Pacific Journal of Mathematics

Volume 263 No. 1 May 2013

PACIFIC JOURNAL OF MATHEMATICS

msp.org/pjm

Founded in 1951 by E. F. Beckenbach (1906–1982) and F. Wolf (1904–1989)

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

The subscription price for 2013 is US $400/year for the electronic version, and $485/year for print and electronic.
Subscriptions, requests for back issues and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and the Science Citation Index.

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published monthly except July and August. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

PJM peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 263, No. 1, 2013
dx.doi.org/10.2140/pjm.2013.263.1

BIHARMONIC HYPERSURFACES IN COMPLETE RIEMANNIAN MANIFOLDS

LUIS J. ALÍAS, S. CAROLINA GARCÍA-MARTÍNEZ AND MARCO RIGOLI

We consider biharmonic hypersurfaces in complete Riemannian manifolds and prove that, under some additional assumptions, they are minimal.

1. Introduction

According to a definition first given by B. Y. Chen [1991], an isometrically immersed oriented hypersurface in Euclidean space, $\varphi : M \to \mathbb{R}^{m+1}$ is biharmonic if its mean curvature vector field $\boldsymbol{H}$ satisfies

$$\Delta \boldsymbol{H} = 0,$$

where Δ denotes the Laplacian on the hypersurface. It is well known that for submanifolds of Euclidean space, $\operatorname{trace}(B) = m\boldsymbol{H} = \Delta\varphi$, where B is the second fundamental form of the immersion. Hence, for any fixed unit vector $\boldsymbol{a}$ of $\mathbb{R}^{m+1}$,

$$m\Delta\langle \boldsymbol{H}, \boldsymbol{a}\rangle = \Delta^2\langle \varphi, \boldsymbol{a}\rangle \tag{1}$$

and the hypersurface is biharmonic if and only if each component of the immersion φ is a biharmonic function. Chen [1991; 1996] conjectured that a biharmonic hypersurface (in fact any biharmonic submanifold) of $\mathbb{R}^{m+1}$ is minimal, the converse being, of course trivially true. This statement is of a local nature and the conjecture holds for hypersurfaces in $\mathbb{R}^3$ [Chen 1991] and $\mathbb{R}^4$ [Hasanis and Vlachos 1995; Defever 1998]. However, in general, it has been shown to be true only under some additional assumptions, sometimes of a global nature: see for instance [Akutagawa and Maeta 2013] and [Nakauchi and Urakawa 2011].

This work was partially supported by MICINN (Ministerio de Ciencia e Innovación) and FEDER (Fondo Europeo de Desarrollo Regional) project MTM2009-10418 and Fundación Séneca project 04540/GERM/06, Spain. This research is a result of the activity developed within the framework of the Programme in Support of Excellence Groups of the Región de Murcia, Spain, by Fundación Séneca, Regional Agency for Science and Technology (Regional Plan for Science and Technology 2007–2010). García-Martínez was supported by a research training grant within the framework of the GERM program (Grupos de Excelencia de la Región de Murcia) at Universidad de Murcia. Rigoli was partially supported by MEC Grant SAB2010-0073.
MSC2010: primary 53C40, 53C42; secondary 58E20.
Keywords: mean curvature vector, biharmonic hypersurfaces, Chen conjecture.

This problem can be considered in a more general perspective. Indeed, let (M, g) and (N, h) be Riemannian manifolds and $\varphi : (M, g) \to (N, h)$ a smooth map. Let $\tau(\varphi)$ denote its tension field, that is,

$$\tau(\varphi) = \operatorname{trace}(\nabla d\varphi) = \sum_{i=1}^{m} (\nabla d\varphi)(e_i, e_i), \quad m = \dim M,$$

where $\nabla d\varphi$ is the generalized second fundamental tensor and $\{e_1, \dots, e_m\}$ is a local orthonormal frame on (M, g). Given a relatively compact domain $\Omega \subset M$ one introduces the bienergy functional $E_\tau^\varphi(\Omega)$ on Ω by setting

$$E_\tau^\varphi(\Omega) = \frac{1}{2} \int_\Omega |\tau(\varphi)|^2,$$

where integration is understood with respect to the volume element of (M, g). Then φ is a biharmonic map (meaning a critical point of this functional on M — i.e., on each relatively compact domain $\Omega \subset M$), if and only if the bitension field

$$\tau_2(\varphi) = \Delta\tau(\varphi) - \sum_i R^N\big(\tau(\varphi), \varphi_*(e_i)\big)\varphi_*(e_i) \tag{2}$$

vanishes identically. Here R^N denotes the $(3,1)$ curvature tensor of (N, h).

When $\varphi : (M^m, g) \to (N^{m+1}, h)$ is an isometric immersion of an m-dimensional hypersurface and ν is a local unit normal vector field along φ, writing the mean curvature vector as

$$\boldsymbol{H} = H\nu \tag{3}$$

and indicating with B the second fundamental form in the direction of ν, a heavy computation shows that (2) is equivalent to the system

$$\Delta H - |B|^2 H + \operatorname{Ric}^N(\nu, \nu)H = 0, \tag{4a}$$

$$2B(\nabla H, \cdot)^\sharp + \tfrac{1}{2} m \nabla H^2 - 2H(\operatorname{Ric}^N(\nu, \cdot)^\sharp)^T = 0, \tag{4b}$$

where $^\sharp : TM^* \to TM$ denotes the musical isomorphism, ^{T} the tangential component and Ric^N the Ricci tensor of (N, h) [Ou 2010, Theorem 2.1].

At this point one easily verifies that a biharmonic hypersurface in $\mathbb{R}^{m+1}$ in the sense of Chen is exactly a biharmonic hypersurface as defined in this more general setting. In this new perspective Chen's conjecture has been generalized to the following [Caddeo et al. 2001; 2002]:

> *Let $\varphi : (M, g) \to (N, h)$ be an isometric immersion into a Riemannian manifold of nonpositive sectional curvature. If φ is biharmonic then it is minimal.*

This new conjecture has been shown to be true if M is compact [Jiang 1986] or if H is constant [Ou 2010], but false in general [Ou and Tang 2012]. Here we restrict ourselves to complete noncompact biharmonic hypersurfaces and in fact we concentrate our efforts on the consequences of (4a) alone.

To avoid confusion with a terminology used for biharmonic submanifolds, we underline that in what follows by a *proper immersion* we mean an immersion that is topologically proper: preimages of compact sets are compact sets.

2. Statement of main results

Our first main result is the following.

Theorem 1. *Let $\varphi : M \to (N, \langle\, ,\rangle)$ be an oriented, proper, isometrically immersed, biharmonic hypersurface in the complete manifold N. For some origin $o \in N$ assume that*

$$\varphi(M) \cap \operatorname{cut}(o) = \varnothing.$$

Having set $\varrho = \operatorname{dist}_N(\,\cdot\,, o)$, suppose that the radial sectional curvature K^N_{rad} of N satisfies

$$K^N_{\mathrm{rad}} \geq -G(\varrho) \tag{5}$$

for $\varrho \gg 1$ and some $G \in \mathscr{C}^2(\mathbb{R}^+_0)$ such that $G(0) > 0$, $G'(t) \geq 0$ and $G(t) = o(t^2)$ as $t \to +\infty$. Let ν be a unit normal vector field along φ and suppose

$$\operatorname{Ric}^N(\nu, \nu) \leq 0 \tag{6}$$

along φ. Then φ is minimal. In particular if the sectional curvature K^N_{sect} is nonpositive, $\varphi(M)$ is unbounded in N.

As an immediate consequence of Theorem 1, using [Mari and Rigoli 2010] and [Alías et al. 2009], we obtain:

Corollary 2. *Let $\varphi : M \to \mathbb{R}^{m+1}$ be an oriented, isometrically immersed, biharmonic hypersurface. If the image $\varphi(M)$ is contained in a nondegenerate open cone of $\mathbb{R}^{m+1}$ or the hypersurface is cylindrically bounded as $\varphi(M) \subset B_r(o) \times \mathbb{R}^{m-1} \subset \mathbb{R}^2 \times \mathbb{R}^{m-1}$, then the immersion cannot be proper.*

We recall here that, fixed an origin $o \in \mathbb{R}^{m+1}$, the nondegenerate cone with vertex o, direction a and width θ is the subset

$$\mathscr{C} = \mathscr{C}_{o,a,\theta} = \left\{ p \in \mathbb{R}^{m+1} \backslash \{o\} : \left\langle \frac{p - o}{|p - o|}, a \right\rangle \geq \cos\theta \right\},$$

where $a \in \mathbb{S}^m$ is a unit vector and $\theta \in (0, \pi/2)$. By nondegenerate we mean that it is strictly smaller than a half-space. On the other hand, following the definition introduced in [Alías et al. 2009], an immersed hypersurface $\varphi : M \to \mathbb{R}^{m+1}$ is said to be cylindrically bounded if $\varphi(M) \subset B_r(o) \times \mathbb{R}^{m+1-p} \subset \mathbb{R}^p \times \mathbb{R}^{m+1-p}$, where $p \geq 2$ and $B_r(o) \subset \mathbb{R}^p$ denotes the ball of radius r. In particular, $p = 2$ gives the weakest requirement.

To introduce the next result we consider the operator

$$L = \Delta + \mathrm{Ric}^N(\nu, \nu) \tag{7}$$

where ν is a unit normal vector field along the hypersurface $\varphi : M \to (N, \langle\, ,\rangle)$ and we let $\lambda_1^L(M)$ denote its spectral radius. Clearly if $\mathrm{Ric}^N(\nu, \nu) \leq 0$ then $\lambda_1^L(M) \geq 0$ but this latter fact can be true even if $\mathrm{Ric}^N(\nu, \nu) > 0$ provided this positivity compensate with the geometry of M. (For a detailed discussion see [Bianchini et al. 2012]). Thus $\lambda_1^L(M) \geq 0$ is weaker than $\mathrm{Ric}^N(\nu, \nu) \leq 0$.

Theorem 3. *Let $\varphi : M \to (N, \langle\, ,\rangle)$ be a biharmonic, complete, oriented hypersurface with mean curvature H. Suppose that the operator L in* (7) *satisfies*

$$\lambda_1^L(M) \geq 0. \tag{8}$$

If $H \in L^2(M)$ then φ is minimal.

This result is extended to a different class of integrability for H in Theorem 7 of Section 3 below.

Next, we consider the case when $(N, \langle\, ,\rangle)$ is a Cartan–Hadamard manifold, that is, N is complete, simply connected and with nonpositive sectional curvature. What follows is a gap theorem.

Theorem 4. *Let $\varphi : M \to (N, \langle\, ,\rangle)$ be an isometrically immersed, oriented, biharmonic hypersurface of dimension $m \geq 3$ into a Cartan–Hadamard manifold. Suppose that the mean curvature H satisfies*

$$\|H\|_{L^m(M)} < \frac{\omega_m^{1/m}}{\pi 2^{m-1}} \frac{m-1}{m(m+1)^{1+\frac{1}{m}}}, \tag{9}$$

where ω_m is the volume of the unit ball of $\mathbb{R}^m$. Then φ is a minimal hypersurface.

3. Proof of the main theorems and some further results

With the notations of Theorem 1 we consider the function $v = \varrho^2 \circ \varphi$. The assumption $\varphi(M) \cap \mathrm{cut}(o) = \varnothing$ implies that v is smooth on M. Clearly,

$$|\nabla v| \leq 2\sqrt{v}. \tag{10}$$

Since M is complete and noncompact and φ is proper we have

$$v(x) \to +\infty \quad \text{as } x \to \infty \quad \text{in } M. \tag{11}$$

To compute Δv we recall (see, for instance, [Jorge and Koutroufiotis 1981]) that

$$\Delta(\varrho^2 \circ \varphi) = (\operatorname{Hess} \varrho^2)\big(\varphi_*(e_i), \varphi_*(e_i)\big) + \langle \nabla \varrho^2, m\boldsymbol{H}\rangle \tag{12}$$

with $\{e_i\}$ a local orthonormal frame on M. Let $G \in C^\infty(\mathbb{R}_0^+)$ satisfy

$$G(0) > 0 \qquad \text{and} \qquad G'(t) \geq 0 \ \text{ on } \mathbb{R}_0^+. \tag{13}$$

(In particular, G can be chosen to agree, for t large, with the function ct^d, where $0 < d < 2$, or with $ct^2(\log t)^{-\varepsilon}$, where $\varepsilon > 0$.)

If $K_{\mathrm{rad}}^N \geq -G$, by the Hessian comparison theorem (see Theorem 2.3 and Remark 2.3 of [Pigola et al. 2008] for the appropriate statement that we are using here) we get

$$\operatorname{Hess}(\varrho^2) \leq C\varrho\sqrt{G(\varrho)}\langle\,,\rangle \tag{14}$$

outside a compact set and for some appropriate constant $C > 0$. Up to modifying C we can assume that (14) is true on M. Hence, from (12) and (14) we deduce that

$$\Delta v \leq C^2\sqrt{v}\sqrt{G(\sqrt{v})} + 2m\sqrt{v}|H| \tag{15}$$

on M. Next, from (4a), letting $u = H^2$ we get

$$\Delta u = 2H\Delta H + 2|\nabla H|^2 = 2|B|^2 u - 2\operatorname{Ric}^N(\nu,\nu)u + 2|\nabla H|^2. \tag{16}$$

Using Newton's inequality,

$$|B|^2 \geq m|H|^2, \tag{17}$$

we obtain

$$\Delta u + 2\operatorname{Ric}^N(\nu,\nu)u - 2mu^2 \geq 2|\nabla H|^2 \geq 0, \tag{18}$$

and we are left with a solution $u \geq 0$ of the differential inequality

$$\Delta u + a(x)u - 2mu^2 \geq 0 \tag{19}$$

with

$$a(x) = 2\operatorname{Ric}^N(\nu,\nu) \circ \varphi(x). \tag{20}$$

Proof of Theorem 1. First observe that since φ is proper and N is complete, the induced metric on M is complete. Next we follow an idea introduced in [Akutagawa and Maeta 2013]. Since φ is proper, for every $T \in \mathbb{R}^+$, the set

$$D_T = v^{-1}([0, T])$$

is compact. Suppose $u \not\equiv 0$. Then there exists $x_0 \in M$ such that $u(x_0) > 0$ and we can suppose to have chosen T sufficiently large that $x_0 \in D_{T/2} \setminus \partial D_{T/2}$.

We define

$$F(x) = (T - v(x))^2 u(x) \tag{21}$$

on D_T. Note that $F \geq 0$, $F \equiv 0$ on ∂D_T and $F(x_0) > 0$. It follows that there exists a positive absolute maximum for $F(x)$ at some point $\bar{x} \in D_T \backslash \partial D_T$. At this point we have

$$\frac{\nabla F}{F}(\bar{x}) = 0 \quad \text{and} \quad \frac{\Delta F}{F}(\bar{x}) \leq 0. \tag{22}$$

From (22), a straightforward computation yields

$$\frac{\nabla u(\bar{x})}{u(\bar{x})} = \frac{2}{T - v(\bar{x})} \nabla v(\bar{x}) \tag{23}$$

and

$$\frac{\Delta u(\bar{x})}{u(\bar{x})} \leq \frac{2}{T - v(\bar{x})} \Delta v(\bar{x}) - \frac{2}{(T - v(\bar{x}))^2} |\nabla v(\bar{x})|^2 + \frac{4}{T - v(\bar{x})} \frac{|\nabla u(\bar{x})|}{u(\bar{x})} |\nabla v(\bar{x})|.$$

We use (23), (15) at $\bar{x}$ with $\sqrt{u} = |H|$, and (10) at $\bar{x}$ into the above inequality to obtain (omitting $\bar{x}$ for the ease of notation)

$$\begin{aligned} \frac{\Delta u}{u} &\leq \frac{2}{T - v} \left[C^2 \sqrt{G(\sqrt{v})} + 2m\sqrt{u} \right] \sqrt{v} + \frac{6}{(T - v)^2} |\nabla v|^2 \\ &\leq \frac{2}{T - v} \left[C^2 \sqrt{G(\sqrt{v})} + 2m\sqrt{u} \right] \sqrt{v} + \frac{24}{(T - v)^2} v. \end{aligned}$$

From (19) we then deduce

$$u \leq \frac{a}{2m} + \frac{C^2 \sqrt{v}}{m(T - v)} \sqrt{G(\sqrt{v})} + \frac{2\sqrt{v}}{T - v} \sqrt{u} + \frac{12}{m(T - v)^2} v. \tag{24}$$

Multiplying by $(T - v(x))^2$ both sides of (24) and using that $a(x) = a_+(x) - a_-(x)$, that G is nondecreasing, and that $\bar{x} \in D_T$ we have

$$\begin{aligned} F(\bar{x}) \leq \frac{a_+(\bar{x})}{2m} (T - v(\bar{x}))^2 + \frac{C^2 \sqrt{v(\bar{x})}}{m} (T - v(\bar{x})) \sqrt{G(\sqrt{v(\bar{x})})} \\ + 2\sqrt{v(\bar{x})} \sqrt{F(\bar{x})} + \frac{12}{m} v(\bar{x}) \\ \leq \frac{T^2}{2m} a_+(\bar{x}) + \frac{C^2 T^{3/2}}{m} \sqrt{G(\sqrt{T})} + 2\sqrt{T} \sqrt{F(\bar{x})} + \frac{12}{m} T. \end{aligned}$$

Therefore

$$F(\bar{x}) - 2\sqrt{T} \sqrt{F(\bar{x})} - T Z(T) \leq 0,$$

where

$$Z(T) = \frac{T}{2m} \sup_{D_T} a_+ + \frac{C^2}{m} \sqrt{T} \sqrt{G(\sqrt{T})} + \frac{12}{m}.$$

Note that $Z(T) \geq 0$. Then

$$F(x_0) \leq F(\bar{x}) \leq T\big(1+\sqrt{1+Z(T)}\big)^2 \leq C^2 T(1+Z(T))$$

and therefore, since $x_0 \in D_{T/2}$,

$$\begin{aligned} u(x_0) &\leq \frac{C^2 T}{(T-v(x_0))^2}\big(T \sup_{D_T} a_+ + \sqrt{T}\sqrt{G(\sqrt{T})}\big) \\ &\leq \frac{C^2}{T}\big(T \sup_{D_T} a_+ + \sqrt{T}\sqrt{G(\sqrt{T})}\big) = C^2\big(\sup_{D_T} a_+ + \frac{1}{\sqrt{T}}\sqrt{G(\sqrt{T})}\big). \end{aligned}$$

However, by assumption $a_+ \equiv 0$ and using $G(t) = o(t^2)$ as $t \to +\infty$ we have

$$T^{-1/2}\sqrt{G(\sqrt{T})} = o(1) \quad \text{as } T \to +\infty.$$

Thus, letting $T \to +\infty$ in (25), we deduce $u(x_0) \leq 0$ which contradicts the assumption $u(x_0) > 0$. The contradiction shows that $u = H^2 \equiv 0$ on M, that is, φ is minimal.

Suppose now that $K^N_{\text{sect}} \leq 0$. Since φ is minimal (15) becomes

$$\Delta v \leq C^2\sqrt{v}\sqrt{G(\sqrt{v})}. \tag{25}$$

This, together with (10) and (11), guarantees the validity of the Omori–Yau maximum principle on M (see Theorem 1.9 of [Pigola et al. 2005]). Now the result follows from Theorem 3.9 of [Pigola et al. 2005]. □

For the proof of Theorem 3 we need the next proposition which is a version, adapted to the present purposes, of Lemma 3.1 in [Brandolini et al. 1998].

Proposition 5. *Let $(M, \langle\, ,\rangle)$ be a complete manifold and let $a(x), b(x) \in \mathscr{C}^0(M)$ and suppose that*

$$b(x) \geq 0 \tag{26}$$

and

$$\lambda_1^L(M) \geq 0 \quad \text{with } L = \Delta + a(x). \tag{27}$$

Let $u \in C^2(M)$ be a solution of

$$\Delta u + a(x)u - b(x)u = 0 \quad \text{on } M. \tag{28}$$

If $u \in L^2(M)$ then $u \equiv 0$ on $\operatorname{supp}(b(x))$. *In particular, if u does not change sign and $b(x) \not\equiv 0$, then $u \equiv 0$.*

Proof. We suppose $b(x) \not\equiv 0$ otherwise there is nothing to prove. Next, we reason by contradiction and we assume the existence of $x_0 \in \operatorname{supp}(b(x)) \subset M$ such that $u(x_0) \neq 0$ and $b(x_0) \neq 0$. (Note that if $u(x_0) \neq 0$ and $b(x_0) = 0$ by continuity we can always find x_0' sufficiently close to x_0 so that $u(x_0') \neq 0$ and $b(x_0') \neq 0$). Choose $R \gg 1$ such that $x_0 \in B_R$. Let ψ be a cut-off function $0 \leq \psi \leq 1$ satisfying

$$\psi \equiv 1 \quad \text{on } B_R, \quad \operatorname{supp}(\psi) \subseteq B_{R+1}, \quad |\nabla \psi| \leq 2.$$

Then $u\psi \in \mathscr{C}_0^2(M)$, $u\psi \neq 0$ and by the variational characterization of $\lambda_1^L(B_{R+1})$ we have

$$\lambda_1^L(B_{R+1}) \leq \frac{\int_{B_{R+1}} \big(|\nabla(u\psi)|^2 - a(x)(u\psi)^2\big)}{\int_{B_{R+1}} (u\psi)^2}. \tag{29}$$

Since $\lambda_1^L(M) \geq 0$ the monotonicity property of eigenvalues yields $\lambda_1^L(B_{R+1}) > 0$. Next, we consider the vector field $W = u\psi^2 \nabla u$. A direct computation using (28) gives

$$\operatorname{div}(W) = b(x)u^2\psi^2 - a(x)u^2\psi^2 + |\nabla(u\psi)|^2 - u^2|\nabla\psi|^2.$$

Hence by (29) and the divergence theorem

$$0 \geq \lambda_1^L(B_{R+1}) \int_{B_{R+1}} u^2\psi^2 - \int_{B_{R+1}} u^2|\nabla\psi|^2 + \int_{B_{R+1}} b(x)u^2\psi^2.$$

Rearranging, using the properties of ψ and (26) we obtain

$$\lambda_1^L(B_{R+1}) \int_{B_R} u^2 - \int_{B_R} b(x)u^2 \leq 4 \int_{B_{R+1}\setminus B_R} u^2.$$

Letting $R \to +\infty$ and using the fact that $u \in L^2(M)$ we deduce

$$\lambda_1^L(M) \int_M u^2 - \int_M b(x)u^2 \leq 0.$$

We reach a contradiction by observing that $\lambda_1^L(M) \geq 0$ and in a neighborhood of x_0, $b(x)$ and $u^2(x)$ are strictly positive.

The last statement follows immediately from the strong maximum principle and (28) (see the remark after the proof of Theorem 3.5 on page 35 of [Gilbarg and Trudinger 1983]). □

Proof of Theorem 3. We apply Proposition 5 to the solution H of (4a) with $a(x) = \operatorname{Ric}^N(\nu, \nu)$ and $b(x) = |B|^2$. By Newton's inequality (17), $\operatorname{supp}(H) \subseteq \operatorname{supp}(b(x))$, which gives a contradiction to the conclusion of Proposition 5 unless $H \equiv 0$; thus $\varphi : M \to (N, \langle\,,\rangle)$ is minimal. □

Corollary 6. *Any biharmonic, isometrically immersed, complete oriented hypersurface M with mean curvature satisfying $H \in L^2(M)$ in a space with nonpositive Ricci tensor is minimal.*

For the proof of this corollary simply observe that since $\mathrm{Ric}^N(\nu,\nu) \le 0$ then $\lambda_1^L(M) \ge 0$ for $L = \Delta + \mathrm{Ric}^N(\nu,\nu)$.

With the aid of Theorem 4.6 in [Pigola et al. 2008] we can extend the range of integrability of H as follows.

Theorem 7. *Let $\varphi : M \to (N, \langle\,,\rangle)$ be a biharmonic, isometrically immersed, oriented hypersurface. For some $\Lambda \ge \frac{1}{2}$ let $L_\Lambda = \Delta + 2\Lambda\,\mathrm{Ric}^N(\nu,\nu)$ and suppose that*

$$\lambda_1^{L_\Lambda}(M) \ge 0. \tag{30}$$

Let $-\frac{1}{2} \le \beta \le \Lambda - 1$ and assume that

$$H \in L^{4(\beta+1)}(M). \tag{31}$$

Then φ is minimal.

Remark 8. If $\Lambda = \frac{1}{2}$, $L_\Lambda = L = \Delta + \mathrm{Ric}^N(\nu,\nu)$ and $\beta = -\frac{1}{2}$ so that condition (31) becomes $H \in L^2(M)$. In this way, we recover Theorem 3.

Proof of Theorem 7. We let $u = H^2$. From the differential inequality (18) and

$$|\nabla H|^2 = \frac{1}{4}\frac{|\nabla u|^2}{u}$$

we deduce that u is a nonnegative solution of

$$u\Delta u + 2\,\mathrm{Ric}^N(\nu,\nu)u^2 - 2mu^3 \ge \tfrac{1}{2}|\nabla u|^2. \tag{32}$$

By Theorem 1 of [Fischer-Colbrie and Schoen 1980], inequality (30) implies the existence of a positive solution ψ on M of

$$\Delta\psi + 2\Lambda\,\mathrm{Ric}^N(\nu,\nu)\psi = 0.$$

We can thus apply Theorem 4.6 of [Pigola et al. 2008] with $\varphi = \psi$, $A = -\frac{1}{2}$, $|\boldsymbol{H}| = \Lambda$, $K = 0$, $a(x) = 2\,\mathrm{Ric}^N(\nu,\nu)$, $b(x) = 2m$ and $\sigma = 2$. Note that assumption (4.43) of Theorem 4.6 of [Pigola et al. 2008] is true by (31). It follows that $u \equiv 0$, that is, $\varphi : M \to (N, \langle\,,\rangle)$ is minimal. □

We remark that if we let $L_{m/4} = \Delta + (m/2)\,\mathrm{Ric}^N(\nu,\nu)$ and we assume

$$\lambda_1^{L_{m/4}}(M) \ge 0, \tag{33}$$

as a consequence of Theorem 7, if $H \in L^m(M)$ then φ is minimal.

As a matter of fact, we can avoid assumption (33) and obtain the same conclusion in case $(N, \langle\,,\rangle)$ is a Cartan–Hadamard manifold. This is the content of Theorem 4. Towards this end, we observe that if $\varphi : M \to (N, \langle\,,\rangle)$ is an isometric immersion of dimension $m \geq 2$, Hoffman and Spruck [1974] have shown the validity of the following L^1-Sobolev inequality: for every $u \in W_0^{1,1}(M)$,

$$S_1(m)^{-1}\left(\int_M |u|^{m/(m-1)}\right)^{(m-1)/m} \leq \int_M \big(|\nabla u| + m|H||u|\big) \tag{34}$$

with

$$S_1(m) = \frac{\pi 2^{m-1}}{\omega_m^{1/m}} \frac{(m+1)^{1+\frac{1}{m}}}{m-1} \tag{35}$$

where ω_m is the volume of the unit ball of $\mathbb{R}^m$ (observe that in [Hoffman and Spruck 1974] the mean curvature vector field is not normalized). Having fixed $\varepsilon > 0$, from (34) we immediately deduce (see for instance [Pigola et al. 2008, pp. 175–176]) that for every $v \in W_0^{1,2}(M)$

$$S_2(m,\varepsilon)^{-1}\left(\int_M |v|^{2m/(m-2)}\right)^{(m-2)/m} \leq \int_M \left(|\nabla v|^2 + \frac{\varepsilon^2}{4}\left(\frac{m-2}{m-1}\right)^2 m^2|H|^2 v^2\right) \tag{36}$$

with

$$S_2(m,\varepsilon) = \frac{4(m-1)^2}{(m-2)^2}\frac{1+\varepsilon^2}{\varepsilon^2} S_1(m)^2. \tag{37}$$

Proof of Theorem 4. In the assumptions of the theorem and by the above discussion we have the validity of (36) on M. Next, for $u = H^2$ we rewrite (16) in the form

$$u\Delta u + 2\operatorname{Ric}^N(\nu,\nu)u^2 - 2|B|^2u^2 = \tfrac{1}{2}|\nabla u|^2. \tag{38}$$

Since N is Cartan–Hadamard,

$$2(\operatorname{Ric}^N(\nu,\nu) - |B|^2) \leq 0. \tag{39}$$

From (9) and the fact that $H \in L^m(M)$ we have

$$u \in L^{m/2}(M) \quad \text{with } m/2 > \tfrac{1}{2}, \tag{40}$$

because $m \geq 3$. Applying Theorem 9.12 of [Pigola et al. 2008] with $\sigma = m/2$, $\alpha = 2/m$ and $A = -\frac{1}{2}$ to (38) we deduce that either u is identically zero or, by formula (9.41) of [Pigola et al. 2008],

$$\left(\int_M |H|^m\right)^{2/m} \geq \frac{1}{(1+\varepsilon^2)m^2 S_1(m)^2}.$$

Note that to obtain this inequality we use (37). Thus, letting $\varepsilon \downarrow 0^+$ we obtain

$$\|H\|_{L^m(M)} \geq \frac{1}{mS_1(m)} = \frac{\omega_m^{1/m}}{\pi 2^{m-1}} \frac{m-1}{m(m+1)^{1+\frac{1}{m}}}.$$

Using (35) in this latter we contradict (9). Thus $u \equiv 0$ and $\varphi : M \to (N, \langle\, , \rangle)$ is minimal. $\square$

Acknowledgments.

The authors thank the anonymous referee for valuable suggestions and corrections which contributed to improve this paper.

References

[Akutagawa and Maeta 2013] K. Akutagawa and S. Maeta, "Biharmonic properly immersed submanifolds in the Euclidean spaces", *Geom. Dedicata* (2013). To appear.

[Alías et al. 2009] L. J. Alías, G. P. Bessa, and M. Dajczer, "The mean curvature of cylindrically bounded submanifolds", *Math. Ann.* **345**:2 (2009), 367–376. MR 2010i:53104 Zbl 1200.53050

[Bianchini et al. 2012] B. Bianchini, L. Mari, and M. Rigoli, "On some aspects of oscillation theory and geometry", preprint, 2012, Available at http://hdl.handle.net/2434/170621. To appear in *Mem. Amer. Math. Soc.*

[Brandolini et al. 1998] L. Brandolini, M. Rigoli, and A. G. Setti, "Positive solutions of Yamabe type equations on complete manifolds and applications", *J. Funct. Anal.* **160**:1 (1998), 176–222. MR 2000a:35051 Zbl 0923.58049

[Caddeo et al. 2001] R. Caddeo, S. Montaldo, and C. Oniciuc, "Biharmonic submanifolds of S^3", *Internat. J. Math.* **12**:8 (2001), 867–876. MR 2002k:53123 Zbl 1111.53302

[Caddeo et al. 2002] R. Caddeo, S. Montaldo, and C. Oniciuc, "Biharmonic submanifolds in spheres", *Israel J. Math.* **130** (2002), 109–123. MR 2003c:53090 Zbl 1038.58011

[Chen 1991] B.-Y. Chen, "Some open problems and conjectures on submanifolds of finite type", *Soochow J. Math.* **17**:2 (1991), 169–188. MR 92m:53091 Zbl 0749.53037

[Chen 1996] B.-Y. Chen, "A report on submanifolds of finite type", *Soochow J. Math.* **22**:2 (1996), 117–337. MR 98b:53043 Zbl 0867.53001

[Defever 1998] F. Defever, "Hypersurfaces of $\mathbf{E}^4$ with harmonic mean curvature vector", *Math. Nachr.* **196** (1998), 61–69. MR 99j:53004 Zbl 0944.53005

[Fischer-Colbrie and Schoen 1980] D. Fischer-Colbrie and R. Schoen, "The structure of complete stable minimal surfaces in 3-manifolds of nonnegative scalar curvature", *Comm. Pure Appl. Math.* **33**:2 (1980), 199–211. MR 81i:53044 Zbl 0439.53060

[Gilbarg and Trudinger 1983] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, 2nd ed., Grundlehren der Mathematischen Wissenschaften **224**, Springer, Berlin, 1983. MR 86c:35035 Zbl 0562.35001

[Hasanis and Vlachos 1995] T. Hasanis and T. Vlachos, "Hypersurfaces in E^4 with harmonic mean curvature vector field", *Math. Nachr.* **172** (1995), 145–169. MR 96c:53085 Zbl 0839.53007

[Hoffman and Spruck 1974] D. Hoffman and J. Spruck, "Sobolev and isoperimetric inequalities for Riemannian submanifolds", *Comm. Pure Appl. Math.* **27** (1974), 715–727. MR 51 #1676 Zbl 0295.53025

[Jiang 1986] G. Y. Jiang, "2-harmonic isometric immersions between Riemannian manifolds", *Chinese Ann. Math. Ser. A* **7**:2 (1986), 130–144. In Chinese; An English summary appears in *Chinese Ann. Math. Ser. B* **7**:2 (1986), 255. MR 87k:53140 Zbl 0596.53046

[Jorge and Koutroufiotis 1981] L. Jorge and D. Koutroufiotis, "An estimate for the curvature of bounded submanifolds", *Amer. J. Math.* **103**:4 (1981), 711–725. MR 83d:53041b Zbl 0472.53055

[Mari and Rigoli 2010] L. Mari and M. Rigoli, "Maps from Riemannian manifolds into non-degenerate Euclidean cones", *Rev. Mat. Iberoam.* **26**:3 (2010), 1057–1074. MR 2012b:53126 Zbl 1205.53006

[Nakauchi and Urakawa 2011] N. Nakauchi and H. Urakawa, "Biharmonic hypersurfaces in a Riemannian manifold with non-positive Ricci curvature", *Ann. Global Anal. Geom.* **40**:2 (2011), 125–131. MR 2012f:53128 Zbl 1222.58010

[Ou 2010] Y.-L. Ou, "Biharmonic hypersurfaces in Riemannian manifolds", *Pacific J. Math.* **248**:1 (2010), 217–232. MR 2011i:53097 Zbl 1205.53066

[Ou and Tang 2012] Y.-L. Ou and L. Tang, "On the generalized Chen's conjecture on biharmonic submanifolds", *Michigan Math. J.* **61** (2012), 531–542.

[Pigola et al. 2005] S. Pigola, M. Rigoli, and A. G. Setti, "Maximum principles on Riemannian manifolds and applications", *Mem. Amer. Math. Soc.* **174**:822 (2005). MR 2006b:53048 Zbl 1075.58017

[Pigola et al. 2008] S. Pigola, M. Rigoli, and A. G. Setti, *Vanishing and finiteness results in geometric analysis: a generalization of the Bochner technique*, Progress in Mathematics **266**, Birkhäuser, Basel, 2008. MR 2009m:58001 Zbl 1150.53001

Received April 27, 2012. Revised September 25, 2012.

LUIS J. ALÍAS
DEPARTAMENTO DE MATEMÁTICAS
UNIVERSIDAD DE MURCIA
CAMPUS DE ESPINARDO
30100 ESPINARDO, MURCIA
SPAIN
ljalias@um.es

S. CAROLINA GARCÍA-MARTÍNEZ
DEPARTAMENTO DE MATEMÀTICAS
UNIVERSIDAD DE MURCIA
CAMPUS DE ESPINARDO
30100 ESPINARDO, MURCIA
SPAIN
sandracarolina.garcia@um.es
and
DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDADE DE SÃO PAULO
RUA DO MATÃO 1010
05508-900 SÃO PAULO, SP
BRAZIL

MARCO RIGOLI
DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ DEGLI STUDI DI MILANO
VIA SALDINI 50
I-20133 MILANO
ITALY
marco.rigoli@unimi.it

PACIFIC JOURNAL OF MATHEMATICS
Vol. 263, No. 1, 2013
dx.doi.org/10.2140/pjm.2013.263.13

HALF-COMMUTATIVE ORTHOGONAL HOPF ALGEBRAS

JULIEN BICHON AND MICHEL DUBOIS-VIOLETTE

A half-commutative orthogonal Hopf algebra is a Hopf $*$-algebra generated by the self-adjoint coefficients of an orthogonal matrix corepresentation $v = (v_{ij})$ that half commute in the sense that $abc = cba$ for any $a, b, c \in \{v_{ij}\}$. The first nontrivial such Hopf algebras were discovered by Banica and Speicher. We propose a general procedure, based on a crossed product construction, that associates to a self-transpose compact subgroup $G \subset U_n$ a half-commutative orthogonal Hopf algebra $\mathcal{A}_*(G)$. It is shown that any half-commutative orthogonal Hopf algebra arises in this way. The fusion rules of $\mathcal{A}_*(G)$ are expressed in term of those of G.

1. Introduction

The half-liberated orthogonal quantum group O_n^* were recently discovered by Banica and Speicher [2009]. These are compact quantum groups in the sense of [Woronowicz 1987], and the corresponding Hopf $*$-algebra $A_o^*(n)$ is the universal $*$-algebra presented by self-adjoint generators v_{ij} submitted to the relations making $v = (v_{ij})$ an orthogonal matrix and to the half-commutation relations

$$abc = cba, \quad a, b, c \in \{v_{ij}\}.$$

The half-commutation relations arose, via Tannaka duality, from a deep study of certain tensor subcategories of the category of partitions; see [Banica and Speicher 2009]. More examples of Hopf algebras with generators satisfying the half-commutation relations were given in [Banica et al. 2010], and the classification of "easy" orthogonal Hopf algebras (which means that the tensor category of corepresentations is spanned by partitions) with generators satisfying the half-commutation relations was very recently done in [Weber 2012].

The representation theory of O_n^* was discussed in [Banica and Vergnioux 2010], where strong links with the representation theory of the unitary group U_n were found. It followed that the fusion rules of O_n^* are noncommutative if $n \geq 3$. Moreover a matrix model $A_o^*(n) \hookrightarrow M_2(\mathcal{R}(U_n))$ was found in [Banica et al. 2011].

MSC2010: 20G42, 22C05, 16T05.
Keywords: Hopf algebras, quantum groups, compact groups.

The aim of this paper is to continue these works by a general study of what we call half-commutative orthogonal Hopf algebras: Hopf $*$-algebras generated by the self-adjoint coefficients of an orthogonal matrix corepresentation $v = (v_{ij})$ whose coefficients satisfy the previous half-commutation relations. Our main results are as follows.

(1) To any self-transpose compact subgroup $G \subset U_n$ we associate a half-commutative orthogonal Hopf algebra $\mathcal{A}_*(G)$, with $\mathcal{A}_*(U_n) \simeq A_o^*(n)$. The Hopf algebra $\mathcal{A}_*(G)$ is a Hopf $*$-subalgebra of the crossed product $\mathcal{R}(G) \rtimes \mathbb{C}\mathbb{Z}_2$, where the action of $\mathbb{Z}_2$ of $\mathcal{R}(G)$ is induced by the transposition.

(2) Conversely, any noncommutative half-commutative orthogonal Hopf algebra arises from the previous construction for some compact group $G \subset U_n$.

(3) The fusion rules of $\mathcal{A}_*(G)$ can be described in terms of those of G.

Therefore it follows from our study that quantum groups arising from half-commutative orthogonal Hopf algebras are objects that are very close from classical groups. This was suggested by the representation theory results from [Banica and Vergnioux 2010], by the matrix model found in the “easy” case in [Banica et al. 2011] and by the results of [Banica et al. 2013] where it was shown that the quantum group inclusion $O_n \subset O_n^*$ is maximal. The techniques from [Banica et al. 2013], and especially the short five lemma for cosemisimple Hopf algebras, are used in essential way here. The use of versions of the five lemma for Hopf algebras was initiated in [Andruskiewitsch and García 2009].

The paper is organized as follows. In Section 2 we fix some notation and recall the necessary background. In Section 3 we formally introduce half-commutative orthogonal Hopf algebras, and recall the early examples from [Banica and Speicher 2009; Banica et al. 2010]. Section 4 is devoted to our main construction, which associates to a self-transpose compact subgroup $G \subset U_n$ a half-commutative orthogonal Hopf algebra $\mathcal{A}_*(G)$, and we show that any half-commutative orthogonal Hopf algebra arises in this way. At the end of the section we use our construction to propose a possible orthogonal half-liberation of the unitary group U_n. In Section 5 we describe the fusion rules of $\mathcal{A}_*(G)$ in terms of those of G.

We assume that the reader is familiar with Hopf algebras [Montgomery 1993], Hopf $*$-algebras and with the algebraic approach (via algebras of representative functions) to compact quantum groups [Dijkhuizen and Koornwinder 1994; Klimyk and Schmüdgen 1997].

2. Preliminaries

Classical groups. We first fix some notation. As usual, the group of complex $n \times n$ unitary matrices is denoted by U_n, while O_n denotes the group of real orthogonal matrices. We denote by $\mathbb{T}$ the subgroup of U_n consisting of scalar matrices, and by PU_n the quotient group $U_n/\mathbb{T}$.

Definition 2.1. Let $G \subset U_n$ be a compact subgroup.

(1) We say that G is *self-transpose* if $g^t \in G$ for all $g \in G$.

(2) We say that G is *nonreal* if $G \not\subset O_n$, i.e., if there exists $g \in G$ with $g_{ij} \notin \mathbb{R}$, for some i, j.

(3) We say that G is *doubly nonreal* if there exists $g \in G$ with $g_{ij}\overline{g_{kl}} \notin \mathbb{R}$, for some i, j, k, l.

Note that the subgroup $\tilde{O}_n = \mathbb{T} O_n \subset U_n$ (considered in [Banica et al. 2013]) is nonreal but is not doubly nonreal.

Orthogonal and unitary Hopf algebras. We next recall some definitions on the algebraic approach to compact quantum groups. We work at the level of Hopf $*$-algebras of representative functions. The following simple key definition arose from [Woronowicz 1987].

Definition 2.2. A *unitary Hopf algebra* is a $*$-algebra A which is generated by elements $\{u_{ij} \mid 1 \le i, j \le n\}$ such that the matrices $u = (u_{ij})$ and $\bar{u} = (u_{ij}^*)$ are unitaries, and such that:

(1) There is a $*$-algebra map $\Delta : A \to A \otimes A$ such that $\Delta(u_{ij}) = \sum_{k=1}^n u_{ik} \otimes u_{kj}$.

(2) There is a $*$-algebra map $\varepsilon : A \to \mathbb{C}$ such that $\varepsilon(u_{ij}) = \delta_{ij}$.

(3) There is a $*$-algebra map $S : A \to A^{op}$ such that $S(u_{ij}) = u_{ji}^*$.

If $u_{ij} = u_{ij}^*$ for $1 \le i, j \le n$, we say that A is an *orthogonal Hopf algebra*.

It follows that Δ, ε, S satisfy the usual Hopf $*$-algebra axioms and that $u = (u_{ij})$ is a matrix corepresentation of A. Note that the definition forces that a unitary Hopf algebra is of Kac type, i.e., $S^2 = \mathrm{id}$. The motivating example of unitary (resp. orthogonal) Hopf algebra is $A = \mathcal{R}(G)$, the algebra of representative functions on a compact subgroup $G \subset U_n$ (resp. $G \subset O_n$). Here the standard generators u_{ij} are the coordinate functions which take a matrix to its (i, j)-entry.

In fact every commutative unitary Hopf algebra is of the form $\mathcal{R}(G)$ for some unique compact group $G \subset U_n$ defined by $G = \mathrm{Hom}_{*\text{-alg}}(A, \mathbb{C})$ (this the Hopf algebra version of the Tannaka–Krein theorem). This motivates the notation "$A = \mathcal{R}(G)$" for any unitary (resp. orthogonal) Hopf algebra, where G is a *unitary* (resp. *orthogonal*) *compact quantum group*.

The universal examples of unitary and orthogonal Hopf algebras are as follows [Wang 1995a].

Definition 2.3. The universal unitary Hopf algebra $A_u(n)$ is the universal $*$-algebra generated by elements $\{u_{ij} \mid 1 \leq i, j \leq n\}$ such that the matrices $u = (u_{ij})$ and $\bar{u} = (u_{ij}^*)$ in $M_n(A_u(n))$ are unitaries.

The universal orthogonal Hopf algebra $A_o(n)$ is the universal $*$-algebra generated by self-adjoint elements $\{u_{ij} \mid 1 \leq i, j \leq n\}$ such that the matrix $u = (u_{ij})_{1\leq i,j\leq n}$ in $M_n(A_o(n))$ is orthogonal.

The existence of the Hopf $*$-algebra structural morphisms follows from the universal properties of $A_u(n)$ and $A_o(n)$. As discussed above, we use the notations $A_u(n) = \mathcal{R}(U_n^+)$ and $A_o(n) = \mathcal{R}(O_n^+)$, where U_n^+ is the *free unitary quantum group* and O_n^+ is the *free orthogonal quantum group*.

The Hopf $*$-algebra $A_u(n)$ was introduced by Wang [1995a], while the Hopf algebra $A_o(n)$ was defined first in [Dubois-Violette and Launer 1990] under the notation $\mathcal{A}(I_n)$, and was then defined independently in [Wang 1995a] in the compact quantum group framework.

Exact sequences of Hopf algebras. In this subsection we recall some facts on exact sequences of Hopf algebras.

Definition 2.4. A sequence of Hopf algebra maps

$$\mathbb{C} \to B \xrightarrow{i} A \xrightarrow{p} L \to \mathbb{C}$$

is called preexact if i is injective, p is surjective and $i(B) = A^{\mathrm{co}\, p}$, where

$$A^{\mathrm{co}\, p} = \{a \in A \mid (\mathrm{id} \otimes p)\Delta(a) = a \otimes 1\}.$$

A preexact sequence as in Definition 2.4 is said to be exact [Andruskiewitsch and Devoto 1995] if in addition we have $i(B)^+A = \ker(p) = Ai(B)^+$, where $i(B)^+ = i(B) \cap \ker(\varepsilon)$. For the kind of sequences to be considered in this paper, preexactness is actually equivalent to exactness.

The following lemma, that we record for future use, is Proposition 3.2 in [Banica et al. 2013].

Lemma 2.5. *Let A be an orthogonal Hopf algebra with generators u_{ij}. Assume that we have surjective Hopf algebra map $p : A \to \mathbb{C}\mathbb{Z}_2$, $u_{ij} \to \delta_{ij} g$, where $< g >= \mathbb{Z}_2$. Let $P_u A$ be the subalgebra generated by the elements $u_{ij}u_{kl}$ with the inclusion $i : P_u A \subset A$. Then the sequence*

$$\mathbb{C} \to P_u A \xrightarrow{i} A \xrightarrow{p} \mathbb{C}\mathbb{Z}_2 \to \mathbb{C}$$

is preexact.

Exact sequences of compact groups induce exact sequences of Hopf algebras. In particular, if $G \subset U_n$ is a compact subgroup, we have an exact sequence of compact groups

$$1 \to G \cap \mathbb{T} \to G \to G/G \cap \mathbb{T} \to 1,$$

which induces an exact sequence of Hopf algebras

$$\mathbb{C} \to \mathcal{R}(G/G \cap \mathbb{T}) \to \mathcal{R}(G) \to \mathcal{R}(G \cap \mathbb{T}) \to \mathbb{C}.$$

We sketch a proof of the next lemma for completeness.

Lemma 2.6. *Let $G \subset U_n$ be a compact subgroup. Then $\mathcal{R}(G/G \cap \mathbb{T})$ is the subalgebra of $\mathcal{R}(G)$ generated by the elements $u_{ij}u_{kl}^*$, $i, j, k, l \in \{1, \dots, n\}$. Moreover, if $G = U_n$, then $\mathcal{R}(PU_n) = \mathcal{R}(U_n/\mathbb{T})$ is isomorphic with the commutative $*$-algebra presented by generators $w_{ij,kl}$, $1 \leq i, j, k, l \leq n$ and submitted to the relations*

$$\sum_{j=1}^{n} w_{ik,jj} = \delta_{ik} = \sum_{j=1}^{n} w_{jj,ik}, \quad w_{ij,kl}^* = w_{ji,lk},$$

$$\sum_{k,l=1}^{n} w_{ij,kl} w_{pq,kl}^* = \delta_{ip}\delta_{jq}.$$

The isomorphism is given by $w_{ij,kl} \mapsto u_{ik}u_{jl}^*$.

Proof. Let $p : \mathcal{R}(G) \to \mathcal{R}(G \cap \mathbb{T})$ be the restriction map. It is clear $\mathrm{Ker}(p)$ is generated as a $*$-ideal by the elements u_{ij}, $i \neq j$, and $u_{ii} - u_{jj}$. Let B be the subalgebra generated by the elements $u_{ij}u_{kl}^*$. Then B is a Hopf $*$-subalgebra of $\mathcal{R}(G)$ and it is clear that $B \subset \mathcal{R}(G)^{\mathrm{co}\, p}$. To prove the reverse inclusion we form the Hopf algebra quotient $\mathcal{R}(G)//B = \mathcal{R}(G)/B^+\mathcal{R}(G)$ and denote by $\rho : \mathcal{R}(G) \to \mathcal{R}(G)//B$ the canonical projection. It is not difficult to see that in $\mathcal{R}(G)//B$ we have $\rho(u_{ij}) = 0$ if $i \neq j$ and $\rho(u_{ii}) = \rho(u_{jj})$ for any i, j. Hence there exists a Hopf $*$-algebra map $p' : \mathcal{R}(G/\mathbb{T}) \to \mathcal{R}(G)//B$ such that $p' \circ p = \rho$. It follows that $\mathcal{R}(G)^{\mathrm{co}\, p} \subset \mathcal{R}(G)^{\mathrm{co}\, \rho}$. But since our algebras are commutative, $\mathcal{R}(G)$ is a faithfully flat B-module and hence by [Takeuchi 1972] (see also [Andruskiewitsch and Devoto 1995]) we have $\mathcal{R}(G)^{\mathrm{co}\, \rho} = B$, and hence $\mathcal{R}(G/G \cap \mathbb{T}) = \mathcal{R}(G)^{\mathrm{co}\, p} = B$.

The last assertion is just the reformulation of the standard fact that PU_n is the automorphism group of the $*$-algebra $M_n(\mathbb{C})$ (see, e.g., [Wang 1998]). □

3. Half-commutative Hopf algebras

We now formally introduce half-commutative orthogonal Hopf algebras. Of course the definition of half-commutativity can be given in a general context, as follows. It was first formalized, in a probabilistic context, in [Banica et al. 2012].

Definition 3.1. Let A be an algebra. We say that a family $(a_i)_{i \in I}$ of elements of A half-commute if $abc = cba$ for any $a, b, c \in \{a_i, i \in I\}$. The algebra A is said to be half-commutative if it has a family of generators that half-commute.

At a Hopf algebra level, a reasonable definition seems to be the following one.

Definition 3.2. A half-commutative Hopf algebra is a Hopf algebra A generated by the coefficients of a matrix corepresentation $v = (v_{ij})$ whose coefficients half-commute.

We will not study half-commutative Hopf algebras in this generality. A reason for this is that it is unclear if the half-commutativity relations outside of the orthogonal case are the natural ones in the categorical framework of [Banica and Speicher 2009]. Thus we will restrict to the following special case.

Definition 3.3. A half-commutative orthogonal Hopf algebra is a Hopf $*$-algebra A generated by the self-adjoint coefficients of an orthogonal matrix corepresentation $v = (v_{ij})$ whose coefficients half-commute.

The first example is the universal one, defined in [Banica and Speicher 2009].

Definition 3.4. The half-liberated orthogonal Hopf algebra $A_o^*(n)$ is the universal $*$-algebra generated by self-adjoint elements $\{v_{ij} \mid 1 \leq i, j \leq n\}$ which half-commute and such that the matrix $v = (v_{ij})_{1 \leq i,j \leq n}$ in $M_n(A_o^*(n))$ is orthogonal.

The existence of the Hopf algebra structural morphisms follows from the universal property of $A_o^*(n)$, and hence $A_o^*(n)$ is a half-commutative orthogonal Hopf algebra. We use the notation $A_o^*(n) = \mathcal{R}(O_n^*)$, where O_n^* is the *half-liberated orthogonal quantum group*. We have $\mathcal{R}(O_n^+) \twoheadrightarrow \mathcal{R}(O_n^*) \twoheadrightarrow \mathcal{R}(O_n)$, i.e., $O_n \subset O_n^* \subset O_n^+$. At $n = 2$ we have $O_2^* = O_2^+$, but for $n \geq 3$ these inclusions are strict.

Another example of half-commutative orthogonal Hopf algebra is the following one, taken from [Banica et al. 2010].

Definition 3.5. The half-liberated hyperoctahedral Hopf algebra $A_h^*(n)$ is the universal $*$-algebra generated by self-adjoint elements $\{v_{ij} \mid 1 \leq i, j \leq n\}$ which half-commute, such that $v_{ij}v_{ik} = 0 = v_{ki}v_{ji}$ for $k \neq j$, and such that the matrix $v = (v_{ij})_{1 \leq i,j \leq n}$ in $M_n(A_o^*(n))$ is orthogonal.

Again the existence of the Hopf algebra structural morphisms follows from the universal property of $A_h^*(n)$, and hence $A_h^*(n)$ is a half-commutative orthogonal Hopf algebra. See [Banica et al. 2010] and [Weber 2012] for further examples.

The following lemma will be an important ingredient in the proof of the structure theorem of half-commutative orthogonal Hopf algebras.

Lemma 3.6. *Let A be a half-commutative orthogonal Hopf algebra generated by the self-adjoint coefficients of an orthogonal matrix corepresentation $v = (v_{ij})$ whose coefficients half-commute. Then $P_v A$ is a commutative Hopf $*$-subalgebra of A. If moreover A is noncommutative then there exists a Hopf $*$-algebra map*

$p : A \to \mathbb{C}\mathbb{Z}_2$ such that for any i, j, $p(v_{ij}) = \delta_{ij}s$, where $\langle s\rangle = \mathbb{Z}_2$, that induces a preexact sequence

$$\mathbb{C} \to P_v A \xrightarrow{i} A \xrightarrow{p} \mathbb{C}\mathbb{Z}_2 \to \mathbb{C}.$$

Proof. The key observation that $P_v A$ is commutative is Proposition 3.2 in [Banica and Vergnioux 2010]. It is clear that $P_v A$ is a normal Hopf $*$-subalgebra of A, and hence we can form the Hopf $*$-algebra quotient $A//P_v A = A/A(P_v A)^+$, with $p : A \to A//P_v A$ the canonical surjection. It is not difficult to see that in $A//P_v A$ we have $p(v_{ij}) = 0$ if $i \neq j$, $p(v_{ii}) = p(v_{jj})$ for any i, j and if we put $g = p(v_{ii})$, $g^2 = 1$. So we have to prove that $g \neq 1$. If $g = 1$, then $A//P_v A$ is trivial and $p = \varepsilon$. We know from [Chirvasitu 2011] that A is faithfully flat as a $P_v A$-module (since orthogonal Hopf algebras are cosemisimple), and hence by [Schneider 1992], we have $A^{\mathrm{co}\,p} = P_v A$. So if $g = 1$ we have $A^{\mathrm{co}\,p} = P_v A = A$ and A is commutative. Thus if A is noncommutative we have $g \neq 1$, the map p satisfies the conditions in the statement and we have the announced exact sequence (Lemma 2.5). □

Remark 3.7. The previous exact sequence is cocentral. Thus it is possible, in principle, to classify the finite-dimensional half-commutative orthogonal Hopf algebras according to the scheme used in [Bichon and Natale 2011]. The classification data will involve in particular pairs (Γ, ω) formed by a finite subgroup $\Gamma \subset PU_n$ and a cocycle $\omega \in H^2(\Gamma, \mathbb{Z}_2)$, see [Bichon and Natale 2011] for details.

4. The main construction

In this section we perform our main construction that associates to any self-transpose compact subgroup $G \subset U_n$ a half-commutative orthogonal Hopf algebra $\mathcal{A}_*(G)$ and we show any half-commutative orthogonal Hopf algebra arises in this way.

We begin with a well-known lemma. We give a proof for the sake of completeness.

Lemma 4.1. *Let $G \subset U_n$ be a compact subgroup, and denote by u_{ij} the coordinate functions on G. The following assertions are equivalent.*

1. *G is self-transpose.*
2. *There is a unique involutive Hopf $*$-algebra automorphism $s : \mathcal{R}(G) \to \mathcal{R}(G)$ such that $s(u_{ij}) = u_{ij}^*$.*

Moreover if G is self-transpose the automorphism is nontrivial if and only G is nonreal.

Proof. Assume that G is self-transpose. Then we have an involutive compact group automorphism

$$\sigma : G \to G, \quad g \mapsto (g^t)^{-1} = \bar{g},$$

which induces an involutive Hopf $*$-algebra automorphism $s : \mathcal{R}(G) \to \mathcal{R}(G)$ such that $s(u_{ij}) = u_{ij}^*$. Uniqueness is obvious since the elements u_{ij} generate $\mathcal{R}(G)$

as a $*$-algebra. Conversely, the existence of s will ensure the existence of the automorphism σ since $G \simeq \mathrm{Hom}_{*\text{-alg}}(\mathcal{R}(G), \mathbb{C})$, and hence G will be self-transpose. The last assertion is immediate. $\square$

Definition 4.2. Let $G \subset U_n$ be a self-transpose nonreal compact subgroup. We denote by $\mathcal{R}(G) \rtimes \mathbb{C}\mathbb{Z}_2$ the crossed product Hopf $*$-algebra associated to the involutive Hopf $*$-algebra automorphism s of Lemma 4.1.

Recall that the Hopf $*$-algebra structure of $\mathcal{R}(G) \rtimes \mathbb{C}\mathbb{Z}_2$ is defined as follows (see, e.g., [Klimyk and Schmüdgen 1997]).

(1) As a coalgebra, $\mathcal{R}(G) \rtimes \mathbb{C}\mathbb{Z}_2 = \mathcal{R}(G) \otimes \mathbb{C}\mathbb{Z}_2$.

(2) We have $(f \otimes s^i) \cdot (g \otimes s^j) = f s^i(g) \otimes s^{i+j}$, for any $f, g \in \mathcal{R}(G)$ and $i, j \in \{0, 1\}$.

(3) We have $(f \otimes s^i)^* = s^i(f)^* \otimes s^i$ for any $f \in \mathcal{R}(G)$ and $i \in \{0, 1\}$.

(4) The antipode is given by $S(u_{ij} \otimes 1) = u_{ji}^* \otimes 1$, $S(u_{ij} \otimes s) = u_{ji} \otimes s$ (in short $S(f \otimes s^i) = s^i(S(f)) \otimes s^i$ for any $f \in \mathcal{R}(G)$ and $i \in \{0, 1\}$).

For notational simplicity we denote, for $f \in \mathcal{R}(G)$, the respective elements $f \otimes 1$ and $f \otimes s$ of $\mathcal{R}(G) \rtimes \mathbb{C}\mathbb{Z}_2$ by f and fs.

Definition 4.3. Let $G \subset U_n$ be a self-transpose compact subgroup. We denote by $\mathcal{A}_*(G)$ the subalgebra of $\mathcal{R}(G) \rtimes \mathbb{C}\mathbb{Z}_2$ generated by the elements $u_{ij}s$, where i, j range over $\{1, \ldots, n\}$.

Proposition 4.4. *Let $G \subset U_n$ be a self-transpose compact subgroup. Then $\mathcal{A}_*(G)$ is a Hopf $*$-subalgebra of $\mathcal{R}(G) \rtimes \mathbb{C}\mathbb{Z}_2$, and there exists a surjective Hopf $*$-algebra morphism*

$$\pi : A_o^*(n) \to \mathcal{A}_*(G), \quad v_{ij} \mapsto u_{ij}s.$$

Hence $\mathcal{A}_*(G)$ is a half-commutative orthogonal Hopf algebra, and is noncommutative if and only if G is doubly nonreal.

Proof. We have $(u_{ij}s)^* = su_{ij}^* = u_{ij}s$ and hence the elements $u_{ij}s$ are self-adjoint and generate a $*$-subalgebra. Moreover, using the coproduct and antipode formula, it is immediate to check that $\Delta(u_{ij}s) = \sum_k u_{ik}s \otimes u_{kj}s$ and $S(u_{ij}s) = u_{ji}s$, and hence $\mathcal{A}_*(G)$ is an orthogonal Hopf $*$-subalgebra of $\mathcal{R}(G) \rtimes \mathbb{C}\mathbb{Z}_2$. We have

$$u_{ij}su_{kl}su_{pq}s = u_{ij}u_{kl}^*u_{pq}s = u_{pq}u_{kl}^*u_{ij}s = u_{pq}su_{kl}su_{ij}s.$$

Hence the coefficients of the orthogonal matrix $(u_{ij}s)$ half-commute, and we get our Hopf $*$-algebra map $\pi : A_o^*(n) \to \mathcal{A}_*(G)$. The algebra $\mathcal{A}_*(G)$ is commutative if and only if the elements $u_{ij}s$ pairwise commute. We have $u_{ij}su_{kl}s = u_{ij}u_{kl}^*$, so $\mathcal{A}_*(G)$ is noncommutative if and only if there exist i, j, k, l with $u_{ij}u_{kl}^* \neq u_{kl}u_{ij}^*$, which precisely means that G is doubly nonreal. $\square$

The Hopf $*$-algebra $\mathcal{A}_*(G)$ is part of a natural preexact sequence.

Proposition 4.5. *Let $G \subset U_n$ be a self-transpose compact subgroup. Then there exists a Hopf $*$-algebra embedding $\mathcal{R}(G/G\cap\mathbb{T}) \hookrightarrow \mathcal{A}_*(G)$ and a preexact sequence*

$$\mathbb{C} \to \mathcal{R}(G/G\cap\mathbb{T}) \xrightarrow{j} \mathcal{A}_*(G) \xrightarrow{q} \mathbb{C}\mathbb{Z}_2 \to \mathbb{C}.$$

Proof. The map q is defined as the restriction to $\mathcal{A}_*(G)$ of the Hopf $*$-algebra map $\varepsilon \otimes \mathrm{id} : \mathcal{R}(G) \rtimes \mathbb{C}\mathbb{Z}_2 \to \mathbb{C}\mathbb{Z}_2$. Hence we have $q(u_{ij}s) = \delta_{ij}s$. Let B be the subalgebra of $\mathcal{A}_*(G)$ generated by the elements $u_{ij}su_{kl}s = u_{ij}u_{kl}^*$. It is clear that $B = \mathcal{A}_*(G)^{\mathrm{co}\,q}$, and hence we have a preexact sequence

$$\mathbb{C} \to B \xrightarrow{j} \mathcal{A}_*(G) \xrightarrow{q} \mathbb{C}\mathbb{Z}_2 \to \mathbb{C}.$$

Consider now the injective Hopf algebra map $\nu : \mathcal{R}(G) \hookrightarrow \mathcal{R}(G) \rtimes \mathbb{C}\mathbb{Z}_2$, $f \mapsto f \otimes 1$. Since $\mathcal{R}(G/G\cap\mathbb{T}) = \mathcal{R}(G)^{G\cap\mathbb{T}}$ is the subalgebra generated by the elements $u_{ij}u_{kl}^*$ (Lemma 2.6), we have $\nu(\mathcal{R}(G/G\cap\mathbb{T})) = B$, and we get our preexact sequence. □

We will prove (Theorem 4.7) that a noncommutative half-commutative orthogonal Hopf algebra is isomorphic to $\mathcal{A}_*(G)$ for some compact group $G \subset U_n$. Before this we first prove that the morphism in Proposition 4.4 is an isomorphism $A_o^*(n) \simeq \mathcal{A}_*(U_n)$. This can be seen as a consequence of the forthcoming Theorem 4.7, but the proof is less technical while it already well enlightens the main ideas.

Theorem 4.6. *We have a Hopf $*$-algebra isomorphism $A_o^*(n) \simeq \mathcal{A}_*(U_n)$.*

Proof. Let $\pi : A_o^*(n) \to \mathcal{A}_*(U_n)$ be the Hopf $*$-algebra map from Proposition 4.4, defined by $\pi(v_{ij}) = u_{ij}s$. It induces a commutative diagram of Hopf algebra maps with preexact rows

$$\begin{array}{ccccccccc}
\mathbb{C} & \longrightarrow & P_v A_o^*(n) & \xrightarrow{i} & A_o^*(n) & \xrightarrow{p} & \mathbb{C}\mathbb{Z}_2 & \longrightarrow & \mathbb{C} \\
 & & \big\downarrow{\scriptstyle \pi_|} & & \big\downarrow{\scriptstyle \pi} & & \Big\| & & \\
\mathbb{C} & \longrightarrow & \mathcal{R}(PU_n) & \xrightarrow{j} & \mathcal{A}_*(U_n) & \xrightarrow{q} & \mathbb{C}\mathbb{Z}_2 & \longrightarrow & \mathbb{C}
\end{array}$$

where the sequence on the top row is the one of Lemma 3.6 and the sequence on the lower row is the one of Proposition 4.5. The standard presentation of $\mathcal{R}(PU_n)$ (Lemma 2.6) ensures the existence of a $*$-algebra map $\mathcal{R}(PU_n) \to P_v A_o^*(n)$, $u_{ij}u_{kl}^* \mapsto v_{ij}v_{kl}$, which is clearly an inverse isomorphism for $\pi_|$. Thus we can invoke the short five lemma from [Banica et al. 2013, Theorem 3.4] to conclude that π is an isomorphism. □

A precursor for the previous isomorphism $A_o^*(n) \simeq \mathcal{A}_*(U_n)$ was the matrix model $A_o^*(n) \hookrightarrow M_2(\mathcal{R}(U_n))$ found in [Banica et al. 2011, Section 8].

Theorem 4.7. *Let A be a noncommutative half-commutative orthogonal Hopf algebra. Then there exists a self-transpose doubly nonreal compact group G with $\mathbb{T} \subset G \subset U_n$ such that $A \simeq \mathcal{A}_*(G)$.*

Proof. Let A be a noncommutative half-commutative orthogonal Hopf algebra.

Step 1. We first write a convenient presentation for A. By Lemma 3.6 there exist surjective Hopf $*$-algebra maps

$$A_o^*(n) \xrightarrow{f} A \xrightarrow{p} \mathbb{C}\mathbb{Z}_2$$

with $pf(v_{ij}) = \delta_{ij}s$. We denote by V the comodule over $A_o^*(n)$ corresponding to the matrix $v = (v_{ij}) \in M_n(A_o^*(n))$, with its standard basis $e_1, \dots, e_n$. To any linear map $\underline{\lambda} : \mathbb{C} \to V^{\otimes m}$, with

$$\underline{\lambda}(1) = \sum_{i_1,\dots,i_m} \lambda(i_1, \dots, i_m) e_{i_1} \otimes \cdots \otimes e_{i_m},$$

we associate families $X(\underline{\lambda})$ and $X'(\underline{\lambda})$ of elements of $A_o^*(n)$ defined by

$$X(\underline{\lambda}) = \Big\{ \sum_{j_1,\dots,j_m} v_{i_1j_1} \cdots v_{i_mj_m} \lambda(j_1, \dots, j_m) - \lambda(i_1, \dots, i_m)1 \;\Big|\; i_1, \dots, i_m \in \{1, \dots, n\} \Big\},$$

$$X'(\underline{\lambda}) = \Big\{ \sum_{j_1,\dots,j_m} v_{j_mi_m} \cdots v_{j_1i_1} \lambda(j_1, \dots, j_m) - \lambda(i_1, \dots, i_m)1 \;\Big|\; i_1, \dots, i_m \in \{1, \dots, n\} \Big\}.$$

These elements generate a $*$-ideal in $A_o^*(n)$, which is in fact a Hopf $*$-ideal, that we denote by $I_{\underline{\lambda}}$. We also view V as an A-comodule via f, and the map $\underline{\lambda}$ is a morphism of A-comodules if and only if $f(I_{\underline{\lambda}}) = 0$. Now given a family $\mathscr{C}$ of linear maps $\mathbb{C} \to V^{\otimes m}$, $m \in \mathbb{N}$, we denote by $I_{\mathscr{C}}$ the Hopf $*$-ideal of $A_o^*(n)$ generated by all the elements of $X(\underline{\lambda})$ and $X'(\underline{\lambda})$, $\underline{\lambda} \in \mathscr{C}$. It follows from Woronowicz Tannaka–Krein duality [Woronowicz 1988] that f induces an isomorphism $A_o^*(n)/I_{\mathscr{C}} \simeq A$ for a suitable set $\mathscr{C}$ of morphisms of A-comodules (typically $\mathscr{C}$ is a family of morphisms that generate the tensor category of corepresentations of A).

Step 2. We now construct a compact group G with $\mathbb{T} \subset G \subset U_n$. We start with a presentation $A_o^*(n)/I_{\mathscr{C}} \simeq A$ as in Step 1. The existence of the map $p : A \to \mathbb{C}\mathbb{Z}_2$ implies that for $\underline{\lambda} : \mathbb{C} \to V^{\otimes m}$, if $\underline{\lambda} \neq 0$ and $\underline{\lambda} \in \mathscr{C}$, then m is even (evaluate p on the elements of $X(\underline{\lambda})$). We associate to $\underline{\lambda} : \mathbb{C} \to V^{\otimes 2m} \in \mathscr{C}$ the following families of elements in $\mathscr{R}(U_n)$, where in each case $i_1, \dots, i_{2m}$ range over $\{1, \dots, n\}$:

$$X_1(\underline{\lambda}) = \Big\{ \sum_{j_1,\dots,j_{2m}} u_{i_1j_1} u^*_{i_2j_2} \cdots u_{i_{2m-1}j_{2m-1}} u^*_{i_{2m}j_{2m}} \lambda(j_1, \dots, j_{2m}) - \lambda(i_1, \dots, i_{2m})1 \Big\},$$

$$X_1'(\underline{\lambda}) = \Big\{ \sum_{j_1,\dots,j_{2m}} u^*_{j_1i_1} u_{j_2i_2} \cdots u^*_{j_{2m-1}i_{2m-1}} u_{j_{2m}i_{2m}} \lambda(j_1, \dots, j_{2m}) - \lambda(i_1, \dots, i_{2m})1 \Big\},$$

$$X_2(\underline{\lambda}) = \Big\{ \sum_{j_1,\dots,j_{2m}} u^*_{i_1j_1} u_{i_2j_2} \cdots u^*_{i_{2m-1}j_{2m-1}} u_{i_{2m}j_{2m}} \lambda(j_1, \dots, j_{2m}) - \lambda(i_1, \dots, i_{2m})1 \Big\},$$

$$X_2'(\underline{\lambda}) = \Big\{ \sum_{j_1,\dots,j_{2m}} u_{j_1i_1} u^*_{j_2i_2} \cdots u_{j_{2m-1}i_{2m-1}} u^*_{j_{2m}i_{2m}} \lambda(j_1, \dots, j_{2m}) - \lambda(i_1, \dots, i_{2m})1 \Big\}.$$

Now denote by $J_{\mathscr{C}}$ the $*$-ideal of $\mathscr{R}(U_n)$ generated by the elements of $X_1(\underline{\lambda})$, $X_1'(\underline{\lambda})$, $X_2(\underline{\lambda})$ and $X_2'(\underline{\lambda})$ for all the elements $\underline{\lambda} \in \mathscr{C}$. In fact $J_{\mathscr{C}}$ is a Hopf $*$-ideal and we define G to be the compact group $G \subset U_n$ such that $\mathscr{R}(G) \simeq \mathscr{R}(U_n)/J_{\mathscr{C}}$. The existence of a Hopf $*$-algebra map $\rho : \mathscr{R}(G) \to \mathbb{C}\mathbb{Z}$, $u_{ij} \mapsto \delta_{ij}t$, where t denotes a generator of $\mathbb{Z}$, is straightforward, and thus $\mathbb{T} \subset G$. Also it is easy to check the existence of a Hopf $*$-algebra map $\mathscr{R}(G) \to \mathscr{R}(G)$, $u_{ij} \mapsto u_{ij}^*$, and this shows that G is self-transpose. We have, by Proposition 4.4, a Hopf $*$-algebra map $\pi : A_o^*(n) \to \mathscr{A}_*(G)$, $v_{ij} \mapsto u_{ij}s$. It is a direct verification to check that π vanishes on $I_{\mathscr{C}}$, so induces a Hopf $*$-algebra map $\overline{\pi} : A \to \mathscr{A}_*(G)$. We still denote by v_{ij} the element $f(v_{ij})$ in A. We get a commutative diagram with preexact rows

$$\begin{array}{ccccccccc}
\mathbb{C} & \longrightarrow & P_vA & \xrightarrow{i} & A & \xrightarrow{p} & \mathbb{C}\mathbb{Z}_2 & \longrightarrow & \mathbb{C} \\
 & & \big\downarrow{\scriptstyle \overline{\pi}_|} & & \big\downarrow{\scriptstyle \overline{\pi}} & & \| & & \\
\mathbb{C} & \longrightarrow & \mathscr{R}(G/\mathbb{T}) & \xrightarrow{j} & \mathscr{A}_*(G) & \xrightarrow{q} & \mathbb{C}\mathbb{Z}_2 & \longrightarrow & \mathbb{C}
\end{array}$$

where the sequence on the top row is the one of Lemma 3.6 and the sequence on the lower row is the one of Proposition 4.5. To prove that $\overline{\pi}$ is an isomorphism, we just have, by the short five-lemma for cosemisimple Hopf algebra [Banica et al. 2013], to prove that $\overline{\pi}_| : P_vA \to \mathscr{R}(G/\mathbb{T})$ is an isomorphism. Let $J_{\mathscr{C}}'$ be the $*$-ideal of $\mathscr{R}(PU_n)$ generated by the elements of $X_1(\underline{\lambda})$, $X_1'(\underline{\lambda})$, $X_2(\underline{\lambda})$ and $X_2(\underline{\lambda})$ for all the elements $\underline{\lambda} \in \mathscr{C}$. It is clear, using the $\mathbb{Z}$-grading on $\mathscr{R}(G)$ induced by the inclusion $\mathbb{T} \subset G$ and the fact that $J_{\mathscr{C}}$ is generated by elements of degree zero, that $J_{\mathscr{C}}' = J_{\mathscr{C}} \cap \mathscr{R}(PU_n)$, so $\mathscr{R}(G/\mathbb{T}) \simeq \mathscr{R}(PU_n)/J_{\mathscr{C}}'$. But then the natural $*$-algebra map $\mathscr{R}(PU_n) \to P_vA$ (Lemma 2.6) vanishes on $J_{\mathscr{C}}'$, and hence induces a $*$-algebra map $\mathscr{R}(G/\mathbb{T}) \to P_vA$, which is an inverse for $\overline{\pi}_|$. Hence $\overline{\pi}$ is an isomorphism, and the algebra A being noncommutative, it follows from Proposition 4.4 that G is doubly nonreal. $\square$

The proof of Theorem 4.7 also provides a method to find the compact group G from the half-commutative orthogonal Hopf algebra A.

Example 4.8. On can check, by following the proof of Theorem 4.7, that the hyperoctahedral Hopf algebra $A_h^*(n)$ is isomorphic to $\mathscr{A}_*(K_n)$, where K_n is the subgroup of U_n formed by matrices having exactly one nonzero element on each column and line (with $K_n \simeq \mathbb{T}^n \rtimes S_n$).

Remark 4.9. Let $H \subset G \subset U_n$ be self-transpose compact subgroups. The inclusion $H \subset G$ induces a surjective Hopf $*$-algebra map $\mathscr{A}_*(G) \to \mathscr{A}_*(H)$, compatible with the exact sequence in Proposition 4.5. Thus if the inclusion $H \subset G$ induces an isomorphism $H/H \cap \mathbb{T} \simeq G/G \cap \mathbb{T}$, the short five lemma ensures that $\mathscr{A}_*(G) \simeq \mathscr{A}_*(H)$. In particular, $\mathscr{A}_*(U_n) \simeq \mathscr{A}_*(SU_n)$.

We now propose a tentative orthogonal half-liberation for the unitary group. In fact another possible half-liberation of U_n has already been proposed in [Bhowmick et al. 2011], using the symbol $A_u^*(n)$. We shall use the notation $A_u^{**}(n)$ for the object we construct, which is different from the one in [Bhowmick et al. 2011].

Example 4.10. Let $A_u^{**}(n)$ be the quotient of $A_u(n)$ by the ideal generated by the elements

$$abc - cba, \quad a, b, c, \in \{u_{ij}, u_{ij}^*\},$$

Then $A_u^{**}(n)$ is isomorphic with $\mathcal{A}_*(U_{2,n})$, where $U_{2,n}$ is the subgroup of U_{2n} consisting of unitary matrices of the form

$$\begin{pmatrix} A & B \\ -B & A \end{pmatrix}, \quad A, B \in M_n(\mathbb{C}),$$

and hence is a half-commutative orthogonal Hopf algebra.

Proof. Let $\omega \in \mathbb{C}$ be a primitive fourth root of unity. We start with the probably well-known surjective Hopf $*$-algebra map

$$\begin{aligned} A_o(2n) &\to A_u(n), \\ x_{i,j}, x_{n+i,n+j} &\mapsto \frac{u_{ij} + u_{ij}^*}{2}, \ i, j \in \{1, \dots, n\}, \\ x_{n+i,j} &\mapsto \frac{u_{ij} - u_{ij}^*}{2\omega}, \ i, j \in \{1, \dots, n\}, \\ x_{i,n+j} &\mapsto \frac{u_{ij}^* - u_{ij}}{2\omega}, \ i, j \in \{1, \dots, n\}, \end{aligned}$$

where $x_{i,j}$ denote the standard generators of $A_o(2n)$. It is clear that it induces a surjective Hopf $*$-algebra map $A_o^*(2n) \to A_u^{**}(n)$, and hence $A_u^{**}(n)$ is a half-commutative orthogonal Hopf algebra.

Let J be the ideal of $A_o^*(2n)$ generated by the elements

$$v_{i,j} - v_{n+i,n+j}, \ v_{n+i,j} + v_{i,n+j}, \ i, j \in \{1, \dots, n\}$$

(where $v_{i,j}$ denotes the class of x_{ij} in $A_o^*(n)$). Then J is a Hopf $*$-ideal in $A_o^*(2n)$ and the previous Hopf $*$-algebra map induces an isomorphism $A_o^*(2n)/J \simeq A_u^{**}(n)$ (the inverse sends u_{ij} to $x_{ij} + \omega x_{n+i,j}$). Now having the presentation $A_o^*(2n)/J \simeq A_u^{**}(n)$, the proof of Theorem 4.7 yields $A_u^{**}(n) \simeq \mathcal{A}_*(U_{2,n})$. □

5. Representation theory

In this section we describe the fusion rules of $\mathcal{A}_*(G)$ for any compact group G (as usual by fusion rules we mean the set of isomorphism classes of simple comodules together with the decomposition of tensor products of simple comodules into simple constituents). Thanks to Theorem 4.7, this gives a description of the fusion rules of any half-commutative orthogonal Hopf algebra.

If A is a cosemisimple Hopf algebra, we denote by $\mathrm{Irr}(A)$ the set of simple (irreducible) comodules over A. If $A = \mathcal{R}(G)$ for some compact group, then $\mathrm{Irr}(\mathcal{R}(G)) = \mathrm{Irr}(G)$, the set of isomorphism classes of irreducible representations of G. By a slight abuse of notation, for a simple A-comodule V, we write $V \in \mathrm{Irr}(A)$.

Let $G \subset U_n$ be a self-transpose compact subgroup. Recall that the transposition induces an involutive compact group automorphism

$$\sigma : G \to G, \quad g \mapsto (g^t)^{-1} = \bar{g}.$$

For $V \in \mathrm{Irr}(G)$, we denote by V^σ the (irreducible) representation of G induced by the composition with σ. If U is the fundamental n-dimensional representation of G, then $U^\sigma \simeq \bar{U}$.

We begin by recalling the description of the fusion rules for the crossed product $\mathcal{R}(G) \rtimes \mathbb{C}\mathbb{Z}_2$. See [Wang 1995b, Theorem 3.7], for example.

Proposition 5.1. *Let $G \subset U_n$ be a self-transpose compact subgroup. Then there is a bijection*

$$\mathrm{Irr}(\mathcal{R}(G) \rtimes \mathbb{C}\mathbb{Z}_2) \simeq \mathrm{Irr}(G) \amalg \mathrm{Irr}(G).$$

More precisely, if $X \in \mathrm{Irr}(\mathcal{R}(G) \rtimes \mathbb{C}\mathbb{Z}_2)$, then there exists a unique $V \in \mathrm{Irr}(G)$ with either $X \simeq V$ or $X \simeq V \otimes s$. For $V, W \in \mathrm{Irr}(G)$, we have

$$V \otimes (W \otimes s) \simeq (V \otimes W) \otimes s,$$
$$(V \otimes s) \otimes W \simeq (V \otimes W^\sigma) \otimes s,$$
$$(V \otimes s) \otimes (W \otimes s) \simeq V \otimes W^\sigma.$$

Proof. The description of the simple comodules follows in a straightforward manner from the fact that $\mathcal{R}(G) \rtimes \mathbb{C}\mathbb{Z}_2 = \mathcal{R}(G) \otimes \mathbb{C}\mathbb{Z}_2$ as coalgebras. The tensor product decompositions are obtained by using character theory; see [Woronowicz 1987] or [Klimyk and Schmüdgen 1997]. □

Remark 5.2. If $G \subset U_n$ is connected and has a maximal torus T of G contained in $\mathbb{T}^n$, it follows from highest weight theory that $V^\sigma \simeq \overline{V}$ for any $V \in \mathrm{Irr}(G)$. We do not know if this is still true without these assumptions.

To express the fusion rules of $\mathcal{A}_*(G)$, we need more notation. Let $G \subset U_n$ be a compact subgroup, and denote by U the fundamental n-dimensional representation of G. For $m \in \mathbb{Z}$, we put

$$\mathrm{Irr}(G)_{[m]} = \{V \in \mathrm{Irr}(G),\ V \subset U^{\otimes m} \otimes (U \otimes \overline{U})^{\otimes l} \text{ for some } l \in \mathbb{N}\},$$

where $U^{\otimes 0} = \mathbb{C}$ and for $m < 0$ $U^{\otimes m} = \overline{U}^{\otimes -m}$.

Now if $V \in \mathrm{Irr}(G)_{[0]}$, then $V \in \mathrm{Irr}(G/G \cap \mathbb{T})$ (see Lemma 2.6), and since $\mathcal{R}(G/G \cap \mathbb{T}) \subset \mathcal{A}_*(G)$, we get an element in $\mathrm{Irr}(\mathcal{A}_*(G))$, still denoted V.

If $V \in \mathrm{Irr}(G)_{[1]}$, then $V \subset U \otimes (U \otimes \bar{U})^{\otimes l}$, for some $l \in \mathbb{N}$, and hence the coefficients of $V \otimes s$ belong to $\mathcal{A}_*(G)$. Thus we get an element of $\mathrm{Irr}(\mathcal{A}_*(G))$, denoted Vs.

Corollary 5.3. *Let $G \subset U_n$ be a self-transpose compact subgroup. Then the map*

$$\mathrm{Irr}(G)_{[0]} \amalg \mathrm{Irr}(G)_{[1]} \to \mathrm{Irr}(\mathcal{A}_*(G))$$

given by

$$V \mapsto \begin{cases} V & \text{if } V \in \mathrm{Irr}(G)_{[0]}, \\ Vs & \text{if } V \in \mathrm{Irr}(G)_{[1]}, \end{cases}$$

is a bijection. Moreover, for $V \in \mathrm{Irr}(G)_{[0]}$, $W, W' \in \mathrm{Irr}(G)_{[1]}$, we have

$$\begin{aligned} V \otimes Ws &\simeq (V \otimes W)s, \\ Ws \otimes V &\simeq (W \otimes V^\sigma)s, \\ Ws \otimes W's &\simeq W \otimes W'^\sigma, \\ \overline{Ws} &\simeq \overline{W}^\sigma s. \end{aligned}$$

Proof. The existence of the map follows from the discussion before the corollary, while injectivity comes from Proposition 5.1. For $V \in \mathrm{Irr}(G)_{[m]}$, $V' \in \mathrm{Irr}(G)_{[m']}$, the simple constituents of $V \otimes V'$ all belong to $\mathrm{Irr}(G)_{[m+m']}$, and that $V^\sigma \in \mathrm{Irr}(G)_{[-m]}$. So the isomorphisms in the statement (that all come from the isomorphisms of Proposition 5.1) yield decompositions into simple $\mathcal{A}_*(G)$-comodules. Thus we have a family of simple $\mathcal{A}_*(G)$-comodules, stable under decompositions of tensor products and conjugation, and that contains the fundamental comodule Us: we conclude (e.g., from the orthogonality relations [Woronowicz 1987; Klimyk and Schmüdgen 1997]) that we have all the simple comodules. □

References

[Andruskiewitsch and Devoto 1995] N. Andruskiewitsch and J. Devoto, "Extensions of Hopf algebras", *Algebra i Analiz* **7**:1 (1995), 22–61. In Russian; translated in *St. Petersburg Math. J.* **7**:1 (1996), 17–52. MR 96f:16044 Zbl 0857.16032

[Andruskiewitsch and García 2009] N. Andruskiewitsch and G. A. García, "Quantum subgroups of a simple quantum group at roots of one", *Compos. Math.* **145**:2 (2009), 476–500. MR 2010a:20102 Zbl 1236.17019

[Banica and Speicher 2009] T. Banica and R. Speicher, "Liberation of orthogonal Lie groups", *Adv. Math.* **222**:4 (2009), 1461–1501. MR 2010j:46125 Zbl 05614878

[Banica and Vergnioux 2010] T. Banica and R. Vergnioux, "Invariants of the half-liberated orthogonal group", *Ann. Inst. Fourier* (*Grenoble*) **60**:6 (2010), 2137–2164. MR 2012f:46146 Zbl 05849155

[Banica et al. 2010] T. Banica, S. Curran, and R. Speicher, "Classification results for easy quantum groups", *Pacific J. Math.* **247**:1 (2010), 1–26. MR 2011k:46099 Zbl 1208.46068

[Banica et al. 2011] T. Banica, S. Curran, and R. Speicher, "Stochastic aspects of easy quantum groups", *Probab. Theory Related Fields* **149**:3-4 (2011), 435–462. MR 2012e:60015 Zbl 1236.60008

[Banica et al. 2012] T. Banica, S. Curran, and R. Speicher, "De Finetti theorems for easy quantum groups", *Ann. Probab.* **40**:1 (2012), 401–435. MR 2917777 Zbl 1242.46073

[Banica et al. 2013] T. Banica, J. Bichon, B. Collins, and S. Curran, "A maximality result for orthogonal quantum groups", *Comm. Algebra* **41**:2 (2013), 656–665. arXiv 1106.5467

[Bhowmick et al. 2011] J. Bhowmick, F. D'Andrea, and L. Dabrowski, "Quantum isometries of the finite noncommutative geometry of the standard model", *Comm. Math. Phys.* **307**:1 (2011), 101–131. MR 2835874 Zbl 1236.81191

[Bichon and Natale 2011] J. Bichon and S. Natale, "Hopf algebra deformations of binary polyhedral groups", *Transform. Groups* **16**:2 (2011), 339–374. MR 2012g:16066 Zbl 1238.16024

[Chirvasitu 2011] A. Chirvasitu, "Cosemisimple Hopf algebras are faithfully flat over Hopf subalgebras", preprint, 2011. arXiv 1110.6701

[Dijkhuizen and Koornwinder 1994] M. S. Dijkhuizen and T. H. Koornwinder, "CQG algebras: a direct algebraic approach to compact quantum groups", *Lett. Math. Phys.* **32**:4 (1994), 315–330. MR 95m:16029 Zbl 0861.17005

[Dubois-Violette and Launer 1990] M. Dubois-Violette and G. Launer, "The quantum group of a nondegenerate bilinear form", *Phys. Lett. B* **245**:2 (1990), 175–177. MR 91j:16049 Zbl 1119.16307

[Klimyk and Schmüdgen 1997] A. Klimyk and K. Schmüdgen, *Quantum groups and their representations*, Springer, Berlin, 1997. MR 99f:17017 Zbl 0891.17010

[Montgomery 1993] S. Montgomery, *Hopf algebras and their actions on rings*, CBMS Regional Conference Series in Mathematics **82**, Amer. Math. Soc., Providence, RI, 1993. MR 94i:16019 Zbl 0793.16029

[Schneider 1992] H.-J. Schneider, "Normal basis and transitivity of crossed products for Hopf algebras", *J. Algebra* **152**:2 (1992), 289–312. MR 93j:16032 Zbl 0789.16026

[Takeuchi 1972] M. Takeuchi, "A correspondence between Hopf ideals and sub-Hopf algebras", *Manuscripta Math.* **7** (1972), 251–270. MR 48 #328 Zbl 0238.16011

[Wang 1995a] S. Wang, "Free products of compact quantum groups", *Comm. Math. Phys.* **167**:3 (1995), 671–692. MR 95k:46104 Zbl 0838.46057

[Wang 1995b] S. Wang, "Tensor products and crossed products of compact quantum groups", *Proc. London Math. Soc.* (3) **71**:3 (1995), 695–720. MR 96i:46074 Zbl 0837.46052

[Wang 1998] S. Wang, "Quantum symmetry groups of finite spaces", *Comm. Math. Phys.* **195**:1 (1998), 195–211. MR 99h:58014 Zbl 1013.17008

[Weber 2012] M. Weber, "On the classification of easy quantum groups: the nonhyperoctahedral and the half-liberated case", preprint, 2012. arXiv 1201.4723

[Woronowicz 1987] S. L. Woronowicz, "Compact matrix pseudogroups", *Comm. Math. Phys.* **111**:4 (1987), 613–665. MR 88m:46079 Zbl 0627.58034

[Woronowicz 1988] S. L. Woronowicz, "Tannaka–Kreĭn duality for compact matrix pseudogroups: twisted SU(N) groups", *Invent. Math.* **93**:1 (1988), 35–76. MR 90e:22033 Zbl 0664.58044

Received February 28, 2012.

JULIEN BICHON
LABORATOIRE DE MATHÉMATIQUES
UNIVERSITÉ BLAISE PASCAL
CAMPUS DES CÉZEAUX BP 80026
63171 AUBIÈRE CEDEX
FRANCE

Julien.Bichon@math.univ-bpclermont.fr

MICHEL DUBOIS-VIOLETTE
LABORATOIRE DE PHYSIQUE THÉORIQUE D'ORSAY
UNIVERSITÉ PARIS-SUD
BÂTIMENT 210
91405 ORSAY CEDEX
FRANCE

Michel.Dubois-Violette@u-psud.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 263, No. 1, 2013
dx.doi.org/10.2140/pjm.2013.263.29

SUPERDISTRIBUTIONS, ANALYTIC AND ALGEBRAIC SUPER HARISH-CHANDRA PAIRS

CLAUDIO CARMELI AND RITA FIORESI

We extend the theory of super Harish-Chandra pairs, originally developed by Kostant and Koszul for smooth Lie supergroups, to algebraic supergroups over a field of characteristic zero. We also review the corresponding complex analytic theory and we give a characterization of the action of an algebraic (resp. complex analytic) super Harish-Chandra pair on a supervariety (resp. complex analytic supermanifold).

1. Introduction

The main purpose of this paper is to extend the theory of super Harish-Chandra pairs, originally developed by Kostant [1977] and Koszul [1983] for smooth Lie supergroups, to algebraic supergroups, enlightening similarities and differences with the complex analytic setting, treated in detail by Vishnyakova [2011]. This approach appears to be especially fruitful in the study of algebraic supergroup representations and more in general supergroup actions on supervarieties.

Roughly speaking, a super Harish-Chandra pair (SHCP for short) consists of a pair $(G_0, \mathfrak{g})$, where G_0 is an ordinary algebraic (resp. analytic or smooth) supergroup and $\mathfrak{g}$ is a Lie superalgebra, with even part $\mathfrak{g}_0 = \mathrm{Lie}(G_0)$. If G is a supergroup (algebraic, analytic or differential), we have a natural SHCP associated with it: $(G_0, \mathrm{Lie}(G))$. What appears to be surprising is the fact that the correspondence between supergroups and SHCP is bijective (up to isomorphism), i.e., starting from a given SHCP $(G_0, \mathfrak{g})$, we can reconstruct a supergroup, which has a corresponding SHCP $(G_0, \mathrm{Lie}(G)) = (G_0, \mathfrak{g})$, and such supergroup is unique. Actually more is true: there is an equivalence of categories between the category of supergroups (algebraic, analytic or differential) and the category of SHCPs (algebraic, analytic or differential), once morphisms are properly defined.

Such equivalence in the smooth context dates back to [Koszul 1983], while the analytic setting is due to Vishnyakova [2011], though a careful reading of [Koszul 1983], shows that the complex theory appeared already, somehow implicitly, in that

MSC2010: primary 14A22, 14M30, 32C11, 58A50; secondary 16S38, 58C50.
Keywords: supergroups, supergeometry.

paper. Vishnyakova applied the result about the equivalence of categories between analytic supergroups and analytic SHCPs to provide a characterization of those complex homogeneous analytic supermanifolds that are split. We take her work a step forward: we characterize the concept of action of an analytic SHCP on an analytic supermanifold, proving it is equivalent to the ordinary notion of action of an analytic super Lie group on an analytic supermanifold. Our result, which is novel, immediately carries over to the affine algebraic category.

After our paper appeared on the web on June 2011, Masuoka [2012] published a more general and very interesting result in which he quoted our work, giving us the credit for being the first authors to treat the algebraic setting for the equivalence of categories between algebraic supergroups and algebraic SHCPs in characteristic zero. Masuoka is able to obtain a generalization of our result through a characteristic free approach, in purely algebraic terms.

In his paper, Masuoka defines a category of SHCPs whose objects are pairs consisting of an Hopf algebra C and a finite dimensional right C–comodule W, together with appropriate compatibility conditions. In the characteristic zero case, the category of Masuoka's SHCPs is anti-isomorphic to the algebraic SHCP category we use in the present paper. He then establishes an equivalence between the category of such SHCPs (C, W) and the category of affine (i.e., super commutative and finitely generated) Hopf superalgebras, which in turn is contravariantly equivalent to the category of affine algebraic supergroups. The functor establishing such an equivalence associates to each pair (C, W) a subalgebra $A(C, W)$ of the completion of the smash product Hopf algebra $C \times' T(W)$ (here $T(W)$ denotes the tensor algebra of W). In this sense, Masuoka's approach seems more related to Kostant's proof of the categorical equivalence between smooth SHCPs and smooth super Lie groups. Indeed in his approach Kostant realizes the structure sheaf of the supergroup as a subalgebra of the algebraic dual of the smash product $\mathbb{R}[G_0] \times' \mathcal{U}(\mathfrak{g})$. We believe that the importance of Koszul's approach relies in the simple geometrical realization of the sheaf as the coinduced module

$$\mathrm{Hom}_{\mathcal{U}(\mathfrak{g}_0)}(\mathcal{U}(\mathfrak{g}), \mathcal{O}_{G_0}(G_0)),$$

which is very explicit. This is particularly important when one tries to deduce general properties of super Lie groups (see, for instance, the characterization of split homogeneous supermanifold in [Vishnyakova 2011], or our Proposition 4.3). Moreover, as far as we understand, it is still an open problem to establish whether the correspondence between SHCPs as we define them and algebraic supergroups is an equivalence of categories in the positive characteristic case.

Since our methods are essentially different from Masuoka's and present a geometric point of view particularly useful for the applications (see our Section 4), we believe that our work still deserves a place in the literature.

Our treatment begins with the definition of *distribution superalgebra*. We keep our discussion general enough to accommodate both the analytic and algebraic category and we believe this is one of the strengths of our paper and it singles it out from the previous treatments of the same subject we quoted above, which usually deal with just one category (algebraic, analytic or differential) at a time. The distribution superalgebra is a key object; its definition in differential supergeometry dates back to Kostant [1977], who first recognized its importance in this context. As we show in our work, the distribution superalgebra $D(G)$ of a supergroup G (algebraic, analytic or differential) is naturally equipped with a Hopf superalgebra structure and it is indeed this Hopf structure, which makes possible the reconstruction of the algebraic, analytic or differentiable supergroup associated with an SHCP. In fact, when the characteristic of the ground field k is zero, $D(G)$ is linearly isomorphic to $k|G| \otimes \mathcal{U}(\mathfrak{g})$ ($k|G|$ denoting the ordinary group algebra associated with the topological group $|G|$ underlying the supergroup G). This allows us to endow $k|G| \otimes \mathcal{U}(\mathfrak{g})$ with an Hopf superalgebra structure, inherited by $D(G)$ via the above mentioned linear isomorphism. The superalgebra of the global sections of the structural sheaf of the algebraic supergroup G, associated (uniquely) with the given SHCP ($|G|$, $\mathfrak{g}$), is then realized inside the dual of $k|G| \otimes \mathcal{U}(\mathfrak{g})$, thus inheriting its Hopf structure. This is essentially the reason why the above mentioned equivalence of categories works, though the proofs and the statements are necessarily more complicated, since of the technicalities involved, which at this point differ depending on the category we consider, for example for the analytic category we cannot take into consideration the global sections only, but we need to look at the whole sheaf.

This paper is organized as follows.

In Section 2 we describe the superalgebra of distributions of an analytic or an algebraic supergroup, establishing its relation with the universal enveloping superalgebra. The material exposed here is general common knowledge, though we are not aware of a treatment as complete and general as ours.

Section 3 contains the main results of our paper, including Theorem 3.6, which establishes the equivalence of categories between SHCPs (algebraic or analytic) and supergroups (algebraic or analytic). For the reader's convenience, this is preceded (starting on page 39) by a brief review of the equivalence between the category of analytic SHCPs and the category of analytic supergroups.[1] Subsequently (page 42) we establish the equivalence between the category of algebraic SHCPs and the category of affine algebraic supergroups under suitable hypothesis for the ground field. The results of this section were generalized in [Masuoka 2012], with totally different methods, posted on the web at a later date than ours.

In Section 4 we provide an equivalent approach to the study of the actions of

[1]The material of this section appeared already, essentially in this form, in [Vishnyakova 2011].

supergroups, via SHCPs. This result extends the result stated in [Deligne and Morgan 1999] for the smooth category (see also [Balduzzi et al. 2009; Carmeli et al. 2011]). These results are novel as far as we know.

We believe the present work is justified, given the importance of the algebraic theory for practical purposes together with the lack of an appropriate and complete available reference.

For all the definitions and main results in supergeometry expressed with our notation, we refer the reader to [Fioresi and Gavarini 2011] or [Fioresi and Gavarini 2012, Chapter 2] or [Carmeli et al. 2011, Chapters 1, 4, 10]. In particular we shall employ both the sheaf-theoretic and the functor of points approach to supergeometry. On this we invite the reader to consult the classical references [Deligne and Morgan 1999; Manin 1988; Varadarajan 2004].

2. The superalgebra of distributions

We start by giving the definition of distribution and distribution superalgebra. Our treatment is general enough to accommodate the two very different categories of supermanifolds and superschemes. For the classical definitions we send the reader to [Jantzen 2003, page 95], [Demazure and Gabriel 1970, Chapter II §4, no. 6], and [Dieudonné 1970]. For the basic definitions of supergeometry we refer the reader to [Manin 1988; Varadarajan 2004; Deligne and Morgan 1999; Fioresi and Gavarini 2012].

Distributions. Let k be the ground field.

Let $X = (|X|, \mathcal{O}_X)$ be an analytic supermanifold or an algebraic superscheme over the field k.[2]

Let $X(k)$ be the k-points of X, that is $X(k) = \mathrm{Hom}(k^{0|0}, X)$ in the functor of points notation. For an analytic supermanifold X we have that its k-points $X(k)$ are identified with the topological points $|X|$, while for X a superscheme the k-points, are in one to one correspondence with the rational points, that is, the points $x \in |X|$ for which $\mathcal{O}_{X,x}/m_{X,x} \cong k$, $m_{X,x}$ being the maximal ideal in the stalk $\mathcal{O}_{X,x}$.

Definition 2.1. A *distribution supported at* $x \in X(k)$ *of order at most* n is a morphism $\phi : \mathcal{O}_{X,x} \to k$, with $m_{X,x}^{n+1} \subset \ker(\phi)$ for some n. The set of all distributions at x of order n is denoted as $D_n(X, x)$, while $D(X, x)$ denotes all distributions supported at x. Both $D_n(X, x)$ and $D(X, x)$ have a natural super vector space structure.

We also define

$$D(X) = \bigcup_{x \in X(k)} D(X, x)$$

[2] If X is an analytic supermanifold, $k = \mathbb{R}$ or $k = \mathbb{C}$ or even $k = \mathbb{Q}_p$, the p-adic numbers (see for example [Serre 1992]). If X is a superscheme, k is a generic field.

as the *distributions of finite order* of X. Also $D(X)$ has a natural super vector space structure.

Observation 2.2. (1) We have

$$D_n(X, x) \cong (\mathcal{O}_{X,x}/m_{X,x}^{n+1})^*,$$

since if $\phi \in D_n(X, x)$, we have $\phi(m_{X,x}^{n+1}) = 0$; hence ϕ factors and becomes an element in $(\mathcal{O}_{X,x}/m_{X,x}^{n+1})^*$. Further notice that

$$D_0(X, x) = k, \quad D_1(X, x) = k \oplus (m_{X,x}/m_{X,x}^2)^*.$$

Hence $D_1(X, x)^+ := (m_{X,x}/m_{X,x}^2)^*$ becomes identified with the tangent space to X at the point x.

(2) If X is an affine algebraic superscheme, $\mathcal{O}(X)$ the superalgebra of the global sections of its structural sheaf, a distribution supported at x of order n can be equivalently seen as a morphism $\phi : \mathcal{O}(X) \to k$, with $m_x^n \subset \ker(\phi)$, where $m_x := \{\phi \in \mathcal{O}(X) \mid \phi(x) = 0\}$ is the maximal ideal of all the functions vanishing at x, where as usual in supergeometry $f(x)$ simply means the image in $\mathcal{O}_{X,x}/m_{X,x}$ of the element $f \in \mathcal{O}(X)$ under the natural morphisms: $\mathcal{O}(X) \to \mathcal{O}_{X,x} \to \mathcal{O}_{X,x}/m_{X,x} \cong k$. (Notice that since x is rational, we have $\mathcal{O}(X) = k \oplus m_x$ and $\mathcal{O}_{X,x}/m_{X,x} \cong k$).

We leave it to the reader to check that the two definitions of distributions given are essentially the same in this case.

(3) If X is a smooth supermanifold, that is, if we are in the differential category, we can view a point supported distribution as a morphism $\phi : \mathcal{O}(X) \to \mathbb{R}$, $m_x^n \subset \ker(\phi)$, where m_x is the maximal ideal corresponding to the point $x \in |X|$ (see [Kostant 1977] and [Carmeli et al. 2011, 4.7]), thus recovering the same definition as in (2) for the affine algebraic category. This is one of the many analogies between the category of affine supervarieties and smooth supermanifolds.

Example 2.3 (distributions on $k^{p|q}$). Here we assume $\mathrm{char}(k) = 0$. Consider the superspace $X = k^{p|q}$ (both in the analytic and affine algebraic context). Let $x_1 \dots x_p$, $\xi_1 \dots \xi_q$ denote the global coordinates and $m_0 = (x_1 \dots x_p, \xi_1 \dots \xi_q)$ the maximal ideal in the stalk $\mathcal{O}_{X,0}$ at the origin. We have

$$\mathcal{O}_{X,0}/m_0^{n+1} \cong \mathrm{span}_k\big\{1,\ x_1^{i_1} \dots x_p^{i_p} \xi_1^{i_{p+1}} \dots \xi_q^{i_{p+q}},\ \textstyle\sum i_k = n\big\}.$$

If $I = (i_1 \dots i_{p+q})$, let X^I denote the monomial $x_1^{i_1} \dots x_p^{i_p} \xi_1^{i_{p+1}} \dots \xi_q^{i_{p+q}}$. Since the distributions at 0 of order n are the dual of the super vector space $\mathcal{O}_{X,0}/m_0^{n+1}$, we have that a basis for the super vector space of distributions at the point 0 is given by ϕ_J such that $\phi_J(X_I) = \delta_{IJ}$, with $I = (i_1 \dots i_{p+q})$, $J = (j_1 \dots j_{p+q})$ multiindices,

$\sum i_k = \sum j_k = n$. So we have

$$\phi_{j_1 \dots j_{p+q}}(f) = \frac{1}{j_1! \dots j_{p+q}!} \left(\frac{\partial}{\partial x_1}\right)^{j_1} \dots \left(\frac{\partial}{\partial x_p}\right)^{j_p} \left(\frac{\partial}{\partial \xi_q}\right)^{j_{p+1}} \dots \left(\frac{\partial}{\partial \xi_1}\right)^{j_{p+q}} (f)(0).$$

The superalgebra of distributions of an analytic supermanifold. In this section we characterize the distributions for an analytic supermanifold $M = (|M|, \mathcal{O}_M)$ in the following way. Distributions at the point $x \in |M|$ are the elements in $\mathcal{O}^*_{M,x}$ whose kernel contains an ideal of finite codimension, in analogy with Kostant's treatment [1977] for the smooth category. We start with a lemma.

Lemma 2.4. *Let $M = (|M|, \mathcal{O}_M)$ be an analytic supermanifold, $x \in |M|$, $m_{X,x}$ the ideal in $\mathcal{O}_{M,x}$ of the sections vanishing at x. For each positive integer p, $m^p_{X,x}$ is an ideal of finite codimension.*

Proof. It follows from the Taylor expansion formula. In fact, every element f in $\mathcal{O}_{M,x}$ can be written as $f = \sum_I f_I \theta^I$, where f_I is an element in the classical stalk of germs of holomorphic functions $\mathcal{H}_{M,x}$. For each positive integer q, a germ f_I can in turn be written as

$$f_I(z) = f_I(x) + \sum_{K\,:\,1 \leq |K| \leq q-1} (\partial_K f_I)(x) z^K + \sum_{J\,:\,|J|=q} z^J h_{I,J}(z)$$

where I, J, K are multiindices. Hence we can write

$$f = \sum_I \left(f_I(x) + \sum_{R\,:\,|R+I|<p} (\partial_R f_I)(x) z^R \right) \theta^I + \sum_{|I+R|=p} h_{I,R}(z) z^R \theta^I.$$

From this formula, it follows that the elements in $m^p_{X,x}$ are generated by the monomials $\{z^K \theta^I\}_{|K+I| \leq p}$, and $\mathcal{O}_{M,x}/m^p_{M,x}$ has finite dimension. $\square$

Proposition 2.5. *An ideal J in $\mathcal{O}_{M,x}$ has finite codimension if and only if there exists an integer $p > 0$ such that $m^p_{M,x} \subseteq J$.*

Proof. The "if" part follows from the previous lemma. For the "only if" part we reason as follows. Consider the descending chain of ideals $J + m^p_{M,x} \supseteq J + m^{p+1}_{M,x}$. Since J has finite codimension there exists q such that $J + m^q_{M,x} = J + m^{q+1}_{M,x}$. From this it follows that $m^q_{M,x} \subseteq J + m^q_{M,x} \cdot m_{M,x}$. Since, by the previous lemma, $m^q_{M,x}$ is finitely generated we can apply the super version of Nakayama lemma (see [Varadarajan 2004]) and we get $m^q_{M,x} \subseteq J$. $\square$

We have then obtained the following result, which establishes a parallelism with the smooth category.

