GEOMETRIC MAPPING THEORY OF THE HEISENBERG GROUP, SUB-RIEMANNIAN MANIFOLDS, AND HYPERBOLIC SPACES

ΒY

ANTON LUKYANENKO

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Mathematics in the Graduate College of the University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Jang-Mei Wu, Chair Professor Jeremy Tyson, Director of Research Assistant Professor Jayadev Athreya Associate Professor Nathan Dunfield Professor Aimo Hinkkanen

Abstract

We discuss the Heisenberg group Φ^n and its mappings from three perspectives. As a nilpotent Lie group, Φ^n can be viewed as a generalization of the real numbers, leading to new notions of base-*b* expansions and continued fractions. As a metric space, Φ^n serves as an infinitesimal model (metric tangent space) of some sub-Riemannian manifolds and allows one to study derivatives of mappings between such spaces. As a subgroup of the isometry group of complex hyperbolic space $\mathbb{H}^{n+1}_{\mathbb{C}}$, Φ^n becomes a large-scale model of a rank-one symmetric space and provides rigidity results in $\mathbb{H}^{n+1}_{\mathbb{C}}$.

After discussing homotheties and conformal mappings of Φ^n , we show the convergence of base-*b* and continued fraction expansions of points in Φ^n , and discuss their dynamical properties.

We then generalize to sub-Riemannian manifolds and their quasi-conformal and quasi-regular mappings. We show that sub-Riemannian lens spaces admit uniformly quasi-regular (UQR) self-mappings, and use Margulis–Mostow derivatives to construct for each UQR self-mapping of an equiregular sub-Riemannian manifold an invariant measurable conformal structure.

Turning next to hyperbolic spaces, we recall the relationship between quasiisometries of Gromov hyperbolic spaces and quasi-symmetries of their boundaries. We show that every quasi-symmetry of Φ^n lifts to a bi-Lipschitz mapping of $\mathbb{H}^{n+1}_{\mathbb{C}}$, providing a rigidity result for quasi-isometries of $\mathbb{H}^{n+1}_{\mathbb{C}}$. We conclude by showing that if Γ is a lattice in the isometry group of a non-compact rank one symmetric space (except $\mathbb{H}^1_{\mathbb{C}} = \mathbb{H}^2_{\mathbb{R}}$), then every quasi-isometric embedding of Γ into itself is, in fact, a quasi-isometry. To everyone whose efforts have made this possible.

Acknowledgments

I would like to first thank my adviser Jeremy Tyson for his encouragement, guidance, and instruction; and my collaborators Noel DeJarnette, Katrin Fässler, Piotr Hajłasz, Ilya Kapovich, Kirsi Peltonen, and Joseph Vandehey for exploring with me the various topics in this thesis, and for the many things I have learned from them.

I would also like to thank my committee members Jayadev Athreya, Nathan Dunfield, Aimo Hinkkanen, and Jang-Mei Wu for their interest in my work. Likewise, many teachers and mentors have contributed to my progress throughout my studies, including William Goldman, Ryan Hoban, Richard Schwartz, and James Schafer.

The faculty and staff of the mathematics department have been amazingly supportive, and I would like to especially thank Jayadev Athreya, Jonathan Manton, Tori Corkery and Wendy Harris, working with whom has been an invaluable experience.

I am additionally thankful for financial support from various sources, including the REGS program, GEAR network, and GAANN fellowship. In part, this work was supported by NSF grants DMS-0838434, DMS-1107452, DMS-0901620, and the hospitality of the Aalto University Department of Mathematics and Systems Analysis.

Last but not least, a big thanks to my family and friends for putting up with me, especially to my wife Cindy!

Table of Contents

Chapter 1 Introduction	1
1.1 Number theory and dynamics	2
1.2 Quasi-conformal analysis and branched covers	3
1.3 Hyperbolic geometry	4
	-
Chapter 2 Heisenberg group	5
2.1 Lattices and base- <i>b</i> expansions	5
2.2 Metrics on the Heisenberg group	9
2.3 Path metrics	12
2.4 Isometries and conformal mappings	13
2.5 Continued fractions	15
2.6 Conformal dynamics	17
2.7 Representing the Heisenberg group	19
2.8 Unitary model \ldots 2	21
2.9 Finite continued fractions and rational points	22
2.10 $PU(n,1)$ and the sub-Riemannian metric $\ldots \ldots \ldots$	24
2.11 Compactification of the Heisenberg group	24
Chapter 3 Heisenberg geometry distorted	26
3.1 Sub-Riemannian manifolds	26
3.2 Carpot groups	20
3.3 Cromov-Hausdorff tangent spaces	20
3.4 Quesi conformal mannings	20
3.5 Modifying quasi conformal mappings	20 20
3.6 Danau differentiability	24 24
3.7 Margulis-Mostow differentiability	25
3.7 Marguns-Mostow unicientiability	27
3.0 Dynamics of UOR mappings	20 20
2.10 Evictorea of UOP mappings)9 41
2.11 Aride: Dange manifolds	±1 49
5.11 Aside: Falisu manifolds	ŧЭ
Chapter 4 Hyperbolic spaces	15
4.1 The hyperbolic plane	15
4.2 Hyperbolic lattices	16
4.3 Real hyperbolic space and horospherical coordinates 4	48
4.4 Complex hyperbolic space	51
4.5 Gromov hyperbolicity and quasi-isometries	53
4.6 Quasi-isometries and quasi-symmetries	55
4.7 Quasi-isometric rigidity and quasi-conformal extension 5	57
4.8 Co-Hopficity of co-compact lattices	30
4.9 Co-Hopficity of non-uniform lattices	32
References	35

Chapter 1 Introduction

In a variety of situations, ranging from parallel parking to neurophysiology, one is interested in modeling *non-holonomic* phenomena that involve local but not global motion constraints. For example, in parallel parking one is interested in controlling the position and angle of the car, but directly controls only the angle and speed of the front wheels. This disconnect is fundamentally different from Euclidean geometry, where all directions of motion are immediately accessible.

The classical study of geometry is based on the properties of Euclidean space: manifolds, Riemannian metrics, and derivatives are all designed to harness the metric and algebraic structure of \mathbb{R}^n . For non-holonomic geometry, it is the Heisenberg group Φ^n and more generally the Carnot groups that provide the intuition and infinitesimal structure. The first Heisenberg group Φ^1 is defined (via geometric coordinates) as follows:

Definition 1.0.1. Φ^1 is the space $\mathbb{R}^3 = \{(x, y, t)\}$ with group structure

$$(x, y, t) * (x', y', t') = (x + x', y + y', t + t' + 2(xy' - xy')).$$



Figure 1.1: Left translates of Altgeld Hall in Φ^1 , with left multiplication acting by shears isometric with respect to the metric d_{sR} . The plane at the base of the building is spanned by the vector fields X and Y.

One gives Φ^1 the sub-Riemannian (or non-holonomic) metric d_{sR} by declar-

ing the left-invariant vector fields

$$X = \frac{\partial}{\partial x} - 2y\frac{\partial}{\partial t}, \qquad \qquad Y = \frac{\partial}{\partial y} + 2x\frac{\partial}{\partial t}$$

orthonormal, and computing distance along curves as in Riemannian geometry. It follows from the relation $[X, Y] = 4 \frac{\partial}{\partial t}$ that, in fact, any two points can be connected by a curve of finite length.

We now describe the structure and main ideas of the thesis. A recurring theme is the generalization of concepts based on Euclidean geometry to similar ones based on Heisenberg geometry.

Remark 1.0.2. The notation Φ^n for the Heisenberg group is non-standard. However, the standard symbol H is competed for by hyperbolic spaces, horospheres, Hausdorff dimension, and the horizontal distribution in sub-Riemannian spaces. On the other hand, the symbol Φ is evocative of the model $\mathbb{C} \times \mathbb{R}$ of the Heisenberg group.

1.1 Number theory and dynamics

One represents points in \mathbb{R} by exploiting the existence of a dilation $\delta_r(x) = rx$ and the integer lattice $\mathbb{Z} \subset \mathbb{R}$ with fundamental domain [0, 1). For a number $x \in$ [0, 1), the first base-10 digit of x is defined by the property $10x - a_1 \in [0, 1)$, and further digits are defined by iterating the digit-removing map $x \mapsto 10x - |10x|$.

An alternate common representation of $x \in [0, 1)$ is given by working with the condition $1/x - a_1 \in [0, 1)$ and the digit-removing map $x \mapsto 1/x - \lfloor 1/x \rfloor$. The corresponding digits a_i are known as *continued fraction* digits of x and satisfy the relation

$$x = \lim_{k \to \infty} \frac{1}{a_1 + \frac{1}{a_2 + \cdot \cdot \frac{1}{a_k}}}$$

Both representations are well-studied in number theory, with strong connections to dynamical systems and hyperbolic geometry. For example, one shows that the digit-removing maps are ergodic, and that (generically) the digit sequences of all real numbers are equally random.

In [31], Joseph Vandehey and I explored analogous notions for the Heisenberg group, showing that base-*b* expansions make sense in Φ^n for all *n*, and that continued fractions make sense for n = 1. Replacing the integer lattices \mathbb{Z} with the Heisenberg integer lattices $\Phi_{\mathbb{Z}}^n$ with a special fundamental domain $K_{S,b}$, we defined base-*b* expansions via the property $-a_1 * \delta_b(p) \in K_{S,b}$ and digit-removing mapping $p \mapsto -\lfloor \delta_b(p) \rfloor * \delta_b(p)$. Likewise, the continued fraction digits satisfy the property $-a_1 * \iota(p) \in K$, with digit-removing mapping $p \mapsto -[\iota(p)] * \iota(p)$, with the Korányi inversion ι given by:

$$\iota(z,t) = \left(\frac{-z}{|z|^2 + it}, \frac{-t}{|z|^4 + t^2}\right).$$

We show that in both cases, the digits can be recombined to produce the original point p. For the base-b case, we further established ergodicity of the digit-removing mapping, while for the continued fractions proving ergodicity appears to be quite complicated.

The results of [31] are summarized in Chapter 2, which also provides a description of the isometries and conformal mappings of Φ^n .

1.2 Quasi-conformal analysis and branched covers

In complex analysis of one variable, one is concerned with conformal mappings, i.e. angle-preserving smooth mappings. In higher dimensions, conformal mappings become rigid, and one generalizes to *quasi-conformal* mappings, defined as homeomorphisms that send infinitesimal balls to ellipsoids of bounded eccentricity. A classical result based on the theorems of Rademacher and Stepanov then states that quasi-conformal mappings in \mathbb{R}^n are *almost everywhere* differentiable. Because Riemannian manifolds are locally diffeomorphic to \mathbb{R}^n , the same differentiability result holds for quasi-conformal mappings between Riemannian manifolds.

In part due to connections to group theory (see §1.3), one is interested in quasi-conformal groups, i.e. groups of quasi-conformal mappings whose dilatation (the maximal stretching of infinitesimal balls) is bounded by the same constant for all group elements. In [46], Tukia showed that every quasi-conformal group Γ leaves invariant a measurable conformal structure. That is, while it need not be conjugate to a group of conformal transformations, there is a notion of angle based on which Γ seems conformal.

More recently, interest has appeared in the dynamics of analogous quasiregular mappings, which are branched covers with a dilatation bound, such as the composition of the maps $f_1(z) = z^2$ and $f_2(x, y) = (2x, y)$. In particular, Iwaniec–Martin showed in [21] that every abelian uniformly quasi-regular (UQR) semigroup admits an invariant measurable conformal structure.

Katrin Fässler, Kirsi Peltonen and I explored the dynamics of quasi-regular mappings of *sub-Riemannian* manifolds in [13]. Specifically, we showed that every lens space with its natural sub-Riemannian metric admits a non-trivial UQR mapping (that is, a non-homeomorphic mapping that generates a UQR semigroup), and that more generally any UQR mapping admits an invariant measurable conformal structure. The existence result is based on the conformal trap method of [13] and the quasi-conformal flow techniques of Libermann and Korányi–Reimann, while the invariant structure is found by generalizing Tukia's result via Margulis–Mostow derivatives and Gromov–Hausdorff tangent spaces.

An exposition of the results of [13] is provided in Chapter 3.

1.3 Hyperbolic geometry

The hyperbolic plane $\mathbb{H}^1_{\mathbb{C}}$ is constructed by giving the upper half-space $\{(x, y) : y > 0\}$ the Riemannian metric with line element

$$ds^2 = \frac{dx^2 + dy^2}{y^2}$$

The hyperbolic plane has an extensive geometric theory, with a key role played by the boundary $\partial \mathbb{H}^1_{\mathbb{C}} = \{(x, y) : y = 0\}$. Despite the extrinsic appearance of $\partial \mathbb{H}^1_{\mathbb{C}}$ in a specific model of $\mathbb{H}^1_{\mathbb{C}}$, it can be defined intrinsically using geodesics of $\mathbb{H}^1_{\mathbb{C}}$, and many interesting mappings of $\mathbb{H}^1_{\mathbb{C}}$ (specifically, the *quasi-isometries*, whose distortion is controlled by a linear function) extend to continuous mappings of $\partial \mathbb{H}^1_{\mathbb{C}}$.

Higher-dimensional analogues of $\mathbb{H}^1_{\mathbb{C}}$ come in four varieties: the real hyperbolic spaces $\mathbb{H}^n_{\mathbb{R}}$ (with $\mathbb{H}^2_{\mathbb{R}}$ isometric to $\mathbb{H}^1_{\mathbb{C}}$), the complex hyperbolic spaces $\mathbb{H}^n_{\mathbb{C}}$, the quaternionic hyperbolic spaces, and the octnionic hyperbolic plane. Together, these are known as the *non-compact rank-one symmetric spaces*. We focus on $\mathbb{H}^n_{\mathbb{C}}$, for $n \geq 2$.

As with $\mathbb{H}^1_{\mathbb{C}}$, the higher-dimensional spaces $\mathbb{H}^n_{\mathbb{C}}$ can be described using upper half-space models $\Phi^{n-1} \times \mathbb{R}^+$, with the Heisenberg group $\Phi^{n-1} \times \{0\}$ now playing the role of the boundary. In particular, if $f: \Phi^{n-1} \to \Phi^{n-1}$ is a homothety with dilation factor r, then F(p, s) = (f(p), rs) is an isometry of $\mathbb{H}^n_{\mathbb{C}}$ in the horospherical model. Through the theory of Gromov hyperbolic spaces, one sees that quasi-isometries of $\mathbb{H}^n_{\mathbb{C}}$ induce quasi-conformal mappings of Φ^{n-1} and furthermore every quasi-conformal mapping of Φ^{n-1} arises in this fashion. We follow Tukia–Väisälä [49] to show in [30] that, in fact, every quasi-conformal mapping of Φ^{n-1} is the boundary of a bi-Lipschitz mapping (i.e. one with *multiplicative* distortion of distances).

Switching directions, we then focus on the lattices Γ in the isometry group of $\mathbb{H}^n_{\mathbb{C}}$. It is a classical result that if $\Gamma \setminus \mathbb{H}^n_{\mathbb{C}}$ is compact, then any quasi-isometric embedding $f : \Gamma \hookrightarrow \Gamma$ is, in fact, a quasi-isometry. Ilya Kapovich and I proved in [22] the same result for non-uniform lattices in rank-one semi-simple Lie groups.

Chapter 4 provides a description of the non-compact rank one symmetric spaces, their lattices and mapping theory, as well as the results of [30] and [22].

Chapter 2

Heisenberg group

The most common *geometric* model of the Heisenberg group Φ^n is given by identifying Φ^n with $\mathbb{C}^n \times \mathbb{R}$, with the group law h * h' given by

$$(z,t) * (z',t') = (z+z',t+t'+2\mathrm{Im}\langle z,z'\rangle),$$

where $\langle z, z' \rangle$ is the standard Hermitian inner product in \mathbb{C}^n . Alternately, one gives \mathbb{C}^n real coordinates and writes

$$(x_i, y_i, t) * (x'_i, y'_i, t') = \left(x_i + x'_i, y_i + y'_i, t + t' + 2\sum_{i=1}^{n} (x_i y'_i - x'_i y_i)\right)$$

The group's center $\{0\} \times \mathbb{R}$ of Φ^n acts by a Euclidean translation, while left translation by other elements causes a shearing of the space (cf. Figure 1.1):

$$(0,1) * (z,t) = (z,t+1)$$
$$(1,0) * (z,t) = (z+1,t-2y)$$

Thus, while we will need a replacement for the Euclidean distance, the usual Lebesgue measure provides a natural notion of volume.

In this chapter, we begin by investigating the essential group-theoretic and number-theoretic properties of Φ^n . This will lead us to a description of several metrics on Φ^n and a study of the conformal, and eventually quasi-conformal, mappings of the space. The number-theoretic results and dynamical-systems considerations in this chapter are based on joint work with Joseph Vandehey [31].

2.1 Lattices and base-b expansions

Consider the subgroup $\Phi_{\mathbb{Z}}^n$ of points in Φ^n all of whose coordinates in the geometric model are integers.

Lemma 2.1.1. The unit cube $K_C = [0, 1]^{2n+1}$ is a fundamental region for $\Phi_{\mathbb{Z}}^n$. That is, the translates $\Phi_{\mathbb{Z}}^n * K_C$ of the cube tile all of Φ^n while overlapping only along the faces.

Proof. Consider a point in Φ^n with coordinates (z, t). Pick $z' \in \mathbb{Z}[i]^n$ such that

 $-z' + z \in [0,1]^{2n}$ and therefore $(-z',0) * (z,t) \in [0,1]^{2n} \times \mathbb{R}$. A further choice of $t' \in \mathbb{Z}$ ensures that $(-z',-t') * (z,t) \in K_C$. We thus have that $(z',t') \in \Phi^n_{\mathbb{Z}}$ and $p \in (z',t') * K_C$. Likewise, only points on the boundary of K_C are related by elements of $\Phi^n_{\mathbb{Z}}$.

We thus have that $\Phi_{\mathbb{Z}}^n$ is a *uniform lattice* in Φ^n , that is, a discrete subgroup with a compact fundamental domain. The quotient $\Phi_{\mathbb{Z}}^n \setminus \Phi^n$ is an example of a *nilmanifold*. While the quotient may resemble a torus, its fundamental group is not abelian. Rather, we have (in the standard group-theoretic multiplicative notation):

 $\pi_1(\Phi_{\mathbb{Z}}^n \setminus \Phi^n) = \Phi_{\mathbb{Z}}^n \cong \langle a_i, b_i, c : [a_i, b_i] = c^4 \text{ for each } i; \text{ others commute} \rangle.$

Remark 2.1.2. The generators are given by translates by x_i , y_i , and t. The relations are easily verified, and are shown to be sufficient by putting every element of $\Phi_{\mathbb{Z}}^n$ in a normal form. Note also the factor of 4 in the presentation, caused by our convention on the group law; adjusting the group law to fix this causes the 4 to reappear elsewhere.

