficulties. We restrict to the space of triples (g, f, τ) such that the measure $(4\pi\tau)^{-1/2}e^{-f} dA_g$ is fixed. If (v, h, σ) is a tangent vector to this space then, by [Kleiner and Lott 2008, Propositions 5.3 and 12.1],

$$\delta\mathcal{W}|_{(g,f,\tau)}(v, h, \sigma) = (4\pi\tau)^{-1} \int_M [\sigma(R_g + |\nabla f|^2) - \tau\langle v, \text{Ric}_g + \nabla^2 f \rangle + h]e^{-f} dA_g$$

This requires justifying the three integrations by parts

$$\begin{aligned} \int_M e^{-f} (-\Delta \text{tr}_g v) dA &= - \int_M \Delta(e^{-f}) \text{tr}_g v dA, \\ \int_M e^{-f} \delta^* \delta^* v dA &= \int_M \langle \nabla^2 e^{-f}, v \rangle dA, \\ \int_M e^{-f} \langle \nabla f, \nabla h \rangle dA &= \int_M \Delta e^{-f} h dA, \end{aligned}$$

which we do in the usual way, using the expansion for f .

Still following [Kleiner and Lott 2008, §12], set $v = -2(\text{Ric}_g + \nabla^2 f)$, so $\text{tr}_g v = -2(R_g + \Delta f)$, and also $h = -\Delta f + |\nabla f|^2 - R_g + 1/(2\tau)$, $\sigma = -1$. Then

$$\delta\mathcal{W}|_{(g,f,\tau)}(v, h, \sigma) = \int_M \tau \left| \text{Ric}_g + \nabla^2 f - \frac{1}{2\tau} g \right|^2 (4\pi\tau)^{-1} e^{-f} dA_g \geq 0.$$

To recover the actual Ricci flow, we add to v and h the Lie derivative terms $\mathcal{L}_V g$ and $\mathcal{L}_V f = Vf$, respectively, where $V = \nabla f$. This new infinitesimal variation corresponds to the flow

$$\partial_t g = -2 \text{Ric}_g, \quad \partial_t f = -\Delta f + |\nabla f|^2 - R_g + \frac{1}{\tau} \quad \text{and} \quad \partial_t \tau = -1,$$

along which we have

$$\delta\mathcal{W}|_{(g,f,\tau)}(v, h, \sigma) = \int_M 2\tau \left| \nabla^2 f + \frac{1}{2} \left(R_g - \frac{1}{\tau} \right) g_{ij} \right|^2 (4\pi\tau)^{-1/2} e^{-f} dA_g.$$

Finally, define

$$\mu(g, \tau) := \inf_f \{ \mathcal{W}(g, f, \tau) : (4\pi\tau)^{-1/2} \int_M e^{-f} dA_g = 1 \}.$$

We have proved that $\mu(g(t), \tau(t))$ increases along the Ricci flow. Using this monotonicity, we follow precisely the same arguments as in Perelman’s proof of volume noncollapsing for the Kähler–Ricci flow (see [Sesum and Tian 2008] for details). □

5B. Entropy estimate. The potential function f satisfies $\Delta f = R - \rho$. Define the symmetric, trace-free 2-tensor

$$\mu = \nabla^2 f - \frac{1}{2} \Delta f g \tag{5-1}$$

and the vector field $X = \nabla R + R \nabla f$. As in the earlier part of this section, g is a gradient Ricci soliton if $\mu \equiv 0$; in fact, one also has $X \equiv 0$ on any soliton. The entropy function introduced by Hamilton [1988] when R_{g_0} is a strictly positive function on M is the quantity

$$N(t) = \int_M R \log R dA.$$

When R changes sign, Chow [1991b] considered the modified entropy

$$N(t) = \int_M (R - s) \log(R - s) dA, \tag{5-2}$$

where $s'(t) = s(s - r)$ with $s(0) < \min_{x \in M} R(x, 0)$. In either case, if M is smooth, these authors showed that $N(t) \leq C$ for $t \geq 0$; in the first case, this is based on the monotonicity of N , which follows from the formula

$$\frac{dN}{dt} = - \int \left(2|\mu|^2 + \frac{|X|^2}{R} \right) dA. \tag{5-3}$$

We now prove that this entropy function, or its modified form, is still bounded above even in the conic setting.

Lemma 5.5. *If $g(t)$ is an angle-preserving solution of the normalized Ricci flow, and if the entropy N is defined by (5-3) if $R > 0$ everywhere and by (5-2) if R changes signs, then $N(t) \leq C$ for all $t < \infty$.*

Proof. The argument proceeds exactly as in the smooth case once we show that the various integrations by parts are justified. We assume that R does change signs, since the two cases are very similar, and follow Chow’s [1991a] proof on orbifolds.

Define $L = \log(R - s)$. The proof relies on the following identities:

$$\begin{aligned} \int \Delta L(R - s) &= - \int \frac{|\nabla R|^2}{R - s}, & \int L \Delta R &= - \int \frac{|\nabla R|^2}{R - s}, \\ \int (\Delta f)^2 &= - \int \langle \nabla f, \nabla \Delta f \rangle, & \int \langle \nabla f, \Delta \nabla f \rangle &= - \int |D^2 f|^2, \\ \int f \Delta f &= - \int |\nabla f|^2, & \int \langle \nabla R, \nabla f \rangle &= - \int R \Delta f = - \int R(R - r), \\ \int \langle \nabla L, \nabla f \rangle &= - \int L \Delta f = - \int L(R - r); \end{aligned}$$

these are all proven using Green’s identity on $M \setminus B(\vec{p}, \epsilon)$ and taking advantage of the expansions of f and R to show that the boundary terms vanish in the limit $\epsilon \rightarrow 0$. □

5C. Harnack estimate and curvature bound. The proof of the Harnack estimate for R , when $R > 0$ everywhere, or for $R - s$ if R changes sign, again proceeds exactly as in the smooth [Chow 1991b] and orbifold [Chow and Wu 1991] cases, although now using the maximum principles from Lemmas 3.15 and 3.16. We outline the main step. Consider $P = Q + sL$, where

$$Q = \partial_t L - |\nabla L|^2 - s = \Delta L + R - \rho \quad \text{and} \quad L = \log(R - s).$$

One computes that

$$\partial_t P \geq \Delta P + 2\nabla L \cdot \nabla P + \frac{1}{2}(P^2 - C^2), \quad (5-4)$$

where C is a constant chosen so that $L \geq -C - Ct$. By Corollary 3.14, R is polyhomogeneous (for $t > 0$) and the only terms in its expansion less than r^2 are r^0 and $r^{1/\beta}$. Using (2-7), the initial terms in the expansion of ΔR have the same exponents. Thus, $|\nabla P|$ satisfies the conditions in these maximum principle lemmas, and we conclude that $Q \geq -C$, independently of t . The usual integration in spatial and time variables leads to the Harnack inequality — see [Chow 1991b] for details — and thus gives:

Lemma 5.6. *If $y \in B_{g(t)}(x, \frac{1}{8}\sqrt{R(x, t)})$, then $R(y, t + 1) \geq CR(x, t)$ for some universal constant $C > 0$.*

Using the entropy bound and area comparison, the boundedness of R follows as in [Hamilton 1988; Chow 1991b].

Corollary 5.7. *There exist constants $c, C > 0$ such that $|R(\cdot, t)| < C$ for all $t > 0$ and $R(\cdot, t) \geq c$ for $t \gg 1$.*

Proof of Proposition 5.3. Consider the following modification of the Ricci flow equation:

$$\frac{\partial}{\partial t} \hat{g}_{ij} = 2\hat{\mu}_{ij} = (\rho - \hat{R})\hat{g}_{ij} - 2\nabla_i \nabla_j f, \quad (5-5)$$

where \hat{R} is the scalar curvature of \hat{g} (the covariant derivatives in the last term are also with respect to $\hat{g}(t)$, but we omit this from the notation for simplicity) and f is the same potential function as before. This differs from the standard flow by the action of the one-parameter family of diffeomorphisms ψ_t generated by ∇f , i.e., $\hat{g}(t) = \psi_t^* g(t)$, where $g(t)$ is a solution of the original normalized (but unmodified) Ricci flow. According to Lemma 5.5, Corollary 5.7, and (5-3), $N(t)$ is monotone (for t sufficiently large) and converges to a finite limit, hence

$$\lim_{t \rightarrow \infty} \frac{dN}{dt} = 0;$$

recalling (5-3), the conclusion follows from this. □

Remark 5.8. Hamilton's original argument showing that the pointwise norm of μ converges exponentially to zero breaks down in our setting for the following reason. As for many of the other quantities we consider here, the function f admits an expansion

$$f = a_0(t) + r^{1/\beta}(a_{11}(t) \cos y + a_{12}(t) \sin y) + a_2(t)r^2 + \mathcal{O}(r^{2+\epsilon}),$$

where the a_i and a_{ij} are smooth in t . This follows from the equation satisfied by R , the equation

$$\Delta_{g(t)} f(t) = R(t) - \rho = -\rho + \frac{R_0 - \Delta_{g(0)} \log u(t)}{u(t)},$$

the asymptotic expansion (1-1) of $u(t)$, and [Jeffres et al. 2014, Corollary 3.5]. However,

$$\mu = \nabla^2 f - \frac{1}{2} \Delta f g \tag{5-6}$$

is a second-order operator applied to f , and not all of these annihilate the troublesome term $r^{1/\beta}$ in the expansion of f . This means that, although μ , and hence $|\mu|$, has an asymptotic expansion, this expansion contains a singular term of the form $r^{1/\beta-2}$. This means that the maximum principle is not applicable, and we cannot conclude the exponential decay of $|\mu|$. Note that there is no difficulty with what we prove above, since this most singular term $r^{1/\beta-2}$ is square-integrable with respect to $r dr dy$.

5D. One concentration point. We now prove the second part of Theorem 1.4, concerning the divergence profile of the unmodified flow. Namely, we show that the conformal factor ϕ blows up at precisely one point q as $t \nearrow \infty$, but tends uniformly to zero on every compact set $K \subset S^2 \setminus \{q\}$. This argument is drawn from methods developed specifically for higher-dimensional complex analysis, so it is convenient to now change to the Kähler formalism.

Fix the initial conic metric g_0 ; since the flow immediately smooths out any initial metric, we may as well assume that g_0 is polyhomogeneous. Denote its associated Kähler form by ω . Define \mathcal{H}_ω to consist of all functions ϕ such that $\omega_\phi := \omega + \sqrt{-1} \partial \bar{\partial} \phi > 0$, and then denote by PSH_ω the L^1 closure of \mathcal{H}_ω . Observe that, since ω and ω_ϕ (or, rather, g_0 and g_ϕ) lie in the same Kähler class, they are conformally related; indeed,

$$\omega_\phi = (1 + \Delta_0 \phi) \omega, \quad \text{and similarly} \quad \omega = (1 - \Delta_\phi \phi) \omega_\phi.$$

Here Δ_0 and Δ_ϕ are the Laplacians for ω and ω_ϕ , respectively. Note that this implies that

$$\Delta_0 \phi > -1 \quad \text{and} \quad \Delta_\phi \phi < 1. \tag{5-7}$$

For any $\phi \in \text{PSH}_\omega$, we define the multiplier ideal sheaf $\mathcal{I}(\phi)$ associated to the presheaf which assigns to any open set U the space of holomorphic functions

$$\mathcal{I}(\phi)(U) = \{h \in \mathcal{O}_{S^2}(U) : |h|^2 e^{-\phi} \in L^1_{\text{loc}}(S^2, \omega)\}.$$

It is proved in [Nadel 1990] that $\mathcal{I}(\phi)$ is always coherent; moreover, it is called proper if it is neither the trivial (zero) sheaf nor the structure sheaf \mathcal{O}_{S^2} .

Definition 5.9 [Nadel 1990, Definition 2.4]. The multiplier ideal sheaf $\mathcal{I}(\phi)$ is called a Nadel sheaf if there exists an $\epsilon > 0$ such that $(1 + \epsilon)\phi \in \text{PSH}_\omega$.

A fundamental result of Nadel’s [1990] is that any Nadel sheaf has connected support. The proof is not hard in this low dimension, so we give it below. This uses an extension (for the one-dimensional case only) of the result [Rubinstein 2009, Theorem 1.3], which in turn extends Nadel’s work from the continuity method to the Ricci flow. Note too that [Rubinstein 2009] provided a new proof of the uniformization

theorem in the smooth case using the Ricci flow (see also [Chen et al. 2006] for an earlier and different flow-based proof), and hence its use here is natural.

We write the flow equation in this setting in terms of the Kähler potential as

$$\omega + \sqrt{-1}\partial\bar{\partial}\phi := \omega_\phi = e^{f_\omega - \phi + \dot{\phi}}\omega, \quad \phi(0) = \phi_0, \quad \text{and} \quad \dot{\phi} = \partial_t\phi, \quad (5-8)$$

where f_ω is the initial value $f(0)$ of the Ricci potential, as defined in (3-24). There is a choice of constant in this initial condition, and it is explained in [Phong et al. 2007] how to choose this additional constant so that $\dot{\phi}$ remains bounded along the flow. We assume henceforth that this initial condition has been set properly. We also write A for the (constant value of the) area of $(S^2, g(t))$.

Theorem 5.10. *Suppose that $(S^2, J, \vec{p}, \vec{\beta})$ does not satisfy (2-14). Fix any $\gamma \in (\frac{1}{2}, 1)$. Then the solution $\phi(t)$, normalized as above, admits a subsequence $\phi_j := \phi_{t_j}$ for which $\hat{\phi}_j := \phi_{t_j} - A^{-1} \int \phi_{t_j}$ converges in L^1 to $\phi_\infty \in \text{PSH}_\omega$. Finally, $\mathcal{I}(\gamma\phi_\infty)$ is a proper Nadel multiplier ideal sheaf with support equal to a single point.*

Proof. We proceed in a series of steps.

Step 1: $\text{diam}(M, g(t)) \leq C$. This is a special case of [Jeffres et al. 2014, Claim 6.4]. Indeed, since $\beta < 1$, if $p, q \in M$ are not conic points then the minimizing geodesic which connects them does not pass through a conic point. Thus we can apply the standard argument for Myers' theorem, using that $R > c > 0$ for large t . This can also be deduced by specializing Perelman's diameter estimate [Sesum and Tian 2008] to our setting, which is possible using Theorem 1.2.

Step 2: $-\inf \phi \leq \sup \phi + C$. The proof of [Rubinstein 2009, Lemma 2.2] carries over without change by using the twisted Berger–Moser–Ding functional

$$D(\phi) = \frac{\sqrt{-1}}{2A} \int \partial\phi \wedge \bar{\partial}\phi - \log\left(\frac{1}{A} \int e^{f_\omega - \phi} \omega\right).$$

This is monotone along the flow, which gives, after some calculations, that [Rubinstein 2009, (15)]

$$\frac{1}{A} \int -\phi\omega_\phi \leq \frac{1}{A} \int \phi\omega + C. \quad (5-9)$$

We next show that the average $A^{-1} \int \phi\omega$ is comparable to $\sup \phi$. Indeed, the inequality $A^{-1} \int \phi\omega \leq \sup \phi$ is trivial. For the converse, recall that the Green function of Δ_0 , normalized so that $\int G_0(q, q')\omega(q') = 0$ for every q and $G \searrow -\infty$ near the diagonal, is bounded from above by a constant E_0 . We then write

$$\begin{aligned} \phi(q) - \frac{1}{A} \int \phi\omega &= \int G(q, q')\Delta_\omega\phi(q')\omega(q') \\ &= \int -(G(q, q') - E_0)(-\Delta_\omega\phi(q'))\omega(q') \leq - \int (G(q, q') - E_0)\omega \leq AE_0, \end{aligned}$$

using the first inequality in (5-7). Taking the supremum over the left side gives $\sup \phi \leq (1/A) \int \phi\omega + C$, as claimed.

To estimate the infimum of ϕ we use a similar trick, but using the upper bound $G_\phi(q, q') \leq E_\phi$ and the second inequality in (5-7). This gives

$$\begin{aligned} \phi(q) - \frac{1}{A} \int \phi \omega_\phi &= \int G_\phi(q, q') \Delta_\phi \phi(q') \omega_\phi(q') \\ &= \int (G(q, q') - E_0) \Delta_\omega \phi(q') \omega_\phi(q') \geq \int (G(q, q') - E_0) \omega_\phi(q') \geq -AE_\phi, \end{aligned}$$

so, taking the infimum, $-\inf \phi \leq -(1/A) \int \phi \omega_\phi + AE_\phi$.

It remains only to observe that, special to this dimension, $G_\phi(q, q') = G_0(q, q')$; this is because, if we write $\omega_\phi = F\omega$, and if $\int f \omega_\phi = 0$, then

$$\Delta_\phi \int G_\phi(q, q') f(q') \omega_\phi(q') = F^{-1} \Delta_0 \int G_\phi(q, q') f(q') F(q') \omega(q'),$$

and this equals $f(q)$ when $G_\phi = G_0$. This means that $E_\phi = E_0$ and the constant in this inequality does not vary along the flow.

Putting these inequalities together completes this step.

Step 3: $\sup_t \|\phi_t\|_0 = \infty$. Indeed, if this supremum were finite, then, by Step 2, ϕ would be bounded in C^0 , and standard regularity estimates would then show that some subsequence of the ϕ_t converges. The limiting metric (or rather, the limit of any one of these subsequences) would then need to have constant curvature. Furthermore, the uniform boundedness of the conformal factor shows that the cone angles do not change in the limit. This is a contradiction, since we are assuming that the Troyanov conditions (2-14) fail.

The construction of the Nadel sheaf now proceeds as in [Rubinstein 2009, p. 5846].

Step 4: If $\gamma \in (\frac{1}{2}, 1)$ and V_γ denotes the support of $\mathcal{I}_\gamma := \mathcal{I}(\gamma\phi_\infty)$, then V_γ is a single point. Recall that a coherent sheaf is locally free away from a complex codimension-two set, so, since we are in complex dimension one, \mathcal{I}_γ is a sheaf of sections of a holomorphic line bundle $\mathcal{O}_{S^2}(-k)$, $k \geq 0$. By the properness assumption, $k \geq 1$. We claim that $k = 1$, which then implies that \mathcal{I}_γ is spanned by a single holomorphic section, which vanishes to order one at precisely one point.

To do this, let U be a small open set and let $h \in \mathcal{I}_\gamma(U)$, and assume that h vanishes exactly to order one at a point $p \in U$. Then $\int_U |h|^2 e^{-\gamma\phi_\infty} \omega < \infty$. Now fix a local holomorphic coordinate z which vanishes at p and assume either that U contains no conic points or, if it does contain one, then p is that point. In the first of these cases, ω is locally equivalent to $|dz|^2$, while $\phi_\infty \geq 4 \log |z|$, and $0 < \gamma < 1$. If $p = p_i$ is a conic point, then, assuming that U contains no other conic points, $\int_U |h|^2 e^{-\gamma\phi_\infty} z^{2\beta_i-2} |dz|^2 < \infty$. This follows just as before but using that ϕ_∞ has a singularity of, at worst, $4 \log |z| - 2(1 - \beta_i) \log |z|$ (recall $[\omega] = \mathcal{O}_{S^2}(2) - \sum(1 - \beta_i)[p_i]$) and $0 < \gamma < 1$. Thus k cannot be greater than one. Since $k > 0$, it follows that $k = 1$, as desired.

There is an alternative proof that does not rely on facts about coherent sheaves, using weighted L^2 estimates for the $\bar{\partial}$ -equation. This proceeds as follows. Let η be a $(0, 1)$ -form such that $\int |\eta|^2 e^{-\gamma\phi_\infty} |dz|^2 < \infty$. It is always possible [Berndtsson 2010, §1] to find a solution ρ to $\bar{\partial}\rho = \eta$ that satisfies $\int |\rho|^2 e^{-\gamma\phi_\infty} |dz|^2 \leq$

$C_\gamma \int |\eta|^2 e^{-\gamma\phi_\infty} |dz|^2 < \infty$, where $C_\gamma = O((1-\gamma)^{-1})$. The same arguments can be used to verify that this estimate also holds with respect to the measure $|z|^{2\beta_i-2} |dz|^2$. This proves that $H^1(S^2, \mathcal{I}_\gamma) = 0$. From the long exact sequence in cohomology corresponding to the short exact sequence of sheaves $0 \rightarrow \mathcal{I}_\gamma \rightarrow \mathcal{O}_{S^2} \rightarrow \mathcal{O}_{V_\gamma} \rightarrow 0$, one concludes that $H^0(V_\gamma, \mathcal{O}_{V_\gamma}) \cong H^0(S^2, \mathcal{O}_{S^2}) \cong \mathbb{C}$, which means once again that the support of \mathcal{I}_γ is connected, i.e., a single point.

These two methods of proof are closely related, of course, by virtue of the identification $H^1(S^2, \mathcal{I}_\gamma) \cong H^0(S^2, \mathcal{O}_{S^2}(K_{S^2} - \mathcal{I}_\gamma))$. \square

Following [Clarke and Rubinstein 2013, Lemma 6.5], we can use Theorem 5.10 to deduce estimates on the conformal factor:

Corollary 5.11. *The conformal factor u blows up at exactly one point. On any compact set K disjoint from that point, $u \rightarrow 0$ uniformly, so, in particular, the area of K with respect to $g(t)$ tends to 0.*

Acknowledgements

The authors gratefully acknowledge support by the NSF under grants DMS-1105050 (Mazzeo), DMS-0802923,1206284 (Rubinstein), and DMS-0905749,1056387 (Sesum). Rubinstein is also grateful to the Stanford MRC and grant DMS-0603355 for support during the academic year 2009–2010. The authors also wish to acknowledge the careful reading by the anonymous referees as well as by J. Song, whose comments and suggestions helped us clarify and correct a number of points. In particular, this updated version of our paper contains a weaker version of what was originally stated as Proposition 5.3.

References

- [Albin et al. 2013] P. Albin, C. L. Aldana, and F. Rochon, “Ricci flow and the determinant of the Laplacian on non-compact surfaces”, *Comm. Partial Differential Equations* **38**:4 (2013), 711–749. MR 3040681 Zbl 1284.53060
- [Bahuaud and Vertman 2014] E. Bahuaud and B. Vertman, “Yamabe flow on manifolds with edges”, *Math. Nachr.* **287**:2-3 (2014), 127–159. MR 3163570 Zbl 1305.53065
- [Bartolucci et al. 2011] D. Bartolucci, F. De Marchis, and A. Malchiodi, “Supercritical conformal metrics on surfaces with conical singularities”, *Int. Math. Res. Not.* **2011**:24 (2011), 5625–5643. MR 2863376 Zbl 1254.30066
- [Berman 2013] R. J. Berman, “A thermodynamical formalism for Monge–Ampère equations, Moser–Trudinger inequalities and Kähler–Einstein metrics”, *Adv. Math.* **248** (2013), 1254–1297. MR 3107540 Zbl 1286.58010
- [Berndtsson 2010] B. Berndtsson, “An introduction to things $\bar{\partial}$ ”, pp. 7–76 in *Analytic and algebraic geometry*, edited by J. McNeal and M. Mustata, IAS/Park City Math. Ser. **17**, Amer. Math. Soc., Providence, RI, 2010. MR 2012c:32052 Zbl 1227.32039
- [Berndtsson 2015] B. Berndtsson, “A Brunn–Minkowski type inequality for Fano manifolds and some uniqueness theorems in Kähler geometry”, preprint, 2015. arXiv 1303.4975
- [Carlotto and Malchiodi 2012] A. Carlotto and A. Malchiodi, “Weighted barycentric sets and singular Liouville equations on compact surfaces”, *J. Funct. Anal.* **262**:2 (2012), 409–450. MR 2854708 Zbl 1232.53035
- [Chen and Wang 2013] X. Chen and Y. Wang, “Bessel functions, heat kernel and the conical Kähler–Ricci Flow”, preprint, 2013. arXiv 1305.0255
- [Chen et al. 2006] X. Chen, P. Lu, and G. Tian, “A note on uniformization of Riemann surfaces by Ricci flow”, *Proc. Amer. Math. Soc.* **134**:11 (2006), 3391–3393. MR 2007d:53109 Zbl 1113.53042

- [Chow 1991a] B. Chow, “On the entropy estimate for the Ricci flow on compact 2-orbifolds”, *J. Differential Geom.* **33**:2 (1991), 597–600. MR 92e:58228 Zbl 0734.53034
- [Chow 1991b] B. Chow, “The Ricci flow on the 2-sphere”, *J. Differential Geom.* **33**:2 (1991), 325–334. MR 92d:53036 Zbl 0734.53033
- [Chow and Wu 1991] B. Chow and L.-F. Wu, “The Ricci flow on compact 2-orbifolds with curvature negative somewhere”, *Comm. Pure Appl. Math.* **44**:3 (1991), 275–286. MR 92g:53035 Zbl 0745.58047
- [Chow et al. 2006] B. Chow, P. Lu, and L. Ni, *Hamilton’s Ricci flow*, Graduate Studies in Mathematics **77**, American Mathematical Society, Providence, RI, 2006. MR 2008a:53068 Zbl 1118.53001
- [Clarke and Rubinstein 2013] B. Clarke and Y. A. Rubinstein, “Ricci flow and the metric completion of the space of Kähler metrics”, *Amer. J. Math.* **135**:6 (2013), 1477–1505. MR 3145001 Zbl 1288.58005
- [Donaldson 2012] S. K. Donaldson, “Kähler metrics with cone singularities along a divisor”, pp. 49–79 in *Essays in mathematics and its applications*, edited by P. M. Pardalos and T. M. Rassias, Springer, Heidelberg, 2012. MR 2975584
- [Gell-Redman 2011] J. Gell-Redman, “Harmonic maps into conic surfaces with cone angles less than 2π ”, preprint, 2011. To appear in *Comm. Anal. Geom.* arXiv 1010.4156
- [Giesen and Topping 2010] G. Giesen and P. M. Topping, “Ricci flow of negatively curved incomplete surfaces”, *Calc. Var. Partial Differential Equations* **38**:3-4 (2010), 357–367. MR 2011d:53155 Zbl 1244.53076
- [Giesen and Topping 2011] G. Giesen and P. M. Topping, “Existence of Ricci flows of incomplete surfaces”, *Comm. Partial Differential Equations* **36**:10 (2011), 1860–1880. MR 2012i:53063 Zbl 1233.35123
- [Gil et al. 2006] J. B. Gil, T. Krainer, and G. A. Mendoza, “Resolvents of elliptic cone operators”, *J. Funct. Anal.* **241**:1 (2006), 1–55. MR 2007j:58032 Zbl 1109.58023
- [Hamilton 1988] R. S. Hamilton, “The Ricci flow on surfaces”, pp. 237–262 in *Mathematics and general relativity* (Santa Cruz, CA, 1986), edited by J. A. Isenberg, Contemp. Math. **71**, Amer. Math. Soc., Providence, RI, 1988. MR 89i:53029 Zbl 0663.53031
- [Isenberg et al. 2011] J. Isenberg, R. Mazzeo, and N. Sesum, “Ricci flow in two dimensions”, pp. 259–280 in *Surveys in geometric analysis and relativity*, edited by H. L. Bray, Adv. Lect. Math. (ALM) **20**, International Press, 2011. MR 2906929 Zbl 1268.53072
- [Jeffres 2005] T. D. Jeffres, “A maximum principle for parabolic equations on manifolds with cone singularities”, *Adv. Geom.* **5**:2 (2005), 319–323. MR 2005m:53055 Zbl 1075.58016
- [Jeffres and Loya 2003] T. D. Jeffres and P. Loya, “Regularity of solutions of the heat equation on a cone”, *Int. Math. Res. Not.* **2003**:3 (2003), 161–178. MR 2003i:58043 Zbl 1183.58023
- [Jeffres et al. 2014] T. D. Jeffres, R. Mazzeo, and Y. A. Rubinstein, “Kähler–Einstein metrics with edge singularities”, preprint, 2014. arXiv 1105.5216
- [Kleiner and Lott 2008] B. Kleiner and J. Lott, “Notes on Perelman’s papers”, *Geom. Topol.* **12**:5 (2008), 2587–2855. MR 2010h:53098 Zbl 1204.53033
- [Lunardi 1995] A. Lunardi, *Analytic semigroups and optimal regularity in parabolic problems*, Progress in Nonlinear Differential Equations and their Applications **16**, Birkhäuser, Basel, 1995. MR 1329547 Zbl 0816.35001
- [Luo and Tian 1992] F. Luo and G. Tian, “Liouville equation and spherical convex polytopes”, *Proc. Amer. Math. Soc.* **116**:4 (1992), 1119–1129. MR 93b:53034 Zbl 0806.53012
- [Mazzeo 1991] R. Mazzeo, “Elliptic theory of differential edge operators, I”, *Comm. Partial Differential Equations* **16**:10 (1991), 1615–1664. MR 93d:58152 Zbl 0745.58045
- [Mazzeo and Rubinstein 2012] R. Mazzeo and Y. A. Rubinstein, “The Ricci continuity method for the complex Monge–Ampère equation, with applications to Kähler–Einstein edge metrics”, *C. R. Math. Acad. Sci. Paris* **350**:13-14 (2012), 693–697. MR 2971382 Zbl 1273.32043
- [Mazzeo and Rubinstein \geq 2015] R. Mazzeo and Y. A. Rubinstein, “Kähler–Einstein metrics with crossing edge singularities”. In preparation.
- [McOwen 1988] R. C. McOwen, “Point singularities and conformal metrics on Riemann surfaces”, *Proc. Amer. Math. Soc.* **103**:1 (1988), 222–224. MR 89m:30089 Zbl 0657.30033

- [McOwen 1993] R. C. McOwen, “Prescribed curvature and singularities of conformal metrics on Riemann surfaces”, *J. Math. Anal. Appl.* **177**:1 (1993), 287–298. MR 94f:53076 Zbl 0806.53040
- [Melrose 1993] R. B. Melrose, *The Atiyah–Patodi–Singer index theorem*, Research Notes in Mathematics **4**, A K Peters, Ltd., Wellesley, MA, 1993. MR 96g:58180 Zbl 0796.58050
- [Mooers 1999] E. A. Mooers, “Heat kernel asymptotics on manifolds with conic singularities”, *J. Anal. Math.* **78** (1999), 1–36. MR 2000g:58039 Zbl 0981.58022
- [Nadel 1990] A. M. Nadel, “Multiplier ideal sheaves and Kähler–Einstein metrics of positive scalar curvature”, *Ann. of Math.* (2) **132**:3 (1990), 549–596. MR 92d:32038 Zbl 0731.53063
- [Phong et al. 2007] D. H. Phong, N. Sesum, and J. Sturm, “Multiplier ideal sheaves and the Kähler–Ricci flow”, *Comm. Anal. Geom.* **15**:3 (2007), 613–632. MR 2009a:32037 Zbl 1143.53064
- [Phong et al. 2014] D. H. Phong, J. Song, J. Sturm, and X. Wang, “The Ricci flow on the sphere with marked points”, preprint, 2014. arXiv 1407.1118
- [Ramos 2011] D. Ramos, “Smoothing cone points with Ricci flow”, preprint, 2011. arXiv 1109.5554
- [Ramos 2013] D. Ramos, “Gradient Ricci solitons on surfaces”, preprint, 2013. arXiv 1304.6391
- [Rochon 2014] F. Rochon, “Polyhomogénéité des métriques asymptotiquement hyperboliques complexes le long du flot de Ricci”, preprint, 2014. arXiv 1305.5457
- [Rochon and Zhang 2012] F. Rochon and Z. Zhang, “Asymptotics of complete Kähler metrics of finite volume on quasiprojective manifolds”, *Adv. Math.* **231**:5 (2012), 2892–2952. MR 2970469 Zbl 1264.53066
- [Ross and Thomas 2011] J. Ross and R. Thomas, “Weighted projective embeddings, stability of orbifolds, and constant scalar curvature Kähler metrics”, *J. Differential Geom.* **88**:1 (2011), 109–159. MR 2819757 Zbl 1244.32013
- [Rubinstein 2009] Y. A. Rubinstein, “On the construction of Nadel multiplier ideal sheaves and the limiting behavior of the Ricci flow”, *Trans. Amer. Math. Soc.* **361**:11 (2009), 5839–5850. MR 2010i:32021 Zbl 1185.53077
- [Rubinstein 2014] Y. A. Rubinstein, “Smooth and singular Kähler–Einstein metrics”, *Contemp. Math.* **630** (2014), 45–138.
- [Sesum and Tian 2008] N. Sesum and G. Tian, “Bounding scalar curvature and diameter along the Kähler Ricci flow (after Perelman)”, *J. Inst. Math. Jussieu* **7**:3 (2008), 575–587. MR 2009c:53092 Zbl 1147.53056
- [Simon 2002] M. Simon, “Deformation of C^0 Riemannian metrics in the direction of their Ricci curvature”, *Comm. Anal. Geom.* **10**:5 (2002), 1033–1074. MR 2003j:53107 Zbl 1034.58008
- [Struwe 2002] M. Struwe, “Curvature flows on surfaces”, *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* **1**:2 (2002), 247–274. MR 2004f:53083 Zbl 1150.53025
- [Tian and Zhang 2013] G. Tian and Z. Zhang, “Regularity of the Kähler–Ricci flow”, *C. R. Math. Acad. Sci. Paris* **351**:15–16 (2013), 635–638. MR 3119891 Zbl 1278.53073
- [Trojanov 1991] M. Troyanov, “Prescribing curvature on compact surfaces with conical singularities”, *Trans. Amer. Math. Soc.* **324**:2 (1991), 793–821. MR 91h:53059 Zbl 0724.53023
- [Yin 2010] H. Yin, “Ricci flow on surfaces with conical singularities”, *J. Geom. Anal.* **20**:4 (2010), 970–995. MR 2012b:53146 Zbl 1201.53049
- [Yin 2013] H. Yin, “Ricci flow on surfaces with conical singularities, II”, preprint, 2013. arXiv 1305.4355

Received 26 May 2014. Revised 27 Jan 2015. Accepted 6 Mar 2015.

RAFE MAZZEO: mazzee@math.stanford.edu

Department of Mathematics, Stanford University, Stanford, CA 94305, United States

YANIR A. RUBINSTEIN: yanir@umd.edu

Department of Mathematics, University of Maryland, College Park, MD 20742, United States

NATASA SESUM: natasas@math.rutgers.edu

Department of Mathematics, Rutgers University, 110 Frelinghuysen road, Piscataway, NJ 10027, United States

GROWTH OF SOBOLEV NORMS FOR THE QUINTIC NLS ON T^2

EMANUELE HAUS AND MICHELA PROCESI

We study the quintic nonlinear Schrödinger equation on a two-dimensional torus and exhibit orbits whose Sobolev norms grow with time. The main point is to reduce to a sufficiently simple toy model, similar in many ways to the one discussed by Colliander et al. for the case of the cubic NLS. This requires an accurate combinatorial analysis.

1. Introduction

We consider the quintic defocusing NLS on the two-dimensional torus $\mathbb{T}^2 = \mathbb{R}^2 / (2\pi\mathbb{Z})^2$

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

which is an infinite-dimensional dynamical system with Hamiltonian

$$H = \int_{\mathbb{T}^2} |\nabla u|^2 + \frac{1}{3} \int_{\mathbb{T}^2} |u|^6 \tag{1-2}$$

having the mass (the L^2 norm) and momentum

$$L = \int_{\mathbb{T}^2} |u|^2 \quad \text{and} \quad M = \int_{\mathbb{T}^2} \Im(u \cdot \nabla u) \tag{1-3}$$

as constants of motion. The well-posedness result of [Bourgain 1993; Burq et al. 2004] for data $u_0 \in H^s(\mathbb{T}^2)$, $s \geq 1$, gives the existence of a global-in-time smooth solution to (1-1) from smooth initial data, and one would like to understand some qualitative properties of solutions.

A fruitful approach to this question is to apply the powerful tools of *singular perturbation theory*, such as KAM theory, the Birkhoff normal form and Arnold diffusion, first developed in order to study finite-dimensional systems.

We are interested in the phenomenon of the growth of Sobolev norms, i.e., we look for solutions which initially oscillate only on scales comparable to the spatial period and eventually oscillate on arbitrarily short spatial scales. This is a natural extension of the results in [Colliander et al. 2010; Guardia and Kaloshin 2015], which prove similar results for the cubic NLS. In the strategy of the proof, we follow [Colliander et al. 2010] — henceforth abbreviated [CKSST] — as closely as possible; therefore our main result is the precise analogue of theirs for the cubic NLS. Namely, we prove:

MSC2010: 35B34, 35Q55, 37K45.

Keywords: nonlinear Schrödinger equation, growth of Sobolev norms, Hamiltonian PDEs, weak turbulence.

Theorem 1.1. *Let $s > 1$, $K \gg 1$ and $0 < \delta \ll 1$ be given parameters. Then there exists a global smooth solution $u(t, x)$ to (1-1) and a time $T > 0$ with*

$$\|u(0)\|_{H^s(\mathbb{T}^2)} \leq \delta \quad \text{and} \quad \|u(T)\|_{H^s(\mathbb{T}^2)} \geq K.$$

Note that we are making no claim regarding the time T over which the growth of Sobolev norms occurs; this is the main difference between the approaches of [CKSTT] and [Guardia and Kaloshin 2015].

1A. Some literature. The growth of Sobolev norms for solutions of the nonlinear Schrödinger equation has been studied widely in the literature, but most of the results regard upper bounds on such growth. In the one-dimensional case with an analytic nonlinearity $\partial_{\bar{u}} P(|u|^2)$, Bourgain [1996b] and Staffilani [1997] proved at-most polynomial growth of Sobolev norms. In the same context, Bourgain [2000] proved a Nekhoroshev-type theorem for a perturbation of the cubic NLS. Namely, for s large and a typical initial datum $u(0) \in H^s(T)$ of small size $\|u(0)\|_s \leq \varepsilon$, he proved

$$\sup_{t \leq T} \|u(t)\|_s \leq C\varepsilon, \quad |t| < T, \quad T \leq \varepsilon^{-A},$$

with $A = A(s) \rightarrow 0$ as $s \rightarrow \infty$. Similar upper bounds on the growth have been obtained also for the NLS equation on \mathbb{R} and \mathbb{R}^2 as well as on compact manifolds.

We finally mention [Faou et al. 2013], which discusses the existence of stability regions for the NLS on tori.

Concerning instability results for the NLS on tori, we mention the work by Kuksin [1997b] (and see his related works [1995; 1996; 1997a; 1999]) who studied the growth of Sobolev norms for the equation

$$-i\partial_t u + \delta\Delta u = |u|^{2p}u, \quad p \in \mathbb{N},$$

and constructed solutions whose Sobolev norms grow by an inverse power of δ . Note that the solutions that he obtains (for $p = 2$) correspond to orbits of (1-1) with large initial data. A big advance appeared in [CKSTT], where the authors prove Theorem 1.1 for cubic NLS. Note that the initial data are *small* in H^s . Finally, [Guardia and Kaloshin 2015] follows the same general strategy of [CKSTT] and constructs orbits whose Sobolev norm grows (by an arbitrary factor) in a time which is polynomial in the growth factor. This is done by a careful analysis of the equation and using in a clever way various tools from diffusion in finite-dimensional systems.

These results do not imply the existence of solutions with diverging Sobolev norm, nor do they claim that the unstable behavior is typical. Recently, Hani [2014] has made remarkable progress towards the existence of unbounded Sobolev orbits: for a class of cubic NLS equations with nonpolynomial nonlinearity, the combination of a result like Theorem 1.1 with some clever topological arguments leads to the existence of solutions with diverging Sobolev norm. Moreover, Hani et al. [2013] prove infinite growth of Sobolev norms for the cubic NLS on $\mathbb{R} \times \mathbb{T}^2$.

Regarding growth of Sobolev norms for other equations, we mention the following papers: [Bourgain 1996b] for the wave equation with a cubic nonlinearity but with a spectrally defined Laplacian; [Gérard and Grellier 2010; Pocovnicu 2011] for the Szegő equation; and [Pocovnicu 2013] for certain nonlinear

wave equations. We also mention the long time stability results obtained in [Bambusi 1997; 1999; 2003; Bambusi and Grébert 2006; Grébert et al. 2009a; 2009b; Wang 2008; 2010].

A dual point of view to instability is to construct quasiperiodic orbits. These are nongeneric solutions which are global-in-time and whose Sobolev norms are approximately constant. Among the relevant literature we mention [Wayne 1990; Pöschel 1996; Kuksin and Pöschel 1996; Bourgain 1998; Berti and Bolle 2013; Eliasson and Kuksin 2010; Geng et al. 2011; Berti and Biasco 2011; Wang 2014; Procesi and Xu 2013; Berti et al. 2015]. Of particular interest are the recent results obtained through KAM theory, which gives information on linear stability close to the quasiperiodic solutions. In particular, [Procesi and Procesi 2015] proves the existence of both stable and unstable tori (of arbitrary finite dimension) for the cubic NLS.

In finite-dimensional systems diffusive orbits are usually constructed by proving that the stable and unstable manifolds of a *chain of unstable tori* intersect. Usually this is done with tori of codimension one, so that the manifolds should intersect for dimensional reasons. Unfortunately, in the infinite-dimensional case one is not able to prove the existence of codimension-one tori. Actually, the construction of almost-periodic orbits is an open problem except for very special cases, such as integrable equations or equations with infinitely many external parameters (see, for instance, [Pöschel 2002; Chierchia and Peretti 1995; Bourgain 1996a]).

In [CKSTT] and [Guardia and Kaloshin 2015] (and the present paper) this problem is avoided by taking advantage of the specific form of the equation. First one reduces to an approximate equation, the *first-order Birkhoff normal form*; see (1-5). Then, for this dynamical system, one proves directly the existence of chains of one-dimensional unstable tori (periodic orbits) together with their heteroclinic connections. Next, one proves the existence of a *slider solution* which shadows the heteroclinic chain in a finite time. Finally, one proves the persistence of the slider solution for the full NLS. In the next section, we describe the strategy more in detail.

1B. Informal description of the results. In order to understand the dynamics of (1-1), it is convenient to pass to the *interaction representation* picture

$$u(t, x) = \sum_{j \in \mathbb{Z}^2} a_j(t) e^{ij \cdot x + |j|^2 t},$$

so that the equations of motion become

$$-i\dot{a}_j = \sum_{\substack{j_1, j_2, j_3, j_4, j_5 \in \mathbb{Z}^2 \\ j_1 + j_2 + j_3 - j_4 - j_5 = j}} a_{j_1} a_{j_2} a_{j_3} \bar{a}_{j_4} \bar{a}_{j_5} e^{i\omega_6 t}, \tag{1-4}$$

where $\omega_6 = |j_1|^2 + |j_2|^2 + |j_3|^2 - |j_4|^2 - |j_5|^2 - |j|^2$.

We define the *resonant truncation* of (1-4) as

$$-i\dot{\beta}_j = \sum_{\substack{j_1, j_2, j_3, j_4, j_5 \in \mathbb{Z}^2 \\ j_1 + j_2 + j_3 - j_4 - j_5 = j \\ |j_1|^2 + |j_2|^2 + |j_3|^2 - |j_4|^2 - |j_5|^2 = |j|^2}} \beta_{j_1} \beta_{j_2} \beta_{j_3} \bar{\beta}_{j_4} \bar{\beta}_{j_5}. \tag{1-5}$$

It is well known that the dynamics of (1-4) is well approximated by the one of (1-5) for finite but long times.¹ Our aim is to first prove Theorem 1.1 for (1-5) and then extend the result to (1-4) by an approximation lemma. The idea of the approximation lemma roughly speaking is that, by integrating in time the left-hand side of (1-4), one sees that the nonresonant terms (those with $\omega_6 \neq 0$) give a contribution of order $O(a^9)$. By scaling $a^{(\lambda)}(t) = \lambda^{-1}a(\lambda^{-4}t)$ with λ arbitrarily small, we see that the nonresonant terms are an arbitrarily small perturbation with respect to the resonant terms appearing in (1-5) and hence they can be ignored for arbitrarily long finite times.

We now outline the strategy used to prove Theorem 1.1 for (1-5).

The equations (1-5) are Hamiltonian with respect to the Hamiltonian function

$$\mathcal{H} = \frac{1}{3} \sum_{\substack{j_1, j_2, j_3, j_4, j_5, j_6 \in \mathbb{Z}^2 \\ j_1 + j_2 + j_3 = j_4 + j_5 + j_6 \\ |j_1|^2 + |j_2|^2 + |j_3|^2 = |j_4|^2 + |j_5|^2 + |j_6|^2}} \beta_{j_1} \beta_{j_2} \beta_{j_3} \bar{\beta}_{j_4} \bar{\beta}_{j_5} \bar{\beta}_{j_6} \tag{1-6}$$

and the symplectic form $\Omega = i d\beta \wedge d\bar{\beta}$.

This is still a very complicated (infinite-dimensional) Hamiltonian system, but it has the advantage of having many invariant subspaces on which the dynamics simplifies significantly. Let us set up some notation.

Definition 1.2 (resonance). A sextuple $(k_1, k_2, k_3, k_4, k_5, k_6) \in (\mathbb{Z}^2)^6$ is a resonance if

$$k_1 + k_2 + k_3 - k_4 - k_5 - k_6 = 0 \quad \text{and} \quad |k_1|^2 + |k_2|^2 + |k_3|^2 - |k_4|^2 - |k_5|^2 - |k_6|^2 = 0. \tag{1-7}$$

A resonance is *trivial* if it is of the form $(k_1, k_2, k_3, k_1, k_2, k_3)$ up to permutations of the last three elements.

Definition 1.3 (completeness). We say that a set $\mathcal{S} \subset \mathbb{Z}^2$ is *complete* if the following holds: for every quintuple $(k_1, k_2, k_3, k_4, k_5) \in \mathcal{S}^5$, if there exists $k_6 \in \mathbb{Z}^2$ such that $(k_1, k_2, k_3, k_4, k_5, k_6)$ is a resonance, then $k_6 \in \mathcal{S}$.

It is easily seen that, for any complete $\mathcal{S} \subset \mathbb{Z}^2$, the subspace defined by requiring $\beta_k = 0$ for all $k \notin \mathcal{S}$ is invariant.

Definition 1.4 (action-preserving). A complete set $\mathcal{S} \subset \mathbb{Z}^2$ is said to be *action-preserving* if all the resonances in \mathcal{S} are trivial.

We remark that, for any complete and action-preserving $\mathcal{S} \subset \mathbb{Z}^2$, the Hamiltonian restricted to \mathcal{S} is given by (see [Procesi and Procesi 2012])

$$\mathcal{H}|_{\mathcal{S}} = \frac{1}{3} \left(\sum_{j \in \mathcal{S}} |\beta_j|^6 + 9 \sum_{\substack{j, k \in \mathcal{S} \\ j \neq k}} |\beta_j|^4 |\beta_k|^2 + 36 \sum_{\substack{j, k, m \in \mathcal{S} \\ j \prec k \prec m}} |\beta_j|^2 |\beta_k|^2 |\beta_m|^2 \right), \tag{1-8}$$

where \preceq is any fixed total ordering of \mathbb{Z}^2 .

¹Actually, passing to the resonant truncation is equivalent to performing the first step of a Birkhoff normal form. However, since we follow closely the proof in [CKSTT], we chose to use similar notation.

If \mathcal{S} is complete and action-preserving, then $\mathcal{H}|_{\mathcal{S}}$ is a function of the actions $|\beta_j|^2$ only, with nonvanishing twist (i.e., the amplitude-to-frequency map is locally one-to-one); therefore, the corresponding motion is periodic, quasiperiodic or almost-periodic, depending on the initial data. In particular, if $\beta_j(0) = \beta_k(0)$ for all $j, k \in \mathcal{S}$, then the motion is periodic. Finally, since all the actions are constants of motion, so are the H^s norms of the solution.

On the other hand, it is easy to give examples of sets \mathcal{S} that are complete but not action-preserving. For instance, one can consider complete sets of the form $\mathfrak{S}^{(1)} = \{k_1, k_2, k_3, k_4\}$, where the k_j are the vertices of a nondegenerate rectangle in \mathbb{Z}^2 , or of the form $\mathfrak{S}^{(2)} = \{k_1, k_2, k_3, k_4, k_5, k_6\}$, where the $k_j \in \mathbb{Z}^2$ are all distinct and satisfy (1-7). Other examples are sets of the form $\mathfrak{S}^{(3)} = \{k_1, k_2, k_3, k_4\}$ with

$$k_1 + 2k_2 - 2k_3 - k_4 = 0 \quad \text{and} \quad |k_1|^2 + 2|k_2|^2 - 2|k_3|^2 - |k_4|^2 = 0, \quad (1-9)$$

studied in [Grébert and Thomann 2012] or, more generally, the sets $\mathfrak{S}^{(4)} = \bigcup_j \mathfrak{S}_j^{(3)}$ studied in [Haus and Thomann 2013]². In all these cases, the variation of the H^s norm of the solution is of order $O(1)$. Note that, while sets of the form $\mathfrak{S}^{(2)}$, $\mathfrak{S}^{(3)}$, $\mathfrak{S}^{(4)}$ exist in \mathbb{Z}^d for all d , the nondegenerate rectangles $\mathfrak{S}^{(1)}$ exist only in dimension $d \geq 2$. Let us briefly describe the dynamics on these sets. By writing the Hamiltonian in symplectic polar coordinates $\beta_j = \sqrt{I_j} e^{i\theta_j}$, one sees that all these systems are integrable. However, their phase portraits are quite different. In $\mathfrak{S}^{(1)}$ one can exhibit two periodic orbits $\mathbb{T}_1, \mathbb{T}_2$ that are linked by a heteroclinic connection. \mathbb{T}_1 is supported on the modes k_1, k_2 and \mathbb{T}_2 on k_3, k_4 . The H^s norm of each periodic orbit is constant in time. By choosing $\mathfrak{S}^{(1)}$ appropriately, one can ensure that these two values are different, and this produces a growth of the Sobolev norms. Moreover, all the energy is transferred from \mathbb{T}_1 to \mathbb{T}_2 . In the other cases, $\mathfrak{S}^{(2)}$, $\mathfrak{S}^{(3)}$, $\mathfrak{S}^{(4)}$, there is no orbit transferring all the energy from some modes to others (see Appendix C).

These heteroclinic connections are the key to the energy transfer. In fact, assume that

$$\mathcal{S}_1 := \{v_1, \dots, v_n\}, \quad \mathcal{S}_2 := \{w_1, \dots, w_n\}$$

with n even are two complete and action-preserving sets. Assume moreover that, for all $1 \leq j \leq n/2$, $\{v_{2j-1}, v_{2j}, w_{2j-1}, w_{2j}\}$ are the vertices of a rectangle as in $\mathfrak{S}^{(1)}$. Finally, assume that $\mathcal{S}_1 \cup \mathcal{S}_2$ is complete and contains no nontrivial resonances except those of the form $(k, v_{2j-1}, v_{2j}, k, w_{2j-1}, w_{2j})$. As in the case of $\mathfrak{S}^{(1)}$, the periodic orbits

$$\mathbb{T}_1: \quad \beta_{v_j}(t) = b_1(t) \neq 0, \quad \beta_{w_j}(t) = 0 \quad \text{for all } j = 1, \dots, n$$

and

$$\mathbb{T}_2: \quad \beta_{w_j}(t) = b_2(t) \neq 0, \quad \beta_{v_j}(t) = 0 \quad \text{for all } j = 1, \dots, n$$

are linked by a heteroclinic connection.

We iterate this procedure constructing a *generation set* $\mathcal{S} = \bigcup_{i=1}^N \mathcal{S}_i$, where each \mathcal{S}_i is complete and action-preserving. The corresponding periodic orbit \mathbb{T}_i is linked by heteroclinic connections to \mathbb{T}_{i-1} and \mathbb{T}_{i+1} . There are two delicate points:

²The papers [Grébert and Thomann 2012; Haus and Thomann 2013] actually consider the one-dimensional case, but of course the construction of complete sets can always be trivially extended to higher dimensions.

- (i) At each step, when adding a new generation \mathcal{S}_i , we need to ensure that the resulting generation set is still complete and contains no nontrivial resonances except for those prescribed and those implied by the prescribed ones. The prescribed resonances are those of the form $(k, v_1, v_2, k, v_3, v_4)$, where $v_1, v_2 \in \mathcal{S}_i$ and $v_3, v_4 \in \mathcal{S}_{i+1}$ for some $1 \leq i \leq N - 1$ and $\{v_1, v_2, v_3, v_4\}$ are the vertices of a rectangle.
- (ii) We need to ensure that the Sobolev norms grow by an arbitrarily large factor K/δ , which requires taking n (the number of elements in each \mathcal{S}_j) and N (the number of generations) to be large.

The point (i) is a question of combinatorics. It requires some careful classification of the possible resonances and it turns out to be significantly more complicated than in the cubic case. We discuss this in Section 3B.

The point (ii) is treated exactly in the same way as in [CKSTT]; we discuss it for completeness in Section 3A, Remark 3.2.

Given a generation set \mathcal{S} as above we proceed in the following way: First we restrict to the finite-dimensional invariant subspace where $\beta_k = 0$ for all $k \notin \mathcal{S}$. To further simplify the dynamics, we restrict to the invariant subspace

$$\beta_v(t) = b_i(t) \quad \text{for all } v \in \mathcal{S}_i, \quad i = 1, \dots, n;$$

this is the so called *toy model*. Note that the periodic solutions \mathbb{T}_i live in this subspace. The toy model is a Hamiltonian system, with Hamiltonian given by (2-3) and with the constant of motion $J = \sum_{i=1}^N |b_i|^2$. We work on the sphere $J = 1$, which contains all the \mathbb{T}_i with action $|b_i|^2 = 1$.

As discussed above, we construct a chain of heteroclinic connections going from \mathbb{T}_1 to \mathbb{T}_N . Then, we prove (see Proposition 2.10) the existence of a *slider solution* which “shadows” this chain, starting at time 0 from a neighborhood of \mathbb{T}_3 and ending at time T in a neighborhood of \mathbb{T}_{N-2} .³

We proceed as follows: First, we perform a symplectic reduction that will allow us to study the local dynamics close to the periodic orbit \mathbb{T}_j , which puts the Hamiltonian in the form (2-6). The new variables c_k are the ones obtained by synchronizing the b_k ($k \neq j$) with the phase of b_j . Then, we diagonalize the linear part of the vector field associated to (2-6). In particular, the eigenvalues are the Lyapunov exponents of the periodic orbit \mathbb{T}_j . As for the cubic case, one obtains that all the eigenvalues are purely imaginary, except for four of them which, due to the symmetries of the problem, are of the form $\lambda, \lambda, -\lambda, -\lambda \in \mathbb{R}$. Note that these hyperbolic directions are directly related to the heteroclinic connections connecting \mathbb{T}_j to \mathbb{T}_{j-1} and to \mathbb{T}_{j+1} . It turns out that the heteroclinic connections are straight lines in the variables c_k . The equations of motion for the reduced system have the form (2-10) (which is very similar to the cubic case); this is crucial in order to be able to apply almost verbatim the proof given in [CKSTT]. Note that it is not obvious a priori that the equations (2-10) hold true: for instance, this turns out to be false for the NLS of degree 7 and above.

³One could ask why we construct a slider solution diffusing from the third mode b_3 to the third-to-last mode b_{N-2} , instead of diffusing from the first mode b_1 to the last mode b_N . The reason is that, since we rely on the proof given in [CKSTT], our statement is identical to their Proposition 2.2 and Theorem 3.1. As there, also in our case, it would be possible to diffuse from the first to the last mode just by overcoming some very small notational issues.

The strategy of the proof, which is exactly the same as in [CKSTT], consists substantially of two parts:

- Studying the linear dynamics close to \mathbb{T}_j , treating the nonlinear terms as a small perturbation; one needs to prove that the flow associated to equations (2-10) maps points close to the *incoming* heteroclinic connection (from \mathbb{T}_{j-1}) to points close to the *outgoing* heteroclinic connection (towards \mathbb{T}_{j+1}) (note that, in order to take advantage of the linear dynamics close to \mathbb{T}_j , we need that almost all the energy is concentrated on S_j).
- Following closely the heteroclinic connection in order to flow from a neighborhood of \mathbb{T}_j to a neighborhood of \mathbb{T}_{j+1} .

The precise statement of these two facts requires the introduction of the notions of *targets* and *covering* and is summarized in Proposition 2.13. The main analytical tool for the proof are repeated applications of Gronwall's lemma. Our proof of Proposition 2.13 follows almost verbatim the proof of the analogous statement, given in Section 3 of [CKSTT]. However, the only way to check that the proof works also in our case is to go through the whole proof in [CKSTT], which is rather long and technical, and make the needed adaptations. Therefore, for the convenience of the reader, in Appendix A we give a summary of the proof of Proposition 2.13, highlighting the points where there are significant differences with [CKSTT].

1C. Comparison with the cubic case and higher-order NLS equations. In the cubic NLS, the only resonant sets of frequencies are rectangles, which makes the choice of using rectangles as building blocks of the generation set \mathcal{S} completely natural. In the quintic and higher-degree NLS many more resonant sets appear, which a priori gives much more freedom in the construction of \mathcal{S} . In particular, in the quintic case, sets of the form $\mathfrak{S}^{(2)}$ are the most generic resonant sets, and therefore it would look reasonable to use them as building blocks. However (see Appendix C), such a choice does not allow full energy transfer from a generation to the next one and is therefore incompatible with our strategy. The same happens if one uses sets of the form $\mathfrak{S}^{(3)}$. This leads us to use rectangles for the construction of \mathcal{S} also in the quintic case.

It is worth remarking that, while nondegenerate rectangles do not exist in one space dimension, sets of the form $\mathfrak{S}^{(2)}$, $\mathfrak{S}^{(3)}$ already exist in one dimension. The equations of the toy model only depend on the combinatorics of the set \mathcal{S} . Therefore, if one were able to prove diffusion in a toy model built with resonant sets of the form $\mathfrak{S}^{(2)}$, $\mathfrak{S}^{(3)}$ (or other resonant sets that exist already in one dimension), then one could hope to prove the same type of result for some one-dimensional (noncubic) NLS.

The use of rectangles as building blocks for the generation set of a quintic or higher-order NLS makes things more complicated, since the rectangles induce many different resonant sets; see Section 2. This leads to combinatorial problems that make it harder to prove the nondegeneracy and completeness of \mathcal{S} . The equations of the toy model also have a more complicated form than in the cubic case. Since these types of difficulties grow with the degree, dealing with the general case will most probably require some careful — and possibly complicated — combinatorics, and one cannot expect to have a completely explicit formula for the toy model Hamiltonian of any degree.

In the quintic case the formula is explicit and relatively simple, and we can explicitly perform the symmetry reduction. After some work, we still get equations of the form (2-10) that resemble the cubic case with some relevant differences: here the Lyapunov exponent λ depends on n and tends to infinity

as $n \rightarrow \infty$; moreover, the nonlinear part of the vector field associated to (2-6) is not homogeneous in the variables c_k , as it contains both terms of order 3 and 5 (in the cubic case, it is homogeneous of order 3).

For the NLS of higher degree, not only the reduced Hamiltonian gets essentially unmanageable, but there also appears a further difficulty. Already for the NLS of degree 7, a toy model built using rectangles (after symplectic reduction and diagonalization) does not satisfy equations like (2-10), meaning that the heteroclinic connections are not straight lines. Such a problem can be probably overcome, but this requires a significant adaptation of the analytical techniques used in order to prove the existence of the slider solution (work in progress with M. Guardia).

1D. Plan of the paper. In Section 2, we assume we have a generation set $\mathcal{S} = \bigcup_{i=1}^N \mathcal{S}_i$ which satisfies all the needed nondegeneracy properties and deduce the form of the toy model Hamiltonian. Then we study this Hamiltonian and prove the existence of slider solutions.

In Section 3 we prove the existence of nondegenerate generation sets such that the corresponding slider solution undergoes the required growth of Sobolev norms.

In Section 4 we prove, via the approximation Lemma 4.1 and a scaling argument, the persistence of solutions with growing Sobolev norm for the full NLS.

Since some of the proofs follow very closely the ones in [CKSTT], we move them to the appendix.

2. The toy model

We now define a finite subset $\mathcal{S} = \bigcup_{i=1}^N \mathcal{S}_i \subset \mathbb{Z}^2$ which satisfies appropriate *nondegeneracy conditions* (Definition 2.8) as explained in the introduction. In the following we assume that such a set exists. This is not obvious and will be discussed in Section 3B.

For reasons that will be clear, and following [CKSTT], the \mathcal{S}_i will be called *generations*. In order to describe the resonances which connect different generations, we introduce some notation.

Definition 2.1 (family). A family (of age $i \in \{1, \dots, N-1\}$) is a list $(v_1, v_2; v_3, v_4)$ of elements of \mathcal{S} such that the points form the vertices of a nondegenerate rectangle, meaning that

$$v_1 + v_2 = v_3 + v_4 \quad \text{and} \quad |v_1|^2 + |v_2|^2 = |v_3|^2 + |v_4|^2,$$

and such that one has $v_1, v_2 \in \mathcal{S}_i$ and $v_3, v_4 \in \mathcal{S}_{i+1}$. Whenever $(v_1, v_2; v_3, v_4)$ form a family, we say that v_1, v_2 are the parents of v_3, v_4 and that v_3, v_4 are the children of v_1, v_2 . Moreover, we say that v_1 is the spouse of v_2 (and vice versa) and that v_3 is the sibling of v_4 (and vice versa). We denote (for instance) $v_1 = v_3^{\text{par}_1}$, $v_2 = v_3^{\text{par}_2}$, $v_1 = v_2^{\text{sp}}$, $v_4 = v_3^{\text{sib}}$, $v_3 = v_1^{\text{ch}_1}$, $v_4 = v_1^{\text{ch}_2}$.

Remark 2.2. If $(v_1, v_2; v_3, v_4)$ is a family of age i , then the same holds for its trivial permutations $(v_2, v_1; v_3, v_4)$, $(v_1, v_2; v_4, v_3)$ and $(v_2, v_1; v_4, v_3)$.

Definition 2.3. An integer vector $\lambda \in \mathbb{Z}^{|\mathcal{S}|}$ such that

$$\sum_i \lambda_i = 0 \quad \text{and} \quad |\lambda| := \sum_i |\lambda_i| \leq 6$$

is resonant for \mathcal{S} if

$$\sum_i \lambda_i v_i = 0 \quad \text{and} \quad \sum_i \lambda_i |v_i|^2 = 0.$$

To a family $\mathcal{F} = (v_1, v_2; v_3, v_4)$ we associate a special resonant vector $\lambda^{\mathcal{F}}$ with $|\lambda| = 4$, through $\sum_i \lambda_i^{\mathcal{F}} v_i = v_1 + v_2 - v_3 - v_4$. Similarly, to the couple of parents in the family \mathcal{F} we associate the vector $\lambda^{\mathcal{F}_p}$ through $\sum_i \lambda_i^{\mathcal{F}_p} v_i = v_1 + v_2$ and to the couple of children we associate $\lambda^{\mathcal{F}_c}$ through $\sum_i \lambda_i^{\mathcal{F}_c} v_i = v_3 + v_4$, so that $\lambda^{\mathcal{F}} = \lambda^{\mathcal{F}_p} - \lambda^{\mathcal{F}_c}$.

Definition 2.4 (generation set). The set \mathcal{S} is said to be a generation set if it satisfies the following:

- (1) For all $i \in \{1, \dots, N - 1\}$, every $v \in \mathcal{S}_i$ is a member of one and only one (up to trivial permutations) family of age i . We denote such a family by \mathcal{F}^v . (Note that $\mathcal{F}^v = \mathcal{F}^w$ if $v = w^{\text{sp}}$.)
- (2) For all $i \in \{2, \dots, N\}$, every $v \in \mathcal{S}_i$ is a member of one and only one (up to trivial permutations) family of age $i - 1$. We denote such a family by \mathcal{F}_v . (Note that $\mathcal{F}_v = \mathcal{F}_w$ if $v = w^{\text{sib}}$.)
- (3) For all $v \in \bigcup_{i=2}^{N-1} \mathcal{S}_i$, one has $v^{\text{sp}} \neq v^{\text{sib}}$.

Remark 2.5. The vectors $\lambda^{\mathcal{F}}$ corresponding to the families of a generation set are linearly independent.

Whenever two families \mathcal{F}_1 and \mathcal{F}_2 have a common member (which must be a child in one family and a parent in the other one), $\lambda^{\mathcal{F}_1} + \lambda^{\mathcal{F}_2}$ is a nontrivial resonant vector whose support has cardinality exactly 6. This motivates the following definition:

Definition 2.6 (resonant vector of type CF). A resonant vector λ is said to be of type CF (couple of families) if there exist two families $\mathcal{F}_1 \neq \mathcal{F}_2$ such that $\lambda = \pm(\lambda^{\mathcal{F}_1} + \lambda^{\mathcal{F}_2})$. (Note that, since $|\lambda| \leq 6$, the two families \mathcal{F}_1 and \mathcal{F}_2 must have a common member.)

Definition 2.7. Given an ordering of \mathcal{S} , we have a one-to-one correspondence $e_i \leftrightarrow v_i$ between the canonical basis of $\mathbb{Z}^{|\mathcal{S}|}$ and the elements of \mathcal{S} .

We say that a generation set is nondegenerate if the following condition is fulfilled:

Definition 2.8 (nondegeneracy). Suppose that there exists $\lambda \in \mathbb{Z}^{|\mathcal{S}|}$, with $\sum_i \lambda_i = 1$ and $|\lambda| \leq 5$, such that

$$\sum_i \lambda_i |v_i|^2 - \left| \sum_i \lambda_i v_i \right|^2 = 0.$$

Then only four possibilities are allowed:

- (1) $|\lambda| = 1$.
- (2) $|\lambda| = 3$ and the support of λ consists of exactly three distinct elements of the same family, and the two λ_i appearing with a positive sign correspond either to the two parents or to the two children of the family.
- (3) $|\lambda| = 5$ and there exist a family \mathcal{F} and an element $v \in \mathcal{S}$ such that $\lambda = \pm\lambda^{\mathcal{F}} + e_i$. Here, e_i is the vector of the canonical basis in $\mathbb{Z}^{|\mathcal{S}|}$ associated to v by Definition 2.7.
- (4) $|\lambda| = 5$ and there exists $v \in \mathcal{S}$ (with $bf e_i \leftrightarrow v$) such that $\lambda - e_i$ is a resonant vector of type CF.

Note that, if \mathcal{S} is a nondegenerate generation set and λ is a resonant vector, then either $\lambda = \pm\lambda^{\mathcal{F}}$ for some family \mathcal{F} or λ is a resonant vector of type CF.

In what follows we will assume that \mathcal{S} is a nondegenerate generation set. This implies that \mathcal{S} is complete and all the subsets \mathcal{S}_i are pairwise disjoint, complete and action-preserving. Finally, the only resonances which appear are those induced by the family relations. Then, the Hamiltonian restricted to \mathcal{S} is

$$\begin{aligned} \mathcal{H}|_{\mathcal{S}} = & \frac{1}{3} \left(\sum_{j \in \mathcal{S}} |\beta_j|^6 + 9 \sum_{\substack{j,k \in \mathcal{S} \\ j \neq k}} |\beta_j|^4 |\beta_k|^2 + 36 \sum_{\substack{j,k,m \in \mathcal{S} \\ j < k < m}} |\beta_j|^2 |\beta_k|^2 |\beta_m|^2 \right) \\ & + 3 \sum_{i=1}^{N-1} \sum_{j \in \mathcal{S}_i} (\beta_j \beta_{j^{sp}} \bar{\beta}_{j^{ch_1}} \bar{\beta}_{j^{ch_2}} + \bar{\beta}_j \bar{\beta}_{j^{sp}} \beta_{j^{ch_1}} \beta_{j^{ch_2}}) \left(2 \sum_{\substack{k \in \mathcal{S} \\ k \notin \mathcal{F}^j}} |\beta_k|^2 + \sum_{m \in \mathcal{F}^j} |\beta_m|^2 \right) \\ & + 12 \sum_{i=2}^{N-1} \sum_{j \in \mathcal{S}_i} (\beta_j^{par_1} \beta_j^{par_2} \beta_{j^{sp}} \bar{\beta}_{j^{sib}} \bar{\beta}_{j^{ch_1}} \bar{\beta}_{j^{ch_2}} + \beta_{j^{ch_1}} \beta_{j^{ch_2}} \beta_{j^{sib}} \bar{\beta}_{j^{sp}} \bar{\beta}_{j^{par_1}} \bar{\beta}_{j^{par_2}}). \end{aligned} \quad (2-1)$$

We restrict to the invariant subspace $\mathcal{D} \subset \mathcal{S}$ where $\beta_k = b_i$ for all $k \in \mathcal{S}_i$ and $i = 1, \dots, N$. Denote by n (which must be an even integer) the cardinality of each generation. Following the construction in [CKSTT], one has $n = 2^{N-1}$. A straightforward computation (involving some easy combinatorics) of the Hamiltonian yields

$$\begin{aligned} \frac{3}{n} \mathcal{H}|_{\mathcal{D}} = & \sum_{k=1}^N |b_k|^6 + 9 \left[(n-1) \sum_{k=1}^N |b_k|^6 + n \sum_{\substack{k,\ell=1 \\ k \neq \ell}}^N |b_k|^4 |b_\ell|^2 \right] \\ & + 6 \left[(n-1)(n-2) \sum_{k=1}^N |b_k|^6 + 3n(n-1) \sum_{\substack{k,\ell=1 \\ k \neq \ell}}^N |b_k|^4 |b_\ell|^2 \right] + 36n^2 \sum_{\substack{k,\ell,m=1 \\ k < \ell < m}}^N |b_k|^2 |b_\ell|^2 |b_m|^2 \\ & + 18 \sum_{k=1}^{N-1} \left(-|b_k|^2 - |b_{k+1}|^2 + n \sum_{\ell=1}^N |b_\ell|^2 \right) (b_k^2 \bar{b}_{k+1}^2 + b_{k+1}^2 \bar{b}_k^2) \\ & + 36 \sum_{k=2}^{N-1} |b_k|^2 (b_{k-1}^2 \bar{b}_{k+1}^2 + b_{k+1}^2 \bar{b}_{k-1}^2). \end{aligned} \quad (2-2)$$

The equations of motion for the toy model can be deduced by considering the effective Hamiltonian $h(b, \bar{b}) := \mathcal{H}|_{\mathcal{D}}(b, \bar{b})/n$, endowed with the symplectic form $\Omega = i db \wedge d\bar{b}$.

Due to the conservation of the total mass L , the quantity

$$J := \frac{L}{n} = \sum_{k=1}^N |b_k|^2$$

is a constant of motion.

We use this conservation law in order to remove from the right-hand side of (2-2) the terms depending on n^2 . We compute the quantity $3h - 6n^2 J^3$: Up to a global phase shift, the subtraction of the constant term $6n^2 J^3$ can be ignored, so with an abuse of notation we keep denoting it by $3h$. We get

$$3h = 4 \sum_{k=1}^N |b_k|^6 - 9n \sum_{h=1}^N |b_h|^2 \left[\sum_{k=1}^N |b_k|^4 - 2 \sum_{k=1}^{N-1} (b_k^2 \bar{b}_{k+1}^2 + b_{k+1}^2 \bar{b}_k^2) \right] + 18 \sum_{k=1}^{N-1} (-|b_k|^2 - |b_{k+1}|^2)(b_k^2 \bar{b}_{k+1}^2 + b_{k+1}^2 \bar{b}_k^2) + 36 \sum_{k=2}^{N-1} |b_k|^2 (b_{k-1}^2 \bar{b}_{k+1}^2 + b_{k+1}^2 \bar{b}_{k-1}^2). \quad (2-3)$$

2A. Invariant subspaces. Since J is a constant of motion, the dynamics is confined to its level sets. For simplicity, we will restrict to $J = 1$, that is, to

$$\Sigma := \left\{ b \in \mathbb{C}^N : \sum_{k=1}^N |b_k|^2 = 1 \right\}.$$

All the monomials in the toy model Hamiltonian have even degree in each of the modes (b_j, \bar{b}_j) , which implies that

$$\text{Supp}(b) := \{1 \leq j \leq N \mid b_j \neq 0\}$$

is invariant in time. This automatically produces many invariant subspaces, some of which will play a specially important role, namely:

(i) The subspaces M_j corresponding to $\text{Supp}(b) = \{j\}$ for some $1 \leq j \leq N$. In this case the dynamics is confined to the circle $|b_j|^2 = J$, with

$$b_j(t) = \sqrt{J} \exp[-i(3n - \frac{4}{3})J^2 t]. \quad (2-4)$$

The intersection of M_j with Σ is a single periodic orbit, which we denote by \mathbb{T}_j .

(ii) The subspaces generated by M_j and M_{j+1} (corresponding to $\text{Supp}(b) = \{j, j + 1\}$) for some $1 \leq j \leq N - 1$. Here, the Hamiltonian becomes

$$3h_{2g} = 4(|b_j|^6 + |b_{j+1}|^6) - 9n(|b_j|^2 + |b_{j+1}|^2)[|b_j|^4 + |b_{j+1}|^4 - 2(b_j^2 \bar{b}_{j+1}^2 + b_{j+1}^2 \bar{b}_j^2)] - 18(|b_j|^2 + |b_{j+1}|^2)(b_j^2 \bar{b}_{j+1}^2 + b_{j+1}^2 \bar{b}_j^2). \quad (2-5)$$

Passing to symplectic polar coordinates

$$b_j = \sqrt{I_1} e^{i\theta_1}, \quad b_{j+1} = \sqrt{I_2} e^{i\theta_2},$$

we have

$$3h_{2g} = (4 - 9n)(I_1 + I_2)^3 + 6(I_1 + I_2)I_1 I_2 (3n - 2 + 6(n - 1) \cos(2(\theta_1 - \theta_2)));$$

since $J = I_1 + I_2$ is a conserved quantity, the dynamics is integrable and easy to study.

We pass to the symplectic variables

$$J, I_1, \theta_2 \quad \text{and} \quad \varphi = \theta_2 - \theta_1$$

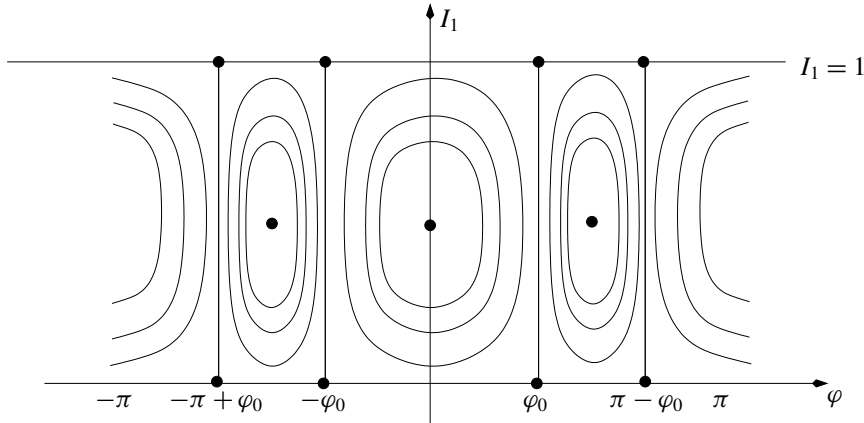


Figure 1. Phase portrait of the two-generation Hamiltonian h_{2g} on Σ .

and obtain the Hamiltonian

$$3h_{2g} = (4 - 9n)J^3 + 6JI_1(J - I_1)(3n - 2 + 6(n - 1) \cos(2\varphi)).$$

The phase portrait (ignoring the evolution of the cyclic variable θ_2) restricted to Σ is described in Figure 1.

Remark 2.9. The coordinates I_1, φ and the domain given by the cylinder $(\varphi, I_1) \in \mathbb{S}^1 \times [0, 1]$ are singular, since the angle $\varphi = \theta_2 - \theta_1$ is ill-defined when $I_1 = 0$ or $I_1 = 1$. In the correct picture for the reduced dynamics, each of the lines $I_1 = 0$ and $I_1 = 1$ should be shrunk to a single point, thus obtaining (topologically) a two-dimensional sphere (see Figure 2).

This can also be seen in the following way. The level set $J = 1$ is a three-dimensional sphere \mathbb{S}^3 , with the gauge symmetry group \mathbb{S}^1 acting freely on it. Due to the Hopf fibration, the topology of the quotient space is \mathbb{S}^2 .

As for the case of the cubic NLS (see [CKSTT] and [Guardia and Kaloshin 2015]), there exist heteroclinic connections linking \mathbb{T}_j to \mathbb{T}_{j+1} . Again as in the cubic case, the orbits have fixed angle

$$\varphi(t) = \varphi_0 = \frac{1}{2} \arccos\left(-\frac{3n - 2}{6(n - 1)}\right), \quad I_1(t) = \frac{e^{2\lambda t}}{1 + e^{2\lambda t}},$$

where $\lambda = 2\sqrt{(9n - 8)(3n - 4)}$. Our aim will be to construct slider solutions that are very concentrated on the mode b_3 at the time $t = 0$ and very concentrated on the mode b_{N-2} at the time $t = T$. These solutions will start very close to the periodic orbit \mathbb{T}_3 and then use the heteroclinic connections in order to slide from \mathbb{T}_3 to \mathbb{T}_4 and so on until \mathbb{T}_{N-2} .

2B. Symplectic reduction. Now, since we are interested in studying the dynamics close to the j -th periodic orbit \mathbb{T}_j , we introduce a set of coordinates that are in phase with it and give a symplectic reduction with respect to the constant of motion J . This procedure is the same as was carried out, for the cubic NLS, in [Guardia and Kaloshin 2015] and, substantially, already in [CKSTT].