Theorem 2.6. *The distributions on an analytic supermanifold M supported at a point x correspond to morphisms $f : \mathcal{O}_{M,x} \to k$ whose kernel contains an ideal of finite codimension.*

The distributions of a supergroup at the identity. We now want to restrict our attention to the distributions of a supergroup (analytic or algebraic) at the identity element $e \in G(k)$.

As a consequence of the Observation 2.2, we have

$$D_1(G, e)^+ \cong (m_{G,e}/m^2_{G,e})^* \cong T_e(G) = \mathrm{Lie}(G).$$

It is only natural to expect $D(G, e)$ to be identified with $\mathcal{U}(\mathfrak{g})$, with $\mathfrak{g} = \mathrm{Lie}(G)$. This is true, as we shall see, provided we exert some care.

As we remarked in the Definition 2.1 the distributions at the identity are a super vector space, however there is a natural additional superalgebra structure that we can associate to the super vector space of distributions, by defining the *convolution product*.

Definition 2.7. Let ϕ, $\psi \in D(G, e)$. We define their *convolution product* as the following morphism:

$$(\phi \star \psi)(f) = (\phi \otimes \psi)\mu^*(f), \quad f \in \mathcal{O}_{G,e}$$

where μ denotes the multiplication in the supergroup G and μ^* the corresponding sheaf morphism.

The following proposition is a straightforward check.

Proposition 2.8. *The convolution product makes $D(G, e)$ into an associative superalgebra, its unit being the evaluation at e, denoted by* $\mathrm{ev}_e : \mathcal{O}_{G,e} \to k$.

We now want to examine the relation of $D(G, e)$ with the universal enveloping superalgebra of the supergroup G. Since $D(G, e) \supset D_1(G, e)^+ \cong \mathrm{Lie}(G)$, by the universal property of the universal enveloping superalgebra $\mathcal{U}(\mathfrak{g})$, we have a superalgebra morphism $\alpha : \mathcal{U}(\mathfrak{g}) \to D(G, e)$.

Observation 2.9. If G is an algebraic supergroup and the characteristic of k is positive, say $\mathrm{char}(k) = p > 0$, then $D(G, e)$ contains more than the elements coming from $\mathcal{U}(\mathfrak{g})$ (refer to Example 2.3). This is because the divided powers $X^m/m!$ are in $D(G, e)$ but not in $\mathcal{U}(\mathfrak{g})$. Again similarly, as in the classical situation, we have that any morphism $\mathcal{U}(\mathfrak{g}) \to D(G, e)$ factors via the *universal enveloping restricted algebra* $\mathcal{U}^r(\mathfrak{g})$:

$$\mathcal{U}(\mathfrak{g}) \to \mathcal{U}^r(\mathfrak{g}) = \mathcal{U}(\mathfrak{g})/(X^p - X^{[p]}) \to D(G, e)$$

where $X^{[p]}$ denotes the derivation in $\mathfrak{g}$ corresponding to p-times the derivation X (which is a derivation here, since we are in characteristic p).

Let $\mathrm{char}(k) = 0$.

Proposition 2.10. *The morphism $\alpha : \mathcal{U}(\mathfrak{g}) \to D(G, e)$ is an isomorphism.*

Proof. This is done essentially in the same way as in the classical setting, which is detailed in [Varadarajan 2004, Chapter I] for the analytic category and [Demazure and Gabriel 1970, Chapter II, 6, 1.1] for the algebraic category. □

Proposition 2.11. *There is an isomorphism of the superalgebra of distributions on a supergroup G and the superalgebra of the left-invariant differential operators on G. In this situation $\mathcal{U}(\mathfrak{g})$ is isomorphic to the superalgebra of the left-invariant differential operators on G.*

Proof. The same remarks as in the previous proof apply. □

The distributions of an affine algebraic supergroup. We now want to restrict ourselves to the case of affine algebraic supergroups. As we shall see, this algebraic setting shares many similarities with the differential one.

Consider the module of distributions $D(G)$ (see Observation 2.2):

$$D(G) = \bigcup_{x \in G(k)} D(G, x) \subset \mathcal{O}(G)^*.$$

Definition 2.12. If $\phi = \sum \phi_{p_i}$ is a distribution with $\phi_{p_i} \in D(G, p_i)$ we say that ϕ is *supported* at $\{p_i\}$. On the whole $D(G)$ we have a well-defined associative product, called the *convolution product*:

$$(\phi_p \star \phi_q)(f) = (\phi_p \otimes \phi_q)\mu^*(f)$$

and its unit is ev_e, the evaluation at the unit element: $\mathrm{ev}_e(f) = f(e)$. Here μ^* denotes (as before) the comultiplication in the Hopf superalgebra $\mathcal{O}(G)$.

Observation 2.13. If ϕ_p and ϕ_q are distributions supported at p and q respectively, then $\phi_p \star \phi_q$ is supported at pq. This is a consequence of the fact that

$$\mu^*(m_{pq}) \subset m_p \otimes \mathcal{O}(G) + \mathcal{O}(G) \otimes m_q$$

where m_x is as usual the maximal ideal of the sections in $\mathcal{O}(G)$ vanishing at $x \in G(k)$. $m_x = m_{x,0} + J_{O(G)}$, that is, m_x is the sum of $m_{x,0}$ the ordinary maximal ideal corresponding to the topological rational point $x \in G(k)$ and the ideal $J_{\mathcal{O}(G)}$ generated by the odd sections in $\mathcal{O}(G)$.

Lemma 2.14. *Let $\phi_g \in D(G, g)$. Then there exists a unique $\phi_e \in D(G, e)$ such that $\phi_e = \mathrm{ev}_{g^{-1}} \star \phi_g$.*

Proof. Since $\phi_g = (\mathrm{ev}_g \star \mathrm{ev}_{g^{-1}}) \star \phi_g$, define $\phi_e = \mathrm{ev}_{g^{-1}} \star \phi_g \in D(G, e)$. □

Proposition 2.15. *$D(G)$ is a super Hopf algebra with comultiplication Δ, counit ϵ and antipode S given by*

$$\Delta(\phi_g)(f \otimes g) := \phi_g(f \cdot g), \quad \epsilon(\phi_g)(f) := \phi_g(\mathrm{ev}_e(f)), \quad S(\phi_g)(f) := \phi_g(i^*(f)),$$

where $i : G \to G$ denotes the inverse morphism.

Proof. Direct check. □

Let $k|G|$ be the group algebra corresponding to the ordinary group $G(k)$, i.e.,

$$k|G| = \Big\{ \sum_{\substack{g\in G(k)\\ \lambda_g\in k}} \lambda_g g \Big\}.$$

Proposition 2.16. *We have a linear isomorphism*

$$\Psi : D(G) \to k|G| \otimes \mathcal{U}(g), \quad \phi_g \mapsto g \otimes \phi_e,$$

which endows $k|G| \otimes \mathcal{U}(\mathfrak{g})$ of a Hopf superalgebra structure. This structure is induced by the natural Hopf structures on the group algebra $k|G|$ and $\mathcal{U}(\mathfrak{g})$:

$$\Delta_{k|G|}(g) = g \otimes g, \quad \Delta_{\mathcal{U}(\mathfrak{g})}(U) = U \otimes 1 + 1 \otimes U, \quad g \in G(k), U \in \mathfrak{g}.$$

The superalgebra structure is defined by

$$(g \otimes X)(h \otimes Y) = gh \otimes (h^{-1}X)Y, \quad g \in G(k), \quad X, Y \in \mathcal{U}(\mathfrak{g}),$$

with $h^{-1}X := \mathrm{ev}_{h^{-1}} \star X \star \mathrm{ev}_h$. (By Proposition 2.10 we identify distributions at e with elements in $\mathcal{U}(\mathfrak{g})$.)

Proof. This is done with a direct check. We just point out that it is enough to do such check just on generators. □

3. Super Harish-Chandra pairs

The theory of super Harish-Chandra Pairs (SHCP) that we shall develop presently provides an equivalent way to approach the analytic or affine algebraic supergroups.

Definition of an SHCP. Any time we say *supergroup* we mean an analytic or an affine algebraic supergroup over a field k of characteristic zero.

Definition 3.1. Let G_0 be a group (complex analytic or affine algebraic) and $\mathfrak{g}$ a super Lie algebra. We make the following assumptions:

(1) $\mathfrak{g}_0 \simeq \mathrm{Lie}(G_0)$.

(2) G_0 acts on $\mathfrak{g}$ and this action restricted to $\mathfrak{g}_0$ is the adjoint representation of G_0 on $\mathrm{Lie}(G_0)$. Moreover, the differential of the action is the Lie bracket. We denote such an action by Ad or as $g.X$, $g \in G_0$, $X \in \mathfrak{g}$.

Then $(G_0, \mathfrak{g})$ is called a *super Harish-Chandra pair* (*SHCP*).

A *morphism* of SHCP is simply a pair of morphisms $\psi = (\psi_0, \rho^\psi)$ preserving the SHCP structure; that is:

(1) $\psi_0 : G_0 \to H_0$ is a group morphism (in the analytic or algebraic category).

(2) $\rho^\psi : \mathfrak{g} \to \mathfrak{h}$ is a super Lie algebra morphism.

(3) ψ_0 and ρ^ψ are compatible in the sense that $\rho^\psi_{|\mathfrak{g}_0} = d\psi_0$ and

$$\mathrm{Ad}(\psi_0(g)) \circ \rho^\psi = \rho^\psi \circ \mathrm{Ad}(g).$$

When G_0 is an analytic group we shall speak of an *analytic SHCP*, when G_0 is an affine algebraic group of an *algebraic SHCP*.

We would like to show that the category of (analytic of algebraic) SHCP, denoted by (shcps), is equivalent to the category of supergroups (analytic or algebraic), denoted by (sgrps). In order to do this we start by associating in a natural way a supergroup to an SHCP.

Definition 3.2. Let $(G_0, \mathfrak{g})$ be an SHCP. The sheaf $\mathcal{O}_{G_0}$ of the ordinary group G_0 carries a natural action of $\mathcal{U}(\mathfrak{g}_0)$, since the elements of $\mathcal{U}(\mathfrak{g}_0)$ act on the sections in $\mathcal{O}_{G_0}(U)$ as left-invariant differential operators. We define $\mathcal{O}_G(U)$ as

$$\mathcal{O}_G(U) := \mathrm{Hom}_{\mathcal{U}(\mathfrak{g}_0)}(\mathcal{U}(\mathfrak{g}), \mathcal{O}_{G_0}(U)), \quad U \subset_{open} G_0.$$

Proposition 3.3. *The assignment $U \mapsto \mathcal{O}_G(U)$ is a sheaf of superalgebras on G_0, where the superalgebra structure on $\mathcal{O}_G(U)$ is given by*

$$f_1 \cdot f_2 = m_{\mathcal{O}_{G_0}} \circ (f_1 \otimes f_2) \circ \Delta_{\mathcal{U}(\mathfrak{g})}$$

and the restriction morphisms $\rho_{UV} : \mathcal{O}_G(U) \to \mathcal{O}_G(V)$ are $\rho_{UV}(f) := \tilde{\rho}_{UV} \circ f$, where $\tilde{\rho}_{UV}$ are the restrictions of the ordinary sheaf $\mathcal{O}_{G_0}$.

Proof. The check $f_1 \cdot f_2$ is an associative product is routine, while the sheaf property comes from the fact $\mathcal{O}_{G_0}$ is an ordinary sheaf. □

We now show that $(G_0, \mathcal{O}_G)$ is a superspace, by showing that is *globally split*; in other words, that

$$\mathcal{O}_G(U) \cong \mathcal{O}_{G_0}(U) \otimes \bigwedge(\mathfrak{g}_1).$$

Theorem 3.4. (1) *Let $\gamma : \bigwedge(\mathfrak{g}_1) \to \mathcal{U}(\mathfrak{g})$ be the symmetrization map, given by*

$$\gamma(X_1 \wedge \cdots \wedge X_p) = \frac{1}{p!} \sum_{\tau \in S_p} (-1)^{|\tau|} X_{\tau(1)} \cdots X_{\tau(p)},$$

where $|\tau|$ denotes the parity of the permutation τ. Then

$$\widehat{\gamma} : \mathcal{U}(\mathfrak{g}_0) \otimes \bigwedge(\mathfrak{g}_1) \to \mathcal{U}(\mathfrak{g}), \quad X \otimes Y \mapsto X \cdot \gamma(Y)$$

is an isomorphism of super left $\mathcal{U}(\mathfrak{g}_0)$-modules.

(2) $(G_0, \mathcal{O}_G)$ *is globally split; i.e., for each open subset $U \subseteq G_0$ there is an isomorphism of superalgebras*

$$\mathcal{O}_G(U) \simeq \mathrm{Hom}\big(\bigwedge(\mathfrak{g}_1), \mathcal{O}_{G_0}(U)\big) \simeq \mathcal{O}_{G_0}(U) \otimes \bigwedge(\mathfrak{g}_1)^*.$$

Hence $\mathcal{O}_G$ carries a natural $\mathbb{Z}$-gradation.

Proof. (1) is an application of Poincaré–Birkhoff–Witt (PBW) theorem (see [Varadarajan 2004]), while for (2) consider the map

$$\phi_U : \mathcal{O}_G(U) \to \mathrm{Hom}\big(\bigwedge(\mathfrak{g}_1), \mathcal{O}_{G_0}(U)\big), \quad f \mapsto f \circ \gamma.$$

Since γ is a supercoalgebra morphism, ϕ_U is a superalgebra morphism. In fact,

$$\phi_U(f_1 \cdot f_2) = m \circ f_1 \otimes f_2 \circ \Delta_{\mathcal{U}(\mathfrak{g})} \circ \gamma = m \circ f_1 \otimes f_2 \circ (\gamma \otimes \gamma)\Delta_{\mathcal{U}(\mathfrak{g})} = \phi_U(f_1)\phi_U(f_2).$$

That ϕ_U is a superalgebra isomorphism follows at once from $\mathcal{U}(\mathfrak{g}_0)$-linearity. □

As an almost immediate consequence of the previous theorem we have:

Corollary 3.5. *If G_0 is an analytic manifold or algebraic scheme, then $(G_0, \mathcal{O}_G)$ is a superspace.*

In the next sections we complete the task of showing $(G_0, \mathcal{O}_G)$ is a supergroup by providing explicit expression for the multiplication, unit and inverse. This will lead to the main result of the paper, namely the equivalence of categories between the SHCP and supergroups. We now state the main result of the paper and then we shall prove it with different methods in the next sections, since at this point the analytic and algebraic categories diverge and require dramatically different treatment.

Theorem 3.6. *Let k be a field of characteristic zero, $k = \mathbb{C}$ if we are in the algebraic category. Define the functors*

$$\begin{array}{rcl} \mathcal{H} : (\text{sgrps}) & \to & (\text{shcps}) \\ G & \mapsto & (G_0, \mathrm{Lie}(G)) \\ \phi & \mapsto & (|\phi|, (d\phi)_e) \end{array}$$

and

$$\begin{array}{rcl} \mathcal{K} : (\text{shcps}) & \to & (\text{sgrps}) \\ (G_0, \mathfrak{g}) & \mapsto & \bar{G} := \big(G_0, \mathrm{Hom}_{\mathcal{U}(\mathfrak{g}_0)}(\mathcal{U}(\mathfrak{g}), \mathcal{O}_{G_0})\big) \\ \psi = (\psi_0, \rho^\psi) & \mapsto & f \mapsto \psi_0^* \circ f \circ \rho_\psi, \end{array}$$

where G and $(G_0, \mathfrak{g})$ are objects and ϕ, ψ are morphisms of the corresponding categories (in the definition of $\mathcal{H}$, G_0 is the ordinary group underlying G). Then $\mathcal{H}$ and $\mathcal{K}$ define an equivalence between the categories of supergroups (analytic or algebraic) and super Harish-Chandra pairs (analytic or algebraic).

Analytic SHCP. Let $k = \mathbb{C}$.

For analytic SHCP it is relatively easy to define a supergroup structure on the superspace $(G_0, \mathcal{O}_G)$ we have defined above, by mimicking what happens in the smooth case. In fact for an analytic ordinary group G_0, the action of $\mathcal{U}(\mathfrak{g}_0)$ on $\mathcal{O}_{G_0}$ is given by

$$(\tilde{D}_Z \cdot f)(g) = f(ge^{tZ}), \quad Z \in \mathfrak{g}_0, \quad f \in \mathcal{O}_{G_0}(U),$$

where e^{tZ} denotes the one-parameter subgroup corresponding to the element $Z \in \mathfrak{g}_0$. Notice that at this point we encounter an important difference with the algebraic setting, since in that case we do not have a result such as the Frobenius theorem available.

Proposition 3.7. *$(G_0, \mathcal{O}_G)$ is an analytic supergroup where the multiplication μ, inverse i and unit e are defined via the corresponding sheaf morphisms by*

$$\begin{aligned}\big[\mu^*(f)(X, Y)\big](g, h) &= \big[f\big((h^{-1}.X)Y\big)\big](gh),\\ \big[i^*(f)(X)\big](g^{-1}) &= \big[f(g^{-1}.\bar{X})\big](g),\\ e^*(f) &= \big[f(1)\big](e),\end{aligned}$$

for $f \in \mathcal{O}_G(U)$ and $g, h \in |G|$, where $|G|$ is the topological space underlying G_0. Here $\bar{X}$ denotes the antipode in $\mathcal{U}(\mathfrak{g})$.

Note. We shall discuss the peculiar form of μ^*, i^*, e^* in Remark 3.14.

Proof. The proof of this result is the same as in the differential smooth setting, where everything is defined in the same way (see [Carmeli et al. 2011, Chapter 7]. In particular to prove that μ^*, i^*, e^* are $\mathcal{U}(\mathfrak{g}_0)$-morphisms is harder than the verification of the compatibility conditions and the Hopf superalgebra properties. As an example, let us verify μ is well-defined the other properties being essentially the same type of calculation. Due to the PBW theorem, it is enough to prove $\mathfrak{g}_0$-linearity. Let $Z \in \mathfrak{g}_0$; then

$$\begin{aligned}\mu^*(f)(ZX, Y)(g, h) &= f(h^{-1}(ZX)Y)(gh)\\ &= f((h^{-1}.Z)(h^{-1}.X)Y)(gh)\\ &= \tilde{D}_{h^{-1}.Z}\left[f((h^{-1}.X)Y)\right](gh).\end{aligned}$$

On other hand,

$$\begin{aligned}\big[(\tilde{D}_Z \otimes \mathrm{id})(\mu^*(f)(X, Y))\big](g, h) &= \frac{d}{dt}_{|t=0} f((h^{-1}X)Y)(ge^{tZ}h)\\ &= \frac{d}{dt}_{|t=0} f((h^{-1}X)Y)(ghe^{t(h^{-1}Z)})\\ &= \tilde{D}_{h^{-1}Z}\left[f((h^{-1}.X)Y)\right](gh).\end{aligned}$$

Similarly, for the left entry, one finds

$$\begin{aligned}\mu^*(f)(X, ZY)(g, h) &= f((h^{-1}X)ZY)(gh)\\ &= f\big(Z(h^{-1}X)Y + [h^{-1}X, Z]Y\big)(gh)\\ &= \tilde{D}_Z\big(f((h^{-1}X)Y)\big)(gh) + f\big([h^{-1}X, Z]Y\big)(gh)\end{aligned}$$

and

$$\begin{aligned}\frac{d}{dt}_{|t=0}\mu^*(f)(X,Y)(g,he^{tZ}) &= \frac{d}{dt}_{|t=0} f\big(((he^{tZ})^{-1}X)Y\big)(ghe^{tZ}) \\ &= \big[\tilde{D}_Z f((h^{-1}X)Y)\big](gh) + f\big([(h^{-1}X),Z]Y\big)(gh).\end{aligned}$$

□

We are now ready for the proof of Theorem 3.6 in the analytic setting.

Theorem 3.8. *There is an equivalence of categories between analytic SHCP and analytic supergroups expressed by the functors $\mathcal{K}$ and $\mathcal{H}$ in Theorem 3.6.*

Proof. Let us first show the correspondence between morphisms. If ϕ is a morphisms of analytic supergroups, it is immediate that $(|\phi|,(d\phi)_e)$ is a morphism of SHCP. Conversely, if $\psi=(\psi_0,\rho_\psi)$ is a morphism of SHCP $(G_0,\mathfrak{g})$, $(H_0,\mathfrak{h})$, then the map $\psi^*:\mathcal{O}_H(U)\to\mathcal{O}_G(\psi_0^{-1}(U))$ defined by $\psi^*(f)=\psi_0^*\circ f\circ\rho_\psi$ is a sheaf morphism and (ψ_0,ψ^*) is a morphism of the supergroups G and H. As one can check, the assignments in Theorem 3.6 establish a one-to-one correspondence between the set of morphisms of SHCPs and the set of morphisms of analytic supergroups.

We now turn to the correspondence between the objects. Let G be a supergroup and $\overline{G}$ the supergroup obtained from the SHCP $(G_0,\mathrm{Lie}(G))$, where G_0 is the ordinary analytic group underlying G. As for the smooth setting, let us define the morphism $\eta\colon\overline{G}\to G$ by

$$\begin{aligned}\eta^*\colon\mathcal{O}_G(U)\to\mathcal{O}_{\overline{G}}(U)&=\mathrm{Hom}_{\mathcal{U}(\mathfrak{g}_0)}\big(\mathcal{U}(\mathfrak{g}),\mathcal{O}_{G_0}(U)\big),\\ s&\mapsto\big(\bar{s}\colon X\to(-1)^{|X|}|(D_Xs)|\big).\end{aligned}$$

Here D_X denotes the left-invariant differential operator on G associated with $X\in\mathcal{U}(\mathfrak{g})$, that is $D_X=(1\otimes X)\mu^*$. The definition is well-posed as one can directly check, moreover η is a SLG morphism, i.e.,

$$\eta\circ\mu_{\overline{G}}=\mu_G\circ(\eta\times\eta).$$

Indeed, for each $s\in\mathcal{O}(G)$, $X,Y\in\mathcal{U}(\mathfrak{g})$, and $g,h\in G_0$,

$$\begin{aligned}\big[\big((\eta^*\otimes\eta^*)\mu_G^*(s)\big)(X,Y)\big](g,h)&=(-1)^{|X|+|Y|}|(D_X\otimes D_Y)\mu_G^*(s)|(g,h)\\ &=(-1)^{|X|+|Y|}|D_{h^{-1}.X}D_Ys|(gh)\\ &=\big[\eta^*(s)\big((h^{-1}.X)Y\big)\big](gh)\\ &=\big[\big(\mu_{\overline{G}}^*\eta^*(s)\big)(X,Y)\big](g,h).\end{aligned}$$

The last thing to check is that η is an isomorphism. This is true because $|\eta|$ is clearly bijective and, for each $g\in G_0$, the differential $(d\eta)_g$ is bijective:

$$\begin{aligned}\big[(d\eta)_g(\overline{D}_{X_g})\big](s)=\overline{D}_{X_g}\eta^*(s)&=\mathrm{ev}_g(\overline{D}_X\eta^*(s))=[\overline{D}_X\eta^*(s)](1)(g)\\ &=(-1)^{|X|}\eta^*(s)(X)(g)=|(D_Xs)|(g)=D_{X_g}(s),\end{aligned}$$

where we denote by $\overline{D}_X$ a left-invariant differential operator on $\overline{G}$ corresponding to $X \in \mathcal{U}(\mathfrak{g})$ while D_X denotes a left-invariant differential operator on G.

We conclude using the inverse function theorem, which holds also for analytic supermanifolds and again this is an important difference with the algebraic setting, where we do not have this tool available. □

Remark 3.9 (p-adic SHCP). One can define p-adic supermanifolds, supergroups and SHCP through the obvious same definitions within the framework described classically in [Serre 1992]. In fact since the category of p-adic manifolds resembles very closely the category of analytic manifolds, it is then only reasonable to expect that one can develop along the same lines the theory of p-adic supermanifolds. Once the basic results, like the inverse function theorem, are established, the equivalence of categories between p-adic supergroups and the p-adic SHCP will then follow through the same proof we have detailed for the analytic category.

Algebraic SHCP. We now prove our main result, Theorem 3.6, in the case of G an affine algebraic supergroup over an algebraically closed field of characteristic zero.

The category of affine algebraic supergroups is equivalent to the category of commutative Hopf superalgebras; hence we need to show that there is a unique commutative Hopf superalgebra $\mathcal{O}(G)$ associated to a SHCP $(G_0, \mathfrak{g})$, namely the superalgebra of the global sections of the sheaf $\mathcal{O}_G$ as in Definition 3.2.

Since the exponential appears for the action of $\mathcal{U}(\mathfrak{g}_0)$ on $\mathcal{O}(G_0)$ (see beginning of previous subsection), the question is entirely classical and it is treated in detail in [Demazure and Gabriel 1970, Chapter 2] for the algebraic setting. We shall briefly review a few key facts, sending the reader to that reference for details.

Let G_0 be an algebraic group and A a commutative algebra, $p : A(t) \to A[t]/(t^2)$ the natural projection, t even. By definition, $\mathrm{Lie}(G_0)(A) = \ker G_0(p)$. Since G_0 is affine we have $G_0 \subset \mathrm{GL}(V)$ for a suitable vector space V; hence we can write

$$\begin{aligned}\mathrm{Lie}(G_0)(A) = \{\, 1 + tZ \,\} &\subset G_0(A(t)) \subset \mathrm{GL}(V)(A(t)) \\ &= \mathrm{GL}(V)(A) + t\mathrm{End}(V)(A)\end{aligned}$$

for suitable $Z \in \mathrm{End}(V)(A)$, where $\mathrm{End}(V)$ is the functor of points of the superscheme of the endomorphisms of the vector space V. Very often $\mathrm{Lie}(G_0)$ is identified with the subspace in $\mathrm{End}(V)$ consisting of the elements Z. As a notation device we define

$$e^{tZ} = 1 + tZ \in G_0(A(t)).$$

Let $g \in G_0(A) = \mathrm{Hom}(\mathcal{O}(G_0), A)$, that is, g is an A-point of G_0, and let $f \in \mathcal{O}(G_0)$. As another common notational device, we denote $g(f)$ with $f(g)$. Since A embeds naturally in $A(t)$ we can view g also as an $A(t)$-point of G_0 and consider $f(ge^{tZ})$.

We then define

$$(*) \qquad \frac{d}{dt}_{|t=0} f(ge^{tZ}) = b,$$

where $f(ge^{tZ}) = (ge^{tZ})(f) = a + bt \in A(t)$. One sees that the left-hand side of (*) corresponds to the natural action of $Z \in \mathrm{Lie}(G_0)$ on $\mathcal{O}(G_0)$ via left-invariant operators, that is,

$$\frac{d}{dt}_{|t=0} f(ge^{tZ}) = (1 \otimes Z)\mu^*(f),$$

which we denoted by $\tilde{D}_Z f$ in the analytic category.

We now go back to the super setting and prove the analogue of Proposition 3.7.

Proposition 3.10. *The superalgebra $\mathcal{O}(G) = \mathrm{Hom}(\mathcal{U}(\mathfrak{g}), \mathcal{O}(G_0))$ associated to the algebraic SHCP $(G_0, \mathfrak{g})$ is an Hopf superalgebra where the comultiplication μ^*, antipode i^* and counit e^** [3] *are defined as follows*:

$$[\mu^*(f)(X, Y)](g, h) = [f((h^{-1}.X)Y)](gh),$$
$$[i^*(f)(X)](g^{-1}) = [f(g^{-1}.\overline{X})](g),$$
$$e^*(f) = [f(1)](e),$$

for $f \in \mathcal{O}(G)$, $g, h \in |G|$. Here $\overline{X}$ denotes the antipode in $\mathcal{U}(\mathfrak{g})$.

Proof. It is the same as for Proposition 3.7. Though the context is different, once the exponential terminology assumes a meaning for the algebraic category, the calculations are the same. □

The next proposition shows a very natural fact: given an SHCP $(G_0, \mathcal{O}_G)$, the sheaf $\mathcal{O}_G$ is the structural sheaf associated with the superalgebra of its global sections $\mathcal{O}(G)$, so that the morphisms μ^*, i^*, e^* are actually defined as the appropriate sheaf morphisms, corresponding to μ, i, e, multiplication, inverse and unit in the algebraic supergroup $G = \underline{\mathrm{Spec}}\, \mathcal{O}(G)$. corresponding to the SHCP $(G_0, \mathfrak{g})$.

Proposition 3.11. *Let $(G_0, \mathfrak{g})$ be an SHCP, with G_0 an affine group scheme and let $\mathcal{O}_G$ as in 3.1. Then $G := (G_0, \mathcal{O}_G)$ is a supergroup scheme.*

Proof. In Proposition 3.10 we have seen that $\mathcal{O}(G) := \mathrm{Hom}_{\mathcal{U}(\mathfrak{g}_0)}(\mathcal{U}(\mathfrak{g}), \mathcal{O}_{G_0}(G_0))$ has an Hopf superalgebra structure, moreover by Theorem 3.4 it is globally split. Hence we only need to prove that $G = \underline{\mathrm{Spec}}\, \mathcal{O}(G)$. Clearly the topological spaces underlying the superspaces $G = (G_0, \mathcal{O}_G)$ and $\underline{\mathrm{Spec}}\, \mathcal{O}(G)$ are homeomorphic. We only need to show that $\mathcal{O}_{\mathcal{O}(G)} \cong \mathcal{O}_G$, where $\mathcal{O}_{\mathcal{O}(G)}$ denotes the structural sheaf associated with the superring $\mathcal{O}(G)$. We set up a morphism

[3] In analogy with Proposition 3.7 we have kept the terminology μ^*, i^*, e^*, though we are not making (yet) any claim on the sheaf morphisms.

$$\phi : \mathcal{O}_G(U) \to \mathcal{O}_{\mathcal{O}(G)}(U)$$

taking $s : \mathcal{U}(\mathfrak{g}) \to \mathcal{O}_{G_0}(U)$ to

$$\phi(s) : U \to \coprod_{x \in U} \mathcal{O}(G)_x,$$

as follows. Any $s \in \mathcal{O}_G(U)$ gives raise naturally to $s_x : \mathcal{U}(\mathfrak{g}) \to \mathcal{O}_{G_0}(U) \to \mathcal{O}_{G_0,x}$. Since as a $\mathcal{U}(\mathfrak{g}_0)$ module, $\mathcal{U}(\mathfrak{g})$ is finitely generated, say by N generators, once we fix those generators, s_x is equivalent to the choice of N elements in $\mathcal{O}_{G_0,x}$. Since likewise $\mathcal{O}(G)_x$ is finitely generated by N elements as free $\mathcal{O}_{G_0,x}$-module (those N elements corresponds dually to the generators of $\mathcal{U}(\mathfrak{g})$ as $\mathcal{U}(\mathfrak{g}_0)$-module), we have that s_x can be viewed as an element of $\mathcal{O}(G)_x$. So we define

$$\phi(s)(x) = s_x, \quad x \in U.$$

We leave to the reader the check that ϕ is a sheaf isomorphism. □

Theorem 3.12. *The category of algebraic SHCP is equivalent to the category of affine algebraic supergroups.*

Proof. We need to establish a one to one correspondence between the objects and the morphisms.

As for the objects, if $(G_0, \mathfrak{g})$ is an algebraic SHCP, we can define an affine algebraic supergroup defining the following Hopf superalgebra (see Proposition 3.10):

$$\mathcal{O}(G_0, \mathfrak{g}) = \underline{\mathrm{Hom}}_{\mathcal{U}(\mathfrak{g}_0)}(\mathcal{U}(\mathfrak{g}), \mathcal{O}(G_0)).$$

Conversely, if we have an algebraic supergroup, we can find right away the SHCP associated to it. What we need to show is that these operations are one the inverse of the other; that is,

$$\mathcal{O}(G_0, \mathfrak{g}) \cong \mathcal{O}(G),$$

where G_0 is the algebraic group underlying G and $\mathfrak{g} = \mathrm{Lie}(G)$. Certainly they are isomorphic as $\mathcal{O}(G_0)$-modules, since they have the same reduced part and, by a result from [Masuoka 2005], they both can be written as $\mathcal{O}(G_0) \otimes \Lambda$ for some exterior algebra Λ, but being their odd dimension the same, the two exterior algebras are isomorphic.

We can set a map

$$\eta^* : \mathcal{O}(G) \to \mathcal{O}(G_0, \mathfrak{g})$$

taking s to $\bar{s} : X \mapsto (-1)^{|X|}|D_X(s)|$, where $D_X(s) = (1 \otimes X)\mu^*$. This is a well-defined morphism of Hopf superalgebras and $X \mapsto (-1)^{|X|}|D_X(s)|$ is a $\mathcal{U}(\mathfrak{g}_0)$-morphism. This is done precisely in the same way as in the proof of Theorem 3.8.

We now want to show that η^* is surjective. This will imply that η^* is an isomorphism. In fact the two given supergroups $G = \underline{\mathrm{Spec}}\, \mathcal{O}(G)$ and $\overline{G} = \underline{\mathrm{Spec}}\, \mathcal{O}(G_0, \mathfrak{g})$

are smooth superschemes, with the same underlying topological space and same Lie superalgebra (hence the same superdimension), and η^* induces an injective morphism $\eta : \overline{G} \to G$ (see [Fioresi and Gavarini 2013, Section 2]).

For the surjectivity of η^*, we need to show that, for each morphism of $\mathcal{U}(\mathfrak{g}_0)$-modules $\bar{s} : \mathcal{U}(\mathfrak{g}) \to \mathcal{O}(G_0)$, there exists $s \in \mathcal{O}(G)$ such that $\bar{s}(X) = (-1)^{|X|}|D_X(s)|$. Since $\mathcal{U}(\mathfrak{g}) \cong \mathcal{U}(\mathfrak{g}_0) \otimes \bigwedge(\mathfrak{g}_1)$ (see Theorem 3.4) and $\bar{s}$ is a morphism of $\mathcal{U}(\mathfrak{g}_0)$-modules, $\bar{s}$ is determined by $\bar{s}(\gamma(X^I))$ for $X^I = X_1^{i_1} \dots X_n^{i_n}$, where the X_i form a basis for $\mathfrak{g}_1$ and $i_j = 0, 1$ (again refer to Theorem 3.4). Notice that $X_i = \gamma(X_i)$. Since $X_1, \dots, X_n$ are linearly independent, also the corresponding left-invariant vector fields $D_{X_1}, \dots, D_{X_n}$ will be linearly independent at each point. Let $D_{\gamma(X)}$ denote the left-invariant differential operator corresponding to $\gamma(X) \in \mathcal{U}(\mathfrak{g})$. Notice that fixing a suitable basis in $\mathcal{U}(\mathfrak{g})$, the linear morphism $X \mapsto \gamma(X)$ corresponds to an upper triangular matrix and sends linearly independent vectors to linearly independent vectors. Consider the equation $(-1)^{|X^I|}|D_{\gamma(X^I)}s| = \bar{s}(X^I)$, for $X^I = X_1^{i_1} \dots X_n{}^{i_n}$ a monomial in $\bigwedge(\mathfrak{g}_1)$. This is an equation where each D_{X_i} appearing in the expression for $D_{\gamma(X^I)}$ can be expressed as

$$D_{X_i} = \sum a_i \, \partial_{x_{ij}}, \qquad p(a_i) \neq p(x_{ij})$$

where the x_{ij} are global coordinates on $\mathrm{GL}_{m|n} \supset G$ (regardless of their parity).

Since the $D_{X_1}^{i_1} \dots D_{X_n}^{i_n}$ are linearly independent by the PBW theorem (see also Proposition 2.11), the $D_{\gamma(X)}$ will also be linearly independent, and the equality

$$(-1)^{|X|}|D_{\gamma(X^I)}| = \bar{s}(X^I)$$

will yield a solution

$$\partial_{x_{i_1 j_1}} \dots \partial_{x_{i_r j_r}} s = a_{i_1 j_1 \dots i_r j_r}$$

for all $i_1 j_1 \dots i_r j_r$ such that

$$s = \sum a_{i_1 j_1 \dots i_r j_r} x_{i_1 j_i} \dots x_{i_r j_r}.$$

We leave to the reader the correspondence between morphisms. □

Example 3.13. We want to verify explicitly the surjectivity of η^* in the case of $\mathrm{GL}(1|1)$ and make a few remarks on how to extend the calculation to the case of $G = \mathrm{GL}(m|n)$. Let $\mathcal{O}(\mathrm{GL}(1|1)) = k[a_{11}, a_{22}, \alpha_{12}, \alpha_{21}][a_{11}^{-1}, a_{22}^{-1}]$. Let

$$D_{12} = (1 \otimes \partial_{\alpha_{12}})\mu^* = a_{11}\, \partial_{\alpha_{12}} + \alpha_{21}\, \partial_{a_{22}},$$
$$D_{21} = (1 \otimes \partial_{\alpha_{21}})\mu^* = \alpha_{12}\, \partial_{a_{11}} + a_{22}\, \partial_{\alpha_{21}},$$

be the left-invariant vector fields corresponding to the generators $\partial_{\alpha_{12}}, \partial_{\alpha_{21}}$ of $\mathrm{Lie}(G)_1$; then

$$\begin{aligned}\gamma(D_{12}D_{21}) &= \tfrac{1}{2}(D_{12}D_{21} - D_{21}D_{12}) \\ &= \tfrac{1}{2}(a_{11}\,\partial_{a_{11}} - a_{22}\,\partial_{a_{22}}) \\ &\qquad\qquad + a_{11}a_{22}\,\partial_{\alpha_{12}}\,\partial_{\alpha_{21}} + \text{terms with coefficients in } J_{\mathcal{O}(\mathrm{GL}(1|1))},\end{aligned}$$

where $J_{\mathcal{O}(\mathrm{GL}(1|1))}$ denotes as usual the ideal generated by the odd elements. Notice that the terms with coefficients in $J_{\mathcal{O}(\mathrm{GL}(1|1))}$ do not contribute in the expression $|D_{\gamma(D_{12}D_{21})}s|$. For the same reason, the term $a_{11}\,\partial_{a_{11}} - a_{22}\,\partial_{a_{22}}$ will make a contribution only if applied to s^0, and consequently can be considered not as unknown, but as a known term. This is important in case one wants to generalize this procedure to $\mathrm{GL}(m|n)$; in fact only the terms containing only odd derivations will produce new quantities to be determined.

Given $\bar{s} : \mathcal{U}(\mathfrak{g}) \to \mathcal{O}(G_0)$ we want to determine $s \in \mathcal{O}(G)$, with $\eta^*(s) = \bar{s}$. Since $\mathrm{Lie}(\mathrm{GL}(1|1)_1 = \langle \partial_{\alpha_{12}}, \partial_{\alpha_{21}} \rangle$, the map $\bar{s}$ is determined once we know its image on $\bigwedge \mathrm{Lie}(\mathrm{GL}(1|1)_1$, that is,

$$s^0 = \bar{s}(1), \quad s^{12} = \bar{s}(\partial_{\alpha_{12}}), \quad s^{21} = \bar{s}(\partial_{\alpha_{21}}), \quad s^{12,21} = \bar{s}(\gamma(\partial_{\alpha_{12}}\partial_{\alpha_{21}})).$$

Consequently the s we want to determine must satisfy the equations

$$\begin{aligned} s^0 &= |1s|, \\ s^{12} &= -|a_{11}\partial_{\alpha_{12}}s + \alpha_{21}\partial_{a_{22}}s|, \\ s^{21} &= -|\alpha_{12}\partial_{a_{11}}s + a_{22}\partial_{\alpha_{21}}s|, \\ s^{12,21} &= \left|\tfrac{1}{2}(a_{11}\partial_{a_{11}}s - a_{22}\partial_{a_{22}}s) + a_{11}a_{22}\partial_{\alpha_{12}}\partial_{\alpha_{21}}s\right|. \end{aligned}$$

A simple calculation gives us

$$s = s^0 + \frac{\alpha_{12}s^{12}}{a_{11}} - \frac{\alpha_{21}s^{21}}{a_{22}} + \left[s^{12,21} - \tfrac{1}{2}\left(a_{11}\partial_{a_{11}}s^0 - a_{22}\partial_{a_{22}}s^0\right)\right]\frac{\alpha_{12}\alpha_{21}}{a_{11}a_{22}}.$$

There is no conceptual obstacle to extending this calculation to the case of $G = \mathrm{GL}(m|n)$. If $\mathcal{O}(G) = k[a_{ij}, \alpha_{kl}][d_1^{-1}, d_2^{-1}]$ where $d_1 = \det(a_{ij})_{\{1 \le i,j \le m\}}$ and $d_2 = \det(a_{ij})_{\{m+1 \le i,j \le m+n\}}$, the left-invariant vector fields are given by

$$X_{ij} = \left(1 \otimes \partial_{x_{ij}}\right)\mu^* = \sum_k x_{ki}\,\partial_{x_{kj}},$$

where x_{ij} denote the coordinates on $\mathrm{GL}(m|n)$ regardless of their parity. We can then repeat the calculation we did above. Notice that any even derivation appearing in the expression $|D_{\gamma(X)}s|$ will affect only $s^0 = |1s|$ since we are taking the reduction modulo the ideal of the odd nilpotents.

Remark 3.14. We clarify the relation between the Hopf superalgebra $\mathcal{O}(G) = \mathrm{Hom}(\mathcal{U}(\mathfrak{g}), \mathcal{O}(G_0))$ associated to the SHCP $(G_0, \mathfrak{g})$ and the distribution superalgebra $D(G)$ of the supergroup G (also naturally associated to the same SHCP).

For an affine supergroup G, the superalgebra of distributions $D(G)$ has a natural Hopf superalgebra structure; see Proposition 2.15. This structure is inherited by $k|G| \otimes \mathcal{U}(\mathfrak{g})$ through the linear isomorphism with $D(G)$ given in Proposition 2.16. The superalgebra of global sections of G, $\mathcal{O}(G) = \mathrm{Hom}(\mathcal{U}(\mathfrak{g}), \mathcal{O}(G_0))$ can then be naturally viewed as a subspace of $D(G)^* \cong (k|G| \otimes \mathcal{U}(\mathfrak{g}))^*$, since elements in $\mathcal{O}(G)$ arise as suitable morphisms $|G| \times \mathcal{U}(\mathfrak{g}) \to k$. One can then immediately verify that the Hopf superalgebra structure on $\mathcal{O}(G) \subset D(G)^*$ is precisely obtained by duality, from the Hopf superalgebra on $D(G)$ suitably restricting the comultiplication, counit and antipode morphisms.

4. Action of supergroups and SHCPs

We now want to relate the action of an analytic of algebraic supergroup G on a supermanifold or superscheme M, with the action of the corresponding SHCP $(G_0, \mathfrak{g})$ on M. In this section, if $g \in |G|$ we denote by $\hat{g} : \mathbb{C}^{0|0} \to G$ the morphism whose pull-back is the evaluation at g. We recall a well-know definition:

Definition 4.1. A morphism $a\colon G \times M \to M$ is called an *action* of G on M if

$$a \circ (\mu \times \mathbb{1}_M) = a \circ (\mathbb{1}_G \times a) \tag{**}$$

and

$$a \circ \langle \hat{e}, \mathbb{1}_M \rangle = \mathbb{1}_M .$$

In the functor of points notation, this is the same as demanding the following, where T is a supermanifold (resp. a superscheme) and $M(T) = \mathrm{Hom}(T, M)$ are the T-points of M:

(1) $1 \cdot x = x$ for all $x \in M(T)$, where 1 the unit in $G(T)$.

(2) $(g_1 g_2) \cdot x = g_1 \cdot (g_2 \cdot x)$ for all $x \in M(T)$ and all $g_1, g_2 \in G(T)$.

Here, as usual, we are writing $a(g, x)$ as $g \cdot x$.

If an action a of G on M is given, then we say that G *acts* on M.

Definition 4.2. An *action* of an analytic SHCP $(G_0, \mathfrak{g})$ on a supermanifold M consists of an action

$$\underline{a}\colon G_0 \times M \to M$$

of the reduced Lie group G_0 on M, with $\underline{a} : a \circ (j_{|G| \to G} \times \mathbb{1}_M)$, plus a representation

$$\begin{aligned} \rho_a \colon \mathfrak{g} &\to \mathrm{Vec}(M)^{\mathrm{op}} \\ X &\mapsto (X \otimes \mathbb{1}_{\mathcal{O}(M)}) a^* \end{aligned}$$

of the super Lie algebra $\mathfrak{g}$ of G on the opposite of the Lie superalgebra of vector fields over M, the whole satisfying the compatibility relations

$$\rho_{a\,|_{\mathfrak{g}_0}}(X) = (X \otimes \mathbb{1}_{\mathcal{O}(M)})\underline{a}^* \qquad \text{for all } X \in \mathfrak{g}_0,$$
$$\rho_a(g.Y) = \big(\underline{a}^{g^{-1}}\big)^* \rho_a(Y)(\underline{a}^g)^* \qquad \text{for all } g \in |G|,\ Y \in \mathfrak{g},$$

where $a^g\colon M \to M$ is given by $a^g := a \circ \langle \hat{g}, \mathbb{1}_M\rangle$.

The next proposition tells us that actions of an SHCP correspond bijectively to actions of the corresponding analytic supergroup.

Proposition 4.3. *Let G be an analytic supergroup acting on a supermanifold M. Then there is an action of the SHCP $(G_0, \mathrm{Lie}(G))$ on M. Conversely, given an action of the SHCP $(G_0, \mathfrak{g})$ on M, there is a unique action $a_\rho\colon G \times M \to M$ of the analytic supergroup G corresponding to the given SHCP on M whose reduced and infinitesimal actions are the given ones. If U is an open subset of M, we have*

$$a_\rho^*\colon \mathcal{O}_M(U) \to \mathrm{Hom}_{\mathcal{U}(\mathfrak{g}_0)}\big(\mathcal{U}(\mathfrak{g}), (\mathcal{O}_{G_0}\hat{\otimes}\mathcal{O}_M)(|a|^{-1}(U))\big),$$
$$f \mapsto \big[X \mapsto (-1)^{|X|}\big(\mathbb{1}_{\mathcal{O}(G_0)} \otimes \rho(X)\big)\underline{a}^*(f)\big].$$

Proof. Let us check that $a_\rho^*(f)$ is $\mathcal{U}(\mathfrak{g}_0)$-linear. For all $X \in \mathcal{U}(\mathfrak{g})$ and $Z \in \mathfrak{g}_0$ we have

$$\begin{aligned}
a_\rho^*(f)(ZX) &= (-1)^{|X|}\big(\mathbb{1} \otimes \rho(ZX)\big)\underline{a}^*(f)\\
&= (-1)^{|X|}\big(\mathbb{1} \otimes \rho(X)\big)(\mathbb{1} \otimes Z_e \otimes \mathbb{1})(\mathbb{1} \otimes \underline{a}^*)\underline{a}^*(f)\\
&= (-1)^{|X|}\big(\mathbb{1} \otimes \rho(X)\big)(\mathbb{1} \otimes Z_e \otimes \mathbb{1})(\tilde{\mu}^* \otimes \mathbb{1})\underline{a}^*(f)\\
&= \big(\tilde{D}_Z \otimes \mathbb{1}\big)\big[a_\rho^*(f)(X)\big].
\end{aligned}$$

We now check that a_ρ^* is a superalgebra morphism.

$$\begin{aligned}
\big[a_\rho^*(f_1) \cdot a_\rho^*(f_2)\big](X) &= m_{\mathcal{O}_{G_0}\hat{\otimes}\mathcal{O}_M}\big[a^*(f_1) \otimes a^*(f_2)\big]\Delta(X)\\
&= (-1)^{|X|} m\big[\big(\mathbb{1} \otimes \rho(X_{(1)})\big)\underline{a}^*(f_1) \otimes \big(\mathbb{1} \otimes \rho(X_{(2)})\big)\underline{a}^*(f_2)\big]\\
&= (-1)^{|X|}\big(\mathbb{1} \otimes \rho(X)\big)\big(\underline{a}^*(f_1) \cdot \underline{a}^*(f_2)\big) = a_\rho^*(f_1 \cdot f_2)(X),
\end{aligned}$$

where $f_i \in \mathcal{O}(M)$ and $X_{(1)} \otimes X_{(2)}$ denotes $\Delta(X)$. Concerning the "associative" property, we have that, for $X, Y \in \mathcal{U}(\mathfrak{g})$ and $g, h \in G_0$,

$$\begin{aligned}
\big[(\mu^* \otimes \mathbb{1})a_\rho^*(f)\big](X, Y)(g, h) &= \big[a_\rho^*(f)\big](h^{-1}.XY)(gh)\\
&= (-1)^{|X|+|Y|+|X||Y|}\rho(Y)\rho(h^{-1}.X)(\underline{a}^{gh})^*(f)\\
&= (-1)^{|X|+|Y|+|X||Y|}\rho(Y)(\underline{a}^h)^*\rho(X)(\underline{a}^g)^*(f)\\
&= \big[(\mathbb{1} \otimes a_\rho^*)a_\rho^*(f)\big](X, Y)(g, h),
\end{aligned}$$

and, finally, $(\mathrm{ev}_e \otimes \mathbb{1})a_\rho^*(f) = \rho(1) = f$.

Uniqueness can be proved as follows. Let a be an action of G on M and let $(\underline{a}, \rho_a)$ be as in Proposition 4.3. If $f \in \mathcal{O}_M(U)$, then

$$\begin{aligned} a^*(f) &\in (\mathrm{Hom}_{\mathcal{U}(\mathfrak{g}_0)}\big(\mathcal{U}(\mathfrak{g}), \mathcal{O}_{G_0}\big)\hat{\otimes}\mathcal{O}_M)(|a|^{-1}(U)) \\ &\cong \mathrm{Hom}_{\mathcal{U}(\mathfrak{g}_0)}\big(\mathcal{U}(\mathfrak{g}), (\mathcal{O}_{G_0}\hat{\otimes}\mathcal{O}_M)(|a|^{-1}(U))\big); \end{aligned}$$

hence, using (**) in Definition 4.1 and the fact that ρ_a is an antihomomorphism, we obtain for all $X \in \mathcal{U}(\mathfrak{g})$

$$\begin{aligned} a^*(f)(X) &= (-1)^{|X|}\big[(D_X \otimes \mathbb{1})a^*(\phi)\big](1) \\ &= (-1)^{|X|}\big(\mathbb{1} \otimes \rho_a(X)\big)\big(a^*(f)(1)\big) = (-1)^{|X|}\big(\mathbb{1} \otimes \rho_a(X)\big)\underline{a}^*(f). \qquad \square \end{aligned}$$

Let us now assume G is an affine algebraic supergroup over a field of characteristic zero and $(G_0, \mathfrak{g})$ is the corresponding SHCP and furthermore assume they are acting on a supervariety M, the Definition 4.2 being the same, taking the morphisms in the appropriate category.

We state the analogue of the Proposition 4.3 in the algebraic setting, its proof being essentially the same.

Proposition 4.4. *Let G be an algebraic supergroup acting on a supervariety M (not necessarily affine). Then there is an action of the SHCP $(G_0, \mathrm{Lie}(G))$ on M. Conversely, given an algebraic action of the algebraic SHCP $(G_0, \mathfrak{g})$ on M, there is a unique action $a_\rho \colon G \times M \to M$ of the algebraic supergroup G corresponding to the given SHCP on M whose reduced and infinitesimal actions are the given ones. If U is an open subset of M, we have*

$$\begin{aligned} a_\rho^* \colon \mathcal{O}_M(U) &\to \mathrm{Hom}_{\mathcal{U}(\mathfrak{g}_0)}\big(\mathcal{U}(\mathfrak{g}), (\mathcal{O}_{G_0} \otimes \mathcal{O}_M)(|a|^{-1}(U))\big), \\ f &\mapsto \big[X \mapsto (-1)^{|X|}\big(\mathbb{1}_{\mathcal{O}(G_0)} \otimes \rho(X)\big)\underline{a}^*(f)\big]. \end{aligned}$$

Acknowledgements

We wish to thank Professor Varadarajan for suggesting the problem and Professors Cassinelli and Gavarini for helpful discussions.

References

[Balduzzi et al. 2009] L. Balduzzi, C. Carmeli, and G. Cassinelli, "Super G-spaces", pp. 159–176 in *Symmetry in mathematics and physics*, edited by D. Babbitt et al., Contemp. Math. **490**, Amer. Math. Soc., Providence, RI, 2009. MR 2010k:58011 Zbl 1236.58012

[Carmeli et al. 2011] C. Carmeli, L. Caston, and R. Fioresi, *Mathematical foundations of supersymmetry*, European Mathematical Society, Zürich, 2011. MR 2012h:58010 Zbl 1226.58003

[Deligne and Morgan 1999] P. Deligne and J. W. Morgan, "Notes on supersymmetry (following Joseph Bernstein)", pp. 41–97 in *Quantum fields and strings: a course for mathematicians* (Princeton, NJ, 1996/1997), vol. 1, edited by P. Deligne et al., Amer. Math. Soc., Providence, RI, 1999. MR 2001g:58007 Zbl 1170.58302

[Demazure and Gabriel 1970] M. Demazure and P. Gabriel, *Groupes algébriques, I: Géométrie algébrique, généralités, groupes commutatifs*, Masson, Paris, 1970. MR 46 #1800 Zbl 0203.23401

[Dieudonné 1970] J. Dieudonné, *Éléments d'analyse, Tome III: Chapitres XVI et XVII*, Cahiers Scientifiques **33**, Gauthier-Villars, Paris, 1970. MR 42 #5266 Zbl 0208.31802

[Fioresi and Gavarini 2011] R. Fioresi and F. Gavarini, "On the construction of Chevalley supergroups", pp. 101–123 in *Supersymmetry in mathematics and physics*, edited by S. Ferrara et al., Lecture Notes in Math. **2027**, Springer, Berlin, 2011. MR 2906339 Zbl 06078896

[Fioresi and Gavarini 2012] R. Fioresi and F. Gavarini, *Chevalley supergroups*, Mem. Amer. Math. Soc. **1014**, Amer. Math. Soc., Providence, RI, 2012. MR 2918543 Zbl 1239.14045

[Fioresi and Gavarini 2013] R. Fioresi and F. Gavarini, "Algebraic supergroups with Lie superalgebras of classical type", *Journal of Lie Group Theory* **23**:1 (2013), 143–158.

[Jantzen 2003] J. C. Jantzen, *Representations of algebraic groups*, 2nd ed., Mathematical Surveys and Monographs **107**, Amer. Math. Soc., Providence, RI, 2003. MR 2004h:20061 Zbl 1034.20041

[Kostant 1977] B. Kostant, "Graded manifolds, graded Lie theory, and prequantization", pp. 177–306 in *Differential geometrical methods in mathematical physics* (Bonn, 1975), edited by K. Bleuler and A. Reetz, Lecture Notes in Math **570**, Springer, Berlin, 1977. MR 58 #28326 Zbl 0358.53024

[Koszul 1983] J.-L. Koszul, "Graded manifolds and graded Lie algebras", pp. 71–84 in *Proceedings of the international meeting on geometry and physics* (Florence, 1982), edited by M. Modugno, Pitagora, Bologna, 1983. MR 85m:58019 Zbl 0548.22012

[Manin 1988] Y. I. Manin, *Gauge field theory and complex geometry*, Grundlehren der Mathematischen Wissenschaften **289**, Springer, Berlin, 1988. MR 89d:32001 Zbl 0641.53001

[Masuoka 2005] A. Masuoka, "The fundamental correspondences in super affine groups and super formal groups", *J. Pure Appl. Algebra* **202**:1–3 (2005), 284–312. MR 2006e:16066 Zbl 1078.16045

[Masuoka 2012] A. Masuoka, "Harish–Chandra pairs for algebraic affine supergroup schemes over an arbitrary field", *Transform. Groups* **17**:4 (2012), 1085–1121. MR 3000482

[Serre 1992] J.-P. Serre, *Lie algebras and Lie groups*, 2nd ed., Lecture Notes in Mathematics **1500**, Springer, Berlin, 1992. MR 93h:17001 Zbl 0742.17008

[Varadarajan 2004] V. S. Varadarajan, *Supersymmetry for mathematicians: an introduction*, Courant Lecture Notes in Mathematics **11**, New York University, 2004. MR 2005g:58011 Zbl 1142.58009

[Vishnyakova 2011] E. G. Vishnyakova, "On complex Lie supergroups and split homogeneous supermanifolds", *Transform. Groups* **16**:1 (2011), 265–285. MR 2012b:58010 Zbl 1218.22013

Received September 5, 2012. Revised September 19, 2012.

CLAUDIO CARMELI
D.I.M.E., UNIVERSITÀ DI GENOVA
VIA CADORNA 2
I-17100 SAVONA
ITALY

and

INFN, SEZIONE DI GENOVA
VIA DODECANESO 33
I-16416 GENOVA
ITALY

claudio.carmeli@gmail.com

RITA FIORESI
DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ DI BOLOGNA
PIAZZA DI PORTA S. DONATO, 5
40126 BOLOGNA
ITALY

fioresi@dm.unibo.it

PACIFIC JOURNAL OF MATHEMATICS
Vol. 263, No. 1, 2013
dx.doi.org/10.2140/pjm.2013.263.53

ORBIFOLDS WITH SIGNATURE $(0; k, k^{n-1}, k^n, k^n)$

ANGEL CAROCCA, RUBÉN A. HIDALGO AND RUBÍ E. RODRÍGUEZ

Two interesting problems that arise in the theory of closed Riemann surfaces are (i) computing algebraic curves representing the surface and (ii) deciding if the field of moduli is a field of definition.

In this paper we consider pairs (S, H), where S is a closed Riemann surface and H is a subgroup of $\mathrm{Aut}(S)$, the group of automorphisms of S, so that S/H is an orbifold with signature $(0; k, k^{n-1}, k^n, k^n)$ where $k, n \geq 2$ are integers.

In the case that S is the highest abelian branched cover of S/H we provide explicit algebraic curves representing S. In the case that k is an odd prime, we also describe algebraic curves for some intermediate abelian covers.

For $k = p \geq 3$ a prime and H a p-group, we prove that H is a p-Sylow subgroup of $\mathrm{Aut}(S)$, and if $p \geq 7$ we prove that H is normal in $\mathrm{Aut}(S)$. Also, when $n \neq 3$ we prove that the field of moduli in such cases is a field of definition. If, moreover, S is the highest abelian branched cover of S/H, then we compute explicitly the field of moduli.

1. Introduction

A closed Riemann surface S of genus $g \geq 2$ may be described by many different objects, for instance, by algebraic curves (by the Riemann–Roch theorem [Farkas and Kra 1992]), by torsion-free cocompact Fuchsian groups (by the Koebe–Poincaré uniformization theorem [Koebe 1907a; 1907b; Poincaré 1907]), by Schottky groups (by the retrosection theorem [Bers 1975; Koebe 1907b]), or by certain principally polarized abelian varieties (by the Torelli theorem [Torelli 1913; Weil 1956]). In general, to provide different explicit representations for the same Riemann surface has been a difficult problem, in spite of huge efforts to solve it. It seems that Burnside [1893] and Klein [1878] provided the first examples of algebraic curves and Fuchsian groups, both representing the same Riemann surfaces. In many cases, the group $\mathrm{Aut}(S)$ of automorphisms of S and its subgroups play

Carocca and Rodríguez were supported by Fondecyt grants 1095165 and 1100767. Hidalgo was supported by Fondecyt grant 1110001 and UTFSM grant 12.11.01.
MSC2000: primary 30F10, 30F40; secondary 14H37.
Keywords: algebraic curves, Riemann Surfaces, automorphisms, field of moduli.

a fundamental role in finding algebraic curves representing S. For instance, if $S/\operatorname{Aut}(S)$ has signature of the form $(0; r, s, t)$, then there are known examples having an explicit Fuchsian group and an explicit algebraic curve, both representing S [Burnside 1893; Klein 1878] (we also recommend reading [Karcher and Weber 1999]).

A *field of definition* of S is a subfield $\mathbb{K}$ of $\mathbb{C}$ for which it is possible to find an irreducible nonsingular projective algebraic curve representing S, defined by polynomials whose coefficients belong to $\mathbb{K}$. If C is an algebraic curve describing S, then the *field of moduli* of S is defined as the fixed field of the group of field automorphisms σ of $\mathbb{C}$ such that C and C^{σ} are isomorphic, where C^{σ} is the algebraic curve defined as the zeroes of the polynomials obtained from the ones defining C after σ acts on their coefficients. The field of moduli is always contained in any field of definition, but it may happen that the field of moduli is not a field of definition.

In this article we study closed Riemann surfaces S admitting subgroups $H < \operatorname{Aut}(S)$ so that S/H has signature $(0; k, k^{n-1}, k^n, k^n)$, where $n, k \geq 2$ are integers. For $k = 2$ this type of surface was considered in [Carvacho 2010; González-Diez and Hidalgo 1997] to provide examples of closed Riemann surfaces admitting topologically equivalent but conformally nonequivalent cyclic groups of order 2^n.

In the general case, if S is the homology cover of S/H, then we compute the field of moduli and we give explicit algebraic curves for S. These explicit algebraic curves for homology covers allow us to find algebraic curves for those Riemann surfaces S admitting an abelian group $G < \operatorname{Aut}(S)$ such that S/G has signature $(0; k, k^{n-1}, k^n, k^n)$. We describe such a situation for the case that k is a prime and $G \cong \mathbb{Z}_k \times \mathbb{Z}_{k^n}$. Also, for k an odd prime, we describe the group $\operatorname{Aut}(S)$ and we prove that the field of moduli of S is in fact a field of definition.

In this article we will use letters such as S, R, $\widetilde{S}$ to denote (closed) Riemann surfaces, orbifolds will usually be denoted using italic letters such as $\mathcal{O}$, $\widetilde{\mathcal{O}}$ or as S/H (where S is a Riemann surface and $H < \operatorname{Aut}(S)$), groups will be denoted by letters such as H, Γ, G, etc.

2. Preliminaries

2.1. ***Orbifolds.*** An *orbifold* is a tuple $\mathcal{O} = (S, \{(p_1, k_1), \dots, (p_n, k_n), \dots\})$ where (i) S is a Riemann surface, called the *Riemann surface structure* of $\mathcal{O}$, (ii) $\{p_1, p_2, \dots\} \subset S$ is a collection of different isolated points, called the *cone points* of $\mathcal{O}$, and (iii) each $k_j \geq 2$ is an integer, called the *cone order* of p_j. An *orbifold* of signature $(\gamma; k_1, \dots, k_n)$ is given by an orbifold $\mathcal{O} = (S, \{(p_1, k_1), \dots, (p_n, k_n)\})$ where S is a closed Riemann surface of genus γ. An orbifold without cone points is just a Riemann surface.

A *conformal homeomorphism* between two orbifolds, say $\mathcal{O}_1 = (S_1, \{(p_1, k_1), \dots, (p_n, k_n), \dots\})$ and $\mathcal{O}_2 = (S_2, \{(q_1, l_1), \dots, (q_n, l_n), \dots\})$, is a conformal homeomorphism between S_1 and S_2 (the corresponding Riemann surface structures), sending cone points to cone points, and preserving the cone point orders. If $\mathcal{O}_1 = \mathcal{O}_2 = \mathcal{O}$, then we speak about a *conformal automorphism* of the orbifold $\mathcal{O}$. We use the notation $\mathcal{O}_1 \cong \mathcal{O}_2$ to indicate that $\mathcal{O}_1$ and $\mathcal{O}_2$ are conformally equivalent orbifolds.

We denote by $\mathrm{Aut}_{\mathrm{orb}}(\mathcal{O})$ the group of conformal automorphisms of the orbifold $\mathcal{O}$. If S is the conformal Riemann surface structure of $\mathcal{O}$, then we denote by $\mathrm{Aut}(S)$ its group of conformal automorphisms. There is a natural inclusion $\mathrm{Aut}_{\mathrm{orb}}(\mathcal{O}) < \mathrm{Aut}(S)$, but in general these two groups are different.

If $\mathcal{O}$ is an orbifold and $H < \mathrm{Aut}_{\mathrm{orb}}(\mathcal{O})$ acts discontinuously on the Riemann surface structure, then the quotient $\mathcal{O}/H$ may be seen again as an orbifold as follows. We denote by $\pi : \mathcal{O} \to \mathcal{O}/H$ the canonical quotient map. A cone point of $\mathcal{O}/H$ may be obtained in two different ways. In the first case, if $p \in \mathcal{O}$ is not a cone point and it has nontrivial H-stabilizer $H(p)$, then $\pi(p)$ is a cone point with order equal to the order of $H(p)$. In the second case, if $p \in \mathcal{O}$ is a cone point of order n and its H-stabilizer has order m, then $\pi(p)$ is a cone point with order equal to nm.

The orbifolds we consider in this paper are the *good orbifolds* in Thurston's terminology; they are obtained as quotient spaces $\mathcal{O} = \widetilde{S}/F$, where $\widetilde{S}$ is a (not necessarily closed) Riemann surface and $F < \mathrm{Aut}(\widetilde{S})$ is a discontinuous group of conformal automorphisms of $\widetilde{S}$. The cone points are those equivalence classes of points of $\widetilde{S}$ with nontrivial F-stabilizer.

2.2. *Homology covers.* Good orbifolds admit as (branched) universal cover either the Riemann sphere, the complex plane or the hyperbolic plane; this is a consequence of the classical uniformization theorem. Let us consider a good orbifold $\mathcal{O} = (S, \{(p_1, k_1), \dots, (p_n, k_n)\})$ of signature $(\gamma; k_1, \dots, k_n)$. The first (orbifold) fundamental group of $\mathcal{O}$ is

$$\text{(2-1)} \quad \pi_1^{\mathrm{orb}}(\mathcal{O}) = \Big\langle \alpha_1, \dots, \alpha_\gamma, \beta_1, \dots, \beta_\gamma, \delta_1, \dots, \delta_n : \prod_{j=1}^{\gamma} [\alpha_j, \beta_j] \prod_{k=1}^{n} \delta_k = \delta_1^{k_1} = \cdots = \delta_n^{k_n} = 1 \Big\rangle,$$

where $\pi_1(S) = \langle \alpha_1, \dots, \alpha_\gamma, \beta_1, \dots, \beta_\gamma : \prod_{j=1}^{\gamma} [\alpha_j, \beta_j] = 1 \rangle$, with $[a,b] = aba^{-1}b^{-1}$, and the element δ_j represents a simple small loop around p_j in $S - \{p_1, \dots, p_n\}$, for each $j = 1, \dots, n$.

It is clear that to each normal subgroup N of finite index of $\pi_1^{\mathrm{orb}}(\mathcal{O})$ there corresponds an orbifold $\widetilde{\mathcal{O}}$ and a finite group $H < \mathrm{Aut}_{\mathrm{orb}}(\widetilde{\mathcal{O}})$, so that $\mathcal{O} = \widetilde{\mathcal{O}}/H$. Observe that H is isomorphic to $\pi_1^{\mathrm{orb}}(\mathcal{O})/N$. When $N = \pi_1^{\mathrm{orb}}(\mathcal{O})'$ (the derived subgroup of $\pi_1^{\mathrm{orb}}(\mathcal{O})$), the corresponding cover orbifold $\widetilde{\mathcal{O}}$ is called the *homology* *orbifold cover* of $\mathbb{O}$. We will be interested only in the particular case when the homology orbifold cover is a closed Riemann surface (i.e., there are no cone points), in which case we call it the *homology cover* of $\mathbb{O}$, and say that $\mathbb{O}$ is a *homology orbifold*.

Clearly, the homology orbifold cover of $\mathbb{O}$ is the homology cover if and only if $\pi_1^{\mathrm{orb}}(\mathbb{O})'$ has finite index in $\pi_1^{\mathrm{orb}}(\mathbb{O})$ and it acts freely on the universal cover space of $\mathbb{O}$. The finite index condition is equivalent to the condition that the underlying Riemann surface structure of $\mathbb{O}$ is the Riemann sphere; that is, $\gamma = 0$. The free action condition is equivalent to the following one.

Theorem 1 [Maclachlan 1965]. *Let $\mathbb{O}$ be an orbifold of signature $(\gamma; k_1, \dots, k_n)$. Then $\pi_1^{\mathrm{orb}}(\mathbb{O})'$ is torsion-free if and only if*

$$\mathrm{lcm}(k_1, \dots, k_{j-1}, k_{j+1}, \dots, k_n) = \mathrm{lcm}(k_1, \dots, k_n) \quad \textit{for all } j = 1, \dots, n. \tag{2-2}$$

The homology cover (when it exists) is the highest abelian Galois cover of $\mathbb{O}$.

2.3. *Fuchsian groups.* The basic theory of Fuchsian groups may be found, for instance, in the classical book [Beardon 1983]. A *cocompact Fuchsian group* acting on the upper half-plane $\mathbb{H}^2$ is a discrete group $\Gamma < \mathrm{PSL}(2, \mathbb{R})$ such that $\mathbb{H}^2/\Gamma$ is an orbifold of some signature; that is, the underlying Riemann surface is a closed Riemann surface. It is known that a cocompact Fuchsian group Γ has a presentation of the form

$$\Gamma = \left\langle a_1, b_1, \dots, a_\gamma, b_\gamma, \delta_1, \dots, \delta_n : \prod_{j=1}^{\gamma} [a_j, b_j] \prod_{j=1}^{n} \delta_j = \delta_1^{k_1} = \cdots = \delta_n^{k_n} = 1 \right\rangle, \tag{2-3}$$

where γ and n are nonnegative integers, the $k_j \geq 2$ are integers, and $2\gamma - 2 + n - \sum_{j=1}^n k_j^{-1} > 0$. The tuple $(\gamma; k_1, \dots, k_n)$ is known as the *signature* of Γ (this is the signature of its quotient orbifold $\mathbb{H}^2/\Gamma$).

An orbifold $\mathbb{O}$ is *of hyperbolic type* if there is a cocompact Fuchsian group Γ so that $\mathbb{O} \cong \mathbb{H}^2/\Gamma$. By the Poincaré–Koebe uniformization theorem [Koebe 1907a; 1907b; Poincaré 1907], every orbifold with signature $(\gamma; k_1, \dots, k_n)$ is of hyperbolic type if and only if $2\gamma - 2 + n - \sum_{j=1}^n k_j^{-1} > 0$.

By the hyperbolic area of a Fuchsian group Γ (respectively, of a hyperbolic orbifold) of signature $(\gamma; k_1, \dots, k_n)$ we refer to the hyperbolic area of a fundamental polygon domain for it; it is given by

$$A(\Gamma) = 2\pi \left(2\gamma - 2 + \sum_{j=1}^{n} \left(1 - \frac{1}{k_j} \right) \right). \tag{2-4}$$

We say that a cocompact Fuchsian group Γ, with presentation (2-3), is a *homology Fuchsian group* if $\gamma = 0$ and it satisfies Maclachlan's conditions (2-2). In other words, homology Fuchsian groups are exactly those cocompact Fuchsian groups providing a Fuchsian uniformization of a hyperbolic homology orbifold of genus zero. If Γ is a homology Fuchsian group of signature $(0;k_1,\dots,k_n)$, then the homology cover of the homology orbifold $\mathbb{O}=\mathbb{H}^2/\Gamma$ is $S=\mathbb{H}^2/\Gamma'$, where Γ' denotes the derived subgroup of Γ.

2.4. ***Fields of moduli and fields of definition.*** As a consequence of the implicit function theorem, every irreducible nonsingular projective algebraic curve defines a closed Riemann surface; conversely, by the Riemann–Roch theorem, every closed Riemann surface may be described by an irreducible nonsingular projective algebraic curve. It is this equivalence which allows the work in the analytical and in the algebraic settings in a parallel way.

Let C be an irreducible nonsingular projective algebraic curve, say defined by homogeneous polynomials $P_1,\dots,P_r$, each one with coefficients in a subfield $\mathbb{K}<\mathbb{C}$. Let g denote the genus of the closed Riemann surface corresponding to C. If $\sigma\in\mathrm{Aut}(\mathbb{C}/\mathbb{Q})$, the group of field automorphisms of $\mathbb{C}$, then we may consider the new polynomials $P_1^{\sigma},\dots,P_r^{\sigma}$, where the coefficients of P_j^{σ} are the corresponding images under σ of the coefficients of the original polynomial P_j. The algebraic curve C^{σ}, defined by these new polynomials, is still an irreducible nonsingular projective algebraic curve, and it defines a new closed Riemann surface of genus g. It is not difficult to see that if $\widetilde{C}$ is another irreducible nonsingular projective algebraic curve that is birationally equivalent to C, then C^{σ} and $\widetilde{C}^{\sigma}$ are also birationally equivalent. Therefore, a natural action of $\mathrm{Aut}(\mathbb{C}/\mathbb{Q})$ is defined on the moduli space of genus g. The stabilizer of the moduli class of C under such action is the subgroup

$$K_C=\{\sigma\in\mathrm{Aut}(\mathbb{C}/\mathbb{Q}):C\cong C^{\sigma}\}<\mathrm{Aut}(\mathbb{C}/\mathbb{Q}).$$

The fixed field of K_C, denoted by $\mathbb{M}(C)$, is called the *field of moduli* of C.

A subfield $\mathbb{K}$ of $\mathbb{C}$ is called a *field of definition* of C if there is an irreducible nonsingular projective algebraic curve $\widetilde{C}$ defined over $\mathbb{K}$ which is birationally equivalent to C. At this point it is important to note that it is not clear that given a field of definition $\mathbb{L}<\mathbb{C}$ of C there is a smaller subfield $\mathbb{F}<\mathbb{L}$ which is again a field of definition of C.

The field of moduli $\mathbb{M}(C)$ is contained in any field of definition of C, and it coincides with the intersection of all fields of definition of C [Koizumi 1972]. Moreover, there is a field of definition of C which is an extension of finite degree of the field of moduli [Dèbes and Emsalem 1999; Hammer and Herrlich 2003].

If $g=0$, then $C\cong\mathbb{P}^1$, so in this case $\mathbb{M}(C)=\mathbb{Q}$ is a field of definition. If $g=1$, then C is equivalent to an (affine) elliptic curve $E_\eta=\{y^2=x(x-1)(x-\eta)\}$, where $\eta\in\mathbb{C}-\{0,1\}$. If $j(\eta)=(1-\eta+\eta^2)^3/\eta^2(\eta-1)^2$ is its j-invariant and

$$a(\eta) = \frac{27j(\eta)}{j(\eta)-1},$$

then E_η is also described by $D_\eta = \{y^2 = 4x^3 - a(\eta)x - a(\eta)\}$. It follows that $\mathbb{Q}(j(\eta))$ is a field of definition for E_η. Moreover, if $\sigma \in \mathrm{Aut}(\mathbb{C}/\mathbb{Q})$ and $E_\eta^\sigma = E_{\sigma(\eta)}$ is conformally equivalent to E_η, then they must have the same j-invariant; that is, $\sigma(j(\eta)) = j(\eta)$. It follows that $\mathbb{M}(C) = \mathbb{M}(E_\eta) = \mathbb{Q}(j(\eta))$ is also a field of definition.

In genus $g \geq 2$, the situation is more difficult. There are examples for which the field of moduli is not a field of definition [Earle 1971; Huggins 2007; Shimura 1972]; all of the examples there are hyperelliptic curves. It is stated in [Earle 1971] that there are examples of nonhyperelliptic Riemann surfaces with the same properties, but no explicit one is given. An explicit example of a nonhyperelliptic Riemann surface of genus $g = 17$ which cannot be defined over $\mathbb{R}$ and whose field of moduli lies inside $\mathbb{R}$ is given in [Hidalgo 2009] (this example is related to the hyperelliptic example in [Earle 1971]).

A. Weil [1956] provided the following sufficient and necessary conditions for the moduli field to be a field of definition.

Theorem 2 [Weil 1956]. *Let C be an irreducible nonsingular projective algebraic curve defined over a finite Galois extension $\mathbb{L}$ of its field of moduli $\mathbb{M}(C)$. If for every $\sigma \in \mathrm{Aut}(\mathbb{L}/\mathbb{M}(C))$ there is a biholomorphism $f_\sigma : C \to C^\sigma$ defined over $\mathbb{L}$ such that the compatibility condition $f_{\tau\sigma} = f_\sigma^\tau \circ f_\tau$ holds for all $\sigma, \tau \in \mathrm{Aut}(\mathbb{L}/\mathbb{M}(C))$, then there exists an irreducible nonsingular projective algebraic curve E defined over $\mathbb{M}(C)$ and there exists a biregular map $F : C \to E$, defined over $\mathbb{L}$, such that $F^\sigma \circ f_\sigma = F$.*

As a consequence of Theorem 2, it follows that if C has no nontrivial automorphism, then it may be defined over its field of moduli. Unfortunately, if C has nontrivial automorphisms, then it is a very difficult task to check whether Weil's conditions hold. But if $C/\mathrm{Aut}(C)$ has signature of the form $(0; a, b, c)$ (quasiplatonic surfaces, or platonic if some cone order is equal to 2), then C may be defined over its field of moduli [Coombes and Harbater 1985; Wolfart 2006].

Consider a (branched) holomorphic covering between closed Riemann surfaces, say $f : S \to R$. Assume S and R are given by fixed algebraic curves and that R is defined over $\mathbb{M}(S)$. For each $\sigma \in \mathrm{Aut}(\mathbb{C}/\mathbb{M}(S))$ we may consider the (branched) holomorphic covering $f^\sigma : S^\sigma \to R^\sigma = R$. We say that they are *equivalent coverings*, denoted by $\{f^\sigma : S^\sigma \to R\} \cong \{f : S \to R\}$, if there is a holomorphic isomorphism $\phi_\sigma : S \to S^\sigma$ so that $f^\sigma \circ \phi_\sigma = f$. The *field of moduli* of $f : S \to R$, denoted by $\mathbb{M}(f : S \to R)$, is the fixed field of the subgroup

$$K(f : S \to R) = \bigl\{\sigma \in \mathrm{Aut}(\mathbb{C}/\mathbb{M}(S)) : \{f^\sigma : S^\sigma \to R\} \cong \{f : S \to R\}\bigr\}.$$

It is clear from the definition that $\mathbb{M}(S) < \mathbb{M}(f : S \to R)$, but in general they may be different fields. For the particular case that $R = S/\operatorname{Aut}(S)$ and S has genus at least two, the following is well known (a direct consequence of Theorem 2).

Theorem 3 [Dèbes and Emsalem 1999]. *If C is an irreducible nonsingular projective algebraic curve of genus $g \geq 2$, then there exists an irreducible nonsingular projective algebraic curve C_1, defined over $\mathbb{M}(C)$, and there exists a Galois cover $f : C \to C_1$, with $\operatorname{Aut}(C)$ as deck group, so that $\mathbb{M}(f : C \to C_1) = \mathbb{M}(C)$. Moreover, if $(C_1)_f$ denotes the branch locus of f and if $C_1 - (C_1)_f$ contains at least one $\mathbb{M}(C)$-rational point, then $\mathbb{M}(C)$ is also a field of definition of C. Such a curve C_1 is called a canonical model of $C/\operatorname{Aut}(C)$.*

3. Main results

Let S be a closed Riemann surface and let H_1, $H_2 < \operatorname{Aut}(S)$. We say that H_1 and H_2 are (weakly) *topologically equivalent* (respectively, *conformally equivalent*) if there is an orientation preserving self-homeomorphism (respectively, conformal automorphism) $h : S \to S$ so that $H_2 = f H_1 f^{-1}$. If $H < \operatorname{Aut}(S)$, then we denote by $\operatorname{Aut}_H(S)$ the normalizer of H in $\operatorname{Aut}(S)$.

3.1. ***p-groups of automorphisms.*** A *regular cover* of an orbifold $\mathbb{O}$ is a closed Riemann surface S together with a group H of conformal automorphisms such that the quotient orbifold S/H is isomorphic to $\mathbb{O}$. In the case that H is an abelian group, we say that the regular cover is an *abelian cover* of the orbifold. In this section we consider regular p^{n+1}-covers of orbifolds of type $(0; p, p^{n-1}, p^n, p^n)$, where $n \geq 2$ and p is an odd prime; that is, H is a p-group of order p^{n+1}. The interest in this type of example is that examples were constructed in [Carvacho 2010; González-Diez and Hidalgo 1997] of closed Riemann surfaces S admitting topologically equivalent but conformally nonequivalent cyclic groups of order 2^{n+1}, where $n \geq 2$, so the quotient of S by the 2-group generated by these two cyclic subgroups is an orbifold with signature $(0; 2, 2^n, 2^{n+1}, 2^{n+1})$.

Let S be a closed Riemann surface and let $H < \operatorname{Aut}(S)$ be a p-group such that S/H has signature of the form $(0; p, p^{n-1}, p^n, p^n)$, with $n \geq 2$, and consider the regular branched cover $P : S \to \widehat{\mathbb{C}}$, with H as deck group.

Since $n \geq 3$, then (up to left composition by a suitable Möbius transformation) we may assume that the branch values of P are ∞ of order p, 0 of order p^{n-1}, and 1 and some $\lambda \in \mathbb{C} - \{0, 1\}$ are the ones of order p^n. The choice of λ is not unique, but the only other possible choice is $1/\lambda$.

Theorem 4. *Let $p \geq 3$ be a prime and let $n \geq 2$ be an integer. Consider a closed Riemann surface S with a subgroup $H < \operatorname{Aut}(S)$ such that H is a p-group with S/H of signature $(0; p, p^{n-1}, p^n, p^n)$. Let $\lambda \in \mathbb{C} - \{0, 1\}$ be as defined above. Then:*

(1) *H is a p-Sylow subgroup of* $\mathrm{Aut}(S)$. *In particular, if $H_1, H_2 < \mathrm{Aut}(S)$ are p-groups with S/H_j of signature* $(0; p, p^{n-1}, p^n, p^n)$, *then H_1 and H_2 are conformally equivalent.*

(2) *If $n \geq 3$, then*
  (a) $\mathrm{Aut}_H(S) = H$ *for* $\lambda \neq -1$,
  (b) $[\mathrm{Aut}_H(S) : H] \in \{1, 2\}$ *for* $\lambda = -1$.

(3) *If $n = 2$, then*
  (a) $[\mathrm{Aut}_H(S) : H] \in \{1, 2\}$ *for* $\lambda \neq -1$,
  (b) $[\mathrm{Aut}_H(S) : H] \in \{1, 2, 4\}$ *for* $\lambda = -1$.

(4) *If $p \geq p_0$, where*
  (a) $p_0 = 7$ *for $n = 2$, and*
  (b) $p_0 = 5$ *for $n \geq 3$,*

  then $\mathrm{Aut}_H(S) = \mathrm{Aut}(S)$.

Remark 5. In the case $\lambda = -1$ and $n \geq 3$, part (2) of Theorem 4 asserts that either $\mathrm{Aut}_H(S) = H$ or $[\mathrm{Aut}_H(S) : H] = 2$. In the last case, $S/\mathrm{Aut}_H(S)$ has signature $(0; 2p, 2p^{n-1}, p^n)$, which is a maximal signature [Singerman 1972], so $\mathrm{Aut}_H(S) = \mathrm{Aut}(S)$.

3.2. *Normality condition.* Let S be a closed Riemann surface and $H < \mathrm{Aut}(S)$. Let $\mathcal{M}(S, H)$ denote the locus in the moduli space $\mathcal{M}(S)$ of S consisting of those classes of Riemann surfaces $\widehat{S}$ admitting a group $\widehat{H}$ of conformal automorphisms, which is topologically equivalent to H. In general, one should expect that $\mathcal{M}(S, H)$ is a singular variety. The following shows that this is not the case if H is a p-group and S/H has signature $(0; p, p^{n-1}, p^n, p^n)$.

Corollary 6. *Let $p \geq 3$ be a prime and let $n \geq 2$ be an integer. Consider a closed Riemann surface S and let $H <$* $\mathrm{Aut}(S)$ *be a p-group such that S/H has signature* $(0; p, p^{n-1}, p^n, p^n)$. *Then $\mathcal{M}(S, H)$ is a normal subvariety of $\mathcal{M}(S)$.*

Proof. The normality condition for $\mathcal{M}(S, H)$ is equivalent to the following property: given any two pairs (S_1, H_1) and (S_2, H_2), where S_j is a closed Riemann surface (of the same genus as S) and H_j is a p-group of conformal automorphisms of S_j so that S_j/H_j has signature $(0; p, p^{n-1}, p^n, p^n)$, and there is an orientation preserving homeomorphism $f : S_1 \to S_2$ with $f H_1 f^{-1} = H_2$, then f may be replaced by a biholomorphism with the same properties. This property is exactly what part (1) of Theorem 4 states. □

3.3. *Homology rigidity.*

Corollary 7. *Every orbifold of signature* $(0; p, p^{n-1}, p^n, p^n)$, *where $p \geq 3$ is a prime and $n \geq 2$ is an integer, is uniquely determined, up to conformal equivalence, by its homology cover Riemann surface.*

Proof. A consequence of part (1) of Theorem 4. □

Remark 8 (Torelli's theorem). Let $\mathcal{O}$ be an orbifold of signature $(0; p, p^{n-1}, p^n, p^n)$, where $p \geq 3$ is a prime and $n \geq 2$ is an integer. As any two homology covers of $\mathcal{O}$ are conformally equivalent Riemann surfaces, we may define the Jacobian of $\mathcal{O}$, denoted by $J(\mathcal{O})$, as the Jacobian of any of these covers. It follows that $J(\mathcal{O})$ is uniquely determined, up to equivalence of principally polarized abelian varieties, by $\mathcal{O}$. As a consequence of Torelli's theorem, $J(\mathcal{O})$ determines the conformal class of the homology cover of $\mathcal{O}$ and, by Corollary 7, it also determines the conformal class of $\mathcal{O}$. In this way, a kind of Torelli's theorem is obtained for this class of orbifolds. We may wonder how to describe the Jacobian of $\mathcal{O}$ in terms of multivalued holomorphic differential forms so that it looks more similar to the construction for the case of Riemann surfaces. In order to do this, we use as homology the orbifold homology group

$$H_1^{\mathrm{orb}}(\mathcal{O}) = \pi_1^{\mathrm{orb}}(\mathcal{O}) / \pi_1^{\mathrm{orb}}(\mathcal{O})',$$

and as holomorphic forms those multivalued holomorphic forms whose liftings to the homology cover define the holomorphic one forms of it.

3.4. ***Algebraic curves in the abelian case.*** Curves for the hyperelliptic homology covers and for the homology covers of homology orbifolds with triangular signature have been described in [Hidalgo 2012]. Algebraic curves for the homology covers of orbifolds with signature of the form $(0; k, \ldots, k)$ have been obtained in [González-Diez et al. 2009]. We next provide the algebraic curves for the homology covers of orbifolds with signature $(0; k, k^{n-1}, k^n, k^n)$, where $k, n \geq 2$ are integers. As a consequence of the results in [Hidalgo 2012], the homology covers of such orbifolds cannot be hyperelliptic. Note that if R is the homology cover of such an orbifold $\mathcal{O}$, then $\mathcal{O} = R/H$, where $H \cong \mathbb{Z}_k \times \mathbb{Z}_{k^{n-1}} \times \mathbb{Z}_{k^n}$.

Theorem 9. *Let k, $n \geq 2$ be integers and let $\mathcal{O}$ be an orbifold with signature $(0; k, k^{n-1}, k^n, k^n)$. Denote by R a homology cover of $\mathcal{O}$, let $H < \mathrm{Aut}(R)$ be so that $R/H = \mathcal{O}$, and let $P : R \to \mathcal{O}$ be the Galois cover with H as deck group. We may assume (up to a Möbius transformation) that the cone points of $\mathcal{O}$ (that is, the branch values of P) are given by the points 0, 1, ∞ and $\lambda \in \mathbb{C} - \{0, 1\}$. We may also assume that ∞ is the cone point of order k, that 0 is the cone point of order k^{n-1} and that 1 and λ are the cone points of order k^n.*

Then R is represented by the (singular) projective algebraic curve

$$C_\lambda : \left\{ \begin{array}{r} z_0^k z_3^{k^n - k} + z_1^{k^{n-1}} z_3^{k^n - k^{n-1}} + z_2^{k^n} = 0 \\ \lambda z_0^k z_3^{k^{n-1} - k} + z_1^{k^{n-1}} + z_3^{k^{n-1}} = 0 \end{array} \right\} \subset \mathbb{P}^3,$$

H is generated by the projective linear transformations

$$\begin{aligned}a_0([z_0 : z_1 : z_2 : z_3]) &= [\rho_1 z_0 : z_1 : z_2 : z_3],\\ b_0([z_0 : z_1 : z_2 : z_3]) &= [z_0 : \rho_{n-1} z_1 : z_2 : z_3],\\ c_0([z_0 : z_1 : z_2 : z_3]) &= [z_0 : z_1 : \rho_n z_2 : z_3],\end{aligned}$$

where $\rho_s = e^{2\pi i/k^s}$, for each positive integer s, and the branched covering map P is represented in this model by

$$P([z_0 : z_1 : z_2 : z_3]) = -\left(\frac{z_1^{k^{n-1}}}{z_0^k z_3^{k^{n-1}-k}}\right).$$

The only singular point of the above curve is $[1 : 0 : 0 : 0]$.

Theorem 9 may be used to find algebraic curves for closed Riemann surfaces S admitting an abelian group $G < \mathrm{Aut}(S)$ whose quotient orbifold S/G has signature of the form $(0; k, k^{n-1}, k^n, k^n)$. In fact, let $Q : S \to S/G = \mathbb{O}$ be a regular abelian branched cover with G as deck group. Let R be the homology cover of $\mathbb{O}$, let $P : R \to \mathbb{O}$ be the regular abelian branched cover, with deck group $H < \mathrm{Aut}(R)$. Then there exists a subgroup $K < H$, acting freely on R and so that $G \cong H/K$, and there exists a regular unbranched cover $F : R \to S$, with K as deck group, satisfying $P = Q \circ F$. As we have explicit curves for R and an explicit presentation for H, the classical invariant theory permits us to obtain explicit algebraic curves for S and an explicit presentation of G. We show an application in the next section.

3.5. ***Families with Galois group of order p^{n+1}.*** As mentioned before, we are interested in regular p^{n+1}-covers of orbifolds of type $(0; p, p^{n-1}, p^n, p^n)$, where $n \geq 2$ and p is an odd prime. In Section 9 we will see that the algebraic structure of the corresponding groups of order p^{n+1} is restricted to only two algebraic types: a direct or a semidirect product of $\mathbb{Z}_{p^n}$ and $\mathbb{Z}_p$. The geometric types (classified by either geometric signature or generating vector for the corresponding action) are more varied: four different types are found in each algebraic case.

We study the corresponding families of Riemann surfaces, giving their algebraic curves in the abelian case.

The next result makes the above more explicit for the case when $G \cong \mathbb{Z}_p \times \mathbb{Z}_{p^n}$, where p is a prime. As we will see in its proof, this is a heavy computational procedure, but not a hard one.

Theorem 10. *Let S be a closed Riemann surface admitting a group $G < \mathrm{Aut}(S)$ such that $G = \langle A, B : A^p = B^{p^n} = [A, B] = 1\rangle \cong \mathbb{Z}_p \times \mathbb{Z}_{p^n}$ and $\mathbb{O} = S/G$ is an orbifold with signature $(0; p, p^{n-1}, p^n, p^n)$, where $n \geq 2$ and p is an odd prime. Let R be a homology cover of $\mathbb{O}$, let $H < \mathrm{Aut}(R)$ be so that $R/H = \mathbb{O}$. Let $K < H$ be the normal subgroup so that $S = R/K$ and $G = H/K$.*

(1) *If* $K \cong \mathbb{Z}_{p^{n-1}}$, *there exist* $\beta \in \{1, 2, \ldots, p^{n-1}-1\}$, $\alpha \in \{0, 1, \ldots, p-1\}$ *and* $q \in \{1, \ldots, [(p^n-1)/p]\}$, *with* $(\beta, p) = 1 = (p, q)$, *such that a (singular) projective algebraic curve representation of* S *is given by one of the following two families.*

(a) *If* $\alpha = 0$, *there exists* λ *in* $\mathbb{C}$, *with* $\lambda \neq 0, 1$, *such that*

$$S : \left\{ \begin{array}{r} (\lambda - 1)w_0^p - w_1^p + w_3^p = 0 \\ (-1)^{q+1}(w_0^p + w_1^p)^q w_1^{p^{n-1}-\beta} + w_2^{p^{n-1}} w_3^{qp-\beta} = 0 \end{array} \right\} \subset \mathbb{P}^3$$

and the action of G *is generated by the projective linear transformations*

$$A([w_0 : w_1 : w_2 : w_3]) = [\rho_1 w_0 : w_1 : w_2 : w_3],$$
$$B([w_0 : w_1 : w_2 : w_3]) = [w_0 : \rho_1 w_1 : \rho_n^{p^{n-1}-\beta} w_2 : w_3],$$

where $\rho_k = e^{2\pi i/p^k}$. *The regular branched covering map* $Q : S \to S/G$ *in this model is represented by*

$$Q([w_0 : w_1 : w_2 : w_3]) = \frac{w_0^p + w_1^p}{w_0^p}.$$

The singular points of the above curve are given by the $(p+1)$ *points* $[0 : 0 : 1 : 0]$ *and* $[1 : 0 : 0 : (1-\lambda)^{1/p}]$.

(b) *If* $\alpha > 0$, *there exists* λ *in* $\mathbb{C}$, *with* $\lambda \neq 0, 1$, *such that*

$$S : \left\{ \begin{array}{r} v_1^{p^{n-1}} + \dfrac{(-1)^{q+1}}{(\lambda-1)^q}(\lambda v_1^p - v_3^p)^q v_1^{p^{n-1}-\beta} v_3^{\beta - pq} = 0 \\ v_2^p v_3^{p(p^r-\beta)+\alpha p - p} + \dfrac{(-1)^{\alpha+1}}{(\lambda-1)^{\alpha+p^r-\beta}}(v_0^p - v_3^p)^{p^r-\beta}(\lambda v_0^p - v_3^p)^\alpha = 0 \end{array} \right\} \subset \mathbb{P}^3$$

and the group G *is generated by the transformations*

$$A([v_0 : v_1 : v_2 : v_3]) = [v_0 : v_1 : \rho_1^{p^r-\beta} v_2 : v_3],$$
$$B([v_0 : v_1 : v_2 : v_3]) = [\rho_n^{p^{n-1}} v_0 : \rho_n^{p^{n-1}-\beta} v_1 : v_2 : v_3].$$

The regular branched covering map $Q : S \to S/G$ *in this model is represented by*

$$Q([v_0 : v_1 : v_2 : v_3]) = \frac{\lambda v_0^p - v_3^p}{v_0^p + v_3^p}.$$

(2) *If* $K \cong \mathbb{Z}_{p^{n-2}} \times \mathbb{Z}_p$, *there exist* λ *in* $\mathbb{C}$, *with* $\lambda \neq 0, 1$, *and integers* $\gamma, \upsilon \in \{1, \ldots, p-1\}$ *such that a (singular) projective algebraic curve representation of* S

is provided by the plane projective curve

$$\left\{\frac{(-1)^{p^{n-1}(p-\gamma)}}{\lambda^{p^{n-1}(p-\gamma)+1}}\left(u_0^p+u_2^p\right)^{p^{n-1}(p-\gamma)}u_1^{p^2v}\left((\lambda-1)u_0^p-u_2^p\right) + u_1^{p^n}u_2^{p^n(p-\gamma-1)+p+p^2v}=0\right\}\subset\mathbb{P}^2,$$

and the group G is generated by the transformations

$$A([u_0:u_1:u_2])=[\rho_1 u_0:u_1:u_2],\quad B([u_0:u_1:u_2])=[u_0:\rho_n u_1:u_2].$$

The regular branched covering map $Q:S\to S/G$ in this model is represented by

$$Q([u_0:u_1:u_2])=\frac{\lambda u_0^p}{u_0^p+u_1^p}.$$

3.6. ***Field of moduli.*** If S is a closed Riemann surface, then it follows from the Riemann–Roch theorem that S may be described by an irreducible nonsingular projective algebraic curve C. It is clear from the definition that we may define the field of moduli of S as the field of moduli of C and a field of definition of S as a field of definition of C.

Theorem 11. *Let $p\geq 3$ be a prime, $n\geq 3$ be an integer, S be a closed Riemann surface, and $H<\mathrm{Aut}(S)$ be a p-group with S/H of signature* $(0;p,p^{n-1},p^n,p^n)$. *Then S may be defined over its field of moduli.*

Remark 12. Under the hypotheses of Theorem 11, if $\mathrm{Aut}_{\mathrm{orb}}(S/H)$ is nontrivial, then S/H admits an extra conformal involution J such that $(S/H)/\langle J\rangle$ is the orbifold whose underlying Riemann surface is $\widehat{\mathbb{C}}$, with exactly three cone points (of orders $2p$, $2p^{n-1}$ and p^n). It follows that S is a Belyi curve and hence it may be defined over a finite extension of $\mathbb{Q}$.

Our next result computes the field of moduli for the homology covers of orbifolds with signature $(0;p,p^{n-1},p^n,p^n)$, where $p\geq 3$ is a prime and $n\geq 2$.

Theorem 13. *Let $p\geq 3$ be a prime and $n\geq 2$ be an integer. For each $\lambda\in\mathbb{C}-\{0,1\}$, let C_λ be as in Theorem 9 with $k=p$. Then*:

(1) $C_\lambda\cong C_\mu$ *for* $\lambda,\mu\in\mathbb{C}-\{0,1\}$ *if and only if* $\mu\in\{\lambda,1/\lambda\}$.

(2) $\mathbb{M}(C_\lambda)=\mathbb{Q}(\lambda+\lambda^{-1})$.

(3) $\mathbb{M}(C_\lambda)$ *is a field of definition for* C_λ.

Theorem 13 will be proved using arguments similar to those given by Dèbes and Emsalem in the proof of Theorem 3. In our case, we do not consider the quotient by the full group of automorphisms, but just the quotient by the abelian group H in Theorem 9.

4. Proof of Theorem 4

Proof of part 4. As previously noted, there is a regular branched cover $P : S \to \widehat{\mathbb{C}}$, with H as deck group, so that its branch values are ∞ of order p, 0 of order p^{n-1}, 1 of order p^n and λ of order p^n. Let us denote by $\mathbb{O}_\lambda$ the orbifold whose underlying Riemann surface is $\widehat{\mathbb{C}}$ and whose cone points are ∞ of order p, 0 of order p^{n-1}, 1 of order p^n and λ of order p^n; that is, $\mathbb{O}_\lambda = S/H$.

If H is not a p-Sylow subgroup, then there is some $H \lhd K < \mathrm{Aut}(S)$, where K is a p-group and $[K : H] = p$. It follows that there is an automorphism of order $p \geq 3$ of the orbifold $\mathbb{O}_\lambda$. As there are no three cone points with the same order, this is impossible. □

Proof of parts (2) *and* (3). If $n \geq 3$, then it is easy to see that

$$\mathrm{Aut}_{\mathrm{orb}}(\mathbb{O}_\lambda) = \begin{cases} \{I\}, & \lambda \in \mathbb{C} - \{0, \pm 1\}, \\ \langle \tau(z) = -z \rangle, & \lambda = -1. \end{cases}$$

Since $\mathrm{Aut}_H(S)/H < \mathrm{Aut}_{\mathrm{orb}}(\mathbb{O}_\lambda)$, it follows that

$$\mathrm{Aut}_H(S) = \begin{cases} H, & \lambda \in \mathbb{C} - \{0, \pm 1\}, \\ K, & \lambda = -1, \end{cases}$$

where $[K : H] \in \{1, 2\}$.

If $n = 2$, then

$$\mathrm{Aut}_{\mathrm{orb}}(\mathbb{O}_\lambda) = \begin{cases} \langle \alpha(z) = \lambda/z \rangle, & \lambda \in \mathbb{C} - \{0, \pm 1\}, \\ \langle \tau(z) = -z, \beta(z) = -1/z \rangle, & \lambda = -1. \end{cases}$$

Again as $\mathrm{Aut}_H(S)/H < \mathrm{Aut}_{\mathrm{orb}}(\mathbb{O}_\lambda)$, it follows that

$$\mathrm{Aut}_H(S) = \begin{cases} \widehat{H}, & \lambda \in \mathbb{C} - \{0, \pm 1\}, \\ \widehat{K}, & \lambda = -1, \end{cases}$$

where $[\widehat{H} : H] \in \{1, 2\}$ and $[\widehat{K} : H] \in \{1, 2, 4\}$. □

Proof of part (4). As a consequence of the results in [Leyton and Hidalgo 2007], there exists a prime p_0 such that the group H is a normal subgroup in $\mathrm{Aut}(S)$ for $p \geq p_0$; that is, $\mathrm{Aut}(S) = \mathrm{Aut}_H(S)$. Next, we proceed to prove that p_0 may be chosen as desired.

Let $p \geq 3$ be any odd prime. We already know that H is a p-Sylow subgroup of $\mathrm{Aut}(S)$ and that S/H has signature $(0; p, p^{n-1}, p^n, p^n)$. If $S/\mathrm{Aut}(S)$ has signature of the form $(0; a, b, c, d)$, then it follows from Singerman's list [1972] of maximal Fuchsian groups that $(0; a, b, c, d) = (0; p, p^{n-1}, p^n, p^n)$ and, in particular, that $H = \mathrm{Aut}(S)$.

Thus we need only take care of the case when $S/\mathrm{Aut}(S)$ has signature of the form $(0; r, s, t)$. In this case, at least one of the values r, s, t should be a multiple of

p^n. We may assume $t = kp^n$, where k is a positive integer. We may also assume that $2 \le r \le s$ and, moreover, that if $r = 2$, then $s \ge 3$. Let

$$D = [\mathrm{Aut}(S) : H].$$

If $D = 2$, then clearly $\mathrm{Aut}_H(S) = \mathrm{Aut}(S)$.

From now on assume that $D \ge 3$. Riemann–Hurwitz (hyperbolic area comparison) asserts that

$$D\left(1 - \frac{1}{r} - \frac{1}{s} - \frac{1}{kp^n}\right) = 2 - \frac{1}{p} - \frac{1}{p^{n-1}} - \frac{2}{p^n}, \tag{4-1}$$

where both sides are necessarily positive.

Lemma 14. *If either*

(1) $p \ge 7$, *or*

(2) $p \in \{3, 5\}$ *and* $n \ge 3$,

then $D \le 11$.

Proof. Assume $D \ge 12$. As $(r, s) \ne (2, 2)$, it follows from (4-1) that

$$D\left(\frac{1}{6} - \frac{1}{kp^n}\right) \le 2 - \frac{1}{p} - \frac{1}{p^{n-1}} - \frac{2}{p^n}.$$

Since the quantity in parentheses is positive, the last inequality implies that

$$k \le \frac{12}{2 + p + p^{n-1}}.$$

Therefore, if $p \ge 7$ then

$$k \le \frac{12}{2 + p + p^{n-1}} \le \frac{12}{2 + 2p} \le \frac{3}{4} < 1,$$

and if $p \in \{3, 5\}$ and $n \ge 3$ then

$$k \le \frac{12}{2 + p + p^{n-1}} \le \frac{12}{2 + 3 + 3^2} \le \frac{6}{7} < 1,$$

obtaining a contradiction in all cases. □

The following proposition gives the desired result.

Proposition 15.

(1) *If* $n \ge 2$, *then* $p_0 \le 7$.

(2) *If* $n \ge 3$, *then* $p_0 \le 5$.

Proof. Let us denote by N_p be the number of p-Sylow subgroups of $\mathrm{Aut}(S)$. We need to prove that $N_p = 1$, if either (i) $p \geq 7$ is prime and $n \geq 2$ or (ii) $p \geq 5$ is a prime and $n \geq 3$.

As $N_p \equiv 1 \mod p$, we may write $N_p = 1 + pL_p$, where L_p is a nonnegative integer.

If we assume that $N_p > 1$, then $N_p \geq 1 + p$. As N_p divides $|\mathrm{Aut}(S)| = D|H|$, it follows that N_p must divide D.

If $p \geq 11$, then $N_p \geq 12$; as $D \leq 11$ by Lemma 14, we obtain a contradiction.

For the remaining cases, we will make use of the following equality, obtained from (4-1):

$$\left(D\left(1 - \frac{1}{r} - \frac{1}{s}\right) - 2\right)p^n + p^{n-1} + p + 2 = \frac{D}{k} \in \{1, \ldots, D\}. \tag{4-2}$$

Note that both sides in this equality are positive integers.

If $p = 7$, since $D \leq 11$ by Lemma 14, we must have that $L_7 = 1$ and $N_7 = D = 8$. If either $r, s \geq 3$ or $r = 2$ and $s \geq 4$, then

$$\left(8\left(1 - \frac{1}{r} - \frac{1}{s}\right) - 2\right) \geq 0$$

and the left side of (4-2) is bigger than 8, a contradiction to the fact that the right side should be less than or equal to D.

We are left with the case $r = 2$ and $s = 3$. But in this case the left side of (4-2) equals

$$\left(8\left(1 - \frac{1}{r} - \frac{1}{s}\right) - 2\right)7^n + 7^{n-1} + 9 < 0,$$

again a contradiction.

Now we consider $p = 5$ and $n \geq 3$. In this case either (i) $L_5 = 1$ and $N_5 = D = 6$ or (ii) $L_5 = 2$ and $N_5 = D = 11$.

For $D = 6$, if either (a) $r, s \geq 3$ or (b) $r = 2$ and $s \geq 6$, then

$$\left(6\left(1 - \frac{1}{r} - \frac{1}{s}\right) - 2\right) \geq 0$$

and the left side of (4-2) is bigger than D, a contradiction. The remaining cases are $r = 2$ and $3 \leq s \leq 5$. But in these cases we have

$$\left(6\left(1 - \frac{1}{r} - \frac{1}{s}\right) - 2\right)5^n + 5^{n-1} + 7 < 0,$$

again a contradiction.

For $D = 11$, if either (a) $r, s \geq 3$ or (b) $r = 2$ and $s \geq 4$, then

$$\left(11\left(1 - \frac{1}{r} - \frac{1}{s}\right) - 2\right) \geq 0$$

and the left side of (4-2) is bigger than D, a contradiction. The remaining cases are $r = 2$ and $s = 3, 4$. But in these cases we have

$$\left(11\left(1-\frac{1}{r}-\frac{1}{s}\right)-2\right)5^n + 5^{n-1} + 7 < 0,$$

again a contradiction. □

5. Proof of Theorem 9

Let R be the homology cover of an orbifold $\mathcal{O}$ with signature $(0; k, k^{n-1}, k^n, k^n)$, where $k, n \geq 2$. The closed Riemann surface R admits a group $H < \mathrm{Aut}(R)$, where $H \cong \mathbb{Z}_k \times \mathbb{Z}_{k^{n-1}} \times \mathbb{Z}_{k^n}$ and such that $R/H = \mathcal{O}$.