We start by finding a way to represent elements of Φ^n in an inherent fashion, without referring extensively to the geometric representation. We first mimic the base-*b* expansions of real numbers, and will develop a continued-fraction representation below. We need the following two tools:

Definition 2.1.3. Let K be a fundamental region for a group Γ acting on some space X. We have an associated "nearest integer" map $[\cdot] : X \to \Gamma$ defined by the property that $p \in [p] * K$. Like the usual nearest integer map, $[\cdot]$ is uniquely defined on the interiors of the tiles $\Gamma \cdot K$.

Definition 2.1.4. For each r > 0, define:

$$\delta_r(p) = \delta_r(z, t) = (rz, r^2 t).$$

The map $\delta_r : \Phi^n \to \Phi^n$ is a group isomorphism. For $r \in \mathbb{N}$, $\delta_r : \Phi_{\mathbb{Z}}^n \to \Phi_{\mathbb{Z}}^n$ is an injective homomorphism. Based on these properties, we will think of δ_r as a dilation map (we will later introduce an appropriate metric for which this is true).

Definition 2.1.5 (base-*b* expansion). Fix a positive integer $b \ge 2$ and suppose that *K* is a fundamental domain for $\Phi_{\mathbb{Z}}^n$ satisfying:

- 1. $0 \in K$,
- 2. $\delta_b K = \bigcup_{d \in D} d * K$ for some finite *digit set* $D \subset \Phi^n_{\mathbb{Z}}$.

Let $[\cdot]$ be the nearest-integer map associated to K. Define $T: K \to K$ be the map $T: p \mapsto [\delta_b p]^{-1} * \delta_b p$, and $a_i := [T^{i-1}p]$. We refer to $\{a_i\}$ as the base-b digits of p. We write $p = 0.a_1 a_2 a_3 \ldots$, for now without justification.

Remark 2.1.6. We are paralleling the familiar base-10 expansions of real numbers. In that case, we have K = [0,1], $x \mapsto [x]$ the floor function, and T(x) = -[10x] + 10x. The digits a_i are then exactly the base-10 digits of the number $x \in [0,1]$. In base-10 notation, the map T simply removes the first digit of x.

The cube K_C does not satisfy the assumptions of Definition 2.1.5. However, the following theorem of Strichartz states that the desired fundamental domain does exist (see Figure 2.1):

Theorem 2.1.7 (Strichartz [44]). For each b > 0, there exists a fundamental domain $K_S = K_{S,b}$ for $\Phi_{\mathbb{Z}}^n$ satisfying the conditions of Definition 2.1.5.



Figure 2.1: The self-similar Strichartz tile $K_{S,2}$ from Theorem 2.1.7.

Note that if $p = 0.a_1a_2a_3\cdots$, then $\delta_{b^{-1}}p = 0.0a_1a_2a_3\cdots$. This property allows us to define the expansion for points outside of K. Namely,

Definition 2.1.8. Fix b, K as above. Suppose $p \in \Phi^n$ and k is the smallest integer so that $\delta_b^{-k}p \in K$, and $\delta_b^{-k}p = 0.a_1a_2\cdots$. Then we write $p = a_1a_2\ldots a_k.a_{k+1}a_{k+2}\cdots$.

Abusing notation, as one does with decimal expansions in \mathbb{R} , we define:

Definition 2.1.9. Let $N \in \mathbb{N}$ and $\{a_i\}_{i=-N}^{\infty}$ elements of $\Phi^n(\mathbb{Z})$. Define:

$$a_{-N}a_{-N+1}\cdots a_0.a_1a_2\cdots := \lim_{n\to\infty}\delta_b^N a_{-N}*\delta_b^{N-1}a_{-N+1}*\cdots*\delta_b^{N-n}a_{-N+n}$$

should this limit exist. Note that this is completely analogous to the meaning of base-*b* numbers in \mathbb{R} .

Theorem 2.1.10 (Lukyanenko–Vandehey [31]). Let $N \in \mathbb{N}$ and $\{a_i\}_{i=-N}^{\infty}$ a bounded sequence of elements of $\Phi^n(\mathbb{Z})$. Then the limit $a_{-N}a_{-N+1}\cdots a_0.a_1a_2\cdots$ exists. Furthermore, if $\{a_i\}$ are the base-b digits of a point p, then we indeed

have that

$$a_{-N}a_{-N+1}\cdots a_0.a_1a_2\cdots = p$$

in the sense of Definition 2.1.9.

To prove Theorem 2.1.10, it is convenient to provide Φ^n with a left-invariant metric for which δ_r is a dilation by factor r.

Definition 2.1.11. The gauge (or Cygan or Korányi) metric d_g on Φ^n is defined, using geometric coordinates, as:

$$||(z,t)||^4 = |z|^4 + t^2$$
 $d_g(p,q) = ||p^{-1} * q|$

It is straightforward to show that d_g is a metric, is left-invariant, and induces the expected Euclidean topology on Φ^n .

Proof of Theorem 2.1.10. By definition,

$$a_{-N}a_{-N+1}\cdots a_0.a_1a_2\cdots = \lim_{n \to \infty} \delta_b^N a_{-N} * \delta_b^{N-1}a_{-N+1} * \cdots * \delta_b^{-N+n}a_{-N+n},$$

if it exists. Now, the sequence of partial sums

$$\{\delta_b^N a_{-N} * \delta_b^{N-1} a_{-N+1} * \dots * \delta_b^{N-n} a_{-N+n}\}_{n=0}^{\infty}$$

is Cauchy because δ_b^{-1} is distance-decreasing and the digits are bounded, hence convergent. Indeed, by the triangle inequality we have for each n < m:

$$\begin{split} d_g(\delta_b^N a_{-N} * \delta_b^{N-1} a_{-N+1} * \cdots * \delta_b^{N-n} a_{-N+n}, \\ \delta_b^N a_{-N} * \delta_b^{N-1} a_{-N+1} * \cdots * \delta_b^{N-m} a_{-N+m}) \\ &= \left\| \delta_b^{N-n-1} a_{-N+n+1} * \cdots * \delta_b^{N-m} a_{-N+m} \right\| \\ &\leq \left\| \delta_b^{N-n-1} a_{-N+n+1} \right\| + \cdots + \left\| \delta_b^{N-m} a_{-N+m} \right\| \\ &= b^{N-n-1} \left\| a_{-N+n+1} \right\| + \cdots + b^{N-m} \left\| a_{-N+m} \right\| \end{split}$$

We assumed that the $\{a_i\}$ are bounded. In particular, their norm is bounded by some $A \ge 0$, and the above sum is bounded above by $Ab^{N-n-1}\frac{1}{1-1/b}$.

The second half of the theorem is given by an analogous estimate. $\hfill \Box$

We finish the section on base-b expansions with an analogue of a classic number theory result.

Definition 2.1.12. Let D_b be the set of possible base-*b* digits in Φ^n , consisting of b^{2n+2} elements. An infinite sequence in the elements of D_b is *normal* if each digit appears equally often, and furthermore, for each m > 1, all strings of length *m* appear equally often in the sequence.

A point $p \in \Phi^n$ is normal in base b if its expansion is normal with respect to D_b . The point is normal if it is normal with respect to any base b. The next result follows from immediately from standard ergodic theory, namely Theorem 2.6.3:

Theorem 2.1.13. Let $T: K_S \to K_S$ be given by $T: p \mapsto [\delta_b p]^{-1} \delta_b p$. Then T is ergodic with respect to Lebesgue measure on K_S .

Corollary 2.1.14. Almost every point of Φ^n is normal.

Proof. Fix $b \ge 2$ and $m \ge 1$. The self-similarity of K_S implies that the points of K_S that start with some sequence $a_1 \cdots a_m$ are represented by a sub-tile of K_S with the same volume as a sub-tile associated to any other starting sequence of the same length. The Birkhoff Ergodic Theorem then guarantees that the orbit of a generic point in K_S visits each sub-tile equally often. That is, the subset $N_{b,m}$ of K_S consisting of (b,m)-normal points has full measure. Since there are countably many choices of b and m, we conclude that the set of normal points $N = \bigcap_{b,m} N_{b,m}$ is of full measure.

Example 2.1.15. Consider points in Φ^1 expressed in base-2. There are 16 possible digits: (x, y, t) with x = 0, 1, y = 0, 1, t = 0, 1, 2, 3. We may number these in base 16 as $0, 1, 2, 3, \ldots, 9, A, B, C, D, E, F$. It is well-known that the *Champernowne sequence* 0123456789*ABCDEF*101112131415161718191*A*... is normal. Thus, the corresponding point in Φ^1 is normal base-2. Similarly, a base-*b* normal point can be constructed in any Φ^n for each base *b*.

Question 2.1.16. Construct a point that is normal (with respect to all b).

Question 2.1.17. Suppose p is a normal point in Φ^n . Are the geometric coordinates of p normal (as real numbers)? Conversely, is a point with normal coordinates normal?

Question 2.1.18. Show that a point in Φ^n has an eventually periodic expansion if and only if its coordinates are rational. What is the relationship between the period of the expansion and the coordinates?

2.2 Metrics on the Heisenberg group

Definition 2.1.11 provides a convenient metric on Φ^n , but is it a natural metric to study? Indeed, it seems far more natural to use a Riemannian metric:

Definition 2.2.1. Consider Φ^n in its geometric model, and consider the standard inner product at the origin. Extend it via left multiplication to a leftinvariant metric tensor g on all of Φ^n . The standard Riemannian metric d_{Riem} on Φ^n is the associated path metric.

On the small scale, the Riemannian Heisenberg group is essentially Euclidean. We will now show that the large-scale geometry of (Φ^n, d_{Riem}) is closer to that of the gauge metric. This will allow us to focus on the metric d_g , which

is both easier to compute and equipped with dilations δ_r , which already arize from the group structure of Φ^n . To facilitate the transition, we define another metric on Φ^n :

Definition 2.2.2. Let α be the left-invariant differential one-form on Φ^n defined by the property (in geometric coordinates) that $\alpha|_0 = dt$. Set $H\Phi^n = \text{Ker } \alpha$, a *horizontal* hyperplane bundle. Lastly, consider the restriction $g_{sR} = g|_{H\Phi^n}$ of the Riemannian inner product g to the horizontal bundle.

A path in Φ^n is said to be *horizontal* or *admissible* if its velocity is almost everywhere in $H\Phi^n$, so that g_{sR} can be used to calculate the length of the path. The *sub-Riemannian* (or *Carnot-Carathéodory*) distance d_{sR} between two points of Φ^n is the infimal length of admissible paths between the two points.

The fact that (Φ^n, d_{sR}) is a metric space and homeomorphic to \mathbb{R}^{2n+1} is established by the following theorem initially studied in the context of PDEs (note that an analogous theorem is immediate for the gauge metric):

Theorem 2.2.3 (Ball–Box Theorem [5]). Let $B(0,r) = \{p \in \Phi^n : d_{sR}(0,p) \le r\}$ and $Box = [-1,1]^{2n+1} \subset \Phi^n$. Then there exist $r_1, r_2 > 0$ such that

$$B(0, r_1) \subset Box \subset B(0, r_2).$$



Figure 2.2: The Ball-Box Theorem 2.2.3 with $r_1 = 1, r_2 = 3$.

The following proposition relates the three metrics (see below for the terminology):

Proposition 2.2.4. (Φ^n, d_g) is equivalent to (Φ^n, d_{sR}) , which is in turn isometric to the asymptotic cone of (Φ^n, d_{Riem}) .

Definition 2.2.5. Recall that a function $f : X \to Y$ between two metric spaces is an (L, C)-quasi-isometric embedding if one has

$$-C + L^{-1}|x - x'| \le |fx - fx'| \le L|x - x'| + C$$

for all points $x, x' \in X$. It is a quasi-isometry if furthermore the *C*-neighborhood of f(X) is all of *Y*. If we have C = 0, then *f* is a bi-Lipschitz embedding, or a bi-Lipschitz homeomorphism, respectively. We say that two metrics on a fixed space are equivalent if the identity map between them is bi-Lipschitz. The first part of Proposition 2.2.4 is therefore formalized as:

Lemma 2.2.6. The identity map $id : (\Phi^n, d_g) \to (\Phi^n, d_{sR})$ is a bi-Lipschitz equivalence.

Proof. Let $p, q \in \Phi^n$. We would like to compare $d_g(p,q)$ to $d_{sR}(p,q)$. Because both metrics are left-invariant, we may assume that q = 0. Furthermore, both metrics rescale in the same way via the dilation δ_r , so we may assume $d_{sR}(p,0) =$ 1. It remains to provide upper and lower bounds for the gauge norm of points pin the sub-Riemannian unit sphere. These follow immediately from the Ball-Box Theorem 2.2.3 by comparing the box with a gauge sphere.

Before defining asymptotic cones (for the statement of Proposition 2.2.4), we first define the notion of asymptotic isometries.

Definition 2.2.7. Let (X_{∞}, d_{∞}) a be a metric space, and (X_i, d_i) a family of compact metric spaces with uniform diameter. A sequence of maps $f_i : X_i \to X_{\infty}$ is an *asymptotically isometric (embedding)* if each f_i is an (L_i, C_i) -quasiisometric (embedding) with

$$\lim_{i \to \infty} L_i = 1, \qquad \qquad \lim_{i \to \infty} C_i = 0.$$

In the spirit of many analytic results requiring uniform convergence on compacts, we extend the Definition 2.2.7 as follows:

Definition 2.2.8. Let (X_{∞}, d_{∞}) be a metric space, and (X_i, d_i) a family of locally compact metric spaces. A sequence of maps $f_i : X_i \to X_{\infty}$ is an *asymptotically isometric (embedding)* if, for any r > 0 and collection of points $x_i \in X_i$, the restriction of each f_i to the ball $B(x_i, r)$ is an (L_i, C_i) -quasi-isometric (embedding) with

$$\lim_{i \to \infty} L_i = 1, \qquad \qquad \lim_{i \to \infty} C_i = 0.$$

Remark 2.2.9. For Definition 2.2.8, one usually works with *pointed* metric spaces. However, we are interested in homogeneous metric spaces, so our definition is equivalent in this context.

We can now say precisely what it means for one space to be a large-scale model of another.

Definition 2.2.10. Let (X, d) and (X_{∞}, d_{∞}) be two locally compact metric spaces. Fix a sequence $r_i > 0$ going to infinity, and set $(X_i, d_i) := (X, r_i^{-1}d)$, so that the identity map from X to X_i is a similarity with dilation r_i . One says that X_{∞} is the *asymptotic cone* of X if there exists an asymptotically isometric sequence of maps $f_i : X_i \to X$.

A priori, the asymptotic cone is guaranteed neither to exist nor to be unique. Indeed, some spaces have multiple asymptotic cones, depending on the choice of rescaling sequence. In our case, this will not be the case: **Proposition 2.2.11** (Pansu [37]). The space (Φ^n, d_{sR}) is the unique (up to isometry) asymptotic cone for (Φ^n, d_{Riem}) .

The proof of Proposition 2.2.11 uses Riemannian "penalty" metrics on Φ^n :

Definition 2.2.12. Let s > 0. The left-invariant Riemannian *penalty metric* d_s with parameter s is the defined by a left-invariant metric tensor g_s characterized as follows. At the origin of Φ^n in geometric coordinates $\mathbb{C}^n \times \mathbb{R}$, g_s agrees with g_{sR} and g_{Riem} along the complex direction; the real direction is declared orthogonal to the complex direction, with the vector (0, 1) assigned length s.

Remark 2.2.13. In terms of tensors, we have $g_1 = g_{Riem}$ and $\lim_{r\to\infty} g_r = g_{sR}$.

Recall now that given a metric space (X, d) and r > 0, the space (X, rd) consists of the same points as (X, d), but with all distances rescaled by factor r.

Lemma 2.2.14. For each r > 0, the rescaled metric space (Φ^n, rd_{Riem}) is isometric to (Φ^n, d_r) .

Proof. Note that (Φ^n, rd_{Riem}) is still Riemannian, defined by the metric tensor rg. Since δ_r is a linear map, it is its own derivative, and have that $(\delta_r^{-1})_*(rg) = g_r$. Thus, $\delta_r^{-1}: (\Phi^n, rd_{Riem}) \to (\Phi^n, d_r)$ is an isometry.

To complete Proposition 2.2.11, we invoke the following lemma:

Lemma 2.2.15 ([8]). The geodesics of (Φ^n, d_{sR}) are limits of geodesics in (Φ^n, d_r) . In particular, one has that for any two points $p, q \in \Phi^n$,

$$d_{sR}(p,q) = \lim_{r \to \infty} d_r(p,q),$$

uniformly on compacts.

Remark 2.2.16. Solving the geodesic equation for the penalty metrics allows us to draw geodesics and spheres in (Φ^n, d_{sR}) , as in Figure 2.2.

2.3 Path metrics

We now explore a closer connection between the metrics d_{sR} and d_g . We start by recalling some standard metric-space definitions.

Definition 2.3.1. Let (X, d) be a metric space. It is said to be *geodesic* if for any $x, x' \in X$ there exists an isometric embedding of the interval [0, d(x, x')] that starts at x and ends in x'.

Definition 2.3.2. Let (X, d) be a metric space (geodesic or not). Let γ : $[0, a] \to X$ be a path. The length of γ is given by

$$\ell(\gamma) := \sup \sum_{i=1}^{N} d(\gamma(a_{i-1}), \gamma(a_i)),$$

where the supremum is taken over all partitions $\{a_i\}$ of the interval [0, a]. A path is *rectifiable* if it has finite length. A metric spaces is *rectifiably connected* if every pair of points is joined by a rectifiable curve.

Definition 2.3.3. Let (X, d) be a rectifiably connected metric space. The *path* metric associated to d is given by

$$d_{path}(x, x') := \inf \ell(\gamma),$$

where the infimum is taken over all rectifiable paths γ joining x and x'.

The metric axioms for the associated path metric follow immediately from those for the original metric. Furthermore, a metric space (X, d) is geodesic if and only if the path metric associated to d is d itself.

Lemma 2.3.4 ([25]). Let $p \in \Phi^n$. We then have, for q approaching p along horizontal curves:

$$\lim_{q \to p} \frac{d_g(p,q)}{d_{sR}(p,q)} = 1$$

Combining Lemma 2.3.4 with the rotational symmetry of Φ^n and the fact that projection onto the x_1 -axis is 1-Lipschitz in both metrics gives:

Corollary 2.3.5. The sub-Riemannian metric d_{sR} is the path metric associated to the gauge metric d_q .

2.4 Isometries and conformal mappings

The classification of the isometries of Φ^n seems to be due to Ursula Hamenstädt in 1990 [17]. The key tool is a generalization of the Myers–Steenrod Theorem [36], which states that isometries of Riemannian manifolds are smooth. Hamenstädt proved the corresponding fact for manifolds whose geodesics satisfy an analogue of the geodesic equation. A generalization to all regular sub-Riemannian manifolds (including ones with *abnormal* geodesics) was only recently provided by Capogna–LeDonne [9].