Let $\vartheta^{(j)}$ be the phase of the complex number b_j . Then, for $k \neq j$, let $c_k^{(j)}$ the variable obtained by conjugating b_k with the phase $\vartheta^{(j)}$, i.e.,

$$c_k^{(j)} = b_k e^{-i\vartheta^{(j)}}.$$

Then, the change of coordinates (which is well-defined on $\{b_j \neq 0\}$) given by

$$(b_1, \dots, b_N, \bar{b}_1, \dots, \bar{b}_N) \mapsto (c_1^{(j)}, \dots, c_{j-1}^{(j)}, J, c_{j+1}^{(j)}, \dots, c_N^{(j)}, \bar{c}_1^{(j)}, \dots, \bar{c}_{j-1}^{(j)}, \vartheta^{(j)}, \bar{c}_{j+1}^{(j)}, \dots, \bar{c}_N^{(j)})$$

is symplectic. Namely, in the new coordinates the symplectic form is given by

$$\Omega = i dc^{(j)} \wedge d\bar{c}^{(j)} + dJ \wedge d\vartheta^{(j)}.$$

Then, we rewrite the Hamiltonian h in terms of the new coordinates (from now on, in order to simplify the notation, we will omit the superscript (j) in the $c^{(j)}$ variables, in their complex conjugates $\bar{c}^{(j)}$ and in the phase $\vartheta^{(j)}$). Thus, we get the expression

$$\begin{aligned} 3h = & 4 \sum_{k \neq j} |c_k|^6 - 4 \left(\sum_{k \neq j} |c_k|^2 \right)^3 + (18n-12)J^2 \sum_{k \neq j} |c_k|^2 - 9nJ \sum_{k \neq j} |c_k|^4 - (9n-12)J \left(\sum_{k \neq j} |c_k|^2 \right)^2 \\ & + 18 \sum_{\substack{k=1 \\ k \neq j-1, j}}^{N-1} (-|c_k|^2 - |c_{k+1}|^2 + nJ)(c_k^2 \bar{c}_{k+1}^2 + c_{k+1}^2 \bar{c}_k^2) \\ & + 18 \left[\sum_{\substack{k=1 \\ k \neq j-1, j}}^N |c_k|^2 + (n-1)J \right] \left(J - \sum_{\substack{\ell=1 \\ \ell \neq j}}^N |c_\ell|^2 \right) (c_{j-1}^2 + \bar{c}_{j-1}^2) \\ & + 18 \left[\sum_{\substack{k=1 \\ k \neq j, j+1}}^N |c_k|^2 + (n-1)J \right] \left(J - \sum_{\substack{\ell=1 \\ \ell \neq j}}^N |c_\ell|^2 \right) (c_{j+1}^2 + \bar{c}_{j+1}^2) \\ & + 36 \sum_{\substack{k=2 \\ k \neq j-1, j, j+1}}^{N-1} |c_k|^2 (c_{k-1}^2 \bar{c}_{k+1}^2 + c_{k+1}^2 \bar{c}_{k-1}^2) + 36 |c_{j-1}|^2 \left(J - \sum_{\substack{k=1 \\ k \neq j}}^N |c_k|^2 \right) (c_{j-2}^2 + \bar{c}_{j-2}^2) \\ & + 36 \left(J - \sum_{\substack{k=1 \\ k \neq j}}^N |c_k|^2 \right) (c_{j-1}^2 \bar{c}_{j+1}^2 + c_{j+1}^2 \bar{c}_{j-1}^2) \\ & + 36 |c_{j+1}|^2 \left(J - \sum_{\substack{k=1 \\ k \neq j}}^N |c_k|^2 \right) (c_{j+2}^2 + \bar{c}_{j+2}^2). \end{aligned} \tag{2-6}$$

Observe that the Hamiltonian h does not depend on ϑ . Since J is a constant of motion, the terms depending only on J can be erased from the Hamiltonian. Up to those constant terms, one has

$$h = h_2 + r_4, \tag{2-7}$$

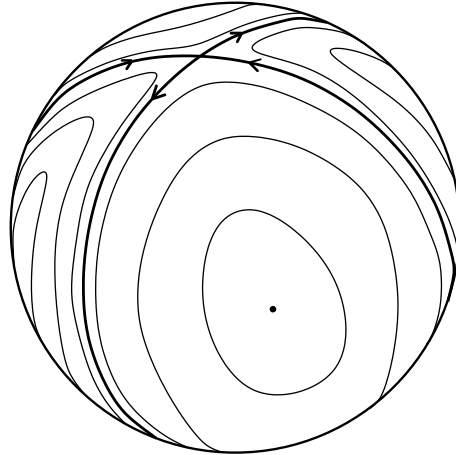


Figure 2. A sketch of the phase portrait of the two-generation Hamiltonian h_{2g} on Σ in the correct topology.

where h_2 is the part of order 2 in (c, \bar{c}) (which corresponds to the linear part of the vector field) and r_4 is of order at least 4 in (c, \bar{c}) . By an explicit computation, one obtains

$$h_2 = 2J^2 \left[(3n - 2) \sum_{\substack{k=1 \\ k \neq j}}^N |c_k|^2 + 3(n - 1)(c_{j-1}^2 + \bar{c}_{j-1}^2 + c_{j+1}^2 + \bar{c}_{j+1}^2) \right]. \tag{2-8}$$

It is easily seen that the dynamics associated to the vector field generated by h_2 is elliptic in the modes c_k with $1 \leq k \leq j - 2$ or $j + 2 \leq k \leq N$, while it is hyperbolic in the modes c_{j-1} and c_{j+1} . In order to put in evidence the hyperbolic dynamics, we perform a change of coordinates which diagonalizes the linear part of the vector field. Namely, for $k = j - 1, j + 1$, we set

$$c_k = \frac{1}{\sqrt{2\Im(\omega^2)}} (\bar{\omega}c_k^- + \omega c_k^+) \quad \text{and} \quad \bar{c}_k = \frac{1}{\sqrt{2\Im(\omega^2)}} (\omega c_k^- + \bar{\omega}c_k^+),$$

where $\omega = e^{i\varphi_0}$ with

$$\varphi_0 = \frac{1}{2} \arccos\left(-\frac{3n - 2}{6(n - 1)}\right).$$

Note that this change of variables affects only the hyperbolic modes, which are expressed in terms of the new variables $(c_{j-1}^+, c_{j-1}^-, c_{j+1}^+, c_{j+1}^-)$. This transformation is symplectic; writing h_2 as a function of the new variables, we get

$$h_2 = 2J^2 \left[(3n - 2) \sum_{\substack{k=1 \\ k \neq j-1, j, j+1}}^N |c_k|^2 + \sqrt{(9n - 8)(3n - 4)}(c_{j-1}^+ c_{j-1}^- + c_{j+1}^+ c_{j+1}^-) \right]. \tag{2-9}$$

We have proved that the periodic orbit (2-4) is hyperbolic and we have explicitly written the quadratic part of the Hamiltonian in the local variables. Similarly to the case of the cubic NLS, these local variables are

actually well adapted to describing also the global dynamics connecting two periodic orbits, as discussed in the previous section.

To this purpose, we study the integrable *two-generation* Hamiltonian (2-5) after all the changes of variables described in this section, i.e., in the variables c_{j+1}^+ and c_{j+1}^- . Direct substitution shows that the Hamiltonian is given by

$$h_{2g} = 2J\sqrt{(9n-8)(3n-4)}c_{j+1}^+c_{j+1}^- \left\{ J - \frac{1}{2\Im(\omega^2)} [(c_{j+1}^+)^2 + (c_{j+1}^-)^2 + 2\Re(\omega^2)c_{j+1}^+c_{j+1}^-] \right\}.$$

It is important to note that all the monomials in h_{2g} contain both c_{j+1}^+ and c_{j+1}^- , so the subspaces $c_{j+1}^+ = 0$ and $c_{j+1}^- = 0$ (which correspond to the heteroclinic connections) are invariant for the 2-generation dynamics. It is useful to let $c^* = \{c_h\}_{h \neq j-1, j, j+1}$, so that the dynamical variables of the Hamiltonian (2-6) become $(c_{j-1}^+, c_{j-1}^-, c_{j+1}^+, c_{j+1}^-, c^*, \bar{c}^*)$.

Now, since

$$h_{2g} = h|_{c_{j-1}^+ = c_{j-1}^- = q_1 = 0, c^* = 0},$$

exploiting also the symmetry between (c_{j-1}^+, c_{j-1}^-) and (c_{j+1}^+, c_{j+1}^-) , this implies that, in h , none of the monomials in $(c_{j-1}^+, c_{j-1}^-, c_{j+1}^+, c_{j+1}^-, c^*, \bar{c}^*)$ depends only on one of the variables $c_{j-1}^+, c_{j-1}^-, c_{j+1}^+, c_{j+1}^-$.

Finally, we recall that all the monomials in $h(c_{j-1}^+, c_{j-1}^-, c_{j+1}^+, c_{j+1}^-, c^*)$ have even degree in each of the couples (c_k^*, \bar{c}_k^*) and in both couples (c_k^+, c_k^-) .

From these observations, and from the bound $O(c^2) \lesssim J = O(1)$, we immediately deduce the following relations about the Hamilton equations associated to h :

$$\begin{aligned} \dot{c}_{j-1}^- &= -2J^2\sqrt{(9n-8)(3n-4)}c_{j-1}^- + O(c^2c_{j-1}^-) + O(c_{\neq j-1}^+c_{j-1}^-), \\ \dot{c}_{j-1}^+ &= 2J^2\sqrt{(9n-8)(3n-4)}c_{j-1}^+ + O(c^2c_{j-1}^+) + O(c_{\neq j-1}^-c_{j-1}^+), \\ \dot{c}_{j+1}^- &= -2J^2\sqrt{(9n-8)(3n-4)}c_{j+1}^- + O(c^2c_{j+1}^-) + O(c_{\neq j+1}^+c_{j+1}^-), \\ \dot{c}_{j+1}^+ &= 2J^2\sqrt{(9n-8)(3n-4)}c_{j+1}^+ + O(c^2c_{j+1}^+) + O(c_{\neq j+1}^-c_{j+1}^+), \\ \dot{c}^* &= 2J^2(3n+2)ic^* + O(c^2c^*), \end{aligned} \tag{2-10}$$

where we denote $c = (c_{j-1}^+, c_{j-1}^-, c_{j+1}^+, c_{j+1}^-, c^*)$, $c_{\neq j-1} = (c_{j+1}^+, c_{j+1}^-, c^*)$, $c_{\neq j+1} = (c_{j-1}^+, c_{j-1}^-, c^*)$. These relations are the precise analogue of [CKSTT, Proposition 3.1]; the factor $2J^2\sqrt{(9n-8)(3n-4)}$ here replaces the factor $\sqrt{3}$ in [CKSTT].

From the equations of motion (2-10), we deduce that

$$-i\dot{c}_{j+1} = \frac{\partial h_{2g}}{\partial \bar{c}_{j+1}} + O(c_{j+1}c_{\neq j+1}^2).$$

We have

$$h_{2g} = 2J\sqrt{(9n-8)(3n-4)}c_{j+1}^+c_{j+1}^-(J - |c_{j+1}|^2),$$

where c_{j+1}^+ and c_{j+1}^- can be thought of as functions of (c_{j+1}, \bar{c}_{j+1}) . Then

$$\dot{c}_{j+1} = 2iJ\sqrt{(9n-8)(3n-4)}(J - |c_{j+1}|^2)\frac{\partial(c_{j+1}^+c_{j+1}^-)}{\partial\bar{c}_{j+1}} + O(c_{j+1}c_{j+1}^+c_{j+1}^-) + O(c_{j+1}c_{\neq j+1}^2).$$

We compute

$$2i\frac{\partial(c_{j+1}^+c_{j+1}^-)}{\partial\bar{c}_{j+1}} = \sqrt{\frac{2}{\Im(\omega^2)}}(\omega c_{j+1}^+ - \bar{\omega}c_{j+1}^-),$$

from which we deduce

$$\dot{c}_{j+1} = J\sqrt{\frac{2(9n-8)(3n-4)}{\Im(\omega^2)}}(\omega c_{j+1}^+ - \bar{\omega}c_{j+1}^-)(J - |c_{j+1}|^2) + O(c_{j+1}c_{j+1}^+c_{j+1}^-) + O(c_{j+1}c_{\neq j+1}^2), \tag{2-11}$$

which is the analogue for c_{j+1} of equation (3.19) in [CKSTT]. In the same way, one deduces

$$\dot{c}_{j-1} = J\sqrt{\frac{2(9n-8)(3n-4)}{\Im(\omega^2)}}(\omega c_{j-1}^+ - \bar{\omega}c_{j-1}^-)(J - |c_{j-1}|^2) + O(c_{j-1}c_{j-1}^+c_{j-1}^-) + O(c_{j-1}c_{\neq j-1}^2), \tag{2-12}$$

which is the analogue of equation (3.19) in [CKSTT] for the evolution of c_{j-1} .

2C. Existence of a “slider solution”. In this section, we are going to prove the following proposition (which is the analogue of Proposition 2.2 in [CKSTT]), which establishes the existence of a slider solution.

Proposition 2.10. *For all $\epsilon > 0$ and $N \geq 6$, there exist a time $T_0 > 0$ and an orbit of the toy model such that*

$$\begin{aligned} |b_3(0)| &\geq 1 - \epsilon, & |b_j(0)| &\leq \epsilon, & j &\neq 3, \\ |b_{N-2}(T_0)| &\geq 1 - \epsilon, & |b_j(T_0)| &\leq \epsilon, & j &\neq N - 2. \end{aligned}$$

Furthermore, one has $\|b(t)\|_{\ell^\infty} \sim 1$ for all $t \in [0, T_0]$.

More precisely, there exists a point x_3 within $O(\epsilon)$ of \mathbb{T}_3 (using the usual metric on Σ), a point x_{N-2} within $O(\epsilon)$ of \mathbb{T}_{N-2} and a time $T_0 \geq 0$ such that $S(T_0)x_3 = x_{N-2}$, where $S(t)x$ is the dynamics at time t of the toy model Hamiltonian with initial datum x .

In order to prove Proposition 2.10, we completely rely on the proof of the analogous Proposition 2.2 in [CKSTT]. In order to keep our notations as close as possible to those of [CKSTT], we rescale the time $t = 2\sqrt{(9n-8)(n-4/3)}\tau$ in our toy model; this means rescaling h to $\sqrt{3}h/2\sqrt{(9n-8)(3n-4)}$, where h is defined in (2-3), so that the Lyapunov exponents of the linear dynamics are $\sqrt{3}$. We hence prove Proposition 2.10 for the rescaled toy model. By formulae (2-10), (2-11), (2-12), we have the analogue of Proposition 3.1 and of equation (3.19) of [CKSTT].

Proposition 2.11. *Let $3 \leq j \leq N - 2$ and let $b(\tau)$ be a solution of the rescaled toy model living on Σ and with $b_j(\tau) \neq 0$. We have the system of equations*

$$\dot{c}_{j-1}^- = -\sqrt{3}c_{j-1}^- + O(c^2c_{j-1}^-) + O(c_{\neq j-1}^2c_{j-1}^+), \tag{2-13a}$$

$$\dot{c}_{j-1}^+ = \sqrt{3}c_{j-1}^+ + O(c^2c_{j-1}^+) + O(c_{\neq j-1}^2c_{j-1}^-), \tag{2-13b}$$

$$\dot{c}_{j+1}^- = -\sqrt{3}c_{j+1}^- + O(c^2c_{j+1}^-) + O(c_{\neq j+1}^2c_{j+1}^+), \tag{2-13c}$$

$$\dot{c}_{j+1}^+ = \sqrt{3}c_{j+1}^+ + O(c^2c_{j+1}^+) + O(c_{\neq j+1}^2c_{j+1}^-), \tag{2-13d}$$

$$\dot{c}^* = i\kappa c^* + O(c^2c^*), \quad \kappa = \frac{\sqrt{3}(3n-2)}{\sqrt{(9n-8)(3n-4)}}. \tag{2-13e}$$

Moreover,

$$\dot{c}_{j+1} = \sqrt{\frac{3}{2\Im(\omega^2)}}(\omega c_{j+1}^+ - \bar{\omega}c_{j+1}^-)(J - |c_{j+1}|^2) + O(c_{j+1}c_{j+1}^+c_{j+1}^-) + O(c_{j+1}c_{\neq j+1}^2) \tag{2-14}$$

and

$$\dot{c}_{j-1} = \sqrt{\frac{3}{2\Im(\omega^2)}}(\omega c_{j-1}^+ - \bar{\omega}c_{j-1}^-)(J - |c_{j-1}|^2) + O(c_{j-1}c_{j-1}^+c_{j-1}^-) + O(c_{j-1}c_{\neq j-1}^2). \tag{2-15}$$

Finally, since the equations (2-13) come from the Hamiltonian (2-6), which is an even polynomial of degree six, one has that all the symbols $O(c^3)$ are actually $\mathcal{O}(c^3) + \mathcal{O}(c^5)$.⁴ For instance,

$$O(c^2c_{j-1}^-) = \mathcal{O}(c^2c_{j-1}^-) + \mathcal{O}(c^4c_{j-1}^-), \quad O(c_{\neq j-1}^2c_{j-1}^+) = \mathcal{O}(c_{\neq j-1}^2c_{j-1}^+) + \mathcal{O}(c^2c_{\neq j-1}^2c_{j-1}^+). \tag{2-16}$$

The only difference with [CKSTT] is that our remainder terms (of type $O(c^2c_{j-1}^-)$, $O(c_{\neq j-1}^2c_{j-1}^+)$, etc.) are not homogeneous of degree three but have also a term of degree five (which is completely irrelevant in the analysis).

We now introduce some definitions and notations of [CKSTT].

Definition 2.12 (targets). A target is a triple (M, d, R) , where M is a subset of Σ , d is a semimetric on Σ and $R > 0$ is a radius. We say that a point $x \in \Sigma$ is within a target (M, d, R) if we have $d(x, y) < R$ for some $y \in M$. Given two points $x, y \in \Sigma$, we say that x hits y , and write $x \mapsto y$, if we have $y = S(t)x$ for some $t \geq 0$. Given an initial target (M_1, d_1, R_1) and a final target (M_2, d_2, R_2) , we say that (M_1, d_1, R_1) can cover (M_2, d_2, R_2) , and write $(M_1, d_1, R_1) \rightarrow (M_2, d_2, R_2)$, if for every $x_2 \in M_2$ there exists an $x_1 \in M_1$ such that, for any point $y_1 \in \Sigma$ with $d(x_1, y_1) < R_1$, there exists a point $y_2 \in \Sigma$ with $d_2(x_2, y_2) < R_2$ such that y_1 hits y_2 .

We refer the reader to pp. 64–66 of [CKSTT] for a presentation of the main properties of targets.

We need a number of parameters: First, an increasing set of exponents

$$1 \ll A_3^0 \ll A_3^+ \ll A_4^- \ll \dots \ll A_{N-2}^- \ll A_{N-2}^0,$$

⁴As in [CKSTT], we use the *schematic notation* $\mathcal{O}(\cdot)$. The symbol $\mathcal{O}(y)$ indicates a linear combination of terms that resemble y up to the presence of multiplicative constants and complex conjugations. So, for instance, a term like $2i\bar{c}_{j+1}|c_{j+2}|^2c_{j+3}^2 - 3c_{j+1}|c_{j+2}|^4$ is of the form $\mathcal{O}(c^5)$ and, more precisely, $\mathcal{O}(c_{j+1}c_{\neq j+1}^4)$

which, for sake of concreteness, we will take to be consecutive powers of 10. Next, we shall need a small parameter $0 < \sigma \ll 1$ depending on N and the exponents A which basically measures the distance to \mathbb{T}_j at which the quadratic Hamiltonian dominates the quartic terms. Then, we need a set of scale parameters

$$1 \ll r_{N-2}^0 \ll r_{N-2}^- \ll r_{N-2}^+ \ll r_{N-3}^- \ll \cdots \ll r_3^+ \ll r_3^0,$$

where each parameter is assumed to be sufficiently large, depending on the preceding parameters and on σ and the A 's. These parameters represent a certain shrinking of each target from the previous one (in order to guarantee that each target can be covered by the previous). Finally, we need a very large time parameter $T \gg 1$ that we shall assume to be as large as necessary depending on all the previous parameters.

Setting

$$\{c_1, \dots, c_h\} := c_{\leq h}, \quad \{c_h, c_{h+1}, \dots, c_N\} = c_{\geq h},$$

we call $c_{\leq j-1}$ the *trailing modes*, $c_{\geq j+1}$ the *leading modes*, $c_{\leq j-2}$ the *trailing peripheral modes*, and finally $c_{\geq j+2}$ the *leading peripheral modes*. We construct a series of targets:

- An incoming target (M_j^-, d_j^-, R_j^-) (located near the stable manifold of \mathbb{T}_j) defined as follows: M_j^- is the subset of Σ where

$$c_{\leq j-2}, c_{j-1}^+ = 0, \quad c_{j-1}^- = \sigma, \quad |c_{\geq j+1}| \leq r_j^- e^{-2\sqrt{3}T},$$

$R_j^- = T^{A_j^-}$ and the semimetric is

$$d_j^-(x, \tilde{x}) := e^{2\sqrt{3}T} |c_{\leq j-2} - \tilde{c}_{\leq j-2}| + e^{\sqrt{3}T} |c_{j-1}^- - \tilde{c}_{j-1}^-| + e^{4\sqrt{3}T} |c_{j-1}^+ + \tilde{c}_{j-1}^+| + e^{3\sqrt{3}T} |c_{\geq j+1} - \tilde{c}_{\geq j+1}|$$

- A ricochet target (M_j^0, d_j^0, R_j^0) (located very near \mathbb{T}_j itself), defined as follows: M_j^0 is the subset of Σ where

$$c_{\leq j-1}, c_{j+1}^- = 0, \quad |c_{j+1}^+| \leq r_j^0 e^{-\sqrt{3}T}, \quad |c_{\geq j+2}| \leq r_j^0 e^{-2\sqrt{3}T},$$

$R_j^0 = T^{A_j^0}$ and the semimetric is

$$d_j^0(x, \tilde{x}) := e^{2\sqrt{3}T} (|c_{\leq j-2} - \tilde{c}_{\leq j-2}| + |c_{j+1}^+ + \tilde{c}_{j+1}^+|) + e^{\sqrt{3}T} |c_{j-1}^- - \tilde{c}_{j-1}^-| \\ + e^{3\sqrt{3}T} (|c_{j-1}^+ + \tilde{c}_{j-1}^+| + |c_{j+1}^- + \tilde{c}_{j+1}^-| + |c_{\geq j+2} - \tilde{c}_{\geq j+2}|)$$

- An outgoing target (M_j^+, d_j^+, R_j^+) (located near the unstable manifold of \mathbb{T}_j) defined as follows: M_j^+ is the subset of Σ where

$$c_{\leq j-1}, c_{j+1}^- = 0, \quad c_{j+1}^+ = \sigma, \quad |c_{\geq j+2}| \leq r_j^+ e^{-2\sqrt{3}T},$$

$R_j^+ = T^{A_j^+}$ and the semimetric is

$$d_j^+(x, \tilde{x}) := e^{2\sqrt{3}T} |c_{\leq j-1} - \tilde{c}_{\leq j-1}| + e^{4\sqrt{3}T} |c_{j+1}^- - \tilde{c}_{j+1}^-| + e^{\sqrt{3}T} |c_{j+1}^+ + \tilde{c}_{j+1}^+| + e^{3\sqrt{3}T} |c_{\geq j+2} - \tilde{c}_{\geq j+2}|.$$

By Section 3.5 of [CKSTT], Proposition 2.10 follows from:

Proposition 2.13. $(M_j^-, d_j^-, R_j^-) \rightarrow (M_j^0, d_j^0, R_j^0)$ for all $3 < j \leq N - 2$,
 $(M_j^0, d_j^0, R_j^0) \rightarrow (M_j^+, d_j^+, R_j^+)$ for all $3 \leq j < N - 2$,
 $(M_j^+, d_j^+, R_j^+) \rightarrow (M_{j+1}^-, d_{j+1}^-, R_{j+1}^-)$ for all $3 \leq j < N - 2$.

Proof. See Appendix A. □

Proof of Proposition 2.10. By [CKSTT, Lemma 3.1] we deduce the covering relations

$$(M_3^0, d_3^0, R_3^0) \rightarrow (M_{N-2}^0, d_{N-2}^0, R_{N-2}^0); \quad (2-17)$$

in turn this implies that there is at least one solution $b(t)$ which starts within the ricochet target (M_3^0, d_3^0, R_3^0) at some time t_0 and ends up within the ricochet target $(M_{N-2}^0, d_{N-2}^0, R_{N-2}^0)$ at some later time $t_1 > t_0$. But, from the definition of these targets, we thus see that $b(t_0)$ lies within a distance $O(r_3^0 e^{-\sqrt{3}T})$ of \mathbb{T}_3 , while $b(t_1)$ lies within a distance $O(r_{N-2}^0 e^{-\sqrt{3}T})$ of \mathbb{T}_{N-2} . The claim follows. □

3. Construction of the set \mathcal{S}

3A. The density argument and the norm explosion property. The perturbative argument for the construction of the frequency set \mathcal{S} works exactly as in [CKSTT, Section 4]. However, for the convenience of the reader, we recall here the main points.

A convenient way to construct a generation set is to first fix a “genealogical tree”, i.e., an abstract combinatorial model of the parenthood and brotherhood relations, and then to choose a placement function, embedding this abstract combinatorial model in \mathbb{R}^2 . Our choice of the abstract combinatorial model is the one described in [CKSTT, pp. 99–100]. Then, once the combinatorial model is fixed, the choice of the embedding in \mathbb{R}^2 is equivalent to the choice of the following free parameters:

- the placement of the first generation \mathcal{S}_1 (which implies the choice of a parameter in \mathbb{R}^{2^N});
- the choice of a procreation angle $\vartheta^{\mathcal{F}}$ for each family of the generation set (which globally implies the choice of a parameter in $\mathbb{T}^{(N-1)2^{N-2}}$, since $(N-1)2^{N-2}$ is the number of families).

We let $\mathcal{S}(\mathcal{S}_1, \vartheta^{\mathcal{F}})$ be the corresponding generation set and write $\mathcal{X} := \mathbb{R}^{2^N} \times \mathbb{T}^{(N-1)2^{N-2}}$ for the space of parameters.

The set of parameters producing *degenerate* generation sets is *small*; more precisely, we have the following:

Proposition 3.1. *There exists a closed set of zero measure $\mathcal{D} \subset \mathcal{X}$ such that the generation set $\mathcal{S}(\mathcal{S}_1, \vartheta^{\mathcal{F}})$ is nondegenerate for all $(\mathcal{S}_1, \vartheta^{\mathcal{F}}) \in \mathcal{X} \setminus \mathcal{D}$.*

For the proof, see Section 3B.

We claim that the set of $(\mathcal{S}_1, \vartheta^{\mathcal{F}}) \in \mathcal{X}$ such that $\mathcal{S}(\mathcal{S}_1, \vartheta^{\mathcal{F}}) \subset \mathbb{Q}^2 \setminus \{0\}$ is dense in \mathcal{X} . This is a consequence of two facts:

- the density of $\mathbb{Q}^2 \setminus \{0\}$ in \mathbb{R}^2 (for the placement of the first generation);
- the density of (nonzero) rational points on circles having a diameter with rational endpoints.

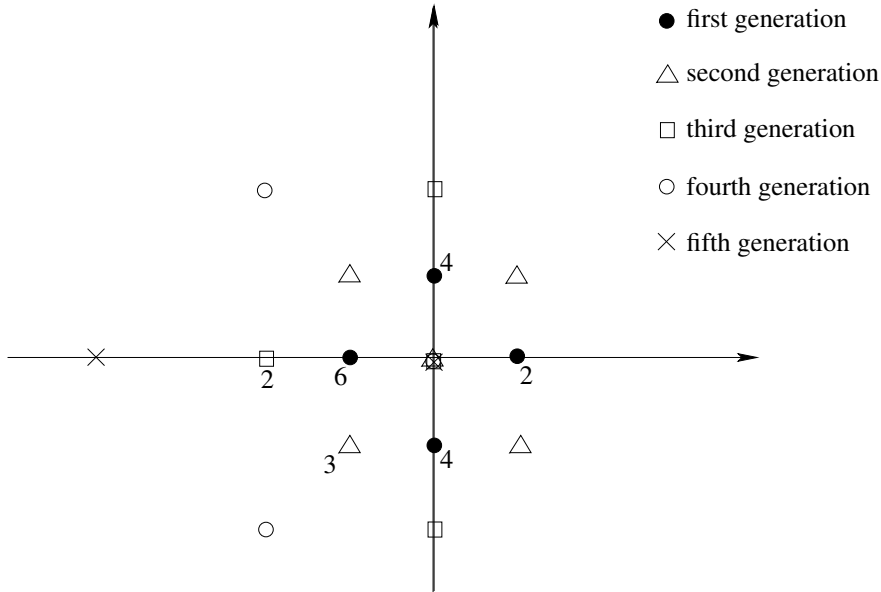


Figure 3. The prototype embedding with five generations. Note that this is a highly degenerate realization of the abstract combinatorial model of [CKSTT]. Since $N = 5$, each generation contains 16 points; we have explicitly written the multiplicity of each point when it is not one. In zero there are: 0 points of the first generation, 8 points of the second, 12 of the third, 14 of the fourth and 15 points of the fifth generation.

These two points imply that the set of $(S_1, \vartheta^{\mathcal{F}}) \in \mathcal{X}$ such that $\mathcal{S}(S_1, \vartheta^{\mathcal{F}})$ is nondegenerate and $\mathcal{S}(S_1, \vartheta^{\mathcal{F}}) \subset \mathbb{Q}^2 \setminus \{0\}$ is dense in \mathcal{X} .

In order to prove the growth of Sobolev norms, we require a further property on the generation set \mathcal{S} , the *norm explosion property*

$$\sum_{k \in \mathcal{S}_{N-2}} |k|^{2s} > \frac{1}{2} 2^{(s-1)(N-5)} \sum_{k \in \mathcal{S}_3} |k|^{2s}. \tag{3-1}$$

Given $N \gg 1$, our aim is to prove the existence of a nondegenerate generation set $\mathcal{S} \subset \mathbb{Q}^2 \setminus \{0\}$ satisfying (3-1). The fact that (3-1) is an open condition on the space of parameters \mathcal{X} , together with the above remarks, implies that it is enough to prove the existence of a (possibly degenerate) generation set $\mathcal{S} \subset \mathbb{R}^2$ satisfying (3-1), which is achieved by the *prototype embedding* described in [CKSTT, pp. 101–102] (see Figure 3).

For the reader’s convenience, we recall the construction of the prototype embedding. Let

$$S_1 = \{1, i\}, \quad S_2 = \{0, i + 1\};$$

then the 2^{N-1} elements of the k -th generation are identified with

$$(z_1, \dots, z_{k-1}, z_k, \dots, z_{N-1}) \in S_2^{k-1} \times S_1^{N-k} := \Sigma_k. \tag{3-2}$$

The union of the Σ_k is denoted by Σ . For all $1 \leq k \leq N-1$, a combinatorial nuclear family with parents in the k -th generation and children in the $(k+1)$ -st generation is a quadruple

$$(z_1, \dots, z_{k-2}, w, z_k, \dots, z_{N-1}), \quad w \in S_1 \cup S_2, \quad (3-3)$$

where all the z_j with $j \neq k-1$ are fixed, with $z_j \in S_2$ if $1 \leq j \leq k-2$ and $z_j \in S_1$ if $k \leq j \leq N-1$. Then, the prototype embedding $f : \Sigma \rightarrow \mathbb{C} \simeq \mathbb{R}^2$ is the one defined by

$$f(z_1, \dots, z_{N-1}) = \prod_{j=1}^{N-1} z_j. \quad (3-4)$$

Remark 3.2. For any given positive integer ℓ , the function $F : \mathbb{S}^{\ell-1} \rightarrow \mathbb{R}$, where

$$\mathbb{S}^{\ell-1} = \left\{ (x_1, \dots, x_\ell) \in \mathbb{R}^\ell \mid \sum_{i=1}^{\ell} x_i^2 = 1 \right\},$$

defined by

$$F(x_1, \dots, x_\ell) = \sum_{i=1}^{\ell} x_i^{2s}$$

attains its minimum (since $s > 1$) at

$$(x_1, \dots, x_\ell) = (\ell^{-1/2}, \dots, \ell^{-1/2})$$

and its maximum at

$$(x_1, x_2, \dots, x_\ell) = (1, 0, \dots, 0).$$

From this, one deduces that, for each family \mathcal{F} with parents v_1, v_2 and children v_3, v_4 , one must have

$$\frac{|v_3|^{2s} + |v_4|^{2s}}{|v_1|^{2s} + |v_2|^{2s}} \leq 2^{s-1}$$

and therefore, for all $1 \leq i \leq N-1$,

$$\frac{\sum_{k \in \mathcal{S}_{i+1}} |k|^{2s}}{\sum_{k \in \mathcal{S}_i} |k|^{2s}} \leq 2^{s-1},$$

which implies

$$\frac{\sum_{k \in \mathcal{S}_j} |k|^{2s}}{\sum_{k \in \mathcal{S}_i} |k|^{2s}} \leq 2^{(s-1)(j-i)}$$

for all $1 \leq i \leq j \leq N$. This means that we have to choose N large if we want the ratio

$$\frac{\sum_{k \in \mathcal{S}_{N-2}} |k|^{2s}}{\sum_{k \in \mathcal{S}_3} |k|^{2s}}$$

to be large.

Moreover, since

$$F(\ell^{-1/2}, \dots, \ell^{-1/2}) = \ell^{-s+1}, \quad F(1, 0, \dots, 0) = 1,$$

we have, for all $1 \leq i, j \leq N$,

$$\frac{\sum_{k \in \mathcal{S}_j} |k|^{2s}}{\sum_{k \in \mathcal{S}_i} |k|^{2s}} \leq n^{s-1},$$

which implies that n (the number of elements in each generation) also has to be chosen large enough.

In this sense, the prototype embedding and the choice $n = 2^{N-1}$ are optimal, because they attain the maximum possible growth of the quantity $\sum_{k \in \mathcal{S}_i} |k|^{2s}$ both at each step and between the first and the last generation.

Trivially there exists a one-to-one map from Σ to \mathcal{S} which preserves the *age*, then the same map identifies Σ with the basis vectors of $\mathbb{Z}^{|\mathcal{S}|}$. This defines the family relations of our generation set. Then, once we are given a nondegenerate generation set contained in $\mathbb{Q}^2 \setminus \{0\}$ and satisfying (3-1), it is enough to multiply by any integer multiple of the least common denominator of its elements in order to get a nondegenerate generation set $\mathcal{S} \in \mathbb{Z}^2 \setminus \{0\}$ and satisfying (3-1) (note that (3-1) is invariant by dilations of the set \mathcal{S}). Note that we can dilate \mathcal{S} as much as we wish, so we can make $\min_{k \in \mathcal{S}} |k|$ as large as desired.

These considerations are summarized by the following proposition (the analogue of Proposition 2.1 in [CKSTT]):

Proposition 3.3. *For all $K, \delta, \mathcal{R} > 0$, there exist $N \gg 1$ and a nondegenerate generation set $\mathcal{S} \subset \mathbb{Z}^2$ such that*

$$\frac{\sum_{k \in \mathcal{S}_{N-2}} |k|^{2s}}{\sum_{k \in \mathcal{S}_3} |k|^{2s}} \gtrsim \frac{K^2}{\delta^2} \tag{3-5}$$

and such that

$$\min_{k \in \mathcal{S}} |k| \geq \mathcal{R}. \tag{3-6}$$

3B. Proof of Proposition 3.1. The proof is composed of several steps. First, we need a lemma ensuring that any linear relation among the elements of the generation set that is not a linear combination of the family relations is generically not fulfilled.

Lemma 3.4. *Let $\mu \in \mathbb{Z}^{N2^{N-1}}$, $i = 1, \dots, M$ be an integer vector, linearly independent from the subspace of $\mathbb{R}^{N2^{N-1}}$ generated by all the vectors $\lambda^{\mathcal{F}}$ associated to the families. Then, for an open set of full measure $\mathfrak{S} \subset \mathcal{X}$, one has that, if $(\mathcal{S}_1, \vartheta^{\mathcal{F}}) \in \mathfrak{S}$, then $\mathcal{S}(\mathcal{S}_1, \vartheta^{\mathcal{F}})$ is such that*

$$\sum_{j=1}^{N2^{N-1}} \mu_j v_j \neq 0. \tag{3-7}$$

Proof. We denote the elements of \mathcal{S} by $v_1, \dots, v_{|\mathcal{S}|}$, with $|\mathcal{S}| = N2^{N-1}$. For simplicity and without loss of generality, we order the v_j so that couples of siblings always have consecutive subindices.

For each family \mathcal{F} , both the linear and the quadratic relations

$$\sum_j \lambda_j^{\mathcal{F}} v_j = 0 \quad \text{and} \quad \sum_j \lambda_j^{\mathcal{F}} |v_j|^2 = 0$$

are satisfied. The coefficients of the linear relations can be collected in a matrix $\Lambda^{\mathcal{F}}$ with $(N - 1)2^{N-2}$ rows (as many as the number of families) and $N2^{N-1}$ columns (as many as the elements of \mathcal{S}), so that the linear relations become

$$\Lambda^{\mathcal{F}}v = 0.$$

We choose to order the rows of $\Lambda^{\mathcal{F}}$ so that the matrix is in *lower row echelon* form (see figure).

$$\begin{array}{cccccccccccc} w_1 & w_2 & w_3 & w_4 & w_5 & p_5 & w_6 & p_6 & w_7 & p_7 & w_8 & p_8 \\ 1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & -1 & -1 \end{array}$$

Each row of a matrix in lower row echelon form has a *pivot*, the first nonzero coefficient of the row starting from the right. Being in lower row echelon form means that the pivot of a row is always strictly to the right of the pivot of the row above it. In the matrix $\Lambda^{\mathcal{F}}$, the pivots are all equal to -1 and they correspond to one and only one of the children from each family. In order to use this fact, we accordingly rename the elements of the generation set by writing $v = (p, w) \in \mathbb{R}^{2a} \times \mathbb{R}^{2b}$ with $a = (N - 1)2^{N-2}$, $b = N2^{N-1} - a = (N + 1)2^{N-2}$, where the $p_j \in \mathbb{R}^2$ are the elements of the generation set corresponding to the pivots and the $w_\ell \in \mathbb{R}^2$ are all the others, that is, all the elements of the first generation and one and only one child (the nonpivot) from each family. Here, the index ℓ ranges from 1 to b , while the index j ranges from $2^{N-1} + 1$ to b (note that $a + 2^{N-1} = b$), so that a couple (p_j, w_ℓ) corresponds to a couple of siblings if and only if $j = \ell$. Then, the linear relations $\Lambda^{\mathcal{F}}v = 0$ can be used to write each p_j as a linear combination of the w_ℓ with $\ell \leq j$ only:

$$p_j = \sum_{\ell \leq j} \eta_\ell w_\ell, \quad \eta_\ell \in \mathbb{Q}. \tag{3-8}$$

Finally, the quadratic relations $\Lambda^{\mathcal{F}}|v|^2 = 0$ constrain each w_ℓ with $\ell > 2^{N-1}$ (i.e., not in the first generation) to a circle depending on the w_j with $j < \ell$; note that this circle has positive radius provided that the parents of w_ℓ are distinct. Then, (3-8) implies that the left-hand side of (3-7) can be rewritten in a unique way as a linear combination of the w_ℓ only, so we have

$$\sum_{\ell=1}^b v_\ell w_\ell = 0. \tag{3-9}$$

Hence, the assumption that μ is linearly independent from the space generated by the $\lambda^{\mathcal{F}}$ is equivalent to the fact that $v \in \mathbb{R}^{2b}$ does not vanish.

Now, let

$$\bar{\ell} := \max\{\ell \mid v_\ell \neq 0\},$$

so that (3-9) is equivalent to

$$w_{\bar{\ell}} = -\frac{1}{v_{\bar{\ell}}} \sum_{\ell < \bar{\ell}} v_\ell w_\ell. \tag{3-10}$$

If $\bar{\ell} \leq 2^{N-1}$, then $w_{\bar{\ell}}$ is in the first generation. Since there are no restrictions (either linear or quadratic) on the first generation, the statement is trivial. Hence, assume $\bar{\ell} > 2^{N-1}$. We can assume (by removing from \mathcal{X} a closed subset of zero measure) that $v_h \neq v_k$ for all $h \neq k$. Then the quadratic constraint on $w_{\bar{\ell}} \in \mathbb{R}^2$ gives a circle of positive radius. By excluding at most one point of this circle, we can ensure that the relation (3-10) is not fulfilled, which proves the thesis of the lemma. \square

In view of Lemma 3.4, those vectors $\mu \in \mathbb{Z}^{|\mathcal{S}|}$ that are linear combinations of the family vectors assume a special importance, since that is the only case in which the relation $\sum \mu_i v_i = 0$ cannot be excluded when constructing the set \mathcal{S} . In that case, we will refer to $\mu \sim 0$ as a *formal identity*. In general, we will write $\mu \sim \lambda$ whenever the vector $\mu - \lambda$ is a linear combination of the family relations.

We introduce some more notation: given a vector $\lambda \in \mathbb{Z}^{|\mathcal{S}|}$, we denote by $\pi_j \lambda$ the projection of λ on the j -th generation, namely the projection of λ on $\mathcal{A}_j \subset \mathbb{Z}^{|\mathcal{S}|}$ defined by

$$\mathcal{A}_j := \text{Span}(\{e_i \mid v_i \text{ belongs to the } j\text{-th generation}\}; \mathbb{Z}).$$

Now, let $R_\alpha = \sum_i \alpha_i \lambda^{F_i}$ be a linear combination with integer coefficients of the family vectors. We denote by n_{R_α} the number of families on which the linear combination is supported, the cardinality of $\{i \mid \alpha_i \neq 0\}$. Moreover, we denote by $n_{R_\alpha}^k$ the number of families of age k on which R_α is supported, the cardinality of

$$\{i \mid \alpha_i \neq 0 \text{ and } F_i \text{ is a family of age } k\}.$$

Finally, we denote respectively by m_{R_α} and M_{R_α} the minimal and the maximal age of families on which R_α is supported. Then, we make the two following simple remarks.

Remark 3.5. If $n_{R_\alpha}^k = n_{R_\alpha}^{k+1} = 1$, then $\pi_{k+1} R_\alpha$ is supported on at least two distinct elements.

Remark 3.6. If $n_{R_\alpha}^k \neq n_{R_\alpha}^{k+1}$, then $\pi_{k+1} R_\alpha$ is supported on at least two distinct elements.

Before proving the main result of this section, we need some lemmas.

Lemma 3.7. *If $n_{R_\alpha} \geq 3$, then R_α is supported on at least 8 distinct elements.*

Proof. For simplicity of notation, here we put $m := m_{R_\alpha}$ and $M := M_{R_\alpha}$. First, observe that $\pi_m R_\alpha$ is supported on $2n_{R_\alpha}^m$ elements and that $\pi_{M+1} R_\alpha$ is supported on $2n_{R_\alpha}^M$ elements. So, if $n_{R_\alpha}^m + n_{R_\alpha}^M \geq 4$, the thesis is trivial.

Up to symmetry between parents and children, we may choose $n_{R_\alpha}^m \leq n_{R_\alpha}^M$. So, the only nontrivial cases to consider are $(n_{R_\alpha}^m, n_{R_\alpha}^M) = (1, 1)$ and $(n_{R_\alpha}^m, n_{R_\alpha}^M) = (1, 2)$.

Case (1,1): We must have $M \geq m + 2$, since there must be at least three families in R_α . Now, let $C := \max_i n_{R_\alpha}^i$. If $C = 1$ then, by Remark 3.5, the support of R_α involves at least 4 generations and at least 2 elements for each generation, so it includes at least 8 elements. If $C > 1$, then there exist $m \leq i, j < M$ with $i \neq j$ such that $n_{R_\alpha}^i < n_{R_\alpha}^{i+1}$ and $n_{R_\alpha}^{j+1} < n_{R_\alpha}^j$. Then, by Remark 3.6, $\pi_{i+1} R_\alpha$ and $\pi_{j+1} R_\alpha$ are supported on at least 2 elements each. Since $\pi_m R_\alpha$ and $\pi_{M+1} R_\alpha$ are supported on exactly 2 elements and since the four indices $m, i + 1, j + 1, M + 1$ are all distinct, then we have the thesis.

Case (1,2): Here, $\pi_m R_\alpha$ is supported on 2 elements and $\pi_M R_\alpha$ is supported on 4 elements. Moreover, there exists $m \leq i < M$ such that $n_{R_\alpha}^{i+1} < n_{R_\alpha}^i$, which, by Remark 3.6, gives us at least 2 elements in the support of $\pi_{i+1} R_\alpha$. Thus, we have the thesis. \square

From Lemma 3.7, the next corollary follows immediately.

Corollary 3.8. *If R_α is supported on at most 7 elements, then R_α is an integer multiple of either a family vector or a resonant vector of type CF.*

Lemma 3.9. *Let $A, B, C \in \mathbb{R}$, $R > 0$ and $p, q \in \mathbb{R}^2 \simeq \mathbb{C}$ be fixed. Let*

$$c_1(\vartheta) := p + Re^{i\vartheta}, \quad c_2(\vartheta) := p - Re^{i\vartheta}.$$

Then, the function $F : \mathbb{S}^1 \rightarrow \mathbb{R}$ defined by

$$F(\vartheta) := A|c_1(\vartheta)|^2 + B|c_2(\vartheta)|^2 + C - |Ac_1(\vartheta) + Bc_2(\vartheta) + q|^2$$

is an analytic function of ϑ , and it is a constant function only if $A = B$ or if $(A + B - 1)p + q = 0$.

Proof. An explicit computation yields

$$F(\vartheta) = 2R(B - A)\langle (A + B - 1)p + q, e^{i\vartheta} \rangle + K,$$

where K is a suitable constant that does not depend on ϑ . \square

Corollary 3.10. *If $A \neq B$ and $(A + B - 1)p + q \neq 0$, then the zeros of F are isolated.*

Lemma 3.11. *Let $\mathcal{F} = (p_1, p_2; c_1, c_2) \equiv (v_{i_1}, v_{i_2}; v_{i_3}, v_{i_4})$ be a family of age i in \mathcal{S} and let $\lambda^{\mathcal{F}^p} := \mathbf{e}_{i_1} + \mathbf{e}_{i_2}$ be the abstract vector corresponding to the sum of the parents of the family \mathcal{F} . Moreover, let $\mu \in \mathbb{Z}^{|\mathcal{S}|}$ be another vector with $|\mu| \leq 5$ such that $\pi_j \mu = 0$ for all $j > i + 1$ and such that the support of μ and the support of the abstract vector $\lambda^{\mathcal{F}^c} := \mathbf{e}_{i_3} + \mathbf{e}_{i_4}$ corresponding to the sum of the children of \mathcal{F} are disjoint. Finally, let $h, k \in \mathbb{Z} \setminus \{0\}$. Assume that the formal identity $h\mu + k\lambda^{\mathcal{F}^p} \sim 0$ holds. Then, only two possibilities are allowed:*

- (1) $h\mu + k\lambda^{\mathcal{F}^p} = 0$;
- (2) $h\mu + k\lambda^{\mathcal{F}^p}$ is an integer multiple of $\lambda^{\tilde{\mathcal{F}}}$, where $\tilde{\mathcal{F}}$ is a family of age $i - 1$, one of whose children is a parent in \mathcal{F} .

Proof. We first remark that $h\mu + k\lambda^{\mathcal{F}^p}$ is supported on at most 7 elements. Moreover, since it is a linear combination of some family vectors (because of the formal identity $h\mu + k\lambda^{\mathcal{F}^p} \sim 0$), we are in a position to apply Corollary 3.8 and conclude that $h\mu + k\lambda^{\mathcal{F}^p}$ must be an integer multiple of either a family vector or a resonant vector of type CF.

Now, assume by contradiction that $h\mu + k\lambda^{\mathcal{F}^p}$ is a nonzero integer multiple of a resonant vector of type CF. Then, the support of $h\mu + k\lambda^{\mathcal{F}^p}$ cannot include both parents of the family \mathcal{F} , since the support of a CF vector including a couple of parents of age i should include also a couple of children of age $i + 2$, but we know by the assumptions of this lemma that the support of $h\mu + k\lambda^{\mathcal{F}^p}$ does not include elements of age greater than $i + 1$. Therefore, at least one of the elements in λ must cancel out with one of the

elements in $\lambda^{\mathcal{F}_p}$, but then the support of $h\mu + k\lambda^{\mathcal{F}_p}$ can include at most 5 elements, and therefore it cannot be a vector of type CF.

Then, if $h\mu + k\lambda^{\mathcal{F}_p}$ is a nonzero integer multiple of a single family vector $\tilde{\mathcal{F}}$, observe that its support must contain one and only one of the parents of \mathcal{F} . In fact, if both canceled out, then the support of $h\mu + k\lambda^{\mathcal{F}_p}$ could contain at most 3 elements, which is absurd. If none of them canceled out, then we should have $\tilde{\mathcal{F}} = \mathcal{F}$, which in turn is absurd since, by the assumptions of this lemma, the support of $h\mu + k\lambda^{\mathcal{F}_p}$ cannot include any of the children of the family \mathcal{F} . This concludes the proof of the lemma. \square

We can now prove the main proposition.

Proof of Proposition 3.1. The proof is based on the following induction procedure. At each step, we assume we have already fixed i generations and say $h < 2^{N-2}$ families with children in the $(i+1)$ -st generation. Our induction hypothesis is that the nondegeneracy condition is satisfied for the vectors μ whose support involves only the elements that we have already fixed. Then, our aim is to show that the nondegeneracy condition holds true also for the set of vectors supported on the already fixed elements plus the two children of a new family (whose procreation angle has to be chosen accordingly) with children in the $(i+1)$ -st generation, up to removing from \mathcal{X} a closed set of null measure.

First, we observe that, at the inductive step zero, that is, when placing the first generation \mathcal{S}_1 , the set of parameters that satisfy both nondegeneracy and nonvanishing of any fixed finite number of linear relations that are not formal identities is obviously open and of full measure.

Then, we have to study what happens when choosing a procreation angle, i.e., when generating the children of a family $\mathcal{F} = (p_1, p_2; c_1, c_2) \equiv (v_{i_1}, v_{i_2}; v_{i_3}, v_{i_4})$ whose parents $(p_1, p_2) \equiv (v_{i_1}, v_{i_2})$ have already been fixed. We need to study the nondegeneracy condition associated to the vector $\lambda(A, B, \mu) \in \mathbb{Z}^{|\mathcal{S}|}$ given by

$$\lambda(A, B, \mu) := A\mathbf{e}_{i_3} + B\mathbf{e}_{i_4} + \mu,$$

where μ satisfies the same properties as in the assumptions of Lemma 3.11, and

$$|A| + |B| + |\mu| \leq 5 \quad \text{and} \quad A + B + \sum_j \mu_j = 1.$$

If $A \neq B$ and if $(A+B-1)(p_1+p_2)+2\sum_j \mu_j v_j \neq 0$, then we are done, because, thanks to Corollary 3.10, the nondegeneracy condition is satisfied for any choice of the generation angle except at most a finite number. Therefore, we have to study separately the case $A = B$ and, for $A \neq B$, we have to prove that $(A+B-1)\lambda^{\mathcal{F}_p} + 2\mu \sim 0$ holds as a formal identity only in the cases allowed by Definition 2.8. Whenever the formal identity $(A+B-1)\lambda^{\mathcal{F}_p} + 2\mu \sim 0$ does not hold, we can impose $(A+B-1)(p_1+p_2)+2\sum_j \mu_j v_j \neq 0$ by just removing from \mathcal{X} a closed set of measure zero, thanks to Lemma 3.4.

Case $A = B$: If $(A, B) = (0, 0)$ there is nothing to prove, thanks to the induction hypothesis. Then we have to study $(A, B) = \pm(1, 1)$. In this case, thanks to the linear relation defining the family \mathcal{F} , we have the formal identity

$$\lambda(A, B, \mu) \sim \pm\lambda^{\mathcal{F}_p} + \mu =: v^\pm$$

with $|\mu| \leq 3$, $\sum_j v_j^+ = -1$ and $\sum_j v_j^- = 3$. The good point is that v^\pm is entirely supported on the elements of the generation set that have already been fixed, so we can apply the induction hypothesis of nondegeneracy to v^\pm and distinguish the 4 cases given by Definition 2.8: we have to verify that $\lambda(A, B, \mu)$ accordingly falls into one of the allowed cases:

- v^\pm satisfies (1) of Definition 2.8. Then one readily verifies that $\lambda(A, B, \mu)$ satisfies either (2) or (3) of Definition 2.8.
- v^\pm satisfies (2) of Definition 2.8. Observe that the family involved by the statement of (2) cannot be \mathcal{F} , since v^\pm cannot be supported on either child of the family \mathcal{F} . Then μ must cancel out one of the two parents appearing in $\pm\lambda^{\mathcal{F}_p}$. It cannot be supported on both parents because that would not be consistent with $|\mu| \leq 3$ and $|v^\pm| = 3$. Then one verifies that $\lambda(A, B, \mu)$ satisfies (4) of Definition 2.8.
- v^\pm satisfies (3) of Definition 2.8. Since $|v^\pm| = 1$, then nothing cancels out, so the support of v^\pm includes both parents of \mathcal{F} . But this is absurd, so this case cannot happen.
- v^\pm satisfies (4) of Definition 2.8. This case is again absurd, since v^\pm should be supported on 5 of the 6 elements of a CF vector, including the two parents of the family \mathcal{F} .

Case $A \neq B$: By symmetry, we may suppose $|A| > |B|$. Assume that $(A + B - 1)\lambda^{\mathcal{F}_p} + 2\mu \sim 0$ holds as a formal identity; we must prove that this can be true only in the cases allowed by Definition 2.8. First, we consider the case $A + B - 1 = 0$: then, we must have the formal identity $\mu \sim 0$ with $|\mu| \leq 5$: so, by Corollary 3.8, either μ is (up to the sign) a family vector (which may happen only if $(A, B) = (1, 0)$ due to the constraint $|A| + |B| + |\mu| \leq 5$) or $\mu = 0$. Consider the case $(A, B) = (1, 0)$: if μ is a family vector, then $\lambda(A, B, \mu)$ falls into case (3) of Definition 2.8; if $\mu = 0$, then $\lambda(A, B, \mu)$ falls into case (1) of Definition 2.8. If $(A, B) = (2, -1)$ or $(A, B) = (3, -2)$, then $\mu = 0$. Then, in both cases, from

$$\sum_j \lambda_j(A, B, \mu) |v_j|^2 - \left| \sum_j \lambda_j(A, B, \mu) v_j \right|^2$$

with some explicit computations one deduces $|c_1 - c_2|^2 = 0$, which is absurd, since the induction hypothesis implies $p_1 \neq p_2$ and since the endpoints of a diameter of a circle with positive radius are distinct.

Now, if $A + B - 1 \neq 0$ we can apply Lemma 3.11 and deduce that $(A + B - 1)\lambda^{\mathcal{F}_p} + 2\mu$ is either zero or an integer multiple of the vector of a family where one of the parents of \mathcal{F} appears as a child. Suppose first $(A + B - 1)\lambda^{\mathcal{F}_p} + 2\mu = 0$. Then $A + B - 1$ must be even. If $(A, B) = (-1, 0)$, then $\mu = \lambda^{\mathcal{F}_p}$ and $\lambda(A, B, \mu)$ falls into case (2) of Definition 2.8. If $(A, B) = (2, 1)$, then $\mu = -\lambda^{\mathcal{F}_p}$ and $\lambda(A, B, \mu)$ falls into case (3) of Definition 2.8. These are the only possible cases if $(A + B - 1)\lambda^{\mathcal{F}_p} + 2\mu = 0$. Finally, assume that $(A + B - 1)\lambda^{\mathcal{F}_p} + 2\mu$ is an integer multiple of the vector of a family where one of the parents of \mathcal{F} appears as a child. Then μ must be such that the other parent of \mathcal{F} is canceled out, so $A + B - 1$ again has to be even. If $(A, B) = (-1, 0)$, then $\mu - \lambda^{\mathcal{F}_p}$ is the vector of a family where one of the parents of \mathcal{F} appears as a child and $\lambda(A, B, \mu)$ falls into case (4) of Definition 2.8. This is the only possible case, since the support of μ must include one parent of \mathcal{F} and the other three members of the family where the other parent of \mathcal{F} appears as a child. This also concludes the proof of the proposition. \square

4. Proof of Theorem 1.1

In the previous sections we have proved the existence of nondegenerate sets \mathcal{S} on which the Hamiltonian is (2-1) and the existence of a *slider solution* for its dynamics. We now turn to the NLS equation (1-4) with the purpose of proving the persistence of this type of solution.

As in [CKSTT], one can easily prove that (1-4) is locally well-posed in $\ell^1(\mathbb{Z}^2)$: to this end, one first introduces the multilinear operator

$$\mathcal{N}(t) : \ell^1(\mathbb{Z}^2) \times \ell^1(\mathbb{Z}^2) \times \ell^1(\mathbb{Z}^2) \times \ell^1(\mathbb{Z}^2) \times \ell^1(\mathbb{Z}^2) \rightarrow \ell^1(\mathbb{Z}^2)$$

defined by

$$(\mathcal{N}(t)(a, b, c, d, f))_j := \sum_{\substack{j_1, j_2, j_3, j_4, j_5 \in \mathbb{Z}^2 \\ j_1 + j_2 + j_3 - j_4 - j_5 = j}} a_{j_1} b_{j_2} c_{j_3} \bar{d}_{j_4} \bar{f}_{j_5} e^{i\omega_6 t}, \quad (4-1)$$

so that (1-4) can be rewritten as

$$-i\dot{a}_j = (\mathcal{N}(t)(a, a, a, a, a))_j.$$

Then, in order to obtain local well-posedness, it is enough to observe that the following multilinear estimate holds:

$$\|\mathcal{N}(t)(a, b, c, d, f)\|_{\ell^1} \lesssim \|a\|_{\ell^1} \|b\|_{\ell^1} \|c\|_{\ell^1} \|d\|_{\ell^1} \|f\|_{\ell^1}. \quad (4-2)$$

Lemma 4.1. *Let $0 < \sigma < 1$ be an absolute constant (all implicit constants in this lemma may depend on σ). Let $B \gg 1$, and let $T \ll B^4 \log B$. Let $g(t) := \{g_j(t)\}_{j \in \mathbb{Z}^2}$ be a solution of the equation*

$$\dot{g}(t) = i(\mathcal{N}(t)(g(t), g(t), g(t), g(t), g(t)) + \mathcal{E}(t)) \quad (4-3)$$

for times $0 \leq t \leq T$, where $\mathcal{N}(t)$ is defined in (4-1) and the initial data $g(0)$ is compactly supported. Assume also that the solution $g(t)$ and the error term $\mathcal{E}(t)$ obey bounds of the form

$$\|g(t)\|_{\ell^1(\mathbb{Z}^2)} \lesssim B^{-1}, \quad (4-4)$$

$$\left\| \int_0^t \mathcal{E}(\tau) d\tau \right\|_{\ell^1(\mathbb{Z}^2)} \lesssim B^{-1}. \quad (4-5)$$

We conclude that, if $a(t)$ denotes the solution to the NLS (1-4) with initial data $a(0) = g(0)$, then we have

$$\|a(t) - g(t)\|_{\ell^1(\mathbb{Z}^2)} \lesssim B^{-1-\sigma/2} \quad (4-6)$$

for all $0 \leq t \leq T$.

Proof. The proof is the transposition to the quintic case of the proof of Lemma 2.3 of [CKSTT] and is postponed to Appendix B. \square

Given δ, K , construct \mathcal{S} as in Proposition 3.3. Note that we are free to specify $\mathcal{R} = \mathcal{R}(\delta, K)$ (which measures the inner radius of the frequencies involved in \mathcal{S}) as large as we wish. This construction fixes $N = N(\delta, K)$ (the number of generations). We introduce a further parameter ϵ (which we are free to specify as a function of δ, K) and construct the slider solution $b(t)$ to the toy model concentrated at

scale ϵ according to Proposition 2.10 above. This proposition also gives us a time $T_0 = T_0(K, \delta)$. Note that the toy model has the scaling

$$b^{(\lambda)}(t) := \lambda^{-1} b\left(\frac{t}{\lambda^4}\right).$$

We choose the initial data for NLS by setting

$$a_j(0) = b_i^{(\lambda)}(0) \quad \text{for all } j \in \mathcal{S}_i \quad (4-7)$$

and $a_j(t) = 0$ when $j \notin \mathcal{S}$. We want to apply the approximation lemma, Lemma 4.1, with a parameter B chosen large enough so that

$$B^4 \log B \gg \lambda^4 T_0. \quad (4-8)$$

We set $g(t) = \{g_j(t)\}_{j \in \mathbb{Z}^2}$ defined by the slider solution as

$$g_j(t) = b_i^{(\lambda)}(t) \quad \text{for all } j \in \mathcal{S}_i,$$

$g_j(t) = 0$ otherwise. Then we set $\mathcal{E}(t) := \{\mathcal{E}_j(t)\}_{j \in \mathbb{Z}^2}$ with

$$\mathcal{E}_j(t) = - \sum_{\substack{k_i \in \mathcal{S}: k_1 + k_2 + k_3 - k_4 - k_5 = j \\ \omega_6 \neq 0}} g_{k_1} g_{k_2} g_{k_3} \bar{g}_{k_4} \bar{g}_{k_5} e^{i\omega_6 t},$$

where $\omega_6 = |k_1|^2 + |k_2|^2 + |k_3|^2 - |k_4|^2 - |k_5|^2 - |j|^2$. We recall that the frequency support of $g(t)$ is in \mathcal{S} for all times. We choose $B = C(N)\lambda$ and then show that, for large enough λ , the required conditions (4-4), (4-5) hold true. Observe that (4-8) holds true with this choice for large enough λ . Note first that, simply by considering its support, the fact that $|\mathcal{S}| = C(N)$, and the fact that $\|b(t)\|_{\ell^\infty} \sim 1$, we can be sure that $\|b(t)\|_{\ell^1(\mathbb{Z})} \sim C(N)$ and therefore

$$\|b^{(\lambda)}(t)\|_{\ell^1(\mathbb{Z})}, \|g(t)\|_{\ell^1(\mathbb{Z}^2)} \leq \lambda^{-1} C(N). \quad (4-9)$$

Thus, (4-4) holds with the choice $B = C(N)\lambda$. For the second condition, (4-5), we claim

$$\left\| \int_0^t \mathcal{E}(\tau) d\tau \right\|_{\ell^1} \lesssim C(N)(\lambda^{-5} + \lambda^{-9}T). \quad (4-10)$$

This implies (4-5) since $B = \lambda C(N)$ and $T = \lambda^4 T_0$.

We now prove (4-10). Since ω_6 does not vanish in the sum defining \mathcal{E} , we can replace $e^{i\omega_6 \tau}$ by $d[e^{i\omega_6 \tau}/(i\omega_6)]/d\tau$ and then integrate by parts. Three terms arise: the boundary terms at $\tau = 0, T$ and the integral term involving

$$\frac{d}{d\tau} [g_{k_1} g_{k_2} g_{k_3} \bar{g}_{k_4} \bar{g}_{k_5}].$$

For the boundary terms, we use (4-9) to obtain an upper bound of $C(N)\lambda^{-5}$. For the integral term, the τ -derivative falls on one of the g factors. We replace this differentiated term using the equation to get

an expression that is 9-linear in g and bounded by $C(N)\lambda^{-9}T$. Once λ has been chosen as above, we choose \mathcal{R} sufficiently large so that the initial data $g(0) = a(0)$ has the right size:

$$\left(\sum_{j \in \mathcal{S}} |g_j(0)|^2 |j|^{2s} \right)^{\frac{1}{2}} \sim \delta. \quad (4-11)$$

This is possible since the quantity on the left scales like λ^{-1} and \mathcal{R}^s respectively in the parameters λ , \mathcal{R} . The issue here is that our choice of frequencies \mathcal{S} only gives us a large factor (that is, K/δ) by which the Sobolev norm of the solution will grow. If our initial datum is much smaller than δ in size, the Sobolev norm of the solution will not grow to be larger than K . It remains to show that we can guarantee

$$\left(\sum_{j \in \mathbb{Z}^2} |a_j(\lambda^4 T_0)|^2 |j|^{2s} \right)^{\frac{1}{2}} \geq K, \quad (4-12)$$

where $a(t)$ is the evolution of the initial datum $g(0)$ under the NLS. We do this by first establishing

$$\left(\sum_{j \in \mathcal{S}} |g_j(\lambda^4 T_0)|^2 |j|^{2s} \right)^{\frac{1}{2}} \gtrsim K, \quad (4-13)$$

and then

$$\sum_{j \in \mathcal{S}} |g_j(\lambda^4 T_0) - a_j(\lambda^4 T_0)|^2 |j|^{2s} \lesssim 1. \quad (4-14)$$

In order to prove (4-13), consider the ratio

$$Q := \frac{\sum_{j \in \mathcal{S}} |g_j(\lambda^4 T_0)|^2 |j|^{2s}}{\sum_{j \in \mathcal{S}} |g_j(0)|^2 |j|^{2s}} = \frac{\sum_{i=1}^N |b_i^{(\lambda)}(\lambda^4 T_0)|^2 \sum_{j \in \mathcal{S}_i} |j|^{2s}}{\sum_{i=1}^N |b_i^{(\lambda)}(0)|^2 \sum_{j \in \mathcal{S}_i} |j|^{2s}}. \quad (4-15)$$

Set $\mathfrak{J}_i := \sum_{j \in \mathcal{S}_i} |j|^{2s}$; by construction, $\mathfrak{J}_i/\mathfrak{J}_j \sim 2^{i-j}$ and, by the choice of N , one has $\mathfrak{J}_3/\mathfrak{J}_{N-2} \lesssim \delta^2 K^{-2}$. Then one has

$$\begin{aligned} \frac{\sum_{i=1}^N |b_i^{(\lambda)}(\lambda^4 T_0)|^2 \mathfrak{J}_i}{\sum_{i=1}^N |b_i^{(\lambda)}(0)|^2 \mathfrak{J}_i} &\gtrsim \frac{\mathfrak{J}_{N-2}(1-\epsilon)}{\epsilon \sum_{i \neq 3} \mathfrak{J}_i + (1-\epsilon)\mathfrak{J}_3} \\ &= \frac{1-\epsilon}{\epsilon \sum_{i \neq 3} \mathfrak{J}_i/\mathfrak{J}_{N-2} + (1-\epsilon)\mathfrak{J}_3/\mathfrak{J}_{N-2}} = \frac{1}{\mathfrak{J}_3/\mathfrak{J}_{N-2} + O(\epsilon)} \gtrsim \frac{K^2}{\delta^2} \end{aligned}$$

provided that $\epsilon = \epsilon(N, K, \delta)$ is sufficiently small.

In order to prove (4-14), we use the approximation lemma, Lemma 4.1, to obtain that

$$\sum_{j \in \mathcal{S}} |g_j(\lambda^4 T_0) - a_j(\lambda^4 T_0)|^2 |j|^{2s} \lesssim \lambda^{-2-\sigma} \sum_{j \in \mathcal{S}} |j|^{2s} \leq \frac{1}{2}. \quad (4-16)$$

The last inequality is obtained by scaling λ by some (big) parameter C and \mathcal{R} by $C^{1/s}$ so that the bound (4-11) still holds while $\lambda^{-2-\sigma} \sum_{n \in \mathcal{S}} |j|^{2s}$ scales as $C^{-\sigma}$.

Appendix A: Proof of Proposition 2.13

This proof is in fact exactly the same as in [CKSTT], however in that paper all the results are stated for the cubic case (even though they are clearly more general) and so we give a schematic overview of the main steps.

Lemma A.1. *Suppose that $[0, \tau]$ is a time interval on which we have the smallness condition*

$$\int_0^\tau |c(s)|^2 ds \lesssim 1;$$

then we have the estimates

$$\begin{aligned} |c_{j\pm 1}^-(\tau)| &\lesssim e^{-\sqrt{3}\tau} |c_{j\pm 1}^-(0)| + \int_0^\tau e^{-\sqrt{3}(\tau-s)} |c_{j\pm 1}^+(s)| |c_{\neq j\pm 1}|^2, \\ |c_{j\pm 1}^+(\tau)| &\lesssim e^{\sqrt{3}\tau} |c_{j\pm 1}^+(0)| + \int_0^\tau e^{\sqrt{3}(\tau-s)} |c_{j\pm 1}^-(s)| |c_{\neq j\pm 1}|^2, \\ |c_{j\pm 1}(\tau)| &\lesssim e^{\sqrt{3}\tau} |c_{j\pm 1}(0)|, \\ |c^*(\tau)| &\lesssim |c^*(0)|. \end{aligned}$$

Proof. As in [CKSTT] this lemma follows from equations (2-13) by Gronwall's inequality and the definition of $O(\cdot)$. \square

We now prove that the incoming target covers the ricochet target. We start from some basic upper bounds on the flow.

Proposition A.2. *Let $b(\tau)$ be a solution to the toy model such that $b(0)$ is within (M_j^-, d_j^-, R_j^-) . Let $c(\tau)$ denote the coordinates of $b(\tau)$ as in (2-13). Then, for all $0 \leq \tau \leq T$, we have the bounds*

$$\begin{aligned} |c^*(\tau)| &= O(T^{A_j^-} e^{-2\sqrt{3}T}), \\ |c_{j-1}^-(\tau)| &= O(\sigma e^{-\sqrt{3}\tau}), \\ |c_{j-1}^+(\tau)| &= O(T^{2A_j^-+1} e^{-4\sqrt{3}T+\sqrt{3}\tau}), \\ |c_{j+1}^-(\tau)| &= O(r_j^-(1+\tau) e^{-2\sqrt{3}T-\sqrt{3}\tau}), \\ |c_{j+1}^+(\tau)| &= O(r_j^- e^{-2\sqrt{3}T+\sqrt{3}\tau}). \end{aligned} \tag{A-1}$$

Proof. This is Proposition 3.2 of [CKSTT]. The proof is an application of the continuity method and of Lemma A.1. \square

Now, from these basic upper bounds and from the equations of motion (2-13) and (2-16), we deduce improved upper bounds on the dynamical variables. We first consider c_{j-1}^- ; we have

$$\dot{c}_{j-1}^- = -\sqrt{3}c_{j-1}^- + \mathcal{O}((c_{j-1}^-)^3) + \mathcal{O}((c_{j-1}^-)^5) + O(T^{A_j^-} e^{-2\sqrt{3}T})$$

for some explicit expression $\mathcal{O}((c_{j-1}^-)^3) + \mathcal{O}((c_{j-1}^-)^5)$. Let g be the solution to the corresponding equation

$$\dot{g} = -\sqrt{3}g + \mathcal{O}(g^3) + \mathcal{O}(g^5)$$

with the same initial datum $g(0) = \sigma$. One has the bound

$$g(\tau) = O(\sigma e^{\sqrt{3}\tau}), \quad (\text{A-2})$$

which is formula (3.35) of [CKSTT]. Then, by estimating the *error term* $E_{j-1}^- := c_{j-1}^- - g$, one has

$$c_{j-1}^-(\tau) = g(\tau) + O(T^{A_j^-+1} e^{-2\sqrt{3}T}), \quad (\text{A-3a})$$

$$\mathcal{O}(c^2) = \mathcal{O}(g^2) + O(T^{A_j^-+1} e^{-2\sqrt{3}T}), \quad (\text{A-3b})$$

$$\mathcal{O}(c_{\neq j+1}^2) = \mathcal{O}(g^2) + O(T^{A_j^-+1} e^{-2\sqrt{3}T - \sqrt{3}\tau}), \quad (\text{A-3c})$$

which are respectively formulae (3.36)–(3.38) of [CKSTT]. Now we control the leading peripheral modes. Inserting (A-3b) in (2-13e), we see that

$$\dot{c}_{\geq j+2} = i\kappa c_{\geq j+2} + \mathcal{O}(c_{\geq j+2} g^2) + \mathcal{O}(c_{\geq j+2} g^4) + O(T^{A_j^-} e^{-2\sqrt{3}T} |c_{\geq j+2}|).$$

We approximate this by the corresponding linear equation

$$\dot{u} = i\kappa u + \mathcal{O}(u g^2) + \mathcal{O}(u g^4),$$

where $u(\tau) \in \mathbb{C}^{N-j-1}$. This equation has a fundamental solution $G_{\geq j+2}(\tau) : \mathbb{C}^{N-j-1} \rightarrow \mathbb{C}^{N-j-1}$. From (A-2) and Gronwall's inequality, we have

$$\int_0^T g^2(\tau) d\tau = O(1) \quad (\text{A-4})$$

and

$$|G_{\geq j+2}|, |G_{\geq j+2}^{-1}| = O(1). \quad (\text{A-5})$$

Setting $c_{\geq j+2}(0) = e^{-2\sqrt{3}T} a_{\geq j+2} + O(T^{A_j^-} e^{-3\sqrt{3}T})$, we define

$$E_{\geq j+2} := c_{\geq j+2} - e^{-2\sqrt{3}T} G_{\geq j+2} a_{\geq j+2}.$$

Applying the bound on $c_{\geq j+2}$ from Proposition A.2 and Gronwall's inequality, we conclude

$$|E_{\geq j+2}(\tau)| = O(T^{A_j^-} e^{-3\sqrt{3}T})$$

for all $0 \leq \tau \leq T$, and thus

$$c_{\geq j+2}(\tau) = e^{-2\sqrt{3}T} G_{\geq j+2}(\tau) a_{\geq j+2} + O(T^{A_j^-} e^{-3\sqrt{3}T}). \quad (\text{A-6})$$

This is formula (3.41) of [CKSTT].

Now we consider the two leading secondary modes c_{j+1}^+ , c_{j+1}^- simultaneously. From (2-13), (A-3) and Proposition A.2, we have the system

$$\begin{pmatrix} \dot{c}_{j+1}^- \\ \dot{c}_{j+1}^+ \end{pmatrix} = \sqrt{3} \begin{pmatrix} -c_{j+1}^- \\ c_{j+1}^+ \end{pmatrix} + M(\tau) \begin{pmatrix} c_{j+1}^- \\ c_{j+1}^+ \end{pmatrix} + \begin{pmatrix} O(T^{A_j^-+1} e^{-4\sqrt{3}T}) \\ O(T^{A_j^-+1} e^{-4\sqrt{3}T + \sqrt{3}\tau}) \end{pmatrix}.$$

Here $M(\tau)$ is a two-by-two matrix with all entries $\mathcal{O}(g^2) + \mathcal{O}(g^4)$. Passing to the variables

$$\tilde{a}_{j+1}(\tau) := \begin{pmatrix} \tilde{a}_{j+1}^-(\tau) \\ \tilde{a}_{j+1}^+(\tau) \end{pmatrix},$$

where

$$\tilde{a}_{j+1}^-(\tau) = e^{2\sqrt{3}T + \sqrt{3}\tau} c_{j+1}^-(\tau), \quad \tilde{a}_{j+1}^+(\tau) = e^{2\sqrt{3}T - \sqrt{3}\tau} c_{j+1}^+(\tau),$$

we get the equation

$$\begin{cases} \partial_\tau \tilde{a}_{j+1}(\tau) = A(\tau) \tilde{a}_{j+1}(\tau) + \mathcal{O}(T^{A_j^- + 1} e^{-2\sqrt{3}T + \sqrt{3}\tau}), \\ \tilde{a}_{j+1}(0) = a_{j+1} + \mathcal{O}(T^{A_j^-} e^{-\sqrt{3}T}), \end{cases} \quad (\text{A-7})$$

where $A(\tau)$ is some known matrix which (by (A-2)) has bounds

$$A(\tau) = \sigma^2 \begin{pmatrix} \mathcal{O}(e^{-2\sqrt{3}\tau}) & \mathcal{O}(1) \\ \mathcal{O}(e^{-4\sqrt{3}\tau}) & \mathcal{O}(e^{-2\sqrt{3}\tau}) \end{pmatrix}.$$

We have obtained formula (3.42) of [CKSTT]. Hence, following verbatim the proof given in [CKSTT], we get

$$\begin{pmatrix} e^{2\sqrt{3}T + \sqrt{3}\tau} c_{j+1}^- \\ e^{2\sqrt{3}T - \sqrt{3}\tau} c_{j+1}^+ \end{pmatrix} = G_{j+1}(\tau) a_{j+1} + \mathcal{O}(T^{A_j^- + 2} e^{-\sqrt{3}T}), \quad (\text{A-8})$$

which is formula (3.45) of [CKSTT].

Then, following Section 3.7 of [CKSTT] verbatim, we deduce that the incoming target covers the ricochet target.

Then, one has to prove that the ricochet target covers the outgoing target. In order to do this, one should adapt Sections 3.8–3.9 of [CKSTT] exactly as we have done in the previous section. Since this is completely straightforward, we will not write it down.