First consider the orbifold $\mathcal{O}^*$ obtained from $\mathcal{O}$, but assuming all cone points are of order k^n. The homology cover of this new orbifold is a closed Riemann surface S admitting a group $H^* < \mathrm{Aut}(S)$, $H^* \cong \mathbb{Z}_{k^n} \times \mathbb{Z}_{k^n} \times \mathbb{Z}_{k^n}$, and such that $\mathcal{O}^* = S/H^*$. It is known (see [González-Diez et al. 2009]) that an algebraic curve representation of S is given by

$$\widehat{C} : \left\{ \begin{array}{rl} x_0^{k^n} + x_1^{k^n} + x_2^{k^n} &= 0 \\ \lambda x_0^{k^n} + x_1^{k^n} + x_3^{k^n} &= 0 \end{array} \right\} \subset \mathbb{P}^3,$$

that H^* is generated by the projective transformations

$$a([x_0 : x_1 : x_2 : x_3]) = [\rho_n x_0 : x_1 : x_2 : x_3], \quad b([x_0 : x_1 : x_2 : x_3]) = [x_0 : \rho_n x_1 : x_2 : x_3],$$
$$c([x_0 : x_1 : x_2 : x_3]) = [x_0 : x_1 : \rho_n x_2 : x_3]$$

and that the holomorphic map

$$\pi : \widehat{C} \to \widehat{\mathbb{C}} : [x_0 : x_1 : x_2 : x_3] \mapsto -\left(\frac{x_1}{x_0}\right)^{k^n}$$

has degree k^{3n} and is a branched regular cover with H^* as deck group. In this case, $\pi(\mathrm{Fix}(a)) = \infty$, $\pi(\mathrm{Fix}(b)) = 0$, $\pi(\mathrm{Fix}(c)) = 1$ and $\pi(\mathrm{Fix}(abc)) = \lambda$.

Now consider the subgroup of H^* given by $K = \langle a^k, b^{k^{n-1}} \rangle \cong \mathbb{Z}_{k^{n-1}} \times \mathbb{Z}_k$, and set $\mathcal{O}_0 = S/K$. The group $H_0 = H^*/K$ is a group of conformal automorphisms of $\mathcal{O}_0$, $H_0 \cong H$, and $\mathcal{O}_0/H_0 = \mathcal{O}^*$.

Clearly, if R_0 denotes the underlying Riemann surface structure of the orbifold $\mathcal{O}_0$, then R_0/H_0 is the orbifold $\mathcal{O}$. In this way, since any two homology covers of $\mathcal{O}$ are conformally equivalent, we may assume $R = R_0$.

In order to find an algebraic curve representation for R_0 we proceed as follows. First, we consider the affine curve representation of S defined by $x = x_0/x_3$,

$y=x_1/x_3$ and $z=x_2/x_3$; that is,

$$\widehat{C}_0=\left\{\begin{array}{r} x^{k^n}+y^{k^n}+z^{k^n}=0 \\ \lambda x^{k^n}+y^{k^n}+1=0 \end{array}\right\}\subset\mathbb{C}^3$$

and the action of H^* is generated by the linear transformations

$$a(x,y,z)=(\rho_n x,y,z),\quad b(x,y,z)=(x,\rho_n y,z),\quad c(x,y,z)=(x,y,\rho_n z).$$

The subalgebra of $\langle a^k,b^{k^{n-1}}\rangle$ invariant polynomials, $\mathbb{C}[x,y,z]^{\langle a^k,b^{k^{n-1}}\rangle}$, is generated by the monomials $x^{k^{n-1}}$, y^k and z. It follows that the holomorphic map

$$\begin{aligned} &F:\mathbb{C}^3\to\mathbb{C}^3,\\ &(x,y,z)\mapsto(x^{k^{n-1}},y^k,z)=(u,v,w)\end{aligned}$$

is a regular branched covering with $\langle a^k,b^{k^{n-1}}\rangle$ as deck group, and therefore $F(\widehat{C}_0)$ provides an affine algebraic curve representation of R, given by

$$F(\widehat{C}_0)=\left\{\begin{array}{r} u^k+v^{k^{n-1}}+w^{k^n}=0 \\ \lambda u^k+v^{k^{n-1}}+1=0 \end{array}\right\}\subset\mathbb{C}^3,$$

where the action of $H=H^*/K$ is generated by

$$a_0(u,v,w)=(\rho_1 u,v,w),\quad b_0(u,v,w)=(u,\rho_{n-1}v,w),$$
$$c_0(u,v,w)=(u,v,\rho_n w).$$

If we consider the projective space $\mathbb{P}^3$ with coordinates $[z_0:z_1:z_2:z_3]$, and we set

$$u=\frac{z_0}{z_3},\quad v=\frac{z_1}{z_3},\quad w=\frac{z_2}{z_3},$$

then we obtain that R is represented by the projective algebraic curve

$$C=\left\{\begin{array}{r} z_0^k z_3^{k^n-k}+z_1^{k^{n-1}}z_3^{k^n-k^{n-1}}+z_2^{k^n}=0 \\ \lambda z_0^k z_3^{k^{n-1}-k}+z_1^{k^{n-1}}+z_3^{k^{n-1}}=0 \end{array}\right\}\subset\mathbb{P}^3.$$

As the branched covering map $P:R\to R/H$ must satisfy $\pi=P\circ F$ and

$$F([x_0:x_1:x_2:x_3])=\left[x_0^{k^{n-1}}:x_1^k x_3^{k^{n-1}-k}:x_2 x_3^{k^{n-1}-1}:x_3^{k^{n-1}}\right],$$

then

$$P([z_0:z_1:z_2:z_3])=-\left(\frac{z_1^{k^{n-1}}}{z_0^k z_3^{k^{n-1}-k}}\right).$$

6. Proof of Theorem 10

Proof. Consider a closed Riemann surface S admitting a group $G < \mathrm{Aut}(S)$ such that $G \cong \mathbb{Z}_p \times \mathbb{Z}_{p^n}$ and $\mathcal{O} = S/G$ is an orbifold with signature $(0; p, p^{n-1}, p^n, p^n)$, where $n \geq 2$ and p is an odd prime. Denote by $P : S \to \mathcal{O}$ the natural holomorphic branched cover with G as deck group.

In this section we will find algebraic curves representing S and the action of G on them.

Let R be the homology cover of $\mathcal{O}$, and let $Q : R \to \mathcal{O} = R/H$ be the branched regular covering with H as deck group, where $H = \mathbb{Z}_p \times \mathbb{Z}_{p^{n-1}} \times \mathbb{Z}_{p^n}$.

Since G is abelian, there is a subgroup $K < H$ such that $S = R/K$ (and hence K acts freely on R), $G = H/K$, and there is a regular holomorphic covering $T : R \to S$ with K as deck group and $Q = P \circ T$.

Consider the affine algebraic curve C_0 representing R, obtained from Theorem 9 by making $z_3 = 1$:

$$C_0 = \left\{ \begin{array}{c} z_0^p + z_1^{p^{n-1}} + z_2^{p^n} = 0 \\ \lambda z_0^p + z_1^{p^{n-1}} + 1 = 0 \end{array} \right\} \subset \mathbb{C}^3,$$

in which case the group H is generated by

$$a_0(z_0, z_1, z_2) = (\rho_1 z_0, z_1, z_2), \quad b_0(z_0, z_1, z_2) = (z_0, \rho_{n-1} z_1, z_2),$$
$$c_0(z_0, z_1, z_2) = (z_0, z_1, \rho_n z_2).$$

6.1. *Algebraic structure of K.* We next describe the algebraic structure of K. At this point we should note that, using the model of R given in Theorem 9, the transformations in H acting with fixed points on S are exactly the ones that belong to $\langle a_0 \rangle \cup \langle b_0 \rangle \cup \langle c_0 \rangle \cup \langle a_0 b_0 c_0 \rangle$.

Proposition 16. *Consider the algebraic model of (R, H) provided by Theorem 9. Let $K < H$ be such that K acts freely on R and $H/K \cong \mathbb{Z}_p \times \mathbb{Z}_{p^n}$. Then, either*

(1) $\mathbb{Z}_{p^{n-1}} \cong K = \langle a_0^{\alpha} b_0 c_0^{pq} \rangle$, *where* $(p, q) = 1$ *and* $0 \leq \alpha \leq p - 1$; *or*

(2) $\mathbb{Z}_{p^{n-2}} \times \mathbb{Z}_p \cong K = \langle b_0^{-p} c_0^{p^2 v} \rangle \times \langle a_0 c_0^{p^{n-1}\gamma} \rangle$, *where* $(p, v) = 1$ *and* $1 \leq \gamma \leq p - 1$.

Proof. Consider a surjective homomorphism

$$\Phi : H \to J = \mathbb{Z}_p \times \mathbb{Z}_{p^n}$$

with $K = \ker(\Phi)$ acting freely on R. Note that the order of K is p^{n-1}. Then:

a) $K \cap \langle a_0 \rangle = \{I\}$, which implies that $\Phi(a_0)$ has order p.

b) $K \cap \langle b_0 \rangle = \{I\}$, which implies that $\Phi(b_0)$ has order p^{n-1}.

c) $K \cap \langle c_0 \rangle = \{I\}$, which implies that $\Phi(c_0)$ has order p^n.

d) $K \cap \langle a_0 b_0 c_0 \rangle = \{I\}$, which implies that $\Phi(a_0)\Phi(b_0)\Phi(c_0)$ has order p^n.

Hence the subgroups of J given by $\langle \Phi(b_0) \rangle$ and $\langle \Phi(c_0) \rangle$ have respective indices p^2 and p, and there are two cases to be considered, as follows.

Case i). Assume $\langle \Phi(b_0) \rangle \subset \langle \Phi(c_0) \rangle$. Then there exists $1 \le u \le p-1$ such that $\Phi(b_0) = \Phi(c_0^{pu})$, in which case $h = b_0 c_0^{-pu}$ is an element of K of order p^{n-1}, and therefore $K = \langle h \rangle$ is cyclic of the form given in case (1).

Case ii). Assume $\langle \Phi(b_0) \rangle \not\subset \langle \Phi(c_0) \rangle$. Then we have the following commutative diagram of subgroup inclusions and corresponding indices:

$$\begin{tikzcd}
& J & \\
\langle \Phi(c_0) \rangle \arrow[ur, "p"] & & \langle \Phi(b_0) \rangle_{p^2} \arrow[ul] \\
& \langle \Phi(c_0) \rangle \cap \langle \Phi(b_0) \rangle_p \arrow[ul, "p^2"] \arrow[ur] &
\end{tikzcd}$$

and it follows that

$$\langle \Phi(c_0) \rangle \cap \langle \Phi(b_0) \rangle = \left\langle \Phi\left(c_0^{p^2}\right) \right\rangle = \langle \Phi(b_0^p) \rangle.$$

Hence there exists v such that $h_0 = c_0^{p^2 v} b_0^{-p}$ is in K, and h_0 has order p^{n-2}. Also note that $(v, p) = 1$, since otherwise an adequate power of h_0 would be a nontrivial power of b_0 in K. It follows that there are two possibilities for K, either $K \cong \mathbb{Z}_{p^{n-1}}$ or $K = \langle h_0 \rangle \times \langle t \rangle \cong \mathbb{Z}_{p^{n-2}} \times \mathbb{Z}_p$.

Subcase K is not cyclic. As previously noted, in this case $K = \langle h_0 \rangle \times \langle t \rangle \cong \mathbb{Z}_{p^{n-2}} \times \mathbb{Z}_p$, where $h_0 = c_0^{p^2 v} b_0^{-p}$ and $(p, v) = 1$. As $t \in H$ has order p, it has the form

$$t = a_0^{\alpha} b_0^{\beta p^{n-2}} c_0^{\gamma p^{n-1}},$$

where $\alpha, \beta, \gamma \in \{0, 1, \ldots, p-1\}$.

Let us assume $\alpha = 0$. If $\gamma = 0$, then $t \in \langle b_0 \rangle$. As K acts freely on R, necessarily $t = 1$ and we get a contradiction. If $(\gamma, p) = 1$, then we may assume $t = b_0^{\beta p^{n-2}} c_0^{p^{n-1}}$ (by considering an appropriate power of the original t); hence

$$\tilde{h} = t h_0^{-p^{n-3}} = b_0^{(\beta+v)p^{n-2}} \in K \cap \langle b_0 \rangle.$$

Again, as K acts freely, $\tilde{h}$ must be trivial, and t would belong to $\langle h_0 \rangle$, again a contradiction. Then we have proved that $\alpha > 0$.

Since t has order p, we may replace t by a suitable power of it in order to assume that $t = a_0 b_0^{\beta p^{n-2}} c_0^{\gamma p^{n-1}}$.

We now claim that we may assume $\beta = 0$. Indeed, if $\beta > 0$, then $t h_0^{\beta p^{n-3}} = a_0 c_0^{p^{n-1}(\gamma+v)}$ is an element of order p in K that does not belong to $\langle h_0 \rangle$.

Therefore we may write $t = a_0 c^{p^{n-1}\gamma}$, and observe that $1 \le \gamma \le p-1$ because $K \cap \langle a_0 \rangle = \{I\}$. This is case (2).

Subcase K is cyclic. In this case, $K = \langle h \rangle \cong \mathbb{Z}_{p^{n-1}}$. Let us write

$$h = a_0^{\alpha} b_0^{\beta} c_0^{\gamma}$$

where $\alpha \in \{0, 1, \dots, p-1\}$, $\beta \in \{0, 1, \dots, p^{n-1}-1\}$, $\gamma \in \{0, 1, \dots, p^n - 1\}$.

The condition $c_0^{\gamma p^{n-1}} = h^{p^{n-1}} = 1$ ensures that $\gamma \equiv 0 \mod p$. It follows that either $\gamma = 0$ or $\gamma = p^s q$, where $s \in \{1, \dots, n-1\}$ and $(p, q) = 1$.

Next, we need to ensure that, for $\delta \in \{1, 2, \dots, p^{n-1}-1\}$, no power h^{δ} acts with fixed points in C; that is, $h^{\delta} \notin \langle a_0 \rangle \cup \langle b_0 \rangle \cup \langle c_0 \rangle \cup \langle a_0 b_0 c_0 \rangle$.

But if $\gamma = 0$ then $h^p = b_0^{p\beta}$ is a nontrivial element of the group generated by b_0, a contradiction. Similarly, if $s > 1$ then $h^{p^{n-s}} = b_0^{\beta p^{n-s}}$ is a nontrivial element of the group generated by b_0, a contradiction.

Therefore $h = a_0^{\alpha} b_0^{\beta} c_0^{pq}$, with $(p, q) = 1$, and it follows that h^{δ} is not in $\langle b_0 \rangle$.

But if $\beta \equiv 0 \mod p$, then $h^{p^{n-2}} = c^{q p^{n-2}}$ is a nontrivial element of the group generated by c_0, a contradiction. Hence $(p, \beta) = 1$, and h^{δ} is not in $\langle c_0 \rangle$.

We note that $h^{\delta} \in \langle a_0 \rangle$ implies that $\beta\delta \equiv 0 \mod p^{n-1}$, and since $(\beta, p) = 1$, to have $\delta \equiv 0 \mod p^{n-1}$, which is not possible by our choice for δ.

The condition $h^{\delta} \in \langle a_0 b_0 c_0 \rangle$ implies that $\beta\delta \equiv pq\delta \mod p^{n-1}$, from which $(\beta - pq)\delta \equiv 0 \mod p^{n-1}$, and then $\delta \equiv 0 \mod p^{n-1}$, which is not possible by our choice for δ.

By taking an appropriate power of h, we may assume that

$$K = \langle a_0^{\alpha} b_0 c_0^{pq} \rangle,$$

where $(p, q) = 1$.

Now note that in this case $1 \le \alpha \le p-1$, since $\alpha = 0$ implies that $\Phi(b_0) = \Phi(c_0)^{-pq}$ is an element of $\langle \Phi(c_0) \rangle$, which is a contradiction, as we are in case ii). This is case (1). $\square$

6.2. ***The cyclic case.*** As a consequence of Proposition 16, we may assume

$$K = \langle a_0^{\alpha} b_0 c_0^{pq} \rangle,$$

where $(p, q) = 1$ and $\alpha \in \{0, 1, \dots, p-1\}$. Note that

$$a_0^{\alpha} b_0 c_0^{pq}(z_0, z_1, z_2) = (\rho_1^{\alpha} z_0, \rho_{n-1} z_1, \rho_{n-1}^{q} z_2).$$

6.2.1. *The case* $\alpha = 0$. We next search for polynomials in $\mathbb{C}[z_0, z_1, z_2]^K$. We first note that $z_0 \in \mathbb{C}[z_0, z_1, z_2]^K$. Next, we search for polynomials of the form $z_1^u z_2^v \in \mathbb{C}[z_0, z_1, z_2]^K$, where $u,\ v \in \{0, 1, \dots, p^{n-1}\}$. The invariance property requires that the values u and v satisfy the relation

$$u + vq \equiv 0 \mod p^{n-1}.$$

As $(p, q) = 1$, we have that some of those polynomials are given by

$$z_1^{p^{n-1}}, \quad z_2^{p^{n-1}}, \quad z_1^q z_2^{p^{n-1}-1}.$$

Let us consider the holomorphic map

$$F : \mathbb{C}^3 \to \mathbb{C}^4,$$
$$F(z_0, z_1, z_2) = (z_0, z_1^{p^{n-1}}, z_2^{p^{n-1}}, z_1^q z_2^{p^{n-1}-1}) = (x_1, x_2, x_3, x_4).$$

Let us note that $x_4/x_3 = z_1^q/z_2$. As $(p^{n-1}, q) = 1$, it follows that there exist integers a, b so that $aq + bp^{n-1} = 1$; that is, $z_1 = (z_1^q)^a (z_1^{p^{n-1}})b = (x_4/x_3)^a x_2^b$. It follows that z_1 is uniquely determined by the tuple (x_1, x_2, x_3, x_4) and a choice for z_2. In particular, as z_0 is uniquely determined by x_1, one sees that the map F has degree p^{n-1} and it is K-invariant. In this way, an affine algebraic curve defining $F(C_0)$ is given by

$$F(C_0) = \left\{ \begin{array}{r} x_1^p + x_2 + x_3^p = 0 \\ \lambda x_1^p + x_2 + 1 = 0 \\ x_4^{p^{n-1}} - x_2^q x_3^{p^{n-1}-1} = 0 \end{array} \right\} \subset \mathbb{C}^4$$

and a projective one is provided by taking $x_1 = y_0/y_4$, $x_2 = y_1/y_4$, $x_3 = y_2/y_4$, $x_4 = y_3/y_4$, where $[y_0 : y_1 : y_2 : y_3, y_4] \in \mathbb{P}^4$, as follows:

$$\left\{ \begin{array}{r} y_0^p + y_1 y_4^{p-1} + y_2^p = 0 \\ \lambda y_0^p + y_1 y_4^{p-1} + y_4^p = 0 \\ y_3^{p^{n-1}} - y_1^q y_2^{p^{n-1}-1} y_4^{1-q} = 0 \end{array} \right\} \subset \mathbb{P}^4.$$

The map F is, in projective coordinates, given as

$$\begin{aligned} F([z_0 : z_1 : z_2 : z_3]) &= \left[z_0 z_3^{p^{n-1}-1} : z_1^{p^{n-1}} : z_2^{p^{n-1}} : z_1^q z_2^{p^{n-1}-1} z_3^{1-q} : z_3^{p^{n-1}} \right] \\ &= [y_0 : y_1 : y_2 : y_3 : y_4]. \end{aligned}$$

As, by the first equality above,

$$y_1 = -\left(\frac{y_0^p + y_2^p}{y_4^{p-1}} \right),$$

the above also provides the (birational) algebraic curve

$$\left\{\begin{aligned} (\lambda-1)y_0^p - y_2^p + y_4^p &= 0\\ (-1)^{q+1}(y_0^p+y_2^p)^q y_2^{p^{n-1}-1} + y_3^{p^{n-1}} y_4^{qp-1} &= 0 \end{aligned}\right\} \subset \mathbb{P}^3.$$

By making the change of coordinates $w_0 = y_0$, $w_1 = y_2$, $w_2 = y_3$, $w_3 = y_4$, the above is written as

$$\left\{\begin{aligned} (\lambda-1)w_0^p - w_1^p + w_3^p &= 0\\ (-1)^{q+1}(w_0^p+w_1^p)^q w_1^{p^{n-1}-1} + w_2^{p^{n-1}} w_3^{qp-1} &= 0 \end{aligned}\right\} \subset \mathbb{P}^3$$

and the map F is given as

$$F([z_0:z_1:z_2:z_3]) = \left[z_0 z_3^{p^{n-1}-1} : z_2^{p^{n-1}} : z_1^q z_2^{p^{n-1}-1} z_3^{1-q} : z_3^{p^{n-1}}\right] = [w_0:w_1:w_2:w_3].$$

In this case, the group $G = H/K$ is generated by the transformations

$$\begin{aligned} A_1([w_0:w_1:w_2:w_3]) &= [\rho_1 w_0 : w_1 : w_2 : w_3],\\ B_1([w_0:w_1:w_2:w_3]) &= [w_0 : w_1 : \rho_{n-1}^q w_2 : w_3],\\ C_1([w_0:w_1:w_2:w_3]) &= [w_0 : \rho_1 w_1 : \rho_n^{p^{n-1}-1} w_2 : w_3]. \end{aligned}$$

Notice that the elements $A = A_1$ and $B = C_1$ also generate G as desired. As the branched covering map $Q : S \to S/G$ must satisfy $P = Q \circ F$, where $P : R \to R/H$ is (as in Theorem 9) given by

$$P([z_0:z_1:z_2:z_3]) = -\left(\frac{z_1^{p^{n-1}}}{z_0^p z_3^{p^{n-1}-p}}\right),$$

and since

$$-\left(\frac{z_1^{p^{n-1}}}{z_0^p z_3^{p^{n-1}-p}}\right) = -\left(\frac{y_1 y_4^{p-1}}{y_0^p}\right) = \frac{y_0^p + y_2^p}{y_0^p} = \frac{w_0^p + w_1^p}{w_0^p},$$

we obtain

$$Q([w_0:w_1:w_2:w_3]) = \frac{w_0^p + w_1^p}{w_0^p}.$$

6.2.2. *The case* $\alpha \in \{1, 2 \ldots, p-1\}$. Next, we search for polynomials of the form $z_0^t z_1^u z_2^v \in \mathbb{C}[z_0, z_1, z_2]^K$, where $t \in \{0, 1, \ldots, p-1\}$ and $u, v \in \{0, 1, \ldots, p^{n-1}\}$. The invariance property requires that the values u and v satisfy the relation

$$t\alpha p^{n-2} + u + vq \equiv 0 \mod p^{n-1}.$$

As $(p, q) = (\alpha, p) = 1$, we have that some of those polynomials are given by

$$z_0^p, \quad z_1^{p^{n-1}}, \quad z_2^{p^{n-1}}, \quad z_1^q z_2^{p^{n-1}-1}, \quad z_0^{p-1} z_1^{\alpha p^{n-2}}.$$

Let us consider the holomorphic map

$$F:\mathbb{C}^3\to\mathbb{C}^5,$$
$$F(z_0,z_1,z_2)=\left(z_0^p,z_1^{p^{n-1}},z_2^{p^{n-1}},z_1^q z_2^{p^{n-1}-1},z_0^{p-1}z_1^{\alpha p^{n-2}}\right)=(x_1,x_2,x_3,x_4,x_5).$$

Let us note that $x_4/x_3=z_1^q/z_2$. As $(p^{n-1},q)=1$, it follows that there exist integers a, b so that $aq+bp^{n-1}=1$, from where $z_1=(z_1^q)^a\left(z_1^{p^{n-1}}\right)^b=(x_4/x_3)^a x_2^b z_2$. It follows that z_1 is uniquely determined by the tuple (x_1,x_2,x_3,x_4,x_5) and a choice for z_2.

As z_0^p is uniquely determined by x_1, and $z_0^{p-1}z_1^{\alpha p^{n-2}}$ is uniquely determined by x_2, x_3, x_4, x_5 and a choice of z_2, we have that z_0 is also uniquely determined by the previous data.

All the above permits us to see that the map F has degree p^{n-1} and it is K-invariant. In this way, an affine algebraic curve defining $F(C_0)$ is given by

$$F(C_0)=\left\{\begin{array}{r} x_1+x_2+x_3^p=0\\ \lambda x_1+x_2+1=0\\ x_4^{p^{n-1}}-x_2^q x_3^{p^{n-1}-1}=0\\ x_5^p-x_1^{p-1}x_2^{\alpha}=0 \end{array}\right\}\subset\mathbb{C}^5.$$

We may write $x_2=-(x_1+x_3^p)$. In this way, writing $u_1=x_1$, $u_2=x_3$, $u_3=x_4$ and $u_4=x_5$, the above curve is

$$\left\{\begin{array}{r} (\lambda-1)u_1-u_2^p+1=0\\ u_3^{p^{n-1}}+(-1)^{q+1}(u_1+u_2^p)^q u_2^{p^{n-1}-1}=0\\ u_4^p+(-1)^{\alpha+1}u_1^{p-1}(u_1+u_2^p)^{\alpha}=0 \end{array}\right\}\subset\mathbb{C}^4.$$

Now, we may write

$$u_1=\frac{1}{\lambda-1}(u_2^p-1),$$

and setting $y_1=u_2$, $y_2=u_3$ and $y_3=u_4$, the above curve is

$$\left\{\begin{array}{r} y_2^{p^{n-1}}+\dfrac{(-1)^{q+1}}{(\lambda-1)^q}(\lambda y_1^p-1)^q y_1^{p^{n-1}-1}=0\\ y_3^p+\dfrac{(-1)^{\alpha+1}}{(\lambda-1)^{\alpha+p-1}}(y_1^p-1)^{p-1}(\lambda y_1^p-1)^{\alpha}=0 \end{array}\right\}\subset\mathbb{C}^3$$

and F is of the form

$$F(z_0,z_1,z_2)=\left(z_2^{p^{n-1}},z_1^q z_2^{p^{n-1}-1},z_0^{p-1}z_1^{\alpha p^{n-2}}\right)=(y_1,y_2,y_3).$$

Writing $y_1 = v_0/v_3$, $y_2 = v_1/v_3$ and $y_3 = v_2/v_3$, we obtain the projective model

$$\left\{\begin{array}{r} v_1^{p^{n-1}} v_3^{pq-1} + \dfrac{(-1)^{q+1}}{(\lambda-1)^q}(\lambda v_0^p - v_3^p)^q v_0^{p^{n-1}-1} = 0 \\ v_2^p v_3^{p^2+p(\alpha-2)} + \dfrac{(-1)^{\alpha+1}}{(\lambda-1)^{\alpha+p-1}}(v_0^p - v_3^p)^{p-1}(\lambda v_0^p - v_3^p)^\alpha = 0 \end{array}\right\} \subset \mathbb{P}^3$$

and for $n \geq 3$ we have that $\max\{p^{n-1}, p^{n-1}+q-1, \alpha p^{n-2}+p-1\} = p^{n-1}+q-1$ and therefore $F : \mathbb{P}^3 \to \mathbb{P}^3$ is given as

$$\begin{aligned} &F([z_0 : z_1 : z_2 : z_3]) \\ &\quad = \left[z_2^{p^{n-1}} z_3^{q-1} : z_1^q z_2^{p^{n-1}-1} : z_0^{p-1} z_1^{\alpha p^{n-2}} z_3^{p^{n-1}+q-p-\alpha p^{n-2}} : z_3^{p^{n-1}+q-1}\right]. \end{aligned}$$

In the case $n = 2$ a similar formula may be given for F; the maximum value above is $p+q-1$ if $q \geq \alpha$ and $p+\alpha-1$ otherwise.

Continuing with $n \geq 3$, the group $G = H/K$ is generated by the transformations

$$\begin{aligned} A_2([v_0 : v_1 : v_2 : v_3]) &= [v_0 : v_1 : \rho_1^{p-1} v_2 : v_3], \\ B_2([v_0 : v_1 : v_2 : v_3]) &= [v_0 : \rho_{n-1}^q v_1 : \rho_1^\alpha v_2 : v_3], \\ C_2([v_0 : v_1 : v_2 : v_3]) &= [\rho_1 v_0 : \rho_n^{p^{n-1}-1} v_1 : v_2 : v_3]. \end{aligned}$$

Notice that the elements $A = A_2$ and $B = C_2$ also generate G as desired. As the branched covering map $Q : S \to S/G$ must satisfy $P = Q \circ F$, where $P : R \to R/H$ is (as in Theorem 9) given by

$$\begin{aligned} P([z_0 : z_1 : z_2 : z_3]) &= -\left(\frac{z_1^{p^{n-1}}}{z_0^p z_3^{p^{n-1}-p}}\right) = -\left(\frac{x_2}{x_1}\right) = \frac{u_1 + u_2^p}{u_1} \\ &= 1 + \frac{(\lambda-1)u_2^p}{(u_2^p - 1)} = 1 + \frac{(\lambda-1)y_1^p}{(y_1^p - 1)} = 1 + \frac{(\lambda-1)v_0^p}{v_0^p - v_3^p}, \end{aligned}$$

we obtain

$$Q([v_0 : v_1 : v_2 : v_3]) = \frac{\lambda v_0^p - v_3^p}{v_0^p + v_3^p}.$$

6.3. ***The noncyclic case.*** In this case,

$$K = \langle b_0^{-p} c_0^{p^2 v}, a_0 c_0^{\gamma p^{n-1}} \rangle,$$

where $(p, v) = 1$ and $\gamma \in \{1, 2, \ldots, p-1\}$.

We have that

$$b_0^{-p}c_0^{p^2v}(z_0,z_1,z_2)=(z_0,\rho_{n-2}^{-1}z_1,\rho_{n-2}^{v}z_2),$$
$$a_0c_0^{\gamma p^{n-1}}(z_0,z_1,z_2)=(\rho_1 z_0,z_1,\rho_1^{\gamma}z_2).$$

Clearly, $z_0^Az_1^Bz_2^C\in\mathbb{C}[z_0,z_1,z_2]^K$ if and only if

$$\begin{cases} A+C\gamma\equiv 0 \mod p,\\ Cv-B\equiv 0 \mod p^{n-2}.\end{cases}$$

In this way,

$$z_0^p,\ z_1^{p^{n-2}},\ z_0^{p-\gamma}z_1^vz_2\in\mathbb{C}[z_0,z_1,z_2]^K.$$

Let us consider the map

$$F:\mathbb{C}^3\to\mathbb{C}^3,$$
$$F(z_0,z_1,z_2)=(z_0^p,z_1^{p^{n-2}},z_0^{p-\gamma}z_1^vz_2)=(x_1,x_2,x_3).$$

If we fix (x_1,x_2,x_3), then we have p choices for z_0 ($z_0^p=x_1$) and p^{n-2} choices for z_1 ($z_1^{p^{n-2}}=x_2$). Once we have made such choices, the value of z_2 is uniquely determined from $z_0^{p-\gamma}z_1^vz_2=x_3$. It follows that F has degree p^{n-1} and is K-invariant as desired.

The algebraic curve $F(C_0)$ is provided by

$$F(C_0)=\left\{\begin{array}{r} x_1^{p^{n-1}(p-\gamma)}x_2^{p^2v}(x_1+x_2^p)+x_3^{p^n}=0\\ \lambda x_1+x_2^p+1=0\end{array}\right\}\subset\mathbb{C}^3.$$

As

$$x_1=-\frac{(1+x_2^p)}{\lambda},$$

this curve is also represented by, taking $y_1=x_2$ and $y_2=x_3$,

$$\left\{\frac{(-1)^{p^{n-1}(p-\gamma)}}{\lambda^{p^{n-1}(p-\gamma)}}\big(1+y_1^p\big)^{p^{n-1}(p-\gamma)}y_1^{p^2v}\left(y_1^p-\frac{(1+y_1^p)}{\lambda}\right)+y_2^{p^n}=0\right\}\subset\mathbb{C}^2.$$

A projectivization of this plane curve is given by, using the projective coordinates $[u_0:u_1:u_2]\in\mathbb{P}^2$ and taking $y_1=u_0/u_2$ and $y_2=u_1/u_2$, the following one:

$$\left\{\frac{(-1)^{p^{n-1}(p-\gamma)}}{\lambda^{p^{n-1}(p-\gamma)+1}}\big(u_0^p+u_2^p\big)^{p^{n-1}(p-\gamma)}u_1^{p^2v}\big((\lambda-1)u_0^p-u_2^p\big)\right.$$
$$\left.+u_1^{p^n}u_2^{p^n(p-\gamma-1)+p+p^2v}=0\right\}\subset\mathbb{P}^2.$$

In this case, the transformations a_0, b_0 and c_0 define the transformations

$$A_3([u_0:u_1:u_2])=[u_0:\rho_1^{p-\gamma}u_1:u_2],\quad B_3([u_0:u_1:u_2])=[\rho_1 u_0:\rho_{n-1}^{v}u_1:u_2],$$
$$C_3([u_0:u_1:u_2])=[u_0:\rho_n u_1:u_2].$$

The elements $A=C_3^{-vp}B_3$ and $B=C_3$ also generate G as desired. And since

$$P(z_0,z_1,z_2)=-\left(\frac{z_1^{p^{n-1}}}{z_0^p}\right)=\frac{\lambda y_1^p}{1+y_1^p},$$

we obtain

$$Q([u_0:u_1:u_2])=\frac{\lambda u_0^p}{u_0^p+u_1^p}.$$ □

7. Proof of Theorem 11

Proof. Let C be a nonsingular projective algebraic curve admitting a p-group H of conformal automorphisms of C with C/H of signature $(0;p,p^{n-1},p^n,p^n)$ and let $P:C\to C/H=\widehat{\mathbb{C}}$ be a holomorphic branched covering with H as deck group. We may assume the branch values of P are given by ∞ or order p, 0 of order p^{n-1}, and 1 and $\lambda\in\mathbb{C}-\{0,1\}$ are the ones of order p^n. We notice that

$$\mathrm{Aut}_{\mathrm{orb}}(S/H)=\begin{cases}\{I\}, & \lambda\neq-1,\\ \langle J(z)=-z\rangle, & \lambda=-1.\end{cases}$$

Let $K_C=\{\sigma\in\mathrm{Aut}(\mathbb{C}/\mathbb{Q}):C^\sigma\cong C\}$. For each $\sigma\in K_C$ there is a biholomorphism $f_\sigma:C\to C^\sigma$. As H^σ is unique up to conjugation in $\mathrm{Aut}(C^\sigma)$, by Theorem 4, we may assume that $f_\sigma Hf_\sigma^{-1}=H^\sigma$. It follows that there is a Möbius transformation M_σ so that $P^\sigma\circ f_\sigma=M_\sigma\circ P$. The transformation M_σ is uniquely determined by f_σ. As M_σ must preserve the cone points and their orders, it follows that $M_\sigma(\infty)=\infty$, $M_\sigma(0)=0$ and that $\{1,\lambda_\sigma\}=\{M_\sigma(1),M_\sigma(\lambda)\}$, where $\lambda_\sigma\in\mathbb{C}-\{0,1\}$ is branch value of order p^n of $P^\sigma:C^\sigma\to\widehat{\mathbb{C}}$ (in fact, $\lambda_\sigma=\sigma(\lambda)$). It follows that either (i) $M_\sigma=I$, in which case $\lambda_\sigma=\lambda$ or (ii) $M_\sigma(z)=z/\lambda$, in which case $\lambda_\sigma=1/\lambda$.

7.1. Let us assume, from now on, that $\lambda\neq-1$.

Lemma 17. *Let $\lambda\neq-1$ and $\sigma\in K_C$. If there is another biholomorphism $\widehat{f}_\sigma$: $C\to C^\sigma$ such that $\widehat{f}_\sigma H\widehat{f}_\sigma^{-1}=H^\sigma$, then $\widehat{f}_\sigma=h\circ f_\sigma$, for some $h\in H$.*

Proof. If there is another biholomorphism $\widehat{f}_\sigma:C\to C^\sigma$ such that $\widehat{f}_\sigma H\widehat{f}_\sigma^{-1}=H^\sigma$, then $f_\sigma^{-1}\circ\widehat{f}_\sigma\in\mathrm{Aut}(C)$ normalizes H. In this way, $f_\sigma^{-1}\circ\widehat{f}_\sigma$ induces an element of $\mathrm{Aut}_{\mathrm{orb}}(S/H)$. As this last group is trivial, we obtain that $f_\sigma^{-1}\circ\widehat{f}_\sigma\in H$. □

As a consequence of Lemma 17, M_σ is uniquely determined by σ and, in particular, the collection $\{M_\sigma : \sigma \in K_C\}$ satisfies Weil's conditions in Theorem 2. Hence, there is an isomorphism $F : \widehat{\mathbb{C}} \to C_1$, where C_1 is defined over $\mathbb{M}(C)$, with the property that $F = F^\sigma \circ M_\sigma$ for every $\sigma \in K_C$.

Let us consider the Galois cover $Q : C \to B$, where $Q = F \circ P$. We note that, for $\sigma \in K_C$, we have (as $P^\sigma = P$)

$$Q^\sigma \circ f_\sigma = F^\sigma \circ P^\sigma \circ f_\sigma = F \circ M_\sigma^{-1} \circ M_\sigma \circ P \circ f_\sigma^{-1} \circ f_\sigma = R \circ P = Q.$$

Now we follow Dèbes and Emsalem's arguments [1999]. Assume we are able to find a point $c_1 \in C_1$ which is $\mathbb{M}(C)$-rational and so that c_1 is not a branch value of the Galois covering Q. Fix a point $c \in C$ so that $Q(c) = c_1$. It follows that the H-stabilizer of c is trivial. We have the points $\sigma(c)$, $f_\sigma(c) \in C^\sigma$. As

$$Q^\sigma(\sigma(c)) = \sigma(Q(c)) = \sigma(c_1) = c_1 \quad \text{and} \quad Q^\sigma(f_\sigma(c)) = Q(c) = c_1,$$

it follows that there is some $h_\sigma \in H$ so that $h_\sigma(f_\sigma(c)) = \sigma(c)$. Moreover, as a consequence of Lemma 17 and the fact that c has trivial stabilizer in H, such $h_\sigma \in H$ is unique. In this way, we may assume that $f_\sigma(c) = \sigma(c)$ and, by the above, such an isomorphism is uniquely determined by σ. Again, by the uniqueness, this new family $\{f_\sigma : \sigma \in K_\lambda\}$ satisfies Weil's conditions and, by Theorem 2, C is definable over its field of moduli.

In this way, in order to finish our proof, we only need find a $\mathbb{M}(C)$-rational point on C_1 outside the branch set. This is equivalent to finding a point $r \in \widehat{\mathbb{C}} - \{\infty, 0, 1, \lambda\}$ with the property that $F(r) = \sigma(F(r))$, for every $\sigma \in K_C$. As $\sigma(F(r)) = F^\sigma(\sigma(r)) = F(M_\sigma^{-1}(\sigma(r)))$, we need to find a point $r \in \mathbb{C} - \{0, 1, \lambda\}$ such that

$$M_\sigma(r) = \sigma(r).$$

In this way, we need to find a point $r \in \mathbb{C} - \{0, 1, \lambda\}$ so that

(1) if $\sigma(\lambda) = \lambda$, then $\sigma(r) = r$; and

(2) if $\sigma(\lambda) = 1/\lambda$, then $\sigma(r) = r/\lambda$.

Condition (1) asserts that we need to find $r \in \mathbb{Q}(\lambda)$. Clearly, any point of the form $r = \alpha(1 + \lambda)$, where $\alpha \in \mathbb{Q}$ satisfies (1) and (2).

7.2. Let us now consider the case $\lambda = -1$. We have (see Remark 5) that either (i) $\mathrm{Aut}_H(C) = H$ or (ii) $\mathrm{Aut}(C) = \mathrm{Aut}_H(C)$ and $[\mathrm{Aut}(C) : H] = 2$.

In case (i) we may proceed as in the case $\lambda \neq -1$ as Lemma 17 is still valid in this situation (the normalizer of H in $\mathrm{Aut}(C)$ is H).

In case (ii) we have that $C/\mathrm{Aut}(C) = (C/H)/\langle J \rangle$; that is, C is quasiplatonic, so it is defined over its field of moduli. $\square$

8. Proof of Theorem 13

Proof. Since

$$C_\lambda = \left\{ \begin{array}{r} z_0^p z_3^{p^n-p} + z_1^{p^{n-1}} z_3^{p^n-p^{n-1}} + z_2^{p^n} = 0 \\ \lambda z_0^p z_3^{p^{n-1}-p} + z_1^{p^{n-1}} + z_3^{p^{n-1}} = 0 \end{array} \right\} \subset \mathbb{P}^3$$

and

$$P([z_0 : z_1 : z_2 : z_3]) = -\left(\frac{z_1^{k^{n-1}}}{z_0^k z_3^{k^{n-1}-k}} \right),$$

then, for each $\sigma \in \mathrm{Aut}(\mathbb{C}/\mathbb{Q})$, one has that $C_\lambda^\sigma = C_{\sigma(\lambda)}$ and $P^\sigma = P$.

Let $K_\lambda = \{\sigma \in \mathrm{Aut}(\mathbb{C}/\mathbb{Q}) : C_\lambda \cong C_{\sigma(\lambda)}\}$, so $\mathbb{M}(C_\lambda) = \mathrm{Fix}(K_\lambda)$.

If $\sigma \in K_\lambda$, then there is an isomorphism $f_\sigma : C_\lambda \to C_{\sigma(\lambda)}$. As a consequence of Theorem 4, we may assume $f_\sigma H f_\sigma^{-1} = H$. So, there is a Möbius transformation M_σ such that $M_\sigma \circ P = P^\sigma \circ f_\sigma$. As M_σ must preserve the cone points and their orders, one has that

$$M_\sigma(\infty) = \infty, \quad M_\sigma(0) = 0, \quad M_\sigma\{1, \lambda\} = \{1, \sigma(\lambda)\}.$$

It follows from the first two equalities above that $M_\sigma(z) = Lz$, for a suitable $L \in \mathbb{C} - \{0\}$. The equality $M_\sigma\{1, \lambda\} = \{1, \sigma(\lambda)\}$ asserts that either (1) $L = 1$ and $\sigma(\lambda) = \lambda$ or (2) $L = \sigma(\lambda)$ and $\sigma(\lambda) = 1/\lambda$. As a consequence, we have proved (1) and (2).

Part (3) is a consequence of Theorem 11. □

9. Galois groups of order p^{n+1}

In this section, we consider those groups G of order $|G| = p^{n+1}$ acting on compact Riemann surfaces with signature $(0; p, p^{n-1}, p^n, p^n)$, for any odd prime p.

The algebraic structure for these groups is determined by the following result.

Proposition 18. *Let p be an odd prime number and let $G < \mathrm{Aut}(S)$ be a group of order $|G| = p^{n+1}$ acting on a compact Riemann surface S with S/G of signature $(0; p, p^{n-1}, p^n, p^n)$.*

Then G is isomorphic to either

(1) $\mathbb{Z}_{p^n} \times \mathbb{Z}_p$, *or*

(2) $\langle x, y : x^{p^n} = y^p = 1, y^{-1}xy = x^{p^{n-1}+1} \rangle$.

Remark 19. in the first case we have provided, in Theorem 10, algebraic curves for S. In the second case explicit algebraic curves are more complicated, but we will study this problem elsewhere.

Proof. First notice that G has a presentation of the form

$$G = \langle x_1, x_2, x_3, x_4 : x_1^{p^n} = x_2^{p^n} = x_3^{p^{n-1}} = x_4^p = x_1x_2x_3x_4 = 1, \mathcal{R}\rangle$$

where $\mathcal{R}$ denotes other relations.

Therefore G cannot be cyclic, since otherwise it could not be generated by elements of the given orders.

Moreover, G has a cyclic subgroup of order p^n, which is normal because it has index p, and therefore G is isomorphic to

$$G \cong \mathbb{Z}_{p^n} \rtimes_\sigma \mathbb{Z}_p = \langle x\rangle \rtimes_\sigma \langle y\rangle$$

where $\sigma(x) = x^u$ with $u^p = 1 \bmod p^n$. The only solutions for u are $u = 1$ and the powers of $u = p^{n-1} + 1$, and the result follows. □

Remark 20. We will denote the groups appearing in Proposition 18 as follows:

$$G_u = \langle x, y : x^{p^n} = y^p = 1, y^{-1}xy = x^u\rangle \tag{9-1}$$

with $u = 1$ or $u = 1 + p^{n-1}$, and we will study the families of algebraic curves admitting G_u actions with signature $(0; p^n, p^n, p^{n-1}, p)$.

Lemma 21. *Consider the groups G_u given by* (9-1) *and*

$$\Gamma = \langle a_0, b_0, c_0 d_0 : a_0^p = b_0^{p^{n-1}} = c_0^{p^n} = d_0^{p^n} = a_0b_0c_0d_0 = 1\rangle. \tag{9-2}$$

Assume $\Phi : \Gamma \twoheadrightarrow G_u$ is an epimorphism such that $K = \ker\Phi$ is torsion-free. Then either

I) $K = \langle\langle b_0c_0^{-pq}a_0^{-\alpha}, a_0^{-1}c_0a_0c_0^{-u^s}\rangle\rangle$, *with* $0 \le \alpha \le p-1$, $0 < s < p$ *and* $(q, p) = 1$, *or*

II) $K = \langle\langle a_0c_0^{-p^{n-1}v}, b_0^pc_0^{-p^2q}, b_0^{-1}c_0b_0c_0^{-u^s}\rangle\rangle$, *with* $1 \le v \le p-1$, $0 < s < p$ *and* $(q, p) = 1$,

where $\langle\langle\cdot\rangle\rangle$ denotes normal closure in Γ.

Proof. Since K is torsion-free, we obtain that

a) $K \cap \langle a_0\rangle = \{1\}$, and it follows that $y_1 = \Phi(a_0)$ has order p;
b) $K \cap \langle b_0\rangle = \{1\}$, and it follows that $y_2 = \Phi(b_0)$ has order p^{n-1};
c) $K \cap \langle c_0\rangle = \{1\}$, and it follows that $y_3 = \Phi(c_0)$ has order p^n;
d) $K \cap \langle a_0b_0c_0\rangle = \{1\}$, and it follows that $y_4 = \Phi(d_0)$ has order p^n.

Since Φ is an epimorphism, $\{y_1, y_2, y_3, y_4\}$ generate G_u. But clearly $y_4 = (y_1y_2y_3)^{-1}$, and therefore $\{y_1, y_2, y_3\}$ generate G_u.

We now examine the following two cases separately.

Case I) Suppose $\langle y_1, y_3\rangle = G_u$. We have $G_u = \langle y_3\rangle \rtimes_{u^s} \langle y_1\rangle$ for some $0 < s < p$. Also $y_2 = y_1^{\alpha} y_3^{pq}$ with $(q, p) = 1$. Hence

$$y_2 y_3^{-pq} y_1^{-\alpha} = \Phi(b_0 c_0^{-pq} a_0^{-\alpha}) = 1$$

and it follows that $b_0 c_0^{-pq} a_0^{-\alpha} \in K$.

Furthermore $\Phi(a_0^{-1} c_0 a_0 c_0^{-u^s}) = y_1^{-1} y_3 y_1 y_3^{-u^s} = 1$; hence $a_0^{-1} c_0 a_0 c_0^{-u^s} \in K$.

Then, checking the order of $\Gamma/\langle\langle b_0 c_0^{-pq} a_0^{-\alpha}, a_0 c_0^{-1} a_0^{-1} c_0^{-u^s}\rangle\rangle$, we obtain, as required,

$$K = \langle\langle b_0 c_0^{-pq} a_0^{-\alpha}, a_0 c_0^{-1} a_0^{-1} c_0^{-u^s}\rangle\rangle.$$

Case II) Suppose $\langle y_1, y_3\rangle < G_u$. Then

$$y_1 = y_3^{p^{n-1} v}$$

with $(v, p) = 1$, since $\langle y_3\rangle$ is a maximal subgroup of G_u. Hence $a_0 c_0^{-p^{n-1} v} \in K$.

In this case,

$$\langle y_2, y_3\rangle = G_u = \langle y_3\rangle \rtimes_{u^s} \langle y_2\rangle$$

for some $0 < s < p$. Hence $y_2^{-1} y_3 y_2 y_3^{-u^s} = 1$ from where $b_0^{-1} c_0 b_0 c_0^{-u^s} \in K$.

Finally, $y_2^p = y_3^{p^2 q}$ with $(q, p) = 1$, from where $b_0^p c_0^{-p^2 q} \in K$.

Again, by checking the order of $\Gamma/\langle\langle a_0 c_0^{-p^{n-1} v}, b_0^p c_0^{-p^2 q}, b_0^{-1} c_0 b_0 c_0^{-u^s}\rangle\rangle$, we obtain

$$K = \langle\langle a_0 c_0^{-p^{n-1} v}, b_0^p c_0^{-p^2 q}, b_0^{-1} c_0 b_0 c_0^{-u^s}\rangle\rangle. \qquad \square$$

Considering the above notation for the elements $y_1 = \Phi(a_0)$, $y_2 = \Phi(b_0)$, $y_3 = \Phi(c_0)$ and $y_4 = \Phi(d_0)$ in G_u, we have the following result, which states that examples for both cases of Proposition 18 exist, by the Riemann existence theorem.

Corollary 22. *If the group G_u, with $u = 1$ or $u = 1 + p^{n-1}$, acts on a compact Riemann surface with signature $(0; p, p^{n-1}, p^n, p^n)$, then a generating vector for the action may be chosen to be exactly of one of the following forms*:

a) $\left(y_1, y_1^{\alpha} y_3^{pq}, y_3, y_3^{-1-pq} y_1^{-1-\alpha}\right)$, *with* $(q, p) = 1$ *and* $1 \le \alpha \le p - 2$;

b) $\left(y_1, y_3^{pq}, y_3, y_3^{-1-pq} y_1^{-1}\right)$, *with* $(q, p) = 1$;

c) $\left(y_1, y_1^{-1} y_3^{pq}, y_3, y_3^{-1-pq}\right)$, *with* $(q, p) = 1$;

d) $\left(y_3^{p^{n-1} v}, y_2, y_3, y_3^{-1-p^{n-1} v} y_2^{-1}\right)$.

In the first three cases the order of y_1 is p, the order of y_3 is p^n and $y_1^{-1} y_3 y_1 = y_3^{u^s}$ with $0 < s < p$. In the last case y_2 has order p^{n-1}, y_3 has order p^n, $y_2^p = y_3^{qp^2}$ and $y_2^{-1} y_3 y_2 = y_3^{u^s}$ with $0 < s < p$.

The following table gives the genera of some intermediate curves, where g_L *denotes the genus of the quotient of* S *by the subgroup* $L \leq \mathrm{Aut}(S)$ *and where* V *is any cyclic maximal subgroup acting freely*:

generating vector	$u = 1 + p^{n-1}$	$u = 1$
$\left(y_1, y_1^{\alpha} y_3^{pq}, y_3, y_3^{-1-pq} y_1^{-1-\alpha}\right)$	$g_{\langle y_3\rangle} = \frac{p-1}{2}$ $g_{\langle y_3^{-1-pq} y_1^{-1-\alpha}\rangle} = \frac{p-1}{2}$ $g_{\langle y_1\rangle} = \frac{2p^n - p^{n-2}(2p-1) - p}{2}$ $g_{\langle y_3^p\rangle} = p^2 - 2p + 1$ $g_{\langle y_3^p, y_1\rangle} = 0$ $g_V = p - 1$	$g_{\langle y_3\rangle} = \frac{p-1}{2}$ $g_{\langle y_3^{-1-pq} y_1^{-1-\alpha}\rangle} = \frac{p-1}{2}$ $g_{\langle y_1\rangle} = \frac{p^n - p}{2}$ $g_{\langle y_3^p\rangle} = p^2 - 2p + 1$ $g_{\langle y_3^p, y_1\rangle} = 0$ $g_V = p - 1$
$\left(y_1, y_3^{pq}, y_3, y_3^{-1-pq} y_1^{-1}\right)$	$g_{\langle y_3\rangle} = 0$ $g_{\langle y_3^{-1-pq} y_1^{-1}\rangle} = 0$ $g_{\langle y_1\rangle} = \frac{2p^n - p^{n-2}(2p-1) - p}{2}$ $g_{\langle y_3^p\rangle} = \frac{p^2 - 3p}{2} + 1$ $g_{\langle y_3^p, y_1\rangle} = 0$ $g_V = \frac{p-1}{2}$	$g_{\langle y_3\rangle} = 0$ $g_{\langle y_3^{-1-pq} y_1^{-1}\rangle} = 0$ $g_{\langle y_1\rangle} = \frac{p^n - p}{2}$ $g_{\langle y_3^p\rangle} = \frac{p^2 - 3p}{2} + 1$ $g_{\langle y_3^p, y_1\rangle} = 0$ $g_V = \frac{p-1}{2}$
$\left(y_1, y_1^{-1} y_3^{pq}, y_3, y_3^{-1-pq}\right)$	$g_{\langle y_3\rangle} = 0$ $g_{\langle y_3^{-1-pq}\rangle} = 0$ $g_{\langle y_1\rangle} = \frac{2p^n - p^{n-2}(2p-1) - p}{2}$ $g_{\langle y_3^p\rangle} = p^2 - 2p + 1$ $g_{\langle y_3^p, y_1\rangle} = 0$ $g_V = p - 1$	$g_{\langle y_3\rangle} = 0$ $g_{\langle y_3^{-1-pq}\rangle} = 0$ $g_{\langle y_1\rangle} = \frac{p^n - p}{2}$ $g_{\langle y_3^p\rangle} = p^2 - 2p + 1$ $g_{\langle y_3^p, y_1\rangle} = 0$ $g_V = p - 1$
$\left(y_3^{p^{n-1}v}, y_2, y_3, y_3^{-1-p^{n-1}v} y_2^{-1}\right)$	$g_{\langle y_3\rangle} = 0$ $g_{\langle y_3^{-1-p^{n-1}v} y_2^{-1}\rangle} = 0$ $g_{\langle y_1\rangle} = \frac{2p^n - p^{n-1} - p}{2}$ $g_{\langle y_3^p\rangle} = \frac{p^2 - 3p}{2} + 1$ $g_{\langle y_3^p, y_1\rangle} = 0$ $g_V = \frac{p-1}{2}$	$g_{\langle y_3\rangle} = 0$ $g_{\langle y_3^{-1-p^{n-1}v} y_2^{-1}\rangle} = 0$ $g_{\langle y_1\rangle} = \frac{2p^n - p^{n-1} - p}{2}$ $g_{\langle y_3^p\rangle} = \frac{p^2 - 3p}{2} + 1$ $g_{\langle y_3^p, y_1\rangle} = 0$ $g_V = \frac{p-1}{2}$

References

[Beardon 1983] A. F. Beardon, *The geometry of discrete groups*, Graduate Texts in Mathematics **91**, Springer, New York, 1983. MR 85d:22026 Zbl 0528.30001

[Bers 1975] L. Bers, "Automorphic forms for Schottky groups", *Advances in Math.* **16** (1975), 332–361. MR 51 #13218 Zbl 0327.32011

[Burnside 1893] W. Burnside, "Note on the equation $y^2 = x(x^4 - 1)$", *Proc. London Math. Soc.* **24**:1 (1893), 17–20. JFM 25.0795.03

[Carvacho 2010] M. P. Carvacho, *Equivalence of group actions on Riemann surfaces*, Ph.D. thesis, Universidad de Chile, 2010. arXiv 1012.4337

[Coombes and Harbater 1985] K. Coombes and D. Harbater, "Hurwitz families and arithmetic Galois groups", *Duke Math. J.* **52**:4 (1985), 821–839. MR 87g:14012 Zbl 0601.14023

[Dèbes and Emsalem 1999] P. Dèbes and M. Emsalem, "On fields of moduli of curves", *J. Algebra* **211**:1 (1999), 42–56. MR 99k:14044 Zbl 0934.14019

[Earle 1971] C. J. Earle, "On the moduli of closed Riemann surfaces with symmetries", pp. 119–130 in *Advances in the theory of Riemann surfaces* (Stony Brook, NY, 1969), edited by L. V. Ahlfors, Ann. of Math. Studies **66**, Princeton Univ. Press, 1971. MR 45 #5343 Zbl 0218.32010

[Farkas and Kra 1992] H. M. Farkas and I. Kra, *Riemann surfaces*, 2nd ed., Graduate Texts in Mathematics **71**, Springer, New York, 1992. MR 93a:30047 Zbl 0764.30001

[González-Diez and Hidalgo 1997] G. González-Diez and R. A. Hidalgo, "Conformal versus topological conjugacy of automorphisms on compact Riemann surfaces", *Bull. London Math. Soc.* **29**:3 (1997), 280–284. MR 97k:14031 Zbl 0865.14011

[González-Diez et al. 2009] G. González-Diez, R. A. Hidalgo, and M. Leyton, "Generalized Fermat curves", *J. Algebra* **321**:6 (2009), 1643–1660. MR 2010i:14050 Zbl 1175.14023

[Hammer and Herrlich 2003] H. Hammer and F. Herrlich, "A remark on the moduli field of a curve", *Arch. Math.* (*Basel*) **81**:1 (2003), 5–10. MR 2004g:14030 Zbl 1048.14013

[Hidalgo 2009] R. A. Hidalgo, "Non-hyperelliptic Riemann surfaces with real field of moduli but not definable over the reals", *Arch. Math.* (*Basel*) **93**:3 (2009), 219–224. MR 2010j:14063 Zbl 1180.14028

[Hidalgo 2012] R. A. Hidalgo, "Homology closed Riemann surfaces", *Q. J. Math.* **63**:4 (2012), 931–952. Zbl 1256.30025

[Huggins 2007] B. Huggins, "Fields of moduli of hyperelliptic curves", *Math. Res. Lett.* **14**:2 (2007), 249–262. MR 2009a:14040 Zbl 1126.14036

[Karcher and Weber 1999] H. Karcher and M. Weber, "The geometry of Klein's Riemann surface", pp. 9–49 in *The eightfold way*, edited by S. Levy, Math. Sci. Res. Inst. Publ. **35**, Cambridge Univ. Press, 1999. MR 2001a:14027 Zbl 0964.14029

[Klein 1878] F. Klein, "Ueber die Transformation siebenter Ordnung der elliptischen Functionen", *Math. Ann.* **14**:3 (1878), 428–471. MR 1509988 JFM 11.0297.01

[Koebe 1907a] P. Koebe, "Über die Uniformisierung beliebiger analytischer Kurven, I", *Nachr. Acad. Wiss. Göttingen* **1907** (1907), 177–190. JFM 38.0453.01

[Koebe 1907b] P. Koebe, "Über die Uniformisierung beliebiger analytischer Kurven, II", *Nachr. Acad. Wiss. Göttingen* **1907** (1907), 633–669. JFM 38.0455.02

[Koizumi 1972] S. Koizumi, "The fields of moduli for polarized abelian varieties and for curves", *Nagoya Math. J.* **48** (1972), 37–55. MR 50 #4582 Zbl 0246.14006

[Leyton and Hidalgo 2007] M. Leyton A. and R. A. Hidalgo, "On uniqueness of automorphisms groups of Riemann surfaces", *Rev. Mat. Iberoam.* **23**:3 (2007), 793–810. MR 2009c:30107 MR 1144: 30017

[Maclachlan 1965] C. Maclachlan, "Abelian groups of automorphisms of compact Riemann surfaces", *Proc. London Math. Soc.* (3) **15** (1965), 699–712. MR 31 #3596 Zbl 0156.08902

[Poincaré 1907] H. Poincaré, "Sur l'uniformisation des fonctions analytiques", *Acta Math.* **31**:1 (1907), 1–63. MR 1555036 JFM 38.0452.02

[Shimura 1972] G. Shimura, "On the field of rationality for an abelian variety", *Nagoya Math. J.* **45** (1972), 167–178. MR 46 #5342 Zbl 0243.14012

[Singerman 1972] D. Singerman, "Finitely maximal Fuchsian groups", *J. London Math. Soc.* (2) **6** (1972), 29–38. MR 48 #529 Zbl 0251.20052

[Torelli 1913] R. Torelli, "Sulle varieta di Jacobi", *Rend. R. Accad. Lincei* **22**:2 (1913), 98–103. JFM 44.0655.03

[Weil 1956] A. Weil, "The field of definition of a variety", *Amer. J. Math.* **78** (1956), 509–524. MR 18,601a Zbl 0072.16001

[Wolfart 2006] J. Wolfart, "ABC for polynomials, dessins d'enfants and uniformization: A survey", pp. 313–345 in *Elementare und analytische Zahlentheorie* (Mainz, Germany, 2004), edited by S. Baier et al., Schr. Wiss. Ges. Johann Wolfgang Goethe Univ. Frankfurt am Main **20**, Steiner, Stuttgart, 2006. MR 2009d:14031 Zbl 1214.11079

Received April 3, 2012. Revised February 3, 2013.

ANGEL CAROCCA
DEPARTAMENTO DE MATEMÁTICA Y ESTADÍSTICA
UNIVERSIDAD DE LA FRONTERA
CASILLA 54-D
4780000 TEMUCO
CHILE
angelcarocca@ufro.cl

RUBÉN A. HIDALGO
DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
CASILLA 110-V
2340000 VALPARAÍSO
CHILE
ruben.hidalgo@usm.cl

RUBÍ E. RODRÍGUEZ
FACULTAD DE MATEMÁTICAS
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
CASILLA 306–22
8320000 SANTIAGO
CHILE
rubi@mat.puc.cl

PACIFIC JOURNAL OF MATHEMATICS
Vol. 263, No. 1, 2013
dx.doi.org/10.2140/pjm.2013.263.87

EXPLICIT ISOGENY THEOREMS FOR DRINFELD MODULES

IMIN CHEN AND YOONJIN LEE

Let $F = \mathbb{F}_q(T)$ and $A = \mathbb{F}_q[T]$. Given two nonisogenous rank-r Drinfeld A-modules ϕ and ϕ' over K, where K is a finite extension of F, we obtain a partially explicit upper bound (dependent only on ϕ and ϕ') on the degree of primes $\wp$ of K such that $P_\wp(\phi) \neq P_\wp(\phi')$, where $P_\wp(*)$ denotes the characteristic polynomial of Frobenius at $\wp$ on a Tate module of $*$. The bounds are completely explicit in terms of the defining coefficients of ϕ and ϕ', except for one term, which can be made explicit in the case of $r = 2$. An ingredient in the proof of the partially explicit isogeny theorem for general rank is an explicit bound for the different divisor of torsion fields of Drinfeld modules, which detects primes of potentially good reduction.

Our results are a Drinfeld module analogue of Serre's work (1981), but the results we obtain are unconditional because the generalized Riemann hypothesis holds for function fields.

1. Introduction

Let $A = \mathbb{F}_q[T]$, $F = \mathbb{F}_q(T)$, and let $\bar{F}$ be a fixed algebraic closure of F, K a finite extension of F in $\bar{F}$, $\bar{K}$ the algebraic closure of K in $\bar{F}$, $\mathbb{O}$ the ring of integers of K, and $\mathbb{F}_q$ a finite field of order q.

By a prime $\wp$ (or place) of K, we mean a discrete valuation ring R with field of fractions K and maximal ideal $\wp$, and v denotes the discrete valuation associated to a prime $\wp$ of K. For each place v of K, we fix a choice of $\bar{K}_v$ and extend v to $\bar{K}_v$, which by abuse of notation we also call v. Also, when we speak of finite extensions of K_v, we assume they are initially given as subfields of $\bar{K}_v$.

Let ∞ be the infinite prime of F with corresponding discrete valuation

$$v_\infty(f/g) = \deg g - \deg f,$$

where $f, g \in A$. Let S_∞^K be the set of the infinite primes of K lying over ∞, and

Chen was supported by the NSERC. Lee was supported by Priority Research Centers Program through the NRF funded by the Ministry of Education, Science and Technology (2009-0093827) and by the NRF grant funded by the Korea government (MEST) (2012-0005432). Lee is the corresponding author.
MSC2010: primary 11G09; secondary 11R58.
Keywords: Drinfeld modules, Galois representations, Chebotarev density theorem, isogeny theorem, different, ramification.

let $\bar{\infty} \in S_\infty^K$ have corresponding discrete valuation $v_{\bar{\infty}}$.

Let τ be the map that raises an element to its q-th power. A *Drinfeld A-module ϕ over K* is given by an $\mathbb{F}_q$-algebra homomorphism $i : A \to K$ and an $\mathbb{F}_q$-algebra homomorphism

$$\phi : A \to K\{\tau\}$$

such that ϕ_a has constant term $i(a)$ for any $a \in A$, and the image of ϕ is not contained in K.

A rank-r Drinfeld A-module ϕ over K is completely determined by

$$\phi_T = i(T) + a_1(\phi)\tau + \cdots + a_{r-1}(\phi)\tau^{r-1} + \Delta(\phi)\tau^r,$$

where $a_i(\phi)$, $a_r = \Delta(\phi) \in K$ for $1 \le i \le r-1$. We call $\Delta(\phi)$ the *discriminant* of ϕ.

For any *monic* $a \in \mathbb{F}_q[T]$, we have

$$\phi_a = i(a) + \sum_{i=1}^{M-1} a_i(\phi, a)\tau^i + \Delta(\phi)^{(q^M-1)/(q^r-1)}\tau^M \tag{1}$$

for some $a_i(\phi, a) \in K$, where $M = r \deg_K a$.

For any $a \in A$, $a \neq 0$, we define the A-module of *a-torsion points* as

$$\phi[a] = \{\lambda \in \overline{K} \mid \phi_a(\lambda) = 0\}.$$

If I is a nonzero ideal of A, we similarly define the A-module of *I-torsion points*:

$$\phi[I] = \{\lambda \in \overline{K} \mid \phi_a(\lambda) = 0 \text{ for every } a \in I\}.$$

We have $\phi[a] \simeq (A/aA)^r$ if ϕ is of rank r [Rosen 2002, Proposition 12.4]. Let $K_{\phi,a} := K(\phi[a])$ be the field obtained by adjoining a-torsion points of ϕ to K, and let $K_{\phi,I} := K(\phi[I])$.

In the following, we briefly explain the definition of good reduction of a Drinfeld module. For more details, refer to [Goss 1996; Thakur 2004]. Let ϕ be a rank-r Drinfeld A-module over K and let $\wp$ be a prime of K. Let $\mathbb{O}_\wp$ be the valuation ring of $\wp$ with the maximal ideal $\wp$ and residue field $\mathbb{F}_\wp := \mathbb{O}_\wp/\wp$. We say that ϕ has *integral coefficients* at $\wp$ if ϕ_a has coefficients in $\mathbb{O}_\wp$ for all $a \in A$ and the reduction modulo $\wp$ of these coefficients defines a Drinfeld module over $\wp$. The reduced Drinfeld module is denoted by $\phi^\wp$.

Let ϕ and ϕ' be Drinfeld A-modules over K. Then a *morphism* f from ϕ to ϕ' over K is a polynomial f in $K\{\tau\}$ with the property that $f\phi_a = \phi'_a f$ for all $a \in A$. A nonzero morphism from ϕ to ϕ' over K is called an *isogeny* from ϕ to ϕ' (over K). If there exists an isogeny from ϕ to ϕ' over K, then we say that ϕ and ϕ' are *isogenous* (over K). An isogeny f from ϕ to ϕ' over K is called an *isomorphism* (over K) if there is an isogeny g from ϕ' to ϕ over K such that $fg = I = gf$, where I denotes the identity morphism. We note that ϕ and ϕ' are *isomorphic* (over K) if and only if there is a $c \in K^*$ such that $c\phi_a = \phi'_a c$ for all $a \in A$ (for more details, refer to [Rosen 2002]).

We say that ϕ has *good reduction* at $\wp$ if there exists a Drinfeld module ψ over K that is isomorphic to ϕ over K, ψ has integral coefficients at $\wp$, and $\psi^{\wp}$ is a Drinfeld module of rank r.

By [Takahashi 1982] (see [Goss 1996, Theorem 4.10.5]; also [Goss 1992, Theorem 3.2.3] for one direction), we have that ϕ has good reduction at $\wp$ if and only if the G_K-module $\phi[\mathfrak{L}^\infty] := \bigcup_{m\geq 1} \phi[\mathfrak{L}^m]$ is unramified at $\wp$, where G_K is the absolute Galois group of K and $\mathfrak{L}$ is a prime ideal of A different from $\wp$. This is the analogue for Drinfeld modules of the classical result of Ogg, Néron, and Shafarevich in the theory of abelian varieties.

If ϕ is a Drinfeld A-module defined over K and all its defining coefficients $a_i(\phi)$ lie in $\mathbb{O}$, then we say that ϕ is *integral over* $\mathbb{O}$. If ϕ is integral over $\mathbb{O}$, then it has good reduction outside any set of primes S of K that includes the primes lying over ∞ and the primes dividing the discriminant $\Delta(\phi)$ of ϕ. In particular, the G_K-modules $\phi[\mathfrak{L}^\infty]$ and $\phi[\mathfrak{L}]$ are unramified outside $S \cup \{\text{primes of } K \text{ lying over } \mathfrak{L}\}$.

Let $\mathfrak{L}$ be a finite prime of A. The $\mathfrak{L}$-torsion points of ϕ in $\overline{K}$ give rise to a representation

$$\rho_{\phi,\mathfrak{L}} : G_K \to \mathrm{Aut}_{A/\mathfrak{L}}(\phi[\mathfrak{L}]) \cong \mathrm{GL}_r(A/\mathfrak{L}A),$$

where G_K is the absolute Galois group of K. For a prime $\wp$ of K, if ϕ has good reduction at $\wp$, then $\rho_{\phi,\mathfrak{L}}$ is unramified at $\wp$ if $\wp$ does not lie over $\mathfrak{L}$.

For an unramified prime $\wp$ of K, let $\mathrm{Frob}_\wp \in G_K$ denote a Frobenius conjugacy class at $\wp$. Let $a_\wp(\phi)$ denote the trace of $\mathrm{Frob}_\wp$ on the $T_{\mathfrak{L}}(\phi)$, and $P_\wp(\phi)(X)$ the characteristic polynomial of $\mathrm{Frob}_\wp$ on the $T_{\mathfrak{L}}(\phi)$. It is known that $a_\wp(\phi)$ and $P_\wp(\phi)(X)$ are independent of $\mathfrak{L}$ [Goss 1996, Theorem 4.12.12].

Serre [1972] proved that if E is an elliptic curve over a number field L without complex multiplication, then there are only finitely many primes p such that the Galois representation $\rho_{E,p}$ on the p-torsion points of E is not surjective. The analogue of Serre's result [1972] for rank-2 Drinfeld A-modules was proved by Gardeyn [2001], that is, if ϕ is a rank-2 Drinfeld module over K without complex multiplication, then there are only finitely many primes $\mathscr{L}$ such that $\rho_{\phi,\mathscr{L}}$ is not surjective. The case of general rank is proven in [Pink and Rütsche 2009a; 2009b].

The following theorem is the Tate conjecture for rank-r Drinfeld A-modules over K, and its generalization to t-motives can be found in [Tamagawa 1994].

Theorem 1.1 [Taguchi 1996]. *Let ϕ, ϕ' be rank-r Drinfeld A-modules over K, and $A_{\mathfrak{L}}$ the $\mathfrak{L}$-adic completion of A. Then the natural homomorphism*

$$\mathrm{Hom}_K(\phi, \phi') \otimes_A A_{\mathfrak{L}} \to \mathrm{Hom}_{A_{\mathfrak{L}}[G_K]}(T_{\mathfrak{L}}(\phi), T_{\mathfrak{L}}(\phi'))$$

is an isomorphism, *where $T_{\mathfrak{L}}(*)$ is the $\mathfrak{L}$-adic Tate module of $*$.*

A consequence of the Tate conjecture is the isogeny theorem [Taguchi 1992, Proposition 3.1] that states that two Drinfeld A-modules ϕ, ϕ' over K are K-isogenous if and only if $P_\wp(\phi)(X) = P_\wp(\phi')(X)$ for all but finitely many primes $\wp$.

We prove the following partially explicit and effective version of the isogeny theorem for rank-r Drinfeld A-modules over K. For a Drinfeld A-module ϕ and a place $\wp$ of K, define

$$\tau_{K,\wp}(\phi) = \inf\left\{ \frac{v_\wp(a_i(\phi))}{q^i - 1} : i = 1, \dots, r \right\}.$$

For any extension L/F, let $\gamma_L = [\mathbb{F}_L : \mathbb{F}_q]$. It is known that the constant field of

$$K_{\phi,\mathrm{tor}} := K(\phi[a] : a \in A \text{ nonzero})$$

is finite over $\mathbb{F}_q$ (see [David 2001, Lemma 3.2]), so we may define $\gamma_\phi = \gamma_{K_{\phi,\mathrm{tor}}}$. More precisely, let $g_{\phi,\bar\infty} = [K_{\bar\infty}(\Lambda_{\phi,\bar\infty}) : K_{\bar\infty}]$, where $\Lambda_{\phi,\bar\infty}$ is the lattice associated to the uniformization of ϕ over $C_{\bar\infty}$. Then we have

$$\gamma_\phi \le g_\phi = \min\{g_{\phi,\bar\infty} : \bar\infty \mid \infty\}.$$

One can bound $g_{\phi,\bar\infty}$ using knowledge of the successive minima of the lattices $\Lambda_{\phi,\bar\infty}$ associated to ϕ [Gardeyn 2002, Proposition 4(i)]. Unfortunately, an explicit bound for these successive minima is not currently known except in the case of rank ≤ 2 [Chen and Lee 2013], so this term is currently inexplicit in general.

Throughout, $\ln x$ denotes the natural logarithm of x, $\log_q x$ denotes the logarithm of x to base q, and $\log_q^* x = \log_q \max\{x, 1\}$.

Theorem 1.2. *Let ϕ_1, ϕ_2 be rank-r Drinfeld A-modules that are integral over $\mathbb{O}$ and not K-isogenous. Let S be the set consisting of the primes of K lying over the prime ∞ and the primes dividing $\Delta(\phi_1)\Delta(\phi_2)$. Suppose $\wp \notin S$ is a prime of K of least degree such that $P_\wp(\phi_1) \ne P_\wp(\phi_2)$. Then*

$$\deg_K \wp \le \max\left\{ \frac{4}{m_0}\big(C_{q,r} + W + c_r s_{q,r} \log_q W\big),\ s \max\big\{1 + 2\log_q s, 7\big\} \right\}, \tag{2}$$

where

$s =$ the geometric extension degree of K/F,

$m_0 = \gamma_K$,

$c_r = 2r^2 + r + 1$,

$d_r = c_r + \log_q 86rs^2(g+1)$,

$$s_{q,r} = \frac{\ln(qd_r)}{\ln(qd_r) - 1},$$

$$C_{q,r} = \log_q 86rs^2(g+1) + c_r\Big(1 + s_{q,r}\log_q \frac{4}{m_0} + \log_q d_r\Big) + c_r s_{q,r} \log_q \log_q d_r,$$

$$a_r(\phi_i) = \Delta(\phi_i), \quad i = 1, 2,$$
$$W = \log_q^*\big(\Lambda_K(\phi_1, \phi_2) + 2D(\phi_1, \phi_2)\big) + g_{\phi_1} g_{\phi_2} m_0,$$
$$\textit{where } D(\phi_1, \phi_2) = \deg_K \operatorname{rad}_K \Delta(\phi_1) + \deg_K \operatorname{rad}_K \Delta(\phi_2),$$
$$\Lambda_K(\phi_1, \phi_2) = -\sum_v \tau_{K,v}(\phi_1) \deg_K v - \sum_v \tau_{K,v}(\phi_2) \deg_K v,$$
$$\deg_K \operatorname{rad}_K x = \sum_{v(x)\neq 0} \deg_K v.$$

(*The sums are over every place* v *of* K.)

Note that any Drinfeld A-module defined over K is isomorphic over K to a Drinfeld A-module that is integral over $\mathbb{O}$. In order to reduce the bounds given by the above theorem, in particular the quantity $\deg_K \operatorname{rad}_K \Delta(\phi_1)\Delta(\phi_2)$, one should use minimal models of ϕ_1 and ϕ_2 (see [Taguchi 1993, Section 2]).

The proof follows the strategy in [Serre 1981] adapted to the Drinfeld module situation, with the notable difference that the effective Chebotarev density theorem we use [Kumar Murty and Scherk 1994] is stronger and unconditional because the general Riemann hypothesis holds for function fields. Also, unlike in the number field case, it is necessary to deal with wild ramification when bounding the different divisor. The bound we obtain on the different divisor is completely explicit in terms of the defining coefficients of the Drinfeld modules involved, unlike the results in [Gardeyn 2002], which are effective but not explicit. Also, the bounds are sensitive to primes of potentially good reduction, unlike the bounds in [Taguchi 1992].

We discuss some of the differences between our method and that of [Gardeyn 2002] in more detail in Section 7. In the rank-2 case, it is possible to make explicit the quantities involved in Gardeyn's bounds for the different divisor of torsion fields by determining the Newton polygons of exponential functions attached to Drinfeld modules [Chen and Lee 2013]. However, the computation of Newton polygons grows in complexity for higher rank, so new techniques using weaker information will likely be required to obtain explicit bounds for successive minima so we can apply the bounds of [Gardeyn 2002] for the different divisor and g_ϕ. Further remarks about this will be made in Section 7.

Cojocaru and David [2008] find upper bounds for the number of primes $\wp$ of degree d such that the field extension of F obtained by adjoining a root of the characteristic polynomial of the Frobenius endomorphism of ϕ over the finite field $A/\wp$ is the fixed field K, where ϕ is a Drinfeld module over K of rank 2 and K is an imaginary quadratic field over F. An ingredient in their proof requires the surjectivity results of Pink [1997] and Gardeyn [2001]. However, they do not require explicit versions of these in order to achieve their results; that is, they use the fact that the Galois representation $\rho_{\phi,\mathscr{L}} : \mathrm{Gal}(F^{\mathrm{sep}}/F) \to \mathrm{GL}_2(A/\mathscr{L}A)$ and its projection in $\mathrm{PGL}_2(A/\mathscr{L}A)$ are surjective for all but finitely many primes $\mathscr{L}$ in A, assuming

$\mathrm{End}_{\bar{F}}(\phi) = A$. As a method, they also use the effective version of the Chebotarev density theorem in [Kumar Murty and Scherk 1994], but for the different divisor bounds they only require the bounds in [Gardeyn 2002, Proposition 6].

2. Preliminaries

Let L be a finite extension of K and let $\mathbb{O}_L$ be the maximal order of L, that is, the integral closure of $\mathbb{O}$ in L. The *constant field* $\mathbb{F}_L$ of L is the algebraic closure of $\mathbb{F}_q$ in L. The *geometric extension degree* of L/K is the degree of L/K', where K' is the maximal constant field extension of K in L (that is, $[L:K]/[\mathbb{F}_L:\mathbb{F}_K]$). We say L/K is a *geometric extension* if $K = K'$.

For a prime ideal $\mathfrak{B}$ of $\mathbb{O}_L$, we let $\deg_L \mathfrak{B}$ be the $\mathbb{F}_L$-dimension of the residue class field $\mathbb{F}_{L,\mathfrak{B}} := \mathbb{O}_L/\mathfrak{B}$ of $\mathfrak{B}$, extending this to a general ideal I of $\mathbb{O}_L$ by additivity on products. For a in $\mathbb{O}_L$, we define the degree of a by $\deg_L a := \deg_L(a)$, where (a) is the principal ideal of $\mathbb{O}_L$ generated by a.

More generally, let $\mathfrak{B}$ be a prime of L, $\mathbb{O}_{L,\mathfrak{B}}$ the valuation ring of $\mathfrak{B}$, and $\mathbb{F}_{L,\mathfrak{B}} := \mathbb{O}_{L,\mathfrak{B}}/\mathfrak{B}$ the residue class field of $\mathfrak{B}$. Then the *degree* of $\mathfrak{B}$ is defined to be $\deg_L \mathfrak{B} := [\mathbb{F}_{L,\mathfrak{B}} : \mathbb{F}_L]$, the $\mathbb{F}_L$-dimension of $\mathbb{F}_{L,\mathfrak{B}}$. We extend the definition by linearity to a divisor $\mathfrak{D} = \sum_{\mathfrak{B}} n_{\mathfrak{B}}\mathfrak{B}$ of L by $\deg_L \mathfrak{D} = \sum_{\mathfrak{B}} n_{\mathfrak{B}} \deg_L \mathfrak{B}$. The *finite part* $\mathfrak{D}_0$ of a divisor $\mathfrak{D} = \sum_{\mathfrak{B}} n_{\mathfrak{B}}\mathfrak{B}$ is the divisor $\sum_{\mathfrak{B}\nmid\infty} n_{\mathfrak{B}}\mathfrak{B}$.

Let $i_{L/K} : \mathrm{Div}(K) \to \mathrm{Div}(L)$ be the *conorm map* from divisors on K to divisors on L, defined by

$$i_{L/K}(\wp) = \sum_{\mathfrak{B}|\wp} e(\mathfrak{B}/\wp)\mathfrak{B}$$

for every prime $\wp$ of K and then extended by linearity, where $e(\mathfrak{B}/\wp)$ denotes the *ramification index* of $\mathfrak{B}$ over $\mathfrak{B}$.

For $\mathfrak{B}$ a prime of L lying over the prime $\wp$ of K, denote by $f(\mathfrak{B}/\wp)$ the inertia degree of $\mathfrak{B}$ over $\wp$.

Lemma 2.1 [Rosen 2002, Proposition 7.7]. *Let L/K be a finite extension, $\mathfrak{D}$ a divisor of K, and $\mathfrak{B}$ a prime of L lying over the prime $\wp$ of K. Then*

$$\deg_L i_{L/K}\mathfrak{D} = n' \deg_K \mathfrak{D}, \quad \deg_L \mathfrak{B} = \frac{f(\mathfrak{B}/\wp)}{[\mathbb{F}_L : \mathbb{F}_K]} \deg_K \wp,$$

where n' is the geometric extension degree of L/K.

Let L/K be a finite extension. Writing divisors in terms of places instead of primes, the *different divisor* $\mathfrak{D}(L/K)$ of L/K is defined as

$$\mathfrak{D}(L/K) = \sum_w w(D(L_w/K_v))w,$$

and its degree is given by

$$\deg_L \mathfrak{D}(L/K) = \sum_w w(D(L_w/K_v)) \deg_L w,$$

where w ranges through all normalized places of L, and $D(L_w/K_v)$ is the *different ideal* of L_w/K_v.

For convenience, we also define *the degree with respect to* K of $\mathfrak{D}(L/K)$ as

$$\deg_K \mathfrak{D}(L/K) = \sum_v \max\bigl\{v(D(L_w/K_v)) : w|v\bigr\} \deg_K v,$$

where v ranges through all normalized places of K. Similarly, we define the degree with respect to K of $\mathfrak{D}_0(L/K)$ as

$$\deg_K \mathfrak{D}_0(L/K) = \sum_{v\nmid\infty} \max\bigl\{v(D(L_w/K_v)) : w|v\bigr\} \deg_K v.$$

Lemma 2.2. *Let L/K be a finite extension. Then*

$$\deg_L \mathfrak{D}(L/K) \le n' \deg_K \mathfrak{D}(L/K),$$

where n' is the geometric extension degree of L/K.

Proof. By the definition, we have

$$\begin{aligned}
\deg_L \mathfrak{D}(L/K) &= \sum_w w(D(L_w/K_v)) \deg_L w \\
&= \sum_v \sum_{w|v} w(D(L_w/K_v)) \deg_L w \\
&= \sum_v \sum_{w|v} v(D(L_w/K_v)) e(w/v) f(w/v) \frac{1}{[\mathbb{F}_L : \mathbb{F}_K]} \deg_K v \\
&\le \frac{1}{[\mathbb{F}_L : \mathbb{F}_K]} \sum_v \max\bigl\{v(D(L_w/K_v)) : w|v\bigr\} \sum_{w|v} e(w/v) f(w/v) \deg_K v \\
&= n' \sum_v \max\bigl\{v(D(L_w/K_v)) : w|v\bigr\} \deg_K v = n' \deg_K \mathfrak{D}(L/K),
\end{aligned}$$

where $\mathbb{F}_L$ and $\mathbb{F}_K$ are the constant fields of L and K respectively, $f(w/v)$ denotes the *relative degree* of w over v, and we use the identity

$$[L : K] = \sum_{w|v} e(w/v) f(w/v),$$

which is valid because our constant fields are finite and hence perfect [Rosen 2002, Proposition 7.4]. □

Lemma 2.3 [Serre 1979, Proposition 8, Chapter III.4]. *Let $M/L/K$ be a tower of finite separable extensions. The different divisor satisfies the transitivity relation*

$$\mathfrak{D}(M/K) = \mathfrak{D}(M/L) + i_{M/L}\mathfrak{D}(L/K).$$

Lemma 2.4. *Let K be a local field with ring of integers $\mathbb{O}$, and let L/K be a finite extension of K with ring of integers $\mathbb{O}_L$. Let $\alpha \in \mathbb{O}_L$ be such that $L = K(\alpha)$, and suppose $f(X) \in \mathbb{O}[X]$ is the minimal polynomial of α over K. Then the different ideal $D(\mathbb{O}_L/\mathbb{O})$ divides the ideal $(f'(\alpha))$, with equality holding if and only if $\mathbb{O}_L = \mathbb{O}[\alpha]$. Furthermore, we may replace $f(X)$ by any monic polynomial $g(X)$ in $\mathbb{O}[X]$ that α satisfies.*

Proof. See [Serre 1979, Corollary 2, III.6]. For the final remark, note that $g(X) = f(X)h(X)$ for some $g(X) \in \mathbb{O}[X]$, so $(g'(\alpha)) = (f'(\alpha)h(\alpha)) \subseteq (f'(\alpha))$. □

Lemma 2.5. *Let E/K and L/K be finite extensions of local fields, with $\mathbb{O}$ the ring of integers of K, $\mathbb{O}_E$ the ring of integers of E, $\mathbb{O}_{EL}$ the ring of integers of EL, and $\mathbb{O}_L$ the ring of integers of L.*

Then the different ideals satisfy $D(EL/L) \mid \mathbb{O}_{EL} \cdot D(E/K)$.

Proof. Suppose that $\mathbb{O}_E = \mathbb{O}_K[x]$ for some $x \in B$, so that $E = K(x)$ (see [Serre 1979, Proposition 12, III.6]). Let $f \in \mathbb{O}_K[X]$ be the minimal polynomial of x over K.

Now $EL = K(x)L = K(x)$ and $x \in \mathbb{O}_{EL}$.

As $f \in \mathbb{O}[X]$ is monic and $x \in \mathbb{O}_{EL}$ is a root of f, we may apply Lemma 2.4 to get that $D(EL/L) \mid \mathbb{O}_{EL} \cdot f'(x)$. But as $\mathbb{O}_E = \mathbb{O}[x]$, we have $D(E/K) = \mathbb{O}_E \cdot f'(x)$. Hence, $\mathbb{O}_{EL} \cdot f'(x) = \mathbb{O}EL \cdot \mathbb{O}_E \cdot f'(x) = \mathbb{O}_{EL} \cdot D(E/K)$. The result thus follows. □

Lemma 2.6. *Let E/K and L/K be finite extensions of global fields. Then*

$$\mathfrak{D}(EL/K) \le i_{EL/E}\mathfrak{D}(E/K) + i_{EL/L}\mathfrak{D}(L/K).$$

Proof. This follows by localization and applying Lemma 2.3 and Lemma 2.5. □

3. Effective Chebotarev density theorem

Lemma 3.1. *Let K be a finite extension of $F = \mathbb{F}_q(T)$ with constant field $\mathbb{F}_q$, where $\mathbb{F}_q$ is a finite field of order q, and let g be the genus of K. Let $S(N)$ be the number of primes $\wp$ of K with $\deg_K \wp = N$. Then*

$$\left| S(N) - \frac{q^N}{N} \right| \le \left(2g + 1 + \left(2g + \tfrac{3}{2}\right)\frac{4}{q} \right) \frac{q^{N/2}}{N}.$$

Proof. From the prime number theorem for L [Rosen 2002, Theorem 5.12], we have that

$$S(N) = \frac{q^N}{N} + O\left(\frac{q^N/2}{N}\right).$$

We recall the proof to make the constant explicit.

Let $Z_K(u)$ be the zeta function of K. Using the Euler product decomposition of $Z_K(u)$ and [Rosen 2002, Theorem 5.9], we obtain

$$Z_K(u) = \frac{\prod_{i=1}^{2g}(1-\pi_i u)}{(1-u)(1-qu)} = \prod_{d=1}^{\infty}(1-u^d)^{-S(d)}.$$

Taking the logarithmic derivative of both sides, multiplying by u, and equating coefficients of u^N yields the relation

$$q^N + 1 - \sum_{i=1}^{2g} \pi_i^N = \sum_{d \mid N} dS(d).$$

Using the Möbius inversion formula yields

$$NS(N) = \sum_{d \mid N} \mu(d) q^{N/d} + 0 - \sum_{d \mid N} \mu(d) \Biggl(\sum_{i=1}^{2g} \pi_i^{N/d} \Biggr).$$

Following the argument in [Rosen 2002, Theorem 2.2], we obtain

$$\Biggl| \sum_{d \mid N} \mu(d) q^{N/d} - q^N \Biggr| \le q^{N/2} + N q^{N/3}.$$

Similarly, using the Riemann hypothesis [Rosen 2002, Theorem 5.10], we obtain

$$\Biggl| \sum_{d \mid N} \mu(d) \Biggl(\sum_{i=1}^{2g} \pi_i^{N/d} \Biggr) \Biggr| \le 2g q^{N/2} + 2gN q^{N/4}.$$

It follows that

$$\bigl| NS(N) - q^N \bigr| \le (2g+1) q^{N/2} + N q^{N/3} + 2gN q^{N/4},$$

so

$$\begin{aligned} \biggl| S(N) - \frac{q^N}{N} \biggr| &\le \frac{2g+1}{N} q^{N/2} + q^{N/3} + 2g q^{N/4} \\ &\le \frac{q^{N/2}}{N} \biggl(2g + 1 + \frac{N}{q^{N/6}} + 2g \frac{N}{q^{N/4}} \biggr). \end{aligned} \tag{3}$$

Since $x/q^x \le 1/q$ for $x \ge 1$, (3) is less than or equal to

$$\frac{q^{N/2}}{N} \biggl(2g + 1 + \bigl(2g + \tfrac{3}{2}\bigr) \frac{4}{q} \biggr). \qquad \square$$

The next theorem follows from the effective Chebotarev density theorem in [Kumar Murty and Scherk 1994, Theorem 1].

Theorem 3.2. *Let K be a finite extension of $F = \mathbb{F}_{q_0}(T)$ with constant field $\mathbb{F}_q$ and genus g, where $q = q_0^{m_0}$. Let E be a finite Galois extension of K with Galois group G, $\mathbb{F}_{q^m}$ the algebraic closure of $\mathbb{F}_q$ in E, and $K' = \mathbb{F}_{q^m} K$ the maximal constant field extension of K in E.*

Let $\mathscr{C} \subseteq G = \mathrm{Gal}(E/K)$ be a nonempty conjugacy class in G whose restriction to $\mathbb{F}_{q^m}/\mathbb{F}_q \cong K'/K$ is τ^k, where τ is the Frobenius map $\tau(x) = x^q$, and let $\mathfrak{D}$ be the different divisor *of E/K'. Let Σ be the divisor of K that is the sum of the primes of K that are ramified in E, and suppose Σ' is a divisor of K such that $\Sigma' \geq \Sigma$. Let $B = \max\{\deg_K \Sigma', \deg_E \mathfrak{D}, 2|\mathrm{Gal}(E/K')| - 2, 1\}$.*

If

$$N \geq \frac{2}{m_0} \log_{q_0} \frac{4}{3}\Big(B^2 + B\Big(2g + \frac{g}{m} + 3\Big) + 2\Big(5g + \frac{g}{m} + 3\Big)\Big)$$

and $N \equiv k \pmod{m}$, there is a prime $\wp \notin \Sigma'$ of K such that $\deg_K \wp = N$ and $\mathrm{Frob}_\wp = \mathscr{C}$.

Proof. The situation at the outset is that we start with $F = \mathbb{F}_{q_0}(T)$ and K a finite extension of F with possibly larger constant field $\mathbb{F}_q$, where $q = q_0^n$. Next, we replace $F = \mathbb{F}_{q_0}(T)$ by $F = \mathbb{F}_q(T)$, so that K is a geometric extension of $F = \mathbb{F}_q(T)$. This allows us to use Lemma 3.1 without modification, but now q_0 is replaced by q.

Another remark is that if there exists a prime $\wp \notin \Sigma'$ of K such that $\deg_K \wp = N$ and $\mathrm{Frob}_\wp = \mathscr{C}$, then it follows that $\mathscr{C}$ restricted to $K'/K \cong \mathbb{F}_{q^m}/\mathbb{F}_q$ is τ^N by [Kumar Murty and Scherk 1994, Lemma 1]. Since $\mathrm{Gal}(\mathbb{F}_{q^m}/\mathbb{F}_q)$ is cyclic of order m, we have that $\tau^N = \tau^k$ in $\mathrm{Gal}(\mathbb{F}_{q^m}/\mathbb{F}_q)$ if and only if $N \equiv k \pmod{m}$.

Let $\mathbb{F}_{q^m}$ be the algebraic closure of $\mathbb{F}_q$ in E, so $K' := \mathbb{F}_{q^m} K$ and E/K' is a geometric extension. Let $\mathscr{D} := \deg_E \mathfrak{D}$ and $\delta' = \deg_K \Sigma'$. Let $\pi(N, \Sigma')$ be the number of primes $\wp \notin \mathrm{Supp}\, \Sigma'$ of K with $\deg_K \wp = N$, and let $\pi_{\mathscr{C}}(N, \Sigma')$ be the number of primes $\wp \notin \mathrm{Supp}\, \Sigma'$ of K such that $\deg_K \wp = N$ and $\mathrm{Frob}_\wp = \mathscr{C}$.

It suffices to find a lower bound N_0 for N such that for $N \geq N_0$, $\pi_{\mathscr{C}}(N, \Sigma')$ is positive.

In fact, the genus g of K over $\mathbb{F}_q$ is the same as that of K' over $\mathbb{F}_{q^m}$ (see [Rosen 2002, Proposition 8.9]). We know that the genus of K' over $\mathbb{F}_{q^m}$ and the genus of E over $\mathbb{F}_{q^m}$ are related by the Riemann–Hurwitz theorem [Rosen 2002, Theorem 7.16]. Thus, letting g_E be the genus of E, we have

$$g_E = 1 + |\mathrm{Gal}(E/K')|(g-1) + \tfrac{1}{2}\mathscr{D}. \tag{4}$$

The effective Chebotarev density theorem in [Kumar Murty and Scherk 1994, Theorem 1] gives

$$\frac{m|\mathscr{C}|}{|G|}\pi(N, \Sigma') - \alpha \leq \pi_{\mathscr{C}}(N, \Sigma') \leq \frac{m|\mathscr{C}|}{|G|}\pi(N, \Sigma') + \alpha,$$

where

$$\alpha = \frac{|\mathscr{C}|}{N} q^{N/2}\left(2g_E \frac{1}{|G|} + 2(2g+1) + \frac{1+N/|\mathscr{C}|}{q^{N/2}}\delta'\right). \tag{5}$$

The condition $N \equiv r \pmod m$ ensures $\mathscr{C}$ restricted to $\mathbb{F}_{q^m}/\mathbb{F}_q$ is τ^N.

Remark 3.3. When $\Sigma' = \Sigma$, this is what is proved in [Kumar Murty and Scherk 1994, Theorem 1]. However, the proof carries over with Σ replaced by Σ'. In particular, the key identity (2.1) still holds with $y \in Y_r$ unramified replaced by $y \in Y_r$ not in the support of $\Sigma' \geq \Sigma$.

We have $\pi(N, \Sigma') \geq S(N) - \dfrac{\deg_K \Sigma'}{N}$. Thus,

$$\frac{m|\mathscr{C}|}{|G|}\left(S(N) - \frac{\deg_K \Sigma'}{N}\right) - \alpha \leq \pi_{\mathscr{C}}(N, \Sigma').$$

It is therefore enough to find a lower bound for N such that

$$\frac{m|\mathscr{C}|}{|G|}\left(S(N) - \frac{\deg_K \Sigma'}{N}\right) - \alpha > 0. \tag{6}$$

From Lemma 3.1, we have

$$\begin{aligned} \frac{q^N}{N} - \left(2g+1+\left(2g+\tfrac{3}{2}\right)\frac{4}{q}\right)\frac{q^{N/2}}{N} &\leq S(N) \\ &\leq \frac{q^N}{N} + \left(2g+1+\left(2g+\tfrac{3}{2}\right)\frac{4}{q}\right)\frac{q^{N/2}}{N}. \end{aligned} \tag{7}$$

Since $|G|/m = |\mathrm{Gal}(E/L')|$ and $(1+N/|\mathscr{C}|)/q^{N/2} \leq 2$, from (5) we have

$$\alpha \;\leq\; \frac{2m|\mathscr{C}|}{N|G|}\; q^{N/2}\left(|\mathrm{Gal}(E/K')|\,(2g+1+\delta') + \frac{g_E}{m}\right). \tag{8}$$

Therefore, combining (6) through (8), we obtain

$$\begin{aligned} &\frac{m|\mathscr{C}|}{|G|}\left(S(N) - \frac{\deg_K \Sigma'}{N}\right) - \alpha \\ &\geq \frac{m|\mathscr{C}|}{N|G|}\, q^{N/2}\left(q^{N/2} - \left(c_0 + \frac{\deg_K \Sigma'}{q^{N/2}} + 2\big|\mathrm{Gal}(E/K')\big|\big(2g+1+\delta'\big) + 2\frac{g_E}{m}\right)\right), \end{aligned} \tag{9}$$

where $c_0 = 2g+1+\left(2g+\tfrac{3}{2}\right)4/q$.

We thus need to find a lower bound of N such that the right-hand side of the inequality in (9) is positive, or equivalently,

$$\begin{aligned} q^{N/2} &> c_0 + \frac{\deg_K \Sigma'}{q^{N/2}} + 2\,|\mathrm{Gal}(E/K')|\,(2g+1+\delta') + 2\frac{g_E}{m} \\ &= c_0 + \frac{\deg_K \Sigma'}{q^{N/2}} + 2\,|\mathrm{Gal}(E/K')|\,(2g+1+\delta') \\ &\qquad\qquad + \frac{2}{m}\big(1 + |\mathrm{Gal}(E/K')|\,(g-1) + \tfrac{1}{2}\mathfrak{D}\big) \\ &= c_0 + \frac{\deg_K \Sigma'}{q^{N/2}} + 2\,|\mathrm{Gal}(E/K')|\Big(2g+1+\delta' + \frac{g-1}{m}\Big) + \frac{2}{m}\big(1+\tfrac{1}{2}\mathfrak{D}\big), \end{aligned}$$

using (4).

Let $1 \le B$, $\delta' \le B$, $\mathfrak{D} \le B$, and $|\mathrm{Gal}(E/K')| \le \frac{1}{2}B + 1$. Note that if $g = 0$, it suffices to take $\delta' \le B$ and $\mathfrak{D} \le B$ only, as it is then automatic that $|\mathrm{Gal}(E/K')| \le \frac{1}{2}\mathfrak{D} + 1 \le \frac{1}{2}B + 1$. Therefore, we have

$$\begin{aligned} c_0 &+ \frac{\deg_K \Sigma'}{q^{N/2}} + 2\,|\mathrm{Gal}(E/K')|\Big(2g+1+\delta' + \frac{g-1}{m}\Big) + \frac{2}{m}\big(1+\tfrac{1}{2}\mathfrak{D}\big) \\ &\le c_0 + \frac{B}{q^{N/2}} + (B+2)\Big(2g+1+B+\frac{g-1}{m}\Big) + \frac{2}{m}\big(1+\tfrac{1}{2}B\big) \\ &\le 2g+1+\big(2g+\tfrac{3}{2}\big)\frac{4}{q} + \frac{B}{q^{N/2}} + (B+2)\Big(2g+1+B+\frac{g-1}{m}\Big) + \frac{2}{m}\big(1+\tfrac{1}{2}B\big) \\ &\le B^2 + B\Big(2g+3+\frac{g}{m}\Big) + 6g+3+\frac{2g}{m} + \big(2g+\tfrac{3}{2}\big)\frac{4}{q} + \frac{B}{q^{N/2}} \\ &\le B^2 + B\Big(2g+3+\frac{g}{m}\Big) + 10g+6+\frac{2g}{m} + \frac{B}{q^{N/2}}, \end{aligned}$$

where the last inequality uses $4/q \le 2$. Therefore, it suffices to have

$$q^{N/2} > \Big(B^2 + B\Big(2g+3+\frac{g}{m}\Big) + 10g+6+\frac{2g}{m}\Big) + \frac{B}{q^{N/2}}.$$

This can be satisfied if the following two inequalities hold:

$$\alpha q^{N/2} \ge B^2 + B\Big(2g+3+\frac{g}{m}\Big) + 10g+6+\frac{2g}{m}, \qquad (1-\alpha)q^{N/2} > \frac{B}{q^{N/2}},$$

where $0 < \alpha < 1$; equivalently,

$$N \ge 2\log_q \frac{1}{\alpha}\Big(B^2 + B\Big(2g+3+\frac{g}{m}\Big) + 10g+6+\frac{2g}{m}\Big), \qquad N > \log_q \frac{1}{1-\alpha}B.$$

Taking $\alpha = \frac{3}{4}$, the required inequalities become

$$N \ge 2\log_q \tfrac{4}{3}\Big(B^2 + B\Big(2g+3+\frac{g}{m}\Big) + 10g+6+\frac{2g}{m}\Big), \qquad N > \log_q 4B.$$

So if

$$N \geq \frac{2}{m_0} \log_{q_0} \frac{4}{3}\Big(B^2 + B\Big(2g+3+\frac{g}{m}\Big) + 2\Big(5g+3+\frac{g}{m}\Big)\Big)$$

and $N \equiv k \pmod{m}$, then there is a prime $\wp \notin \Sigma'$ of K such that $\deg_K \wp = N$ and $\mathrm{Frob}_\wp = \mathscr{C}$. $\square$

Corollary 3.4. *Let the notation and hypotheses be as in Theorem 3.2. Then there exists a prime $\wp \notin \Sigma'$ of K such that* $\mathrm{Frob}_\wp = \mathscr{C}$ *and*

$$\deg_K \wp \leq \frac{4}{m_0} \log_{q_0} \tfrac{4}{3}(B+3g+3) + m. \tag{10}$$

Proof. Let M be the integer such that

$$M = \frac{2}{m_0} \log_{q_0} \frac{4}{3}\Big(B^2 + B\Big(2g+\frac{g}{m}+3\Big) + 2\Big(5g+\frac{g}{m}+3\Big)\Big) + \delta,$$

where $0 \leq \delta < 1$. Let $N = M + k'$, where $0 \leq k' \leq m-1$ is chosen so that $N \equiv k \pmod{m}$. Then

$$N \geq \frac{2}{m_0} \log_{q_0} \frac{4}{3}\Big(B^2 + B\Big(2g+\frac{g}{m}+3\Big) + 2\Big(5g+\frac{g}{m}+3\Big)\Big)$$

and $N \equiv k \pmod{m}$. By Theorem 3.2, there exists a prime $\wp \notin \Sigma'$ of K such that $\deg_K \wp = N$ and $\mathrm{Frob}_\wp = \mathscr{C}$. Now,

$$\begin{aligned}
\deg_K \wp &= N = M + k' \\
&\leq \frac{2}{m_0} \log_{q_0} \frac{4}{3}\Big(B^2 + B\Big(2g+3+\frac{g}{m}\Big) + 10g + 6 + \frac{2g}{m}\Big) + m \\
&\leq \frac{2}{m_0} \log_{q_0} \frac{4}{3}\Big(B + 2g + 3 + \frac{g}{m}\Big)^2 + m \\
&\leq \frac{4}{m_0} \log_{q_0} \tfrac{4}{3}(B+3g+3) + m.
\end{aligned}$$

$\square$

4. Bounds for the different divisor

Proposition 4.1. *Let ϕ be a rank-r Drinfeld A-module that is integral over K, and let $\mathfrak{L} = (a)$ be a finite prime of A with a monic. Let $\mathfrak{D}_0(K_{\phi,\mathfrak{L}}/K)$ be the finite part of the different divisor $\mathfrak{D}(K_{\phi,\mathfrak{L}}/K)$. Then we have*

$$\deg_K \mathfrak{D}_0(K_{\phi,\mathfrak{L}}/K) \leq r\left(\deg_K a + \frac{(\ell^r-2)(\ell^r-1)}{q^r-1} \deg_K \Delta(\phi)\right), \tag{11}$$

where $\ell = q^{\deg_F \mathfrak{L}}$. In addition, if $v(a\Delta(\phi)) = 0$ for a finite place v of K, then

$$v(D(K_{\phi,\mathfrak{L},w}/K_v)) = 0, \tag{12}$$

where $D(K_{\phi,\mathfrak{L},w}/K_v)$ is the different ideal of $K_{\phi,\mathfrak{L},w}/K_v$, and $w|v$ is a place of $K_{\phi,\mathfrak{L},w}$.

Proof. This is a slightly modified version of [David 2001, Lemma 4.2], derived from [Taguchi 1992].

Let $\alpha \in \overline{K}$ be a root of a separable polynomial

$$f(X) = b_0 X + b_1 X^q + \cdots + b_m X^{q^m}$$

with $b_i \in \mathbb{O}$ and $b_0 b_m \neq 0$. Then

$$\begin{aligned} h(X) &= b_m^{q^m-1} f(X/b_m) \\ &= b_0 b_m^{q^m-2} X + b_1 b_m^{q^m-1-q} X^q + \cdots + b_{m-1} b_m^{q^m-1-q^{m-1}} X^{q^{m-1}} + X^{q^m} \in \mathbb{O}[X] \end{aligned}$$

is monic. Since $h(b_m\alpha) = 0$ and $K(\alpha) = K(b_m\alpha)$, we may apply Lemma 2.4 to $b_m\alpha$ and $h(X)$ to show that the *different ideal* $D(K(\alpha)/K)$ divides the principal ideal $(b_0 b_m^{q^m-2})$.

Let $\mathfrak{L} = (a)$ and $f(X) = \phi_a(X)$. Then $f(X) = aX + \cdots + \Delta(\phi)^{(q^m-1)/(q^r-1)} X^{q^m}$, where $m = r \deg_F a$ (see [Rosen 2002, Proposition 13.8]). There are r roots $\beta_1, \ldots, \beta_r$ of $\phi_a(X)$ that generate $K_{\phi,\mathfrak{L}}$ over K. Using the transitivity of the different (see Lemma 2.3), it follows that

$$D(K_{\phi,\mathfrak{L}}/K) \,\big|\, (b_0 b_m^{q^m-2})^r = \big(a\,(\Delta(\phi))^{(q^m-2)(q^m-1)/(q^r-1)}\big)^r. \tag{13}$$

This shows that if $v(a\Delta(\phi)) = 0$ for a finite place v, then $v(D(K_{\phi,\mathfrak{L},w}/K_v)) = 0$. Furthermore, taking the degree with respect to K of (13), we obtain

$$\deg_K \mathfrak{D}_0(K_{\phi,\mathfrak{L}}/K) \le r\left(\deg_K a + \frac{(\ell^r-2)(\ell^r-1)}{q^r-1} \deg_K \Delta(\phi)\right). \qquad \square$$

It is possible to obtain a bound on $\deg_K \mathfrak{D}(K_{\phi,\mathfrak{L}}/K)$ based on Proposition 4.1 and Lemma 4.2, but instead we find a slightly more refined bound in Proposition 4.3, using additional techniques.