Proposition 2.4.1 (Hamenstädt [17]). Let $f : \Phi^n \to \Phi^n$ be an isometry with respect to d_g , d_{sR} , or d_{Riem} . Then f is the composition of the following maps:

- 1. Left translations $\ell_p : q \mapsto p * q$, for some p,
- 2. Linear transformations of the form $A \otimes 1$, for some $A \in U(n)$,
- 3. The map $(z,t) \mapsto (\overline{z},-t)$.

where the splitting refers to geometric coordinates on Φ^n and U(n) is the group of unitary matrices.

Remark 2.4.2. Hamenstädt's result states that an origin-preserving isometry of a Carnot group (e.g. Φ^n) with its sub-Riemannian metric must be a Lie group isomorphism preserving the horizontal distribution. Such isomorphisms are generated by the maps $A \otimes \det(A)$ with A a symplectic matrix. For d_{sR} , Proposition 2.4.1 follows by including the left translations and identifying the distance-preserving Lie group isomorphisms of Φ^n . For the gauge metric d_g , the result follows from Proposition 2.3.5.

Corollary 2.4.3. The homotheties of Φ^n are generated by the mappings in Proposition 2.4.1 and the maps δ_r .

We would now like to identify the *conformal* maps of Φ^n , as their study in Euclidean space leads to topics in both analysis and geometry.

Definition 2.4.4. Let $f : X \to Y$ be a homeomorphism between metric spaces. We say that f is *conformal* if at every point $x \in X$ the following limit exists:

$$\lim_{r \to 0} \sup \left\{ \frac{|fx - fx'|}{|x - x'|} : |x - x'| \le r \right\}$$

where $|\cdot - \cdot|$ denotes the distance in the appropriate space.

Example 2.4.5. Suppose X and Y are Riemannian manifolds with corresponding metric tensors g_X and g_Y , and $f: X \to Y$ is a homeomorphism. Then f is conformal if and only if it is *smooth* and one has $f^*g_Y = \lambda g_X$ for a smoothly-varying non-vanishing λ on X (see [14] and [28]).

Recall that for domains in the plane, one has an extensive theory of conformal mappings. However, in larger spaces the theory is more constrained:

Theorem 2.4.6 (Liouville theorem in \mathbb{R}^n [15]). Let X, Y be domains in \mathbb{R}^n , for $n \geq 3$, or the full plane \mathbb{R}^2 . Any conformal map $f : X \to Y$ is the restriction of a Möbius transformation of $\mathbb{R}^n \cup \{\infty\}$.

Definition 2.4.7. The *Möbius transformations* of $\mathbb{R}^n \cup \{\infty\}$ are generated by homotheties of \mathbb{R}^n and the map $\iota(x) = -x/||x||^2$, which exchanges 0 and ∞ .

Returning to the Heisenberg group, note that we have a notion of a conformal map with respect to both the sub-Riemannian and gauge metrics. The two classes of mappings coincide (which follows from a more general characterization of *quasi*-conformal mappings, $\S3.4$):

Lemma 2.4.8 (Korányi–Reimann [25]). Let U, V be domains in Φ^n . A mapping $f: U \to V$ is conformal with respect to d_g if and only if it is conformal with respect to d_{sR} .

Analogously to classical Möbius transformations of \mathbb{R}^n , we can define Möbius transformations for Φ^n (see [25]):

Definition 2.4.9. The *Möbius transformations* of $\mathbb{R}^n \cup \{\infty\}$ are generated by:

- 1. Left translations $\ell_p: q \mapsto p * q$, for $p \in \Phi^n$,
- 2. Rotations $A \otimes 1$, for $A \in U(n)$,
- 3. Dilations δ_r , for r > 0,
- 4. The reflection $(z,t) \mapsto (\overline{z},-t)$,
- 5. The Korányi inversion

$$\iota(z,t) = \left(\frac{-z}{|z|^2 + it}, \frac{-t}{|z|^4 + t^2}\right)$$
(2.4.1)

Remark 2.4.10. The Korányi inversion has a simpler expression if we extend some previous notation. Recall that we have $||(z,t)||^4 = |z|^4 + t^2$. Write $||(z,t)||^2_{\mathbb{C}} := |z|^2 + it$. Furthermore, allow δ_r to accept $r \in \mathbb{C}$ by writing $\delta_r(z,t) = (rz, |r|^2 t)$. We can then write

$$\iota(z,t) = -\delta_{\|(z,t)\|_{\mathbb{C}}^{2}}^{-1}(z,t).$$

While this is not standard notation, it can be useful. We will link the Korányi inversion to an antipodal map on the sphere in §2.7.

Theorem 2.4.11 (Liouville-type theorem in Φ^n [24, 7]). Let $f : U \to V$ be a conformal mapping between domains in Φ^n (for $n \ge 1$), with respect to the sub-Riemannian or gauge metric. Then f is the restriction of a Möbius transformation.

Remark 2.4.12. Note that the Korányi inversion is *not* conformal with respect to the Riemannian metric on Φ^n , even though it is conformal from the sub-Riemannian and gauge-metric perspective.

2.5 Continued fractions

In §2.1 we discussed an intrinsic notion of base-*b* expansions on Φ^n . The classification of conformal mappings on Φ^n now provides us with the tools to define a continued fraction expansion for points in Φ^1 .

We start with a critical observation concerning the Korányi inversion and the gauge metric d_g (inversion in \mathbb{R}^n satisfies the same relation with respect to the Euclidean metric).

Lemma 2.5.1. The Korányi inversion ι satisfies, for all $p, q \in \Phi^n$:

$$d_g(\iota p, \iota q) = \frac{d_g(p, q)}{\|p\| \, \|q\|}.$$

In particular, ι preserves the gauge unit sphere.

Recall that a classical continued fraction represents a number $x \in [0, 1]$ as a limit of fractions:

$$x = \lim_{k \to \infty} \frac{1}{a_1 + \frac{1}{a_2 + \cdot \cdot \cdot \frac{1}{a_k}}}$$

Interpreting addition as left translation and inversion as ι , we define the analogous concept in Φ^n :

Definition 2.5.2 (Heisenberg continued fraction). Let $\{\gamma_i\} \subset \Phi_{\mathbb{Z}}^n$ be a sequence of integer Heisenberg elements. The associated continued fraction is

$$\mathbb{K}\{\gamma_i\} := \lim_{k \to \infty} \iota \gamma_1 \iota \gamma_2 \cdots \iota \gamma_k,$$

if this limit exists. Note that for reading convenience, we have dropped parentheses and multiplication sign.

Theorem 2.5.3 (Lukyanenko–Vandehey [31]). Suppose $\{\gamma_i\}$ is a sequence of elements of $\Phi_{\mathbb{Z}}^n$ with $\|\gamma_i\| \geq 2$ for each *i*. Then $\mathbb{K}\{\gamma_i\}$ exists. Furthermore, the bound on the elements depends only on *n*.

Proof. It is hard to track the sequence of partial fractions $\gamma_1 \iota \gamma_2 \cdots \iota \gamma_k$. Instead, we consider the region where each partial fraction might be and show that this limits to a single point.

Let *B* be the closed unit ball in the gauge metric (of diameter 2). We have $0 \in B$, and if $||\gamma_1||$ is sufficiently large (say, bigger than 2), all points in $\gamma_1 B$ are at least some fixed distance *C* from the origin.

By Lemma 2.5.1, $\iota \gamma_1 B$ is a set of diameter at most $2/C^2$ and is contained in B. Likewise, $\iota \gamma_1 \iota \gamma_2 B$ has diameter at most $2/C^4$; and furthermore $\iota \gamma_1 \iota \gamma_2 B \subset \iota \gamma_1 B \subset B$. Proceeding recursively, we obtain a sequence of compact nested sets with uniformly shrinking diameter. Their intersection is the point $\mathbb{K}\{\gamma_i\}$. \Box

We now reverse the continued-fraction algorithm, obtaining a sequence of digits that converges to a given point. Note that *not all* of the resulting converging continued fractions satisfy the assumptions of Theorem 2.5.3.

Definition 2.5.4 (Admissible fundamental domain). A fundamental domain K for $\Phi_{\mathbb{Z}}^1$ is *admissible* if it satisfies $0 \in K$, is properly contained in the unit ball with respect to the gauge metric, and tiles Φ^1 without overlap.

Definition 2.5.5 (Continued fraction expansion). Let K be an admissible fundamental domain and $[\cdot]$ the associated nearest-integer mapping. The Gauss map associated to K is given by $T(p) = [\iota p]^{-1} * \iota p$. The continued fraction digits of a point $p \in K$ are the Heisenberg-integer elements that appear under iteration of T. That is, each $\gamma_i \in \Phi^1_{\mathbb{Z}}$ satisfies $T^i p = \gamma_i^{-1} * \iota T^{i-1} p$. We write $CF(p) = \{\gamma_i\}$. The continued fraction contains finitely many steps if and only if $T^k p = 0$ for some k. A variant on the Euclidean algorithm allows us to prove the following result (here $\Phi^n_{\mathbb{O}}$ is the set of points with rational coordinates in the geometric model):

Theorem 2.5.6 (Lukyanenko–Vandehey [31]). A point $p \in \Phi^1$ admits a finite continued fraction expansion if and only if $p \in \Phi^1_{\mathbb{O}}$.

We will provide a proof of Theorem 2.5.6 in §2.7. The proof of the main theorem 2.5.7 follows the same framework but is more technical, see [31].

Theorem 2.5.7 (Lukyanenko–Vandehey [31]). Let $\{\gamma_i\}$ be the digits of a point $p \in K$, with respect to an admissible fundamental domain K in Φ^1 . Then $\mathbb{K}\{\gamma_i\}$ exists and equals p.

Remark 2.5.8. The dimension restriction in the above discussion is critical, as admissible fundamental domains cease to exist in higher dimension. This is true even in the analogous Euclidean case where any fundamental domain for the \mathbb{Z}^n action has volume 1 while the unit ball has volume smaller than 1 in dimensions above 13.

2.6 Conformal dynamics

The study of continued fractions in [31] led to a question in dynamical systems, which we now describe. Recall from §2.5 that the standard continued fraction on \mathbb{R} is defined by means of the Gauss map $T : [0,1] \to [0,1]$ given by $T(x) = 1/x - \lfloor 1/x \rfloor$. It is a classical result that the Gauss map leaves invariant the Gauss measure

$$\mu(A) = \frac{1}{\log 2} \int_A \frac{1}{1+z} dx$$

Here, a measure μ is invariant under a transformation T if $\mu(f^{-1}A) = \mu(A)$ for any measurable A; with f^{-1} denoting the full preimage of A under f.

Furthermore, it is classical that the Gauss map is ergodic with respect to μ . That is, any *T*-invariant set has either measure zero or full measure.

Question 2.6.1 (Lukyanenko–Vandehey [31]). Does the Heisenberg Gauss map admit an invariant measure that is absolutely continuous with respect to Lebesgue measure? If so, is it ergodic with respect to this measure?

The standard machinery for questions similar to Question 2.6.1 is the notion of a fibered system (see Figure 2.3 for some intuition).

Definition 2.6.2 (Fibered System). Consider a topological space K and a piecewise-continuous mapping $T: K \to K$. Let \mathcal{D} be a countable *digit set*, and assume that T is continuous and invertible on sets $C_{\{w_1\}} \subset K$, for various $w_1 \in D$. As with continued fractions, one associates with each $h \in K$ the sequence $\{w_1,\ldots\}$ of digits satisfying $T^{n-1}h \in C_{w_n}$. A sequence arising in



Figure 2.3: The cylinders for the complex Gauss map, given by the inversion of the integer lattice in the unit circle. The large square has width 1 and is centered at the origin.

this way (or any of its initial subsequences) is called *admissible*. To each finite admissible sequence $\{w_1, \ldots, w_n\}$, one associates the cylinder set $C_{\{w_1, \ldots, w_n\}}$, consisting of the points in K whose digit sequence starts with $\{w_1, \ldots, w_n\}$. The collection of cylinders is known as a *fibered system*.

Theorem 2.6.3 (See Theorems 4 and 8 in [42]). Let T give rise to a fibered system over a set K, with digit set \mathcal{D} . Let λ be some measure on K. Suppose

- 1. $\lambda(K) = 1;$
- 2. The system is Markov (that is, all the cylinders are full);
- 3. For any infinite admissible sequence $w = \{w_1, w_2, ...\}$ of digits from \mathcal{D} , we have

$$\lim_{n \to \infty} diam C_{\{w_1, w_2, \dots, w_n\}} = 0;$$

4. There is a constant $C \ge 1$ such that for all finite admissible strings w of length n, the Jacobian $J(T^n)$ of T^n satisfies

$$\frac{\sup_{y \in T^n C_w} J_y(T^n)}{\inf_{y \in T^n C_w} J_y(T^n)} \le C$$

Then T is ergodic and admits a unique finite invariant measure μ absolutely continuous with respect to λ (furthermore, μ is equivalent to λ).

While Theorem 2.6.3 is strong enough to prove that many systems are ergodic, including the base-b system of Theorem 2.1.13, it is not sufficient for our case. Indeed, the cylinders of the Heisenberg Gauss map are not full, as is illustrated in Figure 2.3 for the analogous continued-fraction dynamical system in the complex plane. In this case, it is only known that the dynamical system leaves invariant a measure equivalent to Lebesgue measure, but neither ergodicity nor a nice expression for the measure are available. For complex Gauss maps starting with different fundamental domains, nothing seems to be known.



Figure 2.4: Suspected invariant measure for the Heisenberg Gauss map for the Dirichlet region K_D (left) and cube K_C (right). The radius of each sphere represents the expect value of the measure.

Nonetheless, experimental results offer some hope. Figure 2.6 shows the results of a computer simulation estimating the value of a putative invariant measure at each point. These were computed as follows. A fundamental domain K was chosen (in Figure 2.6 the Dirichlet domain is the region closer to 0 than to any other integer point, in the gauge metric). Then a hopefully generic point was chosen within K, and its forward orbit under T was computed. By the Birkhoff Ergodic Theorem, the number of visits of the point to each subset of K would correspond to the measure of the subset with respect to the invariant measure. Thus, K was broken up into "bins" and visits to the bin were counted. For each bin, a sphere is displayed, centered at the center of the bin, and with radius corresponding to the number of visits to the bin.

2.7 Representing the Heisenberg group

We now return to the general theory of the Heisenberg group, setting up the framework used to prove the results in §2.5. We start by representing the space Φ^n as a subset S of $\mathbb{C}^{n+1} \hookrightarrow \mathbb{CP}^{n+1}$. We then observe that certain matrices in $GL(n+2,\mathbb{C})$, acting by linear fractional transformations on \mathbb{CP}^{n+1} , restrict to the closure \overline{S} of S as conformal mappings of Φ^n . Furthermore, we see that nearly all conformal mappings of Φ^n can be represented in this way, giving a representation $\mathbb{U}: \operatorname{Conf}(\Phi^n) \to PGL(n+1,\mathbb{C}).$

Recall that complex projective space \mathbb{CP}^{n+1} is the space of non-zero vectors in \mathbb{C}^{n+2} , with two vectors considered equivalent if they are multiples of each other. Points in \mathbb{CP}^{n+1} are written as $(z_0 : \ldots : z_{n+1})$, with the coordinates well-defined only up to rescaling. A standard coordinate patch is provided by setting $z_0 = 1$ and interpreting $(1 : z_1 : \cdots : z_{n+1})$ as $(z_1, \ldots, z_{n+1}) \in \mathbb{C}^{n+1}$.

Linear transformations $GL(n+2,\mathbb{C})$ of \mathbb{C}^{n+2} act by linear-fractional transformations $PGL(n+2,\mathbb{C})$ on \mathbb{CP}^{n+1} . Elements of $PGL(n+2,\mathbb{C})$ are, likewise, $(n+2) \times (n+2)$ complex matrices, considered equivalent if they differ by a scalar multiple.



Figure 2.5: A schematic of the Siegel region S defined by $-2\operatorname{Re}(z_{n+1}) = -(|z_1|^2 + \ldots + |z_n|^2)$. The blue (depth) axis represents the coordinates z_1, \ldots, z_n , while the red (width) and green (height) axes represent the real and complex parts of z_{n+1} , respectively.

Definition 2.7.1. The Siegel model S of Φ^n is the space (see Figure 2.5)

$$S = \{(z_1, \dots, z_{n+1}) \in \mathbb{C}^{n+1} : -2\operatorname{Re}(z_{n+1}) = |z_1|^2 + \dots + |z_n|^2\}$$

We identify Φ^n with S via the correspondence $\zeta : (\vec{z}, t) \mapsto (\vec{z}(1+i), |\vec{z}|^2 + it)$, and view $S \subset \mathbb{C}^{n+1} \subset \mathbb{CP}^{n+2}$ via the embedding $(z_1, \ldots, z_{n+1}) \mapsto (1 : z_1 : \cdots : z_{n+1})$.

We now begin to define the map $\mathbb{U}: \operatorname{Conf}(\Phi^n) \to PGL(n+2, \mathbb{C}).$

Definition 2.7.2. For each $p \in \Phi^n$ with geometric coordinates (\vec{z}, t) , define

$$\mathbb{U}(p) := \begin{pmatrix} 1 & 0 & 0 \\ \vec{z}(1+i) & I & 0 \\ |\vec{z}|^2 + it & \vec{z}^{\dagger}(1-i) & 1 \end{pmatrix},$$

where \vec{z}^{\dagger} denotes conjugate transpose of \vec{z} and I is the $n \times n$ identity matrix. The image of Φ^n under \mathbb{U} is the *(Siegel) unitary model* of Φ^n .

It is easy to see that we have correctly represented left multiplication:

Lemma 2.7.3. For all $p, q \in \Phi^n$ we have $\mathbb{U}(p) \cdot \zeta(q) = \zeta(p * q)$, where $\mathbb{U}(p)$ acts by a linear fractional transformation on $S \subset \mathbb{C}^{n+1} \subset \mathbb{CP}^{n+1}$.

Definition 2.7.4. Recall from Theorem 2.4.11 that every conformal mapping is a Möbius transformation (Definition 2.4.9). We extend \mathbb{U} to the rotations, dilations, and Korányi inversion.

$$\mathbb{U}(\delta_r) := \begin{pmatrix} r^{-1} & 0 & 0\\ 0 & I & 0\\ 0 & 0 & r \end{pmatrix} \qquad \qquad \mathbb{U}(A \oplus 1) := \begin{pmatrix} 1 & 0 & 0\\ 0 & A & 0\\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbb{U}(\iota) := \left(\begin{array}{rrrr} 0 & 0 & -1 \\ 0 & I & 0 \\ -1 & 0 & 0 \end{array}\right)$$

Remark 2.7.5. Note that we are not defining \mathbb{U} on the reflection $(z,t) \mapsto (\overline{z}, -t)$, as it does not act on S by a linear fractional transformation (instead, it acts by $(z_1, \ldots, z_{n+1}) \mapsto (\overline{z_1}, \ldots, \overline{z_{n+1}})$). We define $\operatorname{Conf}^+(\Phi^n)$ to be the group of conformal mappings generated by left translation, rotation, dilation, and Korányi inversion.