The last step consists in proving that the outgoing target (M_j^+, d_j^+, r_j^+) covers the next incoming target $(M_{j+1}^-, d_{j+1}^-, r_{j+1}^-)$. An initial datum in the outgoing target has the form

$$\begin{aligned} c_{\leq j-1}(0) &= \mathcal{O}(T^{A_j^+} e^{-2\sqrt{3}T}), \\ c_{j+1}^-(0) &= \mathcal{O}(T^{A_j^+} e^{-4\sqrt{3}T}), \\ c_{j+1}^+(0) &= \sigma + \mathcal{O}(T^{A_j^+} e^{-\sqrt{3}T}), \\ c_{\geq j+2}(0) &= e^{-2\sqrt{3}T} a_{\geq j+2} + \mathcal{O}(T^{A_j^+} e^{-3\sqrt{3}T}) \end{aligned}$$

for some $a_{\geq j+2}$ of magnitude at most r_j^+ . From (2-13e), (2-14) and (2-15) we deduce

$$\dot{c}_{\neq j+1} = \mathcal{O}(|c_{\neq j+1}|).$$

Thus, for all $0 \leq \tau \leq 10 \log(1/\sigma)$, Gronwall's inequality gives

$$c_{\neq j+1}(\tau) = \mathcal{O}\left(\frac{1}{\sigma^{O(1)}} T^{A_j^+} e^{-2\sqrt{3}T}\right). \quad (\text{A-9})$$

The stable leading mode c_{j+1}^- can be controlled by (2-13c), which, by (A-9), becomes

$$\dot{c}_{j+1}^- = O(|c_{j+1}^-|) + O\left(\frac{1}{\sigma^{O(1)}} T^{2A_j^+} e^{-4\sqrt{3}T}\right).$$

By Gronwall's inequality we conclude

$$c_{j+1}^-(\tau) = O\left(\frac{1}{\sigma^{O(1)}} T^{2A_j^+} e^{-4\sqrt{3}T}\right). \quad (\text{A-10})$$

Then, taking the c_{j+1}^+ component of (2-11), we obtain, by (A-9) and (A-10),

$$\dot{c}_{j+1}^+ = \sqrt{3}(1 - |c_{j+1}^+|^2)c_{j+1}^+ + O\left(\frac{1}{\sigma^{O(1)}} T^{2A_j^+} e^{-4\sqrt{3}T}\right).$$

As in [CKSTT], we define \hat{g} to be the solution to the ODE

$$\partial_\tau \hat{g} = \sqrt{3}(1 - |\hat{g}|^2)\hat{g} \quad (\text{A-11})$$

with initial datum $\hat{g}(0) = \sigma$. This solution can easily be computed and is given by

$$\hat{g}(\tau) = \frac{1}{\sqrt{1 + e^{-2\sqrt{3}(\tau - \tau_0)}}},$$

where τ_0 is defined by

$$\frac{1}{\sqrt{1 + e^{2\sqrt{3}\tau_0}}} = \sigma.$$

We note that

$$\hat{g}(2\tau_0) = \frac{1}{\sqrt{1 + e^{-2\sqrt{3}\tau_0}}} = \sqrt{1 - \sigma^2}$$

and that $2\tau_0 \leq 10 \log(1/\sigma)$ if σ is small enough. Then, estimating as in [CKSTT] (via Gronwall's inequality) the error

$$E_{j+1}^+ := c_{j+1}^+ - \hat{g},$$

we get

$$c_{j+1}^+(\tau) = \hat{g}(\tau) + O\left(\frac{1}{\sigma^{O(1)}} T^{A_j^+} e^{-\sqrt{3}T}\right). \quad (\text{A-12})$$

This (together with (A-9) and (A-10)) implies

$$\mathcal{O}(c^2) + \mathcal{O}(c^4) = \mathcal{O}(\hat{g}^2) + \mathcal{O}(\hat{g}^4) + O\left(\frac{1}{\sigma^{O(1)}} T^{A_j^+} e^{-\sqrt{3}T}\right). \quad (\text{A-13})$$

Now, from (2-13c), (A-9) and (A-13), we have

$$\dot{c}_{\geq j+2} = i\kappa c_{\geq j+2} + \mathcal{O}(\hat{g}^2 c_{\geq j+2}) + \mathcal{O}(\hat{g}^4 c_{\geq j+2}) + O\left(\frac{1}{\sigma^{O(1)}} T^{2A_j^+} e^{-3\sqrt{3}T}\right).$$

We approximate this flow by the linear equation

$$\dot{u} = i\kappa u + \mathcal{O}(u\hat{g}^2) + \mathcal{O}(u\hat{g}^4),$$

where $u(\tau) \in \mathbb{C}^{N-j-1}$. This equation has a fundamental solution $\hat{G}_{\geq j+2}(\tau) : \mathbb{C}^{N-j-1} \rightarrow \mathbb{C}^{N-j-1}$ for all $\tau \geq 0$; from the boundedness of \hat{g} and Gronwall's inequality we get

$$|\hat{G}_{\geq j+2}(\tau)|, |\hat{G}_{\geq j+2}^{-1}(\tau)| \lesssim \frac{1}{\sigma^{O(1)}}. \quad (\text{A-14})$$

As in [CKSTT], a Gronwall estimate of the error

$$E_{\geq j+2}(\tau) := c_{\geq j+2}(\tau) - e^{-2\sqrt{3}T} \hat{G}_{\geq j+2}(\tau) a_{\geq j+2}$$

gives

$$c_{\geq j+2}(\tau) = e^{-2\sqrt{3}T} \hat{G}_{\geq j+2}(\tau) a_{\geq j+2} + O\left(\frac{1}{\sigma^{O(1)}} T^{2A_j^+} e^{-3\sqrt{3}T}\right), \quad (\text{A-15})$$

which is equation (3.62) of [CKSTT]. Then, at the time $\tau = 2\tau_0 \leq 10 \log(1/\sigma)$, the estimates become

$$\begin{aligned} c_{\leq j-1}(2\tau_0) &= O\left(\frac{1}{\sigma^{O(1)}} T^{A_j^+} e^{-2\sqrt{3}T}\right), \\ c_{j+1}^-(2\tau_0) &= O\left(\frac{1}{\sigma^{O(1)}} T^{2A_j^+} e^{-4\sqrt{3}T}\right), \\ c_{j+1}^+(2\tau_0) &= \sqrt{1-\sigma^2} + O\left(\frac{1}{\sigma^{O(1)}} T^{A_j^+} e^{-\sqrt{3}T}\right), \\ c_{\geq j+2}(2\tau_0) &= e^{-2\sqrt{3}T} \hat{G}_{\geq j+2}(2\tau_0) a_{\geq j+2} + O\left(\frac{1}{\sigma^{O(1)}} T^{2A_j^+} e^{-3\sqrt{3}T}\right). \end{aligned}$$

From this, we deduce

$$|b_j| = \left(1 - \sum_{k \neq j} |c_k|^2\right)^{\frac{1}{2}} = \sigma + O\left(\frac{1}{\sigma^{O(1)}} T^{A_j^+} e^{-\sqrt{3}T}\right).$$

Moving back to the coordinates b_1, \dots, b_N , this means that we have

$$\begin{aligned} b_{\leq j-1}(2\tau_0) &= O\left(\frac{1}{\sigma^{O(1)}} T^{A_j^+} e^{-2\sqrt{3}T}\right), \\ b_j(2\tau_0) &= \left[\sigma + \Re O\left(\frac{1}{\sigma^{O(1)}} T^{A_j^+} e^{-\sqrt{3}T}\right)\right] e^{i\vartheta^{(j)}(2\tau_0)}, \\ b_{j+1}(2\tau_0) &= \left[\sqrt{1-\sigma^2} + \Re O\left(\frac{1}{\sigma^{O(1)}} T^{A_j^+} e^{-\sqrt{3}T}\right)\right] \bar{\omega} e^{i\vartheta^{(j)}(2\tau_0)} + O\left(\frac{1}{\sigma^{O(1)}} T^{2A_j^+} e^{-4\sqrt{3}T}\right), \\ b_{\geq j+2}(2\tau_0) &= e^{i\vartheta^{(j)}(2\tau_0)} e^{-2\sqrt{3}T} \hat{G}_{\geq j+2}(2\tau_0) a_{\geq j+2} + O\left(\frac{1}{\sigma^{O(1)}} T^{2A_j^+} e^{-3\sqrt{3}T}\right), \end{aligned}$$

where the notation $f = \Re O(\cdot)$ means that both $f = O(\cdot)$ and $f \in \mathbb{R}$. We now have to recast this in terms of the variables $c_1^{(j+1)}, \dots, c_N^{(j+1)}$ in phase with \mathbb{T}_{j+1} . Following [CKSTT], we denote these variables by $\tilde{c}_1, \dots, \tilde{c}_N$. We first note that

$$\vartheta^{(j+1)}(2\tau_0) = \vartheta^{(j)}(2\tau_0) + \bar{\omega} + O\left(\frac{1}{\sigma^{O(1)}} T^{2A_j^+} e^{-4\sqrt{3}T}\right).$$

Then, we deduce our final estimates

$$\begin{aligned}\tilde{c}_{\leq j-1}(2\tau_0) &= O\left(\frac{1}{\sigma^{O(1)}} T^{A_j^+} e^{-2\sqrt{3}T}\right), \\ \tilde{c}_j^-(2\tau_0) &= \sigma + O\left(\frac{1}{\sigma^{O(1)}} T^{A_j^+} e^{-\sqrt{3}T}\right), \\ \tilde{c}_j^+(2\tau_0) &= O\left(\frac{1}{\sigma^{O(1)}} T^{2A_j^+} e^{-4\sqrt{3}T}\right), \\ \tilde{c}_{\geq j+2}(2\tau_0) &= \omega e^{-2\sqrt{3}T} \hat{G}_{\geq j+2}(2\tau_0) a_{\geq j+2} + O\left(\frac{1}{\sigma^{O(1)}} T^{2A_j^+} e^{-3\sqrt{3}T}\right).\end{aligned}$$

This, together with (A-14), shows that the outgoing target (M_j^+, d_j^+, r_j^+) covers the next incoming target $(M_{j+1}^-, d_{j+1}^-, r_{j+1}^-)$ (it is enough to choose $a_{\geq j+2}$ appropriately).

Appendix B: Proof of Lemma 4.1

Proof. First note that, since $a(0) = g(0)$ is assumed to be compactly supported, the solution $a(t)$ to (1-4) exists globally in time, is smooth with respect to time, and is in $\ell^1(\mathbb{Z}^2)$ in space. Write

$$F(t) := -i \int_0^t \mathcal{E}(\tau) d\tau \quad \text{and} \quad d(t) := g(t) + F(t).$$

Observe that

$$-i\dot{d} = \mathcal{N}(d - F, d - F, d - F, d - F, d - F),$$

and $g = O_{\ell^1}(B^{-1})$ and $F = O_{\ell^1}(B^{-1-\sigma})$ by hypothesis. In particular we have $d = O_{\ell^1}(B^{-1})$. By multilinearity and (4-2) we thus have

$$-i\dot{d} = \mathcal{N}(d, d, d, d, d) + O_{\ell^1}(B^{-5-\sigma}). \tag{B-1}$$

Now write $e := a - d$ and recall that a is the solution of the NLS. Then we have

$$-i(\dot{d} + \dot{e}) = \mathcal{N}(d + e, d + e, d + e, d + e, d + e). \tag{B-2}$$

Subtracting (B-2) from (B-1) (and using (4-2)) we get

$$i\dot{e} = O_{\ell^1}(B^{-5-\sigma}) + O_{\ell^1}(B^{-4}\|e\|_1) + O_{\ell^1}(\|e\|_1^5),$$

so, taking the ℓ^1 norm and differentiating in time, we have

$$\frac{d}{dt}\|e\|_1 \lesssim B^{-5-\sigma} + B^{-4}\|e\|_1 + \|e\|_1^5.$$

We make the bootstrap assumption that $\|e\|_1 = O(B^{-1})$ for all $t \in [0, T]$, so that one can absorb the third term on the right-hand side into the second. Gronwall's inequality then gives

$$\|e\|_1 \leq B^{-1-\sigma} \exp(CB^{-4}t)$$

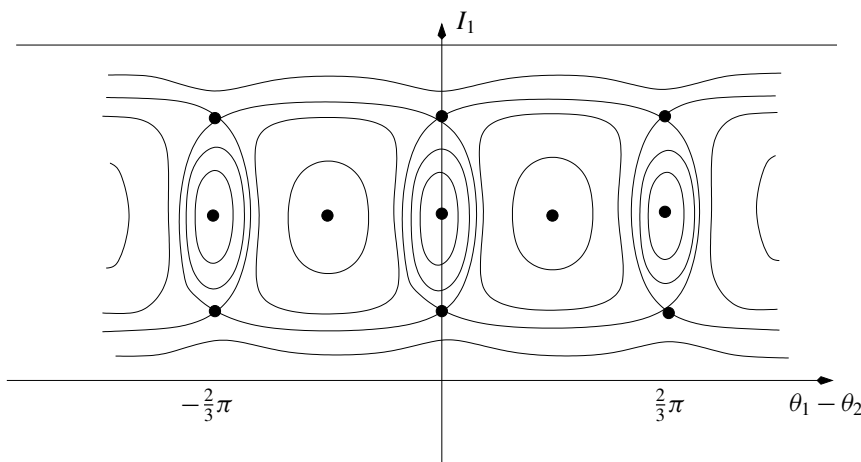


Figure 4. The phase portrait of H on the subspace $I_1 + I_2 = 1$; the dynamical variables are clearly $I_1, \theta_1 - \theta_2$.

for all $t \in [0, T]$. Since $T \ll B^4 \log B$, we have $\|e\|_1 \ll B^{-1-\sigma/2}$ and the result follows by the bootstrap argument. □

The result of Lemma 4.1 is that $g(t)$ is a good approximation of a solution to (1-4) on a time interval of approximate length $B^4 \log B$, a factor $\log B$ larger than the interval $[0, B^4]$ for which the solution is controlled by a straightforward local-in-time argument. We choose the exponent $\sigma/2$ for concreteness, but it could be replaced by any exponent between 0 and σ .

Appendix C: Two-generation sets without full energy transmission

We describe the dynamics associated to the sets $\mathfrak{S}^{(2)}, \mathfrak{S}^{(3)}$ given in the introduction.

In $\mathfrak{S}^{(2)}$ we have six complex variables $\beta_k, k \in \mathfrak{S}^{(2)}$ and correspondingly six constants of motion, so that the system is integrable. Passing to symplectic polar coordinates $\beta_k = \sqrt{J_k} e^{i\theta_k}$, we find that $J_{k_1} - J_{k_2}, J_{k_1} - J_{k_3}, J_{k_4} - J_{k_5}$ and $J_{k_4} - J_{k_6}$ are constant in time. Then one can study the dynamics reduced to the invariant subspace where all these constants are zero. We are left with four degrees of freedom, denoted by $I_1, I_2, \theta_1, \theta_2$, and the Hamiltonian

$$H = 31(I_1 + I_2)^3 - 66I_1 I_2 (I_1 + I_2) + 24I_1^{3/2} I_2^{3/2} \cos(3(\theta_1 - \theta_2))$$

Then we reduce to the subspace⁵ where $I_1 + I_2 = 1$, and get the phase portrait of Figure 4. It is evident from the picture that there is no orbit connecting $I_1 = 0$ to $I_1 = 1$. One could argue that this is due to our choice of invariant subspace. However, if we set, for instance, $J_{k_1} \neq J_{k_2}$, then we cannot transfer all the mass to k_4, k_5, k_6 since this would imply $J_{k_1} = J_{k_2} = J_{k_3} = 0$.

The case of $\mathfrak{S}^{(3)}$ is discussed in detail in [Grébert and Thomann 2012]. Proceeding as above, one gets the phase portrait of Figure 5. One could generalize this approach by taking two complete and

⁵This subspace is invariant due to the conservation of mass.

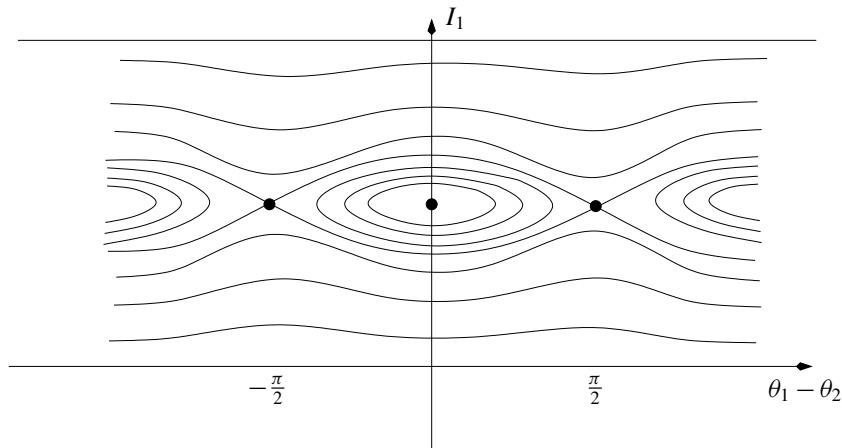


Figure 5. The phase portrait of $\mathfrak{G}^{(3)}$.

action-preserving sets \mathcal{S}_1 , \mathcal{S}_2 and connecting them with resonances as $\mathfrak{G}^{(2)}$ or $\mathfrak{G}^{(3)}$, as we have discussed in introduction for $\mathfrak{G}^{(1)}$. However, the dynamics is in fact qualitatively the same and one does not have full energy transfer.

We have experimented also with higher-order NLS equations. We have not performed a complete classification but it appears that the sets $\mathfrak{G}^{(2)}$, $\mathfrak{G}^{(3)}$ never give full energy transfer, while $\mathfrak{G}^{(1)}$ does.

Acknowledgement

This research was supported by the European Research Council under FP7, project “Hamiltonian PDEs and small divisor problem: a dynamical systems approach” (no. 306414 HamPDEs).

References

- [Bambusi 1997] D. Bambusi, “Long time stability of some small amplitude solutions in nonlinear Schrödinger equations”, *Comm. Math. Phys.* **189**:1 (1997), 205–226. MR 98m:35188 Zbl 0895.35095
- [Bambusi 1999] D. Bambusi, “Nekhoroshev theorem for small amplitude solutions in nonlinear Schrödinger equations”, *Math. Z.* **230**:2 (1999), 345–387. MR 2000h:35146 Zbl 0928.35160
- [Bambusi 2003] D. Bambusi, “Birkhoff normal form for some nonlinear PDEs”, *Comm. Math. Phys.* **234**:2 (2003), 253–285. MR 2003k:37121 Zbl 1032.37051
- [Bambusi and Grébert 2006] D. Bambusi and B. Grébert, “Birkhoff normal form for partial differential equations with tame modulus”, *Duke Math. J.* **135**:3 (2006), 507–567. MR 2007j:37124 Zbl 1110.37057
- [Berti and Biasco 2011] M. Berti and L. Biasco, “Branching of Cantor manifolds of elliptic tori and applications to PDEs”, *Comm. Math. Phys.* **305**:3 (2011), 741–796. MR 2012e:37152 Zbl 1230.37092
- [Berti and Bolle 2013] M. Berti and P. Bolle, “Quasi-periodic solutions with Sobolev regularity of NLS on \mathbb{T}^d with a multiplicative potential”, *J. Eur. Math. Soc. (JEMS)* **15**:1 (2013), 229–286. MR 2998835 Zbl 1260.35196
- [Berti et al. 2015] M. Berti, L. Corsi, and M. Procesi, “An abstract Nash–Moser theorem and quasi-periodic solutions for NLW and NLS on compact Lie groups and homogeneous manifolds”, *Comm. Math. Phys.* **334**:3 (2015), 1413–1454. MR 3312439 Zbl 06410173
- [Bourgain 1993] J. Bourgain, “Fourier transform restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations, I: Schrödinger equations”, *Geom. Funct. Anal.* **3**:2 (1993), 107–156. MR 95d:35160a Zbl 0787.35097

- [Bourgain 1996a] J. Bourgain, “Construction of approximative and almost periodic solutions of perturbed linear Schrödinger and wave equations”, *Geom. Funct. Anal.* **6**:2 (1996), 201–230. MR 97f:35013 Zbl 0872.35007
- [Bourgain 1996b] J. Bourgain, “On the growth in time of higher Sobolev norms of smooth solutions of Hamiltonian PDE”, *Internat. Math. Res. Notices* **6** (1996), 277–304. MR 97k:35016 Zbl 0934.35166
- [Bourgain 1998] J. Bourgain, “Quasi-periodic solutions of Hamiltonian perturbations of 2D linear Schrödinger equations”, *Ann. of Math. (2)* **148**:2 (1998), 363–439. MR 2000b:37087 Zbl 0928.35161
- [Bourgain 2000] J. Bourgain, “On diffusion in high-dimensional Hamiltonian systems and PDE”, *J. Anal. Math.* **80** (2000), 1–35. MR 2001j:35134 Zbl 0964.35143
- [Burq et al. 2004] N. Burq, P. Gérard, and N. Tzvetkov, “Strichartz inequalities and the nonlinear Schrödinger equation on compact manifolds”, *Amer. J. Math.* **126**:3 (2004), 569–605. MR 2005h:58036 Zbl 1067.58027
- [Chierchia and Perfetti 1995] L. Chierchia and P. Perfetti, “Second order Hamiltonian equations on T^∞ and almost-periodic solutions”, *J. Differential Equations* **116**:1 (1995), 172–201. MR 96d:58119 Zbl 0821.34043
- [Colliander et al. 2010] J. Colliander, M. Keel, G. Staffilani, H. Takaoka, and T. Tao, “Transfer of energy to high frequencies in the cubic defocusing nonlinear Schrödinger equation”, *Invent. Math.* **181**:1 (2010), 39–113. MR 2011f:35320 Zbl 1197.35265
- [Eliasson and Kuksin 2010] L. H. Eliasson and S. B. Kuksin, “KAM for the nonlinear Schrödinger equation”, *Ann. of Math. (2)* **172**:1 (2010), 371–435. MR 2011g:37203 Zbl 1201.35177
- [Faou et al. 2013] E. Faou, L. Gauckler, and C. Lubich, “Sobolev stability of plane wave solutions to the cubic nonlinear Schrödinger equation on a torus”, *Comm. Partial Differential Equations* **38**:7 (2013), 1123–1140. MR 3169740 Zbl 1274.35357
- [Geng et al. 2011] J. Geng, X. Xu, and J. You, “An infinite dimensional KAM theorem and its application to the two dimensional cubic Schrödinger equation”, *Adv. Math.* **226**:6 (2011), 5361–5402. MR 2012k:37159 Zbl 1213.37104
- [Gérard and Grellier 2010] P. Gérard and S. Grellier, “The cubic Szegő equation”, *Ann. Sci. Éc. Norm. Supér. (4)* **43**:5 (2010), 761–810. MR 2012b:37188 Zbl 1228.35225
- [Grébert and Thomann 2012] B. Grébert and L. Thomann, “Resonant dynamics for the quintic nonlinear Schrödinger equation”, *Ann. Inst. H. Poincaré Anal. Non Linéaire* **29**:3 (2012), 455–477. MR 2926244 Zbl 1259.37045
- [Grébert et al. 2009a] B. Grébert, R. Imekraz, and É. Paturel, “Normal forms for semilinear quantum harmonic oscillators”, *Comm. Math. Phys.* **291**:3 (2009), 763–798. MR 2010f:81119 Zbl 1185.81073
- [Grébert et al. 2009b] B. Grébert, T. Kappeler, and J. Pöschel, “Normal form theory for the NLS equation”, preprint, 2009. arXiv 0907.39381
- [Guardia and Kaloshin 2015] M. Guardia and V. Kaloshin, “Growth of Sobolev norms in the cubic defocusing nonlinear Schrödinger equation”, *J. Eur. Math. Soc. (JEMS)* **17**:1 (2015), 71–149. MR 3312404 Zbl 06419398
- [Hani 2014] Z. Hani, “Long-time instability and unbounded Sobolev orbits for some periodic nonlinear Schrödinger equations”, *Arch. Ration. Mech. Anal.* **211**:3 (2014), 929–964. MR 3158811 Zbl 1293.35298
- [Hani et al. 2013] Z. Hani, B. Pausader, N. Tzvetkov, and N. Visciglia, “Modified scattering for the cubic Schrödinger equation on product spaces and applications”, preprint, 2013. arXiv 1311.2275
- [Haus and Thomann 2013] E. Haus and L. Thomann, “Dynamics on resonant clusters for the quintic nonlinear Schrödinger equation”, *Dyn. Partial Differ. Equ.* **10**:2 (2013), 157–169. MR 3071705 Zbl 06225369
- [Kuksin 1995] S. B. Kuksin, “On squeezing and flow of energy for nonlinear wave equations”, *Geom. Funct. Anal.* **5**:4 (1995), 668–701. MR 96d:35091 Zbl 0834.35086
- [Kuksin 1996] S. B. Kuksin, “Growth and oscillations of solutions of nonlinear Schrödinger equation”, *Comm. Math. Phys.* **178**:2 (1996), 265–280. MR 97f:35203 Zbl 0862.35112
- [Kuksin 1997a] S. B. Kuksin, “On turbulence in nonlinear Schrödinger equations”, *Geom. Funct. Anal.* **7**:4 (1997), 783–822. MR 2001a:35197 Zbl 0912.35143
- [Kuksin 1997b] S. B. Kuksin, “Oscillations in space-periodic nonlinear Schrödinger equations”, *Geom. Funct. Anal.* **7**:2 (1997), 338–363. MR 98e:35157 Zbl 0874.35113
- [Kuksin 1999] S. B. Kuksin, “Spectral properties of solutions for nonlinear PDEs in the turbulent regime”, *Geom. Funct. Anal.* **9**:1 (1999), 141–184. MR 2000c:35223 Zbl 0929.35145

- [Kuksin and Pöschel 1996] S. Kuksin and J. Pöschel, “Invariant Cantor manifolds of quasi-periodic oscillations for a nonlinear Schrödinger equation”, *Ann. of Math. (2)* **143**:1 (1996), 149–179. MR 96j:58147 Zbl 0847.35130
- [Pocovnicu 2011] O. Pocovnicu, “Explicit formula for the solution of the Szegő equation on the real line and applications”, *Discrete Contin. Dyn. Syst.* **31**:3 (2011), 607–649. MR 2012h:35330 Zbl 1235.35263
- [Pocovnicu 2013] O. Pocovnicu, “First and second order approximations for a nonlinear wave equation”, *J. Dynam. Differential Equations* **25**:2 (2013), 305–333. MR 3054639 Zbl 1270.65060
- [Pöschel 1996] J. Pöschel, “A KAM-theorem for some nonlinear partial differential equations”, *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)* **23**:1 (1996), 119–148. MR 97g:58146 Zbl 0870.34060
- [Pöschel 2002] J. Pöschel, “On the construction of almost periodic solutions for a nonlinear Schrödinger equation”, *Ergodic Theory Dynam. Systems* **22**:5 (2002), 1537–1549. MR 2005a:37142 Zbl 1020.37044
- [Procesi and Procesi 2012] M. Procesi and C. Procesi, “A normal form for the Schrödinger equation with analytic non-linearities”, *Comm. Math. Phys.* **312**:2 (2012), 501–557. MR 2917174 Zbl 1277.35318
- [Procesi and Procesi 2015] C. Procesi and M. Procesi, “A KAM algorithm for the resonant non-linear Schrödinger equation”, *Adv. Math.* **272** (2015), 399–470. MR 3303238 Zbl 06391241
- [Procesi and Xu 2013] M. Procesi and X. Xu, “Quasi-Töplitz functions in KAM theorem”, *SIAM J. Math. Anal.* **45**:4 (2013), 2148–2181. MR 3072759 Zbl 1304.37056
- [Staffilani 1997] G. Staffilani, “On the growth of high Sobolev norms of solutions for KdV and Schrödinger equations”, *Duke Math. J.* **86**:1 (1997), 109–142. MR 98b:35192 Zbl 0874.35114
- [Wang 2008] W.-M. Wang, “Logarithmic bounds on Sobolev norms for time dependent linear Schrödinger equations”, *Comm. Partial Differential Equations* **33**:10-12 (2008), 2164–2179. MR 2009k:35247 Zbl 1154.35450
- [Wang 2010] W.-M. Wang, “Supercritical nonlinear Schrödinger equations, II: Almost global existence”, preprint, 2010. arXiv 1007.0154
- [Wang 2014] W.-M. Wang, “Supercritical nonlinear Schrödinger equations: quasi-periodic solutions”, preprint, 2014. arXiv 1007.0156
- [Wayne 1990] C. E. Wayne, “Periodic and quasi-periodic solutions of nonlinear wave equations via KAM theory”, *Comm. Math. Phys.* **127**:3 (1990), 479–528. MR 91b:58236 Zbl 0708.35087

Received 9 Jun 2014. Revised 19 Jan 2015. Accepted 6 Mar 2015.

EMANUELE HAUS: emanuele.haus@unina.it

Dipartimento di Matematica e Applicazioni “R. Caccioppoli”, Università di Napoli “Federico II”, Via Cintia, Monte S. Angelo, I-80126 Napoli, Italy

MICHELA PROCESI: mprocesi@mat.uniroma1.it

Dipartimento di Matematica, Università di Roma “La Sapienza”, Piazzale Aldo Moro, 5, I-00185 Roma, Italy

POWER SPECTRUM OF THE GEODESIC FLOW ON HYPERBOLIC MANIFOLDS

SEMYON DYATLOV, FRÉDÉRIC FAURE AND COLIN GUILLARMOU

We describe the complex poles of the power spectrum of correlations for the geodesic flow on compact hyperbolic manifolds in terms of eigenvalues of the Laplacian acting on certain natural tensor bundles. These poles are a special case of Pollicott–Ruelle resonances, which can be defined for general Anosov flows. In our case, resonances are stratified into bands by decay rates. The proof also gives an explicit relation between resonant states and eigenstates of the Laplacian.

1. Introduction	923
2. Outline and structure	929
3. Geometry of the hyperbolic space	933
4. Horocyclic operators	941
5. Pollicott–Ruelle resonances	952
6. Properties of the Laplacian	966
7. Expansions of eigenstates of the Laplacian	978
Appendix A. Some technical calculations	988
Appendix B. The special case of dimension 2	993
Appendix C. Eigenvalue asymptotics for symmetric tensors	994
Acknowledgements	997
References	998

1. Introduction

In this paper, we consider the characteristic frequencies of correlations,

$$\rho_{f,g}(t) = \int_{SM} (f \circ \varphi_{-t}) \cdot \bar{g} \, d\mu, \quad f, g \in C^\infty(SM), \quad (1-1)$$

for the geodesic flow φ_t on a compact hyperbolic manifold M of dimension $n + 1$ (that is, M has constant sectional curvature -1). Here φ_t acts on SM , the unit tangent bundle of M , and μ is the natural smooth probability measure. Such φ_t are classical examples of *Anosov flows*; for this family of examples, we are able to prove much more precise results than in the general Anosov case.

An important question, expanding on the notion of mixing, is the behavior of $\rho_{f,g}(t)$ as $t \rightarrow +\infty$. Following [Ruelle 1986], we take the *power spectrum*, which in our convention is the Laplace transform $\hat{\rho}_{f,g}(\lambda)$ of $\rho_{f,g}$ restricted to $t > 0$. The long-time behavior of $\rho_{f,g}(t)$ is related to the properties of

MSC2010: 37D40.

Keywords: Pollicott–Ruelle resonances, hyperbolic manifolds.

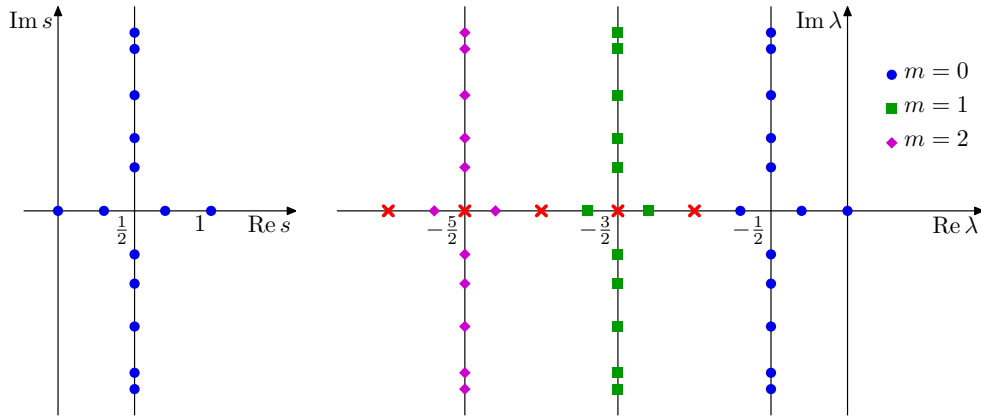


Figure 1. An illustration of Theorem 1, with eigenvalues of the Laplacian on the left and the resonances of geodesic flow on the right. The red crosses mark exceptional points where the theorem does not apply.

the meromorphic extension of $\hat{\rho}_{f,g}(\lambda)$ to the entire complex plane. The poles of this extension, called *Pollicott–Ruelle resonances* (see [Pollicott 1986; Ruelle 1986; Faure and Sjöstrand 2011] and (1-7) below), are the complex characteristic frequencies of $\rho_{f,g}$, describing its decay and oscillation and not depending on f, g .

For the case of dimension $n + 1 = 2$, the following connection between resonances and the spectrum of the Laplacian was announced in [Faure and Tsujii 2013b, Section 4] (see [Flaminio and Forni 2003] for a related result and the remarks below regarding the zeta function techniques).

Theorem 1. *Assume that M is a compact hyperbolic surface ($n = 1$) and the spectrum of the positive Laplacian on M is (see Figure 1)*

$$\text{Spec}(\Delta) = \{s_j(1 - s_j)\}, \quad s_j \in [0, 1] \cup \left(\frac{1}{2} + i\mathbb{R}\right).$$

Then Pollicott–Ruelle resonances for the geodesic flow on SM in $\mathbb{C} \setminus (-1 - \frac{1}{2}\mathbb{N}_0)$ are

$$\lambda_{j,m} = -m - 1 + s_j, \quad m \in \mathbb{N}_0. \tag{1-2}$$

Remark. We use the Laplace transform (which has poles in the left half-plane) rather than the Fourier transform as in [Ruelle 1986; Faure and Sjöstrand 2011] to simplify the relation to the parameter s used for Laplacians on hyperbolic manifolds.

Our main result concerns the case of higher dimensions $n + 1 > 2$. The situation is considerably more involved than in the case of Theorem 1, featuring the spectrum of the Laplacian on certain tensor bundles. More precisely, for $\sigma \in \mathbb{R}$, denote

$$\text{Mult}_\Delta(\sigma, m) := \dim \text{Eig}^m(\sigma),$$

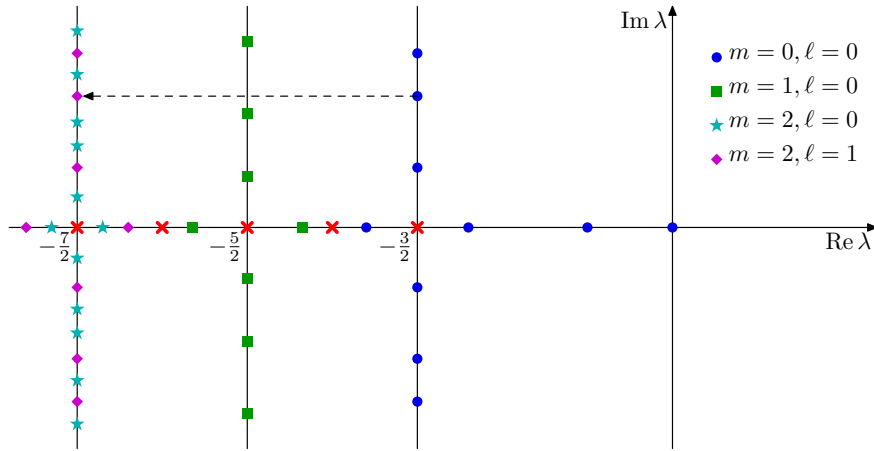


Figure 2. An illustration of Theorem 2 for $n = 3$. The red crosses mark exceptional points where the theorem does not apply. Note that the points with $m = 2, \ell = 1$ are simply the points with $m = 0, \ell = 0$ shifted by -2 (modulo exceptional points), as illustrated by the arrow.

where $\text{Eig}^m(\sigma)$, defined in (5-19), is the space of *trace-free, divergence-free symmetric sections* of $\otimes^m T^*M$ satisfying $\Delta f = \sigma f$. Denote by $\text{Mult}_R(\lambda)$ the geometric multiplicity of λ as a Pollicott–Ruelle resonance of the geodesic flow on M (see Theorem 3 and the remarks preceding it for a definition).

Theorem 2. *Let M be a compact hyperbolic manifold of dimension $n + 1 \geq 2$. Assume $\lambda \in \mathbb{C} \setminus (-\frac{1}{2}n - \frac{1}{2}\mathbb{N}_0)$. Then, for $\lambda \notin -2\mathbb{N}$, we have (see Figure 2)*

$$\text{Mult}_R(\lambda) = \sum_{m \geq 0} \sum_{\ell=0}^{\lfloor m/2 \rfloor} \text{Mult}_\Delta(-(\lambda + m + \frac{1}{2}n)^2 + \frac{1}{4}n^2 + m - 2\ell, m - 2\ell) \tag{1-3}$$

and, for $\lambda \in -2\mathbb{N}$, we have

$$\text{Mult}_R(\lambda) = \sum_{\substack{m \geq 0 \\ m \neq -\lambda}} \sum_{\ell=0}^{\lfloor m/2 \rfloor} \text{Mult}_\Delta(-(\lambda + m + \frac{1}{2}n)^2 + \frac{1}{4}n^2 + m - 2\ell, m - 2\ell). \tag{1-4}$$

Remark. (i) If $\text{Mult}_\Delta(-(\lambda + m + \frac{1}{2}n)^2 + \frac{1}{4}n^2 + m - 2\ell, m - 2\ell) > 0$, then Lemma 6.1 and the fact that $\Delta \geq 0$ on functions imply that either $\lambda \in -m - \frac{1}{2}n + i\mathbb{R}$ or

$$\begin{aligned} \lambda &\in [-1 - m, -m] && \text{if } n = 1, m > 2\ell, \\ \lambda &\in [1 - n - m, -1 - m] && \text{if } n > 1, m > 2\ell, \\ \lambda &\in [-n - m, -m] && \text{if } m = 2\ell. \end{aligned} \tag{1-5}$$

In particular, we confirm that resonances lie in $\{\text{Re } \lambda \leq 0\}$ and the only resonance on the imaginary axis is $\lambda = 0$ with $\text{Mult}_R(0) = 1$, corresponding to $m = \ell = 0$. We call the set of resonances corresponding

to some m the m -th band. This is a special case of the band structure for general contact Anosov flows established in the work of Faure and Tsujii [2013a; 2013b; 2014].

(ii) The case $n = 1$ fits into Theorem 2 as follows: for $m \geq 2$, the spaces $\text{Eig}^m(\sigma)$ are trivial unless σ is an exceptional point (since the corresponding spaces $\text{Bd}^{m,0}(\lambda)$ of Lemma 5.6 would have to be trace-free sections of a one-dimensional vector bundle), and the spaces $\text{Eig}^1(\sigma + 1)$ and $\text{Eig}^0(\sigma)$ are isomorphic, as shown in Appendix C2.

(iii) The band with $m = 0$ corresponds to the spectrum of the scalar Laplacian; the band with $m = 1$ corresponds to the spectrum of the Hodge Laplacian on coclosed 1-forms; see Appendix C2.

(iv) As seen from (1-3) and (1-4), for $m \geq 2$ the m -th band of resonances contains shifted copies of bands $m - 2, m - 4, \dots$. The special case (1-4) means that the resonance 0 of the $m = 0$ band is not copied to other bands.

(v) A Weyl law holds for the spaces $\text{Eig}^m(\sigma)$; see Appendix C1. It implies the following Weyl law for resonances in the m -th band:

$$\sum_{\lambda \in -n/2 - m + i[-R, R]} \text{Mult}_R(\lambda) = \frac{2^{-n} \pi^{-(n+1)/2}}{\Gamma(\frac{1}{2}(n+3))} \cdot \frac{(m+n-1)!}{m!(n-1)!} \text{Vol}(M) R^{n+1} + \mathcal{O}(R^n). \quad (1-6)$$

The power R^{n+1} agrees with the Weyl law of [Faure and Tsujii 2013b, (5.3)] and with the earlier upper bound of [Datchev et al. 2014]. We also see that, if $n > 1$, then each m and $\ell \in [0, \frac{1}{2}m]$ produce a nontrivial contribution to the set of resonances. The factor $(m+n-1)!/m!(n-1)!$ is the dimension of the space of homogeneous polynomials of order m in n variables; it is natural in light of [Faure and Tsujii 2013a, Proposition 5.11], which locally reduces resonances to such polynomials.

The proof of Theorem 2 is outlined in Section 2. We use in particular the microlocal method of Faure and Sjöstrand [2011], defining Pollicott–Ruelle resonances as the points $\lambda \in \mathbb{C}$ for which the (unbounded nonselfadjoint) operator

$$X + \lambda : \mathcal{H}^r \rightarrow \mathcal{H}^r, \quad r > -C_0 \text{Re } \lambda, \quad (1-7)$$

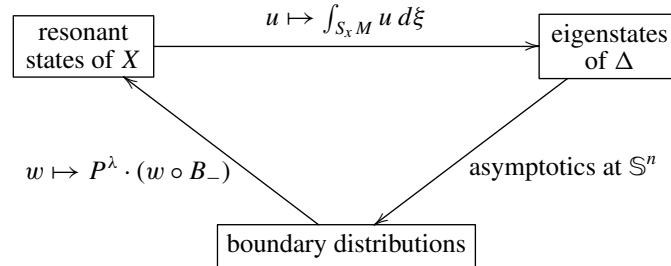
is not invertible. Here X is the vector field on SM generating the geodesic flow, so that $\varphi_t = e^{tX}$, \mathcal{H}^r is a certain *anisotropic Sobolev space*, and C_0 is a fixed constant independent of r ; see Section 5A for details. Resonances do not depend on the choice of r . The relation to correlations (1-1) is given by the formula

$$\hat{\rho}_{f,g}(\lambda) = \int_0^\infty e^{-\lambda t} \rho_{f,g}(t) dt = \int_0^\infty e^{-\lambda t} \langle e^{-tX} f, g \rangle dt = \langle (X + \lambda)^{-1} f, g \rangle_{L^2(SM)},$$

valid for $\text{Re } \lambda > 0$ and $f, g \in C^\infty(SM)$. See also Theorem 4 below.

We stress that our method provides an *explicit relation between classical and quantum states*, that is, between Pollicott–Ruelle resonant states (elements of the kernel of (1-7)) and eigenstates of the Laplacian; namely, in addition to the poles of $\hat{\rho}_{f,g}(\lambda)$, we describe its residues. For instance, for the $m = 0$ band, if $u(x, \xi), x \in M, \xi \in S_x M$, is a resonant state, then the corresponding eigenstate of the Laplacian, $f(x)$, is obtained by integration of u along the fibers $S_x M$; see (2-3). On the other hand, to obtain u from f one needs to take the *boundary distribution* w of f , which is a distribution on the conformal boundary \mathbb{S}^n of

the hyperbolic space \mathbb{H}^{n+1} appearing as the leading coefficient of a weak asymptotic expansion at \mathbb{S}^n of the lift of f to \mathbb{H}^{n+1} . Then u is described by w via an explicit formula, (2-4); this formula features the Poisson kernel P and the map $B_- : S\mathbb{H}^{n+1} \rightarrow \mathbb{S}^n$ mapping a tangent vector to the endpoint in negative infinite time of the corresponding geodesic of \mathbb{H}^{n+1} . The explicit relation can be schematically described as follows:



For $m > 0$, one needs to also use *horocyclic differential operators*; see Section 2.

Theorem 2 used the notion of *geometric multiplicity* of a resonance λ , that is, the dimension of the kernel of $X + \lambda$ on \mathcal{H}^r . For nonselfadjoint problems, it is often more natural to consider the *algebraic multiplicity*, the dimension of the space of elements of \mathcal{H}^r which are killed by some power of $X + \lambda$.

Theorem 3. *If $\lambda \notin -\frac{1}{2}n - \frac{1}{2}\mathbb{N}_0$, then the algebraic and geometric multiplicities of λ as a Pollicott–Ruelle resonance coincide.*

Theorem 3 relies on a *pairing formula* (Lemma 5.10), which states that

$$\langle u, u^* \rangle_{L^2(SM)} = F_{m,\ell}(\lambda) \langle f, f^* \rangle_{L^2(M; \otimes^{m-2\ell} T^*M)},$$

where u is a resonant state at some resonance λ corresponding to some m, ℓ in Theorem 2, u^* is a coresonant state (that is, an element of the kernel of the adjoint of $(X + \lambda)$), f, f^* are the corresponding eigenstates of the Laplacian, and $F_{m,\ell}(\lambda)$ is an explicit function. Here $\langle u, u^* \rangle_{L^2}$ refers to the integral $\int u \bar{u}^*$, which is well-defined despite the fact that neither u nor u^* lie in L^2 ; see (5-6). This pairing formula is of independent interest as a step towards understanding the high frequency behavior of resonant states and attempting to prove *quantum ergodicity of resonant states* in the present setting. Anantharaman and Zelditch [2007] obtained the pairing formula in dimension 2 and studied concentration of Patterson–Sullivan distributions, which are directly related to resonant states; see also [Hansen et al. 2012].

To motivate the study of Pollicott–Ruelle resonances, we also apply to our setting the following *resonance expansion*, proved by Tsujii [2010, Corollary 1.2] and Nonnenmacher and Zworski [2015, Corollary 5]:

Theorem 4. *Fix $\varepsilon > 0$. Then, for N large enough and f, g in the Sobolev space $H^N(SM)$,*

$$\rho_{f,g}(t) = \int f d\mu \int g d\mu + \sum_{\lambda \in (-n/2, 0)} \sum_{k=1}^{\text{Mult}_R(\lambda)} e^{\lambda t} \langle f, u_{\lambda,k}^* \rangle_{L^2} \langle u_{\lambda,k}, g \rangle_{L^2} + \mathcal{O}_{f,g}(e^{-(n/2-\varepsilon)t}), \quad (1-8)$$

where $u_{\lambda,k}$ is any basis of the space of resonant states associated to λ and $u_{\lambda,k}^*$ is the dual basis of the space of coresonant states (so that $\sum_k u_{\lambda,k} \otimes_{L^2} u_{\lambda,k}^*$ is the spectral projector of $-X$ at λ).

Here we use Theorem 3 to see that there are no powers of t in the expansion and that there exists the dual basis of coresonant states to a basis of resonant states.

Combined with Theorem 2, the expansion (1-8) in particular gives the optimal exponent in the decay of correlations in terms of the small eigenvalues of the Laplacian; more precisely, the difference between $\rho_{f,g}(t)$ and the product of the integrals of f and g is $\mathcal{O}(e^{-\nu_0 t})$, where

$$\nu_0 = \min_{0 \leq m < n/2} \min \{ \nu + m \mid \nu \in (0, \frac{1}{2}n - m), \nu(n - \nu) + m \in \text{Spec}^m(\Delta) \},$$

or $\mathcal{O}(e^{-(n/2-\varepsilon)t})$ for each $\varepsilon > 0$ if the set above is empty. Here $\text{Spec}^m(\Delta)$ denotes the spectrum of the Laplacian on trace-free, divergence-free symmetric tensors of order m . Using (1-5), we see that in fact one has $\nu \in [1, \frac{1}{2}n - m)$ for $m > 0$.

In order to go beyond the $\mathcal{O}(e^{-(n/2-\varepsilon)t})$ remainder in (1-8), one would need to handle the infinitely many resonances in the $m = 0$ band. This is thought to be impossible in the general context of scattering theory, as the scattering resolvent can grow exponentially near the bands; however, there exist cases, such as Kerr–de Sitter black holes, where a resonance expansion with infinitely many terms holds; see [Bony and Häfner 2008; Dyatlov 2012]. The case of black holes is somewhat similar to the one considered here, because in both cases the trapped set is normally hyperbolic; see [Dyatlov 2015; Faure and Tsujii 2014]. What is more, one can try to prove a resonance expansion with remainder $\mathcal{O}(e^{-(n/2+1-\varepsilon)t})$, where the sum over resonances in the first band is replaced by $\langle (\Pi_0 f) \circ \varphi^{-t}, g \rangle$ and Π_0 is the projector onto the space of resonant states with $m = 0$, having the microlocal structure of a Fourier integral operator — see [Dyatlov 2015] for a similar result in the context of black holes.

Previous results. In the constant curvature setting in dimension $n+1=2$, the spectrum of the geodesic flow on L^2 was studied by Fomin and Gelfand [1952] using representation theory. An exponential rate of mixing was proved by Ratner [1987] and it was extended to higher dimensions by Moore [1987]. In variable negative curvature for surfaces and, more generally, for Anosov flows with stable/unstable jointly nonintegrable foliations, exponential decay of correlations was first shown by Dolgopyat [1998] and then by Liverani [2004] for contact flows. The work of Tsujii [2010; 2012] established the asymptotic size of the resonance-free strip and the work of Nonnenmacher and Zworski [2015] extended this result to general normally hyperbolic trapped sets. Faure and Tsujii [2013a; 2013b; 2014] established the band structure for general smooth contact Anosov flows and proved an asymptotic for the number of resonances in the first band.

In dimension 2, the study of resonant states in the first band ($m = 0$) — that is, distributions which lie in the spectrum of X and are annihilated by the horocyclic vector field U_- — appears already in the works of Guillemin [1977, Lecture 3] and Zelditch [1987], both using the representation theory of $\text{PSL}(2; \mathbb{R})$, albeit without explicitly interpreting them as Pollicott–Ruelle resonant states. A more general study of the elements in the kernel of U_- was performed by Flaminio and Forni [2003].

An alternative approach to resonances involves the *Selberg and Ruelle zeta functions*. The singularities (zeros and poles) of the Ruelle zeta function correspond to Pollicott–Ruelle resonances on differential forms (see [Fried 1986; 1995; Giulietti et al. 2013; Dyatlov and Zworski 2015]), while the singularities of the Selberg zeta function correspond to eigenvalues of the Laplacian. The Ruelle and Selberg zeta

functions are closely related; see [Leboeuf 2004, Section 5.1, Figure 1; Dyatlov and Zworski 2015, (1.2)] in dimension 2 and [Fried 1986; Bunke and Olbrich 1995, Proposition 3.4] in arbitrary dimensions. However, the Ruelle zeta function does not recover all resonances on functions, due to cancellations with singularities coming from differential forms of different orders. For example, [Juhl 2001, Theorem 3.7] describes the spectral singularities of the Ruelle zeta function for $n = 3$ in terms of the spectrum of the Laplacian on functions and 1-forms, which is much smaller than the set obtained in Theorem 2.

The book of Juhl [2001] and the works of Bunke and Olbrich [1995; 1997; 1999; 2001] study Ruelle and Selberg zeta functions corresponding to various representations of the orthogonal group. They also consider general locally symmetric spaces and address the question of what happens at the exceptional points (which in our case are contained in $-\frac{1}{2}n - \frac{1}{2}\mathbb{N}_0$), relating the behavior of the zeta functions at these points to topological invariants. It is possible that the results [Juhl 2001; Bunke and Olbrich 1995; 1997; 1999; 2001] together with an appropriate representation-theoretic calculation recover our description of resonances, even though no explicit description featuring the spectrum of the Laplacian on trace-free, divergence-free symmetric tensors as in (1-3), (1-4) seems to be available in the literature. The direct spectral approach used in this paper, unlike the zeta function techniques, gives an explicit relation between resonant states and eigenstates of the Laplacian (see the remarks following (1-7)) and is a step towards a more quantitative understanding of decay of correlations.

An essential component of our work is the analysis of the correspondence between eigenstates of the Laplacian on \mathbb{H}^{n+1} and distributions on the conformal infinity \mathbb{S}^n . In the scalar case, such a correspondence for hyperfunctions on \mathbb{S}^n is due to Helgason [1970; 1974] (see also [Minemura 1975]); the correspondence between tempered eigenfunctions of Δ and distributions (instead of hyperfunctions) was shown by Oshima and Sekiguchi [1980] and van den Ban and Schlichtkrull [1987] (see also [Grellier and Otal 2005]). Olbrich [1995] studied Poisson transforms on general homogeneous vector bundles, which include the bundles of tensors used in the present paper. The question of regularity of equivariant distributions on \mathbb{S}^n by certain Kleinian groups of isometries of \mathbb{H}^{n+1} (geometrically finite groups) is interesting since it determines the regularity of resonant states for the flow; precise regularity was studied by Otal [1998] in the 2-dimensional cocompact case, Grellier and Otal [2005] in higher dimensions, and Bunke and Olbrich [1999] for geometrically finite groups. In dimension 2, the correspondence between the eigenfunctions of the Laplacian on the hyperbolic plane and distributions on the conformal boundary \mathbb{S}^1 appeared in [Pollicott 1989; Bunke and Olbrich 1997]; it is also an important tool in the theory developed by [Bunke and Olbrich 2001] to study Selberg zeta functions on convex cocompact hyperbolic manifolds (see also [Juhl 2001] in the compact setting). These distributions on the conformal boundary \mathbb{S}^n , of Patterson–Sullivan type, are also the central object of the recent work of Anantharaman and Zelditch [2007; 2012] studying quantum ergodicity on hyperbolic compact surfaces; a generalization to higher-rank, locally symmetric spaces was provided by Hansen, Hilgert and Schröder [Hansen et al. 2012].

2. Outline and structure

In this section, we give the ideas of the proof of Theorem 2, first in dimension 2 and then in higher dimensions, and describe the structure of the paper.

2A. Dimension 2. We start by using the following criterion (Lemma 5.1): $\lambda \in \mathbb{C}$ is a Pollicott–Ruelle resonance if and only if the space

$$\text{Res}_X(\lambda) := \{u \in \mathcal{D}'(SM) \mid (X + \lambda)u = 0, \text{WF}(u) \subset E_u^*\}$$

is nontrivial. Here $\mathcal{D}'(SM)$ is the space of *distributions* on SM (see [Hörmander 1983]), $\text{WF}(u) \subset T^*(SM)$ is the *wavefront set* of u (see [Hörmander 1983, Chapter 8]), and $E_u^* \subset T^*(SM)$ is the dual unstable foliation described in (3-15). It is more convenient to use the condition $\text{WF}(u) \subset E_u^*$ rather than $u \in \mathcal{H}'$, because this condition is invariant under differential operators of any order.

The key tools for the proof are the *horocyclic vector fields* U_\pm on SM , pictured in Figure 3(a) below. To define them, we represent $M = \Gamma \backslash \mathbf{H}^2$, where $\mathbf{H}^2 = \{z \in \mathbb{C} \mid \text{Im } z > 0\}$ is the hyperbolic plane and $\Gamma \subset \text{PSL}(2; \mathbb{R})$ is a cocompact Fuchsian group of isometries acting by Möbius transformations. (See Appendix B for the relation of the notation we use in dimension 2, based on the half-plane model of the hyperbolic space, to the notation used elsewhere in the paper that is based on the hyperboloid model.) Then SM is covered by $S\mathbf{H}^2$, which is isomorphic to the group $G := \text{PSL}(2; \mathbb{R})$ by the map $\gamma \in G \mapsto (\gamma(i), d\gamma(i) \cdot i)$. Consider the left-invariant vector fields on G corresponding to the following elements of its Lie algebra:

$$X = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}, \quad U_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad U_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}; \tag{2-1}$$

then X, U_\pm descend to vector fields on SM , with X becoming the generator of the geodesic flow. We have the commutation relations

$$[X, U_\pm] = \pm U_\pm \quad \text{and} \quad [U_+, U_-] = 2X. \tag{2-2}$$

For each λ and $m \in \mathbb{N}_0$, define the spaces

$$V_m(\lambda) := \{u \in \mathcal{D}'(SM) \mid (X + \lambda)u = 0, U_-^m u = 0, \text{WF}(u) \subset E_u^*\},$$

and put

$$\text{Res}_X^0(\lambda) := V_1(\lambda).$$

By (2-2), $U_-^m(\text{Res}_X(\lambda)) \subset \text{Res}_X(\lambda + m)$. Since there are no Pollicott–Ruelle resonances in the right half-plane, we conclude that

$$\text{Res}_X(\lambda) = V_m(\lambda) \quad \text{for } m > -\text{Re } \lambda.$$

We now use the diagram (writing $\text{Id} = U_\pm^0, U_\pm = U_\pm^1$ for uniformity of notation)

$$\begin{array}{ccccccc} 0 = V_0(\lambda) & \xrightarrow{\iota} & V_1(\lambda) & \xrightarrow{\iota} & V_2(\lambda) & \xrightarrow{\iota} & V_3(\lambda) & \xrightarrow{\iota} & \dots, \\ & & \begin{array}{c} U_+^0 \uparrow \downarrow U_-^0 \\ \text{Res}_X^0(\lambda) \end{array} & & \begin{array}{c} U_+^1 \uparrow \downarrow U_-^1 \\ \text{Res}_X^0(\lambda + 1) \end{array} & & \begin{array}{c} U_+^2 \uparrow \downarrow U_-^2 \\ \text{Res}_X^0(\lambda + 2) \end{array} & & \\ & & & & & & & & \end{array}$$

where ι denotes the inclusion maps and, unless $\lambda \in -1 - \frac{1}{2}\mathbb{N}_0$, we have

$$V_{m+1}(\lambda) = V_m(\lambda) \oplus U_+^m(\text{Res}_X^0(\lambda + m)),$$

and U_+^m is one-to-one on $\text{Res}_X^0(\lambda + m)$; indeed, using (2-2) we calculate

$$U_-^m U_+^m = m! \left(\prod_{j=1}^m (2\lambda + m + j) \right) \text{Id} \quad \text{on } \text{Res}_X^0(\lambda + m)$$

and the coefficient above is nonzero when $\lambda \notin -1 - \frac{1}{2}\mathbb{N}_0$. We then see that

$$\text{Res}_X(\lambda) = \bigoplus_{m \geq 0} U_+^m(\text{Res}_X^0(\lambda + m)).$$

It remains to describe the space of resonant states in the first band,

$$\text{Res}_X^0(\lambda) = \{u \in \mathcal{D}'(SM) \mid (X + \lambda)u = 0, U_-u = 0, \text{WF}(u) \subset E_u^*\}.$$

We can remove the condition $\text{WF}(u) \subset E_u^*$ as it follows from the other two; see the remark following Lemma 5.6. We claim that the pushforward map

$$u \mapsto f(x) := \int_{S_x M} u(x, \xi) dS(\xi) \tag{2-3}$$

is an isomorphism from $\text{Res}_X^0(\lambda)$ onto $\text{Eig}(-\lambda(1 + \lambda))$, where $\text{Eig}(\sigma) = \{u \in C^\infty(M) \mid \Delta u = \sigma u\}$; this would finish the proof. In other words, the eigenstate of the Laplacian corresponding to u is obtained by integrating u over the fibers of SM .

To show that (2-3) is an isomorphism, we reduce the elements of $\text{Res}_X^0(\lambda)$ to the conformal boundary \mathbb{S}^1 of the ball model \mathbb{B}^2 of the hyperbolic space as follows:

$$\text{Res}_X^0(\lambda) = \{P(y, B_-(y, \xi))^\lambda w(B_-(y, \xi)) \mid w \in \text{Bd}(\lambda)\}, \tag{2-4}$$

where $P(y, v)$ is the Poisson kernel: $P(y, v) = (1 - |y|^2)/|y - v|^2$, $y \in \mathbb{B}^2$, $v \in \mathbb{S}^1$; $B_- : S\mathbb{B}^2 \rightarrow \mathbb{S}^1$ maps (y, ξ) to the limiting point of the geodesic $\varphi_t(y, \xi)$ as $t \rightarrow -\infty$ — see Figure 3(a) — and $\text{Bd}(\lambda) \subset \mathcal{D}'(\mathbb{S}^1)$ is the space of distributions satisfying a certain equivariance property with respect to Γ . Here we lifted $\text{Res}_X^0(\lambda)$ to distributions on $S\mathbb{H}^2$ and used the fact that the map B_- is invariant under both X and U_- ; see Lemma 5.6 for details.

It remains to show that the map $w \mapsto f$ defined via (2-3) and (2-4) is an isomorphism from $\text{Bd}(\lambda)$ to $\text{Eig}(-\lambda(1 + \lambda))$. This map is given by (see Lemma 6.6)

$$f(y) = \mathcal{P}_\lambda^- w(y) := \int_{\mathbb{S}^1} P(y, v)^{1+\lambda} w(v) dS(v) \tag{2-5}$$

and is the Poisson operator for the (scalar) Laplacian corresponding to the eigenvalue $s(1 - s)$, $s = 1 + \lambda$. This Poisson operator is known to be an isomorphism for $\lambda \notin -1 - \mathbb{N}$ — see the remark following Theorem 6 in Section 5B — finishing the proof.

2B. Higher dimensions. In higher dimensions, the situation is made considerably more difficult by the fact we can no longer define the vector fields U_{\pm} on SM . To get around this problem, we remark that, in dimension 2, U_-u is the derivative of u along a certain canonical vector in the one-dimensional *unstable foliation* $E_u \subset T(SM)$ and, similarly, U_+u is the derivative along an element of the stable foliation E_s ; see Section 4B. In dimension $n + 1 > 2$, the foliations E_u, E_s are n -dimensional and one cannot trivialize them. However, each of these foliations is canonically parametrized by the following vector bundle \mathcal{E} over SM :

$$\mathcal{E}(x, \xi) = \{\eta \in T_x M \mid \eta \perp \xi\}, \quad (x, \xi) \in SM.$$

This makes it possible to define *horocyclic operators*

$$\mathcal{U}_{\pm}^m : \mathcal{D}'(SM) \rightarrow \mathcal{D}'(SM; \otimes_S^m \mathcal{E}^*),$$

where \otimes_S^m stands for the m -th symmetric tensor power, and we have the diagram

$$\begin{array}{ccccccc} 0 = V_0(\lambda) & \xrightarrow{\iota} & V_1(\lambda) & \xrightarrow{\iota} & V_2(\lambda) & \xrightarrow{\iota} & V_3(\lambda) \xrightarrow{\iota} \dots, \\ & & \mathcal{V}_+^0 \updownarrow \mathcal{U}^0 & & \mathcal{V}_+^1 \updownarrow \mathcal{U}^1 & & \mathcal{V}_+^2 \updownarrow \mathcal{U}^2 \\ & & \text{Res}_{\mathcal{X}}^0(\lambda) & & \text{Res}_{\mathcal{X}}^1(\lambda + 1) & & \text{Res}_{\mathcal{X}}^2(\lambda + 2) \end{array}$$

where $\mathcal{V}_+^m = (-1)^m (\mathcal{U}_+^m)^*$ and we put, for a certain extension \mathcal{X} of X to $\otimes_S^m \mathcal{E}^*$,

$$\begin{aligned} V_m(\lambda) &:= \{u \in \mathcal{D}'(SM) \mid (X + \lambda)u = 0, \mathcal{U}_-^m u = 0, \text{WF}(u) \subset E_u^*\}, \\ \text{Res}_{\mathcal{X}}^m(\lambda) &:= \{v \in \mathcal{D}'(SM; \otimes_S^m \mathcal{E}^*) \mid (\mathcal{X} + \lambda)v = 0, \mathcal{U}_- v = 0, \text{WF}(v) \subset E_u^*\}. \end{aligned}$$

Similarly to in dimension 2, we reduce the problem to understanding the spaces $\text{Res}_{\mathcal{X}}^m(\lambda)$, and an operator similar to (2-3) maps these spaces to eigenspaces of the Laplacian on divergence-free symmetric tensors. However, to make this statement precise, we have to further decompose $\text{Res}_{\mathcal{X}}^m(\lambda)$ into terms coming from traceless tensors of degrees $m, m - 2, m - 4, \dots$, explaining the appearance of the parameter ℓ in the theorem. (Here the trace of a symmetric tensor of order m is the result of contracting two of its indices with the metric, yielding a tensor of order $m - 2$.) The procedure of reducing elements of $\text{Res}_{\mathcal{X}}^m(\lambda)$ to the conformal boundary \mathbb{S}^n is also made more difficult because the boundary distributions w are now sections of $\otimes_S^m(T^*\mathbb{S}^n)$.

A significant part of the paper is dedicated to proving that the higher-dimensional analog of (2-5) on symmetric tensors is indeed an isomorphism between appropriate spaces. To show that the Poisson operator \mathcal{P}_{λ}^- is injective, we prove a weak expansion of $f(y) \in C^\infty(\mathbb{B}^{n+1})$ in powers of $1 - |y|$ as $y \in \mathbb{B}^{n+1}$ approaches the conformal boundary \mathbb{S}^n ; since w appears as the coefficient in one of the terms of the expansion, $\mathcal{P}_{\lambda}^- w = 0$ implies $w = 0$. To show the surjectivity of \mathcal{P}_{λ}^- , we prove that the lift to \mathbb{H}^{n+1} of every trace-free, divergence-free eigenstate f of the Laplacian admits a weak expansion at the conformal boundary (this requires a fine analysis of the Laplacian and divergence operators on symmetric tensors); putting w to be the coefficient next to one of the terms of this expansion, we can prove that $f = \mathcal{P}_{\lambda}^- w$.

2C. Structure of the paper. In Section 3, we study in detail the geometry of the hyperbolic space \mathbb{H}^{n+1} , which is the covering space of M . In Section 4, we introduce and study the horocyclic operators. In Section 5, we prove Theorems 2 and 3, modulo properties of the Poisson operator. In Sections 6 and 7, we show the injectivity and the surjectivity of the Poisson operator. Appendix A contains several technical lemmas. Appendix B shows how the discussion of Section 2A fits into the framework of the remainder of the paper. Appendix C shows a Weyl law for divergence-free symmetric tensors and relates the $m = 1$ case to the Hodge Laplacian.

3. Geometry of the hyperbolic space

In this section, we review the structure of the hyperbolic space and its geodesic flow and introduce various objects to be used later, including:

- the isometry group G of the hyperbolic space and its Lie algebra, including the horocyclic vector fields U_i^\pm (Section 3B);
- the stable/unstable foliations E_s, E_u (Section 3C);
- the conformal compactification of the hyperbolic space, the maps B_\pm , the coefficients Φ_\pm , and the Poisson kernel (Section 3D);
- parallel transport to conformal infinity and the maps \mathcal{A}_\pm (Section 3F).

3A. Models of the hyperbolic space. Consider the Minkowski space $\mathbb{R}^{1,n+1}$ with the Lorentzian metric

$$g_M = dx_0^2 - \sum_{j=1}^{n+1} dx_j^2.$$

The corresponding scalar product is denoted $\langle \cdot, \cdot \rangle_M$. We denote by e_0, \dots, e_{n+1} the canonical basis of $\mathbb{R}^{1,n+1}$.

The hyperbolic space of dimension $n + 1$ is defined to be one sheet of the two-sheeted hyperboloid

$$\mathbb{H}^{n+1} := \{x \in \mathbb{R}^{1,n+1} \mid \langle x, x \rangle_M = 1, x_0 > 0\}$$

equipped with the Riemannian metric

$$g_H := -g_M|_{T\mathbb{H}^{n+1}}.$$

We denote the unit tangent bundle of \mathbb{H}^{n+1} by

$$S\mathbb{H}^{n+1} := \{(x, \xi) \mid x \in \mathbb{H}^{n+1}, \xi \in \mathbb{R}^{1,n+1}, \langle \xi, \xi \rangle_M = -1, \langle x, \xi \rangle_M = 0\}. \tag{3-1}$$

Another model of the hyperbolic space is the unit ball $\mathbb{B}^{n+1} = \{y \in \mathbb{R}^{n+1} \mid |y| < 1\}$, which is identified with $\mathbb{H}^{n+1} \subset \mathbb{R}^{1,n+1}$ via the map (here $x = (x_0, x') \in \mathbb{R} \times \mathbb{R}^{n+1}$)

$$\psi : \mathbb{H}^{n+1} \rightarrow \mathbb{B}^{n+1}, \quad \psi(x) = \frac{x'}{x_0 + 1}, \quad \psi^{-1}(y) = \frac{1}{1 - |y|^2}(1 + |y|^2, 2y). \tag{3-2}$$

and the metric g_H pulls back to the following metric on \mathbb{B}^{n+1} :

$$(\psi^{-1})^* g_H = \frac{4 dy^2}{(1 - |y|^2)^2}. \tag{3-3}$$

We will also use the upper half-space model $\mathbb{U}^{n+1} = \mathbb{R}_{z_0}^+ \times \mathbb{R}_z^n$ with the metric

$$(\psi^{-1} \psi_1^{-1})^* g_H = \frac{dz_0^2 + dz^2}{z_0^2}, \tag{3-4}$$

where the diffeomorphism $\psi_1 : \mathbb{B}^{n+1} \rightarrow \mathbb{U}^{n+1}$ is given by (here $y = (y_1, y') \in \mathbb{R} \times \mathbb{R}^n$)

$$\psi_1(y_1, y') = \frac{(1 - |y|^2, 2y')}{1 + |y|^2 - 2y_1}, \quad \psi_1^{-1}(z_0, z) = \frac{(z_0^2 + |z|^2 - 1, 2z)}{(1 + z_0)^2 + |z|^2}. \tag{3-5}$$

3B. Isometry group. We consider the group

$$G = \text{PSO}(1, n + 1) \subset \text{SL}(n + 2; \mathbb{R})$$

of all linear transformations of $\mathbb{R}^{1, n+1}$ preserving the Minkowski metric, the orientation, and the sign of x_0 on timelike vectors. For $x \in \mathbb{R}^{1, n+1}$ and $\gamma \in G$, denote by $\gamma \cdot x$ the result of multiplying x by the matrix γ . The group G is exactly the group of orientation-preserving isometries of \mathbb{H}^{n+1} ; under the identification (3-2), it corresponds to the group of direct Möbius transformations of \mathbb{R}^{n+1} preserving the unit ball.

The Lie algebra of G is spanned by the matrices

$$X = E_{0,1} + E_{1,0}, \quad A_k = E_{0,k} + E_{k,0}, \quad R_{i,j} = E_{i,j} - E_{j,i} \tag{3-6}$$

for $i, j \geq 1$ and $k \geq 2$, where $E_{i,j}$ is the elementary matrix if $0 \leq i, j \leq n + 1$ (that is, $E_{i,j} e_k = \delta_{jk} e_i$). Denote for $i = 1, \dots, n$

$$U_i^+ := -A_{i+1} - R_{1,i+1}, \quad U_i^- := -A_{i+1} + R_{1,i+1} \tag{3-7}$$

and observe that $X, U_i^+, U_i^-, R_{i+1,j+1}$ (for $1 \leq i < j \leq n$) also form a basis. Henceforth we identify elements of the Lie algebra of G with left-invariant vector fields on G .

We have the commutator relations (for $1 \leq i, j, k \leq n$ and $i \neq j$)

$$\begin{aligned} [X, U_i^\pm] &= \pm U_i^\pm, & [U_i^\pm, U_j^\pm] &= 0, & [U_i^+, U_i^-] &= 2X, & [U_i^\pm, U_j^\mp] &= 2R_{i+1,j+1}, \\ [R_{i+1,j+1}, X] &= 0, & [R_{i+1,j+1}, U_k^\pm] &= \delta_{jk} U_i^\pm - \delta_{ik} U_j^\pm. \end{aligned} \tag{3-8}$$

The Lie algebra elements U_i^\pm are very important in our argument, since they generate horocyclic flows; see Section 4B. The flows of U_\pm^1 in the case $n = 1$ are shown in Figure 3(a); for $n > 1$, the flows of U_\pm^j do not descend to $S\mathbb{H}^{n+1}$.

The group G acts on \mathbb{H}^{n+1} transitively, with the isotropy group of $e_0 \in \mathbb{H}^{n+1}$ isomorphic to $\text{SO}(n + 1)$. It also acts transitively on the unit tangent bundle $S\mathbb{H}^{n+1}$, by the rule $\gamma.(x, \xi) = (\gamma \cdot x, \gamma \cdot \xi)$, with the isotropy group of $(e_0, e_1) \in S\mathbb{H}^{n+1}$ being

$$H = \{\gamma \in G \mid \gamma \cdot e_0 = e_0, \gamma \cdot e_1 = e_1\} \simeq \text{SO}(n). \tag{3-9}$$

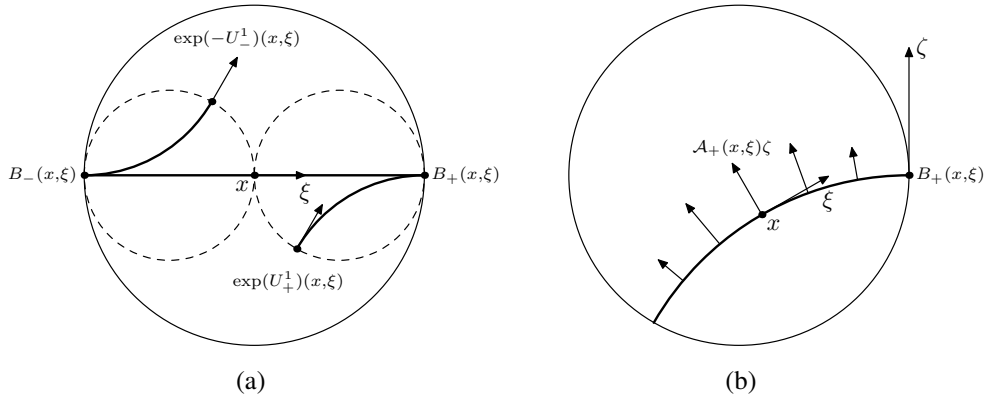


Figure 3. (a) The horocyclic flows $\exp(\pm U_1^\pm)$ in dimension $n + 1 = 2$, pulled back to the ball model by the map ψ from (3-2). The thick lines are geodesics and the dashed lines are horocycles. (b) The map \mathcal{A}_+ and the parallel transport of an element of \mathcal{E} along a geodesic.

Note that H is the connected Lie subgroup of G with Lie algebra spanned by $R_{i+1, j+1}$ for $1 \leq i, j \leq n$. We can then write $S\mathbb{H}^{n+1} \simeq G/H$, where the projection $\pi_S : G \rightarrow S\mathbb{H}^{n+1}$ is given by

$$\pi_S(\gamma) = (\gamma \cdot e_0, \gamma \cdot e_1) \in S\mathbb{H}^{n+1}, \quad \gamma \in G. \tag{3-10}$$

3C. Geodesic flow. The geodesic flow,

$$\varphi_t : S\mathbb{H}^{n+1} \rightarrow S\mathbb{H}^{n+1}, \quad t \in \mathbb{R},$$

is given in the parametrization (3-1) by

$$\varphi_t(x, \xi) = (x \cosh t + \xi \sinh t, x \sinh t + \xi \cosh t). \tag{3-11}$$

We note that, with the projection $\pi_S : G \rightarrow S\mathbb{H}^{n+1}$ defined in (3-10),

$$\varphi_t(\pi_S(\gamma)) = \pi_S(\gamma \exp(tX)),$$

where X is as defined in (3-6). This means that the generator of the geodesic flow can be obtained by pushing forward the left-invariant field on G generated by X by the map π_S (which is possible since X is invariant under right multiplications by elements of the subgroup H defined in (3-9)). By abuse of notation, we then denote by X also the generator of the geodesic flow on $S\mathbb{H}^{n+1}$:

$$X = \xi \cdot \partial_x + x \cdot \partial_\xi. \tag{3-12}$$

We now provide the stable/unstable decomposition for the geodesic flow, demonstrating that it is hyperbolic (and thus the flow on a compact quotient by a discrete group will be Anosov). For $\rho = (x, \xi) \in S\mathbb{H}^{n+1}$, the tangent space $T_\rho(S\mathbb{H}^{n+1})$ can be written as

$$T_\rho(S\mathbb{H}^{n+1}) = \{(v_x, v_\xi) \in (\mathbb{R}^{1, n+1})^2 \mid \langle x, v_x \rangle_M = \langle \xi, v_\xi \rangle_M = \langle x, v_\xi \rangle_M + \langle \xi, v_x \rangle_M = 0\}.$$

The differential of the geodesic flow acts by

$$d\varphi_t(\rho) \cdot (v_x, v_\xi) = (v_x \cosh t + v_\xi \sinh t, v_x \sinh t + v_\xi \cosh t).$$

We have $T_\rho(S\mathbb{H}^{n+1}) = E^0(\rho) \oplus \tilde{T}_\rho(S\mathbb{H}^{n+1})$, where $E^0(\rho) := \mathbb{R}X$ is the flow direction and

$$\tilde{T}_\rho(S\mathbb{H}^{n+1}) = \{(v_x, v_\xi) \in (\mathbb{R}^{1,n+1})^2 \mid \langle x, v_x \rangle_M = \langle x, v_\xi \rangle_M = \langle \xi, v_x \rangle_M = \langle \xi, v_\xi \rangle_M = 0\},$$

and this splitting is invariant under $d\varphi_t$. A natural norm on $\tilde{T}_\rho(S\mathbb{H}^{n+1})$ is given by the formula

$$|(v_x, v_\xi)|^2 := -\langle v_x, v_x \rangle_M - \langle v_\xi, v_\xi \rangle_M, \tag{3-13}$$

using the fact that v_x and v_ξ are Minkowski orthogonal to the timelike vector x and thus must be spacelike or zero. Note that this norm is invariant under the action of G .

We now define the *stable/unstable decomposition* $\tilde{T}_\rho(S\mathbb{H}^{n+1}) = E_s(\rho) \oplus E_u(\rho)$, where

$$\begin{aligned} E_s(\rho) &:= \{(v, -v) \mid \langle x, v \rangle_M = \langle \xi, v \rangle_M = 0\}, \\ E_u(\rho) &:= \{(v, v) \mid \langle x, v \rangle_M = \langle \xi, v \rangle_M = 0\}. \end{aligned} \tag{3-14}$$

Then $T_\rho(S\mathbb{H}^{n+1}) = E_0(\rho) \oplus E_s(\rho) \oplus E_u(\rho)$, this splitting is invariant under φ_t and under the action of G , and, using the norm from (3-13),

$$|d\varphi_t(\rho) \cdot w| = e^{-t}|w|, \quad w \in E_s(\rho), \quad \text{and} \quad |d\varphi_t(\rho) \cdot w| = e^t|w|, \quad w \in E_u(\rho).$$

Finally, we remark that the vector subbundles E_s and E_u are spanned by the left-invariant vector fields U_1^+, \dots, U_n^+ and U_1^-, \dots, U_n^- from (3-7) in the sense that

$$\pi_S^* E_s = \text{span}(U_1^+, \dots, U_n^+) \oplus \mathfrak{h}, \quad \pi_S^* E_u = \text{span}(U_1^-, \dots, U_n^-) \oplus \mathfrak{h}.$$

Here $\pi_S^* E_s := \{(\gamma, w) \in TG \mid (\pi_S(\gamma), d\pi_S(\gamma) \cdot w) \in E_s\}$ and $\pi_S^* E_u$ is defined similarly; \mathfrak{h} is the left translation of the Lie algebra of H , or equivalently the kernel of $d\pi_S$. Note that, while the individual vector fields U_1^\pm, \dots, U_n^\pm are not invariant under right multiplications by elements of H in dimensions $n + 1 > 2$ (and thus do not descend to vector fields on $S\mathbb{H}^{n+1}$ by the map π_S), their spans are invariant under H , by (3-8).

The dual decomposition, used in the construction of Pollicott–Ruelle resonances, is

$$T_\rho^*(S\mathbb{H}^{n+1}) = E_0^*(\rho) \oplus E_s^*(\rho) \oplus E_u^*(\rho), \tag{3-15}$$

where $E_0^*(\rho), E_s^*(\rho), E_u^*(\rho)$ are dual to $E_0(\rho), E_u(\rho), E_s(\rho)$ in the original decomposition (that is, for instance, $E_s^*(\rho)$ consists of all covectors annihilating $E_0(\rho) \oplus E_s(\rho)$). The switching of the roles of E_s and E_u is due to the fact that the flow on the cotangent bundle is $(d\varphi_t^{-1})^*$.

3D. Conformal infinity. The metric (3-3) in the ball model \mathbb{B}^{n+1} is conformally compact; namely, the metric $(1 - |y|^2)^2(\psi^{-1})^*g_H$ continues smoothly to the closure $\bar{\mathbb{B}}^{n+1}$, which we call the *conformal compactification* of \mathbb{H}^{n+1} ; note that \mathbb{H}^{n+1} embeds into the interior of $\bar{\mathbb{B}}^{n+1}$ by the map (3-2). The boundary $\mathbb{S}^n = \partial\bar{\mathbb{B}}^{n+1}$, endowed with the standard metric on the sphere, is called *conformal infinity*. On the hyperboloid model, it is natural to associate to a point at conformal infinity $v \in \mathbb{S}^n$ the lightlike ray

$\{(s, s\nu) \mid s > 0\} \subset \mathbb{R}^{1,n+1}$; this ray is asymptotic to the curve $\{(\sqrt{1+s^2}, s\nu) \mid s > 0\} \subset \mathbb{H}^{n+1}$, which converges to ν in $\overline{\mathbb{B}^{n+1}}$.