Lemma 4.2. *Let $\bar{\infty}$ be an infinite prime of K, $K_{\bar{\infty}}$ the completion of K at $\bar{\infty}$, $\mathbb{O}_{\bar{\infty}}$ the valuation ring of $\bar{\infty}$, $v_{\bar{\infty}}$ the valuation associated to $\bar{\infty}$, and e the ramification index of $\bar{\infty}$ over ∞.*

Let $\phi_T(X) = TX + a_1X^q + \cdots + a_iX^{q^i} + \cdots + a_rX^{q^r}$ be a rank-r Drinfeld A-module defined over K, and write

$$\phi_{T^n}(X) = T^n X + b_1 X^q + \cdots + b_i X^{q^i} + \cdots + b_{rn} X^{q^{rn}},$$

where $n \ge 1$.

Let $\omega_1 = \max\left\{e, -\frac{v_{\bar{\infty}}(a_i)}{q^i} : i = 1, \ldots, r\right\}$ and $\omega_n = n\omega_1$. Then

$$\omega_n \ge \max\left\{ne, -\frac{v_{\bar{\infty}}(b_i)}{q^i} : i = 1, \ldots, rn\right\}.$$

Proof. We use induction on n. First note that

$$\phi_{T^n}(\lambda_n X) = T^n \lambda_n X + b_1 \lambda_n^q X^q + \cdots + b_i \lambda_n^{q^i} X^{q^i} + \cdots + b_{rn} \lambda_n^{q^{rn}} X^{q^{rn}},$$

so taking $\lambda_n \in K$ with

$$v_{\bar\infty}(\lambda_n) = \omega_n \geq \max\left\{ne, -\frac{v_{\bar\infty}(b_i)}{q^i} : i = 1, \ldots, rn\right\}$$

implies that $\phi_{T^n}(\lambda_n X) \in \mathbb{O}_{\bar\infty}[X]$.

The result is true for $n = 1$, as

$$\omega_1 = \max\left\{e, -\frac{v_{\bar\infty}(a_i)}{q^i} : i = 1, \ldots, r\right\}.$$

Assume

$$\omega_n = n\omega_1 \geq \max\left\{ne, -\frac{v_{\bar\infty}(b_i)}{q^i} : i = 1, \ldots, rn\right\}.$$

Now consider the terms in the product

$$\phi_{T^{n+1}} = \phi_{T^n} \circ \phi_T = (T^n + b_1\tau + \cdots + b_{rn}\tau^{rn}) \circ (T + a_1\tau + \cdots + a_r\tau^r),$$

where there are $2(r+1)$ types of terms to consider:

$$\begin{aligned} b_i\tau^i T &= b_i T^{q^i}\tau^i, \qquad 1 \leq i \leq rn,\\ b_i\tau^i a_1\tau &= b_i a_1^{q^i}\tau^{i+1}, \quad 1 \leq i \leq rn,\\ &\vdots\\ b_i\tau^i a_r\tau^r &= b_i a_r^{q^i}\tau^{i+r}, \quad 1 \leq i \leq rn,\\ T^{n+1}&, T^n a_1\tau, T^n a_2\tau^2, \ldots, T^n a_r\tau^r. \end{aligned}$$

We need to show that ω_{n+1} is greater than or equal to the valuations of the coefficients of each type of term, that is, for each i with $1 \leq i \leq rn$,

$$\omega_{n+1} \geq -\frac{v_{\bar\infty}(b_i)}{q^i} + e, \tag{14}$$

$$\omega_{n+1} \geq -\frac{v_{\bar\infty}(b_i)}{q^{i+j}} - \frac{v_{\bar\infty}(a_j)}{q^j}, \quad 1 \leq j \leq r, \tag{15}$$

$$\omega_{n+1} \geq ne + 1, \tag{16}$$

$$\omega_{n+1} \geq ne - \frac{v_{\bar\infty}(a_j)}{q^j}, \quad 1 \leq j \leq r. \tag{17}$$

As $\omega_n \geq -v_{\bar\infty}(b_i)/q^i$ for $1 \leq i \leq 2n$, we have

$$\omega_{n+1} = \omega_n + \omega_1 \geq \frac{\omega_n}{q^j} + \omega_1 \geq -\frac{v_{\bar\infty}(b_i)}{q^{i+j}} + \omega_1$$

for $j = 0, 1, \dots, r$ and $i = 1, 2, \dots, rn$, so (14) and (15) are satisfied. Since $\omega_1 = \max\{e, -v_{\bar{\infty}}(a_j)/q^j : j = 1, \dots, r\}$,

$$\begin{aligned}\omega_{n+1} = (n+1)\omega_1 = n\omega_1 + \ \omega_1 \geq ne + \omega_1 \\ \geq \max\left\{(n+1)e,\ ne - \frac{v_{\bar{\infty}}(a_j)}{q^j} : j = 1, \dots, r, \right\}\end{aligned}$$

so the last inequalities in (16) and (17) are satisfied. □

In the following proposition, we obtain an upper bound on the degree of the different divisor of $K_{\phi,\mathfrak{L}}/K$ that uses mild information from the Newton polygons of $\phi_a(X)$ and takes into account primes of potentially good reduction.

Proposition 4.3. *Let ϕ be a rank-r Drinfeld A-module that is integral over K, and let $\mathfrak{L} = (a)$ be a finite prime of A with monic. Let $\mathfrak{D}(K_{\phi,\mathfrak{L}}/K)$ be the different divisor of $K_{\phi,\mathfrak{L}}/K$. Then we have*

$$\deg_K \mathfrak{D}(K_{\phi,\mathfrak{L}}/K) \leq r\left(\frac{\ell^r - 1}{q - 1}\big(s \deg_K a + \Lambda(\phi)\big) + 2 \deg_K a \operatorname{rad}_K \Delta(\phi)\right),$$

where s denotes the geometric extension degree of K/F, $\ell = q^{\deg_F \mathfrak{L}}$, $\Lambda(\phi) = -\sum_v \tau_v(\phi) \deg_K v$, and for $x \in K$ we let $\deg_K \operatorname{rad}_K x := \sum_{v(x)\neq 0} \deg_K v$ (the sums are over every place v of K).

Proof. Let $\phi_T(X) = TX + a_1X^q + \cdots + a_rX^{q^r}$, where $a_i \in \mathbb{O}$. Let

$$f(X) = \phi_a(X) = b_0X + b_1X^q + \cdots + b_{rn}X^{q^{rn}} = b_{rn}\prod_{i=1}^{q^{rn}}(X - \alpha_i),$$

where $b_0 = a$, $b_{rn} = a_r^{(q^{rn}-1)/(q^r-1)}$, and $n = \deg_K a = \deg_K \mathfrak{L}$. Let α be any one of the α_i.

Let $\wp$ be a finite place of K with corresponding discrete valuation $v_\wp$, and let

$$\tau_\wp = \inf\left\{\frac{v_\wp(a_i)}{q^i - 1},\ i = 1, \dots, r.\right\}$$

Note that $\tau_\wp \geq 0$. Let $K_\wp$ be the completion of K at $\wp$, and $K'_\wp/K_\wp$ a totally tamely ramified extension with ramification index $1/(q^{rn} - 1)$ and ring of integers $\mathbb{O}'_\wp$.

Over $K'_\wp$, ϕ_T is isomorphic to a Drinfeld A-module

$$\phi'_T(X) = TX + a'_1X^q + \cdots + a'_rX^{q^r},$$

where $a'_i = a_i/\lambda^{q^i-1}$, $v_\wp(a'_i) \geq 0$ for $1 \leq i \leq r$, $v_\wp(\lambda) = \tau_\wp$, and $\lambda \in K'_\wp$.

Let $\phi'_a(X) = b'_0X + b'_1X^q + \cdots + b'_{rn}X^{q^{rn}}$. As $b'_i = b_i/\lambda^{q^i-1}$, we have

$$v_\wp(b_i) \geq (q^i - 1)v_\wp(\lambda) = (q^i - 1)\tau_\wp.$$

From the Newton polygon of $f(X)$, we have

$$v_\wp(\alpha) \geq -\frac{v_\wp(a_r)\dfrac{q^{rn}-1}{q^r-1} - (q^{rn-1}-1)\tau_\wp}{q^{rn}-q^{rn-1}} =: -\delta.$$

Pick a $\mu \in K'_\wp$ such that $v_\wp(\mu) = \delta + \epsilon$, where $0 \leq \epsilon < \dfrac{1}{q^{rn}-1}$. Now

$$f(X/\mu) = b_{rn}/\mu^{q^{rn}} \prod_{i=1}^{q^{rn}} (X - \mu\alpha_i),$$

and we know that $g(X) = \prod_i (X - \mu\alpha_i)$ is monic and lies in $\mathbb{O}'_\wp[X]$, where $\mathbb{O}'_\wp$ is the ring of integers of $K'_\wp$. Thus, $g'(X) = \mu^{q^{rn}-1} a/b_{rn}$. Hence,

$$\begin{aligned}
v_\wp(g'(\mu\alpha)) &= v_\wp(\mu)(q^{rn}-1) + v_\wp(a) - v_\wp(b_{rn}) \\
&\leq \delta(q^{rn}-1) + 1 + v_\wp(a) - v_\wp(a_r)\frac{q^{rn}-1}{q^r-1} \\
&\leq v_\wp(a_r)\frac{q^{rn}-1}{q^r-1}\left(\frac{q^{rn}-1}{q^{rn}-q^{rn-1}} - 1\right) - \frac{(q^{rn-1}-1)(q^{rn}-1)}{q^{rn}-q^{rn-1}}\tau_\wp + 1 + v_\wp(a) \\
&\leq v_\wp(a_r)\frac{q^{rn}-1}{q^r-1} \cdot \frac{1-q^{1-rn}}{q-1} - \frac{q^{2rn-1}-q^{rn}-q^{rn-1}+1}{q^{rn}-q^{rn-1}}\tau_\wp + 1 + v_\wp(a) \\
&= v_\wp(a_r)\frac{q^{rn}-1}{(q^r-1)(q-1)} - \frac{q^{rn}-q-1+q^{1-rn}}{q-1}\tau_\wp + 1 + v_\wp(a).
\end{aligned}$$

It follows that

$$v_\wp\big(D(K'_\wp(\mu\alpha)/K'_\wp)\big) \leq v_\wp(a_r)\frac{q^{rn}-1}{(q^r-1)(q-1)} - \frac{q^{rn}-q-1+q^{1-rn}}{q-1}\tau_\wp + 1 + v_\wp(a)$$

and

$$\begin{aligned}
&v_\wp\big(D(K_\wp(\alpha)/K_\wp)\big) \\
&\qquad \leq v_\wp\big(D(K'_\wp(\mu\alpha)/K'_\wp)\big) + v_\wp(D(K'_\wp/K_\wp)) \\
&\qquad \leq v_\wp(a_r)\frac{q^{rn}-1}{(q^r-1)(q-1)} - \frac{q^{rn}-q-1+q^{1-rn}}{q-1}\tau_\wp + 2 + v_\wp(a).
\end{aligned}$$

Since $\tau_\wp \leq v_\wp(a_r)/(q^r-1)$, we have

$$\begin{aligned}
&v_\wp(a_r)\frac{q^{rn}-1}{(q^r-1)(q-1)} - \frac{q^{rn}-q-1+q^{1-rn}}{q-1}\tau_\wp + 2 + v_\wp(a) \\
&\qquad \geq v_\wp(a_r)\frac{q^{rn}-1}{(q^r-1)(q-1)} - \frac{q^{rn}-q-1+q^{1-rn}}{q-1}\frac{v_\wp(a_r)}{q^r-1} + 2 + v_\wp(a) \\
&\qquad = v_\wp(a_r)\frac{q-q^{1-rn}}{(q^r-1)(q-1)} + 2 + v_\wp(a) \geq 2.
\end{aligned}$$

From Proposition 4.1, we know that for a finite place $v_\wp$ of K, $v_\wp(D(K(\alpha)/K))=0$ if $v_\wp(aa_r)=0$. It follows that

$$\text{(18)}\quad v_\wp\big(D(K_\wp(\alpha)/K_\wp)\big) \le v_\wp(a_r)\frac{q^{rn}-1}{(q^r-1)(q-1)} - \frac{q^{rn}-q-1+q^{1-rn}}{q-1}\tau_\wp + 2\nu + v_\wp(a),$$

where $\nu=1$ if $v_\wp(aa_r)>0$ and $\nu=0$ if $v_\wp(aa_r)=0$.

Let $\bar\infty \in S_\infty^K$ be an infinite prime of K with corresponding valuation $v_{\bar\infty}$, and let $K'_{\bar\infty}/K_{\bar\infty}$ be a totally tamely ramified extension with ramification index $1/(q^{rn}-1)$ and ring of integers $\mathbb{O}'_{\bar\infty}$.

Let

$$\tau_{\bar\infty}(\phi)=\inf\left\{\frac{v_{\bar\infty}(a_i)}{q^i-1},\ i=1,\dots,r.\right\}$$

Note that $\tau_{\bar\infty}\le 0$.

Over $K'_{\bar\infty}$, ϕ_T is isomorphic to a Drinfeld A-module

$$\phi'_T(X)=TX+a'_1X^q+\cdots+a'_rX^{q^r},$$

where $a'_i=a_i/\lambda^{q^i-1}$, $v_{\bar\infty}(a'_i)\ge 0$ for $1\le i\le r$, $v_{\bar\infty}(\lambda)=\tau_{\bar\infty}$, and $\lambda\in K'_{\bar\infty}$.

Let $\phi'_a(X)=b'_0X+b'_1X^q+\cdots+b'_{rn}X^{q^{rn}}$. Set

$$\omega_1=\max\left\{e,-\frac{v_{\bar\infty}(a'_i)}{q^i}:i=1,\dots,r\right\}=1.$$

From Lemma 4.2, we know that

$$\omega_n=n\omega_1\ge\max\left\{ne,-\frac{v_{\bar\infty}(b'_i)}{q^i}:i=1,\dots,rn\right\}.$$

Thus, $v_{\bar\infty}(b'_i)\ge -q^ine$ for $i=1,\dots,rn$. As $b'_i=b_i/\lambda^{q^i-1}$, we have

$$v_{\bar\infty}(b_i)\ge -q^ine+(q^i-1)v_{\bar\infty}(\lambda)=-q^ine+(q^i-1)\tau_{\bar\infty}.$$

From the Newton polygon of $f(X)$, it follows that

$$v_{\bar\infty}(\alpha)\ge -\frac{v_{\bar\infty}(a_r)\dfrac{q^{rn}-1}{q^r-1}+neq^{rn-1}-(q^{rn-1}-1)\tau_{\bar\infty}}{q^{rn}-q^{rn-1}} =: -\delta_{\bar\infty}.$$

Let $\mu_{\bar\infty}$ be such that $v_{\bar\infty}(\mu_{\bar\infty})=\delta_{\bar\infty}+\epsilon_\infty$, where $0\le\epsilon_\infty<1/(q^{rn}-1)$. Now

$$f(X/\mu_{\bar\infty})=b_{rn}/\mu_{\bar\infty}^{q^{rn}}\prod_{i=1}^{q^{rn}}(X-\mu_{\bar\infty}\alpha_i),$$

and we know that $g(X)=\prod_{i=1}^{q^{rn}}(X-\mu_{\bar\infty}\alpha_i)$ is monic and lies in $\mathbb{O}'_{\bar\infty}[X]$, where $\mathbb{O}'_{\bar\infty}$ is the ring of integers of $K'_{\bar\infty}$. Thus, $g'(X)=\mu_{\bar\infty}^{q^{rn}-1}a/b_{rn}$. Hence,

$$\begin{aligned}
&v_{\bar\infty}(g'(\mu_{\bar\infty}\alpha))\\
&=v_{\bar\infty}(\mu_{\bar\infty})(q^{rn}-1)+v_{\bar\infty}(a)-v_{\bar\infty}(b_{rn})\\
&\le \delta_{\bar\infty}(q^{rn}-1)+1+v_{\bar\infty}(a)-v_{\bar\infty}(a_r)\frac{q^{rn}-1}{q^r-1}\\
&\le v_{\bar\infty}(a_r)\frac{q^{rn}-1}{q^r-1}\left(\frac{q^{rn}-1}{q^{rn}-q^{rn-1}}-1\right)\\
&\qquad\qquad +ne\frac{q^{rn}-1}{q-1}-\frac{(q^{rn-1}-1)(q^{rn}-1)}{q^{rn}-q^{rn-1}}\tau_{\bar\infty}+1+v_{\bar\infty}(a)\\
&=v_{\bar\infty}(a_r)\frac{q^{rn}-1}{q^r-1}\cdot\frac{1-q^{1-rn}}{q-1}+ne\frac{q^{rn}-1}{q-1}-\frac{q^{2rn-1}-q^{rn}-q^{rn-1}+1}{q^{rn}-q^{rn-1}}\tau_{\bar\infty}+1+v_{\bar\infty}(a)\\
&=v_{\bar\infty}(a_r)\frac{q^{rn}-1}{(q^r-1)(q-1)}+ne\frac{q^{rn}-1}{q-1}-\frac{q^{rn}-q-1+q^{1-rn}}{q-1}\tau_{\bar\infty}+1+v_{\bar\infty}(a).
\end{aligned}$$

It follows that

$$\begin{aligned}
(19)\quad v_{\bar\infty}\big(D(K_{\bar\infty}(\alpha)/K_{\bar\infty})\big)\le v_{\bar\infty}(a_r)&\frac{q^{rn}-1}{(q^r-1)(q-1)}+ne\frac{q^{rn}-1}{q-1}\\
&-\frac{q^{rn}-q-1+q^{1-rn}}{q-1}\tau_{\bar\infty}+2+v_{\bar\infty}(a).
\end{aligned}$$

Let $\mathfrak{D}(K(\alpha)/K)$ be the different divisor of $K(\alpha)$ over K, and Ω_P the set of conjugates of α over K_P. Using (18) and (19), we obtain

$$\begin{aligned}
\deg_K\mathfrak{D}(K(\alpha)/K)&=\sum_P\max\big\{v_P(D(K_P(\alpha)/K_P)):\alpha\in\Omega_P\big\}\deg_K P\\
&\le n\frac{q^{rn}-1}{q-1}\sum_{\bar\infty\in S_\infty^K}e(\bar\infty/\infty)\deg_K\bar\infty\\
&\qquad\qquad -\frac{q^{rn}-q-1+q^{1-rn}}{q-1}\sum_v\tau_P\deg_K P+2\deg_K\operatorname{rad}_K aa_r\\
&=n\frac{q^{rn}-1}{q-1}\sum_{\bar\infty\in S_\infty^K}e(\bar\infty/\infty)\frac{f(\bar\infty/\infty)}{[\mathbb{F}_K:\mathbb{F}_F]}\deg_F\infty\\
&\qquad\qquad -\frac{q^{rn}-q-1+q^{1-rn}}{q-1}\sum_v\tau_P\deg_K P+2\deg_K\operatorname{rad}_K aa_r\\
&\le n\frac{q^{rn}-1}{q-1}s-\frac{q^{rn}-q-1+q^{1-rn}}{q-1}\sum_v\tau_P\deg_K P+2\deg_K\operatorname{rad}_K aa_r,
\end{aligned}$$

where the summation runs through all the primes P of K, s is the geometric extension degree of K/F, and we use the fact that $\sum_P v_P(x)\deg_K P = 0$ for $x \in K$. Remark that $\sum_P \tau_P \deg_K P \le 0$; so we finally get

$$\begin{aligned}
&\deg_K \mathfrak{D}(K(\alpha)/K)\\
&\quad\le ns\frac{q^{rn}-1}{q-1}+\frac{q^{rn}-q-1+q^{1-rn}}{q-1}\Bigl(-\sum_v \tau_P \deg_K P\Bigr)+2\deg_K \operatorname{rad}_K aa_r\\
&\quad\le ns\frac{q^{rn}-1}{q-1}+\frac{q^{rn}-1}{q-1}\Bigl(-\sum_P \tau_P \deg_K P\Bigr)+2\deg_K \operatorname{rad}_K aa_r\\
&\quad\le \frac{q^{rn}-1}{q-1}\Bigl(ns-\sum_v \tau_P \deg_K P\Bigr)+2\deg_K \operatorname{rad}_K aa_r\\
&\quad\le \frac{\ell^r-1}{q-1}(ns+\Lambda(\phi))+2\deg_K \operatorname{rad}_K aa_r\\
&\quad\le \frac{\ell^r-1}{q-1}(s\deg_K a+\Lambda(\phi))+2\deg_K \operatorname{rad}_K a\Delta(\phi).
\end{aligned}$$

Using transitivity of the different (see Lemma 2.3) and the fact that $K_{\phi,\mathfrak{L}}$ is generated by r of the roots α_i, the result follows. □

Corollary 4.4. *Assume the notation of Proposition 4.3. Let ϕ_1 and ϕ_2 be rank-r Drinfeld A-modules that are integral over $\mathbb{O}$. Let $\mathfrak{D}(\tilde{K}/K)$ be the different divisor of $\tilde{K}/K$, where $\tilde{K} = K_{\phi_1,\mathfrak{L}}K_{\phi_2,\mathfrak{L}}$. Then we have*

$$\deg_K \mathfrak{D}(\tilde{K}/K) \le r\biggl(\frac{\ell^r-1}{q-1}\bigl(2s\deg_K a+\Lambda(\phi_1,\phi_2)\bigr)+2D(\phi_1,\phi_2)+4\deg_K a\biggr),$$

where $\Lambda(\phi_1,\phi_2)=\Lambda(\phi_1)+\Lambda(\phi_2)$.

Proof. The result follows from Lemma 2.6 and Proposition 4.3. □

5. Proof of Theorem 1.2

We first recall some intermediate results, which are function field analogues of those found in [Serre 1981] (see [Gardeyn 2002]).

Lemma 5.1. *We have*

$$\sum_{1\le \deg_F \mathfrak{L}\le N} \deg_F \mathfrak{L} \ge q^N$$

for all positive integers N, where the sum is over finite primes $\mathfrak{L}$ of F.

Proof. The product of all finite primes $\mathfrak{L}$ of F such that $\deg \mathfrak{L}$ divides N is equal to $T^{q^N}-T$, so the inequality follows. □

Lemma 5.2. *For any nonzero $n \in A$, there exists a finite prime $\mathfrak{L}$ of F such that $n \not\equiv 0 \pmod{\mathfrak{L}}$ with $\deg_F \mathfrak{L} \leq 1 + \log_q \deg_F n$.*

Proof. Suppose $n \equiv 0 \pmod{\mathfrak{L}}$ for all the primes $\mathfrak{L}$ such that

$$1 \leq \deg_F \mathfrak{L} \leq 1 + \log_q \deg_F n.$$

Choose $k := \lfloor 1 + \log_q \deg_F n \rfloor$, so that $k - 1 \leq \log_q \deg_F n < k$, and hence $q^{k-1} \leq \deg_F n < q^k$.

Then $\prod_{1 \leq \deg_F \mathfrak{L} \leq k}$ divides n, so $q^k \leq \deg_F n$, by Lemma 5.1. But $\deg_F n < q^k$, which is a contradiction. □

For the proof of Theorem 1.2, we will require an estimate of the form

$$\gamma x^t \leq \frac{x}{1 + \log_q x}, \tag{20}$$

for $x \geq C$.

Lemma 5.3. *Let $c^* \geq 1$ be given and set $t^* = 1 - 1/\ln(qc^*)$. Then we have*

$$\gamma x^{t^*} \leq \frac{x}{1 + \log_q x} \tag{21}$$

for $x \geq c^*$, where

$$\gamma = \frac{(c^*)^{1-t^*}}{1 + \log_q c^*} = \frac{(c^*)^1/\ln(qc^*)}{1 + \log_q c^*}.$$

Proof. The inequality

$$\gamma x^t \leq \frac{x}{1 + \log_q x}$$

is equivalent to

$$f(x, t) = \frac{x^{1-t}}{1 + \log_q x} \geq \gamma.$$

For a fixed t, taking the derivative of f with respect to x,

$$f'(x, t) = x^{-t}\Big((1-t)(1 + \log_q x) - \frac{1}{\ln q}\Big)/*^2,$$

where $* = (1 + \log_q x)$. Hence, $f'(x, t) \geq 0$ is equivalent to

$$(1-t)(1 + \log_q x) - \frac{1}{\ln q} \geq 0,$$

or equivalently,

$$(1-t)(\ln q + \ln x) \geq 1. \tag{22}$$

Assuming $t < 1$, (22) is equivalent to

$$x \geq \frac{e^{1/(1-t)}}{q} =: \beta(t).$$

Thus, for a fixed $t < 1$, $f(x,t)$ is increasing with respect to x, when $x \geq \beta(t)$; that is, $f(x,t) \geq f(\beta(t),t)$ if $x \geq \beta(t)$. Now, $\beta(t^*) = c^*$ and $t^* < 1$, so we obtain

$$x^{t^*} f(c^*, t^*) \leq \frac{x}{1+\log_q x},$$

for $x \geq c^*$. □

Lemma 5.4.

$$\log_q(x+y) \leq \max\{\log_q(2x), \log_q(2y)\}, \tag{23}$$

$$\log_q(x+y) \leq \log_q x + \log_q y \quad \textit{if } x, y \geq 2. \tag{24}$$

Proof. In order to have $z \geq \log_q(x+y)$, it suffices to have

$$\tfrac{1}{2}q^z \geq x \quad \text{and} \quad \tfrac{1}{2}q^z \geq y,$$

which is equivalent to

$$z \geq \log_q(2x) \quad \text{and} \quad z \geq \log_q(2y).$$

Thus, taking $z = \max\{\log_q(2x),\ \log_q(2y)\}$, we have

$$\log_q(x+y) \leq \max\{\log_q(2x),\ \log_q(2y)\}.$$ □

Conclusion of the proof of Theorem 1.2. Let $\wp \notin S$ be a prime of K with least degree such that $P_\wp(\phi_1) \neq P_\wp(\phi_2)$, where S is the given finite set of primes of K outside of which both ϕ_1 and ϕ_2 have good reduction. Let α_0 be a nonzero coefficient of $P_\wp(\phi_1) - P_\wp(\phi_2)$.

It is known that a root γ of $P_\wp(\phi_1)$ or $P_\wp(\phi_2)$ satisfies

$$v_\infty(\gamma) = -\frac{1}{r}\deg_K \wp$$

(see [Goss 1992, Theorem 3.2.3(c)(d); Gardeyn 2002, Proposition 9]). This implies that each coefficient β of $P_\wp(\phi_1)$ and $P_\wp(\phi_2)$ satisfies $\deg_F \beta \leq \deg_K \wp$, and hence each coefficient α of $P_\wp(\phi_1) - P_\wp(\phi_2)$ also satisfies $\deg_F \alpha \leq \deg_K \wp$; in particular $\deg_F \alpha_0 \leq \deg_K \wp$.

We choose a finite prime $\mathfrak{L}$ of F by Lemma 5.2 such that

$$\alpha_0 \not\equiv 0 \pmod{\mathfrak{L}} \quad \text{and} \quad \deg_F \mathfrak{L} \leq 1 + \log_q \deg_K \wp, \tag{25}$$

and write $\mathfrak{L} = (a)$, where a is monic in A.

Suppose $\wp$ lies above the prime $\mathfrak{p}$ of F. For $x \geq 7$, we have $\log_q x < \frac{1}{2}(x-1)$ (since if we let $f(x) = \frac{1}{2}(x-1) - \log_q x$, then $f'(x) > 0$ for $x \geq 7$ and $f(7) > 0$). Hence, we obtain that $x < q^{(1/2)(x-1)}$, so $q^{(1/2)(x-1)}/x > 1$; hence, $q^{x-1}/x > q^{(1/2)(x-1)}$ for $x \geq 7$. Thus, noting that

$$s \geq \frac{f(\wp/\mathfrak{p})}{[\mathbb{F}_K : \mathbb{F}_F]},$$

if $x \geq \max\{1 + 2\log_q s, 7\}$, we get that

$$\frac{q^{x-1}}{x} > q^{(1/2)(x-1)} \geq s \geq \frac{f(\wp/\mathfrak{p})}{[\mathbb{F}_K : \mathbb{F}_F]}.$$

But then if $\mathfrak{L} = \wp$, we would have

$$\deg_F \mathfrak{p} \leq 1 + \log_q \deg_K \wp = 1 + \log_q \frac{f(\wp/\mathfrak{p})}{[\mathbb{F}_K : \mathbb{F}_F]} \deg_F \mathfrak{p};$$

in other words,

$$\frac{q^{x-1}}{x} \leq \frac{f(\wp/\mathfrak{p})}{[\mathbb{F}_K : \mathbb{F}_F]},$$

where $x = \deg_F \mathfrak{p} = \deg_F \mathfrak{L}$. Therefore, we either have that

$$\deg_F \mathfrak{p} \leq \max\{1 + 2\log_q s, 7\},$$

or $\mathfrak{L} \neq \mathfrak{p}$ by the above inequality. In the former case, it follows that $\deg_K \wp \leq s \max\{1 + 2\log_q s, 7\}$.

Suppose we are now in the latter case, where $\mathfrak{L} \neq \mathfrak{p}$. Consider the representation

$$\psi_{\mathfrak{L}} : G_K \to \mathrm{Aut}_{A/\mathfrak{L}}(\phi_1[\mathfrak{L}]) \times \mathrm{Aut}_{A/\mathfrak{L}}(\phi_2[\mathfrak{L}]) \cong \mathrm{GL}_r(A/\mathfrak{L}) \times \mathrm{GL}_r(A/\mathfrak{L}),$$

where $\psi_{\mathfrak{L}} = \rho_{\phi_1,\mathfrak{L}} \times \rho_{\phi_2,\mathfrak{L}}$. Let $G_{\mathfrak{L}}$ be the image of this homomorphism. Let $C_{\mathfrak{L}}$ be the subset of $G_{\mathfrak{L}}$ consisting of pairs (a, b) such that the characteristic polynomials of a and b are not equal. Note that $C_{\mathfrak{L}}$ is invariant under conjugation, so it is a union of conjugacy classes in $G_{\mathfrak{L}}$. Since $\mathfrak{L} \neq \mathfrak{p}$, we have $C_{\mathfrak{L}} \neq \varnothing$; in particular, there is some conjugacy class $\mathscr{C} \subseteq C_{\mathfrak{L}}$ in $G_{\mathfrak{L}}$ with $\mathscr{C} \neq \varnothing$.

Let $S_{\mathfrak{L}} = S \cup \{\text{primes } \mathfrak{l} \text{ of } K \text{ lying over } \mathfrak{L}\}$. Then the Galois representation $\psi_{\mathfrak{L}}$ is unramified outside $S_{\mathfrak{L}}$. We have $A/\mathfrak{L} \cong \mathbb{F}_\ell$, where $\ell = q^{\deg_F \mathfrak{L}}$.

Let $\tilde{K}/K$ be the field extension associated to $\psi_{\mathfrak{L}}$, and let n (resp. n') be its degree (resp. geometric extension degree). Applying Corollary 3.4 to $\tilde{K}/K$, and using Lemma 2.2 together with the bound for the degree with respect to K of $\mathfrak{D} = \mathfrak{D}(\tilde{K}/K)$ given in Corollary 4.4, we deduce that there is a prime $P \notin S_{\mathfrak{L}}$ such that $\mathrm{Frob}_P = \mathscr{C} \subseteq C_{\mathfrak{L}}$ and

$$\deg_K P \leq \frac{4}{m_0} \log_q \tfrac{4}{3}(B + 3g + 3) + m,$$

where

$$\Sigma' = \sum_{\mathfrak{p}\in S_{\mathfrak{L}}} \mathfrak{p} \geq \Sigma = \sum_{\mathfrak{p}\in S} \mathfrak{p}, \quad m = [\mathbb{F}_{\tilde{K}} : \mathbb{F}_K], \quad m_0 = [\mathbb{F}_K : \mathbb{F}_F],$$

$$\deg_K \Sigma' = \deg_K \operatorname{rad}_K \Delta(\phi_1)\Delta(\phi_2) + \deg_K \mathfrak{L},$$

$$B = \max\{\deg_K \Sigma', \deg_{\tilde{K}} \mathfrak{D}, 2|\mathrm{Gal}(E/K')| - 2, 2\},$$

$$\deg_{\tilde{K}} \mathfrak{D} \leq rn'\left(\frac{\ell^r - 1}{q-1}\big(2s \deg_K a + \Lambda(\phi_1, \phi_2)\big) + 2D(\phi_1, \phi_2) + 4 \deg_K a\right).$$

Then

$$\deg_K P \leq \frac{4}{m_0} \log_q \tfrac{4}{3}(B + 3g + 3) + m \leq \frac{4}{m_0}\big(\log_q \tfrac{4}{3} B + \log_q 4(g+1)\big) + m, \tag{26}$$

using $B \geq 2$ and Lemma 5.4. Note that regarding B, the terms $\deg_K \Sigma'$ and $2|\mathrm{Gal}(E/K')| - 2$ are less than the bound we use for $\deg_{\tilde{K}} \mathfrak{D}$, so we can ignore them later on when we bound B.

Using Lemma 5.4, we obtain

$$\begin{aligned}
&\log_q \deg_{\tilde{K}} \mathfrak{D} \\
&\quad = \log_q rn'\left(\frac{\ell^r-1}{q-1}\Lambda(\phi_1,\phi_2) + 2D(\phi_1,\phi_2) + \left(2s\frac{\ell^r-1}{q-1} + 4\right)\deg_K a\right) \\
&\quad \leq \log_q rn' + \log_q\left(\frac{\ell^r-1}{q-1}(\Lambda(\phi_1,\phi_2) + 2D(\phi_1,\phi_2)) + \left(2s\frac{\ell^r-1}{q-1} + 4\right)\deg_K a\right) \\
&\quad \leq \log_q rn' + \max\{V_1,\ V_2\},
\end{aligned}$$

where

$$\begin{aligned}
V_1 &:= \log_q 2\frac{\ell^r-1}{q-1}\big(\Lambda(\phi_1, \phi_2) + 2\,D(\phi_1, \phi_2)\big) \\
&= \log_q 2 + \log_q \frac{\ell^r-1}{q-1} + \log_q\big(\Lambda(\phi_1, \phi_2) + 2\,D(\phi_1, \phi_2)\big), \\
V_2 &:= \log_q 2\Big(2s\frac{\ell^r-1}{q-1} + 4\Big)\deg_K a \\
&\leq \log_q 2 + \log_q 8s + \log_q \frac{\ell^r-1}{q-1} + \log_q \deg_K a \leq V_1 + \log_q 8s + \log_q \deg_K a.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\log_q B \\
&\leq \log_q rn' + V_1 + \log_q 8s + \log_q \deg_K a \\
&= \log_q rn' + \log_q 16s + \log_q \frac{\ell^r-1}{q-1} + \log_q \deg_K a + \log_q\big(\Lambda(\phi_1, \phi_2) + 2\,D(\phi_1, \phi_2)\big).
\end{aligned}$$

Since $n' \leq n = |G_{\mathfrak{L}}| < \ell^{2r^2}$, $\log_q \ell = \deg_F \mathfrak{L} = \deg_F a$, and $\deg_K a \leq s \deg_F a = s \log_q \ell$, we finally obtain

$$\begin{aligned}(27)\quad &\log_q B\\ &\le \log_q 16rs^2 + (2r^2+r)\log_q \ell + \log_q \log_q \ell + \log_q\bigl(\Lambda(\phi_1,\phi_2) + 2\,D(\phi_1,\phi_2)\bigr).\end{aligned}$$

Note that if $\log_q\bigl(\Lambda(\phi_1,\phi_2) + 2\,D(\phi_1,\phi_2)\bigr) = 0$, the derivation of the bound (27) above can be modified so as to obtain

$$\log_q B \le \log_q 16rs^2 + (2r^2+r)\log_q \ell + \log_q \log_q \ell. \tag{28}$$

Thus, we have

$$\log_q \tfrac{4}{3}B \le \log_q \tfrac{64}{3}rs^2 + (2r^2+r+1)\log_q \ell + \log_q^* \left(\Lambda(\phi_1,\phi_2) + 2\,D(\phi_1,\phi_2)\right). \tag{29}$$

Returning to (26), we obtain

$$\begin{aligned}\deg_K P \le \frac{4}{m_0}\bigl(&\log_q 86rs^2(g+1) + (2r^2+r+1)\log_q \ell\\ &+ \log_q^*(\Lambda(\phi_1,\phi_2) + 2\,D(\phi_1,\phi_2))\bigr) + m.\end{aligned} \tag{30}$$

By construction of $C_{\mathfrak{L}}$, we have $P_P(\phi_1) \not\equiv P_P(\phi_2) \pmod{\mathfrak{L}}$. Thus, $\deg_K \wp \le \deg_K P$, and from (25), it follows that

$$\begin{aligned}\deg_K \wp &\le \frac{4}{m_0}\bigl(\log_q 86rs^2(g+1) + (2r^2+r+1)\log_q \ell\\ &\qquad + \log_q^*(\Lambda(\phi_1,\phi_2) + 2\,D(\phi_1,\phi_2))\bigr) + m\\ &\le \frac{4}{m_0}\bigl(\log_q 86rs^2(g+1) + (2r^2+r+1)(1+\log_q \deg_K \wp)\\ &\qquad + \log_q^*(\Lambda(\phi_1,\phi_2) + 2\,D(\phi_1,\phi_2))\bigr) + m.\end{aligned} \tag{31}$$

As $1+\log_q x \ge 1$, $\dfrac{\log_q x}{x} \le 1$, we have

$$\frac{\deg_K \wp}{1+\log_q(\deg_K \wp)} \le \frac{4}{m_0}(d_r + W),$$

where $c_r = 2r^2+r+1$, $d_r := c_r + \log_q 86rs^2(g+1)$, and

$$W := \log_q^*\bigl(\Lambda(\phi_1,\phi_2) + 2\,D(\phi_1,\phi_2)\bigr) + mm_0.$$

If $x \ge d_r$, then using Lemma 5.3 with $c^* = d_r$ and $x = \deg_K \wp$, we obtain

$$\gamma x^{t^*} \le \frac{x}{1+\log_q x} \le \frac{4}{m_0}(d_r + W),$$

where γ is as in Lemma 5.3. This implies that

$$x^{t^*} \le \frac{4}{m_0}\frac{(d_r + W)}{\gamma},$$

so that

$$\log_q \deg_K \wp = \log_q x \leq \frac{1}{t^*}\log_q \frac{4}{m_0}(d_r+W)\cdot\frac{1+\log_q d_r}{(d_r)^{1/\ln(qd_r)}} \tag{32}$$
$$\leq s^*\Big(\log_q\frac{4}{m_0}+\log_q(d_r+W)+\log_q(1+\log_q d_r)-\frac{1}{\ln(qd_r)}\log_q d_r\Big)$$
$$\leq s^*\Big(\log_q\frac{4}{m_0}+\log_q d_r+\log_q W+\log_q(1+\log_q d_r)-\frac{1}{\ln(qd_r)}\log_q d_r\Big)$$
$$\leq s^*\Big(\log_q\frac{4}{m_0}+\log_q W+\log_q\log_q d_r\Big)+\log_q d_r,$$

using d_r, $W \geq 2$, and where

$$t^* = \frac{\ln(qd_r)-1}{\ln(qd_r)} \quad \text{and} \quad s^* = s^*_{q,r} = \frac{1}{t^*} = \frac{\ln(qd_r)}{\ln(qd_r)-1}.$$

We note that when q or r is large, $s^*_{q,r}$ tends to 1 from above.

Substitution of (32) into (31) yields

$$\tfrac{1}{4}\deg_K \wp \leq \log_q 86rs^2(g+1)+c_r(1+\log_q \deg_K \wp)+W \tag{33}$$
$$\leq \log_q 86rs^2(g+1)$$
$$+\; c_r\Big(1+s^*\Big(\log_q\frac{4}{m_0}+\log_q W+\log_q\log_q d_r\Big)+\log_q d_r\Big)+W$$
$$=\log_q 86rs^2(g+1)+c_r\Big(1+s^*\log_q\frac{4}{m_0}+\log_q d_r\Big)$$
$$+\; c_r s^*\log_q\log_q d_r+W+c_r s^*\log_q W$$
$$=C_{q,r}+W+c_r s^*_{q,r}\log_q W,$$

where

$$C_{q,r}=\log_q 86rs^2(g+1)+c_r\Big(1+s^*_{q,r}\log_q\frac{4}{m_0}+\log_q d_r\Big)+c_r s^*_{q,r}\log_q\log_q d_r.$$

Therefore, we either have the above upper bound (33) on $\deg_K \wp$ or $\deg_K \wp \leq d_r \leq C_{q,r}$, so in the end we get

$$\deg_K \wp \leq \frac{4}{m_0}\big(C_{q,r}+W+c_r s_{q,r}\log_q W\big). \tag{34}$$

Finally, we note from the discussion in the introduction that $m \leq g_{\phi_1}g_{\phi_2}$.

6. The case of rank 2

In this section, we consider the case of rank 2 and $K = F$, and explain how to make all the terms explicit in our isogeny theorem.

For a Drinfeld A-module ϕ of rank 2 over $K = F = \mathbb{F}_q(T)$, the successive minima of the lattices associated to the uniformizations of ϕ are determined in [Chen and Lee 2013], and this is used to obtain an explicit bound for the valuation $v_\infty(D(K_\infty(\phi[a])/K_\infty))$ of the different of $K_{\phi,a} = K(\phi[a])$ over K at the infinite prime ∞ of K and $v_\mathfrak{p}(D(K_\mathfrak{p}(\phi[a])/K_\mathfrak{p}))$ at a finite prime $\mathfrak{p}$ of K, following the work of Goss [1996].

The infinite prime case is obtained using the explicit information about the Newton polygon of the exponential map $e_{\phi,\infty}$ attached to ϕ from its uniformization over C_∞.

Assume the same notation as in the proof and statement of Proposition 4.3, taking $K = F = \mathbb{F}_q(T)$ and $\bar{\infty} = \infty$; the explicit bounds given in [Chen and Lee 2013] are as follows.

Let $\phi_T = T + a_1\tau + a_2\tau^2$ and $j(\phi) = a_1^{q+1}/a_2$, and let m be the least positive integer such that $-v_\infty(j(\phi)) \le q^{m+1}$. Then we have

$$v_\infty\big(D(K_\infty(\phi[a])/K_\infty)\big) \le \begin{cases} 1 & \text{if } -v_\infty(j(\phi)) \le q, \\ 1+\kappa(q^{\kappa+1}-1) & \text{if } q < -v_\infty(j(\phi)) \le q^{m+1}, \end{cases}$$

where

$$\kappa = \frac{-v_\infty(j(\phi)) - q^m}{q^m(q-1)} + m - 1,$$

and

$$v_\mathfrak{p}\big(D(K_\mathfrak{p}(\phi[a])/K_\mathfrak{p})\big) \le \begin{cases} 2v_\mathfrak{p}(a) & \text{if } \phi \text{ has good reduction over } K_\mathfrak{p}, \\ 2v_\mathfrak{p}(a)+1 & \text{if } v_\mathfrak{p}(j(\phi)) \ge 0 \text{ and } \phi \text{ has bad reduction over } K_\mathfrak{p}, \\ 2v_\mathfrak{p}(a)+1-\dfrac{2}{q-1}v_\mathfrak{p}(j(\phi)) & \text{if } v_\mathfrak{p}(j(\phi)) < 0. \end{cases}$$

Putting this together yields the following explicit bound on the different divisor of $F(\phi[a])/F$ when ϕ has rank 2, which can be used in place of the more general bound that we use in this paper. See Section 7 for a comparison of the two bounds in the context of our application.

Theorem 6.1. *Let ϕ be a Drinfeld A-module of rank 2 over F, and $\mathfrak{D}(F(\phi[a])/F)$ the different divisor of $F(\phi[a])/F$. Then*

$$\deg_F \mathfrak{D}(F(\phi[a])/F) \le 2\deg_F a + \deg_F \eta + \frac{2}{q-1}\deg_F \delta + v_\infty\big(D(F_\infty(\phi[a])/F_\infty)\big),$$

where δ is the (monic) denominator of $j(\phi)$ as represented by a fraction in reduced form, and η is the product of finite primes $\mathfrak{p}$ such that ϕ has bad reduction over $F_\mathfrak{p}$.

Concerning the term g_ϕ, we have from [Gardeyn 2002] that

$$g_\phi = g_{\phi,\infty} \le (q^2-1)(q^2-q)\nu_{2,\phi,\infty}/\nu_{1,\phi,\infty},$$

where $\nu_{i,\phi,\infty}$ is the i-th successive minimum of ϕ associated to its uniformization over C_∞. In [Chen and Lee 2013], the $\nu_{i,\phi,\infty}$ are determined as follows.

Case 1: If $-v(j(\phi)) \le q$, then $\nu_{1,\phi,\infty} = \nu_{2,\phi,\infty} = -s_1$.

Case 2: If $q < -v(j(\phi)) \le q^{m+1}$, then $\nu_{1,\phi,\infty} = -s_1$, $\nu_{2,\phi,\infty} = -s_1 - \kappa$, where $s_1 = (v(a_2)+q^2)/(q^2-1)$ in Case 1 and $s_1 = (v(a_1)+q)/(q-1)$ in Case 2, and m, κ are as above.

7. Comparison with work of Gardeyn

In this section we make some detailed comparisons with the work in [Gardeyn 2002], where an effective isogeny theorem is proven.

For the proof of our Theorem 1.2, an essential ingredient is the bound on the different divisor given in Proposition 4.3,

$$\begin{aligned}(35)\quad & \deg_K \mathfrak{D}(K_{\phi,\mathfrak{L}}/K) \\ & \le r\left(\frac{\ell^r-1}{q-1}\big(s\deg_K a + \Lambda(\phi)\big) + 2\deg_K \operatorname{rad}_K \Delta(\phi) + 2\deg_K a\right),\end{aligned}$$

where we recall that $\Lambda(\phi) = -\sum_v \tau_v(\phi)\deg_K v$. The counterpart of (35) in [Gardeyn 2002] is

$$\deg_K \mathfrak{D}(K_{\phi,\mathfrak{L}}/K) \le r\deg_K a + \deg_K \Delta_\phi, \tag{36}$$

where Δ_ϕ is a divisor of K that is determined from the Newton polygons of the exponential functions associated to uniformizations of ϕ over $C_\wp$, where $\wp$ is a prime of K.

Although there is a larger dependence on ℓ in our different bounds when we take degrees with respect to K, what is required in the application is the degree with respect to $K_{\phi,\mathfrak{L}}$, which necessitates multiplying the degree with respect to K by $n' < \ell^{r^2}$. This means both bounds end up being comparable in their dependence on ℓ, as we later take the $\log_q$ of this degree with respect to $K_{\phi,\mathfrak{L}}$.

The quantity Δ_ϕ is more difficult to make explicit and compare, as we saw in Section 6, where its determination in the case of rank 2 and $K = F = \mathbb{F}_q(T)$ is recalled from [Chen and Lee 2013]. The method in [Chen and Lee 2013] yields the entire Newton polygon and uses Gekeler's theory of Drinfeld modular forms as well as Rosen's theory of formal Drinfeld modules. It may be possible to obtain weaker information using the more elementary approach of Chen and Lee [2012] in the infinite prime case, and to generalize Rosen's work to higher rank in the finite prime case, in such a way that Gardeyn's bounds can be made explicit.

As for the terms g_ϕ, it would seem that this also requires some knowledge relating to the successive minima of the lattices associated to the uniformization of ϕ over infinite primes.

Finally, two other places of difference are in our use of [Kumar Murty and Scherk 1994] for the Chebotarev density theorem instead of [Geyer and Jarden 1998], and in our analytic estimation methods, which differ slightly from [Gardeyn 2002; Serre 1981] because we have attempted to reduce the size of the constants in the different divisor bound, especially in front of the dominating terms.

Acknowledgements

We would like to thank A. Cojocaru and C. David for useful discussions pertaining to the subject of this paper.

References

[Chen and Lee 2012] I. Chen and Y. Lee, "Coefficients of exponential functions attached to Drinfeld modules of rank 2", *Manuscripta Math.* **139**:1-2 (2012), 123–136. MR 2959673 Zbl pre06084034

[Chen and Lee 2013] I. Chen and Y. Lee, "Newton polygons, successive minima, and different bounds for Drinfeld modules of rank 2", *Proc. Amer. Math. Soc.* **141**:1 (2013), 83–91. MR 2988712 Zbl 06146598

[Cojocaru and David 2008] A. C. Cojocaru and C. David, "Frobenius fields for Drinfeld modules of rank 2", *Compos. Math.* **144**:4 (2008), 827–848. MR 2009g:11071 Zbl 1158.11029

[David 2001] C. David, "Frobenius distributions of Drinfeld modules of any rank", *J. Number Theory* **90**:2 (2001), 329–340. MR 2002k:11084 Zbl 1012.11048

[Gardeyn 2001] F. Gardeyn, *t-motives and Galois representations*, Ph.D. thesis, Universiteit Gent, 2001, https://biblio.ugent.be/publication/522143. MR 2715545

[Gardeyn 2002] F. Gardeyn, "Une borne pour l'action de l'inertie sauvage sur la torsion d'un module de Drinfeld", *Arch. Math.* (*Basel*) **79**:4 (2002), 241–251. MR 2003i:11073 Zbl 1019.11015

[Geyer and Jarden 1998] W.-D. Geyer and M. Jarden, "Bounded realization of l-groups over global fields: the method of Scholz and Reichardt", *Nagoya Math. J.* **150** (1998), 13–62. MR 99d:12001 Zbl 0906.12002

[Goss 1992] D. Goss, "L-series of t-motives and Drinfeld modules", pp. 313–402 in *The arithmetic of function fields* (Columbus, OH, 1991), edited by D. Goss et al., Ohio State Univ. Math. Res. Inst. Publ. **2**, de Gruyter, Berlin, 1992. MR 94b:11048 Zbl 0806.11028

[Goss 1996] D. Goss, *Basic structures of function field arithmetic*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **35**, Springer, Berlin, 1996. MR 97i:11062 Zbl 0874.11004

[Kumar Murty and Scherk 1994] V. Kumar Murty and J. Scherk, "Effective versions of the Chebotarev density theorem for function fields", *C. R. Acad. Sci. Paris Sér. I Math.* **319**:6 (1994), 523–528. MR 95j:11104 Zbl 0822.11077

[Pink 1997] R. Pink, "The Mumford–Tate conjecture for Drinfeld-modules", *Publ. Res. Inst. Math. Sci.* **33**:3 (1997), 393–425. MR 98f:11062 Zbl 0895.11025

[Pink and Rütsche 2009a] R. Pink and E. Rütsche, "Adelic openness for Drinfeld modules in generic characteristic", *J. Number Theory* **129**:4 (2009), 882–907. MR 2010f:11092 Zbl 05542339

[Pink and Rütsche 2009b] R. Pink and E. Rütsche, "Image of the group ring of the Galois representation associated to Drinfeld modules", *J. Number Theory* **129**:4 (2009), 866–881. MR 2010f:11091 Zbl 05542338

[Rosen 2002] M. Rosen, *Number theory in function fields*, Graduate Texts in Mathematics **210**, Springer, New York, 2002. MR 2003d:11171 Zbl 1043.11079

[Serre 1972] J.-P. Serre, "Propriétés Galoisiennes des points d'ordre fini des courbes elliptiques", *Invent. Math.* **15**:4 (1972), 259–331. MR 52 #8126 Zbl 0235.14012

[Serre 1979] J.-P. Serre, *Local fields*, Graduate Texts in Mathematics **67**, Springer, New York, 1979. MR 82e:12016 Zbl 0423.12016

[Serre 1981] J.-P. Serre, "Quelques applications du théorème de densité de Chebotarev", *Inst. Hautes Études Sci. Publ. Math.* **54** (1981), 323–401 (123–201). MR 83k:12011 Zbl 0496.12011

[Taguchi 1992] Y. Taguchi, "Ramifications arising from Drinfeld modules", pp. 171–187 in *The arithmetic of function fields* (Columbus, OH, 1991), edited by D. Goss et al., Ohio State Univ. Math. Res. Inst. Publ. **2**, de Gruyter, Berlin, 1992. MR 94b:11049 Zbl 0801.11027

[Taguchi 1993] Y. Taguchi, "Semi-simplicity of the Galois representations attached to Drinfeld modules over fields of 'infinite characteristics'", *J. Number Theory* **44**:3 (1993), 292–314. MR 94k:11064 Zbl 0781.11024

[Taguchi 1996] Y. Taguchi, "On ϕ-modules", *J. Number Theory* **60**:1 (1996), 124–141. MR 97e:11067 Zbl 0862.11038

[Takahashi 1982] T. Takahashi, "Good reduction of elliptic modules", *J. Math. Soc. Japan* **34**:3 (1982), 475–487. MR 83h:12023 Zbl 0476.14010

[Tamagawa 1994] A. Tamagawa, "Generalization of Anderson's t-motives and Tate conjecture", *Sūrikaisekikenkyūsho Kōkyūroku* **884** (1994), 154–159. MR 96g:11061 Zbl 0900.11036

[Thakur 2004] D. S. Thakur, *Function field arithmetic*, World Scientific, River Edge, NJ, 2004. MR 2005h:11115 Zbl 1061.11001

Received December 20, 2011.

IMIN CHEN
DEPARTMENT OF MATHEMATICS
SIMON FRASER UNIVERSITY
BURNABY, BRITISH COLUMBIA V5A 1S6
CANADA

ichen@math.sfu.ca

YOONJIN LEE
DEPARTMENT OF MATHEMATICS
EWHA WOMANS UNIVERSITY
11-1 DAEHYUN-DONG, SEODAEMUN-GU
SEOUL 120-750
SOUTH KOREA

yoonjinl@ewha.ac.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 263, No. 1, 2013
dx.doi.org/10.2140/pjm.2013.263.117

TOPOLOGICAL PRESSURES FOR ϵ-STABLE AND STABLE SETS

XIANFENG MA AND ERCAI CHEN

In this paper, topological pressures of the preimages of ϵ-stable sets and certain closed subsets of stable sets in positive entropy systems are investigated. It is shown that the topological pressure of any topological system can be calculated in terms of the topological pressure of the preimages of ϵ-stable sets. For the constructed closed subset (W. Huang, Commun. Math. Phys. 279, 535–557 (2008)) of the stable set or the unstable set of any point in a measure-theoretic "rather big" set of a topological system with positive entropy, especially for the weakly mixing subset contained in the closure of the stable and unstable sets, it is proved that topological pressures of these subsets can be no less than the measure-theoretic pressure.

1. Introduction

Let (X, T) be a *topological dynamical system* (TDS) in the sense that X is a compact metric space with a compatible metric d and $T : X \to X$ is a homeomorphism. A TDS is said to be *noninvertible* if the map is surjective and continuous but not one-to-one. For $x \in X$ and $\epsilon > 0$, the ϵ-*stable* set of x under T is the set of points whose forward orbit ϵ-shadows that of x:

$$W^s_\epsilon(x, T) = \{y \in X : d(T^n x, T^n y) \le \epsilon \text{ for all } n \ge 0\}.$$

The preimages of these sets can be nontrivial and hence disperse at a nonzero exponent rate. the dispersal rate function $h_s(T, x, \epsilon)$ was introduced in [Fiebig et al. 2003]. The relationship between $h_s(T, x, \epsilon)$ and the topological entropy $h_{\text{top}}(T)$ was also investigated. It was proved that when X has finite covering dimension, for all $\epsilon > 0$,

$$\sup_{x \in X} h_s(T, x, \epsilon) = h_{\text{top}}(T).$$

The authors are supported by the National Natural Science Foundation of China (grant numbers 10971100 and 11271191). The first author is supported by the China Scholarship Council and the Fundamental Research Funds for the Central Universities. The second author is partially supported by the National Basic Research Program of China (973 Program) (grant number 2013CB834100).
MSC2010: 37A35.
Keywords: topological pressure, stable sets, ϵ-stable sets, weak mixing sets.

In [Huang 2008], the finite-dimensionality hypothesis turns out to be redundant. This equality is proved to be always true for any noninvertible TDS.

It is known that certain results concerning topological entropy can be generalized to topological pressure. For any $f \in C(X, \mathbb{R})$, consider the *topological pressure of the preimages of the ϵ-stable set of x*:

$$P(T, f, x, \epsilon) = \lim_{\delta \to 0} \limsup_{n \to +\infty} \frac{1}{n} \log P_n(T, f, \delta, T^{-n} W^s_\epsilon(x, T)),$$

where

$$P_n(T, f, \delta, T^{-n} W^s_\epsilon(x, T)) \\ = \sup \left\{ \sum_{x \in E} \exp f_n(x) : E \text{ is an } (n, \delta)\text{-separated subset of } T^{-n} W^s_\epsilon(x, T) \right\},$$

and $f_n(x) = \sum_{i=0}^{n-1} f \circ T^i(x)$. We show that the topological pressure of any noninvertible TDS with positive metric entropy can be calculated in terms of the topological pressure of the preimages of ϵ-stable sets. That is, for all $\epsilon > 0$,

$$\sup_{x \in X} P(T, f, x, \epsilon) = P(T, f),$$

where $P(T, f)$ is the standard notion of the topological pressure. For the null function f, this equality is the above one for the topological entropy.

For $x \in X$, the *stable set* $W^s(x, T)$ and the *unstable set* $W^u(x, T)$ of x are defined as

$$W^s(x, T) = \{y \in X : \lim_{n \to +\infty} d(T^n x, T^n y) = 0\},$$
$$W^u(x, T) = \{y \in X : \lim_{n \to +\infty} d(T^{-n} x, T^{-n} y) = 0\}.$$

For Anosov diffeomorphisms on a compact manifold, pairs belonging to the stable set are asymptotic under T and tend to diverge under T^{-1}. However, Blanchard et al. [2002] showed that in most case, this phenomenon does not happen in a TDS with positive metric entropy. N. Sumi [2003] investigated the stable and unstable sets of C^2 diffeomorphisms of C^∞ manifolds with positive metric entropy. He showed that the closure of the stable set $W^s(x, T)$ of "many points" is a perfect $*$-chaotic set and the closure of the unstable set $W^u(x, T)$ contains a perfect $*$-chaotic set. W. Huang [2008] got further information in the general noninvertible TDS with positive metric entropy. He proved that there exists a measure-theoretically "rather big" set such that the closure of the stable or unstable sets of points in the set contains a weakly mixing set. The Bowen entropies of these sets were also estimated there. It was proved that the lower bound is the usual metric entropy $h_\mu(T)$ for the ergodic invariant measure μ.

By introducing the topological pressure for the closed subset and using the excellent partition formed in Lemma 4 of [Blanchard et al. 2002], we show that, for the constructed closed subsets of stable and unstable sets in [Huang 2008], the topological pressure of these sets can also be estimated. More precisely, we prove that if μ is an ergodic invariant measure of a TDS (X, T) with $h_\mu(T) > 0$, then, for μ-a.e. $x \in X$, the closed subsets

$$A(x) \subseteq W^s(x, T), \quad B(x) \subseteq W^u(x, T)$$

and the weakly mixing subset

$$E(x) \subseteq \overline{W^s(x, T)} \cap \overline{W^u(x, T)}$$

constructed in [Huang 2008] have the following properties:

(a) $\lim_{n\to+\infty} \operatorname{diam} T^n A(x) = 0$ and $P(T^{-1}, f, A(x)) \geq P_\mu(T, f)$,

(b) $\lim_{n\to+\infty} \operatorname{diam} T^{-n} B(x) = 0$ and $P(T, f, B(x)) \geq P_\mu(T, f)$,

(c) $P(T, f, E(x)) \geq P_\mu(T, f)$ and $P(T^{-1}, f, E(x)) \geq P_\mu(T, f)$,

where $P_\mu(T, f)$ is the measure-theoretic pressure.

The paper is organized as follows. In Section 2, the topological pressure for the closed subset of a TDS is introduced. Some related notions and results about entropy are also listed. In Section 3, the topological pressure of the preimages of an ϵ-stable set is introduced. Using the tool formed in [Blanchard et al. 2002], we show that the topological pressure of any TDS can be calculated in terms of the topological pressure of the preimages of an ϵ-stable set. As a generalization of the entropy point, the notion of the pressure point is also introduced. In Section 4, results (a)–(c) above are proved. In Section 5, the results in sections 3 and 4 are stated and proved for the noninvertible TDS.

2. Preliminaries

Let (X, T) be a TDS and $\mathcal{B}_X$ be the σ-algebra of all Borel subsets of X. Recall that a *cover* of X is a finite family of Borel subsets of X whose union is X, and a *partition* of X is a cover of X whose elements are pairwise disjoint. We denote the set of covers, partitions, and open covers, of X by $\mathcal{C}_X$, $\mathcal{P}_X$, and $\mathcal{C}_X^o$. Given a partition α of X and $x \in X$, denote by $\alpha(x)$ the atom of α containing x. For two given covers $\mathcal{U}, \mathcal{V} \in \mathcal{C}_X$, $\mathcal{U}$ is said to be *finer* than $\mathcal{V}$ (denoted by $\mathcal{U} \succeq \mathcal{V}$) if each element of $\mathcal{U}$ is contained in some element of $\mathcal{V}$. Let

$$\mathcal{U} \vee \mathcal{V} = \{U \cap V : U \in \mathcal{U}, V \in \mathcal{V}\}.$$

Given integers M, N with $0 \leq M \leq N$ and $\mathcal{U} \in \mathcal{C}_X$, we set

$$\mathcal{U}_M^N = \bigvee_{n=M}^{N} T^{-n}\mathcal{U}.$$

Given $\mathcal{U} \in \mathcal{C}_X$ and $K \subset X$, put

$$N(\mathcal{U}, K) = \min\Big\{\text{the cardinality of } \mathcal{F} : \mathcal{F} \subset \mathcal{U}, \bigcup_{F\in\mathcal{F}} F \supset K\Big\}$$

and $H(\mathcal{U}, K) = \log N(\mathcal{U}, K)$. Then the topological entropy of $\mathcal{U}$ with respect to T for the compact subset K is

$$h_{\text{top}}(T, \mathcal{U}, K) = \lim_{n\to\infty} \frac{1}{n} H(\mathcal{U}_0^{n-1}, K) = \inf_{n\geq 1} \frac{1}{n} H(\mathcal{U}_0^{n-1}, K).$$

The topological entropy of T for the compact subset K is defined by $h_{\text{top}}(T, K) = \sup_{\mathcal{U}\in\mathcal{C}_X^o} h_{\text{top}}(T, \mathcal{U}, K)$; and the topological entropy of T is defined by $h_{\text{top}}(T) = \sup_K h_{\text{top}}(T, K)$.

Let (X, T) be a TDS, K a closed subset of X, $\mathcal{U} \in \mathcal{C}_X^o$, and $f \in C(X, \mathbb{R})$, where $C(X, \mathbb{R})$ is the Banach space of all continuous, real-valued functions on X endowed with the supremum norm. We set

$$P_n(T, f, \mathcal{U}, K) = \inf\Big\{\sum_{V\in\mathcal{V}} \sup_{x\in V\cap K} \exp f_n(x) : \mathcal{V} \in \mathcal{C}_X \text{ and } \mathcal{V} \succeq \mathcal{U}_0^{n-1}\Big\}, \tag{1}$$

where $f_n(x) = \sum_{j=0}^{n-1} f(T^j x)$. When $V \cap K = \varnothing$, we let $\sup_{x\in V\cap K} \exp f_n(x) = 0$. Then the above definition is well defined. It is clear that if f is the null function, $P_n(T, 0, \mathcal{U}, K) = N(\mathcal{U}_0^{n-1}, K)$.

For $\mathcal{V} \in \mathcal{C}_X$, we let α be the Borel partition generated by $\mathcal{V}$ and define

$$\mathcal{P}^*(\mathcal{V}) = \{\beta \in \mathcal{P}_X : \beta \succeq \mathcal{V} \text{ and each atom of } \beta \text{ is the union of some atoms of } \alpha\}.$$

Lemma 2.1 [Ma et al. 2010, Lemma 2.1]. *Let M be a compact subset of X and let $f \in C(X, \mathbb{R})$, $\mathcal{V} \in \mathcal{C}_X$. Then*

$$\inf_{\substack{\beta\in\mathcal{C}_X\\ \beta\succeq\mathcal{V}}} \sum_{B\in\beta} \sup_{x\in B\cap M} f(x) = \min\Big\{\sum_{B\in\beta} \sup_{x\in B\cap M} f(x) : \beta \in \mathcal{P}^*(\mathcal{V})\Big\}.$$

Let $\mathcal{K}(X)$ be the collection of all nonempty closed subsets of X. For any nonempty subset A of X and $\epsilon > 0$, let $N(A, \epsilon) = \{x \in X : \text{dist}(x, A) < \epsilon\}$, where $\text{dist}(x, A) = \inf\{d(x, y) : y \in A\}$. The *Hausdorff metric* H_d on the space $\mathcal{K}(X)$ induced by the metric d is defined as

$$H_d(A, B) = \inf\{\epsilon : A \subset N(B, \epsilon) \text{ and } B \subset N(A, \epsilon)\} \quad \text{for any } A, B \subset X.$$

Then $(\mathcal{K}(X), H_d)$ constitutes a compact metric space.

Lemma 2.2. *Let (X, T) be a TDS, $\mathcal{U} \in \mathcal{C}_X^o$, and $f \in C(X, \mathbb{R}^+)$. Then the function*

$$F : K \to \inf\Big\{\sum_{V\in\mathcal{V}} \sup_{x\in V\cap K} f(x) : \mathcal{V} \in \mathcal{C}_X \textit{ and } \mathcal{V} \succeq \mathcal{U}\Big\}$$

is measurable from $\mathcal{K}(X)$ to $\mathbb{R}^+$, where $\sup_{x\in V\cap K} f(x) = 0$ for $V \cap K = \varnothing$.

Proof. By Lemma 2.1 it suffices to prove that for each $B \in \beta$, where $\beta \in \mathcal{P}^*(\mathcal{U})$, the function $F_B : K \to \sup_{x \in B \cap K} f(x)$ is measurable.

For each $r \in \mathbb{R}$, let $\mathcal{E}_r = \{K : \sup_{x \in B \cap K} f(x) > r\}$. Let $U = f^{-1}(r, +\infty)$. Then U is an open subset of X. For $r \geq 0$, if $B \cap U = \emptyset$, $\mathcal{E}_r = \emptyset$. If $B \cap U \neq \emptyset$, $\mathcal{E}_r = \{K : K \cap (B \cap U) \neq \emptyset\}$. Let α be the Borel partition generated by the open cover $\mathcal{U} = \{U_i\}_{i=1}^s$. Then each $A \in \alpha$ has the form $\big(\bigcap_{i \in L} U_i\big) \cap \big(\bigcap_{j \in M} U_j^c\big)$, where $L, M \subset \{1, \dots, s\}$ and $L \cap M = \emptyset$. Note that, for each open subset W of X, the sets $\{K : K \cap (W \cap U) \neq \emptyset\}$ and $\{K : K \cap (W^c \cap U) \neq \emptyset\}$ — which equals $\{K : K \cap U \neq \emptyset\} \cap \big(\mathcal{K}(X) \backslash \{K : K \subset W\}\big)$ — are both measurable subsets of $\mathcal{K}(X)$. Then the set $\{K : K \cap (A \cap U) \neq \emptyset\}$ is measurable for each $A \in \alpha$. Since each atom B of β is the finite union of elements of α, it follows that $\mathcal{E}_r$ is a measurable subset of $\mathcal{K}(X)$. For $r < 0$, $\mathcal{E}_r = \mathcal{E}_0 \cup \{K : \sup_{x \in B \cap K} f(x) = 0\} = \mathcal{E}_0 \cup \{K : B \cap K = \emptyset\}$. Since $\{K : B \cap K = \emptyset\} = \mathcal{K}(X) \backslash \{K : B \cap K \neq \emptyset\}$ and $\{K : B \cap K \neq \emptyset\}$ is measurable, $\mathcal{E}_r$ is also measurable. Thus F_B is a measurable function. $\square$

Let $K \in \mathcal{K}(X)$, $\mathcal{U} \in \mathcal{C}_X^o$, and $f \in C(X, \mathbb{R})$. We define $P(T, f, \mathcal{U}, K) = \limsup_{n \to \infty} (1/n) \log P_n(T, f, \mathcal{U}, K)$.

Let (X, T) be a TDS. Denote by $\mathcal{M}(X)$ the set of all Borel probability measures on X, by $\mathcal{M}(X, T)$ the set of T-invariant measures, and by $\mathcal{M}^e(X, T)$ the set of ergodic measures. Then $\mathcal{M}^e(X, T) \subset \mathcal{M}(X, T) \subset \mathcal{M}(X)$, and $\mathcal{M}(X)$, $\mathcal{M}(X, T)$ are convex, compact metric spaces endowed with the weak*-topology.

Since the map f is a homeomorphism, it induces in a natural way a homeomorphism $\widehat{T} : \mathcal{K}(X) \to \mathcal{K}(X)$ by $\widehat{T}(A) = T(A)$ for each $A \in \mathcal{K}(X)$. Then $(\mathcal{K}(X), \widehat{T})$ constitutes a TDS induced by (X, T).

For each $\hat{\mu} \in \mathcal{M}(\mathcal{K}(X), \widehat{T})$, the following lemma shows that the limit superior in the above definition can be obtained by the limit for $\hat{\mu}$-a.e. $K \in \mathcal{K}(X)$.

Lemma 2.3. *Let (X, T) be a TDS, $\mathcal{U} \in \mathcal{C}_X^o$, $f \in C(X, \mathbb{R})$, and $\hat{\mu} \in \mathcal{M}(\mathcal{K}(X), \widehat{T})$. Then, for $\hat{\mu}$-a.e. $K \in \mathcal{K}(X)$, $P(T, f, \mathcal{U}, K) = \lim_{n \to +\infty} (1/n) \log P_n(T, f, \mathcal{U}, K)$ exists.*

Proof. For any $n, m \in \mathbb{N}$, $\mathcal{V}_1 \succeq \mathcal{U}_0^{n-1}$, $\mathcal{V}_2 \succeq \mathcal{U}_0^{m-1}$, we have $\mathcal{V}_1 \vee T^{-n}\mathcal{V}_2 \succeq \mathcal{U}_0^{n+m-1}$. It follows that

$$
\begin{aligned}
P_{n+m}(T, f, \mathcal{U}, K) &\leq \sum_{V_1 \in \mathcal{V}_1} \sum_{V_1 \in \mathcal{V}_2} \sup_{x \in V_1 \cap T^{-n} V_2 \cap K} \exp f_{n+m}(x) \\
&= \sum_{V_1 \in \mathcal{V}_1} \sum_{V_2 \in \mathcal{V}_2} \sup_{x \in V_1 \cap T^{-n} V_2 \cap K} \exp(f_n(x) + f_m(T^n x)) \\
&\leq \sum_{V_1 \in \mathcal{V}_1} \sum_{V_2 \in \mathcal{V}_2} \big(\sup_{x \in V_1 K} \exp f_n(x) \cdot \sup_{z \in V_2 \cap T^n K} \exp f_m(z)\big) \\
&= \Bigg(\sum_{V_1 \in \mathcal{V}_1} \sup_{x \in V_1 \cap K} \exp f_n(x) \Bigg) \Bigg(\sum_{V_2 \in \mathcal{V}_2} \sup_{z \in V_2 \cap T^n K} \exp f_m(z) \Bigg).
\end{aligned}
$$

Since $\mathcal{V}_i$, $i = 1, 2$ is arbitrary,

$$P_{n+m}(T, f, \mathcal{U}, K) \le P_n(T, f, \mathcal{U}, K) \cdot P_m(T, f, \mathcal{U}, T^n K).$$

By the definition of $\widehat{T}$ and Lemma 2.2, we have that

$$\log P_n(T, f, \mathcal{U}, K) : \mathcal{K}(X) \to \mathbb{R} \cup \{-\infty\}$$

is a subadditive sequence of measurable functions. Then, by Kingman's subadditive ergodic theorem (see [Walters 1982]), we complete the proof. □

When $K = X$, $P(T, f, \mathcal{U}, X) = P(T, f, \mathcal{U})$, which is the local topological pressure defined by Huang and Yi [2007], clearly, $P(T, 0, \mathcal{U}, K) = h_{\text{top}}(T, \mathcal{U}, K)$.

Given a partition $\alpha \in \mathcal{P}(X)$, $\mu \in \mathcal{M}(X)$ and a sub-σ-algebra $\mathcal{C} \subseteq \mathcal{B}_\mu$, let

$$H_\mu(\alpha) = \sum_{A\in\alpha} -\mu(A) \log \mu(A),$$

$$H_\mu(\alpha \mid \mathcal{C}) = \sum_{A\in\alpha} \int_X -\mathbb{E}(1_A \mid \mathcal{C}) \log \mathbb{E}(1_A \mid \mathcal{C})\, d\mu,$$

where $\mathbb{E}(1_A \mid \mathcal{C})$ is the expectation of 1_A with respect to $\mathcal{C}$. One standard fact states that $H_\mu(\alpha \mid \mathcal{C})$ increases with respect to α and decreases with respect to $\mathcal{C}$. The measure-theoretic entropy of μ is defined as

$$h_\mu(T) = \sup_{\alpha\in\mathcal{P}_X} h_\mu(T, \alpha),$$

where

$$h_\mu(T, \alpha) = \lim_{n\to+\infty} \frac{1}{n} H_\mu(\alpha_0^{n-1}) = \inf_{n\ge1} H_\mu(\alpha_0^{n-1}).$$

For each $f \in C(X, \mathbb{R})$, the *measure-theoretic pressure of* μ is defined as

$$P_\mu(T, f) = h_\mu(T) + \int_X f\, d\mu.$$

For a given $\mathcal{U} \in \mathcal{C}_X$, set

$$H_\mu(\mathcal{U}) = \inf_{\beta\in\mathcal{P}_X,\, \beta\succeq\mathcal{U}} H_\mu(\beta) \quad \text{and} \quad H_\mu(\mathcal{U} \mid \mathcal{C}) = \inf_{\beta\in\mathcal{P}_X,\, \beta\succeq\mathcal{U}} H_\mu(\beta \mid \mathcal{C}).$$

When $\mu \in \mathcal{M}(X, T)$ and $\mathcal{C}$ is T-invariant (that is, $T^{-1}\mathcal{C} = \mathcal{C}$), $H_\mu(\mathcal{U}_0^{n-1} \mid \mathcal{C})$ is a nonnegative subadditive sequence for a given $\mathcal{U} \in \mathcal{U}$. Let

$$h_\mu(T, \mathcal{U} \mid \mathcal{C}) = \lim_{n\to+\infty} \frac{1}{n} H_\mu(\mathcal{U}_0^{n-1} \mid \mathcal{C}) = \inf_{n\ge1} H_\mu(\mathcal{U}_0^{n-1} \mid \mathcal{C}).$$

For $\mathcal{C} = \{\varnothing, X\}$ (mod μ), we write $H_\mu(\mathcal{U} \mid \mathcal{C})$ and $h_\mu(T, \mathcal{U} \mid \mathcal{C})$ as $H_\mu(\mathcal{U})$ and $h_\mu(T, \mathcal{U})$, respectively. Romagnoli [2003] proved that

$$h_\mu(T) = \sup_{\mathcal{U}\in\mathcal{C}_X^o} h_\mu(T, \mathcal{U}).$$

It is well known that, for $\beta \in \mathcal{P}_X$, $h_\mu(T, \beta) = h_\mu(T, \beta \mid P_\mu(T)) \le H_\mu(\beta \mid P_\mu(T))$, where $P_\mu(T)$ is the Pinsker σ-algebra of $(X, \mathcal{B}_\mu, \mu, T)$.

Lemma 2.4 [Huang 2008, Lemma 2.1]. *Let (X, T) be a TDS, $\mu \in \mathcal{M}(X, T)$, and $\mathcal{U} \in \mathcal{C}_X$. Then*

$$h_\mu(T, \mathcal{U}) = h_\mu(T, \mathcal{U} \mid P_\mu(T)).$$

For $\mathcal{U} \in \mathcal{C}_X^o$, $\mu \in \mathcal{M}(X, T)$ and $f \in C(X, \mathbb{R})$, we define the *measure-theoretic pressure for T with respect to $\mathcal{U}$* as

$$P_\mu(T, f, \mathcal{U}) = h_\mu(T, \mathcal{U}) + \int_X f \, d\mu.$$

Obviously,

$$P_\mu(T, f) = h_\mu(T) + \int_X f \, d\mu = \sup_{\mathcal{U} \in \mathcal{C}_X^o} h_\mu(T, \mathcal{U}) + \int_X f \, d\mu = \sup_{\mathcal{U} \in \mathcal{C}_X^o} P_\mu(T, f, \mathcal{U}).$$

Let (X, T) be a TDS, $\mu \in \mathcal{M}(X, T)$, and $\mathcal{B}_\mu$ be the completion of $\mathcal{B}_X$ under μ. Then $(X, \mathcal{B}_\mu, \mu, T)$ is a Lebesgue system. If $\{\alpha_i\}_{i \in I}$ is a countable family of finite partitions of X, the partition $\alpha = \bigvee_{i \in I} \alpha_i$ is called a *measurable partition*. The sets $A \in \mathcal{B}_\mu$, which are unions of atoms of α, form a sub-σ-algebra of $\mathcal{B}_\mu$ by $\hat{\alpha}$ or α if there is no ambiguity. Every sub-σ-algebra of $\mathcal{B}_\mu$ coincides with a σ-algebra constructed in this way (mod μ).

Given a measurable partition α, put $\alpha^- = \bigvee_{n=1}^{\infty} T^{-n}\alpha$ and $\alpha^T = \bigvee_{n=-\infty}^{+\infty} T^{-n}\alpha$. Define in the same way $\mathcal{F}^-$ and $\mathcal{F}^T$ if $\mathcal{F}$ is a sub-σ-algebra of $\mathcal{B}_\mu$. It is clear that for a measurable partition α of X, we have

$$\widehat{\alpha^-} = (\hat{\alpha})^- \quad \text{and} \quad \widehat{\alpha^T} = (\hat{\alpha})^T \pmod{\mu}.$$

Let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{B}_\mu$ and α be the measurable partition of X with $\alpha^- = \mathcal{F}$ (mod μ). μ can be disintegrated over $\mathcal{F}$ as $\mu = \int_X \mu_x \, d\mu(x)$, where $\mu_x \in \mathcal{M}(X)$ and $\mu_x(\alpha(x)) = 1$ for μ-a.e. $x \in X$. The disintegration is characterized by two properties:

(a) For every $f \in L^1(X, \mathcal{B}_X, \mu)$, $f \in L^1(X, \mathcal{B}_X, \mu_x)$ for μ-a.e. $x \in X$, and the map $x \mapsto \int_X f(y) \, d\mu_x(y)$ is in $L^1(X, \mathcal{F}, \mu)$.

(b) For every $f \in L^1(X, \mathcal{B}_X, \mu)$, $\mathbb{E}_\mu(f \mid \mathcal{F})(x) = \int_X f \, d\mu_x$ for μ a.e. $x \in X$.

Then, for any $f \in L^1(X, \mathcal{B}_X, \mu)$,

$$\int_X \left(\int_X f \, d\mu_x \right) d\mu(x) = \int_X f \, d\mu.$$

Lemma 2.5 [Huang 2008, Lemma 2.2]. *Let (X, T) be a TDS, $\mu \in \mathcal{M}(X, T)$, and $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{B}_\mu$. If $\mu = \int_X \mu_x \, d\mu(x)$ is the disintegration of μ over $\mathcal{F}$,*

(a) *for $\mathcal{V} \in \mathcal{C}_X$, $H_\mu(\mathcal{V} \mid \mathcal{F}) = \int_X H_{\mu_x}(\mathcal{V}) \, d\mu(x)$,*

(b) *for $\mathcal{U}, \mathcal{V} \in \mathcal{C}_X$, $H_\mu(\mathcal{U} \vee \mathcal{V} \mid \mathcal{F}) \le H_\mu(\mathcal{U} \mid \mathcal{F}) + H_\mu(\mathcal{V} \mid \mathcal{F})$.*

Let K be a nonempty closed subset of X. For $\epsilon > 0$, a subset of X is called an (n, ϵ)-*spanning set* of K, if for any $x \in K$ there exists $y \in F$ with $d_n(x, y) \le \epsilon$, where $d_n(x, y) = \max_{i=0}^{n-1} d(T^i x, T^i y)$; a subset E of K is called an (n, ϵ)-*separated set* of K, if $x, y \in E$, $x \ne y$ implies $d_n(x, y) > \epsilon$. Let $r_n(d, T, \epsilon, K)$ denote the smallest cardinality of any (n, ϵ)-spanning subset for K and $s_n(d, T, \epsilon, K)$ denote the largest cardinality of any (n, ϵ)-separated subset of K.

For each $\epsilon > 0$ and $f \in C(X, \mathbb{R})$, we define

$$P_n(T, f, \epsilon, K) = \sup\Big\{\sum_{x\in E} \exp f_n(x) : E \text{ is an } (n, \epsilon)\text{-separated subset of } K\Big\}.$$

The topological pressure of T for the closed subset K is defined as

$$P(T, f, K) = \lim_{\epsilon\to 0} \limsup_{n\to+\infty} \frac{1}{n} \log P_n(T, f, \epsilon, K).$$

Clearly, for $f \equiv 0$, we can write $P_n(T, 0, \epsilon, K) = s_n(d, T, \epsilon, K)$. It follows that $P(T, f, K) = h(T, K)$, where $h(T, K)$ is the Bowen entropy for the closed subset K defined in [Walters 1982]; see also [Huang 2008]. When $K = X$, $P(T, f, X) = P(T, f)$, where $P(T, f)$ is the standard notion of topological pressure defined in [Walters 1982]. Moreover, it is not hard to verify that $P(T, f, K) = \sup_{\mathcal{U}\in\mathscr{C}_X^o} P(T, f, \mathcal{U}, K)$.

3. ϵ-stable sets

Let (X, T) be a TDS with a compatible metric d. Given $\epsilon > 0$, the ϵ-*stable set* of x under T is the set of points whose forward orbit ϵ-shadows that of x:

$$W^s_\epsilon(x, T) = \{y \in X : d(T^n x, T^n y) \le \epsilon \text{ for all } n = 0, 1, \ldots\}.$$

Since the preimages of these sets can be nontrivial, we can consider the following function. For each $x \in X$, $f \in C(X, \mathbb{R})$, and $\epsilon > 0$, let

$$P_s(T, f, x, \epsilon) := \lim_{\delta\to 0} \limsup_{n\to+\infty} \frac{1}{n} \log P_n(T, f, \delta, T^{-n} W^s_\epsilon(x, T)).$$

$P_s(T, f, x, \epsilon)$ is called the *topological pressure of the preimages of the ϵ-stable sets of x*. For $f \equiv 0$, $P_s(T, 0, x, \epsilon) = h_s(T, x, \epsilon)$, where the latter is the dispersal rate function defined in [Fiebig et al. 2003]. It was proved in [Huang 2008] that $\sup_{x\in X} h_s(T, x, \epsilon) = h_{\text{top}}(T)$ for all $\epsilon > 0$. In the present section, we show that this is also true for the functions $P_s(T, f, x, \epsilon)$ and $P(T, f)$. By proving that, for any $\mu \in \mathcal{M}^e(X, T)$ with positive entropy, $\lim_{\epsilon\to 0} P_s(T, f, x, \epsilon) \ge P_\mu(T, f)$ for μ-a.e. $x \in X$, we can obtain the result. We need the following lemmas.

Lemma 3.1. *Let (X, T) be a TDS, $f \in C(X, \mathbb{R})$, and $\{K_n\}$ be a sequence of nonempty closed subsets of X. Then*

$$\lim_{\delta\to 0}\limsup_{n\to+\infty}\frac{1}{n}\log P_n(T, f, \delta, K_n) = \sup_{\mathcal{U}\in\mathcal{C}_X^0}\limsup_{n\to+\infty}\frac{1}{n}\log P_n(T, f, \mathcal{U}, K_n).$$

Proof. For a fixed $\delta > 0$, choose $\mathcal{V} \in \mathcal{C}_X^o$ with $\operatorname{diam}\mathcal{V} < \delta$. For $n \in \mathbb{N}$ let A be an (n, δ)-separated set of K_n. Since $B \cap K_n$ contains at most one element of A for each B of $\bigvee_{i=0}^{n-1} T^{-i}\mathcal{V}$, for every $\mathcal{W} \in \mathcal{C}_X$ with $\mathcal{W} \succeq \mathcal{V}_0^{n-1}$, each element of $\mathcal{W}$ also contains at most one element of A. We get $\sum_{x\in A}\exp f_n(x) \le P_n(T, f, \mathcal{V}, K_n)$. That is $P_n(T, f, \delta, K_n) \le P_n(T, f, \mathcal{V}, K_n)$. Then

$$\begin{aligned}\limsup_{n\to+\infty}\frac{1}{n}\log P_n(T, f, \delta, K_n) &\le \limsup_{n\to+\infty}\frac{1}{n}\log P_n(T, f, \mathcal{V}, K_n)\\ &\le \sup_{\mathcal{U}\in\mathcal{C}_X^o}\limsup_{n\to+\infty}\frac{1}{n}\log P_n(T, f, \mathcal{U}, K_n).\end{aligned}$$

Letting $\delta \to 0$, we get

$$\lim_{\delta\to 0}\limsup_{n\to+\infty}\frac{1}{n}\log P_n(T, f, \delta, K_n) \le \sup_{\mathcal{U}\in\mathcal{C}_X^0}\limsup_{n\to+\infty}\frac{1}{n}\log P_n(T, f, \mathcal{U}, K_n).$$

In the following, we show the converse inequality. For any fixed $\mathcal{U} \in \mathcal{C}_X^o$, let δ be the Lebesgue number of $\mathcal{U}$. For $n \in \mathbb{N}$, let E be an $(n, \delta/2)$-separated set of K_n with the largest cardinality. Then E is also an $(n, \delta/2)$-spanning set of K_n. From the definition of spanning sets, we know that

$$\bigcup_{x\in E}\bigcap_{i=0}^{n-1} T^{-i}\overline{B_{\delta/2}(T^i x)} \supset K_n, \quad \text{where } \overline{B_{\delta/2}(T^i x)} = \Big\{y \in X : d(T^i x, y) \le \frac{\delta}{2}\Big\}.$$

Now, for each $x \in E$ and $0 \le i \le n-1$, $\overline{B_{\delta/2}(T^i x)}$ is contained in some element of $\mathcal{U}$ since δ is the Lebesgue number of the open cover $\mathcal{U}$. Hence, for each $x \in E$, the intersection $\bigcap_{i=0}^{n-1} T^{-i}\overline{B_{\delta/2}(T^i x)}$ is contained in some element of $\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}$. Let $\mathcal{W} = \big\{\bigcap_{i=0}^{n-1} T^{-i}\overline{B_{\delta/2}(T^i x)} : x \in E\big\}$. Then $\mathcal{W} \in \mathcal{C}_X$ and $\mathcal{W} \succeq \mathcal{U}_0^{n-1}$. Let

$$Q_n(T, f, \mathcal{U}, K_n) = \inf\Big\{\sum_{V\in\mathcal{V}}\inf_{x\in V\cap K_n}\exp f_n(x) : \mathcal{V} \in \mathcal{C}_X \text{ and } \mathcal{V} \succeq \mathcal{U}_0^{n-1}\Big\}.$$

Then

$$Q_n(T, f, \mathcal{U}, K_n) \le \sum_{x\in E} f_n(x) \le P_n\Big(T, f, \frac{\delta}{2}, K_n\Big).$$

Let $\tau_{\mathcal{U}} = \sup\{| f(x) - f(y) | : d(x, y) \le \operatorname{diam}\mathcal{U}\}$. Then

$$\exp(-n\tau_{\mathcal{U}})P_n(T, f, \mathcal{U}, K_n) \le Q_n(T, f, \mathcal{U}, K_n).$$

So

$$-\tau_{\mathcal{U}} + \limsup_{n\to+\infty} \frac{1}{n} \log P_n(T, f, \mathcal{U}, K_n) \le \limsup_{n\to+\infty} \frac{1}{n} \log P_n\Big(T, f, \frac{\delta}{2}, K_n\Big)$$
$$\le \lim_{\delta\to 0} \limsup_{n\to+\infty} \frac{1}{n} \log P_n\Big(T, f, \frac{\delta}{2}, K_n\Big).$$

Since $\mathcal{U}$ is arbitrary, we get

$$\sup_{\mathcal{U}\in\mathscr{C}_X^0} \limsup_{n\to+\infty} \frac{1}{n} \log P_n(T, f, \mathcal{U}, K_n) \le \lim_{\delta\to 0} \limsup_{n\to+\infty} \frac{1}{n} \log P_n(T, f, \delta, K_n). \qquad \square$$

An immediate consequence of Lemma 3.1 is the following.

Lemma 3.2. *Let (X, T) be a TDS and $f \in C(X, \mathbb{R})$. Then, for each $x \in X$ and $\epsilon > 0$,*

$$P_s(T, f, x, \epsilon) = \sup_{\mathcal{U}\in\mathscr{C}_X^o} \limsup_{n\to+\infty} \frac{1}{n} \log P_n(T, f, \mathcal{U}, T^{-n} W_\epsilon^s(x, T)).$$

Lemma 3.3 [Walters 1982, Lemma 9.9]. *Let $a_1, \ldots, a_k$ be given real numbers. If $p_i \ge 0$, $i = 1, \ldots, k$, and $\sum_{i=1}^k p_i = 1$,*

$$\sum_{i=1}^{k} p_i(a_i - \log p_i) \le \log \sum_{i=1}^{k} e^{a_i},$$

and equality holds if and only if

$$p_i = \frac{e^{a_i}}{\sum_{i=1}^k e^{a_i}} \quad \text{for all } i = 1, \ldots, k.$$

Let (X, T) be a TDS, $\mu \in \mathcal{M}(X, T)$, and $\mathcal{B}_\mu$ be the completion of $\mathcal{B}_X$ under μ. The *Pinsker σ-algebra $P_\mu(T)$* is defined as the smallest sub-σ-algebra of $\mathcal{B}_\mu$ containing $\{\xi \in \mathcal{P}_X : h_\mu(T, \xi) = 0\}$. It is well known that $P_\mu(T) = P_\mu(T^{-1})$ and $P_\mu(T)$ is T-invariant, that is, $T^{-1}(P_\mu(T)) = P_\mu(T)$.

Lemma 3.4 [Huang 2008, Lemma 3.5]. *Let (X, T) be a TDS, $\mu \in \mathcal{M}(X, T)$, and $\delta > 0$. Then there exist $\{W_i\}_{i=1}^\infty \subset \mathcal{P}_X$ and $0 = k_1 < k_2 < \cdots$ such that*

(a) diam $W_1 < \delta$ *and* $\lim_{i\to+\infty}$ diam $W_i = 0$,

(b) $\lim_{k\to+\infty} H_\mu(P_k \mid \mathcal{P}^-) = h_\mu(T)$, *where* $P_k = \bigvee_{i=1}^k T^{-k_i} W_i$ *and* $\mathcal{P} = \bigvee_{k=1}^\infty P_k$,

(c) $\bigcap_{n=0}^\infty \widehat{T^{-n}\mathcal{P}^-} = P_\mu(T)$.

Lemma 3.5. *Let (X, T) be a TDS, $\mathcal{U} \in \mathscr{C}_X^o$, $f \in C(X, \mathbb{R})$, and $K \in \mathcal{K}(X)$. Then, for each $n \in \mathbb{N}$,*

$$P_n(T, f, \mathcal{U}, T^{-n} K) = P_n(T, f \circ T^{-n}, T^n \mathcal{U}, K).$$

Proof. For each $\mathcal{V} \in \mathcal{C}_X$ and $\mathcal{V} \succeq \bigvee_{i=1}^{n} T^i \mathcal{U}$, obviously, $T^{-n}\mathcal{V} \in \mathcal{C}_x$ and $T^{-n}\mathcal{V} \succeq \bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}$.

Since for each $V \in \mathcal{V}$,

$$\sup_{x \in T^{-n}V \cap T^{-n}K} \exp f_n(x) = \sup_{x \in V \cap K} \exp f_n(T^{-n}x),$$

it is easy to see that $P_n(T, f, \mathcal{U}, T^{-n}K) \leq P_n(T, f \circ T^{-n}, T^n\mathcal{U}, K)$. From the homeomorphism of T, the inverse inequality holds. Then $P_n(T, f, \mathcal{U}, T^{-n}K) = P_n(T, f \circ T^{-n}, T^n\mathcal{U}, K)$. □

Recall that a set-valued map F from X to $\mathcal{K}(X)$ is said to be *measurable* if $\{x \in X : F(x) \cap A \neq \varnothing\} \in \mathcal{B}_X$ for every Borel (open or closed) subset A of X.

Lemma 3.6. *Let $G : X \to \mathcal{K}(X)$ be a measurable set-valued map, $f \in C(X, \mathbb{R}^+)$, and $\mathcal{U} \in \mathcal{C}_X^o$. Then*

$$F : x \to \inf\Big\{\sum_{V \in \mathcal{V}} \sup_{y \in V \cap G(x)} f(y) : \mathcal{V} \in \mathcal{C}_X \text{ and } \mathcal{V} \succeq \mathcal{U}\Big\}$$

is Borel-measurable, where $\sup_{y \in V \cap G(x)} f(y) = 0$ for $V \cap G(x) = \varnothing$.

Proof. By Lemma 2.1, for each $x \in X$, we have

$$\inf\Big\{\sum_{V \in \mathcal{V}} \sup_{y \in V \cap G(x)} f(y) : \mathcal{V} \in \mathcal{C}_X, \mathcal{V} \succeq \mathcal{U}\Big\} = \min\Big\{\sum_{V \in \mathcal{V}} \sup_{y \in V \cap G(x)} f(y) : \mathcal{V} \in \mathcal{P}^*(\mathcal{U})\Big\}.$$

It is sufficient to prove that, for each $V \in \mathcal{V}$, where $\mathcal{V} \in \mathcal{P}^*(\mathcal{U})$, the function $H_V : x \to \sup_{y \in V \cap G(x)} f(y)$ is Borel-measurable.

For each $r \in \mathbb{R}$, let $E_r = \{x : \sup_{y \in V \cap G(x)} f(y) > r\}$. Note that $U = f^{-1}(r, +\infty)$ is an open subset of X. For $r \geq 0$, if $V \cap U = \varnothing$, $E_r = \varnothing$. If $V \cap U \neq \varnothing$, then $E_r = \{x : V \cap G(x) \cap U \neq \varnothing\}$. Since $V \cap U \in \mathcal{B}(X)$, by the set-valued measurability of G, it is clear that E_r is a Borel subset of X. For $r < 0$, $E_r = E_0 \cup F$, where $F = \{x : \sup_{y \in V \cap G(x)} f(y) = 0\}$. Since

$$F = \{x : V \cap G(x) = \varnothing\} = X \backslash \{x : V \cap G(x) \neq \varnothing\}$$

is Borel-measurable, E_r is also a Borel subset of X; thus H_V is Borel-measurable. □

The next theorem clearly implies the main result of this paper.

Theorem 3.7. *Let (X, T) be a TDS, $f \in C(X, \mathbb{R})$, and $\mu \in \mathcal{M}^e(X, T)$ with $h_\mu(T) > 0$. Then, for μ-a.e. $x \in X$, $\lim_{\epsilon \to 0} P_s(T, f, x, \epsilon) \geq P_\mu(T, f)$.*

Proof. It suffices to prove that, for a given $\epsilon > 0$, $P_s(T, f, x, \epsilon) \geq P_\mu(T, f)$ for μ-a.e. $x \in X$.

Fix $\epsilon > 0$. Since T is a homeomorphism on X, there exists $\delta \in (0, \epsilon)$ such that $d(T^{-1}x, T^{-1}y) < \epsilon$ when $d(x, y) < \delta$. By Lemma 3.4, there exists $\{P_i\}_{i=1}^{\infty} \subset \mathcal{P}_X$ satisfying diam $P_1 \leq \delta$, $\bigcap_{n=0}^{\infty} \widehat{T^{-n}\mathcal{P}^-} = P_\mu(T)$, and $H_\mu(P_k \mid \mathcal{P}^-) \to h_\mu(T)$ when

$k \to +\infty$, where $\mathcal{P} = \bigvee_{i=1}^{\infty} P_i$. Since diam $P_1 \leq \delta$, it is clear that $\mathcal{P}^-(x) \subseteq W_\epsilon^s(x, T)$ for each $x \in X$.

Let $\mu = \int_X \mu_x \, d\mu(x)$ be the disintegration of μ over $\mathcal{P}^-$. Then

$$\operatorname{supp}(\mu_x) \subseteq \overline{\mathcal{P}^-(x)} \subseteq W_\epsilon^s(x, T) \quad \text{for } \mu\text{-a.e. } x \in X.$$

Let $k \in \mathbb{N}$. By inequality (3.3) in [Huang 2008], we know that there exists $\mathcal{U}_k \in \mathcal{C}_X^o$ such that

$$\limsup_{n \to +\infty} \frac{1}{n} H_\mu\biggl(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}_k \mid T^{-n}\mathcal{P}^-\biggr) \geq H_\mu(P_k \mid \mathcal{P}^-) - \frac{1}{k}. \tag{2}$$

For $n \in \mathbb{N}$, let $F_n(x) = (1/n) \log P_n(T, f \circ T^{-n}, T^n\mathcal{U}_k, W_\epsilon^s(x, T))$. Noting that the map $x \to W_\epsilon^s(x, T)$ is upper semicontinuous, it follows from Lemma 3.6 that F_n is a Borel-measurable function. Let $F(x) = \limsup_{n \to +\infty} F_n(x)$ for $x \in X$. Then F is also Borel-measurable. Since $TW_\epsilon^s(x, T) \subseteq W_\epsilon^s(Tx, T)$ for each $x \in X$, we have

$$\begin{aligned}
&P_n(T, f \circ T^{-n}, T^n\mathcal{U}_k, W_\epsilon^s(x, T)) \\
&\qquad \leq \inf\biggl\{\sum_{V \in \mathcal{V}} \sup_{y \in V \cap TW_\epsilon^s(x,T)} \exp f_n \circ T^{-(n+1)}(y) : \mathcal{V} \in \mathcal{C}_X \text{ and } \mathcal{V} \succeq \bigvee_{i=2}^{n+1} T^i\mathcal{U}_k\biggr\} \\
&\qquad \leq \inf\biggl\{\sum_{V \in \mathcal{V}} \sup_{y \in V \cap W_\epsilon^s(Tx,T)} \exp f_n \circ T^{-(n+1)}(y) : \mathcal{V} \in \mathcal{C}_X \text{ and } \mathcal{V} \succeq \bigvee_{i=1}^{n+1} T^i\mathcal{U}_k\biggr\} \\
&\qquad = P_{n+1}(T, f \circ T^{-(n+1)}, T^{n+1}\mathcal{U}_k, W_\epsilon^s(Tx, T)).
\end{aligned}$$

Then

$$\begin{aligned}
F(x) &= \limsup_{n \to +\infty} \frac{1}{n} \log P_n(T, f \circ T^{-n}, T^n\mathcal{U}_k, W_\epsilon^s(x, T)) \\
&\leq \limsup_{n \to +\infty} \frac{n+1}{n} \cdot \frac{1}{n+1} \log P_{n+1}(T, f \circ T^{-(n+1)}, T^{n+1}\mathcal{U}_k, W_\epsilon^s(Tx, T)) \\
&= F(Tx).
\end{aligned}$$

Thus, $F(x) \leq F(Tx)$ for each $x \in X$. Since $\mu \in \mathcal{M}(X, T)$, $\int_X F(Tx) \, d\mu(x) = \int_X F(x) \, d\mu(x)$, we have, $F(Tx) = F(x)$ for μ-a.e. $x \in X$. Moreover, $F(x) \equiv a_k$ for μ-a.e. $x \in X$ as μ is ergodic, where $a_k \geq 0$ is a constant.

From Lemma 2.1, there exists a finite partition

$$\beta \in \mathcal{P}^*\biggl(\bigvee_{i=1}^{n} T^i\mathcal{U}_k\biggr)$$

such that

$$P_n(T, f \circ T^{-n}, T^n\mathcal{U}_k, W_\epsilon^s(x, T)) = \sum_{B \in \beta} \sup_{x \in B \cap W_\epsilon^s(x,T)} \exp f_n \circ T^{-n}(x).$$

It follows from Lemma 3.3 that

$$\begin{aligned}
&\log P_n(T, f\circ T^{-n}, T^n\mathcal{U}_k, W^s_\epsilon(x,T))\\
&\quad = \log \sum_{B\in\beta} \sup_{x\in B\cap W^s_\epsilon(x,T)} \exp f_n\circ T^{-n}(x)\\
&\quad \geq \sum_{B\in\beta} \mu_x(B\cap W^s_\epsilon(x,T))\big(\sup_{x\in B\cap W^s_\epsilon(x,T)} \exp f_n\circ T^{-n}(x) - \log \mu_x(B\cap W^s_\epsilon(x,T))\big)\\
&\quad = H_{\mu_x}(\beta) + \sum_{B\in\beta} \sup_{x\in B\cap W^s_\epsilon(x,T)} f_n\circ T^{-n}(x)\cdot \mu_x(B) \quad (\mathrm{supp}(\mu_x)\subseteq W^s_\epsilon(x,T)\\
&\qquad\qquad \text{for } \mu\text{-a.e. } x\in X)\\
&\quad \geq H_{\mu_x}\Big(\bigvee_{i=1}^{n} T^i\mathcal{U}_k\Big) + \int_X f_n\circ T^{-n}\,d\mu_x
\end{aligned}$$

Then

$$\begin{aligned}
a_k &= \int_X F(x)\,d\mu = \int_X \limsup_{n\to+\infty} F_n(x)\,d\mu \geq \limsup_{n\to+\infty}\int_X F_n(x)\,d\mu\\
&\geq \limsup_{n\to+\infty}\int_X \frac{1}{n}\Big(H_{\mu_x}\Big(\bigvee_{i=1}^{n} T^i\mathcal{U}_k\Big) + \int f_n\circ T^{-n}\,d\mu_x\Big)\,d\mu(x)\\
&= \limsup_{n\to+\infty}\Big(\int_X \frac{1}{n}H_{\mu_x}\Big(\bigvee_{i=1}^{n} T^i\mathcal{U}_k\Big)\,d\mu(x) + \frac{1}{n}\int_X\int f_n\circ T^{-n}\,d\mu_x d\mu(x)\Big)\\
&= \limsup_{n\to+\infty}\Big(\int_X \frac{1}{n}H_{\mu_x}\Big(\bigvee_{i=1}^{n} T^i\mathcal{U}_k\Big)\,d\mu(x) + \frac{1}{n}\int_X f_n\circ T^{-n}\,d\mu(x)\Big)\\
&= \limsup_{n\to+\infty}\int_X \frac{1}{n}H_{\mu_x}\Big(\bigvee_{i=1}^{n} T^i\mathcal{U}_k\Big)\,d\mu(x) + \int_X f\,d\mu(x) \quad (\text{since } \mu\in\mathcal{M}(X,T))\\
&= \limsup_{n\to+\infty}\frac{1}{n}H_{\mu}\Big(\bigvee_{i=1}^{n} T^i\mathcal{U}_k \mid \mathcal{P}^-\Big) + \int_X f\,d\mu(x) \quad (\text{by Lemma 2.5(a)})\\
&= \limsup_{n\to+\infty}\frac{1}{n}H_{\mu}\Big(\bigvee_{i=1}^{n-1} T^{-i}\mathcal{U}_k \mid T^{-n}\mathcal{P}^-\Big) + \int_X f\,d\mu(x)\\
&\geq H_\mu(P_k\mid\mathcal{P}^-) - \frac{1}{k} + \int_X f\,d\mu(x) \quad (\text{by inequality (2)}).
\end{aligned}$$

Since $P_s(T, f, x, \epsilon)\geq F(x)$ for each $x\in X$, we have

$$\begin{aligned}
P_s(T,f,x,\epsilon) &\geq \lim_{k\to+\infty}\Big(H_\mu(P_k\mid\mathcal{P}^-) - \frac{1}{k} + \int_X f\,d\mu(x)\Big)\\
&= h_\mu(T) + \int_X f\,d\mu(x) = P_\mu(T,f)
\end{aligned}$$

for μ-a.e. $x\in X$. □

We introduce the ϵ-pressure point and pressure point for a TDS. Let (X, T) be a TDS, $f \in C(X, \mathbb{R})$. For $\epsilon > 0$, we call $x \in X$ an *ϵ-pressure point* for T if $P_s(T, f, x, \epsilon) = P(T, f)$, and we call it a *pressure point* if $\lim_{\epsilon\to o} P_s(T, f, x, \epsilon) = P(T, f)$. The function $P_s(T, f, x, \epsilon)$ is decreasing in ϵ. It follows that every pressure point is also an ϵ-pressure point for each $\epsilon > 0$. Note that, while the notion of an ϵ-pressure point depends on the choice of the metric, that of pressure point does not. Denote by $\mathcal{P}(T, f)$ the set of all pressure points of (X, T) for $f \in C(X, \mathbb{R})$. For $f \equiv 0$, the ϵ-pressure point and pressure point are the ϵ-entropy point and entropy point, respectively, which are introduced in [Fiebig et al. 2003]. Moreover, $\mathcal{P}(T, 0) = \mathcal{E}(T)$, where $\mathcal{E}$ is the set of all entropy points of (X, T).

Remark 3.8. Let (X, T) be a TDS, $f \in C(X, \mathbb{R})$. If there exists $\mu \in \mathcal{M}^e(X, T)$ such that $P(T, f) = P_\mu(T, f)$, $\mathcal{P}(T, f) \neq \varnothing$.

4. Stable sets

The main results of the present section are Theorems 4.1 and 4.5. Recall that, for a TDS (X, T) and $x \in X$,

$$W^s(x, T) = \{y \in X : \lim_{n\to+\infty} d(T^n x, T^n y) = 0\},$$
$$W^u(x, T) = \{y \in X : \lim_{n\to+\infty} d(T^{-n} x, T^{-n} y) = 0\}.$$

$W^s(x, T)$ is called the *stable set* of x for T, and $W^u(x, T)$ is called the *unstable set* of x for T. Obviously, $W^s(x, T) = W^u(x, T^{-1})$ and $W^u(x, T) = W^s(x, T^{-1})$.

Theorem 4.1. *Let (X, T) be a TDS, $f \in C(X, \mathbb{R})$, and $\mu \in \mathcal{M}^e(X, T)$ with $h_\mu(T) > 0$. Then, for μ-a.e. $x \in X$,*

(a) *there exists a closed subset $A(x) \subseteq W^s(x, T)$ such that*

$$\lim_{n\to+\infty} \operatorname{diam} T^n A(x) = 0 \quad \textit{and} \quad P(T^{-1}, f, A(x)) \geq P_\mu(T, f);$$

(b) *there exists a closed subset $B(x) \subseteq W^u(x, T)$ such that*

$$\lim_{n\to+\infty} \operatorname{diam} T^{-n} B(x) = 0 \quad \textit{and} \quad P(T, f, B(x)) \geq P_\mu(T, f).$$

Proof. Since $\mu \in \mathcal{M}^e(X, T)$, $P_\mu(T^{-1}, f) = P_\mu(T, f)$, and $W^s(x, T^{-1}) = W^u(x, T)$, (a) implies (b). It remains to prove (a).