It is straightforward to check that we have correctly represented the conformal maps in $\operatorname{Conf}^+(\Phi^n)$:

Lemma 2.7.6. Let $p \in \Phi^n$ and $f \in Conf^+(\Phi^n)$. Then $\mathbb{U}(f) \cdot \zeta(p) = \zeta(fp)$.

2.8 Unitary model

We now explain the notation \mathbb{U} and the terminology "unitary model".

Recall that a complex matrix is *unitary* if $M^{\dagger}IM = I$, or equivalently if M preserves the standard Hermitian inner product $\langle z, w \rangle = z^{\dagger}w$.

More generally, let J be a complex matrix satisfying $J^{\dagger} = J$. A matrix M is unitary with respect to J if $M^{\dagger}JM = J$, or equivalently if M preserves the inner product $\langle z, w \rangle_J = z^{\dagger}Jw$. With the restriction $J^{\dagger} = J$, the matrix J must have real eigenvalues. If J has a positive eigenvalues and b negative eigenvalues, it has signature (a, b). It is well-known that up to a change of coordinates there is a unique matrix of each signature, so that one can speak of the unitary groups U(a, b) for each a, b. If a specific J is chosen, one can also speak of U(J).

Working in dimension n + 2, fix a matrix J. The inner product $\langle \cdot, \cdot \rangle$ splits \mathbb{C}^{n+2} into three types of vectors: those of positive square norm, negative square norm, and zero square norm. The classification is invariant under rescaling, so that one obtains positive, negative, and null points in \mathbb{CP}^{n+1} . The matrices U(J) preserve the inner product and in particular the separation into positive, negative, and null vectors. Projectivizing, we see that the group PU(J) preserves the separation of \mathbb{CP}^{n+1} into positive, negative, and null points.

Lemma 2.8.1. Fix the matrix $J = J_3$ of signature (n, 1) given by

$$J_3 = \left(\begin{array}{ccc} 0 & 0 & -1 \\ 0 & I & 0 \\ -1 & 0 & 0 \end{array} \right).$$

The Siegel model S of the Heisenberg group consists entirely of null points with respect to J_3 . Furthermore, $\mathbb{U}(Conf^+(\Phi^n)) \subset PU(J)$.

Proof. The first statement is part of the definition of S. The second is a quick computation.

The converse of Lemma 2.8.1 is almost true. In fact:

Lemma 2.8.2. The set of null points in \mathbb{CP}^{n+1} is exactly the closure $\overline{S} = S \cup \{(0 : 0 : \cdots : 1)\}$. Furthermore, $\mathbb{U} : Conf^+(\Phi^n) \to PU(J_3)$ is a group isomorphism.

2.9 Finite continued fractions and rational points

We now sketch a proof of Theorem 2.5.6, which states that it is the rational points in Φ^1 that have finite continued fraction expansions. Our proof (joint with Joseph Vandehey [31]) is motivated by the work of Falbel–Francsics–Lax–Parker [12].

We will require the following lemma:

Lemma 2.9.1. The gauge metric on S is given by

$$||h||_q = ||(z_1, \dots, z_{n+1})||_q = \sqrt{|z_{n+1}|}.$$

The action of the unitary model on \mathbb{CP}^n preserves S, acting by isometries in the gauge metric.

Now, there are three ways to interpret the phrase "rational point". They coincide:

Lemma 2.9.2. The following are equivalent characterizations of rational points in Φ^n :

- 1. Points (x_i, y_i, t) in the geometric model with all coordinates rational.
- 2. Points of the form $\delta_r h$, for $r \in \mathbb{Q}$ and $h \in \Phi^n_{\mathbb{Z}}$.
- 3. Points (z_1, \ldots, z_{n+1}) in the Siegel model S with all coordinates in $\mathbb{Q}[i]$.

For the remainder of the section, we work with Φ^1 in the Siegel model, and fix an admissible fundamental domain K for $\Phi^1_{\mathbb{Z}}$ with respect to which continued fractions are defined. We furthermore fix a point $h = (u, v) \in K$. Lifting to \mathbb{C}^3 , h is represented by the vector (1, u, v).

It is clear from Lemma 2.9.2 that if h is given by a finite continued fraction, then it is rational (this is all the more clear if we allow $r \in \mathbb{Q}[i]$ for δ_r). We now focus on the opposite implication. Namely, if $\{\gamma_i\}$ are the digits of the continued fraction expansion of (u, v), we would like to prove that for sufficiently high iwe have $\gamma_i = 0$. **Definition 2.9.3.** Given an element $\gamma \in \Phi^1_{\mathbb{Z}}$ with Siegel coordinates $(\alpha, \beta) \in (\mathbb{Z}[i] \times \mathbb{Z}[i]) \cap S$, define

$$A_{\gamma} := \mathbb{U}(\iota)\mathbb{U}(\gamma) = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ \alpha & 1 & 0 \\ \beta & \overline{\alpha} & 1 \end{pmatrix}$$
$$= \begin{pmatrix} -\beta & -\overline{\alpha} & -1 \\ \alpha & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}.$$

Lemma 2.9.4. In projective coordinates for the Siegel model, we have

$$\mathbb{K}\{\gamma_i\}_{i=1}^n = A_{\gamma_1} \cdots A_{\gamma_n}(1:0:0).$$

Proof. Abstractly, we have the definition $\mathbb{K}\{\gamma_i\}_{i=1}^n = \iota \gamma_1 \cdots \iota \gamma_n$. Using the identity element $0 \in \Phi$, we may also write $\mathbb{K}\{\gamma_i\}_{i=1}^n = \iota \gamma_1 \cdots \iota \gamma_n 0$. We now apply Lemmas 2.7.6 and 2.8.2 to switch to $\mathbb{U}(\iota)\mathbb{U}(\gamma_1)\cdots\mathbb{U}(\iota)\mathbb{U}(\gamma_n)\cdot(1:0:\cdots:0)$.

The continued fraction algorithm terminates after *i* steps exactly if the *i*th forward iterate $h_i = T^i h$ is equal to zero. The idea of the proof is to now show that the points h_i can be written as fractions whose denominators are strictly decreasing with *i*.

Recall that h is rational and write $h = \left(\frac{r}{q}, \frac{p}{q}\right)$, with $q, r, p \in \mathbb{Z}[i]$. Because $h \in K$, we have by Lemma 2.9.1 that $|p/q| \leq \operatorname{rad}(K)^2 < 1$, where $\operatorname{rad}(K)$ is the maximal gauge norm of the points in K, bounded by 1 for admissible fundamental domains.

Consider the first forward iterate $h_1 = Th = \gamma_1^{-1}\iota h$ as a vector in \mathbb{C}^3 :

$$\begin{pmatrix} q^{(1)} \\ r^{(1)} \\ p^{(1)} \end{pmatrix} := A_{\gamma_1}^{-1} \begin{pmatrix} q \\ r \\ p \end{pmatrix}$$
$$= \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & \alpha_1 \\ -1 & -\overline{\alpha_1} & -\overline{\beta_1} \end{pmatrix} \begin{pmatrix} q \\ r \\ p \end{pmatrix} = \begin{pmatrix} -p \\ r + \alpha_1 p \\ -q - \overline{\alpha_1} r - \overline{\beta_1} p \end{pmatrix}$$

Thus, h_1 is a rational point with planar Siegel coordinates $h_1 = \left(\frac{r^{(1)}}{q^{(1)}}, \frac{p^{(1)}}{q^{(1)}}\right)$. Furthermore, we have $q^{(1)} = -p$, so that

$$\left|\frac{q^{(1)}}{q}\right| = \left|\frac{p}{q}\right| = \|h\|^2 < \operatorname{rad}(K)^2 < 1.$$
(2.9.1)

Repeating this procedure recursively, we have rational coordinates $h_i = \left(\frac{r^{(i)}}{q^{(i)}}, \frac{p^{(i)}}{q^{(i)}}\right)$ for each forward iterate h_i , satisfying $|q^{(i)}| = |p^{(i-1)}|$. Since $h_i \in K$

for each i, we obtain for each n:

$$\left|q^{(n)}\right| \le |q| \,(\mathrm{rad}(K))^{2n}$$
 (2.9.2)

For sufficiently large n, we conclude $|q^{(n)}| < 1$, which implies that $q^{(n)} = 0$, but that is only possible if $h_{n-1} = 0$ and CF(h) is, in fact, finite.

2.10 PU(n, 1) and the sub-Riemannian metric

In §2.7, we defined the Siegel model S of Φ^n as the set of points (z_1, \ldots, z_{n+1}) satisfying the condition $-2\operatorname{Re}(z_{n+1}) = |z_1|^2 + \cdots + |z_n|^2$. In this model, the gauge distance from a point to the origin simplified to $||(z_1 : \cdots : z_{n+1})||_g = \sqrt{|z_{n+1}|}$. The sub-Riemannian metric likewise admits a straightforward explanation.

We start by restricting the tangent space of S based on its embedding in \mathbb{C}^{n+1} . Note first that \mathbb{CP}^{n+1} is a complex manifold, so that at any point $p \in \mathbb{CP}^n$ the tangent space $T_p \mathbb{CP}^{n+1}$ is a complex vector space. We denote multiplication by $i = \sqrt{-1}$ in this vector space by \mathbb{J} (not to be confused with the matrices J that give us Hermitian inner products).

Definition 2.10.1. Let $S \subset \mathbb{CP}^n$ be a smooth submanifold of codimension 1. The standard contact form α on S is given by $\alpha(p) = (\mathbb{J}_* d\vec{n}(p))|_{T_pS}$, where $\vec{n}(p)$ is the normal vector to S at p. The *complex tangent space* $T_{\mathbb{C},p}S$ to S at p is the space Ker $\alpha = (\mathbb{C}n(p))^{\perp}$.

Thus, the complex tangent space is a complex vector space consisting of tangent vectors to S whose \mathbb{C} -span is tangent to S. It is a reasonable subspace to select, since any complex-analytic operation on S that ignores the normal vector must also ignore its complex multiples.

Recall now that the sub-Riemannian metric on Φ^n was defined by selecting a subbundle $H\Phi^n \subset T\Phi^n$.

Lemma 2.10.2. In the Siegel model, we have $H\Phi^n = T_{\mathbb{C}}S$.

Proof. Via the unitary representation, Φ^n acts on S by conformal complexanalytic transformations, so it suffices to check the claim at the origin. Recall that the geometric model of Φ^n embeds into \mathbb{CP}^{n+1} via $(z,t) \mapsto ((1+i)z, |z|^2 +$ $it) \subset \mathbb{C}^{n+1} \hookrightarrow \mathbb{CP}^{n+1}$. Thus, at the origin $H\Phi^n$ corresponds to the space spanned by the first n complex coordinates. The space S is given by the equation $|z_1|^2 + \ldots + |z_n|^2 - 2\operatorname{Re} z_{n+1} = 0$, and so has normal vector $\vec{n}(0) = (0, \ldots, 1)$. The orthogonal complement of $\mathbb{C}\vec{n}(0)$ coincides exactly with $H\Phi^n|_0$.

2.11 Compactification of the Heisenberg group

Recall that sterographic projection relates the spaces \mathbb{S}^n and \mathbb{R}^n . A similar mapping is available between the sphere $\mathbb{S}^{2n+1} \subset \mathbb{C}^{n+1}$ and the Heisenberg

group Φ^n , if we give the sphere a sub-Riemannian metric.

Definition 2.11.1. Let J_1 be the diagonal $(n + 2) \times (n + 2)$ matrix with entries $-1, 1, \ldots, 1$. The set of null points in \mathbb{CP}^1 with respect to the associated Hermitian form is exactly the unit sphere $\mathbb{S}^{2n+1} \subset \mathbb{C}^{n+1} \subset \mathbb{CP}^{n+1}$. We give \mathbb{S}^{2n+1} a sub-Riemannian metric (see §2.10) d_{sR} by restricting the Euclidean inner product on \mathbb{C}^{n+1} to $H\mathbb{S}^{2n+1} := T_{\mathbb{C}}\mathbb{S}^{2n+1}$.

The sub-Riemannian sphere is related to the Heisenberg group by a change of coordinates. Namely, we can define a generalized stereographic projection from \mathbb{S}^{2n+1} to Φ^n by first relating J_1 and J_3 , and then \mathcal{S} and Φ^n . More precisely,

Definition 2.11.2. The Cayley transform on \mathbb{CP}^{n+1} is the map given by the matrix

$$\mathcal{C} = \left(egin{array}{ccc} 1 & 0 & 1 \\ 0 & (1-\mathrm{i})I & 0 \\ \mathrm{i} & 0 & -\mathrm{i} \end{array}
ight),$$

which satisfies (projectively) $C^{\dagger}J_1C = J_3$ and sends S to S^{2n+1} . The stereographic map from Φ^n to S^{2n+1} is the composition of $(z,t) \mapsto (z(1+i), |z|^2 + it)$ sending Φ^n to S with the action of the Cayley transform:

$$(z,t)\mapsto \left(\frac{2z}{1+\left|z\right|^{2}+\mathrm{i}t},\frac{\mathrm{i}(1-\left|z\right|^{2}-\mathrm{i}t)}{1+\left|z\right|^{2}+\mathrm{i}t}\right)$$

The stereographic unitary representation of Φ^n is, likewise, the image of Φ^n in $U(n+1,1;J_1)$ under the composition of \mathcal{C} with $\mathbb{U}: \Phi^n \to U(n+1,1;J_3)$.

Remark 2.11.3. Stereographic projection provides an explicit Darboux chart for the contact structure on \mathbb{S}^{2n+1} .

We thus have the group Φ^n sitting inside the conformal automorphism group of the sub-Riemannian \mathbb{S}^{2n+1} , just as Euclidean space sits within the conformal group of the Riemannian sphere. We finish the chapter with the following characterization of Φ^n from this perspective.

Proposition 2.11.4. Let U(n + 1, 1) be the unitary group of signature (n + 1, 1) and $p \in \mathbb{CP}^{n+1}$ a null point for the associated Hermitian from. Then the subgroup of U(n+1, 1) of transformations fixing exactly the point p is isomorphic to Φ^n .

Proof. Using the symmetries of \mathbb{S}^{2n+1} , we may assume p is the north pole. After applying the Cayley transform, the task becomes to identify the conformal transformations of S that have no fixed points. These are exactly the maps $\mathbb{U}(\Phi^n)$.

Chapter 3

Heisenberg geometry distorted

In Chapter 2, we focused on the Heisenberg group Φ^n from a relatively rigid algebraic perspective. We saw that the study of conformal mappings of Φ^n led to interesting directions in number theory and dynamical systems.

We now expand our point of view to include a much wider class of spaces and mappings. We will be interested in *sub-Riemannian manifolds*, which share much of the structure of Φ^n while being flexible enough to allow applications ranging from robotics to neurophysiology. As in the Riemannian context, isometries and conformal mappings become far more rare in the study of sub-Riemannian manifolds. We are therefore led to consider quasi-conformal homeomorphisms and quasi-regular branched covering maps between sub-Riemannian manifolds.

Our primary focus will be on uniformly quasi-regular (UQR) mappings of sub-Riemannian manifolds, whose distortion does not build up under iteration. After building up a theory of sub-Riemannian manifolds and the appropriate notion of differentiation, we will show that every UQR mapping leaves invariant a measurable conformal structure in §3.9 and provide a family of examples of non-trivial UQR mappigns in §3.10.

The material in this chapter is based on joint work with Katrin Fässler and Kirsi Peltonen [30].

3.1 Sub-Riemannian manifolds

Definition 3.1.1. Let M be a smooth manifold, and $HM \subset TM$ a smooth subbundle (that is, a smoothly-varying choice of subspace of each tangent space; though we allow its dimension to vary). Define, inductively:

$$H_1M = HM \qquad \qquad H_{i+1}M = [H_1M, H_iM]$$

If there exists an s > 0 such that $H_s M = TM$, then one says that HM is completely non-integrable, and M is a sub-Riemannian manifold. If the dimension of each $H_i M$ is constant on M, then M is equiregular. A curve $\gamma \subset M$ is admissible (or horizontal) if $\dot{\gamma} \in HM$ almost everywhere.

Note that the condition that HM is completely non-integrable is the opposite

of the assumption of the Frobenius theorem, which states that a distribution that is closed under the Lie bracket is tangent to a foliation of the space by submanifolds.

Theorem 3.1.2 (Chow's Theorem). Connected sub-Riemannian manifolds are path-connected by admissible curves. Furthermore, a choice of inner product on HM induces a path metric that generates the standard topology on M.

Example 3.1.3. The metric d_{sR} on Φ^n is a sub-Riemannian metric.

Example 3.1.4. The sub-Riemannian sphere in Definition 2.11.1 is a sub-Riemannian manifold.

Example 3.1.5. Consider the following vector fields on $M = \mathbb{R}^2$:

$$X = \frac{\partial}{\partial x} \qquad \qquad Y = x \frac{\partial}{\partial y}$$

Let $HM = \langle X, Y \rangle$. It is easy to see that the distribution is bracket-generating, so choosing X and Y to be unit vectors (except along x = 0) gives M a sub-Riemannian metric. The space is called the *Grushin plane*. While it is the most straight-forward non-trivial sub-Riemannian metric space, it is not homogeneous and is less-studied than Φ^n .

Example 3.1.6. The unit tangent bundle $T^1\mathbb{R}^2$ of the Euclidean plane admits a sub-Riemannian metric. Namely, a vector is allowed to turn or to move forward in \mathbb{R}^2 in the direction it is facing. The resulting *roto-translation space* may also be thought of as the orientation-preserving isometry group $Isom^+(\mathbb{R}^2)$ of \mathbb{R}^2 , or as the space $\mathbb{C} \times \mathbb{S}^1$. One can also consider the universal cover $\widetilde{\mathcal{RT}} = \mathbb{C} \times \mathbb{R}$ on which the distribution is easiest to write down:

$$H\widetilde{\mathcal{RT}} = \left\langle \cos\theta \frac{\partial}{\partial x} + \sin\theta \frac{\partial}{\partial y}, \frac{\partial}{\partial \theta} \right\rangle$$

One may think of the sub-Riemannian \mathcal{RT} as specifying the positions of a wheelbarrow, with the restriction to $H\mathcal{RT}$ signifying that one can push the wheelbarrow forward or rotate it, but not push it over sideways.