Take $(x, \xi) \in S\mathbb{H}^{n+1}$. Then $\langle x \pm \xi, x \pm \xi \rangle_M = 0$ and $x_0 \pm \xi_0 > 0$, so we can write

$$x \pm \xi = \Phi_{\pm}(x, \xi)(1, B_{\pm}(x, \xi))$$

for some maps

$$\Phi_{\pm} : S\mathbb{H}^{n+1} \rightarrow \mathbb{R}^+, \quad B_{\pm} : S\mathbb{H}^{n+1} \rightarrow \mathbb{S}^n. \tag{3-16}$$

Then $B_{\pm}(x, \xi)$ is the limit as $t \rightarrow \pm\infty$ of the x -projection of the geodesic $\varphi_t(x, \xi)$ in $\overline{\mathbb{B}^{n+1}}$:

$$B_{\pm}(x, \xi) = \lim_{t \rightarrow \pm\infty} \pi(\varphi_t(x, \xi)), \quad \pi : S\mathbb{H}^{n+1} \rightarrow \mathbb{H}^{n+1}.$$

This implies that, for X defined in (3-12), $dB_{\pm} \cdot X = 0$, since $B_{\pm}(\varphi_s(x, \xi)) = B_{\pm}(x, \xi)$ for all $s \in \mathbb{R}$. Moreover, since $\Phi_{\pm}(\varphi_t(x, \xi)) = e^{\pm t}(x_0 + \xi_0) = e^{\pm t}\Phi_{\pm}(x, \xi)$ from (3-11), we find

$$X\Phi_{\pm} = \pm\Phi_{\pm}. \tag{3-17}$$

For $(x, \nu) \in \mathbb{H}^{n+1} \times \mathbb{S}^n$ (in the hyperboloid model), define the function

$$P(x, \nu) = (x_0 - x' \cdot \nu)^{-1} = (\langle x, (1, \nu) \rangle_M)^{-1} \quad \text{if } x = (x_0, x') \in \mathbb{H}^{n+1}. \tag{3-18}$$

Note that $P(x, \nu) > 0$ everywhere, and in the Poincaré ball model \mathbb{B}^{n+1} we have

$$P(\psi^{-1}(y), \nu) = \frac{1 - |y|^2}{|y - \nu|^2}, \quad y \in \mathbb{B}^{n+1}, \tag{3-19}$$

which is the usual Poisson kernel. Here ψ is as defined in (3-2).

For $(x, \nu) \in \mathbb{H}^{n+1} \times \mathbb{S}^n$, there exist unique $\xi_{\pm} \in S_x\mathbb{H}^{n+1}$ such that $B_{\pm}(x, \xi_{\pm}) = \nu$: these are given by

$$\xi_{\pm}(x, \nu) = \mp x \pm P(x, \nu)(1, \nu), \tag{3-20}$$

and we have

$$\Phi_{\pm}(x, \xi_{\pm}(x, \nu)) = P(x, \nu). \tag{3-21}$$

Notice that the equations $B_{\pm}(x, \xi_{\pm}(x, \nu)) = \nu$ imply that B_{\pm} are submersions. The map $\nu \rightarrow \xi_{\pm}(x, \nu)$ is conformal with the standard choice of metrics on \mathbb{S}^n and $S_x\mathbb{H}^{n+1}$; in fact, for $\zeta_1, \zeta_2 \in T_{\nu}\mathbb{S}^n$,

$$\langle \partial_{\nu}\xi_{\pm}(x, \nu) \cdot \zeta_1, \partial_{\nu}\xi_{\pm}(x, \nu) \cdot \zeta_2 \rangle_M = -P(x, \nu)^2 \langle \zeta_1, \zeta_2 \rangle_{\mathbb{R}^{n+1}}. \tag{3-22}$$

Using that $\langle x + \xi, x - \xi \rangle_M = 2$, we see that

$$\Phi_+(x, \xi)\Phi_-(x, \xi)(1 - B_+(x, \xi) \cdot B_-(x, \xi)) = 2. \tag{3-23}$$

One can parametrize $S\mathbb{H}^{n+1}$ by

$$(\nu_-, \nu_+, s) = \left(B_-(x, \xi), B_+(x, \xi), \frac{1}{2} \log \frac{\Phi_+(x, \xi)}{\Phi_-(x, \xi)} \right) \in (\mathbb{S}^n \times \mathbb{S}^n)_{\Delta} \times \mathbb{R}, \tag{3-24}$$

where $(\mathbb{S}^n \times \mathbb{S}^n)_\Delta$ is $\mathbb{S}^n \times \mathbb{S}^n$ minus the diagonal. In fact, the geodesic $\gamma(t) = \varphi_t(x, \xi)$ goes from v_- to v_+ in $\bar{\mathbb{B}}^{n+1}$ and $\gamma(-s)$ is the point of γ closest to $e_0 \in \mathbb{H}^{n+1}$ (corresponding to $0 \in \mathbb{B}^{n+1}$). In the parametrization (3-24), the geodesic flow φ_t is simply

$$(v_-, v_+, s) \mapsto (v_-, v_+, s + t).$$

We finally remark that the stable/unstable subspaces of the cotangent bundle, $E_s^*, E_u^* \subset T^*(S\mathbb{H}^{n+1})$, defined in (3-15), are in fact the conormal bundles of the fibers of the maps B_\pm :

$$E_s^*(\rho) = N^*(B_+^{-1}(B_+(\rho))), \quad E_u^*(\rho) = N^*(B_-^{-1}(B_-(\rho))), \quad \rho \in S\mathbb{H}^{n+1}. \tag{3-25}$$

This is equivalent to saying that the fibers of B_+ integrate (that is, are tangent to) the subbundle $E_0 \oplus E_s \subset T(S\mathbb{H}^{n+1})$, while the fibers of B_- integrate the subbundle $E_0 \oplus E_u$. To see the latter statement, say for B_+ , it is enough to note that $dB_+ \cdot X = 0$ and differentiation along vectors in E_s annihilates the function $x + \xi$ and thus the map B_+ ; therefore, the kernel of dB_+ contains $E_0 \oplus E_s$, and this containment is an equality since the dimensions of both spaces are equal to $n + 1$.

3E. Action of G on the conformal infinity. For $\gamma \in G$ and $v \in \mathbb{S}^n$, $\gamma \cdot (1, v)$ is a lightlike vector with positive zeroth component. We can then define $N_\gamma(v) > 0$, $L_\gamma(v) \in \mathbb{S}^n$ by

$$\gamma \cdot (1, v) = N_\gamma(v)(1, L_\gamma(v)). \tag{3-26}$$

The map L_γ gives the action of G on the conformal infinity $\mathbb{S}^n = \partial\bar{\mathbb{B}}^{n+1}$. This action is transitive and the isotropy groups of $\pm e_1 \in \mathbb{S}^n$ are given by

$$H_\pm = \{\gamma \in G \mid \exists s > 0 \ \gamma \cdot (e_0 \pm e_1) = s(e_0 \pm e_1)\}. \tag{3-27}$$

The isotropy groups H_\pm are the connected subgroups of G with the Lie algebras generated by $R_{i+1, j+1}$ for $1 \leq i < j \leq n$, X , and U_i^\pm for $1 \leq i \leq n$. To see that H_\pm are connected, for $n = 1$ we can check directly that every $\gamma \in H_\pm$ can be written as a product $e^{tX} e^{sU_1^\pm}$ for some $t, s \in \mathbb{R}$, and for $n > 1$ we can use the fact that $\mathbb{S}^n \simeq G/H_\pm$ is simply connected and G is connected, and the homotopy long exact sequence of a fibration.

The differentials of N_γ and L_γ (in v) can be written as

$$dN_\gamma(v) \cdot \zeta = \langle dx_0, \gamma \cdot (0, \zeta) \rangle, \quad (0, dL_\gamma(v) \cdot \zeta) = \frac{\gamma \cdot (0, \zeta) - (dN_\gamma(v) \cdot \zeta)(1, L_\gamma(v))}{N_\gamma(v)};$$

here $\zeta \in T_v\mathbb{S}^n$. We see that the map $v \mapsto L_\gamma(v)$ is conformal with respect to the standard metric on \mathbb{S}^n ; in fact, for $\zeta_1, \zeta_2 \in T_v\mathbb{S}^n$,

$$\langle dL_\gamma(v) \cdot \zeta_1, dL_\gamma(v) \cdot \zeta_2 \rangle_{\mathbb{R}^{n+1}} = N_\gamma(v)^{-2} \langle \zeta_1, \zeta_2 \rangle_{\mathbb{R}^{n+1}}.$$

The maps $B_\pm : S\mathbb{H}^{n+1} \rightarrow \mathbb{S}^n$ are equivariant under the action of G :

$$B_\pm(\gamma \cdot (x, \xi)) = L_\gamma(B_\pm(x, \xi)).$$

Moreover, the functions $\Phi_{\pm}(x, \xi)$ and $P(x, v)$ enjoy the following properties:

$$\Phi_{\pm}(\gamma \cdot (x, \xi)) = N_{\gamma}(B_{\pm}(x, \xi))\Phi_{\pm}(x, \xi), \quad P(\gamma \cdot x, L_{\gamma}(v)) = N_{\gamma}(v)P(x, v). \tag{3-28}$$

3F. The bundle \mathcal{E} and parallel transport to the conformal infinity. Consider the vector bundle \mathcal{E} over $S\mathbb{H}^{n+1}$ defined as follows:

$$\mathcal{E} = \{(x, \xi, v) \in S\mathbb{H}^{n+1} \times T_x\mathbb{H}^{n+1} \mid g_H(\xi, v) = 0\},$$

i.e., the fibers $\mathcal{E}(x, \xi)$ consist of all tangent vectors in $T_x\mathbb{H}^{n+1}$ orthogonal to ξ ; equivalently, $\mathcal{E}(x, \xi)$ consists of all vectors in $\mathbb{R}^{1,n+1}$ orthogonal to x and ξ with respect to the Minkowski inner product. Note that G naturally acts on \mathcal{E} , by putting $\gamma \cdot (x, \xi, v) := (\gamma \cdot x, \gamma \cdot \xi, \gamma \cdot v)$.

The bundle \mathcal{E} is invariant under parallel transport along geodesics. Therefore, one can consider the first-order differential operator

$$\mathcal{X} : \mathcal{C}^{\infty}(S\mathbb{H}^{n+1}; \mathcal{E}) \rightarrow \mathcal{C}^{\infty}(S\mathbb{H}^{n+1}; \mathcal{E}), \tag{3-29}$$

which is the generator of parallel transport; namely, if v is a section of \mathcal{E} and $(x, \xi) \in S\mathbb{H}^{n+1}$, then $\mathcal{X}v(x, \xi)$ is the covariant derivative at $t = 0$ of the vector field $v(t) := v(\varphi_t(x, \xi))$ on the geodesic $\varphi_t(x, \xi)$. Note that $\mathcal{E}(\varphi_t(x, \xi))$ is independent of t as a subspace of $\mathbb{R}^{1,n+1}$, and, under this embedding, \mathcal{X} just acts as X on each coordinate of v in $\mathbb{R}^{1,n+1}$. The operator $\frac{1}{i}\mathcal{X}$ is a symmetric operator with respect to the standard volume form on $S\mathbb{H}^{n+1}$ and the inner product on \mathcal{E} inherited from $T\mathbb{H}^{n+1}$.

We now consider parallel transport of vectors along geodesics going off to infinity. Let $(x, \xi) \in S\mathbb{H}^{n+1}$ and $v \in T_x\mathbb{H}^{n+1}$. We let $(x(t), \xi(t)) = \varphi_t(x, \xi)$ be the corresponding geodesic and $v(t) \in T_{x(t)}\mathbb{H}^{n+1}$ be the parallel transport of v along this geodesic. We embed $v(t)$ into the unit ball model \mathbb{B}^{n+1} by defining

$$w(t) = d\psi(x(t)) \cdot v(t) \in \mathbb{R}^{n+1},$$

where ψ is as defined in (3-2). Then $w(t)$ converges to 0 as $t \rightarrow \pm\infty$, but the limits $\lim_{t \rightarrow \pm\infty} x_0(t)w(t)$ are nonzero for nonzero v ; we call the transformation mapping v to these limits the *transport to conformal infinity* as $t \rightarrow \pm\infty$. More precisely, if

$$v = c\xi + u, \quad u \in \mathcal{E}(x, \xi),$$

then we calculate

$$\lim_{t \rightarrow \pm\infty} x_0(t)w(t) = \pm cB_{\pm}(x, \xi) + u' - u_0B_{\pm}(x, \xi), \tag{3-30}$$

where $B_{\pm}(x, \xi) \in \mathbb{S}^n$ is as defined in Section 3D. We will in particular use the inverse of the map $\mathcal{E}(x, \xi) \ni u \mapsto u' - u_0B_{\pm}(x, \xi) \in T_{B_{\pm}(x, \xi)}\mathbb{S}^n$: for $(x, \xi) \in S\mathbb{H}^{n+1}$ and $\zeta \in T_{B_{\pm}(x, \xi)}\mathbb{S}^n$, define (see Figure 3(b))

$$\mathcal{A}_{\pm}(x, \xi)\zeta = (0, \zeta) - \langle (0, \zeta), x \rangle_M(x \pm \xi) = \pm \frac{\partial_v \xi_{\pm}(x, B_{\pm}(x, \xi)) \cdot \zeta}{P(x, B_{\pm}(x, \xi))} \in \mathcal{E}(x, \xi). \tag{3-31}$$

Here ξ_{\pm} is as defined in (3-20). Note that, by (3-22), \mathcal{A}_{\pm} is an isometry:

$$|\mathcal{A}_{\pm}(x, \xi)\zeta|_{g_H} = |\zeta|_{\mathbb{R}^n}, \quad \zeta \in T_{B_{\pm}(x, \xi)}\mathbb{S}^n. \tag{3-32}$$

Also, \mathcal{A}_\pm is equivariant under the action of G :

$$\mathcal{A}_\pm(\gamma \cdot x, \gamma \cdot \xi) \cdot dL_\gamma(B_\pm(x, \xi)) \cdot \zeta = N_\gamma(B_\pm(x, \xi))^{-1} \gamma \cdot (\mathcal{A}_\pm(x, \xi)\zeta). \tag{3-33}$$

We now write the limits (3-30) in terms of the 0-tangent bundle of Mazzeo and Melrose [1987]. Consider the boundary defining function $\rho_0 := 2(1 - |y|)/(1 + |y|)$ on $\bar{\mathbb{B}}^{n+1}$; note that in the hyperboloid model, with the map ψ defined in (3-2),

$$\rho_0(\psi(x)) = 2 \frac{\sqrt{x_0 + 1} - \sqrt{x_0 - 1}}{\sqrt{x_0 + 1} + \sqrt{x_0 - 1}} = x_0^{-1} + \mathcal{O}(x_0^{-2}) \quad \text{as } x_0 \rightarrow \infty. \tag{3-34}$$

The hyperbolic metric can be written near the boundary as $g_H = (d\rho_0^2 + h_{\rho_0})/\rho_0^2$ with h_{ρ_0} a smooth family of metrics on \mathbb{S}^n , and $h_0 = d\theta^2$ is the canonical metric on the sphere (with curvature 1).

Define the 0-tangent bundle ${}^0T\bar{\mathbb{B}}^{n+1}$ to be the smooth bundle over $\bar{\mathbb{B}}^{n+1}$ whose smooth sections are the elements of the Lie algebra $\mathcal{V}_0(\bar{\mathbb{B}}^{n+1})$ of smooth vectors fields vanishing at $\mathbb{S}^n = \bar{\mathbb{B}}^{n+1} \cap \{\rho_0 = 0\}$; near the boundary, this algebra is locally spanned over $\mathcal{C}^\infty(\bar{\mathbb{B}}^{n+1})$ by the vector fields $\rho_0 \partial_{\rho_0}, \rho_0 \partial_{\theta_1}, \dots, \rho_0 \partial_{\theta_n}$ if θ_i are local coordinates on \mathbb{S}^n . There is a natural map ${}^0T\bar{\mathbb{B}}^{n+1} \rightarrow T\bar{\mathbb{B}}^{n+1}$, which is an isomorphism when restricted to the interior \mathbb{B}^{n+1} . We denote by ${}^0T^*\bar{\mathbb{B}}^{n+1}$ the dual bundle to ${}^0T\bar{\mathbb{B}}^{n+1}$, generated locally near $\rho_0 = 0$ by the covectors $d\rho_0/\rho_0, d\theta_1/\rho_0, \dots, d\theta_n/\rho_0$. Note that $T^*\bar{\mathbb{B}}^{n+1}$ naturally embeds into ${}^0T^*\bar{\mathbb{B}}^{n+1}$ and this embedding is an isomorphism in the interior. The metric g_H is a smooth, nondegenerate, positive definite quadratic form on ${}^0T\bar{\mathbb{B}}^{n+1}$, that is, $g_H \in \mathcal{C}^\infty(\bar{\mathbb{B}}^{n+1}; \otimes_S^2({}^0T^*\bar{\mathbb{B}}^{n+1}))$, where \otimes_S^2 denotes the space of symmetric 2-tensors. We refer the reader to [Mazzeo and Melrose 1987] for further details (in particular, for an explanation of why 0-bundles are smooth vector bundles); see also [Melrose 1993, §2.2] for the similar b -setting.

We can then interpret (3-30) as follows: for each $(y, \eta) \in S\bar{\mathbb{B}}^{n+1}$ and each $w \in T_y\bar{\mathbb{B}}^{n+1}$, the parallel transport $w(t)$ of w along the geodesic $\varphi_t(y, \eta)$ (this geodesic extends smoothly to a curve on $\bar{\mathbb{B}}^{n+1}$, as it is part of a line or a circle) has limits as $t \rightarrow \pm\infty$ in the 0-tangent bundle ${}^0T\bar{\mathbb{B}}^{n+1}$. In fact (see [Guillarmou et al. 2010, Appendix A]), the parallel transport

$$\tau(y', y) : {}^0T_y\bar{\mathbb{B}}^{n+1} \rightarrow {}^0T_{y'}\bar{\mathbb{B}}^{n+1}$$

from y to $y' \in \bar{\mathbb{B}}^{n+1}$ along the geodesic starting at y and ending at y' extends smoothly to the boundary $(y, y') \in \bar{\mathbb{B}}^{n+1} \times \bar{\mathbb{B}}^{n+1} \setminus \text{diag}(\mathbb{S}^n \times \mathbb{S}^n)$ as an endomorphism ${}^0T_y\bar{\mathbb{B}}^{n+1} \rightarrow {}^0T_{y'}\bar{\mathbb{B}}^{n+1}$, where $\text{diag}(\mathbb{S}^n \times \mathbb{S}^n)$ denotes the diagonal in the boundary; this parallel transport is an isometry with respect to g_H . The same properties hold for parallel transport of covectors in ${}^0T^*\bar{\mathbb{B}}^{n+1}$, using the duality provided by the metric g_H . An explicit relation to the maps \mathcal{A}_\pm is given by the following formula:

$$\mathcal{A}_\pm(x, \xi) \cdot \zeta = d\psi(x)^{-1} \cdot \tau(\psi(x), B_\pm(x, \xi)) \cdot (\rho_0\zeta), \tag{3-35}$$

where $\rho_0\zeta \in {}^0T_{B_\pm(x, \xi)}\bar{\mathbb{B}}^{n+1}$ is tangent to the conformal boundary \mathbb{S}^n .

4. Horocyclic operators

In this section, we build on the results of Section 3 to construct horocyclic operators

$$\mathcal{U}_\pm : \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes^j \mathcal{E}^*) \rightarrow \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes^{j+1} \mathcal{E}^*).$$

4A. Symmetric tensors. In this subsection, we assume that E is a vector space of finite dimension N , equipped with an inner product g_E , and let E^* denote the dual space, which has a scalar product induced by g_E (also denoted g_E). (In what follows, we shall take either $E = \mathcal{E}(x, \xi)$ or $E = T_x \mathbb{H}^{n+1}$ for some $(x, \xi) \in S\mathbb{H}^{n+1}$, and the scalar product g_E in both cases is given by the hyperbolic metric g_H on those vector spaces.) We will work with tensor powers of E^* , but the constructions also apply to tensor powers of E by swapping E with E^* .

We introduce some notation for finite sequences to simplify the calculations below. Denote by \mathcal{A}^m the space of all sequences $K = k_1 \dots k_m$ with $1 \leq k_\ell \leq N$. For $k_1 \dots k_m \in \mathcal{A}^m$, $j_1 \dots j_r \in \mathcal{A}^r$, and a sequence of distinct numbers $1 \leq \ell_1, \dots, \ell_r \leq m$, denote by

$$\{\ell_1 \rightarrow j_1, \dots, \ell_r \rightarrow j_r\} K \in \mathcal{A}^m$$

the result of replacing the ℓ_p -th element of K by j_p for all p . We can also replace some j_p by blank space, which means that the corresponding indices are removed from K .

For $m \geq 0$ denote by $\otimes^m E^*$ the m -th tensor power of E^* and by $\otimes_S^m E^*$ the subset of those tensors which are symmetric, i.e., $u \in \otimes_S^m E^*$ if $u(v_{\sigma(1)}, \dots, v_{\sigma(m)}) = u(v_1, \dots, v_m)$ for all $\sigma \in \Pi_m$ and all $v_1, \dots, v_m \in E$, where Π_m is the permutation group of $\{1, \dots, m\}$. There is a natural linear projection $\mathcal{S} : \otimes^m E^* \rightarrow \otimes_S^m E^*$ defined by

$$\mathcal{S}(\eta_1^* \otimes \dots \otimes \eta_m^*) = \frac{1}{m!} \sum_{\sigma \in \Pi_m} \eta_{\sigma(1)}^* \otimes \dots \otimes \eta_{\sigma(m)}^*, \quad \eta_k^* \in E^*. \tag{4-1}$$

The metric g_E induces a scalar product on $\otimes^m E^*$,

$$\langle v_1^* \otimes \dots \otimes v_m^*, w_1^* \otimes \dots \otimes w_m^* \rangle_{g_E} = \prod_{j=1}^m \langle v_j^*, w_j^* \rangle_{g_E}, \quad w_i^*, v_i^* \in E^*.$$

The operator \mathcal{S} is selfadjoint and thus an orthogonal projection with respect to this scalar product.

Using the metric g_E , one can decompose the vector space $\otimes_S^m E^*$ as follows. Let $(e_i)_{i=1}^N$ be an orthonormal basis of E for the metric g_E and (e_i^*) be the dual basis. First of all, introduce the trace map $\mathcal{T} : \otimes^m E^* \rightarrow \otimes^{m-2} E^*$ contracting the first two indices by the metric: for $v_i \in E$, define

$$\mathcal{T}(u)(v_1, \dots, v_{m-2}) := \sum_{i=1}^N u(e_i, e_i, v_1, \dots, v_{m-2}) \tag{4-2}$$

(the result is independent of the choice of the basis). For $m < 2$, we define \mathcal{T} to be zero on $\otimes^m E^*$. Note that \mathcal{T} maps $\otimes_S^{m+2} E^*$ onto $\otimes_S^m E^*$. Set

$$e_K^* := e_{k_1}^* \otimes \dots \otimes e_{k_m}^* \in \otimes^m E^*, \quad K = k_1 \dots k_m \in \mathcal{A}^m.$$

Then

$$\mathcal{T}\left(\sum_{K \in \mathcal{A}^{m+2}} f_K \mathbf{e}_K^*\right) = \sum_{K \in \mathcal{A}^m} \sum_{q \in \mathcal{A}} f_{qK} \mathbf{e}_K^*.$$

The adjoint of $\mathcal{T} : \otimes_S^{m+2} E^* \rightarrow \otimes_S^m E^*$ with respect to the scalar product g_E is given by the map $u \mapsto \mathcal{S}(g_E \otimes u)$. To simplify computations, we define a scaled version of it: let $\mathcal{I} : \otimes_S^m E^* \rightarrow \otimes_S^{m+2} E^*$ be defined by

$$\mathcal{I}(u) = \frac{(m+2)(m+1)}{2} \mathcal{S}(g_E \otimes u) = \frac{(m+2)(m+1)}{2} \mathcal{T}^*(u). \quad (4-3)$$

Then

$$\mathcal{I}\left(\sum_{K \in \mathcal{A}^m} f_K \mathbf{e}_K^*\right) = \sum_{K \in \mathcal{A}^{m+2}} \sum_{\substack{\ell, r=1 \\ \ell < r}}^{m+2} \delta_{k_\ell k_r} f_{\{\ell \rightarrow, r \rightarrow\} K} \mathbf{e}_K^*.$$

Note that, for $u \in \otimes_S^m E^*$,

$$\mathcal{T}(\mathcal{I}u) = (2m + N)u + \mathcal{I}(\mathcal{T}u). \quad (4-4)$$

By (4-3) and (4-4), the homomorphism $\mathcal{TI} : \otimes_S^m E^* \rightarrow \otimes_S^m E^*$ is positive definite and thus an isomorphism. Therefore, for $u \in \otimes_S^m E^*$, we can decompose $u = u_1 + \mathcal{I}(u_2)$, where $u_1 \in \otimes_S^m E^*$ satisfies $\mathcal{T}(u_1) = 0$ and $u_2 = (\mathcal{TI})^{-1} \mathcal{T}u \in \otimes_S^{m-2} E^*$. Iterating this process, we can decompose any $u \in \otimes_S^m E^*$ into

$$u = \sum_{r=0}^{\lfloor m/2 \rfloor} \mathcal{I}^r(u_r), \quad u_r \in \otimes_S^{m-2r} E^*, \quad \mathcal{T}(u_r) = 0, \quad (4-5)$$

with u_r determined uniquely by u .

Another operation on tensors which will be used is the interior product: if $v \in E$ and $u \in \otimes_S^m E^*$, we denote by $\iota_v(u) \in \otimes_S^{m-1} E^*$ the interior product of u by v given by

$$\iota_v u(v_1, \dots, v_{m-1}) := u(v, v_1, \dots, v_{m-1}).$$

If $v^* \in E^*$, we write $\iota_{v^*} u$ for the tensor $\iota_v u$ with $g_E(v, \cdot) = v^*$.

We conclude this subsection with a correspondence which will be useful in certain calculations later. There is a linear isomorphism between $\otimes_S^m E^*$ and the space $\text{Pol}^m(E)$ of homogeneous polynomials of degree m on E : to a tensor $u \in \otimes_S^m E^*$ we associate the function on E given by $x \rightarrow P_u(x) := u(x, \dots, x)$. If we write $x = \sum_{i=1}^N x_i \mathbf{e}_i$ in a given orthonormal basis, then

$$P_{\mathcal{S}(\mathbf{e}_K^*)}(x) = \prod_{j=1}^m x_{k_j}, \quad K = k_1 \cdots k_m \in \mathcal{A}^m.$$

The flat Laplacian associated to g_E is given by $\Delta_E = -\sum_{i=1}^N \partial_{x_i}^2$ in the coordinates induced by the basis (\mathbf{e}_i) . Then it is direct to see that

$$\Delta_E P_u(x) = -m(m-1) P_{\mathcal{T}(u)}(x), \quad u \in \otimes_S^m E^*, \quad (4-6)$$

which means that the trace corresponds to applying the Laplacian (see [Dairbekov and Sharafutdinov 2010, Lemma 2.4]). In particular, trace-free symmetric tensors of order m correspond to homogeneous harmonic polynomials, and thus restrict to spherical harmonics on the sphere $|x|_{g_E} = 1$ of E . We also have

$$P_{\mathcal{I}(u)}(x) = \frac{1}{2}(m+2)(m+1)|x|^2 P_u(x), \quad u \in \otimes_S^m E^*. \tag{4-7}$$

4B. Horocyclic operators. We now consider the left-invariant vector fields $X, U_{\pm}^i, R_{i+1,j+1}$ on the isometry group G , identified with the elements of the Lie algebra of G introduced in (3-6), (3-7). Recall that G acts on $S\mathbb{H}^{n+1}$ transitively with the isotropy group $H \simeq \text{SO}(n)$ and this action gives rise to the projection $\pi_S : G \rightarrow S\mathbb{H}^{n+1}$; see (3-10). Note that, with the maps $\Phi_{\pm} : S\mathbb{H}^{n+1} \rightarrow \mathbb{R}^+$ and $B_{\pm} : S\mathbb{H}^{n+1} \rightarrow \mathbb{S}^n$ defined in (3-16), we have

$$B_{\pm}(\pi_S(\gamma)) = L_{\gamma}(\pm e_1) \quad \text{and} \quad \Phi_{\pm}(\pi_S(\gamma)) = N_{\gamma}(\pm e_1), \quad \gamma \in G,$$

where $N_{\gamma} : \mathbb{S}^n \rightarrow \mathbb{R}^+$ and $L_{\gamma} : \mathbb{S}^n \rightarrow \mathbb{S}^n$ are defined in (3-26). Since H_{\pm} , the isotropy group of $\pm e_1$ under the action L_{γ} , contains X and U_i^{\pm} in its Lie algebra (see (3-27) and Figure 3(a)), we find

$$d(B_{\pm} \circ \pi_S) \cdot U_i^{\pm} = 0 \quad \text{and} \quad d(B_{\pm} \circ \pi_S) \cdot X = 0. \tag{4-8}$$

We also calculate

$$d(\Phi_{\pm} \circ \pi_S) \cdot U_i^{\pm} = 0. \tag{4-9}$$

Define the differential operator on G

$$U_K^{\pm} := U_{k_1}^{\pm} \cdots U_{k_m}^{\pm}, \quad K = k_1 \cdots k_m \in \mathcal{A}^m.$$

Note that the order in which k_1, \dots, k_m are listed does not matter, by (3-8). Moreover, by (3-8),

$$[R_{i+1,j+1}, U_K^{\pm}] = \sum_{\ell=1}^m (\delta_{jk_{\ell}} U_{\{\ell \rightarrow i\}K}^{\pm} - \delta_{ik_{\ell}} U_{\{\ell \rightarrow j\}K}^{\pm}). \tag{4-10}$$

Since H is generated by the vector fields $R_{i+1,j+1}$, we see that in dimensions $n+1 > 2$ the horocyclic vector fields U_i^{\pm} , and more generally the operators U_K^{\pm} , are not invariant under right multiplication by elements of H and therefore do not descend to differential operators on $S\mathbb{H}^{n+1}$ —in other words, if $u \in \mathcal{D}'(S\mathbb{H}^{n+1})$, then $U_K^{\pm}(\pi_S^* u) \in \mathcal{D}'(G)$ is not in the image of π_S^* .

However, in this section we will show how to differentiate distributions on $S\mathbb{H}^{n+1}$ along the horocyclic vector fields, resulting in sections of the vector bundle \mathcal{E} introduced in Section 3F and its tensor powers. First of all, we note that by (3-14), the stable and unstable bundles $E_s(x, \xi)$ and $E_u(x, \xi)$ are canonically isomorphic to $\mathcal{E}(x, \xi)$, by the maps

$$\theta_+ : \mathcal{E}(x, \xi) \rightarrow E_s(x, \xi), \quad \theta_- : \mathcal{E}(x, \xi) \rightarrow E_u(x, \xi), \quad \theta_{\pm}(v) = (-v, \pm v).$$

For $u \in \mathcal{D}'(S\mathbb{H}^{n+1})$, we then define the horocyclic derivatives $\mathcal{U}_{\pm} u \in \mathcal{D}'(S\mathbb{H}^{n+1}; \mathcal{E}^*)$ by restricting the differential $du \in \mathcal{D}'(S\mathbb{H}^{n+1}; T^*(S\mathbb{H}^{n+1}))$ to the stable/unstable foliations and pulling it back by θ_{\pm} :

$$\mathcal{U}_{\pm} u(x, \xi) := du(x, \xi) \circ \theta_{\pm} \in \mathcal{E}^*(x, \xi). \tag{4-11}$$

To relate \mathcal{U}_\pm to the vector fields U_j^\pm on the group G , consider the orthonormal frame e_1^*, \dots, e_n^* of the bundle $\pi_S^* \mathcal{E}^*$ over G defined by

$$e_j^*(\gamma) := \gamma^{-*}(e_{j+1}^*) \in \mathcal{E}^*(\pi_S(\gamma)),$$

where the $e_j^* = dx_j$ form the dual basis to the canonical basis $(e_j)_{j=0, \dots, n+1}$ of $\mathbb{R}^{1, n+1}$, and $\gamma^{-*} = (\gamma^{-1})^* : (\mathbb{R}^{1, n+1})^* \rightarrow (\mathbb{R}^{1, n+1})^*$. More generally, we can define the orthonormal frame e_K^* of $\pi_S^*(\otimes^m \mathcal{E}^*)$ by

$$e_K^* := e_{k_1}^* \otimes \dots \otimes e_{k_m}^*, \quad K = k_1 \dots k_m \in \mathcal{A}^m.$$

We compute, for $u \in \mathcal{D}'(S\mathbb{H}^{n+1})$, $du(\pi_S(\gamma)) \cdot \theta_\pm(\gamma(e_{j+1})) = U_j^\pm(\pi_S^* u)(\gamma)$, and thus

$$\pi_S^*(\mathcal{U}_\pm u) = \sum_{j=1}^n U_j^\pm(\pi_S^* u) e_j^*. \tag{4-12}$$

We next use the formula (4-12) to define \mathcal{U}_\pm as an operator

$$\mathcal{U}_\pm : \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes^m \mathcal{E}^*) \rightarrow \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes^{m+1} \mathcal{E}^*) \tag{4-13}$$

as follows: for $u \in \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes^m \mathcal{E}^*)$, define $\mathcal{U}_\pm u$ by

$$\pi_S^*(\mathcal{U}_\pm u) = \sum_{r=1}^n \sum_{K \in \mathcal{A}^m} (U_r^\pm u_K) e_{rK}^*, \quad \pi_S^* u = \sum_{K \in \mathcal{A}^m} u_K e_K^*. \tag{4-14}$$

This definition makes sense (that is, the right-hand side of the first formula in (4-14) lies in the image of π_S^*) since a section

$$f = \sum_{K \in \mathcal{A}^m} f_K e_K^* \in \mathcal{D}'(G; \pi_S^*(\otimes^m \mathcal{E}^*)), \quad f_K \in \mathcal{D}'(G),$$

lies in the image of π_S^* if and only if $R_{i+1, j+1} f = 0$ for $1 \leq i < j \leq n$ (the differentiation is well defined since the fibers of $\pi_S^*(\otimes^m \mathcal{E}^*)$ are the same along each integral curve of $R_{i+1, j+1}$), and this translates to

$$R_{i+1, j+1} f_K = \sum_{\ell=1}^m (\delta_{jk_\ell} f_{\{\ell \rightarrow i\}K} - \delta_{ik_\ell} f_{\{\ell \rightarrow j\}K}), \quad 1 \leq i < j \leq n, \quad K \in \mathcal{A}^m; \tag{4-15}$$

to show (4-15) for $f_{rK} = U_r^\pm u_K$, we use (3-8):

$$\begin{aligned} R_{i+1, j+1} f_{rK} &= [R_{i+1, j+1}, U_r^\pm] u_K + U_r^\pm R_{i+1, j+1} u_K \\ &= \delta_{jr} U_i^\pm u_K - \delta_{ir} U_j^\pm u_K + \sum_{\ell=1}^m \delta_{jk_\ell} U_r^\pm u_{\{\ell \rightarrow i\}K} - \delta_{ik_\ell} U_r^\pm u_{\{\ell \rightarrow j\}K}. \end{aligned}$$

To interpret the operator (4-13) in terms of the stable/unstable foliations in a manner similar to (4-11), consider the connection ∇^S on the bundle \mathcal{E} over $S\mathbb{H}^{n+1}$ defined as follows: for $(x, \xi) \in S\mathbb{H}^{n+1}$, $(v, w) \in T_{(x, \xi)}(S\mathbb{H}^{n+1})$, and $u \in \mathcal{D}'(S\mathbb{H}^{n+1}; \mathcal{E})$, let $\nabla_{(v, w)}^S u(x, \xi)$ be the orthogonal projection of

$$\nabla_{(v, w)}^{\mathbb{R}^{1, n+1}} u(x, \xi) \quad \text{onto} \quad \mathcal{E}(x, \xi) \subset \mathbb{R}^{1, n+1},$$

where $\nabla^{\mathbb{R}^{1,n+1}}$ is the canonical connection on the trivial bundle $S\mathbb{H}^{n+1} \times \mathbb{R}^{1,n+1}$ over $S\mathbb{H}^{n+1}$ (corresponding to differentiating the coordinates of u in $\mathbb{R}^{1,n+1}$). Then ∇^S naturally induces a connection on $\otimes^m \mathcal{E}^*$, also denoted ∇^S , and we have, for $v, v_1, \dots, v_m \in \mathcal{E}(x, \xi)$ and $u \in \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes^m \mathcal{E}^*)$,

$$\mathcal{U}_{\pm} u(x, \xi)(v, v_1, \dots, v_m) = (\nabla_{\theta_{\pm}(v)}^S u)(v_1, \dots, v_m). \tag{4-16}$$

Indeed, if $\gamma(t) = \gamma(0)e^{tU_j^{\pm}}$ is an integral curve of U_j^{\pm} on G , then $\gamma(t)e_2, \dots, \gamma(t)e_{n+1}$ form a parallel frame of \mathcal{E} over the curve $(x(t), \xi(t)) = \pi_S(\gamma(t))$ with respect to ∇^S , since the covariant derivative of $\gamma(t)e_k$ in t with respect to $\nabla^{\mathbb{R}^{1,n+1}}$ is simply $\gamma(t)U_j^{\pm}e_k$; by (3-7) this is a linear combination of $x(t) = \gamma(t)e_0$ and $\xi(t) = \gamma(t)e_1$ and thus $\nabla_t^S(\gamma(t)e_k) = 0$.

Note also that the operator \mathcal{X} defined in (3-29) can be interpreted as the covariant derivative on \mathcal{E} along the generator X of the geodesic flow by the connection ∇^S . One can naturally generalize \mathcal{X} to a first-order differential operator

$$\mathcal{X} : \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes^m \mathcal{E}^*) \rightarrow \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes^m \mathcal{E}^*), \tag{4-17}$$

and $\frac{1}{i}\mathcal{X}$ is still symmetric with respect to the natural measure on $S\mathbb{H}^{n+1}$ and the inner product on $\otimes^m \mathcal{E}^*$ induced by the Minkowski metric. A characterization of X in terms of the frame e_K^* is given by

$$\pi_S^*(\mathcal{X}u) = \sum_{K \in \mathcal{A}^m} (Xu_K)e_K^*, \quad \pi_S^*u = \sum_{K \in \mathcal{A}^m} u_K e_K^*. \tag{4-18}$$

It follows from (3-8) that, for $u \in \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes^m \mathcal{E}^*)$,

$$\mathcal{X}\mathcal{U}_{\pm}u - \mathcal{U}_{\pm}\mathcal{X}u = \pm\mathcal{U}_{\pm}u. \tag{4-19}$$

We also observe that, since $[U_i^{\pm}, U_j^{\pm}] = 0$, for each scalar distribution $u \in \mathcal{D}'(S\mathbb{H}^{n+1})$ and $m \in \mathbb{N}$ we have $\mathcal{U}_{\pm}^m u \in \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes_S^m \mathcal{E}^*)$, where $\otimes_S^m \mathcal{E}^* \subset \otimes^m \mathcal{E}^*$ denotes the space of all symmetric cotensors of order m . Inversion of the operator \mathcal{U}_{\pm}^m is the topic of the next subsection. We conclude with the following lemma, describing how the operator \mathcal{U}_{\pm}^m acts on distributions invariant under the left action of an element of G :

Lemma 4.1. *Let $\gamma \in G$ and $u \in \mathcal{D}'(S\mathbb{H}^{n+1})$. Assume also that u is invariant under left multiplications by γ , namely $u(\gamma.(x, \xi)) = u(x, \xi)$ for all¹ $(x, \xi) \in S\mathbb{H}^{n+1}$. Then $v = \mathcal{U}_{\pm}^m u$ is equivariant under left multiplication by γ in the following sense:*

$$v(\gamma.(x, \xi)) = \gamma.v(x, \xi), \tag{4-20}$$

where the action of γ on $\otimes_S^m \mathcal{E}^*$ is naturally induced by its action on \mathcal{E} (by taking inverse transposes), which in turn comes from the action of γ on $\mathbb{R}^{1,n+1}$.

Proof. We have, for $\gamma' \in G$,

$$\mathcal{U}_{\pm}^m u(\pi_S(\gamma')) = \sum_{K \in \mathcal{A}^m} (U_K^{\pm}(u \circ \pi_S)(\gamma'))e_K^*(\gamma').$$

¹Strictly speaking, this statement should be formulated in terms of the pullback of the distribution u by the map $(x, \xi) \mapsto \gamma.(x, \xi)$.

Therefore, since U_j^\pm are left-invariant vector fields on G ,

$$\mathcal{U}_\pm^m u(\gamma.\pi_S(\gamma')) = \mathcal{U}_\pm^m u(\pi_S(\gamma\gamma')) = \sum_{K \in \mathcal{A}^m} (U_K^\pm(u \circ \pi_S)(\gamma')) \mathbf{e}_K^*(\gamma\gamma').$$

It remains to note that $\mathbf{e}_K^*(\gamma\gamma') = \gamma.\mathbf{e}_K^*(\gamma')$. \square

4C. Inverting horocyclic operators. In this subsection, we show that distributions $v \in \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes_S^m \mathcal{E}^*)$ satisfying certain conditions are in fact in the image of \mathcal{U}_\pm^m acting on $\mathcal{D}'(S\mathbb{H}^{n+1})$. This is an important step in our construction of Pollicott–Ruelle resonances, as it will make it possible to recover a scalar resonant state corresponding to a resonance in the m -th band. More precisely, we prove:

Lemma 4.2. *Assume that $v \in \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes_S^m \mathcal{E}^*)$ satisfies $\mathcal{U}_\pm v = 0$, and $\mathcal{X}v = \pm\lambda v$ for $\lambda \notin \frac{1}{2}\mathbb{Z}$. Then there exists $u \in \mathcal{D}'(S\mathbb{H}^{n+1})$ such that $\mathcal{U}_\pm^m u = v$ and $Xu = \pm(\lambda - m)u$. Moreover, if v is equivariant under left multiplication by some $\gamma \in G$ in the sense of (4-20), then u is invariant under left multiplication by γ .*

The proof of Lemma 4.2 is modeled on the following well-known formula recovering a homogeneous polynomial of degree m from its coefficients: given constants a_α for each multiindex α of length m , we have

$$\partial_x^\beta \sum_{|\alpha|=m} \frac{1}{\alpha!} x^\alpha a_\alpha = a_\beta, \quad |\beta| = m. \quad (4-21)$$

The formula recovering u from v in Lemma 4.2 is morally similar to (4-21), with U_j^\pm taking the role of ∂_{x_j} , the condition $\mathcal{U}_\pm v = 0$ corresponding to a_α being constants, and U_j^\mp taking the role of the multiplication operators x_j . However, the commutation structure of U_j^\pm , given by (3-8), is more involved than that of ∂_{x_j} and x_j , and in particular it involves the vector field X , explaining the need for the condition $\mathcal{X}v = \pm\lambda v$ (which is satisfied by resonant states).

To prove Lemma 4.2, we define the operator

$$\mathcal{V}_\pm : \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes^{m+1} \mathcal{E}^*) \rightarrow \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes^m \mathcal{E}^*), \quad \mathcal{V}_\pm := \mathcal{T}\mathcal{U}_\pm,$$

where \mathcal{T} is as defined in Section 4A. Then, by (4-14),

$$\pi_S^*(\mathcal{V}_\pm u) = \sum_{K \in \mathcal{A}^m} \sum_{q \in \mathcal{A}} (U_q^\pm u_{qK}) \mathbf{e}_K^*, \quad u = \sum_{K \in \mathcal{A}^{m+1}} u_K \mathbf{e}_K^*.$$

For later use, we record the following fact:

Lemma 4.3. $\mathcal{U}_\pm^* = -\mathcal{V}_\pm$, where the adjoint is understood in the formal sense.

Proof. If $u \in \mathcal{C}_0^\infty(S\mathbb{H}^{n+1}; \otimes^m \mathcal{E}^*)$, $v \in \mathcal{C}^\infty(S\mathbb{H}^{n+1}; \otimes^{m+1} \mathcal{E}^*)$, and u_K, v_J are the coordinates of $\pi_S^* u$ and $\pi_S^* v$ in the bases $(\mathbf{e}_K^*)_{K \in \mathcal{A}^m}$ and $(\mathbf{e}_J^*)_{J \in \mathcal{A}^{m+1}}$, then, by (4-14), we compute the following pointwise identity on $S\mathbb{H}^{n+1}$:

$$\langle \mathcal{U}_\pm u, \bar{v} \rangle + \langle u, \overline{\mathcal{V}_\pm v} \rangle = \mathcal{V}_\pm w, \quad w \in \mathcal{C}_0^\infty(S\mathbb{H}^{n+1}; \mathcal{E}^*), \quad \pi_S^* w = \sum_{\substack{K \in \mathcal{A}^m \\ q \in \mathcal{A}}} u_K \overline{v_{qK}} \mathbf{e}_q^*.$$

It remains to show that, for each w , the integral of $\mathcal{V}_\pm w$ is equal to zero. Since \mathcal{V}_\pm is a differential operator of order 1, we must have

$$\int_{S\mathbb{H}^{n+1}} \mathcal{V}_\pm w = \int_{S\mathbb{H}^{n+1}} \langle w, \eta_\pm \rangle$$

for all w and some $\eta_\pm \in C^\infty(S\mathbb{H}^{n+1}; \mathcal{E}^*)$ independent of w . Then η_\pm is equivariant under the action of the isometry group G and, in particular, $|\eta_\pm|$ is a constant function on $S\mathbb{H}^{n+1}$. Moreover, using that $\int Xf = 0$ for all $f \in C_0^\infty(S\mathbb{H}^{n+1})$ and $\mathcal{V}_\pm(\mathcal{X}w) = (X \mp 1)\mathcal{V}_\pm w$, we get, for all $w \in C_0^\infty$,

$$\mp \int_{S\mathbb{H}^{n+1}} \langle w, \eta_\pm \rangle = \int_{S\mathbb{H}^{n+1}} \mathcal{V}_\pm(\mathcal{X}w) = - \int_{S\mathbb{H}^{n+1}} \langle w, \mathcal{X}\eta_\pm \rangle.$$

This implies that $\mathcal{X}\eta_\pm = \pm\eta_\pm$ and, in particular,

$$X|\eta_\pm|^2 = 2\langle \mathcal{X}\eta_\pm, \eta_\pm \rangle = \pm 2|\eta_\pm|^2.$$

Since $|\eta_\pm|^2$ is a constant function, this implies $\eta_\pm = 0$, finishing the proof. \square

To construct u from v in Lemma 4.2, we first handle the case when $\mathcal{T}(v) = 0$; this condition is automatically satisfied when $m \leq 1$.

Lemma 4.4. *Assume that $v \in \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes_S^m \mathcal{E}^*)$ and $\mathcal{U}_\pm v = 0$, $\mathcal{T}(v) = 0$. Define $u = \mathcal{V}_\mp^m v \in \mathcal{D}'(S\mathbb{H}^{n+1})$. Then*

$$\mathcal{U}_\pm^m u = 2^m m! \left(\prod_{\ell=n-1}^{n+m-2} (\ell \pm \mathcal{X}) \right) v. \tag{4-22}$$

Proof. Assume that

$$\pi_S^* v = \sum_{K \in \mathcal{A}^m} f_K \mathbf{e}_K^*, \quad f_K \in \mathcal{D}'(G).$$

Then

$$\pi_S^* u = \sum_{K \in \mathcal{A}^m} U_K^\mp f_K, \quad \pi_S^*(\mathcal{U}_\pm^m u) = \sum_{K, J \in \mathcal{A}^m} U_J^\pm U_K^\mp f_K \mathbf{e}_J^*.$$

For $0 \leq r < m$, $J \in \mathcal{A}^{m-1-r}$, and $p \in \mathcal{A}$, we have, by (3-8),

$$\sum_{\substack{K \in \mathcal{A}^r \\ q \in \mathcal{A}}} [U_p^\pm, U_q^\mp] U_K^\mp f_{qKJ} = \pm 2X \sum_{K \in \mathcal{A}^r} U_K^\mp f_{pKJ} + 2 \sum_{\substack{K \in \mathcal{A}^r \\ q \in \mathcal{A}}} R_{p+1, q+1} U_K^\mp f_{qKJ}.$$

To compute the second term on the right-hand side, we commute $R_{p+1, q+1}$ with U_K^\mp by (4-10) and use (4-15) to get

$$\begin{aligned} \sum_{\substack{K \in \mathcal{A}^r \\ q \in \mathcal{A}}} R_{p+1, q+1} U_K^\mp f_{qKJ} &= \sum_{\substack{K \in \mathcal{A}^r \\ q \in \mathcal{A}}} \left(\sum_{\ell=1}^r (\delta_{qk_\ell} U_{\{\ell \rightarrow p\}K}^\mp f_{qKJ} - \delta_{pk_\ell} U_{\{\ell \rightarrow q\}K}^\mp f_{qKJ}) + U_K^\mp f_{pKJ} - \delta_{pq} U_K^\mp f_{qKJ} \right. \\ &\quad + \sum_{\ell=1}^r (\delta_{qk_\ell} U_K^\mp f_{q(\{\ell \rightarrow p\}K)J} - \delta_{pk_\ell} U_K^\mp f_{q(\{\ell \rightarrow q\}K)J}) \\ &\quad \left. + \sum_{\ell=1}^{m-1-r} (\delta_{qj_\ell} U_K^\mp f_{qK(\{\ell \rightarrow p\}J)} - \delta_{pj_\ell} U_K^\mp f_{qK(\{\ell \rightarrow q\}J)}) \right). \end{aligned}$$

Since v is symmetric and $\mathcal{T}(v) = 0$, the expressions $\sum_{K \in \mathcal{A}^r, q \in \mathcal{A}} \delta_{qk\ell} U_{\{\ell \rightarrow p\}K}^\mp f_{qKJ}$, $\sum_{q \in \mathcal{A}} f_{q(\{\ell \rightarrow q\}K)J}$, and $\sum_{q \in \mathcal{A}} f_{qK(\{\ell \rightarrow q\}J)}$ are zero. Further using the symmetry of v , we find

$$\sum_{\substack{K \in \mathcal{A}^r \\ q \in \mathcal{A}}} R_{p+1, q+1} U_K^\mp f_{qKJ} = (n+m-r-2) \sum_{K \in \mathcal{A}^r} U_K^\mp f_{pKJ},$$

and thus

$$\sum_{\substack{K \in \mathcal{A}^r \\ q \in \mathcal{A}}} [U_p^\pm, U_q^\mp] U_K^\mp f_{qKJ} = 2 \sum_{K \in \mathcal{A}^r} U_K^\mp (\pm X + n + m - 2r - 2) f_{pKJ}. \quad (4-23)$$

Then, using that $\mathcal{U}_\pm v = 0$, we find

$$\begin{aligned} \sum_{\substack{K \in \mathcal{A}^{r+1} \\ q \in \mathcal{A}}} U_p^\pm U_K^\mp f_{KJ} &= \sum_{\substack{K \in \mathcal{A}^r \\ q \in \mathcal{A}}} \sum_{\ell=1}^{r+1} U_{k_\ell \dots k_r}^\mp [U_p^\pm, U_q^\mp] U_{k_1 \dots k_{\ell-1}}^\mp f_{qKJ} \\ &= 2 \sum_{K \in \mathcal{A}^r} \sum_{\ell=1}^{r+1} U_K^\mp (\pm X + n + m - 2\ell) f_{pKJ} \\ &= 2(r+1) \sum_{K \in \mathcal{A}^r} U_K^\mp (\pm X + n + m - r - 2) f_{pKJ}. \end{aligned} \quad (4-24)$$

By iterating (4-24) we obtain (using also that v is symmetric), for $J \in \mathcal{A}^m$,

$$\begin{aligned} U_J^\pm \sum_{K \in \mathcal{A}^m} U_K^\mp f_K &= 2m U_{j_1 \dots j_{m-1}}^\pm \sum_{K \in \mathcal{A}^{m-1}} U_K^\mp (\pm X + n - 1) f_{Kj_m} \\ &= 4m(m-1) U_{j_1 \dots j_{m-2}}^\pm \sum_{K \in \mathcal{A}^{m-2}} U_K^\mp (\pm X + n) (\pm X + n - 1) f_{Kj_{m-1}j_m} \\ &\quad \vdots \\ &= 2^m m! \prod_{\ell=n-1}^{n+m-2} (\pm X + \ell) f_J, \end{aligned}$$

which achieves the proof. \square

To handle the case $\mathcal{T}(v) \neq 0$, define also the horocyclic Laplacians

$$\Delta_\pm := -\mathcal{T}\mathcal{U}_\pm^2 = -\mathcal{V}_\pm \mathcal{U}_\pm : \mathcal{D}'(S\mathbb{H}^{n+1}) \rightarrow \mathcal{D}'(S\mathbb{H}^{n+1}),$$

so that, for $u \in \mathcal{D}'(S\mathbb{H}^{n+1})$,

$$\pi_S^* \Delta_\pm u = - \sum_{q=1}^n U_q^\pm U_q^\pm (\pi_S^* u).$$

Note that, by the commutation relation (3-8),

$$[X, \Delta_\pm] = \pm 2\Delta_\pm. \quad (4-25)$$

Also, by Lemma 4.3, Δ_\pm are symmetric operators.

Lemma 4.5. *Assume that $u \in \mathcal{D}'(S\mathbb{H}^{n+1})$ and $U_{\pm}^{m+1}u = 0$. Then*

$$U_{\pm}^{m+2}\Delta_{\mp}u = -4(\mathcal{X} \mp m)(2\mathcal{X} \pm (n-2))\mathcal{I}(U_{\pm}^m u) - 4\mathcal{I}^2(\mathcal{T}(U_{\pm}^m u)).$$

Proof. We have

$$\pi_S^*(U_{\pm}^{m+2}\Delta_{\mp}u) = - \sum_{\substack{K \in \mathcal{A}^{m+2} \\ q \in \mathcal{A}}} U_K^{\pm} U_q^{\mp} U_q^{\mp} u \mathbf{e}_K^*.$$

Using (3-8), we compute, for $K \in \mathcal{A}^{m+2}$ and $q \in \mathcal{A}$,

$$\begin{aligned} & [U_K^{\pm}, U_q^{\mp}] \\ &= \sum_{\ell=1}^{m+2} U_{k_1 \dots k_{\ell-1}}^{\pm} [U_{k_{\ell}}^{\pm}, U_q^{\mp}] U_{k_{\ell+1} \dots k_{m+2}}^{\pm} \\ &= 2 \sum_{\ell=1}^{m+2} (\delta_{qk_{\ell}} U_{\{\ell \rightarrow\}K}^{\pm} (\pm X + m - \ell + 2) + U_{k_1 \dots k_{\ell-1}}^{\pm} R_{k_{\ell+1}, q+1} U_{k_{\ell+1} \dots k_{m+2}}^{\pm}) \\ &= 2 \sum_{\ell=1}^{m+2} \left(U_{\{\ell \rightarrow\}K}^{\pm} (\delta_{qk_{\ell}} (\pm X + m - \ell + 2) + R_{k_{\ell+1}, q+1}) + \sum_{r=\ell+1}^{m+2} (\delta_{qk_r} U_{\{r \rightarrow\}K}^{\pm} - \delta_{k_{\ell}k_r} U_{\{\ell \rightarrow, r \rightarrow\}K}^{\pm}) \right) \\ &= 2 \sum_{\ell=1}^{m+2} \left(U_{\{\ell \rightarrow\}K}^{\pm} (\delta_{qk_{\ell}} (\pm X + m + 1) + R_{k_{\ell+1}, q+1}) - \sum_{r=\ell+1}^{m+2} \delta_{k_{\ell}k_r} U_{\{\ell \rightarrow, r \rightarrow\}K}^{\pm} \right). \end{aligned}$$

Since $U_{\pm}^{m+1}u = 0$, for $K \in \mathcal{A}^{m+2}$ and $q \in \mathcal{A}$ we have $U_K^{\pm}u = [U_K^{\pm}, U_q^{\mp}]u = 0$, and thus

$$U_K^{\pm} U_q^{\mp} U_q^{\mp} u = [[U_K^{\pm}, U_q^{\mp}], U_q^{\mp}]u.$$

We calculate

$$\sum_{q \in \mathcal{A}} [\delta_{qk_{\ell}} (\pm X + m + 1) + R_{k_{\ell+1}, q+1}, U_q^{\mp}] = (n-2)U_{k_{\ell}}^{\mp},$$

and thus, for $K \in \mathcal{A}^{m+2}$,

$$\sum_{q \in \mathcal{A}} U_K^{\pm} U_q^{\mp} U_q^{\mp} u = 2 \sum_{\ell=1}^{m+2} \left([U_{\{\ell \rightarrow\}K}^{\pm}, U_{k_{\ell}}^{\mp}] (\pm X + m + n - 1) - \sum_{r=\ell+1}^{m+2} \delta_{k_{\ell}k_r} \sum_{q \in \mathcal{A}} [U_{\{\ell \rightarrow, r \rightarrow\}K}^{\pm}, U_q^{\mp}] \right) u.$$

Now, for $K \in \mathcal{A}^{m+2}$,

$$\begin{aligned} & \sum_{\ell=1}^{m+2} [U_{\{\ell \rightarrow\}K}^{\pm}, U_{k_{\ell}}^{\mp}] (\pm X + m + n - 1) u \\ &= 2 \sum_{\substack{\ell, s=1 \\ \ell \neq s}}^{m+2} \left(\delta_{k_{\ell}k_s} U_{\{\ell \rightarrow, s \rightarrow\}K}^{\pm} (\pm X + m) - \sum_{\substack{r=s+1 \\ r \neq \ell}}^{m+2} \delta_{k_{\ell}k_r} U_{\{s \rightarrow, r \rightarrow\}K}^{\pm} \right) (\pm X + m + n - 1) u \\ &= 2 \sum_{\substack{\ell, r=1 \\ \ell < r}}^{m+2} \delta_{k_{\ell}k_r} U_{\{\ell \rightarrow, r \rightarrow\}K}^{\pm} (\pm 2X + m) (\pm X + m + n - 1) u. \end{aligned}$$

Furthermore, we have, for $K \in \mathcal{A}^m$,

$$\sum_{q \in \mathcal{A}} [U_{qK}^\pm, U_q^\mp] u = 2U_K^\pm ((m+n)(\pm X + m) - m)u - 2 \sum_{q \in \mathcal{A}} \sum_{\substack{s,p=1 \\ s < p}}^m \delta_{k_s k_p} U_{qq\{s \rightarrow, p \rightarrow\}}^\pm K u.$$

We finally compute

$$\begin{aligned} & \sum_{q \in \mathcal{A}} U_K^\pm U_q^\mp U_q^\mp u \\ &= 4 \sum_{\substack{\ell, r=1 \\ \ell < r}}^{m+2} \delta_{k_\ell k_r} U_{\{\ell \rightarrow, r \rightarrow\}}^\pm X(2X \pm (n+2m-2))u + 4 \sum_{q \in \mathcal{A}} \sum_{\substack{\ell, r=1 \\ \ell < r}}^{m+2} \sum_{\substack{s,p=1 \\ s < p \\ \{s,p\} \cap \{\ell,r\} = \emptyset}}^{m+2} \delta_{k_\ell k_r} \delta_{k_s k_p} U_{qq\{\ell \rightarrow, r \rightarrow, s \rightarrow, p \rightarrow\}}^\pm K u, \end{aligned}$$

which finishes the proof. \square

Arguing by induction using (4-4) and applying Lemma 4.5 to $\Delta_{\mp}^r u$, we get:

Lemma 4.6. *Assume that $u \in \mathcal{D}'(S\mathbb{H}^{n+1})$ and $\mathcal{U}_\pm^{m+1} u = 0$, $\mathcal{T}(\mathcal{U}_\pm^m u) = 0$. Then, for each $r \geq 0$,*

$$\mathcal{U}_\pm^{m+2r} \Delta_{\mp}^r u = (-1)^r 2^{2r} \left(\prod_{j=0}^{r-1} (\mathcal{X} \mp (m+j)) \right) \left(\prod_{j=1}^r (2\mathcal{X} \pm (n-2j)) \right) \mathcal{I}^r (\mathcal{U}_\pm^m u).$$

Moreover, for $r \geq 1$,

$$\mathcal{T}(\mathcal{U}_\pm^{m+2r} \Delta_{\mp}^r u) = (-1)^r 2^{2r} r(n+2m+2r-2) \left(\prod_{j=0}^{r-1} (\mathcal{X} \mp (m+j)) \right) \left(\prod_{j=1}^r (2\mathcal{X} \pm (n-2j)) \right) \mathcal{I}^{r-1} (\mathcal{U}_\pm^m u).$$

We are now ready to complete the proof of Lemma 4.2. Following (4-5), we decompose v as $v = \sum_{r=0}^{\lfloor m/2 \rfloor} \mathcal{I}^r (v_r)$ with $v_r \in \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes_S^{m-2r} \mathcal{E}^*)$ and $\mathcal{T}(v_r) = 0$. Since X commutes with \mathcal{T} and \mathcal{I} , we find $Xv_r = \pm \lambda v_r$. Moreover, since $\mathcal{U}_\pm v = 0$, we have $\mathcal{U}_\pm v_r = 0$. Put

$$u_r := (-\Delta_{\mp})^r \mathcal{V}_{\mp}^{m-2r} v_r \in \mathcal{D}'(S\mathbb{H}^{n+1}).$$

By Lemma 4.4 (applied to v_r) and Lemma 4.6 (applied to $\mathcal{V}_{\mp}^{m-2r} v_r$ and with m replaced by $m-2r$),

$$\begin{aligned} \mathcal{U}_\pm^m u_r &= 2^{2r} \left(\prod_{j=0}^{r-1} (\lambda - (m-2r+j)) \right) \left(\prod_{j=1}^r (2\lambda + n - 2j) \right) \mathcal{I}^r (\mathcal{U}_\pm^{m-2r} \mathcal{V}_{\mp}^{m-2r} v_r) \\ &= 2^m (m-2r)! \left(\prod_{j=n-1}^{n+m-2r-2} (\lambda + j) \right) \left(\prod_{j=m-2r}^{m-r-1} (\lambda - j) \right) \left(\prod_{j=1}^r (2\lambda + n - 2j) \right) \mathcal{I}^r (v_r). \end{aligned}$$

Since $\lambda \notin \frac{1}{2}\mathbb{Z}$, we see that $v = \mathcal{U}_\pm^m u$, where u is a linear combination of $u_0, \dots, u_{\lfloor m/2 \rfloor}$. The relation $Xu = \pm(\lambda - m)u$ follows immediately from (4-19) and (4-25). Finally, the equivariance property under G follows similarly to Lemma 4.1.

4D. Reduction to the conformal boundary. We now describe the tensors $v \in \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes_S^m \mathcal{E}^*)$ that satisfy $\mathcal{U}_\pm v = 0$ and $Xv = 0$ via symmetric tensors on the conformal boundary \mathbb{S}^n . For that we define the operators

$$\mathcal{Q}_\pm : \mathcal{D}'(\mathbb{S}^n; \otimes^m(T^*\mathbb{S}^n)) \rightarrow \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes^m \mathcal{E}^*)$$

by the following formula: if $w \in \mathcal{C}^\infty(\mathbb{S}^n; \otimes^m(T^*\mathbb{S}^n))$, we set, for $\eta_i \in \mathcal{E}(x, \xi)$,

$$\mathcal{Q}_\pm w(x, \xi)(\eta_1, \dots, \eta_m) := (w \circ B_\pm(x, \xi))(\mathcal{A}_\pm^{-1}(x, \xi)\eta_1, \dots, \mathcal{A}_\pm^{-1}(x, \xi)\eta_m), \tag{4-26}$$

where $\mathcal{A}_\pm(x, \xi) : T_{B_\pm(x, \xi)}\mathbb{S}^n \rightarrow \mathcal{E}(x, \xi)$ is the parallel transport defined in (3-31), and we see that the operator (4-26) extends continuously to $\mathcal{D}'(\mathbb{S}^n; \otimes_S^m(T^*\mathbb{S}^n))$, since the map $B_\pm : S\mathbb{H}^{n+1} \rightarrow \mathbb{S}^n$ defined in (3-16) is a submersion; see [Hörmander 1983, Theorem 6.1.2]; the result can be written as $\mathcal{Q}_\pm w = (\otimes^m(\mathcal{A}_\pm^{-1})^T).w \circ B_\pm$, where T denotes the transpose.

Lemma 4.7. *The operator \mathcal{Q}_\pm is a linear isomorphism from $\mathcal{D}'(\mathbb{S}^n; \otimes_S^m(T^*\mathbb{S}^n))$ onto the space*

$$\{v \in \mathcal{D}'(S\mathbb{H}^{n+1}; \otimes_S^m \mathcal{E}^*) \mid \mathcal{U}_\pm v = 0, \mathcal{X}v = 0\}. \tag{4-27}$$

Proof. It is clear that \mathcal{Q}_\pm is injective. Next, we show that the image of \mathcal{Q}_\pm is contained in (4-27). For that it suffices to show that, for $w \in \mathcal{C}^\infty(\mathbb{S}^n; \otimes_S^m(T^*\mathbb{S}^n))$, we have $\mathcal{U}_\pm(\mathcal{Q}_\pm w) = 0$ and $\mathcal{X}(\mathcal{Q}_\pm w) = 0$. We prove the first statement; the second one is established similarly. Let $\gamma \in G$, $w_1, \dots, w_m \in \mathcal{C}^\infty(\mathbb{S}^n; T\mathbb{S}^n)$, and $w_i^* = \langle w_i, \cdot \rangle_{g_{\mathbb{S}^n}}$ be the duals through the metric. Then

$$\begin{aligned} \mathcal{Q}_\pm(w_1^* \otimes \dots \otimes w_m^*)(\pi_S(\gamma)) &= \sum_{k_1, \dots, k_m=1}^n \left(\prod_{j=1}^m (w_j^* \circ B_\pm \circ \pi_S(\gamma))(\mathcal{A}_\pm^{-1}(\pi_S(\gamma))\gamma \cdot e_{k_j+1}) \right) \mathbf{e}_K^*(\gamma) \\ &= (-1)^m \sum_{k_1, \dots, k_m=1}^n \left(\prod_{j=1}^m \langle (\mathcal{A}_\pm.w_j \circ B_\pm) \circ \pi_S(\gamma), \gamma \cdot e_{k_j+1} \rangle_M \right) \mathbf{e}_K^*(\gamma), \end{aligned}$$

where we have used (3-32) in the second identity. Now we have, from (3-31),

$$\mathcal{A}_\pm(\pi_S(\gamma))\zeta = (0, \zeta) - \langle (0, \zeta), \gamma \cdot e_0 \rangle_M \gamma (e_0 + e_1);$$

thus

$$\mathcal{Q}_\pm(w_1^* \otimes \dots \otimes w_m^*)(\pi_S(\gamma)) = \sum_{k_1, \dots, k_m=1}^n \left(\prod_{j=1}^m \langle (0, -w_j(B_\pm(\pi_S(\gamma)))) \rangle_M, \gamma \cdot e_{k_j+1} \right) \mathbf{e}_K^*(\gamma).$$

Since $d(B_\pm \circ \pi_S) \cdot U_\ell^\pm = 0$ by (4-8) and $U_\ell^\pm(\gamma \cdot e_{k_j+1}) = \gamma \cdot U_\ell^\pm \cdot e_{k_j+1}$ is a multiple of $\gamma \cdot (e_0 \pm e_1) = \Phi_\pm(\pi_S(\gamma))(1, B_\pm(\pi_S(\gamma)))$, we see that $\mathcal{U}_\pm(\mathcal{Q}_\pm w) = 0$ for all w .

It remains to show that, for v in (4-27), we have $v = \mathcal{Q}_\pm(w)$ for some w . For that, define

$$\tilde{v} = (\otimes^m \mathcal{A}_\pm^T) \cdot v \in \mathcal{D}'(S\mathbb{H}^{n+1}; B_\pm^*(\otimes_S^m T^*\mathbb{S}^n)),$$

where \mathcal{A}_\pm^T denotes the transpose of \mathcal{A}_\pm . Then $\mathcal{U}_\pm v = 0$ and $\mathcal{X}v = 0$ imply that $U_\ell^\pm(\pi_S^* \tilde{v}) = 0$ and $X\tilde{v} = 0$ (where, to define differentiation, we embed $T^*\mathbb{S}^n$ into \mathbb{R}^{n+1}). Additionally, $R_{i+1, j+1}(\pi_S^* \tilde{v}) = 0$; therefore $\pi_S^* \tilde{v}$ is constant on the right cosets of the subgroup $H_\pm \subset G$ defined in (3-27). Since

$(B_{\pm} \circ \pi_S)^{-1}(B_{\pm} \circ \pi_S(\gamma)) = \gamma H_{\pm}$, we see that \tilde{v} is the pullback under B_{\pm} of some $w \in \mathcal{D}'(\mathbb{S}^n; \otimes_S^m T^*\mathbb{S}^n)$, and it follows that $v = \mathcal{Q}_{\pm}(w)$. □

In fact, using (3-31) and the expression of $\xi_{\pm}(x, \nu)$ in (3-20) in terms of the Poisson kernel, it is not difficult to show that $\mathcal{Q}_{\pm}(w)$ belongs to a smaller space of *tempered* distributions: in the ball model, this can be described as the dual space to the Fréchet space of smooth sections of $\otimes^m({}^0S\bar{\mathbb{B}}^{n+1})$ over $\bar{\mathbb{B}}^{n+1}$ which vanish to infinite order at the conformal boundary $\mathbb{S}^n = \partial\bar{\mathbb{B}}^{n+1}$.

We finally give a useful criterion for invariance of $\mathcal{Q}_{\pm}(w)$ under the left action of an element of G :

Lemma 4.8. *Take $\gamma \in G$ and let $w \in \mathcal{D}'(\mathbb{S}^n; \otimes_S^m(T^*\mathbb{S}^n))$. Take $s \in \mathbb{C}$ and define $v = \Phi_{\pm}^s \mathcal{Q}_{\pm}(w)$. Then v is equivariant under left multiplication by γ , in the sense of (4-20), if and only if w satisfies the condition*

$$L_{\gamma}^* w(v) = N_{\gamma}(v)^{-s-m} w(v), \quad v \in \mathbb{S}^n. \tag{4-28}$$

Here $L_{\gamma}(v) \in \mathbb{S}^n$ and $N_{\gamma}(v) > 0$ are defined in (3-26).

Proof. The lemma follows by a direct calculation from (3-28) and (3-33). □

5. Pollicott–Ruelle resonances

In this section, we first recall the results of Butterley and Liverani [2007] and Faure and Sjöstrand [2011] on the Pollicott–Ruelle resonances for Anosov flows. We next state several useful microlocal properties of these resonances and prove Theorem 2, modulo properties of Poisson kernels (Lemma 5.8 and Theorem 6), which will be proved in Sections 6 and 7. Finally, we prove a pairing formula for resonances and Theorem 3.

5A. Definition and properties. We follow the presentation of [Faure and Sjöstrand 2011]; a more recent treatment using different technical tools is given in [Dyatlov and Zworski 2015]. We refer the reader to these two papers for the necessary notions of microlocal analysis.

Let \mathcal{M} be a smooth compact manifold of dimension $2n + 1$ and $\varphi_t = e^{tX}$ be an Anosov flow on \mathcal{M} , generated by a smooth vector field X . (In our case, $\mathcal{M} = SM$, $M = \Gamma \backslash \mathbb{H}^{n+1}$, and φ_t is the geodesic flow — see Section 5B.) The Anosov property is defined as follows: there exists a continuous splitting

$$T_y \mathcal{M} = E_0(y) \oplus E_u(y) \oplus E_s(y), \quad y \in \mathcal{M}, \quad E_0(y) := \mathbb{R}X(y), \tag{5-1}$$

invariant under $d\varphi_t$ and such that the stable/unstable subbundles $E_s, E_u \subset T\mathcal{M}$ satisfy, for some fixed smooth norm $|\cdot|$ on the fibers of $T\mathcal{M}$ and some constants C and $\theta > 0$,

$$\begin{aligned} |d\varphi_t(y)v| &\leq C e^{-\theta t} |v|, & v \in E_s(y), \\ |d\varphi_{-t}(y)v| &\leq C e^{-\theta t} |v|, & v \in E_u(y). \end{aligned} \tag{5-2}$$

We make an additional assumption that \mathcal{M} is equipped with a smooth measure μ which is invariant under φ_t , that is, $\mathcal{L}_X \mu = 0$.

We will use the dual decomposition to (5-1), given by

$$T_y^* \mathcal{M} = E_0^*(y) \oplus E_u^*(y) \oplus E_s^*(y), \quad y \in \mathcal{M}, \tag{5-3}$$

where E_0^*, E_u^*, E_s^* are dual to E_0, E_s, E_u respectively (note that E_u and E_s switch places), so for example $E_u^*(y)$ consists of covectors annihilating $E_0(y) \oplus E_u(y)$.

Following [Faure and Sjöstrand 2011, (1.24)], we now consider, for each $r \geq 0$, an *anisotropic Sobolev space*

$$\mathcal{H}^r(\mathcal{M}), \quad \text{where } \mathcal{C}^\infty(\mathcal{M}) \subset \mathcal{H}^r(\mathcal{M}) \subset \mathcal{D}'(\mathcal{M}).$$

Here we put $u := -r, s := r$ in [Faure and Sjöstrand 2011, Lemma 1.2]. Microlocally near E_u^* , the space \mathcal{H}^r is equivalent to the Sobolev space H^{-r} , in the sense that, for each pseudodifferential operator A of order 0 whose wavefront set is contained in a small enough conic neighborhood of E_u^* , the operator A is bounded, $\mathcal{H}^r \rightarrow H^{-r}$ and $H^{-r} \rightarrow \mathcal{H}^r$. Similarly, microlocally near E_s^* , the space \mathcal{H}^r is equivalent to the Sobolev space H^r . We also have $\mathcal{H}^0 = L^2$. The first-order differential operator X admits a unique closed unbounded extension from \mathcal{C}^∞ to \mathcal{H}^r ; see [Faure and Sjöstrand 2011, Lemma A.1].

The following theorem, defining Pollicott–Ruelle resonances associated to φ_t , is due to Faure and Sjöstrand [2011, Theorems 1.4 and 1.5]; see also [Dyatlov and Zworski 2015, Section 3.2].

Theorem 5. *Fix $r \geq 0$. Then the closed unbounded operator*

$$-X : \mathcal{H}^r(\mathcal{M}) \rightarrow \mathcal{H}^r(\mathcal{M})$$

has discrete spectrum in the region $\{\text{Re } \lambda > -r/C_0\}$ for some constant C_0 independent of r . The eigenvalues of $-X$ on \mathcal{H}^r , called Pollicott–Ruelle resonances, and taken with multiplicities, do not depend on the choice of r as long as they lie in the appropriate region.

We have the following criterion for Pollicott–Ruelle resonances which does not use the \mathcal{H}^r spaces explicitly:

Lemma 5.1. *A number $\lambda \in \mathbb{C}$ is a Pollicott–Ruelle resonance of X if and only the space*

$$\text{Res}_X(\lambda) := \{u \in \mathcal{D}'(\mathcal{M}) \mid (X + \lambda)u = 0, \text{WF}(u) \subset E_u^*\} \tag{5-4}$$

is nontrivial. Here WF denotes the wavefront set; see, for instance, [Faure and Sjöstrand 2011, Definition 1.6]. The elements of $\text{Res}_X(\lambda)$ are called resonant states associated to λ and the dimension of this space is called the geometric multiplicity of λ .

Proof. Assume first that λ is a Pollicott–Ruelle resonance. Take $r > 0$ such that $\text{Re } \lambda > -r/C_0$. Then λ is an eigenvalue of $-X$ on \mathcal{H}^r , which implies that there exists nonzero $u \in \mathcal{H}^r$ such that $(X + \lambda)u = 0$. By [Faure and Sjöstrand 2011, Theorem 1.7], we have $\text{WF}(u) \subset E_u^*$; thus u lies in (5-4).

Assume now that $u \in \mathcal{D}'(\mathcal{M})$ is a nonzero element of (5-4). For large enough r , we have $\text{Re } \lambda > -r/C_0$ and $u \in H^{-r}(\mathcal{M})$. Since $\text{WF}(u) \subset E_u^*$ and \mathcal{H}^r is equivalent to H^{-r} microlocally near E_u^* , we have $u \in \mathcal{H}^r$. Together with the identity $(X + \lambda)u$, this shows that λ is an eigenvalue of $-X$ on \mathcal{H}^r and thus a Pollicott–Ruelle resonance. □

For each λ with $\text{Re } \lambda > -r/C_0$, the operator $X + \lambda : \mathcal{H}^r \rightarrow \mathcal{H}^r$ is Fredholm of index zero on its domain; this follows from the proof of Theorem 5. Therefore, $\dim \text{Res}_X(\lambda)$ is equal to the dimension of the kernel

of the adjoint operator $X^* + \bar{\lambda}$ on the L^2 dual of \mathcal{H}^r , which we denote by \mathcal{H}^{-r} . Since $\frac{1}{i}X$ is symmetric on L^2 , we see that $\text{Res}_X(\lambda)$ has the same dimension as the following space of *coresonant states* at λ :

$$\text{Res}_{X^*}(\lambda) := \{u \in \mathcal{D}'(\mathcal{M}) \mid (X - \bar{\lambda})u = 0, \text{WF}(u) \subset E_s^*\}. \quad (5-5)$$

The main difference of (5-5) from (5-4) is that the subbundle E_s^* is used instead of E_u^* ; this can be justified by applying Lemma 5.1 to the vector field $-X$ instead of X , since the roles of the stable/unstable spaces for the corresponding flow φ_{-t} are reversed.

Note also that, for any $\lambda, \lambda^* \in \mathbb{C}$, one can define a pairing

$$\langle u, u^* \rangle \in \mathbb{C}, \quad u \in \text{Res}_X(\lambda), \quad u^* \in \text{Res}_{X^*}(\lambda^*). \quad (5-6)$$

One way to do that is to use the fact that wavefront sets of u and u^* intersect only at the zero section and apply [Hörmander 1983, Theorem 8.2.10]. An equivalent definition is obtained by noting that u is in \mathcal{H}^r and u^* is in \mathcal{H}^{-r} for $r > 0$ large enough and using the duality of \mathcal{H}^r and \mathcal{H}^{-r} . Note that, for $\lambda \neq \lambda^*$, we have $\langle u, u^* \rangle = 0$; indeed, $X(u\bar{u}^*) = (\lambda^* - \lambda)u\bar{u}^*$ integrates to 0. The question of computing the product $\langle u, u^* \rangle$ for $\lambda = \lambda^*$ is much more subtle and related to algebraic multiplicities; see Section 5C.