By Lemma 3.4, there exist $\{W_i\}_{i=1}^\infty \subset \mathcal{P}_X$ and $0 = k_1 < k_2 < \cdots$ satisfying

(a) $\operatorname{diam} W_1 < \delta$ and $\lim_{i\to+\infty} \operatorname{diam} W_i = 0$,

(b) $\lim_{k\to+\infty} H_\mu(P_k \mid \mathcal{P}^-) = h_\mu(T)$, where $P_k = \bigvee_{i=1}^k T^{-k_i} W_i$ and $\mathcal{P} = \bigvee_{k=1}^\infty P_k$,

(c) $\bigcap_{n=0}^\infty \widehat{T^{-n}\mathcal{P}^-} = P_\mu(T)$.

Let $Q_i = \bigvee_{j=1}^{i} T^{-j}(P_1 \vee P_2 \vee \cdots \vee P_i)$ for $i \in \mathbb{N}$. Then $Q_i \in \mathcal{P}_X$, $Q_1 \preceq Q_2 \preceq \cdots$, and $\bigvee_{i=1}^{\infty} Q_i = \mathcal{P}^-$.

For $x \in X$, let $A(x) = \bigcap_{i=1}^{\infty} \overline{Q_i(x)}$. Then $A(x)$ is a closed set and $A(x) \supseteq \overline{\mathcal{P}^-(x)}$. The set $A(x)$ also has the properties $\lim_{n\to+\infty} \operatorname{diam} T^n A(x) = 0$ and $A(x) \subseteq W^s(x, T)$ (see the proof of [Huang 2008, Theorem 4.2] for details).

Moreover, the set-valued map $A : x \to A(x)$ is measurable. In fact, for each open set U of X,

$$\left\{x : \bigcap_{n=1}^{\infty} \overline{Q_i(x)} \subseteq U\right\} = \bigcup_{n\geq 1} \bigcap_{k\geq n} \bigcap \{A \in Q_k : \bar{A} \subseteq U\}$$

is a Borel set of X. Then, for each closed set V of X, $\{x : \overline{Q_i(x)} \subseteq X \backslash V\}$ is a Borel set. It follows that $\{x : \overline{Q_i(x)} \cap V \neq \varnothing\}$ is Borel and then $A : x \to A(x)$ is set-valued measurable.

Let $\mu = \int_X \mu_x \, d\mu(x)$ be the disintegration of μ over $\mathcal{P}^-$. Then

$$\operatorname{supp}(\mu_x) \subseteq \overline{\mathcal{P}^-(x)} \subseteq A(x) \quad \text{for } \mu\text{-a.e. } x \in X. \tag{3}$$

We now prove that, for μ-a.e. $x \in X$, $P(T^{-1}, f, A(x)) \geq P_\mu(T, f)$. Since $\lim_{k\to+\infty} H_\mu(P_k \mid \mathcal{P}^-) = h_\mu(T)$, it is sufficient to prove that, for each $k \in \mathbb{N}$, $P(T^{-1}, f, A(x)) \geq H_\mu(P_k \mid \mathcal{P}^-) - 1/k + \int_X f \, d\mu(x)$ for μ-a.e. $x \in X$.

For a given $k \in \mathbb{N}$, there exists $\mathcal{U}_k \in \mathcal{C}_X^o$ such that

$$\limsup_{n\to+\infty} \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}_k \mid T^{-n}\mathcal{P}^-\right) \geq H_\mu(P_k \mid \mathcal{P}^-) - \frac{1}{k} \quad \text{for each } n \in \mathbb{N} \tag{4}$$

(see [Huang 2008] for details).

Let $F_n(x) = (1/n) \log P_n(T^{-1}, f, \mathcal{U}_k, A(x))$, where

$$P_n(T^{-1}, f, \mathcal{U}_k, A(x)) = \inf\left\{\sum_{V\in\mathcal{V}} \sup_{y\in V\cap A(x)} \exp f_n \circ T^{-(n-1)}(y) : \mathcal{V} \in \mathcal{C}_X \text{ and } \mathcal{V} \succeq \bigvee_{i=0}^{n-1} T^i\mathcal{U}_k\right\},$$

and $f_n(z) = \sum_{i=0}^{n-1} f(T^i z)$. By Lemma 3.6, F_n is a Borel-measurable function. Let $F(x) = \limsup_{n\to+\infty} F_n(x)$ for each $x \in X$. Then F is also a Borel-measurable function on X.

For each $\mathcal{V} \succeq \bigvee_{i=0}^{n-1} T^i\mathcal{U}_k$, $T^{-1}\mathcal{V} \succeq \bigvee_{i=0}^{n-1} T^i\mathcal{U}_k$. Since $T(A(x)) \subseteq A(T(x))$ (see the proof of [Huang 2008, Theorem 4.2]), for each $V \in \mathcal{V}$,

$$\begin{aligned}\sup_{y\in T^{-1}V\cap A(x)} \sum_{i=0}^{n-1} f(T^{-i}y) &\leq \sup_{y\in T^{-1}(V\cap A(Tx))} \sum_{i=0}^{n-1} f(T^{-i}y) \\ &= \sup_{y\in V\cap A(Tx)} \sum_{i=1}^{n} f(T^{-i}y) \leq \sup_{y\in V\cap A(Tx)} \sum_{i=0}^{n} f(T^{-i}y),\end{aligned}$$

it is not hard to see that $P_n(T^{-1}, f, \mathcal{U}_k, A(x)) \le P_{n+1}(T^{-1}, f, \mathcal{U}_k, A(Tx))$. Hence

$$\begin{aligned}F(x) &= \limsup_{n\to+\infty} \frac{1}{n} \log P_n(T^{-1}, f, \mathcal{U}_k, A(x))\\ &\le \limsup_{n\to+\infty} \frac{n+1}{n} \cdot \frac{1}{n+1} \log P_n(T^{-1}, f, \mathcal{U}_k, A(Tx)) = F(Tx).\end{aligned}$$

Thus $F(x) \le F(Tx)$ for each $x \in X$. Since $\mu \in \mathcal{M}(X, T)$, we have

$$\int_X (f(Tx) - f(x))\, d\mu(x) = 0.$$

Then $F(Tx) = F(x)$ for μ-a.e. $x \in X$. From the ergodicity of μ, there exists a constant $a_k \ge 0$ such that $F(x) \equiv a_k$ for μ-a.e. $x \in X$.

By Lemma 2.1, there exists a partition $\beta \in \mathcal{P}^*(\bigvee_{i=0}^{n-1} T^i \mathcal{U}_k)$ such that, for μ-a.e. $x \in X$,

$$\begin{aligned}&\log P_n(T^{-1}, f, \mathcal{U}_k, A(x))\\ &\quad = \log \sum_{B\in\beta} \sup_{y\in B\cap A(x)} \exp \sum_{i=0}^{n-1} f(T^{-i}y)\\ &\quad \ge \sum_{B\in\beta} \mu_x(B) \Big(\sup_{y\in B\cap A(x)} \exp \sum_{i=0}^{n-1} f(T^{-i}y) - \log \mu_x(B) \Big) \quad \text{(by (3) and Lemma 3.3)}\\ &\quad = H_{\mu_x}(\beta) + \sum_{B\in\beta} \sup_{y\in B\cap A(x)} \exp \sum_{i=0}^{n-1} f(T^{-i}y) \cdot \mu_x(B)\\ &\quad \ge H_{\mu_x}\Big(\bigvee_{i=0}^{n-1} T^i \mathcal{U}_k \Big) + \int_X f_n \circ T^{-(n-1)}\, d\mu_X.\end{aligned}$$

Then

$$\begin{aligned}a_k &= \int_X F(x)\, d\mu = \int_X \limsup_{n\to+\infty} F_n(x)\, d\mu(x) \ge \limsup_{n\to+\infty} \int_X F_n(x)\, d\mu(x)\\ &\ge \limsup_{n\to+\infty} \frac{1}{n} \int_X \Big(H_{\mu_x}\Big(\bigvee_{i=0}^{n-1} T^i \mathcal{U}_k \Big) + \int_X f_n \circ T^{-(n-1)}\, d\mu_x \Big) d\mu(x)\\ &= \limsup_{n\to+\infty} \frac{1}{n} \Big(\int_X H_{\mu_x}\Big(\bigvee_{i=0}^{n-1} T^i \mathcal{U}_k \Big) d\mu(x) + \int_X f_n \circ T^{-(n-1)}\, d\mu(x) \Big)\\ &= \limsup_{n\to+\infty} \frac{1}{n} \int_X H_{\mu_x}\Big(\bigvee_{i=0}^{n-1} T^i \mathcal{U}_k \Big) d\mu(x) + \int_X f\, d\mu(x) \quad \text{(since } \mu \in \mathcal{M}(X, T))\end{aligned}$$

$$
\begin{aligned}
&= \limsup_{n\to+\infty} \frac{1}{n} H_\mu\Big(\bigvee_{i=0}^{n-1} T^i \mathcal{U}_k \mid \mathcal{P}^-\Big) + \int_X f\,d\mu(x) \quad \text{(by Lemma 2.5(a))}\\
&= \limsup_{n\to+\infty} \frac{1}{n} H_\mu\Big(\bigvee_{i=0}^{n-1} T^i \mathcal{U}_k \mid T^{-(n-1)}\mathcal{P}^-\Big) + \int_X f\,d\mu(x)\\
&\geq H_\mu(P_k \mid \mathcal{P}^-) - \frac{1}{k} + \int_X f\,d\mu(x) \quad \text{(by (4))}.
\end{aligned}
$$

Therefore, for μ-a.e. $x \in X$,

$$
P(T^{-1}, f, A(x)) \geq P(T^{-1}, f, \mathcal{U}_k, A(x)) = F(x) \geq H_\mu(P_k \mid \mathcal{P}^-) - \frac{1}{k} + \int_X f\,d\mu(x)
$$

for each $k \in \mathbb{N}$.

Then

$$
\begin{aligned}
P(T^{-1}, f, A(x)) &\geq \lim_{n\to+\infty} \Big(H_\mu(P_k \mid \mathcal{P}^-) - \frac{1}{k}\Big) + \int_X f\,d\mu(x)\\
&= H_\mu(T) + \int_X f\,d\mu(x) = P_\mu(T, f). \qquad \square
\end{aligned}
$$

This completes the proof of Theorem 4.1.

A direct consequence of Theorem 4.1 is the following.

Corollary 4.2. *Let (X, T) be a TDS, $f \in C(X, \mathbb{R})$. If there exists $\mu \in \mathcal{M}^e(X, T)$ with $P_\mu(T, f) = P(T, f)$, there exists $x \in X$, a closed subset $A(x) \subseteq W^s(x, T)$, and a closed subset $B(x) \subseteq W^u(x, T)$ such that*

(a) $\lim_{n\to+\infty} \operatorname{diam} T^n A(x) = 0$ *and* $P(T^{-1}, f, A(x)) = P(T, f)$;

(b) $\lim_{n\to+\infty} \operatorname{diam} T^{-n} B(x) = 0$ *and* $P(T, f, B(x)) = P(T, f)$.

A TDS (X, T) is transitive if, for each pair of nonempty open subsets U and V of X, there exists $n \geq 0$ such that $U \cap T^{-n} V \neq \varnothing$; and it is weakly mixing if $(X \times X, T \times T)$ is transitive. These notions describe the global properties of the whole TDS. Blanchard and Huang [2008] give a new criterion to picture “a certain amount of weakly mixing” in some consistent sense. The notion of a weakly mixing set was introduced as follows.

If X, Y are topological spaces, denote by $\mathscr{C}(X, Y)$ the set of all continuous maps from X to Y.

Definition 4.3. Let (X, T) be a TDS and $A \in 2^X$. The set A is said to be weakly mixing for T if there exists $B \subset A$ satisfying

(a) B is the union of countably many Cantor sets;

(b) the closure of B equals A;

(c) for any $C \in B$ and $g \in \mathscr{C}(C, A)$, there exists an increasing sequence of natural numbers $\{n_i\} \subset \mathbb{N}$ such that $\lim_{i\to+\infty} T^{n_i} x = g(x)$ for any $x \in C$.

Denote by $WM_s(X, T)$ the family of weakly mixing subsets of (X, T). The system (X, T) itself is called partially mixing when it contains a weakly mixing set. The whole space X is a weakly mixing set if and only if TDS (X, T) is weakly mixing [Xiong and Yang 1991]. The following result (See [Blanchard and Huang 2008, Proposition 4.2]) gives an equivalent characterization of the weakly mixing set in another way.

Proposition 4.4. *Let (X, T) be a TDS and A be a nonsingleton closed subset of X. Then A is a weakly mixing subset of X if and only if, for any $k \in \mathbb{N}$ and any choice of nonempty open subsets $V_1, \dots, V_k$ of A and nonempty open subsets $U_1, \dots, U_k$ of X with $A \cap U_i \neq \varnothing$, $i = 1, 2, \dots, k$, there exists $m \in \mathbb{N}$ such that $T^m V_i \cap U_i \neq \varnothing$ for each $1 \leq i \leq k$.*

Now we prove the following theorem. Part (a) of Theorem 4.5 was already proved in [Huang 2008]. For completeness, we state it in the theorem.

Theorem 4.5. *Let (X, T) be a TDS and $\mu \in \mathcal{M}^e(X, T)$ with $h_\mu(T) > 0$. Then, for μ-a.e. $x \in X$, there exists a closed subset*

$$E(x) \subseteq \overline{W^s(x, T)} \cap \overline{W^u(x, T)}$$

such that

(a) $E(x) \in WM_s(X, T) \cap WM_s(X, T^{-1})$, *i.e., $E(x)$ is weakly mixing for T, T^{-1};*

(b) $P(T, f, E(x)) \geq P_\mu(T, f)$ *and* $P(T^{-1}, f, E(x)) \geq P_\mu(T, f)$.

Proof. Let $\mathcal{B}_\mu$ be the completion of $\mathcal{B}_X$ under μ. Then $(X, \mathcal{B}_\mu, \mu, T)$ is a Lebesgue system. Let $P_\mu(T)$ be the Pinsker σ-algebra of $(X, \mathcal{B}_\mu, \mu, T)$. Let $\mu = \int_X \mu_x d\mu(x)$ be the disintegration of μ over $P_\mu(T)$. Then, for μ-a.e. $x \in X$,

$$\operatorname{supp}(\mu_x) \subseteq \overline{W^s(x, T)} \cap \overline{W^u(x, T)}$$

and

$$\operatorname{supp}(\mu_x) \in WM_s(X, T) \cap WM_s(X, T^{-1})$$

(see [Huang 2008, Theorem 4.6] for details).

We now prove that, for μ-a.e. $x \in X$,

$$P(T, f, \operatorname{supp}(\mu_x)) \geq P_\mu(T, f) \quad \text{and} \quad P(T^{-1}, f, \operatorname{supp}(\mu_x)) \geq P_\mu(T, f).$$

By the symmetry of T and T^{-1}, $P_\mu(T, f) = P_\mu(T^{-1}, f)$. It remains to prove that, for μ-a.e. $x \in X$, $P(T, f, \operatorname{supp}(\mu_x)) \geq P_\mu(T, f)$. Since $P_\mu(T)$ is T-invariant, $T\mu_x = \mu_{Tx}$ for μ-a.e. $x \in X$. Therefore, there exists a T-invariant measurable set $X_0 \subset X$ with $\mu(X_0) = 1$ and $T\mu_x = \mu_{Tx}$ for $x \in X_0$.

For each $\mathcal{U} \in \mathcal{C}_X^o$, $x \in X_0$, and $n \in \mathbb{N}$, by Lemma 2.1, there exists a $\beta \in \mathcal{P}^*(\mathcal{U}_0^{n-1})$ such that

$$
\begin{aligned}
(5)\quad & \log P_n(T, f, \mathcal{U}, \mathrm{supp}(\mu_x)) \\
&= \log\inf\Big\{\sum_{V\in\mathcal{V}} \sup_{y\in V\cap \mathrm{supp}(\mu_x)} \exp f_n(x) : \mathcal{V} \in \mathcal{C}_X \text{ and } \mathcal{V} \succeq \mathcal{U}_0^{n-1}\Big\} \\
&= \log \sum_{B\in\beta} \sup_{y\in B\cap \mathrm{supp}(\mu_x)} \exp f_n(x) \\
&\geq \sum_{B\in\beta} \mu_x(B)\Big(\sup_{y\in B\cap \mathrm{supp}(\mu_x)} f_n(x) - \log \mu_x(B)\Big) \quad \text{(by Lemma 3.3)} \\
&= H_{\mu_x}(\beta) + \sum_{B\in\beta} \mu_x(B) \sup_{y\in B\cap \mathrm{supp}(\mu_x)} f_n(x) \\
&\geq H_{\mu_x}(\mathcal{U}_0^{n-1}) + \int_X f_n \, d\mu_x
\end{aligned}
$$

Fix $\mathcal{U} \in \mathcal{C}_X^o$ and $n \in \mathbb{N}$. Denote $F_n(x) = H_{\mu_x}(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}) + \int_X f_n \, d\mu_x$ for each $x \in X_0$. Then

$$
\begin{aligned}
F_{n+m}(x) &= H_{\mu_x}\Big(\bigvee_{i=0}^{n+m-1} T^{-i}\mathcal{U}\Big) + \int_X f_{n+m}\, d\mu_x \\
&\leq H_{\mu_x}\Big(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}\Big) + H_{\mu_x}\Big(T^{-n}\bigvee_{i=0}^{m-1} T^{-i}\mathcal{U}\Big) + \int_X f_n\, d\mu_x + \int_X f_m \circ T^n \, d\mu_x \\
&\leq F_n(x) + H_{T^n\mu_x}\Big(\bigvee_{i=0}^{m-1} T^{-i}\mathcal{U}\Big) + \int_X f_m \circ T^n \, d\mu_x \\
&= F_n(x) + H_{T^n\mu_x}\Big(\bigvee_{i=0}^{m-1} T^{-i}\mathcal{U}\Big) + \int_X f_m dT^n\mu_x \\
&= F_n(x) + H_{\mu_{T^n x}}\Big(\bigvee_{i=0}^{m-1} T^{-i}\mathcal{U}\Big) + \int_X f_m\, d\mu_{T^n x} \\
&= F_n(x) + F_m(T^n x),
\end{aligned}
$$

that is, $\{F_n\}$ is subadditive. Since the map $x \to \mu_x(A)$ for each $A \in \mathcal{B}$ is measurable on X_0, it follows that $F_n(x)$ is measurable on X_0. By Kingman's subadditive ergodic theorem, $\lim_{n\to\infty}(1/n)F_n(x) \equiv a_{\mathcal{U}}$ for μ-a.e. $x \in X$, where $a_{\mathcal{U}}$ is a constant. Then, by (5),

$$P(T, f, \mathcal{U}, \mathrm{supp}(\mu_x)) \geq a_{\mathcal{U}}$$

for each $\mathcal{U} \in \mathcal{C}_X^0$ and μ-a.e. $x \in X$. Therefore

$$\begin{aligned}
a_{\mathcal{U}} &= \int_X \lim_{n\to\infty} \frac{1}{n} F_n(x)\, d\mu = \lim_{n\to\infty} \frac{1}{n} \int_X F_n(x)\, d\mu \\
&= \lim_{n\to\infty} \frac{1}{n} \int_X \left(H_{\mu_x}(\mathcal{U}_0^{n-1}) + \int_X f_n \, d\mu_x \right) d\mu(x) \\
&= \lim_{n\to\infty} \frac{1}{n} H_\mu(\mathcal{U}_0^{n-1} \mid P_\mu(T)) + \int_X f\, d\mu \\
&= h_\mu(T, \mathcal{U} \mid P_\mu(T)) + \int_X f\, d\mu \\
&= P_\mu(T, f, \mathcal{U}) \qquad \text{(by Lemma 2.4).}
\end{aligned}$$

It follows that

$$P(T, f, \mathcal{U}, \operatorname{supp}(\mu_x)) \geq P_\mu(T, f, \mathcal{U})$$

for each $\mathcal{U} \in \mathcal{C}_X^o$ and μ-a.e. $x \in X$.

Choose a sequence of open covers $\{\mathcal{U}_m\}_{m=1}^\infty$ with $\lim \operatorname{diam}\{\mathcal{U}_m\} = 0$. Then

$$\begin{aligned}
\lim_{n\to\infty} P_\mu(T, f, \mathcal{U}_m) &= \lim_{n\to\infty} \left(h_\mu(T, \mathcal{U}_m) + \int_X f\, d\mu \right) \\
&= h_\mu(T) + \int_X f\, d\mu = P_\mu(T, f).
\end{aligned}$$

Since for each $m \in \mathbb{N}$ and μ-a.e. $x \in X$, $P(T, f, \mathcal{U}_m, \operatorname{supp}(\mu_x)) \geq P_\mu(T, f, \mathcal{U}_m)$, we have

$$P(T, f, \operatorname{supp}(\mu_x)) = \sup_{m\in\mathbb{N}} P(T, f, \mathcal{U}_m, \operatorname{supp}(\mu_x)) \geq \sup_{m\geq 1} P_\mu(T, f, \mathcal{U}_m) = P_\mu(T, f)$$

for each μ-a.e. $x \in X$. □

It is not hard to see that the following corollary holds.

Corollary 4.6. *Let (X, T) be a TDS and $f \in C(X, \mathbb{R})$. Then*

(a) $\sup_{x\in X} P(T, f, \overline{W^s(x, T)} \cap \overline{W^u(x, T)}) = P(T, f)$;

(b) *if there exists $\mu \in \mathcal{M}^e(X, T)$ with $P_\mu(T, f) = P(T, f)$, then, for μ-a.e. $x \in X$, there exists a closed subsets $E(x) \subseteq \overline{W^s(x, T)} \cap \overline{W^u(x, T)}$ such that*

(i) $E(x) \in WM_s(X, T) \cap WM_s(X, T^{-1})$,

(ii) $P(T, f, E(x)) = P(T^{-1}, f, E(x)) = P(T, f)$.

5. Noninvertible case

In this section, we generalize the results in Sections 3 and 4 to the noninvertible case. Let (X, T) be a noninvertible TDS, that is, X is a compact metric space, and $T : X \to X$ is a surjective continuous map but not one-to-one.

Set $\widetilde{X} = \{(x_1, x_2, \ldots) : T(x_{i+1}) = x_i, x_i \in X, i \in \mathbb{N}\}$. It is clear that $\widetilde{X}$ is a subspace of the product space $\Pi_{i=1}^{\infty} X$ with the metric d_T defined by

$$d_T((x_1, x_2, \ldots), (y_1, y_2, \ldots)) = \sum_{i=1}^{\infty} \frac{d(x_i, y_i)}{2^i}.$$

Let $\widetilde{T} : \widetilde{X} \to \widetilde{X}$ be the shift homeomorphism, that is,

$$\widetilde{T}(x_1, x_2, \ldots) = (T(x_1), x_1, x_2, \ldots).$$

We refer to the TDS $(\widetilde{X}, \widetilde{T})$ as the inverse limit of (X, T). Let $\pi_i : \widetilde{X} \to X$ be the natural projection onto the i-th coordinate. Then $\pi_i : (\widetilde{X}, \widetilde{T}) \to (X, T)$ is a factor map.

Lemma 5.1. *Let (X, T) be a noninvertible TDS, $f \in C(X, \mathbb{R})$. Then, for each $\mathcal{U} \in \mathcal{C}_X^o$ and $K \in \mathcal{K}(X)$,*

$$P_{n+m}(T, f, \mathcal{U}, K) \le P_m(T, f, \mathcal{U}, K) \cdot P_n(T, f \circ T^m, T^{-m}\mathcal{U}, K)$$

for each $n, m \in \mathbb{N}$.

Proof. Since for each $\mathcal{V}_1 \succeq \mathcal{U}_0^{m-1}$ and $\mathcal{V}_2 \succeq \mathcal{U}_0^{n-1}$ we have $\mathcal{V}_1 \vee T^{-m}\mathcal{V}_2 \succeq \mathcal{U}_0^{n+m-1}$, it follows that

$$\begin{aligned}
P_{n+m}(T, f, \mathcal{U}, K) &\le \sum_{V_1 \in \mathcal{V}_1} \sum_{V_2 \in \mathcal{V}_2} \sup_{x \in V_1 \cap T^{-m} V_2 \cap K} \exp f_{n+m}(x) \\
&= \sum_{V_1 \in \mathcal{V}_1} \sum_{V_2 \in \mathcal{V}_2} \sup_{x \in V_1 \cap T^{-m} V_2 \cap K} \exp(f_m(x) + f_n(T^m x)) \\
&\le \sum_{V_1 \in \mathcal{V}_1} \sum_{V_2 \in \mathcal{V}_2} \sup_{x \in V_1 \cap K} \exp f_m(x) \cdot \sup_{x \in T^{-m} V_2 \cap K} \exp f_n(T^m x) \\
&= \sum_{V_1 \in \mathcal{V}_1} \sup_{x \in V_1 \cap K} \exp f_m(x) \cdot \sum_{V_2 \in \mathcal{V}_2} \sup_{x \in T^{-m} V_2 \cap K} \exp(f \circ T^m)_n(x).
\end{aligned}$$

By the arbitrariness of $\mathcal{V}_1$ and $\mathcal{V}_2$, we have

$$P_{n+m}(T, f, \mathcal{U}, K) \le P_m(T, f, \mathcal{U}, K) \cdot P_n(T, f \circ T^m, T^{-m}\mathcal{U}, K). \qquad \square$$

Lemma 5.2. *Let (X, T) be a noninvertible TDS, $f \in C(X, \mathbb{R})$. Then, for each $\mathcal{U} \in \mathscr{C}_X^o$ and $K \in \mathscr{K}(X)$,*

$$P_n(T, f \circ T^m, T^{-m}\mathcal{U}, T^{-m}K) = P_n(T, f, \mathcal{U}, K) \quad \textit{for each } n, m \in \mathbb{N}.$$

Proof. Fix $n, m \in \mathbb{N}$. For each $\mathcal{V} \succeq (T^{-m}\mathcal{U})_0^{n-1}$,

$$\begin{aligned}\sum_{V \in \mathcal{V}} \sup_{x \in V \cap T^{-m}K} \exp(f \circ T^m)_n(x) &= \sum_{V \in \mathcal{V}} \sup_{x \in V \cap T^{-m}K} \exp f_n(T^m x)\\ &= \sum_{V \in \mathcal{V}} \sup_{x \in T^m V \cap K} \exp f_n(x).\end{aligned}$$

Since $T^m\mathcal{V} \succeq \mathcal{U}_0^{n-1}$,

$$P_n(T, f \circ T^m, T^{-m}\mathcal{U}, T^{-m}K) \le P_n(T, f, \mathcal{U}, K).$$

Conversely, for each $\mathcal{V} \succeq \mathcal{U}_0^{n-1}$, $T^{-m}\mathcal{V} \succeq (T^{-m}\mathcal{U})_0^{n-1}$ and

$$\begin{aligned}\sum_{V \in \mathcal{V}} \sup_{x \in V \cap K} \exp f_n(x) &= \sum_{V \in \mathcal{V}} \sup_{x \in T^{-m}(V \cap K)} \exp f_n(T^m x)\\ &= \sum_{V \in \mathcal{V}} \sup_{x \in T^{-m}V \cap T^{-m}K} \exp(f \circ T^m)_n(x).\end{aligned}$$

Then

$$P_n(T, f \circ T^m, T^{-m}\mathcal{U}, T^{-m}K) \ge P_n(T, f, \mathcal{U}, K),$$

which completes the proof. □

Lemma 5.3. *Let $(\widetilde{X}, \widetilde{T})$ be the inverse limit of a noninvertible TDS (X, T). Let $f \in C(X, \mathbb{R})$ and let $\pi_1 : \widetilde{X} \to X$ be the projection to the first coordinate. Then, for any sequence of nonempty closed subsets K_n of $\widetilde{X}$,*

$$\lim_{\delta \to 0} \limsup_{n \to +\infty} \frac{1}{n} \log P_n(\widetilde{T}, f \circ \pi_1, \delta, K_n) = \lim_{\delta \to 0} \limsup_{n \to +\infty} \frac{1}{n} \log P_n(T, f, \delta, \pi_1(K_n)).$$

Proof. Let $\mathcal{U} \in \mathscr{C}_X^o$. For each $\mathcal{V} \in \mathscr{C}_X$ with $\mathcal{V} \succeq \mathcal{U}_0^{n-1}$ and $x \in V \cap \pi_1(K_n)$, obviously, $\pi_1^{-1}\mathcal{V} \succeq (\pi_i^{-1}\mathcal{U})_0^{n-1}$ and

$$(f \circ \pi_1)_n(\tilde{x}) = \sum_{j=0}^{n-1} (f \circ \pi_1)(\widetilde{T}^j(\tilde{x})) = \sum_{j=0}^{n-1} f \circ T^j(\pi_1 \tilde{x}) = f_n(\pi_1 \tilde{x}) = f_n(x),$$

where $x = \pi_1 \tilde{x}$. Then

$$\sum_{V \in \mathcal{V}} \sup_{\tilde{x} \in \pi_1^{-1}V \cap K_n} \exp(f \circ \pi_1)_n(\tilde{x}) = \sum_{V \in \mathcal{V}} \sup_{x \in V \cap \pi_1(K_n)} \exp f_n(x).$$

It follows that

$$P_n(\widetilde{T}, f \circ \pi_1, \pi_1^{-1}\mathcal{U}, K_n) \le P_n(T, f, \mathcal{U}, \pi_1(K_n)). \tag{6}$$

On the other hand, for each $\widetilde{\mathcal{V}} \in \mathscr{C}_X^{\mathbb{N}}$ with $\widetilde{\mathcal{V}} \succeq (\pi_1^{-1}\mathcal{U})_0^{n-1}$, $\tilde{x} \in \widetilde{V} \cap K_n$, $\pi_1 \widetilde{V} \succeq \mathcal{U}_0^{n-1}$, and

$$\begin{aligned}\sum_{\widetilde{V}\in\widetilde{\mathcal{V}}} \sup_{\tilde{x}\in\widetilde{V}\cap K_n} \exp(f\circ\pi_1)_n(\tilde{x}) &= \sum_{\widetilde{V}\in\widetilde{\mathcal{V}}} \sup_{x\in\pi_1(\widetilde{V}\cap K_n)} \exp f_n(x)\\ &= \sum_{V\in\pi_1\widetilde{V}} \sup_{x\in\pi_1\widetilde{V}\cap\pi_1 K_n} \exp f_n(x),\end{aligned}$$

where $x = \pi_1\tilde{x}$. Then we get the opposite part of the inequality of (6), and consequently

$$P_n(\widetilde{T}, f\circ\pi_1, \pi_1^{-1}\mathcal{U}, K_n) = P_n(T, f, \mathcal{U}, \pi_i(K_n)). \tag{7}$$

Now we have

$$\limsup_{n\to\infty}\frac{1}{n}\log P_n(\widetilde{T}, f\circ\pi_1, \pi_1^{-1}\mathcal{U}, K_n) = \limsup_{n\to\infty}\frac{1}{n}\log P_n(T, f, \mathcal{U}, \pi_1(K_n)).$$

From Lemma 3.1, we get

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log P_n(\widetilde{T}, f\circ\pi_1, \delta, K_n) \geq \lim_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log P_n(T, f, \delta, \pi_1(K_n)).$$

Conversely, let $\pi_i : \widetilde{X} \to X$ be the projection to the i$^{\text{th}}$ coordinate and $\widetilde{\mathcal{U}} \in \mathscr{C}_{\widetilde{X}}^{o}$. By the definition of $\widetilde{X}$, it is easy to see that there exists some $\mathcal{U} \in \mathscr{C}_X^o$ such that $\pi_i^{-1}(\mathcal{U}) \succeq \widetilde{U}$. Since for any two closed subsets C and D of X, $P_n(T, f, \mathcal{U}, C) \leq P_n(T, f, \mathcal{U}, D)$ and $\pi_i(K_n) \succeq T^{-(i-1)}\pi_i(K_n)$, by (7), we have

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{n}\log P_n(\widetilde{T}, f\circ\pi_1, \widetilde{\mathcal{U}}, K_n)\\ &\leq \limsup_{n\to\infty}\frac{1}{n}\log P_n(\widetilde{T}, f\circ\pi_1, \pi_i^{-1}\mathcal{U}, K_n)\\ &= \limsup_{n\to\infty}\frac{1}{n}\log P_n(T, f, \mathcal{U}, \pi_i(K_n))\\ &\leq \limsup_{n\to\infty}\frac{1}{n}\log P_n(T, f, \mathcal{U}, T^{-(i-1)}\pi_i(K_n))\\ &= \limsup_{n\to\infty}\frac{1}{n+i-1}\log P_{n+i-1}(T, f, \mathcal{U}, T^{-(i-1)}\pi_i(K_n))\\ &\leq \limsup_{n\to\infty}\frac{1}{n+i-1}\log\big(P_{i-1}(T, f, \mathcal{U}, T^{-(i-1)}\pi_i(K_n))\\ &\qquad\cdot P_n(T, f\circ T^{i-1}, T^{-(i-1)}\mathcal{U}, T^{-(i-1)}\pi_i(K_n))\big) \quad \text{(by Lemma 5.1)}\\ &= \limsup_{n\to\infty}\frac{1}{n}P_n(T, f, \mathcal{U}, \pi_1(K_n)) \qquad \text{(by Lemma 5.2)}\\ &\leq \lim_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}P_n(T, f, \delta, \pi_1(K_n)).\end{aligned}$$

By Lemma 3.1, we get

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log P_n(\widetilde{T}, f\circ\pi_1, \delta, K_n) \le \lim_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log P_n(T, f, \delta, \pi_1(K_n)). \quad \square$$

Now we can prove the following theorem.

Theorem 5.4. *Let (X, T) be a noninvertible TDS, $f \in C(X, \mathbb{R})$, and $\mu \in \mathcal{M}^e(X, T)$ with $h_\mu(T) > 0$. Then, for μ-a.e. $x \in X$, $\lim_{\epsilon\to 0} P_s(T, f, x, \epsilon) \ge P_\mu(T, f)$.*

Proof. Let $(\widetilde{X}, \widetilde{T})$ be the inverse limit of (X, T). For $\epsilon > 0$, $n \in \mathbb{N}$, and $\tilde{x} \in \widetilde{X}$, denote $K_n = \widetilde{T}^{-n} W^s_{\epsilon/2}(\tilde{x}, \widetilde{T})$. Then, from the definitions of d_T and $\widetilde{X}$, it is easy to see that $\pi_1(K_n) \subseteq T^{-n} W^s_\epsilon(x, T)$, where $x = \pi_1(\tilde{x})$. By Lemma 5.3, we have

$$\begin{aligned} P_s(T, f, x, \epsilon) &= \lim_{\delta\to 0}\limsup_{n\to+\infty}\frac{1}{n}\log P_n(T, f, \delta, T^{-n}W^s_\epsilon(x, T)) \\ &\ge \lim_{\delta\to 0}\limsup_{n\to+\infty}\log P_n(T, f, \delta, \pi_1(K_n)) \\ &= \lim_{\delta\to 0}\limsup_{n\to+\infty}\log P_n(\widetilde{T}, f\circ\pi_1, \delta, K_n) \\ &= P_s\Big(\widetilde{T}, f\circ\pi_1, \tilde{x}, \frac{\epsilon}{2}\Big). \end{aligned}$$

It follows that, for each $\tilde{x} \in \widetilde{X}$,

$$\lim_{\epsilon\to 0} P_s(T, f, \pi_1(\tilde{x}), \epsilon) \ge \lim_{\epsilon\to 0} P_s\Big(\widetilde{T}, f\circ\pi_1, \tilde{x}, \frac{\epsilon}{2}\Big). \tag{8}$$

Let $\tilde{\mu} \in \mathcal{M}^e(\widetilde{X}, \widetilde{T})$ with $\pi_1(\tilde{\mu}) = \mu$. Then, by Theorem 3.7, there exists a Borel subset $\widetilde{X}_0 \subseteq \widetilde{X}$ with $\tilde{\mu}(\widetilde{X}_0) = 1$ such that, for any $\tilde{x} \in \widetilde{X}_0$,

$$\begin{aligned} \lim_{\epsilon\to 0} P_s\Big(\widetilde{T}, f\circ\pi_1, \tilde{x}, \frac{\epsilon}{2}\Big) &\ge P_{\tilde{\mu}}(\widetilde{T}, f\circ\pi_1) = h_{\tilde{\mu}}(\widetilde{T}) + \int_{\widetilde{X}} f\circ\pi_1 d\tilde{\mu} \\ &\ge h_\mu(T) + \int_X f\, d\mu = P_\mu(T, f). \end{aligned} \tag{9}$$

Let $X_0 = \pi_1(\widetilde{X}_0)$. Then $X_0 \in \mathcal{B}_\mu$ and $\mu(X_0) = 1$. By the inequality (8) and (9), we have

$$\lim_{\epsilon\to 0} P_s(T, f, x, \epsilon) \ge P_\mu(T, f) \quad \text{for each } x \in X_0, \qquad \square$$

Theorem 5.4 immediately leads to the following corollary.

Corollary 5.5. *Let (X, T) be a noninvertible TDS and $f \in C(X, \mathbb{R})$. If there exists a $\mu \in \mathcal{M}^e(X, T)$ such that $P_\mu(T, f) = P(T, f)$, $\mathcal{P}(T, f) \ne \varnothing$.*

Lemma 5.6. *Let $(\widetilde{X}, \widetilde{T})$ be the inverse limit of a noninvertible TDS (X, T). If $A \subseteq \widetilde{E}$ is weak mixing, so is $\pi_1(A)$ and $P(\widetilde{T}, f\circ\pi_1, A) = P(T, f, \pi_1(A))$.*

Proof. The fact that $\pi_1(A)$ is weak mixing follows from Lemma 4.8 in [Blanchard and Huang 2008]. The latter follows from Lemmas 5.3 and 3.1. $\square$

The following theorem shows that Theorem 4.5 also holds for noninvertible TDS.

Theorem 5.7. *Let (X, T) be a noninvertible TDS and $\mu \in \mathcal{M}^e(X, T)$ with $h_\mu(T) > 0$. Then, for μ-a.e. $x \in X$, there exists a closed subset $E(x) \subseteq \overline{W^s(x, T)}$ such that $P(T, f, E(x)) \geq P_\mu(T, f)$ and $E(x) \in WM_s(X, T)$.*

Proof. Let $(\widetilde{X}, \widetilde{T})$ be the inverse limit of (X, T). Then there exists $\tilde{\mu} \in \mathcal{M}^e(\widetilde{X}, \widetilde{T})$ with $\pi_1(\tilde{\mu}) = \mu$, where π_1 is the projection to the first coordinate. Obviously,

$$P_{\tilde{\mu}}(\widetilde{T}, f \circ \pi_1) = h_{\tilde{\mu}}(\widetilde{T}) + \int_{\widetilde{X}} f \circ \pi_1 \, d\tilde{\mu} \geq h_\mu(T) + \int_X f \, d\mu = P(T, f).$$

By Theorem 4.5, there exists a Borel set $\widetilde{X}_0 \subseteq \widetilde{X}$ with $\tilde{\mu}(\widetilde{X}_0) = 1$ such that, for each $\tilde{x} \in \widetilde{X}_0$, there exists a closed subset $E(\tilde{x}) \subseteq \overline{W^s(\tilde{x}, \widetilde{T})}$ such that

$$P(\widetilde{T}, f \circ \pi_1, E(\tilde{x})) \geq P_{\tilde{\mu}}(\widetilde{T}, f \circ \pi_1) \quad \text{and} \quad E(\tilde{x}) \in WM_s(\widetilde{X}, \widetilde{T}).$$

Let $(X_0) = \pi_1(\widetilde{X}_0)$. Then $X_0 \in \mathcal{B}_\mu$ and $\mu(X_0) = 1$. For each $x \in X_0$ let $E(x) = \pi_1(E(\tilde{x}))$, where $x = \pi_1(\tilde{x})$. Then $E(x) \subseteq \pi_1(\overline{W^s(\tilde{x}, \widetilde{T})}) \subseteq \overline{W^s(x, T)}$. By Lemma 5.6, we have

$$P(T, f, E(x)) = P(\widetilde{T}, f \circ, E(\tilde{x})) \geq P_{\tilde{\mu}}(\widetilde{T}, f \circ \pi_1) \geq P_\mu(T, f)$$

and $E(x) \in WM_s(X, T)$. $\square$

The following result is immediate.

Corollary 5.8. *Let (X, T) be a noninvertible TDS. Then*

(a) $\sup_{x \in X} P(T, f, \overline{W^s(x, T)}) = P(T, f)$;

(b) *if there exists $\mu \in \mathcal{M}^e(X, T)$ with $P_\mu(T, f) = P(T, f)$, then, for μ-a.e. $x \in X$, there exists a closed subset $E(x) \subseteq \overline{W^s(x, T)}$ such that $E(x) \in WM_s(X, T)$ and $P(T, f, E(x)) = P(T, f)$.*

Remark 5.9. From the proof of Theorem 4.5, we know that $E(x) = \operatorname{supp}(\mu_x)$, where μ_x is a probability measure determined by the disintegration of $\mu \in \mathcal{M}^e(X, T)$ over the Pinsker σ-algebra $P_\mu(T)$.

Acknowledgements

We thank the reviewers for constructive comments and valuable suggestions.

References

[Blanchard and Huang 2008] F. Blanchard and W. Huang, “Entropy sets, weakly mixing sets and entropy capacity”, *Discrete Contin. Dyn. Syst.* **20**:2 (2008), 275–311. MR 2009a:37026 Zbl 1151.37019

[Blanchard et al. 2002] F. Blanchard, B. Host, and S. Ruette, “Asymptotic pairs in positive-entropy systems”, *Ergodic Theory Dynam. Systems* **22**:3 (2002), 671–686. MR 2003d:37004 Zbl 1018.37005

[Fiebig et al. 2003] D. Fiebig, U.-R. Fiebig, and Z. H. Nitecki, “Entropy and preimage sets”, *Ergodic Theory Dynam. Systems* **23**:6 (2003), 1785–1806. MR 2004k:37025 Zbl 1063.37017

[Huang 2008] W. Huang, “Stable sets and ϵ-stable sets in positive-entropy systems”, *Comm. Math. Phys.* **279**:2 (2008), 535–557. MR 2009f:37017 Zbl 1167.37013

[Huang and Yi 2007] W. Huang and Y. Yi, “A local variational principle of pressure and its applications to equilibrium states”, *Israel J. Math.* **161** (2007), 29–74. MR 2008i:37013 Zbl 1137.37008

[Ma et al. 2010] X. Ma, E. Chen, and A. Zhang, “A relative local variational principle for topological pressure”, *Sci. China Math.* **53**:6 (2010), 1491–1506. MR 2011m:37006 Zbl 1201.37008

[Romagnoli 2003] P.-P. Romagnoli, “A local variational principle for the topological entropy”, *Ergodic Theory Dynam. Systems* **23**:5 (2003), 1601–1610. MR 2004i:37030 Zbl 1056.37017

[Sumi 2003] N. Sumi, “Diffeomorphisms with positive entropy and chaos in the sense of Li–Yorke”, *Ergodic Theory Dynam. Systems* **23**:2 (2003), 621–635. MR 2004b:37040 Zbl 1042.37019

[Walters 1982] P. Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics **79**, Springer, New York, 1982. MR 84e:28017 Zbl 0475.28009

[Xiong and Yang 1991] J. C. Xiong and Z. G. Yang, “Chaos caused by a topologically mixing map”, pp. 550–572 in *Dynamical systems and related topics* (Nagoya, 1990), edited by K. Shiraiwa, Adv. Ser. Dynam. Systems **9**, World Scientific, River Edge, NJ, 1991. MR 93c:58153

Received April 27, 2012. Revised October 27, 2012.

XIANFENG MA
DEPARTMENT OF MATHEMATICS
EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY
SHANGHAI 200237
CHINA

xianfengma@gmail.com

ERCAI CHEN
SCHOOL OF MATHEMATICAL SCIENCES AND INSTITUTE OF MATHEMATICS
NANJING NORMAL UNIVERSITY
NANJING 210023
CHINA

and

CENTER OF NONLINEAR SCIENCE
NANJING UNIVERSITY
NANJING, 210093
CHINA

ecchen@njnu.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 263, No. 1, 2013
dx.doi.org/10.2140/pjm.2013.263.143

LIPSCHITZ AND BILIPSCHITZ MAPS ON CARNOT GROUPS

WILLIAM MEYERSON

Suppose A is an open subset of a Carnot group G and H is another Carnot group. We show that a Lipschitz function from A to H whose image has positive Hausdorff measure in the appropriate dimension is bilipschitz on a subset of A of positive Hausdorff measure. We also construct Lipschitz maps from open sets in Carnot groups to Euclidean space that do not decrease dimension. Finally, we discuss two counterexamples to explain why Carnot group structure is necessary for these results.

1. Introduction

Guy David [1988] proved that if f is a Lipschitz function from the unit cube in $\mathbb{R}^n$ to a subset of some Euclidean space with positive n-dimensional Hausdorff measure, there exists a subset K of the domain of f with positive n-dimensional Hausdorff measure such that f is bilipschitz on K.

Shortly thereafter, Peter Jones [1988] proved the following stronger result: if f is a Lipschitz function from the unit cube in $\mathbb{R}^n$ to a subset of some Euclidean space, then the unit cube can be broken up into the union of a "garbage" set (whose image under f has arbitrarily small n-dimensional Hausdorff content) and a finite number of sets $K_1, \dots, K_N$ such that f is bilipschitz on each K_i.

David [1991] later translated this proof into the language of wavelets, which are more readily generalizable to Heisenberg and other Carnot groups. The proof as written in [David 1991] only depends on a few general properties, all but one of which hold for Heisenberg (and other Carnot) groups.

This story has further generalizations: for example, [David and Semmes 1993] generalizes Jones' argument to work with Lipschitz functions that are only defined on Ahlfors d-regular subsets of a Euclidean space $\mathbb{R}^N$, with d possibly less than N, while [Semmes 2000] allows the domain and range to be metric spaces subject to a specific condition.

In Section 2 we adapt some of the ideas in [David 1991; Jones 1988] to Carnot groups and prove that a Lipschitz function between such groups having an image of positive Hausdorff measure in the appropriate dimension is bilipschitz on a subset

MSC2010: 43A80.
Keywords: analysis on Carnot groups, Heisenberg groups, Grushin plane, subriemannian, wavelets.

of the domain of positive Hausdorff measure. Section 3 investigates how big, in terms of dimension, Lipschitz images of Carnot groups in Euclidean space can be. Finally, Section 4 explores two counterexamples explaining why Carnot group structure is necessary for these results. In particular, neither Ahlfors regularity nor subriemannian manifold structure would be sufficient.

2. Jones-type decomposition for Carnot groups

2A. ***Brief outline.*** This section is organized as follows. In Section 2B we give some definitions concerning Carnot groups and set up some notational conventions. In Section 2C we state the five properties of Euclidean space on which David's argument rests and show how the first four of them work for Heisenberg groups. In Section 2D we explain why these properties also work for other Carnot groups. In Section 2E, we prove our main result (Theorem 2.12): If A is an appropriate subset of the k-th Heisenberg group H_k corresponding roughly to the unit cube in $\mathbb{R}^n$, and F is a Lipschitz function from A to another Heisenberg group whose image has positive Hausdorff $(2k+2)$-dimensional measure, then there exists $B \subset A$ with positive Hausdorff $(2k+2)$-dimensional measure such that F is bilipschitz on B. Finally, in Section 2F we derive some corollaries of Theorem 2.12.

Although our main focus is on the Heisenberg groups (especially H_1), all of the results in this paper apply equally well to Carnot groups in general. To exploit this fact, the results in Section 2E will be stated and proved in the more general context of Carnot groups.

2B. ***Definitions.***

Definition 2.1. The *n-th Heisenberg group H_n* is defined as the set

$$\{(z_1, \ldots, z_n, t) : z_j \in \mathbb{C},\ t \in \mathbb{R}\}$$

equipped with the following group law:

$$(z_1, \ldots, z_n, t)(w_1, \ldots, w_n, s) = \Big(z_1 + w_1, \ldots, z_n + w_n, t + s + \Im \sum_{j=1}^{n} z_j \overline{w}_j\Big),$$

where $\Im$ denotes imaginary part.

For $n = 1$, we often write z_1 in terms of its real components as $z_1 = x + iy$ and refer to the point (z_1, t) as (x, y, t), so H_1 inherits a natural Euclidean coordinate structure from $\mathbb{R}^3$.

The Heisenberg group is a special example of a Carnot group:

Definition 2.2. A *Carnot group G* is a connected, simply connected, nilpotent Lie group whose Lie algebra $\mathfrak{g}$ is graded, i.e.,

$$\mathfrak{g} = \bigoplus_{j=1}^{d} \mathfrak{g}_j,$$

where

$$[\mathfrak{g}_1, \mathfrak{g}_j] = \mathfrak{g}_{j+1} \quad \text{and} \quad \mathfrak{g}_{d+1} = \{0\}.$$

We call $\mathfrak{g}_1$ the *horizontal component* of $\mathfrak{g}$.

By standard results of Lie group theory (see, for example, [Varadarajan 1984]), the exponential map gives a diffeomorphism between a Carnot group and its Lie algebra. Further, the standard definition of a Lie algebra in terms of vector fields provides a canonical identification between the tangent space of a Lie group at a given point and the Lie group itself. (When $g \in G$ is fixed, for every tangent vector v there is a unique $X \in \mathfrak{g}$ such that $X(g) = v$ and we can identify $\exp(X)$ with v.)

We shall freely use these canonical identifications between a Carnot group, its Lie algebra, and its tangent space throughout this paper. For example, every Carnot group has a coordinate structure induced by its Lie algebra. For H_n, this coordinate structure was already mentioned in Definition 2.1, where $\mathfrak{g}_1$ consists of the points of the form $(z_1, \dots, z_n, 0)$ with final coordinate equal to zero.

Every Carnot group has a family of dilation homomorphisms $\{\delta_\lambda : \lambda > 0\}$ and a metric called the Carnot–Carathéodory metric. They are defined as follows:

Definition 2.3. Let $\lambda > 0$, let G be a Carnot group and let $g \in G$, where $g = \sum_i g_i$ with $g_i \in \mathfrak{g}_i$. Define the *dilation*

$$\delta_\lambda(g) = \sum_i \lambda^i g_i.$$

Definition 2.4. Let G be a Carnot group, let $g, h \in G$, and let $\Gamma_{g,h}$ be the set of all curves

$$\gamma : [0, 1] \to G$$

with $\gamma(0) = g$, $\gamma(1) = h$, and $\gamma'(t) \in \mathfrak{g}_1$ for each $t \in [0, 1]$. Define the *Carnot–Carathéodory distance* between g and h to be

$$d_{CC}(g, h) = \inf_{\gamma \in \Gamma_{g,h}} \int_0^1 |\gamma'(t)|\, dt,$$

where $|\gamma'(t)|$ is the length of $\gamma'(t)$ in a fixed Euclidean metric on the real vector space $\mathfrak{g}_1$.

Because $\Gamma_{g,h}$ in this definition is nonempty — see [Montgomery 2002] — we have $d_{CC}(g, h) < \infty$ whenever $g, h \in G$.

Note 2.5. It is often easier to work with a comparable L^∞ quasidistance function d based on the Carnot metric. For the first Heisenberg group H_1, this is done by defining distance to the origin as

$$d((x, y, z), (0, 0, 0)) = \max(|x|, |y|, |z|^{1/2})$$

and for an arbitrary g, h in this group, defining

$$d(g,h) = d(h^{-1}g, (0,0,0)).$$

There is of course a completely analogous construction in an arbitrary Carnot group: if G is a Carnot group, we use the grading of its Lie algebra $\mathfrak{g}$ as in the definition of Carnot groups:

$$\mathfrak{g} = \bigoplus_{j=1}^{d} \mathfrak{g}_j.$$

Because the identity element in a Carnot group is the image of the origin under the exponential map, we shall refer to it as 0. Now, letting g be an arbitrary point in G we first define its quasidistance to the identity element, $d(g,0)$, by recalling the direct sum decomposition

$$\exp^{-1}(g) = \sum_j g_j$$

with $g_j \in \mathfrak{g}_\mathrm{j}$ and setting

$$d(g,0) = \max_{1 \le j \le d} (\|g_j\|_j)^{1/j}.$$

where $\|\cdot\|_j$ is a norm on $\mathfrak{g}_j$ for $j = 1, \dots, d$. Finally, for an arbitrary $g, h \in G$, we finish by setting

$$d(g,h) = d(h^{-1}g, 0).$$

For the duration of this paper, d_{CC} shall refer to Carnot–Carathéodory distance and d shall refer to quasidistance.

A fundamental operation for Carnot groups is the Pansu differential, defined as follows (see [Capogna et al. 2007], for example):

Definition 2.6. Let $F : G \to H$ be a function from one Carnot group G to another Carnot group H. The *Pansu differential* $DF(g)$ of F at $g \in G$ is the map

$$DF(g) : G \to H$$

defined at $g' \in G$ as the limit

$$DF(g)(g') = \lim_{s \to 0} \delta_{s^{-1}}[F(g)^{-1}F(g\delta_s g')]$$

whenever the limit exists.

Using the canonical identifications stated above, we can view the Pansu differential as a map between Lie algebras or as a map from the tangent space at $g \in G$ to the tangent space at $F(g)$. We shall take advantage of this fact throughout.

In the tangent vector interpretation, the Pansu differential $DF(g)$ induces a linear map between the horizontal component of the tangent space of G at g and the horizontal component of the tangent space of H at $F(g)$ [Pansu 1989]. Calling this linear map $MF(g)$, we can view MF as a matrix-valued map sending g to $MF(g)$.

2C. *Five key properties.*

2C1. *Dyadic decomposition.* There exists a dyadic decomposition for Euclidean space defined as follows: For each nonnegative integer k we let $\mathcal{Q}_k$ be the set of all cubes of the form

$$\left(a_1 \cdot 2^{-k}, (a_1+1) \cdot 2^{-k}\right) \times \cdots \times \left(a_n \cdot 2^{-k}, (a_n+1) \cdot 2^{-k}\right)$$

contained in the unit cube, where the a_i are all integers. Then the elements of $\mathcal{Q}_k$ are disjoint open sets. Further, each element of $\mathcal{Q}_k$ is (up to a set of measure zero) a disjoint union of elements of $\mathcal{Q}_{k+1}$, the $\mathcal{Q}_k$ are all translates of each other, and one can transform an arbitrary element of $\mathcal{Q}_k$ into an arbitrary element of $\mathcal{Q}_{k+1}$ by a dilation (by a factor of 2^{-1}) followed by translation. Finally, fixing a cube $Q \in \mathcal{Q}_k$ and letting d be its diameter (i.e., $d = \sqrt{n}2^{-k}$), the number of cubes in $\mathcal{Q}_k$ whose distance from Q is at most d is bounded above by a constant depending only on n.

Our immediate goal is to generalize this decomposition to the Heisenberg group H_1. To do this we loosely follow Christ's construction of Theorem 11 in [Christ 1990]. First we let B_0 denote the discrete subgroup of H_1 generated by $(1, 0, 0)$ and $(0, 1, 0)$ and call it the *discrete Heisenberg group*. We then define B_n, for each positive integer n, to be the image of B_0 under the dilation $\delta_{10^{-n}}$ (in particular, the first 2 coordinates are multiplied by 10^{-n}; the final coordinate is multiplied by 10^{-2n}). Equivalently, B_n is the subgroup of the first Heisenberg group generated by $(10^{-n}, 0, 0)$ and $(0, 10^{-n}, 0)$. *If x is a point in B_n, we give it the label* (x, n) and note that x has a different label for each B_n containing x. We form a tree by defining an order relation $\leq$ on the set of all such pairs (x, n). We start this procedure with the following definition.

Definition 2.7. We say that (x, α) is a *parent* of (y, β) if $\beta = \alpha + 1$ and $y = xg$, where the first two components of g all lie in $\left(-\frac{1}{2}10^{-\alpha}, \frac{1}{2}10^{-\alpha}\right]$ and the final component lies in $\left(-\frac{1}{2} \cdot 10^{-2\alpha}, \frac{1}{2} \cdot 10^{-2\alpha}\right]$.

Using the obvious analogies from family trees ("ancestor", "descendant", "grandparent", "sibling", etc.) for both the tree points and corresponding dyadic cubes (to be defined momentarily), we say $(x, \alpha) \leq (y, \beta)$ if (y, β) is an ancestor of (x, α). Following along exactly as in Definition 14 of [Christ 1990], we create from this tree a family of dyadic "cubes". In particular, we define

$$Q(x, \alpha) = \bigcup_{(y,\beta) \leq (x,\alpha)} B_{CC}(y, \tfrac{1}{10}10^{-\beta}),$$

where $B_{CC}(z, \epsilon)$ is the ball centered at z of radius ϵ with respect to Carnot–Carathéodory distance. We will say that each cube $Q(x, \alpha)$ is a *cube at scale* α and we define $\mathcal{Q}_\alpha$ to be the set of all the cubes of scale α. All the cubes in $\mathcal{Q}_\alpha$ are translates of each other by elements of the discrete Heisenberg group of the appropriate scale; further, each member of each $\mathcal{Q}_\alpha$ is an open set while each element of $\mathcal{Q}_\alpha$ is (up to a set of measure zero) the disjoint union of elements of $\mathcal{Q}_{\alpha+1}$. Also, one can transform an arbitrary element of $\mathcal{Q}_\alpha$ into an arbitrary element of $\mathcal{Q}_{\alpha+1}$ by a dilation (by a factor of 10^{-1}) followed by translation. Finally, the number of cubes in $\mathcal{Q}_\alpha$ within $\operatorname{diam}(Q(x,\alpha))$ of $Q(x,\alpha)$ is bounded by a constant independent of α.

Analogously, for the k-th Heisenberg group, we begin by rewriting the elements of H_k to mirror the above construction for H_1: in other words, writing $z_j = x_{2j-1} + i x_{2j}$ where $x_{2j-1}, x_{2j} \in \mathbb{R}$, we let B_0 be the subgroup of H_k generated by

$$\{(x_1, \dots, x_{2k}, 0) : x_j = \pm\delta_{j,l},\ 1 \le l \le 2k\}$$

where $\delta_{j,l}$ is the Kronecker delta. In this setting, B_n would be the subgroup of H_k generated by

$$\{(x_1, \dots, x_{2k}, 0) : x_j = \pm 10^{-n}\delta_{j,l},\ 1 \le l \le 2k\}$$

and the construction for H_1 goes through for H_k with only minor changes. In particular, the definition of (x,α) being a parent of (y,β) would now require $y = xg$ where the first $2k$ components of g all lie in $\left(-\frac{1}{2}10^{-\alpha}, \frac{1}{2}10^{-\alpha}\right]$ and the final component lies in $\left(-\frac{1}{2}\cdot 10^{-2\alpha}, \frac{1}{2}\cdot 10^{-2\alpha}\right]$.

In this construction, the analogue to the unit cube in Euclidean space is the unique cube of scale 0 containing the identity element; according to the notation defined in the preceding paragraph, the name for this cube is $Q(0,0)$.

Remark. In making this decomposition we are saying nothing about the boundaries of the elements of the $\mathcal{Q}_\alpha$ other than that they are closed sets of Hausdorff measure zero in the appropriate dimension. Also, this decomposition is not the same as the decomposition of the Heisenberg group found in [Strichartz 1992].

2C2. *Orthogonal decomposition of L^2.* Looking back at Euclidean space $\mathbb{R}^n$ for inspiration, we note that the Hilbert space $L^2([0,1]^n)$ of square-integrable functions on the unit cube can be decomposed into orthogonal subspaces as follows: if β is a positive integer, we define $C_\beta \subset L^2([0,1]^n)$ as

$$\left\{ f \in L^2([0,1]^n) : f|_Q \text{ is constant for } Q \in \mathcal{Q}_\beta \text{ and } \int_Q f = 0 \text{ for } Q \in \mathcal{Q}_{\beta-1} \right\},$$

while $C_0 \subset L^2([0,1]^n)$ is defined as

$$\{ f \in L^2([0,1]^n) : f \text{ is constant}\}.$$

This yields the orthogonal decomposition

$$L^2([0,1]^n) = \bigoplus_{\beta=0}^{\infty} C_\beta.$$

In other words, if $f \in C_\beta$, $g \in C_\gamma$ with $\beta \neq \gamma$, $\int_{[0,1]^n} fg = 0$ while for each $h \in L^2([0,1]^n)$ there exists $h_\beta \in C_\beta$ for β a nonnegative integer with $h = \sum_{\beta=0}^{\infty} h_\beta$, the sum in question converging in $L^2([0,1]^n)$ to h.

For the Heisenberg groups we can mimic this procedure as follows: here, our "base" cube shall be denoted as $Q(0,0)$ where the first zero denotes the origin and the second zero denotes scale. Similarly, we define the C_β (as subspaces of the Hilbert space $L^2(Q(0,0))$ of real-valued, square-integrable functions) identically to the way we did with Euclidean space. In other words, if β is a positive integer, we define $C_\beta \subset L^2(Q(0,0))$ as

$$\left\{ f \in L^2(Q(0,0)) : f|_Q \text{ is constant for } Q \in \mathfrak{Q}_\beta \text{ and } \int_Q f = 0 \text{ for } Q \in \mathfrak{Q}_{\beta-1} \right\},$$

while $C_0 \subset L^2(Q(0,0))$ is defined as

$$\{ f \in L^2(Q(0,0)) : f \text{ is constant}\}.$$

This yields the orthogonal decomposition

$$L^2(Q(0,0)) = \bigoplus_{\beta=0}^{\infty} C_\beta.$$

For $\beta > 0$, C_β has a spanning set consisting of the functions $f_{Q,Q'}$ for Q, Q' sibling cubes in $\mathfrak{Q}_\beta$ defined as follows: $f_{Q,Q'}$ is equal to 1 on Q, -1 on Q', and 0 everywhere else; we shall call this spanning set S_β. S_β is *approximately orthogonal* in the following sense: there exists some universal constant K (independent of β) such that for each $f \in S_\beta$ we have

$$\#\left\{ g \in S_\beta : \int_{Q(0,0)} fg \neq 0 \right\} \leq K,$$

where the # symbol denotes cardinality.

When we proceed to the proof, we will wish to find a fixed $Y > 0$ such that if $g, g' \in Q(0,0)$ with $g \neq g'$, there exists some dyadic cube Q such that

$$\operatorname{diam}(Q) < Y d_{CC}(g, g') \text{ and } g, g' \in Q.$$

This is arranged by considering not just the cube families $\mathfrak{Q}_\alpha$ discussed in the previous section but expanding each cube family $\mathfrak{Q}_\alpha$ to a larger family $\mathfrak{Q}'_\alpha$.

If $\alpha > 0$, we define $\mathfrak{Q}'_\alpha$ to consist of the cubes of the form

$$\{ gQ : Q \in \mathfrak{Q}_\alpha,\ g \in B_{\alpha+2}\};$$

remember that $B_{\alpha+2}$ was defined in the previous subsection as the discrete Heisenberg group of scale $\alpha + 2$.

This does not cause the number of dyadic cubes of a given scale to multiply unreasonably because writing $g \in B_{\alpha+2}$ in coordinate form as $(z_1, \ldots, z_n, t)$, every element of $\mathcal{Q}'_\alpha$ can be written as gQ for some $Q \in \mathcal{Q}_\alpha$ and $g \in B_{\alpha+2}$ with

$$z_1, \ldots, z_n \in [-10^{-\alpha}, 10^{-\alpha}] \text{ and } t \in [-10^{-2\alpha}, 10^{-2\alpha}].$$

Letting L_g denote left translation by g whenever $g \in H^k$, we then define

$$C'_\beta = \{f \circ L_{g^{-1}} : f \in C_\beta, g \in B_{\beta+2}\}.$$

In fact, writing $g \in B_{\beta+2}$ in coordinate form as $(z_1, \ldots, z_n, t)$, every element of C'_β can be written as $f \circ L_{g^{-1}}$ for some $f \in C_\beta$ and $g \in B_{\beta+2}$ with

$$z_1, \ldots, z_n \in [-10^{-\beta}, 10^{-\beta}] \quad \text{and} \quad t \in [-10^{-2\beta}, 10^{-2\beta}].$$

Fixing β, we can construct an approximately orthogonal basis for C'_β analogously to the way we did for each C_β: we simply construct an approximately orthogonal basis for $C_\beta \circ L_{g^{-1}}$ for each g separately.

Finally, for both Euclidean space and the Heisenberg group, it is occasionally necessary to work with sets on a slightly larger scale than the unit cube. To do this, one fixes some integer $k \leq 0$, denotes our base cube to be the cube of scale k which contains $Q(0, 0)$, and then defines C_β and C'_β appropriately for $\beta \leq 0$ (for example, C_k will consist of the constant functions on our new base cube here).

2C3. *Differentiability.* On the Euclidean unit cube, there exists a Jacobian map that sends each Lipschitz function f (which may be either scalar-valued or Euclidean vector-valued) on the unit cube to the almost-everywhere-defined function Jf, the Jacobian of f. At almost every point, the Jacobian is a linear map from the tangent space of the domain to the tangent space of the image. Further, the partial derivative of each component is bounded above by the Lipschitz coefficient of f. Finally, a Lipschitz function f with almost everywhere constant Jacobian defined on a connected open set is uniquely determined by this Jacobian and its value at a single point: if Jf is equal to the linear map T almost everywhere and $f(x_0) = y_0$ then

$$f(x) = T(x - x_0) + y_0 \text{ for all } x \text{ where } f(x) \text{ is defined.}$$

Similarly, if G and H are two Heisenberg groups and $F : G \to H$ is Lipschitz, then by [Pansu 1989] the Pansu differential DF (which, for almost every $g \in G$ induces a map $DF(g) : G \to H$) satisfies these three properties:

(i) At almost every $g \in G$, the differential of the Lie group map $DF(g)$ at the identity induces a Lie algebra homomorphism from the tangent space of G at g to the tangent space of H at $F(g)$.

(ii) The magnitude of each component of DF is bounded above (up to a constant depending on normalization) by the Lipschitz coefficient of F.

(iii) If for almost every g with respect to Haar measure on G, $DF(g)$ (which was defined as an H-valued function defined on G) is equal to the Lie group homomorphism $\phi : G \to H$ and $g_0 \in G$, $h_0 \in H$ with $F(g_0) = h_0$ then

$$F(g) = h_0\phi(g_0^{-1}g) \quad \text{for all } g \text{ where } F(g) \text{ is defined.}$$

Of the properties, only (iii) is not a simple consequence of [Pansu 1989]. However, (iii) is a direct consequence of this fact concerning uniqueness of Lipschitz maps:

Fact 2.8. *Suppose G and H are Carnot groups, $U \subset G$ is connected and open, $g_0 \in U$ and $F_1 : U \to G$ and $F_2 : U \to G$ are two Lipschitz maps such that $DF_1(g) = DF_2(g)$ for almost all $g \in U$ with respect to Haar measure and $F_1(g_0) = F_2(g_0)$. Then $F_1 = F_2$.*

Proof. Suppose there exists $u \in U$ with $F_1(u) \neq F_2(u)$. Fix $\epsilon > 0$ such that

$$d_{CC}(F_1(u), F_2(u)) > \epsilon.$$

Let γ be a piecewise horizontal curve in U joining g_0 to u. There exists $g' \in G$ sufficiently close to the identity such that the left translation of γ by g' lies in U (which implies that $g'g_0, g'u \in U$) with $d_{CC}(F_1(g'g_0), F_2(g'u)) > \epsilon$ and almost everywhere on this translation, $DF_1 = DF_2$. However, integration then implies

$$F_1(g'u)F_1(g'g_0)^{-1} = F_2(g'u)F_2(g'g_0)^{-1}.$$

Therefore, we know that $d_{CC}(F_1(g'u), F_2(g'u)) = d_{CC}(F_1(g'g_0), F_2(g'g_0)) > \epsilon$; since g' can be made arbitrarily close to the identity this gives us that

$$\epsilon \leq d_{CC}(F_1(u), F_2(u)) = d_{CC}(F_1(g_0), F_2(g_0)) = 0$$

producing a contradiction, so we conclude that $F_1 = F_2$ as desired. □

In fact, because each linear map ψ from the horizontal component of G to the horizontal component of H has at most one extension (which we call $\tilde{\psi}$) to a Lie group homomorphism from G to H, we can go one step further and say that if MF is equal to the linear map ψ almost everywhere and $g_0 \in G$, $h_0 \in H$ with $F(g_0) = h_0$ then

$$F(g) = h_0\tilde{\psi}(g_0^{-1}g)$$

for all g where $F(g)$ is defined.

2C4. *Weak convergence.* If a sequence f_n of uniformly Lipschitz functions on a bounded Euclidean region converges uniformly to some function f then f is Lipschitz, and moreover the Jacobians Jf_n converge weakly in L^2 to the Jacobian of f. In other words:

Fact 2.9. *Let $U \subset \mathbb{R}^k$ be a bounded open set and let $\{f_n\} : U \to \mathbb{R}^m$ be a sequence of uniformly Lipschitz functions which converges uniformly to the function $f : U \to \mathbb{R}^m$. If $g : U \to \mathbb{R}$ is an L^2 function and D represents partial differentiation with respect to a fixed vector in $\mathbb{R}^k$ then*

$$\int_U (Df_n)g \to \int_U (Df)g,$$

where the integrals are with respect to Lebesgue measure and the derivatives in question are defined almost everywhere.

As will be stated shortly, Fact 2.9 generalizes to Heisenberg groups when the map MF induced by the Pansu differential (see the definitions section) is used in place of the Jacobian. In particular, one notes that because MF consists of derivatives of horizontal components of F with respect to horizontal tangent vectors, MF can be viewed as an array of horizontal derivatives of real-valued Lipschitz functions (after postcomposing with the appropriate coordinate functions). Then, the weak convergence in question is the following fact:

Fact 2.10. *Let $U \subset H_k$ be a bounded open set and let $\{f_n\} : U \to H_m$ be a sequence of uniformly Lipschitz functions which converges uniformly to the function $f : U \to H_m$. If $g : U \to \mathbb{R}$ is an L^2 function* (*with respect to Haar measure*) *and D represents partial differentiation with respect to a fixed left-invariant horizontal vector field in H_k then*

$$\int_U (Df_n)g \to \int_U (Df)g,$$

where the integrals are with respect to Haar measure and the derivatives in question are defined almost everywhere.

Facts 2.9 and 2.10 have the same classical proof, which involves approximating g by sufficiently smooth test functions with compact support and integrating by parts.

2C5. *Lipschitz extension.* If A is a subset of the unit cube of $\mathbb{R}^n$ and f is a Lipschitz function from A to some Euclidean space, then f can be extended to a Lipschitz function on the entire unit cube (or, in fact, to all of $\mathbb{R}^n$ for that matter). It is not known whether this extension property also holds for maps from a subset of a Heisenberg group G into the same group G [Balogh and Fässler 2009; Brudnyi and Brudnyi 2007], and for that reason we assume the Lipschitz map in Corollary 2.17 below is defined on an open subset of G. It has been shown recently in [Balogh and Fässler 2009] that this extension property does not hold for maps from $\mathbb{R}^k$ to H_n with $n < k$. Also, [Rigot and Wenger 2010] shows that the property does not hold for maps from $\mathbb{R}^k$ to any jet space on H_n whenever $n < k$. However, this property does hold for maps from any Carnot group to any $\mathbb{R}^k$. It also holds for maps from $\mathbb{R}^2$ to H_n for $n \geq 2$, as was shown in [Fässler 2007; Magnani 2010].

More generally, based on recent results in [Wenger and Young 2010] the Lipschitz extension property holds for maps from any set with Assouad–Nagata dimension less than or equal to n to any jet space group on $\mathbb{R}^n$. Notably, this implies the Lipschitz extension property for maps from H_k to H_{2k+1}.

2D. ***General Carnot groups.*** We now explain how the constructions performed in Section 2C on the Heisenberg group can be generalized to work on other Carnot groups. We need a notion of discretization (already used implicitly in the decomposition in Section 2C1).

Definition 2.11. Let G be a Carnot group whose Lie algebra $\mathfrak{g}$ is graded as

$$\mathfrak{g} = \bigoplus_{j=1}^{d} \mathfrak{g}_j.$$

Write m_j as the vector space dimension of $\mathfrak{g}_j$ for $1 \le j \le d$. We say that G is *discretizable* if for $1 \le j \le d$ there exist collections

$$\{X_{(j,i)}\}_{i=1}^{m_j} \in \mathfrak{g}_j \quad \text{and} \quad \{g_{(j,i)}\}_{i=1}^{m_j} \in G, \quad \text{with } \exp(X_{(j,i)}) = g_{(j,i)},$$

and subgroups

$$H_j \le G, \quad \text{where } H_j = \langle \{g_{(j',i)}\}_{1 \le i \le m_{j'},\, j \le j' \le d} \rangle,$$

such that $\{X_{(j,i)}\}_{i=1}^{m_j}$ spans $\mathfrak{g}_j$ as a vector space and, writing $G' = \langle \{g_{(1,i)}\}_{i=1}^{m_1} \rangle$ and $G_j = \langle \{\exp(\mathfrak{g}_{j'})\}_{j' \ge j} \rangle$,

$$G' \text{ is a discrete subgroup with } G' \cap G_j = H_j.$$

In this setting, we say that G' is the *discretization* of G.

Examples of discretizable Carnot groups include Heisenberg groups, Euclidean spaces, and jet spaces. For example, we can take the discrete Heisenberg groups as the discretization of the Heisenberg groups.

If G is discretizable, the method in [Christ 1990] can be followed as in Section 2C1 to create a dyadic decomposition, with B_0 now defined to be the discretization G'. Although the scaling constant used to create B_n from B_0 (which was 10^{-n} in the case of Heisenberg groups) depends on the specific Carnot group itself (in particular, it depends on the relationship between the coordinates of an arbitrary point g and $d_{CC}(g, 0)$; compare Theorem 2.10 in [Montgomery 2002]), the procedure for Heisenberg groups can otherwise be copied exactly to create a dyadic decomposition for G into dyadic "cubes". Because the base cube for our construction will still be a cube based at the origin of scale zero, we can still refer to it as $Q(0, 0)$. With our new dyadic decomposition in hand, we can also copy the construction of the C_β and C'_β in Section 2C2 in the setting of our discretizable group G.

Finally, we observe that the results in Sections 2C3 and 2C4 (which involved differentiability and weak convergence) used no properties specific to Heisenberg groups. Therefore, the results in Sections 2C3 and 2C4 carry over just as well to maps from one Carnot group to another. Actually, Fact 2.8 in Section 2C3 was already stated and proved in terms of Carnot groups.

2E. ***Proof of main theorem.*** In what follows, H^k and h^k shall refer to Hausdorff k-dimensional measure and Hausdorff k-dimensional content, respectively (both of which we define with respect to Carnot–Carathéodory distance).

Our goal is to prove the following theorem.

Theorem 2.12. *Let G be a discretizable Carnot group of homogeneous dimension k and let H be another Carnot group. Suppose $F : Q(0,0) \subset G \to H$ is Lipschitz. If $\delta > 0$, there exists a positive integer N and subsets $Z, X_1, \ldots, X_N$ of $Q(0,0)$ such that*

$$h^k(F(Z)) < \delta,$$
$$Z \cup X_1 \cup \cdots \cup X_N = Q(0,0),$$

and F is bilipschitz on each X_i. Furthermore, N and the bilipschitz coefficients of the $F|X_i$ depend only on the groups G and H, δ, and the Lipschitz coefficient of F, and not on the map F itself.

Before beginning the proof, we shall introduce two notions of nearness.

Definition 2.13. Suppose $Q(x,\alpha)$ and $Q(y,\alpha)$ are elements of the decomposition from 2C1 of a discretizable Carnot group into cubes of the same scale. We say that $Q(x,\alpha)$ and $Q(y,\alpha)$ are *adjacent* if the distance from $Q(x,\alpha)$ to $Q(y,\alpha)$ is bounded above by the diameter of $Q(x,\alpha)$.

Note that two coincident cubes of the same scale are considered adjacent.

Definition 2.14. Suppose $Q(x,\alpha)$ and $Q(y,\alpha)$ are elements of the decomposition of a discretizable Carnot group into cubes of the same scale. We say that $Q(x,\alpha)$ and $Q(y,\alpha)$ are *semiadjacent* if $Q(x,\alpha)$ and $Q(y,\alpha)$ are not adjacent and the parents of $Q(x,\alpha)$ and $Q(y,\alpha)$ are not adjacent, but the grandparents of $Q(x,\alpha)$ and $Q(y,\alpha)$ are adjacent.

Turning to the proof of Theorem 2.12, we begin by establishing some further notation and normalizations.

Let E be the ratio of the diameter of an arbitrary "cube" to the diameter of one of its "children" using Carnot–Carathéodory distance. For example, if G is a Heisenberg group (using exactly the "cube" decomposition from Section 2C1), then $E = 10$.

Using the Carnot–Carathéodory distance, we set

$$\theta = \operatorname{diam}(Q(0,0)).$$

Also, we let $0 < L_1 < L_2 < \infty$ be constants such that if Q and Q' are semiadjacent,

$$L_1 \operatorname{diam}(Q) < d(Q, Q') < L_2 \operatorname{diam}(Q).$$

We note that L_1 and L_2 only depend on G, not Q or Q'.

In addition, we may assume that F is 1-Lipschitz and that there exists $\eta > 0$ such that F is defined on the dilation $\delta_{1+\eta} Q(0,0)$. For convenience we scale Hausdorff measure so that $|Q(0,0)| = 1$ where $|S|$ denotes the Hausdorff measure of S.

Finally, we let W be a positive integer such that every cube Q' of scale $W - 10$ such that $Q' \cap Q(0,0) \neq \varnothing$ satisfies $Q' \subset \delta_{1+\eta} Q(0,0)$. Throughout this proof, we will be focusing primarily on subcubes of $Q(0,0)$ of scale at least W.

With our notation and normalizations set up, we prove the following proposition, which provides a partial wavelet decomposition of the linear map MF induced by the Pansu differential DF of F.

Proposition 2.15. *Suppose $1 \geq \epsilon > 0$. There exists $n, C > 0$ such that if $\alpha \geq W$ and $Q := Q(a, \alpha)$ and $Q' := Q(b, \alpha)$ are semiadjacent cubes with*

$$h^k(F(Q)) > \epsilon E^{-k\alpha} \tag{1}$$

and

$$h^k(F(Q')) > \epsilon E^{-k\alpha} \tag{2}$$

but

$$d_{CC}(F(Q), F(Q')) \leq \tfrac{1}{2} \epsilon L_1 \theta E^{-\alpha} \tag{3}$$

then there exists $\beta \in [\alpha - 4, \alpha + n]$ and $f_{Q,Q'} \in C'_\beta$ and integers i, j such that

$$\frac{|\langle (MF)_{i,j}, f_{Q,Q'} \rangle|}{|\langle f_{Q,Q'}, f_{Q,Q'} \rangle|} \geq C |Q|^{1/2}, \tag{4}$$

where C'_β is the space defined in Section 2C2 and MF is the matrix of horizontal components of the Pansu differential DF.

Further, C only depends on G, H, and ϵ (and, in particular, not on the specific choice of F).

Also, the inner product in (4) is taken with respect to $L^2(G)$; it equals

$$\int_G (MF)_{i,j} f_{Q,Q'} \, d\mu$$

where μ is Haar measure on G scaled so that $\mu(Q(0,0)) = 1$.

We also note that the number of possible candidates for $f_{Q,Q'}$ for a given Q is uniformly bounded, with a bound that depends only on the specific groups G and H.

Proof. Assume the contrary. Then, for each n there exists a 1-Lipschitz map F_n and semiadjacent cubes $Q(a_n, \alpha_n)$ and $Q(b_n, \alpha_n)$ such that

- $h^k(F_n(Q(a_n, \alpha_n))) > \epsilon E^{-k\alpha_n}$,
- $h^k(F_n(Q(b_n, \alpha_n))) > \epsilon E^{-k\alpha_n}$,
- $d_{CC}\big(F_n(Q(a_n, \alpha_n)), F_n(Q(b_n, \alpha_n))\big) < \frac{1}{2}\epsilon L_1 \theta E^{-\alpha_n}$, and
- $\int_{Q(0,0)} \psi f \leq 2^{-n} |Q(a_n, \alpha_n)|^{1/2} \|f\|^2_{L^2(Q(0,0))}$ whenever ψ is a matrix entry of MF_n and $f \in C'_\beta$, where $\beta \in [\alpha_n - 4, \alpha_n + n]$.

By rescaling and translating we may suppose

$$Q(a_n, \alpha_n) = Q(a, \alpha)$$

for all n and by passing to a subsequence we suppose

$$Q(b_n, \alpha_n) = Q(b, \alpha)$$

for all n. Further, the Arzelà–Ascoli theorem lets us pass to another subsequence such that F_n converges uniformly on $Q(0, 0)$ to some Lipschitz map F. Moreover, by translation (we can do this because of the expanded C' families) we can suppose $Q(a, \alpha)$ and $Q(b, \alpha)$ have the same great-great-grandparent $Q(z, \alpha - 4)$. By weak-star convergence, the restriction of each component of MF to $Q(z, \alpha - 4)$ is orthogonal to C_β for $\beta > \alpha - 4$ which implies that MF is constant almost everywhere on $Q(z, \alpha - 4)$. From this, the discussion in Section 2C3 lets us conclude that there exists a Lie group homomorphism ϕ such that $DF = \phi$ on $Q(z, \alpha - 4)$ and further, there exist elements $g_0 \in G$, $h_0 \in H$ such that

$$F(g) = h_0 \phi(g_0^{-1} g) \tag{5}$$

for all $g \in Q(z, \alpha - 4)$. Further,

$$h^k(F(Q(a, \alpha)) \geq \liminf_n h^k(F_n(Q(a, \alpha))) \geq \epsilon E^{-k\alpha}$$

because $F_n(Q(a, \alpha))$ is eventually contained in an arbitrarily small neighborhood of the closure $\overline{F(Q(a, \alpha))}$; such a neighborhood can have Hausdorff content arbitrarily close to $h^k(F(Q(a, \alpha)))$.

Working towards a contradiction, we next define the sequences of points $\{X_n\}$ and $\{Y_n\}$ such that

$$X_n \in Q(a, \alpha), Y_n \in Q(b, \alpha)$$

and

$$d_{CC}(F_n(X_n), F_n(Y_n)) \leq \tfrac{1}{2}\epsilon L_1 \theta E^{-\alpha}.$$

By the definition of sequential compactness, there exist points $a' \in Q(a, \alpha)$, $b' \in Q(b, \alpha)$ such that

$$d_{CC}(F(a'), F(b')) \le \tfrac{1}{2}\epsilon L_1 \theta E^{-\alpha}.$$

However, because $Q(a, \alpha)$ and $Q(b, \alpha)$ are semiadjacent,

$$d_{CC}(a', b') \ge L_1 \theta E^{-\alpha}.$$

Therefore, since (5) implies that the Pansu differential DF of F is defined everywhere and, in fact, is constant, the image of the Pansu differential DF of F in the direction of the tangent vector from a' to b' has magnitude at most $\frac{1}{2}\epsilon$. As F is Lipschitz with coefficient 1, this implies that

$$h^k(F(Q(a, \alpha))) \le |F(Q(a, \alpha))| \le \tfrac{1}{2}\epsilon E^{-k\alpha}. \tag{6}$$

The first inequality in (6) follows immediately from the fact that Hausdorff content is bounded above by Hausdorff measure. The second inequality is a direct consequence of the change-of-variables formula for Carnot groups (see the proof of Theorem 7 of [Vodopyanov and Ukhlov 1996], which can be directly adapted to this case).

As (6) contradicts our hypotheses, the proposition follows. □

Armed with this proposition, our next goal is to show that a sufficiently large portion of our domain lies in finitely many such semiadjacent pairs.

Proposition 2.16. *Let Ω be the set of all pairs of cubes which satisfy the hypotheses of Proposition 2.15 and let*

$$\phi(x) = \#\{\omega = (Q, Q') \in \Omega : x \in Q \cup Q'\}.$$

Suppose $N > 0$; then there exists a constant K' depending only G, H, and ϵ such that

$$|\{x : \phi(x) \ge N\}| \le K' N^{-1}.$$

Proof. If $(Q, Q') \in \Omega$, Proposition 2.15 gives us a wavelet function $f_{Q,Q'}$ corresponding to (Q, Q') such that the projection of MF onto $f_{Q,Q'}$ had L^2 magnitude at least $C\epsilon|Q|^{1/2}$. However, only a bounded number of pairs of cubes can be assigned a given wavelet function in this way. This is because of the control that Proposition 2.15 gives to both the scale and support of $f_{Q,Q'}$ in terms of the scale and location of Q. Now, we seek to show that

$$1 \succeq \sum_{(Q,Q') \in \Omega} |Q|$$

where the implied multiplicative constant only depends on G, H, and ϵ.

Because F is 1-Lipschitz, we can replace the constant 1 on the left hand side with $\|MF\|_2^2$. Next, for any specific pair (Q, Q') in our sum, we let $\pi_{(Q,Q')}(MF)$

be the orthogonal projection of MF onto $f_{Q,Q'}$. By Proposition 2.15,

$$\|\pi_{(Q,Q')}(MF)\|_2 \geq C\epsilon|Q|^{1/2};$$

in other words,

$$\int |\pi_{(Q,Q')}(MF)|^2 \geq C^2\epsilon^2|Q|.$$

Summing this over Ω gives us indeed that

$$1 \succeq \|MF\|_2^2 \succeq \sum_{(Q,Q')\in\Omega} |Q|$$

because the $f_{Q,Q'}$ are approximately orthogonal and a given wavelet function can only appear in the sum a bounded number of times. However,

$$\int \phi = \sum_{(Q,Q')\in\Omega} |Q|,$$

so Chebyshev's inequality therefore tells us that

$$S_N = \{x : \phi(x) \geq N\}$$

has

$$|S_N| \preceq N^{-1},$$

which proves the proposition. □

Proof of theorem. We complete the theorem through an infinite series of iterations as in [Jones 1988]. This process is divided into stages (indexed by $\alpha \geq 0$); at stage α we assign each point x of each subcube of $Q(0, 0)$ of scale α a label x_α, i.e., a finite string of zeroes and ones, such that every point in a fixed cube of scale α has the same label.

At stage 0 we apply a leading digit of 0 to every point in the base cube. In other words, for each $x \in Q(0, 0)$, we set $x_0 = 0$. Also, we define $Z_0 = \varnothing$ for future reference.

For $0 < \alpha$, we begin by defining the garbage set Z_α by letting S_α be the collection of all cubes Q of scale $\alpha + W$ such that

$$|F(Q)| \leq \delta E^{-k(\alpha+W)}$$

and set $Z_\alpha = S_\alpha \cup Z_{\alpha-1}$.

Next, we run through each pair of cubes at scale $\alpha + W$ which lie in $Q(0, 0) \setminus Z_\alpha$ and which satisfy the hypotheses of Proposition 2.16 with $\epsilon = \frac{1}{100}\delta$. Supposing that there are n_α such pairs $(Q_1, Q'_1), \ldots, (Q_{n_\alpha}, Q'_{n_\alpha})$, we will inductively define the labels $x_{(\alpha,m)}$ for $m = 0, 1, \ldots, n_\alpha$ as follows:

First, $x_{(\alpha,0)} = x_{\alpha-1}$ for each $x \in Q(0, 0) \setminus Z_\alpha$. Then, for $m > 0$ we define $x_{(\alpha,m)} = x_{(\alpha,m-1)}$ for $x \notin Q_m \cup Q'_m$. We note that $x_{(\alpha,m-1)}$, when viewed as a function on $Q(0,0) \setminus Z_\alpha$, is constant at a value (call it z_1, and let y_1 be its length) on Q_m and at a possibly different value (call it z_2, and let y_2 be its length) on Q'_m; without loss of generality we may assume that $y_1 \geq y_2$. There are several cases to consider:

(I) If $y_1 = y_2$ and $z_1 \neq z_2$ we simply define $x_{(\alpha,m)} = x_{(\alpha,m-1)}$ on both Q_m and Q'_m.

(II) If $y_1 = y_2$ and $z_1 = z_2$ we then let $x_{(\alpha,m)}$ be equal to the string created by adding a 0 to the end of $x_{(\alpha,m-1)}$ on Q_m and the string created by adding a 1 to the end of $x_{(\alpha,m-1)}$ on Q'_m.

(III) If $y_1 > y_2$ and z_2 **is not** the first y_2 digits of z_1 we simply define $x_{(\alpha,m)} = x_{(\alpha,m-1)}$ on both Q_m and Q'_m.

(IV) If $y_1 > y_2$ and z_2 **is** the first y_2 digits of z_1, we let define $x_{(\alpha,m)} = x_{(\alpha,m-1)}$ on Q_m; on Q'_m we let y' be the element of $\{0, 1\}$ that **is not** the $(y_2 + 1)$-th digit of z_1 and define $x_{(\alpha,m)}$ on Q'_m to be the string created by adding y' to the end of $x_{(\alpha,m-1)}$.

Once we have finished this process for each cube, we define $x_\alpha = x_{(\alpha,n_\alpha)}$ on $Q(0,0) \setminus Z_\alpha$.

Now, defining Y_n to be the set of all points x such that x_α has length at least n for some α, we conclude from Proposition 2.16 that there exists N such that

$$\Big|\Big\{x \in Q(0,0) \setminus \bigcup_\alpha Z_\alpha : x \in Y_N\Big\}\Big| < \tfrac{1}{100}\delta;$$

we now define the set $Z = \bigcup_\alpha Z_\alpha \cup Y_N$.

If $x \in Q(0,0) \setminus Z$, the sequence $\{x_\alpha\}$ is eventually constant; denote its limiting value by x_∞. Since there are at most 2^n strings of length n, there are at most 2^N possible values of x_∞.

We finish by setting

$$X_w = \{x \in Q(0,0) \setminus Z : x_\infty = w\}$$

whenever w is a string of zeroes and ones of length less than N. For each such w, $F|X_w$ must be bilipschitz (if not, there exist $x_1, x_2 \in X_w$ and a pair of cubes (Q, Q') satisfying the hypotheses of Proposition 2.15 such that $x_1 \in Q$, $x_2 \in Q'$, contradicting the definition of X_w), proving the theorem. □

2F. *Consequences.*

Corollary 2.17. *Suppose A is an open subset of a discretizable Carnot group G (with homogeneous dimension k), H is another Carnot group, and $F : A \to H$ is Lipschitz, and $H^k(F(A)) > 0$. Then there exists a subset $B \subset A$ of positive k-dimensional Hausdorff measure such that F restricted to B is bilipschitz.*

Proof. We can express A as a countable union of translates and dilates of the base cube $Q(0,0)$; by countable additivity of Hausdorff measure one of these cubes, which we call C, is sent by F to a set $F(C)$ with $H^k(F(C)) > 0$. By rescaling we can suppose C is the base cube $Q(0,0)$. The previous theorem divides this cube into the union of a "garbage" set Z (consisting of those cubes whose image has measure too small, as well as those cubes which are in too many bad pairs), where $F(Z)$ can be taken to be arbitrarily small (say, with $h^k(F(Z)) < \frac{1}{2}h^k(F(A))$) and a finite union of sets F_j such that $F|F_j$ is bilipschitz for each j. Since $H^k(F(\bigcup_j F_j)) > 0$, there exists some j where $|F_j| > 0$ and we let $B = F_j$. □

If one assumed that $H^k(A) < \infty$, looking closely at the shape of A would allow us to conclude above that the measure of B and the bilipschitz constant of F would depend only on G, H, A, the Lipschitz coefficient of F, and the k dimensional Hausdorff content of $F(A)$.

Restricting attention to the first Heisenberg group H_1, we use this corollary to show that if we only consider maps whose domains are open, two questions from [Heinonen and Semmes 1997] are equivalent. To begin we need two more definitions.

Definition 2.18. Suppose Q_1 and Q_2 are metric spaces with Hausdorff dimension k. We say that Q_1 *looks down on* Q_2 if there exists a Lipschitz function f from some subset of Q_1 to Q_2 such that the image of f has nonzero Hausdorff k-measure.

Definition 2.19. Suppose Q is a metric space with Hausdorff dimension k. We say that Q is *minimal in looking down* if whenever Q' is a metric space with Hausdorff dimension k such that Q looks down on Q', Q' also looks down on Q.

(Note that this definition is formulated differently from the one in [Heinonen and Semmes 1997].)

Question 22 in [Heinonen and Semmes 1997] asks whether the first Heisenberg group is minimal in looking down and Question 24 asks if every Lipschitz map from H_1 to a metric space with nontrivial Hausdorff 4-measure is bilipschitz on some subset with positive Hausdorff 4-measure.

Clearly 24 implies 22. However, we now know from the corollary that 22 implies 24 when only looking at maps from open sets. This is true because (assuming H_1 is minimal in looking down) if $F : E \subset H_1 \to X$ is Lipschitz and $H^4(F(E)) > 0$ then, letting $G : X \to H_1$ be another Lipschitz map with $H^4(G(X)) > 0$ (and supposing, by restricting images, that $X = F(E)$), $G \circ F$ satisfies the conditions of the corollary and therefore is bilipschitz on some subset $E' \subset E$ with $|E'| > 0$. On this set, we therefore have that F is invertible with inverse $(G \circ F)^{-1} \circ G$, which is clearly Lipschitz, which therefore implies that $F|E'$ is bilipschitz. Because F was arbitrary, we can conclude that Question 24, when restricted to maps defined on open sets, is equivalent to Question 22.

Raanan Schul recently proved a statement corresponding to Question 24 for maps where the domain is Euclidean in [Schul 2009]. In particular, he showed that if F is a Lipschitz function from the k-dimensional unit cube $[0, 1]^k$ into a general metric space, one has the decomposition

$$[0, 1]^k = G \cup \bigcup_{j=1}^{n} F_j,$$

where $F(G)$ has arbitrarily small k-dimensional Hausdorff content and F is bilipschitz on each of the F_j. The main reason why Schul's argument does not generalize to this setting is the dearth of rectifiable curves passing through a given point in a general Carnot group. For example, although the first Heisenberg group has Hausdorff dimension 4, the space of horizontal tangents to rectifiable curves through a given point in that group has dimension two.

We finish this section by discussing the question of Jones-style decompositions for Lipschitz maps on Carnot groups. Just as in [Jones 1988], my argument for the main theorem actually implies the following stronger statement:

Corollary 2.20. *Suppose U is a bounded open subset of a discretizable Carnot group G with Hausdorff dimension Q, H is another Carnot group, $F : U \to H$ is Lipschitz, and $\epsilon > 0$. Then there exists a finite collection $\{A_i\}$ of subsets of U such that each restriction $F|A_i$ is bilipschitz and*

$$h^Q\big(F\big(U \setminus \bigcup_i A_i\big)\big) < \epsilon.$$

For unbounded open subsets of discretizable Carnot groups a diagonalization argument yields the following.

Corollary 2.21. *Suppose U is an open subset of a discretizable Carnot group G with Hausdorff dimension Q, H is another Carnot group and $F : U \to H$ is Lipschitz. Then there exists a countable collection $\{A_i\}$ of subsets of U such that each restriction $F|A_i$ is bilipschitz and*

$$h^Q\big(F\big(U \setminus \bigcup_i A_i\big)\big) = 0.$$

A natural generalization of the above results is in the setting of subriemannian manifolds, defined below.

Definition 2.22. A *subriemannian manifold* is a triple (M, Δ, g) where M is a smooth manifold, Δ is a distribution (i.e., subbundle of the tangent bundle TM) on M which is smooth and satisfies the property that for each $p \in M$, $(TM)_p$ is generated as a Lie algebra by Δ_p, and g is a smooth section of positive-definite quadratic forms on Δ (i.e., g_p defines an inner product on Δ_p which varies smoothly in p).

Recall [Varadarajan 1984] that the set S is said to generate a Lie algebra $\mathfrak{g}$ if the set of finite Lie brackets of elements of S spans $\mathfrak{g}$ as a vector space.

We shall consider M to be naturally equipped with a metric d_{CC} defined as follows: for $x, y \in M$,

$$d_{CC}(x, y) = \inf_{\gamma \in \Gamma_{x,y}} \int_0^1 \sqrt{g(\gamma'(t), \gamma'(t))}\, dt$$

where $\Gamma_{x,y}$ is the family of all curves

$$\gamma : [0, 1] \to M$$

with $\gamma(0) = x$, $\gamma(1) = y$, and $\gamma'(t) \in \Delta_{\gamma(t)}$ for all t.

Now, suppose M and N are subriemannian manifolds such that M is locally bilipschitz equivalent to a discretizable Carnot group G and N is locally bilipschitz equivalent to a Carnot group H. Then Corollary 2.21 still holds if G is replaced by M and H is replaced by N.

For example, M and N could both be ordinary riemannian manifolds. Because riemannian manifolds are locally bilipschitz equivalent to Euclidean spaces, where we have all five properties from Section 2, we can consider arbitrary subsets of M instead of just open subsets. Thus we have the following corollary: if M is a riemannian manifold, $A \subset M$ has Hausdorff dimension k, N is another riemannian manifold, and $F : A \to N$ is Lipschitz with $H^k(F(A)) > 0$, then there exists a subset $B \subset A$ with $H^k(B) > 0$ such that $f|B$ is bilipschitz.

Not all subriemannian manifolds are locally bilipschitz equivalent to Carnot groups, and hence we cannot replace G and H by arbitrary subriemannian manifolds in Corollary 2.21. In particular, we will show in Section 4B that Corollary 2.21 becomes false if G and H are replaced by the Grushin plane and the Euclidean plane, respectively.

3. Hausdorff dimension of Lipschitz images

We begin by observing the following corollary of the results in Section 2.

Corollary 3.1. *Assume A is an open subset of a discretizable Carnot group G with homogeneous dimension k, assume H is another Carnot group, and let $f : A \to H$ be a Lipschitz map such that $H^k(f(A)) > 0$. Then there exists an injective Lie group homomorphism from G to H.*

Proof. By the preceding results, f is bilipschitz on some $B \subset A$ with positive k-dimensional Hausdorff measure. Then the Pansu differential of f at any Lebesgue point of B gives the desired homomorphism. $\square$

Because the converse of this result is trivial (the Lie group homomorphism in question is locally Lipschitz), Corollary 3.1 reduces the question of whether one Carnot group "looks down" on another to a question about the groups' Lie algebras.

An easy consequence of Corollary 3.1 is that if G is a discretizable nonabelian Carnot group with homogeneous dimension k and $U \subset G$ then every Lipschitz image of U in any Euclidean space has zero k-dimensional Hausdorff measure. This follows because there are no injective group homomorphisms from a nonabelian group to an abelian group. In fact, for this consequence we need not assume U is open here because the image space, Euclidean space, has the Lipschitz extension property.

Despite having Hausdorff measure k-measure zero, the Lipschitz image of U in $\mathbb{R}^k$ can still be quite large. For example, we have the following theorem, which answers a question asked by Le Donne (personal communication, 2009):

Theorem 3.2. *Suppose that G is a discretizable Carnot group with homogeneous dimension k, and let $\epsilon > 0$. There exists a bounded open $U \subset G$ and a Lipschitz map $F : U \to \mathbb{R}^k$ such that $H^{k-\epsilon}(F(U)) > 0$.*

Proof. As in our results in Section 2, we illustrate the case $G = H^1$ in detail and remark that the construction is analogous for the general case. The construction is based on the procedure from [Kaufman 1979].
We begin by setting

$$\gamma = 16^{1/(\epsilon-4)},$$

which tells us that $\gamma < \frac{1}{2}$ and $\log_{\gamma^{-1}} 16 = 4 - \epsilon$. We next fix $\beta \in \left[\gamma, \frac{1}{2}\right)$ and define

$$\lambda = \frac{20}{\frac{1}{4} - \beta^2};$$

in particular,

$$\lambda\left(\tfrac{1}{4} - \beta^2\right) = 20 > 10.$$

With this data, we then set our initial box

$$I^0 = [-1, 1] \times [-1, 1] \times [-\lambda, \lambda] \subset H^1$$

and define I^1 to be the union of the sixteen boxes

$$(a, b, c) \cdot \delta_\beta I^0$$

where

$$a \in \left\{-\tfrac{1}{2}, \tfrac{1}{2}\right\}, \quad b \in \left\{-\tfrac{1}{2}, \tfrac{1}{2}\right\}, \quad c \in \left\{-\tfrac{3}{4}\lambda, -\tfrac{1}{4}\lambda, \tfrac{1}{4}\lambda, \tfrac{3}{4}\lambda\right\}.$$

We arbitrarily label these boxes I^1_j for $j = 1, \ldots, 16$.

The point of this construction is to find $\eta > 0$ such that

$$d_{CC}(I_j^1, I_k^1) > \eta \quad \text{for } j \neq k$$

and

$$d_{CC}(I_j^1, \delta(I^0)) > \eta \quad \text{for all } j.$$

Clearly, if two of the boxes in I^1 have different horizontal components, then they are at least $1 - 2\beta$ apart; similarly, every box in I^1 is at a distance of exactly $\frac{1}{2} - \beta$ away from the nearest horizontal edge of I^0.

The only issue is vertical distance. To find the minimum distance between a vertical edge of I^0 and a box in I^1, it suffices to consider a box in I^1 where $c = -\frac{3}{4}\lambda$ and look at the bottom edge of I^0. Every point on the bottom edge of such a box has a vertical coordinate which is at least

$$-\tfrac{3}{4}\lambda - \beta^2\lambda - 2 \cdot \tfrac{1}{2}\beta > -\lambda + 10 - 2 = -\lambda + 8.$$

Now, we recall that if $g = (x_1, y_1, 0)$ and $h = (x_2, y_2, 0)$ are points in H^1 with $x_1, y_1, x_2, y_2 \in [-1, 1]$, then writing the product $g^{-1}h$ as (x_3, y_3, z_3) we note that $|z_3| < 2$.

Consequently, if $p = (p_1, p_2, p_3)$ is a point in I^1 and $q = (q_1, q_2, -\lambda)$ is a point on the bottom edge of I^0, we note that the vertical coordinate of $p^{-1}q$ is at most

$$-(-\lambda + 8) - \lambda + 2 = -6,$$

which implies that vertical edges of I^0 will be separated from boxes in I^1 by at least 6 units.

Similarly, looking at two boxes in I^1 with the same horizontal component (e.g., let A be such a box with $c = -\frac{3}{4}\lambda$ and B be such a box with $c = -\frac{1}{4}\lambda$), the vertical coordinate of points in A are at most $-\frac{1}{2}\lambda - 8$ and the vertical coordinate of points in B are at least $-\frac{1}{2}\lambda + 8$. Therefore, whenever $a \in A$ and $b \in B$, the vertical coordinate of $a^{-1}b$ is at least

$$\left(-\tfrac{1}{2}\lambda + 8\right) - \left(-\tfrac{1}{2}\lambda - 8\right) - 2 = 14,$$

implying a separation of 14 between any two such boxes.

In subsequent stages we replace each box of the form

$$p \cdot \delta_\mu I^0$$

(there are 16^k such boxes in stage k; at this stage $\mu = \beta^k$) with the sixteen boxes

$$p \cdot \delta_\mu(a, b, c) \cdot \delta_{\beta\mu}(I^0)$$

and denote by I^k the union of all the boxes produced in stage k.

In stage k, each box has a label of the form $I^k_{(a_1,\dots,a_k)}$ where each a_i ranges from 1 to 16; we extend this process to stage $k+1$ by labeling the subboxes from $I^k_{(a_1,\dots,a_k)}$ as $I^{k+1}_{(a_1,\dots,a_k,v)}$ where $v = 1, 2, \dots, 16$. The intersection of the I^k's, to be defined as I, is a Cantor set in H^1 of dimension

$$\log_{\beta^{-1}} 16 \geq 4 - \epsilon.$$

Each point $x \in I$ has a unique label of the form $(a_1, \dots, a_n, \dots)$ (where each a_i ranges from 1 to 16) such that for each $n \in \mathbb{N}$, $x \in I^n_{(a_1,\dots,a_n)}$; if $v = (a_1, \dots, a_n, \dots)$ and $w = (b_1, \dots, b_n, \dots)$ with m being the smallest integer where $a_m \neq b_m$, the distance between the points corresponding to v and w is (up to a multiplicative constant independent of m) equal to β^m.

Similarly, we set J^0 to be the box $[-1, 1]^4$ in Euclidean space $\mathbb{R}^4$ and J^1 to be the union of the sixteen boxes

$$(a, b, c, d) + \gamma I^0$$

where a, b, c, d can each equal $-\frac{1}{2}$ or $\frac{1}{2}$. We arbitrarily label these boxes J^1_j for $j = 1, \dots, 16$.

The point of this construction is now to find $\eta' > 0$ such that

$$d(J^1_j, J^1_k) > \eta' \quad \text{for } j \neq k$$

and

$$d(J^1_j, \delta(J^0)) > \eta' \quad \text{for all } j,$$

where the distance above is Euclidean.

Clearly, any two of the boxes in J^1 are at least $1 - 2\gamma$ apart; similarly, each such box is at a distance of exactly $\frac{1}{2} - \gamma$ away from the boundary of J^0.

In subsequent stages we replace the box

$$p + \nu J^0$$

with the sixteen boxes

$$p + \nu((a, b, c, d) + \gamma J^0)$$

and denote by J^k the union of all boxes produced in stage k. Note that at stage k, $\nu = \gamma^k$.

In stage k, each box has a label of the form $J^k_{(a_1,\dots,a_k)}$ where each a_i ranges from 1 to 16; we extend this process to stage $k+1$ by labeling the subboxes from $J^k_{(a_1,\dots,a_k)}$ as $J^{k+1}_{(a_1,\dots,a_k,v)}$ where $v = 1, 2, \dots, 16$. The intersection of the J^k's, to be defined as J, is a Cantor set in $\mathbb{R}^4$ of dimension

$$\log_{\gamma^{-1}} 16 = 4 - \epsilon.$$

Each point $x \in J$ has a unique label of the form $(a_1, \dots, a_n, \dots)$ (where each a_i ranges from 1 to 16) such that for each $n \in \mathbb{N}$, $x \in J^n_{(a_1,\dots,a_n)}$; if $v = (a_1, \dots, a_n, \dots)$ and $w = (b_1, \dots, b_n, \dots)$ with m being the smallest integer where $a_m \neq b_m$, the distance between the points corresponding to v and w is (up to a multiplicative constant independent of m) equal to γ^m.

We can define a Lipschitz map F from $I^0 \subset H^1$ to $\mathbb{R}^4$ whose image contains J (and therefore has Hausdorff dimension $\log_{\gamma^{-1}}(16)$) via the following three-step process.

Step 1. Map I to J. This is done by mapping a point in I with a label of the form $(a_1, \dots, a_n, \dots)$ to the point with the same label in J. By construction, one notes that if $\beta = \gamma$ then this map is bilipschitz.