Example 3.1.7. Let M be a complete Riemannian manifold and T^1M its unit tangent bundle. One can give T^1M a sub-Riemannian metric by allowing a vector $(p, v) \in T^1M$ to change v while fixing p, or to change p in the direction that v is pointing.

3.2 Carnot groups

Riemannian manifolds can be thought of on the small scale as a distorted version of \mathbb{R}^n . Indeed, this is formalized by the notion of tangent space and the exponential map. For sub-Riemannian manifolds, the natural infinitesimal model is given by a class of Lie groups called *Carnot groups*. We make this precise in §3.3 and provide a corresponding notion of derivative in §3.7.

Enrico LeDonne recently characterized Carnot groups as exactly the spaces where one can intrinsically make sense of differentiation. Indeed, he shows in [26] that any locally-compact geodesic metric space with a transitive isometry group and admitting non-trivial homotheties is a Carnot group with a sub-Finsler metric. We now give the classical definition of Carnot groups:

Definition 3.2.1 (Carnot group). A connected and simply-connected Lie group G with Lie algebra \mathfrak{g} is a *Carnot group* if there exists a splitting

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \cdots \oplus \mathfrak{g}_s$$

of the Lie algebra into sub-vector-spaces satisfying $[\mathfrak{g}_i, \mathfrak{g}_1] = \mathfrak{g}_{i+1}$ for all $1 \leq i < s$. We then have that \mathfrak{g} is generated by \mathfrak{g}_1 as a Lie algebra, and the group G is nilpotent of step s.

It is common to equip a Carnot group with a left-invariant sub-Riemannian metric by choosing an inner product on \mathfrak{g}_1 , or with a sub-Finsler metric by choosing a norm on \mathfrak{g}_1 . One then obtains a distance function by measuring distances along horizontal curves, where the horizontal subspace is given by $HG = \mathfrak{g}_1$.

Example 3.2.2. A Carnot group with step 1 is simply \mathbb{R}^n with the standard group law. In this case, $H\mathbb{R}^n$ is the full tangent space, so a choice of left-invariant inner product gives \mathbb{R}^n a metric isometric to the usual one. A choice of left-invariant *norm* on the tangent space induces a metric that is bi-Lipschitz to the usual one.

Example 3.2.3. The simplest step-2 Carnot groups are the Heisenberg groups Φ^n . Letting i = 1, ..., n, consider the vector fields

$$X_i = \frac{\partial}{\partial x_i} - 2y_i \frac{\partial}{\partial t} \qquad \qquad Y_i = \frac{\partial}{\partial y_i} + 2x_i \frac{\partial}{\partial t} \qquad \qquad T = \frac{\partial}{\partial t}.$$

Computing the Lie brackets, one sees that the Lie algebra of Φ^n has the presentation $\mathfrak{h} = \langle X_i, Y_i, T : [X_i, Y_i] = 4T \rangle$. The desired splitting $\mathfrak{h} = \mathfrak{h}_1 \oplus \mathfrak{h}_2$ is given by the vector spaces $\mathfrak{h}_1 = \langle X_i, Y_i \rangle, \ \mathfrak{h}_2 = \langle T \rangle$.

The exponential mapping $\exp : \mathfrak{g} \to \mathbb{G}$ on a Carnot group is a diffeomorphism. This allows us to conflate \mathfrak{g} and G, as one does with $T_0\mathbb{R}^n$ and \mathbb{R}^n . Indeed, given a nilpotent \mathfrak{g} with the desired splitting, one can write the corresponding group law using the Baker–Campbell–Hausdorff formula, using the same *exponential coordinates* for both \mathfrak{g} and G. For the Heisenberg group, the exponential coordinates agree with the geometric model.

We mentioned in Remark 2.4.2 that the automorphisms of the Heisenberg group are generated by the dilations δ_r and the transformations $A \oplus 1$, where Ais a symplectic linear transformation of the \mathbb{C}^n component. Here, the constraint on the linear transformation A comes from the need to preserve the Lie algebra of Φ^n . One does have additional Lie group automorphisms such as $(x, y, t) \mapsto$ (x+t, y, t), but these do not preserve the splitting $\mathfrak{h} = \mathfrak{h}_1 \oplus \mathfrak{h}_2$ or the horizontal distribution $H\Phi^n$.

We say that an automorphism of a Carnot group G is grading-preserving if the induced map on the Lie algebra \mathfrak{g} preserves the splitting $\mathfrak{g} = \mathfrak{g}_1 \oplus \cdots \oplus \mathfrak{g}_s$.

While some Carnot groups admit a large number of grading-preserving automorphisms (such as the maps $A \oplus 1$ of Φ^n), for generic Carnot groups, one has fewer automorphisms. Namely, one is limited to the dilations

$$\delta_r(g_1 \oplus \cdots \oplus g_s) = (rg_1) \oplus \cdots \oplus (r^s g_s),$$

where g_i is a vector in the \mathfrak{g}_i layer of \mathfrak{g} .

This rigidity of Carnot group automorphisms leads to more general rigidity phenomena, see Theorem 4.7.1.

3.3 Gromov–Hausdorff tangent spaces

We now explain the sense in which Carnot groups serve as models of equiregular sub-Riemannian manifolds. A corresponding differentiation theorem will be provided in §3.7.

Informally, the tangent space to a metric space X at a point x is another metric space X_{∞} with a choice of point x_{∞} such that small balls in X centered at x look (after rescaling) like balls in X_{∞} centered at x_0 . More precisely:

Definition 3.3.1 (Gromov-Hausdorff tangent space). Let (X, x) be a pointed metric space. Then (X_{∞}, x_{∞}) is the tangent space to X at x if there exists a sequence of rescalings r_i and mappings $f_i : (X_{\infty}, x_{\infty}) \to (X, x; r_i d_X)$ that is asymptotically isometric on compacts (see Definition 2.2.8). Note that the compacts (or balls) are chosen in X_{∞} , but represent arbitrarily small sets in X due to the rescaling of the metric.

Example 3.3.2. The tangent space to a Riemannian manifold of dimension n at any point is unique and isometric to \mathbb{R}^n . Indeed, one takes $f_i(x) = \exp(r_i^{-1}x)$.

Example 3.3.3. The tangent space to a Carnot group \mathbb{G} at any point $x \in \mathbb{G}$ is unique and isometric to \mathbb{G} .

Remark 3.3.4. The rescaling sequence r_i can affect the tangent space. Indeed, there exist metric spaces that have multiple tangent spaces, depending on the choice of the rescaling sequence. Furthermore, the tangent space can vary discontinuously from point to point [50], see also the discussion in [27].

Theorem 3.3.5 (Mitchell [34]). Let M be an equiregular sub-Riemannian manifold and $p \in M$. Then there exists a Carnot group \mathbb{G} with a sub-Riemannian metric such that the Gromov-Hausdorff tangent space \mathcal{T}_pM is isometric to \mathbb{G} . **Example 3.3.6.** Let M be a Riemannian manifold, and $p \in M$. Then $\mathcal{T}_p M$ is isometric to \mathbb{R}^n .

Example 3.3.7. Let $M = \mathbb{S}^{2n+1}$ be a sub-Riemannian sphere, and $p \in M$. One shows using stereographic projection that $\mathcal{T}_p M$ is isometric to (Φ^n, d_{sR}) .

Example 3.3.8. Let $M = \mathcal{RT}$, and $p \in M$. Using Taylor approximation of the vector fields generating the horizontal distribution, one shows that $\mathcal{T}_p M$ is isometric to (Φ^1, d_{sR}) .

Example 3.3.9. The Grushin plane is not equiregular. Away from the *y*-axis, it is Riemannian, hence approximated by \mathbb{R}^2 . At the origin (and similarly along the whole *y*-axis) the Grushin plane is self-similar via the dilations $\delta_r(x, y) = (rx, r^2x)$, and it follows that it serves as its own tangent space. Note that the Grushin plane is *not* a Carnot group.

At the moment, the Gromov–Hausdorff tangent spaces at different points on a metric space are defined independently of each other. We will see in §3.7 that for equiregular sub-Riemannian manifolds a canonical "exponential map" allows us to speak of differentiability, and one can arrange the spaces into a (Gromov– Hausdorff) tangent bundle and speak of continuous or measurable derivatives. Note, however, that this exponential map does not have the properties one expects from Riemannian geometry. Most critically, it need not be locally bi-Lipschitz.

3.4 Quasi-conformal mappings

We now turn to the mapping theory of sub-Riemannian manifolds.

Recall that a conformal mapping $f : X \to Y$ is a homeomorphism that, infinitesimally, sends balls to balls (see Definition 2.4.4). Quasi-conformal mappings, infinitesimally, send balls to ellipsoids.

Definition 3.4.1. Let X, Y be two metric topological manifolds, and $f: X \to Y$ a homeomorphism. One says that f is K-quasi-conformal (K-qc), with $K \ge 1$, if one has

$$H_f(x) = \limsup_{r \to 0} \frac{\sup_y |fx - fy|}{\inf_y |fx - fy|} \le K$$

at all points $x \in X$. Here, the supremum $L(x) = \sup_{y} |fx - fy|$ and infimum $l(x) = \inf_{y} |fx - fy|$ are taken over all points y with |x - y| = r. If f is K-qc for some K, then it is quasi-conformal (QC)

Remark 3.4.2. Definition 3.4.1 is the *metric* definition of quasi-conformality relevant when X and Y are topological manifolds. For arbitrary metric spaces X and Y (such as the Sierpinski carpet) one has to be more careful about the existence of spheres of radius r to establish a meaningful theory.

Subclasses of quasi-conformal mappings include isometries, bi-Lipschitz, and conformal mappings. In \mathbb{R}^n and Φ^n , 1-quasi-conformal mappings coincide with conformal mappings. When K > 1, however, quasi-conformal mappings become less rigid. For example, any diffeomorphism between compact Riemannian manifolds is quasi-conformal. Indeed, the Teichmuller distance between two hyperbolic surfaces S_1, S_2 is the logarithm of the infimal K such that there exists a K-qc mapping from S_1 to S_2 .

We now provide some examples of quasi-conformal mappings in \mathbb{R}^n .

Example 3.4.3. The planar map f(x, y) = (x, 2y) is 2-quasi-conformal.

A large class of QC mappings of \mathbb{R}^n is provided by

Lemma 3.4.4. Every diffeomorphism $f : \mathbb{R}^n \to \mathbb{R}^n$ is locally quasi-conformal onto its image. More generally, let $\lambda_+(\vec{x})$ and $\lambda_-(\vec{x})$ be the largest and smallest singular values of $df|_x$, respectively. If λ_+/λ_- is bounded, then f is quasi-conformal.

Proof. By definition of the singular values, an infinitesimal unit ball at \vec{x} is sent to an ellipse with major axis of length $\lambda_+(x)$ and minor axis $\lambda_-(x)$. Quasi-conformality requires that the ratios of these be globally bounded, which is always true locally.

Example 3.4.5. Quasi-conformal mappings are a priori only homeomorphisms. Indeed, the mapping

$$z \mapsto z |z|^s$$

of the complex plane is QC for s > -1 but is not differentiable at the origin for s < 0 (the definition is checked directly at 0 and via derivatives elsewhere).

The simplest non-smooth quasi-conformal mapping of \mathbb{R}^n is given by:

Example 3.4.6. The inversion map on $\mathbb{R}^n \setminus \{0\}$ given by $\iota(\vec{x}) = -\vec{x}/|\vec{x}|^2$ is (quasi-)conformal, but sends arbitrarily small sets to arbitrarily large ones.

So far, only Example 3.4.5 has been non-smooth.

Example 3.4.7. Let f(x) be a Lipschitz mapping of \mathbb{R} . Then the shear F(x,y) = (x, y + f(x)) is bi-Lipschitz, hence quasi-conformal. It is possible that f is not differentiable on an uncountable set (say, if f is the Cantor staircase), in which case F is also not differentiable on an uncountable set.

Although this is not immediate from Definition 3.4.1, the class of quasiconformal mappings are closed under composition and inversion, so the above examples generate a large class of quasi-conformal mappings. We will obtain yet more in §3.5.

We now turn to quasi-conformality in Φ^n , where Definition 3.4.1 continues to make sense.
Example 3.4.8. According to the Liouville Theorem, every 1-quasi-conformal mapping between domains in Φ^n is a Möbius transformation.

An immediate counterpart to Lemma 3.4.4 is

Lemma 3.4.9. Let $f : \Phi^n \to \Phi^n$ be a smooth transformation preserving the contact structure $H\Phi^n$. Then f is locally quasi-conformal. That is, the restriction of f to any bounded domain is quasi-conformal onto its image.

Although Lemma 3.4.9 seems like a good way to produce quasi-conformal mappings, it is in practice difficult to come up with contact mappings of Φ^n . We will describe the *quasi-conformal flow* approach in §3.5.

A large class of QC mappings is immediately provided by the following lifting theorem, in conjunction with Example 3.4.7.

Theorem 3.4.10 (Capogna–Tang [10]). Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be a K-quasiconformal homeomorphism that preserves Euclidean area. Then there exists a K-quasi-conformal homeomorphism $F : \Phi^1 \to \Phi^1$ that lifts f.

That is, F satisfies $\pi \circ f = F \circ \pi$ and $a_r \circ F = F \circ a_r$, where $\pi : \Phi^1 \to \mathbb{C}$ is the standard projection in geometric coordinates, and $a_r(z,t) = (z,t+r)$.

In higher dimensions, symplectic quasi-conformal mappings of \mathbb{C}^n lift to \mathbb{R} invariant quasi-conformal mappings of Φ^n .

The following questions concerning QC mappings of Φ^n seem to be open at the moment (see also the related Question 3.11.5).

Question 3.4.11. Let $f : \Phi^n \to \Phi^n$ be quasi-conformal. Is it almost-everywhere differentiable (in the classical sense)?

Question 3.4.12 ([20]). Does there exist a quasi-conformal mapping $f : \Phi^n \to \Phi^n$ such that $f(\{(x, 0, 0) : x \in \mathbb{R}\}) = \{(0, 0, t) : t \in \mathbb{R}\}$?

Recall that the stereographic projection $\mathcal{C}: \Phi^n \to \mathbb{S}^{2n+1}$ is conformal. Conjugating with \mathcal{C} , every quasi-conformal mapping of Φ^n yields a quasi-conformal mapping of the sub-Riemannian \mathbb{S}^{2n+1} .

It is more difficult to find quasi-conformal mappings between more complicated sub-Riemannian manifolds, or to show that they do not exist. The theory in this chapter is largely motivated by this task.

3.5 Modifying quasi-conformal mappings

We are now interested in ways in which one can perturb a quasi-conformal mapping. In particular, we would like to extend mappings defined on a subset of a larger space, or reconcile quasi-conformal mappings defined on two different subsets.

In the case of extension, notice that a QC mapping $f : \mathbb{R}^n \to \mathbb{R}^n$ does not readily extend to a QC mapping of \mathbb{R}^{n+1} . For example, because one has control only on the *relative* distance distortions, the trivial extension $f \oplus id : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ is quasi-conformal if and only if f is in fact bi-Lipschitz, in which case $f \oplus id$ is in fact bi-Lipschitz.

Likewise, suppose one has two regions $U, V \subset \mathbb{R}^{n+1}$ and maps $f_U : U \to \mathbb{R}^n$, $f_V : V \to \mathbb{R}^n$ each of which is quasi-conformal onto its image. One would like to paste together f_U and f_V to create a unified mapping $F : U \cup V \to \mathbb{R}^n$ that is quasi-conformal onto its image. If f_U and f_V embed in flows (see below), this is straightforward, as one can use a bump function to trivialize the mappings on $U \cap V$.

The general answer in \mathbb{R}^n for both concerns is based on Sullivan's deep results in [45]. These are based, in turn, on the investigation of algebraic topology and algebraic geometry of hyperbolic manifolds explored by Deligne–Sullivan in [11].

Theorem 3.5.1 (Tukia–Väisälä [49]). Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a quasi-conformal mapping (quasi-symmetric for n = 1). Then there exists a quasi-conformal mapping $F : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ whose restriction to $\mathbb{R}^n \times \{0\}$ equals f.

Theorem 3.5.2 (Sullivan's annulus extension [45, 47]). Let $f : \mathbb{R}^n \setminus A \to \mathbb{R}^n$ be a quasi-conformal embedding, for A an annulus. Then there exists a quasiconformal $F : \mathbb{R}^n \to \mathbb{R}^n$ extending f in the sense that $F|_{\mathbb{R}^n \setminus A'} = f$, for an annulus A' containing the closure of A.

Remark 3.5.3. For our extension of Theorem 3.5.1 see Theorem 4.7.4 and [30].

Analogues of Theorems 3.5.1 and Theorem 3.5.2 are not known in general metric spaces. In sub-Riemannian spaces, we are only able to perturb quasi-conformal mappings that embed in flows.

Recall that a *flow* (on a smooth manifold) is a family f_t of diffeomorphisms satisfying $f_{t_1+t_2} = f_{t_1} \circ f_{t_2}$. A flow is generated by a vector field ξ if one has $\xi(x) = \left. \frac{df_t(x)}{dt} \right|_{t=0}$ at each x in the manifold.

In the Riemannian setting, a flow can be adjusted by directly modifying the vector field ξ . For example, one can use a bump function to make the vector field zero on some region. For sub-Riemannian manifolds, such a modification can produce mappings that do not preserve the horizontal direction and are therefore not quasi-conformal. For sub-Riemannian manifolds with a contact structure, a characterization of appropriate flows was given by Libermann and adapted to the quasi-conformal theory by Korányi–Reimann:

Theorem 3.5.4 (Libermann [29], Korányi–Reimann [25]). Let X_i, Y_i, T be the standard left-invariant vector fields on Φ^n (see Example 3.2.3). A vector field ξ on Φ^n generates a contact flow if and only if it is of the form

$$\frac{1}{4} \sum_{i=1}^{n} (X_i(\rho)Y_i - Y_i(\rho)X_i) + \rho T,$$

where ρ is a smooth potential function. Furthermore, the diffeomorphisms f_t

generated by the flow are quasi-conformal if the second derivatives of ρ with respect to the vector fields X_i, Y_i are bounded (in a quantitative way).

We can thus specify a quasi-conformal mapping of Φ^n by specifying a reasonable function ρ on Φ^n . We can also mimic the annulus extension theorem for some mappings that embed in flows.

Lemma 3.5.5. Let $f : \Phi^n \to \Phi^n$ be a mapping that embeds in a quasi-conformal flow, i.e. $f = f_1$ for a flow f_t generated by a vector field ξ as in Theorem 3.5.4. For each neighborhood U of the origin, there exists a radius r and mapping F such that:

- 1. F is quasi-conformal,
- 2. F = f outside of U,
- 3. F is the identity on the ball B(0,r).

Proof. Let P(x) be a potential function defined by $g(x)\rho(x)$ where ρ is the potential associated to the flow f_t and g is a smooth bump function equal to zero at the origin and one far from it. Let F_t be the flow associated to P. The mapping $F = F_1$ is the desired mapping.