Since $\frac{1}{i}X$ is selfadjoint on $L^2 = \mathcal{H}^0$ (see [Faure and Sjöstrand 2011, Appendix A.1]), it has no eigenvalues on this space away from the real line; this implies that there are no Pollicott–Ruelle resonances in the right half-plane. In other words, we have:

Lemma 5.2. *The spaces $\text{Res}_X(\lambda)$ and $\text{Res}_{X^*}(\lambda)$ are trivial for $\text{Re } \lambda > 0$.*

Finally, we note that the results above apply to certain operators on vector bundles. More precisely, let \mathcal{E} be a smooth vector bundle over \mathcal{M} and assume that \mathcal{X} is a first-order differential operator on $\mathcal{D}'(\mathcal{M}; \mathcal{E})$ whose principal part is given by X , namely

$$\mathcal{X}(f\mathbf{u}) = f\mathcal{X}(\mathbf{u}) + (Xf)\mathcal{X}(\mathbf{u}), \quad f \in \mathcal{D}'(\mathcal{M}), \quad \mathbf{u} \in \mathcal{C}^\infty(\mathcal{M}; \mathcal{E}). \quad (5-7)$$

Assume moreover that \mathcal{E} is endowed with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{E}}$ and $\frac{1}{i}\mathcal{X}$ is symmetric on L^2 with respect to this inner product and the measure μ . By an easy adaptation of the results of [Faure and Sjöstrand 2011] (see [Faure and Tsujii 2014; Dyatlov and Zworski 2015]), one can construct anisotropic Sobolev spaces $\mathcal{H}^r(\mathcal{M}; \mathcal{E})$ and Theorem 5 and Lemmas 5.1 and 5.2 apply to \mathcal{X} on these spaces.

5B. Proof of the main theorem. We now concentrate on the case

$$\mathcal{M} = SM = \Gamma \backslash (S\mathbb{H}^{n+1}), \quad M = \Gamma \backslash \mathbb{H}^{n+1},$$

with φ_t the geodesic flow. Here $\Gamma \subset G = \text{PSO}(1, n+1)$ is a cocompact discrete subgroup with no fixed points, so that M is a compact smooth manifold. Henceforth we identify functions on the sphere bundle SM with functions on $S\mathbb{H}^{n+1}$ invariant under Γ , and similar identifications will be used for other geometric objects. It is important to note that *the constructions of the previous sections, except those involving the conformal infinity, are invariant under left multiplication by elements of G and thus descend naturally to SM .*

The lift of the geodesic flow on SM is the generator of the geodesic flow on $S\mathbb{H}^{n+1}$ (see Section 3C); both are denoted X . The lifts of the stable/unstable spaces E_s, E_u to $S\mathbb{H}^{n+1}$ are given in (3-14), and we see that (5-1) holds with $\theta = 1$. The invariant measure μ on SM is just the product of the volume measure on M and the standard measure on the fibers of SM induced by the metric.

Consider the bundle \mathcal{E} on SM defined in Section 3F. Then, for each m , the operator

$$\mathcal{X} : \mathcal{D}'(SM; \otimes_S^m \mathcal{E}^*) \rightarrow \mathcal{D}'(SM; \otimes_S^m \mathcal{E}^*)$$

defined in (4-17) satisfies (5-7) and $\frac{1}{i}\mathcal{X}$ is symmetric. The results of Section 5A apply both to X and \mathcal{X} .

Recall the operator \mathcal{U}_- introduced in Section 4B and its powers, for $m \geq 0$,

$$\mathcal{U}_-^m : \mathcal{D}'(SM) \rightarrow \mathcal{D}'(SM; \otimes_S^m \mathcal{E}^*).$$

The significance of \mathcal{U}_-^m for Pollicott–Ruelle resonances is explained by the following:

Lemma 5.3. *Assume that $\lambda \in \mathbb{C}$ is a Pollicott–Ruelle resonance of X and $u \in \text{Res}_X(\lambda)$ is a corresponding resonant state as defined in (5-4). Then*

$$\mathcal{U}_-^m u = 0 \quad \text{for } m > -\text{Re } \lambda.$$

Proof. By (4-19),

$$(\mathcal{X} + \lambda + m)\mathcal{U}_-^m u = 0.$$

Note also that $\text{WF}(\mathcal{U}_-^m u) \subset E_u^*$, since $\text{WF}(u) \subset E_u^*$ and \mathcal{U}_-^m is a differential operator. Since $\lambda + m$ lies in the right half-plane, it remains to apply Lemma 5.2 to $\mathcal{U}_-^m u$. \square

We can then use the operators \mathcal{U}_-^m to split the resonance spectrum into bands:

Lemma 5.4. *Assume that $\lambda \in \mathbb{C} \setminus \frac{1}{2}\mathbb{Z}$. Then*

$$\dim \text{Res}_X(\lambda) = \sum_{m \geq 0} \dim \text{Res}_{\mathcal{X}}^m(\lambda + m), \tag{5-8}$$

where

$$\text{Res}_{\mathcal{X}}^m(\lambda) := \{v \in \mathcal{D}'(SM; \otimes_S^m \mathcal{E}^*) \mid (\mathcal{X} + \lambda)v = 0, \mathcal{U}_- v = 0, \text{WF}(v) \subset E_u^*\}. \tag{5-9}$$

The space $\text{Res}_{\mathcal{X}}^m(\lambda)$ is trivial for $\text{Re } \lambda > 0$ (by Lemma 5.2). If $\lambda \in \frac{1}{2}\mathbb{Z}$, then we have

$$\dim \text{Res}_X(\lambda) \leq \sum_{m \geq 0} \dim \text{Res}_{\mathcal{X}}^m(\lambda + m). \tag{5-10}$$

Proof. Denote, for $m \geq 1$,

$$V_m(\lambda) := \{u \in \mathcal{D}'(SM) \mid (X + \lambda)u = 0, \mathcal{U}_-^m u = 0, \text{WF}(u) \subset E_u^*\}.$$

Clearly, $V_m(\lambda) \subset V_{m+1}(\lambda)$. Moreover, by Lemma 5.3 we have $\text{Res}_X(\lambda) = V_m(\lambda)$ for m large enough depending on λ . By (4-19), the operator \mathcal{U}_-^m acts as

$$\mathcal{U}_-^m : V_{m+1}(\lambda) \rightarrow \text{Res}_{\mathcal{X}}^m(\lambda + m), \tag{5-11}$$

and the kernel of (5-11) is exactly $V_m(\lambda)$, with the convention that $V_0(\lambda) = 0$. Therefore,

$$\dim V_{m+1}(\lambda) \leq \dim V_m(\lambda) + \dim \text{Res}_{\mathcal{X}}^m(\lambda + m)$$

and (5-10) follows.

To show (5-8), it remains to prove that the operator (5-11) is onto; this follows from Lemma 4.2 (which does not enlarge the wavefront set of the resulting distribution, since it only employs differential operators in the proof). \square

The space $\text{Res}_{\mathcal{X}}^m(\lambda + m)$ is called the space of *resonant states at λ associated to the m -th band*; later we see that most of the corresponding Pollicott–Ruelle resonances satisfy $\text{Re } \lambda = -\frac{1}{2}n - m$. Similarly, we can describe $\text{Res}_{\mathcal{X}^*}^m(\lambda)$ via the spaces $\text{Res}_{\mathcal{X}^*}^m(\lambda + m)$, where

$$\text{Res}_{\mathcal{X}^*}^m(\lambda) := \{v \in \mathcal{D}'(SM; \otimes_S^m \mathcal{E}^*) \mid (\mathcal{X} - \bar{\lambda})v = 0, \mathcal{U}_+ v = 0, \text{WF}(v) \subset E_S^*\}; \tag{5-12}$$

note that here \mathcal{U}_+ is used in place of \mathcal{U}_- .

We further decompose $\text{Res}_{\mathcal{X}}^m(\lambda)$ using *trace-free* tensors:

Lemma 5.5. *Recall the homomorphisms $\mathcal{T} : \otimes_S^m \mathcal{E}^* \rightarrow \otimes_S^{m-2} \mathcal{E}^*$, $\mathcal{I} : \otimes_S^m \mathcal{E}^* \rightarrow \otimes_S^{m-2} \mathcal{E}^*$ defined in Section 4A (we put $\mathcal{T} = 0$ for $m = 0, 1$). Define the space*

$$\text{Res}_{\mathcal{X}}^{m,0}(\lambda) := \{v \in \text{Res}_{\mathcal{X}}^m(\lambda) \mid \mathcal{T}(v) = 0\}. \tag{5-13}$$

Then for all $m \geq 0$ and λ ,

$$\dim \text{Res}_{\mathcal{X}}^m(\lambda) = \sum_{\ell=0}^{\lfloor m/2 \rfloor} \dim \text{Res}_{\mathcal{X}}^{m-2\ell,0}(\lambda). \tag{5-14}$$

In fact,

$$\text{Res}_{\mathcal{X}}^{m,0}(\lambda) = \bigoplus_{\ell=0}^{\lfloor m/2 \rfloor} \mathcal{I}^\ell(\text{Res}_{\mathcal{X}}^{m-2\ell,0}(\lambda)). \tag{5-15}$$

Proof. The identity (5-15) follows immediately from (4-5); it is straightforward to see that the defining properties of $\text{Res}_{\mathcal{X}}^m(\lambda)$ are preserved by the canonical tensorial operations involved. The identity (5-14) then follows since \mathcal{I} is one-to-one by the paragraph following (4-4). \square

The elements of $\text{Res}_{\mathcal{X}}^{m,0}(\lambda)$ can be expressed via distributions on the conformal boundary \mathbb{S}^n :

Lemma 5.6. *Let \mathcal{Q}_- be the operator defined in (4-26); recall that it is injective. If $\pi_\Gamma : S\mathbb{H}^{n+1} \rightarrow SM$ is the natural projection map, then*

$$\pi_\Gamma^* \text{Res}_{\mathcal{X}}^{m,0}(\lambda) = \Phi_-^\lambda \mathcal{Q}_-(\text{Bd}^{m,0}(\lambda)),$$

where $\text{Bd}^{m,0}(\lambda) \subset \mathcal{D}'(\mathbb{S}^n; \otimes_S^m(T^*\mathbb{S}^n))$ consists of all distributions w such that $\mathcal{T}(w) = 0$ and

$$L_\gamma^* w(v) = N_\gamma(v)^{-\lambda-m} w(v), \quad v \in \mathbb{S}^n, \gamma \in \Gamma; \tag{5-16}$$

L_γ and N_γ are as defined in (3-26). Similarly

$$\pi_\Gamma^* \text{Res}_{\mathcal{X}^*}^{m,0}(\lambda) = \Phi_+^{\bar{\lambda}} \mathcal{Q}_+(\text{Bd}^{m,0}(\bar{\lambda})), \quad \text{Bd}^{m,0}(\bar{\lambda}) = \overline{\text{Bd}^{m,0}(\lambda)}.$$

Proof. Assume first that $w \in \text{Bd}^{m,0}(\lambda)$ and put $\tilde{v} = \Phi_-^\lambda Q_-(w)$. Then, by Lemma 4.8 and (5-16), \tilde{v} is invariant under Γ and thus descends to a distribution $v \in \mathcal{D}'(SM; \otimes_S^m \mathcal{E}^*)$. Since $X\Phi_-^\lambda = -\lambda\Phi_-^\lambda$ and $U_j^-(\Phi_-^\lambda \circ \pi_S) = 0$ by (3-17) and (4-8), and \mathcal{X} and \mathcal{U}_- annihilate the image of Q_- by Lemma 4.7, we have $(\mathcal{X} + \lambda)v = 0$ and $\mathcal{U}_-v = 0$. Moreover, by [Hörmander 1983, Theorem 8.2.4] the wavefront set of \tilde{v} is contained in the conormal bundle to the fibers of the map B_- ; by (3-25), we see that $\text{WF}(v) \subset E_u^*$. Finally, $\mathcal{T}(v) = 0$ since the map $\mathcal{A}_-(x, \xi)$ used in the definition of Q_- is an isometry. Therefore, $v \in \text{Res}_\chi^{m,0}(\lambda)$ and we have proved the containment $\pi_\Gamma^* \text{Res}_\chi^{m,0}(\lambda) \supset \Phi_-^\lambda Q_-(\text{Bd}^{m,0}(\lambda))$. The opposite containment is proved by reversing this argument. \square

Remark. It follows from the proof of Lemma 5.6 that the condition $\text{WF}(v) \subset E_u^*$ in (5-9) is unnecessary. This could also be seen by applying [Hörmander 1994, Theorem 18.1.27] to the equations $(\mathcal{X} + \lambda)v = 0$ and $\mathcal{U}_-v = 0$, since \mathcal{X} differentiates along the direction E_0 , \mathcal{U}_- differentiates along the direction E_u (see (4-11) and (4-16)), and the annihilator of $E_0 \oplus E_u$ (that is, the joint critical set of $\mathcal{X} + \lambda, \mathcal{U}_-$) is exactly E_u^* .

It now remains to relate the space $\text{Bd}^{m,0}(\lambda)$ to an eigenspace of the Laplacian on symmetric tensors. For that, we introduce the following operator, obtained by integrating the corresponding elements of $\text{Res}_\chi^{m,0}(\lambda)$ along the fibers of \mathbb{S}^n :

Definition 5.7. Take $\lambda \in \mathbb{C}$. The *Poisson operators*

$$\mathcal{P}_\lambda^\pm : \mathcal{D}'(\mathbb{S}^n; \otimes^m T^*\mathbb{S}^n) \rightarrow \mathcal{C}^\infty(\mathbb{H}^{n+1}; \otimes^m T^*\mathbb{H}^{n+1})$$

are defined by the formulas

$$\begin{aligned} \mathcal{P}_\lambda^- w(x) &= \int_{S_x \mathbb{H}^{n+1}} \Phi_-(x, \xi)^\lambda Q_-(w)(x, \xi) dS(\xi), \\ \mathcal{P}_\lambda^+ w(x) &= \int_{S_x \mathbb{H}^{n+1}} \Phi_+(x, \xi)^{\bar{\lambda}} Q_+(w)(x, \xi) dS(\xi). \end{aligned} \tag{5-17}$$

Here, integration of elements of $\otimes^m \mathcal{E}^*(x, \xi)$ is performed by embedding them in $\otimes^m T_x^* \mathbb{H}^{n+1}$ using composition with the orthogonal projection $T_x \mathbb{H}^{n+1} \rightarrow \mathcal{E}(x, \xi)$.

The operators \mathcal{P}_λ^\pm are related by the identity

$$\overline{\mathcal{P}_\lambda^\pm w} = \mathcal{P}_\lambda^\mp \bar{w}. \tag{5-18}$$

By Lemma 5.6, \mathcal{P}_λ^- maps $\text{Bd}^{m,0}(\lambda)$ onto symmetric Γ -equivariant tensors, which can thus be considered as elements of $\mathcal{C}^\infty(M; \otimes_S^m T^*M)$. The relation with the Laplacian is given by the following fact, proved in Section 6C:

Lemma 5.8. *The image of $\text{Bd}^{m,0}(\lambda)$ under \mathcal{P}_λ^- is contained in the eigenspace $\text{Eig}^m(-\lambda(n + \lambda) + m)$ for each λ , where*

$$\text{Eig}^m(\sigma) := \{f \in \mathcal{C}^\infty(M; \otimes_S^m T^*M) \mid \Delta f = \sigma f, \nabla^* f = 0, \mathcal{T}(f) = 0\}. \tag{5-19}$$

Here the trace \mathcal{T} was defined in Section 4A and the Laplacian Δ and the divergence ∇^* are introduced in Section 6A. (A similar result for \mathcal{P}_λ^+ follows from (5-18).)

Furthermore, in Sections 6C and 7 we show the crucial:

Theorem 6. *Assume that $\lambda \notin \mathcal{R}_m$, where*

$$\mathcal{R}_m = \begin{cases} -\frac{1}{2}n - \frac{1}{2}\mathbb{N}_0 & \text{if } n > 1 \text{ or } m = 0, \\ -\frac{1}{2}\mathbb{N}_0 & \text{if } n = 1 \text{ and } m > 0. \end{cases} \tag{5-20}$$

Then the map $\mathcal{P}_\lambda^- : \text{Bd}^{m,0}(\lambda) \rightarrow \text{Eig}^m(-\lambda(n + \lambda) + m)$ is an isomorphism.

Remark. In Theorem 6, the set of exceptional points where we do not show isomorphism is not optimal but is sufficient for our application (we only need $\mathcal{R}_m \subset m - \frac{1}{2}n - \frac{1}{2}\mathbb{N}_0$); we expect the exceptional set to be contained in $-n + 1 - \mathbb{N}_0$. This result is known for functions, that is for $m = 0$, with the exceptional set being $-n - \mathbb{N}$. This was proved by Helgason [1974] and Minemura [1975] in the case of hyperfunctions on \mathbb{S}^n and by Oshima and Sekiguchi [1980] and Schlichtkrull and van den Ban [1987] for distributions; Grellier and Otal [2005] studied the sharp functional spaces on \mathbb{S}^n of the boundary values of bounded eigenfunctions on \mathbb{H}^{n+1} . The extension to $m > 0$ does not seem to be known in the literature and is not trivial: it takes most of Sections 6 and 7.

We finally provide the following refinement of Lemma 5.4, needed to handle the case $\lambda \in (-\frac{1}{2}n, \infty) \cap \frac{1}{2}\mathbb{Z}$:

Lemma 5.9. *Assume that $\lambda \in -\frac{1}{2}n + \frac{1}{2}\mathbb{N}$. If $\lambda \in -2\mathbb{N}$, then*

$$\dim \text{Res}_X(\lambda) = \sum_{\substack{m \geq 0 \\ m \neq -\lambda}} \dim \text{Res}_X^m(\lambda + m).$$

If $\lambda \notin -2\mathbb{N}$, then (5-8) holds.

Proof. We use the proof of Lemma 5.4. We first show that, for m odd or $\lambda \neq -m$,

$$\mathcal{U}_-^m(V_{m+1}(\lambda)) = \text{Res}_X^m(\lambda + m). \tag{5-21}$$

Using (5-15), it suffices to prove that, for $0 \leq \ell \leq \frac{1}{2}m$, the space $\mathcal{I}^\ell(\text{Res}_X^{m-2\ell,0}(\lambda + m))$ is contained in $\mathcal{U}_-^m(V_{m+1}(\lambda))$. This follows from the proof of Lemma 4.2 as long as

$$\begin{aligned} \lambda + m &\notin \mathbb{Z} \cap ([2\ell + 2 - n - m, 1 - n] \cup [m - 2\ell, m - \ell - 1]), \\ \lambda + m + \frac{1}{2}n &\notin \mathbb{Z} \cap [1, \ell]; \end{aligned}$$

using that $\lambda > -\frac{1}{2}n$, it suffices to prove that

$$\lambda \notin \mathbb{Z} \cap [-2\ell, -\ell - 1]. \tag{5-22}$$

On the other hand, by Lemma 5.6, Theorem 6, and Lemma 6.1, if $\ell < \frac{1}{2}m$ and the space $\text{Res}_X^{m-2\ell,0}(\lambda + m)$ is nontrivial, then

$$-(\lambda + m + \frac{1}{2}n)^2 + \frac{1}{4}n^2 + m - 2\ell \geq m - 2\ell + n - 1,$$

implying

$$|\lambda + m + \frac{1}{2}n| \leq |\frac{1}{2}n - 1|, \tag{5-23}$$

and (5-22) follows. For the case $\ell = \frac{1}{2}m$, since $\Delta \geq 0$ on functions we have

$$-(\lambda + m + \frac{1}{2}n)^2 + \frac{1}{4}n^2 \geq 0,$$

which implies that $\lambda \leq -m$ and thus (5-22) holds unless $\lambda = -m$.

It remains to consider the case when $m = 2\ell$ is even and $\lambda = -m$. We have

$$\text{Res}_{\mathcal{X}}^m(0) = \mathcal{I}^\ell(\text{Res}_{\mathcal{X}}^{0,0}(0));$$

that is, $\text{Res}_{\mathcal{X}}^{m-2\ell',0}(0)$ is trivial for $\ell' < \frac{1}{2}m$. For $n > 1$, this follows immediately from (5-23), and, for $n = 1$, since the bundle \mathcal{E}^* is one-dimensional, we get $\text{Res}_{\mathcal{X}}^{m',0}(\lambda) = 0$ for $m' \geq 2$. Now, $\text{Res}_{\mathcal{X}}^{0,0}(0) = \text{Res}_{\mathcal{X}}^0(0)$ corresponds via Lemma 5.6 and Theorem 6 to the kernel of the scalar Laplacian, that is, to the space of constant functions. Therefore, $\text{Res}_{\mathcal{X}}^{0,0}$ is one-dimensional and it is spanned by the constant function 1 on SM ; it follows that $\text{Res}_{\mathcal{X}}^m(0)$ is spanned by $\mathcal{I}^\ell(1)$. However, by Lemma 4.3, for each $u \in \mathcal{D}'(SM)$,

$$\langle \mathcal{I}^\ell(1), \mathcal{U}_-^m u \rangle_{L^2} = (-1)^m \langle \mathcal{V}_-^m \mathcal{I}^\ell(1), u \rangle_{L^2} = 0.$$

Since $\mathcal{U}_-^m(V_{m+1}(\lambda)) \subset \text{Res}_{\mathcal{X}}^m(0)$, we have $\mathcal{U}_-^m = 0$ on $V_{m+1}(\lambda)$, which implies that $V_{m+1}(\lambda) = V_m(\lambda)$, finishing the proof. □

To prove Theorem 2, it now suffices to combine Lemmas 5.4–5.9 with Theorem 6.

5C. Resonance pairing and algebraic multiplicity. In this section, we prove Theorem 3. The key component is a pairing formula which states that the inner product between a resonant and a coresonant state, defined in (5-6), is determined by the inner product between the corresponding eigenstates of the Laplacian. The nondegeneracy of the resulting inner product as a bilinear operator on $\text{Res}_X(\lambda) \times \text{Res}_{X^*}(\lambda)$ for $\lambda \notin \frac{1}{2}\mathbb{Z}$ immediately implies the fact that the algebraic and geometric multiplicities of λ coincide (that is, $X + \lambda$ does not have any nontrivial Jordan cells).

To state the pairing formula, we first need a decomposition of the space $\text{Res}_X(\lambda)$, which is an effective version of the formulas (5-8) and (5-14). Take $m \geq 0$, $\ell \leq \lfloor m/2 \rfloor$ and $w \in \text{Bd}^{m-2\ell,0}(\lambda)$. Let \mathcal{I} be the operator defined in Section 4A. Then (5-15) and Lemma 5.6 show that

$$\text{Res}_{\mathcal{X}}^m(\lambda) = \bigoplus_{\ell=0}^{\lfloor m/2 \rfloor} \mathcal{I}^\ell(\text{Res}_{\mathcal{X}}^{m-2\ell,0}(\lambda)) = \bigoplus_{\ell=0}^{\lfloor m/2 \rfloor} \mathcal{I}^\ell(\Phi_-^\lambda \mathcal{Q}_-(\text{Bd}^{m-2\ell,0}(\lambda))).$$

Next, let

$$\mathcal{V}_\pm^m : \mathcal{D}'(SM; \otimes_S^m \mathcal{E}^*) \rightarrow \mathcal{D}'(SM) \quad \text{and} \quad \Delta_\pm : \mathcal{D}'(SM) \rightarrow \mathcal{D}'(SM)$$

be the operators introduced in Section 4C. Then the proofs of Lemma 5.4 and Lemma 4.2 show that, for $\lambda \notin \frac{1}{2}\mathbb{Z}$,

$$\begin{aligned} \text{Res}_X(\lambda) &= \bigoplus_{m \geq 0} \bigoplus_{\ell=0}^{\lfloor m/2 \rfloor} V_{m\ell}(\lambda), & V_{m\ell}(\lambda) &:= \Delta_+^\ell \mathcal{V}_+^{m-2\ell}(\Phi_-^{\lambda+m} \mathcal{Q}_-(\text{Bd}^{m-2\ell,0}(\lambda+m))), \\ \text{Res}_{X^*}(\lambda) &= \bigoplus_{m \geq 0} \bigoplus_{\ell=0}^{\lfloor m/2 \rfloor} V_{m\ell}^*(\lambda), & V_{m\ell}^*(\lambda) &:= \Delta_-^\ell \mathcal{V}_-^{m-2\ell}(\overline{\Phi_+^{\lambda+m} \mathcal{Q}_+(\text{Bd}^{m-2\ell,0}(\lambda+m))}), \end{aligned} \tag{5-24}$$

and the operators in the definitions of $V_{m\ell}(\lambda)$, $V_{m\ell}^*(\lambda)$ are one-to-one on the corresponding spaces. By the proof of Lemma 5.9, the decomposition (5-24) is also valid for $\lambda \in (-\frac{1}{2}n, \infty) \setminus (-2\mathbb{N})$; for $\lambda \in (-\frac{1}{2}n, \infty) \cap (-2\mathbb{N})$, we have

$$\text{Res}_X(\lambda) = \bigoplus_{\substack{m \geq 0 \\ m \neq -\lambda}} \bigoplus_{\ell=0}^{\lfloor m/2 \rfloor} V_{m\ell}(\lambda), \quad \text{Res}_{X^*}(\lambda) = \bigoplus_{\substack{m \geq 0 \\ m \neq -\lambda}} \bigoplus_{\ell=0}^{\lfloor m/2 \rfloor} V_{m\ell}^*(\lambda). \tag{5-25}$$

We can now state the pairing formula:

Lemma 5.10. *Let $\lambda \notin -\frac{1}{2}n - \frac{1}{2}\mathbb{N}_0$ and $u \in \text{Res}_X(\lambda)$, $u^* \in \text{Res}_{X^*}(\lambda)$. Let $\langle u, u^* \rangle_{L^2(SM)}$ be defined by (5-6). Then:*

- (1) *If $u \in V_{m\ell}(\lambda)$, $u^* \in V_{m'\ell'}^*(\lambda)$, and $(m, \ell) \neq (m', \ell')$, then $\langle u, u^* \rangle_{L^2(SM)} = 0$.*
- (2) *If $u \in V_{m\ell}(\lambda)$, $u^* \in V_{m\ell}^*(\lambda)$, and $w \in \text{Bd}^{m-2\ell,0}(\lambda+m)$ and $w^* \in \overline{\text{Bd}^{m-2\ell,0}(\lambda+m)}$ are the elements generating u and u^* according to (5-24), then*

$$\langle u, u^* \rangle_{L^2(SM)} = c_{m\ell}(\lambda) \langle \mathcal{P}_{\lambda+m}^-(w), \mathcal{P}_{\lambda+m}^+(w^*) \rangle_{L^2(M)}, \tag{5-26}$$

where

$$c_{m\ell}(\lambda) = 2^{m+2\ell-n} \pi^{-1-n/2} \ell! (m-2\ell)! \sin(\pi(\frac{1}{2}n + \lambda)) \times \frac{\Gamma(m + \frac{1}{2}n - \ell) \Gamma(\lambda + n + 2m - 2\ell) \Gamma(-\lambda - \ell) \Gamma(-\lambda - m - \frac{1}{2}n + \ell + 1)}{\Gamma(m + \frac{1}{2}n - 2\ell) \Gamma(-\lambda - 2\ell)},$$

and, under the conditions (i) either $\lambda \notin -2\mathbb{N}$ or $m \neq -\lambda$ and (ii) $V_{m\ell}(\lambda)$ is nontrivial, we have $c_{m\ell}(\lambda) \neq 0$.

Remark. (i) The proofs below are rather technical, and it is suggested that the reader start with the case of resonances in the first band, $m = \ell = 0$, which preserves the essential analytic difficulties of the proof but considerably reduces the amount of calculation needed (in particular, one can go immediately to Lemma 5.11, and the proof of this lemma for the case $m = \ell = 0$ does not involve the operator \mathcal{E}_η). We have

$$c_{00}(\lambda) = (4\pi)^{-n/2} \frac{\Gamma(n + \lambda)}{\Gamma(\frac{1}{2}n + \lambda)}.$$

(ii) In the special case of $n = 1$, $m = \ell = 0$, Lemma 5.10 is a corollary of [Anantharaman and Zelditch 2007, Theorem 1.2], where the product $uu^* \in \mathcal{D}'(SM)$ lifts to a Patterson–Sullivan distribution on $S\mathbb{H}^2$. In general, if $|\text{Re } \lambda| \leq C$ and $\text{Im } \lambda \rightarrow \infty$, then $c_{m\ell}(\lambda)$ grows like $|\lambda|^{n/2+m}$.

Lemma 5.10 immediately gives:

Proof of Theorem 3. By Theorem 6, we know that

$$\mathcal{P}_\lambda^- : \text{Bd}^{m-2\ell,0}(\lambda+m) \rightarrow \text{Eig}^{m-2\ell}(-(\lambda+m + \frac{1}{2}n)^2 + \frac{1}{4}n^2 + m - 2\ell)$$

is an isomorphism. Given (5-18), we also get the isomorphism

$$\mathcal{P}_\lambda^+ : \overline{\text{Bd}^{m-2\ell,0}(\lambda+m)} \rightarrow \text{Eig}^{m-2\ell}(-(\lambda+m + \frac{1}{2}n)^2 + \frac{1}{4}n^2 + m - 2\ell).$$

Here we used that the target space is invariant under complex conjugation. By Lemma 5.10, the bilinear product

$$\text{Res}_X(\lambda) \times \text{Res}_{X^*}(\lambda) \rightarrow \mathbb{C}, \quad (u, u^*) \mapsto \langle u, u^* \rangle_{L^2(SM)} \tag{5-27}$$

is nondegenerate, since the $L^2(M)$ inner product restricted to $\text{Eig}^{m-2\ell}(-(\lambda + m + \frac{1}{2}n)^2 + \frac{1}{4}n^2 + m - 2\ell)$ is nondegenerate for all m, ℓ .

Assume now that $\tilde{u} \in \mathcal{D}'(SM)$ satisfies $(X + \lambda)^2 \tilde{u} = 0$ and $\tilde{u} \in \mathcal{H}^r$ for some r , $\text{Re } \lambda > -r/C_0$; we need to show that $(X + \lambda)\tilde{u} = 0$. Put $u := (X + \lambda)\tilde{u}$. Then $u \in \text{Res}_X(\lambda)$. However, u also lies in the image of $X + \lambda$ on \mathcal{H}^r ; therefore we have $\langle u, u^* \rangle = 0$ for each $u^* \in \text{Res}_{X^*}(\lambda)$. Since the product (5-27) is nondegenerate, we see that $u = 0$, finishing the proof. \square

In the remaining part of this section, we prove Lemma 5.10. Take some $m, m', \ell, \ell' \geq 0$ such that $2\ell \leq m, 2\ell' \leq m'$, and consider $u \in V_{m\ell}(\lambda), u^* \in V_{m'\ell'}(\lambda)$ given by

$$u = \Delta_+^\ell \mathcal{V}_+^{m-2\ell} v, \quad u^* = \Delta_-^{\ell'} \mathcal{V}_-^{m'-2\ell'} v^*,$$

where, for some $w \in \text{Bd}^{m-2\ell,0}(\lambda + m)$ and $w^* \in \overline{\text{Bd}^{m'-2\ell',0}(\lambda + m')}$,

$$v = \Phi_-^{\lambda+m} \mathcal{Q}_-(w) \in \text{Res}_X^{m-2\ell,0}(\lambda + m), \quad v^* = \Phi_+^{\bar{\lambda}+m'} \mathcal{Q}_+(w^*) \in \text{Res}_{X^*}^{m'-2\ell',0}(\lambda + m').$$

Using Lemma 4.3 and the fact that Δ_\pm are symmetric, we get

$$\langle u, u^* \rangle_{L^2(SM)} = (-1)^{m'} \langle \mathcal{U}_-^{m'-2\ell'} \Delta_-^{\ell'} \Delta_+^\ell \mathcal{V}_+^{m-2\ell} v, v^* \rangle_{L^2(SM; \otimes^{m'-2\ell'} \mathcal{E}^*)}.$$

By Lemmas 4.4 and 4.6, $\mathcal{U}_-^{m+1} \Delta_+^\ell \mathcal{V}_+^{m-2\ell} v = 0$. Therefore, if $m' > m$, we derive that $\langle u, u^* \rangle_{L^2(SM)} = 0$; by swapping u and u^* , one can similarly handle the case $m' < m$. We therefore assume that $m = m'$. Then, by Lemmas 4.4 and 4.6 (see the proof of Lemma 4.2),

$$\begin{aligned} & (-1)^{\ell+\ell'} \mathcal{U}_-^{m-2\ell'} \Delta_-^{\ell'} \Delta_+^\ell \mathcal{V}_+^{m-2\ell} v \\ &= \mathcal{T}^{\ell'} \mathcal{U}_-^m (-\Delta_+)^{\ell'} \mathcal{V}_+^{m-2\ell} v \\ &= 2^{m+\ell} (m-2\ell)! \frac{\Gamma(\lambda + n + 2m - 2\ell - 1) \Gamma(-\lambda - \ell) \Gamma(-\lambda - m - \frac{1}{2}n + \ell + 1)}{\Gamma(\lambda + m + n - 1) \Gamma(-\lambda - 2\ell) \Gamma(-\lambda - m - \frac{1}{2}n + 1)} \mathcal{T}^{\ell'} \mathcal{I}^\ell v. \end{aligned}$$

If $\ell' > \ell$, this implies that $\langle u, u^* \rangle_{L^2(SM)} = 0$, and the case $\ell' < \ell$ is handled similarly. (Recall that $\mathcal{T}(v) = 0$.) We therefore assume that $m = m', \ell = \ell'$. In this case, by (4-4),

$$\mathcal{T}^\ell \mathcal{I}^\ell v = 2^\ell \ell! \frac{\Gamma(m + \frac{1}{2}n - \ell)}{\Gamma(m + \frac{1}{2}n - 2\ell)} v,$$

which implies that

$$\begin{aligned} \langle u, u^* \rangle_{L^2(SM)} &= (-2)^{m+2\ell} \ell! (m-2\ell)! \frac{\Gamma(m + \frac{1}{2}n - \ell) \Gamma(\lambda + n + 2m - 2\ell - 1)}{\Gamma(m + \frac{1}{2}n - 2\ell) \Gamma(\lambda + n + m - 1)} \\ &\quad \times \frac{\Gamma(-\lambda - \ell) \Gamma(-\lambda - m - \frac{1}{2}n + \ell + 1)}{\Gamma(-\lambda - 2\ell) \Gamma(-\lambda - m - \frac{1}{2}n + 1)} \langle v, v^* \rangle_{L^2(SM; \otimes^{m-2\ell} \mathcal{E}^*)}. \end{aligned}$$

Note that, under assumptions (i) and (ii) of Lemma 5.10, the coefficient in the formula above is nonzero; see the proof of Lemma 5.9.

It then remains to prove the following identity (note that the coefficient there is nonzero for $\lambda \notin \mathbb{Z}$ or $\operatorname{Re} \lambda > m - \frac{1}{2}n$):

Lemma 5.11. *Assume that $v \in \operatorname{Res}_{\chi}^{m,0}(\lambda)$ and $v^* \in \operatorname{Res}_{\chi^*}^{m,0}(\lambda)$. Define*

$$f(x) := \int_{S_x M} v(x, \xi) dS(\xi), \quad f^*(x) := \int_{S_x M} v^*(x, \xi) dS(\xi),$$

where integration of tensors is understood as in Definition 5.7. If $\lambda \notin -(\frac{1}{2}n + \mathbb{N}_0)$, then

$$\langle f, f^* \rangle_{L^2(M; \otimes^m T^*M)} = 2^n \pi^{n/2} \frac{\Gamma(\frac{1}{2}n + \lambda)}{(n + \lambda + m - 1)\Gamma(n - 1 + \lambda)} \langle v, v^* \rangle_{L^2(SM; \otimes^m \mathcal{E}^*)}.$$

Proof. We write

$$\langle f, f^* \rangle_{L^2(M; \otimes^m T^*M)} = \int_{S^2 M} \langle v(y, \eta_-), \overline{v^*(y, \eta_+)} \rangle_{\otimes^m T_y^* M} dy d\eta_- d\eta_+, \tag{5-28}$$

where the bundle $S^2 M$ is given by

$$S^2 M = \{(y, \eta_-, \eta_+) \mid y \in M, \eta_{\pm} \in S_y M\}.$$

Define also

$$S^2_{\Delta} M = \{(y, \eta_-, \eta_+) \in S^2 M \mid \eta_- + \eta_+ \neq 0\}.$$

On the other hand,

$$\langle v, v^* \rangle_{L^2(SM; \otimes^m \mathcal{E}^*)} = \int_{SM} \langle v(x, \xi), \overline{v^*(x, \xi)} \rangle_{\otimes^m \mathcal{E}^*(x, \xi)} dx d\xi. \tag{5-29}$$

The main idea of the proof is to reduce (5-28) to (5-29) by applying the coarea formula to a correctly chosen map $S^2_{\Delta} M \rightarrow SM$. More precisely, consider the following map $\Psi : \mathcal{E} \rightarrow S^2_{\Delta} \mathbb{H}^{n+1}$: for $(x, \xi) \in S\mathbb{H}^{n+1}$ and $\eta \in \mathcal{E}(x, \xi)$, define $\Psi(x, \xi, \eta) := (y, \eta_-, \eta_+)$, with

$$\begin{pmatrix} y \\ \eta_- \\ \eta_+ \end{pmatrix} = A(|\eta|^2) \begin{pmatrix} x \\ \xi \\ \eta \end{pmatrix}, \quad A(s) = \begin{pmatrix} \sqrt{s+1} & 0 & 1 \\ \frac{s}{\sqrt{s+1}} & \frac{1}{\sqrt{s+1}} & 1 \\ -\frac{s}{\sqrt{s+1}} & \frac{1}{\sqrt{s+1}} & -1 \end{pmatrix}.$$

Note that, with $|\eta|$ denoting the Riemannian length of η (that is, $|\eta|^2 = -\langle \eta, \eta \rangle_M$),

$$\Phi_{\pm}(y, \eta_{\pm}) = \frac{\Phi_{\pm}(x, \xi)}{\sqrt{1 + |\eta|^2}}, \quad B_{\pm}(y, \eta_{\pm}) = B_{\pm}(x, \xi), \quad |\eta_+ + \eta_-| = \frac{2}{\sqrt{1 + |\eta|^2}}.$$

Also,

$$\det A(s) = -\frac{2}{s+1}, \quad A(s)^{-1} = \begin{pmatrix} \sqrt{s+1} & -\frac{1}{2}\sqrt{s+1} & \frac{1}{2}\sqrt{s+1} \\ 0 & \frac{1}{2}\sqrt{s+1} & \frac{1}{2}\sqrt{s+1} \\ -s & \frac{1}{2}(s+1) & -\frac{1}{2}(s+1) \end{pmatrix}.$$

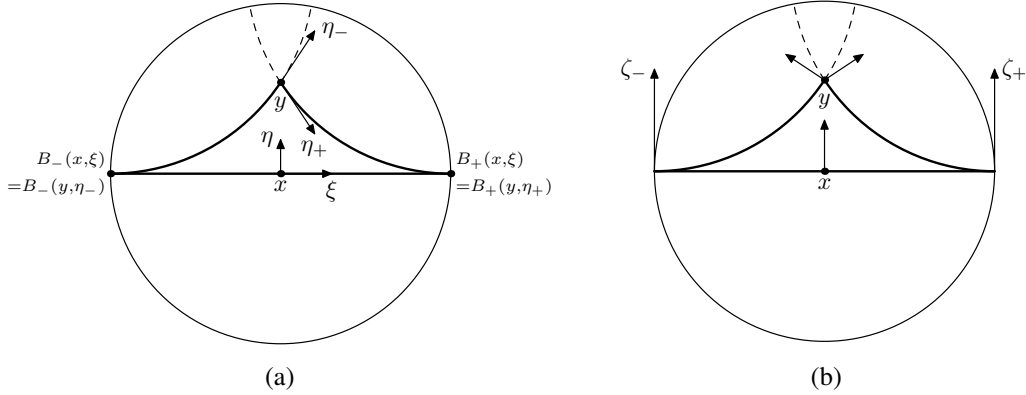


Figure 4. (a) The map $\Psi : (x, \xi, \eta) \mapsto (y, \eta_-, \eta_+)$. (b) The vectors $\mathcal{A}_\pm(x, \xi)\zeta_\pm$ (which are equal in the case drawn) and $\mathcal{A}_\pm(y, \eta_\pm)\zeta_\pm$.

The map Ψ is a diffeomorphism; the inverse is given by the formulas

$$x = \frac{2y + \eta_+ - \eta_-}{|\eta_+ + \eta_-|}, \quad \xi = \frac{\eta_+ + \eta_-}{|\eta_+ + \eta_-|}, \quad \eta = \frac{2(\eta_- - \eta_+) - |\eta_+ - \eta_-|^2 y}{|\eta_+ + \eta_-|^2}.$$

The map Ψ^{-1} can be visualized as follows (see Figure 4(a)): given (y, η_-, η_+) , the corresponding tangent vector (x, ξ) is the closest to y on the geodesic going from $v_- = B_-(y, \eta_-)$ to $v_+ = B_+(y, \eta_+)$ and the vector η measures both the distance between x and y and the direction of the geodesic from x to y . The exceptional set $\{\eta_+ + \eta_- = 0\}$ corresponds to $|\eta| = \infty$.

A calculation using (3-31) shows that, for $\zeta_\pm \in T_{B_\pm(x, \xi)}\mathbb{S}^n$,

$$\mathcal{A}_\pm(y, \eta_\pm)\zeta_\pm = \mathcal{A}_\pm(x, \xi)\zeta_\pm + \frac{(\mathcal{A}_\pm(x, \xi)\zeta_\pm) \cdot \eta}{\sqrt{1 + |\eta|^2}}(x \pm \xi).$$

Here, \cdot stands for the Riemannian inner product on \mathcal{E} , which is equal to $-\langle \cdot, \cdot \rangle_M$ restricted to \mathcal{E} . Then (see Figure 4(b))

$$\begin{aligned} & (\mathcal{A}_+(y, \eta_+)\zeta_+) \cdot (\mathcal{A}_-(y, \eta_-)\zeta_-) \\ &= (\mathcal{A}_+(x, \xi)\zeta_+) \cdot (\mathcal{A}_-(x, \xi)\zeta_-) - \frac{2}{1 + |\eta|^2} ((\mathcal{A}_+(x, \xi)\zeta_+) \cdot \eta) ((\mathcal{A}_-(x, \xi)\zeta_-) \cdot \eta) \\ &= (\mathcal{C}_\eta(\mathcal{A}_+(x, \xi)\zeta_+)) \cdot (\mathcal{A}_-(x, \xi)\zeta_-), \end{aligned}$$

where $\mathcal{C}_\eta : \mathcal{E}(x, \xi) \rightarrow \mathcal{E}(x, \xi)$ is given by

$$\mathcal{C}_\eta(\tilde{\eta}) = \tilde{\eta} - \frac{2}{1 + |\eta|^2} (\tilde{\eta} \cdot \eta)\eta.$$

We can similarly define $\mathcal{C}_\eta^* : \mathcal{E}(x, \xi)^* \rightarrow \mathcal{E}(x, \xi)^*$. Then, for $\zeta_\pm \in \otimes^m T_{B_\pm(x, \xi)}^*\mathbb{S}^n$,

$$\begin{aligned} & \langle \otimes^m (\mathcal{A}_+^{-1}(y, \eta_+)^T)\zeta_+, \otimes^m (\mathcal{A}_-^{-1}(y, \eta_-)^T)\zeta_- \rangle_{\otimes^m T_y^*\mathbb{H}^{n+1}} \\ &= \langle \otimes^m \mathcal{C}_\eta^* \otimes^m (\mathcal{A}_+^{-1}(x, \xi)^T)\zeta_+, \otimes^m (\mathcal{A}_-^{-1}(x, \xi)^T)\zeta_- \rangle_{\otimes^m \mathcal{E}^*(x, \xi)}. \end{aligned} \quad (5-30)$$

The Jacobian of Ψ with respect to naturally arising volume forms on \mathcal{E} and $S^2_\Delta \mathbb{H}^{n+1}$ is given by (see Appendix A2 for the proof)

$$J_\Psi(x, \xi, \eta) = 2^n (1 + |\eta|^2)^{-n}. \tag{5-31}$$

Now, Ψ is equivariant under G , therefore it descends to a diffeomorphism

$$\Psi : \mathcal{E}_M \rightarrow S^2_\Delta M, \quad \mathcal{E}_M := \{(x, \xi, \eta) \mid (x, \xi) \in SM, \eta \in \mathcal{E}(x, \xi)\}.$$

Using Lemma 5.6 and (5-30), we calculate, for $(x, \xi, \eta) \in \mathcal{E}_M$ and $(y, \eta_-, \eta_+) = \Psi(x, \xi, \eta)$,

$$\langle v(y, \eta_-), \overline{v^*(y, \eta_+)} \rangle_{\otimes^m T^*_y M} = (1 + |\eta|^2)^{-\lambda} \langle \otimes^m \mathcal{C}_\eta^* v(x, \xi), \overline{v^*(x, \xi)} \rangle_{\otimes^m \mathcal{E}^*(x, \xi)}. \tag{5-32}$$

We would now like to plug this expression into (5-28), make the change of variables from (y, η_-, η_+) to (x, ξ, η) , and integrate η out, obtaining a multiple of (5-29). However, this is not directly possible because (i) the integral in η typically diverges and (ii) since the expression integrated in (5-28) is a distribution, one cannot simply replace $S^2 M$ by $S^2_\Delta M$ in the integral.

We will instead use the asymptotic behavior of both integrals as one approaches the set $\{\eta_+ + \eta_- = 0\}$, and Hadamard regularization in η in the (x, ξ, η) variables. For that, fix $\chi \in C_0^\infty(\mathbb{R})$ such that $\chi = 1$ near 0, and define, for $\varepsilon > 0$,

$$\chi_\varepsilon(y, \eta_-, \eta_+) = \chi(\varepsilon |\eta(y, \eta_-, \eta_+)|),$$

where $\eta(y, \eta_-, \eta_+)$ is the corresponding component of Ψ^{-1} ; we can write

$$\chi_\varepsilon(y, \eta_-, \eta_+) = \chi\left(\varepsilon \frac{|\eta_+ - \eta_-|}{|\eta_+ + \eta_-|}\right).$$

Then $\chi_\varepsilon \in \mathcal{D}'(S^2 M)$. In fact, χ_ε is supported inside $S^2_\Delta M$; by making the change of variables $(y, \eta_-, \eta_+) = \Psi(x, \xi, \eta)$ and, using (5-31) and (5-32), we get

$$\begin{aligned} \int_{S^2 M} \chi_\varepsilon(y, \eta_-, \eta_+) \langle v(y, \eta_-), \overline{v^*(y, \eta_+)} \rangle_{\otimes^m T^*_y M} dy d\eta_- d\eta_+ \\ = 2^n \int_{\mathcal{E}_M} \chi(\varepsilon |\eta|) (1 + |\eta|^2)^{-\lambda-n} \langle \otimes^m \mathcal{C}_\eta^* v(x, \xi), \overline{v^*(x, \xi)} \rangle_{\otimes^m \mathcal{E}^*(x, \xi)} dx d\xi d\eta. \end{aligned} \tag{5-33}$$

By Lemma A.4, (5-33) has the asymptotic expansion

$$2^n \pi^{n/2} \frac{\Gamma(\frac{1}{2}n + \lambda)}{(n + \lambda + m - 1)\Gamma(n - 1 + \lambda)} \langle v, v^* \rangle_{L^2(SM; \otimes^m \mathcal{E}^*)} + \sum_{0 \leq j \leq -\operatorname{Re} \lambda - n/2} c_j \varepsilon^{n+2\lambda+2j} + o(1) \tag{5-34}$$

for some constants c_j .

It remains to prove the following asymptotic expansion as $\varepsilon \rightarrow 0$:

$$\int_{S^2 M} (1 - \chi_\varepsilon(y, \eta_-, \eta_+)) \langle v(y, \eta_-), \overline{v^*(y, \eta_+)} \rangle_{\otimes^m T^*_y M} dy d\eta_- d\eta_+ \sim \sum_{j=0}^\infty c'_j \varepsilon^{n+2\lambda+2j}, \tag{5-35}$$

where the c'_j are some constants. Indeed, $\langle f, f^* \rangle_{L^2(M; \otimes^m T^* M)}$ is equal to the sum of (5-33) and (5-35); since (5-35) does not have a constant term, $\langle f, f^* \rangle$ is equal to the constant term in the expansion (5-34).

To show (5-35), we use the dilation vector field $\eta \cdot \partial_\eta$ on \mathcal{E} , which under Ψ becomes the following vector field on $S^2_\Delta M$ extending smoothly to $S^2 M$:

$$L_{(y, \eta_-, \eta_+)} = \left(\frac{1}{2}(\eta_- - \eta_+), \frac{1}{4}|\eta_+ - \eta_-|^2 y - \frac{1}{2}\eta_+ + \frac{1}{2}(\eta_- \cdot \eta_+)\eta_-, -\frac{1}{4}|\eta_+ - \eta_-|^2 y - \frac{1}{2}\eta_- + \frac{1}{2}(\eta_- \cdot \eta_+)\eta_+ \right).$$

The vector field L is tangent to the submanifold $\{\eta_+ + \eta_- = 0\}$; in fact,

$$L(|\eta_+ - \eta_-|^2) = -L(|\eta_+ + \eta_-|^2) = \frac{1}{2}|\eta_+ - \eta_-|^2 \cdot |\eta_+ + \eta_-|^2.$$

We can then compute (following the identity $L|\eta| = |\eta|$)

$$L\left(\frac{|\eta_+ - \eta_-|}{|\eta_+ + \eta_-|}\right) = \frac{|\eta_+ - \eta_-|}{|\eta_+ + \eta_-|} \quad \text{on } S^2_\Delta M.$$

Using the (x, ξ, η) coordinates and (5-31), we can compute the divergence of L with respect to the standard volume form on $S^2 M$:

$$\text{Div } L = n(\eta_+ \cdot \eta_-).$$

Moreover, $B_\pm(y, \eta_\pm)$ are constant along the trajectories of L , and

$$L(\Phi_\pm(y, \eta_\pm)) = -\frac{1}{4}|\eta_+ - \eta_-|^2 \Phi_\pm(y, \eta_\pm).$$

We also use (3-31) to calculate, for $\zeta_\pm \in T_{B_\pm(y, \eta_\pm)} S^n$,

$$\begin{aligned} L((\mathcal{A}_+(y, \eta_+)\zeta_+) \cdot (\mathcal{A}_-(y, \eta_-)\zeta_-)) &= ((\mathcal{A}_+(y, \eta_+)\zeta_+) \cdot \eta_-)((\mathcal{A}_-(y, \eta_-)\zeta_-) \cdot \eta_+), \\ L((\mathcal{A}_\pm(y, \eta_\pm)\zeta_\pm) \cdot \eta_\mp) &= (\eta_+ \cdot \eta_-)((\mathcal{A}_\pm(y, \eta_\pm)\zeta_\pm) \cdot \eta_\mp). \end{aligned}$$

Combining these identities and using Lemma 5.6, we get

$$(L + \frac{1}{2}\lambda|\eta_+ - \eta_-|^2)\langle v(y, \eta_-), \overline{v^*(y, \eta_+)} \rangle_{\otimes^m T_y^* M} = m\langle \iota_{\eta_+} v(y, \eta_-), \iota_{\eta_-} \overline{v^*(y, \eta_+)} \rangle_{\otimes^{m-1} T_y^* M}. \quad (5-36)$$

Integrating by parts, we find

$$\begin{aligned} \varepsilon \partial_\varepsilon \int_{S^2 M} (1 - \chi_\varepsilon(y, \eta_-, \eta_+)) \langle v(y, \eta_-), \overline{v^*(y, \eta_+)} \rangle_{\otimes^m T_y^* M} dy d\eta_- d\eta_+ \\ = \int_{S^2 M} L(1 - \chi_\varepsilon(y, \eta_-, \eta_+)) \langle v(y, \eta_-), \overline{v^*(y, \eta_+)} \rangle_{\otimes^m T_y^* M} dy d\eta_- d\eta_+ \\ = \int_{S^2 M} \left(\frac{1}{2}\lambda|\eta_+ - \eta_-|^2 - n(\eta_+ \cdot \eta_-) \right) (1 - \chi_\varepsilon(y, \eta_-, \eta_+)) \langle v(y, \eta_-), \overline{v^*(y, \eta_+)} \rangle_{\otimes^m T_y^* M} dy d\eta_- d\eta_+ \\ \quad - m \int_{S^2 M} (1 - \chi_\varepsilon(y, \eta_-, \eta_+)) \langle \iota_{\eta_+} v(y, \eta_-), \iota_{\eta_-} \overline{v^*(y, \eta_+)} \rangle_{\otimes^{m-1} T_y^* M} dy d\eta_- d\eta_+. \end{aligned}$$

Arguing similarly, we see that if, for integers $0 \leq r \leq m$, $p \geq 0$, we put

$$I_{r,p}(\varepsilon) := \int_{S^2 M} |\eta_- + \eta_+|^{2p} (1 - \chi_\varepsilon(y, \eta_-, \eta_+)) \langle \iota_{\eta_+}^r v(y, \eta_-), \iota_{\eta_-}^r \overline{v^*(y, \eta_+)} \rangle_{\otimes^{m-r} T_y^* M} dy d\eta_- d\eta_+,$$

then $(\varepsilon\partial_\varepsilon - 2\lambda - n - 2(r + p))I_{r,p}(\varepsilon)$ is a finite linear combination of $I_{r',p'}(\varepsilon)$, where $r' \geq r$, $p' \geq p$, and $(r', p') \neq (r, p)$. For example, the calculation above shows that

$$(\varepsilon\partial_\varepsilon - 2\lambda - n)I_{0,0}(\varepsilon) = -\frac{1}{2}(\lambda + n)I_{0,1}(\varepsilon) - mI_{1,0}(\varepsilon).$$

Moreover, if N is fixed and p is large enough depending on N , then $I_{r,p}(\varepsilon) = \mathcal{O}(\varepsilon^N)$; to see this, note that $I_{r,p}(\varepsilon)$ is bounded by some fixed C^∞ -seminorm of $|\eta_- + \eta_+|^{2p}(1 - \chi_\varepsilon(y, \eta_-, \eta_+))$. It follows that, if N is fixed and \tilde{N} is large depending on N , then

$$\left(\prod_{j=0}^{\tilde{N}} (\varepsilon\partial_\varepsilon - 2\lambda - n - 2j) \right) I_{0,0}(\varepsilon) = \mathcal{O}(\varepsilon^N),$$

which implies the existence of the decomposition (5-35) and finishes the proof. □

6. Properties of the Laplacian

In this section, we introduce the Laplacian and study its basic properties (Section 6A). We then give formulas for the Laplacian on symmetric tensors in the half-plane model (Section 6B), which will be the basis for the analysis of the following sections. Using these formulas, we study the Poisson kernel and in particular prove Lemma 5.8 and the injectivity of the Poisson kernel (Section 6C).

6A. Definition and Bochner identity. The Levi-Civita connection associated to the hyperbolic metric g_H is the operator

$$\nabla : \mathcal{C}^\infty(\mathbb{H}^{n+1}, T\mathbb{H}^{n+1}) \rightarrow \mathcal{C}^\infty(\mathbb{H}^{n+1}, T^*\mathbb{H}^{n+1} \otimes T\mathbb{H}^{n+1}),$$

which induces a natural covariant derivative, still denoted ∇ , on sections of $\otimes^m T^*\mathbb{H}^{n+1}$. We can work in the ball model \mathbb{B}^{n+1} and use the 0-tangent structure (see Section 3F), and nabla can be viewed as a differential operator of order 1:

$$\nabla : \mathcal{C}^\infty(\bar{\mathbb{B}}^{n+1}; \otimes^m ({}^0T^*\bar{\mathbb{B}}^{n+1})) \rightarrow \mathcal{C}^\infty(\bar{\mathbb{B}}^{n+1}, \otimes^{m+1} ({}^0T^*\bar{\mathbb{B}}^{n+1})).$$

We denote by ∇^* its adjoint with respect to the L^2 scalar product, called the *divergence*; it is given by $\nabla^*u = -\mathcal{T}(\nabla u)$, where \mathcal{T} denotes the trace; see Section 4A. Define the rough Laplacian acting on $\mathcal{C}^\infty(\bar{\mathbb{B}}^{n+1}; \otimes^m ({}^0T^*\bar{\mathbb{B}}^{n+1}))$ by

$$\Delta := \nabla^*\nabla; \tag{6-1}$$

this operator maps symmetric tensors to symmetric tensors. It also extends to $\mathcal{D}'(\mathbb{B}^{n+1}; \otimes_S^m ({}^0T^*\bar{\mathbb{B}}^{n+1}))$ by duality. The operator Δ commutes with \mathcal{T} and \mathcal{I} :

$$\Delta\mathcal{T}(u) = \mathcal{T}(\Delta u) \quad \text{and} \quad \Delta\mathcal{I}(u) = \mathcal{I}(\Delta u) \tag{6-2}$$

for all $u \in \mathcal{D}'(\mathbb{B}^{n+1}; \otimes_S^m ({}^0T^*\bar{\mathbb{B}}^{n+1}))$.

There is another natural operator given by

$$\Delta_D = D^*D$$

if

$$D : C^\infty(\bar{\mathbb{B}}^{n+1}; \otimes_S^m({}^0T^*\bar{\mathbb{B}}^{n+1})) \rightarrow C^\infty(\bar{\mathbb{B}}^{n+1}; \otimes_S^{m+1}({}^0T^*\bar{\mathbb{B}}^{n+1}))$$

is defined by $D := \mathcal{S} \circ \nabla$, where \mathcal{S} is the symmetrization defined by (4-1), and $D^* = \nabla^*$ is the formal adjoint. There is a Bochner–Weitzenböck formula relating Δ and Δ_D , and, using that the curvature is constant, we have on trace-free symmetric tensors of order m , by [Dairbekov and Sharafutdinov 2010, Lemma 8.2],

$$\Delta_D = \frac{1}{m+1}(mDD^* + \Delta + m(m+n-1)). \tag{6-3}$$

In particular, since $|\mathcal{S}\nabla u|^2 \leq |\nabla u|^2$ pointwise by the fact that \mathcal{S} is an orthogonal projection, we see that, for u smooth and compactly supported, $\|Du\|_{L^2}^2 \leq \|\nabla u\|_{L^2}^2$ and thus, for $m \geq 1$, $u \in C_0^\infty(\mathbb{H}^{n+1}; \otimes_S^m(T^*\mathbb{H}^{n+1}))$, and $\mathcal{T}u = 0$,

$$\langle \Delta u, u \rangle_{L^2} \geq (m+n-1)\|u\|^2. \tag{6-4}$$

Since the Bochner identity is local, the same inequality clearly descends to cocompact quotients $\Gamma \backslash \mathbb{H}^{n+1}$ (where Δ is selfadjoint and has compact resolvent by standard theory of elliptic operators, as its principal part is given by the scalar Laplacian), and this implies:

Lemma 6.1. *The spectrum of Δ acting on trace-free symmetric tensors of order $m \geq 1$ on hyperbolic compact manifolds of dimension $n + 1$ is bounded below by $m + n - 1$.*

We finally define

$$E^{(m)} := \otimes_S^m({}^0T^*\bar{\mathbb{B}}^{n+1}) \cap \ker \mathcal{T} \tag{6-5}$$

to be the bundle of trace-free symmetric m -cotensors over the ball model of hyperbolic space.

6B. Laplacian in the half-plane model. We now give concrete formulas concerning the Laplacian on symmetric tensors in the half-space model \mathbb{U}^{n+1} (see (3-4)). We fix $\nu \in \mathbb{S}^n$ and map \mathbb{B}^{n+1} to \mathbb{U}^{n+1} by a composition of a rotation of \mathbb{B}^{n+1} and the map (3-5); the rotation is chosen so that ν is mapped to $0 \in \bar{\mathbb{U}}^{n+1}$ and $-\nu$ is mapped to infinity.

The 0-cotangent and tangent bundles ${}^0T^*\bar{\mathbb{B}}^{n+1}$ and ${}^0T\bar{\mathbb{B}}^{n+1}$ pull back to the half-space; we denote them ${}^0T^*\mathbb{U}^{n+1}$ and ${}^0T\mathbb{U}^{n+1}$. The coordinates on \mathbb{U}^{n+1} are $(z_0, z) \in \mathbb{R}^+ \times \mathbb{R}^n$ and $z = (z_1, \dots, z_n)$. We use the following orthonormal bases of ${}^0T\mathbb{U}^{n+1}$ and ${}^0T^*\mathbb{U}^{n+1}$:

$$Z_i = z_0 \partial_{z_i} \quad \text{and} \quad Z_i^* = \frac{dz_i}{z_0}, \quad 0 \leq i \leq n.$$

Note that in the compactification $\bar{\mathbb{B}}^{n+1}$ this basis is smooth only on $\bar{\mathbb{B}}^{n+1} \setminus \{-\nu\}$.

Let $\mathcal{A} := \{1, \dots, n\}$. We can decompose the vector bundle $\otimes_S^m({}^0T^*\mathbb{U}^{n+1})$ into an orthogonal direct sum

$$\otimes_S^m({}^0T^*\mathbb{U}^{n+1}) = \bigoplus_{k=0}^m E_k^{(m)}, \quad E_k^{(m)} = \text{span}(\mathcal{S}((Z_0^*)^{\otimes k} \otimes Z_I^*)_{I \in \mathcal{A}^{m-k}}),$$

and we let π_i be the orthogonal projection onto $E_i^{(m)}$. Now, each tensor $u \in \otimes_S^m ({}^0T^*\mathbb{U}^{n+1})$ can be decomposed as $u = \sum_{i=0}^m u_i$ with $u_i = \pi_i(u) \in E_i^{(m)}$ which we can write as

$$u = \sum_{i=0}^m u_i, \quad u_i = \mathcal{S}((Z_0^*)^{\otimes i} \otimes u'_i), \quad u'_i \in E_0^{(m-i)}. \tag{6-6}$$

We can therefore identify $E_k^{(m)}$ with $E_0^{(m-k)}$ and view $E^{(m)}$ as a direct sum $E^{(m)} = \bigoplus_{k=0}^m E_0^{(m-k)}$. The trace-free condition, $\mathcal{T}(u) = 0$, is equivalent to the relations

$$\mathcal{T}(u'_r) = -\frac{(r+2)(r+1)}{(m-r)(m-r-1)} u'_{r+2}, \quad 0 \leq r \leq m-2, \tag{6-7}$$

and, in particular, all u_i are determined by u_0 and u_1 by iterating the trace map \mathcal{T} . The u'_i are related to the elements in the decomposition (4-5) of u_0 and u_1 viewed as a symmetric m -cotensor on the bundle $(Z_0)^\perp$ using the metric $z_0^{-2}h = \sum_i Z_i^* \otimes Z_i^*$. We see that a nonzero trace-free tensor u on \mathbb{U}^{n+1} must have a nonzero u_0 or u_1 component.

The Koszul formula gives us, for $i, j \geq 1$,

$$\nabla_{Z_i} Z_j = \delta_{ij} Z_0, \quad \nabla_{Z_0} Z_j = 0, \quad \nabla_{Z_i} Z_0 = -Z_i, \quad \nabla_{Z_0} Z_0 = 0, \tag{6-8}$$

which implies

$$\nabla Z_0^* = -\sum_{j=1}^n Z_j^* \otimes Z_j^* = -\frac{h}{z_0}, \quad \nabla Z_j^* = Z_j^* \otimes Z_0^*. \tag{6-9}$$

We shall use the following notations: If Π_m denotes the set of permutations of $\{1, \dots, m\}$, we write $\sigma(I) := (i_{\sigma(1)}, \dots, i_{\sigma(m)})$ if $\sigma \in \Pi_m$. If $S = S_1 \otimes \dots \otimes S_\ell$ is a tensor in $\otimes^\ell ({}^0T^*\mathbb{U}^{n+1})$, we denote by $\tau_{i \leftrightarrow j}(S)$ the tensor obtained by permuting S_i with S_j in S , and by $\rho_{i \rightarrow V}(S)$ the operation of replacing S_i by $V \in {}^0T^*\mathbb{U}^{n+1}$ in S .

The Laplacian and ∇^* acting on $E_0^{(m)}$ and $E_1^{(m)}$. We start by computing the action of Δ on sections of $E_0^{(m)}$ and $E_1^{(m)}$, and we will later deduce from this computation the action on $E_k^{(m)}$. Let us consider the tensor $Z_I^* := Z_{i_1}^* \otimes \dots \otimes Z_{i_m}^* \in E_0^{(m)}$, where $I = (i_1, \dots, i_m) \in \mathcal{A}^m$ and $Z_{\sigma(I)}^* := Z_{i_{\sigma(1)}}^* \otimes \dots \otimes Z_{i_{\sigma(m)}}^*$. The symmetrization of Z_I^* is given by $\mathcal{S}(Z_I^*) = (1/m!) \sum_{\sigma \in \Pi_m} Z_{\sigma(I)}^*$ and those elements form a basis of the space $E_0^{(m)}$ when I ranges over all combinations of m -tuples in $\mathcal{A} = \{1, \dots, n\}$.

Lemma 6.2. *Let $u_0 = \sum_{I \in \mathcal{A}^m} f_I \mathcal{S}(Z_I^*)$ with $f_I \in C^\infty(\mathbb{U}^{n+1})$. Then one has*

$$\Delta u_0 = \sum_{I \in \mathcal{A}^m} ((\Delta + m) f_I) \mathcal{S}(Z_I^*) + 2m \mathcal{S}(\nabla^* u_0 \otimes Z_0^*) + m(m-1) \mathcal{S}(\mathcal{T}(u_0) \otimes Z_0^* \otimes Z_0^*), \tag{6-10}$$

while, denoting $d_z f_I = \sum_{i=1}^n Z_i(f_I) Z_i^*$, the divergence is given by

$$\nabla^* u_0 = -(m-1) \mathcal{S}(\mathcal{T}(u_0) \otimes Z_0^*) - \sum_{I \in \mathcal{A}^m} \iota_{d_z f_I} \mathcal{S}(Z_I^*). \tag{6-11}$$

Proof. Using (6-9), we compute

$$\nabla(f_I \mathcal{S}(Z_I^*)) = \sum_{i=0}^n (Z_i f_I)(z) Z_i^* \otimes \mathcal{S}(Z_I^*) + \frac{f_I(z)}{m!} \sum_{k=1}^m \sum_{\sigma \in \Pi_m} \tau_{1 \leftrightarrow k+1}(Z_0^* \otimes Z_{\sigma(I)}^*).$$

Then, taking the trace of $\nabla(f_I \mathcal{S}(Z_I^*))$ gives

$$\begin{aligned} \nabla^*(f_I \mathcal{S}(Z_I^*)) &= -\frac{f_I}{m!} \sum_{k=2}^m \sum_{\sigma \in \Pi_m} \delta_{i_{\sigma(1)}, i_{\sigma(k)}} \rho_{k-1 \rightarrow Z_0^*}(Z_{i_{\sigma(2)}}^* \otimes \cdots \otimes Z_{i_{\sigma(m)}}^*) \\ &\quad - \sum_{i=1}^n (Z_i f_I) \frac{1}{m!} \sum_{\sigma \in \Pi_m} \delta_{i, i_{\sigma(1)}}(Z_{i_{\sigma(2)}}^* \otimes \cdots \otimes Z_{i_{\sigma(m)}}^*). \end{aligned} \quad (6-12)$$

We notice that $\mathcal{S}(\mathcal{T}(\mathcal{S}(Z_I^*)) \otimes Z_0^*)$ is given by

$$\mathcal{S}(\mathcal{T}(\mathcal{S}(Z_I^*)) \otimes Z_0^*) = \frac{1}{m!(m-1)} \sum_{\sigma \in \Pi_m} \sum_{k=1}^{m-1} \delta_{i_{\sigma(1)}, i_{\sigma(2)}} \tau_{1 \leftrightarrow k}(Z_0^* \otimes Z_{i_{\sigma(3)}}^* \otimes \cdots \otimes Z_{i_{\sigma(m)}}^*),$$

which implies (6-11). Let us now compute $\nabla^2(f_I \mathcal{S}(Z_I^*))$:

$$\begin{aligned} &\nabla^2(f_I \mathcal{S}(Z_I^*)) \\ &= \sum_{i,j=0}^n Z_j Z_i (f_I) Z_j^* \otimes Z_i^* \otimes \mathcal{S}(Z_I^*) - Z_0(f_I) z_0^{-2} h \otimes \mathcal{S}(Z_I^*) \\ &\quad + \sum_{j=1}^n Z_j (f_I) Z_j^* \otimes Z_0^* \otimes \mathcal{S}(Z_I^*) + \frac{Z_0(f_I)}{m!} \sum_{\sigma \in \Pi_m} \sum_{k=1}^m \tau_{1 \leftrightarrow k+2}(Z_0^* \otimes Z_0^* \otimes Z_{\sigma(I)}^*) \\ &\quad + \sum_{i=1}^n \frac{Z_i(f_I)}{m!} \sum_{\sigma \in \Pi_m} \sum_{k=1}^m \tau_{1 \leftrightarrow k+2}(Z_0^* \otimes Z_i^* \otimes Z_{\sigma(I)}^*) + \sum_{i=1}^n \frac{Z_i(f_I)}{m!} \sum_{\sigma \in \Pi_m} \sum_{k=1}^m \tau_{2 \leftrightarrow k+2}(Z_i^* \otimes Z_0^* \otimes Z_{\sigma(I)}^*) \\ &\quad + \frac{Z_0(f_I)}{m!} Z_0^* \otimes \sum_{k=1}^m \sum_{\sigma \in \Pi_m} \tau_{1 \leftrightarrow k+1}(Z_0^* \otimes Z_{\sigma(I)}^*) \\ &\quad - \frac{f_I}{m!} \sum_{j=1}^n Z_j^* \otimes \sum_{\sigma \in \Pi_m} \sum_{k=1}^m \tau_{1 \leftrightarrow k+1}(Z_j^* \otimes Z_{\sigma(I)}^*) + \frac{f_I}{m!} \sum_{k=1}^m \sum_{\substack{\ell=1 \\ \ell \neq k+1}}^{m+1} \tau_{1 \leftrightarrow \ell+1}(Z_0^* \otimes \tau_{1 \leftrightarrow k+1}(Z_0^* \otimes Z_{\sigma(I)}^*)). \end{aligned}$$

We then take the trace: the first line on the right-hand side has trace $-(\Delta f_I) \mathcal{S}(Z_I^*)$, the second and fourth lines have vanishing trace, the first term on the last line has trace $-m f_I \mathcal{S}(Z_I^*)$, the last term has trace

$$\frac{2f_I}{m!} \sum_{\sigma \in \Pi_m} \sum_{1 \leq k < \ell \leq m} \delta_{i_{\sigma(k)}, i_{\sigma(\ell)}} \rho_{k \rightarrow Z_0^*} \rho_{\ell \rightarrow Z_0^*}(Z_{\sigma(I)}^*), \quad (6-13)$$

and the third line has total trace

$$2 \sum_{i=1}^n \frac{Z_i(f_I)}{m!} \sum_{\sigma \in \Pi_m} \sum_{k=1}^m \delta_{i, i_{\sigma(k)}} \rho_{k \rightarrow Z_0^*}(Z_{\sigma(I)}^*). \quad (6-14)$$

Computing $\mathcal{S}(\mathcal{T}(\mathcal{S}(Z_I^*)) \otimes Z_0^* \otimes Z_0^*)$ gives

$$\begin{aligned} &\mathcal{S}(\mathcal{T}(\mathcal{S}(Z_I^*)) \otimes Z_0^* \otimes Z_0^*) \\ &= \frac{2}{m!m(m-1)} \sum_{1 \leq k < \ell \leq m} \sum_{\sigma \in \Pi_m} \delta_{i_{\sigma(1)}, i_{\sigma(2)}} \tau_{1 \leftrightarrow k+2} \tau_{2 \leftrightarrow \ell+2} (Z_0^* \otimes Z_0^* \otimes Z_{i_{\sigma(3)}}^* \otimes \cdots \otimes Z_{i_{\sigma(m)}}^*); \end{aligned}$$

therefore the term (6-13) can be simplified to

$$m(m-1) f_I \mathcal{S}(\mathcal{T}(\mathcal{S}(Z_I^*)) \otimes Z_0^* \otimes Z_0^*).$$

Similarly, to simplify (6-14), we compute

$$\begin{aligned} &\mathcal{S}(\nabla^*(f_I \mathcal{S}(Z_I^*)) \otimes Z_0^*) \\ &= -(m-1) \mathcal{S}(\mathcal{T}(f_I \mathcal{S}(Z_I^*)) \otimes Z_0^* \otimes Z_0^*) - \sum_{i=1}^n (Z_i f_I) \frac{1}{m!m} \sum_{k=1}^m \sum_{\sigma \in \Pi_m} \delta_{i, i_{\sigma(1)}} \tau_{1 \leftrightarrow k} (Z_0^* \otimes Z_{i_{\sigma(2)}}^* \otimes \cdots \otimes Z_{i_{\sigma(m)}}^*), \end{aligned}$$

so that

$$\begin{aligned} &2 \sum_{i=1}^n \frac{Z_i(f_I)}{m!} \sum_{\sigma \in \Pi_m} \sum_{k=1}^m \delta_{i, i_{\sigma(k)}} \rho_{k \rightarrow Z_0^*} (Z_{\sigma(I)}^*) \\ &= -2m \mathcal{S}(\nabla^*(f_I \mathcal{S}(Z_I^*)) \otimes Z_0^*) - 2m(m-1) \mathcal{S}(\mathcal{T}(f_I \mathcal{S}(Z_I^*)) \otimes Z_0^* \otimes Z_0^*), \end{aligned}$$

and this achieves the proof of (6-10). □

A similarly tedious calculation, omitted here, yields:

Lemma 6.3. *Let $u_1 = \mathcal{S}(Z_0^* \otimes u'_1)$, $u'_1 = \sum_{J \in \mathcal{A}^{m-1}} g_J \mathcal{S}(Z_J^*)$ with $g_J \in \mathcal{C}^\infty(\mathbb{U}^{n+1})$; then the $E_0^{(m)} \oplus E_1^{(m)}$ components of the Laplacian of u_1 are*

$$\Delta u_1 = \sum_{J \in \mathcal{A}^{m-1}} ((\Delta + n + 3(m-1))g_J) \mathcal{S}(Z_0^* \otimes Z_J^*) + 2 \sum_{J \in \mathcal{A}^{m-1}} \mathcal{S}(d_z g_J \otimes Z_J^*) + \text{Ker}(\pi_0 + \pi_1) \quad (6-15)$$

and the $E_0^{(m)} \oplus E_1^{(m)}$ components of divergence of u_1 are

$$\nabla^* u_1 = \frac{1}{m} \sum_{J \in \mathcal{A}^{m-1}} ((n+m-1)g_J - Z_0(g_J)) \mathcal{S}(Z_J^*) - \frac{m-1}{m} \sum_{J \in \mathcal{A}^{m-1}} \mathcal{S}(Z_0^* \otimes \iota_{d_z g_J} \mathcal{S}(Z_J^*)) + \text{Ker}(\pi_0 + \pi_1). \quad (6-16)$$

General formulas for Laplacian and divergence. Armed with Lemmas 6.2 and 6.3, we can show the following fact, which, together with (6-7), completely determines the Laplacian on trace-free symmetric tensors.

Lemma 6.4. *Assume that $u \in \mathcal{D}'(\mathbb{U}^{n+1}; \otimes_S^m T^* \mathbb{U}^{n+1})$ satisfies $\mathcal{T}(u) = 0$ and is written in the form (6-6).*

Let

$$u_0 = \sum_{I \in \mathcal{A}^m} f_I \mathcal{S}(Z_I^*), \quad u_1 = \sum_{J \in \mathcal{A}^{m-1}} g_J \mathcal{S}(Z_0 \otimes Z_J^*).$$

Then the projection of Δu onto $E_0^{(m)} \oplus E_1^{(m)}$ can be written

$$\pi_0(\Delta u) = \sum_{I \in \mathcal{A}^m} ((\Delta + m)f_I) \mathcal{S}(Z_I^*) + 2 \sum_{J \in \mathcal{A}^{m-1}} \mathcal{S}(d_z g_J \otimes Z_J^*) + m(m-1) \mathcal{S}(z_0^{-2} h \otimes \mathcal{T}(u_0)), \quad (6-17)$$

$$\begin{aligned} \pi_1(\Delta u) = & \sum_{J \in \mathcal{A}^{m-1}} ((\Delta + n + 3(m-1))g_J) \mathcal{S}(Z_0^* \otimes Z_J^*) - 2m \sum_{I \in \mathcal{A}^m} \mathcal{S}(Z_0^* \otimes \iota_{d_z f_I} \mathcal{S}(Z_I^*)) \\ & + (m-1)(m-2) \mathcal{S}(Z_0^* \otimes z_0^{-2} h \otimes \mathcal{T}(u_1')) - 2m(m-1) \sum_{I \in \mathcal{A}^m} \mathcal{S}(Z_0^* \otimes d_z f_I \otimes \mathcal{T}(\mathcal{S}(Z_I^*))). \end{aligned} \quad (6-18)$$

Proof. First, it is easily seen from (6-9) that Δu_k is a section of $\bigoplus_{j=k-2}^{k+2} E_j^{(m)}$. From Lemmas 6.2 and 6.3, we have

$$\pi_0(\Delta(u_0 + u_1)) = \sum_{I \in \mathcal{A}^m} ((\Delta + m)f_I) \mathcal{S}(Z_I^*) + 2 \sum_{J \in \mathcal{A}^{m-1}} \mathcal{S}(d_z g_J \otimes Z_J^*). \quad (6-19)$$

Then, for u_2 , using $\mathcal{S}((Z_0^*)^{\otimes 2} \otimes u_2') = \mathcal{S}(g_H \otimes u_2') - \mathcal{S}(z_0^{-2} h \otimes u_2')$ and $\Delta \mathcal{I} = \mathcal{I} \Delta$,

$$\pi_0(\Delta u_2) = \pi_0(\mathcal{S}(z_0^{-2} h \otimes \Delta u_2')) - \pi_0(\Delta(\mathcal{S}(z_0^{-2} h \otimes u_2')))$$

and, writing $u_2' = -\frac{1}{2}m(m-1)\mathcal{T}(u_0)$ by (6-7), we obtain, using (6-10),

$$\pi_0(\Delta u_2) = m(m-1) \mathcal{S}(z_0^{-2} h \otimes \mathcal{T}(u_0)). \quad (6-20)$$

We therefore obtain (6-17).