Step 2. For each ordered n-tuple $(a_1, \dots, a_n)$ with each a_i in $\{1, \dots, 16\}$ (this includes the zero-tuple, where we would be mapping the boundary of I^0) we choose a point $p_{(a_1,\dots,a_n)}$ in $J^n_{(a_1,\dots,a_n)}$ and then send all of the points in the boundary of $I^n_{(a_1,\dots,a_n)}$ to $p_{(a_1,\dots,a_n)}$.

Step 3. The remaining region of I^0 consists of sets of the form $S^n_{(a_1,\dots,a_n)}$ defined as the set of all points in $I^n_{(a_1,\dots,a_n)}$ which do not lie in $I^{n+1}_{(a_1,\dots,a_n,v)}$ for $v = 1, 2, \dots, 16$. The closure of this region includes the boundary of $I^n_{(a_1,\dots,a_n)}$ and of $I^{n+1}_{(a_1,\dots,a_n,v)}$ for $v = 1, \dots, 16$. Fixing $(a_1, \dots, a_n)$ (we may work on each $S^n_{(a_1,\dots,a_n)}$ separately) we define the map f from the interval $[0, 16]$ to $\mathbb{R}^4$ to be a smooth function sending 0 to $p_{(a_1,\dots,a_n)}$ and $v = 1, \dots, 16$ to $p_{(a_1,\dots,a_n,v)}$. We can suppose f has Lipschitz norm comparable to γ^n. We then define g to be a smooth, real-valued, Lipschitz function (with Lipschitz coefficient comparable to β^{-n}) on the closure of $S^n_{(a_1,\dots,a_n)}$ which sends the boundary of $I^n_{(a_1,\dots,a_n)}$ to 0 and the boundary of $I^{n+1}_{(a_1,\dots,a_n,v)}$ to v. We can create such a g by the Whitney extension theorem (the construction is more straightforward if we do not require smoothness). On the closure of $S^n_{(a_1,\dots,a_n)}$ (the construction merely repeats the existing one on the boundary) set $F = f \circ g$; then $F|\overline{S^n_{(a_1,\dots,a_n)}}$ has Lipschitz norm comparable to $(\frac{\gamma}{\beta})^n$.

Note that if $\gamma < \beta$, $(\frac{\gamma}{\beta})^n$ goes to zero as n goes to infinity, which means that F is differentiable (in the Pansu sense) at each point of I with derivative zero. Further, by construction F is C^1 outside of I where the Pansu differential always has rank zero or one (and this differential approaches zero as we approach points of I); in fact, it is locally constant near the boundaries of the relevant cubes if we use the Whitney extension, so the construction here is indeed an appropriate analogue of [Kaufman 1979]. □

In fact, because the constructed map is constant on the boundary of I^0, nesting appropriately rescaled examples of this form inside each other yield the following corollary.

Corollary 3.3. *Suppose that G is a discretizable Carnot group with homogeneous dimension k. There exists a bounded open $U \subset G$ and a Lipschitz map $F : U \to \mathbb{R}^k$ such that $F(U)$ has Hausdorff dimension k.*

4. Counterexamples

In this section we develop two counterexamples to show why Carnot group structure, or something close to it, is necessary for the results of the previous two sections.

4A. *A space-filling curve.*

Theorem 4.1. *There exists an Ahlfors 2-regular metric space X and a Lipschitz map $F : X \to \mathbb{R}^2$ such that $F(X)$ has positive 2-dimensional Hausdorff measure but F is not bilipschitz on any set of positive 2-dimensional measure.*

Proof. The function in question will be the space-filling curve F from $[0, 1]$ (equipped with the square root distance metric) to the unit square in $\mathbb{R}^2$ mentioned in Section 7.3 of [Stein and Shakarchi 2005]. Although this function is a surjective map of spaces with Hausdorff dimension 2 and Lipschitz, it is not bilipschitz on any subset with positive Hausdorff 2-measure. To see this, suppose that the space-filling curve F is bilipschitz on a set A with $H^2(A) > 0$. As $F(A)$ has positive Lebesgue measure, it contains a point x of Lebesgue density one. Letting $\epsilon > 0$ there exists $\delta > 0$ such that

$$|B(x;\delta) \cap F(A)| > (1-\epsilon)|B(x;\delta)|.$$

Writing out the binary expansion of the components of x and of δ, $B(x;\delta)$ contains a dyadic cube Q of side at least $\frac{1}{10}\delta$; since $\epsilon|B(x;\delta)| \le 1000\epsilon|Q|$, we have

$$|Q \cap F(A)| > (1-1000\epsilon)|Q|.$$

As F is measure-preserving, letting J be the preimage of Q we conclude

$$|J \cap A| > (1-1000\epsilon)|J|.$$

By rescaling and translating we can suppose F is therefore bilipschitz on a set A of Hausdorff 2-measure arbitrarily close to 1 (although the rescaled F is not identical to our space-filling curve, it preserves all the relevant properties, such as being Lipschitz in the appropriate metric, measure-preserving, and sending a pair of points whose "square root" distance is at least $\frac{1}{2}$ to the same point).

Let x, x' be two points which are at least $\frac{1}{4}$ apart in Euclidean distance (and therefore $\frac{1}{2}$ away with respect to square root distance) such that $F(x) = F(x')$. We can suppose that $y, y' \in A$ are arbitrarily close to x, x' respectively; therefore, $|y - y'| \ge \frac{1}{4}$; however,

$$|F(y) - F(y')| \le |F(x) - F(y)| + |F(x') - F(y')|$$

which can be made arbitrarily small by the Lipschitz property (all distances use the square root metric in the domain and the Euclidean metric in the image) showing that F cannot be bilipschitz on A with any coefficient. □

In this example, the third and fourth properties (involving differentiability) from Section 2C fail. This suggests that some notion of differentiability is necessary for the results in [Jones 1988] to extend to other spaces.

4B. ***The Grushin plane.***

Theorem 4.2. *There exists a* 2*-dimensional subriemannian manifold* M *with Hausdorff dimension* 2, *an open* $U \subset M$, *and a Lipschitz map*

$$F : U \to \mathbb{R}^2$$

which is not decomposable in the following sense: *There does not exist a countable collection* $\{A_i\}$ *of sets such that*

$$H^2\big(F\big(U \setminus \bigcup_i A_i\big)\big) = 0$$

and $F|A_i$ *is bilipschitz for each* i.

Proof. We use the Grushin plane M as our subriemannian manifold.

To construct the Grushin plane we define a riemannian metric on the following region of $\mathbb{R}^2$: $\{(x, y) : y \neq 0\}$.

This metric is defined as $ds^2 = dx^2 + x^{-2}dy^2$. We then use this metric to induce a geodesic structure on all of $\mathbb{R}^2$, where a rectifiable curve must have horizontal tangent at each point that it crosses the y-axis.

One can observe that off of the vertical axis, the Grushin plane is locally bilipschitz to Euclidean space (but with a constant that blows up as we get closer to the axis). However, the distance between two points on the vertical axis is proportional to the square root of their Euclidean distance.

In other words, the Grushin plane is a union of a (disconnected) riemannian manifold and a line of Hausdorff dimension two, making it a subriemannian manifold of both Euclidean and Hausdorff dimension two.

To construct our counterexample, we consider an open neighborhood of the segment S joining $(0, 0)$ to $(0, 1)$, say: $U_\epsilon = (-\epsilon, \epsilon) \times (-\epsilon, 1 + \epsilon)$ for $\epsilon > 0$. The space-filling curve previously constructed as in Chapter 7 of [Stein and Shakarchi 2005] has already been shown to be Lipschitz when defined as a function from a set which is bilipschitz to S with image the unit square. We can extend this mapping to a Lipschitz mapping F from U_ϵ to $\mathbb{R}^2$ by standard constructions (note the importance of having a Euclidean target space here).

However, there does not exist a countable collection of sets $A_1, \dots, A_n, \dots$ such that $G := U_\epsilon \setminus \bigcup_n A_n$ is sent to a set of arbitrarily small Hausdorff content by F and

F is bilipschitz when restricted to the A_n. This is because $A_n \cap S$ must be a nullset (by the previous arguments concerning the space-filling curve for each G) which implies that G must contain almost all of S, in the sense of Hausdorff measure. Therefore, $F(G)$ must contain almost all of the unit square in the sense of Hausdorff measure (or Hausdorff content, which is equivalent in this case), producing our desired contradiction. $\square$

In this example, the first and second properties (involving homogeneity) from Section 2C fail, which suggests that some notion of homogeneity is also necessary for the results in [Jones 1988] to extend to other spaces.

Acknowledgements

The author would like to thank Raanan Schul for showing him Guy David's proof of the decomposition of Lipschitz functions from subsets of $\mathbb{R}^m$ to $\mathbb{R}^n$ into bilipschitz pieces. Further, the author would like to thank Jeremy Tyson for showing him the usefulness of Cantor set constructions for creating maps on Carnot groups.

References

[Balogh and Fässler 2009] Z. M. Balogh and K. S. Fässler, "Rectifiability and Lipschitz extensions into the Heisenberg group", *Math. Z.* **263**:3 (2009), 673–683. MR 2010j:53049 Zbl 1177.53032

[Brudnyi and Brudnyi 2007] A. Brudnyi and Y. Brudnyi, "Linear and nonlinear extensions of Lipschitz functions from subsets of metric spaces", *Algebra i Analiz* **19**:3 (2007), 106–118. MR 2008m:26004 Zbl 1213.54040

[Capogna et al. 2007] L. Capogna, D. Danielli, S. D. Pauls, and J. T. Tyson, *An introduction to the Heisenberg group and the sub-Riemannian isoperimetric problem*, Progress in Mathematics **259**, Birkhäuser, Basel, 2007. MR 2009a:53053 Zbl 1138.53003

[Christ 1990] M. Christ, "A $T(b)$ theorem with remarks on analytic capacity and the Cauchy integral", *Colloq. Math.* **60/61**:2 (1990), 601–628. MR 92k:42020 Zbl 0758.42009

[David 1988] G. David, "Morceaux de graphes lipschitziens et intégrales singulières sur une surface", *Rev. Mat. Iberoam.* **4**:1 (1988), 73–114. MR 90h:42026 Zbl 0696.42011

[David 1991] G. David, *Wavelets and singular integrals on curves and surfaces*, Lecture Notes in Mathematics **1465**, Springer, Berlin, 1991. MR 92k:42021 Zbl 0764.42019

[David and Semmes 1993] G. David and S. Semmes, "Quantitative rectifiability and Lipschitz mappings", *Trans. Amer. Math. Soc.* **337**:2 (1993), 855–889. MR 93h:42015 Zbl 0792.49029

[Fässler 2007] K. Fässler, "Extending Lipschitz maps from Euclidean spaces into Heisenberg groups", Master's thesis, Universität Bern, 2007.

[Heinonen and Semmes 1997] J. Heinonen and S. Semmes, "Thirty-three yes or no questions about mappings, measures, and metrics", *Conform. Geom. Dyn.* **1** (1997), 1–12. MR 99h:28012 Zbl 0885.00006

[Jones 1988] P. W. Jones, "Lipschitz and bi-Lipschitz functions", *Rev. Mat. Iberoam.* **4**:1 (1988), 115–121. MR 90h:26016 Zbl 0782.26007

[Kaufman 1979] R. Kaufman, "A singular map of a cube onto a square", *J. Differential Geom.* **14**:4 (1979), 593–594. MR 82a:26013 Zbl 0463.57011

[Magnani 2010] V. Magnani, "Contact equations, Lipschitz extensions and isoperimetric inequalities", *Calc. Var. Partial Differential Equations* **39**:1-2 (2010), 233–271. MR 2012g:49100 Zbl 1195.49053

[Montgomery 2002] R. Montgomery, *A tour of subRiemannian geometries, their geodesics and applications*, Mathematical Surveys and Monographs **91**, Amer. Math. Soc., Providence, RI, 2002. MR 2002m:53045 Zbl 1044.53022

[Pansu 1989] P. Pansu, "Métriques de Carnot–Carathéodory et quasiisométries des espaces symétriques de rang un", *Ann. of Math.* (2) **129**:1 (1989), 1–60. MR 90e:53058 Zbl 0678.53042

[Rigot and Wenger 2010] S. Rigot and S. Wenger, "Lipschitz non-extension theorems into jet space Carnot groups", *Int. Math. Res. Not.* **2010**:18 (2010), 3633–3648. MR 2011f:53056 Zbl 1203.53028

[Schul 2009] R. Schul, "Bi-Lipschitz decomposition of Lipschitz functions into a metric space", *Rev. Mat. Iberoam.* **25**:2 (2009), 521–531. MR 2011e:28007 Zbl 1228.28004

[Semmes 2000] S. Semmes, "Measure-preserving quality within mappings", *Rev. Mat. Iberoam.* **16**:2 (2000), 363–458. MR 2001m:30029 Zbl 1041.42014

[Stein and Shakarchi 2005] E. M. Stein and R. Shakarchi, *Real analysis: measure theory, integration, and Hilbert spaces*, Princeton Lectures in Analysis **3**, Princeton University Press, Princeton, NJ, 2005. MR 2005k:28024 Zbl 1081.28001

[Strichartz 1992] R. S. Strichartz, "Self-similarity on nilpotent Lie groups", pp. 123–157 in *Geometric analysis* (Philadelphia, PA, 1991), edited by E. L. Grinberg, Contemp. Math. **140**, Amer. Math. Soc., Providence, RI, 1992. MR 94e:43011 Zbl 0797.43004

[Varadarajan 1984] V. S. Varadarajan, *Lie groups, Lie algebras, and their representations*, Graduate Texts in Mathematics **102**, Springer, New York, 1984. MR 85e:22001 Zbl 0955.22500

[Vodopyanov and Ukhlov 1996] S. K. Vodopyanov and A. D.-O. Ukhlov, "Approximately differentiable transformations and change of variables on nilpotent groups", *Sibirsk. Mat. Zh.* **37**:1 (1996), 70–89. In Russian; translated in *Siberian Math. J.* **37**:1 (1996), 62–78. MR 97f:22011 Zbl 0870.43005

[Wenger and Young 2010] S. Wenger and R. Young, "Lipschitz extensions into jet space Carnot groups", *Math. Res. Lett.* **17**:6 (2010), 1137–1149. MR 2011j:53048 Zbl 1222.53037

Received March 27, 2012. Revised August 14, 2012.

WILLIAM MEYERSON
DEPARTMENT OF MATHEMATICS AND STATISTICS
HELSINGIN YLIOPISTO
FI-00014 HELSINKI
FINLAND

william.meyerson@helsinki.fi

PACIFIC JOURNAL OF MATHEMATICS
Vol. 263, No. 1, 2013
dx.doi.org/10.2140/pjm.2013.263.171

GEOMETRIC INEQUALITIES IN CARNOT GROUPS

FRANCESCOPAOLO MONTEFALCONE

Let $\mathbb{G}$ be a subriemannian k-step Carnot group of homogeneous dimension Q. We prove several geometric inequalities concerning smooth hypersurfaces (submanifolds of codimension one) immersed in $\mathbb{G}$, endowed with the H-perimeter measure.

1. Introduction

During the last years there has been an increasing interest in studying analysis and geometric measure theory in metric spaces (see [Ambrosio 2001; Ambrosio and Kirchheim 2000a; 2000b; Cheeger and Kleiner 2010; David and Semmes 1997; Garofalo and Nhieu 1996; Varopoulos et al. 1992] and bibliographic references therein, but this list is far from being exhaustive). In this regard, important examples of highly noneuclidean geometries are represented by the so-called Carnot–Carathéodory (or subriemannian) geometries; see [Capogna et al. 1994; Gromov 1996; Montgomery 2002; Pansu 1982; 1989; 2005; Strichartz 1986; Vershik and Gershkovich 1994]. In this context, Carnot groups play the role of modeling the tangent space (in a suitable generalized sense, which is related to the Gromov–Hausdorff convergence) of a subriemannian manifold; see [Gromov 1996; Montgomery 2002]. For this and many other reasons, Carnot groups are an intriguing field of research; see [Ambrosio et al. 2006; Balogh 2003; Balogh et al. 2009; Capogna et al. 2010; Cheng et al. 2005; Danielli et al. 2007; 2010; Franchi

Partially supported by the Fondazione CaRiPaRo Project "Nonlinear partial differential equations: models, analysis, and control-theoretic problems".
MSC2010: primary 49Q15; secondary 46E99, 43A80.
Keywords: Carnot groups, subriemannian geometry, hypersurfaces, geometric inequalities.

et al. 2001; 2003a; 2003b; 2007; Hladky and Pauls 2008; Magnani 2002; Magnani and Vittone 2008; Montefalcone 2005; 2007a; Ritoré and Rosales 2008].

A *k-step Carnot group* $(\mathbb{G}, \bullet)$ is an n-dimensional, connected, simply connected, nilpotent, stratified Lie group (with respect to the group multiplication $\bullet$) whose Lie algebra $\mathfrak{g} \cong \mathbb{R}^n$ satisfies

$$\mathfrak{g} = H_1 \oplus \cdots \oplus H_k, \quad [H_1, H_{i-1}] = H_i \quad (i = 2, \dots, k), \quad H_{k+1} = \{0\}.$$

We assume that $h_i = \dim H_i$ $(i = 1, \dots, k)$ so that $n = \sum_{i=1}^{k} h_i$. Any Carnot group $\mathbb{G}$ has a 1-parameter family of dilations, adapted to the stratification, that makes it a *homogeneous group*, in the sense of Stein's definition [1993]. We refer the reader to Section 1.1 for a more detailed introduction to Carnot groups.

In this paper, we shall prove some geometric inequalities concerning smooth hypersurfaces immersed in a subriemannian k-step Carnot group $\mathbb{G}$ of homogeneous dimension $Q := \sum_{i=1}^{k} i h_i$. We have to stress that hypersurfaces will be endowed with the so-called H-perimeter measure σ_H^{n-1}, which is a natural substitute for the intrinsic $(Q-1)$-dimensional CC Hausdorff measure. In Section 1.2, we will discuss some preliminary notions concerning homogeneous measures and the horizontal geometry of hypersurfaces. Then we will recall some tools which will be important in the sequel, such as a coarea-type formula and the horizontal integration by parts theory; see Section 1.3.

In Section 2 we will extend to this setting some isoperimetric-type constants, introduced in [Cheeger 1970] for compact riemannian manifolds and later studied in [Yau 1975].

In particular, we shall prove the validity of some global inequalities for smooth compact hypersurfaces with (or without) boundary, immersed into $\mathbb{G}$. Here, we would like to remark that there is a strong relationship between these inequalities and some eigenvalue problems related to the second-order differential operator $\mathcal{L}_{HS}$ (which is nothing but a horizontal version of the Laplace–Beltrami operator); see, more precisely, Definition 21 in Section 1.2.

Roughly speaking, after defining the isoperimetric constants (in purely geometric terms), we will show that they are equal to the infimum of some Rayleigh quotients. More precisely, let $S \subset \mathbb{G}$ be a smooth hypersurface and assume $\partial S \neq \varnothing$. Furthermore, set

$$\mathrm{Isop}(S) := \inf \frac{\sigma_H^{n-2}(N)}{\sigma_H^{n-1}(S_1)},$$

where $N \subset S$ is a smooth hypersurface of S such that $N \cap \partial S = \varnothing$ and S_1 is the unique $(n-2)$-dimensional submanifold of S such that $N = \partial S_1$. We have to stress that σ_H^{n-1} and σ_H^{n-2} denote homogeneous measures on S_1 and N, respectively. These measures can be thought of, respectively, as the $(Q-1)$-dimensional and the $(Q-2)$-dimensional CC Hausdorff measures on S_1 and N; see Section 1.2. Then, it will be shown that

$$\operatorname{Isop}(S) = \inf \frac{\int_S |\operatorname{grad}_{HS} \psi| \, \sigma_H^{n-1}}{\int_S |\psi| \, \sigma_H^{n-1}},$$

where the infimum is taken over suitably smooth functions on S such that $\psi|_{\partial S} = 0$. As mentioned, this constant is related to the first nonzero eigenvalue λ_1 of the following Dirichlet-type problem:

$$\begin{cases} -\mathcal{L}_{HS}\psi = \lambda\psi, \\ \quad \psi|_{\partial S} = 0; \end{cases}$$

see Definition 21. Indeed, we shall see that

$$\lambda_1 \geq \frac{(\operatorname{Isop}(S))^2}{4};$$

see Corollary 28. Some similar results concerning another isoperimetric constant will be proved; see Theorem 30 and Corollary 31. The proofs of these results follow the scheme of the riemannian case, for which we refer the reader to [Yau 1975]; see also [Cheeger 1970] and [Chavel 1984; 1993]. We also remark that the main technical tool in the original proofs is the coarea formula.

In Section 3 we shall prove two geometric inequalities involving volume, H-perimeter and the first eigenvalue of the operator $\mathcal{L}_{HS}$ on S. These results generalize an inequality of Chavel [1978] and an inequality of Reilly [1977], respectively.

In Section 4 we will extend to the Carnot setting some classical differential-geometric results (such as linear isoperimetric inequalities); see, for instance, [Burago and Zalgaller 1988] and references therein. The starting point is an integral formula very similar to the euclidean Minkowski formula; see Corollary 20 for a precise statement. In particular, we will show that

$$(h-1)\, \sigma_H^{n-1}(S) \leq R\biggl(\int_S \bigl(|\mathcal{H}_H| + |C_H \nu_H|\bigr)\, \sigma_H^{n-1} + \sigma_H^{n-2}(\partial S) \biggr),$$

where $S \subset \mathbb{G}$ is a compact hypersurface with boundary and R denotes the radius of a homogeneous ϱ-ball circumscribed about S. From this linear (isoperimetric) inequality, it is possible to infer some geometric consequences and, among them, we prove a weak monotonicity inequality for the H-perimeter; see Section 4.1, Proposition 38.

Section 5 contains a theorem about nonhorizontal graphs in 2-step Carnot groups. This generalizes a classical result of Heinz [1955]; see also [Chern 1965].

Let us describe this result in the simpler case of the Heisenberg group $\mathbb{H}^1$. So let $S \subset \mathbb{H}^1$ be a T-graph associated with a function $t = f(x, y)$ of class C^2 over the

xy-plane. If the horizontal mean curvature $\mathcal{H}_H$ of S satisfies a bound $|\mathcal{H}_H| \geq C > 0$, then

$$C\mathcal{H}^2_{\mathrm{Eu}}(\mathcal{P}_{xy}(\mathcal{U})) \leq \mathcal{H}^1_{\mathrm{Eu}}(\mathcal{P}_{xy}(\partial\mathcal{U}))$$

for every C^1-smooth relatively compact open set $\mathcal{U} \subset S$, where $\mathcal{H}^i_{\mathrm{Eu}}$ ($i = 1, 2$) is the usual i-dimensional euclidean Hausdorff measure and $\mathcal{P}_{xy}$ is the orthogonal projection onto the xy-plane. Hence, taking $\mathcal{U} := S \cap C_r(\mathcal{T})$, where $C_r(\mathcal{T})$ denotes a vertical cylinder of radius r around the T-axis of $\mathbb{H}^1$, yields

$$r \leq \frac{2}{C}$$

for every $r > 0$. It follows that any entire xy-graph of class C^2, having either constant or only bounded horizontal mean curvature $\mathcal{H}_H$, must be necessarily a H-minimal surface. An analogous result holds true in the framework of step 2 Carnot groups; see Theorem 42.

In Section 6 we shall study some (local) Poincaré-type inequalities, depending on the local geometry of the hypersurface S and, more precisely, on its characteristic set C_S; see Theorems 44 and 45.

For instance, let $S \subset \mathbb{G}$ be a C^2-smooth hypersurface with bounded horizontal mean curvature $\mathcal{H}_H$. Then, we shall prove that for every $x \in S$ there exists $R_0 \leq \mathrm{dist}_\varrho(x, \partial S)$ (which explicitly depends on C_S) such that

$$\left(\int_{S_R} |\psi|^p\, \sigma_H^{n-1}\right)^{\frac{1}{p}} \leq C_p R\left(\int_{S_R} |\mathrm{grad}_{HS}\,\psi|^p\, \sigma_H^{n-1}\right)^{\frac{1}{p}}, \quad p \in [1, +\infty[,$$

for all $\psi \in C^1_0(S_R)$ and all $R \leq R_0$, where $S_R := S \cap B_\varrho(x, R)$.

These results are obtained by means of elementary "linear" estimates starting from the horizontal integration by parts formula, together with a simple analysis of the role played by the characteristic set. Finally, in Section 6.1 we will prove the validity of a Caccioppoli-type inequality for weak solutions of the operator $\mathcal{L}_{HS}$.

1.1. *Carnot groups.* A *k-step Carnot group* $(\mathbb{G}, \bullet)$ is a finite-dimensional connected, simply connected, nilpotent and stratified Lie group with respect to a polynomial group law $\bullet$. The Lie algebra $\mathfrak{g} \cong \mathbb{R}^n$ fulfills the conditions $\mathfrak{g} = H_1 \oplus \cdots \oplus H_k$, $[H_1, H_{i-1}] = H_i$ for all $i = 2, \ldots, k+1$, $H_{k+1} = \{0\}$, where $[\cdot, \cdot]$ denotes the Lie bracket and each H_i is a vector subspace of $\mathfrak{g}$. In particular, we denote by 0 the identity of $\mathbb{G}$ and assume that $\mathfrak{g} \cong T_0\mathbb{G}$. We also use the notation $H := H_1$ and $V := H_2 \oplus \cdots \oplus H_k$. The subspaces H and V are smooth subbundles of $T\mathbb{G}$ called *horizontal* and *vertical*, respectively.

Notation 1. Throughout this paper, we denote by $\mathcal{P}_{H_i} : T\mathbb{G} \to H_i$ the orthogonal projection map from $T\mathbb{G}$ onto H_i for any $i = 1, \ldots, k$. In particular, we set

$\mathcal{P}_H := \mathcal{P}_{H_i}$. Analogously, we set $\mathcal{P}_V : T\mathbb{G} \to V$ to denote the orthogonal projection map from $T\mathbb{G}$ onto V.

Let $h_i := \dim H_i$ for any $i = 1, \dots, k$. Set $n_0 := 0$ and $n_i := \sum_{j=1}^{i} h_j$ for any $i = 1, \dots, k$. Note that $n_1 = h_1$, $n_2 = h_1 + h_2$, ..., and $n_k = n$.

Notation 2. Throughout this paper, we set $I_{H_i} := \{n_{i-1} + 1, \dots, n_i\}$ for any $i = 1, \dots, k$. We also set $I_V := \{h_1 + 1, \dots, n\}$ and use Greek letters $\alpha, \beta, \gamma, \dots,$ for indices in I_V. For the sake of simplicity, we set $h := h_1$ and $I_H := I_{H_1}$.

The horizontal bundle H is generated by a frame $\underline{X_H} := \{X_1, \dots, X_h\}$ of left-invariant vector fields. This frame can be completed to a global graded, left-invariant frame $\underline{X} := \{X_1, \dots, X_n\}$ for $T\mathbb{G}$. Note that the standard basis $\{\mathrm{e}_i : i = 1, \dots, n\}$ of $\mathbb{R}^n$ can be relabeled to be *graded* or *adapted to the stratification*. Any left-invariant vector field of the frame $\underline{X}$ is given by $X_i(x) = L_{x*}\mathrm{e}_i$ $(i = 1, \dots, n)$, where L_{x*} denotes the differential of the left-translation L_x, defined by $L_x y := x \bullet y$ for all $y \in \mathbb{G}$. We also fix a euclidean metric on $\mathfrak{g} = T_0\mathbb{G}$ such that $\{\mathrm{e}_i : i = 1, \dots, n\}$ is an orthonormal basis. This metric $g = \langle \cdot, \cdot \rangle$ extends to the whole tangent bundle by left-translations and makes $\underline{X}$ an orthonormal left-invariant frame. Therefore $(\mathbb{G}, g)$ is a riemannian manifold.

Let $\exp : \mathfrak{g} \to \mathbb{G}$ be the exponential map. Hereafter, we will use *exponential coordinates of the first kind*; see [Varadarajan 1974, Chapter 2, p. 88].

As for the case of nilpotent Lie groups, the multiplication $\bullet$ of $\mathbb{G}$ is uniquely determined by the "structure" of the Lie algebra $\mathfrak{g}$. This is the content of the *Baker–Campbell–Hausdorff formula*; see [Corwin and Greenleaf 1990]. More precisely,

$$\exp(X) \bullet \exp(Y) = \exp(X \star Y) \quad \text{for all } X, Y \in \mathfrak{g},$$

where $\star : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ denotes the *Baker–Campbell–Hausdorff product*, given by

$$X \star Y = X + Y + \tfrac{1}{2}[X, Y] + \tfrac{1}{12}[X, [X, Y]] - \tfrac{1}{12}[Y, [X, Y]] + \text{brackets of length} \geq 3. \tag{1}$$

Using exponential coordinates and (1), the group multiplication $\bullet$ turns out to be polynomial and explicitly computable; see [Corwin and Greenleaf 1990]. Moreover, $0 = \exp(0, \dots, 0)$ and the inverse of $x \in \mathbb{G}$ $(x = \exp(x_1, \dots, x_n))$ is $x^{-1} = \exp(-x_1, \dots, -x_n)$.

A *subriemannian metric* g_H is a symmetric positive bilinear form on the horizontal bundle H. The *CC-distance* $d_{CC}(x, y)$ between $x, y \in \mathbb{G}$ is given by

$$d_{CC}(x, y) := \inf \int \sqrt{g_H(\dot{\gamma}, \dot{\gamma})}\, dt,$$

where the infimum is taken over all piecewise-smooth horizontal paths γ joining x to y. Later, we shall choose $g_H := g_{|H}$.

Carnot groups are *homogeneous groups*; that is, they admit a 1-parameter group of automorphisms $\delta_t : \mathbb{G} \to \mathbb{G}$ $(t \geq 0)$ defined by $\delta_t x := \exp\bigl(\sum_{j,i_j} t^j x_{i_j} \mathrm{e}_{i_j}\bigr)$, where $x = \exp\bigl(\sum_{j,i_j} x_{i_j} \mathrm{e}_{i_j}\bigr) \in \mathbb{G}$. As already said, the *homogeneous dimension* of $\mathbb{G}$ is the integer $Q := \sum_{i=1}^{k} i h_i$ coinciding with the *Hausdorff dimension* of $(\mathbb{G}, d_{\mathrm{CC}})$ as a metric space; see [Montgomery 2002].

We recall that a continuous distance $\varrho : \mathbb{G} \times \mathbb{G} \to \mathbb{R}_+ \cup \{0\}$ is a *homogeneous distance* if, and only if,

$$\varrho(x, y) = \varrho(z \bullet x, z \bullet y) \ \text{ for all } x, y, z \in \mathbb{G}; \quad \varrho(\delta_t x, \delta_t y) = t\varrho(x, y) \ \text{ for all } t \geq 0.$$

The *structural constants* of $\mathfrak{g}$ (see [Chavel 1993]) associated with the frame $\underline{X}$ are defined by $C_{ij}^r := \langle [X_i, X_j], X_r \rangle$ for all $i, j, r = 1, \ldots, n$. They are skew-symmetric and satisfy Jacobi's identity. The stratification of the Lie algebra $\mathfrak{g}$ implies a fundamental "structural" property of Carnot groups: if $X_i \in H_l$, $X_j \in H_m$, then $[X_i, X_j] \in H_{l+m}$. Note that, if $i \in I_{H_s}$ and $j \in I_{H_r}$, then

$$C_{ij}^m \neq 0 \Longrightarrow m \in I_{H_{s+r}}. \tag{2}$$

Equivalently, if $C_{ij}^r \neq 0$, then $\mathrm{ord}(i) + \mathrm{ord}(j) = \mathrm{ord}(r)$, where $\mathrm{ord} : \{1, \ldots, n\} \to \{1, \ldots, k\}$ is the function defined as $\mathrm{ord}(l) = i \iff l \in I_{H_i}$.

Notation 3. Henceforth, we shall set

- $C_H^\alpha := [C_{ij}^\alpha]_{i,j=1,\ldots,h} \in \mathcal{M}_{h \times h}(\mathbb{R})$ for all $\alpha \in I_{H_2} = \{h+1, \ldots, h+h_2\}$;
- $C^\alpha := [C_{ij}^\alpha]_{i,j=1,\ldots,n} \in \mathcal{M}_{n \times n}(\mathbb{R})$ for all $\alpha \in I_V = \{h+1, \ldots, n\}$.

Remark 4. It is important to observe that (2) immediately implies that the matrices just defined are the only ones which can be nonzero.

Let us define the left-invariant coframe $\underline{\omega} := \{\omega_i : i = 1, \ldots, n\}$ dual to $\underline{X}$; i.e., $\omega_i = X_i^*$ for every $i = 1, \ldots, n$. The *left-invariant 1-forms* ω_i for $i = 1, \ldots, n$ are uniquely determined by the condition $\omega_i(X_j) = \langle X_i, X_j \rangle = \delta_i^j$ for all $i, j = 1, \ldots, n$, where δ_i^j denotes Kronecker delta.

Definition 5. We shall denote by ∇ the unique left-invariant Levi-Civita connection on $\mathbb{G}$ associated with the left-invariant metric $g = \langle \cdot, \cdot \rangle$. Moreover, if $X, Y \in \mathfrak{X}(H) := C^\infty(\mathbb{G}, H)$, we shall set

$$\nabla_X^H Y := \mathcal{P}_H(\nabla_X Y).$$

Let $\underline{X} = \{X_1, \ldots, X_n\}$ be the global left-invariant frame on $T\mathbb{G}$. Then

$$\nabla_{X_i} X_j = \frac{1}{2} \sum_{r=1}^{n} \bigl(C_{ij}^r - C_{jr}^i + C_{ri}^j\bigr) X_r \quad \text{for all } i, j = 1, \ldots, n; \tag{3}$$

see, for instance, [Milnor 1976, Section 5, pp. 310–311]. Furthermore, we stress that ∇^H is a partial connection, called *horizontal H-connection*; see [Ge 1992] or [Koiller et al. 2001]; see also [Montefalcone 2007a] and references therein. Using Definition 5 together with (3) and (2), it is not difficult to show the following:

- ∇^H is flat; i.e.,
$$\nabla^H_{X_i} X_j = 0 \quad \text{for all } i, j \in I_H;$$
- ∇^H is compatible with the subriemannian metric g_H; i.e.,
$$X\langle Y, Z\rangle = \langle \nabla^H_X Y, Z\rangle + \langle Y, \nabla^H_X Z\rangle \quad \text{for all } X, Y, Z \in \mathfrak{X}(H);$$
- ∇^H is torsion-free; i.e.,
$$\nabla^H_X Y - \nabla^H_Y X - \mathcal{P}_H[X, Y] = 0 \quad \text{for all } X, Y \in \mathfrak{X}(H).$$

Definition 6. If $\psi \in C^\infty(\mathbb{G})$ we define the horizontal gradient of ψ as the unique horizontal vector field $\operatorname{grad}_H \psi$ such that $\langle \operatorname{grad}_H \psi, X\rangle = d\psi(X) = X\psi$ for every $X \in \mathfrak{X}(H)$. The horizontal divergence of $X \in \mathfrak{X}(H)$, $\operatorname{div}_H X$, is defined, at each point $x \in \mathbb{G}$, by
$$\operatorname{div}_H X(x) := \operatorname{Trace}\big(Y \to \nabla^H_Y X\big)(x) \quad (Y \in H_x).$$
For any $Y = \sum_{j \in I_H} y_j X_j \in \mathfrak{X}(H)$, we denote by $\mathcal{J}_H Y$ the horizontal Jacobian matrix of Y; i.e.,
$$\mathcal{J}_H Y := \big[X_i(y_j)\big]_{j, i \in I_H}.$$

Example 7 (Heisenberg group $\mathbb{H}^n$ ($n \geq 1$)). The Lie algebra $\mathfrak{h}_n \cong \mathbb{R}^{2n+1}$ of the n-th Heisenberg group $\mathbb{H}^n$ can be described by means of a left-invariant frame $\underline{Z} := \{X_1, Y_1, \dots, X_i, Y_i, \dots, X_n, Y_n, T\}$, where, at each $p = \exp(x_1, y_1, x_2, y_2, \dots, x_n, y_n, t) \in \mathbb{H}^n$, we have set $X_i(p) := \partial/\partial x_i - \frac{1}{2} y_i \partial/\partial t$, $Y_i(p) := \partial/\partial y_i + \frac{1}{2} x_i \partial/\partial t$ for every $i = 1, \dots, n$; $T(p) := \partial/\partial t$. One has $[X_i, Y_i] = T$ for every $i = 1, \dots, n$, and all other commutators vanish, so that T is the *center* of $\mathfrak{h}_n$ and $\mathfrak{h}_n$ turns out to be a nilpotent and stratified Lie algebra of step 2; i.e., $\mathfrak{h}_n = H \oplus H_2$. The structural constants of $\mathfrak{h}_n$ are described by the skew-symmetric $(2n \times 2n)$-matrix
$$C_H^{2n+1} := \begin{vmatrix} 0 & 1 & \cdot & 0 & 0 \\ -1 & 0 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 0 & 1 \\ 0 & 0 & \cdot & -1 & 0 \end{vmatrix}.$$

1.2. ***Hypersurfaces.*** The (riemannian) left-invariant volume form of any Carnot group $\mathbb{G}$ is defined as $\sigma^n_R := \bigwedge_{i=1}^n \omega_i \in \bigwedge^n(T^*\mathbb{G})$. By integration of the n-form σ^n_R, one obtains the Haar measure of $\mathbb{G}$, which equals the push-forward of the n-dimensional Lebesgue measure $\mathcal{L}^n$ on $\mathfrak{g} \cong \mathbb{R}^n$. The symbols $\mathcal{H}^s_{CC}$, $\mathcal{H}^s_{Eu}$ will denote the intrinsic CC s-dimensional Hausdorff measure and the euclidean s-dimensional Hausdorff measure, respectively. (Sometimes we will use the notation $\sigma_R^n = \mathcal{V}\mathrm{ol}^n$). Let $S \subset \mathbb{G}$ be a hypersurface (i.e., a codimension 1 submanifold of $\mathbb{G}$) of class C^i ($i \geq 1$). Let ν denote the (riemannian) unit normal vector along S. Then $x \in S$ is a *characteristic point* if and only if $\dim H_x = \dim(H_x \cap T_x S)$. The *characteristic set* of S is given by $C_S := \{x \in S : \dim H_x = \dim(H_x \cap T_x S)\}$. In other words, a point $x \in S$ is noncharacteristic (hereafter abbreviated as NC) if and only if H is transversal to S at x. Hence, one has $C_S := \{x \in S : |\mathcal{P}_H \nu(x)| = 0\}$, where $\mathcal{P}_H$ denotes orthogonal projection onto H. It is of fundamental importance that the $(Q-1)$-dimensional CC Hausdorff measure of the characteristic set C_S vanishes; i.e., $\mathcal{H}_{\mathrm{CC}}^{Q-1}(C_S) = 0$; see, for instance, Theorem 6.6.2 in [Magnani 2002]. We also stress that if S is a hypersurface of class C^2, then precise estimates of the riemannian Hausdorff dimension of C_S can be found in [Balogh et al. 2010]; see also [Balogh 2003] for the case of the Heisenberg group $\mathbb{H}^n$ ($n \geq 1$).

The $(n-1)$-dimensional riemannian measure along S is defined by integration of the $(n-1)$-differential form $\sigma_R^{n-1} \llcorner S := (\nu \lrcorner \sigma_R^n)|_S$, where $\lrcorner$ denotes the "contraction" operator on differential forms; see [Federer 1969]. We recall that $\lrcorner : \bigwedge^k (T^*\mathbb{G}) \to \bigwedge^{k-1}(T^*\mathbb{G})$ is defined, for $X \in T\mathbb{G}$ and $\alpha \in \bigwedge^k(T^*\mathbb{G})$, by setting $(X \lrcorner \alpha)(Y_1, \dots, Y_{k-1}) := \alpha(X, Y_1, \dots, Y_{k-1})$.

At each NC point $x \in S \setminus C_S$ the *unit H-normal* is defined as

$$\nu_H := \frac{\mathcal{P}_H \nu}{|\mathcal{P}_H \nu|}.$$

Similarly to the riemannian case, we define an $(n-1)$-differential form $\sigma_H^{n-1} \in \bigwedge^{n-1}(T^* S)$ by setting

$$\sigma_H^{n-1} \llcorner S := (\nu_H \lrcorner \sigma_R^n)|_S.$$

By integration of $\sigma_H^{n-1} \llcorner S$, one gets a left-invariant and $(Q-1)$-homogeneous measure, which is called *H-perimeter measure*. This measure can be extended to the whole of S by setting $\sigma_H^{n-1} \llcorner C_S = 0$. Note that $\sigma_H^{n-1} \llcorner S = |\mathcal{P}_H \nu| \, \sigma_R^{n-1} \llcorner S$. Furthermore, denoting by $\mathcal{S}_{\mathrm{CC}}^{Q-1}$ the $(Q-1)$-dimensional spherical intrinsic CC Hausdorff measure (i.e., associated with the CC-distance d_{CC}), then

$$\sigma_H^{n-1}(S \cap B) = k(\nu_H)\, \mathcal{S}_{\mathrm{CC}}^{Q-1} \llcorner (S \cap B) \quad \text{for all } B \in \mathcal{B}\mathrm{or}(\mathbb{G}),$$

where the density-function $k(\nu_H)$, called *metric factor*, explicitly depends on ν_H and d_{CC}; see [Magnani 2002].

At each NC point $x \in S \setminus C_S$, the *horizontal tangent bundle* $HS := H \cap TS \subset TS$ and the *horizontal normal bundle* $\nu_H S \subset H$ split the horizontal bundle H into an orthogonal direct sum; i.e., $H = \nu_H \oplus HS$. The stratification of $\mathfrak{g}$ induces a stratification of $TS := \oplus_{i=1}^{k} H_i S$, where we have set $HS := H_1 S$; see [Gromov 1996]. Note that at any characteristic point $x \in C_S$ one has $H_x = H_x S$, so that

$$\dim(H_x S) = \begin{cases} h-1 & \text{if } x \in S \setminus C_S, \\ h & \text{if } x \in C_S. \end{cases}$$

Notation 8. Throughout this paper, we denote by $\mathcal{P}_{HS} : TS \to HS$ the orthogonal projection map from TS onto HS.

Now let $S \subset \mathbb{G}$ be a hypersurface of class C^2 and let ∇^{TS} denote the induced connection on S from ∇. The tangential connection ∇^{TS} induces a partial connection on HS defined by

$$\nabla_X^{HS} Y := \mathcal{P}_{HS}\big(\nabla_X^{TS} Y\big) \quad \text{for all } X, Y \in \mathfrak{X}^1(HS) := C^1(S, HS).$$

It turns out that

$$\nabla_X^{HS} Y = \nabla_X^H Y - \big\langle \nabla_X^H Y, \nu_H \big\rangle \nu_H \quad \text{for every } X, Y \in \mathfrak{X}^1(HS);$$

see [Montefalcone 2007a].

Definition 9 (see [Montefalcone 2007a]). We call HS-gradient of $\psi \in C^1(S)$ the unique horizontal tangent vector field $\operatorname{grad}_{HS} \psi$ such that

$$\langle \operatorname{grad}_{HS} \psi, X \rangle = d\psi(X) = X\psi \quad \text{for all } X \in \mathfrak{X}^1(HS).$$

We denote by div_{HS} the HS-divergence; i.e., if $X \in \mathfrak{X}^1(HS)$ and $x \in S$, then

$$\operatorname{div}_{HS} X(x) := \operatorname{Trace}\big(Y \to \nabla_Y^{HS} X\big)(x) \quad (Y \in H_x S).$$

The *HS-Laplacian* Δ_{HS} is the second-order differential operator defined as

$$\Delta_{HS}\psi := \operatorname{div}_{HS}(\operatorname{grad}_{HS}\psi) \quad \text{for every } \psi \in C^2(S).$$

The horizontal second fundamental form of $S \setminus C_S$ is the map given by

$$B_H(X, Y) := \big\langle \nabla_X^H Y, \nu_H \big\rangle \quad \text{for all } X, Y \in \mathfrak{X}^1(HS).$$

The horizontal mean curvature $\mathcal{H}_H$ is the trace of B_H; i.e., $\mathcal{H}_H := \operatorname{Tr} B_H = -\operatorname{div}_H \nu_H$.

It is worth observing that the HS-connection admits, in general, a nonzero torsion because B_H is *not symmetric*; see [Montefalcone 2007a].

Definition 10. Let $\mathcal{U} \subseteq S$ be an open set. We shall denote by $C_{HS}^i(\mathcal{U})$ $(i = 1, 2)$ the space of functions whose HS-derivatives up to i-th order are continuous on $\mathcal{U}$.

We stress that the previous definitions concerning the horizontal second fundamental form $B_H(\cdot\,,\cdot)$ and the HS-connection can also be reformulated by using the function space $C_{HS}^i(\mathcal{U})$ $(i = 1, 2)$ and, more precisely, by replacing $\mathfrak{X}^1(HS) = C^1(S, HS)$ with $\mathfrak{X}_{HS}^1(HS) := C_{HS}^1(S, HS)$.

Let $S \subset \mathbb{G}$ be a hypersurface of class C^i $(i \geq 1)$ and let ν be the outward-pointing unit normal vector field along S. We need to define some important geometric objects. To this end, we first note that $\nu = \mathcal{P}_H\nu + \mathcal{P}_V\nu$. By using the left-invariant frame $\underline{X} = \{X_1, \dots, X_n\}$, we see that $\mathcal{P}_V\nu = \sum_{\alpha \in I_V} \nu_\alpha X_\alpha$, where $\nu_\alpha := \langle \nu, X_\alpha \rangle$; see Notation 2.

Notation 11. Hereafter we shall set

- $\varpi_\alpha := \dfrac{\nu_\alpha}{|\mathcal{P}_H\nu|}$ for all $\alpha \in I_V$;
- $\varpi := \sum_{\alpha \in I_V} \varpi_\alpha X_\alpha$;
- $C_H := \sum_{\alpha \in I_{H_2}} \varpi_\alpha \, C_H^\alpha$;

see, for instance, Notation 3 and Remark 4.

1.3. ***Other tools.*** Let $S \subset \mathbb{G}$ be a hypersurface of class C^i $(i \geq 1)$. Let ∂S be an $(n-2)$-dimensional submanifold of S of class C^1, oriented by the outward pointing unit normal vector $\eta \in TS \cap \mathrm{Nor}(\partial S)$. We shall denote by σ_R^{n-2} the riemannian measure on ∂S; i.e., $\sigma_R^{n-2} \llcorner \partial S = (\eta \lrcorner \sigma_R^{n-1})|_{\partial S}$. In particular, note that $(X \lrcorner \sigma_H^{n-1})|_{\partial S} = \langle X, \eta \rangle |\mathcal{P}_H\nu| \, \sigma_R^{n-2} \llcorner \partial S$ for every $X \in \mathfrak{X}^1(TS) := C^1(S, TS)$. The *unit HS-normal* along ∂S is given by $\eta_{HS} := \mathcal{P}_{HS}\eta / |\mathcal{P}_{HS}\eta|$. In this way, we can define a homogeneous $(n-2)$-dimensional measure $\sigma_H^{n-2} \in \bigwedge^{n-2}(T^*\partial S)$ by setting $\sigma_H^{n-2} \llcorner \partial S := (\eta_{HS} \lrcorner \sigma_H^{n-1})|_{\partial S}$. It follows that

$$\sigma_H^{n-2} \llcorner \partial S = |\mathcal{P}_H\nu| \, |\mathcal{P}_{HS}\eta| \, \sigma_R^{n-2} \llcorner \partial S$$

and that $(X \lrcorner \sigma_H^{n-1})|_{\partial S} = \langle X, \eta_{HS} \rangle \, \sigma_H^{n-2} \llcorner \partial S$ for all $X \in \mathfrak{X}^1(HS) := C^1(S, HS)$.

Now let $\nu \wedge \eta \in \Lambda^2(TS)$ be a unit 2-vector orienting ∂S, where $\nu \in \mathrm{Nor}(S)$ and $\eta \in TS \cap \mathrm{Nor}(\partial S)$. Then, the *characteristic set* of ∂S is defined as

$$C_{\partial S} := \{p \in \partial S : |\mathcal{P}_H(\nu \wedge \eta)| = 0\},$$

where the orthogonal projection operator $\mathcal{P}_H$ is extended to 2-vectors in the standard way.

Proposition 12. *Let $S \subset \mathbb{G}$ be a compact hypersurface of class C^1 and let $\phi \in C^1_{HS}(S)$. Then*

$$\int_S |\mathrm{grad}_{HS}\phi(x)| \, \sigma_H^{n-1}(x) = \int_{\mathbb{R}} \sigma_H^{n-2}\{\phi^{-1}[s] \cap S\} \, ds. \tag{4}$$

Proof. This formula follows from the riemannian coarea formula; see [Burago and Zalgaller 1988], [Chavel 2001] or [Montefalcone 2009]. We have

$$\int_S \phi(x) |\mathrm{grad}_{TS}\varphi(x)| \, \sigma_R^{n-1}(x) = \int_{\mathbb{R}} ds \int_{\varphi^{-1}[s] \cap S} \phi(y) \, \sigma_R^{n-2}(y)$$

for every $\phi \in L^1(S, \sigma_R^{n-1})$; see [Burago and Zalgaller 1988; Chavel 2001]. Choosing

$$\phi = \frac{|\mathrm{grad}_{HS}\varphi|}{|\mathrm{grad}_{TS}\varphi|}|\mathcal{P}_H\nu|$$

yields

$$\int_S \phi|\mathrm{grad}_{TS}\varphi|\,\sigma_R^{n-1} = \int_S \frac{|\mathrm{grad}_{HS}\varphi|}{|\mathrm{grad}_{TS}\varphi|}|\mathrm{grad}_{TS}\varphi|\,\underbrace{|\mathcal{P}_H\nu|\,\sigma_R^{n-1}}_{=\sigma_H^{n-1}} = \int_S |\mathrm{grad}_{HS}\varphi|\,\sigma_H^{n-1}.$$

The (riemannian) unit normal η along $\varphi^{-1}[s]$ is given by $\eta = \mathrm{grad}_{TS}\varphi/|\mathrm{grad}_{TS}\varphi|$. Hence $|\mathcal{P}_{HS}\eta| = |\mathrm{grad}_{HS}\varphi|/|\mathrm{grad}_{TS}\varphi|$ and it turns out that

$$\begin{aligned}\int_{\mathbb{R}} ds \int_{\varphi^{-1}[s]\cap S} \phi(y)\,\sigma_R^{n-2} &= \int_{\mathbb{R}} ds \int_{\varphi^{-1}[s]\cap S} \frac{|\mathrm{grad}_{HS}\varphi|}{|\mathrm{grad}_{TS}\varphi|}|\mathcal{P}_H\nu|\,\sigma_R^{n-2}\\ &= \int_{\mathbb{R}} ds \int_{\varphi^{-1}[s]\cap S} \underbrace{|\mathcal{P}_{HS}\eta|\,|\mathcal{P}_H\nu|\,\sigma_R^{n-2}}_{=\sigma_H^{n-2}}\\ &= \int_{\mathbb{R}} ds \int_{\varphi^{-1}[s]\cap S} \sigma_H^{n-2}. \qquad \square\end{aligned}$$

Below, we recall a basic integration by parts formula for horizontal vector fields; see [Montefalcone 2007a].

Definition 13. Let $\mathcal{D}_{HS} : \mathfrak{X}^1_{HS}(HS) \to C(S)$ be the first-order differential operator given by

$$\mathcal{D}_{HS}X := \mathrm{div}_{HS}\, X + \langle C_H \nu_H, X\rangle \quad \text{for all } X \in \mathfrak{X}^1_{HS}(HS) \ \big(:= C^1_{HS}(S, HS)\big).$$

Furthermore, let $\mathcal{L}_{HS} : C^2_{HS}(S) \to C(S)$ be the second-order differential operator given by

$$\mathcal{L}_{HS}\varphi := \Delta_{HS}\varphi + \langle C_H\nu_H, \mathrm{grad}_{HS}\varphi\rangle \quad \text{for all } \varphi \in C^2_{HS}(S);$$

see Definition 9 and Notation 11.

The horizontal matrix C_H is a key object, related to the skew-symmetric part of the horizontal second fundamental form B_H. Note that $\mathcal{D}_{HS}(\varphi X) = \varphi\mathcal{D}_{HS}X + \langle \mathrm{grad}_{HS}\varphi, X\rangle$ for every $X \in \mathfrak{X}^1_{HS}(HS)$ and every $\varphi \in C^1_{HS}(S)$. Moreover, one has $\mathcal{L}_{HS}\varphi = \mathcal{D}_{HS}(\mathrm{grad}_{HS}\varphi)$ for every $\varphi \in C^2_{HS}(S)$. These definitions are motivated by Theorem 3.17, Corollary 3.18 and Corollary 3.19 in [Montefalcone 2007a].

Theorem 14 (see [Montefalcone 2007a]). *Let S be a compact NC hypersurface of class C^2 with boundary ∂S of class C^1. Then*

$$(5)\quad \int_S \mathcal{D}_{HS}X\,\sigma_H^{n-1} = -\int_S \mathcal{H}_H\langle X, \nu_H\rangle\,\sigma_H^{n-1} + \int_{\partial S}\langle X, \eta_{HS}\rangle\,\sigma_H^{n-2} \quad \textit{for all } X \in \mathfrak{X}^1(H).$$

Remark 15. We note that, in general, $\mathcal{H}_H \notin L^1_{\mathrm{loc}}(S;\sigma_R^{n-1})$; see [Danielli et al. 2012]. However, it is always true that $\mathcal{H}_H \in L^1_{\mathrm{loc}}(S;\sigma_H^{n-1})$; see, for instance, [Montefalcone 2012].

Remark 16. Let $S \subset \mathbb{G}$ be a hypersurface of class C^2 and ν the outward-pointing unit normal vector along S. For any $X \in \mathfrak{X}(\mathbb{G})$ let us set $X^{\perp} := \langle X, \nu\rangle \nu$ and $X^{\top} := X - X^{\perp}$ to denote the riemannian normal and tangential components of X at any point of S. We would like to stress that formula (5) can be seen as a particular case of a general integral formula, the so-called first variation formula of the H-perimeter. More precisely, the first variation formula is given by

$$(6)\quad I_S(X,\sigma_H^{n-1}) = \int_S \bigl(-\mathcal{H}_H\langle X^{\perp},\nu\rangle + \mathrm{div}_{TS}\bigl(X^{\top}|\mathcal{P}_H\nu| - \langle X^{\perp},\nu\rangle \nu_H^{\top}\bigr)\bigr)\,\sigma_R^{n-1},$$

where $I_S(X,\sigma_H^{n-1})$ denotes the first derivative of the H-perimeter under a smooth variation of S with initial velocity X; see [Montefalcone 2012, Theorem 4.6]. Formula (6) also holds if $C_S \neq \varnothing$, but in this case we need to assume $\mathcal{H}_H \in L^1_{\mathrm{loc}}(S;\sigma_R^{n-1})$. We observe that, in the case of the first Heisenberg group $\mathbb{H}^1$, this formula coincides with that of Ritoré and Rosales [2008, Lemma 4.3, p. 14]. Note that, if $X = X_H \in \mathfrak{X}(H)$, then

$$\begin{aligned} X_H^{\top}|\mathcal{P}_H\nu| - \langle X_H^{\perp},\nu\rangle\nu_H^{\top} &\\ &= \bigl(X_H - |\mathcal{P}_H\nu|\langle X_H,\nu_H\rangle\nu\bigr)|\mathcal{P}_H\nu| - |\mathcal{P}_H\nu|\langle X_H,\nu\rangle\bigl(\nu_H - |\mathcal{P}_H\nu|\nu\bigr)\\ &= \bigl(X_H - \langle X_H,\nu\rangle\nu_H\bigr)|\mathcal{P}_H\nu| = \mathcal{P}_{HS}(X_H)\,|\mathcal{P}_H\nu|, \end{aligned}$$

where we have used the fact that $\nu = |\mathcal{P}_H\nu|\nu_H + \sum_{\alpha\in I_V}\nu_\alpha X_\alpha$ at each NC point. Finally, inserting this into (6), we obtain an equivalent form of (5). In particular, for any $X \in \mathfrak{X}(H)$ the function $\mathcal{D}_{HS}X$ turns out to be the Lie derivative of the differential $(n-1)$-form $\sigma_H^{n-1} \llcorner S$ with respect to the initial velocity X of a smooth variation of S. Roughly speaking, this can be rephrased by saying that the differential $(n-1)$-form $(\mathcal{D}_{HS}X)\,\sigma_H^{n-1} \in \Lambda^{n-1}(T^*S)$ is the "infinitesimal" first variation of S.

Formula (5) holds true even if $C_S \neq \varnothing$, at least under suitable assumptions.

Definition 17. Let $X \in C^1(S \setminus C_S, HS)$ and set $\alpha_X := (X \lrcorner\, \sigma_H^{n-1})|_S$. We say that X is *admissible (for the horizontal divergence formula)* if the differential forms α_X and $d\alpha_X$ are continuous on all of S, or, more generally, if $\alpha, d\alpha \in L^{\infty}(S)$ and $\iota_S^*\alpha \in L^{\infty}(\partial S)$. We say that $\phi \in C^2_{HS}(S \setminus C_S)$ is *admissible* if $\mathrm{grad}_{HS}\phi$ is admissible for the horizontal divergence formula.

We stress that, if the differential forms α_X and $d\alpha_X$ are continuous on all of S (or, more generally, if α, $d\alpha \in L^{\infty}(S)$ and $\iota_S^*\alpha \in L^{\infty}(\partial S)$, where $\iota_S : \partial M \to \overline{M}$ is the natural inclusion), then Stokes' formula holds true; see, for instance, [Taylor 2006]. This fact motivates the following:

Corollary 18. *Let $S \subset \mathbb{G}$ be a compact hypersurface of class C^2 with boundary ∂S of class C^1. Then*:

(i) $\int_S \mathcal{D}_{HS} X\, \sigma_H^{n-1} = \int_{\partial S} \langle X, \eta_{HS}\rangle\, \sigma_H^{n-2}$ *for every admissible $X \in C^1(S \setminus C_S, HS)$.*

(ii) $\int_S \mathcal{L}_{HS}\phi\, \sigma_H^{n-1} = \int_{\partial S} \langle \mathrm{grad}_{HS}\phi, \eta_{HS}\rangle\, \sigma_H^{n-2}$ *for every admissible $\phi \in C^2_{HS}(S \setminus C_S)$.*

(iii) *If $\partial S = \varnothing$, then*

$$-\int_S \varphi \mathcal{L}_{HS}\varphi\, \sigma_H^{n-1} = \int_S |\mathrm{grad}_{HS}\varphi|^2\, \sigma_H^{n-1}$$

for every $\varphi \in C^2_{HS}(S \setminus C_S)$ such that φ^2 is admissible.

The last formula holds even if $\partial S \neq \varnothing$, but for compactly supported functions. One can show that φ^2 is admissible if and only if $\varphi \in C^2_{HS}(S \setminus C_S) \cap W^{1,2}_{HS}(S, \sigma_H^{n-1})$, where we have set $W^{1,2}_{HS}(S, \sigma_H^{n-1}) := \{\varphi \in L^2(S, \sigma_H^{n-1}) : |\mathrm{grad}_{HS}\varphi| \in L^2(S, \sigma_H^{n-1})\}$. We also remark that any vector field $X \in C^1(S, HS)$ turns out to be admissible. Analogously, any $\varphi \in C^2_{HS}(S)$ is admissible.

Lemma 19. *Let $x_H := \sum_{i \in I_H} x_i X_i$ be the "horizontal position vector" and let g_H denote its component along the H-normal ν_H; i.e., $g_H := \langle x_H, \nu_H\rangle$. In the sequel, the function g_H will be called "horizontal support function" of x_H. Then, we have*:

(i) $\mathrm{div}_H\, x_H = h$;

(ii) $\mathcal{D}_{HS}(x_{HS}) = (h-1) + g_H \mathcal{H}_H + \langle C_H \nu_H, x_{HS}\rangle$ *at each NC point $x \in S \setminus C_S$, where $x_{HS} := x_H - g_H \nu_H$.*

Proof. We have $\mathrm{div}_H\, x_H = \sum_{i=1}^h \langle \nabla_{X_i} x_H, X_i\rangle = \sum_{i,\, j=1}^h \big(X_i(x_j) + \langle \nabla_{X_i} X_j, X_i\rangle\big) = \sum_{i,\, j=1}^h \delta_i^j = h$, where δ_i^j denotes Kronecker's delta; here we have used $\mathcal{J}_H(x_H) = \mathbf{Id}_h$ and $\langle \nabla_{X_i} X_j, X_i\rangle = 0$ for all $i,\ j \in I_H$; see Definition 6 and formula (6). Furthermore, by definition, one has $\mathrm{div}_{HS}\, x_H = \mathrm{div}_H\, x_H - \langle \nabla_{\nu_H} x_H, \nu_H\rangle$. Hence $\mathrm{div}_{HS}\, x_H = h - \langle \nu_H, \nu_H\rangle = h - 1$. Furthermore, by definition, we have

$$\mathrm{div}_{HS}\, x_{HS} = \sum_{i=2}^h \big\langle \nabla_{\tau_i}(x_H - g_H \nu_H), \tau_i\big\rangle, \tag{7}$$

where we have used an orthonormal horizontal frame $\underline{\tau}_H := \{\tau_1, \dots, \tau_h\}$ in an open neighborhood $U \subset \mathbb{G}$ of S such that $\tau_1(x) = \nu_H(x)$ at any $x \in S \setminus C_S$; see, for instance, Definition 3.4 in [Montefalcone 2007a]. Starting from (7), we compute

$$\mathrm{div}_{HS}\, x_{HS} = \sum_{i=2}^h \big(\langle \tau_i, \tau_i\rangle - g_H \langle \nabla^H_{\tau_i} \nu_H, \tau_i\rangle\big) = (h-1) - g_H\, \mathrm{div}_H\, \nu_H = (h-1) + g_H \mathcal{H}_H$$

for every $x \in S \setminus C_S$. The thesis easily follows from the definition of $\mathcal{D}_{HS}$. $\square$

A simple consequence of Corollary 18 and Lemma 19 is given by the following:

Corollary 20 (Minkowski-type formula). *Let $S \subset \mathbb{G}$ be a compact hypersurface of class C^2. Let $x_H = \sum_{i \in I_H} x_i X_i$ be the horizontal position vector. Furthermore, set $g_H = \langle x_H, \nu_H \rangle$ and $x_{HS} = x - g_H \nu_H$ for every $x \in S \setminus C_S$. Then*

$$\int_S \big((h-1) + g_H \mathcal{H}_H + \langle C_H \nu_H, x_{HS} \rangle\big)\, \sigma_H^{n-1} = 0.$$

Proof. It is enough to apply Corollary 18 to the horizontal tangent vector field $x_{HS} \in C^1(S \setminus C_S, HS)$. Using Remark 15 and Lemma 19 the thesis easily follows. $\square$

Definition 21 (eigenvalue problems for $\mathcal{L}_{HS}$). Let $S \subset \mathbb{G}$ be a compact hypersurface of class C^2 without boundary. Then we look for solutions of class $C^2_{HS}(S \setminus C_S) \cap W^{1,2}_{HS}(S, \sigma_H^{n-1})$ to the problem

$$(\mathrm{P}_1) \qquad \begin{cases} -\mathcal{L}_{HS}\psi = \lambda\psi; \\ \int_S \psi\, \sigma_H^{n-1} = 0. \end{cases}$$

If $\partial S \neq \varnothing$, we look for solutions of class $C^2_{HS}(S \setminus C_S) \cap W^{1,2}_{HS}(S, \sigma_H^{n-1})$ to the problems

$$(\mathrm{P}_2) \quad \begin{cases} -\mathcal{L}_{HS}\psi = \lambda\psi; \\ \psi|_{\partial S} = 0; \end{cases} \qquad (\mathrm{P}_3) \quad \begin{cases} -\mathcal{L}_{HS}\psi = \lambda\psi; \\ \dfrac{\partial\psi}{\partial\eta_{HS}}\Big|_{\partial S} = 0. \end{cases}$$

We explicitly remark that $\partial\psi/\partial\eta_{HS} = \langle \mathrm{grad}_{HS}\psi, \eta_{HS} \rangle$.

The problems (P_1), (P_2) and (P_3) generalize to our context the classical *closed, Dirichlet and Neumann eigenvalue problems* for the Laplace–Beltrami operator on riemannian manifolds; see [Chavel 1984; 1993].

Finally, we recall a recent general result about the size of horizontal tangencies to noninvolutive distributions, which applies to our Carnot setting; see Theorem 4.5 in [Balogh et al. 2010].

Theorem 22 (generalized Derridj's theorem). *Let $\mathbb{G}$ be a k-step Carnot group.*

(i) *If $S \subset \mathbb{G}$ is a hypersurface of class C^2, the euclidean-Hausdorff dimension of the characteristic set C_S of S satisfies* $\dim_{\text{Eu-Hau}}(C_N) \leq n-2$.

(ii) *If $V = H^\perp \subset T\mathbb{G}$ satisfies* $\dim V \geq 2$ *and $N \subset \mathbb{G}$ is an $(n-2)$-dimensional submanifold of class C^2, then the euclidean-Hausdorff dimension of the characteristic set C_N of N satisfies* $\dim_{\text{Eu-Hau}}(C_N) \leq n-3$.

Remark 23. Let $N \subset \mathbb{G}$ be an $(n-2)$-dimensional submanifold of class C^2. This smoothness condition is sharp; see [Balogh et al. 2010]. Moreover, we stress that $\dim V = 1$ just for Heisenberg groups and 2-step Carnot groups having 1-dimensional center. For Heisenberg groups $\mathbb{H}^n$, $n > 1$, using Frobenius' theorem yields $\dim_{\text{Eu-Hau}}(C_N) \leq n$, where $n = \frac{1}{2}\dim H$; see also [Balogh et al. 2010]. On the contrary, 1-dimensional curves in $\mathbb{H}^1$ can be horizontal or transversal to H. For

2-step groups having 1-dimensional center (or, equivalently, horizontal bundle H of codimension 1) a simple analysis shows that $\dim_{\text{Eu-Hau}}(C_N) = n-2$ if, and only if, $\mathbb{G}$ reduces to the direct product of $\mathbb{H}^1$ and of a euclidean space $\mathbb{R}^{h-2}$.

2. Isoperimetric constants and the first eigenvalue of $\mathscr{L}_{HS}$ on compact hypersurfaces

As a consequence of the coarea formula (4) we may generalize to the Carnot groups setting some results about isoperimetric constants and global Poincaré inequalities for which we refer the reader to [Chavel 1984; 1993]; see also [Cheeger 1970; Yau 1975].

Let $S \subset \mathbb{G}$ be a compact hypersurface of class C^2 with (or without) boundary. Similarly as in the riemannian setting (see [Cheeger 1970; Yau 1975]), we may give the following:

Definition 24. The *isoperimetric constant* $\operatorname{Isop}(S)$ of S is defined as follows:

- If $\partial S = \varnothing$, we set

$$\operatorname{Isop}(S) := \inf \frac{\sigma_H^{n-2}(N)}{\min\{\sigma_H^{n-1}(S_1), \sigma_H^{n-1}(S_2)\}},$$

  where the infimum is taken over all C^2-smooth $(n-2)$-dimensional submanifolds N of S which divide S into two hypersurfaces S_1, S_2 with common boundary $N = \partial S_1 = \partial S_2$.

- If $\partial S \neq \varnothing$, we set

$$\operatorname{Isop}(S) := \inf \frac{\sigma_H^{n-2}(N)}{\sigma_H^{n-1}(S_1)},$$

  where $N \subset S$ is a smooth hypersurface of S such that $N \cap \partial S = \varnothing$ and S_1 is the unique C^2-smooth $(n-2)$-dimensional submanifold of S such that $N = \partial S_1$.

Here ∂S, S_1, S_2 and $N = \partial S_i$ $(i = 1, 2)$ are not assumed to be connected.

This definition requires some comments. As recalled in the introduction, in the riemannian setting analogous isoperimetric constants were introduced in [Cheeger 1970], in order to give a geometric lower bound for the smallest eigenvalue of the Laplace–Beltrami operator on smooth compact riemannian manifolds. This definition was somewhat motivated by an example of Calabi, the so-called *dumbbell* manifold, homeomorphic to $\mathbf{S}^2$. Actually, an analysis of this example shows that, in order to bound λ from below, the diameter and the volume are not enough.

We also have to recall that these isoperimetric constants turn out to be strictly positive. Although this claim turns out to be (more or less) elementary in dimension

$n = 2$, it becomes a bit more difficult when $n > 2$; see [Cheeger 1970]. Some years after Cheeger's result, Yau [1975] reconsidered the isoperimetric constants and demonstrated that λ has a bound in terms of volume, diameter and (of a lower bound of the) Ricci curvature. See the survey [Li 1982] for a glimpse on this topic.

Below we shall generalize some of the results of [Yau 1975]. Our results will follow the original scheme, which is based mainly on a suitable use of the coarea formula for smooth functions. Note also that, instead of C^∞-smooth hypersurfaces, here we are considering hypersurfaces of class C^2. We have to observe that all the results could also be stated for C^1 hypersurfaces. But the delicate matter here is that in our setting, new difficulties come from the presence of characteristic points and, in the C^1 case, it is not simple to prove that isoperimetric constants are strictly positive. Actually, the following further hypothesis seems to be unavoidable in order to have nonzero isoperimetric constants:

(H) *Every C^2-smooth $(n-2)$-dimensional submanifold $N \subset S$ satisfies*

$$\dim C_N < n-2.$$

This assumption can be overcome by using the generalized Derridj's theorem, (Theorem 22); see also Remark 23. As a consequence, the results of this section are "meaningful" (in the sense that the isoperimetric constants do not vanish) at least for any Carnot group $\mathbb{G}$ such that $\dim V \geq 2$ and for all Heisenberg groups $\mathbb{H}^n$, with $n > 1$.

Theorem 25. *Let $S \subset \mathbb{G}$ be a compact hypersurface of class C^2.*

(i) *If $\partial S = \varnothing$, then*

$$\mathrm{Isop}(S) = \inf \frac{\int_S |\mathrm{grad}_{HS}\,\psi|\,\sigma_H^{n-1}}{\int_S |\psi|\,\sigma_H^{n-1}},$$

where the infimum is taken over all C^2-smooth functions on S such that $\int_S \psi\,\sigma_H^{n-1} = 0$.

(ii) *If $\partial S \neq \varnothing$, then*

$$\mathrm{Isop}(S) = \inf \frac{\int_S |\mathrm{grad}_{HS}\,\psi|\,\sigma_H^{n-1}}{\int_S |\psi|\,\sigma_H^{n-1}},$$

where the infimum is taken over all C^2-smooth functions on S such that $\psi|_{\partial S} = 0$.

Warning 26. The definition of $\mathrm{Isop}(S)$ can be weakened. For instance, part (i) of Definition 24 can be given by assuming S of class C^1 and then by taking the infimum over all $(n-2)$-dimensional submanifolds N of S of class C^1 which divide S into two hypersurfaces S_1, S_2 with common boundary $N = \partial S_1 = \partial S_2$.

In this case, Theorem 25(i) holds, without modifications, by taking the infimum over C^1_{HS}-smooth functions. If $\partial S \neq \varnothing$ an analogous claim holds, for the other isoperimetric constant. Furthermore, equivalent remarks can be given for all the results of this section. Nevertheless, as already said, this weaker formulation seems to be less meaningful because of the presence of characteristic points.

Warning 27. Throughout this section, we shall fix a homogeneous distance ϱ on $\mathbb{G}$ of class C^1 outside the diagonal of $\mathbb{G}$.

Proof of Theorem 25. The proof repeats almost verbatim the arguments of Theorem 1 in [Yau 1975]. We just prove the theorem for $\partial S = \varnothing$ since the other case is analogous. First, let us prove the inequality

$$\mathrm{Isop}(S) \leq \inf \frac{\int_S |\mathrm{grad}_{HS}\,\psi|\,\sigma_H^{n-1}}{\int_S |\psi|\,\sigma_H^{n-1}},$$

where $\psi \in C^2(S)$ and $\int_S \psi\,\sigma_H^{n-1} = 0$. To prove this inequality let us consider the auxiliary functions $\psi^+ = \max\{0, \psi\}$, $\psi^- = \max\{0, -\psi\}$. By applying the coarea formula (4) and the definition of $\mathrm{Isop}(S)$ we get that

$$\int_S |\mathrm{grad}_{HS}\,\psi^\pm|\,\sigma_H^{n-1} = \int_0^{+\infty} \sigma_H^{n-2}\{x \in S : \psi^\pm = t\}\,dt \geq \mathrm{Isop}(S) \int_S |\psi^\pm|\,\sigma_H^{n-1}.$$

Now we shall prove the reversed inequality. So let us assume that $\sigma_H^{n-1}(S_1) \leq \sigma_H^{n-1}(S_2)$ and let $\epsilon > 0$. By making use of the fixed homogeneous distance ϱ on $\mathbb{G}$, we now define a function $\psi_\epsilon : S \to \mathbb{R}$ by setting

$$\psi_\epsilon(x)|_{S_1} := \begin{cases} \dfrac{\varrho(x, N)}{\epsilon} & \text{if } \varrho(x, N) \leq \epsilon, \\ 1 & \text{if } \varrho(x, N) > \epsilon, \end{cases} \tag{8}$$

$$\psi_\epsilon(x)|_{S_2} := \begin{cases} -\alpha\dfrac{\varrho(x, N)}{\epsilon} & \text{if } \varrho(x, N) \leq \epsilon, \\ -\alpha & \text{if } \varrho(x, N) > \epsilon, \end{cases}$$

where the constant α depends on ϵ and is chosen in a way that $\int_S \psi_\epsilon\,\sigma_H^{n-1} = 0$. Obviously

$$\lim_{\epsilon \to 0} \alpha = \frac{\sigma_H^{n-1}(S_1)}{\sigma_H^{n-1}(S_2)}.$$

Since

$$\begin{aligned} \int_S |\mathrm{grad}_{HS}\,\psi_\epsilon|\,\sigma_H^{n-1} &= \frac{1+\alpha}{\epsilon} \int_{N_\epsilon := \{x \in S : \varrho(x, N) \leq \epsilon\}} |\mathrm{grad}_{HS}\varrho(x, N)|\,\sigma_H^{n-1} \\ &= \frac{1+\alpha}{\epsilon} \int_0^\epsilon \sigma_H^{n-2}\{x \in N_\epsilon : \varrho(x, N) = t\}\,dt, \end{aligned}$$

one gets

$$\lim_{\epsilon\to 0}\int_S |\mathrm{grad}_{HS}\,\psi_\epsilon|\,\sigma_H^{n-1} = (1+\alpha)\,\sigma_H^{n-2}(N).$$

Moreover $\lim_{\epsilon\to 0}\int_S |\psi_\epsilon|\,\sigma_H^{n-1} = \sigma_H^{n-1}(S_1) + \alpha\,\sigma_H^{n-1}(S_2)$. Putting it all together we get

$$\inf_\psi \frac{\int_S |\mathrm{grad}_{HS}\,\psi|\,\sigma_H^{n-1}}{\int_S |\psi|\,\sigma_H^{n-1}} \le \lim_{\epsilon\to 0}\frac{\int_S |\mathrm{grad}_{HS}\,\psi_\epsilon|\,\sigma_H^{n-1}}{\int_S |\psi_\epsilon|\,\sigma_H^{n-1}} \le \frac{\sigma_H^{n-1}(N)}{\sigma_H^{n-2}(S_1)}.$$

If we take the infimum over N and S_1, the inequality follows. □

Corollary 28. *Let λ_1 be the first nonzero eigenvalue of either the closed eigenvalue problem or the Dirichlet eigenvalue problem* (see Definition 21). *Then we have* $\lambda_1 \ge \frac{1}{4}(\mathrm{Isop}(S))^2$.

Proof. We just prove the first claim, as the second claim is similar. Let ψ be an eigenfunction of $\mathcal{L}_{HS}$ corresponding to λ_1. Then

$$\begin{aligned}\lambda_1 &= -\frac{\int_S \psi\,\mathcal{L}_{HS}\psi\,\sigma_H^{n-1}}{\int_S |\psi|^2\sigma_H^{n-1}} = \frac{\int_S |\mathrm{grad}_{HS}\,\psi|^2\,\sigma_H^{n-1}}{\int_S |\psi|^2\,\sigma_H^{n-1}} = \frac{\int_S |\mathrm{grad}_{HS}\,\psi|^2\,\sigma_H^{n-1}}{\left(\int_S |\psi|^2\sigma_H^{n-1}\right)^2}\int_S |\psi|^2\,\sigma_H^{n-1}\\ &\ge \frac{\left(\int_S |\psi|\,|\mathrm{grad}_{HS}\,\psi|\,\sigma_H^{n-1}\right)^2}{\left(\int_S |\psi|^2\sigma_H^{n-1}\right)^2} = \frac{1}{4}\frac{\left(\int_S |\mathrm{grad}_{HS}\,\psi^2|\,\sigma_H^{n-1}\right)^2}{\left(\int_S \psi^2\sigma_H^{n-1}\right)^2} \ge \frac{(\mathrm{Isop}(S))^2}{4},\end{aligned}$$

where we have used Theorem 25 together with the Cauchy–Schwarz inequality. □

We now extend, to Carnot groups, another isoperimetric constant and some related facts which, in the riemannian case, were studied in [Yau 1975].

Definition 29. The *isoperimetric constant* $\mathrm{Isop}_0(S)$ of any C^2-smooth compact hypersurface $S \subset \mathbb{G}$ with boundary ∂S is given by

$$\mathrm{Isop}_0(S) := \inf\left\{\frac{\sigma_H^{n-2}(\partial S_1 \cap \partial S_2)}{\min\{\sigma_H^{n-1}(S_1), \sigma_H^{n-1}(S_2)\}}\right\},$$

where the infimum is taken over all decompositions $S = S_1 \cup S_2$ such that $\sigma_H^{n-1}(S_1 \cap S_2) = 0$.

Theorem 30. *Let $S \subset \mathbb{G}$ be a compact hypersurface of class C^2 with boundary. Then*

$$\mathrm{Isop}_0(S) = \inf\left\{\frac{\int_S |\mathrm{grad}_{HS}\,\psi|\,\sigma_H^{n-1}}{\inf_{\beta\in\mathbb{R}}\int_S |\psi-\beta|\,\sigma_H^{n-1}}\right\},$$

where the infimum is taken over all C^2-functions defined on S.

Proof. The proof is analogous to that of Theorem 6 in [Yau 1975]. First, let us prove the inequality

$$\mathrm{Isop}(S) \leq \inf \frac{\int_S |\mathrm{grad}_{HS}\psi|\,\sigma_H^{n-1}}{\int_S |\psi|\,\sigma_H^{n-1}}.$$

To this end, let us define the functions $\psi^+ := \max\{0, \psi-k\}$, $\psi^- := -\min\{0, \psi-k\}$, where $k \in \mathbb{R}$ is any constant such that

$$\sigma_H^{n-1}\{x \in S : \psi^+ > 0\} \leq \tfrac{1}{2}\sigma_H^{n-1}(S), \quad \sigma_H^{n-1}\{x \in S : \psi^- > 0\} \leq \tfrac{1}{2}\sigma_H^{n-1}(S).$$

By using again the coarea formula (4) together with the definition of $\mathrm{Isop}_0(S)$ we get that

$$\int_S |\mathrm{grad}_{HS}\psi^\pm|\,\sigma_H^{n-1} = \int_0^{+\infty} \sigma_H^{n-2}\{x \in S : \psi^\pm = t\}\,dt \geq \mathrm{Isop}(S) \int_S |\psi^\pm|\,\sigma_H^{n-1}.$$

We prove the other inequality. Assuming $\sigma_H^{n-1}(S_1) \leq \sigma_H^{n-1}(S_2)$ and $\epsilon > 0$, we define the function

$$\text{(9)}\quad \psi_\epsilon(x)|_{S_1} := 1, \quad \psi_\epsilon(x)|_{S_2} := \begin{cases} 1 - \dfrac{\varrho(x, \partial S_1 \cap \partial S_2)}{\epsilon} & \text{if } \varrho(x, \partial S_1 \cap \partial S_2)) \leq \epsilon, \\ 0 & \text{if } \varrho(x, \partial S_1 \cap \partial S_2)) > \epsilon. \end{cases}$$

Furthermore, one can find a constant $k(\epsilon)$ satisfying

$$\int_S |\psi_\epsilon - k(\epsilon)|\,\sigma_H^{n-1} = \inf_{\beta\in\mathbb{R}} \int_S |\psi_\epsilon - \beta|\,\sigma_H^{n-1}$$

and such that $k(\epsilon) \to 0$ for $\epsilon \to 0^+$. Hence

$$\lim_{\epsilon\to 0} \left\{ \frac{\int_S |\mathrm{grad}_{HS}\psi_\epsilon|\,\sigma_H^{n-1}}{\inf_{\beta\in\mathbb{R}} \int_S |\psi_\epsilon - \beta|\,\sigma_H^{n-1}} \right\} \leq \frac{\sigma_H^{n-2}(\partial S_1 \cap \partial S_2)}{\min\{\sigma_H^{n-1}(S_1), \sigma_H^{n-1}(S_2)\}}. \qquad \square$$

Corollary 31. *Let $S \subset \mathbb{G}$ be a compact hypersurface of class C^2. Then*

$$\text{(10)}\qquad \int_S |\psi - k|^2\,\sigma_H^{n-1} \leq \frac{4}{(\mathrm{Isop}_0(S))^2} \int_S |\mathrm{grad}_{HS}\psi|^2\,\sigma_H^{n-1}$$

for every $\psi \in C^2(S)$ and every $k \in \mathbb{R}$ such that

$$\sigma_H^{n-1}\{x \in S : \psi \geq k\} \geq \tfrac{1}{2}\sigma_H^{n-1}(S), \quad \sigma_H^{n-1}\{x \in S : \psi \leq k\} \geq \tfrac{1}{2}\sigma_H^{n-1}(S).$$

Furthermore, if $\psi \in C^2(S)$ and $\int_S \psi\,\sigma_H^{n-1} = 0$, then

$$\text{(11)}\qquad \int_S |\psi|^2\,\sigma_H^{n-1} \leq \frac{4}{(\mathrm{Isop}_0(S))^2} \int_S |\mathrm{grad}_{HS}\psi|^2\,\sigma_H^{n-1}.$$

Proof. One has $\int_S (\psi^+ \cdot \psi^-)\, \sigma_H^{n-1} = 0$, where the functions $\psi^\pm$ are defined as in the proof of Theorem 30. Moreover, by using once more coarea formula, we get

$$\begin{aligned}\int_S |\psi-k|^2 \sigma_H^{n-1} &= \int_S |\psi^+ + \psi^-|^2 \sigma_H^{n-1} \leq \int_S |\psi^+|^2 \sigma_H^{n-1} + \int_S |\psi^-|^2 \sigma_H^{n-1} \\ &\leq \frac{1}{\mathrm{Isop}_0(S)} \left(\int_S |\mathrm{grad}_{HS} (\psi^+)^2| \sigma_H^{n-1} + \int_S |\mathrm{grad}_{HS} (\psi^-)^2| \sigma_H^{n-1} \right) \\ &\leq \frac{2}{\mathrm{Isop}_0(S)} \int_S (\psi^+ + \psi^-) |\mathrm{grad}_{HS} \psi| \sigma_H^{n-1} \\ &\leq \frac{2}{\mathrm{Isop}_0(S)} \|\psi^+ + \psi^-\|_{L^2(S;\sigma_H^{n-1})} \|\mathrm{grad}_{HS} \psi\|_{L^2(S;\sigma_H^{n-1})}.\end{aligned}$$

This proves (10). In order to prove (11) we note that the hypothesis $\int_S \psi\, \sigma_H^{n-1} = 0$ actually implies that

$$\int_S \psi^2\, \sigma_H^{n-1} = \inf_{k\in\mathbb{R}} \int_S (\psi - k)^2\, \sigma_H^{n-1},$$

which, together with (10), implies the thesis of the theorem. □

3. Two upper bounds on λ_1

Below we shall extend two (nowadays classical) inequalities obtained, respectively, by Chavel and Reilly in the euclidean/riemannian setting. An important feature of these results is that they give explicit upper bounds for the first nontrivial eigenvalue (of the Laplacian) of a compact submanifold of $\mathbb{R}^n$. For further details we refer to [Chavel 1978] and [Reilly 1977]; see also [Heintze 1988]. To begin with, let $\Omega \subsetneq \mathbb{G}$ be a bounded domain and assume that $S := \partial\Omega$ is a connected hypersurface of class C^2, with orientation given by the outward normal vector ν. Moreover, let x_H be the horizontal position vector field and let us apply the usual divergence formula. We also set $\sigma_R^n = \mathcal{V}\mathrm{ol}^n$. We have

$$h\,\mathcal{V}\mathrm{ol}^n(\Omega) = \int_\Omega \mathrm{div}_H\, x_H\, \sigma_R^n = \int_{\partial\Omega} \langle x_H, \nu\rangle\, \sigma_R^{n-1} = \int_S \langle x_H, \nu_H\rangle\, \sigma_H^{n-1},$$

where we have used Lemma 19(i). Furthermore, we may further assume that the "center of mass" of $\partial\Omega$ (with respect to the H-perimeter) is placed at the identity $0 \in \mathbb{G}$. In other words, let us assume that $\int_S x_i\, \sigma_H^{n-1} = 0$ for every $i \in I_H = \{1, \dots, h\}$, where $x_H \equiv (x_1, \dots, x_i, \dots, x_h)$ is the horizontal position vector; see Lemma 19.

The last assumption is justified by the following:

Lemma 32. *Let $S \subset \mathbb{G}$ be a compact hypersurface of class C^i ($i \geq 1$). We can always choose a system of exponential coordinates $x = \exp(x_1, \dots, x_n)$ on $\mathbb{G}$ such that $\int_S x_i\, \sigma_H^{n-1}(x) = 0$ for every $i \in I_H = \{1, \dots, h\}$.*

Proof. Let

$$a_i := \frac{\int_S x_i\, \sigma_H^{n-1}(x)}{\sigma_H^{n-1}(S)} \quad \text{for all } i \in I_H = \{1, \dots, h\}$$

and $a_H \equiv (a_1, \dots, a_i, \dots, a_h)$. Set $a := \exp(a_H, 0_V)$, where the symbol 0_V denotes the origin of $V \subset \mathfrak{g}$. Consider the change of variables $y := \Phi(x) = a^{-1} \bullet x$ ($x \in \mathbb{G}$). Equivalently, we have $\Phi(x) = L_{a^{-1}}(x)$, where $L_{a^{-1}}$ is the left-translation by $a^{-1} = -a$; see Section 1.1. The usual change of variables formula together with standard properties of the pull-back imply the following chain of equalities:

$$\begin{aligned}\int_{\Phi(S)} f(y)\, \sigma_H^{n-1}(y) &= \int_S f(\Phi(x))\, \mathcal{J}\mathrm{ac}(\Phi)(x)\, \sigma_H^{n-1}(x) \\ &= \int_S \Phi^*\big(f\, \sigma_H^{n-1}\big) = \int_S (f \circ \Phi)\big(\Phi^*\, \sigma_H^{n-1}\big)\end{aligned} \tag{12}$$

for every smooth function $f : S \to \mathbb{R}$; see, for instance, [Lee 2003, Lemma 9.11, p. 214]. Using the left-invariance of the H-perimeter yields $\mathcal{J}\mathrm{ac}(\Phi) = 1$, or equivalently, $\Phi^*\sigma_H^{n-1} = \sigma_H^{n-1}$. Now, let us assume that $f(y) := y_i$ for any $i \in I_H$. Equivalently, let f be the i-th exponential coordinate of the variable $y \in \mathbb{G}$. Note also that $(f \circ \Phi)(x) = \Phi_i(x) = -a_i + x_i$ for any $i \in I_H$. Actually, this follows from the fact that the group law $\bullet$ acts linearly on the horizontal layer; see (1). Then, using (12) yields

$$\int_{\Phi(S)} y_i\, \sigma_H^{n-1}(y) = \int_S (-a_i + x_i)\, \sigma_H^{n-1}(x) = 0 \quad \text{for all } i \in I_H,$$

which achieves the proof. $\square$

We therefore get that

$$\begin{aligned} h\, \mathcal{V}\mathrm{ol}^n(\Omega) &= \int_S \langle x_H, \nu_H \rangle\, \sigma_H^{n-1} \le \int_S |x_H|\, \sigma_H^{n-1} \le \sqrt{\sigma_H^{n-1}(S)} \sqrt{\int_S |x_H|^2\, \sigma_H^{n-1}} \\ &= \sqrt{\sigma_H^{n-1}(S)} \sqrt{\int_S \sum_{i \in I_H} x_i^2\, \sigma_H^{n-1}} \le \sqrt{\frac{\sigma_H^{n-1}(S)}{\lambda_1}} \sqrt{\int_S \sum_{i \in I_H} |\mathrm{grad}_{HS}\, x_i|^2\, \sigma_H^{n-1}},\end{aligned}$$

where the last identity follows from Lord Rayleigh's characterization of the first nontrivial eigenvalue λ_1 of the operator $\mathcal{L}_{HS}$ on S. Now a direct computation gives the pointwise identity $\sum_{i \in I_H} |\mathrm{grad}_{HS}\, x_i|^2 = h - 1$. Hence, putting it all together, we have shown the following:

Theorem 33. *Let $\Omega \subsetneq \mathbb{G}$ be a bounded domain with C^2 boundary $S = \partial D$. Let λ_1 be the first (nontrivial) eigenvalue of the operator $\mathcal{L}_{HS}$ on S. Then*

$$\sqrt{\lambda_1}\, \frac{\mathcal{V}\mathrm{ol}^n(\Omega)}{\sigma_H^{n-1}(S)} \le \frac{\sqrt{h-1}}{h}.$$

We now discuss another geometric inequality, which looks very similar to the last one. More precisely, let S be a C^2-smooth compact hypersurface without boundary. So let us make use of Rayleigh's principle:

$$\lambda_1 \int_S \varphi^2\,\sigma_H^{n-1} \le \int_S |\mathrm{grad}_{HS}\varphi|^2\,\sigma_H^{n-1}$$

for any function $\varphi \in C^2(S\setminus C_S)\cap W^{1,2}{}_{HS}(S,\sigma_H^{n-1})$ satisfying $\int_S \varphi\,\sigma_H^{n-1}=0$. Again, we assume that the center of mass of $S=\partial\Omega$ is placed at $0\in\mathbb{G}$ so that $\int_S x_i\,\sigma_H^{n-1}=0$ for every $i\in I_H$. Hence, similarly as above, we get that

$$\lambda_1\int_S |x_H|^2\,\sigma_H^{n-1}=\lambda_1\sum_{i\in I_H}\int_S x_i^2\,\sigma_H^{n-1}\le\lambda_1\sum_{i\in I_H}\int_S|\mathrm{grad}_{HS}x_i|^2\,\sigma_H^{n-1}=(h-1)\,\sigma_H^{n-1}(S).$$

At this point, we reformulate Corollary 20 as follows:

$$\int_S\big((h-1)+\big\langle(\mathcal{H}_H\nu_H+C_H\nu_H),x_H\big\rangle\big)\,\sigma_H^{n-1}=0.$$

From this identity and the Cauchy–Schwarz inequality, we easily get that

$$\begin{aligned}(h-1)\,\sigma_H^{n-1}(S)&\le\sqrt{\int_S|x_H|^2\,\sigma_H^{n-1}}\sqrt{\int_S\big|\mathcal{H}_H\nu_H+C_H\nu_H\big|^2\,\sigma_H^{n-1}}\\&\le\sqrt{\int_S|x_H|^2\,\sigma_H^{n-1}}\sqrt{\int_S\big(\mathcal{H}_H^2+|C_H\nu_H|^2\big)\,\sigma_H^{n-1}}.\end{aligned}$$

Therefore

$$\frac{\big((h-1)\,\sigma_H^{n-1}(S)\big)^2}{\int_S\big(\mathcal{H}_H^2+|C_H\nu_H|^2\big)\,\sigma_H^{n-1}}\le\int_S|x_H|^2\,\sigma_H^{n-1}$$

and hence

$$\lambda_1\frac{\big((h-1)\,\sigma_H^{n-1}(S)\big)^2}{\int_S\big(\mathcal{H}_H^2+|C_H\nu_H|^2\big)\,\sigma_H^{n-1}}\le(h-1)\,\sigma_H^{n-1}(S),$$

which proves the following:

Theorem 34. *Let $\Omega\subsetneq\mathbb{G}$ be a bounded domain with C^2 boundary $S=\partial D$ and ν the outward-pointing unit normal vector along S. Let λ_1 be the first eigenvalue of the operator $\mathcal{L}_{HS}$ on S. The following upper bound for λ_1 holds*:

$$\lambda_1\le\frac{\int_S\big(\mathcal{H}_H^2+|C_H\nu_H|^2\big)\,\sigma_H^{n-1}}{(h-1)\,\sigma_H^{n-1}(S)}=\frac{\fint_S\big(\mathcal{H}_H^2+|C_H\nu_H|^2\big)\,\sigma_H^{n-1}}{h-1}.$$

4. Horizontal linear isoperimetric inequalities

Let $S \subset \mathbb{G}$ be a compact hypersurface of class C^2 with (or without) boundary. Let x_H be the horizontal position vector of S and set $x_{HS} := x_H - g_H \nu_H$ where $g_H = \langle x_H, \nu_H \rangle$ is the horizontal support function of S; see Lemma 19. We recall that

$$\int_S \big((h-1) + g_H \mathcal{H}_H + \langle C_H \nu_H, x_{HS} \rangle\big)\, \sigma_H^{n-1} = \int_{\partial S} \langle x_H, \eta_{HS} \rangle\, \sigma_H^{n-2}; \tag{13}$$

see Corollary 20. Note that, if $\partial S = \varnothing$, then the boundary integral vanishes. From this we easily get that

$$(h-1)\, \sigma_H^{n-1}(S) \leq \int_S \big(|g_H|\,|\mathcal{H}_H| + |\langle C_H \nu_H, x_{HS} \rangle|\big)\, \sigma_H^{n-1} + \int_{\partial S} |\langle x_H, \eta_{HS} \rangle|\, \sigma_H^{n-2}. \tag{14}$$

Remark 35 (assumptions on ϱ). Let $\varrho(x) = \varrho(0, x) = \|x\|_\varrho$ be a homogeneous norm on $\mathbb{G}$ and let $\varrho(x, y) = \|y^{-1} \bullet x\|_\varrho$ be the associated (homogeneous) distance on $\mathbb{G}$. In this section we assume the following:

(i) ϱ is piecewise C^1 outside the diagonal of $\mathbb{G}$;

(ii) $|\mathrm{grad}_H \varrho| \leq 1$ at each regular point of ϱ;

(iii) $|x_H| \leq \varrho(x, 0)$ for all $x \in \mathbb{G}$.

Example 36. On the Heisenberg group $\mathbb{H}^n$, the CC-distance d_{CC} satisfies these assumptions. Another example is the distance associated with the Koranyi norm defined by $\|x\|_\varrho := \varrho(x) = \sqrt[4]{|x_H|^4 + 16t^2}$ for $x = \exp(x_H, t) \in \mathbb{H}^n$. This norm is homogeneous and C^∞-smooth out of $0 \in \mathbb{H}^n$ and satisfies conditions (ii) and (iii). This example can easily be generalized to any Carnot group having step 2 and satisfying $C_H^\alpha C_{H_2}^\beta = -\mathbf{1}_{H_i} \delta_\alpha^\beta$, $(\alpha, \beta \in I_{H_2})$. Actually, in this case, one can show that the homogeneous norm $\|\cdot\|_\varrho$, defined by $\|x\|_\varrho := \sqrt[4]{|x_H|^4 + 16|x_{H_2}|^2}$ for all $x = \exp(x_H, x_{H_2})$ satisfies all the conditions in Remark 35.

Let R be the radius of the ϱ-ball $B_\varrho(0, R)$, centered at the identity 0 of the group $\mathbb{G}$ and circumscribed about S. It is important to remark that, because of the left-invariance of the H-perimeter, we may replace 0 with any $x \in \mathbb{G}$. Below, we shall estimate (by the Cauchy–Schwarz inequality) the right-hand side of (14). To this aim, note that $g_H \leq |x_H| \leq \|x\|_\varrho$. So we have

$$(h-1)\, \sigma_H^{n-1}(S) \leq R\bigg(\int_S \big(|\mathcal{H}_H| + |C_H \nu_H|\big)\, \sigma_H^{n-1} + \sigma_H^{n-2}(\partial S) \bigg), \tag{15}$$

which is a linear inequality. Obviously, if S is H-minimal, i.e., $\mathcal{H}_H = 0$, we have

$$(h-1)\, \sigma_H^{n-1}(S) \leq R\bigg(\int_S |C_H \nu_H|\, \sigma_H^{n-1} + \sigma_H^{n-2}(\partial S) \bigg). \tag{16}$$

Furthermore, if $\mathcal{H}_H^0 := \max\{\mathcal{H}_H(x) | x \in S\}$, one gets

$$\sigma_H^{n-1}(S)\big((h-1) - R\mathcal{H}_H^0\big) \leq R\bigg(\int_S |C_H\nu_H|\,\sigma_H^{n-1} + \sigma_H^{n-2}(\partial S)\bigg). \tag{17}$$

Equivalently, we have

$$R \geq \frac{(h-1)\,\sigma_H^{n-1}(S)}{\mathcal{H}_H^0\sigma_H^{n-1}(S) + \big(\int_S |C_H\nu_H|\,\sigma_H^{n-1} + \sigma_H^{n-2}(\partial S)\big)}, \tag{18}$$

and, by assuming $R\mathcal{H}_H^0 < h-1$, we also get

$$\sigma_H^{n-1}(S) \leq \frac{R\big(\int_S |C_H\nu_H|\,\sigma_H^{n-1} + \sigma_H^{n-2}(\partial S)\big)}{(h-1) - R\mathcal{H}_H^0}. \tag{19}$$

Here, we just remark that there are no closed compact H-minimal hypersurfaces immersed in Carnot groups. This fact can be proved by using the first variation formula of the H-perimeter; see [Montefalcone 2012]. The previous formulae have been proved for hypersurfaces with boundary, but they hold even if $\partial S = \varnothing$. More precisely:

Proposition 37. *Let $S \subset \mathbb{G}$ be a compact hypersurface of class C^2 without boundary. Let R be the radius of the ϱ-ball $B_\varrho(0, R)$, centered at the identity 0 of the group $\mathbb{G}$ and circumscribed about S. Then*

$$(h-1)\,\sigma_H^{n-1}(S) \leq R\int_{\mathcal{U}} \big(|\mathcal{H}_H| + |C_H\nu_H|\big)\,\sigma_H^{n-1}; \tag{20}$$

$$R \geq \frac{(h-1)\,\sigma_H^{n-1}(S)}{\mathcal{H}_H^0\sigma_H^{n-1}(S) + \int_S |C_H\nu_H|\,\sigma_H^{n-1}}; \tag{21}$$

$$\sigma_H^{n-1}(S) \leq \frac{R\int_S |C_H\nu_H|\,\sigma_H^{n-1}}{(h-1) - R\mathcal{H}_H^0}. \tag{22}$$

4.1. ***Application: a weak monotonicity formula.*** In the sequel, we shall set $S_t = S \cap B_\varrho(x,t)$. The "natural" monotonicity formula which can be deduced from the inequality (15) is contained in:

Proposition 38. *The following inequality holds for $\mathcal{L}^1$-a.e. $t > 0$:*

$$-\frac{d}{dt}\frac{\sigma_H^{n-1}(S_t)}{t^{h-1}} \leq \frac{1}{t^{h-1}}\bigg(\int_{S_t} \big(|\mathcal{H}_H| + |C_H\nu_H|\big)\,\sigma_H^{n-1} + \sigma_H^{n-2}(\partial S \cap B_\varrho(x,t))\bigg). \tag{23}$$

Proof. Since we are assuming that the homogeneous distance ϱ is smooth (at least piecewise C^1), by applying the classical Sard's theorem we get that S_t is a C^2-smooth manifold with boundary for $\mathcal{L}^1$-a.e. $t > 0$ (or, equivalently, this claim follows by intersecting S with the boundary of a ϱ-ball $B_\varrho(x,t)$ centered at x and of radius t). So let us apply formula (13) for the set S_t. We have

$$(h-1)\,\sigma_H^{n-1}(S_t) \le t\biggl(\int_{S_t} \bigl(|\mathcal{H}_H| + |C_H\nu_H|\bigr)\,\sigma_H^{n-1} + \sigma_H^{n-2}(\partial S_t)\biggr),$$

where t is the radius of a ϱ-ball centered at x and intersecting S. Since

$$\partial S_t = \{\partial S \cap B_\varrho(x,t)\} \cup \{\partial B_\varrho(x,t) \cap S\},$$

we get

$$(h-1)\,\sigma_H^{n-1}(S_t) \le t\biggl(\underbrace{\int_{S_t} \bigl(|\mathcal{H}_H| + |C_H\nu_H|\bigr)\,\sigma_H^{n-1}}_{=:\mathcal{A}(t)} + \underbrace{\sigma_H^{n-2}\bigl(\partial S \cap B_\varrho(x,t)\bigr)}_{=:\mathcal{B}(t)} + \sigma_H^{n-2}\bigl(\partial B_\varrho(x,t) \cap S\bigr)\biggr). \tag{24}$$

Now let us consider the function $\psi(y) := \|y^{-1} \bullet x\|_\varrho$ for all $y \in S$. By hypothesis, ψ is a C^1-smooth function — at least piecewise — satisfying $|\mathrm{grad}_H\,\psi| \le 1$; see Remark 35. So we may apply the coarea formula to this function. Since $|\mathrm{grad}_{HS}\,\psi| \le |\mathrm{grad}_H\,\psi|$, we easily get that

$$\begin{aligned}\sigma_H^{n-1}(S_{t_1}) - \sigma_H^{n-1}(S_t) &\ge \int_{S_{t_1}\setminus S_t} |\mathrm{grad}_{HS}\,\psi|\,\sigma_H^{n-1} = \int_t^{t_1} \sigma_H^{n-2}\{\psi^{-1}[s] \cap S\}\,ds \\ &= \int_t^{t_1} \sigma_H^{n-2}\bigl(\partial B_\varrho(x,s) \cap S\bigr)\,ds.\end{aligned}$$

From the last inequality we infer that

$$\frac{d}{dt}\sigma_H^{n-1}(S_t) \ge \sigma_H^{n-2}\bigl(\partial B_\varrho(x,t) \cap S\bigr)$$

for $\mathcal{L}^1$-a.e. $t > 0$. Hence, from this inequality and (24), we obtain

$$(h-1)\,\sigma_H^{n-1}(S_t) \le t\biggl(\mathcal{A}(t) + \mathcal{B}(t) + \frac{d}{dt}\sigma_H^{n-1}(S_t)\biggr),$$

which is an equivalent form of (23). □

We have to notice however that, in order to prove an "intrinsic" isoperimetric inequality, the number $(h-1)$ in the previous differential inequality *is not the correct one*, which is $(Q-1)$. This fact motivates a further study, made by the author in [Montefalcone 2009; 2010].

5. A theorem about nonhorizontal graphs in 2-step Carnot groups

We begin by describing our result in the simpler setting of the first Heisenberg group $\mathbb{H}^1$; see also [Montefalcone 2007b]. For the notation, see Example 7.

Theorem 39 (Heinz's estimate for T-graphs). *Let $S = \{p = \exp(x, y, t) \in \mathbb{H}^1 : t = f(x, y)$ for all $(x, y) \in \mathbb{R}^2\}$ be a T-graph of class C^2 over the xy-plane. If $|\mathcal{H}_H| \geq C > 0$, then*

$$C\, \mathcal{H}^2_{\mathrm{Eu}}(\mathcal{P}_{xy}(\mathcal{U})) \leq \mathcal{H}^1_{\mathrm{Eu}}(\mathcal{P}_{xy}(\partial\mathcal{U}))$$

for every C^1-smooth relatively compact open set $\mathcal{U} \subset S$. Hence, taking $\mathcal{U} := S \cap C_r(\mathcal{T})$, where $C_r(\mathcal{T})$ denotes a vertical cylinder of radius r around the T-axis $\mathcal{T} := \{p = \exp(0, 0, t) \in \mathbb{H}^1,\ t \in \mathbb{R}\}$, yields, for every $r > 0$,

$$r \leq \frac{2}{C}.$$

It follows that any entire xy-graph of class C^2 having constant (or just bounded) horizontal mean curvature $\mathcal{H}_H$ must be necessarily a H-minimal surface. To see this fact, it is enough to send $r \to +\infty$. The proof of the previous theorem is elementary. More precisely, one uses the following identity:

$$-\int_{\mathcal{U}} \mathcal{H}_H \varpi\, \sigma^2_H = \int_{\partial\mathcal{U}} \nu_H \lrcorner\ d\theta,$$

where $\theta = T^* = dt + \frac{y dx - x dy}{2}$ denotes the dual 1-form to the vertical direction T. We also have to remark that $\varpi \sigma^2_H = -d\theta = dx \wedge dy$. The previous theorem is a generalization to our context of a classical result obtained in [Heinz 1955]. This was generalized in [Chern 1965] and then by other authors in a number of different directions.

Below, we shall restrict ourselves to consider only 2-step Carnot groups.

Definition 40 (nonhorizontal graphs in 2-step Carnot groups). Let $\mathbb{G}$ be a 2-step Carnot group and let $Z = \sum_{\alpha \in I_V} z_\alpha X_\alpha \in V$ be a constant vertical vector. In this case, for the sake of simplicity, we reorder the variables in $\mathfrak{g}$ as $x \equiv (x_{Z^\perp}, x_Z)$, where $x_Z := \langle x, Z\rangle \in \mathbb{R}$ and $x_{Z^\perp} := x - x_Z Z \in Z^\perp$. Then, we say that $S \subset \mathbb{G}$ is a Z-graph (over the hyperplane $Z^\perp$) if there exists a function $\psi : Z^\perp \to \mathbb{R}$ such that $S = \{p = \exp(x_{Z^\perp}, \psi(x_{Z^\perp})) \in \mathbb{G},\ x_{Z^\perp} \in Z^\perp\}$.

Let us fix a constant vertical vector $Z \in V$ and let $S = \{p = \exp(x_{Z^\perp}, \psi(x_{Z^\perp})) \in \mathbb{G},\ x_{Z^\perp} \in Z^\perp\}$ be a Z-graph of class C^2 over the $Z^\perp$-hyperplane. For the sake of simplicity and without loss of generality, we may take $Z = X_\alpha$ for a fixed index $\alpha \in I_V = \{h+1, \dots, n\}$.

Now let us define a differential $(n-2)$-form on $S \subset \mathbb{G}$ by setting

$$\xi^\alpha := \left(\nu_H \lrcorner X_\alpha \lrcorner \sigma^n_R\right)\big|_{S \setminus C_S} \in \Lambda^2(T^* S).$$

This differential $(n-2)$-form ξ^α is well-defined out of C_S and we have to compute its exterior derivative. Below we will briefly sketch a proof, which can also be found in [Montefalcone 2007a]; see Claim 3.22.