3.6 Pansu differentiability

The quasi-conformal mappings in sections 3.4 and 3.5 are all smooth or at least almost-everywhere differentiable. In the Euclidean case, this is true in general:

Theorem 3.6.1 (Stepanov–Rademacher [19]). Let $f : M \to N$ be a mapping between Riemannian manifolds. Then f is almost everywhere differentiable on the set $\{x : Lip_f(x) < \infty\}$, where $Lip_f(x) = \limsup_{y\to x} d(fx, fy)/d(x, y)$.

Corollary 3.6.2. Let $f: M \to N$ be a quasi-conformal mapping between Riemannian manifolds. Then f is almost everywhere differentiable, with the derivative df varying measurably.

It is *not* known whether every quasi-conformal mapping of sub-Riemannian manifolds is almost-everywhere differentiable. For Carnot groups (§3.2), a differentiability result is due to Pansu. We describe a more general differentiability theorem in §3.7.

Recall that a map f of \mathbb{R}^n is differentiable at x_0 if there exists a linear mapping df of \mathbb{R}^n such that

$$\lim_{x \to x_0} \frac{|fx - (fx_0 + df(x - x_0))|}{|x - x_0|} = 0$$

Equivalently, one has that the following limit converges uniformly-on-compacts in $v \in \mathbb{R}^n$:

$$\lim_{r \to 0} \frac{|f(x_0 + rv) - (fx_0 + df(rv))|}{r} = 0.$$

Definition 3.6.3. A mapping $f : \mathbb{G}_1 \to \mathbb{G}_2$ is Pansu differentiable at a point $x_0 \in \mathbb{G}$ if there exists a graded homomorphism $df : \mathbb{G}_1 \to \mathbb{G}_2$ that approximates f at x_0 . That is the following limit converges uniformly-on-compacts in $v \in \mathbb{G}_1$.

$$\lim_{r \to 0} \frac{d(f(x_0 * \delta_r v), fx_0 * df(\delta_r v))}{r} = 0,$$

where d denotes the distance in \mathbb{G}_2 .

Theorem 3.6.4 (Pansu–Rademacher Theorem [38]). Let f be a Lipschitz or quasi-conformal mapping between domains in Carnot groups. Then at almost every point in the domain of f, the Pansu derivative of f exists.

Corollary 3.6.5. There does not exist a quasi-conformal mapping $f : \Phi^n \to \mathbb{R}^m$ for any n, m.

Proof. By topological considerations, one would require m = 2n + 1. Thus, a QC mapping $f : \Phi^n \to \mathbb{R}^{2n+1}$ would have a derivative df at some point, which would be a Lie group isomorphism $df : \Phi^n \to \mathbb{R}^{2n+1}$. But the groups are not isomorphic (in particular, Φ^n has a one-dimensional center).

Similar arguments can be used to restrict Lipschitz mappings between Φ^n and \mathbb{R}^n , or between more general Carnot groups. In a similar manner, Pansu proves:

Corollary 3.6.6. Suppose \mathbb{G} is a Carnot group with Lie algebra \mathfrak{g} . Suppose further that every Lie algebra automorphism of \mathfrak{g} is a dilation δ_r . Then any quasi-conformal mapping $U \to V$, for $U, V \subset \mathbb{G}$ is induced by a homothety of all of \mathbb{G} .

Remark 3.6.7. One can use Pansu differentiability to define "Pansu manifolds" analogous to smooth manifolds. We explore this idea in §3.11.

3.7 Margulis–Mostow differentiability

The idea of the derivative is to zoom in to a space and observe infinitesimal behavior. That is, given a map $f: X \to Y$ between metric spaces and a point $p \in X$, we would like to dilate X around p and Y around fp to get a "zoomed in" version of f. We would then like to take a limit as we zoom in more and more.

The exact way one zooms into a point can make a difference as to what maps are differentiable, even in Euclidean space, and one has to make a choice of zooming sequence to discuss differentiability. This is possible in Riemannian manifolds via any smooth coordinate chart, and in equiregular sub-Riemannian manifolds via *privileged coordinates*.

Recall that an equiregular sub-Riemannian manifold is a smooth manifold with a smooth distribution $HM \subset TM$ and a smooth choice of inner product on

HM, with the further restriction that the higher brackets of HM have constant dimension and eventually span TM. Denote the dimensions of the horizontal distribution and its brackets by:

$$q_1 = \dim H_1 M = \dim HM$$

$$q_2 = \dim H_2 M - q_1 = \dim \langle HM, [HM, HM] \rangle - q_1$$
...
$$q_s = \dim H_s M - \sum_{i=1}^{s-1} q_i = \dim TM - \sum_{i=1}^{s-1} q_i$$

Definition 3.7.1 (Priveleged coordinates). Let M be a sub-Riemannian manifold of dimension n with horizontal distribution HM and inner product g_{sR} . A privileged coordinate chart at a point $p \in M$ is a smooth chart $\Psi : U \to M$, for $0 \in U \subset \mathbb{R}^n$ satisfying:

- 1. $\Psi(0) = p$,
- 2. The vectors $d\Psi|_0(\frac{\partial}{\partial x_1}), \ldots, d\Psi|_0(\frac{\partial}{\partial x_n})$ are in H_pM and orthonormal with respect to g_{sR} ,
- 3. For each *i*, the vectors $d\Psi|_0(\frac{\partial}{\partial x_1}), \ldots, d\Psi|_0(\frac{\partial}{\partial x_{q_1}+\cdots+q_i})$ span H_iM .

Example 3.7.2. The geometric coordinates for Φ^n are (after the standard identification with \mathbb{R}^{2n+1}) privileged coordinates.

Example 3.7.3. For Riemannian manifolds, privileged coordinates are provided by the exponential map.

Priveleged coordiantes are easily constructed on any sub-Riemannian manifold. Namely, let v_1, \ldots, v_n be vectors in T_pM such that v_1, \ldots, v_{q_1} are orthonormal vectors spanning HM, and such that for each i the vectors $v_1, \ldots, v_{q_1+\ldots+q_i}$ span the first i layers H_iM of TM. Let Ψ be any smooth chart. Up to a composition with a translation normalizing $\Psi^{-1}(p)$ and linear transformation normalizing each $d\Psi(v_j)$, Ψ is a privileged coordinate chart.

In privileged coordinates, as in the geometric model of Φ^n , one can mimic Carnot group dilations to define a mapping

$$\delta_r(\vec{x}_1,\ldots,\vec{x}_s) := (r\vec{x}_1,\ldots,r^s\vec{x}_s).$$

Except in special cases, this is not a homothety, but it is an asymptotic homothety (that is, an asymptotic isometry up to a rescaling of the metrics by factor r). That is, the sub-Riemannian metrics on \mathbb{R}^n (defined near 0) limit to a metric on \mathbb{R}^n that happens to be isometric to $\mathcal{T}_p M$. Indeed, this is the key to Mitchell's approach in [34] for identifying the Gromov-Hausdorff tangent spaces of sub-Riemannian manifolds. **Definition 3.7.4.** Let $f : M \to N$ a mapping between equiregular sub-Riemannian manifolds, and $p \in M$. Suppose there is a grading-preserving isomorphism $d_p f : \mathcal{T}_p M \to \mathcal{T}_{fp} N$ such that, in some (any) privileged coordinates one has that

$$\lim_{r \to \infty} \delta_r \circ f \circ \delta_r^{-1} = d_p f,$$

where the limit denotes uniform convergence on compacts. Then $d_p f$ is the Margulis–Mostow derivative of f at p, and f is Margulis–Mostow differentiable at p.

Theorem 3.7.5 (Margulis–Mostow [33]). Let $f: M \to N$ be a quasi-conformal mapping between equiregular sub-Riemannian manifolds. Then $d_p f$ exists at almost every p.

Indeed, the map $df: \mathcal{T}M \to \mathcal{T}N$ is measurable, see Lemma 6.3 of [13].

Remark 3.7.6. In Theorem 3.7.5, "almost everywhere" refers to the standard notion of zero-measure on smooth manifolds. By Mitchell [34], this further agrees with the notion of Hausdorff measure zero, in the appropriate dimension $Q = \sum_{i} iq_{i}$.

Example 3.7.7. If M and N are Riemannian manifolds, the Margulis–Mostow derivative agrees with the classical derivative (with the classical tangent spaces interpreted as Gromov–Hausdorff tangent spaces).

Example 3.7.8. If M and N are Carnot groups, then the Margulis–Mostow derivative agrees with the Pansu derivative (with the Lie algebra interpreted as a Gromov–Hausdorff tangent space).

3.8 Uniformly quasi-regular mappings

Quasi-conformal mappings often come in groups, and one is often able to use this fact to show that they are, in fact, conformal. This is key to the proof of Mostow's celebrated rigidity theorem [35]. We are interested in the following theorem of Tukia:

Theorem 3.8.1 (Tukia [46]). Let Γ be a group of K-quasi-conformal transformations acting on \mathbb{S}^n . Then there exists a measurable conformal structure on \mathbb{S}^n preserved by every element of Γ .

Here, a measurable conformal structure is a choice of inner product g on \mathbb{S}^n , defined up to rescaling and varying measurably (alternately one can think of a conformal structure as a choice of an ellipsoid in the tangent space over each point). A measurable conformal structure represented by an inner product gis invariant under a quasi-conformal mapping f if one has $f_*g = \lambda g$, for some measurable λ . Here, f_* is the pullback of the inner product defined using the Rademacher theorem.

Our goal now is to state an analogue of Theorem 3.8.1 for *quasi-regular* mappings, which allow for some non-injectivity while maintaining the general idea of quasi-conformality.

Definition 3.8.2. Let $f : M \to N$ be a continuous mapping between two oriented topological manifolds of dimension at least two. The mapping is a *branched cover* if it is discrete (every point has a discrete preimage) and orientation-preserving. The *branch set* of f is the minimal closed set B_f such that $f|_{M\setminus B_f}$ is a covering map onto its image (that is, a local homeomorphism onto its image).

The following is a classical topological result (see Rickman [39] for this lemma and the general theory of QR mappings in \mathbb{R}^n).

Lemma 3.8.3. The branch set B_f has topological codimension at most 2.

Definition 3.8.4. Let M, N be equiregular sub-Riemannian manifolds. A branched mapping $f: M \to N$ is K-quasi-regular if the dilatation of f at every point is at most K (see Definition 3.4.1), and furthermore B_f has measure zero.

A map $f: M \to M$ is uniformly K-quasi-regular if every composition $f \circ \cdots \circ f$ is K-quasi-regular for the same K.

Differentiability is a key step for even stating an analogue of Theorem 3.8.1.

Lemma 3.8.5 (Fässler–Lukyanenko–Peltonen [13]). Let $f : M \to N$ be a quasi-regular mapping between two equiregular sub-Riemannian manifolds. Then $d_p f : \mathcal{T}_p M \to \mathcal{T}_{fp} N$ exists and is furthermore a K-quasi-conformal mapping for almost every $p \in M$.

Sketch of proof. Away from the branch set, f is locally K-quasi-conformal onto its image. Thus, the Margulis–Mostow derivative exists at almost every point (Theorem 3.7.5). Furthermore, one shows that the quasi-conformality condition passes to the tangent space under the Gromov–Hausdorff limit.

Definition 3.8.6. A measurable conformal structure on a sub-Riemannian manifold M is a choice g of inner product on the horizontal distribution HM, up to rescaling. It is invariant under a quasi-regular mapping $f: M \to M$ if one has that $f_*g = \lambda g$ for some measurable λ .

Here, we must specify what we mean by f_*g , since the inner product g exists on the horizontal distribution HM, while f_* must be defined using the Margulis–Mostow derivative df on the Gromov–Hausdorff tangent space. The key is to use privileged coordinates at each point p to identify \mathbb{R}^{q_1} with both the horizontal distribution of H_pM and the horizontal distribution of the Carnot group \mathcal{T}_pM (see Definition 3.7.1).

We can now state our generalization of Theorem 3.8.1:

Theorem 3.8.7 (Fassler–Lukyanenko–Peltonen [13]). Let $f : M \to M$ be a uniformly quasi-regular mapping on a sub-Riemannian manifold. Then M admits a measure conformal structure preserved by f.

3.9 Dynamics of UQR mappings

We now sketch the proof of Theorem 3.8.7, which relies primarily on a reinterpretation in terms of vector bundles, followed by the use of Tukia's central result in [46].

Let M be a manifold and B a finite-dimensional vector bundle over M, with a fixed choice of inner product (on the fibers). For $p \in M$, denote the fiber over M by F_p . Recall that a map $f: B \to B$ is a bundle map if f sends each fiber F_p to another fiber (which we denote F_{fp}), and the restriction $f_p := f|_{F_p} : F_p \to F_{fp}$ is linear. Since f takes fibers to fibers, it induces a map $f|_M : M \to M$, which we continue to denote as f.

We say that a vector bundle map $f: B \to B$ has bounded distortion if

$$\sup_{p \in M} \frac{\sup_{v \in F_p} |fv| / |v|}{\inf_{v \in F_p} |fv| / |v|} = C < \infty.$$
(3.9.1)

Furthermore, f has uniformly bounded distortion if (3.9.1) holds for all iterates f^n of f, with C independent of n.

The bundle *B* admits an *f*-invariant measurable conformal structure if there exists an inner product $\langle \cdot, \cdot \rangle$ on the fibers of *B*, varying measurably, such that for some positive measurable function λ on *M* we have

$$\langle fu, fv \rangle = \lambda \langle u, v \rangle \tag{3.9.2}$$

for all $u, v \in F_p$ for almost every fiber F_p . Typically, one also assumes a boundedness condition for $\langle \cdot, \cdot \rangle$, which can be expressed in terms of the matrix-valued function s associated to $\langle \cdot, \cdot \rangle$.

The following theorem is a rephrasing of Tukia's core result in [46]. We sketch its proof to adapt it to our terminology.

Theorem 3.9.1 (Tukia [46], cf. Theorem 6.1 of [13]). Let $f : B \to B$ be a bundle map of uniformly bounded distortion. Then B admits an f-invariant measurable conformal structure.

Sketch of proof. Denote the given inner product on B by $\langle \cdot, \cdot \rangle_0$ and furthermore fix an orthonormal basis \mathcal{B}_p at each point p, varying measurably. Let d be the dimension of the fibers of B.

We would like to show that there exists an inner product $\langle \cdot, \cdot \rangle$ on *B* that is *f*-invariant, up to a multiplicative factor. That is, we would like the property

that for all u, v in each fiber F_p and for some positive function λ on M,

$$\langle f_p u, f_p v \rangle = \lambda(p) \langle u, v \rangle.$$
 (3.9.3)

Using the basis \mathcal{B}_p , we may find a positive definite matrix $s_p \in GL(d, \mathbb{R})$ so that

$$\langle u, v \rangle = \langle u, s_p v \rangle_0 \tag{3.9.4}$$

and the invariance relation (3.9.3) becomes (taking transposes using \mathcal{B})

$$\langle f_p u, s_{fp} f_p v \rangle_0 = \lambda(p) \langle u, s_p v \rangle_0, \qquad (3.9.5)$$

$$\langle u, f_p^t s_{fp} f_p v \rangle_0 = \lambda(p) \langle u, s_p v \rangle_0, \qquad (3.9.6)$$

$$f_p^t s_{fp} f_p = \lambda(p) s_p. \tag{3.9.7}$$

We now assume s_p has determinant one, so that (3.9.3) further reduces to

$$(\det f_p)^{-2/d} f_p^t s_{fp} f_p = s_p.$$
(3.9.8)

Note now that s_p needs to be an element of the space $S \subset SL(d, \mathbb{R})$ of positive definite symmetric matrices. Tukia points out that S may be identified with $SL(d, \mathbb{R})/SO(d)$, a non-positively curved symmetric space. Normalized transpose-conjugation by an element of $GL(d, \mathbb{R})$, as in (3.9.8), is an isometry of S.

We are now ready to construct s_p . Consider first the orbit of p under f:

$$\mathcal{O}(p) = \{ f^n p : n \in \mathbb{N} \}.$$
(3.9.9)

Note that we have $\mathcal{O}(fp) = f\mathcal{O}(p)$. Now, consider the transformations $(f^n)_p$: $F_p \to F_{f^n p}$ as elements of S:

$$\mathcal{S}(p) = \{ (\det \ (f^n)_p)^{-2/d} (f^n)_p^t (f^n)_p \ : \ n \in \mathbb{N} \} \subset S.$$
(3.9.10)

We obtain the invariance equation $\mathcal{S}(fp) = f_p \cdot \mathcal{S}(p) = (\det f_p)^{-2/d} f_p^t \mathcal{S}(p) f_p.$

Under the standard metric on S, the uniformly-bounded condition on f gives us that S(p) is a bounded set in S. Tukia shows that every bounded set in S is contained in a *unique* ball of minimum radius. Take $s_p \in S$ to be the center of the unique ball of minimum radius containing S(p). Because transpose-conjugation is an isometry in S, the invariance relation on S(p) turns into (3.9.8).

We thus have that the conformal class of $\langle u, v \rangle := \langle u, s_p v \rangle$ is *f*-invariant. It remains to show that it is also measurable. It is clear that the map $p \mapsto S(p)$ is measurable, and Tukia shows that "averaging" operation $S(p) \mapsto s_p$ is continuous with respect to the Hausdorff topology on subsets of S.

Theorem 3.8.7 follows Theorem 3.9.1 by considering the bundle HM, with

the bundle map derived from the quasi-regular mapping via the Margulis– Mostow derivative. The fact that the derived map is measurable and has bounded distortion is intuitively clear, and is formalized in the full proof of Theorem 3.8.7 provided in [13].

3.10 Existence of UQR mappings

Theorem 3.8.7 provides a restriction on the existence of uniformly quasi-regular mappings on a sub-Riemannian manifold M. We finish the chapter with a construction of such a mapping:

Theorem 3.10.1 (Fassler–Lukyanenko–Peltonen [13]). Let $L_{p,\mathbf{q}}$ be a lens space with parameters p, \mathbf{q} , with the natural sub-Riemannian metric. Then there exists a uniformly quasi-regular mapping $f : L_{p,\mathbf{q}} \to L_{p,\mathbf{q}}$ with non-empty branch set.

We first recall the definition of a lens space and its natural sub-Riemannian structure. As a special case, Theorem 3.10.1 applies to the sub-Riemannian sphere \mathbb{S}^{2n-1} (= $L_{1,\mathbf{q}}$), and we will focus on it for the purposes of this section.

Definition 3.10.2. Recall that $\mathbb{S}^{2n-1} \subset \mathbb{C}^n$ has a sub-Riemannian structure given by $H_x \mathbb{S}^{2n-1} = (\mathbb{C}\vec{n}(x))^{\perp}$ and inner product induced by the standard inner product on \mathbb{C}^n . Multirotations of \mathbb{C}^n , of the form

$$R_{\theta_1,\ldots,\theta_n}(z_1,\ldots,z_n) = (e^{i\theta_1}z_1,\ldots,e^{i\theta_n}z_n)$$

preserve \mathbb{S}^{2n-1} and act on it by isometries.