Now we consider the projection on $E_1^{(m)}$ of the equation $(\Delta - s)T = 0$. We have, from (6-10),

$$\pi_1(\Delta u_0) = -2m \sum_{I \in \mathcal{A}^m} \mathcal{S}(Z_0^* \otimes \iota_{d_z f_I} \mathcal{S}(Z_I^*)),$$

where $\iota_{d_z f_I}$ means $\sum_{j=1}^n Z_j(f_I) \iota_{Z_j}$. Then, from (6-15),

$$\pi_1(\Delta u_1) = \sum_{J \in \mathcal{A}^{m-1}} ((\Delta + n + 3(m-1))g_J) \mathcal{S}(Z_0^* \otimes Z_J^*).$$

Using again $\mathcal{S}((Z_0^*)^{\otimes 2} \otimes u_2') = \mathcal{S}(g_H \otimes u_2') - \mathcal{S}(z_0^{-2} h \otimes u_2')$ and $\Delta \mathcal{I} = \mathcal{I} \Delta$, (6-10) gives

$$\pi_1(\Delta u_2) = -2m(m-1) \sum_{I \in \mathcal{A}^m} \mathcal{S}(Z_0^* \otimes d_z f_I \otimes \mathcal{T}(\mathcal{S}(Z_I^*))).$$

Finally, we compute $\pi_1(\Delta u_3)$: using the computation (6-15), we get

$$\begin{aligned} \pi_1(\Delta u_3) &= \pi_1(\mathcal{S}(z_0^{-2} h \otimes \Delta \mathcal{S}(Z_0^* \otimes u_3')) - \pi_1(\Delta \mathcal{S}(Z_0^* \otimes z_0^{-2} h \otimes u_3')) \\ &= (m-1)(m-2) \mathcal{S}(Z_0^* \otimes z_0^{-2} h \otimes \mathcal{T}(u_1')). \end{aligned}$$

We conclude that $\pi_1(\Delta u)$ is given by (6-18). \square

Similarly, we also have:

Lemma 6.5. *Let u be as in Lemma 6.4. Then the projection onto $E_0^{(m-1)} \oplus E_1^{(m-1)}$ of the divergence of u is given by*

$$\pi_0(\nabla^* u) = - \sum_{I \in \mathcal{I}^m} \iota_{d_z f_I} \mathcal{S}(Z_I^*) + \frac{1}{m} \sum_{J \in \mathcal{I}^{m-1}} ((n+m-1)g_J - Z_0(g_J)) \mathcal{S}(Z_J^*), \tag{6-21}$$

$$\pi_1(\nabla^* u) = (m-1) \sum_{I \in \mathcal{I}^m} (Z_0 f_I - (m+n-1)f_I) \mathcal{S}(\mathcal{T}(\mathcal{S}(Z_I^*)) \otimes Z_0^*) - \frac{m-1}{m} \sum_{J \in \mathcal{I}^{m-1}} \mathcal{S}(Z_0^* \otimes \iota_{d_z g_J} \mathcal{S}(Z_J^*)). \tag{6-22}$$

Proof. The π_0 part follows from (6-11) and (6-16). For the π_1 part, we also use (6-11) and (6-16) but we need to see the contribution from $\nabla^* u_2$ as well. For that, we write $u'_2 = -\frac{1}{2}m(m-1) \sum_{I \in \mathcal{I}^m} f_I \mathcal{T}(\mathcal{S}(Z_I^*))$, as before, and a direct calculation shows that

$$\pi_1(\nabla^* u_2) = (m-1) \sum_{I \in \mathcal{I}^m} (Z_0 f_I - (m+n-2)f_I) \mathcal{S}(\mathcal{T}(\mathcal{S}(Z_I^*)) \otimes Z_0^*),$$

implying the desired result. □

6C. Properties of the Poisson kernel. In this section, we study the Poisson kernel \mathcal{P}_λ^- defined by (5-17).

Pairing on the sphere. We start by proving the following formula:

Lemma 6.6. *Let $\lambda \in \mathbb{C}$ and $w \in \mathcal{D}'(\mathbb{S}^n; \otimes_S^m(T^*\mathbb{S}^n))$. Then*

$$\mathcal{P}_\lambda^- w(x) = \int_{\mathbb{S}^n} P(x, \nu)^{n+\lambda} (\otimes^m (\mathcal{A}_-^{-1}(x, \xi_-(x, \nu)))^T) w(\nu) dS(\nu),$$

where the map ξ_- is as defined in (3-20).

Proof. Making the change of variables $\xi = \xi_-(x, \nu)$ defined in (3-20), and using (3-21) and (3-22), we have

$$\begin{aligned} \mathcal{P}_\lambda^- w(x) &= \int_{S_x \mathbb{H}^{n+1}} \Phi_-(x, \xi)^\lambda (\otimes^m (\mathcal{A}_-^{-1}(x, \xi))^T) w(B_-(x, \xi)) dS(\xi) \\ &= \int_{\mathbb{S}^n} P(x, \nu)^{n+\lambda} (\otimes^m (\mathcal{A}_-^{-1}(x, \xi_-(x, \nu)))^T) w(\nu) dS(\nu), \end{aligned}$$

as required. □

Poisson maps to eigenstates. To show that $\mathcal{P}_\lambda^- w(x)$ is an eigenstate of the Laplacian, we use:

Lemma 6.7. *Assume that $w \in \mathcal{D}'(\mathbb{S}^n; \otimes^m(T^*\mathbb{S}^n))$ is the delta function centered at $e_1 = \partial_{x_1} \in \mathbb{S}^n$ with the value $e_{j_1+1}^* \otimes \dots \otimes e_{j_m+1}^*$, where $1 \leq j_1, \dots, j_m \leq n$. Then, under the identifications (3-2) and (3-5), we have*

$$\mathcal{P}_\lambda^- w(z_0, z) = z_0^{n+\lambda} Z_{j_1}^* \otimes \dots \otimes Z_{j_m}^*.$$

Proof. We first calculate

$$P(z, e_1) = z_0.$$

It remains to show the identity in the half-space model

$$\mathcal{A}_-^{-T}(z, \xi_-(z, \nu)) e_{j+1}^* = Z_j^*, \quad 1 \leq j \leq n. \tag{6-23}$$

One can verify (6-23) by a direct computation; since \mathcal{A}_- is an isometry, one can instead calculate the image of e_{j+1} under \mathcal{A}_- , and then apply to it the differentials of the maps ψ and ψ_1 defined in (3-2) and (3-5).

Another way to show (6-23) is to use the interpretation of \mathcal{A}_- as parallel transport to conformal infinity; see (3-35). Note that under the diffeomorphism $\psi_1 : \mathbb{B}^{n+1} \rightarrow \mathbb{U}^{n+1}$, $\nu = e_1$ is sent to infinity and geodesics terminating at ν to straight lines parallel to the z_0 axis. By (6-9), the covector field Z_j^* is parallel along these geodesics and orthogonal to their tangent vectors. It remains to verify that the limit of the field $\rho_0 Z_j^*$ along these geodesics as $z \rightarrow \infty$, considered as a covector in the ball model, is equal to e_{j+1}^* . \square

Proof of Lemma 5.8. It suffices to show that, for each $\nu \in \mathbb{S}^n$, if w is a delta function centered at ν with value some symmetric trace-free tensor in $\otimes_S^m T_\nu^* \mathbb{S}^n$, then

$$(\Delta + \lambda(n + \lambda) - m)\mathcal{P}_\lambda^- w = 0, \quad \nabla^* \mathcal{P}_\lambda^- w = 0, \quad \mathcal{T}(\mathcal{P}_\lambda^- w) = 0.$$

Since the group of symmetries G of \mathbb{H}^{n+1} acts transitively on \mathbb{S}^n , we may assume that $\nu = \partial_1$. Applying Lemma 6.7, we write in the upper half-plane model

$$\mathcal{P}_\lambda^- w = z_0^{n+\lambda} u_0, \quad u_0 \in E_0^{(m)}, \quad \mathcal{T}(u_0) = 0.$$

It immediately follows that $\mathcal{T}(\mathcal{P}_\lambda^- w) = 0$. To see the other two identities, it suffices to apply Lemma 6.2 together with the formula

$$\Delta z_0^{n+\lambda} = -\lambda(n + \lambda)z_0^{n+\lambda}.$$

Injectivity of Poisson. Notice that \mathcal{P}_λ^- is an analytic family of operators in λ . We define the set

$$\mathcal{R}_m = \begin{cases} -\frac{1}{2}n - \frac{1}{2}\mathbb{N}_0 & \text{if } n > 1 \text{ or } m = 0, \\ -\frac{1}{2}\mathbb{N}_0 & \text{if } n = 1 \text{ and } m > 0, \end{cases} \tag{6-24}$$

and we will prove that, if $\lambda \notin \mathcal{R}_m$ and $w \in \mathcal{D}'(\mathbb{S}^n; \otimes_S^m T^* \mathbb{S}^n)$ is trace-free, then $\mathcal{P}_\lambda^-(w)$ has a weak asymptotic expansion at the conformal infinity with the leading term given by a multiple of w , proving injectivity of \mathcal{P}_λ^- . We shall use the 0-cotangent bundle approach in the ball model and rewrite $\mathcal{A}_\pm^{-1}(x, \xi_\pm(x, \nu))$ as the parallel transport $\tau(y', y)$ in ${}^0T\bar{\mathbb{B}}^{n+1}$ with $\psi(x) = y$ and $y' = \nu$, as explained in (3-35). Let $\rho \in C^\infty(\bar{\mathbb{B}}^{n+1})$ be a smooth boundary defining function which satisfies $\rho > 0$ in \mathbb{B}^{n+1} , $|d\rho|_{\rho^2 g_H} = 1$ near $\mathbb{S}^n = \{\rho = 0\}$, where g_H is the hyperbolic metric on the ball. We can for example take the function $\rho = \rho_0$ defined in (3-34) and smooth it near the center $y = 0$ of the ball. Such a function is called a *geodesic boundary defining function* and induces a diffeomorphism

$$\theta : [0, \epsilon)_t \times \mathbb{S}^n \rightarrow \bar{\mathbb{B}}^{n+1} \cap \{\rho < \epsilon\}, \quad \theta(t, \nu) := \theta_t(\nu), \tag{6-25}$$

where θ_t is the flow at time t of the gradient $\nabla^{\rho^2 g_H} \rho$ of ρ (denoted also ∂_ρ) with respect to the metric $\rho^2 g_H$. For ρ given in (3-34), we have, for t small,

$$\theta(t, \nu) = \frac{2-t}{2+t} \nu, \quad \nu \in \mathbb{S}^n.$$

For a fixed geodesic boundary defining function ρ , one can identify, over the boundary \mathbb{S}^n of $\bar{\mathbb{B}}^{n+1}$, the bundle $T^*\mathbb{S}^n$ and $T\mathbb{S}^n$ with the bundles ${}^0T^*\mathbb{S}^n := {}^0T^*\bar{\mathbb{B}}^{n+1} \cap \ker \iota_{\rho\partial\rho}$ simply by the isomorphism $v \mapsto \rho^{-1}v$ (and we identify their duals $T\mathbb{S}^n$ and ${}^0T\mathbb{S}^n$ as well). Similarly, over \mathbb{S}^n , $E^{(m)} \cap \ker \iota_{\rho\partial\rho}$ identifies with $\otimes_S^m T^*\mathbb{S}^n \cap \ker \mathcal{T}$ by the map $v \mapsto \rho^{-m}v$. We can then view the Poisson operator as an operator

$$\mathcal{P}_\lambda^- : \mathcal{D}'(\mathbb{S}^n; E^{(m)} \cap \ker \iota_{\rho\partial\rho}) \rightarrow \mathcal{C}^\infty(\mathbb{B}^{n+1}; \otimes_S^m ({}^0T^*\bar{\mathbb{B}}^{n+1})).$$

Lemma 6.8. *Let $w \in \mathcal{D}'(\mathbb{S}^n; E^{(m)} \cap \ker \iota_{\rho_0\partial\rho_0})$ and assume that $\lambda \notin \mathcal{R}_m$. Then $\mathcal{P}_\lambda^-(w)$ has a weak asymptotic expansion at \mathbb{S}^n as follows: for each $v \in \mathbb{S}^n$, there exists a neighborhood $V_v \subset \bar{\mathbb{B}}^{n+1}$ of v and a boundary defining function $\rho = \rho_v$ such that, for any $\varphi \in \mathcal{C}^\infty(V_v \cap \mathbb{S}^n; \otimes_S^m ({}^0T\mathbb{S}^n))$, there exist $F_\pm \in \mathcal{C}^\infty([0, \epsilon))$ such that, for $t > 0$ small,*

$$\begin{aligned} \int_{\mathbb{S}^n} \left\{ \mathcal{P}_\lambda^-(w)(\theta(t, v)), \otimes^m(\tau(\theta(t, v), v)) \cdot \varphi(v) \right\} dS_\rho(v) \\ = \begin{cases} t^{-\lambda} F_-(t) + t^{n+\lambda} F_+(t), & \lambda \notin -\frac{1}{2}n + \mathbb{N}, \\ t^{-\lambda} F_-(t) + t^{n+\lambda} \log(t) F_+(t), & \lambda \in -\frac{1}{2}n + \mathbb{N}. \end{cases} \end{aligned} \tag{6-26}$$

using the product collar neighborhood (6-25) associated to ρ , and, moreover, one has

$$F_-(0) = C \frac{\Gamma(\lambda + \frac{1}{2}n)}{(\lambda + n + m - 1)\Gamma(\lambda + n - 1)} \langle e^{\lambda f} \cdot w, \varphi \rangle \tag{6-27}$$

for some $f \in \mathcal{C}^\infty(\mathbb{S}^n)$ satisfying $\rho = \frac{1}{4}e^f \rho_0 + \mathcal{O}(\rho)$ near $\rho = 0$ and $C \neq 0$ a constant depending only on n . Here dS_ρ is the Riemannian measure for the metric $(\rho^2 g_H)|_{\mathbb{S}^n}$ and the distributional pairing on \mathbb{S}^n is with respect to this measure.

Proof. First we split w into $w_1 + w_2$, where w_1 is supported near $v \in \mathbb{S}^n$ and w_2 is zero near v . For the case where w_2 has support at positive distance from the support of φ , we have, for any geodesic boundary defining function ρ , that

$$t \mapsto t^{-n-\lambda} \int_{\mathbb{S}^n} \left\{ \mathcal{P}_\lambda^-(w_2)(\theta(t, v)), \otimes^m(\tau(\theta(t, v), v)) \cdot \varphi(v) \right\} dS_\rho(v) \in \mathcal{C}^\infty([0, \epsilon));$$

this is a direct consequence of Lemma 6.6 and the following smoothness properties:

$$\begin{aligned} (y, v) \mapsto \log\left(\frac{P(\psi^{-1}(y), v)}{\rho(y)}\right) &\in \mathcal{C}^\infty(\bar{\mathbb{B}}^{n+1} \times \mathbb{S}^n \setminus \text{diag}(\mathbb{S}^n \times \mathbb{S}^n)), \\ \tau(\cdot, \cdot) &\in \mathcal{C}^\infty(\bar{\mathbb{B}}^{n+1} \times \bar{\mathbb{B}}^{n+1} \setminus \text{diag}(\mathbb{S}^n \times \mathbb{S}^n); {}^0T^*\bar{\mathbb{B}}^{n+1} \otimes {}^0T\bar{\mathbb{B}}^{n+1}). \end{aligned}$$

This reduces the consideration of the lemma to the case where w is w_1 , supported near v , and to simplify we shall keep the notation w instead of w_1 . We thus consider now w and φ to have support near v . For convenience of calculations and as we did before, we work in the half-space model $\mathbb{R}_{z_0}^+ \times \mathbb{R}_z^n$ by mapping v to $(z_0, z) = (0, 0)$ (using the composition of a rotation on the ball model with the map defined in (3-5)), and we choose a neighborhood V_v of v which is mapped to $z_0^2 + |z|^2 < 1$ in \mathbb{U}^{n+1} and choose the geodesic defining function $\rho = z_0$ (and thus $\theta(z_0, z) = (z_0, z)$). (See Figure 5.) The geodesic boundary defining

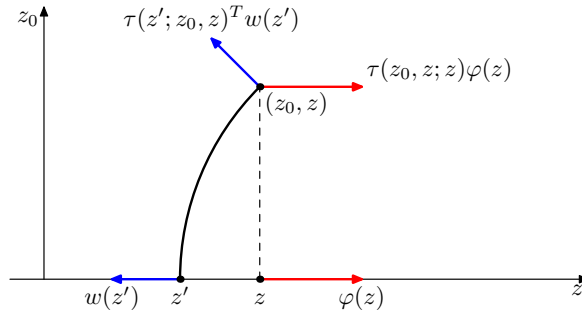


Figure 5. The covector $w(z')$, the vector $\varphi(z)$, and their parallel transports to (z_0, z) , viewed in the 0-bundles for the case $m = 1$.

function $\rho_0 = 2(1 - |y|)/(1 + |y|)$ in the ball equals

$$\rho_0(z_0, z) = \frac{4z_0}{1 + z_0^2 + |z|^2} \tag{6-28}$$

in the half-space model. The metric dS_ρ becomes the Euclidean metric dz on \mathbb{R}^n near 0 and w has compact support in \mathbb{R}^n . By (3-5) and (3-19), the Poisson kernel in these coordinates becomes

$$\tilde{P}(z_0, z; z') = e^{f(z')} P(z_0, z; z') \quad \text{with} \quad P(z_0, z; z') := \frac{z_0}{z_0^2 + |z - z'|^2}, \quad f(z') = \log(1 + |z'|^2),$$

where $z, z' \in \mathbb{R}^n$ and $z_0 > 0$. One has $\rho = \frac{1}{4}e^f \rho_0 + \mathcal{O}(\rho)$ near $\rho = 0$.

In [Guillarmou et al. 2010, Appendix], the parallel transport $\tau(z_0, z; 0, z')$ is computed for $z' \in \mathbb{R}^n$ in a neighborhood of 0: in the local orthonormal basis $Z_0 = z_0 \partial_{z_0}$, $Z_i = z_0 \partial_{z_i}$ of the bundle ${}^0T\mathbb{U}^{n+1}$, near ν , the matrix of $\tau(z_0, z; z') := \tau(z_0, z; 0, z')$ is given by

$$\begin{aligned} \tau_{00} &= 1 - 2P(z_0, z; z') \frac{|z - z'|^2}{z_0}, \\ \tau_{0i} &= -\tau_{i0} = -2z_0(z_i - z'_i) \frac{P(z_0, z; z')}{z_0}, \\ \tau_{ij} &= \delta_{ij} - 2P(z_0, z; z') \frac{(z_i - z'_i) \cdot (z_j - z'_j)}{z_0}. \end{aligned}$$

In particular, we see that $\tau(z_0, z; z)$ is the identity matrix in the basis $(Z_i)_i$ and thus $\tau(\theta(z_0, z), z)$ as well. We denote by $(Z_j^*)_j$ the dual basis to $(Z_j)_j$ as before.

Now, we use the correspondence between symmetric tensors and homogeneous polynomials to facilitate computations, as explained in Section 4A. To $\mathcal{S}(Z_j^*)_j$, we associate the polynomial on \mathbb{R}^n given by

$$P_I(x) = \mathcal{S}(Z_j^*)_j \left(\sum_{i=1}^n x_i Z_i, \dots, \sum_{i=1}^n x_i Z_i \right) = x_I,$$

where $x_I = \prod_{k=1}^m x_{i_k}$ if $I = (i_1, \dots, i_m)$. We denote by $\text{Pol}^m(\mathbb{R}^n)$ the space of homogeneous polynomials of degree m on \mathbb{R}^n and $\text{Pol}_0^m(\mathbb{R}^n)$ those which are harmonic (thus corresponding to trace-free symmetric

tensors in $E_0^{(m)}$. Then we can write $w = \sum_{\alpha} w_{\alpha} p_{\alpha}(x)$ for some $w_{\alpha} \in \mathcal{D}'(\mathbb{R}^n)$ supported near 0 and $p_{\alpha}(x) \in \text{Pol}_0^m(\mathbb{R}^n)$. Each $p_{\alpha}(x)$ composed with the linear map $\tau(z'; z_0, z)|_{Z_0^{\perp}}$ becomes the homogeneous polynomial in x

$$p_{\alpha}\left(x - 2(z - z')\langle z - z', x \rangle \cdot \frac{P(z_0, z; z')}{z_0}\right),$$

where $\langle \cdot, \cdot \rangle$ just denotes the Euclidean scalar product. To prove the desired asymptotic expansion, it suffices to take $\varphi \in C_0^{\infty}([0, \infty)_{z_0} \times \mathbb{R}^n)$ and to analyze the following homogeneous polynomial in x as $z_0 \rightarrow 0$:

$$\int_{\mathbb{R}^n} \sum_{\alpha} \left\langle e^{(n+\lambda)f} w_{\alpha}, \varphi(z_0, z) P(z_0, z; \cdot)^{n+\lambda} p_{\alpha}\left(x - 2(z - \cdot)\langle z - \cdot, x \rangle \cdot \frac{P(z_0, z; \cdot)}{z_0}\right) \right\rangle dz, \tag{6-29}$$

where the bracket $\langle w_{\alpha}, \cdot \rangle$ means the distributional pairing coming from pairing with respect to the canonical measure dS on \mathbb{S}^n , which in \mathbb{R}^n becomes the measure $4^n e^{-nf} dz$, and so the e^{nf} in (6-29) cancels out if one works with the Euclidean measure dz , which we do now. We have a convolution kernel in z and thus apply the Fourier transform in z (denoted \mathcal{F}): writing $P(z_0; |z - z'|)$ for $P(z_0, z; z')$, the integral (6-29) becomes (up to nonzero multiplicative constant)

$$I(z_0, x) := \sum_{\alpha} \left\langle \mathcal{F}^{-1}(e^{\lambda f} w_{\alpha}), \mathcal{F}(\varphi) \cdot \mathcal{F}_{\zeta \rightarrow \cdot} \left(P(z_0; |\zeta|)^{n+\lambda} p_{\alpha}\left(x - 2 \frac{\zeta \langle \zeta, x \rangle}{z_0} P(z_0; |\zeta|)\right) \right) \right\rangle_{\mathbb{R}^n}.$$

We can expand $p_{\alpha}(x - (2\zeta \langle \zeta, x \rangle / z_0) P(z_0; |\zeta|))$ so that

$$P(z_0; |\zeta|)^{n+\lambda} p_{\alpha}\left(x - 2 \frac{\zeta \langle \zeta, x \rangle}{z_0} P(z_0; |\zeta|)\right) = \sum_{r=0}^m Q_{r,\alpha}(\zeta, x) z_0^{-r} 2^r P(z_0; |\zeta|)^{n+\lambda+r},$$

where $Q_{r,\alpha}(\zeta)$ is homogeneous of degree m in x and degree $2r$ in ζ . Now we have (for some $C \neq 0$ independent of λ, r, α)

$$\frac{2^r}{z_0^r} \mathcal{F}_{\zeta \rightarrow \xi} \left(P^{n+\lambda+r}(z_0; |\zeta|) Q_{r,\alpha}(\zeta, x) \right) = \frac{C 2^{-\lambda} z_0^{-\lambda}}{\Gamma(\lambda + n + r)} \left[Q_{r,\alpha}(i \partial_{\zeta}, x) (|\zeta|^{\lambda+n/2+r} K_{\lambda+n/2+r}(|\zeta|)) \right]_{\zeta=z_0 \xi},$$

where $K_{\nu}(\cdot)$ is the modified Bessel function (see [Abramowitz and Stegun 1964, Chapter 9]) defined by

$$K_{\nu}(z) := \frac{\pi (I_{-\nu}(z) - I_{\nu}(z))}{2 \sin(\nu\pi)} \quad \text{if } I_{\nu}(z) := \sum_{\ell=0}^{\infty} \frac{1}{\ell! \Gamma(\ell + \nu + 1)} \left(\frac{z}{2}\right)^{2\ell+\nu}, \tag{6-30}$$

satisfying that $|K_{\nu}(z)| = \mathcal{O}(e^{-z}/\sqrt{z})$ as $z \rightarrow \infty$, and, for $s \notin \mathbb{N}_0$,

$$\mathcal{F}((1 + |z|^2)^{-s})(\xi) = \frac{2^{-s+1} (2\pi)^{n/2}}{\Gamma(s)} |\xi|^{s-n/2} K_{s-n/2}(|\xi|).$$

When $\lambda \notin (-\frac{1}{2}n + \mathbb{Z}) \cup (-n - \frac{1}{2}\mathbb{N}_0)$, we have

$$\begin{aligned}
 & 2^{-\lambda} z_0^{-\lambda} \mathcal{Q}_{r,\alpha}(i\partial_\zeta, x)(|\zeta|^{\lambda+n/2+r} K_{\lambda+n/2+r}(|\zeta|))|_{\zeta=z_0\xi} \\
 &= \frac{2^{r+n/2} \pi z_0^{-\lambda}}{2 \sin(\pi(\lambda + \frac{1}{2}n + r))} \\
 & \quad \times \left(\sum_{\ell=0}^{\infty} \frac{z_0^{2(\ell-r)} \mathcal{Q}_{r,\alpha}(i\partial_\xi, x)(|\frac{1}{2}\xi|^{2\ell})}{\ell! \Gamma(\ell - \lambda - \frac{1}{2}n - r + 1)} - z_0^{2\lambda+n} \sum_{\ell=0}^{\infty} \frac{z_0^{2\ell} \mathcal{Q}_{r,\alpha}(i\partial_\xi, x)(|\frac{1}{2}\xi|^{2(\lambda+r+\ell)+n})}{\ell! \Gamma(\ell + \lambda + \frac{1}{2}n + r + 1)} \right). \tag{6-31}
 \end{aligned}$$

Here the powers of $|\xi|$ are homogeneous distributions (note that, for $\lambda \notin \mathcal{R}_m$, the exceptional powers $|\xi|^{-n-j}$, $j \in \mathbb{N}_0$, do not appear) and the pairing of (6-31) with $\mathcal{F}^{-1}(e^{\lambda f} w_\alpha) \mathcal{F}(\varphi)$ makes sense since this distribution is Schwartz, as w_α has compact support. We deduce from this expansion that, for any $w_\alpha \in \mathcal{D}'(\mathbb{R}^n)$ supported near 0 and $\varphi \in C_0^\infty(\mathbb{R}^n)$, when $\lambda \notin (-\frac{1}{2}n + \mathbb{Z}) \cup (-n - \frac{1}{2}\mathbb{N}_0)$,

$$I(z_0, x) = z_0^{-\lambda} F_-(z_0, x) + z_0^{n+\lambda} F_+(z_0, x)$$

for some smooth functions $F_\pm \in C^\infty([0, \epsilon) \times \mathbb{R}^n)$ homogeneous of degree m in x . We need to analyze $F_-(0, x)$, which is obtained by computing the term of order 0 in ξ in the expansion (6-31) (that is, the terms with $\ell = r$ in the first sum; note that the terms with $\ell < r$ in this sum are zero): we obtain, for some universal constant $C \neq 0$,

$$F_-(0, x) = C \sum_\alpha \langle e^{\lambda f} w_\alpha, \varphi \rangle_{\mathbb{R}^n} \sum_{r=0}^m \frac{(-1)^r 2^{-r} \Gamma(\lambda + \frac{1}{2}n)}{r! \Gamma(\lambda + n + r)} \mathcal{Q}_{r,\alpha}(i\partial_\xi, x)(|\xi|^{2r}),$$

where we have used the inversion formula $\Gamma(1-z)\Gamma(z) = \pi / \sin(\pi z)$ and $\mathcal{Q}_{r,\alpha}(i\partial_\xi, x)(|\xi|^{2r})$ is constant in ξ . Using the Fourier transform, we notice that

$$\mathcal{Q}_{r,\alpha}(i\partial_\xi, x)(|\xi|^{2r}) = \Delta_\zeta^r \mathcal{Q}_{r,\alpha}(\zeta, x)|_{\zeta=0} = \Delta_\zeta^r (p_\alpha(x - \zeta \langle \zeta, x \rangle))|_{\zeta=0}.$$

We use Lemma A.5 to deduce that

$$F_-(0, x) = C \sum_\alpha \langle e^{\lambda f} w_\alpha, \varphi \rangle_{\mathbb{R}^n} p_\alpha(x) m! \frac{\Gamma(\lambda + \frac{1}{2}n)}{\Gamma(\lambda + n + m)} \sum_{r=0}^m \frac{(-1)^r \Gamma(\lambda + n + m)}{(m-r)! \Gamma(\lambda + n + r)}.$$

The sum over r is a nonzero polynomial of order m in λ , and, using the binomial formula, we see that its roots are $\lambda = -n - m + 2, \dots, -n + 1$; therefore, we deduce that

$$F_-(0, x) = C \langle e^{\lambda f} w, \varphi \rangle_{\mathbb{R}^n} \frac{\Gamma(\lambda + \frac{1}{2}n)}{(\lambda + n + m - 1) \Gamma(\lambda + n - 1)}.$$

We obtain the claimed result except for $\lambda \in -\frac{1}{2}n + \mathbb{N}$ by using that the volume measure on \mathbb{S}^n is $4^{-n} e^{nf}$. Now assume that $\lambda = -\frac{1}{2}n + j$ with $j \in \mathbb{N}$. The Bessel function satisfies, for $j \in \mathbb{N}$,

$$|\xi|^j K_j(|\xi|) = - \sum_{\ell=0}^{j-1} \frac{(-1)^\ell 2^{j-1-2\ell} (j-\ell-1)!}{\ell!} |\xi|^{2\ell} + |\xi|^{2j} (\log(|\xi|) L_j(|\xi|) + H_j(|\xi|))$$

for some function $L_j, H_j \in C^\infty(\mathbb{R}^+) \cap L^2(\mathbb{R}^+)$ with $L_j(0) \neq 0$. Then we apply the same arguments as before, and this implies the desired statement. \square

We obtain as a corollary:

Corollary 6.9. *For $m \in \mathbb{N}_0$ and $\lambda \notin \mathcal{R}_m$, the operator*

$$\mathcal{P}_\lambda^- : \mathcal{D}'(\mathbb{S}^n; \otimes_S^m(T^*\mathbb{S}^n) \cap \ker \mathcal{T}) \rightarrow C^\infty(\mathbb{H}^{n+1}; \otimes_S^m(T^*\mathbb{H}^{n+1}))$$

is injective.

This corollary immediately implies the injectivity part of Theorem 6 in Section 5B.

7. Expansions of eigenstates of the Laplacian

In this section, we show the surjectivity of the Poisson operator \mathcal{P}_λ^- (see Theorem 6 in Section 5B). For that, we take an eigenstate u of the Laplacian on M and lift it to \mathbb{H}^{n+1} . The resulting tensor is tempered and thus expected to have a weak asymptotic expansion at the conformal boundary \mathbb{S}^n ; a precise form of this expansion is obtained by a careful analysis of both the Laplacian and the divergence-free condition. We then show that $u = \mathcal{P}_\lambda^- w$, where w is some constant times the coefficient of $\rho^{-\lambda}$ in the expansion of u (compare with Lemma 6.8).

7A. Indicial calculus and general weak expansion. Recall the bundle $E^{(m)}$ defined in (6-5). The operator Δ acting on $C^\infty(\bar{\mathbb{B}}^{n+1}; E^{(m)})$ is an elliptic differential operator of order 2 that lies in the 0-calculus of [Mazzeo and Melrose 1987], which essentially means that it is an elliptic polynomial in elements of the Lie algebra $\mathcal{V}_0(\bar{\mathbb{B}}^{n+1})$ of smooth vector fields vanishing at the boundary of the closed unit ball $\bar{\mathbb{B}}^{n+1}$. Let $\rho \in C^\infty(\bar{\mathbb{B}}^{n+1})$ be a smooth geodesic boundary defining function (see the paragraph preceding (6-25)). The theory developed by Mazzeo [1991] shows that solutions of $\Delta u = su$ which are in $\rho^{-N}L^2(\bar{\mathbb{B}}^{n+1}; E^{(m)})$ for some N have weak asymptotic expansions at the boundary $\mathbb{S}^n = \partial\bar{\mathbb{B}}^{n+1}$, where ρ is any geodesic boundary defining function. To make this more precise, we introduce the *indicial family* of Δ : if $\lambda \in \mathbb{C}$, $\nu \in \mathbb{S}^n$, then there exists a family $I_{\lambda,\nu}(\Delta) \in \text{End}(E^{(m)}(\nu))$ depending smoothly on $\nu \in \mathbb{S}^n$ and holomorphically on λ such that, for all $u \in C^\infty(\bar{\mathbb{B}}^{n+1}; E^{(m)})$,

$$t^{-\lambda} \Delta(\rho^\lambda u)(\theta(t, \nu)) = I_{\lambda,\nu}(\Delta)u(\theta(0, \nu)) + \mathcal{O}(t)$$

near \mathbb{S}^n , where the remainder is estimated with respect to the metric g_H . Notice that $I_{\lambda,\nu}(\Delta)$ is independent of the choice of boundary defining function ρ .

For $\sigma \in \mathbb{C}$, the *indicial set* $\text{spec}_b(\Delta - \sigma; \nu)$ at $\nu \in \mathbb{S}^n$ of $\Delta - \sigma$ is the set

$$\text{spec}_b(\Delta - \sigma; \nu) := \{\lambda \in \mathbb{C} \mid I_{\lambda,\nu}(\Delta) - \sigma \text{ Id is not invertible}\}.$$

Then [Mazzeo 1991, Theorem 7.3] gives the following:²

²The full power of [Mazzeo 1991] is not needed for this lemma. In fact, it can be proved in a direct way by viewing the equation $(\Delta - \sigma)u = 0$ as an ordinary differential equation in the variable $\log \rho$. The indicial operator gives the constant coefficient principal part and the remaining terms are exponentially decaying; an iterative argument shows the needed asymptotics.

Lemma 7.1. Fix σ and assume that $\text{spec}_b(\Delta - \sigma; \nu)$ is independent of $\nu \in \mathbb{S}^n$. If $u \in \rho^\delta L^2(\bar{\mathbb{B}}^{n+1}; E^{(m)})$ with respect to the Euclidean measure for some $\delta \in \mathbb{R}$, and $(\Delta - \sigma)u = 0$, then u has a weak asymptotic expansion at $\mathbb{S}^n = \{\rho = 0\}$ of the form

$$u = \sum_{\substack{\lambda \in \text{spec}_b(\Delta - \sigma) \\ \text{Re}(\lambda) > \delta - \frac{1}{2}}} \sum_{\substack{\ell \in \mathbb{N}_0 \\ \text{Re}(\lambda) + \ell < \delta - \frac{1}{2} + N}} \sum_{p=0}^{k_{\lambda,\ell}} \rho^{\lambda+\ell} (\log \rho)^p w_{\lambda,\ell,p} + \mathcal{O}(\rho^{\delta+N-\frac{1}{2}-\epsilon})$$

for all $N \in \mathbb{N}$ and all $\epsilon > 0$ small, where $k_{\lambda,\ell} \in \mathbb{N}_0$, and $w_{\lambda,\ell,p}$ are in the Sobolev spaces

$$w_{\lambda,\ell,p} \in H^{-\text{Re}(\lambda)-\ell+\delta-\frac{1}{2}}(\mathbb{S}^n; E^{(m)}).$$

Here the weak asymptotic means that, for any $\varphi \in C^\infty(\mathbb{S}^n)$, as $t \rightarrow 0$,

$$\int_{\mathbb{S}^n} u(\theta(t, \nu)) \varphi(\nu) dS_\rho(\nu) = \sum_{\substack{\lambda \in \text{spec}_b(\Delta - \sigma) \\ \text{Re}(\lambda) > \delta - \frac{1}{2}}} \sum_{\substack{\ell \in \mathbb{N}_0 \\ \text{Re}(\lambda) + \ell < \delta - \frac{1}{2} + N}} \sum_{p=0}^{k_{\lambda,\ell}} t^{\lambda+\ell} \log(t)^p \langle w_{\lambda,\ell,p}, \varphi \rangle + \mathcal{O}(t^{\delta+N-\frac{1}{2}-\epsilon}), \tag{7-1}$$

where dS_ρ is the measure on \mathbb{S}^n induced by the metric $(\rho^2 g_H)|_{\mathbb{S}^n}$ and the distributional pairing is with respect to this measure. Moreover, the remainder $\mathcal{O}(t^{\delta+N-1/2-\epsilon})$ is conormal in the sense that it remains $\mathcal{O}(t^{\delta+N-1/2-\epsilon})$ after applying the operator $t\partial_t$ any finite number of times, and it depends on some Sobolev norm of φ .

Remark. The existence of the expansion (7-1) proved by Mazzeo [1991, Theorem 7.3] is independent of the choice of ρ , but the coefficients in the expansion depend on the choice of ρ . Let $\lambda_0 \in \text{spec}_b(\Delta - \sigma)$ with $\text{Re}(\lambda_0) > \delta - \frac{1}{2}$ be an element in the indicial set and assume that $k_{\lambda_0,0} = 0$, which means that the exponent ρ^{λ_0} in the weak expansion (7-1) has no log term. Assume also that there is no element $\lambda \in \text{spec}_b(\Delta - \sigma)$ with $\text{Re}(\lambda_0) > \text{Re}(\lambda) > \delta - \frac{1}{2}$ such that $\lambda \in \lambda_0 - \mathbb{N}$. Then it is direct to see from the weak expansion that, for a fixed function $\chi \in C^\infty(\mathbb{B}^{n+1})$ equal to 1 near \mathbb{S}^n and supported close to \mathbb{S}^n and for each $\varphi \in C^\infty(\mathbb{B}^{n+1})$, the Mellin transform

$$h(\zeta) := \int_{\mathbb{B}^{n+1}} \rho(y)^\zeta \chi(y) \varphi(y) u(y) d\text{Vol}_{g_H}(y), \quad \text{Re } \zeta > n + \frac{1}{2} - \delta,$$

(with values in E^m) has a meromorphic extension to $\zeta \in \mathbb{C}$ with a simple pole at $\zeta = n - \lambda_0$ and residue

$$\text{Res}_{\zeta=n-\lambda_0} h(\zeta) = \langle w_{\lambda_0,0,0}, \varphi|_{\mathbb{S}^n} \rangle. \tag{7-2}$$

As an application, if ρ' is another geodesic boundary defining function, one has $\rho = e^f \rho' + \mathcal{O}(\rho')$ for some $f \in C^\infty(\mathbb{S}^n)$ and we deduce that, if $w'_{\lambda_0,0,0}$ is the coefficient of $(\rho')^{\lambda_0}$ in the weak expansion of u using ρ' , then, as a distribution on \mathbb{S}^n ,

$$w'_{\lambda_0,0,0} = e^{\lambda_0 f} w_{\lambda_0,0,0}. \tag{7-3}$$

In particular, under the assumption above for λ_0 (this assumption can similarly be seen to be independent of the choice of ρ), if one knows the exponents of the asymptotic expansion, then proving that the

coefficient of ρ^{λ_0} term is nonzero can be done locally near any point of \mathbb{S}^n and with any choice of geodesic boundary defining function.

Finally, if $w_{\lambda_0,0,0}$ is the coefficient of $\rho_0^{\lambda_0}$ in the weak expansion with boundary defining function ρ_0 defined in (3-34) and if $\gamma^*u = u$ for some hyperbolic isometry $\gamma \in G$, we can use that $\rho_0 \circ \gamma = N_\gamma^{-1} \cdot \rho_0 + \mathcal{O}(\rho_0^2)$ near \mathbb{S}^n , together with (7-2) to get

$$L_\gamma^* w_{\lambda_0,0,0} = N_\gamma^{\lambda_0} w_{\lambda_0,0,0} \in \mathcal{D}'(\mathbb{S}^n; E^{(m)}) \tag{7-4}$$

as distributions on \mathbb{S}^n (with respect to the canonical measure on \mathbb{S}^n) with values in $E^{(m)}$. Here N_γ, L_γ are as defined in Section 3E. If we view $w_{\lambda_0,0,0}$ as a distribution with values in $\otimes_S^m T^*\mathbb{S}^n$, the covariance becomes

$$L_\gamma^* w_{\lambda_0,0,0} = N_\gamma^{\lambda_0 - m} w_{\lambda_0,0,0} \in \mathcal{D}'(\mathbb{S}^n; \otimes_S^m T^*\mathbb{S}^n). \tag{7-5}$$

Using the calculations of Section 6B, we will compute the indicial family of the Laplacian on $E^{(m)}$:

Lemma 7.2. *Let Δ be the Laplacian on sections of $E^{(m)}$. Then the indicial set $\text{spec}_b(\Delta - \sigma, \nu)$ does not depend on $\nu \in \mathbb{S}^n$ and is equal to³*

$$\bigcup_{k=0}^{\lfloor m/2 \rfloor} \{ \lambda \mid -\lambda^2 + n\lambda + m + 2k(2m + n - 2k - 2) = \sigma \} \cup \bigcup_{k=0}^{\lfloor (m-1)/2 \rfloor} \{ \lambda \mid -\lambda^2 + n\lambda + n + 3(m - 1) + 2k(n + 2m - 2k - 4) = \sigma \}.$$

Proof. We consider an isometry mapping the ball model \mathbb{B}^{n+1} to the half-plane model \mathbb{U}^{n+1} which also maps ν to 0 and do all the calculations in \mathbb{U}^{n+1} with the geodesic boundary defining function z_0 near 0. By (6-7), each tensor $u \in E^{(m)}$ is determined uniquely by its $E_0^{(m)}$ and $E_1^{(m)}$ components, which are denoted u_0 and u_1 ; therefore, it suffices to understand how the corresponding components of $I_{\lambda,\nu}(\Delta)u$ are determined by u_0 and u_1 . We can use the geodesic boundary defining function $\rho = z_0$; note that $\Delta z_0^\lambda = \lambda(n - \lambda)z_0^\lambda$ for all $\lambda \in \mathbb{C}$.

Assume first that u satisfies $u_1 = 0$ and u_0 is constant in the frame $\mathcal{S}(Z_I^*)$. Then, by Lemma 6.4,

$$\begin{aligned} \pi_0(z_0^{-\lambda} \Delta(z_0^\lambda u)) &= R_0 u_0 = (\lambda(n - \lambda) + m)u_0 + m(m - 1)\mathcal{S}(z_0^{-2}h \otimes \mathcal{T}(u_0)), \\ \pi_1(z_0^{-\lambda} \Delta(z_0^\lambda u)) &= 0. \end{aligned}$$

Assume now that u satisfies $u_0 = 0$ and u_1 is constant in the frame $\mathcal{S}(Z_0^* \otimes Z_J^*)$. Then, by Lemma 6.4,

$$\begin{aligned} \pi_0(z_0^{-\lambda} \Delta(z_0^\lambda u)) &= 0, \\ \pi_1(z_0^{-\lambda} \Delta(z_0^\lambda u)) &= R_1 u_1 = (\lambda(n - \lambda) + n + 3(m - 1))u_1 + (m - 1)(m - 2)\mathcal{S}(Z_0^* \otimes z_0^{-2}h \otimes \mathcal{T}(u_1')). \end{aligned}$$

We see that the indicial operator does not intertwine the u_0 and u_1 components and it remains to understand for which λ the number s is a root of R_0 or R_1 .

³Our argument in the next section does not actually use the precise indicial roots, as long as they are independent of ν and form a discrete set.

Next, we consider the decomposition (4-5), where we define $\mathcal{I}(u) = \frac{1}{2}(m+2)(m+1)\mathcal{S}(z_0^{-2}h \otimes u)$ for $u \in E_0^{(m)}$; we have

$$u_0 = \sum_{k=0}^{\lfloor m/2 \rfloor} \mathcal{I}^k(\otimes u_0^k), \quad u_1 = \sum_{k=0}^{\lfloor (m-1)/2 \rfloor} \mathcal{S}(Z_0^* \otimes \mathcal{I}^k(u_1^k)),$$

where $u_0^k \in E_0^{(m-2k)}$ and $u_1^k \in E_0^{(m-2k-1)}$ are trace-free tensors. Using (4-4), we calculate

$$\begin{aligned} R_0(\mathcal{I}^k(u_0^k)) &= (\lambda(n-\lambda) + m)\mathcal{I}^k(u_0^k) + 2\mathcal{I}(\mathcal{T}(\mathcal{I}^k(u_0^k))) \\ &= (-\lambda^2 + n\lambda + m + 2k(2m+n-2k-2))\mathcal{I}^k(u_0^k), \\ R_1(\mathcal{S}(Z_0^* \otimes \mathcal{I}^k(u_1^k))) &= (\lambda(n-\lambda) + n + 3(m-1))\mathcal{S}(Z_0^* \otimes \mathcal{I}^k(u_1^k)) + 2\mathcal{S}(Z_0^* \otimes \mathcal{I}(\mathcal{T}(\mathcal{I}^k(u_1^k)))) \\ &= (-\lambda^2 + n\lambda + n + 3(m-1) + 2k(n+2m-2k-4))\mathcal{S}(Z_0^* \otimes \mathcal{I}^k(u_1^k)), \end{aligned}$$

which finishes the proof of the lemma. □

7B. Weak expansions in the divergence-free case. By Lemma 7.1, we now know that solutions of $\Delta u = \sigma u$ that are trace-free symmetric tensors of order m in some weighted L^2 space have weak asymptotic expansions at the boundary of $\overline{\mathbb{B}^{n+1}}$ with exponents obtained from the indicial set of Lemma 7.2. In fact, we can be more precise about the exponents which really appear in the weak asymptotic expansion if we ask that u also be divergence-free:

Lemma 7.3. *Let $u \in \rho^\delta L^2(\overline{\mathbb{B}^{n+1}}; E^{(m)})$ be a trace-free symmetric m -cotensor with ρ a geodesic boundary defining function and $\delta \in (-\infty, \frac{1}{2})$, where the measure is the Euclidean Lebesgue measure on the ball. Assume that u is a nonzero divergence-free eigentensor for the Laplacian on hyperbolic space:*

$$\Delta u = \sigma u, \quad \nabla^* u = 0 \tag{7-6}$$

for some $\sigma = m + \frac{1}{4}n^2 - \mu^2$ with $\text{Re}(\mu) \in [0, \frac{1}{2}(n+1) - \delta)$ and $\mu \neq 0$. Then the following weak expansion holds: for all $r \in [0, m]$, $N > 0$, and $\epsilon > 0$ small,

$$\begin{aligned} &(\iota_{\rho\partial\rho})^r u \\ &= \sum_{\substack{\ell \in \mathbb{N}_0 \\ \text{Re}(-\mu) + \ell < N - \epsilon}} \rho^{n/2 - \mu + r + \ell} w_{-\mu, \ell}^r + \sum_{\substack{\ell \in \mathbb{N}_0 \\ \text{Re}(\mu) + \ell < N - \epsilon}} \sum_{p=0}^{k_{\mu, \ell}} \rho^{n/2 + \mu + r + \ell} \log(\rho)^p w_{\mu, \ell, p}^r + \mathcal{O}(\rho^{n/2 + N + r - \epsilon}) \end{aligned} \tag{7-7}$$

with $w_{-\mu, \ell}^r \in H^{-n/2 + \text{Re}(\mu) - r - \ell + \delta - 1/2}(\mathbb{S}^n; E^{(m-r)})$, $w_{\mu, \ell, p}^r \in H^{-n/2 - \text{Re}(\mu) - r - \ell + \delta - 1/2}(\mathbb{S}^n; E^{(m-r)})$. Moreover, if $\mu \notin \frac{1}{2}\mathbb{N}_0$, then $k_{\mu, \ell} = 0$.

Remark. (i) If u is the lift to \mathbb{H}^{n+1} of an eigentensor on a compact quotient $M = \Gamma \backslash \mathbb{H}^{n+1}$, then $u \in L^\infty(\mathbb{B}^{n+1}; E^{(m)})$ and so, for all $\epsilon > 0$, the following regularity holds:

$$w_{-\mu, 0} \in H^{-n/2 + \text{Re}(\mu) - \epsilon}(\mathbb{S}^n; E^{(m)}), \quad w_{\mu, 0, 0} \in H^{-n/2 - \text{Re}(\mu) - \epsilon}(\mathbb{S}^n; E^{(m)}).$$

(ii) The existence of the expansion (7-7) does not depend on the choice of ρ . For $r = 0$, this follows from analyzing the Mellin transform of u as in the remark following Lemma 7.1. For $r > 0$, we additionally use that, if ρ' is another geodesic boundary defining function, then $\rho\partial\rho - \rho'\partial\rho' \in \rho \cdot {}^0T\overline{\mathbb{B}^{n+1}}$

(indeed, the dual covector by the metric is $\rho^{-1}d\rho - (\rho')^{-1}d\rho'$ and we have $\rho' = e^f \rho$ for some smooth function f on $\bar{\mathbb{B}}^{n+1}$). Therefore, $(\iota_{\rho'\partial\rho'})^r u$ is a linear combination of contractions with 0-vector fields of $\rho^{r-r'}(\iota_{\rho\partial\rho})^{r'} u$ for $0 \leq r' \leq r$, which have the desired asymptotic expansion. Moreover, as follows from (7-3), for each $r \in [0, m]$, the condition that $w_{-\mu,0}^{r'} = 0$ for all $r' \in [0, r]$ also does not depend on the choice of ρ , and the same can be said about $w_{\mu,0}^{r'}$ when $\mu \notin \frac{1}{2}\mathbb{N}_0$.

Proof. It suffices to describe the weak asymptotic expansion of u near any point $\nu \in \mathbb{S}^n$. For that, we work in the half-space model \mathbb{U}^{n+1} by sending $-\nu$ to ∞ and ν to 0 as we did before (composing a rotation of the ball model with the map (3-5)). Since the choice of geodesic boundary defining function does not change the nature of the weak asymptotic expansion (but only the coefficients), we can take the geodesic boundary defining function ρ to be equal to $\rho(z_0, z) = z_0$ inside $|z| + z_0 < 1$ (which corresponds to a neighborhood of ν in the ball model). Considering the weak asymptotic (7-1) of u near 0 amounts to taking φ supported near ν in \mathbb{S}^n in (7-1); for instance, if we work in the half-space model, we shall consider $\varphi(z)$ supported in $|z| < 1$ in the boundary of \mathbb{U}^{n+1} .

We have the decomposition $u = \sum_{k=0}^m u_k$ with $u_k \in \rho^\delta L^2(\mathbb{U}^{n+1}; E_k^{(m)})$ and we write $u_k = \mathcal{S}((Z_0^*)^{\otimes k} \otimes u'_k)$ for some $u'_k \in \rho^\delta L^2(\mathbb{U}^{n+1}; E_0^{(m-k)})$ following what we did in (6-6). Now, since $u \in \rho^\delta L^2(\bar{\mathbb{B}}^{n+1}) = \rho_0^\delta L^2(\bar{\mathbb{B}}^{n+1})$ satisfies $\Delta u = \sigma u$, we deduce from the form of the Laplacian near $\rho = 0$ that u is in $\rho_0^{\delta-2k} H^{2k}(\bar{\mathbb{B}}^{n+1}; E^{(m)})$ for all $k \in \mathbb{N}$, where H^k denotes the Sobolev space of order k associated to the Euclidean Laplacian on the closed unit ball. Then, by Sobolev embedding, one has that, for each $t > 0$, $u|_{z_0=t}$ belongs to $(1 + |z|)^N L^2(\mathbb{R}_z^n; E^{(m)})$ for some $N \in \mathbb{N}$ and we can consider its Fourier transform in z , as a tempered distribution.⁴ Then Fourier transforming the equation $(\pi_0 + \pi_1)(\Delta u - \sigma u) = 0$ in the z variable (recall that π_i is the orthogonal projection on $E_i^{(m)}$), and writing the Fourier variable ξ as $\xi = \sum_{i=1}^n \xi_i dz_i = \sum_{i=1}^n z_0 \xi_i Z_i^*$, with the notations of Lemma 6.4, we get

$$\sum_{I \in \mathcal{A}^m} ((-Z_0)^2 + nZ_0 + z_0^2 |\xi|^2 + m - \sigma) \hat{f}_I \mathcal{S}(Z_I^*) + 2i \sum_{J \in \mathcal{A}^{m-1}} \hat{g}_J \mathcal{S}(\xi \otimes Z_J^*) + m(m-1) \sum_I \hat{f}_I \mathcal{S}(z_0^{-2} h \otimes \mathcal{T}(\mathcal{S}(Z_I^*))) = 0. \quad (7-8)$$

and

$$\sum_{J \in \mathcal{A}^{m-1}} ((-Z_0)^2 + nZ_0 + z_0^2 |\xi|^2 + n + 3(m-1) - \sigma) \hat{g}_J \mathcal{S}(Z_J^*) - 2im \sum_{I \in \mathcal{A}^m} \hat{f}_I \iota_\xi \mathcal{S}(Z_I^*) - 2im(m-1) \sum_{I \in \mathcal{A}^m} \hat{f}_I \mathcal{S}(\xi \otimes \mathcal{T}(\mathcal{S}(Z_I^*))) + (m-1)(m-2) \sum_{J \in \mathcal{A}^{m-1}} \hat{g}_J \mathcal{S}(z_0^{-2} h \otimes \mathcal{T}(\mathcal{S}(Z_J^*))) = 0, \quad (7-9)$$

where hat denotes Fourier transform in z and ι_ξ means $\sum_{j=1}^n z_0 \xi_j \iota_{Z_j}$.

⁴Unlike in Lemma 6.8, we only use Fourier analysis here for convenience of notation — all the calculations below could be done with differential operators in z instead.

Similarly, we Fourier transform in z the equation $(\pi_0 + \pi_1)(\nabla^*u) = 0$ using Lemma 6.5 to obtain

$$\begin{aligned} \sum_{I \in \mathcal{A}^m} i \hat{f}_I \iota_\xi \mathcal{S}(Z_I^*) &= \frac{1}{m} \sum_{J \in \mathcal{A}^{m-1}} ((n+m-1)\hat{g}_J - Z_0(\hat{g}_J))\mathcal{S}(Z_J^*), \\ \sum_{I \in \mathcal{A}^m} (Z_0 \hat{f}_I - (n+m-1)\hat{f}_I)\mathcal{T}(\mathcal{S}(Z_I^*)) &= \frac{1}{m} \sum_{J \in \mathcal{A}^{m-1}} i \hat{g}_J \iota_\xi \mathcal{S}(Z_J^*). \end{aligned} \tag{7-10}$$

Now, we use the correspondence between symmetric tensors and homogeneous polynomials to facilitate computations, as explained in Section 4A and in the proof of Lemma 6.8; that is, to $\mathcal{S}(Z_I^*)$, we associate the polynomial x_I on \mathbb{R}^n . If $\xi \in \mathbb{R}^n$ is a fixed element and $u \in \text{Pol}^m(\mathbb{R}^n)$, we write $\partial_\xi u = du \cdot \xi \in \text{Pol}^{m-1}(\mathbb{R}^n)$ for the derivative of u in the direction of ξ and ξ^*u for the element $\langle \xi, \cdot \rangle_{\mathbb{R}^n} u \in \text{Pol}^{m+1}(\mathbb{R}^n)$. The trace map \mathcal{T} becomes $-(1/(m(m-1)))\Delta_x$. We define $\hat{u}_0 := \sum_{I \in \mathcal{A}^m} \hat{f}_I x_I$ and $\hat{u}_1 = \sum_{J \in \mathcal{A}^{m-1}} \hat{g}_J x_J$. The elements $\hat{f}_I(z_0, \xi)$ and $\hat{g}_J(z_0, \xi)$ belong to the space $\mathcal{C}^\infty(\mathbb{R}_{z_0}^+; \mathcal{S}'(\mathbb{R}_\xi^n))$. We decompose them as

$$\hat{u}_0 = \sum_{j=0}^{\lfloor m/2 \rfloor} |x|^{2j} \hat{u}_0^{2j} \quad \text{and} \quad \hat{u}_1 = \sum_{j=0}^{\lfloor (m-1)/2 \rfloor} |x|^{2j} \hat{u}_1^{2j} \tag{7-11}$$

for some $\hat{u}_i^{2j} \in \text{Pol}_0^{m-i-2j}(\mathbb{R}^n)$ (harmonic in x , that is, trace-free).

Using the homogeneous polynomial description of u_0 , (7-8) becomes

$$(-(Z_0)^2 + nZ_0 + z_0^2|\xi|^2 + m - \sigma)\hat{u}_0 + 2iz_0\xi^*\hat{u}_1 - |x|^2\Delta_x\hat{u}_0 = 0. \tag{7-12}$$

First, if W is a harmonic homogeneous polynomial in x of degree j , one has $\Delta_x(\xi^*W) = -2\partial_\xi W$ and $\Delta_x^2(\xi^*W) = 0$; thus one can write

$$\xi^*W = \left(\xi^*W - \frac{\partial_\xi W}{n+2(j-1)}|x|^2 \right) + \frac{\partial_\xi W}{n+2(j-1)}|x|^2 \tag{7-13}$$

for the decomposition (4-5) of ξ^*W . In particular, one can write the decomposition (4-5) of $\xi^*\hat{u}_1$ as

$$\xi^*\hat{u}_1 = \sum_{j=0}^{\lfloor (m-1)/2 \rfloor} |x|^{2j} \left(\xi^*\hat{u}_1^{2j} - \frac{\partial_\xi \hat{u}_1^{2j}}{n+2(m-2-2j)}|x|^2 + \frac{\partial_\xi \hat{u}_1^{2(j-1)}}{n+2(m-2j)} \right).$$

We can write $\Delta_x\hat{u}_0 = \sum_{j=0}^{\lfloor m/2 \rfloor} \lambda_j |x|^{2j-2} \hat{u}_0^{2j}$ for $\lambda_j = -2j(n+2(m-j-1))$. Thus (7-12) gives, for $j \leq \lfloor m/2 \rfloor$,

$$(-(Z_0)^2 + nZ_0 + z_0^2|\xi|^2 + m - \sigma - \lambda_j)\hat{u}_0^{2j} + 2iz_0 \left(\xi^*\hat{u}_1^{2j} - \frac{|x|^2 \partial_\xi \hat{u}_1^{2j}}{n+2(m-2-2j)} + \frac{\partial_\xi \hat{u}_1^{2(j-1)}}{n+2(m-2j)} \right) = 0. \tag{7-14}$$

Notice that $\iota_\xi(\mathcal{S}(Z_I^*))$ corresponds to the polynomial $(z_0/m) dx_I \cdot \xi = (z_0/m) \partial_\xi \cdot x_I$ if $I \in \mathcal{A}^m$. From (7-10) we thus have, for $c_m := n+m-1$,

$$\begin{aligned} -iz_0\partial_\xi\hat{u}_0 &= (Z_0 - c_m)\hat{u}_1, \\ -iz_0\partial_\xi\hat{u}_1 &= (Z_0 - c_m)\Delta_x\hat{u}_0. \end{aligned} \tag{7-15}$$

Next, (7-9) implies

$$-(Z_0)^2 + nZ_0 + z_0^2|\xi|^2 + n + 3(m - 1) - \sigma \hat{u}_1 - 2iz_0\partial_\xi \hat{u}_0 + 2iz_0\xi^* \Delta_x \hat{u}_0 - |x|^2 \Delta_x \hat{u}_1 = 0.$$

Using (7-15), this can be rewritten as

$$-(Z_0)^2 + (n + 2)Z_0 + z_0^2|\xi|^2 - n + m - 1 - \sigma \hat{u}_1 + 2iz_0\xi^* \Delta_x \hat{u}_0 - |x|^2 \Delta_x \hat{u}_1 = 0. \tag{7-16}$$

We can write $\Delta_x \hat{u}_1 = \sum_{j=0}^{\lfloor (m-1)/2 \rfloor} \lambda'_j |x|^{2j-2} \hat{u}_1^{2j}$ for $\lambda'_j = -2j(n + 2(m - j - 2))$. From (7-16), we get

$$\begin{aligned} &(-(Z_0)^2 + (n + 2)Z_0 + z_0^2|\xi|^2 - n + m - 1 - \sigma - \lambda'_j) \hat{u}_1^{2j} \\ &+ 2iz_0 \left(\lambda_{j+1} \xi^* \hat{u}_0^{2(j+1)} - \frac{\lambda_{j+1} \partial_\xi \hat{u}_0^{2(j+1)}}{n + 2(m - 3 - 2j)} |x|^2 + \frac{\lambda_j \partial_\xi \hat{u}_0^{2j}}{n + 2(m - 1 - 2j)} \right) = 0. \end{aligned} \tag{7-17}$$

We shall now partially uncouple the system of equations for \hat{u}_0^{2j} and \hat{u}_1^{2j} . Using (7-13) and applying the decomposition (4-5), we have

$$\begin{aligned} \partial_\xi (|x|^{2j} \hat{u}_0^{2j}) &= |x|^{2j} \partial_\xi \hat{u}_0^{2j} \frac{n + 2(m - j - 1)}{n + 2(m - 2j - 1)} + 2j |x|^{2j-2} \left(\xi^* \hat{u}_0^{2j} - \frac{|x|^2 \partial_\xi \hat{u}_0^{2j}}{n + 2(m - 2j - 1)} \right), \\ \partial_\xi (|x|^{2j} \hat{u}_1^{2j}) &= |x|^{2j} \partial_\xi \hat{u}_1^{2j} \frac{n + 2(m - j - 2)}{n + 2(m - 2j - 2)} + 2j |x|^{2j-2} \left(\xi^* \hat{u}_1^{2j} - \frac{|x|^2 \partial_\xi \hat{u}_1^{2j}}{n + 2(m - 2j - 2)} \right), \end{aligned}$$

and, from (7-15), this implies that, for $j \geq 0$,

$$(Z_0 - c_m) \hat{u}_1^{2j} = -iz_0 \left(\partial_\xi \hat{u}_0^{2j} \frac{n + 2(m - j - 1)}{n + 2(m - 2j - 1)} + 2(j + 1) \left(\xi^* \hat{u}_0^{2(j+1)} - \frac{|x|^2 \partial_\xi \hat{u}_0^{2(j+1)}}{n + 2(m - 2j - 3)} \right) \right), \tag{7-18}$$

and, for $j > 0$,

$$(Z_0 - c_m) \hat{u}_0^{2j} = iz_0 \left(\frac{\partial_\xi \hat{u}_1^{2(j-1)}}{2j(n + 2(m - 2j))} + \frac{1}{n + 2(m - j - 1)} \left(\xi^* \hat{u}_1^{2j} - \frac{|x|^2 \partial_\xi \hat{u}_1^{2j}}{n + 2(m - 2j - 2)} \right) \right). \tag{7-19}$$

Combining with (7-14) and (7-17) we get, for $j \geq 0$,

$$\begin{aligned} &(-(Z_0)^2 + (n + 4j)Z_0 + z_0^2|\xi|^2 + m - \sigma - \lambda_j - 4jc_m) \hat{u}_0^{2j} \\ &+ 2iz_0 \frac{n + 2(m - 2j - 1)}{n + 2(m - j - 1)} \left(\xi^* \hat{u}_1^{2j} - \frac{|x|^2 \partial_\xi \hat{u}_1^{2j}}{n + 2(m - 2 - 2j)} \right) = 0, \end{aligned} \tag{7-20}$$

$$\begin{aligned} &(-(Z_0)^2 + (n + 2 + 4j)Z_0 + z_0^2|\xi|^2 - n + m - 1 - \sigma - \lambda'_j - 4jc_m) \hat{u}_1^{2j} \\ &+ 2iz_0(\lambda_{j+1} + 4j(j + 1)) \left(\xi^* \hat{u}_0^{2(j+1)} - \frac{|x|^2 \partial_\xi \hat{u}_0^{2(j+1)}}{n + 2(m - 3 - 2j)} \right) = 0, \end{aligned} \tag{7-21}$$

$$\begin{aligned} &\left(-(Z_0)^2 + \left(n + 2 - \frac{\lambda_{j+1}}{j + 1} \right) Z_0 + z_0^2|\xi|^2 - n + m - 1 - \sigma + \frac{\lambda_{j+1}}{j + 1} (c_m - j) \right) \hat{u}_1^{2j} \\ &+ 2iz_0 \frac{(n + 2(m - j - 1))(n + 2(m - 2j - 2))}{n + 2(m - 2j - 1)} \partial_\xi \hat{u}_0^{2j} = 0, \end{aligned} \tag{7-22}$$

and, for $j > 0$,

$$\left(-Z_0^2 + \left(n - \frac{\lambda_j}{j}\right)Z_0 + z_0^2|\xi|^2 + m - \sigma + \frac{\lambda_j}{j}(c_m - j)\right)\hat{u}_0^{2j} - iz_0 \frac{2(m-1-2j)+n}{j(n+2(m-2j))} \partial_\xi \hat{u}_1^{2(j-1)} = 0. \tag{7-23}$$

To prove the lemma, we will show the following weak asymptotic expansion for $i = 0, 1$:

$$\begin{aligned} \langle \hat{u}_i^{2j}(z_0, \cdot), \hat{\varphi} \rangle &= \sum_{\substack{\ell \in \mathbb{N}_0, \\ \operatorname{Re}(-\mu) + \ell < N - \epsilon}} z_0^{n/2 - \mu + 2j + i + \ell} \langle \tilde{w}_{i; -\mu, \ell}^{2j}, \varphi \rangle \\ &+ \sum_{\substack{\ell \in \mathbb{N}_0, \\ \operatorname{Re}(\mu) + \ell < N - \epsilon}} \sum_{p=0}^{k_{\mu, \ell}} z_0^{n/2 + \mu + 2j + i + \ell} \log(z_0)^p \langle \tilde{w}_{i; \mu, \ell, p}^{2j}, \varphi \rangle + \mathcal{O}(z_0^{n/2 + 2j + i + N - \epsilon}), \end{aligned} \tag{7-24}$$

where $\tilde{w}_{i; -\mu, \ell}^{2j}$ and $\tilde{w}_{i; \mu, \ell, p}^{2j}$ are distributions in some Sobolev spaces in $\{|z| < 1\} \subset \mathbb{R}^n$ and, for $\mu \notin \frac{1}{2}\mathbb{N}_0$, we have $k_{\mu, \ell} = 0$.

Define, for $0 \leq r \leq m$ and $\varphi \in C_0^\infty(\mathbb{R}^n)$ supported in $\{|z| < 1\}$,

$$F^r(\varphi)(z_0) := \begin{cases} \langle \hat{u}_0^r(z_0, \cdot), \hat{\varphi} \rangle & \text{if } r \text{ is even,} \\ \langle \hat{u}_1^{r-1}(z_0, \cdot), \hat{\varphi} \rangle & \text{if } r \text{ is odd.} \end{cases}$$

Since \hat{u}_i^{r-i} is the Fourier transform in z of iterated traces of u_i , Lemma 7.1 gives that the function $F^r(\varphi)(z_0)$ satisfies, for all $N \in \mathbb{N}$, $\epsilon > 0$,

$$F^r(\varphi)(z_0) = \sum_{\substack{\lambda \in \operatorname{spec}_b(\Delta - \sigma) \\ \operatorname{Re}(\lambda) > \delta - \frac{1}{2}}} \sum_{\substack{\ell \in \mathbb{N}_0, \\ \operatorname{Re}(\lambda) + \ell < N - \epsilon}} \sum_{p=0}^{k_{\lambda, \ell}^r} z_0^{\lambda + \ell} \log(z_0)^p \langle w_{\lambda, \ell, p}^r, \varphi \rangle + \mathcal{O}(z_0^{N - \epsilon}) \tag{7-25}$$

as $z_0 \rightarrow 0$ for some $w_{\lambda, \ell, p}^r$ in some Sobolev space on $\{|z| < 1\}$. We pair (7-20), (7-21) with $\hat{\varphi}$, and it is direct to see that we obtain a differential equation in z_0 of the form

$$P^r(Z_0)F^r(\varphi)(z_0) = -z_0^2 F^r(\Delta\varphi)(z_0) + z_0 F^{r+1}(Q^r\varphi)(z_0) \tag{7-26}$$

for $Z_0 = z_0 \partial_{z_0}$,

$$P^r(\lambda) := -\lambda^2 + (n + 2r)\lambda - r(n + r) - \frac{1}{4}n^2 + \mu^2 = -\left(\lambda - \frac{1}{2}n - r\right)^2 + \mu^2,$$

and Q^r some differential operator of order 1 with values in homomorphisms on the space of polynomials in x . Here we denote $F^{m+1} = 0$.

We now show the expansion (7-24) by induction on $r = 2j + i = m, m - 1, \dots, 0$. By plugging the expansion (7-25) in (7-26) and using

$$P^r(Z_0)z_0^\lambda \log(z_0)^p = z_0^\lambda (P_0^r(\lambda)(\log z_0)^p + p \partial_\lambda P_0^r(\lambda)(\log z_0)^{p-1} + \mathcal{O}((\log z_0)^{p-2})), \tag{7-27}$$

we see that if, for some p , $z_0^\lambda (\log z_0)^p$ is featured in the asymptotic expansion of $F^r(\varphi)(z_0)$, then either $\lambda \in \frac{1}{2}n + r - \mu + \mathbb{N}_0$, or $\lambda \in \frac{1}{2}n + r + \mu + \mathbb{N}_0$, or $z_0^{\lambda-2} (\log z_0)^p$ is featured in the expansion of $F^r(\Delta\varphi)(z_0)$. Moreover, if $p > 0$ and $\lambda \notin \{\frac{1}{2}n + r \pm \mu\}$, then either $z_0^\lambda (\log z_0)^{p-1}$ is featured in $F^r(\varphi)(z_0)$ for

some $p' > p$, or $z_0^{\lambda-2}(\log z_0)^p$ is featured in $F^r(\Delta\varphi)(z_0)$, or $z_0^{\lambda-1}(\log z_0)^p$ is featured in $F^{r+1}(Q^r\varphi)(z_0)$. If $p > 0$ and $\lambda = \frac{1}{2}n + r \pm \mu$, then (since $\mu \neq 0$ and thus $\partial_\lambda P_0^r(\lambda) \neq 0$) either $z_0^\lambda(\log z_0)^{p'}$ is featured in $F^r(\varphi)(z_0)$ for some $p' > p$, or $z_0^{\lambda-2}(\log z_0)^{p-1}$ is featured in $F^r(\Delta\varphi)(z_0)$, or $z_0^{\lambda-1}(\log z_0)^{p-1}$ is featured in $F^{r+1}(Q^r\varphi)(z_0)$, however the latter two cases are only possible when $\lambda = \frac{1}{2}n + r + \mu$ and $\mu \in \frac{1}{2}\mathbb{N}_0$. Together, these facts (applied to φ as well as its images under combinations of Δ and Q^r), imply that the weak expansion of u_i^{2j} has the form (7-24).

The asymptotic expansions (7-7) now follow from (7-24), since $\rho\partial_\rho = Z_0$ for our choice of ρ and, for each $r \in [0, m]$, by (6-7) and (7-11), we see that (identifying symmetric tensors with homogeneous polynomials in (x_0, x))

$$(tZ_0)^r u(x_0, x) = \sum_{r'=r}^m \sum_{\substack{s \geq 0 \\ r'+2s \leq m}} c_{m,r,r',s} x_0^{r'-r} |x|^{2s} u_{r'-2[r'/2]}^{2[r'/2]+2s}(x) \tag{7-28}$$

for some constants $c_{m,r,r',s}$; for later use, we also note that $c_{m,r,r,0} \neq 0$. □

7C. Surjectivity of the Poisson operator. In this section, we prove the surjectivity part of Theorem 6 in Section 5B (together with the injectivity part established in Corollary 6.9, this finishes the proof of that theorem). The remaining essential component of the proof is showing that, unless $u \equiv 0$, a certain term in the asymptotic expansion of Lemma 7.3 is nonzero (in particular we will see that u cannot be vanishing to infinite order on \mathbb{S}^n in the weak sense). We start with:

Lemma 7.4. *Take some u satisfying (7-6). Assume that, for all $r \in [0, m]$, the coefficient $w_{-\mu,0}^r$ of the weak expansion (7-7) is zero. (By Remark (ii) following Lemma 7.3, this condition is independent of the choice of ρ .) Then $u \equiv 0$. If $\mu \notin \frac{1}{2}\mathbb{N}_0$, then we can replace $w_{-\mu,0}^r$ by $w_{\mu,0,0}^r$ in the assumption above.*

Proof. We choose some $v \in \mathbb{S}^n$ and transform \mathbb{B}^{n+1} to the half-space model as explained in the proof of Lemma 7.3, and use the notation of that proof. Define the function $f \in C^\infty(\mathbb{B}^{n+1})$ in the half-space model as follows:

$$f = \begin{cases} z_0^{-m} u_0^{2m} & \text{if } m \text{ is even,} \\ z_0^{-m} u_1^{2m-1} & \text{if } m \text{ is odd.} \end{cases}$$

Here u_0^{2j} and u_1^{2j} are obtained by taking the inverse Fourier transforms of \hat{u}_0^{2j} and \hat{u}_1^{2j} . By (7-20) and (7-21) (see also (7-26)) we have

$$(\Delta_{\mathbb{H}^{n+1}} - \frac{1}{4}n^2 + \mu^2)f = 0. \tag{7-29}$$

Denote by $\mathcal{C}_{\text{temp}}^\infty(\mathbb{B}^{n+1})$ the set of smooth functions f in \mathbb{B}^{n+1} which are tempered in the sense that there exists $N \in \mathbb{R}$ such that $\rho_0^N f \in L^2(\mathbb{B}^{n+1})$. Set $\lambda := -\frac{1}{2}n + \mu$; it is proved in [van den Ban and Schlichtkrull 1987; Ōshima and Sekiguchi 1980] (see also [Grellier and Otal 2005] for a simpler presentation in the case $|\text{Re}(\lambda) + \frac{1}{2}n| < \frac{1}{2}n$) that the Poisson operator acting on distributions on hyperbolic space is an isomorphism

$$\mathcal{P}_\lambda^- : \mathcal{D}'(\mathbb{S}^n) \rightarrow \ker(\Delta_{\mathbb{H}^{n+1}} + \lambda(n + \lambda)) \cap \mathcal{C}_{\text{temp}}^\infty(\mathbb{B}^{n+1})$$

for $\lambda \notin -n - \mathbb{N}_0$, and, if $\operatorname{Re}(\lambda) \geq -\frac{1}{2}n$ with $\lambda \neq 0$, any element $v \in C_{\text{temp}}^\infty(\mathbb{B}^{n+1})$ with $(\Delta_{\mathbb{H}^{n+1}} + \lambda(n + \lambda))v = 0$ and $v \neq 0$ satisfies a weak expansion for any $N \in \mathbb{N}$,

$$v = \mathcal{P}_\lambda^-(v_{-\mu,\ell}) = \sum_{\ell=0}^N \left(\rho_0^{n/2-\mu+\ell} v_{-\mu,\ell} + \sum_{p=1}^{k_{\mu,\ell}} \rho_0^{n/2+\mu+\ell} \log(\rho_0)^p v_{\mu,\ell,p} \right) + \mathcal{O}(\rho_0^{n/2-\mu+N})$$

with $v_{-\mu,0} \neq 0$; moreover, $k_{\mu,\ell} = 0$ if $\lambda \notin -\frac{1}{2}n + \frac{1}{2}\mathbb{N}_0$, and $v_{\mu,0,0} \neq 0$ for such λ (here $v_{-\mu,\ell}, v_{\mu,\ell,p}$ are distributions on \mathbb{S}^n as before).⁵

Next, by (7-28), for some nonzero constant c we have

$$f = c(z_0^{-1} \iota_{z_0})^m u = c \langle u, \otimes^m \partial_{z_0} \rangle.$$

A calculation using (3-5) shows that in the ball model, using the geodesic boundary defining function ρ_0 from (3-34),

$$\partial_{z_0} = -\left(\frac{1}{2}(1 - |y|^2)v + (1 + y \cdot v)y\right) \partial_y \tag{7-30}$$

is a $C^\infty(\overline{\mathbb{B}^{n+1}})$ -linear combination of ∂_{ρ_0} and a 0-vector field. It follows from the form of the expansion (7-7) and the assumption of this lemma that the coefficient of $\rho_0^{n/2-\mu}$ of the weak expansion of f is zero. (If $\mu \notin \frac{1}{2}\mathbb{N}_0$, then we can also consider instead the coefficient of $\rho_0^{n/2+\mu}$.)

By (7-29) and the surjectivity of the scalar Poisson kernel discussed above, we now see that $f \equiv 0$. Now, for each fixed $y \in \mathbb{B}^{n+1}$ and each $\eta \in T_y \mathbb{B}^{n+1}$, we can choose v such that η is a multiple of (7-30) at y ; in fact, it suffices to take v such that the geodesic $\varphi_t(y, \eta)$ converges to $-v$ as $t \rightarrow +\infty$. Therefore, for each y, η , we have $\langle u, \otimes^m \eta \rangle = 0$ at y . Since u is a symmetric tensor, this implies $u \equiv 0$. \square

We now relax the assumptions of Lemma 7.4 to only include the term with $r = 0$:

Lemma 7.5. *Take some u satisfying (7-6). If $n = 1$ and $m > 0$, then we additionally assume that $\mu \neq \frac{1}{2}$. Assume that the coefficient $w_{-\mu,0}^0$ of the weak expansion (7-7) is zero. (By Remark (ii) following Lemma 7.3, this condition is independent of the choice of ρ .) Then $u \equiv 0$. If $\mu \notin \frac{1}{2}\mathbb{N}_0$, then we can replace $w_{-\mu,0}^0$ by $w_{\mu,0,0}^0$ in our assumption.*

Proof. Assume that $w_{\pm\mu,0}^0 = 0$; here we consider the case of $w_{\mu,0}^0 := w_{\mu,0,0}^0$ only when $\mu \notin \frac{1}{2}\mathbb{N}_0$. By Lemma 7.4, it suffices to prove that $w_{\pm\mu,0}^r = 0$ for $r = 0, \dots, m$. This is a local statement and we use the half-plane model and the notation of the proof of Lemma 7.3. By (7-28), it then suffices to show that, if $\tilde{w}_{0;\pm\mu,0}^0 = 0$ in the expansion (7-24), then $\tilde{w}_{i;\pm\mu,0}^{2j} = 0$ for all i, j .