Lemma 41. *At each NC point,*

$$d\xi^\alpha|_{S\setminus C_S} = -\mathcal{H}_H \varpi_\alpha \sigma_H^{n-1}|_{S\setminus C_S}.$$

Proof. Let us set $\zeta_j := (X_\alpha \lrcorner X_j \lrcorner \sigma_R^n)|_S$ for any $\alpha \in I_V$ and $j \in I_H$ and compute $d\zeta_j := d(X_\alpha \lrcorner X_j \lrcorner \sigma_R^n)|_S$. Let $\mathbb{G}$ be a k-step Carnot group. We claim that

$$d\zeta_j|_{S\setminus C_S} = \sum_{k=\alpha+1}^{n} C_{\alpha j}^k \,(X_k \lrcorner \sigma_R^n)|_{S\setminus C_S} = \sum_{k=\alpha+1}^{n} C_{\alpha j}^k \,\nu_k\, \sigma_R^{n-1}|_{S\setminus C_S}. \tag{25}$$

The proof of this claim is just a long, but elementary, calculation. Since we are assuming that $\mathbb{G}$ has step 2, using the properties of the Carnot structural constants yields $C_{\alpha j}^k = 0$ whenever $j, k \in I_H$ and $\alpha \in I_V$. Hence $d\zeta_j = 0$ for every $j \in I_H$. By linearity $\xi^\alpha = -\sum_{j\in I_H} \nu_H^j \zeta_j$, where $\nu_H^j = \langle \nu_H, X_j\rangle$ for any $j \in I_H$. It follows easily that $d\xi^\alpha = -\mathcal{H}_H \varpi_\alpha \sigma_H^{n-1}$, as wished. □

Theorem 42 (Heinz's estimate for nonhorizontal graphs in 2-step Carnot groups). *Let $\mathbb{G}$ be a 2-step Carnot group and let $Z \in V$ be a constant vertical vector. Furthermore, let S be a Z-graph of class C^2 over the $Z^\perp$-hyperplane. If $|\mathcal{H}_H| \geq C > 0$, then*

$$C\, \mathcal{H}_{\mathrm{Eu}}^{n-1}(\mathcal{P}_{Z^\perp}(\mathcal{U})) \leq \mathcal{H}_{\mathrm{Eu}}^{n-2}(\mathcal{P}_{Z^\perp}(\partial\mathcal{U})) \tag{26}$$

for every C^1-smooth relatively compact open set $\mathcal{U} \subset S$. Hence, taking $\mathcal{U} := S \cap C_r(\mathfrak{X})$, where $C_r(\mathfrak{X})$ denotes a euclidean cylinder of radius r around the Z-axis given by $\mathfrak{X} := \{p = \exp(0_{Z^\perp}, t) \in \mathbb{G},\ t \in \mathbb{R}\}$, yields, for every $r > 0$,

$$r \leq \frac{n-1}{C}. \tag{27}$$

Proof. Without loss of generality, we may assume $-\mathcal{H}_H \geq C > 0$ and take $Z = X_\alpha$ for some fixed index $\alpha \in I_V$. In this case, one has

$$\varpi_\alpha\, \sigma_H^{n-1}|_S = \nu_\alpha\, \sigma_R^{n-1}|_S = \big(X_\alpha \lrcorner \sigma_R^n\big)\big|_S = d\mathcal{H}_{\mathrm{Eu}}^{n-1} \llcorner X_\alpha^\perp,$$

where the last identity follows from our assumption that S is a X_α-graph. By using Lemma 41 and Stokes' formula, we obtain the integral identity

$$-\int_{\mathcal{U}} \mathcal{H}_H \varpi_\alpha\, \sigma_H^{n-1} = \int_{\partial\mathcal{U}} \nu_H \lrcorner X_\alpha \lrcorner \sigma_R^n.$$

Furthermore, we have

$$-\int_{\mathcal{U}} \mathcal{H}_H \varpi_\alpha \, \sigma_H^{n-1} = -\int_{\mathcal{P}_{X_\alpha^\perp}(\mathcal{U})} \mathcal{H}_H \, d\mathcal{H}_{\mathrm{Eu}}^{n-1},$$

$$\int \big(\nu_H \lrcorner \, d\mathcal{H}_{\mathrm{Eu}}^{n-1}\big)\big|_{\mathcal{P}_{X_\alpha^\perp}(\partial\mathcal{U})} = \int \langle \nu_H, \eta \rangle \, d\mathcal{H}_{\mathrm{Eu}}^{n-2} \llcorner \mathcal{P}_{X_\alpha^\perp}(\partial\mathcal{U}).$$

Putting it all together, we get $C\mathcal{H}_{\mathrm{Eu}}^{n-1}(\mathcal{P}_{X_\alpha^\perp}(\mathcal{U})) \leq \mathcal{H}_{\mathrm{Eu}}^{n-2}\big(\mathcal{P}_{X_\alpha^\perp}(\partial\mathcal{U})\big)$, which proves (26) when $Z = X_\alpha$. The thesis follows by linearity. Finally, (27) follows from (26) and the elementary calculation

$$\frac{\mathcal{H}_{\mathrm{Eu}}^{n-2}(\partial B_{\mathrm{Eu}}^{n-1})}{\mathcal{H}_{\mathrm{Eu}}^{n-1}(B_{\mathrm{Eu}}^{n-1})} = n-1,$$

where B_{Eu}^{n-1} denotes a euclidean unit ball in $Z^\perp \cong \mathbb{R}^{n-1}$. □

It follows that an entire Z-graph of class C^2 over the $Z^\perp$-hyperplane having constant (or bounded) horizontal mean curvature $\mathcal{H}_H$ must be necessarily a H-minimal hypersurface.

6. Local Poincaré-type inequality

By using an elementary technique, somehow analogous to the one used in Section 4, we will state a local Poincaré-type inequality for smooth compactly supported functions on NC domains. First we need the following:

Definition 43. Let $S \subset \mathbb{G}$ be a hypersurface of class C^2 and let $\mathcal{U} \subseteq S$ be an open domain. We say that $\mathcal{U}$ is *uniformly noncharacteristic* (abbreviated UNC) if

$$\sup_{x\in\mathcal{U}} |\varpi(x)| = \sup_{x\in\mathcal{U}} \frac{|\mathcal{P}_V \nu(x)|}{|\mathcal{P}_H \nu(x)|} < +\infty.$$

We stress that

$$|C_H \nu_H| = \Big|\sum_{\alpha\in I_V} \omega_\alpha C_H^\alpha \nu_H\Big| \leq \sum_{\alpha\in I_V} |\omega_\alpha| \|C_H^\alpha\|_{\mathrm{Gr}} \leq \frac{C}{|\mathcal{P}_H \nu|}, \tag{28}$$

where $C := \sum_{\alpha\in I_V} \|C_H^\alpha\|_{\mathrm{Gr}}$ only depends on the structural constants of $\mathfrak{g}$. Set

$$R_{\mathcal{U}} := \frac{1}{2\big[\|\mathcal{H}_H\|_{L^\infty(\mathcal{U})} + C\|\varpi\|_{L^\infty(\mathcal{U})}\big]}.$$

From (28), we have $|C_H \nu_H| \leq C \max_{\alpha\in I_V} |\varpi_\alpha|$. Moreover, $\int_B |\varpi_\alpha| \, \sigma_H^{n-1} = \int_B |\nu_\alpha| \, \sigma_R^{n-1} \leq \sigma_R^{n-1}(B)$ for every Borel set $B \subseteq S$.

Theorem 44. *Let $S \subset \mathbb{G}$ be a hypersurface of class C^2. Let $\mathcal{U} \subset S$ be a uniformly NC open domain. Then, for all $x \in \mathcal{U}$ and for all $R \leq \min\{\mathrm{dist}_\varrho(x, \partial\mathcal{U}), R_{\mathcal{U}}\}$,*

$$\left(\int_{\mathcal{U}_R} |\psi|^p\, \sigma_H^{n-1}\right)^{\frac{1}{p}} \leq C_p\, R\left(\int_{\mathcal{U}_R} |\mathrm{grad}_{HS}\psi|^p\, \sigma_H^{n-1}\right)^{\frac{1}{p}}, \quad p\in[1,+\infty[, \tag{29}$$

for every $\psi \in C^1_{HS}(\mathcal{U}_R)\cap C_0(\mathcal{U}_R)$. *More generally, let* $\widetilde{\mathcal{U}}\subset\mathcal{U}$ *be a bounded open subset of* $\mathcal{U}$ *with smooth boundary such that* $\mathrm{diam}_\varrho(\widetilde{\mathcal{U}})\leq 2\min\{\mathrm{dist}_\varrho(x,\partial\mathcal{U}), R_{\mathcal{U}}\}$. *Then*

$$\left(\int_{\widetilde{\mathcal{U}}} |\psi|^p\, \sigma_H^{n-1}\right)^{\frac{1}{p}} \leq C_p\, \mathrm{diam}_\varrho(\widetilde{\mathcal{U}})\left(\int_{\widetilde{\mathcal{U}}} |\mathrm{grad}_{HS}\psi|^p\, \sigma_H^{n-1}\right)^{\frac{1}{p}}, \quad p\in[1,+\infty[, \tag{30}$$

for every $\psi \in C^1_{HS}(\widetilde{\mathcal{U}})\cap C_0(\widetilde{\mathcal{U}})$.

In this theorem one can take $C_p := \dfrac{2p}{2h-3}$.

Proof. Let us set $\psi_\varepsilon := \sqrt{\varepsilon^2+\psi^2}$ ($\varepsilon\geq 0$). By applying Theorem 14 with $X=\psi_\varepsilon x_H$ we get

$$\int_{\mathcal{U}_R} \{\psi_\varepsilon\big((h-1)+g_H\mathcal{H}_H+\langle C_H\nu_H, x_{HS}\rangle\big)+\langle \mathrm{grad}_{HS}\psi_\varepsilon, x_H\rangle\}\,\sigma_H^{n-1} = \int_{\partial\mathcal{U}_R} \psi_\varepsilon\langle x_H,\eta_{HS}\rangle\,\sigma_H^{n-2},$$

and so

$$\begin{aligned}(h-1)\int_{\mathcal{U}_R}\psi_\varepsilon\sigma_H^{n-1} &\leq R\left(\int_{\mathcal{U}_R}\big[\psi_\varepsilon\big(|\mathcal{H}_H|+|C_H\nu_H|\big)+|\mathrm{grad}_{HS}\psi_\varepsilon|\big]\sigma_H^{n-1}+\int_{\partial\mathcal{U}_R}\psi_\varepsilon\sigma_H^{n-2}\right)\\ &\leq R\big(\|\mathcal{H}_H\|_{L^\infty(\mathcal{U}_R)}+C\|\varpi\|_{L^\infty(\mathcal{U}_R)}\big)\int_{\mathcal{U}_R}\psi_\varepsilon\sigma_H^{n-1}\\ &\quad + R\left(\int_{\mathcal{U}_R}|\mathrm{grad}_{HS}\psi_\varepsilon|\sigma_H^{n-1}+\int_{\partial\mathcal{U}_R}\psi_\varepsilon\sigma_H^{n-2}\right).\end{aligned}$$

By using Fatou's lemma and the estimate $R\leq R_{\mathcal{U}}$ we get that

$$\begin{aligned}&(h-1)\int_{\mathcal{U}_R}|\psi|\,\sigma_H^{n-1}\\ &\quad\leq (h-1)\liminf_{\varepsilon\to 0^+}\int_{\mathcal{U}_R}\psi_\varepsilon\,\sigma_H^{n-1}\\ &\quad\leq \frac{1}{2}\lim_{\varepsilon\to 0^+}\int_{\mathcal{U}_R}\psi_\varepsilon\,\sigma_H^{n-1}+R\lim_{\varepsilon\to 0^+}\left(\int_{\mathcal{U}_R}|\mathrm{grad}_{HS}\psi_\varepsilon|\,\sigma_H^{n-1}+\int_{\partial\mathcal{U}_R}\psi_\varepsilon\,\sigma_H^{n-2}\right).\end{aligned}$$

Obviously, $\psi_\varepsilon\to|\psi|$ and $|\mathrm{grad}_{HS}\psi_\varepsilon|\to|\mathrm{grad}_{HS}\psi|$ as long as $\varepsilon\to 0$; moreover $|\psi|=0$ along $\partial\mathcal{U}_R$. Now since, as it is well-known, $\big|\mathrm{grad}_{HS}|\psi|\big|\leq|\mathrm{grad}_{HS}\psi|$, we easily get the claim by Lebesgue's dominated convergence theorem. So we have shown that

$$\int_{\mathcal{U}_R}|\psi|\,\sigma_H^{n-1}\leq\frac{2R}{2h-3}\int_{\mathcal{U}_R}|\mathrm{grad}_{HS}\psi|\,\sigma_H^{n-1}$$

for every $\psi \in C^1_{HS}(\mathcal{U}_R) \cap C_0(\mathcal{U}_R)$. Finally, the general case follows by Hölder's inequality. More precisely, let us use the last inequality with $|\psi|$ replaced by $|\psi|^p$. This implies

$$\begin{aligned}\int_{\mathcal{U}_R} |\psi|^p\,\sigma_H^{n-1} &\leq \frac{2R}{(2h-3)}\int_{\mathcal{U}_R} p|\psi|^{p-1}|\mathrm{grad}_{HS}\psi|\,\sigma_H^{n-1}\\ &\leq \frac{2pR}{(2h-3)}\left(\int_{\mathcal{U}_R}|\psi|^{(p-1)q}\,\sigma_H^{n-1}\right)^{\frac{1}{q}}\left(\int_{\mathcal{U}_R}|\mathrm{grad}_{HS}\psi|^p\,\sigma_H^{n-1}\right)^{\frac{1}{p}},\end{aligned}$$

where $\frac{1}{p}+\frac{1}{q}=1$. This achieves the proof of (29). Finally, (30) can be proved by repeating the same arguments as above, just by replacing R with $\mathrm{diam}(\widetilde{\mathcal{U}})$. $\square$

With some extra hypotheses one can show that (29) still holds up to the characteristic set.

Theorem 45. *Let $S \subset \mathbb{G}$ be a hypersurface of class C^2 with (or without) boundary ∂S. We assume that S has bounded horizontal mean curvature $\mathcal{H}_H$ and that* $\dim C_S < n-2$. *Furthermore, let $\mathcal{U}_\epsilon$ $(\epsilon > 0)$ be a family of open subsets of S with C^1 boundaries, such that:*

(i) $C_S \subset \mathcal{U}_\epsilon$ *for every* $\epsilon > 0$;

(ii) $\sigma_R^{n-1}(\mathcal{U}_\epsilon) \to 0$ *for* $\epsilon \to 0^+$;

(iii) $\int_{\mathcal{U}_\epsilon} |\mathcal{P}_H\nu|\,\sigma_R^{n-2} \to 0$ *for* $\epsilon \to 0^+$.

Then, for every $x \in S$ and every (small enough) $\epsilon > 0$ there exists $R_0 := R_0(x,\epsilon) \leq$ $\mathrm{dist}_\varrho(x,\partial S)$ *such that*

$$\left(\int_{S_R}|\psi|^p\,\sigma_H^{n-1}\right)^{\frac{1}{p}} \leq C_p R\left(\int_{S_R}|\mathrm{grad}_{HS}\psi|^p\,\sigma_H^{n-1}\right)^{\frac{1}{p}},\quad p\in[1,+\infty[, \tag{31}$$

holds for every $\psi \in C^1_{HS}(S_R)\cap C_0(S_R)$ and every $R \leq R_0$, where

$$R_0 := \min\left\{\mathrm{dist}_\varrho(x,\partial S),\ \frac{1}{2\big[C\big(1+\|\varpi\|_{L^\infty(S_R\setminus\mathcal{U}_\epsilon)}\big)+\|\mathcal{H}_H\|_{L^\infty(S_R)}\big]}\right\}.$$

Proof. Set $\psi_\varepsilon := \sqrt{\varepsilon^2+\psi^2}$ $(0\leq\varepsilon<1)$. We shall prove the theorem for $p=1$. The general case will follow by using Hölder's inequality. Let $\mathcal{U}_\epsilon$ $(\epsilon>0)$ be as above. Fix $\epsilon_0>0$. For every $\epsilon\leq\epsilon_0$ one has

$$\int_{\mathcal{U}_\epsilon}\psi_\varepsilon|C_H\nu_H|\,\sigma_H^{n-1} \leq 2C\,\|\psi\|_{L^\infty(\mathcal{U}_{\epsilon_0})}\sigma_R^{n-1}(\mathcal{U}_\epsilon),$$

where we have put $C := \sum_{\alpha\in I_V}\|C_H^\alpha\|_{\mathrm{Gr}}$. Furthermore (ii) implies that for every $\delta>0$ there exists $\epsilon_\delta>0$ such that $\sigma_R^{n-1}(\mathcal{U}_\epsilon)<\delta$ whenever $\epsilon<\epsilon_\delta$. Taking

$\widetilde{\delta} \leq \dfrac{\int_{S_R} \psi_\varepsilon\, \sigma_H^{n-1}}{2\|\psi\|_{L^\infty(\mathcal{U}_{\epsilon_0})}}$, one gets

$$\int_{\mathcal{U}_\epsilon} \psi_\varepsilon |C_H \nu_H|\, \sigma_H^{n-1} \leq C \int_{S_R} \psi_\varepsilon\, \sigma_H^{n-1}$$

for every $\epsilon \leq \min\{\epsilon_{\widetilde{\delta}}, \epsilon_0\}$. Moreover, for any $\epsilon \in]0, \min\{\epsilon_{\widetilde{\delta}}, \epsilon_0\}[$, one has

$$\int_{S_R \setminus \mathcal{U}_\epsilon} \psi_\varepsilon |C_H \nu_H|\, \sigma_H^{n-1} \leq C \|\varpi\|_{L^\infty(S_R \setminus \mathcal{U}_\epsilon)} \int_{S_R} \psi_\varepsilon\, \sigma_H^{n-1}.$$

It follows that

$$\int_{S_R} \psi_\varepsilon |C_H \nu_H|\, \sigma_H^{n-1} \leq C\big(1 + \|\varpi\|_{L^\infty(S_R \setminus \mathcal{U}_\epsilon)}\big) \int_{S_R} \psi_\varepsilon\, \sigma_H^{n-1}.$$

Since, by hypothesis, the horizontal mean curvature is bounded, we clearly have

$$\int_{S_R} \psi_\varepsilon |\mathcal{H}_H|\, \sigma_H^{n-1} \leq \|\mathcal{H}_H\|_{L^\infty(S_R)} \int_{S_R} \psi_\varepsilon\, \sigma_H^{n-1}.$$

Applying Theorem 14 with $X = \psi_\varepsilon x_H$ (and arguing as in the proof of Theorem 44) yields

$$\begin{aligned}(h-1)\int_{S_R} \psi_\varepsilon\, \sigma_H^{n-1} &\leq R\bigg(\int_{S_R} \{\psi_\varepsilon\big(|\mathcal{H}_H| + |C_H \nu_H|\big) + |\mathrm{grad}_{HS}\, \psi_\varepsilon|\}\, \sigma_H^{n-1} + \int_{\partial S_R} \psi_\varepsilon\, \sigma_H^{n-2}\bigg)\\ &\leq R\big[C\big(1 + \|\varpi\|_{L^\infty(S_R \setminus \mathcal{U}_\epsilon)}\big) + \|\mathcal{H}_H\|_{L^\infty(S_R)}\big] \int_{S_R} \psi_\varepsilon\, \sigma_H^{n-1}\\ &\qquad + R\bigg(\int_{S_R} |\mathrm{grad}_{HS}\, \psi_\varepsilon|\, \sigma_H^{n-1} + \int_{\partial S_R} \psi_\varepsilon\, \sigma_H^{n-2}\bigg).\end{aligned}$$

So if $R \leq R_0$, one gets

$$\int_{S_R} \psi_\varepsilon\, \sigma_H^{n-1} \leq \frac{2R}{2h-3}\bigg(\int_{S_R} |\mathrm{grad}_{HS}\, \psi_\varepsilon|\, \sigma_H^{n-1} + \int_{\partial S_R} \psi_\varepsilon\, \sigma_H^{n-2}\bigg).$$

We have $\psi_\varepsilon \to |\psi|$ and $|\mathrm{grad}_{HS}\, \psi_\varepsilon| \to |\mathrm{grad}_{HS}\, \psi|$ as long as $\varepsilon \to 0$ and $|\psi| = 0$ along ∂S_R. Since $\big|\mathrm{grad}_{HS} |\psi|\big| \leq |\mathrm{grad}_{HS}\, \psi|$, the thesis follows from Fatou's lemma and Lebesgue's dominated convergence theorem. □

6.1. ***A Caccioppoli-type inequality.*** Our final result is a generalization of the classical *Caccioppoli inequality* (see, for instance, [Ambrosio 1997]) for the operator $\mathcal{L}_{HS}$ on smooth hypersurfaces.

Let $S \subset \mathbb{G}$ be a hypersurface of class C^2 and set $S_R := S \cap B_\varrho(x, R)$ for any $x \in \mathbb{G}$. We are going to consider the functions satisfying, in the distributional sense,

$$-\mathcal{L}_{HS}\phi = \psi \quad \text{on } S_R, \tag{32}$$

whenever $\psi \in L^2(S_R, \sigma_H^{n-1})$.

So let us take a function $\zeta \in C^1_{HS}(S_R) \cap C_0(S_R)$ such that $0 \leq \zeta \leq 1$, $\zeta = 1$ on $S_{R/2} = S \cap B_\varrho(0, R/2)$ and $|\mathrm{grad}_{HS}\zeta| \leq C_0/R$. Inserting into the above equation the function $\varphi = \zeta^2(\phi - \phi_0)$, where $\phi_0 \in \mathbb{R}$ is a fixed constant, and then integrating over S_R, yields

$$\underbrace{\int_{S_R} \zeta^2 |\mathrm{grad}_{HS}\phi|^2\, \sigma_H^{n-1}}_{=: I_1} + \underbrace{2\int_{S_R} \zeta(\phi - \phi_0)\langle \mathrm{grad}_{HS}\zeta, \mathrm{grad}_{HS}\phi\rangle\, \sigma_H^{n-1}}_{=: I_2} = \underbrace{\int_{S_R} \psi\zeta^2(\phi-\phi_0)\, \sigma_H^{n-1}}_{=: I_3}. \tag{33}$$

We have

$$I_2 \leq \frac{1}{2}\int_{S_R} |\zeta|^2 |\mathrm{grad}_{HS}\phi|^2\, \sigma_H^{n-1} + 2\underbrace{\int_{S_R} |\phi-\phi_0|^2 |\mathrm{grad}_{HS}\phi|^2\, \sigma_H^{n-1}}_{=: I_4}.$$

Moreover $I_4 \leq 2C_0^2/R^2 \|\phi - \phi_0\|_{L^2(S_R)}$. Now let us estimate the third integral I_3:

$$\begin{aligned}
\int_{S_R} \psi\zeta^2(\phi-\phi_0)\, \sigma_H^{n-1} &= \int_{S_R} 2\left((2R\psi)\frac{\zeta^2(\phi-\phi_0)}{4R}\right)\sigma_H^{n-1} \\
&\leq 4R^2 \int_{S_R} \psi^2\, \sigma_H^{n-1} + \frac{1}{16R^2}\int_{S_R} \zeta^4|\phi-\phi_0|^2\, \sigma_H^{n-1} \\
&\leq 4R^2 \int_{S_R} 2\psi^2\, \sigma_H^{n-1} + \frac{1}{R^2}\int_{S_R} |\phi-\phi_0|^2\, \sigma_H^{n-1}.
\end{aligned}$$

Since $\zeta = 1$ on $S_{R/2}$, using the previous estimates yields

$$\int_{S_{R/2}} |\mathrm{grad}_{HS}\phi|^2\, \sigma_H^{n-1} \leq \frac{2C_0^2+1}{R^2}\int_{S_R} |\phi-\phi_0|^2\, \sigma_H^{n-1} + 4R^2\int_{S_R} \psi^2\, \sigma_H^{n-1}.$$

We summarize these calculations, as follows:

Theorem 46. *Let $S \subset \mathbb{G}$ be a hypersurface of class C^2; let $\phi_0 \in \mathbb{R}$ and let ϕ be a distributional solution to the equation $-\mathcal{L}_{HS}\phi = \psi$ on S_R, where $\psi \in L^2(S_R, \sigma_H^{n-1})$. Then, there exists a positive constant $C > 0$ such that the following "Caccioppoli-type" inequality holds*:

$$\int_{S_{R/2}} |\mathrm{grad}_{HS}\phi|^2\, \sigma_H^{n-1} \leq C\left(\frac{1}{R^2}\int_{S_R} |\phi-\phi_0|^2\, \sigma_H^{n-1} + R^2\int_{S_R} \psi^2\, \sigma_H^{n-1}\right)$$

for every (small enough) $R > 0$, where $S_R := S \cap B_\varrho(x, R)$, for any $x \in S$.

Acknowledgements

The authors wish to thank the anonymous referee for helpful comments that improved the paper.

References

[Ambrosio 1997] L. Ambrosio, *Corso introduttivo alla teoria geometrica della misura ed alle superfici minime*, Scuola Normale Superiore, Pisa, 1997. MR 2000k:49001 Zbl 0977.49028

[Ambrosio 2001] L. Ambrosio, "Some fine properties of sets of finite perimeter in Ahlfors regular metric measure spaces", *Adv. Math.* **159**:1 (2001), 51–67. MR 2002b:31002 Zbl 1002.28004

[Ambrosio and Kirchheim 2000a] L. Ambrosio and B. Kirchheim, "Currents in metric spaces", *Acta Math.* **185**:1 (2000), 1–80. MR 2001k:49095 Zbl 0984.49025

[Ambrosio and Kirchheim 2000b] L. Ambrosio and B. Kirchheim, "Rectifiable sets in metric and Banach spaces", *Math. Ann.* **318**:3 (2000), 527–555. MR 2003a:28009 Zbl 0966.28002

[Ambrosio et al. 2006] L. Ambrosio, F. Serra Cassano, and D. Vittone, "Intrinsic regular hypersurfaces in Heisenberg groups", *J. Geom. Anal.* **16**:2 (2006), 187–232. MR 2007g:49072 Zbl 1085.49045

[Balogh 2003] Z. M. Balogh, "Size of characteristic sets and functions with prescribed gradient", *J. Reine Angew. Math.* **564** (2003), 63–83. MR 2005d:43007 Zbl 1051.53024

[Balogh et al. 2009] Z. M. Balogh, J. T. Tyson, and B. Warhurst, "Sub-Riemannian vs. Euclidean dimension comparison and fractal geometry on Carnot groups", *Adv. Math.* **220**:2 (2009), 560–619. MR 2009k:53064 Zbl 1155.22011

[Balogh et al. 2010] Z. M. Balogh, C. Pintea, and H. Rohner, "Size of tangencies to non-involutive distributions", preprint, 2010, http://org.uib.no/hcaa/BaloghPinteaRohner.pdf. Zbl 06137664

[Burago and Zalgaller 1988] Y. D. Burago and V. A. Zalgaller, *Geometric inequalities*, Grundlehren Math. Wiss. **285**, Springer, Berlin, 1988. MR 89b:52020 Zbl 0633.53002

[Capogna et al. 1994] L. Capogna, D. Danielli, and N. Garofalo, "The geometric Sobolev embedding for vector fields and the isoperimetric inequality", *Comm. Anal. Geom.* **2**:2 (1994), 203–215. MR 96d:46032 Zbl 0864.46018

[Capogna et al. 2010] L. Capogna, G. Citti, and M. Manfredini, "Smoothness of Lipschitz minimal intrinsic graphs in Heisenberg groups $\mathbb{H}^n$, $n > 1$", *J. Reine Angew. Math.* **648** (2010), 75–110. MR 2012g:35367 Zbl 1217.53060

[Chavel 1978] I. Chavel, "On A. Hurwitz' method in isoperimetric inequalities", *Proc. Amer. Math. Soc.* **71**:2 (1978), 275–279. MR 58 #12845 Zbl 0395.52007

[Chavel 1984] I. Chavel, *Eigenvalues in Riemannian geometry*, Pure and Applied Mathematics **115**, Academic Press, Orlando, FL, 1984. MR 86g:58140 Zbl 0551.53001

[Chavel 1993] I. Chavel, *Riemannian geometry: A modern introduction*, Cambridge Tracts in Mathematics **108**, Cambridge University Press, 1993. MR 95j:53001 Zbl 0810.53001

[Chavel 2001] I. Chavel, *Isoperimetric inequalities: Differential geometric and analytic perspectives*, Cambridge Tracts in Mathematics **145**, Cambridge University Press, 2001. MR 2002h:58040 Zbl 0988.51019

[Cheeger 1970] J. Cheeger, "A lower bound for the smallest eigenvalue of the Laplacian", pp. 195–199 in *Problems in analysis: A symposium in honor of Salomon Bochner* (Princeton, NJ, 1969), edited by R. C. Gunning, Princeton University Press, 1970. MR 53 #6645 Zbl 0212.44903

[Cheeger and Kleiner 2010] J. Cheeger and B. Kleiner, "Differentiating maps into L^1, and the geometry of BV functions", *Ann. of Math.* (2) **171**:2 (2010), 1347–1385. MR 2012a:30132 Zbl 1194.22009

[Cheng et al. 2005] J.-H. Cheng, J.-F. Hwang, A. Malchiodi, and P. Yang, "Minimal surfaces in pseudohermitian geometry", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **4**:1 (2005), 129–177. MR 2006f:53008 Zbl 1158.53306

[Chern 1965] S.-S. Chern, "On the curvatures of a piece of hypersurface in euclidean space", *Abh. Math. Sem. Univ. Hamburg* **29** (1965), 77–91. MR 32 #6376 Zbl 0147.20901

[Corwin and Greenleaf 1990] L. J. Corwin and F. P. Greenleaf, *Representations of nilpotent Lie groups and their applications, I: Basic theory and examples*, Cambridge Studies in Advanced Mathematics **18**, Cambridge University Press, 1990. MR 92b:22007 Zbl 0704.22007

[Danielli et al. 2007] D. Danielli, N. Garofalo, and D. M. Nhieu, "Sub-Riemannian calculus on hypersurfaces in Carnot groups", *Adv. Math.* **215**:1 (2007), 292–378. MR 2009h:53061 Zbl 1129.53017

[Danielli et al. 2010] D. Danielli, N. Garofalo, and D. M. Nhieu, "Sub-Riemannian calculus and monotonicity of the perimeter for graphical strips", *Math. Z.* **265**:3 (2010), 617–637. MR 2011e:53033 Zbl 1206.53034

[Danielli et al. 2012] D. Danielli, N. Garofalo, and D. M. Nhieu, "Integrability of the sub-Riemannian mean curvature of surfaces in the Heisenberg group", *Proc. Amer. Math. Soc.* **140**:3 (2012), 811–821. MR 2012m:53060 Zbl 06024089

[David and Semmes 1997] G. David and S. Semmes, *Fractured fractals and broken dreams: Self-similar geometry through metric and measure*, Oxford Lecture Ser. Math. Appl. **7**, Oxford University Press, New York, 1997. MR 99h:28018 Zbl 0887.54001

[Federer 1969] H. Federer, *Geometric measure theory*, Die Grundlehren der mathematischen Wissenschaften **153**, Springer, New York, NY, 1969. MR 41 #1976 Zbl 0176.00801

[Franchi et al. 2001] B. Franchi, R. Serapioni, and F. Serra Cassano, "Rectifiability and perimeter in the Heisenberg group", *Math. Ann.* **321**:3 (2001), 479–531. MR 2003g:49062 Zbl 1057.49032

[Franchi et al. 2003a] B. Franchi, R. Serapioni, and F. Serra Cassano, "On the structure of finite perimeter sets in step 2 Carnot groups", *J. Geom. Anal.* **13**:3 (2003), 421–466. MR 2004i:49085 Zbl 1064.49033

[Franchi et al. 2003b] B. Franchi, R. Serapioni, and F. Serra Cassano, "Regular hypersurfaces, intrinsic perimeter and implicit function theorem in Carnot groups", *Comm. Anal. Geom.* **11**:5 (2003), 909–944. MR 2004m:28008 Zbl 1077.22008

[Franchi et al. 2007] B. Franchi, R. Serapioni, and F. Serra Cassano, "Regular submanifolds, graphs and area formula in Heisenberg groups", *Adv. Math.* **211**:1 (2007), 152–203. MR 2008h:49030 Zbl 1125.28002

[Garofalo and Nhieu 1996] N. Garofalo and D.-M. Nhieu, "Isoperimetric and Sobolev inequalities for Carnot–Carathéodory spaces and the existence of minimal surfaces", *Comm. Pure Appl. Math.* **49**:10 (1996), 1081–1144. MR 97i:58032 Zbl 0880.35032

[Ge 1992] Z. Ge, "Betti numbers, characteristic classes and sub-Riemannian geometry", *Illinois J. Math.* **36**:3 (1992), 372–403. MR 94f:53038 Zbl 0769.53014

[Gromov 1996] M. Gromov, "Carnot–Carathéodory spaces seen from within", pp. 79–323 in *Sub-Riemannian geometry*, edited by A. Bellaïche and J.-J. Risler, Progr. Math. **144**, Birkhäuser, Basel, 1996. MR 2000f:53034 Zbl 0864.53025

[Heintze 1988] E. Heintze, "Extrinsic upper bounds for λ_1", *Math. Ann.* **280**:3 (1988), 389–402. MR 89f:53091 Zbl 0628.53044

[Heinz 1955] E. Heinz, "Über Flächen mit eindeutiger Projektion auf eine Ebene, deren Krümmungen durch Ungleichungen eingeschränkt sind", *Math. Ann.* **129** (1955), 451–454. MR 17,189a Zbl 0065.3720

[Hladky and Pauls 2008] R. K. Hladky and S. D. Pauls, "Constant mean curvature surfaces in sub-Riemannian geometry", *J. Differential Geom.* **79**:1 (2008), 111–139. MR 2009m:53070 Zbl 1156.53038

[Koiller et al. 2001] J. Koiller, P. R. Rodrigues, and P. Pitanga, "Non-holonomic connections following Élie Cartan", *An. Acad. Brasil. Ciênc.* **73**:2 (2001), 165–190. MR 2002i:37097 Zbl 0999.70016

[Lee 2003] J. M. Lee, *Introduction to smooth manifolds*, Graduate Texts in Mathematics **218**, Springer, New York, 2003. MR 2003k:58001 Zbl 1030.53001

[Li 1982] P. Li, "Poincaré inequalities on Riemannian manifolds", pp. 73–83 in *Seminar on differential geometry*, edited by S. T. Yau, Annals of Mathematics Studies **102**, Princeton University Press, 1982. Zbl 0492.5303

[Magnani 2002] V. Magnani, *Elements of geometric measure theory on sub-Riemannian groups*, Scuola Normale Superiore, Pisa, 2002. MR 2005m:28002 Zbl 1064.28007

[Magnani and Vittone 2008] V. Magnani and D. Vittone, "An intrinsic measure for submanifolds in stratified groups", *J. Reine Angew. Math.* **619** (2008), 203–232. MR 2009g:53046 Zbl 1146.53022

[Milnor 1976] J. Milnor, "Curvatures of left invariant metrics on Lie groups", *Advances in Math.* **21**:3 (1976), 293–329. MR 54 #12970 Zbl 0341.53030

[Montefalcone 2005] F. Montefalcone, "Some relations among volume, intrinsic perimeter and one-dimensional restrictions of BV functions in Carnot groups", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **4**:1 (2005), 79–128. MR 2006e:49088 Zbl 1150.49022

[Montefalcone 2007a] F. Montefalcone, "Hypersurfaces and variational formulas in sub-Riemannian Carnot groups", *J. Math. Pures Appl.* (9) **87**:5 (2007), 453–494. MR 2008d:53035 Zbl 1123.49035

[Montefalcone 2007b] F. Montefalcone, "Ipersuperfici e formule variazionali in gruppi di Carnot sub-Riemanniani", lecture notes of Bruno Pini Mathematical Analysis Seminar, Università di Bologna, 2007, http://www.dm.unibo.it/seminario-pini/pdf/SemPiniMont.pdf.

[Montefalcone 2009] F. Montefalcone, "Isoperimetric, Sobolev and Poincaré inequalities on hypersurfaces in sub-Riemannian Carnot groups", preprint, 2009. arXiv 0910.5656

[Montefalcone 2010] F. Montefalcone, "Isoperimetric and Sobolev inequalities on hypersurfaces in sub-Riemannian Carnot groups", preprint, 2010. arXiv 1012.2442

[Montefalcone 2012] F. Montefalcone, "Stable H-minimal hypersurfaces", 2012. arXiv 1203.5972

[Montgomery 2002] R. Montgomery, *A tour of subriemannian geometries, their geodesics and applications*, Mathematical Surveys and Monographs **91**, American Mathematical Society, Providence, RI, 2002. MR 2002m:53045 Zbl 1044.53022

[Pansu 1982] P. Pansu, *Gèometrie du group d'Heisenberg*, Ph.D. thesis, Université Paris Diderot, 1982.

[Pansu 1989] P. Pansu, "Métriques de Carnot–Carathéodory et quasiisométries des espaces symétriques de rang un", *Ann. of Math.* (2) **129**:1 (1989), 1–60. MR 90e:53058 Zbl 0678.53042

[Pansu 2005] P. Pansu, "Submanifolds and differential forms in Carnot manifolds, after M. Gromov et M. Rumin", Université Paris-Sud, 2005, http://www.math.u-psud.fr/~pansu/trento_pansu.pdf.

[Reilly 1977] R. C. Reilly, "On the first eigenvalue of the Laplacian for compact submanifolds of Euclidean space", *Comment. Math. Helv.* **52**:4 (1977), 525–533. MR 58 #2657 Zbl 0382.53038

[Ritoré and Rosales 2008] M. Ritoré and C. Rosales, "Area-stationary surfaces in the Heisenberg group $\mathbb{H}^1$", *Adv. Math.* **219**:2 (2008), 633–671. MR 2009h:49075 Zbl 1158.53022

[Stein 1993] E. M. Stein, *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, Princeton Mathematical Series **43**, Princeton University Press, 1993. MR 95c:42002 Zbl 0821.42001

[Strichartz 1986] R. S. Strichartz, "Sub-Riemannian geometry", *J. Differential Geom.* **24**:2 (1986), 221–263. Correction in **30**:2 (1989), 595–596. MR 88b:53055 Zbl 0609.53021

[Taylor 2006] M. E. Taylor, *Measure theory and integration*, Graduate Studies in Mathematics **76**, American Mathematical Society, Providence, RI, 2006. MR 2007g:28002 Zbl 1139.28001

[Varadarajan 1974] V. S. Varadarajan, *Lie groups, Lie algebras, and their representations*, Prentice-Hall, Englewood Cliffs, NJ, 1974. MR 51 #13113 Zbl 0371.22001

[Varopoulos et al. 1992] N. T. Varopoulos, L. Saloff-Coste, and T. Coulhon, *Analysis and geometry on groups*, Cambridge Tracts in Mathematics **100**, Cambridge University Press, 1992. MR 95f:43008 Zbl 0813.22003

[Vershik and Gershkovich 1994] A. M. Vershik and V. Y. Gershkovich, "Nonholonomic dynamical systems: Geometry of distributions and variational problems", pp. 1–81 in *Dynamical systems*, vol. 7, edited by S. P. Novikov, Encycl. Math. Sci. **16**, Springer, Berlin, 1994. Zbl 0797.58007

[Yau 1975] S. T. Yau, "Isoperimetric constants and the first eigenvalue of a compact Riemannian manifold", *Ann. Sci. École Norm. Sup.* (4) **8**:4 (1975), 487–507. MR 53 #1478 Zbl 0325.53039

Received March 28, 2012. Revised October 1, 2012.

FRANCESCOPAOLO MONTEFALCONE
DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ DEGLI STUDI DI PADOVA
VIA TRIESTE, 63
35121 PADOVA
ITALY

montefal@math.unipd.it

PACIFIC JOURNAL OF MATHEMATICS
Vol. 263, No. 1, 2013
dx.doi.org/10.2140/pjm.2013.263.207

FIXED POINTS OF ENDOMORPHISMS OF VIRTUALLY FREE GROUPS

PEDRO V. SILVA

A fixed point theorem is proved for inverse transducers, which leads to an automata-theoretic proof of the fixed point subgroup of an endomorphism of a finitely generated virtually free group being finitely generated. If the endomorphism is uniformly continuous for the hyperbolic metric, it is proved that the set of regular fixed points in the hyperbolic boundary has finitely many orbits under the action of the finite fixed points. In the automorphism case, it is shown that these regular fixed points are either exponentially stable attractors or exponentially stable repellers.

1. Introduction

Throughout the paper, the ambient groups are assumed to be finitely generated.

Gersten [1987] proved that the fixed point subgroup of a free group automorphism φ is finitely generated. Using a different approach, Cooper [1987] gave an alternative proof, proving also that the fixed points of the continuous extension of φ to the boundary of the free group is, in some sense, finitely generated. Bestvina and Handel [1992] achieved a major breakthrough with their innovative train track techniques, bounding the rank of the fixed point subgroup and the generating set for the infinite fixed points. Their approach was pursued by Maslakova [2003], who considered the problem of effectively computing a basis for the fixed point subgroup. The paper turned out to contain some errors, and subsequently a new paper by Bogopolski and Maslakova [2012] was posted on arXiv with the purpose of correcting these errors.

Gersten's result was generalized to further classes of groups and endomorphisms in subsequent years. Goldstein and Turner extended it to monomorphisms of free groups [1985] and to arbitrary endomorphisms [1986]. Collins and Turner extended it to automorphisms of free products of freely indecomposable groups [1994];

The author acknowledges support from the European Regional Development Fund through the COMPETE program and from the Portuguese government through Fundação para a Ciência e a Tecnologia, under the project PEst-C/MAT/UI0144/2011.

MSC2010: primary 20E05, 20E36, 20F67; secondary 68Q45, 37B25.

Keywords: virtually free groups, endomorphisms, fixed points, hyperbolic boundary, classification of fixed points.

see the survey by Ventura [2002]. With respect to automorphisms, the widest generalization is to hyperbolic groups and is due to Paulin [1989].

Sykiotis [2002] extended Collins and Turner's result to arbitrary endomorphisms of virtually free groups using symmetric endomorphisms; see also [Sykiotis 2007] for further results on symmetric endomorphisms. In [Silva 2012], we generalized Goldstein and Turner's automata-theoretic proof to arbitrary endomorphisms of free products of cyclic groups. In the present paper, this result is extended to arbitrary endomorphisms of virtually free groups, providing an automata-theoretic alternative to Sykiotis' result.

This is done by reducing the problem to the rationality of some languages associated to a finite inverse transducer, and subsequent application of Anisimov and Seifert's theorem.

Infinite fixed points of automorphisms of free groups were discussed by Gaboriau, Jaeger, Levitt, and Lustig [Gaboriau et al. 1998], where it is remarked in particular that some of the results would hold for virtually free groups with some adaptations.

In [Silva 2010], we discussed infinite fixed points for monomorphisms of free products of cyclic groups, the group case of a more general setting based on the concept of special confluent rewriting system. These results are now extended to endomorphisms with finite kernel of virtually free groups (which are precisely the uniformly continuous endomorphisms for the hyperbolic metric), and we discuss the dynamical nature of the regular fixed points in the automorphism case, generalizing the results of [Gaboriau et al. 1998] on free groups.

The paper is organized as follows. Section 2 is devoted to preliminaries on groups and automata. We discuss inverse transducers in Section 3, proving a useful fixed point theorem. In Section 4 we prove that the fixed point subgroup is finitely generated for arbitrary endomorphisms of a (finitely generated) virtually free group G.

In Section 5 we get a rewriting system with good properties to represent the elements of G, and in Section 6 we use it to construct a simple model for the hyperbolic boundary of G. We study uniformly continuous endomorphisms in Section 7 and in Section 8 we prove that the infinite fixed points of such endomorphisms are, in some sense, finitely generated.

The classification of the infinite fixed points of automorphisms is performed in Section 9, and Section 10 includes an example and some open problems.

2. Preliminaries

Throughout the paper, we assume alphabets to be *finite*. We start with some group-theoretic definitions. Given an alphabet A, we denote by A^{-1} a set of *formal inverses* of A, and write $\widetilde{A} = A \cup A^{-1}$. We extend the mapping $a \mapsto a^{-1}$ to an involution of the free monoid $\widetilde{A}^*$ in the obvious way. As usual, the *free group on A* is the quotient of $\widetilde{A}^*$ by the congruence generated by the relation $\{(aa^{-1}, 1) : a \in \widetilde{A}\}$. We denote by $\theta : \widetilde{A}^* \to F_A$ the canonical morphism.

Let

$$R_A = \widetilde{A}^* \setminus \left(\bigcup_{a \in \widetilde{A}} \widetilde{A}^* aa^{-1} \widetilde{A}^* \right)$$

be the subset of all *reduced words* in $\widetilde{A}^*$. It is well known that, for every $g \in F_A$, $g\theta^{-1}$ contains a unique reduced word, denoted by $\bar{g}$. We also write $\bar{u} = \overline{u\theta}$ for every $u \in \widetilde{A}^*$. Note that the equivalence $u\theta = v\theta \Leftrightarrow \bar{u} = \bar{v}$ holds for all $u, v \in \widetilde{A}^*$.

A group G is *virtually free* if G has a free subgroup F of finite index. In view of Nielsen's theorem, it is well-known that F can be assumed to be normal, and is finitely generated if G is finitely generated itself. Therefore every finitely generated virtually free group G admits a decomposition as a disjoint union

$$G = F \cup Fb_1 \cup \cdots \cup Fb_m,$$

where $F \trianglelefteq G$ is a free group of finite rank and $b_1, \ldots, b_m \in G$.

We shall need also some basic concepts from automata theory.

Let A be a (finite) alphabet. A subset of A^* is called an *A-language*. We say that $\mathscr{A} = (Q, q_0, T, \delta)$ is a (finite) *deterministic A-automaton* if

- Q is a (finite) set,
- $q_0 \in Q$ and $T \subseteq Q$,
- $\delta : Q \times A \to Q$ is a partial mapping.

We extend δ to a partial mapping $Q \times A^* \to Q$ by induction through

$$(q, 1)\delta = q, \quad (q, ua)\delta = ((q, u)\delta, a)\delta \quad (u \in A^*, a \in A).$$

When the automaton is clear from the context, we write $qu = (q, u)\delta$. We can view $\mathscr{A}$ as a directed graph with edges labeled by letters $a \in A$ by identifying $(p, a)\delta = q$ with the edge $p \xrightarrow{a} q$. We denote by $E(\mathscr{A}) \subseteq Q \times A \times Q$ the set of all such edges.

A *finite nontrivial path* in $\mathscr{A}$ is a sequence

$$p_0 \xrightarrow{a_1} p_1 \xrightarrow{a_2} \cdots \xrightarrow{a_n} p_n$$

with $(p_{i-1}, a_i, p_i) \in E(\mathscr{A})$ for $i = 1, \ldots, n$. Its *label* is the word $a_1 \cdots a_n \in A^*$. It is said to be a *successful* path if $p_0 = q_0$ and $p_n \in T$. We also consider the *trivial path* $p \xrightarrow{1} p$ for $p \in Q$. It is successful if $p = q_0 \in T$.

The *language $L(\mathscr{A})$ recognized by $\mathscr{A}$* is the set of all labels of successful paths in $\mathscr{A}$. Equivalently, $L(\mathscr{A}) = \{u \in A^* : q_0u \in T\}$. If $(p_{i-1}, a_i, p_i) \in E(\mathscr{A})$ for every

$i \in \mathbb{N}$, we may also consider the *infinite path*

$$p_0 \xrightarrow{a_1} p_1 \xrightarrow{a_2} p_2 \xrightarrow{a_3} \cdots .$$

Its label is the (right) infinite word $a_1a_2a_3\cdots$. We denote by A^ω the set of all (right) infinite words on the alphabet A, and also write $A^\infty = A^* \cup A^\omega$. We denote by $L_\omega(\mathscr{A})$ the set of labels of all infinite paths $q_0 \longrightarrow \cdots$ in $\mathscr{A}$.

Given $u \in A^*$ and $\alpha \in A^\infty$, we say that u is a *prefix* of α and write $u \leq \alpha$ if $\alpha = u\beta$ for some $\beta \in A^\infty$. By convention, this includes the case $\alpha \leq \alpha$ for $\alpha \in A^\omega$. For every $n \in \mathbb{N}$, we denote by $\alpha^{[n]}$ the prefix of length n of α, applying the convention that $\alpha^{[n]} = \alpha$ if $n > |\alpha|$.

It is immediate that $(A^\infty, \leq)$ is a complete $\wedge$-semilattice: given $\alpha, \beta \in A^\infty$, $\alpha \wedge \beta$ is the longest common prefix of α and β (or α if $\alpha = \beta \in A^\omega$). The $\wedge$ operator will play a crucial role in later sections of the paper.

The *star* operator on A-languages is defined by

$$L^* = \bigcup_{n \geq 0} L^n,$$

where $L^0 = \{1\}$. An A-language L is said to be *rational* if L can be obtained from finite A-languages using finitely many times the union, product, and star operators (this is called a *rational expression*). Alternatively, by Kleene's theorem [Berstel 1979, Section III], L is rational if and only if it is recognized by a finite deterministic A-automaton $\mathscr{A}$. The definition through rational expressions generalizes to subsets of an arbitrary group in the obvious way. Moreover, if we fix a homomorphism $\pi : A^* \to G$, the rational subsets of G are the images by π of the rational A-languages. For obvious reasons, we shall be dealing mostly with matched homomorphisms. A homomorphism $\pi : \widetilde{A}^* \to G$ is said to be *matched* if $a^{-1}\pi = (a\pi)^{-1}$ for every $a \in A$. For details on rational languages and subsets, the reader is referred to [Berstel 1979; Sakarovitch 2003].

We shall need also the following classical result of Anisimov and Seifert.

Proposition 2.1 [Sakarovitch 2003, Proposition II.6.2]. *Let H be a subgroup of a group G. Then H is a rational subset of G if and only if H is finitely generated.*

We end this section with an elementary observation that helps us to establish that fixed point subgroups are finitely generated.

Proposition 2.2. *Let $\pi : \widetilde{A}^* \to G$ be a matched epimorphism and let $X \subseteq G$. Let $\mathscr{A}$ be a finite $\widetilde{A}$-automaton such that*

(i) $L(\mathscr{A}) \subseteq X\pi^{-1}$,

(ii) $L(\mathscr{A}) \cap x\pi^{-1} \neq \varnothing$ *for every* $x \in X$.

Then X is a rational subset of G.

Proof. It follows immediately that $X = (L(\mathscr{A}))\pi$, so X is a rational subset of G. $\square$

3. Inverse transducers

Given a finite alphabet A, we say that $\mathscr{T} = (Q, q_0, \delta, \lambda)$ is a (finite) *deterministic A-transducer* if

- Q is a (finite) set,
- $q_0 \in Q$,
- $\delta : Q \times A \to Q$ and $\lambda : Q \times A \to A^*$ are mappings.

As in the automaton case, we may extend δ to a mapping $Q \times A^* \to Q$. Similarly, we extend λ to a mapping $Q \times A^* \to A^*$ through

$$(q, 1)\lambda = 1, \quad (q, ua)\lambda = (q, u)\lambda((q, u)\delta, a)\lambda \quad (u \in A^*, a \in A).$$

When the transducer is clear from the context, we write $qa = (q, a)\delta$. We can view $\mathscr{T}$ as a directed graph with edges labeled by elements of $A \times A^*$ (represented in the form $a|w$) by identifying $(p, a)\delta = q$, $(p, a)\lambda = w$ with the edge $p \xrightarrow{a|w} q$. The set of all such edges is denoted by $E(\mathscr{T}) \subseteq Q \times A \times A^* \times Q$. If $pu = q$ and $(p, u)\lambda = v$, we also write $p \xrightarrow{u|v} q$ and call it a path in $\mathscr{T}$.

It is immediate that, given $u \in A^*$, there exists exactly one path in $\mathscr{T}$ of the form $q_0 \xrightarrow{u|v} q$. We write $u\widehat{\mathscr{T}} = v$, thus defining a mapping $\widehat{\mathscr{T}} : A^* \to A^*$.

Assume now that $\mathscr{T} = (Q, q_0, \delta, \lambda)$ is a deterministic $\widetilde{A}$-transducer such that

$$p \xrightarrow{a|u} q \text{ is an edge of } \mathscr{T} \text{ if and only if } q \xrightarrow{a^{-1}|u^{-1}} p \text{ is an edge of } \mathscr{T}.$$

Then $\mathscr{T}$ is said to be *inverse*.

Proposition 3.1. *Let $\mathscr{T} = (Q, q_0, \delta, \lambda)$ be an inverse $\widetilde{A}$-transducer. Then*

(i) *$\delta : Q \times \widetilde{A}^* \to Q$ induces a mapping $\hat{\delta} : Q \times F_A \to Q$ by $(q, u\theta)\hat{\delta} = (q, u)\delta$,*

(ii) *$\widehat{\mathscr{T}} : \widetilde{A}^* \to \widetilde{A}^*$ induces a mapping $\widetilde{\mathscr{T}} : F_A \to F_A$ by $u\theta\widetilde{\mathscr{T}} = u\widehat{\mathscr{T}}\theta$.*

Proof. (i) Since the free group congruence $\sim$ is generated by the pairs $(aa^{-1}, 1)$, it suffices to show that $(q, vaa^{-1}w)\delta = (q, vw)\delta$ for all $q \in Q$; $v, w \in \widetilde{A}^*$ and $a \in \widetilde{A}$.

Since δ is a full mapping, we have a path

$$q \xrightarrow{v|v'} q_1 \xrightarrow{a|u} q_2 \xrightarrow{a^{-1}|u'} q_3 \xrightarrow{w|w'} q_4 \tag{1}$$

in $\mathscr{T}$. Since $\mathscr{T}$ is inverse (in particular deterministic), we must have $u' = u^{-1}$ and $q_3 = q_1$. Hence we also have a path

$$q \xrightarrow{v|v'} q_1 \xrightarrow{w|w'} q_4$$

and so $(q, vaa^{-1}w)\delta = q_4 = (q, vw)\delta$ as required.

(ii) Similarly to part (i), it suffices to show that $(vaa^{-1}w)\widehat{\mathcal{T}}\theta = (vw)\widehat{\mathcal{T}}\theta$ for all $v, w \in \widetilde{A}^*$ and $a \in \widetilde{A}$.

We consider the path (1) for $q = q_0$. Since $u' = u^{-1}$ and $q_3 = q_1$, we get

$$(vaa^{-1}w)\widehat{\mathcal{T}}\theta = (v'uu^{-1}w')\theta = (v'w')\theta = (vw)\widehat{\mathcal{T}}\theta,$$

as required. □

We now prove one of our main results, generalizing Goldstein and Turner's proof [1986] to mappings induced by inverse transducers.

Theorem 3.2. *Let $\mathcal{T}$ be a finite inverse $\widetilde{A}$-transducer and let $z \in F_A$. Then*

$$L = \{g \in F_A : g\widetilde{\mathcal{T}} = gz\}$$

is rational.

Proof. Write $\mathcal{T} = (Q, q_0, \delta, \lambda)$. For every $g \in F_A$, let $P_1(g) = g^{-1}(g\widetilde{\mathcal{T}}) \in F_A$ and write $q_0g = (q_0, g)\hat{\delta}$, $P(g) = (P_1(g), q_0g)$. Note that $g \in L$ if and only if $P_1(g) = z$. We define a deterministic $\widetilde{A}$-automaton $\mathcal{A}_{\mathcal{T}} = (P, (1, q_0), S, E)$ by

$$P = \{P(g) : g \in F_A\};$$
$$S = P \cap (\{z\} \times Q);$$
$$E = \{(P(g), a, P(ga)) : g \in F_A,\ a \in \widetilde{A}\}.$$

Clearly, $\mathcal{A}_{\mathcal{T}}$ is a possibly infinite automaton. Note that, since $\mathcal{T}$ is inverse, we have $qaa^{-1} = q$ for all $q \in Q$ and $a \in \widetilde{A}$. It follows that, whenever $(p, a, p') \in E$, $(p', a^{-1}, p) \in E$. We say that such edges are the *inverses* of each other.

Since every $w \in \widetilde{A}^*$ labels a unique path $P(1) \xrightarrow{w} P(w\theta)$, it follows that

$$L(\mathcal{A}_{\mathcal{T}}) = L\theta^{-1}.$$

In view of Proposition 2.2, to prove that L is rational it suffices to construct a finite subautomaton $\mathcal{B}_{\mathcal{T}}$ of $\mathcal{A}_{\mathcal{T}}$ such that $\bar{L} \subseteq L(\mathcal{B}_{\mathcal{T}})$.

We now fix

$$M = \max\{|(q, a)\lambda| : q \in Q,\ a \in \widetilde{A}\}, \quad N = \max\{2M + 1, |z|\}$$

and

$$P' = \{P(g) \in P : |P_1(g)| \leq N\}.$$

Since A and $\mathcal{T}$ are finite, so is P'. However, infinitely many $g \in F_A$ may yield the same state $P(g)$.

Given $g \in F_A$, write $g\iota = \bar{g}^{[1]}$. Given $p = (g, q) \in P$, we also write $p\iota = g\iota$. We say that an edge $(p_1, a, p_2) \in E$ is

- *central* if $p_1, p_2 \in P'$,
- *compatible* if it is not central and $p_1\iota = a$.

Lemma 3.3. (i) *There are only finitely many central edges in* $\mathcal{A}_{\mathcal{T}}$.

(ii) *If* $(p_1, a, p_2) \in E$ *is not central, either* (p_1, a, p_2) *or* (p_2, a^{-1}, p_1) *is compatible.*

(iii) *For every* $p \in P$, *there is at most one compatible edge leaving* p.

Proof. (i) A and P' are both finite.

(ii) Assume that (p_1, a, p_2) is neither central nor compatible. Write $p_1 = (g_1, q_1)$ and $p_2 = (g_2, q_2)$. Suppose that $g_1 = 1$. Then $g_2 = P_1(a) = a^{-1}(a\widetilde{\mathcal{T}})$ and so $|g_2| \leq 1 + M \leq N$, in contradiction with (p_1, a, p_2) being noncentral.

Thus $\bar{g}_1 = bu$ for some $b \in \widetilde{A} \setminus \{a\}$ and $u \in R_A$. On the other hand, we have $g_2 = a^{-1} g_1(q_1, a)\lambda$, and so

$$\bar{g}_2 = \overline{a^{-1}bu(q_1, a)\lambda}.$$

If $|u| < M$, then $|g_1|, |g_2| \leq 2M + 1 \leq N$ and (p_1, a, p_2) is central, a contradiction. Thus $|u| \geq M \geq |(q_1, a)\lambda|$ and so $g_2\iota = a^{-1}$. Thus (p_2, a^{-1}, p_1) is compatible.

(iii) Any compatible edge leaving p must be labeled by $p\iota$, and $\mathcal{A}_{\mathcal{T}}$ is deterministic. □

A (possibly infinite) path $q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} \cdots$ in $\mathcal{A}_{\mathcal{T}}$ is

- *central* if all the vertices in it are in P',
- *compatible* if all the edges in it are compatible and no intermediate vertex is in P'.

Lemma 3.4. *Let* $u \in \bar{L}$. *Then there exists a path*

$$(1, q_0) = p_0' \xrightarrow{u_0} p_0'' \xrightarrow{v_1} p_1 \xrightarrow{w_1^{-1}} p_1' \xrightarrow{u_1} \cdots \xrightarrow{v_n} p_n \xrightarrow{w_n^{-1}} p_n' \xrightarrow{u_n} p_n'' \in S$$

in $\mathcal{A}_{\mathcal{T}}$ *such that*

(i) $u = u_0 v_1 w_1^{-1} u_1 \cdots v_n w_n^{-1} u_n$,

(ii) *the paths* $p_j' \xrightarrow{u_j} p_j''$ *are central,*

(iii) *the paths* $p_{j-1}'' \xrightarrow{v_j} p_j$ *and* $p_j' \xrightarrow{w_j} p_j$ *are compatible,*

(iv) $p_j \notin P'$ *if both* v_j *and* w_j *are nonempty.*

Proof. Since $S \subseteq P'$ by definition of N, there exists a path

$$(1, q_0) = p_0' \xrightarrow{u_0} p_0'' \xrightarrow{x_1} p_1' \xrightarrow{u_1} \cdots \xrightarrow{x_n} p_n' \xrightarrow{u_n} p_n'' \in S \tag{2}$$

in $\mathcal{A}_{\mathcal{T}}$ such that $u = u_0 x_1 u_1 \cdots x_n u_n$ and the paths $p_j' \xrightarrow{u_j} p_j''$ (which may be trivial) collect all the occurrences of vertices in P' (and are therefore central).

By Lemma 3.3(ii), if (p, a, r) occurs in a path $p''_{j-1} \xrightarrow{x_j} p'_j$, either (p, a, r) or (r, a^{-1}, p) is compatible. On the other hand, since x_j is reduced, it follows from Lemma 3.3(iii) that $p''_{j-1} \xrightarrow{x_j} p'_j$ can be factored as

$$p''_{j-1} \xrightarrow{v_j} p_j \xrightarrow{w_j^{-1}} p'_j$$

with $p''_{j-1} \xrightarrow{v_j} p_j$ and $p'_j \xrightarrow{w_j} p_j$ compatible. Clearly (iv) holds since no intermediate vertex of $p''_{j-1} \xrightarrow{x_j} p'_j$ belongs to P' by construction. □

We say that a compatible path is *maximal* if it is infinite or cannot be extended (to the right) to produce another compatible path.

Lemma 3.5. *For every $p \in P'$, there exists in $\mathcal{A}_{\mathcal{T}}$ a unique maximal compatible path M_p starting at p.*

Proof. Clearly, every compatible path can be extended to a maximal compatible path. Uniqueness follows from Lemma 3.3(iii). □

We now define

$$P'_1 = \{p \in P' : M_p \text{ has finitely many distinct edges }\}$$

and $P'_2 = P' \setminus P'_1$. Hence M_p contains no cycles if $p \in P'_2$. By Lemma 3.5, if M_p and $M_{p'}$ intersect at vertex $r_{pp'}$, they coincide from $r_{pp'}$ onwards. In particular, if M_p and $M_{p'}$ intersect, then $p \in P'_1$ if and only if $p' \in P'_1$. Let

$$Y = \{(p, p') \in P'_2 \times P'_2 : \ M_p \text{ intersects } M_{p'}\}.$$

For every $(p, p') \in Y$, let $M_p \setminus M_{p'}$ denote the (finite) subpath $p \longrightarrow r_{pp'}$ of M_p. In particular, if $p' = p$, $M_p \setminus M_{p'}$ is the trivial path at p.

Let $\mathcal{B}_{\mathcal{T}}$ be the subautomaton of $\mathcal{A}_{\mathcal{T}}$ containing

- all vertices in P' and all central edges,
- all vertices and edges in the paths M_p $(p \in P'_1)$ and their inverses,
- all vertices and edges in the paths $M_p \setminus M_{p'}$ $((p, p') \in Y)$ and their inverses.

It follows easily from Lemma 3.3(i) and the definitions of P'_1 and $M_p \setminus M_{p'}$ that $\mathcal{B}_{\mathcal{T}}$ is a finite subautomaton of $\mathcal{A}_{\mathcal{T}}$. As remarked before, it suffices to show that $\bar{L} \subseteq L(\mathcal{B}_{\mathcal{T}})$.

Let $u \in \bar{L}$. Since $\mathcal{B}_{\mathcal{T}}$ contains all the central edges of $\mathcal{A}_{\mathcal{T}}$, it suffices to show that all subpaths

$$p''_{j-1} \xrightarrow{v_j} p_j \xrightarrow{w_j^{-1}} p'_j$$

appearing in the factorization provided by Lemma 3.4 are paths in $\mathcal{B}_{\mathcal{T}}$.

Without loss of generality, we may assume that $v_j \neq 1$. If $w_j = 1$, $p''_{j-1} \in P'_1$ and we are done. Hence we may also assume that $w_j \neq 1$. Now, if one of the vertices p''_{j-1}, p'_j is in P'_1, so is the other and we are done, since $\mathcal{B}_{\mathcal{T}}$ contains all the edges in the paths M_p $(p \in P_1')$ and their inverses. Hence we may assume that $p''_{j-1}, p'_j \in P'_2$. It follows that $p_j = r_{p''_{j-1},p'_j}$. (Since $v_j w_j^{-1} \in R_A$, the paths $M_{p''_{j-1}}$ and $M_{p'_j}$ cannot meet before p_j.) Thus $p''_{j-1} \xrightarrow{v_j} p_j$ is $M_{p''_{j-1}} \setminus M_{p'_j}$ and $p'_j \xrightarrow{w_j} p_j$ is $M_{p'_j} \setminus M_{p''_{j-1}}$, and so these are also paths in $\mathcal{B}_{\mathcal{T}}$ as required. □

4. The fixed point subgroup

We can now produce an automata-theoretic proof to Sykiotis' theorem.

Theorem 4.1 [Sykiotis 2002, Proposition 3.4]. *Let φ be an endomorphism of a finitely generated virtually free group. Then* Fix φ *is finitely generated.*

Proof. We consider a decomposition of G as a disjoint union

$$G = Fb_0 \cup Fb_1 \cup \cdots \cup Fb_m, \tag{3}$$

where $F = F_A \trianglelefteq G$ is a free group with A finite and $b_0, \ldots, b_m \in G$ with $b_0 = 1$.

Let $\varphi_0 : F_A \to F_A$ and $\eta : F_A \to \{0, \ldots, m\}$ be defined by

$$g\varphi = (g\varphi_0)b_{g\eta} \quad (g \in F_A).$$

Since the decomposition (3) is disjoint, $g\varphi_0$ and $g\eta$ are both uniquely determined by $g\varphi$, and so both mappings are well defined.

Write $Q = \{0, \ldots, m\}$. For all $i \in Q$ and $a \in \widetilde{A}$, we have $b_i(a\varphi) = h_{i,a}b_{(i,a)\delta}$ for some (unique) $h_{i,a} \in F_A$ and $(i, a)\delta \in Q$. It follows that, for every $j \in Q$, $\mathcal{A}_j = (Q, 0, j, \delta)$ is a well-defined finite deterministic $\widetilde{A}$-automaton. We define also a finite deterministic $\widetilde{A}$-transducer $\mathcal{T} = (Q, 0, \delta, \lambda)$ by taking $(i, a)\lambda = \overline{h_{i,a}}$ for all $i \in Q$ and $a \in \widetilde{A}$.

Assume that

$$i \xrightarrow{a|\overline{h_{i,a}}} (i, a)\delta = j$$

is an edge of $\mathcal{T}$. Then $b_i(a\varphi) = h_{i,a}b_j$ and so

$$b_i = b_i(a\varphi)(a^{-1}\varphi) = h_{i,a}b_j(a^{-1}\varphi) = h_{i,a}h_{j,a^{-1}}b_{(j,a^{-1})\delta}.$$

This yields $h_{i,a}h_{j,a^{-1}} = 1$ and $(j, a^{-1})\delta = i$. Thus there is an edge

$$j \xrightarrow{a^{-1}|\overline{h_{i,a}}^{-1}} (j, a^{-1})\delta = i$$

in $\mathcal{T}$ and so $\mathcal{T}$ is an inverse transducer. We claim that $\widetilde{\mathcal{T}} = \varphi_0$. Indeed, let $g = a_1 \cdots a_n$ $(a_i \in \widetilde{A}_i)$. Then there exists a (unique) path in $\mathcal{T}$ of the form

$$0 = i_0 \xrightarrow{a_1|\overline{h_{i_0,a_1}}} i_1 \xrightarrow{a_2|\overline{h_{i_1,a_2}}} \cdots \xrightarrow{a_n|\overline{h_{i_{n-1},a_n}}} i_n.$$

Moreover, $i_j = (i_{j-1}, a_j)\delta$ for $j = 1, \ldots, n$. It follows that

$$\begin{aligned} g\varphi &= b_{i_0}(a_1\varphi)\cdots(a_n\varphi) = h_{i_0,a_1}b_{i_1}(a_2\varphi)\cdots(a_n\varphi) \\ &= h_{i_0,a_1}h_{i_1,a_2}b_{i_2}(a_3\varphi)\cdots(a_n\varphi) = \cdots = h_{i_0,a_1}\cdots h_{i_{n-1},a_n}b_{i_n} \end{aligned}$$

and so

$$g\varphi_0 = h_{i_0,a_1} \cdots h_{i_{n-1},a_n} = (\overline{h_{i_0,a_1}} \cdots \overline{h_{i_{n-1},a_n}})\theta = g\widetilde{\mathcal{T}}.$$

Thus $\widetilde{\mathcal{T}} = \varphi_0$.

Note that we have also shown that $g\eta = i_n = (0, a_1 \cdots a_n)\delta$. Hence

$$L(\mathcal{A}_j) = \{u \in \widetilde{A}^* : u\theta\eta = j\}. \tag{4}$$

Next let

$$Y = \{(i, j) \in Q \times Q : b_j(b_i\varphi) \in F_A b_i\}.$$

For every $(i, j) \in Y$, let $z_{i,j} \in F_A$ be such that $b_j(b_i\varphi) = z_{i,j}b_i$ and define

$$X_{i,j} = \{g \in F_A : gb_i \in \operatorname{Fix}\varphi \text{ and } g\eta = j\}.$$

We claim that $X_{i,j}$ is a rational subset of F_A for every $(i, j) \in Y$. Indeed, $(gb_i)\varphi = (g\varphi)(b_i\varphi) = (g\varphi_0)b_{g\eta}(b_i\varphi)$. Hence

$$\begin{aligned} X_{i,j} &= \{g \in F_A : (g\varphi_0)b_j(b_i\varphi) = gb_i \text{ and } g\eta = j\} \\ &= \{g \in F_A : (g\varphi_0)z_{i,j}b_i = gb_i \text{ and } g\eta = j\} \\ &= \{g \in F_A : g\varphi_0 = gz_{i,j}^{-1}\} \cap \{g \in F_A : g\eta = j\}. \end{aligned}$$

Writing

$$L_{i,j} = \{g \in F_A : g\varphi_0 = gz_{i,j}^{-1}\},$$

it follows from (4) that $X_{i,j} = L_{i,j} \cap (L(\mathcal{A}_j))\theta$. Since $\varphi_0 = \widetilde{\mathcal{T}}$, it follows from Theorem 3.2 that $X_{i,j}$ is an intersection of two rational subsets of F_A, and is hence rational itself; see [Berstel 1979, Corollary III.2.10].

Now it is easy to check that

$$\operatorname{Fix}\varphi = \bigcup_{i \in Q} \Big(\bigcup\{X_{i,j} : (i, j) \in Y\}\Big)b_i. \tag{5}$$

Indeed, for every $(i, j) \in Y$, we have $X_{i,j}b_i \subseteq \operatorname{Fix}\varphi$ by definition of $X_{i,j}$. Conversely, let $gb_i \in \operatorname{Fix}\varphi$ for some $g \in F_A$ and $i \in Q$. Then $gb_i = (gb_i)\varphi = (g\varphi_0)b_{g\eta}(b_i\varphi)$ and so $b_{g\eta}(b_i\varphi) \in F_A b_i$. Hence $(i, g\eta) \in Y$. Since $g \in X_{i,g\eta}$, (5) holds. Since the $X_{i,j}$ are rational subsets of F_A and therefore of G, it follows that $\operatorname{Fix}\varphi$ is a rational subset of G and is thus finitely generated by Proposition 2.1. □

Unfortunately, our approach does not lead directly to an algorithm to compute a basis of $\operatorname{Fix}\varphi$ (see [Bogopolski and Maslakova 2012]) because it is not clear how to decide in Section 3 whether $p \in P'$ belongs to P'_1 or P'_2 and how to compute the paths M_p and $M_p \setminus M_{p'}$.

5. A good rewriting system

We recall that a (finite) *rewriting system* on A is a (finite) subset $\mathcal{R}$ of $A^* \times A^*$. Given $u, v \in A^*$, we write $u \longrightarrow_{\mathcal{R}} v$ if there exist $(r, s) \in \mathcal{R}$ and $x, y \in A^*$ such that $u = xry$ and $v = xsy$. The reflexive and transitive closure of $\longrightarrow_{\mathcal{R}}$ is denoted by $\longrightarrow^*_{\mathcal{R}}$.

We say that $\mathcal{R}$ is

- *length-reducing* if $|r| > |s|$ for every $(r, s) \in \mathcal{R}$,
- *length-nonincreasing* if $|r| \geq |s|$ for every $(r, s) \in \mathcal{R}$,
- *noetherian* if, for every $u \in A^*$, there is a bound on the length of a chain
$$u \longrightarrow_{\mathcal{R}} v_1 \longrightarrow_{\mathcal{R}} \cdots \longrightarrow_{\mathcal{R}} v_n,$$
- *confluent* if, whenever $u \longrightarrow^*_{\mathcal{R}} v$ and $u \longrightarrow^*_{\mathcal{R}} w$, there exists some $z \in A^*$ such that $v \longrightarrow^*_{\mathcal{R}} z$ and $w \longrightarrow^*_{\mathcal{R}} z$.

A word $u \in A^*$ is *irreducible* if no $v \in A^*$ satisfies $u \longrightarrow_{\mathcal{R}} v$. We denote by $\operatorname{Irr}\mathcal{R}$ the set of all irreducible words in A^* with respect to $\mathcal{R}$.

We introduce now some basic concepts and results from the theory of hyperbolic groups. For details on this class of groups, the reader is referred to [Ghys and de la Harpe 1990].

Let $\pi : \widetilde{A}^* \to G$ be a matched epimorphism with A finite. The *Cayley graph* $\Gamma_A(G)$ of G with respect to π has vertex set G and edges $(g, a, g(a\pi))$ for all $g \in G$ and $a \in \widetilde{A}$. We say that a path $p \xrightarrow{u} q$ in $\Gamma_A(G)$ is a *geodesic* if it has shortest length among all the paths connecting p to q in $\Gamma_A(G)$. We denote by $\operatorname{Geo}_A(G)$ the set of labels of all geodesics in $\Gamma_A(G)$. Note that, since $\Gamma_A(G)$ is vertex-transitive, it is irrelevant whether or not we fix a basepoint.

The *geodesic distance* d_1 on G is defined by taking $d_1(g, h)$ to be the length of a geodesic from g to h. Given $X \subseteq G$ nonempty and $g \in G$, we define
$$d_1(g, X) = \min\{d_1(g, x) : x \in X\}.$$

A *geodesic triangle* in $\Gamma_A(G)$ is a collection of three geodesics
$$P_1 : g_1 \longrightarrow g_2, \quad P_2 : g_2 \longrightarrow g_3, \quad P_3 : g_3 \longrightarrow g_1$$
connecting three vertices $g_1, g_2, g_3 \in G$. Let $V(P_i)$ denote the set of vertices occurring in the path P_i. We say that $\Gamma_A(G)$ is *δ-hyperbolic* for some $\delta \geq 0$ if
$$\forall g \in V(P_1) : d_1(g, V(P_2) \cup V(P_3)) < \delta$$
for every geodesic triangle $\{P_1, P_2, P_3\}$ in $\Gamma_A(G)$. If this happens for some δ, we say that G is *hyperbolic*. It is well known that the concept is independent from both

alphabet and matched epimorphism, but the hyperbolicity constant δ may change. Virtually free groups are among the most important examples of hyperbolic groups.

We now use a theorem of Gilman, Hermiller, Holt, and Rees [Gilman et al. 2007] to prove the following result.

Lemma 5.1. *Let G be a finitely generated virtually free group. Then there exist a finite alphabet A, a matched epimorphism $\pi : \widetilde{A}^* \to G$, and a positive integer N_0 such that, for all $u \in \mathrm{Geo}_A(G)$ and $v \in \widetilde{A}^*$,*

(i) *there exists some $w \in \mathrm{Geo}_A(G)$ such that $w\pi = (uv)\pi$ and*

$$|u \wedge w| \geq |u| - N_0|v|;$$

(ii) *there exists some $z \in \mathrm{Geo}_A(G)$ such that $z\pi = (vu)\pi$ and $|u^{-1} \wedge z^{-1}| \geq |u| - N_0|v|$.*

Proof. (i) By [Gilman et al. 2007, Theorem 1], there exists a finite alphabet A, a matched epimorphism $\pi : \widetilde{A}^* \to G$, and a finite length-reducing rewriting system $\mathcal{R}$ such that $\mathrm{Geo}_A(G) = \mathrm{Irr}\,\mathcal{R}$. The authors also prove that this property characterizes (finitely generated) virtually free groups.

Let $N_0 = 2\max\{|r| : (r, s) \in \mathcal{R}\}$. Suppose that

$$uv = w_0 \longrightarrow_{\mathcal{R}} w_1 \longrightarrow_{\mathcal{R}} \cdots \longrightarrow_{\mathcal{R}} w_n = w$$

is a sequence of reductions leading to a geodesic w. Then $(wv^{-1})\pi = u\pi$ and since u is a geodesic we get $|u| \leq |v| + |w|$ and so $|u| - |w| \leq |v|$. On the other hand, since $\mathcal{R}$ is length-reducing, we get $|u| + |v| = |uv| \geq |w| + n$ and so $n - |v| \leq |u| - |w| \leq |v|$. Thus $n \leq 2|v|$.

Trivially, $|u \wedge w_0| \geq |u|$. Since $u \wedge w_{i-1} \in \mathrm{Geo}_A(G)$, it is immediate that $|u \wedge w_i| > |u \wedge w_{i-1}| - N_0/2$ and so

$$|u \wedge w| = |u \wedge w_n| \geq |u| - n\frac{N_0}{2} \geq |u| - N_0|v|.$$

(ii) The inverse of a geodesic is still a geodesic. By applying (i) to u^{-1} and v^{-1}, we get $(u^{-1}v^{-1})\pi = x\pi$ for some $x \in \mathrm{Geo}_A(G)$ satisfying $|u^{-1} \wedge x| \geq |u^{-1}| - N_0|v^{-1}|$. Then we take $z = x^{-1}$. □

We assume for the remainder of the paper that G is a finitely generated virtually free group, $\pi : \widetilde{A}^* \to G$ a matched epimorphism, and N_0 a positive integer satisfying the conditions of Lemma 5.1. Since G is hyperbolic, it follows from [Epstein et al. 1992, Theorem 3.4.5] that $\mathrm{Geo}_A(G)$ is an automatic structure for G with respect to π (see [Epstein et al. 1992] for definitions), and so the *fellow traveler property* holds for some constant $K_0 > 0$ (which can be taken as $2(\delta + 1)$, if δ is the hyperbolicity constant). This amounts to saying that

$$\forall u, v \in \mathrm{Geo}_A(G) : d_1(u\pi, v\pi) \leq 1 \ \Rightarrow\ \forall n \in \mathbb{N} : d_1(u^{[n]}\pi, v^{[n]}\pi) \leq K_0).$$

We fix a total ordering of $\widetilde{A}$. The *shortlex ordering* of $\widetilde{A}^*$ is defined by

$$u \leq_{sl} v \text{ if } \begin{cases} |u| < |v|, & \text{or} \\ |u| = |v| \text{ and } u = wau',\ v = wbv' \text{ with } a < b \text{ in } \widetilde{A}. \end{cases}$$

This is a well-known well-ordering of $\widetilde{A}^*$, compatible with multiplication on the left and on the right. Let

$$L = \{u \in \mathrm{Geo}_A(G) : u \leq_{sl} v \text{ for every } v \in u\pi\pi^{-1}\}. \tag{6}$$

By [Epstein et al. 1992, Theorem 2.5.1], L is also an automatic structure for G with respect to π, and therefore rational. We note that L is *factorial* (a factor of a word in L is still in L).

Given $g \in G$, let $\bar{g}$ denote the unique word of L representing g. This corresponds precisely to free group reduction if $G = F_A$ and $\pi = \theta$. Since we shall not need free group reduction from now on, we also write $\bar{u} = \overline{u\pi}$ for every $u \in \widetilde{A}^*$ to simplify notation.

Theorem 5.2. *Consider the finite rewriting system $\mathcal{R}'$ on A defined by*

$$\mathcal{R}' = \{(u, \bar{u}) : u \in \widetilde{A}^*,\ |u| \leq K_0N_0 + 1,\ u \neq \bar{u}\}.$$

Then

(i) *$\mathcal{R}'$ is length-nonincreasing, noetherian and confluent,*

(ii) $\mathrm{Irr}\mathcal{R}' = L$.

Proof. (i) $\mathcal{R}'$ is trivially length-nonincreasing, and that it is noetherian follows from

$$(u, \bar{u}) \in \mathcal{R}' \Rightarrow u >_{sl} \bar{u} \tag{7}$$

and $\widetilde{A}^*$ being well-ordered by $\leq_{sl}$, plus compatibility of $\leq_{sl}$ with multiplication.

Next we show that

$$u \longrightarrow^*_{R'} \bar{u} \text{ holds for every } u \in \widetilde{A}^*. \tag{8}$$

We use induction on $|u|$. The case $|u| \leq K_0N_0 + 1$ follows from the definition of $\mathcal{R}'$. Hence assume that $|u| > K_0N_0 + 1$ and (8) holds for shorter words. Write $u = avb$ with $a, b \in \widetilde{A}$. If $av \notin L$, we have $u \longrightarrow^*_{R'} \overline{av}b$ and $\bar{u} = \overline{\overline{av}b}$. Hence $u \longrightarrow^*_{R'} \bar{u}$ follows from $\overline{av}b \longrightarrow^*_{R'} \overline{\overline{av}b}$. Hence we may assume that $av \in L$.

Suppose that $u \notin \mathrm{Geo}_A(G)$. By Lemma 5.1(i), there exists some $w \in \mathrm{Geo}_A(G)$ such that $w\pi = (avb)\pi$ and $|av \wedge w| \geq |av| - N_0 \geq K_0N_0 + 1 - N_0 > 0$. Hence we may write $w = aw'$ and we get $(vb)\pi = (a^{-1}w)\pi = w'\pi$. Since $|w'| < |vb|$ due to $u \notin \mathrm{Geo}_A(G)$, we get $|\overline{vb}| < |vb|$, and so we may apply the induction hypothesis twice to get

$$u = avb \longrightarrow^*_{R'} a\overline{vb} \longrightarrow^*_{R'} \overline{a\overline{vb}} = \bar{u}.$$

Thus we may assume that $u \in \mathrm{Geo}_A(G)$. We claim that $\bar{u}^{[1]} = a$. Let $p = K_0N_0+1 < |u|$. Since $u, \bar{u} \in \mathrm{Geo}_A(G)$ and $u\pi = \bar{u}\pi$, the fellow traveler property yields $d_1(u^{[p]}\pi, \bar{u}^{[p]}\pi) \leq K_0$, and so $u^{[p]}\pi = (\bar{u}^{[p]}x)\pi$ for some x of length $\leq K_0$. Thus, by Lemma 5.1(i), there exists some $w \in \mathrm{Geo}_A(G)$ such that $w\pi = (\bar{u}^{[p]}x)\pi = u^{[p]}\pi$ and

$$|\bar{u}^{[p]} \wedge w| \geq |\bar{u}^{[p]}| - N_0|x| \geq p - K_0N_0 = 1.$$

Hence $\bar{u}^{[1]} = w^{[1]}$. Now $av \in L$ by assumption; hence $u^{[p]} \in L$, and so $u^{[p]} = \overline{u^{[p]}}$. Since $w\pi = u^{[p]}\pi$ and $w \in \mathrm{Geo}_A(G)$, we get $a = u^{[1]} \leq w^{[1]} = \bar{u}^{[1]}$ in $(\widetilde{A}, \leq)$. On the other hand, $\bar{u} \leq_{sl} u$ yields $\bar{u}^{[1]} \leq a$ in $(\widetilde{A}, \leq)$, and so $\bar{u}^{[1]} = a$ as claimed.

Now it follows easily that $\bar{u} = a\overline{a^{-1}u} = a\overline{vb}$ and the induction hypothesis yields $vb \longrightarrow^*_{R'} \overline{vb}$ and therefore $u = avb \longrightarrow^*_{R'} a\overline{vb} = \bar{u}$. Therefore (8) holds.

Assume now that $u \longrightarrow^*_{R'} v$ and $u \longrightarrow^*_{R'} w$. By (8), we get $v \longrightarrow^*_{R'} \bar{v} = \bar{u}$ and $w \longrightarrow^*_{R'} \bar{w} = \bar{u}$. Hence $\mathcal{R}'$ is confluent.

(ii) It follows from (8) that $\mathrm{Irr}\,\mathcal{R}' \subseteq L$. The converse inclusion follows from the implication

$$u \longrightarrow_{R'} v \Rightarrow u >_{sl} v,$$

which follows in turn from (7). □

We now establish some technical results which are useful in later sections.

Lemma 5.3. *Let $u, v \in L$ and let $w \in \widetilde{A}^*$ be such that $vw \in \mathrm{Geo}_A(G)$ and $(vw)\pi = u\pi$. Then $|u \wedge v| \geq |v| - K_0N_0$.*

Proof. Let $k = |v|$ and write $u = u^{[k]}u'$. Since $v = (vw)^{[k]}$, it follows from the fellow traveler property that $d_1(v\pi, u^{[k]}\pi) \leq K_0$. Hence we may write $v\pi = (u^{[k]}z)\pi$ with $|z| \leq K_0$. Since $u^{[k]}$ is itself a geodesic, it follows from Lemma 5.1(i) that there exists a geodesic $u^{[p]}z'$ satisfying $(u^{[p]}z')\pi = (u^{[k]}z)\pi = v\pi$ and

$$p = |u^{[k]} \wedge u^{[p]}z'| \geq |u^{[k]}| - N_0|z| \geq |v| - K_0N_0.$$

Now $v \in L$ yields $v \leq_{sl} u^{[p]}z'$, and so $v^{[p]} \leq_{sl} u^{[p]}$. On the other hand, $u \in L$ yields $u \leq_{sl} vw$, and so $u^{[p]} \leq_{sl} v^{[p]}$. Thus $u^{[p]} = v^{[p]}$, and so $|u \wedge v| \geq p \geq |v| - K_0N_0$. □

Proposition 5.4. (i) *Let $uv \in L$ and let $w \in \widetilde{A}^*$ be such that $|v| \geq K_0N_0 + N_0|w|$. Then $\overline{uvw} = u\overline{vw}$.*

(ii) *Let $u \in \widetilde{A}^*$ and let $vw, vw' \in L$. Then $|\overline{uvw} \wedge \overline{uvw'}| \geq |v| - K_0N_0 - N_0|u|$.*

Proof. (i) Write $v = v_1v_2$ with $|v_2| = N_0|w|$. By Lemma 5.1(i), there exists some $uv_1z \in \mathrm{Geo}_A(G)$ such that $(uv_1z)\pi = (uvw)\pi$. Let $x = \overline{uvw}$. By Lemma 5.3, we get $|x \wedge uv_1| \geq |uv_1| - K_0N_0$. Since $|v_1| = |v| - |v_2| \geq K_0N_0$, $u \leq x$ and we may write $x = uy$ for some y. Since L is factorial, we have $y \in L$. In view of $y\pi = (u^{-1}x)\pi = (vw)\pi$, we get $y = \overline{vw}$ and so $\overline{uvw} = u\overline{vw}$.

(ii) We may assume that $|v| > K_0N_0 + N_0|u|$. Write $v = v_1v_2$ with $|v_1| = N_0|u|$. Let $x = \overline{uv_1}$ and write $p = |x| + |v_2|$. By the proof of Lemma 5.1, we have $xv_2w, xv_2w' \in \mathrm{Geo}_A(G)$.

Let $y = \overline{uvw}$. Since $(xv_2w)\pi = y\pi$, it follows from the fellow traveler property that $d_1((xv_2)\pi, y^{[p]}\pi) \le K_0$. Hence we may write $(xv_2)\pi = (y^{[p]}s)\pi$ with $|s| \le K_0$. Since $y^{[p]}$ is itself a geodesic, it follows from Lemma 5.1(i) that there exists a geodesic $y^{[p-K_0N_0]}s'$ satisfying $(y^{[p-K_0N_0]}s')\pi = (y^{[p]}s)\pi = (xv_2)\pi$. To complete the proof, it suffices to show that

$$|y \wedge \overline{xv_2}| \ge p - K_0N_0. \tag{9}$$

Indeed, together with the corresponding inequality for $y' = \overline{uvw'}$, this implies

$$|\overline{uvw} \wedge \overline{uvw'}| \ge p - K_0N_0 \ge |v_2| - K_0N_0 = |v| - K_0N_0 - N_0|u|$$

and we obtain the desired inequality.

To prove (9), we consider the geodesic $y^{[p-K_0N_0]}s'$. Since $(y^{[p-K_0N_0]}s')\pi = (xv_2)\pi$, we get $\overline{xv_2} \le_{sl} y^{[p-K_0N_0]}s'$, and so $\overline{xv_2}^{[p-K_0N_0]} \le_{sl} y^{[p-K_0N_0]}$. On the other hand, xv_2w is also a geodesic. Hence $y = \overline{uvw} = \overline{xv_2w} \le_{sl} \overline{xv_2}w$ yields $y^{[p-K_0N_0]} \le_{sl} \overline{xv_2}^{[p-K_0N_0]}$. Therefore $y^{[p-K_0N_0]} = \overline{xv_2}^{[p-K_0N_0]}$, so (9) holds. □

6. A new model for the boundary

We can now present a new model for the boundary of a finitely generated virtually free group which proves useful in studying infinite fixed points. The notion of boundary is indeed one of the important features associated to hyperbolic groups. To present it, we define a second distance in G by means of the *Gromov product* (taking 1 as basepoint). We keep all the notation introduced in Section 5. In particular, G is a finitely generated virtually free group and $L = \mathrm{Irr}\,\mathcal{R}'$.

Given $g, h \in G$, we define

$$(g|h) = \tfrac{1}{2}(d_1(1, g) + d_1(1, h) - d_1(g, h)).$$

Fix $\varepsilon > 0$ such that $\varepsilon\delta \le 1/5$, where δ is the hyperbolicity constant from Section 5. Write $z = e^{\varepsilon}$ and define

$$\rho(g, h) = \begin{cases} z^{-(g|h)} & \text{if } g \ne h, \\ 0 & \text{otherwise} \end{cases}$$

for all $g, h \in G$. In general, ρ is not a distance because it fails the triangular inequality. This problem is overcome by defining

$$d_2(g, h) = \inf\{\rho(g_0, g_1) + \cdots + \rho(g_{n-1}, g_n) : g_0 = g,\ g_n = h;\ g_1, \ldots, g_{n-1} \in G\}.$$

By [Väisälä 2005, Proposition 5.16] (see also [Ghys and de la Harpe 1990, Proposition 7.10]), d_2 is a distance on G and the inequalities

$$\tfrac{1}{2}\rho(g,h) \le d_2(g,h) \le \rho(g,h) \tag{10}$$

hold for all $g, h \in G$.

In general, the metric space (G, d_2) is not complete. Its completion $(\widehat{G}, \hat{d}_2)$ is essentially unique, and $\partial G = \widehat{G} \setminus G$ is the *boundary* of G. The elements of the boundary admit several standard descriptions, such as equivalence classes of rays (infinite words whose finite factors are geodesics) where two rays are equivalent if the Hausdorff distance between them is finite [Ghys and de la Harpe 1990, Section 7.1]. We won't need precise definitions for these concepts or $\hat{d}_2$ since, as we shall see next, we can get a simpler description of $\widehat{G}$ for virtually free groups.

Lemma 6.1. *There exists some $M_0 > 0$ such that, for all $g, h \in G$,*

(i) $|\bar{g}| \le |\bar{g} \wedge \overline{gh}| + K_0 N_0 + N_0|\bar{h}|$,

(ii) $d_1(g,h) \ge \dfrac{|\bar{g}| - |\bar{g} \wedge \bar{h}|}{N_0} - K_0$,

(iii) $|\bar{g} \wedge \bar{h}| \le (g|h) \le |\bar{g} \wedge \bar{h}| + M_0$.

Proof. (i) By applying Lemma 5.1 to the product $\bar{g}\bar{h}$, there exists some factorization $\bar{g} = vz$ and some geodesic $vw \in (gh)\pi^{-1}$ such that $|v| \ge |\bar{g}| - N_0|\bar{h}|$. Now we apply Lemma 5.3 to $u = \overline{gh}$ and vw to get $|u \wedge v| \ge |v| - K_0N_0$. Hence

$$|\bar{g} \wedge \overline{gh}| = |u \wedge v| \ge |v| - K_0N_0 \ge |\bar{g}| - N_0|\bar{h}| - K_0N_0.$$

(ii) Let $u = \bar{g} \wedge \bar{h}$. Applying (i) to g and $g^{-1}h$, and in view of $d_1(g,h) = |\overline{g^{-1}h}|$, we get

$$|\bar{g}| \le |\bar{g} \wedge \bar{h}| + K_0N_0 + N_0 d_1(g,h).$$

(iii) We define $M_0 = \delta + (2\delta + 1 + K_0)N_0 - 1/2$, assuming that $\Gamma_A(G)$ is δ-hyperbolic. Let $u = \bar{g} \wedge \bar{h}$, and write $\bar{g} = uv$, $\bar{h} = uw$. It is easy to check that

$$(g|h) = \tfrac{1}{2}(d_1(1,g) + d_1(1,h) - d_1(g,h)) = \tfrac{1}{2}(|u| + d_1(u\pi, g) + |u| + d_1(u\pi, h) - d_1(g,h)).$$

Since $d_1(g,h) \le d_1(g, u\pi) + d_1(u\pi, h)$, we get $|\bar{g} \wedge \bar{h}| = |u| \le (g|h)$.

Consider now the geodesic triangle determined by the paths

$$P_1 : u\pi \xrightarrow{v} g, \quad P_2 : u\pi \xrightarrow{w} h, \quad P_3 : g \xrightarrow{\overline{g^{-1}h}} h.$$

Since $\Gamma_A(G)$ is δ-hyperbolic,

$$d_1(q, V(P_1) \cup V(P_2)) < \delta \text{ for every } q \in V(P_3). \tag{11}$$

Assume that $P_3 : g = q_0 \xrightarrow{a_1} \cdots \xrightarrow{a_n} q_n = h$ with $a_i \in \widetilde{A}$. Since

$$d_1(q_0, V(P_1)) = 0 < \delta \quad \text{and} \quad d_1(q_n, V(P_2)) = 0 < \delta,$$

it follows from (11) that there exist some $j \in \{0, \dots, n-1\}$ and $p_1 \in V(P_1)$, $p_2 \in V(P_2)$ such that $d_1(q_j, p_1), d_1(q_{j+1}, p_2) \leq \delta$. Since P_1 and P_2 are geodesics, we get

$$\begin{aligned}(g|h) &= \tfrac{1}{2}(d_1(1,g) + d_1(1,h) - d_1(g,h)) \\ &= \tfrac{1}{2}(|u| + d_1(u\pi, p_1) + d_1(p_1, g) \\ &\qquad + |u| + d_1(u\pi, p_2) + d_1(p_2, h) - d_1(g, q_j) - 1 - d_1(q_{j+1}, h)) \\ &= |\bar{g} \wedge \bar{h}| + \tfrac{1}{2}(d_1(u\pi, p_1) + d_1(u\pi, p_2)) \\ &\qquad + \tfrac{1}{2}(d_1(p_1, g) - d_1(g, q_j)) + \tfrac{1}{2}(d_1(p_2, h) - d_1(q_{j+1}, h)) - \tfrac{1}{2}.\end{aligned}$$

Since $d_1(p_1, g) \leq d_1(p_1, q_j) + d_1(q_j, g) \leq \delta + d_1(q_j, g)$, we have

$$\tfrac{1}{2}(d_1(p_1, g) - d_1(g, q_j)) \leq \frac{\delta}{2}.$$

Similarly,

$$\tfrac{1}{2}(d_1(p_2, h) - d_1(q_{j+1}, h)) \leq \frac{\delta}{2}.$$

Out of symmetry, it suffices to show that $d_1(u\pi, p_1) \leq (2\delta + 1 + K_0)N_0$.

Applying (ii) to p_1 and p_2, we get

$$d_1(p_1, p_2) \geq \frac{|\overline{p_1}| - |\overline{p_1} \wedge \overline{p_2}|}{N_0} - K_0.$$

Since $\overline{p_1}$ (respectively $\overline{p_2}$) is a prefix of $\bar{g}$ (respectively $\bar{h}$), it follows easily that $\overline{p_1} \wedge \overline{p_2} = u$ and $|\overline{p_1}| - |\overline{p_1} \wedge \overline{p_2}| = d_1(u\pi, p_1)$. Hence

$$\begin{aligned}&d_1(u\pi, p_1) \\ &\quad \leq (d_1(p_1, p_2) + K_0)N_0 \leq (d_1(p_1, q_j) + d_1(q_j, q_{j+1}) + d_1(q_{j+1}, p_2) + K_0)N_0 \\ &\quad \leq (2\delta + 1 + K_0)N_0. \qquad \square\end{aligned}$$

The language L introduced in (6) was noted to be rational. We recall that an automaton is said to be trim if every vertex occurs in some successful path. Let $\mathscr{A} = (Q, q_0, T, E)$ be a finite trim deterministic $\widetilde{A}$-automaton recognizing L (for example, the minimal automaton of L; see [Berstel 1979]). Since L is factorial, we must have $T = Q$. Let

$$\partial L = \{\alpha \in \widetilde{A}^{\omega} : \alpha^{[n]} \in L \text{ for every } n \in \mathbb{N}\}.$$

Equivalently, since $\mathscr{A}$ is trim and deterministic and $T = Q$, we have $\partial L = L_{\omega}(\mathscr{A})$. Write $\hat{L} = L \cup \partial L$. We define a mapping $d_3 : \hat{L} \times \hat{L} \to \mathbb{R}_0^+$ by

$$d_3(\alpha, \beta) = \begin{cases} 2^{-|\alpha \wedge \beta|} & \text{if } \alpha \neq \beta, \\ 0 & \text{otherwise.} \end{cases}$$

It is immediate that d_3 is a distance in $\hat{L}$. Indeed, an ultrametric distance since

$$|\alpha \wedge \gamma| \geq \min\{|\alpha \wedge \beta|, |\beta \wedge \gamma|\}$$

holds for all $\alpha, \beta, \gamma \in \hat{L}$. We commit a slight abuse of notation by also denoting by d_3 the restriction of d_3 to $L \times L$.

Proposition 6.2. (i) *The mutually inverse mappings* $(G, d_2) \to (L, d_3) : g \mapsto \bar{g}$ *and* $(L, d_3) \to (G, d_2) : u \mapsto u\pi$ *are uniformly continuous*;

(ii) $(\hat{L}, d_3)$ *is the completion of* (L, d_3);

(iii) $(\partial L, d_3)$ *is homeomorphic to the boundary of* G.

Proof. (i) In view of (10), it suffices to show that

$$\forall M > 0 \ \exists N > 0 : ((g|h) > N \Rightarrow |\bar{g} \wedge \overline{h}| > M),$$

$$\forall M > 0 \ \exists N > 0 : (|\bar{g} \wedge \overline{h}| > N \Rightarrow (g|h) > M).$$

Now we apply Lemma 6.1(iii).

(ii) Let $(\alpha_n)_n$ be a Cauchy sequence in $(\hat{L}, d_3)$. For every $k \in \mathbb{N}$, the sequence $(\alpha_n^{[k]})_n$ stabilizes when $n \to +\infty$. Moreover, $\lim_{n\to+\infty} \alpha_n^{[k]}$ is a prefix of

$$\lim_{n\to+\infty} \alpha_n^{[k+1]}.$$

Let $\beta \in A^\infty$ be the unique word satisfying $\beta^{[k]} = \lim_{n\to+\infty} \alpha_n^{[k]}$ for every $k \in \mathbb{N}$. It is immediate that $\beta \in \hat{L}$ and $\beta = \lim_{n\to+\infty} \alpha_n$. Hence $(\hat{L}, d_3)$ is complete. Since $\alpha = \lim_{n\to+\infty} \alpha^{[n]}$ for every $\alpha \in \partial L$, $(\hat{L}, d_3)$ is the completion of (L, d_3).

(iii) By (i) and (ii), the uniformly continuous mappings $(G, d_2) \to (L, d_3) : g \mapsto \bar{g}$ and $(L, d_3) \to (G, d_2) : u \mapsto u\pi$ admit (unique) continuous extensions to their completions (see [Dugundji 1966, Section XIV.6]), say

$$\Phi : \widehat{G} \to \hat{L}, \quad \Psi : \hat{L} \to \widehat{G}.$$

Hence $\Phi\Psi$ is a continuous extension of the identity on G to its completion $\widehat{G}$. Since such an extension is unique, $\Phi\Psi$ must be the identity mapping on $\widehat{G}$. Similarly, $\Psi\Phi$ must be the identity mapping on $\hat{L}$, and so Φ and Ψ are mutually inverse homeomorphisms. Therefore the restriction $\Phi|_{\partial G} : \partial G \to \partial L$ must also be a homeomorphism. □

We have just proved that our construction of $\hat{L}$ constitutes a model for the hyperbolic completion of G. But we must also import to $\hat{L}$ the algebraic operations of $\widehat{G}$ since we shall be considering homomorphisms soon. Clearly, the binary operation on L is defined as

$$L \times L \to L : (u, v) \mapsto \overline{uv},$$

so that $(G, d_2) \to (L, d_3) : g \mapsto \bar{g}$ is also a group isomorphism. But there is another important algebraic operation involved. Indeed, for every $g \in G$, the left translation $\tau_g : G \to G : x \mapsto gx$ is uniformly continuous for d_2 and so admits a continuous extension $\hat{\tau}_g : \widehat{G} \to \widehat{G}$. It follows that the left action of G in its boundary, $G \times \partial G \to \partial G : (g, \alpha) \mapsto \alpha\hat{\tau}_g$, is continuous. We can also replicate this operation in $\hat{L}$ as follows.

Proposition 6.3. *Let $u \in L$. Then $\tau_u : L \to L : v \mapsto \overline{uv}$ is uniformly continuous.*

Proof. It suffices to show that

$$\forall M > 0 \ \exists N > 0 : (|v \wedge w| > N \Rightarrow |\overline{uv} \wedge \overline{uv}| > M).$$

By Proposition 5.4(ii), we can take $N = M + K_0N_0 + N_0|u|$. □

Therefore τ_u admits a continuous extension $\hat{\tau}_u : \hat{L} \to \hat{L}$ and the left action $L \times \partial L \to \partial L : (u, \alpha) \mapsto \alpha\hat{\tau}_u$ is continuous. Write $\overline{u\alpha} = \alpha\hat{\tau}_u$. For every $\alpha \in \partial L$, we have

$$\overline{u\alpha} = \overline{u \lim_{n\to+\infty} \alpha^{[n]}} = \lim_{n\to+\infty} \overline{u\alpha^{[n]}}.$$

Hence $(\hat{L}, d_3)$ serves as a model for $(\widehat{G}, \hat{d}_2)$ both topologically and algebraically. From now on, we pursue our work within $(\hat{L}, d_3)$.

7. Uniformly continuous endomorphisms

We keep all the notation introduced in Section 5. In particular, G is a finitely generated virtually free group and $L = \operatorname{Irr}\mathcal{R}'$. Following the program announced above, we work within $(\hat{L}, d_3)$.

Given an endomorphism φ of G, we denote by $\bar{\varphi}$ the corresponding endomorphism of L for the binary operation induced by the product in G, that is, $u\bar{\varphi} = \overline{(u\pi)\varphi}$. To simplify notation, we often write $u\varphi$ instead of $u\pi\varphi$ for $u \in \widetilde{A}^*$.

We say that φ satisfies the *bounded cancellation property* if

$$\{|u\bar{\varphi}| - |u\bar{\varphi} \wedge (uv)\bar{\varphi}| : uv \in L\}$$

is bounded. In that case, we denote its maximum by B_φ. This property was considered originally for free group automorphisms by Cooper [1987].

We also fix the notation $D_\varphi = \max\{|\overline{a\varphi}| : a \in \widetilde{A}\}$ and recall that a homomorphism with finite kernel is called *virtually injective*.

Theorem 7.1. *Let φ be a virtually injective endomorphism φ of G. Then φ satisfies the bounded cancellation property.*

Proof. Suppose that φ does not satisfy the bounded cancellation property. Then

$$\forall m \in \mathbb{N} \ \exists u_m v_m \in L : |u_m\bar{\varphi}| - |u_m\bar{\varphi} \wedge (u_m v_m)\bar{\varphi}| > m.$$

Let $X_0 = (K_0 + D_\varphi)N_0$. We claim that

$$\text{(12)} \quad \forall m \in \mathbb{N}\ \exists u'_m v'_m \in L : (|u'_m \bar{\varphi}| - |(u'_m v'_m)\bar{\varphi}| > m \\ \text{and } |(u'_m v'_m)\bar{\varphi}| - |u'_m \bar{\varphi} \wedge (u'_m v'_m)\bar{\varphi}| \leq X_0).$$

Indeed, let $m \in \mathbb{N}$. Take $n = m + X_0$ and write $v_n = a_1 \cdots a_k$ ($a_i \in \widetilde{A}$). For $i = 0, \dots, k$, let $w_i = (u_n a_1 \cdots a_i)\bar{\varphi}$. Let j denote the smallest i such that

$$|u_n \bar{\varphi} \wedge w_i| \leq |u_n \bar{\varphi} \wedge (u_n v_n)\bar{\varphi}|.$$

Take $u'_m = u_n$ and $v'_m = a_1 \cdots a_{j-1}$ (since $j > 0$). Since L is factorial, we have $u'_m v'_m \in L$.

Now, by the minimality of j, we get

$$|u_n \bar{\varphi} \wedge w_{j-1}| > |u_n \bar{\varphi} \wedge (u_n v_n)\bar{\varphi}|.$$

Since $|u_n \bar{\varphi} \wedge w_j| \leq |u_n \bar{\varphi} \wedge (u_n v_n)\bar{\varphi}|$, it follows that

$$|w_{j-1} \wedge w_j| \leq |u_n \bar{\varphi} \wedge (u_n v_n)\bar{\varphi}|.$$

Applying Lemma 6.1(i) to $w_{j-1}\pi$ and $a_j\varphi$, we get

$$\begin{aligned} |w_{j-1}| &\leq |w_{j-1} \wedge w_j| + K_0 N_0 + N_0 |\overline{a_j \varphi}| \leq |w_{j-1} \wedge w_j| + X_0 \\ &\leq |u_n \bar{\varphi} \wedge (u_n v_n)\bar{\varphi}| + X_0 < |u_n \bar{\varphi}| - n + X_0 = |u_n \bar{\varphi}| - m, \end{aligned}$$

and so $|u'_m \bar{\varphi}| - |(u'_m v'_m)\bar{\varphi}| = |u_n \bar{\varphi}| - |w_{j-1}| > m$.