Let $p \in \mathbb{N}$, and $\vec{q} = (q_1, \ldots, q_n) \subset \mathbb{N}^n$, all nonzero and each q_i relatively prime to p. The lens space $L_{p,\mathbf{q}}$ is the quotient space $\mathbb{S}^{2n-1}/\langle R_{2\pi q_1/p,\ldots,2\pi q_n/p}\rangle$. The group generated by the rotation is cyclic of order p, and serves as the fundamental group of the manifold $L_{p,\mathbf{q}}$. Furthermore, the contact structure and sub-Riemannian metric on \mathbb{S}^{2n-1} descends to a contact structure and sub-Riemannian metric on $L_{p,\mathbf{q}}$.

The mapping f in Theorem 3.10.1 is built from the following family of mappings:

Lemma 3.10.3. Let $a \ge 1$ and $f_a(r_1e^{i\theta_1}, \ldots, r_ne^{i\theta_n}) = (r_1e^{ai\theta_1}, \ldots, r_ne^{ai\theta_n})$. Then f_a is a quasi-regular mapping of \mathbb{S}^{2n+1} . Furthermore, if p divides a, then f induces a quasi-regular self-map of $L_{p,\mathbf{q}}$ for any \mathbf{q} .

Sketch of proof. The proof of Lemma 3.10.3 splits into two parts. It is immediately clear that f_a is a branched cover with a null branch set. Indeed, the branch set is exactly the set of points for which some $r_i = 0$. Next, one shows that f_a is "radially bi-Lipschitz mappings," that is it sends all spheres to quasi-spheres but is not necessarily locally homeomorphic. One shows that f_a is radially bi-Lipschitz (this is easier away from the branch set), and that, in turn, such mappings are quasi-regular.

The proof of Theorem 3.10.1 proceeds by turning f_a into a *uniformly* quasiregular mapping by using the *conformal trap method*, which is best explained using an example:

Example 3.10.4 (Conformal trap construction). Suppose we have a quasiregular planar map $f : \mathbb{R}^2 \to \mathbb{R}^2$ with the following properties:

- 1. The point 0 has two preimages $f^{-1}(0) = \{z_1, z_2\}.$
- 2. The closures of the unit balls $B_0 := B(0,1), B_1 := B(z_1,1), B_2 := B(z_2,1)$, and $B_3 := B(f(0),1)$ are disjoint and do not intersect the branch set.
- 3. The map f respects the balls: $f(B_1) = f(B_2) = B_0$, $f^{-1}(B_0) = B_1 \cup B_2$, and $f(B_0) = B_3$.
- 4. The restriction $f|_{\partial B_i}$ is a Euclidean translation for i = 0, 1, 2.

Under these (idealized) conditions, we can apply the conformal trap method. First, we define a new map g_1 that agrees with f outside of the balls B_i , and inside the balls B_i is given by a Euclidean translation provided by Condition (4). Next, we write $g = \iota \circ g_1$, where $\iota(z) = \frac{1}{z}$ is inversion in the unit circle.

It is now straightforward to check that g is uniformly quasi-regular. Namely, a point inside B_0 is trapped in B_0 and sees no distortion. A point inside B_1 or B_2 likewise sees no distortion, while a point outside of $B_0 \cup B_1 \cup B_2$ is sent into B_0 and does not see any distortion under further iterations. Furthermore, because we are working in the Euclidean plane, the map g obtained in this way is quasiregular provided that f was quasiregular. One also notes that the branch sets of f and g agree.

Sketch of proof of Theorem 3.10.1. Theorem 3.10.1 modifies the mapping f_a from Lemma 3.10.3 along the lines of Example 3.10.4. One chooses a point $x \in \mathbb{S}^{2n-1}$ away from the branch set B_{f_a} and its preimage, and modifies f_a in a neighborhood of x and its preimages. This provides analogues of conditions 1 and 2 in the example.

Analogues of conditions 3 and 4 are not satisfied by f_a . In the Euclidean setting, such issues are remedied using the Sullivan Annulus Extension Theorem [21, 45]. The theorem is not available in the sub-Riemannian setting, so instead we embed f_a in a local quasi-conformal flow and modify it to a mapping satisfying conditions 3 and 4 analogously to Lemma 3.5.5.

Once the mapping f_a is satisfied to fit the analogue of conditions 1–4, it remains to identify an appropriate inversion. Recall that the conformal mappings of \mathbb{S}^{2n-1} are given by the group of projective unitary transformations PU(n-1,1). Using these transformations, one can conjugate the inversion $\vec{z} \mapsto -\vec{z}$ to a transformation that inverts in the boundary of an arbitrarily small region of \mathbb{S}^{2n-1} . The remainder of the proof proceeds as in Example 3.10.4. The case of more general lens spaces is achieved by manipulating f_a more symmetrically so that the final mapping descends to a mapping of the lens space to itself.

3.11 Aside: Pansu manifolds

One would like to extend the quasi-conformal theory on Carnot groups to more general manifolds. This can be done directly for spaces such as the torus $\mathbb{R}^n/\mathbb{Z}^n$ or the nilmanifold $\Phi_{\mathbb{Z}}^n \setminus \Phi^n$ with the induced metric. However, if one is to carry over the Pansu derivative to spaces that are not Carnot groups, it is convenient to have a theory of manifolds that is designed to accomodate it.

Recall the definition of a manifold: it is a topological space covered by countably many open sets homeomorphic to domains (*coordinate charts*) in \mathbb{R}^n . The transition maps between coordinate charts are then homeomorphisms, and topological invariants of \mathbb{R}^n (such as dimension) become topological invariants of the manifold. Stronger conditions on the transition maps (piecewise-linear, bi-Lipschitz, C^1 , C^∞) lead to corresponding classes of manifolds. For example, C^1 charts allow one to speak of the tangent space to the manifold, while C^∞ charts allow for all of differential geometry.

Definition 3.11.1. A Pansu manifold modeled on a Carnot group \mathbb{G} is a topological manifold M covered by countably many open sets homeomorphic to domains in \mathbb{G} such that the transition maps between the coordinate charts are continuously Pansu differentiable. That is, the Pansu derivative of the transition map is required to exist *everywhere* and to vary continuously.

We provide two examples of Pansu manifolds, beyond the obvious case of $M = \mathbb{G}$.

Example 3.11.2. Let M be a Pansu manifold and Γ a discrete torsion-free group acting on M properly discontinuously and by Pansu- C^1 mappings. Then the quotient space M/Γ is also a Pansu manifold, with the Pansu structure induced in the same way as the topological-manifold structure.

Example 3.11.3. The sub-Riemannian sphere \mathbb{S}^{2n+1} is a Pansu manifold modeled on Φ^n . Indeed, coordinates are provided by stereographic projection, with the transition map given by the Korányi inversion ι . The fact that ι is Pansu- C^1 where defined follows from Theorem 3.11.4. More generally, the Darboux theorem implies that every pseudoconvex submanifold of \mathbb{C}^{n+1} has an induced Pansu manifold structure modeled on Φ^n .

Theorem 3.11.4 (Magnani [32]). Let $U, V \subset G$ be domains in a Carnot group G, and $f: U \to V$ a (classically) C^1 mapping preserving the horizontal distribution HG. Then f is Pansu- C^1 .

It is unclear whether the converse holds. Namely,

Question 3.11.5. Does the converse to Theorem 3.11.4 hold? Equivalently, does every Pansu manifold have a natural smooth structure?

In view of Question 3.11.5 we do not delve deeply into Pansu manifolds. However, one can develop some notions parallel to the smooth setting. For example, every Pansu manifold M is automatically equipped with a Pansu tangent space $T^P M$ defined by gluing together the trivial bundles $U \times \mathfrak{g}$ on the coordinate charts (here, \mathfrak{g} is the Lie algebra of \mathbb{G} , on which M is modeled). Likewise, the splitting $\mathfrak{g} = \mathfrak{g}_1 \oplus \cdots \oplus \mathfrak{g}_s$ leads to a splitting $T_P M = T_1^P M \oplus \cdots \oplus T_s^P M$. Furthermore, vectors in $T_1^P M$ can be interpreted as equivalence classes of curves $\gamma : [0, \infty) \to M$ whose Pansu derivative at 0 lies in $T_1 M$.

As with Riemannian manifolds, one can give Pansu manifolds a metric by defining an inner product on $T_1^P M$, and considering the induced path metric (connectivity follows from the connectivity of G. If M happens to also be a smooth manifold (that is, the defining charts are smooth), the metric will agree with a sub-Riemannian metric on M (§3.1).

Lastly, we can speak of Pansu-smooth mappings $f: M \to N$ between Pansu manifolds, with the differential $d^P f: T^P M \to T^P N$ being a continuous gradingpreserving map between the tangent bundles. For bi-Lipschitz or conformal mappings between Pansu manifolds, and we can apply the Pansu–Rademacher theorem locally to obtain the existence of an almost-everywhere defined differential.

Chapter 4

Hyperbolic spaces

4.1 The hyperbolic plane

The hyperbolic plane is characterized as the unique (up to isometry) simplyconnected Riemannian two-manifold of constant curvature -1, and is the negatively curved counterpart to the curvature 0 Euclidean plane and curvature 1 sphere \mathbb{S}^2 . More concretely, the hyperbolic plane embeds in $\mathbb{C} \subset \mathbb{CP}^1$ as the *upper half-plane* $\operatorname{Im}(z) > 0$ with the line element $ds^2 = \frac{dx^2 + dy^2}{y^2}$. Since we are working with a complex model, we denote the hyperbolic plane by $\mathbb{H}^1_{\mathbb{C}}$.



Figure 4.1: The dyadic decomposition of the upper half-plane. The squares are isometric with respect to the hyperbolic metric.

Immediately, one sees that $\mathbb{H}^1_{\mathbb{C}}$ admits a homogeneous action by the group of isometries generated by the "parabolic" maps $z \mapsto z + a$, for $a \in \mathbb{R}$, and "hyperbolic maps" $z \mapsto rz$, for r > 0. In fact, any linear fractional transformation of \mathbb{CP}^1 preserving the upper half-plane is an isometry of $\mathbb{H}^1_{\mathbb{C}}$. These linear fractional transformations form the group $PSL(2,\mathbb{R})$, and account for all orientation-preserving isometries of $\mathbb{H}^1_{\mathbb{C}}$.

The isotropic nature of $\mathbb{H}^1_{\mathbb{C}}$ is not visible in the upper half-plane but becomes evident in the *Poincare disk* model of $\mathbb{H}^1_{\mathbb{C}}$, namely, the unit disk with the rotation-invariant metric $ds^2 = 4 \frac{dx^2 + dy^2}{(1 - x^2 - y^2)^2}$. The isometry from the upper half-plane to the unit disk is given by the *Cayley transform* $z \mapsto i \frac{z - i}{z + i}$.

Remark 4.1.1. Note the factor of 4 in the line element of the Poincare model. It is necessary to make the Poincare disk isometric to the upper half-plane and to normalize the space to have constant sectional curvature -1. This rescaling



Figure 4.2: The dyadic decomposition viewed in the unit disk via the Cayley transform.

will not be necessary for the *real hyperbolic* model of the hyperbolic plane, which leads to some lack of symmetry in $\mathbb{H}^n_{\mathbb{C}}$ for $n \geq 2$.

4.2 Hyperbolic lattices

Recall that a discrete subgroup Γ of a Lie group G is a *lattice* if $\Gamma \backslash G$ has finite volume. If $\Gamma \backslash G$ is compact, G is *uniform* or *cocompact*; otherwise it is *non-uniform*. Selberg's Lemma [43] states that if G is a matrix group, then Γ contains a finite-index subgroup Γ' with no finite-order elements, so that $\Gamma' \backslash G$ is a manifold.

In our case, the Lie group $G = PSL(2, \mathbb{R})$ acts on the hyperbolic plane by isometries, and lattices Γ in $PSL(2, \mathbb{R})$ yield tilings of $\mathbb{H}^1_{\mathbb{C}}$, as illustrated in Figure 4.3. Indeed, fix a point $p_0 \in \mathbb{H}^1_{\mathbb{C}}$ and consider the orbit $\Gamma \cdot p_0$ of p_0 under Γ . Consider the "nearest integer" map $p \mapsto [p]$ assigning to each point the closest point in the orbit $\Gamma \cdot p_0$ (if the closest point is unique). The fibers of $[\cdot]$ are called *Dirichlet fundamental domains* for the action of Γ on $\mathbb{H}^1_{\mathbb{C}}$, and the collection of disjoint tiles is invariant under the action of Γ . Note that the boundaries of the tiles are exactly the places where $[\cdot]$ is not uniquely defined.



Figure 4.3: Tilings associated with co-compact (left) and non-uniform (right) lattices.

One obtains a stronger relationship between the lattice Γ and the space on which it acts by giving Γ a metric. This is provided for finitely generated groups by the word metric:

Definition 4.2.1. Let Γ be a group generated by a finite set of elements Γ_0 . The *Cayley graph* of Γ with respect to Γ_0 is the graph whose vertices are exactly the elements of Γ and for which one has an edge between γ_1 and γ_2 if either $\gamma_1^{-1}\gamma_2 \in \Gamma_0$ or $\gamma_2^{-1}\gamma_1 \in \Gamma_0$. Assigning each edge of the Cayley graph unit distance, one obtains a metric graph and in particular the *word metric* on the vertices Γ .

The word metric a priori depends on the choice of generating set. However, any two word metrics are quasi-isometrically equivalent. That is, the identity map $id : (\Gamma, d_1) \rightarrow (\Gamma, d_2)$ distorts distances by a bounded amount (see Definition 2.2.5). Thus, we think of the word metric as defined only up to quasi-isometry.

The use of the word metric on co-compact lattices is justified by the Milnor– Švarc Theorem. Recall that a group acts geometrically on a metric space if the action is proper, free, and by isometries.

Theorem 4.2.2 (Milnor–Švarc [4] p. 140). Let Γ be a group acting geometrically and co-compactly on a metric space X. Then Γ is finitely generated and quasiisometric to X.

Sketch of proof. Fix $x_0 \in X$ and let $[\cdot]$ be the associated "nearest-integer" map from X to the orbit of Γ (or equivalently to Γ itself). Let K be the Dirichlet region containing x_0 . By co-compactness, the closure of K is compact. By properness, K has finitely many neighbors, say $\gamma_i K$. One shows that the elements γ_i generate Γ and furthermore that the adjacency graph of the tiling is the Cayley graph associated to these generators. Thus, one has interpreted the map $x \mapsto [x]$ as a quasi-isometry from X to Γ .



Figure 4.4: The group $SL(2,\mathbb{R})$ acting on a truncated space given by removing horoballs from $\mathbb{H}^1_{\mathbb{C}}$.

In the case of a non-uniform lattice Γ , the story is more complicated. One can once again define the Dirichlet region as the set of points closer to a base point x_0 than to other points in its orbit $\Gamma \cdot x_0$. However, this region is no longer compact, as it contains "cusps" escaping towards the circle bounding $\mathbb{H}^1_{\mathbb{C}}$ (see Figure 4.3).

The following theorem can be proven directly in $\mathbb{H}^1_{\mathbb{C}}$, but becomes harder in higher dimensions:

Theorem 4.2.3 (Thick-Thin Decomposition [23, 2]). Let Γ be a lattice in Isom(H), where H is a non-compact rank one symmetric space. Then there exists a collection of disjoint horoballs \mathcal{B} such that:

- 1. \mathcal{B} is invariant under Γ ,
- 2. Γ acts co-compactly on $H \setminus \cup \mathcal{B}$,
- 3. every parabolic element of Γ preserves a unique horoball in \mathcal{B} .

In the case of $\mathbb{H}^1_{\mathbb{C}}$, Theorem 4.2.3 states that every non-uniform lattice looks somewhat like Figure 4.4. Namely, one is able to find circles tangent to the boundary (called the *horocycles* or *horospheres*, with *horoball* interior) that can be removed to obtain a new metric space, the *truncated space* $\mathbb{H}^1_{\mathbb{C}} \setminus \mathcal{B}$, on which Γ acts geometrically and co-compactly. One concludes that Γ is finitely generated and quasi-isometric to the truncated space.

Theorem 4.2.3 is known as the Thick-Thin Decomposition since it provides the following description of the quotient $\Gamma \setminus H$: it consists of the *thick part* – namely the compact quotient of the truncated space $H \setminus \cup \mathcal{B}$ by Γ – and the *thin part* – one or more cups formed by taking the quotient of a horoball by a subgroup of Γ .

Note that every non-identity element of Γ fixing a horoball is *parabolic*. That is, it fixes exactly one point of $H \cup \partial H$ and the fixed point is in the boundary. The last part of Theorem 4.2.3 states that only the horoballs in \mathcal{B} are interesting when Γ is concerned: every parabolic element of Γ fixes one of the horoballs \mathcal{B} .

The proof of Theorem 4.2.3 is non-trivial. For $\mathbb{H}^1_{\mathbb{C}}$ it can be shown directly using hyperbolic geometry, but in the more general case it relies on the Margulis lemma. Informally, the result states that if some collection Γ_0 of isometries of H do not move some point very far and yet generate a discrete group, then the resulting group cannot be very complicated.

Theorem 4.2.4 (Margulis Lemma [2]). Let H be a non-compact rank-one symmetric space and Γ_0 a finite subset of Isom(H). Then there exists a Margulis constant $\epsilon = \epsilon(H)$ depending only on H such that if for some point x one has that $d(x, \gamma x) < \epsilon$ for every $\gamma \in \Gamma_0$ and furthermore Γ_0 generates a discrete group Γ , then Γ is virtually nilpotent.

4.3 Real hyperbolic space and horospherical coordinates

In §4.1, we built the hyperbolic plane $\mathbb{H}^1_{\mathbb{C}}$ by viewing the unit disk as a subset of \mathbb{CP}^1 and giving it a metric that is invariant under linear fractional transfor-

mations. We now build the real hyperbolic plane $\mathbb{H}^2_{\mathbb{R}}$ by viewing the unit disk as a subset of \mathbb{RP}^2 . By a coincidence, the two spaces $\mathbb{H}^1_{\mathbb{C}}$ and $\mathbb{H}^2_{\mathbb{R}}$ are isometric.

Recall that \mathbb{RP}^2 is the space of lines through the origin in \mathbb{R}^3 . A point in \mathbb{RP}^2 has homogeneous coordinates $(x_1 : x_2 : x_3)$, well-defined up to rescaling. One embeds $\mathbb{R}^2 \hookrightarrow \mathbb{RP}^2$ by sending (x, y) to (x : y : 1). The linear transformations $GL(3,\mathbb{R})$ induce the "linear fractional" transformations $PGL(3,\mathbb{R})$ of \mathbb{RP}^2 .