We argue by induction on $r = 2j + i = 0, \dots, m$. Assume first that $i = 0, j > 0$, and $\tilde{w}_{1;\pm\mu,0}^{2(j-1)} = 0$. Then we plug (7-24) into (7-23) and consider the coefficient next to $z_0^{n/2 \pm \mu + 2j}$; this gives $\tilde{w}_{0;\pm\mu,0}^{2j} = 0$ if, for $\lambda = \frac{1}{2}n \pm \mu + 2j$, the following constant is nonzero:

$$-\lambda^2 + \left(n - \frac{\lambda j}{j}\right) \lambda + m - \sigma + \frac{\lambda j}{j} (c_m - j) = (n + 2m - 2 - 4j)(\pm 2\mu - n - 2m + 2 + 4j). \tag{7-31}$$

⁵The existence of the weak expansion with known coefficients for elements in the image of \mathcal{P}_λ^- is directly related to the special case $m = 0$ of Lemma 6.8 and the existence of a weak expansion for scalar eigenfunctions of the Laplacian follows from the $m = 0$ case of Lemma 7.3. However, neither the surjectivity of the scalar Poisson operator nor the fact that eigenfunctions have nontrivial terms in their weak expansions follows from these statements.

We see immediately that (7-31) is nonzero unless $m = 2j$. For the case $m = 2j$, we can use (7-19) directly; taking the coefficient next to $z_0^{n/2 \pm \mu + m}$, we get $\tilde{w}_{0; \pm \mu, 0}^{2j} = 0$ as long as $\frac{1}{2}n \pm \mu + m \neq c_m$, or equivalently $\pm \mu \neq \frac{1}{2}n - 1$; the latter inequality is immediately true unless $n = 1$, and it is explicitly excluded by the statement of the present lemma when $n = 1$.

Similarly, assume that $i = 1$, $0 \leq 2j < m$, and $\tilde{w}_{0; \pm \mu, 0}^{2j} = 0$. Then we plug (7-24) into (7-22) and consider the coefficient next to $z_0^{n/2 \pm \mu + 2j + 1}$; this gives $\tilde{w}_{1; \pm \mu, 0}^{2j} = 0$ if, for $\lambda = \frac{1}{2}n \pm \mu + 2j + 1$, the following constant is nonzero:

$$-\lambda^2 + \left(n + 2 - \frac{\lambda_{j+1}}{j+1} \right) \lambda - n + m - 1 - \sigma + \frac{\lambda_{j+1}}{j+1} (c_m - j) = (n + 2m - 4 - 4j)(\pm 2\mu - n - 2m + 4 + 4j). \tag{7-32}$$

We see immediately that (7-32) is nonzero unless $m = 2j + 1$. For the case $m = 2j + 1$, we can use (7-18) directly; taking the coefficient next to $z_0^{n/2 \pm \mu + m}$, we get $\tilde{w}_{1; \pm \mu, 0}^{2j} = 0$ as long as $\frac{1}{2}n \pm \mu + m \neq c_m$, which we have already established is true. \square

We finish the section by the following statement, which immediately implies the surjectivity part of Theorem 6. Note that, for the lifts of elements of $\text{Eig}^m(-\lambda(n + \lambda) + m)$, we can take any $\delta < \frac{1}{2}$ below. The condition $\text{Re } \lambda < \frac{1}{2} - \delta$ for $m > 0$ follows from Lemma 6.1.

Corollary 7.6. *Let $u \in \rho^\delta L^2(\bar{\mathbb{B}}^{n+1}; E^{(m)})$ be a trace-free symmetric m -cotensor with ρ a geodesic boundary defining function and $\delta \in (-\infty, \frac{1}{2})$, where the measure is the Euclidean Lebesgue measure on the ball. Assume that u is a nonzero divergence-free eigentensor for the Laplacian on hyperbolic space:*

$$\Delta u = (-\lambda(n + \lambda) + m)u, \quad \nabla^* u = 0, \tag{7-33}$$

with $\text{Re}(\lambda) < \frac{1}{2} - \delta$ and $\lambda \notin \mathcal{R}_m$, where \mathcal{R}_m is as defined in (5-20). Then, $u = \mathcal{P}_\lambda^-(w)$ for some $w \in H^{\text{Re}(\lambda) + \delta - 1/2}(\mathbb{S}^n; \otimes_S^m T^* \mathbb{S}^n)$. Moreover, if $\gamma^* u = u$ for some $\gamma \in G$, then $L_\gamma^* w = N_\gamma^{-\lambda - m} w$.

Proof. For the case $\text{Re}(\lambda) \geq -\frac{1}{2}n$ we set $\mu = \frac{1}{2}n + \lambda$ and apply Lemma 7.3; the distribution w will be given by $C(\lambda)w_{-\mu, 0}$ for some constant $C(\lambda)$ to be chosen, and this has the desired covariance with respect to elements of G by using (7-5) from the remark after Lemma 7.1.

To see that $u = \mathcal{P}_\lambda^-(w)$ for a certain $C(\lambda)$, it suffices to use the weak expansion in Lemma 6.8 and the identity (7-3) from the remark following Lemma 7.1, to deduce that $C(\lambda)B(\lambda)w_{-\mu, 0}$ appears as the leading coefficient of the power $\rho_0^{-\lambda}$ in the expansion of u , where $B(\lambda)$ is a nonzero constant times the factor appearing in (6-27); here ρ_0 is as defined in (3-34). (The factor $B(\lambda)$ does not depend on the point $\nu \in \mathbb{S}^n$ since the Poisson operator is equivariant under rotations of $\bar{\mathbb{B}}^{n+1}$.) Then, choosing $C(\lambda) := B(\lambda)^{-1}$, we observe that u and $\mathcal{P}_\lambda^-(w)$ both satisfy (7-33) and have the same asymptotic coefficient of $\rho_0^{-\lambda}$ in their weak expansion (7-7); thus from Lemma 7.5 we have $u = \mathcal{P}_\lambda^-(w)$. Finally, for $\text{Re}(\lambda) < -\frac{1}{2}n$ with $\lambda \notin -\frac{1}{2}n - \frac{1}{2}\mathbb{N}_0$ we do the same thing but setting $\mu := -\frac{1}{2}n - \lambda$ in Lemma 7.3. \square

Appendix A: Some technical calculations

A1. Asymptotic expansions for certain integrals. In this subsection, we prove the following version of Hadamard regularization:

Lemma A.1. Fix $\chi \in C_0^\infty(\mathbb{R})$ and define for $\operatorname{Re} \alpha > 0$, $\beta \in \mathbb{C}$, and $\varepsilon > 0$,

$$F_{\alpha\beta}(\varepsilon) := \int_0^\infty t^{\alpha-1}(1+t)^{-\beta} \chi(\varepsilon t) dt.$$

If $\alpha - \beta \notin \mathbb{N}_0$, then $F_{\alpha\beta}(\varepsilon)$ has the following asymptotic expansion as $\varepsilon \rightarrow +0$:

$$F_{\alpha\beta}(\varepsilon) = \frac{\Gamma(\alpha)\Gamma(\beta - \alpha)}{\Gamma(\beta)} \chi(0) + \sum_{0 \leq j \leq \operatorname{Re}(\alpha - \beta)} c_j \varepsilon^{\beta - \alpha + j} + o(1) \tag{A-1}$$

for some constants c_j depending on χ .

Proof. We use the following identity obtained by integrating by parts:

$$\varepsilon \partial_\varepsilon F_{\alpha\beta}(\varepsilon) = \int_0^\infty t^\alpha (1+t)^{-\beta} \partial_t (\chi(\varepsilon t)) dt = (\beta - \alpha) F_{\alpha\beta}(\varepsilon) - \beta F_{\alpha, \beta+1}(\varepsilon). \tag{A-2}$$

By using the Taylor expansion of χ at zero, we also see that

$$\chi(\varepsilon t) = \chi(0) + \mathcal{O}(\varepsilon t);$$

given the following formula, obtained by the change of variables $s = (1+t)^{-1}$ and using the beta function,

$$\int_0^\infty t^{\alpha-1}(1+t)^{-\beta} dt = \frac{\Gamma(\alpha)\Gamma(\beta - \alpha)}{\Gamma(\beta)} \quad \text{if } \operatorname{Re} \beta > \operatorname{Re} \alpha > 0,$$

we see that

$$F_{\alpha\beta}(\varepsilon) = \frac{\Gamma(\alpha)\Gamma(\beta - \alpha)}{\Gamma(\beta)} \chi(0) + \mathcal{O}(\varepsilon) \quad \text{if } \operatorname{Re}(\beta - \alpha) > 1.$$

By applying this asymptotic expansion to $F_{\alpha, \beta+M}$ for a large integer M and iterating (A-2), we derive the expansion (A-1). □

For the next result, we need the following two calculations (see Section 4A for some of the notation used):

Lemma A.2. For each $\ell \geq 0$,

$$\int_{\mathbb{S}^{n-1}} (\otimes^{2\ell} \eta) dS(\eta) = \frac{2\pi^{(n-1)/2} \Gamma(\ell + \frac{1}{2})}{\Gamma(\ell + \frac{1}{2}n)} \mathcal{S}(\otimes^\ell I),$$

where $I = \sum_{j=1}^n \partial_j \otimes \partial_j$.

Proof. Since both sides are symmetric tensors, it suffices to show that, for each $x \in \mathbb{R}^n$,

$$\int_{\mathbb{S}^{n-1}} (x \cdot \eta)^{2\ell} dS(\eta) = \frac{2\pi^{(n-1)/2} \Gamma(\ell + \frac{1}{2})}{\Gamma(\ell + \frac{1}{2}n)} |x|^{2\ell}.$$

Without loss of generality (using homogeneity and rotational invariance), we may assume that $x = \partial_1$. Then, using polar coordinates and Fubini's theorem, we have

$$\frac{1}{2} \Gamma(\ell + \frac{1}{2}n) \int_{\mathbb{S}^{n-1}} \eta_1^{2\ell} dS(\eta) = \int_{\mathbb{R}^n} e^{-|\eta|^2} \eta_1^{2\ell} d\eta = \pi^{(n-1)/2} \Gamma(\ell + \frac{1}{2}),$$

finishing the proof. □

Lemma A.3. For each $\eta \in \mathbb{R}^n$, define the linear map $\mathcal{C}_\eta : \mathbb{R}^n \rightarrow \mathbb{R}^n$ by

$$\mathcal{C}_\eta(\tilde{\eta}) = \tilde{\eta} - \frac{2}{1 + |\eta|^2}(\tilde{\eta} \cdot \eta)\eta.$$

Then, for each $A_1, A_2 \in \otimes_S^m \mathbb{R}^n$ with $\mathcal{T}(A_1) = \mathcal{T}(A_2) = 0$, and each $r \geq 0$, we have

$$\int_{\mathbb{S}^{n-1}} \langle (\otimes^m \mathcal{C}_{r\eta}) A_1, A_2 \rangle dS(\eta) = 2\pi^{n/2} \sum_{\ell=0}^m \frac{m!}{(m-\ell)! \Gamma(\frac{1}{2}n + \ell)} \left(-\frac{r^2}{1+r^2} \right)^\ell \langle A_1, A_2 \rangle.$$

Proof. We have

$$\mathcal{C}_{r\eta} = \text{Id} - \frac{2r^2}{1+r^2} \eta^* \otimes \eta,$$

where $\eta^* \in (\mathbb{R}^n)^*$ is the dual to η by the standard metric. Then

$$\int_{\mathbb{S}^{n-1}} \langle (\otimes^m \mathcal{C}_{r\eta}) A_1, A_2 \rangle dS(\eta) = \int_{\mathbb{S}^{n-1}} \left\langle \otimes^m \left(I - \frac{2r^2}{1+r^2} \eta \otimes \eta \right), \sigma(A_1 \otimes A_2) \right\rangle dS(\eta),$$

where σ is the operator defined by

$$\sigma(\eta_1 \otimes \dots \otimes \eta_m \otimes \eta'_1 \otimes \dots \otimes \eta'_m) = \eta_1 \otimes \eta'_1 \otimes \dots \otimes \eta_m \otimes \eta'_m.$$

We use Lemma A.2, a binomial expansion, and the fact that the A_j are symmetric, to calculate

$$\begin{aligned} & \int_{\mathbb{S}^{n-1}} \left\langle \otimes^m \left(I - \frac{2r^2}{1+r^2} \eta \otimes \eta \right), \sigma(A_1 \otimes A_2) \right\rangle dS(\eta) \\ &= \sum_{\ell=0}^m \frac{m!}{\ell!(m-\ell)!} \left(-\frac{2r^2}{1+r^2} \right)^\ell \int_{\mathbb{S}^{n-1}} \langle (\otimes^{2\ell} \eta) \otimes (\otimes^{m-\ell} I), \sigma(A_1 \otimes A_2) \rangle dS(\eta) \\ &= 2\pi^{(n-1)/2} \sum_{\ell=0}^m \frac{m!}{\ell!(m-\ell)!} \cdot \frac{\Gamma(\ell + \frac{1}{2})}{\Gamma(\ell + \frac{1}{2}n)} \left(-\frac{2r^2}{1+r^2} \right)^\ell \langle S(\otimes^\ell I) \otimes (\otimes^{m-\ell} I), \sigma(A_1 \otimes A_2) \rangle. \end{aligned}$$

Since $\mathcal{T}(A_1) = \mathcal{T}(A_2) = 0$, we can compute

$$\langle S(\otimes^\ell I) \otimes (\otimes^{m-\ell} I), \sigma(A_1 \otimes A_2) \rangle = \frac{2^\ell (\ell!)^2}{(2\ell)!} \langle A_1, A_2 \rangle.$$

Here $2^\ell (\ell!)^2 / (2\ell)!$ is the proportion of permutations τ of 2ℓ elements that satisfy, for each j , that $\tau(2j-1) + \tau(2j)$ is odd. It remains to calculate

$$\sum_{\ell=0}^m \frac{m!}{\ell!(m-\ell)!} \cdot \frac{\Gamma(\ell + \frac{1}{2})}{\Gamma(\ell + \frac{1}{2}n)} \cdot \frac{2^\ell (\ell!)^2}{(2\ell)!} t^\ell = \sum_{\ell=0}^m \frac{\sqrt{\pi} m!}{(m-\ell)! \Gamma(\ell + \frac{1}{2}n)} \left(\frac{1}{2}t \right)^\ell. \quad \square$$

We can now state the following asymptotic formula, used in the proof of Lemma 5.11:

Lemma A.4. *Let $\chi \in C_0^\infty(\mathbb{R})$ be equal to 1 near 0, and take $A_1, A_2 \in \otimes_S^m \mathbb{R}^n$ satisfying $\mathcal{T}(A_1) = \mathcal{T}(A_2) = 0$. Then, for $\lambda \in \mathbb{C}, \lambda \notin -(\frac{1}{2}n + \mathbb{N}_0)$, we have, as $\varepsilon \rightarrow +0$,*

$$\int_{\mathbb{R}^n} \chi(\varepsilon|\eta|)(1 + |\eta|^2)^{-\lambda-n} \langle (\otimes^m \mathcal{C}_\eta) A_1, A_2 \rangle d\eta$$

$$= \pi^{n/2} \frac{\Gamma(\frac{1}{2}n + \lambda)}{(n + \lambda + m - 1)\Gamma(n - 1 + \lambda)} \langle A_1, A_2 \rangle + \sum_{0 \leq j \leq -\operatorname{Re} \lambda - n/2} c_j \varepsilon^{n+2\lambda+2j} + o(1)$$

for some constants c_j .

Proof. We write, using the change of variables $\eta = \sqrt{t}\theta, \theta \in \mathbb{S}^n$, and $\chi(s) = \tilde{\chi}(s^2)$, and by Lemma A.3,

$$\int_{\mathbb{R}^n} \chi(\varepsilon|\eta|)(1 + |\eta|^2)^{-\lambda-n} \langle (\otimes^m \mathcal{C}_\eta) A_1, A_2 \rangle d\eta$$

$$= \frac{1}{2} \int_0^\infty \tilde{\chi}(\varepsilon^2 t) t^{\frac{1}{2}n-1} (1+t)^{-\lambda-n} \int_{\mathbb{S}^{n-1}} \langle (\otimes^m \mathcal{C}_{\sqrt{t}\theta}) A_1, A_2 \rangle dS(\theta) dt$$

$$= \pi^{n/2} \sum_{\ell=0}^m \frac{(-1)^\ell m!}{(m-\ell)! \Gamma(\frac{1}{2}n + \ell)} \langle A_1, A_2 \rangle \int_0^\infty \tilde{\chi}(\varepsilon^2 t) t^{n/2+\ell-1} (1+t)^{-\lambda-n-\ell} dt.$$

We now apply Lemma A.1 to get the required asymptotic expansion. The constant term in the expansion is $\langle A_1, A_2 \rangle$ times

$$\pi^{n/2} \Gamma(\frac{1}{2}n + \lambda) \sum_{\ell=0}^m \frac{(-1)^\ell m!}{(m-\ell)! \Gamma(n + \lambda + \ell)} = \pi^{n/2} (-1)^m m! \Gamma(\frac{1}{2}n + \lambda) \sum_{\ell=0}^m \frac{(-1)^\ell}{\ell! \Gamma(n + \lambda + m - \ell)}. \tag{A-3}$$

We now use the binomial expansion

$$\frac{(1-t)^{n+\lambda+m-1}}{\Gamma(n + \lambda + m)} = \sum_{\ell=0}^\infty \frac{(-1)^\ell}{\ell! \Gamma(n + \lambda + m - \ell)} t^\ell$$

and the sum in the last line of (A-3) is the t^m coefficient of

$$(1-t)^{-1} \frac{(1-t)^{n+\lambda+m-1}}{\Gamma(n + \lambda + m)} = \frac{(1-t)^{n+\lambda+m-2}}{\Gamma(n + \lambda + m)} = \frac{1}{n + \lambda + m - 1} \sum_{j=0}^\infty \frac{(-1)^j}{j! \Gamma(n + \lambda + m - j - 1)} t^j;$$

this finishes the proof. □

A2. The Jacobian of Ψ . Here we compute the Jacobian of the map $\Psi : \mathcal{E} \rightarrow S_\Delta^2 \mathbb{H}^{n+1}$ appearing in the proof of Lemma 5.11, proving (5-31). By the G -equivariance of Ψ , we may assume that $x = \partial_0, \xi = \partial_1, \eta = \sqrt{s} \partial_2$ for some $s \geq 0$. We then consider the following volume 1 basis of $T_{(x,\xi,\eta)} \mathcal{E}$:

$$X_1 = (\partial_1, \partial_0, 0), \quad X_2 = (\partial_2, 0, \sqrt{s} \partial_0), \quad X_3 = (0, \partial_2, -\sqrt{s} \partial_1), \quad X_4 = (0, 0, \partial_2);$$

$$\partial_{x_j}, \partial_{\xi_j}, \partial_{\eta_j}, \quad 3 \leq j \leq n + 1.$$

We have $\Psi(x, \xi, \eta) = (y, \eta_-, \eta_+)$, where

$$y = (\sqrt{s+1}, 0, \sqrt{s}, 0, \dots, 0), \quad \eta_{\pm} = \left(\mp \frac{s}{\sqrt{s+1}}, \frac{1}{\sqrt{s+1}}, \mp \sqrt{s}, 0, \dots, 0 \right).$$

Then we can consider the following volume 1 basis for $T_{(y, \eta_-, \eta_+)} S_{\Delta}^2 \mathbb{H}^{n+1}$:

$$\begin{aligned} Y_1 &= \left(\partial_1, \frac{y}{\sqrt{s+1}}, \frac{y}{\sqrt{s+1}} \right), & Y_2 &= \left(\sqrt{s} \partial_0 + \sqrt{s+1} \partial_2, \frac{\sqrt{s}}{\sqrt{s+1}} y, -\frac{\sqrt{s}}{\sqrt{s+1}} y \right), \\ Y_3 &= \frac{(0, \sqrt{s} \partial_0 - \sqrt{s} \partial_1 + \sqrt{s+1} \partial_2, 0)}{\sqrt{s+1}}, & Y_4 &= \frac{(0, 0, \sqrt{s} \partial_0 + \sqrt{s} \partial_1 + \sqrt{s+1} \partial_2)}{\sqrt{s+1}}; \\ & & & \partial_{y_j}, \partial_{v_{-j}}, \partial_{v_{+j}}, \quad 3 \leq j \leq n+1. \end{aligned}$$

Then the differential $d\Psi(x, \xi, \eta)$ maps

$$\begin{aligned} X_1 &\mapsto \sqrt{s+1} Y_1 - \sqrt{s} Y_3 - \sqrt{s} Y_4, \\ X_2 &\mapsto Y_2, \\ X_3 &\mapsto -\sqrt{s} Y_1 + \sqrt{s+1} Y_3 + \sqrt{s+1} Y_4, \\ X_4 &\mapsto \frac{1}{\sqrt{s+1}} Y_2 + \frac{1}{s+1} Y_3 - \frac{1}{s+1} Y_4. \end{aligned}$$

Moreover, for $3 \leq j \leq n+1$, $d\Psi(x, \xi, \eta)$ maps linear combinations of $\partial_{x_j}, \partial_{\xi_j}, \partial_{\eta_j}$ to linear combinations of $\partial_{y_j}, \partial_{v_{-j}}, \partial_{v_{+j}}$ by the matrix $A(s)$. The identity (5-31) now follows by a direct calculation.

A3. An identity for harmonic polynomials. We give a technical lemma which is used in the proof of Lemma 6.8 (injectivity of the Poisson kernel).

Lemma A.5. *Let P be a harmonic homogeneous polynomial of order m in \mathbb{R}^n ; then, for $r \leq m$, we have for all $x \in \mathbb{R}^n$ that*

$$\Delta_{\zeta}^r P(x - \zeta \langle \zeta, x \rangle) |_{\zeta=0} = 2^r \frac{m!r!}{(m-r)!} P(x).$$

Proof. By homogeneity, it suffices to choose $|x| = 1$. We set $t = \langle \zeta, x \rangle$ and $u = \zeta - tx$, and $P(x - \zeta \langle \zeta, x \rangle)$, viewed in the (t, u) coordinates, is the homogeneous polynomial $(t, u) \mapsto P((1-t^2)x - tu)$. Now, we write, for all $u \in (\mathbb{R}x)^{\perp}$ and $t > 0$,

$$P(tx - u) = \sum_{j=0}^m t^{m-j} P_j(u),$$

where P_j is a homogeneous polynomial of degree j in $u \in (\mathbb{R}x)^{\perp}$, and, since the Laplacian Δ_{ζ} written in the t, u coordinates is $-\partial_t^2 + \Delta_u$, the condition $\Delta_x P = 0$ can be rewritten

$$\Delta_u P_j(u) = (m-j+2)(m-j+1)P_{j-2}(u), \quad \Delta_u P_1(u) = \Delta_u P_0 = 0,$$

which gives, for all j and $\ell \geq 1$,

$$\Delta_u^{\ell} P_{2\ell}(u) = m(m-1) \cdots (m-2\ell+1)P_0, \quad \Delta^j P_{2\ell-1}(u) |_{u=0} = 0.$$

We write $\Delta_\zeta^r = \sum_{k=0}^r (r!/(k!(r-k)!))(-1)^k \partial_t^{2k} \Delta_u^{r-k}$ and, using parity and homogeneity considerations, we have

$$\begin{aligned} \Delta_\zeta^r P(x - \zeta \langle \zeta, x \rangle)|_{\zeta=0} &= \sum_{k=0}^r \frac{(-1)^k r!}{k!(r-k)!} \sum_{2j \leq m} [\partial_t^{2k} ((1-t^2)^{m-2j} t^{2j}) \Delta_u^{r-k} P_{2j}(u)]|_{(t,u)=0} \\ &= \sum_{\max(0, r-m/2) \leq k \leq r} \frac{(-1)^k r!}{k!(r-k)!} (\partial_t^{2k} ((1-t^2)^{m-2(r-k)} t^{2(r-k)}))|_{t=0} \Delta_u^{r-k} P_{2(r-k)} \\ &= P_0 \cdot \frac{m!r!}{(m-r)!} \sum_{r/2 \leq k \leq r} \frac{(-1)^{k+r} (2k)!}{k!(r-k)!(2k-r)!} = 2^r \frac{m!r!}{(m-r)!} P_0 \end{aligned}$$

and P_0 is the constant given by $P(x)$. Here we used the identity

$$\sum_{r/2 \leq k \leq r} \frac{(-1)^{k+r} (2k)!}{k!(r-k)!(2k-r)!} = \sum_{0 \leq k \leq r/2} (-1)^k \frac{r!}{k!(r-k)!} \cdot \frac{(2r-2k)!}{r!(r-2k)!} = 2^r,$$

which holds because both sides are equal to the t^r coefficient of the product

$$(1-t^2)^r \cdot (1-t)^{-1-r} = \frac{(1+t)^r}{1-t};$$

since

$$(1-t)^{-1-r} = \frac{1}{r!} d_t^r (1-t)^{-1} = \sum_{j=0}^{\infty} \frac{(j+r)!}{j!r!} t^j,$$

the t^r coefficient of $(1+t)^r/(1-t)$ equals the sum of the t^0, t^1, \dots, t^r coefficients of $(1+t)^r$, or simply $(1+1)^r = 2^r$. □

Appendix B: The special case of dimension 2

We explain how the argument of Section 2A fits into the framework of Sections 3 and 4. In dimension 2 it is more standard to use the upper half-plane model

$$\mathbf{H}^2 := \{w \in \mathbb{C} \mid \text{Im } w > 0\},$$

which is related to the half-space model of Section 3A by the formula $w = -z_1 + iz_0$.

The group of all isometries of \mathbf{H}^2 is $\text{PSL}(2; \mathbb{R})$, the quotient of $\text{SL}(2; \mathbb{R})$ by the group generated by the matrix $-\text{Id}$, and the action of $\text{PSL}(2; \mathbb{R})$ on \mathbf{H}^2 is by Möbius transformations:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d}, \quad z \in \mathbf{H}^2 \subset \mathbb{C}.$$

Under the identifications (3-2) and (3-5), this action corresponds to the action of $\text{PSO}(1, 2)$ on $\mathbb{H}^2 \subset \mathbb{R}^{1,2}$ by the group isomorphism $\text{PSL}(2; \mathbb{R}) \rightarrow \text{PSO}(1, 2)$ defined by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} \frac{1}{2}(a^2 + b^2 + c^2 + d^2) & \frac{1}{2}(a^2 - b^2 + c^2 - d^2) & -ab - cd \\ \frac{1}{2}(a^2 + b^2 - c^2 - d^2) & \frac{1}{2}(a^2 - b^2 - c^2 + d^2) & cd - ab \\ -ac - bd & bd - ac & ad + bc \end{pmatrix}. \tag{B-1}$$

The induced Lie algebra isomorphism maps the vector fields X, U_-, U_+ of (2-1) to the fields X, U_1^-, U_1^+ of (3-6), (3-7).

The horocyclic operators $\mathcal{U}_\pm : \mathcal{D}'(S\mathbb{H}^2) \rightarrow \mathcal{D}'(S\mathbb{H}^2; \mathcal{E}^*)$ of Section 4B (and analogously horocyclic operators of higher orders) then take the form

$$\mathcal{U}_\pm u = (U_\pm u)\eta^*,$$

where η^* is the dual to the section $\eta \in C^\infty(S\mathbb{H}^2; \mathcal{E})$ defined as follows: for $(x, \xi) \in S\mathbb{H}^2$, $\eta(x, \xi)$ is the unique vector in $T_x\mathbb{H}^2$ such that (ξ, η) is a positively oriented orthonormal frame. Note also that $\eta(x, \xi) = \pm \mathcal{A}_\pm(x, \xi) \cdot \zeta(B_\pm(x, \xi))$, where $\mathcal{A}_\pm(x, \xi)$ is as defined in Section 3F and $\zeta(v) \in T_v\mathbb{S}^1$, $v \in \mathbb{S}^1$, is the result of rotating v counterclockwise by $\frac{1}{2}\pi$; therefore, if we use η and ζ to trivialize the relevant vector bundles, then the operators \mathcal{Q}_\pm of (4-26) are simply the pullback operators by B_\pm , up to multiplication by ± 1 .

Appendix C: Eigenvalue asymptotics for symmetric tensors

C1. Weyl law. In this section, we prove the following asymptotic of the counting function for trace-free, divergence-free tensors (see Sections 4A and 6A for the notation):

Proposition C.1. *If (M, g) is a compact Riemannian manifold of dimension $n + 1$ and constant sectional curvature -1 , and if*

$$\text{Eig}^m(\sigma) = \{u \in C^\infty(M; \otimes_S^m T^*M) \mid \Delta u = \sigma u, \nabla^* u = 0, \mathcal{T}(u) = 0\},$$

then the following Weyl law holds as $R \rightarrow \infty$:

$$\sum_{\sigma \leq R^2} \dim \text{Eig}^m(\sigma) = c_0(n)(c_1(n, m) - c_1(n, m - 2)) \text{Vol}(M)R^{n+1} + \mathcal{O}(R^n),$$

where $c_0(n) = ((2\sqrt{\pi})^{-n-1} / \Gamma(\frac{1}{2}(n + 3)))$ and $c_1(n, m) = (m + n - 1)! / (m!(n - 1)!)$ is the dimension of the space of homogeneous polynomials of order m in n variables. (We put $c_1(n, m) := 0$ for $m < 0$.)

Remark. The constant $c_2(n, m) := c_1(n, m) - c_1(n, m - 2)$ is the dimension of the space of harmonic homogeneous polynomials of order m in n variables. We have

$$c_2(n, 0) = 1, \quad c_2(n, 1) = n.$$

For $m \geq 2$, we have $c_2(n, m) > 0$ if and only if $n > 1$.

The proof of Proposition C.1 uses the following two technical lemmas:

Lemma C.2. *Take $u \in \mathcal{D}'(M; \otimes_S^m T^*M)$. Then, denoting $D = \mathcal{S} \circ \nabla$ as in Section 6A,*

$$[\Delta, \nabla^*]u = (2 - 2m - n)\nabla^*u - 2(m - 1)D(\mathcal{T}(u)), \tag{C-1}$$

$$[\Delta, D]u = (2m + n)Du + 2m\mathcal{S}(g \otimes \nabla^*u). \tag{C-2}$$

Proof. We have

$$\Delta \nabla^*u = \mathcal{T}^2(\nabla^3u), \quad \nabla^* \Delta u = \mathcal{T}^2(\tau_{1 \leftrightarrow 3} \nabla^3u),$$

where $\tau_{j \leftrightarrow k}v$ denotes the result of swapping the j -th and k -th indices in a cotensor v . We have

$$\text{Id} - \tau_{1 \leftrightarrow 3} = (\text{Id} - \tau_{1 \leftrightarrow 2}) + \tau_{1 \leftrightarrow 2}(\text{Id} - \tau_{2 \leftrightarrow 3}) + \tau_{1 \leftrightarrow 2}\tau_{2 \leftrightarrow 3}(\text{Id} - \tau_{1 \leftrightarrow 2});$$

therefore (using that $\mathcal{T}\tau_{1 \leftrightarrow 2} = \mathcal{T}$)

$$[\Delta, \nabla^*]u = \mathcal{T}^2(\nabla(\text{Id} - \tau_{1 \leftrightarrow 2})\nabla^2u + \tau_{2 \leftrightarrow 3}(\text{Id} - \tau_{1 \leftrightarrow 2})\nabla^3u).$$

Since M has sectional curvature -1 , we have, for any cotensor v of rank m ,

$$(\text{Id} - \tau_{1 \leftrightarrow 2})\nabla^2v = \sum_{\ell=1}^m (\tau_{1 \leftrightarrow \ell+2} - \tau_{2 \leftrightarrow \ell+2})(g \otimes v).$$

Then we compute (using that $\mathcal{T}(\tau_{2 \leftrightarrow 3}\tau_{1 \leftrightarrow 3}) = \mathcal{T}(\tau_{2 \leftrightarrow 3})$)

$$[\Delta, \nabla^*]u = \mathcal{T}^2\left(\tau_{2 \leftrightarrow 3} - \text{Id} + \sum_{\ell=1}^m ((\tau_{2 \leftrightarrow \ell+3} - \tau_{3 \leftrightarrow \ell+3})\tau_{1 \leftrightarrow 3} + \tau_{2 \leftrightarrow 3}(\tau_{1 \leftrightarrow \ell+3} - \tau_{2 \leftrightarrow \ell+3}))\right)(g \otimes \nabla u).$$

Now,

$$\begin{aligned} \mathcal{T}^2(g \otimes \nabla u) &= \mathcal{T}^2(\tau_{2 \leftrightarrow 4}\tau_{1 \leftrightarrow 3}(g \otimes \nabla u)) = \mathcal{T}^2(\tau_{2 \leftrightarrow 3}\tau_{1 \leftrightarrow 4}(g \otimes \nabla u)) = -(n + 1)\nabla^*u, \\ \mathcal{T}^2(\tau_{2 \leftrightarrow 3}(g \otimes \nabla u)) &= \mathcal{T}^2(\tau_{3 \leftrightarrow 4}\tau_{1 \leftrightarrow 3}(g \otimes \nabla u)) = \mathcal{T}^2(\tau_{2 \leftrightarrow 3}\tau_{2 \leftrightarrow 4}(g \otimes \nabla u)) = -\nabla^*u, \end{aligned}$$

and, since u is symmetric, for $1 < \ell \leq m$,

$$\begin{aligned} \mathcal{T}^2(\tau_{2 \leftrightarrow \ell+3}\tau_{1 \leftrightarrow 3}(g \otimes \nabla u)) &= \mathcal{T}^2(\tau_{2 \leftrightarrow 3}\tau_{1 \leftrightarrow \ell+3}(g \otimes \nabla u)) = -\nabla^*u, \\ \mathcal{T}^2(\tau_{3 \leftrightarrow \ell+3}\tau_{1 \leftrightarrow 3}(g \otimes \nabla u)) &= \mathcal{T}^2(\tau_{2 \leftrightarrow 3}\tau_{2 \leftrightarrow \ell+3}(g \otimes \nabla u)) = \tau_{1 \leftrightarrow \ell-1}\nabla(\mathcal{T}(u)). \end{aligned}$$

We then compute

$$[\Delta, \nabla^*]u = (2 - 2m - n)\nabla^*u - 2 \sum_{\ell=1}^{m-1} \tau_{1 \leftrightarrow \ell} \nabla(\mathcal{T}(u)),$$

finishing the proof of (C-1). The identity (C-2) follows from (C-1) by taking the adjoint on the space of symmetric tensors. \square

Lemma C.3. *Denote by $\tilde{\pi}_m : \otimes_S^m T^*M \rightarrow \otimes_S^m T^*M$ the orthogonal projection onto the space $\ker \mathcal{T}$ of trace-free tensors. Then, for each m , the space*

$$F^m := \{v \in C^\infty(M; \otimes_S^m T^*M) \mid \mathcal{T}(v) = 0, \tilde{\pi}_{m+1}(Dv) = 0\} \tag{C-3}$$

is finite-dimensional.

Proof. The space F^m is contained in the kernel of the operator

$$P_m := \nabla^* \tilde{\pi}_{m+1} D$$

acting on trace-free sections of $\otimes_S^m T^*M$. By [Dairbekov and Sharafutdinov 2010, Lemma 5.2], the operator P_m is elliptic; therefore, its kernel is finite-dimensional. \square

We now prove Proposition C.1. For each $m \geq 0$ and $s \in \mathbb{R}$, denote

$$W^m(\sigma) := \{u \in \mathcal{D}'(M; \otimes_S^m T^*M) \mid \Delta u = \sigma u, \mathcal{T}(u) = 0\}.$$

The operator Δ acting on trace-free symmetric tensors is elliptic and, in fact, its principal symbol coincides with that of the scalar Laplacian: $p(x, \xi) = |\xi|_g^2$. It follows that the $W^m(\sigma)$ are finite-dimensional and consist of smooth sections. By the general argument of [Hörmander 1994, Section 17.5] (see also [Dimassi and Sjöstrand 1999, Theorem 10.1; Zworski 2012, Theorem 6.8] — all of these arguments adapt straightforwardly to the case of operators with diagonal principal symbols acting on vector bundles), we have the following Weyl law:

$$\sum_{\sigma \leq R^2} \dim W^m(\sigma) = c_0(n)(c_1(n+1, m) - c_1(n+1, m-2)) \text{Vol}(M) R^{n+1} + \mathcal{O}(R^n); \quad (\text{C-4})$$

here $c_1(n+1, m) - c_1(n+1, m+2)$ is the dimension of the vector bundle on which we consider the operator Δ .

By (C-1), for $m \geq 1$ the divergence operator acts as

$$\nabla^* : W^m(\sigma) \rightarrow W^{m-1}(\sigma + 2 - 2m - n). \quad (\text{C-5})$$

This operator is surjective except at finitely many points σ :

Lemma C.4. *Let $C_1 = \dim F^{m-1}$, where F^{m-1} is as defined in (C-3). Then the number of values σ such that (C-5) is not surjective does not exceed C_1 .*

Proof. Assume that (C-5) is not surjective for some σ . Then there exists nonzero $v \in W^{m-1}(\sigma + 2 - 2m - n)$ which is orthogonal to $\nabla^*(W^m(\sigma))$. Since the spaces $W^{m-1}(\sigma)$ are mutually orthogonal, we see from (C-5) that v is also orthogonal to $\nabla^*(W^m(\sigma))$ for all $\sigma \neq \sigma$. It follows that, for each σ and each $u \in W^m(\sigma)$, we have $\langle Dv, u \rangle_{L^2} = 0$. Since $\bigoplus_{\sigma} W^m(\sigma)$ is dense in the space of trace-free tensors, we see that, for each $u \in C^\infty(M; \otimes_S^m T^*M)$ with $\mathcal{T}(u) = 0$, we have $\langle Dv, u \rangle_{L^2} = 0$, which implies $v \in F^{m-1}$. It remains to note that F^{m-1} can have a nontrivial intersection with at most C_1 of the spaces $W^{m-1}(\sigma + 2 - 2m - n)$. \square

Since $\text{Eig}^m(\sigma)$ is the kernel of (C-5), we have

$$\dim \text{Eig}^m(\sigma) \geq \dim W^m(\sigma) - \dim W^{m-1}(\sigma + 2 - 2m - n),$$

and this inequality is an equality if (C-5) is surjective. We then see that, for some constant C_2 independent of R ,

$$\begin{aligned} \sum_{\sigma \leq R^2} \dim W^m(\sigma) - \sum_{\sigma \leq R^2+2-2m-n} \dim W^{m-1}(\sigma) &\leq \sum_{\sigma \leq R^2} \dim \text{Eig}^m(\sigma) \\ &\leq C_2 + \sum_{\sigma \leq R^2} \dim W^m(\sigma) - \sum_{\sigma \leq R^2+2-2m-n} \dim W^{m-1}(\sigma), \end{aligned}$$

and Proposition C.1 now follows from (C-4) and the identity $c_1(n + 1, m) - c_1(n + 1, m - 1) = c_1(n, m)$.

C2. The case $m = 1$. In this section, we describe the space $\text{Eig}^1(\sigma)$ in terms of Hodge theory; see, for instance, [Petersen 2006, Section 7.2] for the notation used. Note that symmetric cotensors of order 1 are exactly differential 1-forms on M . Since the operator $\nabla : C^\infty(M) \rightarrow C^\infty(M; T^*M)$ is equal to the operator d on 0-forms, we have

$$\text{Eig}^1(\sigma) = \{u \in \Omega^1(M) \mid \Delta u = \sigma u, \delta u = 0\}.$$

Here $\Delta = \nabla^* \nabla$; using that M has sectional curvature -1 , we write Δ in terms of the Hodge Laplacian $\Delta_\Omega := d\delta + \delta d$ on 1-forms using the following Weitzenböck formula [Petersen 2006, Corollary 7.21]:

$$\Delta u = (\Delta_\Omega + n)u, \quad u \in \Omega^1(M).$$

We then see that

$$\text{Eig}^1(\sigma) = \{u \in \Omega^1(M) \mid \Delta_\Omega u = (\sigma - n)u, \delta u = 0\}. \tag{C-6}$$

Finally, let us consider the case $n = 1$. The Hodge star operator acts from $\Omega^1(M)$ to itself, and we see that, for $\sigma \neq 1$,

$$\begin{aligned} \text{Eig}^1(\sigma) &= \{*u \mid u \in \Omega^1(M), \Delta_\Omega u = (\sigma - 1)u, du = 0\} \\ &= \{*(df) \mid f \in C^\infty(M), \Delta f = (\sigma - 1)f\}. \end{aligned} \tag{C-7}$$

Note that $*(df)$ can be viewed as the Hamiltonian field of f with respect to the naturally induced symplectic form (that is, volume form) on M .

Acknowledgements

We would like to thank Maciej Zworski, Richard Melrose, Steve Zelditch, Rafe Mazzeo, and Kiril Datchev for many useful discussions and suggestions regarding this project, and Tobias Weich for corrections and references. Dyatlov would like to thank MSRI (NSF grant 0932078 000, Fall 2013), where part of this work was done; another part was supported by the NSF grant DMS-1201417, and part of this project was completed during the period Dyatlov served as a Clay Research Fellow. Guillarmou and Faure have been supported by “Agence Nationale de la Recherche” under the grant ANR-13-BS01-0007-01.

References

- [Abramowitz and Stegun 1964] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, National Bureau of Standards Applied Mathematics Series **55**, U.S. Government Printing Office, Washington, D.C., 1964. MR 29 #4914 Zbl 0171.38503
- [Anantharaman and Zelditch 2007] N. Anantharaman and S. Zelditch, “Patterson–Sullivan distributions and quantum ergodicity”, *Ann. Henri Poincaré* **8**:2 (2007), 361–426. MR 2008f:58028 Zbl 1187.81175
- [Anantharaman and Zelditch 2012] N. Anantharaman and S. Zelditch, “Intertwining the geodesic flow and the Schrödinger group on hyperbolic surfaces”, *Math. Ann.* **353**:4 (2012), 1103–1156. MR 2944025 Zbl 06065228
- [van den Ban and Schlichtkrull 1987] E. P. van den Ban and H. Schlichtkrull, “Asymptotic expansions and boundary values of eigenfunctions on Riemannian symmetric spaces”, *J. Reine Angew. Math.* **380** (1987), 108–165. MR 89g:43010 Zbl 0631.58028
- [Bony and Häfner 2008] J.-F. Bony and D. Häfner, “Decay and non-decay of the local energy for the wave equation on the de Sitter–Schwarzschild metric”, *Comm. Math. Phys.* **282**:3 (2008), 697–719. MR 2010h:58041 Zbl 1159.35007
- [Bunke and Olbrich 1995] U. Bunke and M. Olbrich, *Selberg zeta and theta functions: a differential operator approach*, Mathematical Research **83**, Akademie-Verlag, Berlin, 1995. MR 97c:11088 Zbl 0831.58001
- [Bunke and Olbrich 1997] U. Bunke and M. Olbrich, “Fuchsian groups of the second kind and representations carried by the limit set”, *Invent. Math.* **127**:1 (1997), 127–154. MR 98e:11065 Zbl 0880.30035
- [Bunke and Olbrich 1999] U. Bunke and M. Olbrich, “Group cohomology and the singularities of the Selberg zeta function associated to a Kleinian group”, *Ann. of Math. (2)* **149**:2 (1999), 627–689. MR 2000f:11110 Zbl 0969.11019
- [Bunke and Olbrich 2001] U. Bunke and M. Olbrich, “Regularity of invariant distributions”, preprint, 2001. arXiv 0103144
- [Butterley and Liverani 2007] O. Butterley and C. Liverani, “Smooth Anosov flows: correlation spectra and stability”, *J. Mod. Dyn.* **1**:2 (2007), 301–322. MR 2008a:37035 Zbl 1144.37011
- [Dairbekov and Sharafutdinov 2010] N. S. Dairbekov and V. A. Sharafutdinov, “Conformal Killing symmetric tensor fields on Riemannian manifolds”, *Mat. Tr.* **13**:1 (2010), 85–145. In Russian; translated in *Sib. Adv. Math.* **21**:1 (2011), 1–41. MR 2011j:53052 Zbl 1249.53050
- [Datchev et al. 2014] K. Datchev, S. Dyatlov, and M. Zworski, “Sharp polynomial bounds on the number of Pollicott–Ruelle resonances”, *Ergodic Theory Dynam. Systems* **34**:4 (2014), 1168–1183. MR 3227152 Zbl 06359408
- [Dimassi and Sjöstrand 1999] M. Dimassi and J. Sjöstrand, *Spectral asymptotics in the semi-classical limit*, London Mathematical Society Lecture Note Series **268**, Cambridge University Press, 1999. MR 2001b:35237 Zbl 0926.35002
- [Dolgopyat 1998] D. Dolgopyat, “On decay of correlations in Anosov flows”, *Ann. of Math. (2)* **147**:2 (1998), 357–390. MR 99g:58073 Zbl 0911.58029
- [Dyatlov 2012] S. Dyatlov, “Asymptotic distribution of quasi-normal modes for Kerr–de Sitter black holes”, *Ann. Henri Poincaré* **13**:5 (2012), 1101–1166. MR 2935116 Zbl 1246.83111
- [Dyatlov 2015] S. Dyatlov, “Asymptotics of linear waves and resonances with applications to black holes”, *Comm. Math. Phys.* **335**:3 (2015), 1445–1485. MR 3320319 Zbl 06421372
- [Dyatlov and Zworski 2015] S. Dyatlov and M. Zworski, “Dynamical zeta functions for Anosov flows via microlocal analysis”, preprint, 2015. To appear in *Ann. Sci. École Norm. Sup.* arXiv 1306.4203
- [Faure and Sjöstrand 2011] F. Faure and J. Sjöstrand, “Upper bound on the density of Ruelle resonances for Anosov flows”, *Comm. Math. Phys.* **308**:2 (2011), 325–364. MR 2851145 Zbl 1260.37016
- [Faure and Tsujii 2013a] F. Faure and M. Tsujii, “Prequantum transfer operator for symplectic Anosov diffeomorphism”, preprint, 2013. To appear in the Astérisque series. arXiv 1206.0282
- [Faure and Tsujii 2013b] F. Faure and M. Tsujii, “Band structure of the Ruelle spectrum of contact Anosov flows”, *C. R. Math. Acad. Sci. Paris* **351**:9–10 (2013), 385–391. MR 3072166 Zbl 06192619
- [Faure and Tsujii 2014] F. Faure and M. Tsujii, “The semiclassical zeta function for geodesic flows on negatively curved manifolds”, preprint, 2014. arXiv 1311.4932
- [Flaminio and Forni 2003] L. Flaminio and G. Forni, “Invariant distributions and time averages for horocycle flows”, *Duke Math. J.* **119**:3 (2003), 465–526. MR 2004g:37039 Zbl 1044.37017

- [Fried 1986] D. Fried, “Analytic torsion and closed geodesics on hyperbolic manifolds”, *Invent. Math.* **84**:3 (1986), 523–540. MR 87g:58118 Zbl 0621.53035
- [Fried 1995] D. Fried, “Meromorphic zeta functions for analytic flows”, *Comm. Math. Phys.* **174** (1995), 161–190. MR 97a:58148 Zbl 0841.58053
- [Gel’fand and Fomin 1952] I. M. Gel’fand and S. V. Fomin, “Geodesic flows on manifolds of constant negative curvature”, *Uspehi Matem. Nauk (N.S.)* **7**:1(47) (1952), 118–137. In Russian; translated in *Amer. Math. Soc. Transl. (2)* **1** (1955), 49–65. MR 14,660f Zbl 0048.09203
- [Giulietti et al. 2013] P. Giulietti, C. Liverani, and M. Pollicott, “Anosov flows and dynamical zeta functions”, *Ann. of Math. (2)* **178**:2 (2013), 687–773. MR 3071508 Zbl 06203676
- [Grellier and Otal 2005] S. Grellier and J.-P. Otal, “Bounded eigenfunctions in the real hyperbolic space”, *Int. Math. Res. Not.* **2005**:62 (2005), 3867–3897. MR 2007a:58035 Zbl 1139.58024
- [Guillarmou et al. 2010] C. Guillarmou, S. Moroianu, and J. Park, “Eta invariant and Selberg zeta function of odd type over convex co-compact hyperbolic manifolds”, *Adv. Math.* **225**:5 (2010), 2464–2516. MR 2011m:58038 Zbl 1209.58017
- [Guillemin 1977] V. Guillemin, “Lectures on spectral theory of elliptic operators”, *Duke Math. J.* **44**:3 (1977), 485–517. MR 56 #6758 Zbl 0463.58024
- [Hansen et al. 2012] S. Hansen, J. Hilgert, and M. Schröder, “Patterson–Sullivan distributions in higher rank”, *Math. Z.* **272**:1–2 (2012), 607–643. MR 2968245 Zbl 1252.53062
- [Helgason 1970] S. Helgason, “A duality for symmetric spaces with applications to group representations”, *Advances in Math.* **5** (1970), 1–154. MR 41 #8587 Zbl 0209.25403
- [Helgason 1974] S. Helgason, “Eigenspaces of the Laplacian; integral representations and irreducibility”, *J. Functional Analysis* **17** (1974), 328–353. MR 51 #3353 Zbl 0303.43021
- [Hörmander 1983] L. Hörmander, *The analysis of linear partial differential operators, I: Distribution theory and Fourier analysis*, Grundlehren der Math. Wissenschaften **256**, Springer, Berlin, 1983. MR 85g:35002a Zbl 0521.35001
- [Hörmander 1994] L. Hörmander, *The analysis of linear partial differential operators, III: Pseudo-differential operators*, Grundlehren der Math. Wissenschaften **274**, Springer, Berlin, 1994. MR 95h:35255 Zbl 0601.35001
- [Juhl 2001] A. Juhl, *Cohomological theory of dynamical zeta functions*, Progress in Mathematics **194**, Birkhäuser, Basel, 2001. MR 2001j:37046 Zbl 1099.37017
- [Leboeuf 2004] P. Leboeuf, “Periodic orbit spectrum in terms of Ruelle–Pollicott resonances”, *Phys. Rev. E* (3) **69**:2 (2004), 026204, 13. MR 2005g:37120
- [Liverani 2004] C. Liverani, “On contact Anosov flows”, *Ann. of Math. (2)* **159**:3 (2004), 1275–1312. MR 2005k:37048 Zbl 1067.37031
- [Mazzeo 1991] R. Mazzeo, “Elliptic theory of differential edge operators, I”, *Comm. Partial Differential Equations* **16**:10 (1991), 1615–1664. MR 93d:58152 Zbl 0745.58045
- [Mazzeo and Melrose 1987] R. R. Mazzeo and R. B. Melrose, “Meromorphic extension of the resolvent on complete spaces with asymptotically constant negative curvature”, *J. Funct. Anal.* **75**:2 (1987), 260–310. MR 89c:58133 Zbl 0636.58034
- [Melrose 1993] R. B. Melrose, *The Atiyah–Patodi–Singer index theorem*, Research Notes in Mathematics **4**, A K Peters, Ltd., Wellesley, MA, 1993. MR 96g:58180 Zbl 0796.58050
- [Minemura 1975] K. Minemura, “Eigenfunctions of the Laplacian on a real hyperbolic space”, *J. Math. Soc. Japan* **27**:1 (1975), 82–105. MR 52 #3432 Zbl 0292.35023
- [Moore 1987] C. C. Moore, “Exponential decay of correlation coefficients for geodesic flows”, pp. 163–181 in *Group representations, ergodic theory, operator algebras, and mathematical physics* (Berkeley, CA, 1984), edited by C. C. Moore, Math. Sci. Res. Inst. Publ. **6**, Springer, New York, 1987. MR 89d:58102 Zbl 0625.58023
- [Nonnenmacher and Zworski 2015] S. Nonnenmacher and M. Zworski, “Decay of correlations for normally hyperbolic trapping”, *Invent. Math.* **200**:2 (2015), 345–438. MR 3338007
- [Olbrich 1995] M. Olbrich, *Die Poisson-Transformation für homogene Vektorbündel*, Ph.D. thesis, Humboldt-Universität zu Berlin, 1995, <http://www.uni-math.gwdg.de/olbrich/>.

- [Ōshima and Sekiguchi 1980] T. Ōshima and J. Sekiguchi, “Eigenspaces of invariant differential operators on an affine symmetric space”, *Invent. Math.* **57**:1 (1980), 1–81. MR 81k:43014 Zbl 0434.58020
- [Otal 1998] J.-P. Otal, “Sur les fonctions propres du laplacien du disque hyperbolique”, *C. R. Acad. Sci. Paris Sér. I Math.* **327**:2 (1998), 161–166. MR 99e:35161 Zbl 0919.31005
- [Petersen 2006] P. Petersen, *Riemannian geometry*, 2nd ed., Graduate Texts in Mathematics **171**, Springer, New York, 2006. MR 2007a:53001 Zbl 1220.53002
- [Pollicott 1986] M. Pollicott, “Meromorphic extensions of generalised zeta functions”, *Invent. Math.* **85**:1 (1986), 147–164. MR 87k:58218 Zbl 0604.58042
- [Pollicott 1989] M. Pollicott, “Distributions at infinity for Riemann surfaces”, pp. 91–100 in *Dynamical systems and ergodic theory* (Warsaw, 1986), edited by K. Krzyżewski, Banach Center Publ. **23**, PWN, Warsaw, 1989. MR 92e:58162 Zbl 0696.58037
- [Ratner 1987] M. Ratner, “The rate of mixing for geodesic and horocycle flows”, *Ergodic Theory Dynam. Systems* **7**:2 (1987), 267–288. MR 88j:58103 Zbl 0623.22008
- [Ruelle 1986] D. Ruelle, “Resonances of chaotic dynamical systems”, *Phys. Rev. Lett.* **56**:5 (1986), 405–407. MR 87f:58087
- [Tsujii 2010] M. Tsujii, “Quasi-compactness of transfer operators for contact Anosov flows”, *Nonlinearity* **23**:7 (2010), 1495–1545. MR 2012e:37040 Zbl 1200.37025
- [Tsujii 2012] M. Tsujii, “Contact Anosov flows and the Fourier–Bros–Iagolnitzer transform”, *Ergodic Theory Dynam. Systems* **32**:6 (2012), 2083–2118. MR 2995886 Zbl 1276.37020
- [Zelditch 1987] S. Zelditch, “Uniform distribution of eigenfunctions on compact hyperbolic surfaces”, *Duke Math. J.* **55**:4 (1987), 919–941. MR 89d:58129 Zbl 0643.58029
- [Zworski 2012] M. Zworski, *Semiclassical analysis*, Graduate Studies in Mathematics **138**, American Mathematical Society, Providence, RI, 2012. MR 2952218 Zbl 1252.58001

Received 28 Oct 2014. Revised 30 Jan 2015. Accepted 6 Mar 2015.

SEMYON DYATLOV: dyatlov@math.mit.edu

Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, United States

FRÉDÉRIC FAURE: frederic.faure@ujf-grenoble.fr

Institut Fourier, Université Joseph Fourier, 100, rue des Maths, BP74, 38402 St Martin d’Heres, France

COLIN GUILLARMOU: cguillar@dma.ens.fr

DMA, U.M.R. 8553 CNRS, École Normale Supérieure, 45 rue d’Ulm, DMA, 75230 Paris, France

PAVING OVER ARBITRARY MASAS IN VON NEUMANN ALGEBRAS

SORIN POPA AND STEFAAN VAES

We consider a paving property for a maximal abelian $*$ -subalgebra (MASA) A in a von Neumann algebra M , that we call *so-paving*, involving approximation in the *so*-topology, rather than in norm (as in classical Kadison–Singer paving). If A is the range of a normal conditional expectation, then *so-paving* is equivalent to norm paving in the ultrapower inclusion $A^\omega \subset M^\omega$. We conjecture that any MASA in any von Neumann algebra satisfies *so-paving*. We use work of Marcus, Spielman and Srivastava to check this for all MASAs in $\mathcal{B}(\ell^2\mathbb{N})$, all Cartan subalgebras in amenable von Neumann algebras and in group measure space II_1 factors arising from profinite actions. By earlier work of Popa, the conjecture also holds true for singular MASAs in II_1 factors, and we obtain here an improved paving size $C\varepsilon^{-2}$, which we show to be sharp.

1. Introduction

A famous problem of R. V. Kadison and I. M. Singer [1959] asked whether the diagonal MASA (maximal abelian $*$ -subalgebra) \mathcal{D} in the algebra $\mathcal{B}(\ell^2\mathbb{N})$ of all linear bounded operators on the Hilbert space $\ell^2\mathbb{N}$ satisfies the *paving property*, requiring that, for any $x \in \mathcal{B}(\ell^2\mathbb{N})$ with 0 on the diagonal and any $\varepsilon > 0$, there exists a partition of 1 with projections $p_1, \dots, p_n \in \mathcal{D}$ such that $\|\sum_i p_i x p_i\| \leq \varepsilon \|x\|$.

In striking work, A. Marcus, D. Spielman and N. Srivastava [Marcus et al. 2015] have settled this question in the affirmative, while also obtaining an estimate for the minimal number of projections necessary for such ε paving, $n(x, \varepsilon) \leq 12^4 \varepsilon^{-4}$ for all $x = x^* \in \mathcal{B}(\ell^2\mathbb{N})$.

On the other hand, in [Popa 2014] the paving property for $\mathcal{D} \subset \mathcal{B}(\ell^2\mathbb{N})$ has been shown to be equivalent to the paving property for the ultrapower inclusion $D^\omega \subset R^\omega$, where R is the hyperfinite II_1 factor, D is its Cartan subalgebra and ω is a free ultrafilter on \mathbb{N} . (Recall from [Dixmier 1954; Feldman and Moore 1977] that a subalgebra A in a von Neumann algebra M is a *Cartan subalgebra* if it is a MASA, there exists a normal conditional expectation of M onto A , and the normalizer of A in M , $\mathcal{N}_M(A) = \{u \in \mathcal{U}(M) \mid uAu^* = A\}$, generates M .) It was also shown in [Popa 2014] that if A is a singular MASA in R , or, more generally, in an arbitrary II_1 factor M , then $A^\omega \subset M^\omega$ has the paving property, with corresponding paving size majorized by $C\varepsilon^{-3}$. (Recall from [Dixmier 1954] that a MASA $A \subset M$ is *singular* in M if its normalizer is trivial, that is, $\mathcal{N}_M(A) \subset A$.)

Inspired by these results, we consider in this paper a new, weaker paving property for an arbitrary MASA A in a von Neumann algebra M that we call *so-paving*, which requires that, for any

Popa is supported in part by NSF Grant DMS-1401718. Vaes is supported by ERC Consolidator Grant 614195 from the European Research Council under the European Union's Seventh Framework Programme.

MSC2010: primary 46L10; secondary 46A22, 46L30.

Keywords: Kadison–Singer problem, paving, von Neumann algebra, maximal abelian subalgebra.

$x \in M_{\text{sa}} = \{x \in M \mid x = x^*\}$ and $\varepsilon > 0$, there exists n such that x can be (ε, n) *so-paved*, that is, for any *so*-neighborhood \mathcal{V} of 0 there exists a partition of 1 with projections p_1, \dots, p_n in A and an element $a \in A$ satisfying $\|a\| \leq \|x\|$ and $\|q(\sum_i p_i x p_i - a)q\| \leq \varepsilon \|x\|$ for some projection $q \in M$ with $1 - q \in \mathcal{V}$ (see Section 2). We prove that, if there exists a normal conditional expectation from M onto A , then *so*-paving is equivalent to the property that, for any $x \in M_{\text{sa}}$ and $\varepsilon > 0$, there exists n such that x can be approximated in the *so*-topology with elements that can be (ε, n) norm paved (see Theorem 2.7). If in addition A is countably decomposable, then *so*-paving with uniform bound on the number n necessary to (ε, n) *so*-pave any $x \in M_{\text{sa}}$ is equivalent to the ultrapower inclusion $A^\omega \subset M^\omega$ satisfying norm paving (with M^ω as defined in [Ocneanu 1985]). In particular, this shows that *so*-paving amounts to norm paving in the case $\mathcal{D} \subset \mathcal{B}(\ell^2\mathbb{N})$.

We conjecture that any MASA in any von Neumann algebra satisfies the *so*-paving property (see Conjecture 2.8). We use [Marcus et al. 2015] to check this conjecture for all MASAs in $\mathcal{B}(\ell^2\mathbb{N})$ (i.e., for the remaining case of the diffuse MASA $L^\infty([0, 1]) \subset \mathcal{B}(L^2([0, 1]))$; see Section 3), for all Cartan subalgebras in amenable von Neumann algebras, as well as for any Cartan subalgebra in a group measure space II_1 factor arising from a free ergodic measure-preserving profinite action (see Section 4). At the same time, we prove that, for a von Neumann algebra M with separable predual, norm paving over a MASA $A \subset M$ occurs if and only if M is of type I and there exists a normal conditional expectation of M onto A (see Theorem 3.3).

For singular MASAs $A \subset M$, where the conjecture already follows from results in [Popa 2014], we improve upon the paving size obtained there, by showing that any finite number of elements in M^ω can be simultaneously ε paved over A^ω with $n < 1 + 16\varepsilon^{-2}$ projections (see Theorem 5.1). Moreover, this estimate is sharp: given any MASA in a finite factor, $A \subset M$, and any $\varepsilon > 0$, there exists $x \in M_{\text{sa}}$ with zero expectation onto A such that, if $\|\sum_{i=1}^n p_i x p_i\| \leq \varepsilon \|x\|$ for some partition of 1 with projections in A , then n must be at least ε^{-2} (see Proposition 5.4). We include a discussion on the multipaving size for $\mathcal{D} \subset \mathcal{B}(\ell^2\mathbb{N})$ and, more generally, for Cartan subalgebras (see Remark 5.2).

2. A paving conjecture for MASAs

We will consider several paving properties for a MASA A in a von Neumann algebra M . For convenience we first recall the initial Kadison–Singer paving property [1959], for which we use the following terminology.

Definition 2.1. We say an element $x \in M$ is (ε, n) *pavable over* A if there exist projections $p_1, \dots, p_n \in A$ and $a \in A$ such that $\|a\| \leq \|x\|$, $\sum_{i=1}^n p_i = 1$ and $\|\sum_{i=1}^n p_i x p_i - a\| \leq \varepsilon \|x\|$. We denote by $\mathfrak{n}(A \subset M; x, \varepsilon)$ (or just $\mathfrak{n}(x, \varepsilon)$, if no confusion is possible), the smallest such n . Also, we say that x is *pavable* (over A) if, for every $\varepsilon > 0$, there exists an n such that x is (ε, n) pavable. We say that $A \subset M$ has the *paving property* if any $x \in M$ is pavable. We will sometimes use the terminology *norm pavable/paving* instead of just pavable/paving, when we need to underline the difference with other paving properties.

It is not really crucial to impose $\|a\| \leq \|x\|$. Indeed, without that assumption, the element $a \in A$ in an (ε, n) norm paving of x satisfies $\|a\| \leq (1 + \varepsilon)\|x\|$, so that, replacing a by $a' = (1 + \varepsilon)^{-1}a$, we have $\|a'\| \leq \|x\|$ and $\|\sum_i p_i x p_i - a'\| \leq 2\varepsilon \|x\|$.

Also note that, if there exists a normal conditional expectation E of M onto A , then the element $a \in A$ in an (ε, n) norm paving of x satisfies $\|E(x) - a\| \leq \varepsilon\|x\|$, so that $\|\sum_i p_i x p_i - E(x)\| \leq 2\varepsilon\|x\|$. In the presence of a normal conditional expectation, one often defines (ε, n) norm pavability by requiring the partition $p_1, \dots, p_n \in A$ to satisfy $\|\sum_i p_i x p_i - E(x)\| \leq \varepsilon\|x\|$.

Finally note that, if $y_1, y_2 \in M_{sa}$ are (ε, n) pivable, then $y_1 + iy_2$ is $(2\varepsilon, n^2)$ pivable. Thus, in order to obtain the paving property for $A \subset M$, it is sufficient to check pavability of self-adjoint elements in M .

We next define two weaker notions of paving, involving approximation in the so -topology rather than in norm.

Definition 2.2. An element $x \in M$ is (ε, n) *so-pavable over A* if, for every strong neighborhood \mathcal{V} of 0 in M , there exist projections $p_1, \dots, p_n \in A$, an element $a \in A$ and a projection $q \in M$ such that $\|a\| \leq \|x\|$, $\sum_{i=1}^n p_i = 1$, $\|q(\sum_i p_i x p_i - a)q\| \leq \varepsilon\|x\|$ and $1 - q \in \mathcal{V}$. We denote by $n_s(x, \varepsilon)$ the smallest such n . An element $x \in M$ is *so-pavable over A* if, for any $\varepsilon > 0$, there exists n such that x is (ε, n) *so-pavable*. We say that $A \subset M$ has the *so-paving property* if any $x \in M_{sa}$ is *so-pavable*.

It is easy to see that, if M is a finite von Neumann algebra with a faithful normal trace τ and $x \in M_{sa}$, then x is (ε, n) *so-pavable* if and only if, given any $\delta > 0$, there exist a partition of 1 with projections $p_1, \dots, p_n \in A$ and $a \in A_{sa}$, $\|a\| \leq \|x\|$, such that the spectral projection q of $\sum_i p_i x p_i - a$ corresponding to $[-\varepsilon\|x\|, \varepsilon\|x\|]$ satisfies $\tau(1 - q) \leq \delta$. As pointed out in [Popa 2014, Remark 2.4.1°], if ω is a free ultrafilter on \mathbb{N} , then $x \in M_{sa}$ has this latter property if and only if, when viewed as an element in M^ω , it is pivable over the ultrapower MASA A^ω of M^ω .

Definition 2.3. An element $x \in M$ is $(\varepsilon, n; \kappa)$ *app-pavable over A* if it can be approximated in the so -topology by a net of (ε, n) pivable elements in M bounded in norm by $\kappa\|x\|$. An element $x \in M$ is *app-pavable over A* if there exists κ_0 such that, for any $\varepsilon > 0$, there exists n such that x is $(\varepsilon, n; \kappa_0)$ *app-pavable*. We say that $A \subset M$ has the *app-paving property* if any $x \in M_{sa}$ is *app-pavable*.

Obviously, norm paving implies *so*- and *app*-paving, with $n(x, \varepsilon) \geq n_s(x, \varepsilon)$ for all x . The next result shows that, if a MASA is the range of a normal conditional expectation, then *so*- and *app*-pavability are in fact equivalent.

Proposition 2.4. *Let M be a von Neumann algebra and $A \subset M$ a MASA with the property that there exists a normal conditional expectation $E : M \rightarrow A$. Let $x \in M_{sa}$, $n \in \mathbb{N}$, $\varepsilon > 0$.*

- (1) *If x is $(\varepsilon, n; \kappa)$ app-pavable for some $\kappa \geq 1$, then x is $(2\kappa\varepsilon', n)$ so-pavable for any $\varepsilon' > \varepsilon$.*
- (2) *If x is (ε, n) so-pavable, then x is $(\varepsilon', n; 3)$ app-pavable for any $\varepsilon' > \varepsilon$.*

Proof. (1) Let $x_j \in M_{sa}$ with $\|x_j\| \leq \kappa\|x\|$ for all j be such that x_j is (ε, n) pivable for all j and x_j converges to x in the so -topology. Fix $\varepsilon' > \varepsilon$. We prove that x is $(2\kappa\varepsilon', n)$ *so-pavable*, i.e., that, given any so -neighborhood \mathcal{V} of 0, there exist a partition of 1 with projections $p_1, \dots, p_n \in A$, an element $a \in A$ and $q \in \mathcal{P}(M)$ such that $1 - q \in \mathcal{V}$ and $\|q(\sum_i p_i x p_i - a)q\| \leq 2\kappa\varepsilon'\|x\|$.

Note that, if necessary by changing the multiplicity of the representation of M on the Hilbert space \mathcal{H} , we may assume that the given neighborhood \mathcal{V} is of the form $\mathcal{V} = \{x \in M_{sa} \mid \|x\xi\| \leq \alpha\}$ for some unit vector $\xi \in \mathcal{H}$ and $\alpha > 0$.

For every j , choose a partition of 1 by projections $p_{j,1}, \dots, p_{j,n} \in A$ and an element $a_j \in A$ such that

$$\left\| \sum_{i=1}^n p_{j,i} x_j p_{j,i} - a_j \right\| \leq \varepsilon \|x_j\| \leq \kappa \varepsilon \|x\|.$$

Applying the conditional expectation E , it also follows that $\|E(x_j) - a_j\| \leq \kappa \varepsilon \|x\|$. Therefore,

$$\left\| \sum_{i=1}^n p_{j,i} (x_j - E(x_j)) p_{j,i} \right\| \leq 2\kappa \varepsilon \|x\|.$$

Define the self-adjoint elements

$$T_j = \sum_{i=1}^n p_{j,i} (x - E(x)) p_{j,i} \quad \text{and} \quad S_j = \sum_{i=1}^n p_{j,i} (x_j - E(x_j)) p_{j,i}.$$

Let $\delta = 2(\varepsilon' - \varepsilon)\kappa \|x\|$. Recall that the normal conditional expectation E is automatically faithful because its support is a projection in $A' \cap M = A$ and thus equal to 1. So, we can apply Lemma 2.5 and, since $x_j \rightarrow x$ strongly, we get that $T_j - S_j \rightarrow 0$ strongly. Thus, there exists j large enough such that $\|(T_j - S_j)\xi\| < \alpha\delta$.

We claim that, if we denote by q the spectral projection of $|T_j - S_j|$ corresponding to the interval $[0, \delta]$, then $\|(1-q)\xi\| < \alpha$, and so $1-q \in \mathcal{V}$. Indeed, if not, then $\|(1-q)\xi\| \geq \alpha$ and thus $\||T_j - S_j|(1-q)\xi\| \geq \alpha\delta$, implying that

$$\|(T_j - S_j)\xi\| \geq \||T_j - S_j|(1-q)\xi\| \geq \alpha\delta > \|(T_j - S_j)\xi\|,$$

a contradiction.

On the other hand, $a = E(x)$ satisfies $\|a\| \leq \|x\|$ and we also have the estimates

$$\left\| q \left(\sum_{i=1}^n p_{j,i} (x - E(x)) p_{j,i} \right) q \right\| = \|qT_jq\| \leq \|q(T_j - S_j)q\| + \|qS_jq\| \leq \delta + 2\kappa\varepsilon\|x\| = 2\kappa\varepsilon'\|x\|.$$

This finishes the proof of (1).

(2) Note that if $\varepsilon' \geq 2$ then there is nothing to prove. So, without any loss of generality, we may assume $0 < \varepsilon < \varepsilon' < 2$. Let $\alpha = 1 - (\varepsilon' - \varepsilon)/2$ and $\gamma = 1 - (\alpha\varepsilon' - \varepsilon)/6$. Note that $\varepsilon' < 2$ implies $\alpha\varepsilon' > \varepsilon$, so $\gamma < 1$. We clearly also have $\gamma > \alpha$.

Let $x \in M_{sa}$ be (ε, n) so -pavable. Fix an open so -neighborhood \mathcal{W} of 0 in M . We construct an (ε', n) pavable element $y \in M_{sa}$ with $\|y\| \leq 3\|x\|$ and $x - y \in \mathcal{W}$. We may assume that $x \neq 0$.

By the lower semicontinuity of the norm with respect to the so -topology, it follows that the set

$$\mathcal{W}_1 = \mathcal{W} \cap \{h \in M \mid \|x - h\| > \gamma\|x\|\}$$

is an open so -neighborhood of 0 in M . Choose an open so -neighborhood \mathcal{W}_0 of 0 such that $\mathcal{W}_0 + \mathcal{W}_0 \subset \mathcal{W}_1$.

Using Lemma 2.5 below to realize the second point, we can fix an so -neighborhood \mathcal{V}_1 of 0 such that, for every projection $q \in M$ with $1 - q \in \mathcal{V}_1$, we have that:

- $x - q x q \in \mathcal{W}_0$;
- $q a q - a \in \mathcal{W}_0$ for all $a \in A$ with $\|a\| \leq \|x\|$.

Again using Lemma 2.5 below, we can fix an so -neighborhood $\mathcal{V}_0 \subset \mathcal{V}_1$ of 0 such that, for every projection $q \in M$ with $1 - q \in \mathcal{V}_0$, we have the following property:

- For any partition of 1 with projections $p_1, \dots, p_n \in A$, the spectral projection q' of $\sum_i p_i q p_i$ corresponding to the interval $(1 - ((\alpha\varepsilon' - \varepsilon)/(6n^2))^2, 1]$ satisfies $1 - q' \in \mathcal{V}_1$.

Since x is (ε, n) so -pavable, we can choose projections $p_1, \dots, p_n \in A$, an element $a \in A$ and a projection $q \in M$ such that $\|a\| \leq \|x\|$, $\sum_{i=1}^n p_i = 1$, $\|q(\sum_i p_i x p_i - a)q\| \leq \varepsilon\|x\|$ and $1 - q \in \mathcal{V}_0$.

Let e_i be the spectral projection of $p_i q p_i$ corresponding to the interval $(1 - ((\alpha\varepsilon' - \varepsilon)/(6n^2))^2, 1]$ for each i , and let $q' = \sum_i e_i$. By the last of the above properties, we have $1 - q' \in \mathcal{V}_1$. Define $y = q'(x - a)q' + a$ and note that $\|y\| \leq \|x - a\| + \|a\| \leq 3\|x\|$. We will prove that $x - y \in \mathcal{W}$ and that y is (ε', n) pavable.

Indeed, because $1 - q' \in \mathcal{V}_1$, we have

$$x - y = (x - q' x q') + (q' a q' - a) \in \mathcal{W}_0 + \mathcal{W}_0 \subset \mathcal{W}_1.$$

So, $x - y \in \mathcal{W}$ and $\|y\| \geq \gamma\|x\|$. Since this implies $\|\gamma a\| \leq \|y\|$, in order to prove that y is (ε', n) pavable it is sufficient to prove that $\|\sum_i p_i y p_i - \gamma a\| \leq \varepsilon'\|y\|$. To see this, note first that we have

$$\sum_i p_i y p_i - \gamma a = \sum_i p_i q'(x - a)q' p_i + (1 - \gamma)a = \sum_i e_i(x - a)e_i + (1 - \gamma)a,$$

and thus

$$\left\| \sum_i p_i y p_i - \gamma a \right\| \leq \left\| \sum_i e_i(x - a)e_i \right\| + (1 - \gamma)\|x\|.$$

Since, by the definition of e_i , we have

$$\|e_i - e_i q\|^2 = \|e_i - e_i q e_i\| = \|e_i - e_i(p_i q p_i)\| \leq \left(\frac{\alpha\varepsilon' - \varepsilon}{6n^2} \right)^2,$$

it follows that $\|q' - q' q\| \leq \sum_i \|e_i - e_i q\| \leq n(\alpha\varepsilon' - \varepsilon)/(6n^2) = (\alpha\varepsilon' - \varepsilon)/(6n)$. Thus, since $e_i = q' p_i$, we get that

$$\|e_i - q' q p_i\| = \|(q' - q' q)p_i\| \leq \|q' q - q'\| \leq \frac{\alpha\varepsilon' - \varepsilon}{6n},$$

implying that

$$\begin{aligned} & \left\| \sum_i p_i y p_i - \gamma a \right\| \\ & \leq \left\| \sum_i e_i(x - a)e_i \right\| + (1 - \gamma)\|x\| \\ & \leq \sum_i \|e_i - q' q p_i\| \|x - a\| + \|q' q\| \left(\sum_i p_i x p_i - a \right) q' q' + \sum_i \|x - a\| \|e_i - p_i q q'\| + (1 - \gamma)\|x\| \end{aligned}$$

$$\begin{aligned} &\leq \frac{\alpha\varepsilon' - \varepsilon}{3} \|x - a\| + \varepsilon\|x\| + (1 - \gamma)\|x\| \leq \frac{5\alpha\varepsilon' + \varepsilon}{6} \|x\| \\ &\leq \frac{5\alpha\varepsilon' + \varepsilon}{6} \gamma^{-1} \|y\| \leq \alpha\gamma^{-1} \varepsilon' \|y\| < \varepsilon' \|y\| \end{aligned}$$

where the two last inequalities hold true because $\varepsilon < \alpha\varepsilon'$ and $\alpha\gamma^{-1} < 1$. □

In the proof of the above Proposition 2.4, we used the following elementary lemma:

Lemma 2.5. *Let $M \subset \mathcal{B}(H)$ be a von Neumann algebra and $P \subset M$ a von Neumann subalgebra. Assume that P is finite and that $E : M \rightarrow P$ is a normal faithful conditional expectation. If (x_k) is a bounded net in M that strongly converges to 0, then the nets $(x_k a)$ converge strongly to 0 uniformly over all $a \in (P)_1$:*

$$\lim_k \left(\sup_{a \in (P)_1} \|x_k a \xi\| \right) = 0 \quad \text{for every } \xi \in H.$$

Proof. Since P is finite, we can fix a normal semifinite faithful (nsf) trace Tr on P with the property that the restriction of Tr to the center $\mathcal{Z}(P)$ is still semifinite. Define the nsf weight $\varphi = \text{Tr} \circ E$ on M and the corresponding space $\mathcal{N}_\varphi = \{x \in M \mid \varphi(x^*x) < \infty\}$. We complete \mathcal{N}_φ into a Hilbert space H_φ : to every $x \in \mathcal{N}_\varphi$ corresponds a vector $\hat{x} \in H_\varphi$, and M is faithfully represented on H_φ by $\pi_\varphi(x)\hat{y} = \widehat{xy}$.

Whenever $z \in \mathcal{Z}(P)$ is a projection with $\text{Tr}(z) < \infty$, we consider the normal positive functional $\varphi_z \in M_*$ given by $\varphi_z(x) = \varphi(zxz)$. Since these φ_z form a faithful family of normal positive functionals on M , it suffices to prove that

$$\lim_k \left(\sup_{a \in (P)_1} \varphi_z(a^* x_k^* x_k a) \right) = 0 \quad \text{for all projections } z \in \mathcal{Z}(P) \text{ with } \text{Tr}(z) < \infty. \tag{2-1}$$

We denote by J_φ the modular conjugation on H_φ . Since P belongs to the centralizer of the weight φ , we have that $\widehat{xa} = J_\varphi \pi_\varphi(a)^* J_\varphi \hat{x}$ for all $x \in \mathcal{N}_\varphi$ and $a \in P$. For $z \in \mathcal{Z}(P)$ with $\text{Tr}(z) < \infty$ and $a \in P$, we then find that

$$\varphi_z(a^* x_k^* x_k a) = \|\widehat{x_k a z}\|^2 = \|J_\varphi \pi_\varphi(a)^* J_\varphi \widehat{x_k z}\|^2 \leq \|a\|^2 \varphi_z(x_k^* x_k).$$

Since $\lim_k \varphi_z(x_k^* x_k) = 0$, we get (2-1) and the lemma is proved. □

Remark 2.6. For Lemma 2.5 to hold, both the finiteness of P and the existence of the normal faithful conditional expectation $E : M \rightarrow P$ are crucial. First note that the lemma fails for the diffuse MASA in $\mathcal{B}(H)$. It suffices to take $M = \mathcal{B}(L^2(\mathbb{T}))$ and $P = L^\infty(\mathbb{T})$, with respect to the normalized Lebesgue measure on \mathbb{T} . Consider the unitary operators $a_n \in P$ given by $a_n(z) = z^n$. We can also consider the $(a_n)_{n \in \mathbb{Z}}$ as an orthonormal basis of $L^2(\mathbb{T})$ and define x_k as the orthogonal projection onto the closure of $\text{span}\{a_n \mid n \geq k\}$. Then, $x_k \rightarrow 0$ strongly. With $\xi(z) = 1$ for all $z \in \mathbb{T}$, we find that $\sup_n \|x_k a_n \xi\|_2 = 1$ for every k . So, the existence of the conditional expectation E is essential.