Suppose that $|w_{j-1}| - |u_n \bar{\varphi} \wedge w_{j-1}| > X_0$. Since we have seen above that $|w_{j-1}| \leq |w_{j-1} \wedge w_j| + X_0$, we get $|u_n \bar{\varphi} \wedge w_{j-1}| < |w_{j-1} \wedge w_j|$, in contradiction with $|w_{j-1} \wedge w_j| \leq |u_n \bar{\varphi} \wedge (u_n v_n)\bar{\varphi}| < |u_n \bar{\varphi} \wedge w_{j-1}|$. Thus

$$|(u'_m v'_m)\bar{\varphi}| - |u'_m \bar{\varphi} \wedge (u'_m v'_m)\bar{\varphi}| = |w_{j-1}| - |u_n \bar{\varphi} \wedge w_{j-1}| \leq X_0,$$

and so (12) holds.

We prove that

$$\text{(13)} \qquad \forall m \in \mathbb{N}\ \exists u''_m v''_m \in L : |u''_m \bar{\varphi}| > m \text{ and } |(u''_m v''_m)\bar{\varphi}| \leq X_0 + N_0 D_\varphi.$$

Indeed, let $m \in \mathbb{N}$. We have in $\Gamma_A(G)$ geodesics

$$\begin{tikzcd} 1 \arrow[r, "p"] & g \arrow[r, "q"] \arrow[dr, "r"] & u'_m\varphi \\ & & (u'_m v'_m)\varphi, \end{tikzcd}$$

where $pq = u'_m\bar{\varphi}$, $pr = (u'_m v'_m)\bar{\varphi}$, and $p = u'_m\bar{\varphi} \wedge (u'_m v'_m)\bar{\varphi}$. Assume that $u'_m = a_1 \cdots a_k$ ($a_i \in \widetilde{A}$). Let

$$I = \{i \in \{0, \dots, k\} : \text{ there exists a geodesic } (a_1 \cdots a_i)\varphi \longrightarrow g \xrightarrow{q} u'_m\varphi \text{ in } \Gamma_A(G)\}.$$

Clearly, $0 \in I$. We claim that

$$\big(i - 1 \in I \text{ and } d_1((a_1 \cdots a_{i-1})\varphi, g) > N_0 D_\varphi\big) \Rightarrow i \in I \tag{14}$$

holds for $i = 1, \dots, k$. Indeed, assume $i-1 \in I$ and $(a_1 \cdots a_{i-1})\varphi \xrightarrow{y} g \xrightarrow{q} u'_m\varphi$ is a geodesic with $y \in L$. Applying Lemma 5.1(ii) to the word $a_i^{-1}\bar{\varphi}$ and the geodesic $u = yq$, it follows that there exists some geodesic $(a_1 \cdots a_i)\varphi \xrightarrow{z} u'_m\varphi$ such that z and u share a suffix of length $\geq |yq| - N_0|a_i^{-1}\bar{\varphi}| \geq |yq| - N_0 D_\varphi > |q|$. Since $\Gamma_A(G)$ is deterministic, our geodesic $(a_1 \cdots a_i)\varphi \xrightarrow{z} u'_m\varphi$ factors through g, and so (14) holds.

Since $k \notin I$ due to $|q| > 0$, it follows from (14) that $d_1((a_1 \cdots a_i)\varphi, g) \leq N_0 D_\varphi$ for some $i \in \{1, \dots, k\}$. Let j denote the smallest such i. We define $u''_m = a_{j+1} \cdots a_k$ and $v''_m = v'_m$. Since L is factorial and $u'_m v'_m \in L$, we also have $u''_m v''_m \in L$.

By minimality of j, we have $d_1((a_1 \cdots a_i)\varphi, g) > N_0 D_\varphi$ for $i = 0, \dots, j-1$. By (14), we get $1, \dots, j \in I$ and so there exists a geodesic $(a_1 \cdots a_j)\varphi \longrightarrow g \xrightarrow{q} u'_m\varphi$ in $\Gamma_A(G)$. Hence

$$|u''_m\bar{\varphi}| = d_1(1, u''_m\varphi) = d_1((a_1 \cdots a_j)\varphi, u'_m\varphi) \geq |q| \geq |u'_m\bar{\varphi}| - |(u'_m v'_m)\bar{\varphi}| > m.$$

Finally,

$$\begin{aligned} |(u''_m v''_m)\bar{\varphi}| &= d_1(1, (u''_m v''_m)\varphi) = d_1((a_1 \cdots a_j)\varphi, (u'_m v'_m)\varphi) \\ &\leq d_1((a_1 \cdots a_j)\varphi, g) + d_1(g, (u'_m v'_m)\varphi) \leq N_0 D_\varphi + |r| \\ &= N_0 D_\varphi + |(u'_m v'_m)\bar{\varphi}| - |u'_m\bar{\varphi} \wedge (u'_m v'_m)\bar{\varphi}| \leq N_0 D_\varphi + X_0 \end{aligned}$$

and so (13) holds.

Now, since $|(u''_m v''_m)\bar{\varphi}|$ is bounded, $u''_m v''_m \in L$, and $\operatorname{Ker}\varphi$ is finite, $|u''_m v''_m|$ must be bounded and so must be $|u''_m|$. This implies that $|u''_m\bar{\varphi}|$ must be bounded, contradicting $|u''_m\bar{\varphi}| > m$. Thus φ satisfies the bounded cancellation property. □

Proposition 7.2. *The following conditions are equivalent for a nontrivial endomorphism φ of G:*

(i) *φ is uniformly continuous for d_2;*

(ii) *φ is virtually injective.*

Proof. (i) ⇒ (ii). Suppose that $\operatorname{Ker}\varphi$ is infinite. In view of (10), it suffices to show that there exists some $\eta > 0$ such that

$$\forall \xi > 0 \ \exists g, h \in G : (\rho(g, h) < \xi \text{ and } \rho(g\varphi, h\varphi) \geq \eta).$$

By (10), we only need to show that there exists some $M \in \mathbb{N}$ such that

$$\forall N \in \mathbb{N}\ \exists g, h \in G : \big((g|h) > N \text{ and } g\varphi \neq h\varphi \text{ and } ((g\varphi)|(h\varphi)) \leq M\big).$$

Take $M = (g_0\varphi|1) = 0$ and fix $g_0 \in G \setminus \operatorname{Ker}\varphi$. We prove the claim by showing that

$$\forall N \in \mathbb{N}\ \exists h \in \operatorname{Ker}\varphi : ((hg_0)|h) > N. \tag{15}$$

Let $N \in \mathbb{N}$. By Lemma 6.1(iii), we have $|\overline{hg_0} \wedge \overline{h}| \leq ((hg_0)|h)$ for every $h \in G$; hence we only need to find out $h \in \operatorname{Ker}\varphi$ satisfying $|\overline{hg_0} \wedge \overline{h}| > N$. By Lemma 6.1(i), we have $|\overline{hg_0} \wedge \overline{h}| \geq |\overline{h}| - K_0N_0 - N_0|\overline{g_0}|$. Hence it suffices that $|\overline{h}| > N + K_0N_0 + N_0|\overline{g_0}|$ for some $h \in \operatorname{Ker}\varphi$, and that is ensured by $\operatorname{Ker}\varphi$ being infinite. Thus (15) holds as required.

(ii) $\Rightarrow$ (i). Suppose that φ is not uniformly continuous for d_2. In view of (10), there exists some $\eta > 0$ such that

$$\forall \xi > 0\ \exists g, h \in G : (\rho(g, h) < \xi \text{ and } \rho(g\varphi, h\varphi) \geq \eta).$$

Hence, by (10), there exists some $M \in \mathbb{N}$ such that

$$\forall N \in \mathbb{N}\ \exists g, h \in G : \big((g|h) > N \text{ and } g\varphi \neq h\varphi \text{ and } ((g\varphi)|(h\varphi)) \leq M\big).$$

In view of Lemma 6.1(iii), we have that

$$\forall n \in \mathbb{N}\ \exists u_n, v_n \in L : (|u_n \wedge v_n| > n \text{ and } u_n\bar{\varphi} \neq v_n\bar{\varphi} \text{ and } |u_n\bar{\varphi} \wedge v_n\bar{\varphi}| \leq M).$$

Let $w_n = u_n \wedge v_n \in L$. Then either $w_n\bar{\varphi} \neq u_n\bar{\varphi}$ or $w_n\bar{\varphi} \neq v_n\bar{\varphi}$. Without loss of generality, we may assume that $w_n\bar{\varphi} \neq u_n\bar{\varphi}$. Suppose that $|w_n\bar{\varphi}| > M + B_\varphi$. By definition of B_φ, we get $|w_n\bar{\varphi}| - |w_n\bar{\varphi} \wedge u_n\bar{\varphi}| \leq B_\varphi$, and so $|w_n\bar{\varphi} \wedge u_n\bar{\varphi}| > M$. Similarly, $|w_n\bar{\varphi} \wedge v_n\bar{\varphi}| > M$, and so $|u_n\bar{\varphi} \wedge v_n\bar{\varphi}| > M$, a contradiction. Therefore $|w_n\bar{\varphi}| \leq M + B_\varphi$ for every n. Since $|w_n| > n$ and L is a cross-section for π, it follows that $\operatorname{Ker}\varphi$ is infinite. $\square$

Given a uniformly continuous endomorphism φ of (G, d_2), $\bar{\varphi} : L \to L$ is uniformly continuous for d_3. Since $\hat{L}$ is the completion of (L, d_3), $\bar{\varphi}$ admits a unique continuous extension $\Phi : \hat{L} \to \hat{L}$. By continuity, we have

$$\alpha\Phi = (\lim_{n\to+\infty} \alpha^{[n]})\Phi = \lim_{n\to+\infty} \alpha^{[n]}\bar{\varphi}. \tag{16}$$

Corollary 7.3. *Let φ be a uniformly continuous endomorphism of G and $u\alpha \in \partial L$. Then $|u\bar{\varphi}| - |u\bar{\varphi} \wedge (u\alpha)\Phi| \leq B_\varphi$.*

Proof. We have $(u\alpha)\Phi = \lim_{n\to+\infty}(u\alpha^{[n]})\bar{\varphi}$ by (16). In view of Proposition 7.2, we have $\lim_{n\to+\infty} |(u\alpha^{[n]})\bar{\varphi}| = +\infty$. Hence $|u\bar{\varphi} \wedge (u\alpha)\Phi| = |u\bar{\varphi} \wedge (u\alpha^{[m]})\bar{\varphi}|$ for sufficiently large m. Since $u\alpha^{[m]} \in L$, the claim follows by definition of B_φ. $\square$

8. Infinite fixed points

Keeping all the notation and assumptions introduced in the preceding sections, we fix now a uniformly continuous endomorphism φ of the finitely generated virtually free group G. We adapt notation introduced in [Ladra and Silva 2011] for free groups, and the proofs are adaptations of proofs in [Silva 2010].

Given $u \in L$, let $u\sigma = u \wedge u\bar{\varphi}$ and write

$$u = (u\sigma)(u\tau), \quad u\bar{\varphi} = (u\sigma)(u\rho).$$

Also define

$$u\sigma' = \bigwedge\{(uv)\sigma : uv \in L\}$$

and write $u\sigma = (u\sigma')(u\sigma'')$.

Lemma 8.1. *Let $uv \in L$. Then*

(i) $|u\sigma''| \leq B_\varphi$,

(ii) $|u\sigma| - |u\sigma \wedge (uv)\bar{\varphi}| \leq |u\sigma''|$,

(iii) $(uv)\bar{\varphi} = (u\sigma')\overline{(u\sigma'')(u\rho)(v\bar{\varphi})}$,

(iv) $(uv)\sigma' = (u\sigma')\big(\bigwedge_{uvz \in L} (\overline{(u\sigma'')(u\rho)((vz)\bar{\varphi})} \wedge (u\sigma'')(u\tau)vz)\big)$.

Proof. (i) We may assume that $|u\sigma| > B_\varphi$. Let v denote the suffix of length B_φ of $u\sigma$ and write $u\sigma = u'v$. Suppose that $uw \in L$. It suffices to show that u' is a prefix of $(uw)\bar{\varphi}$, and this follows from

$$|u'v(u\rho)| - |u'v(u\tau) \wedge (uw)\bar{\varphi}| = |u\bar{\varphi}| - |u\bar{\varphi} \wedge (uw)\bar{\varphi}| \leq B_\varphi$$

and $|v| = B_\varphi$.

(ii) $u\sigma'$ is a prefix of $u\sigma \wedge (uv)\bar{\varphi}$.

(iii) $u\sigma'$ is a prefix of $(uv)\bar{\varphi}$ and both sides of the equality are equivalent in G.

(iv) $u\sigma'$ is a prefix of $(uv)\sigma'$ by (iii). □

For every $u \in L$, we define

$$u\xi = (u\sigma'', u\tau, u\rho, q_0u).$$

Note that there exists precisely one path of the form $q_0 \xrightarrow{u} q_0u$ in $\mathcal{A}$.

Lemma 8.2. *Let $u, v \in L$ be such that $u\xi = v\xi$ and let $a \in \widetilde{A}$, $\alpha \in \widetilde{A}^\infty$. Then*

(i) *$ua \in L$ if and only if $va \in L$;*

(ii) *if $ua \in L$, $(ua)\xi = (va)\xi$;*

(iii) $\overline{uv^{-1}} \in \operatorname{Fix}\bar{\varphi}$;

(iv) *$u\alpha \in \hat{L}$ if and only if $v\alpha \in \hat{L}$;*

(v) $u\alpha \in \operatorname{Fix}\Phi$ *if and only if* $v\alpha \in \operatorname{Fix}\Phi$;

(vi) *if* $\alpha \in \hat{L}$, $\alpha = \lim_{n\to+\infty} \overline{\alpha^{[n]}u}$.

Proof. (i) $u\xi = v\xi$ implies $q_0u = q_0v$.

(ii) Clearly, $q_0u = q_0v$ yields $q_0ua = q_0va$. Considering $v = a$ in Lemma 8.1(iii), we may write $(ua)\sigma = (u\sigma')u'$ and deduce that u', $(ua)\tau$, and $(ua)\rho$ are all determined by $u\xi$. Hence $(ua)\tau = (va)\tau$, $(ua)\rho = (va)\rho$, and $u' = v'$.

Finally, since $q_0u = q_0v$, we have $uaz \in L$ if and only if $vaz \in L$. It follows from Lemma 8.1(iv) that there exists a word $x \in L$ which satisfies both $(ua)\sigma' = (u\sigma')x$ and $(va)\sigma' = (v\sigma')x$. Now $(u\sigma')u' = (ua)\sigma = ((ua)\sigma')((ua)\sigma'') = (u\sigma')x((ua)\sigma'')$. Hence $u' = x((ua)\sigma'')$. Similarly, $v' = x((va)\sigma'')$. Since $u' = v'$, we get $(ua)\sigma'' = (va)\sigma''$, and so $(ua)\xi = (va)\xi$.

(iii)
$$\begin{aligned}\overline{(uv^{-1})\varphi} &= \overline{(u\varphi)(v\varphi)^{-1}} = \overline{(u\sigma)(u\rho)(v\rho)^{-1}(v\sigma)^{-1}} = \overline{(u\sigma)(v\sigma)^{-1}}\\ &= \overline{(u\sigma)(u\tau)(v\tau)^{-1}(v\sigma)^{-1}} = \overline{uv^{-1}}.\end{aligned}$$

(iv) We have $u\alpha \in \hat{L}$ if and only if $u\alpha^{[n]} \in L$ for every $n \in \mathbb{N}$. Now we use (i) and induction on n.

(v) We have $u\alpha = (u\sigma')(u\sigma'')(u\tau)\alpha$ and, in view of Corollary 7.3 and (16), also

$$(u\alpha)\Phi = (u\sigma') \lim_{n\to+\infty} \overline{(u\sigma'')(u\rho)(\alpha^{[n]}\bar{\varphi})}.$$

Hence $u\alpha \in \operatorname{Fix}\Phi$ depends just on $u\xi$ and α.

(vi) Let $m = K_0N_0 + N_0|u|$. By Lemma 6.1(i), we have $|\alpha^{[n]} \wedge \overline{\alpha^{[n]}u}| \geq n - m$ for every n. Hence $\alpha = \lim_{n\to+\infty} \alpha^{[n-m]} = \lim_{n\to+\infty} \overline{\alpha^{[n]}u}$. □

Given $X \subseteq A^\infty$, write

$$\operatorname{Pref} X = \{u \in A^* : u\alpha \in X \text{ for some } \alpha \in A^\infty\}.$$

Recall the finite trim deterministic $\widetilde{A}$-automaton $\mathcal{A} = (Q, q_0, Q, E)$ recognizing L. We build a (possibly infinite) $\widetilde{A}$-automaton $\mathcal{A}'_\varphi = (Q', q'_0, T', E')$ by taking

$$\begin{aligned}Q' &= \{u\xi : u \in \operatorname{Pref}\operatorname{Fix}\Phi\},\\ q'_0 &= 1\xi,\\ T' &= \{u\xi \in Q' : u\tau = u\rho = 1\},\\ E' &= \{(u\xi, a, v\xi) \in Q' \times \widetilde{A} \times Q' : v = ua \in \operatorname{Pref}\operatorname{Fix}\Phi\}.\end{aligned}$$

We note that $\mathcal{A}'_\varphi$ is deterministic by Lemma 8.2(ii) and is also *accessible*: if $u \in \operatorname{Pref}\operatorname{Fix}\Phi$, there exists a path $q'_0 \xrightarrow{u} u\xi$, and so every vertex can be reached from the initial vertex.

Let S denote the set of all vertices $q \in Q'$ such that there exist at least two edges in $\mathcal{B}'_\varphi$ leaving q. Let Q'' denote the set of all vertices $q \in Q'$ such that there exists some path $q \longrightarrow p \in S \cup T'$. We define $\mathscr{A}''_\varphi = (Q'', q''_0, T'', E'')$ by taking $q''_0 = q'_0$, $T'' = T'$, and $E'' = E' \cap (Q'' \times \widetilde{A} \times Q'')$.

Lemma 8.3. *S is finite.*

Proof. In view of Lemma 8.1, the unique components of $u\xi$ that may assume infinitely many values are $u\tau$ and $u\rho$. Moreover, we claim that

$$u\tau \neq 1 \Rightarrow |u\rho| \leq B_\varphi \tag{17}$$

holds for every $u \in \operatorname{Pref}\operatorname{Fix}\Phi$. Indeed, suppose that $u\tau \neq 1$ and $|u\rho| > B_\varphi$. Write $\alpha = u\beta$ for some $\alpha \in \operatorname{Fix}\Phi$. In view of Corollary 7.3, $|u\rho| > B_\varphi$ yields $|(u\beta)\Phi \wedge u\bar{\varphi}| > |u\sigma|$ and now $u\tau \neq 1$ yields $((u\beta)\Phi \wedge u\beta) = (u\bar{\varphi} \wedge u) = u\sigma$. Since $\beta \neq 1$, this contradicts $\alpha \in \operatorname{Fix}\Phi$. Therefore (17) holds.

It is also easy to see that

$$|u\rho| > B_\varphi \Rightarrow u\xi \notin S \tag{18}$$

for every $u \in \operatorname{Pref}\operatorname{Fix}\Phi$. Indeed, if $|u\rho| > B_\varphi$ and a is the first letter of $u\rho$, then, by definition of B_φ, $(u\sigma)a$ is a prefix of $(u\alpha)\Phi$ whenever $u\alpha \in \operatorname{Fix}\Phi$. Therefore any edge leaving $u\xi$ in $\mathscr{A}'_\varphi$ must have label a, and so (18) holds.

In view of Proposition 7.2, we can define

$$W_0 = \max\{|u| : u \in L,\ |u\bar{\varphi}| \leq 2(B_\varphi + D_\varphi - 1)\}.$$

Let $Z_0 = B_\varphi + N_0(K_0 + W_0)D_\varphi$. To complete the proof, it suffices to prove that

$$|u\tau| > Z_0 \Rightarrow u\xi \notin S \tag{19}$$

for every $u \in \operatorname{Pref}\operatorname{Fix}\Phi$.

Suppose that $|u\tau| > Z_0$ and

$$(u\xi, a, (ua)\xi), (u\xi, b, (ub)\xi) \in E'$$

for some $u \in \operatorname{Pref}\operatorname{Fix}\Phi$, where $a, b \in \widetilde{A}$ are distinct. We have $(ua)\xi = v\xi$ for some $v \in \operatorname{Pref}\operatorname{Fix}\Phi$. By Lemma 8.2(v), we get $ua\alpha \in \operatorname{Fix}\Phi$ for some $\alpha \in \hat{L}$. By (16), we get $ua\alpha = \lim_{n\to+\infty}(ua\alpha^{[n]})\bar{\varphi}$, and so $|(ua\alpha^{[n]})\bar{\varphi}| \geq |u|$ for sufficiently large n. Let

$$p = \min\{n \in \mathbb{N} : |(ua\alpha^{[n]})\bar{\varphi}| \geq |u|\}.$$

Note that $p > 0$ since $|u\tau| > Z_0$ and by (17). Since $|(ua\alpha^{[p-1]})\bar{\varphi}| < |u|$ by the minimality of p, we get

$$|(ua\alpha^{[p]})\bar{\varphi}| \leq |(ua\alpha^{[p-1]})\bar{\varphi}| + D_\varphi < |u| + D_\varphi. \tag{20}$$

On the other hand,

$$|u| - |(ua\alpha^{[p]})\bar{\varphi} \wedge u| \leq B_\varphi. \tag{21}$$

Otherwise, by definition of B_φ, $ua\alpha$ and $(ua\alpha)\Phi$ would differ at position

$$|(ua\alpha^{[p]})\bar{\varphi} \wedge u| + 1.$$

Similarly, $ub\beta \in \text{Fix}\,\Phi$ for some $\beta \in \hat{L}$. Defining

$$q = \min\{n \in \mathbb{N} : |(ub\beta^{[n]})\bar{\varphi}| \geq |u|\},$$

we get

$$|(ub\beta^{[q]})\bar{\varphi}| < |u| + D_\varphi \tag{22}$$

and

$$|u| - |(ub\beta^{[q]})\bar{\varphi} \wedge u| \leq B_\varphi. \tag{23}$$

Write $u = u_1 u_2$ with $|u_2| = B_\varphi$. Then by (20) and (21) we may write $(ua\alpha^{[p]})\bar{\varphi} = u_1 x$ for some x such that $|x| < B_\varphi + D_\varphi$. Similarly, (22) and (23) yield $(ub\beta^{[q]})\bar{\varphi} = u_1 y$ for some x such that $|x| < B_\varphi + D_\varphi$. Writing $w = \overline{(\beta^{[q]})^{-1} b^{-1} a\alpha^{[p]}}$, it follows that $w\varphi = (y^{-1}x)\pi$, and so $|w\bar{\varphi}| \leq 2(B_\varphi + D_\varphi - 1)$. Hence $|w| \leq W_0$. Applying Lemma 6.1(i) to $g = (ub\beta^{[q]})\pi$ and $h = w\pi$, we get

$$|ub\beta^{[q]}| \leq |ub\beta^{[q]} \wedge ua\alpha^{[p]}| + N_0(K_0 + |w|) \leq |u| + N_0(K_0 + W_0),$$

and so $q < N_0(K_0 + W_0)$. Hence, in view of (17), we get

$$\begin{aligned} |u\tau| = |u| - |u\sigma| &\leq |(ub\beta^{[q]})\bar{\varphi}| - |u\sigma| \leq |u\bar{\varphi}| + |(b\beta^{[q]})\bar{\varphi}| - |u\sigma| \\ &\leq |u\rho| + N_0(K_0 + W_0)D_\varphi \leq B_\varphi + N_0(K_0 + W_0)D_\varphi, \end{aligned}$$

contradicting $|u\tau| > Z_0$. Thus (19) holds and the lemma is proved. □

We say that an infinite fixed point $\alpha \in \text{Fix}\,\Phi \cap \partial L$ is *singular* if α belongs to the topological closure $(\text{Fix}\,\varphi)^c$ of $\text{Fix}\,\varphi$. Otherwise, α is said to be *regular*. We denote by $\text{Sing}\,\Phi$ (respectively $\text{Reg}\,\Phi$) the set of all singular (respectively regular) infinite fixed points of Φ.

Theorem 8.4. *Let φ be a uniformly continuous endomorphism of a finitely generated virtually free group G. Then*

(i) *the automaton $\mathscr{A}''_\varphi$ is finite*;

(ii) $L(\mathscr{A}''_\varphi) = \text{Fix}\,\bar{\varphi}$;

(iii) $L_\omega(\mathscr{A}''_\varphi) = \text{Sing}\,\Phi$.

Proof. (i) The set T' is finite and S is finite by Lemma 8.3. On the other hand, by definition of S, there are only finitely many paths in $\mathscr{A}'_\varphi$ of the form $\nu_j : p' \longrightarrow q'$ with $p', q' \in S \cup T' \cup \{q'_0\}$ and no intermediate vertex in $S \cup T' \cup \{q'_0\}$. Now recall that $\mathscr{A}'_\varphi$ is accessible. Hence every path of the form $q \xrightarrow{u} p \in S \cup T'$ can be extended to some path $q_0' \xrightarrow{v} q \xrightarrow{u} p \in S \cup T'$ which is itself a concatenation of the finitely many paths ν_j. Therefore Q'' is finite and so is $\mathscr{A}''_\varphi$.

(ii) Every $u \in L$ labels at most a unique path $q_0' = 1\xi \xrightarrow{u} u\xi$ out of the initial vertex in $\mathscr{A}'_\varphi$. On the other hand, if $q_0' = 1\xi \xrightarrow{u} q'$ is a path in $\mathscr{A}'_\varphi$, the fourth component of ξ yields a path $q_0 \xrightarrow{u} q$ in $\mathscr{A}$, and so $u \in L$. Hence

$$L(\mathscr{A}'_\varphi) = \{u \in L : u\xi \in T'\} = \{u \in L : u\tau = u\rho = 1\} = \operatorname{Fix} \bar{\varphi}.$$

Since $L(\mathscr{A}''_\varphi) = L(\mathscr{A}'_\varphi)$, (ii) holds.

(iii) Let $\alpha \in L_\omega(\mathscr{A}''_\varphi)$. Then there exists some $q'' \in Q''$ and some infinite sequence $(i_n)_n$ such that $q_0'' \xrightarrow{\alpha^{[i_n]}} q''$ is a path in $\mathscr{A}''_\varphi$ for every n. Write $u = \alpha^{[i_1]}$ and let $v_n = \overline{\alpha^{[i_n]}u^{-1}}$. By Lemma 8.2(iii), we have $v_n \in \operatorname{Fix} \bar{\varphi}$ for every n. It follows from Lemma 8.2(vi) that $\alpha = \lim_{n\to+\infty} v_n$. Thus $\alpha \in \operatorname{Sing} \Phi$.

Conversely, let $\alpha \in \operatorname{Sing} \Phi$. Then we may write $\alpha = \lim_{n\to+\infty} v_n$ for some sequence $(v_n)_n$ in $\operatorname{Fix} \bar{\varphi}$. Let $k \in \mathbb{N}$. For large enough n, we have $\alpha^{[k]} = v_n^{[k]}$, and so there is some path

$$q_0'' \xrightarrow{\alpha^{[k]}} q_k'' \xrightarrow{w} t_k'' \in T'',$$

where $\alpha^{[k]}w = v_n$. Thus $\alpha \in L_\omega(\mathscr{A}''_\varphi)$ as required. □

Recall now the continuous extensions $\hat{\tau}_u : \hat{L} \to \hat{L}$ of the uniformly continuous mappings $\tau_u : L \to L : v \mapsto \overline{uv}$ defined for each $u \in L$ (see Proposition 6.3). As remarked before, this is equivalent to saying that the left action

$$L \times \partial L \to \partial L : (u, \alpha) \mapsto \overline{u\alpha}$$

is continuous. Identifying L with G and ∂L with ∂G, we have a continuous action (on the left) of G on ∂G. Clearly, this action restricts to a left action of $\operatorname{Fix} \varphi$ on $\operatorname{Fix} \Phi \cap \partial G$: if $g \in \operatorname{Fix} \varphi$ and $\alpha \in \operatorname{Fix} \Phi \cap \partial G$ with $\alpha = \lim_{n\to+\infty} g_n$ $(g_n \in G)$,

$$\begin{aligned}(g\alpha)\Phi &= (g \lim_{n\to+\infty} g_n)\Phi = (\lim_{n\to+\infty} gg_n)\Phi = \lim_{n\to+\infty} (gg_n)\varphi \\ &= \lim_{n\to+\infty} (g\varphi)(g_n\varphi) = (g\varphi) \lim_{n\to+\infty} g_n\varphi = g(\lim_{n\to+\infty} g_n)\Phi \\ &= g(\alpha\Phi) = g\alpha.\end{aligned}$$

Moreover, the $(\operatorname{Fix} \varphi)$-orbits of $\operatorname{Sing} \Phi$ and $\operatorname{Reg} \Phi$ are disjoint: if $\alpha \in \operatorname{Sing} \Phi$, we can write $\alpha = \lim_{n\to+\infty} g_n$ with the $g_n \in \operatorname{Fix} \varphi$ and get $g\alpha = \lim_{n\to+\infty} gg_n$ with $gg_n \in \operatorname{Fix} \varphi$ for every n; hence $\alpha \in \operatorname{Sing} \Phi \Rightarrow g\alpha \in \operatorname{Sing} \Phi$ and the action of g^{-1} yields the converse implication.

We can now prove the main result of this section.

Theorem 8.5. *Let φ be a uniformly continuous endomorphism of a finitely generated virtually free group G. Then* $\operatorname{Reg} \Phi$ *has finitely many* $(\operatorname{Fix} \varphi)$*-orbits.*

Proof. Let P be the set of all infinite paths $s'_0 \xrightarrow{a_1} s'_1 \xrightarrow{a_2} \cdots$ in $\mathscr{A}'_\varphi$ such that

$$s'_0 \in S \cup \{q_0\}, \quad s'_n \notin S \cup \{q_0\} \text{ for every } n > 0, \quad s'_n \neq s'_m \text{ whenever } n \neq m.$$

By Lemma 8.3, there are only finitely many choices for s'_0. Since A is finite and $\mathscr{A}'_\varphi$ is deterministic, there are only finitely many choices for s'_1, and from that vertex onwards, the path is uniquely determined due to $s'_n \notin S$ $(n \geq 1)$. Hence P is finite, and we may assume that it consists of paths $p'_i \xrightarrow{\alpha_i} \cdots$ for $i = 1, \dots, m$. Fix a path $q'_0 \xrightarrow{u_i} p'_i$ for each i and let $X = \{u_1\alpha_1, \dots, u_m\alpha_m\} \subseteq \partial L$. We claim that $X \subseteq \operatorname{Reg} \Phi$.

Let $i \in \{1, \dots, m\}$ and write $\beta = u_i\alpha_i$. To show that $\beta \in \operatorname{Fix} \Phi$, it suffices to show that $\lim_{n\to+\infty} \beta^{[n]}\bar{\varphi} = \beta$. Let $k \in \mathbb{N}$. We must show that there exists some $r \in \mathbb{N}$ such that

$$n \geq r \Rightarrow |\beta^{[n]}\bar{\varphi} \wedge \beta| > k. \tag{24}$$

In view of Proposition 7.2, there exists some $r > k$ such that

$$n \geq r \Rightarrow |\beta^{[n]}\bar{\varphi}| > k + B_\varphi.$$

Suppose that $|\beta^{[n]}\bar{\varphi} \wedge \beta| \leq k$ for some $n \geq r$. Then $|\beta^{[n]}\sigma| \leq k$. Since $k < r \leq n$, it follows that $\beta^{[n]}\tau \neq 1$. On the other hand, since $|\beta^{[n]}\bar{\varphi}| > k + B_\varphi$, we get $|\beta^{[n]}\rho| > B_\varphi$. In view of (17), this contradicts $\beta^{[n]}\xi \in Q'$. Therefore (24) holds for our choice of r and so $X \subseteq \operatorname{Fix} \Phi$. Since the path

$$q'_0 \xrightarrow{\beta} \cdots$$

can visit only finitely often a given vertex, $\beta \notin L_\omega(\mathscr{A}''_\varphi)$, and so $X \subseteq \operatorname{Reg} \Phi$ by Theorem 8.4(iii).

By the previous comments on $(\operatorname{Fix} \varphi)$-orbits, the $(\operatorname{Fix} \varphi)$-orbits of the elements of X must be contained in $\operatorname{Reg} \Phi$. We complete the proof of the theorem by proving the opposite inclusion.

Let $\beta \in \operatorname{Reg} \Phi$. By Theorem 8.4(iii), we have $\beta \notin L_\omega(\mathscr{A}''_\varphi)$, and so there exists a factorization $\beta = u\alpha$ and a path

$$q'_0 \xrightarrow{u} p' \xrightarrow{\alpha} \cdots$$

in $\mathscr{A}'_\varphi$ such that p' signals the last occurrence of a vertex from $S \cup \{q'_0\}$. We claim that no vertex is repeated after p'. Otherwise, since no vertex of S appears after p', we would get a factorization of $p' \xrightarrow{\alpha} \cdots$ as

$$p' \xrightarrow{v} q' \xrightarrow{w} q' \xrightarrow{w} \cdots$$

and by Lemma 8.2(iii) and (iv) we would get $(uvw^nv^{-1}u^{-1})\pi \in \operatorname{Fix} \varphi$ and

$$\beta = \lim_{n\to+\infty} \overline{uvw^nv^{-1}u^{-1}},$$

contradicting $\beta \in \operatorname{Reg}\Phi$. Thus no vertex is repeated after p', and so we must have $p' = p'_i$ and $\alpha = \alpha_i$ for some $i \in \{1, \dots, m\}$. It follows that $\beta = u\alpha_i$. By Lemma 8.2(iii), we get

$$\overline{uu_i^{-1}} \in \operatorname{Fix}\bar{\varphi},$$

and we are done. □

Theorem 8.5 is somehow a version for infinite fixed points of Theorem 4.1, which we proved before for finite fixed points. Note however that $\operatorname{Sing}\Phi$ does *not* in general have finitely many $(\operatorname{Fix}\varphi)$-orbits since $\operatorname{Sing}\Phi$ may be uncountable (take for instance the identity automorphism on a free group of rank 2).

Since every finite set is closed in a metric space, we obtain the following corollary from Theorem 8.5.

Corollary 8.6. *Let φ be a uniformly continuous endomorphism of a finitely generated virtually free group G with* $\operatorname{Fix}\varphi$ *finite. Then* $\operatorname{Fix}\Phi$ *is finite.*

9. Classification of the infinite fixed points

We can now investigate the nature of the infinite fixed points of Φ when φ is an automorphism. Since, by Proposition 7.2, both φ and φ^{-1} are then uniformly continuous, they extend to continuous mappings Φ and Ψ which turn out to be mutually inverse in view of the uniqueness of continuous extensions to the completion. Therefore Φ is a bijection. We say that $\alpha \in \operatorname{Reg}\Phi$ is

- an *attractor* if $\exists \varepsilon > 0\ \forall \beta \in \hat{L} : (d_3(\alpha, \beta) < \varepsilon \Rightarrow \lim_{n\to+\infty} \beta\Phi^n = \alpha)$;
- a *repeller* if $\exists \varepsilon > 0\ \forall \beta \in \hat{L} : (d_3(\alpha, \beta) < \varepsilon \Rightarrow \lim_{n\to+\infty} \beta\Phi^{-n} = \alpha)$.

The latter amounts to saying that α is an attractor for Φ^{-1}. There exist other types, but they do not occur in our context as we shall see.

We say that an attractor $\alpha \in \operatorname{Reg}\Phi$ is *exponentially stable* if

$$\exists \varepsilon, k, \ell > 0\ \forall \beta \in \hat{L}\ \forall n \in \mathbb{N} : (d_3(\alpha, \beta) < \varepsilon \Rightarrow d_3(\alpha, \beta\Phi^n) \le k2^{-\ell n} d_3(\alpha, \beta)).$$

This is equivalent to saying that

$$\text{(25)} \quad \exists M, N, \ell > 0\ \forall \beta \in \hat{L}\ \forall n \in \mathbb{N} : \quad (|\alpha \wedge \beta| > M \Rightarrow |\alpha \wedge \beta\Phi^n| + N > \ell n + |\alpha \wedge \beta|).$$

A repeller $\alpha \in \operatorname{Reg}\Phi$ is exponentially stable if it is an exponentially stable attractor for Φ^{-1}.

Theorem 9.1. *Let φ be an automorphism of a finitely generated virtually free group G. Then* $\operatorname{Reg}\Phi$ *contains only exponentially stable attractors and exponentially stable repellers.*

Proof. Let $\alpha \in \operatorname{Reg}\Phi$ and write $\alpha = a_1a_2\cdots$ with $a_i \in \widetilde{A}$. Then there exists a path

$$1\xi \xrightarrow{a_1} \alpha^{[1]}\xi \xrightarrow{a_2} \alpha^{[2]}\xi \xrightarrow{a_3} \cdots$$

in $\mathscr{A}'_\varphi$. Let $Y_0 = B_\varphi(D_{\varphi^{-1}}+1) + B_{\varphi^{-1}}(D_\varphi + 1)$ and let

$$V = \{u\xi \in Q' : |u\tau| > Y_0 \text{ or } |u\rho| > Y_0\}.$$

It is easy to see that $Q' \setminus V$ is finite. We saw in the proof of Theorem 8.5 that there are only finitely many repetitions of vertices in a path in $\mathscr{A}'_\varphi$ labeled by a regular fixed point. Hence there exists some $n_0 \in \mathbb{N}$ such that

$$\alpha^{[n]}\xi \in V \text{ for every } n \geq n_0. \tag{26}$$

Now we consider two cases.

Case I: $\alpha^{[n_0]}\tau = 1$. We claim that

$$\alpha^{[n]}\tau = 1 \text{ for every } n \geq n_0. \tag{27}$$

The case $n = n_0$ holds in Case I, so assume that $\alpha^{[n]}\tau = 1$ for some $n \geq n_0$. Then $\alpha^{[n]}\xi \in V$, and so $|\alpha^{[n]}\rho| > Y_0 > 2B_\varphi$. Since $|\alpha^{[n+1]}\bar{\varphi}| \geq |\alpha^{[n]}\bar{\varphi}| - B_\varphi$ by definition of B_φ,

$$\begin{aligned}|\alpha^{[n+1]}\rho| &\geq |\alpha^{[n+1]}\bar{\varphi}| - |\alpha^{[n+1]}| \geq |\alpha^{[n]}\bar{\varphi}| - B_\varphi - |\alpha^{[n]}| - 1 = |\alpha^{[n]}\rho| - B_\varphi - 1 \\ &> Y_0 - B_\varphi - 1 > B_\varphi.\end{aligned}$$

By (17), we get $\alpha^{[n+1]}\tau = 1$, and so (27) holds.

Next we show that

$$((\alpha^{[n]}\gamma)\Phi)^{[n+1]} = \alpha^{[n+1]} \tag{28}$$

if $n \geq n_0$ and $\alpha^{[n]}\gamma \in \hat{L}$. Indeed, by (27) we have $\alpha^{[n]}\bar{\varphi} = \alpha^{[n]}(\alpha^{[n]}\rho)$ and $|\alpha^{[n]}\rho| > Y_0 > B_\varphi$. By the definition of B_φ and Corollary 7.3, we get $((\alpha^{[n]}\gamma)\Phi)^{[n+1]} = \alpha^{[n]}(\alpha^{[n]}\rho)^{[1]}$. Considering the particular case $\gamma = a_{n+1}$, we also get

$$(\alpha^{[n+1]}\bar{\varphi})^{[n+1]} = \alpha^{[n]}(\alpha^{[n]}\rho)^{[1]} = ((\alpha^{[n]}\gamma)\Phi)^{[n+1]}.$$

Since $\alpha^{[n+1]}\tau = 1$ by (27), we have $(\alpha^{[n+1]}\bar{\varphi})^{[n+1]} = \alpha^{[n+1]}$, and so (28) holds.

Hence we may write $(\alpha^{[n]}\gamma)\Phi = \alpha^{[n+1]}\gamma'$ whenever $\alpha^{[n]}\gamma \in \hat{L}$. Iterating, it follows that, for all $k \geq n_0$ and $n \in \mathbb{N}$, $\alpha^{[k]}\gamma \in \hat{L}$ implies $(\alpha^{[k]}\gamma)\Phi^n = \alpha^{[k+n]}\gamma'$ for some γ'. By considering $\beta = \alpha^{[k]}\gamma$ and $\alpha^{[k]} = \alpha \wedge \beta$, we deduce that

$$|\alpha \wedge \beta| \geq n_0 \Rightarrow |\alpha \wedge \beta\Phi^n| \geq n + |\alpha \wedge \beta|$$

holds for all $\beta \in \hat{L}$ and $n \in \mathbb{N}$. Therefore (25) holds, and so α is an exponentially stable attractor in this case.

Now, if $|\alpha^{[t]}\tau| = 1$ for some $t > n_0$, we can always replace n_0 by t and deduce by Case I that α is an exponentially stable attractor. Thus we may assume the following.

Case II: $\alpha^{[n]}\tau \neq 1$ *for every* $n \geq n_0$. By replacing n_0 by a larger integer if necessary, we may assume that (26) is also satisfied when we consider the equivalents of ξ and V for φ^{-1}.

Since φ is injective, there exists some $n_1 \geq n_0$ such that

$$|\alpha^{[n_1]}\bar{\varphi}| \geq n_0 + B_\varphi.$$

Since $\alpha^{[n_1]}\tau \neq 1$, it follows from (17) that $|\alpha^{[n_1]}\rho| \leq B_\varphi$; hence $\alpha^{[n_1]}\sigma = \alpha^{[n_2]}$ for some $n_2 \geq n_0$. Write $x = \alpha^{[n_1]}\rho$. Then $\alpha^{[n_1]}\bar{\varphi} = \alpha^{[n_2]}x$ yields

$$\alpha^{[n_1]} = \overline{(\alpha^{[n_2]}\overline{\varphi^{-1}})(x\overline{\varphi^{-1}})},$$

and so

$$n_1 = |\alpha^{[n_1]}| \leq |\alpha^{[n_2]}\overline{\varphi^{-1}}| + |x\overline{\varphi^{-1}}| \leq |\alpha^{[n_2]}\overline{\varphi^{-1}}| + B_\varphi D_{\varphi^{-1}}.$$

On the other hand, $|\alpha^{[n_1]}\rho| \leq B_\varphi < Y_0$ and $\alpha^{[n_1]} \in V$ together yield $Y_0 < |\alpha^{[n_1]}\tau| = n_1 - n_2$, and so

$$n_2 + B_{\varphi^{-1}} < n_1 - Y_0 + B_{\varphi^{-1}} < n_1 - B_\varphi D_{\varphi^{-1}} \leq |\alpha^{[n_2]}\overline{\varphi^{-1}}|.$$

In view of (17), we can apply Case I to φ^{-1}. Hence α is an exponentially stable attractor for φ^{-1} and, therefore, an exponentially stable repeller for φ. $\square$

10. Example and open problems

We include a simple example which illustrates some of the constructions introduced earlier.

Example. Let $G = \mathbb{Z} \times \mathbb{Z}_2$ and let $A = \{a, b, c\}$. We note that this is not the canonical set of generators, which would not work. Then the matched homomorphism $\pi : \widetilde{A}^* \to G$ defined by

$$a\pi = (1, 0), \quad b\pi = (0, 1), \quad c\pi = (1, 1)$$

yields

$$\mathrm{Geo}_A(G) = (a \cup c)^* \cup (a^{-1} \cup c^{-1})^* \cup \{b, b^{-1}\},$$

and we can take

$$\begin{aligned}\mathcal{R} = \{(xx^{-1}, 1) : x \in \widetilde{A}\} \cup \{(a^\varepsilon b^\delta, c^\varepsilon), (b^\delta a^\varepsilon, c^\varepsilon), (c^\varepsilon b^\delta, a^\varepsilon), (b^\delta c^\varepsilon, a^\varepsilon) : \delta, \varepsilon = \pm 1\}\\ \cup \{(ac^{-1}, b), (c^{-1}a, b), (a^{-1}c, b), (ca^{-1}, b), (b^2, 1), (b^{-2}, 1)\}\end{aligned}$$

to get $\mathrm{Geo}_A(G) = \mathrm{Irr}\,\mathscr{R}$. Ordering $\widetilde{A}$ by $a < c < a^{-1} < c^{-1} < b < b^{-1}$, we get

$$L = a^*(1 \cup c) \cup (a^{-1})^*(1 \cup c^{-1}) \cup b,$$

recognized by the automaton $\mathscr{A}$ depicted by

$$\begin{tikzcd}
{} \arrow[r, leftrightarrow] & q_0 \arrow[rr, "a"] \arrow[d, "a^{-1}"'] \arrow[drr, "{b,c,c^{-1}}"] & & q_1 \arrow[loop above, "a"] \arrow[r] \arrow[d, "c"] & {} \\
{} & q_2 \arrow[l] \arrow[rr, "c^{-1}"'] \arrow[loop below, "a^{-1}"] & & q_3 \arrow[r] & {}
\end{tikzcd} .$$

Hence $\partial L = L_\omega(\mathscr{A}) = \{a^\omega, (a^{-1})^\omega\}$.

Let φ be the endomorphism of G defined by $(m,n)\varphi = (2m,n)$. Then φ is injective and therefore uniformly continuous, admitting a continuous extension Φ to $\hat{L}$. Since $B_\varphi = 0$, it is easy to check that $\mathscr{A}'_\varphi$ is the automaton

$$\begin{tikzcd}
 & & {} & b\xi \arrow[r] & {} \\
\cdots & a^{-2}\xi \arrow[l, "a^{-1}"] & a^{-1}\xi \arrow[l, "a^{-1}"] & 1\xi \arrow[l, "a^{-1}"] \arrow[u, "b"'] \arrow[ul, leftrightarrow] \arrow[r, "a"'] & a\xi \arrow[r, "a"'] & a^2\xi \arrow[r, "a"'] & \cdots
\end{tikzcd}$$

and

$$1\xi = (1,1,1,q_0), \quad b\xi = (1,1,1,q_3), \quad a^n\xi = (1,1,a^n,q_1), \quad a^{-n}\xi = (1,1,a^{-n},q_2)$$

for $n \geq 1$. Note that, in general, we ignore how to compute $\mathscr{A}'_\varphi$, our proofs being far from constructive!

It is immediate that $\mathrm{Fix}\,\Phi = \{1, b, a^\omega, (a^{-1})^\omega\}$. Moreover, the regular infinite fixed points a^ω and $(a^{-1})^\omega$ are both exponentially stable attractors.

Finally, we end the paper with some easily predictable open problems.

Problem 10.1. Is it possible to generalize Theorems 4.1, 8.5, and 9.1 to arbitrary finitely generated hyperbolic groups?

Paulin proved [1989] that Theorem 4.1 holds for automorphisms of hyperbolic groups.

Problem 10.2. Is $\mathrm{Fix}\,\varphi$ effectively computable when φ is an endomorphism of a finitely generated virtually free group?

For the moment, only the case of free group automorphisms is known; see [Bogopolski and Maslakova 2012].

Another natural question to ask in this context is whether similar results hold for *equalizers*. Given homomorphisms $\varphi, \psi : G \to G'$, let

$$\mathrm{Eq}(\varphi, \psi) = \{x \in G : x\varphi = x\psi\}.$$

Problem 10.3. Given homomorphisms $\varphi, \psi : G \to G'$ of finitely generated virtually free groups with φ injective, is $\mathrm{Eq}(\varphi, \psi)$ finitely generated?

This question has been solved by Goldstein and Turner for free groups [1986]. The restriction to the case where at least one of the homomorphisms is injective is required even in the free group case (see [Gersten 1987] and [Ventura 2002, Section 3] for counterexamples).

References

[Berstel 1979] J. Berstel, *Transductions and context-free languages*, Leitfäden der Angewandten Mathematik und Mechanik **38**, B. G. Teubner, Stuttgart, 1979. MR 80j:68056 Zbl 0424.68040

[Bestvina and Handel 1992] M. Bestvina and M. Handel, "Train tracks and automorphisms of free groups", *Ann. of Math.* (2) **135**:1 (1992), 1–51. MR 92m:20017 Zbl 0757.57004

[Bogopolski and Maslakova 2012] O. Bogopolski and O. Maslakova, "A basis of the fixed point subgroup of an automorphism of a free group", preprint, 2012. arXiv 1204.6728

[Collins and Turner 1994] D. J. Collins and E. C. Turner, "Efficient representatives for automorphisms of free products", *Michigan Math. J.* **41**:3 (1994), 443–464. MR 95k:20039 Zbl 0820.20035

[Cooper 1987] D. Cooper, "Automorphisms of free groups have finitely generated fixed point sets", *J. Algebra* **111**:2 (1987), 453–456. MR 89a:20024 Zbl 0628.20029

[Dugundji 1966] J. Dugundji, *Topology*, Allyn and Bacon, Boston, MA, 1966. MR 33 #1824 Zbl 0144.21501

[Epstein et al. 1992] D. B. A. Epstein, J. W. Cannon, D. F. Holt, S. V. F. Levy, M. S. Paterson, and W. P. Thurston, *Word processing in groups*, Jones and Bartlett, Boston, 1992. MR 93i:20036 Zbl 0764.20017

[Gaboriau et al. 1998] D. Gaboriau, A. Jaeger, G. Levitt, and M. Lustig, "An index for counting fixed points of automorphisms of free groups", *Duke Math. J.* **93**:3 (1998), 425–452. MR 99f:20051 Zbl 0946.20010

[Gersten 1987] S. M. Gersten, "Fixed points of automorphisms of free groups", *Adv. in Math.* **64**:1 (1987), 51–85. MR 88f:20042 Zbl 0616.20014

[Ghys and de la Harpe 1990] E. Ghys and P. de la Harpe (editors), *Sur les groupes hyperboliques d'après Mikhael Gromov*, Progress in Mathematics **83**, Birkhäuser, Boston, 1990. MR 92f:53050 Zbl 0731.20025

[Gilman et al. 2007] R. H. Gilman, S. Hermiller, D. F. Holt, and S. Rees, "A characterisation of virtually free groups", *Arch. Math.* (*Basel*) **89**:4 (2007), 289–295. MR 2009a:20039 Zbl 1139.20029

[Goldstein and Turner 1985] R. Z. Goldstein and E. C. Turner, "Monomorphisms of finitely generated free groups have finitely generated equalizers", *Invent. Math.* **82**:2 (1985), 283–289. MR 88b:57007 Zbl 0582.20023

[Goldstein and Turner 1986] R. Z. Goldstein and E. C. Turner, "Fixed subgroups of homomorphisms of free groups", *Bull. London Math. Soc.* **18**:5 (1986), 468–470. MR 87m:20096 Zbl 0576.20016

[Ladra and Silva 2011] M. Ladra and P. V. Silva, "The generalized conjugacy problem for virtually free groups", *Forum Math.* **23**:3 (2011), 447–482. MR 2012f:20101 Zbl 1230.20033

[Maslakova 2003] O. S. Maslakova, "The fixed point group of a free group automorphism", *Algebra Logika* **42**:4 (2003), 422–472. In Russian; translated in *Algebra and Logic* **42** (2003), 237–265. MR 2004i:20044 Zbl 1031.20014

[Paulin 1989] F. Paulin, "Points fixes des automorphismes de groupe hyperbolique", *Ann. Inst. Fourier* (*Grenoble*) **39**:3 (1989), 651–662. MR 91a:20030 Zbl 0674.20022

[Sakarovitch 2003] J. Sakarovitch, *Eléments de Théorie des Automates*, Vuibert, Paris, 2003. In French. Zbl 1178.68002

[Silva 2010] P. V. Silva, "Fixed points of endomorphisms over special confluent rewriting systems", *Monatsh. Math.* **161**:4 (2010), 417–447. MR 2011k:20118 Zbl 1236.20063

[Silva 2012] P. V. Silva, "Fixed points of endomorphisms of certain free products", *RAIRO Theor. Inform. Appl.* **46**:1 (2012), 165–179. MR 2904968 Zbl 06074750

[Sykiotis 2002] M. Sykiotis, "Fixed points of symmetric endomorphisms of groups", *Internat. J. Algebra Comput.* **12**:5 (2002), 737–745. MR 2003k:20049 Zbl 1010.20016

[Sykiotis 2007] M. Sykiotis, "Fixed subgroups of endomorphisms of free products", *J. Algebra* **315**:1 (2007), 274–278. MR 2008g:20052 Zbl 1130.20032

[Väisälä 2005] J. Väisälä, "Gromov hyperbolic spaces", *Expo. Math.* **23**:3 (2005), 187–231. MR 2006j:53055 Zbl 1087.53039

[Ventura 2002] E. Ventura, "Fixed subgroups in free groups: A survey", pp. 231–255 in *Combinatorial and geometric group theory* (New York, 2000/Hoboken, NJ, 2001), edited by S. Cleary et al., Contemp. Math. **296**, Amer. Math. Soc., Providence, RI, 2002. MR 2003f:20037 Zbl 1025.20012

Received April 12, 2012. Revised November 7, 2012.

PEDRO V. SILVA
CENTRO DE MATEMÁTICA, FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO
R. CAMPO ALEGRE 687
4169-007 PORTO
PORTUGAL

pvsilva@fc.up.pt

PACIFIC JOURNAL OF MATHEMATICS
Vol. 263, No. 1, 2013
dx.doi.org/10.2140/pjm.2013.263.241

THE SHARP LOWER BOUND FOR THE FIRST POSITIVE EIGENVALUE OF THE FOLLAND–STEIN OPERATOR ON A CLOSED PSEUDOHERMITIAN $(2n+1)$-MANIFOLD

CHIN-TUNG WU

In this paper, we obtain a sharp lower bound estimate for the first nonzero eigenvalue of the Folland–Stein operator $\mathscr{L}_c$, $|c| \le n$, on a closed pseudohermitian $(2n+1)$-manifold M. This generalizes the first nonzero eigenvalue estimates of the sublaplacian and Kohn Laplacian.

1. Introduction

Let (M, J, θ) be a closed pseudohermitian $(2n+1)$-manifold (see the next section for basic notions in pseudohermitian geometry). A. Greenleaf [1985], S.-Y. Li and H.-S. Luk [2004], and H.-L. Chiu [2006] proved the sharp lower bound of the first positive eigenvalue λ_1^0 of the sublaplacian Δ_b on a pseudohermitian $(2n+1)$-manifold M. More precisely, it was proved that

$$\lambda_1^0 \ge \frac{nk}{n+1}$$

if $[\mathrm{Ric} - \frac{n+1}{2}\,\mathrm{Tor}](Z, Z) \ge k\langle Z, Z\rangle$ for all $Z \in T_{1,0}$, some positive constant k, on a closed pseudohermitian $(2n+1)$-manifold with the nonnegative CR Paneitz operator P_0 if $n = 1$ (also see [Chang and Wu 2010]).

Very recently, S. Chanillo, H.-L. Chiu and P. Yang [Chanillo et al. 2012] obtained the sharp lower bound of the first positive eigenvalue λ_1^n of the Kohn Laplacian $\Box_b$ on a pseudohermitian $(2n+1)$-manifold M with $n = 1, 2$. Later, S.-C. Chang and the author [Chang and Wu $\ge$ 2013] proved the same result for $n \ge 3$. They showed that

$$\lambda_1^n \ge \frac{2nk}{n+1}$$

if $\mathrm{Ric}(Z, Z) \ge k\langle Z, Z\rangle$ for all $Z \in T_{1,0}$, some positive constant k, on a closed pseudohermitian $(2n+1)$-manifold M with nonnegative CR Paneitz operator P_0 if $n = 1$. Note that there is no assumption involving the pseudohermitian torsion.

Research supported in part by NSC.
MSC2010: primary 32V05, 32V20; secondary 53C56.
Keywords: Folland–Stein operator, sublaplacian, Kohn Laplacian, CR Paneitz operator, pseudohermitian manifold, pseudohermitian Ricci curvature, pseudohermitian torsion.

In this paper, we generalize the first nonzero eigenvalue estimates of the sublaplacian Δ_b and Kohn Laplacian $\Box_b$ to the Folland–Stein operator $\mathscr{L}_c$. First we need some definitions.

Definition 1.1. Let (M, J, θ) be a closed pseudohermitian $(2n+1)$-manifold. Define

$$P\varphi = \sum_{\alpha=1}^{n} \big(\varphi_{\bar{\alpha}}{}^{\bar{\alpha}}{}_{\beta} + inA_{\beta\alpha}\varphi^{\alpha}\big)\theta^{\beta} = (P_\beta\varphi)\theta^{\beta}, \quad \beta = 1, 2, \ldots, n,$$

which is an operator that characterizes CR-pluriharmonic functions ([Lee 1988] for $n = 1$ and [Graham and Lee 1988] for $n \geq 2$). Here $P_\beta\varphi = \sum_{\alpha=1}^{n}(\varphi_{\bar{\alpha}}{}^{\bar{\alpha}}{}_{\beta} + inA_{\beta\alpha}\varphi^{\alpha})$ and $\bar{P}\varphi = (\bar{P}_\beta\varphi)\theta^{\bar{\beta}}$, the conjugate of P. Moreover, we define

$$P_0\varphi = \delta_b(P\varphi),$$

which is the so-called CR Paneitz operator P_0. Here δ_b is the divergence operator that takes $(1, 0)$-forms to functions by $\delta_b(\sigma_\alpha\theta^\alpha) = \sigma_\alpha{,}^{\alpha}$ and $\bar{\delta}_b(\sigma_{\bar{\alpha}}\theta^{\bar{\alpha}}) = \sigma_{\bar{\alpha}}{,}^{\bar{\alpha}}$. If we define $\partial_b\varphi = \varphi_\alpha\theta^\alpha$ and $\bar{\partial}_b\varphi = \varphi_{\bar{\alpha}}\theta^{\bar{\alpha}}$, then the formal adjoint of ∂_b on functions (with respect to the Levi form and the volume form $\theta \wedge (d\theta)^n$) is $\partial_b^* = -\delta_b$.

We observe that P_0 is a real and symmetric operator and

$$\int \langle P\varphi, \partial_b\varphi\rangle = -\int (P_0\varphi)\bar{\varphi}.$$

Definition 1.2. We say that the Paneitz operator P_0 with respect to (J, θ) is nonnegative if, for all C^∞ smooth functions φ,

$$\int (P_0\varphi)\bar{\varphi} \geq 0.$$

Remark 1.3. When (M, J, θ) is a closed pseudohermitian 3-manifold with vanishing pseudohermitian torsion, the corresponding CR Paneitz operator is nonnegative [Chang et al. 2007]. Unlike $n = 1$, let (M, J, θ) be a closed pseudohermitian $(2n+1)$-manifold with $n \geq 2$. The corresponding CR Paneitz operator is always nonnegative as in (3-4).

Definition 1.4 [Graham and Lee 1988]. Let (M, J, θ) be a closed pseudohermitian $(2n+1)$-manifold. We define the purely holomorphic second-order operator Q by

$$Q\varphi = 2i(A^{\alpha\beta}\varphi_\alpha)_{,\beta}\,.$$

Note that $[T, \Delta_b] = 2\,\mathrm{Im}\,Q$ and

$$\begin{aligned} 4P_0 &= \Delta_b^2 + n^2T^2 - 2n\,\mathrm{Re}\,Q = (\Delta_b + inT)(\Delta_b - inT) - 2nQ \\ &= (\Delta_b - inT)(\Delta_b + inT) - 2n\bar{Q}. \end{aligned} \tag{1-1}$$

Now we consider, for $c \in \mathbb{R}$, the self-adjoint operators

$$\mathscr{L}_c = \Delta_b + icT,$$

with $|c| \le n$. By a result in [Folland and Stein 1974], each $\mathscr{L}_c$, $|c| < n$, is a subelliptic operator of order $\frac{1}{2}$; hence $\mathscr{L}_c$ has a discrete spectrum tending to $+\infty$.

In the following we can obtain a sharp lower bound for the first nonzero eigenvalue λ_1^c of the Folland–Stein operator $\mathscr{L}_c$, $c \in \mathbb{R}$ with $|c| \le n$, on a closed pseudohermitian $(2n+1)$-manifold.

Theorem 1.5. *Let (M, J, θ) be a closed pseudohermitian $(2n+1)$-manifold. Suppose that*

$$\begin{cases} \left[\mathrm{Ric} - \dfrac{(n-c)(n+1)}{2(n+c)}\,\mathrm{Tor}\right](Z, Z) \ge k\langle Z, Z\rangle & \text{if } c \ge 0, \\ \left[\mathrm{Ric} - \dfrac{(n+c)(n+1)}{2(n-c)}\,\mathrm{Tor}\right](\overline{Z}, \overline{Z}) \ge k\langle Z, Z\rangle & \text{if } c < 0, \end{cases} \tag{1-2}$$

for a positive constant k and for all $Z \in T_{1,0}$. In addition we assume the Paneitz operator P_0 is nonnegative if $n = 1$. Then the first nonzero eigenvalue of $\mathscr{L}_c$, $|c| \le n$, must satisfy

$$\lambda_1^c \ge \frac{n+|c|}{n+1}k.$$

Note that the constant in the torsion tensor term in assumption (1-2) depends on the variable c. In the standard pseudohermitian $(2n+1)$-sphere $(S^{2n+1}, \hat{J}, \hat{\theta})$ with the induced CR structure $\hat{J}$ from $\mathbb{C}^{n+1}$ and the standard contact form $\hat{\theta}$, we can show that the lower bound in Theorem 1.5 is sharp (see Section 4).

In particular, when (M, J, θ) is a closed pseudohermitian 3-manifold with vanishing pseudohermitian torsion, the corresponding CR Paneitz operator P_0 is nonnegative.

Corollary 1.6. *Let (M, J, θ) be a closed pseudohermitian $(2n+1)$-manifold with vanishing pseudohermitian torsion. Suppose that*

$$\begin{cases} \mathrm{Ric}(Z, Z) \ge k\langle Z, Z\rangle & \text{if } c \ge 0, \\ \mathrm{Ric}(\overline{Z}, \overline{Z}) \ge k\langle Z, Z\rangle & \text{if } c < 0, \end{cases}$$

for a positive constant k and for all $Z \in T_{1,0}$. Then the first nonzero eigenvalue of $\mathscr{L}_c$, $|c| \le n$, must satisfy

$$\lambda_1^c \ge \frac{n+|c|}{n+1}k.$$

Moreover, when $c = n$, the operator $\mathscr{L}_n$ is just the Kohn Laplacian: $\mathscr{L}_n = \Box_b$.

Corollary 1.7. *Let (M, J, θ) be a closed pseudohermitian $(2n+1)$-manifold. Suppose that*

$$\mathrm{Ric}(Z, Z) \ge k\langle Z, Z\rangle$$

for a positive constant k and for all $Z \in T_{1,0}$. In addition we assume the Paneitz operator P_0 is nonnegative if $n = 1$. Then the first nonzero eigenvalue of the Kohn Laplacian $\Box_b$ must satisfy

$$\lambda_1^n \geq \frac{2nk}{n+1}.$$

When $c = 0$, the operator $\mathscr{L}_0$ is just the sublaplacian Δ_b; i.e., $\mathscr{L}_0 = \Delta_b$.

Corollary 1.8. *Let (M, J, θ) be a closed pseudohermitian $(2n+1)$-manifold. Suppose that*

$$\left[\operatorname{Ric} - \frac{n+1}{2}\operatorname{Tor}\right](Z, Z) \geq k\langle Z, Z\rangle$$

for a positive constant k and for all $Z \in T_{1,0}$. In addition we assume the Paneitz operator P_0 is nonnegative if $n = 1$. Then the first nonzero eigenvalue of the sublaplacian Δ_b must satisfy

$$\lambda_1^0 \geq \frac{nk}{n+1}.$$

Further, we study the case when a sharp lower bound estimate of $\mathscr{L}_c$, $|c| \leq n$, is achieved in Section 4.

Proposition 1.9. *Under the same conditions as in Theorem 1.5, if we assume the first nonzero eigenvalue of $\mathscr{L}_c$, $0 < |c| \leq n$, satisfies*

$$\lambda_1^c = \frac{n+|c|}{n+1}k,$$

$$\int A^{\alpha\beta}\varphi_{c\alpha}\bar{\varphi}_{c\beta} = 0 \tag{1-3}$$

for a corresponding eigenfunction φ_c of $\mathscr{L}_c$ with respect to λ_1^c and with $\int\langle\varphi_c, \varphi_c\rangle = 1$, then the eigenfunction φ_c will satisfy

$$\int|\bar{\partial}_b\varphi_c|^2 = \frac{n(n+c)}{2(n^2+c^2)}\lambda_1^c \quad \textit{and} \quad \int|\partial_b\varphi_c|^2 = \frac{n(n-c)}{2(n^2+c^2)}\lambda_1^c; \tag{1-4}$$

thus we also have

$$\int\langle\Delta_b\varphi_c, \varphi_c\rangle = \frac{n^2}{n^2+c^2}\lambda_1^c \quad \textit{and} \quad \int i\langle T\varphi_c, \varphi_c\rangle = \frac{c}{n^2+c^2}\lambda_1^c.$$

Letting $c \to 0^+$, we see that $\int|\bar{\partial}_b\varphi_c|^2 = \int|\partial_b\varphi_c|^2 = \frac{1}{2}\lambda_1^0$ and $\int i\langle T\varphi_c, \varphi_c\rangle = 0$ for $c = 0$. When $c = n$, from (1-4), we get that $\partial_b\varphi_n = 0$ and thus $\overline{\Box}_b\varphi_n = 0$. This implies that the corresponding eigenfunction φ_n of $\mathscr{L}_n = \Box_b$ with respect to λ_1^n will also satisfy

$$\Delta_b\varphi_n = \frac{nk}{n+1}\varphi_n.$$

This yields that φ_n achieves a sharp lower bound for the first nonzero eigenvalue of the sublaplacian Δ_b. Furthermore, it can be showed the pseudohermitian torsion $A_{\alpha\beta}$ of M is zero; thus (M, J, θ) is the standard pseudohermitian $(2n+1)$-sphere $(S^{2n+1}, \hat{J}, \hat{\theta})$ (see [Chang and Wu $\geq$ 2013] for details).

2. Basic materials

Let us give a brief introduction to pseudohermitian geometry (see [Lee 1988] for more details). Let (M, ξ) be a $(2n+1)$-dimensional, orientable, contact manifold with contact structure ξ, $\dim_R \xi = 2n$. A CR structure compatible with ξ is an endomorphism $J : \xi \to \xi$ such that $J^2 = -1$. We also assume that J satisfies the following integrability condition: if X and Y are in ξ, then so is $[JX, Y]+[X, JY]$, and $J([JX, Y]+[X, JY]) = [JX, JY]-[X, Y]$. A CR structure J can extend to $\mathbb{C}\otimes\xi$ and decomposes $\mathbb{C}\otimes\xi$ into the direct sum of $T_{1,0}$ and $T_{0,1}$, which are eigenspaces of J with respect to i and $-i$, respectively. A pseudohermitian structure compatible with ξ is a CR structure J compatible with ξ together with a choice of contact form θ. Such a choice determines a unique real vector field T transverse to ξ, called the characteristic vector field of θ, such that $\theta(T) = 1$ and $\mathscr{L}_T\theta = 0$ or $d\theta(T, \cdot) = 0$. Let $\{T, Z_\alpha, Z_{\bar{\alpha}}\}$ be a frame of $TM\otimes\mathbb{C}$, where Z_α is any local frame of $T_{1,0}$, $Z_{\bar{\alpha}} = \overline{Z}_\alpha \in T_{0,1}$ and T is the characteristic vector field. Then $\{\theta, \theta^\alpha, \theta^{\bar{\alpha}}\}$, which is the coframe dual to $\{T, Z_\alpha, Z_{\bar{\alpha}}\}$, satisfies

$$d\theta = i h_{\alpha\bar{\beta}} \theta^\alpha \wedge \theta^{\bar{\beta}}$$

for some positive definite hermitian matrix of functions $(h_{\alpha\bar{\beta}})$. Actually we can always choose Z_α such that $h_{\alpha\bar{\beta}} = \delta_{\alpha\beta}$; hence, throughout this paper, we assume $h_{\alpha\bar{\beta}} = \delta_{\alpha\beta}$.

The Levi form $\langle\, ,\rangle$ is the Hermitian form on $T_{1,0}$ defined by

$$\langle Z, W\rangle = -i\big\langle d\theta, Z\wedge\overline{W}\big\rangle.$$

We can extend $\langle\, ,\rangle$ to $T_{0,1}$ by defining $\langle\overline{Z}, \overline{W}\rangle = \overline{\langle Z, W\rangle}$ for all $Z, W \in T_{1,0}$. The Levi form induces naturally a Hermitian form on the dual bundle of $T_{1,0}$, also denoted by $\langle\, ,\rangle$, and hence on all the induced tensor bundles.

The pseudohermitian connection of (J, θ) is the connection ∇ on $TM\otimes\mathbb{C}$ (and extended to tensors) given in terms of a local frame $Z_\alpha \in T_{1,0}$ by

$$\nabla Z_\alpha = \omega_\alpha{}^\beta \otimes Z_\beta, \quad \nabla Z_{\bar{\alpha}} = \omega_{\bar{\alpha}}{}^{\bar{\beta}} \otimes Z_{\bar{\beta}}, \quad \nabla T = 0,$$

where $\omega_\alpha{}^\beta$ are the 1-forms uniquely determined by the following equations:

$$d\theta^\beta = \theta^\alpha\wedge\omega_\alpha{}^\beta + \theta\wedge\tau^\beta, \quad \tau_\alpha\wedge\theta^\alpha = 0, \quad \omega_\alpha{}^\beta + \omega_{\bar{\beta}}{}^{\bar{\alpha}} = 0.$$

We can write $\tau_\alpha = A_{\alpha\beta}\theta^\beta$ with $A_{\alpha\beta} = A_{\beta\alpha}$. The curvature of the Webster–Stanton connection, expressed in terms of the coframe $\{\theta = \theta^0, \theta^\alpha, \theta^{\bar\alpha}\}$, is

$$\Pi_\beta{}^\alpha = \overline{\Pi_{\bar\beta}{}^{\bar\alpha}} = d\omega_\beta{}^\alpha - \omega_\beta{}^\gamma \wedge \omega_\gamma{}^\alpha,$$
$$\Pi_0{}^\alpha = \Pi_\alpha{}^0 = \Pi_0{}^{\bar\beta} = \Pi_{\bar\beta}{}^0 = \Pi_0{}^0 = 0.$$

Webster showed that $\Pi_\beta{}^\alpha$ can be written as

$$\Pi_\beta{}^\alpha = R_\beta{}^\alpha{}_{\rho\bar\sigma}\theta^\rho \wedge \theta^{\bar\sigma} + W_\beta{}^\alpha{}_\rho \theta^\rho \wedge \theta - W^\alpha{}_{\beta\bar\rho}\theta^{\bar\rho} \wedge \theta + i\theta_\beta \wedge \tau^\alpha - i\tau_\beta \wedge \theta^\alpha,$$

where the coefficients satisfy

$$R_{\beta\bar\alpha\rho\bar\sigma} = \overline{R_{\alpha\bar\beta\sigma\bar\rho}} = R_{\bar\alpha\beta\bar\sigma\rho} = R_{\rho\bar\alpha\beta\bar\sigma}, \quad W_{\beta\bar\alpha\gamma} = W_{\gamma\bar\alpha\beta}.$$

We will denote components of covariant derivatives with indices preceded by comma; thus write $A_{\alpha\beta,\gamma}$. The indices $\{0, \alpha, \bar\alpha\}$ indicate derivatives with respect to $\{T, Z_\alpha, Z_{\bar\alpha}\}$. For derivatives of a function, we will often omit the comma, for instance, $\varphi_\alpha = Z_\alpha\varphi$, $\varphi_{\alpha\bar\beta} = Z_{\bar\beta}Z_\alpha\varphi - \omega_\alpha{}^\gamma(Z_{\bar\beta})Z_\gamma\varphi$, $\varphi_0 = T\varphi$ for a (smooth) function φ. Let the Cauchy–Riemann operator ∂_b be defined locally by $\partial_b\varphi = \varphi_\alpha\theta^\alpha$, and let $\bar\partial_b$ be the conjugate of ∂_b. For a function φ, the subgradient ∇_b is defined locally by $\nabla_b\varphi = \varphi^\alpha Z_\alpha + \varphi^{\bar\alpha}Z_{\bar\alpha}$. The sublaplacian Δ_b, the Kohn Laplacian $\Box_b$, and the Folland–Stein operator $\mathscr{L}_c$ on functions are defined by

$$\Delta_b\varphi = -(\varphi_\alpha{}^\alpha + \varphi_{\bar\alpha}{}^{\bar\alpha}), \quad \Box_b\varphi = (\Delta_b + inT)\varphi, \quad \mathscr{L}_c\varphi = (\Delta_b + icT)\varphi.$$

The Webster–Ricci tensor and the torsion tensor on $T_{1,0}$ are defined by

$$\mathrm{Ric}(X, Y) = R_{\alpha\bar\beta}X^\alpha Y^{\bar\beta},$$
$$\mathrm{Tor}(X, Y) = i\sum_{\alpha,\beta}\big(A_{\bar\alpha\bar\beta}X^{\bar\alpha}Y^{\bar\beta} - A_{\alpha\beta}X^\alpha Y^\beta\big),$$

where $X = X^\alpha Z_\alpha$, $Y = Y^\beta Z_\beta$, $R_{\alpha\bar\beta} = R_\gamma{}^\gamma{}_{\alpha\bar\beta}$. The Webster scalar curvature is $R = R_\alpha{}^\alpha = h^{\alpha\bar\beta}R_{\alpha\bar\beta}$.

3. Proof of Theorem 1.5

Let (M, J, θ) be a closed pseudohermitian $(2n+1)$-manifold. In this section, we can obtain lower bound estimates for the first nonzero eigenvalue of the Folland–Stein operator $\mathscr{L}_c$, $|c| \le n$, on a closed pseudohermitian $(2n+1)$-manifold.

First we need the following Bochner formula for the Kohn Laplacian [Chanillo et al. 2012, Equation (2.8)]).

Lemma 3.1. *For any complex-valued function φ, we have*

$$-\frac{1}{2}\Box_b\left|\bar\partial_b\varphi\right|^2=\sum_{\alpha,\beta}\left(\varphi_{\bar\alpha\bar\beta}\overline{\varphi}_{\alpha\beta}+\varphi_{\bar\alpha\beta}\overline{\varphi}_{\alpha\bar\beta}\right)+\mathrm{Ric}\big((\nabla_b\varphi)_{\mathbb{C}},(\nabla_b\varphi)_{\mathbb{C}}\big)\\ -\frac{1}{2n}\langle\bar\partial_b\varphi,\bar\partial_b\Box_b\varphi\rangle-\frac{n+1}{2n}\langle\bar\partial_b\Box_b\varphi,\bar\partial_b\varphi\rangle\\ -\frac{1}{n}\langle\overline{P}\varphi,\bar\partial_b\varphi\rangle+\frac{n-1}{n}\langle P\overline{\varphi},\partial_b\overline{\varphi}\rangle, \tag{3-1}$$

where $(\nabla_b\varphi)_{\mathbb{C}}=\varphi^\alpha Z_\alpha$ is the corresponding complex $(1,0)$-vector field of $\nabla_b\varphi$.

First we derive some useful identities which we need in the proof of Theorem 1.5. Let φ be a smooth complex-valued function on M. By integrating the Bochner formula (3-1), we have

$$0=\int\sum_{\alpha,\beta}\left(\varphi_{\bar\alpha\bar\beta}\overline{\varphi}_{\alpha\beta}+\varphi_{\bar\alpha\beta}\overline{\varphi}_{\alpha\bar\beta}\right)-\frac{n+2}{2n}\int\langle\Box_b\varphi,\Box_b\varphi\rangle\\ +\frac{2-n}{n}\int(P_0\varphi)\overline{\varphi}+\int\mathrm{Ric}\big((\nabla_b\varphi)_{\mathbb{C}},(\nabla_b\varphi)_{\mathbb{C}}\big). \tag{3-2}$$

We also have

$$\int\sum_{\alpha,\beta}\varphi_{\bar\alpha\beta}\overline{\varphi}_{\alpha\bar\beta}=\int\sum_{\alpha,\beta}\left|\overline{\varphi}_{\alpha\bar\beta}-\frac{1}{n}\overline{\varphi}_\gamma{}^\gamma h_{\alpha\bar\beta}\right|^2+\frac{1}{4n}\int\langle\Box_b\varphi,\Box_b\varphi\rangle\\ =\frac{n-1}{n}\int(P_0\varphi)\overline{\varphi}+\frac{1}{4n}\int\langle\Box_b\varphi,\Box_b\varphi\rangle. \tag{3-3}$$

Here we used the following divergence formula [Graham and Lee 1988] for the trace-free part of $\overline{\varphi}_{\alpha\bar\beta}$:

$$B_{\alpha\bar\beta}\overline{\varphi}=\overline{\varphi}_{\alpha\bar\beta}-\frac{1}{n}\overline{\varphi}_\gamma{}^\gamma h_{\alpha\bar\beta}.$$

That is,

$$\big(B^{\alpha\bar\beta}\varphi\big)\big(B_{\alpha\bar\beta}\overline{\varphi}\big)=\varphi^{\alpha\bar\beta}\big(B_{\alpha\bar\beta}\overline{\varphi}\big)=\big(\varphi^\alpha B_{\alpha\bar\beta}\overline{\varphi}\big),^{\bar\beta}-\frac{n-1}{n}\varphi^\alpha P_\alpha\overline{\varphi}\\ =\big(\varphi^\alpha B_{\alpha\bar\beta}\overline{\varphi}\big),^{\bar\beta}-\frac{n-1}{n}(\varphi P_\alpha\overline{\varphi}),^\alpha+\frac{n-1}{n}(P_0\overline{\varphi})\varphi.$$

Then we integrate both sides to get

$$\int\sum_{\alpha,\beta}\left|B_{\alpha\bar\beta}\overline{\varphi}\right|^2=\frac{n-1}{n}\int(P_0\varphi)\overline{\varphi}. \tag{3-4}$$

Taking together the two formulas (3-2) and (3-3), we get

$$\frac{n+1}{4n}\int\langle\Box_b\varphi,\Box_b\varphi\rangle=\int\sum_{\alpha,\beta}\varphi_{\bar\alpha\bar\beta}\overline{\varphi}_{\alpha\beta}+\frac{1}{n}\int(P_0\varphi)\overline{\varphi}+\int\mathrm{Ric}\big((\nabla_b\varphi)_{\mathbb{C}},(\nabla_b\varphi)_{\mathbb{C}}\big). \tag{3-5}$$

By taking complex conjugate to (3-5) and replacing $\overline{\varphi}$ by φ, one obtains

$$\frac{n+1}{4n}\int\bigl\langle\overline{\Box}_b\varphi,\overline{\Box}_b\varphi\bigr\rangle=\int\sum_{\alpha,\beta}\varphi_{\alpha\beta}\overline{\varphi}_{\overline{\alpha}\overline{\beta}}+\frac{1}{n}\int(P_0\varphi)\overline{\varphi}+\int\mathrm{Ric}\bigl((\nabla_b\overline{\varphi})_{\mathbb{C}},(\nabla_b\overline{\varphi})_{\mathbb{C}}\bigr). \tag{3-6}$$

From the formula (1-1), we have

$$\begin{aligned}4\int(P_0\varphi)\overline{\varphi}&=\int\bigl\langle(\Delta_b+inT)(\Delta_b-inT)\varphi-2nQ\varphi,\varphi\bigr\rangle\\&=\int\bigl\langle\overline{\Box}_b\varphi,\Box_b\varphi\bigr\rangle-2n\int\langle Q\varphi,\varphi\rangle.\end{aligned} \tag{3-7}$$

By (1-1), we can also obtain

$$4\int(P_0\varphi)\overline{\varphi}=\int\bigl\langle\Box_b\varphi,\overline{\Box}_b\varphi\bigr\rangle-2n\int\bigl\langle\overline{Q}\varphi,\varphi\bigr\rangle. \tag{3-8}$$

Proof of Theorem 1.5. Let φ_c be an eigenfunction of the Folland–Stein operator $\mathscr{L}_c$, $c\in\mathbb{R}$ with $|c|\le n$, with respect to the first nonzero eigenvalue λ_1^c; i.e., $\mathscr{L}_c\varphi_c=\lambda_1^c\varphi_c$.

When $0\le c\le n$, from (3-6) and (3-7) for

$$\mathscr{L}_c=\frac{n+c}{2n}\Box_b+\frac{n-c}{2n}\overline{\Box}_b,$$

we have

$$\begin{aligned}\frac{1}{2}\int\bigl\langle\Box_b\varphi_c,\mathscr{L}_c\varphi_c\bigr\rangle&=\frac{n+c}{4n}\int\bigl\langle\Box_b\varphi_c,\Box_b\varphi_c\bigr\rangle+\frac{n-c}{4n}\int\bigl\langle\Box_b\varphi_c,\overline{\Box}_b\varphi_c\bigr\rangle\\&=\frac{n+c}{n+1}\int\sum_{\alpha,\beta}\varphi_{c\overline{\alpha}\overline{\beta}}\overline{\varphi}_{c\alpha\beta}+\frac{n+2-c}{n+1}\int(P_0\varphi_c)\overline{\varphi}_c\\&\qquad+\frac{n+c}{n+1}\int\mathrm{Ric}\bigl((\nabla_b\varphi_c)_{\mathbb{C}},(\nabla_b\varphi_c)_{\mathbb{C}}\bigr)+\frac{n-c}{2}\int\bigl\langle\overline{Q}\varphi_c,\varphi_c\bigr\rangle\\&=\frac{n+c}{n+1}\int\sum_{\alpha,\beta}\varphi_{c\overline{\alpha}\overline{\beta}}\overline{\varphi}_{c\alpha\beta}+\frac{n+2-c}{n+1}\int(P_0\varphi_c)\overline{\varphi}_c\\&\qquad+\frac{n+c}{n+1}\int\Bigl[\mathrm{Ric}-\frac{(n-c)(n+1)}{2(n+c)}\,\mathrm{Tor}\Bigr]\bigl((\nabla_b\varphi_c)_{\mathbb{C}},(\nabla_b\varphi_c)_{\mathbb{C}}\bigr),\end{aligned}$$

where we used the equation

$$\int\bigl\langle\overline{Q}\varphi_c,\varphi_c\bigr\rangle=-\int\mathrm{Tor}\bigl((\nabla_b\varphi_c)_{\mathbb{C}},(\nabla_b\varphi_c)_{\mathbb{C}}\bigr),$$

since $\int\langle\overline{Q}\varphi_c,\varphi_c\rangle$ is real, and thus $\int\langle\overline{Q}\varphi_c,\varphi_c\rangle=2\int iA^{\overline{\alpha}\overline{\beta}}\varphi_{c\overline{\alpha}}\overline{\varphi}_{c\overline{\beta}}=-2\int iA^{\alpha\beta}\varphi_{c\alpha}\overline{\varphi}_{c\beta}$.

Hence, if P_0 is nonnegative and

$$[\mathrm{Ric}-\frac{(n-c)(n+1)}{2(n+c)}\,\mathrm{Tor}]((\nabla_b\varphi_c)_{\mathbb{C}},(\nabla_b\varphi_c)_{\mathbb{C}})\ge k|\overline{\partial}_b\varphi_c|^2,$$

we have

$$\begin{aligned}
\text{(3-9)}\quad \lambda_1^c\int\left|\bar\partial_b\varphi_c\right|^2 &= \frac{n+c}{n+1}\int\sum_{\alpha,\beta}\varphi_{c\bar\alpha\bar\beta}\bar\varphi_{c\alpha\beta}+\frac{n+2-c}{n+1}\int(P_0\varphi_c)\bar\varphi_c\\
&\quad+\frac{n+c}{n+1}\int\left[\mathrm{Ric}-\frac{(n-c)(n+1)}{2(n+c)}\mathrm{Tor}\right]\big((\nabla_b\varphi_c)_{\mathbb{C}},(\nabla_b\varphi_c)_{\mathbb{C}}\big)\\
&\geq\frac{n+c}{n+1}k\int|\bar\partial_b\varphi_c|^2,
\end{aligned}$$

which shows that $\lambda_1^c\geq\frac{n+c}{n+1}k$.

When $-n\leq c<0$, from (3-5) and (3-8), the same computation shows that

$$\begin{aligned}
\frac{1}{2}\int\langle\overline{\Box}_b\varphi_c,\mathscr{L}_c\varphi_c\rangle &= \frac{n+c}{4n}\int\langle\overline{\Box}_b\varphi_c,\Box_b\varphi_c\rangle+\frac{n-c}{4n}\int\langle\overline{\Box}_b\varphi_c,\overline{\Box}_b\varphi_c\rangle\\
&=\frac{n-c}{n+1}\int\sum_{\alpha,\beta}\varphi_{c\alpha\beta}\bar\varphi_{c\bar\alpha\bar\beta}+\frac{n+2+c}{n+1}\int(P_0\varphi_c)\bar\varphi_c\\
&\quad+\frac{n-c}{n+1}\int\left[\mathrm{Ric}-\frac{(n+c)(n+1)}{2(n-c)}\mathrm{Tor}\right]\big((\nabla_b\bar\varphi_c)_{\mathbb{C}},(\nabla_b\bar\varphi_c)_{\mathbb{C}}\big).
\end{aligned}$$

Thus, if P_0 is nonnegative and

$$[\mathrm{Ric}-\frac{(n+c)(n+1)}{2(n-c)}\mathrm{Tor}]((\nabla_b\bar\varphi_c)_{\mathbb{C}},(\nabla_b\bar\varphi_c)_{\mathbb{C}})\geq k|\partial_b\varphi_c|^2,$$

we get

$$\begin{aligned}
\lambda_1^c\int|\partial_b\varphi_c|^2 &= \frac{n-c}{n+1}\int\sum_{\alpha,\beta}\varphi_{c\alpha\beta}\bar\varphi_{c\bar\alpha\bar\beta}+\frac{n+2+c}{n+1}\int(P_0\varphi_c)\bar\varphi_c\\
&\quad+\frac{n-c}{n+1}\int\left[\mathrm{Ric}-\frac{(n+c)(n+1)}{2(n-c)}\mathrm{Tor}\right]\big((\nabla_b\bar\varphi_c)_{\mathbb{C}},(\nabla_b\bar\varphi_c)_{\mathbb{C}}\big)\\
&\geq\frac{n-c}{n+1}k\int|\partial_b\varphi_c|^2,
\end{aligned}$$

which implies that $\lambda_1^c\geq\dfrac{n-c}{n+1}k$. This completes the proof of Theorem 1.5. □

4. Example and proof of Proposition 1.9

In this section, we calculate the eigenvalues of sublaplacian Δ_b, Kohn Laplacian $\Box_b$, and the Folland–Stein operator $\mathscr{L}_c$, $|c|\leq n$, of the standard pseudohermitian $(2n+1)$-sphere S^{2n+1}. We show that the lower bound in Theorem 1.5 is sharp. We also study the case when a sharp lower bound estimate of $\mathscr{L}_c$, $|c|\leq n$, is achieved.

Let $S^{2n+1}=\left\{(z_0,z_1,\dots,z_n)\mid\sum_{j=0}^n z_j\bar z_j=1\right\}\subset\mathbb{C}^{n+1}$ with the induced CR structure from $\mathbb{C}^{n+1}$ and the contact form $\theta=\frac{i}{2}(\bar\partial u-\partial u)|_{S^{2n+1}}$ where $u=\left(\sum_{j=0}^n z_j\bar z_j\right)-1$ is a defining function. It can be shown that the pseudohermitian torsion is free and the Webster–Ricci tensor is given by $R_{\alpha\bar{\beta}} = (n+1)h_{\alpha\bar{\beta}}$.

We write

$$\partial_j = \frac{\partial}{\partial z_j},\ \bar{\partial}_j = \frac{\partial}{\partial \bar{z}_j} \quad (0 \le j \le n), \qquad \partial_{j\bar{k}} = \partial_j \partial_{\bar{k}} \quad (0 \le j, k \le n),$$

and $z = (z_0, z_1, \dots, z_n)$, $\delta = (\partial_0, \partial_1, \dots, \partial_n)$. We let $\cdot$ denote the dot product. Then, by the computation in Section 1 of [Geller 1980], we have

$$\mathscr{L}_c = 2\biggl(-\Delta + \sum_{j,k=0}^{n} z_j \bar{z}_k \partial_j \partial_{\bar{k}}\biggr) + (n+c)\bar{z}\cdot\bar{\delta} + (n-c) z \cdot \delta,$$

where $\Delta = \sum_{j=0}^{n} \partial_j \partial_{\bar{j}}$ is the standard Laplacian on $\mathbb{C}^{n+1}$. In particular, we have

$$\Delta_b = 2\biggl(-\Delta + \sum_{j,k=0}^{n} z_j \bar{z}_k \partial_j \partial_{\bar{k}}\biggr) + n\bigl(\bar{z}\cdot\bar{\delta} + z\cdot\delta\bigr),$$

$$\Box_b = 2\biggl(-\Delta + \sum_{j,k=0}^{n} z_j \bar{z}_k \partial_j \partial_{\bar{k}}\biggr) + 2n\bar{z}\cdot\bar{\delta}.$$

If Y is a bigraded spherical harmonic of type (p,q) on $\mathbb{C}^{n+1}$ (a harmonic polynomial which is a linear combination in terms of the form $z^{\rho}\bar{z}^{\gamma}$, where ρ, γ are multiindices with $|\rho| = p$, $|\gamma| = q$), then $\mathscr{L}_c Y = \bigl(2pq + (n+c)q + (n-c)p\bigr)Y$. Similarly,

$$\Delta_b Y = (2pq + n(p+q))Y, \quad \Box_b Y = 2q(p+n)Y.$$

This example shows that the lower bound in Theorem 1.5 is sharp.

Now we study the case when a sharp lower bound estimate for the first nonzero eigenvalue of the Folland–Stein operator $\mathscr{L}_c$, $|c| \le n$, on a pseudohermitian $(2n+1)$-manifold M is achieved. We only consider the case when the constant c is nonnegative. The same computation follows when c is negative.

First, from (3-9), we have the following observation.

Lemma 4.1. *Under the same conditions as in Theorem 1.5, when the first nonzero eigenvalue of $\mathscr{L}_c$, $0 \le c \le n$, satisfies*

$$\lambda_1^c = \frac{n+c}{n+1}k,$$

then the corresponding eigenfunction φ_c will satisfy

$$\varphi_{c\bar{\alpha}\bar{\beta}} = 0 \quad \textit{for all } \alpha, \beta, \tag{4-1}$$

$$\biggl[\mathrm{Ric} - \frac{(n-c)(n+1)}{2(n+c)}\,\mathrm{Tor}\biggr]\bigl((\nabla_b\varphi_c)_{\mathbb{C}}, (\nabla_b\varphi_c)_{\mathbb{C}}\bigr) = k\bigl|\bar{\partial}_b\varphi_c\bigr|^2, \tag{4-2}$$

$$P_0\varphi_c = 0. \tag{4-3}$$

Proof of Proposition 1.9. The integral condition (1-3) says that

$$\int \langle Q\varphi_c, \varphi_c\rangle = -2i \int A^{\alpha\beta}\varphi_{c\alpha}\bar{\varphi}_{c\beta} = 0,$$

and then by integration by parts, we obtain

$$\int \langle \overline{Q}\varphi_c, \varphi_c\rangle = \int \langle \varphi_c, Q\varphi_c\rangle = \int \langle Q\varphi_c, \varphi_c\rangle = 0. \tag{4-4}$$

From (1-1), one can see that

$$4P_0 = [\Delta_b - i(n^2/c)T][\Delta_b + icT] - \frac{1}{2c}\big[(2nc+n+c)\overline{Q} + (2nc-n-c)Q\big].$$

Then, from (4-3) and (4-4), one obtains

$$\begin{aligned} 0 = 4\int (P_0\varphi_c)\bar{\varphi}_c &= \lambda_1^c \int \big\langle [\Delta_b - i(n^2/c)T]\varphi_c, \varphi_c\big\rangle \\ &= \frac{1}{2}\lambda_1^c \int \big\langle [(1-n/c)\Box_b + (1+n/c)\overline{\Box}_b]\varphi_c, \varphi_c\big\rangle \\ &= \lambda_1^c \int \big[(1-n/c)|\bar{\partial}_b\varphi_c|^2 + (1+n/c)|\partial_b\varphi_c|^2\big], \end{aligned}$$

which is

$$(n-c)\int \big|\bar{\partial}_b\varphi_c\big|^2 = (n+c)\int |\partial_b\varphi_c|^2. \tag{4-5}$$

On the other hand, the equation $\mathscr{L}_c\varphi_c = (\Delta_b + icT)\varphi_c = \lambda_1^c\varphi_c$ yields

$$\begin{aligned} \lambda_1^c = \lambda_1^c \int \langle \varphi_c, \varphi_c\rangle &= \int \langle \mathscr{L}_c\varphi_c, \varphi_c\rangle \\ &= \frac{1}{2n}\int \big\langle \big[(n+c)\Box_b + (n-c)\overline{\Box}_b\big]\varphi_c, \varphi_c\big\rangle \\ &= \int (1+n/c)\big|\bar{\partial}_b\varphi_c\big|^2 + (1-n/c)|\partial_b\varphi_c|^2. \end{aligned} \tag{4-6}$$

The equations (1-4) follow from (4-5) and (4-6) easily. □

References

[Chang and Wu 2010] S.-C. Chang and C.-T. Wu, "The entropy formulas for the CR heat equation and their applications on pseudohermitian $(2n+1)$-manifolds", *Pacific J. Math.* **246**:1 (2010), 1–29. MR 2011i:58038 Zbl 1206.32016

[Chang and Wu ≥ 2013] S.-C. Chang and C.-T. Wu, "On the CR Obata theorem for Kohn Laplacian in a closed pseudo-Hermitian hypersurface in $\mathbb{C}^{n+1}$", in preparation.

[Chang et al. 2007] S.-C. Chang, J.-H. Cheng, and H.-L. Chiu, "A fourth order curvature flow on a CR 3-manifold", *Indiana Univ. Math. J.* **56**:4 (2007), 1793–1826. MR 2009a:53111 Zbl 1129.53046

[Chanillo et al. 2012] S. Chanillo, H.-L. Chiu, and P. Yang, "Embeddability for 3-dimensional Cauchy–Riemann manifolds and CR Yamabe invariants", *Duke Math. J.* **161**:15 (2012), 2909–2921. MR 2999315 Zbl 06121401

[Chiu 2006] H.-L. Chiu, "The sharp lower bound for the first positive eigenvalue of the sublaplacian on a pseudohermitian 3-manifold", *Ann. Global Anal. Geom.* **30**:1 (2006), 81–96. MR 2007j:58034 Zbl 1098.32017

[Folland and Stein 1974] G. B. Folland and E. M. Stein, "Estimates for the $\bar{\partial}_b$ complex and analysis on the Heisenberg group", *Comm. Pure Appl. Math.* **27** (1974), 429–522. MR 51 #3719 Zbl 0293.35012

[Geller 1980] D. Geller, "The Laplacian and the Kohn Laplacian for the sphere", *J. Differential Geom.* **15**:3 (1980), 417–435. MR 82i:35132 Zbl 0507.58049

[Graham and Lee 1988] C. R. Graham and J. M. Lee, "Smooth solutions of degenerate Laplacians on strictly pseudoconvex domains", *Duke Math. J.* **57**:3 (1988), 697–720. MR 90c:32031 Zbl 0699.35112

[Greenleaf 1985] A. Greenleaf, "The first eigenvalue of a sub-Laplacian on a pseudo-Hermitian manifold", *Comm. Partial Differential Equations* **10**:2 (1985), 191–217. MR 86f:58157 Zbl 0563.58034

[Lee 1988] J. M. Lee, "Pseudo-Einstein structures on CR manifolds", *Amer. J. Math.* **110**:1 (1988), 157–178. MR 89f:32034 Zbl 0638.32019

[Li and Luk 2004] S.-Y. Li and H.-S. Luk, "The sharp lower bound for the first positive eigenvalue of a sub-Laplacian on a pseudo-Hermitian manifold", *Proc. Amer. Math. Soc.* **132**:3 (2004), 789–798. MR 2005c:58056 Zbl 1041.32024

Received May 8, 2011. Revised November 29, 2012.

CHIN-TUNG WU
DEPARTMENT OF APPLIED MATHEMATICS
NATIONAL PINGTUNG UNIVERSITY OF EDUCATION
NO. 4-18 MINSHENG RD
PINGTUNG CITY 90003
TAIWAN

PACIFIC JOURNAL OF MATHEMATICS
Vol. 263, No. 1, 2013
dx.doi.org/10.2140/pjm.2013.263.253

REMARK ON "MAXIMAL FUNCTIONS ON THE UNIT n-SPHERE" BY PETER M. KNOPF (1987)

HONG-QUAN LI

Volume **129**:1 (1987), 77–84

The article in question contains an important result on the behavior of the Hardy–Littlewood maximal function M_{S^n} on the unit n-sphere, providing a weak-type linear bound that has not been improved on in the intervening decades. Unfortunately, the proof has a gap, since it relies on an incorrect intermediate result (Lemma 3). We correct the proof by providing a sharper lower bound for a trigonometry integral than the one used by Knopf.

1. Introduction

Let S^{n-1} ($n \geq 2$) denote the unit sphere of dimension $n-1$, i.e., the $n-1$ dimensional, simply connected Riemannian manifold of constant sectional curvature 1. Let $d_{S^{n-1}}$ be the induced distance and $\mu_{S^{n-1}}$ be the induced measure.

Consider the centered Hardy–Littlewood maximal function, $M_{S^{n-1}}$, on S^{n-1}, i.e.,

$$M_{S^{n-1}} f(x) = \sup_{0<r\leq\pi} \frac{1}{\mu_{S^{n-1}}\big(B_{S^{n-1}}(x,r)\big)} \int_{B_{S^{n-1}}(x,r)} |f(y)|\, d\mu_{S^{n-1}}(y), \qquad x \in S^{n-1},\ f \in L^1(S^{n-1}),$$

where $B_{S^{n-1}}(x,r)$ is the open ball with center x and radius $r>0$.

In [Knopf 1987], the following theorem is presented:

Theorem 1.1. *There exists a constant $A>0$ such that*

$$\|M_{S^{n-1}}\|_{L^1 \longrightarrow L^{1,\infty}} \leq An \quad \textit{for all } n \geq 2. \tag{1-1}$$

The author is partially supported by the NSF of China (grant no. 11171070), NCET-09-0316 and "The Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning".
MSC2010: 42B25, 43A85.
Keywords: Hardy–Littlewood maximal function, sphere.

For other results concerning the estimates of type (1-1), see for example [Stein and Strömberg 1983] in the setting of $\mathbb{R}^n$, [Li 2009; Li and Qian 2011] in the setting of H-type groups, [Li 2010] for Grushin operators, [Li and Lohoué 2012] for the case of real hyperbolic spaces and [Naor and Tao 2010]. There is also a bound of type

$$\lim_{n \longrightarrow +\infty} \|M_{\text{Cube}}\|_{L^1 \longrightarrow L^{1,\infty}} = +\infty$$

about the centered maximal function associated to cubes in $\mathbb{R}^n$; see [Aldaz 2011] or [Aubrun 2009] for details.

Let ω_{n-1} denote the area of the unit sphere of $\mathbb{R}^n$; i.e., $\omega_{n-1} = 2\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}$. Recall that, for $x \in S^{n-1}$, $0 < t \le 2$,

$$\begin{aligned} S(x,t) &= \{y \in S^{n-1} \subset \mathbb{R}^n;\ |x-y| \le t\}, \\ |S(x,t)| &= \mu_{S^{n-1}}(S(x,r)). \end{aligned}$$

There exist some mistakes in [Knopf 1987]. For example, near the end of the proof of Lemma 3, take

$$t = \sqrt{2\left(1 - n^{-\frac{1}{2}}\right)},$$

and we find that Lemma 3 is wrong. Knopf uses the estimate that

$$|S(x,t)| = \omega_{n-2} \int_0^{2\arcsin(t/2)} \sin^{n-2} u\, du \ge \omega_{n-2} \int_0^{2\arcsin(t/2)} \sin^{n-2} u \cos u\, du,$$

which gives the lower bound

$$|S(x,t)| \ge \frac{c\omega_{n-1}}{\sqrt{n}} \left[t^2\left(1 - \frac{t^2}{4}\right)\right]^{\frac{n-1}{2}} \quad \text{for all } 0 < t \le \sqrt{2},\ n \ge 2. \tag{1-2}$$

This estimate is not sharp enough to obtain the desired result. In order to make the proof in [Knopf 1987] effective, we need the sharper and sufficient lower bound:

Lemma 1.2. *There exists a constant* $c > 0$ *such that*, *for all* $n \ge 2$ *and* $0 < t \le \sqrt{2}$, *we have*

$$|S(x,t)| \ge c\omega_{n-1} \left[n\left(1 - t\sqrt{1 - \frac{t^2}{4}}\right) + t\sqrt{1 - \frac{t^2}{4}}\right]^{-\frac{1}{2}} \left[t^2\left(1 - \frac{t^2}{4}\right)\right]^{\frac{n-1}{2}}. \tag{1-3}$$

More specifically, using the bound (1-3) instead of (1-2) in the proof of Knopf's Lemma 1 yields an improved result to replace Lemma 1:

(1-4)
$$M_{S^{n-1}} f(x) \le c \max\Bigg\{ \sup_{\substack{n^{-\frac{1}{2}} \le t \\ \le \sqrt{2(1-n^{-1})}}} \frac{\sqrt{n\Big(1-t\sqrt{1-\frac{t^2}{4}}\Big)+t\sqrt{1-\frac{t^2}{4}}}}{t} u\Big(\Big(1-\frac{t^2}{2}\Big)x\Big), \quad n \sup_{0<t\le n^{-\frac{1}{2}}} u\Big(\Big(1-\frac{t}{\sqrt{n}}\Big)x\Big), \quad u\big(n^{-1}x\big)\Bigg\}.$$

Using (1-4) instead of the original Lemma 1 estimate at the end of the proof of Lemma 3 in [Knopf 1987] gives

(1-5)
$$M_{S^{n-1}} f(x) \le c \max\Bigg\{ \sup_{\substack{n^{-\frac{1}{2}} \le t \\ \le \sqrt{2(1-n^{-1})}}} \frac{\sqrt{n\Big(1-t\sqrt{1-\frac{t^2}{4}}\Big)+t\sqrt{1-\frac{t^2}{4}}}}{t} \Bigg(1+\sqrt{n\ln\Big(1-\frac{t^2}{2}\Big)^{-1}}\Bigg), \quad n \sup_{0<t\le n^{-\frac{1}{2}}} \Bigg(1+\sqrt{n\ln\Big(1-\frac{t}{\sqrt{n}}\Big)^{-1}}\Bigg), \quad 1+\sqrt{n\ln n}\Bigg\} M_T f(x).$$

It is trivial to check that the right side of (1-5) is at most $cnM_T f(x)$, and using this inequality the rest of the original proof works and gives the correct result.

2. Proof of Equation (1-3)

For $0 < t \le \sqrt{2}$, set $r = 2\arcsin(t/2)$; then

$$\begin{aligned} |S(x,t)| &= \int_0^r \omega_{n-2}(\sin s)^{n-2}\, ds = \omega_{n-2}\int_0^{\sin r} y^{n-2}\frac{dy}{\sqrt{1-y^2}} \\ &\ge \frac{\omega_{n-2}}{\sqrt{2}}\int_0^{\sin r} y^{n-2}\frac{dy}{\sqrt{1-y}} = \frac{\omega_{n-2}}{\sqrt{2}}(\sin r)^{n-1}\int_0^1 \frac{u^{n-2}}{\sqrt{1-u\sin r}}\, du. \end{aligned}$$

Observe that

$$\begin{aligned} \int_0^1 \frac{u^{n-2}}{\sqrt{1-u\sin r}}\, du &\ge \Big(1-\frac{1}{n}\Big)^{n-2}\int_{1-\frac{1}{n}}^1 \frac{du}{\sqrt{1-u\sin r}} \\ &= 2e^{(n-2)\ln(1-\frac{1}{n})}\frac{1}{n}\frac{1}{\sqrt{1-\sin r}+\sqrt{1-\big(1-\frac{1}{n}\big)\sin r}} \\ &> c\frac{1}{\sqrt{n}}\frac{1}{\sqrt{n(1-\sin r)+\sin r}}. \end{aligned}$$

Then Stirling’s formula implies (1-3). □

Remark. By (1-3), a simple computation then leads to

$$|S(x,t)| \geq c\omega_{n-1} \quad \text{whenever } \sqrt{2(1-n^{-1})} \leq t \leq 2 \text{ and } n \geq 2. \tag{2-1}$$

Acknowledgement

The author would like to thank Jian-Gang Ying for the help in English and the referee for helpful suggestions.

References

[Aldaz 2011] J. M. Aldaz, “The weak type $(1, 1)$ bounds for the maximal function associated to cubes grow to infinity with the dimension”, *Ann. of Math.* (2) **173**:2 (2011), 1013–1023. MR 2012b:42028 Zbl 1230.42025

[Aubrun 2009] G. Aubrun, “Maximal inequality for high-dimensional cubes”, *Confluentes Math.* **1**:2 (2009), 169–179. MR 2010j:42039 Zbl 1177.42015

[Knopf 1987] P. M. Knopf, “Maximal functions on the unit n-sphere”, *Pacific J. Math.* **129**:1 (1987), 77–84. MR 88i:42031 Zbl 0634.42018

[Li 2009] H.-Q. Li, “Fonctions maximales centrées de Hardy–Littlewood sur les groupes de Heisenberg”, *Studia Math.* **191**:1 (2009), 89–100. MR 2009k:42039 Zbl 1187.42018

[Li 2010] H.-Q. Li, “Fonctions maximales centrées de Hardy–Littlewood pour les opérateurs de Grushin”, preprint, 2010. arXiv 1207.3128

[Li and Lohoué 2012] H.-Q. Li and N. Lohoué, “Fonction maximale centrée de Hardy–Littlewood sur les espaces hyperboliques”, *Ark. Mat.* **50**:2 (2012), 359–378. MR 2961327 Zbl 06112914

[Li and Qian 2011] H.-Q. Li and B. Qian, “Centered Hardy–Littlewood maximal functions on Heisenberg type groups”, 2011. To appear in *Trans. Amer. Math. Soc.*

[Naor and Tao 2010] A. Naor and T. Tao, “Random martingales and localization of maximal inequalities”, *J. Funct. Anal.* **259**:3 (2010), 731–779. MR 2011k:42040 Zbl 1196.42018

[Stein and Strömberg 1983] E. M. Stein and J.-O. Strömberg, “Behavior of maximal functions in $\mathbf{R}^n$ for large n”, *Ark. Mat.* **21**:2 (1983), 259–269. MR 86a:42027 Zbl 0537.42018

Received April 3, 2012. Revised September 6, 2012.

HONG-QUAN LI
SCHOOL OF MATHEMATICAL SCIENCES
FUDAN UNIVERSITY
220 HANDAN ROAD
SHANGHAI 200433
CHINA

hongquan_li@fudan.edu.cn

Guidelines for Authors

Authors may submit manuscripts at msp.berkeley.edu/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text (“the curve looks like this:”). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as “Place Figure 1 here”. The same considerations apply to tables, which should be used sparingly.

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.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 263 No. 1 May 2013

0030-8730(201305)263:1;1-7