The previous paragraph implies in particular that the lemma fails if $M = P = \mathcal{B}(H)$. So, also, the finiteness of P is essential.

We will now relate *so-* and *app-*pavability properties for a MASA $A \subset M$ having a normal conditional expectation $E_A : M \rightarrow A$, with the norm-pavability for the associated inclusion of ultrapower

algebras $A^\omega \subset M^\omega$. We will only consider the case when A is countably decomposable, i.e., when there exists a normal faithful state φ on A . We still denote by φ its extension to M given by $\varphi \circ E_A$.

For the reader's convenience, we recall Ocneanu's [1985] definition of the ultrapower of a von Neumann algebra. Given a free ultrafilter ω on \mathbb{N} , one lets I_ω be the C^* -algebra of all bounded sequences $(x_n)_n \in \ell^\infty(\mathbb{N}, M)$ that are s^* -convergent to 0 along the ultrafilter ω . One denotes by $M^{0,\omega}$ the multiplier (also called the binormalizer) of I_ω in $\ell^\infty(\mathbb{N}, M)$ (which is easily seen to be a C^* -algebra) and one defines M^ω to be the quotient $M^{0,\omega}/I_\omega$. This is shown in [Ocneanu 1985] to be a von Neumann algebra, called the ω -ultrapower of M . Since the constant sequences are in the multiplier $M^{0,\omega}$, we have a natural embedding $M \subset M^\omega$. It is easy to see that, if M is an atomic von Neumann algebra, then $M^\omega = M$; in particular, $\mathfrak{B}(\ell^2\mathbb{N})^\omega = \mathfrak{B}(\ell^2\mathbb{N})$.

To define the ultrapower MASA $A^\omega \subset M^\omega$, one proceeds as in [Popa 1995, Section 1.3]. One lets $E_A^{0,\omega} : \ell^\infty(\mathbb{N}, M) \rightarrow \ell^\infty(\mathbb{N}, A)$ be the conditional expectation defined by $E_A^{0,\omega}((x_n)_n) = (E_A(x_n))_n$. One notices that $E_A^{0,\omega}(I_\omega) = I_\omega \cap \ell^\infty(\mathbb{N}, A) = \{(a_n) \in \ell^\infty(\mathbb{N}, A) \mid \lim_\omega \varphi(a_n^* a_n) = 0\}$ and that $\ell^\infty(\mathbb{N}, A) \subset M^{0,\omega}$. Finally, one defines $A^\omega = (\ell^\infty(\mathbb{N}, A) + I_\omega) / I_\omega \simeq \ell^\infty(\mathbb{N}, A) / I_\omega \cap \ell^\infty(\mathbb{N}, A)$. It follows that A^ω , defined this way, is a von Neumann subalgebra of M^ω , with $E_A^{0,\omega}$ implementing a normal conditional expectation E_{A^ω} that sends the class of $(x_n)_n$ to the class of $(E_A(x_n))_n$. Moreover, by [Popa 1995, Theorem A.1.2], it follows that A^ω is a MASA in M^ω . Note also that E_{A^ω} coincides with E_A when restricted to constant sequences in $M \subset M^\omega$. From the above remark, the ultrapower of $\mathfrak{D} \subset \mathfrak{B}(\ell^2\mathbb{N})$ coincides with $\mathfrak{D} \subset \mathfrak{B}(\ell^2\mathbb{N})$ itself.

Theorem 2.7. *Let M be a von Neumann algebra and $A \subset M$ a MASA with the property that there exists a normal conditional expectation $E_A : M \rightarrow A$. Let ω be a free ultrafilter on \mathbb{N} and denote by $A^\omega \subset M^\omega$ the corresponding ultrapower inclusion.*

- (1) *An element $x \in M_{sa}$ is so-pavable over A if and only if x is app-pavable over A . So, $A \subset M$ has the so-paving property if and only if it has the app-paving property.*
- (2) *Assume that A is countably decomposable. Then $x \in M_{sa}$ is so-pavable over A if and only if x is norm pavable over A^ω . More precisely, if $x \in M_{sa}$ is (ε, n) so-pavable, then x is (ε, n) norm pavable over A^ω ; conversely, if $x \in M_{sa}$ is (ε, n) norm pavable over A^ω , then x is (ε', n) so-pavable for all $\varepsilon' > \varepsilon$.*
- (3) *Still assume that A is countably decomposable. Then the uniform so-paving property of $A \subset M$ is equivalent to the uniform paving property of $A^\omega \subset M^\omega$. More precisely, if every $x \in M_{sa}$ is (ε, n) so-pavable, then every $x \in M_{sa}^\omega$ is (ε, n) norm pavable.*

Proof. (1) follows immediately from Proposition 2.4.

To prove (2) and (3), we assume that A is countably decomposable and it suffices to prove the following two statements for given $0 < \varepsilon < \varepsilon'$ and $n \in \mathbb{N}$:

- If $x \in M_{sa}^\omega$ is represented by the sequence $(x_m) \in M^{0,\omega}$ of self-adjoint elements $x_m \in M_{sa}$ satisfying $\|x_m\| \leq \|x\|$, and if every x_m is (ε, n) so-pavable, then x is (ε, n) norm pavable over A^ω .
- If $x \in M_{sa}$ is (ε, n) norm pavable over A^ω , then x is (ε', n) so-pavable.

Since A is countably decomposable, we can fix a normal faithful state φ on A and still denote by φ its extension $\varphi \circ E_A$ to M . Note that the s^* -topology on the unit ball of M_{sa} coincides with the so -topology, both being implemented by the norm $\|\cdot\|_\varphi$.

We start by proving the first of the two statements above. For every m , the self-adjoint element x_m is (ε, n) so -pavable. So we can take a partition of 1 with projections $p_1^m, \dots, p_n^m \in A$, a projection $q_m \in M$ and an element $a_m \in A$ such that $\|a_m\| \leq \|x_m\| \leq \|x\|$ and such that $\|q_m(\sum_i p_i^m x p_i^m - a_m)q_m\| \leq \varepsilon\|x\|$ and $\varphi(1 - q_m) \leq 2^{-m}$. Since (x_m) and $\ell^\infty(\mathbb{N}, A)$ are both contained in $M^{0,\omega}$, the sequences $((1 - q_m)p_i^m(x_m - a_m)p_i^m)_m$ and $(p_i^m(x_m - a_m)p_i^m(1 - q_m))_m$ belong to I_ω .

Thus, if we let $a = (a_m)$ and $p_i = (p_i^m)_m \in A^\omega$, $1 \leq i \leq n$, then p_1, \dots, p_n is a partition of 1 with projections in A^ω and $p_i(x - a)p_i$ coincides with $(q_m p_i^m(x_m - a_m)p_i^m q_m)_m$ in M^ω . It follows that $\sum_i p_i(x - a)p_i$ coincides with $(q_m \sum_i p_i^m(x_m - a_m)p_i^m q_m)_m$ in M^ω , and thus has norm majorized by $\varepsilon\|x\|$. So we have proved that x is (ε, n) norm pavable over A^ω .

To prove the second of the two statements above, let $x \in M_{\text{sa}}$ be (ε, n) norm pavable over A^ω (as an element in M^ω). Let $\delta > 0$ be arbitrary. We have to prove that there exists an $a' \in A$ with $\|a'\| \leq \|x\|$, a partition of 1 with projections $e_1, \dots, e_n \in A$ and a projection $q \in M$ such that $\varphi(1 - q) \leq \delta$ and $\|q \sum_i e_i(x - a')e_i q\| \leq \varepsilon'\|x\|$.

Take projections $p_1, \dots, p_n \in A^\omega$ and $a \in A_{\text{sa}}^\omega$ such that $\|a\| \leq \|x\|$, $\sum_i p_i = 1$ and $\|\sum_i p_i x p_i - a\| \leq \varepsilon\|x\|$. Represent the p_i by sequences $(p_i^m)_m$ with projections $p_i^m \in A$ such that $\sum_i p_i^m = 1$ for all m , and represent a by a sequence $(a_m)_m$ with $a_m \in A_{\text{sa}}$ and $\|a_m\| \leq \|a\|$ for all m .

We conclude that there exists a sequence of self-adjoint elements $(y_m)_m \in I_\omega$ of norm at most $3\|x\|$ such that the sequence $(b_m)_m = (\sum_i p_i^m(x - a_m)p_i^m - y_m)_m$ satisfies $\|b_m\| \leq \varepsilon\|x\|$ for all m . Since $(y_m)_m \in I_\omega$, we have $\lim_\omega \varphi(|y_m|) = 0$, so that there exists a neighborhood \mathcal{V} of ω such that the spectral projection q_m of $|y_m|$ corresponding to $[0, (\varepsilon' - \varepsilon)\|x\|]$ satisfies $\varphi(1 - q_m) \leq \delta$ for any $m \in \mathcal{V}$. Thus, for any such m , if we let $a' = a_m$, $e_i = p_i^m$ and $q = q_m$, then we have

$$\left\| q \sum_i e_i(x - a')e_i q \right\| \leq \|q_m b_m q_m\| + \|q_m y_m q_m\| \leq \varepsilon\|x\| + (\varepsilon' - \varepsilon)\|x\| \leq \varepsilon'\|x\|. \quad \square$$

Conjecture 2.8. (1) Any MASA in a von Neumann algebra, $A \subset M$, with the property that there exists a normal conditional expectation of M onto A has the so -paving property (equivalently the app -paving property). Also, while the equivalence between so - and app -pavability for an arbitrary MASA A in a von Neumann algebra M is still to be clarified, any MASA $A \subset M$ (not necessarily the range of a normal expectation) ought to satisfy both these properties.

(2) Going even further, we expect that the paving size satisfies the estimate $n_s(x, \varepsilon) \leq C\varepsilon^{-2}$ for all $x \in M_{\text{sa}}$ for some universal constant $C > 0$, independent of $A \subset M$.

Remark 2.9. (i) There is much evidence for 1° in the above conjecture. By Theorem 2.7(3) and the fact that the ultrapower of $\mathcal{D} \subset \mathcal{B}(\ell^2\mathbb{N})$ coincides with $\mathcal{D} \subset \mathcal{B}(\ell^2\mathbb{N})$, so -pavability for this inclusion is equivalent to Kadison–Singer paving, proved to hold true in [Marcus et al. 2015]. It was already noticed in [Popa 2014] that so -pavability over the Cartan MASA of the hyperfinite II_1 factor $D \subset R$ is equivalent to pavability of $\mathcal{D} \subset \mathcal{B}(\ell^2\mathbb{N})$, and thus holds true by [Marcus et al. 2015]. In fact, more cases of the

conjecture can be deduced from [Marcus et al. 2015]. Thus, we note in Section 3 that any MASA in a type I von Neumann algebra (such as a diffuse MASA in $\mathcal{B}(\ell^2\mathbb{N})$) satisfy both *so*- and *app*-pavability. Then in Section 4, we use [Marcus et al. 2015] to prove that any Cartan MASA in an amenable von Neumann algebra, or in a group measure space II_1 factor arising from a free ergodic profinite action, has the *so*-pavability property. On the other hand, the conjecture had already been checked for singular MASAs in II_1 factors in [Popa 2014], and Cyril Houdayer and Yusuke Isono pointed out that, modulo some obvious modifications, the proof in [Popa 2014] works as well for any singular MASA A in an arbitrary von Neumann algebra M , once A is the range of normal conditional expectation from M . Finally, in Remark 5.3, we prove that *so*-pavability also holds for a certain class of MASAs that are neither Cartan, nor singular.

(ii) The estimate on the paving size $n_s(x, \varepsilon) \sim \varepsilon^{-2}$ for all $x \in M_{\text{sa}}$ in point (2) of the conjecture is more speculative, and there is less evidence for it. Based on results in [Popa 2014], we will show in Theorem 5.1 that this estimate does hold true for singular MASAs. We will also show in Proposition 5.4 that this is the best one can expect for the *so*-paving size of any MASA in a II_1 factor, and thus, since $n_s(D \subset R, \varepsilon) = n(\mathcal{D} \subset \mathcal{B}(\ell^2\mathbb{N}), \varepsilon)$, the best one can expect for the paving size in the Kadison–Singer problem as well (a fact already shown in [Casazza et al. 2007]). For the inclusions $\mathcal{D} \subset \mathcal{B}(\ell^2\mathbb{N})$, the order of magnitude of the ε pavings obtained in [Marcus et al. 2015] is $C\varepsilon^{-4}$, but the techniques used there seem to allow obtaining the paving size $C\varepsilon^{-2}$. However, in order to prove Conjecture 2.8 in its full generality, in particular unifying the singular and the Cartan MASA cases (including the diagonal inclusions $D_k \subset \mathcal{B}(\ell_k^2)$, $2 \leq k \leq \infty$), which are quite different in nature, a new idea may be needed.

(iii) The (ε, n) *so*-paving in the case of a MASA $A \subset M$ with a normal conditional expectation $E_A : M \rightarrow A$ and a normal faithful state φ on M with $\varphi \circ E_A = \varphi$ should be compared with (ε, n) L^2 -paving in the Hilbert norm $\|\cdot\|_\varphi$, which, for $x \in M$, $E_A(x) = 0$, requires the existence of a partition of 1 with projections $p_1, \dots, p_n \in A$ such that $\|\sum_i p_i x p_i\|_\varphi \leq \varepsilon \|x\|_\varphi$. This condition is obviously weaker than *so*-paving, with $n(x, \varepsilon) \geq n_s(x, \varepsilon)$ bounded from below by the L^2 -paving size of x for all $x \in M_{\text{sa}}$. It was shown in [Popa 2014, Theorem 3.9] to always occur, with paving size majorised by ε^{-2} (in fact the proof in [Popa 2014] is for MASAs in II_1 factors, but the same proof works in the general case; see also [Popa 1995, Theorem A.1.2] in this respect). The proof of Proposition 5.4 at the end of this paper shows that the paving size is bounded from below by ε^{-2} for all MASAs in II_1 factors.

3. Paving over MASAs in type I von Neumann algebras

Marcus et al. [2015] proved that, for every self-adjoint matrix $T \in M_k(\mathbb{C})$ with zeros on the diagonal and for every $\varepsilon > 0$, there exist r projections $p_1, \dots, p_r \in D_k(\mathbb{C})$ with $r \leq (6/\varepsilon)^4$, $\sum_{i=1}^r p_i = 1$ and $\|p_i T p_i\| \leq \varepsilon \|T\|$ for all i (see also [Tao 2013; Valette 2015] for alternative presentations of the proof). Thus, if \mathcal{D} is the diagonal MASA in $\mathcal{B} = \mathcal{B}(\ell^2\mathbb{N})$, then $\mathcal{D} \subset \mathcal{B}$ has the paving property, with $n(\mathcal{D} \subset \mathcal{B}; x, \varepsilon) \leq 12^4 \varepsilon^{-4}$ for all $x = x^* \in \mathcal{B}$.

In this section, we deduce from this that any MASA A in a type I von Neumann algebra M has the *so*- and *app*-paving property.

We also prove that a MASA A in a von Neumann algebra M with separable predual has the norm paving property if and only if M is of type I and there exists a normal conditional expectation of M onto A .

We start by deducing the following lemma from [Marcus et al. 2015]:

Lemma 3.1. *Let (X, μ) be a standard probability space and $\mathcal{B} = M_k(\mathbb{C})$ or $\mathcal{B} = \mathcal{B}(\ell^2\mathbb{N})$ with the diagonal MASA $\mathcal{D} \subset \mathcal{B}$. Consider the unique normal conditional expectation E of $\mathcal{B} \bar{\otimes} L^\infty(X)$ onto $\mathcal{D} \bar{\otimes} L^\infty(X)$. If $T \in \mathcal{B} \bar{\otimes} L^\infty(X)$ is a self-adjoint element with $E(T) = 0$ and if $\varepsilon > 0$, there exist r projections $p_1, \dots, p_r \in \mathcal{D} \bar{\otimes} L^\infty(X)$ with $r \leq (6/\varepsilon)^4$, $\sum_{i=1}^r p_i = 1$ and $\|p_i T p_i\| \leq \varepsilon \|T\|$ for all i .*

Proof. It suffices to consider $\mathcal{B} = \mathcal{B}(\ell^2\mathbb{N})$. Fix a self-adjoint $T \in \mathcal{B} \bar{\otimes} L^\infty(X)$ with $E(T) = 0$ and $\varepsilon > 0$. Denote by r the largest integer satisfying $r \leq (6/\varepsilon)^4$. We represent T as a Borel function $T : X \rightarrow \mathcal{B}$ satisfying $\|T(x)\| \leq \|T\|$ and $E(T(x)) = 0$ for all $x \in X$. Define Y as the compact Polish space $Y := \{1, \dots, r\}^\mathbb{N}$. For every $y \in Y$ and $i \in \{1, \dots, r\}$, we denote by $p_i^y \in \mathcal{D}$ the projection given by $p_i^y(k) = 1$ if $y(k) = i$ and $p_i^y(k) = 0$ if $y(k) \neq i$. Clearly, the projections p_1^y, \dots, p_r^y with $y \in Y$ describe precisely all partitions of \mathcal{D} . Also, for every $i \in \{1, \dots, r\}$, the map $y \mapsto p_i^y$ is strongly continuous.

Define the Borel map

$$\mathcal{V} : Y \times X \rightarrow [0, +\infty), \quad \mathcal{V}(y, x) = \max_{i=1, \dots, r} \|p_i^y T(x) p_i^y\|$$

and the Borel set $Z \subset Y \times X$ given by $Z := \{(y, x) \in Y \times X \mid \mathcal{V}(y, x) \leq \varepsilon \|T\|\}$. For every $x \in X$, we have that $T(x) \in \mathcal{B}$ with $\|T(x)\| \leq \|T\|$ and $E(T(x)) = 0$. So, by [Marcus et al. 2015], for every $x \in X$ there exists a $y \in Y$ such that $(y, x) \in Z$. Defining $\pi : Y \times X \rightarrow X$ by $\pi(y, x) = x$, this means that $\pi(Z) = X$. By von Neumann’s measurable selection theorem [1949] (or see [Kechris 1995, Theorem 18.1]), we can take a Borel set $X_0 \subset X$ and a Borel function $F : X_0 \rightarrow Y$ such that $\mu(X \setminus X_0) = 0$ and $(F(x), x) \in Z$ for all $x \in X_0$.

The Borel functions $p_i : X_0 \rightarrow \mathcal{D}$, $p_i(x) = p_i^{F(x)}$, then define a partition p_1, \dots, p_r of $\mathcal{D} \bar{\otimes} L^\infty(X)$ with the property that $\|p_i T p_i\| \leq \varepsilon \|T\|$ for all i . □

Proposition 3.2. *Let M be a von Neumann algebra of type I with separable predual and $A \subset M$ an arbitrary MASA. Then $A \subset M$ has both the so- and the app-paving properties.*

More precisely, for every $x \in M_{sa}$ and $\varepsilon > 0$, we have that $n_s(x, \varepsilon) \leq 12^4 \varepsilon^{-4}$. Also, there exists a strongly dense $$ -subalgebra $M_0 \subset M$ with $A \subset M_0$ such that, for every $x \in (M_0)_{sa}$ and $\varepsilon > 0$, we have that $n(x, \varepsilon) \leq 12^4 \varepsilon^{-4}$.*

Proof. Fix an arbitrary MASA $A \subset M$. There exist standard probability spaces $(X_k, \mu_k)_{k \in \mathbb{N}}$ and (X_d, μ_d) , (X_c, μ_c) such that, writing $A_k = L^\infty(X_k)$, and A_d, A_c similarly, the MASA $A \subset M$ is isomorphic to a direct sum of MASAs of the form

$$\begin{aligned} D_k(\mathbb{C}) \otimes A_k &\subset M_k(\mathbb{C}) \otimes A_k, \\ \ell^\infty(\mathbb{N}) \bar{\otimes} A_d &\subset \mathcal{B}(\ell^2(\mathbb{N})) \bar{\otimes} A_d, \\ \text{and } L^\infty([0, 1]) \bar{\otimes} A_c &\subset \mathcal{B}(L^2([0, 1])) \bar{\otimes} A_c. \end{aligned} \tag{3-1}$$

For the first two of these MASAs, by Lemma 3.1, we get that $n(x, \varepsilon) \leq 12^4 \varepsilon^{-4}$ for every self-adjoint element x .

For the rest of the proof, we consider $M = \mathcal{B}(L^2([0, 1])) \bar{\otimes} L^\infty(X)$ and $A = L^\infty([0, 1]) \bar{\otimes} L^\infty(X)$ for some standard probability space (X, μ) . Fix $x \in M_{sa}$ and $\varepsilon > 0$. Let n be the largest integer satisfying $n \leq 12^4 \varepsilon^{-4}$. We prove that x is (ε, n) so-pavable. Choose an so-neighborhood \mathcal{V} of 0 in M . For every $r > 0$, denote by $q_r \in \mathcal{B}(L^2([0, 1]))$ the orthogonal projection on to the subspace $H_r \subset L^2([0, 1])$ defined as

$$H_r = \{\xi \in L^2([0, 1]) \mid \xi \text{ is constant on every interval } [r^{-1}(i-1), r^{-1}i] \text{ for } i = 1, \dots, r\}.$$

Define $\xi_{r,i} = \sqrt{r} \chi_{[r^{-1}(i-1), r^{-1}i]}$, so that $(\xi_{r,i})_{i=1, \dots, r}$ is an orthonormal basis of H_r .

When $r \rightarrow \infty$, we have that $q_r \rightarrow 1$ strongly. So we can fix r large enough such that $1 - (q_r \otimes 1) \in \mathcal{V}$. Denote by $e_i \in L^\infty([0, 1])$ the projection $e_i = \chi_{[r^{-1}(i-1), r^{-1}i]}$. Define the vector functionals ω_{ij} in $\mathcal{B}(L^2([0, 1]))_*$ by $\omega_{ij}(T) = \langle T \xi_{r,i}, \xi_{r,j} \rangle$. Define $a \in A$ by

$$a = \sum_{i=1}^r e_i \otimes (\omega_{ii} \otimes \text{id})(x).$$

By construction, $\|a\| \leq \|x\|$.

Define the isometry $V \in \mathcal{B}(C^r, L^2([0, 1]))$ by $V(\delta_i) = \xi_{r,i}$ for $i = 1, \dots, r$. Define $y \in M_r(\mathbb{C}) \otimes L^\infty(X)$ by $y := (V^* \otimes 1)x(V \otimes 1)$. We also put $b = (V^* \otimes 1)a(V \otimes 1)$. Denoting the natural conditional expectation by $E : M_r(\mathbb{C}) \otimes L^\infty(X) \rightarrow D_r(\mathbb{C}) \otimes L^\infty(X)$, we have $E(y) = b$. By Lemma 3.1, we thus find projections $f_1, \dots, f_n \in D_r(\mathbb{C}) \otimes L^\infty(X)$ such that $f_1 + \dots + f_n = 1$ and $\|f_k(y - b)f_k\| \leq \varepsilon \|y\| \leq \varepsilon \|x\|$ for all $k = 1, \dots, n$.

Define the projections $a_{ki} \in L^\infty(X)$ such that $f_k = \sum_{i=1}^r E_{ii} \otimes a_{ki}$. Then, let $p_k \in A$ be the projections given by $p_k = \sum_{i=1}^r e_i \otimes a_{ki}$. By construction, we have

$$(V^* \otimes 1)p_k x p_k (V \otimes 1) = f_k y f_k \quad \text{for all } k = 1, \dots, n.$$

Therefore,

$$\left\| (q_r \otimes 1) \left(\sum_{k=1}^n p_k x p_k - a \right) (q_r \otimes 1) \right\| = \left\| \sum_{k=1}^n (V^* \otimes 1) p_k x p_k (V \otimes 1) - b \right\| = \left\| \sum_{k=1}^n f_k y f_k - b \right\| \leq \varepsilon \|x\|.$$

Since $1 - (q_r \otimes 1) \in \mathcal{V}$, we have shown that x is (ε, n) so-pavable.

For the final part of the proof, for notational convenience, we replace the interval $[0, 1]$ by the circle \mathbb{T} . We define $M_0 \subset \mathcal{B}(L^2(\mathbb{T}))$ as the $*$ -algebra generated by $L^\infty(\mathbb{T})$ and the periodic rotation unitaries. By construction, $M_0 \subset M$ is a dense $*$ -subalgebra containing A . By Lemma 3.1, every $x \in (M_0)_{sa}$ is $(\varepsilon, 12^4 \varepsilon^{-4})$ pavable for all $\varepsilon > 0$. □

We finally prove that for a MASA A in a von Neumann algebra M with separable predual, the classical Kadison–Singer paving holds if and only if M is of type I and A is the range of a normal conditional expectation.

Theorem 3.3. *Let M be a von Neumann algebra with separable predual and $A \subset M$ a MASA. Then, $A \subset M$ satisfies the norm paving property if and only if M is of type I and A is the range of a normal conditional expectation.*

Also, unless M is of type I and A is the range of a normal conditional expectation, there exist singular conditional expectations of M onto A .

Proof. If M is of type I and A is the range of a normal conditional expectation, then $A \subset M$ is isomorphic to a direct sum of the first two types of MASAs given by (3-1). It then follows from Lemma 3.1 that $A \subset M$ satisfies the norm paving property.

Conversely, assume that $A \subset M$ satisfies the norm paving property. Then there is a unique conditional expectation $E : M \rightarrow A$. By [Akemann and Sherman 2012, Corollary 3.3], this unique conditional expectation E is normal.

Decomposing M as a direct sum of von Neumann algebras of different types, it remains to prove the following: if M has a separable predual and is of type II, type III₁ or type III without a type III₁ direct summand, and if $A \subset M$ is a MASA that is the range of a normal conditional expectation $E : M \rightarrow A$, then there also exists a singular conditional expectation of M onto A . When M is of type II, the existence of a normal conditional expectation of M onto A implies that A is generated by finite projections. By reducing with a projection in A , we may thus assume that M is of type II₁, and, in this case, singular conditional expectations were constructed in [Popa 2014, Remark 2.4.3°] (see also [Popa 1999, Proof of Corollary 4.1.(iii) and Remark 4.3.3°]).

To settle the type III cases, fix a normal faithful state φ on M satisfying $\varphi = \varphi \circ E$. First assume that M is of type III₁ and fix $n \in \mathbb{N}$. We prove that there exist matrix units $\{e_{ij} \mid 1 \leq i, j \leq 2^n\}$ in M such that $\|[\varphi, e_{ij}]\| \leq 8^{-n}$ for all i, j . To prove this statement, we use the following nonfactorial version of the Connes–Størmer transitivity theorem [1978, Theorem 4]: if φ and ρ are normal positive functionals on a type III₁ von Neumann algebra M with separable predual and if $\varphi(a) = \rho(a)$ for all $a \in \mathcal{K}(M)$, then, for every $\varepsilon > 0$, there exists a unitary $u \in M$ such that $\|\varphi - u\rho u^*\| < \varepsilon$.

Since A is diffuse relative to $\mathcal{K}(M) \subset A$, we can choose a partition $e_{ii}, i = 1, \dots, 2^n$, of A satisfying $\varphi(ae_{ii}) = 2^{-n}\varphi(a)$ for all $a \in \mathcal{K}(M)$ and $i = 1, \dots, 2^n$. In particular, the projections e_{ii} have central support 1 and are thus equivalent in M . Put $v_1 = e_{11}$ and choose partial isometries $v_i, i = 2, \dots, 2^n$, such that $v_i v_i^* = e_{11}$ and $v_i^* v_i = e_{ii}$ for all i . Define the positive functionals ψ_i on $e_{11} M e_{11}$ given by $\psi_i(x) = \varphi(v_i^* x v_i)$. Whenever $z \in \mathcal{K}(e_{11} M e_{11})$, write $z = a e_{11}$ with $a \in \mathcal{K}(M)$, so that

$$\psi_i(z) = \varphi(v_i^* a v_i) = \varphi(a v_i^* v_i) = \varphi(a e_{ii}) = 2^{-n} \varphi(a) = \varphi(a e_{11}) = \psi_1(z).$$

By the Connes–Størmer transitivity theorem, we can take unitaries $u_i \in e_{11} M e_{11}$ such that $\|\psi_1 - u_i \psi_i u_i^*\| \leq 8^{-n-1}$ for all i . Replacing v_i by $u_i v_i$, this means that we may assume that $\|\psi_1 - \psi_i\| \leq 8^{-n-1}$ for all i . Define the matrix units $e_{ij} = v_i^* v_j$. Since $\varphi = \varphi \circ E$, we know that $[\varphi, e_{ii}] = 0$ for all i . We then find that $\|[\varphi, e_{ij}]\| \leq 8^{-n}$ for all i, j .

We now proceed as in [Popa 2014, Remark 2.4.3°]. Define the projection $p_n = 2^{-n} \sum_{i,j} e_{ij}$. Since all e_{ii} belong to A , we get that $E(e_{ij}) = \delta_{i,j} e_{ii}$ and thus $E(p_n) = 2^{-n} 1$. Since $\|[\varphi, e_{ij}]\| \leq 8^{-n}$ for all i, j , we also have $\|[\varphi, p_n]\| \leq 4^{-n}$. Define the normal states φ_n on M given by $\varphi_n(x) = 2^n \varphi(p_n x p_n)$, $x \in M$.

Also define the normal functionals η_n on M given by $\eta_n(x) = 2^n \varphi(xp_n)$. Note that $\|\varphi_n - \eta_n\| \leq 2^{-n}$ and that $\eta_n(a) = \varphi(a)$ for all $a \in A$. So, if ψ denotes a weak* limit point of the sequence φ_n in M^* , it follows that ψ is a state on M satisfying $\psi(a) = \varphi(a)$ for all $a \in A$. Defining the projection $q_n = \bigvee_{k=n+1}^\infty p_k$, we get that $\varphi(q_n) \leq 2^{-n}$ and thus $q_n \rightarrow 0$ strongly. By construction, $\psi(1 - q_n) = 0$ for every n . Therefore, ψ is a singular state. Then, averaging ψ by a countable subgroup $\mathcal{U}_0 \subset \mathcal{U}(A)$ with the property that $\mathcal{U}'_0 = A$, we get, as in the proof of [Popa 1999, Corollary 4.1.(iii)], a singular state ψ_0 on M that is A -central and whose restriction to A equals φ . Then $\psi_0 = \varphi \circ \mathcal{E}$, where $\mathcal{E} : M \rightarrow A$ is a singular conditional expectation (see, e.g., [de Korvin 1971]).

Finally, assume that M is of type III but without a direct summand of type III₁. We prove that there exists an intermediate von Neumann algebra $A \subset P \subset M$ such that P is of type II and P is the range of a normal conditional expectation $M \rightarrow P$. (We are grateful to Masamichi Takesaki for useful discussions on the discrete decomposition involved in this part of the proof.) The first part of the proof then shows the existence of singular conditional expectations $P \rightarrow A$, which, composed with the normal expectation of M onto P , provides singular conditional expectations $M \rightarrow A$.

The intermediate type II von Neumann algebra $A \subset P \subset M$ can be constructed using the discrete decomposition for von Neumann algebras of type III _{λ} , $\lambda \in [0, 1)$ (see [Takesaki 2003, Theorems XII.2.1 and XII.3.7]). To avoid the measure-theoretic complications of a direct integral decomposition of M , we use the following “global” discrete decomposition. Denote by $(\sigma_t)_{t \in \mathbb{R}}$ the modular automorphism group of φ and by $N = M \rtimes_\sigma \mathbb{R}$ the continuous core of M (see [Takesaki 2003, Theorem XII.1.1]). Denote by $(\theta_t)_{t \in \mathbb{R}}$ the dual action of \mathbb{R} on N . Write $\mathcal{X}(N) = L^\infty(Z, \mu)$, where (Z, μ) is a standard probability space. Note that θ restricts to a nonsingular action of \mathbb{R} on (Z, μ) . The assumption that M has no direct summand of type III₁ is reflected by the possibility of choosing Z in such a way that no $x \in Z$ is stabilized by all $t \in \mathbb{R}$. This means that the flow $\mathbb{R} \curvearrowright (Z, \eta)$ can be built as a flow under a ceiling function (i.e., a nonergodic version of [Takesaki 2003, Theorem XII.3.2]). More concretely, we find a nonsingular action of $\mathbb{Z} \times \mathbb{R}$ on a standard probability space Ω with the following properties:

- The actions of \mathbb{Z} and \mathbb{R} on Ω are separately free and proper, that is, $\mathbb{Z} \curvearrowright \Omega$ is conjugate with $\mathbb{Z} \curvearrowright \Omega_0 \times \mathbb{Z}$ given by $n \cdot (x, m) = (x, n + m)$, and $\mathbb{R} \curvearrowright \Omega$ is conjugate with $\mathbb{R} \curvearrowright \Omega_1 \times \mathbb{R}$ given by $t \cdot (y, s) = (y, t + s)$.
- The action $\mathbb{R} \curvearrowright Z$ is conjugate with $\mathbb{R} \curvearrowright \Omega/\mathbb{Z}$. So, we can identify $\Omega_0 = Z$ and thus $\Omega = Z \times \mathbb{Z}$ with the action $\mathbb{R} \curvearrowright \Omega$ given by $t \cdot (x, n) = (t \cdot x, \omega(t, x) + n)$, where $\omega : \mathbb{R} \times Z \rightarrow \mathbb{Z}$ is a 1-cocycle.

Since $L^\infty(Z) = \mathcal{X}(N)$, the 1-cocycle ω gives rise to a natural action $\mathbb{R} \curvearrowright N \bar{\otimes} \ell^\infty(\mathbb{Z})$. We define $\mathcal{N} := (N \bar{\otimes} \ell^\infty(\mathbb{Z})) \rtimes \mathbb{R}$ and consider the action $\mathbb{Z} \curvearrowright \mathcal{N}$ given by translation on $\ell^\infty(\mathbb{Z})$ and the identity on N and $L(\mathbb{R})$. As in [Takesaki 2003, Lemma XII.3.5], it follows that \mathcal{N} is of type II and that $\mathcal{N} \rtimes \mathbb{Z}$ is naturally isomorphic with $M \bar{\otimes} \mathcal{B}(L^2(\mathbb{R})) \bar{\otimes} \mathcal{B}(l^2(\mathbb{Z}))$.

Since $\varphi = \varphi \circ E$, we get that every $a \in A$ belongs to the centralizer of φ . We can then view $A \bar{\otimes} L(\mathbb{R})$ as a MASA of $N = M \rtimes_\sigma \mathbb{R}$. Also $\mathcal{X}(N) \subset A \bar{\otimes} L(\mathbb{R})$. So, the above action $\mathbb{R} \curvearrowright N \bar{\otimes} \ell^\infty(\mathbb{Z})$ globally preserves $A \bar{\otimes} L(\mathbb{R}) \bar{\otimes} \ell^\infty(\mathbb{Z})$. We can then define $\mathcal{A} := (A \bar{\otimes} L(\mathbb{R}) \bar{\otimes} \ell^\infty(\mathbb{Z})) \rtimes \mathbb{R}$ as a von Neumann subalgebra of \mathcal{N} .

The dual action $\mathbb{R} \curvearrowright L(\mathbb{R})$ is conjugate with the translation action $\mathbb{R} \curvearrowright L^\infty(\mathbb{R})$. Therefore, the 1-cocycle ω trivializes on $A \bar{\otimes} L(\mathbb{R})$. This yields the natural surjective $*$ -isomorphism

$$\Psi : A \bar{\otimes} \mathcal{B}(L^2(\mathbb{R})) \bar{\otimes} \ell^\infty(\mathbb{Z}) \rightarrow \mathcal{A}.$$

Choose a minimal projection $q \in \mathcal{B}(L^2(\mathbb{R})) \bar{\otimes} \ell^\infty(\mathbb{Z})$ and put $p = \Psi(1 \otimes q)$. We then get that $A \subset p\mathcal{N}p \subset p(\mathcal{N} \rtimes \mathbb{Z})p$. Using the natural isomorphism of $\mathcal{N} \rtimes \mathbb{Z}$ with $M \bar{\otimes} \mathcal{B}(L^2(\mathbb{R})) \bar{\otimes} \mathcal{B}(\ell^2(\mathbb{Z}))$, we can identify $p(\mathcal{N} \rtimes \mathbb{Z})p = M$ and have found $p\mathcal{N}p$ as an intermediate type II von Neumann algebra sitting between A and M . Because there is a natural normal conditional expectation of $\mathcal{N} \rtimes \mathbb{Z}$ onto \mathcal{N} , we also have a normal conditional expectation of M onto $p\mathcal{N}p$. \square

4. Paving over Cartan subalgebras

The paving property for the diagonal MASA $\mathcal{D} \subset \mathcal{B}(\ell^2\mathbb{N})$ was shown in [Popa 2014] to be equivalent to the paving property for the ultrapower inclusion $D^\omega \subset R^\omega$, where D is the Cartan MASA in the hyperfinite II_1 factor R . As we have seen in Theorem 2.7, this is equivalent, in turn, to the (uniform) *so*-paving property for $D \subset R$. Thus, [Marcus et al. 2015] implies that *so*-paving holds true for $D \subset R$. We will now use [Marcus et al. 2015] to prove that, in fact, *so*-paving holds true for any Cartan subalgebra of an amenable von Neumann algebra as well as for Cartan inclusions arising from a free ergodic profinite probability measure-preserving (pmp) action of a countable group, $\Gamma \curvearrowright X$, i.e., $A = L^\infty(X) \subset L^\infty(X) \rtimes \Gamma = M$.

Theorem 4.1. (1) *If M is an amenable von Neumann algebra and $A \subset M$ is a Cartan MASA of M , then $A \subset M$ has the *so*-paving property, with $n_s(A \subset M; x, \varepsilon) \leq 25^4 \varepsilon^{-4}$ for all $x \in M_{\text{sa}}$.*

(2) *Let Γ be a countable group and $\Gamma \curvearrowright (X, \mu)$ an essentially free, ergodic, pmp action that is profinite. Then $A = L^\infty(X) \subset L^\infty(X) \rtimes \Gamma = M$ is *so*-pavable and, for every $x \in M_{\text{sa}}$, $n_s(A \subset M; x, \varepsilon) \leq 13^4 \varepsilon^{-4}$. So, also, $A^\omega \subset M^\omega$ satisfies the norm paving property and, for every $x \in M_{\text{sa}}^\omega$, $n(A^\omega \subset M^\omega; x, \varepsilon) \leq 13^4 \varepsilon^{-4}$.*

Proof. (1) By [Connes et al. 1981], given any $x \in M_{\text{sa}}$ and any *so*-neighborhood \mathcal{V} of 0, there exists a finite-dimensional von Neumann subalgebra $B_0 \subset M$, having the diagonal A_0 contained in A and $\mathcal{N}_{B_0}(A_0) \subset \mathcal{N}_M(A)$, and an element $y_0 = y_0^* \in B_0$, $\|y_0\| \leq \|x\|$, such that $x - y_0 \in \mathcal{V}$. But, by [Marcus et al. 2015] (see Lemma 3.1), y_0 can be (ε_0, n) paved over A_0 (thus also over $A \supset A_0$) for some ε_0 slightly smaller than $\varepsilon/2$ and $n \leq 25^4 \varepsilon^{-4}$. By Proposition 2.4, we conclude that x can be (ε, n) *so*-paved for every $\varepsilon > 0$.

(2) Take a decreasing sequence of finite-index subgroups $\Gamma_n < \Gamma$ such that (X, μ) is the inverse limit of the spaces Γ/Γ_n equipped with the normalized counting measure. Write $r_n : X \rightarrow \Gamma/\Gamma_n$. The essential freeness of $\Gamma \curvearrowright (X, \mu)$ means that, for every $g \in \Gamma - \{e\}$, we have

$$\lim_n \frac{|\{x \in \Gamma/\Gamma_n \mid gx = x\}|}{[\Gamma : \Gamma_n]} = 0. \tag{4-1}$$

Write $A_n = \ell^\infty(\Gamma/\Gamma_n)$. View $A_1 \subset A_2 \subset \dots$ as an increasing sequence of subalgebras of A with dense union. Fix a free ultrafilter ω on \mathbb{N} . For every $n \in \mathbb{N}$, define $M_n \cong M_{[\Gamma, \Gamma_n]}(\mathbb{C})$ as the matrix algebra with entries indexed by elements of Γ/Γ_n . Consider $A_n \subset M_n$ as the diagonal subalgebra. For $g \in \Gamma$, denote

by $u_{g,n} \in M_n$ the corresponding permutation unitary. Denote by τ_n the normalized trace on M_n and by $\|\cdot\|_2$ the corresponding 2-norm. By (4-1), we have that $\|E_{A_n}(u_{g,n})\|_2 \rightarrow 0$ for all $g \in \Gamma - \{e\}$.

Denote by $\mathcal{M} = \prod_{\omega}(M_n, \tau_n)$ the ultraproduct of the matrix algebras M_n , with MASA $\mathcal{A} \subset \mathcal{M}$ defined as $\mathcal{A} = \prod_{\omega} A_n$. We can then define a normal faithful $*$ -homomorphism $\pi : M \rightarrow \mathcal{M}$, where $\pi(au_g) \in \mathcal{M}$ is represented by the sequence $(au_{g,n})_{n \geq m}$ whenever $a \in A_m$.

Fix $\varepsilon > 0$ and denote by r the largest integer that is smaller than or equal to $(12/\varepsilon)^4$. We claim that, for every self-adjoint $x \in M^{\omega}$, there exists a partition p_1, \dots, p_r of A^{ω} such that $\|p_i(x - E_{A^{\omega}}(x))p_i\| \leq \varepsilon\|x\|$ for all i . To prove this claim, it suffices to prove the following local statement: for every self-adjoint $x \in M$ with $\|x\| \leq 1$, and for all $\delta > 0$, $m \in \mathbb{N}$, there exists a partition p_1, \dots, p_r of A (thus independent of m and δ , since r was fixed in the beginning) such that the element $y = \sum_{i=1}^r p_i(x - E_A(x))p_i$ satisfies

$$|\tau(y^k)| \leq \varepsilon^k + \delta \quad \text{for all } k = 1, \dots, m. \tag{4-2}$$

Indeed, once this local statement is proved and given a self-adjoint element $x \in M^{\omega}$ represented by a sequence $(x_m)_m$ with $x_m = x_m^*$ and $\|x_m\| \leq \|x\|$ for all m , we find partitions p_1^m, \dots, p_r^m of A such that the elements $y_m = \sum_{i=1}^r p_i^m(x_m - E_A(x_m))p_i^m$ satisfy

$$|\tau(y_m^k)| \leq (\varepsilon\|x_m\|)^k + \frac{1}{m} \leq (\varepsilon\|x\|)^k + \frac{1}{m} \quad \text{for all } k = 1, \dots, m.$$

Defining the projections $p_i \in A^{\omega}$ by the sequences $p_i = (p_i^m)_m$ and putting $y = \sum_{i=1}^r p_i(x - E_{A^{\omega}}(x))p_i$, this means that $|\tau(y^k)| \leq (\varepsilon\|x\|)^k$ for all $k \in \mathbb{N}$. Since y is self-adjoint, it follows from the spectral radius formula that $\|y\| \leq \varepsilon\|x\|$, so that the claim is proved. This means that every self-adjoint $x \in M^{\omega}$ can be (ε, n) paved for some $n \leq 12^4\varepsilon^{-4}$. So, by Theorem 2.7, also every $x \in M_{\text{sa}}$ can be (ε, n) so-paved for some $n \leq 13^4\varepsilon^{-4}$.

We now deduce the above local statement from [Marcus et al. 2015]. Fix $x \in M_{\text{sa}}$ with $\|x\| \leq 1$ and fix $\delta > 0$ and $m \in \mathbb{N}$. By the Kaplansky density theorem, we can take $n_0 \in \mathbb{N}$, a finite subset $\mathcal{F} \subset \Gamma$ and a self-adjoint $x_0 \in \text{span}\{au_g \mid a \in A_{n_0}, g \in \mathcal{F}\}$ with $\|x_0\| \leq 1$ and $\|x - x_0\|_2 \leq \delta/(m2^m)$. We may assume that $e \in \mathcal{F}$. We prove below that we can find a partition p_1, \dots, p_r of A such that the element $y_0 := \sum_{i=1}^r p_i(x_0 - E_A(x_0))p_i$ satisfies $|\tau(y_0^k)| \leq \varepsilon^k + \delta/2$ for all $k = 1, \dots, m$. Writing $y := \sum_{i=1}^r p_i(x - E_A(x))p_i$, we find that $\|y - y_0\|_2 \leq \|x - x_0\|_2$ and also $\|y\| \leq 2$, $\|y_0\| \leq 2$. Therefore,

$$\|y^k - y_0^k\|_2 \leq m2^{m-1}\|x - x_0\|_2 \leq \frac{\delta}{2} \quad \text{for all } k = 1, \dots, m.$$

Thus $|\tau(y^k) - \tau(y_0^k)| \leq \delta/2$, so that (4-2) follows.

We now must find a good paving for x_0 . For this, we use the ultraproduct \mathcal{M} and the injective homomorphism $\pi : M \rightarrow \mathcal{M}$ defined above. Write $x_0 = \sum_{g \in \mathcal{F}} a_g u_g$ with $a_g \in A_{n_0}$. Then, $\pi(x_0)$ is represented by the bounded sequence of self-adjoint elements $T_n := \sum_{g \in \mathcal{F}} a_g u_{g,n}$. Since $\|\pi(x_0)\| = \|x_0\| \leq 1$, we can take a bounded sequence of self-adjoint elements $S_n \in M_n$ such that $\lim_{n \rightarrow \omega} \|S_n\|_2 = 0$ and $\|T_n - S_n\| \leq 1$ for all n . Take $K > 0$ such that $\|T_n\| \leq K$ and $\|S_n\| \leq K$ for all n . Take $n_1 \geq n_0$ close enough to ω such that $\|S_{n_1}\|_2 \leq \delta/(4m(2K)^{m-1})$ and such that (using (4-1)) the projection $q \in A_{n_1}$

defined by the set

$$\{x \in \Gamma / \Gamma_{n_1} \mid gx \neq x \text{ for all } g \in \mathcal{F}^m \setminus \{e\}\}$$

satisfies $\|1 - q\|_2 \leq \delta/2^{m+2}$. Write $R = T_{n_1} - S_{n_1}$. Since $R = R^*$ and $\|R\| \leq 1$, by [Marcus et al. 2015], there exists a partition p_1, \dots, p_r of A_{n_1} such that the element $Y := \sum_{i=1}^r p_i (R - E_{A_{n_1}}(R)) p_i$ satisfies

$$\|Y\| \leq \frac{1}{2} \varepsilon \|R - E_{A_{n_1}}(R)\| \leq \varepsilon.$$

We define $Z := \sum_{i=1}^r p_i (T_{n_1} - E_{A_{n_1}}(T_{n_1})) p_i$. Note that $\|Y\| \leq 2$ and $\|Z\| \leq 2K$. Also, $\|Y - Z\|_2 \leq \|S_{n_1}\|_2$, so that, for all $k = 1, \dots, m$, we have

$$\|Y^k - Z^k\|_2 \leq m(2K)^{m-1} \|S_{n_1}\|_2 \leq \frac{\delta}{4}.$$

Then also $\|Y^k q - Z^k q\|_2 \leq \delta/4$. Because $\|Y^k q\| \leq \|Y\|^k \leq \varepsilon^k$, we conclude that

$$|\tau_{n_1}(Z^k q)| \leq \varepsilon^k + \frac{\delta}{4} \text{ for all } k = 1, \dots, m.$$

By our choice of q , whenever $1 \leq k \leq m$, $a_1, \dots, a_k \in A_{n_1}$ and $g_1, \dots, g_k \in \mathcal{F}$, we have

$$\tau_{n_1}(a_1 u_{g_1, n_1} \cdots a_k u_{g_k, n_k} q) = \tau(a_1 u_{g_1} \cdots a_k u_{g_k} q),$$

where the left-hand side uses the trace in M_{n_1} , while the right-hand side uses the trace in M . Writing $y_0 = \sum_{i=1}^r p_i (x_0 - E_A(x_0)) p_i$, we find that

$$|\tau(y_0^k q)| = |\tau_{n_1}(Z^k q)| \leq \varepsilon^k + \frac{\delta}{4} \text{ for all } k = 1, \dots, m.$$

Since $\|y_0^k q - y_0^k\|_2 \leq 2^m \|q - 1\|_2 \leq \delta/4$, we get the required estimate

$$|\tau(y_0^k)| \leq \varepsilon^k + \frac{\delta}{2} \text{ for all } k = 1, \dots, m. \quad \square$$

Remark 4.2. We believe that [Marcus et al. 2015] can be used to settle Conjecture 2.8 (*so*-pavability) for all Cartan subalgebras in II_1 factors $A \subset M$, and in fact for any Cartan subalgebra in a von Neumann algebra. The following could be an approach to a solution, but we could not make it work. Consider the abelian von Neumann algebra $\mathcal{A} = A \vee JAJ$ acting on $L^2(M)$. This is a MASA in $\mathcal{M} = \langle M, e_A \rangle = (JAJ)' \cap \mathcal{B}(L^2(M))$ and there exists a normal conditional expectation from the type I von Neumann algebra \mathcal{M} onto \mathcal{A} (see [Feldman and Moore 1977]). Therefore, $\mathcal{A} \subset \mathcal{M}$ satisfies the norm-paving property. If, now, $x \in M$, we can pave x by a partition $p_i \in A \vee JAJ$. Taking a very fine partition $q_j \in A$, we can *so*-approximate p_i by $\sum_j p_{i,j} Jq_j J$. It should be possible to choose the $p_{i,j}$ as “almost partitions” of 1 in A such that, for many j (or at least one j), the $p_{1,j}, \dots, p_{r,j}$ approximately pave x (in the *so*-paving sense).

In relation to the approach to proving *so*-pavability for Cartan subalgebras suggested above, let us mention that the [Marcus et al. 2015] paving property for discrete MASAs in type I von Neumann algebras allows the following new characterization for a MASA to be Cartan:

Corollary 4.3. *Let M be a von Neumann algebra with separable predual and $A \subset M$ a MASA in M that is the range of a normal conditional expectation. Let $\mathcal{M} = \langle M, e_A \rangle = (JAJ)' \cap \mathcal{B}(L^2 M)$ and $\mathcal{A} = A \vee JAJ$.*

The following conditions are equivalent:

- (1) A is a Cartan subalgebra of M .
- (2) \mathcal{A} is a Cartan subalgebra of \mathcal{M} .
- (3) $\mathcal{A} \subset \mathcal{M}$ has the paving property.

Proof. The equivalence of (1) and (2) follows from [Feldman and Moore 1977]. Since \mathcal{M} is of type I, a MASA in \mathcal{M} is a Cartan subalgebra if and only if it is the range of a normal conditional expectation. Also, an abelian subalgebra of \mathcal{M} can only satisfy the paving property if it is maximal abelian. Therefore, the equivalence of (2) and (3) follows from Theorem 3.3 (and, thus, uses [Marcus et al. 2015]). \square

5. Paving size for one or more elements

In [Marcus et al. 2015], it is shown that every self-adjoint element T in $\mathcal{B}(\ell_k^2)$, $1 \leq k \leq \infty$, can be $(\varepsilon, 12^4\varepsilon^{-4})$ paved over its diagonal MASA. In the previous section, we have used this result to prove that any amenable von Neumann algebra M with a Cartan subalgebra $A \subset M$ is $(\varepsilon, 25^4\varepsilon^{-4})$ *so*-pavable over A ; equivalently, any self-adjoint element in M^ω is $(\varepsilon, 25^4\varepsilon^{-4})$ norm pavable over A^ω .

On the other hand, it has been shown in [Popa 2014] that, if A is a singular MASA in a II_1 factor M , then $n(A^\omega \subset M^\omega; x, \varepsilon) \leq 25^2\varepsilon^{-2}(\varepsilon^{-1} + 1) \leq 1250\varepsilon^{-3}$ for all $x \in M_{\text{sa}}^\omega$. Or, equivalently, $n_s(A \subset M; x, \varepsilon) \leq 1250\varepsilon^{-3}$ for all $x \in M_{\text{sa}}$ (see [Popa 2014], Corollary 4.3 and the last lines of the proof of Proposition 2.3). This is shown by first proving that, given any $\varepsilon > 0$ and any finite set of projections in M that have scalar expectation onto A , one can find a simultaneous *so*-paving for all of them with at most $2\varepsilon^{-2}$ projections in A (see [Popa 2014, Corollary 4.2]), then using a dilation argument to deduce it for arbitrary self-adjoint elements.

We will now show that, in fact, the *so*-paving size for self-adjoint elements over singular MASAs, and respectively the norm-paving size over an ultraproduct of singular MASAs, can be improved to $4^2\varepsilon^{-2}$ (the order of magnitude ε^{-2} for the paving size is optimal; see Proposition 5.4 below). Moreover, we show that one can (ε, n) *so*-pave simultaneously any number of self-adjoint elements with $n < 1 + 4^2\varepsilon^{-2}$ many projections over a singular MASA, a phenomenon that does not occur in the classical Kadison–Singer case $\mathcal{D} \subset \mathcal{B}(\ell^2\mathbb{N})$, nor in fact for any Cartan subalgebra in a II_1 factor $A \subset M$ (see Remark 5.2 below). The proof combines the uniform paving of projections that have scalar expectation onto A in [Popa 2014, Corollary 4.2] with a better dilation argument that allows us not to lose on the paving size, while still dealing simultaneously with several self-adjoint elements.

Theorem 5.1. *Let $A_n \subset M_n$ be a sequence of singular MASAs in finite von Neumann algebras. Put $\mathbf{M} = \prod_\omega M_n$ and $\mathbf{A} = \prod_\omega A_n$.*

Let $\varepsilon > 0$. For every finite set of self-adjoint elements $\mathcal{F} \subset \mathbf{M} \ominus \mathbf{A}$, there exists a decomposition of the identity $1 = p_1 + \cdots + p_n$ with $n < 1 + 16\varepsilon^{-2}$ projections $p_j \in \mathbf{A}$ such that

$$\left\| \sum_{j=1}^n p_j x p_j \right\| \leq \varepsilon \|x\| \quad \text{for all } x \in \mathcal{F}.$$

Proof. Fix $\varepsilon > 0$ and let n be the unique integer satisfying $16\varepsilon^{-2} \leq n < 1 + 16\varepsilon^{-2}$. Also fix a finite subset $\{x_1, \dots, x_m\} \subset M \ominus A$ of self-adjoint elements. We may assume that $\|x_k\| = 1$ for all k . Define $y_k = (1 + x_k)/2$. Note that $0 \leq y_k \leq 1$ and $E_A(y_k) = \frac{1}{2}$. Let (B, τ) be any diffuse abelian von Neumann algebra. Write

$$\tilde{M} = \prod_{\omega} (M_2(\mathbb{C}) \otimes (M_n * B))$$

and consider the von Neumann subalgebra $\tilde{A} \subset \tilde{M}$ given by

$$\tilde{A} = \prod_{\omega} (A_n \oplus B) = A \oplus B^{\omega}.$$

Note that, for every n , we have that $A_n \oplus B \subset M_2(\mathbb{C}) \otimes (M_n * B)$ is a singular MASA. Therefore, $\tilde{A} \subset \tilde{M}$ is the ultraproduct of a sequence of singular MASAs.

Define the orthogonal projections $Q_k \in \tilde{M}$ given by

$$Q_k = \begin{pmatrix} y_k & \sqrt{y_k - y_k^2} \\ \sqrt{y_k - y_k^2} & 1 - y_k \end{pmatrix}.$$

Note that $E_{\tilde{A}}(Q_k) = \frac{1}{2}$.

Applying [Popa 2014, Theorem 4.1.(a)] to $X = \{Q_k - \frac{1}{2} \mid k = 1, \dots, m\}$, we find a diffuse von Neumann subalgebra $B_0 \subset \tilde{A}$ such that every product with factors alternatingly from $B_0 \ominus \mathbb{C}1$ and X has zero expectation on \tilde{A} . In particular, for all k , we have that B_0 and $\mathbb{C}1 + \mathbb{C}Q_k$ are free von Neumann subalgebras of (\tilde{M}, τ) .

Choose any decomposition of the identity $1 = P_1 + \dots + P_n$ with n projections $P_j \in B_0$ satisfying $\tau(P_j) = 1/n$. Fix $j \in \{1, \dots, n\}$ and $k \in \{1, \dots, m\}$. Since the projections P_j and Q_k are free, with traces respectively given by $1/n$ and $\frac{1}{2}$, it follows from [Voiculescu 1987, Example 2.8] that

$$\|P_j Q_k P_j - \frac{1}{2} P_j\| \leq \frac{2}{\sqrt{n}}.$$

Write $P_j = p_j \oplus q_j$, where $p_j \in A$ and $q_j \in B^{\omega}$ are projections. The upper left corner of $P_j Q_k P_j - \frac{1}{2} P_j$ equals $p_j(x_k/2)p_j$, and we conclude that

$$\|p_j x_k p_j\| \leq \frac{4}{\sqrt{n}} \leq \varepsilon.$$

This ends the proof. □

Remark 5.2. (1) As shown in Theorem 5.1 above, in the case that $A \subset M$ is singular, any finite number of elements can be simultaneously (ε, n) norm paved over A^{ω} with $n < 1 + 16\varepsilon^{-2}$. By [Popa 2014, Theorem 3.7], any finite number of elements can also be simultaneously (ε, n) L^2 -paved over A^{ω} with $n < 1 + \varepsilon^{-2}$. But this is no longer true for norm paving over a MASA that has “large normalizer”. For instance, one cannot pave multiple matrices in $\mathcal{B}(\ell^2\mathbb{N})$ over its diagonal \mathcal{D} . This can be seen as follows: Assume M is a finite von Neumann algebra and $A \subset M$ is a MASA whose normalizer $\mathcal{N}_M(A)$ generates a II_1 von Neumann algebra. Then, for any $m \geq 1$, there exists a unitary $u \in \mathcal{N}_M(A)$ such that $E_A(u^k) = 0$ for

all $1 \leq k \leq m - 1$, $u^m = 1$. Denote by σ the automorphism $\text{Ad}(u)$ of A . Assume now that p_1, \dots, p_n is a partition of A that simultaneously c paves the set of $m - 1$ unitaries $\{u^k \mid k = 1, \dots, m - 1\}$ for some $0 < c < 1$. Then $\|p_i u^k p_i\| \leq c$ for all $i = 1, \dots, n$ and all $k = 1, \dots, m - 1$. But $\|p_i u^k p_i\| = \|p_i \sigma^k(p_i)\|$ and $p_i \sigma^k(p_i)$ is a projection. Thus, $p_i \sigma^k(p_i)$ must be zero for all i and k . So, for every fixed i , we find that $p_i, \sigma(p_i), \dots, \sigma^{m-1}(p_i)$ are orthogonal. Thus, $\tau(p_i) \leq 1/m$. Since $\sum_i p_i = 1$, it follows that $n \geq m$. Note that, by replacing the cyclic group $\mathbb{Z}/m\mathbb{Z} \simeq \{u^k \mid 0 \leq k \leq m - 1\} \subset \mathcal{N}_M(A)$ with the group $(\mathbb{Z}/2\mathbb{Z})^t \hookrightarrow \mathcal{N}_M(A)$, acting freely on A , one gets the same result for $m = 2^t$, but with a set of $m - 1$ self-adjoint unitaries.

We conclude that if the normalizer of a MASA generates a type II_1 von Neumann algebra then, given any m , there exists a set of $m - 1$ unitaries in M such that, in order to simultaneously c pave all of them, with $c < 1$, we need at least m projections (in the case $m = 2^t$, the set can be taken of self-adjoint unitaries). Note that, if $u \in \mathcal{N}_M(A)$ is as before and we let $X = \{(u^k + u^{-k})/2, (u^k - u^{-k})/(2i) \mid 1 \leq k \leq m - 1\}$, then any partition of 1 with projections $p_1, \dots, p_n \subset A$ that simultaneously $(c/2)$ paves all $x \in X$ must satisfy $n \geq m = |X|/2 + 1$. Thus, under the same assumptions on $A \subset M$ as before, given any m_0 and any $c_0 < \frac{1}{2}$, there exists a set $X_0 \subset M_{\text{sa}}$ with $|X_0| = m_0$ such that, in order to simultaneously c_0 pave all $x \in X_0$, we need at least $m_0/2$ projections.

(2) If $A \subset M$ is a MASA in a von Neumann algebra, $X \subset M$ and $\varepsilon > 0$, we define $n(A \subset M; X, \varepsilon)$ in the obvious way. Also, for m a positive integer, we let $n(A \subset M; m, \varepsilon) = \sup\{n(A \subset M; X, \varepsilon) \mid X \subset M_{\text{sa}}, |X| = m\}$, and call it the *multipaving size* of $A \subset M$. One always has the estimate $n(A \subset M; m, \varepsilon) \leq n(A \subset M; \varepsilon)^m$. By Theorem 5.1, if A is a singular MASA in a II_1 factor M , then $n(A^\omega \subset M^\omega; m, \varepsilon) < 1 + 16\varepsilon^{-2}$ for all $m \geq 1$, $\varepsilon > 0$. By 5.2.1° above, if $\mathcal{N}_M(A)''$ is of type II_1 , then $n(A \subset M; m - 1, c) \geq m$ for all $m = 2^t$, $0 < c < 1$, while for arbitrary m_0 (not of the form 2^t) and $c_0 < \frac{1}{2}$, we have $n(A \subset M; m_0, c_0) \geq m_0/2$. At the same time, by [Marcus et al. 2015], we have $n(A \subset M; m, \varepsilon) \leq (12/\varepsilon)^{4m}$.

It would be interesting to find estimates for this multipaving size in this last case (when $\mathcal{N}_M(A)$ is large). By arguing as in the proof of [Popa 2014, Theorem 2.2], we see that $n(\mathfrak{D} \subset \mathfrak{B}; m, \varepsilon) = n(D^\omega \subset R^\omega; m, \varepsilon) = n(\mathbf{D} \subset \mathbf{M}; m, \varepsilon)$ for all $\varepsilon > 0$, $m \in \mathbb{N}$, where $\mathbf{D} \subset \mathbf{M}$ denotes the ultraproduct inclusion $\prod_\omega D_k \subset \prod_\omega M_{k \times k}(\mathbb{C})$. Thus, estimating the multipaving size for $D^\omega \subset R^\omega$, or for $\mathbf{D} \subset \mathbf{M}$, is the same as doing it for $\mathfrak{D} \subset \mathfrak{B}$. From Remark 5.2(1) and [Marcus et al. 2015], for each fixed $1 > \varepsilon > 0$, the growth in m of the multiple paving size $n(\mathfrak{D} \subset \mathfrak{B}; m, \varepsilon)$ is between m and $(\varepsilon^{-4})^m$. Calculating its order of magnitude seems a very challenging problem. It would already be interesting to decide whether this growth is linear (more generally, polynomial), or exponential.

Remark 5.3. Exactly the same proof as that of [Popa 2014, Theorem 4.1.(a)] shows the following more general result. Let (M, τ) be a von Neumann algebra with a normal faithful tracial state, $A \subset M$ a MASA in M and $A \subset N \subset M$ an intermediate von Neumann subalgebra with the following malnormality property: the only A - N -subbimodule of $L^2(M \ominus N)$ that is finitely generated as a right N -module is $\{0\}$. Then, given any $\|\cdot\|_2$ -separable subspace $X \subset M \ominus N$ and any free ultrafilter ω on \mathbb{N} , there exists a diffuse von Neumann subalgebra $B_0 \subset A^\omega$ such that every “word” with alternating “letters” from $B_0 \ominus \mathbb{C}1$ and X has trace zero. Note that [Popa 2014, Theorem 4.1.(a)] corresponds to the case $N = A$ because,

by [Popa 2006, Section 1.4], the singularity of A in M implies that $L^2(M \ominus A)$ contains no nonzero A - A -subbimodule that is finitely generated as a right A -module.

By combining this result with the dilation argument as in the proof of Theorem 5.1 above, it follows that any $x \in M \ominus N$ can be so -paved, with $n_s(A \subset M; x, \varepsilon) < 5^2 \varepsilon^{-2}$. Thus, if $A \subset N$ satisfies the so -paving property, then so does $A \subset M$, and we have the estimate $n_s(A \subset M; \varepsilon) \leq 20^2 \varepsilon^{-2} n_s(A \subset N; \varepsilon/2)$.

This observation allows us to derive the so -paving property (and thus the validity of Conjecture 2.8(1) for a class of MASAs that are neither singular nor Cartan. More precisely, assume that $A \subset M$ is a MASA in a II_1 factor such that the normalizer $\mathcal{N}_M(A)$ generates a von Neumann algebra N satisfying the conditions: (1) either N is amenable, or $A \subset N$ can be obtained as a group measure space construction from a free ergodic profinite action of a countable group; (2) $N \subset M$ satisfies the above malnormality condition. Then, $A \subset M$ has the so -paving property.

Concrete such examples can be easily derived from [Popa 1983]. For instance, [Popa 1983, Theorem 5.1] provides an example of a MASA A in the hyperfine II_1 factor $M \simeq R$ such that the normalizer of A in M generates a subfactor $N \subset M$ with the property that ${}_N L^2(M \ominus N)_N$ is an infinite multiple of the coarse N - N -bimodule $L^2(N) \otimes L^2(N)$, and thus $N \subset M$ satisfies the malnormality condition. Other examples come from free product constructions: let $A \subset N$ be a Cartan subalgebra of a (separable) amenable von Neumann algebra of type II_1 (e.g., the hyperfinite II_1 factor, $N \simeq R$); let (B, τ) be a diffuse finite von Neumann algebra and denote $M = N * B$; then, A is a MASA in M , the normalizer of A in M generates N , and again, by [Popa 1983, Remark 6.3], ${}_N L^2(M \ominus N)_N$ is an infinite multiple of the coarse N - N -bimodule, so that $N \subset M$ satisfies the malnormality condition.

We end with a result showing that the order of magnitude of the paving size obtained in Theorem 5.1 is optimal. More generally, we show that, for any MASA in any II_1 factor, the ε paving size is at least ε^{-2} , i.e., $\sup\{n(\varepsilon, x) \mid x \in M_{sa}\} \geq \varepsilon^{-2}$. The proof is very similar to [Casazza et al. 2007, Theorem 6], where it was shown that one needs at least ε^{-2} projections to ε pave self-adjoint unitary matrices.

Proposition 5.4. *Let M be a II_1 factor and $A \subset M$ a diffuse abelian von Neumann subalgebra. Let $\varepsilon > 0$ and $n < \varepsilon^{-2}$. There exists a self-adjoint unitary $x \in M$ with $E_A(x) = 0$ and*

$$\left\| \sum_{k=1}^n p_k x p_k \right\| \geq \left\| \sum_{k=1}^n p_k x p_k \right\|_2 > \varepsilon \tag{5-1}$$

for every decomposition of the identity $1 = p_1 + \dots + p_n$ with n projections $p_k \in A$.

So, if $A \subset M$ is a MASA in a II_1 factor, then the uniform L^2 paving size of $A^\omega \subset M^\omega$ is exactly equal to the smallest integer that is greater than or equal to ε^{-2} .

Proof. Fix $\varepsilon > 0$ and $n < \varepsilon^{-2}$. Take r large enough such that

$$\frac{r}{r-1} \frac{1}{n} - \frac{1}{r-1} > \varepsilon^2 \tag{5-2}$$

and such that there exists a conference matrix $C \in M_r(\mathbb{R})$ of size r , that is,

$$C_{ij} = \pm 1 \quad \text{if } i \neq j, \quad C_{ii} = 0 \quad \text{for all } i, \quad \text{and} \quad (r-1)^{-1/2} C \quad \text{is a self-adjoint unitary.}$$

Since A is diffuse, we can choose projections $e_1, \dots, e_r \in A$ with $1 = e_1 + \dots + e_r$ and $\tau(e_i) = 1/r$ for every i . Since M is a II_1 factor, we can choose partial isometries $v_1, \dots, v_r \in M$ such that $v_i v_i^* = e_1$ and $v_i^* v_i = e_i$ for all i . Define

$$x = \frac{1}{\sqrt{r-1}} \sum_{i,j=1}^r C_{ij} v_i^* v_j.$$

Note that x is a self-adjoint unitary. Since A is abelian, we have for all $i \neq j$ that

$$0 = e_i e_j E_A(v_i^* v_j) = e_i E_A(v_i^* v_j) e_j = E_A(e_i v_i^* v_j e_j) = E_A(v_i^* v_j).$$

Since $C_{ii} = 0$ for all i , we get that $E_A(x) = 0$.

Choose an arbitrary decomposition of the identity $1 = p_1 + \dots + p_n$ with n projections $p_k \in A$. We prove that (5-1) holds. First note that

$$\left\| \sum_{k=1}^n p_k x p_k \right\|_2^2 = \sum_{k=1}^n \|p_k x p_k\|_2^2 = \sum_{k=1}^n \tau(p_k x p_k x). \tag{5-3}$$

Since A is abelian, we can define the projections $p_{ik} = e_i p_k$. Writing $p_k = \sum_{i=1}^r p_{ik}$, we get for every $k \in \{1, \dots, n\}$ that

$$\begin{aligned} \tau(p_k x p_k x) &= \sum_{i,j=1}^r \tau(p_{ik} x p_{jk} x) = \sum_{i,j=1}^r \tau(p_{ik} x p_{jk} x e_i) \\ &= \frac{1}{r-1} \sum_{i,j=1}^r C_{ij}^2 \tau(p_{ik} v_i^* v_j p_{jk} v_j^* v_i) \\ &= \frac{1}{r-1} \left(\sum_{i,j=1}^r \tau(v_i p_{ik} v_i^* v_j p_{jk} v_j^*) - \sum_{i=1}^r \tau(v_i p_{ik} v_i^* v_i p_{ik} v_i^*) \right) \\ &= \frac{1}{r-1} (\tau(T_k^2) - \tau(p_k)), \quad \text{where } T_k = \sum_{i=1}^r v_i p_{ik} v_i^*. \end{aligned}$$

In combination with (5-3), it follows that

$$\left\| \sum_{k=1}^n p_k x p_k \right\|_2^2 = \frac{1}{r-1} \tau \left(\sum_{k=1}^n T_k^2 \right) - \frac{1}{r-1}. \tag{5-4}$$

We next observe that, as positive operators, we have

$$\sum_{k=1}^n T_k^2 \geq \frac{1}{n} \left(\sum_{k=1}^n T_k \right)^2. \tag{5-5}$$

Indeed, defining the elements $T, R \in M_{1,n}(\mathbb{C}) \otimes M$ given by

$$T = (T_1 \ T_2 \ \dots \ T_n) \quad \text{and} \quad R = (1 \ 1 \ \dots \ 1),$$

we get that

$$\left(\sum_{k=1}^n T_k \right)^2 = TR^*RT^* \leq \|R\|^2 TT^* = n \sum_{k=1}^n T_k^2.$$

So, (5-5) follows. By construction, we have that $\sum_{k=1}^n T_k = r e_1$. So, in combination with (5-4) and (5-2), we find that

$$\left\| \sum_{k=1}^n p_k x p_k \right\|_2^2 \geq \frac{1}{r-1} \frac{1}{n} \tau(r^2 e_1) - \frac{1}{r-1} = \frac{1}{r-1} \frac{r}{n} - \frac{1}{r-1} > \varepsilon^2.$$

Thus we have proved (5-1).

Now assume that $A \subset M$ is a MASA in the II_1 factor M . It follows that the uniform L^2 -paving size of $A^\omega \subset M^\omega$ is at least ε^{-2} . On the other hand, if n is an integer and $n \geq \varepsilon^{-2}$, it was proved in [Popa 2014, Section 3] that every element $x \in M^\omega$ can be (ε, n) L^2 -paved. \square

References

- [Akemann and Sherman 2012] C. A. Akemann and D. Sherman, “Conditional expectations onto maximal abelian $*$ -subalgebras”, *J. Operator Theory* **68**:2 (2012), 597–607. MR 2995737 Zbl 1274.46117
- [Casazza et al. 2007] P. Casazza, D. Edidin, D. Kalra, and V. I. Paulsen, “Projections and the Kadison–Singer problem”, *Oper. Matrices* **1**:3 (2007), 391–408. MR 2009a:46105 Zbl 1132.46037
- [Connes and Størmer 1978] A. Connes and E. Størmer, “Homogeneity of the state space of factors of type III₁”, *J. Functional Analysis* **28**:2 (1978), 187–196. MR 57 #10435 Zbl 0408.46048
- [Connes et al. 1981] A. Connes, J. Feldman, and B. Weiss, “An amenable equivalence relation is generated by a single transformation”, *Ergodic Theory Dynamical Systems* **1**:4 (1981), 431–450. MR 84h:46090 Zbl 0491.28018
- [Dixmier 1954] J. Dixmier, “Sous-anneaux abéliens maximaux dans les facteurs de type fini”, *Ann. of Math. (2)* **59** (1954), 279–286. MR 15,539b Zbl 0055.10702
- [Feldman and Moore 1977] J. Feldman and C. C. Moore, “Ergodic equivalence relations, cohomology, and von Neumann algebras, II”, *Trans. Amer. Math. Soc.* **234**:2 (1977), 325–359. MR 58 #28261b Zbl 0369.22010
- [Kadison and Singer 1959] R. V. Kadison and I. M. Singer, “Extensions of pure states”, *Amer. J. Math.* **81** (1959), 383–400. MR 23 #A1243 Zbl 0086.09704
- [Kechris 1995] A. S. Kechris, *Classical descriptive set theory*, Graduate Texts in Mathematics **156**, Springer, New York, 1995. MR 96e:03057 Zbl 0819.04002
- [de Korvin 1971] A. de Korvin, “Complete sets of expectations on von Neumann algebras”, *Quart. J. Math. Oxford Ser. (2)* **22** (1971), 135–142. MR 45 #913 Zbl 0207.44404
- [Marcus et al. 2015] A. W. Marcus, D. A. Spielman, and N. Srivastava, “Interlacing families, II: Mixed characteristic polynomials and the Kadison–Singer problem”, *Ann. of Math.* **182**:1 (2015), 327–350.
- [von Neumann 1949] J. von Neumann, “On rings of operators: reduction theory”, *Ann. of Math. (2)* **50** (1949), 401–485. MR 10,548a Zbl 0034.06102
- [Ocneanu 1985] A. Ocneanu, *Actions of discrete amenable groups on von Neumann algebras*, Lecture Notes in Mathematics **1138**, Springer, Berlin, 1985. MR 87e:46091 Zbl 0608.46035
- [Popa 1983] S. Popa, “Orthogonal pairs of $*$ -subalgebras in finite von Neumann algebras”, *J. Operator Theory* **9**:2 (1983), 253–268. MR 84h:46077 Zbl 0521.46048
- [Popa 1995] S. Popa, *Classification of subfactors and their endomorphisms*, CBMS Regional Conference Series in Mathematics **86**, American Mathematical Society, Providence, RI, 1995. MR 96d:46085 Zbl 0865.46044
- [Popa 1999] S. Popa, “The relative Dixmier property for inclusions of von Neumann algebras of finite index”, *Ann. Sci. École Norm. Sup. (4)* **32**:6 (1999), 743–767. MR 2000k:46084 Zbl 0966.46036

- [Popa 2006] S. Popa, “On a class of type II_1 factors with Betti numbers invariants”, *Ann. of Math.* (2) **163**:3 (2006), 809–899. MR 2006k:46097 Zbl 1120.46045
- [Popa 2014] S. Popa, “A II_1 factor approach to the Kadison–Singer problem”, *Comm. Math. Phys.* **332**:1 (2014), 379–414. MR 3253706 Zbl 1306.46060
- [Takesaki 2003] M. Takesaki, *Theory of operator algebras, II*, Encyclopaedia of Mathematical Sciences **125**, Springer, Berlin, 2003. MR 2004g:46079 Zbl 1059.46031
- [Tao 2013] T. Tao, “Real stable polynomials and the Kadison–Singer problem”, electronic resource, 2013, available at <https://terrytao.wordpress.com/tag/kadison-singer-problem/>.
- [Valette 2015] A. Valette, “Le problème de Kadison–Singer”, pp. 451–476 in *Séminaire Bourbaki 2013/2014* (Exposé 1088), Astérisque **367–368**, Société Mathématique de France, Paris, 2015.
- [Voiculescu 1987] D. Voiculescu, “Multiplication of certain noncommuting random variables”, *J. Operator Theory* **18**:2 (1987), 223–235. MR 89b:46076 Zbl 0662.46069

Received 12 Jan 2015. Revised 18 Feb 2015. Accepted 25 Mar 2015.

SORIN POPA: popa@math.ucla.edu

Department of Mathematics, University of California Los Angeles, Los Angeles, CA 90095-1555, United States

STEFAN VAES: stefaan.vaes@wis.kuleuven.be

Department of Mathematics, KU Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium

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 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 8 No. 4 2015

Inequality for Burkholder's martingale transform PAATA IVANISVILI	765
Classification of blowup limits for SU(3) singular Toda systems CHANG-SHOU LIN, JUN-CHENG WEI and LEI ZHANG	807
Ricci flow on surfaces with conic singularities RAFE MAZZEO, YANIR A. RUBINSTEIN and NATASA SESUM	839
Growth of Sobolev norms for the quintic NLS on T^2 EMANUELE HAUS and MICHELA PROCESI	883
Power spectrum of the geodesic flow on hyperbolic manifolds SEMYON DYATLOV, FRÉDÉRIC FAURE and COLIN GUILLARMOU	923
Paving over arbitrary MASAs in von Neumann algebras SORIN POPA and STEFAAN VAES	1001



2157-5045(2015)8:4;1-A