Via the above embedding of $\mathbb{R}^2 \hookrightarrow \mathbb{RP}^2$, the unit disk in \mathbb{R}^2 becomes the set of lines within the solid cone $x^2 + y^2 - z^2 < 0$. The set of linear transformations preserving this cone is the (Lorentz) orthogonal group O(2, 1). Indeed, these are exactly the matrices preserving the quadratic form

$$\langle (x_1, y_1, z_1), (x_2, y_2, z_2) \rangle = x_1 x_2 + y_1 y_2 - z_1 z_2.$$

We have thus provided the unit disk with a group PO(2, 1) of transformations. Because the stabilizer of the origin is exactly the orthogonal group O(2), we are able to find a metric for which these transformations are isometries: there exists a PO(2, 1)-invariant inner product on the unit disk that agrees with the standard inner product at the origin.

The resulting space is the *Klein model* of the hyperbolic plane (we denote it $\mathbb{H}^2_{\mathbb{R}}$), and is isometric to the Poincare unit disk model of $\mathbb{H}^1_{\mathbb{C}}$ via the transformation

$$(x,y) \mapsto \frac{2(x+iy)}{x^2+y^2+1}.$$
 (4.3.1)

Remark 4.3.1. The derivative of (4.3.1) is not the identity at the origin. This accounts for the different normalizations: at the origin the line element of the Klein model is $ds_K^2(0) = dx^2 + dy^2$, while for the Poincare model we have $ds_P^2(0) = 4(dx^2 + dy^2)$.

We now redraw figures from §4.2 in the real model. The geodesics in $\mathbb{H}^2_{\mathbb{R}}$ are, in fact, straight lines, but the metric is not conformally equivalent to that of the plane.



Figure 4.5: Co-compact and non-uniform lattices in $\mathbb{H}^2_{\mathbb{R}}$, and the associated truncated space. Compare Figures 4.3 and 4.4.

In §4.2, we defined horospheres in $\mathbb{H}^1_{\mathbb{C}}$ as circles tangent to the boundary of

 $\mathbb{H}^1_{\mathbb{C}}$. The definition is not intrinsic and furthermore not applicable to $\mathbb{H}^2_{\mathbb{R}}$. We provide two more general definitions, the second of which leads to the notion of *horospherical coordinates*.

Definition 4.3.2. Let X be a metric space. A *horosphere* is a limit $S_{\infty} = \lim_{r \to \infty} S_i$, where the S_i are metric spheres with radius increasing to infinity (with no restriction on centers), and the limit is "Hausdorff convergence on compacts". That is, for each compact $K \subset X$, the Hausdorff distance between $S_i \cap K$ and $S_{\infty} \cap K$ is required to converge to zero. A *horoball* is likewise the limit of arbitrarily large balls in X, and is bounded by a horosphere. (Of course, the trivial limits \emptyset and X are excluded in both definitions.)

Remark 4.3.3. In Euclidean space \mathbb{R}^n , all horospheres are flat embeddings of \mathbb{R}^{n-1} . To see this, think of inflating a balloon for infinite time, so that it converges to the tangent plane at the point where the air is coming in.

In non-compact rank-one symmetric spaces such as $\mathbb{H}^2_{\mathbb{R}}$, one can again think of horospheres as over-inflated balloons. Indeed, every horosphere is constructed by holding on to a sphere at one point and letting the other end go off to infinity along a geodesic. The following definition is equivalent to Definition 4.3.2:

Definition 4.3.4. Let X be a non-compact rank-one symmetric space and $\xi \in \partial X$ a point on the boundary. Denote by G_{ξ} the group of isometries of X fixing only ξ (as well as the identity mapping). This is the *parabolic subgroup* associated to ξ .

A horosphere is the orbit of some point $x \in X$ under the action of G_{ξ} , for some $\xi \in \partial X$. Let $\gamma_{x,\xi}$ be a geodesic ray from x to ξ . A horoball is the union $\sqcup_{g \in G_{\xi}} g \cdot \gamma_{x,\xi}$.

Definiton 4.3.4 allows us to define a new coordinate system for the space X (cf. the case of \mathbb{R}^n below and also Figure 4.7.4):

Definition 4.3.5. Fix a point $\xi \in \partial X$ and a geodesic $\gamma_{\xi}(s)$ ending in ξ . For each $s \in \mathbb{R}$, $G_{\xi} \cdot \gamma_{\xi}(s)$ is a horosphere, while for each $g \in G_{\xi}$, $g \cdot \gamma_{\xi}$ is a geodesic. In particular, the map $G_{\xi} \times \mathbb{R}^+ \to X$ given by $(g_{\xi}, y) \mapsto g_{\xi} \cdot \gamma_{\xi}(\log y)$ is bijective and provides the *horospherical coordinates* on X.

Example 4.3.6. Definitions 4.3.4 and 4.3.4 also work for \mathbb{R}^n if we think of $\partial \mathbb{R}^n$ if we take $\partial \mathbb{R}^n$ to be the set of geodesic rays starting at the origin, or equivalently the unit vectors at the origin. For a unit vector $\vec{v} \in \partial \mathbb{R}^n$, the parabolic subgroup $G_{\vec{v}}$ is the subset of $\operatorname{Isom}(\mathbb{R}^n)$ that preserves \vec{v} and no other vector - i.e. the translations perpendicular to \vec{v} . For a point $x \in \mathbb{R}^n$, the orbit $G_{\vec{v}} \cdot x$ is then a hyperplane perpendicular to \vec{v} . The half-space on the \vec{v} direction of the hyperplane is a horoball, and is foliated by geodesic rays in the \vec{v} direction. We thus obtain for n = 2 a somewhat unnatural representation of the Euclidean

plane as the upper half-plane with line element

$$ds^2 = dx^2 + \frac{dy^2}{y^2}.$$

Note that in \mathbb{R}^n the complement of a horoball is also a horoball; this is not true for hyperbolic spaces.

Example 4.3.7. In the case of the hyperbolic plane, parabolic subgroups are isomorphic to \mathbb{R} , and choosing an identification of G_{ξ} with \mathbb{R} turns the horospherical coordinates into the upper half-plane model, with ξ sent off to infinity.

Example 4.3.8. The construction of $\mathbb{H}^2_{\mathbb{R}}$ as the unit disk with a PO(2, 1)invariant metric immediately generalizes to the construction of $\mathbb{H}^n_{\mathbb{R}}$ as the unit
ball in $\mathbb{R}^n \subset \mathbb{RP}^{n+1}$ with a PO(n, 1)-invariant metric. As for $\mathbb{H}^2_{\mathbb{R}}$, the geodesics
of $\mathbb{H}^n_{\mathbb{R}}$ take the form of straight lines, and the isometries correspond precisely to PO(n, 1).

There does *not* exist a complex model of $\mathbb{H}^n_{\mathbb{R}}$ for $n \geq 3$. One can, however, use horospherical coordinates to write $\mathbb{H}^n_{\mathbb{R}}$ as the *upper half-space* $\mathbb{R}^{n-1} \times \mathbb{R}^+$ with the metric

$$ds^2 = \frac{dx_1^2 + \ldots + dx_n^2}{x_n^2}.$$

Alternately, one may use the mapping

$$(x_1, \dots, x_n) \mapsto \frac{2}{x_1^2 + \dots + x_n^2 + 1} (x_1, \dots, x_n)$$

to turn the Klein model of $\mathbb{H}^n_{\mathbb{R}}$ into a *Poincare ball* model, whose geodesics are the familiar circles perpendicular to the boundary.

4.4 Complex hyperbolic space

We are now ready to work with the *complex* hyperbolic plane $\mathbb{H}^2_{\mathbb{C}}$ (for additional details, see [16]). The construction is identical to that of $\mathbb{H}^2_{\mathbb{R}}$, except that one works with complex coordinates. Namely, $\mathbb{H}^2_{\mathbb{C}}$ is the unit ball in $\mathbb{C}^2 \hookrightarrow \mathbb{CP}^2$, with a metric that is invariant under the group of automorphisms of \mathbb{CP}^2 that preserves the unit ball.

The unit ball in \mathbb{C}^2 is defined by the equation $||(z_1, z_2)|| < 1$, where $||(z_1, z_2)||$ is given by the Hermitian form $\langle (z_1, z_2), (w_1, w_2) \rangle = \overline{z_1}w_1 + \overline{z_2}w_2$. Using the standard embedding $(z_1, z_2) \mapsto (z_1 : z_2 : 1)$ of \mathbb{C}^2 into \mathbb{CP}^2 , one exchanges the condition $||(z_1, z_2)|| < 1$ for the homogeneous condition

$$|z_1|^2 + |z_2|^2 - |z_3|^2 < 0,$$

so that, as in the real case, the unit ball in \mathbb{C}^2 is the projectivization of a cone in \mathbb{C}^3 . Likewise, the linear fractional transformations preserving the unit ball coincide with the projectivization PU(2,1) of the unitary group U(2,1) of linear transformations of \mathbb{C}^3 that preserve the above form.

We would now like to give the unit ball a metric for which PU(2,1) is an isometry group. The subgroup of PU(2,1) fixing the origin is the group U(2), a subgroup of O(2). We may therefore choose, at the origin, $ds^2(0) =$ $4(dx_1^2 + \cdots + dx_4^2)$ and extend it to a PU(2,1)-invariant inner product on the full ball. With this metric, the ball becomes the *complex hyperbolic space* $\mathbb{H}^2_{\mathbb{C}}$.

An analogous construction with 2 replaced by n provides us with complex hyperbolic space $\mathbb{H}^n_{\mathbb{C}}$.

Remark 4.4.1. One can go further and define, in addition to $\mathbb{H}^n_{\mathbb{R}}$ and $\mathbb{H}^n_{\mathbb{C}}$, analogous spaces over the quaternions, and the octonionic plane. Together, these form the *non-compact rank-one symmetric spaces*.

Remark 4.4.2. Note that the normalization $ds^2(0) = 4(dx_1^2 + \cdots + dx_4^2)$ agrees with the Poincare model, but not with the Klein model. As a result, the subspace $\{z_2 = 0\} \subset \mathbb{H}^2_{\mathbb{C}}$ is isometric to $\mathbb{H}^1_{\mathbb{C}}$, but the subspace $\{\operatorname{Re}(z_1) = \operatorname{Re}(z_2) = 0\} \subset$ $\mathbb{H}^2_{\mathbb{C}}$ is isometric to a *rescaled* version of $\mathbb{H}^2_{\mathbb{R}}$. Each of these is a totally geodesic subspace, the first with sectional curvature -1, and the second with sectional curvature -4. We thus have that $\mathbb{H}^2_{\mathbb{C}}$ has *pinched* curvature: every value in the range [-4, -1] is realized.

Recall from Definition 4.3.5 that the horospherical coordinates for $\mathbb{H}^n_{\mathbb{C}}$ are defined by a diffeomorphism from $G_{\xi} \times \mathbb{R}^+$ to $\mathbb{H}^n_{\mathbb{C}}$, where G_{ξ} is the group of isometries of $\mathbb{H}^n_{\mathbb{C}}$ fixing only the point ξ in the boundary of $\mathbb{H}^n_{\mathbb{C}}$.

Lemma 4.4.3. Let $\xi \in \partial \mathbb{H}^n_{\mathbb{C}}$ be a point on the boundary of \mathbb{H}^n_C . Then the parabolic subgroup G_{ξ} is isomorphic to Φ^{n-1} .

Proof. Any two Hermitian forms of signature (n, 1) on \mathbb{C}^{n+1} are related by a linear transformation. We can therefore switch from the Hermitian form $|z_1|^2 + \ldots + |z_n|^2 - |z_{n+1}|^2$ to the Hermitian form given by the matrix J_3 (see Definition 2.7.2). The ball becomes the "paraboloid" $-2\operatorname{Re}(z_n) = |z_1|^2 + \ldots + |z_{n-1}|^2$ (Figure 2.5), and we may further assume that the point ξ is mapped to the point "at infinity" with homogeneous coordinates $(0 : \ldots : 0 : -1)$.

One then sees the Siegel unitary model of the Heisenberg group within G_{ξ} , and furthermore that these are the only elements of G_{ξ} .

Corollary 4.4.4. The horospherical model of $\mathbb{H}^n_{\mathbb{C}}$ is given by $\Phi^{n-1} \times \mathbb{R}^+$.

The geometry of the horospherical model of $\mathbb{H}^n_{\mathbb{C}}$ is not as clear-cut as that of the upper half-space models of $\mathbb{H}^n_{\mathbb{R}}$. However, we can still make some observations.

Lemma 4.4.5. The isometries of the horospherical model of $\mathbb{H}^n_{\mathbb{C}} = \Phi^{n-1} \times \mathbb{R}^+$ are generated by the transformations

• $(p,s) \mapsto (f(p),s), \text{ for } f \in Isom(\Phi^{n-1}),$

• $(p,s) \mapsto (\delta_r p, rs), \text{ for } r \in \mathbb{R}^+,$

where $Isom(\Phi^{n-1})$ is the isometry group of Φ^{n-1} with respect to any of the metrics we have considered.

Proof. One shows using an explicit form of the distance function on $\mathbb{H}^n_{\mathbb{C}}$ that every isometry of $\mathbb{H}^n_{\mathbb{C}}$ is either an element of PU(n, 1), or the same followed by complex conjugation. The isometries of the horospherical model are those isometries that preserve the point in $\partial \mathbb{H}^n_{\mathbb{C}}$ used to define the model. Comparing the classification of these isometries with a classification of the isometries of Φ^{n-1} in Proposition 2.4.1 yields the result (note that the isometries of the Riemannian Φ^{n-1} are again those listed in Proposition 2.4.1).

Based on Lemma 4.4.5, we can write down a description of the Riemannian metric on the horospherical model.

Lemma 4.4.6. The restriction of the Riemannian metric tensor to each horosphere $\Phi^{n-1} \times \{s\}$ provides a Riemannian metric g_s on $\Phi^{n-1} = \Phi^{n-1} \times \{s\}$. Letting $s \to 0$, the rescaled metric $s^{-1}g_s$ limits to the sub-Riemannian metric g_{sR} ; indeed, the metrics $s^{-1}d_s$ are the Riemannian penalty metrics (Definition 2.2.12).

4.5 Gromov hyperbolicity and quasi-isometries

We defined $\mathbb{H}^n_{\mathbb{C}}$ as the unit ball with a PU(n, 1)-invariant metric and have described its horospherical model as $\Phi^{n-1} \times \mathbb{R}^+$. Both spaces come with natural boundaries, respectively the sphere \mathbb{S}^{2n-1} and the Heisenberg group $\Phi^{n-1} \times \{0\}$. How intrinsic are these boundaries? A straightforward answer for the ball model is given by the following lemma; compare Example 4.3.6.

Lemma 4.5.1. Let X be a non-compact rank-one symmetric space viewed in its ball model. Fix $x_0 \in X$. Then the exponential map $\exp_{x_0} : T_{x_0}X \to X$ is a diffeomorphism. Furthermore, \exp_{x_0} induces a bijection between unit vectors in $T_{x_0}X$ and points in ∂X by looking at the endpoint of the corresponding geodesic ray.

Proof. Consider first the case $x_0 = 0$. By symmetry and uniqueness of geodesics (the latter given by negative curvature) one sees that the geodesics emanating from x_0 are precisely line segments joining x_0 and a point on ∂X . Furthermore, such a line segment is immediately constructed to any point in ∂X . For $x_0 \neq 0$, one reduces to the previous case by an isometry. By construction, the isometry is given by a global homeomorphism of the whole ambient space, in particular the closure of the ball. Thus, one has the desired pencil of curves emanating from x_0 and terminating at each point on the boundary.

Lemma 4.5.1 allows us to identify the boundary of the ball and the unit tangent bundle at any point x_0 . The identification is made via geodesics starting at x_0 , and this point of view allows one to work with the boundary in a more general setting, namely that of Gromov hyperbolic spaces. We start by defining *geodesic* Gromov-hyperbolic spaces.

Definition 4.5.2. A geodesic metric space is δ -hyperbolic, for $\delta \geq 0$, if its triangles are δ -thin. That is, given three points a, b, c connected by geodesics [a, b], [b, c], [a, c], the δ -neighborhood of $[a, b] \cup [b, c]$ contains [a, c]. A geodesic space is *(Gromov) hyperbolic* if it is δ -hyperbolic for some δ .

Complete Riemannian manifolds of negative sectional curvature bounded away from zero, including the non-compact rank-one symmetric spaces, are Gromov hyperbolic. The hyperbolicity condition is both robust and powerful, as is demonstrated by the following two lemmas:

Lemma 4.5.3. Let $f : X \to Y$ be a quasi-isometry between geodesic spaces. Then X is hyperbolic if and only if Y is hyperbolic.

Remark 4.5.4. In Definition 4.6.1 we will broaden the definition of hyperbolic spaces to non-geodesic metric spaces. Note, however, that the condition in Lemma 4.5.3 that X and Y are geodesic can be loosened slightly, but not entirely removed. For example, let X be the union of the positive coordinate axes in \mathbb{R}^2 with the restricted metric function, and $Y = \mathbb{R}$. While the two spaces are quasi-isometric, only the second is Gromov-hyperbolic.

Definition 4.5.5. Two sets are *bounded (Hausdorff) distance from each other* if each set is contained in a bounded neighborhood of the other. The minimum size of the neighborhood is the Hausdorff distance (and satisfies the distance axioms if both sets are compact).

Lemma 4.5.6. Let $f : \mathbb{R} \to X$ be a quasi-isometric embedding from \mathbb{R} into a proper geodesic hyperbolic space X (a quasi-geodesic). Then there exists a geodesic $F : \mathbb{R} \to X$ such that $F(\mathbb{R})$ and $f(\mathbb{R})$ are bounded distance from each other.

Lemma 4.5.6 allows us to define the boundary ∂X of a Gromov-hyperbolic space in a way that is quasi-isometry invariant and easy to compute.

Definition 4.5.7. Let X be a geodesic Gromov-hyperbolic space. Then ∂X is the space of equivalence classes of quasi-geodesic rays in X, with two quasi-geodesic rays considered equivalent if they are bounded distance from each other.

Given a quasi-isometry $f : X \to Y$ between two hyperbolic spaces one immediately obtains a map on the set of equivalence classes of quasi-geodesics. We denote this by $\partial f : \partial X \to \partial Y$.

Remark 4.5.8. While the definition of the boundary via quasi-geodesic rays makes sense in non-hyperbolic spaces, it is not as useful there. For example,

 \mathbb{R}^2 contains bounded quasi-geodesic rays, and one would prefer to not think of them as going off to infinity. The definition also does not capture the desired intuition for the truncated spaces (seen in Figure 4.4), which are bounded by horospheres isometric to Euclidean space.

Definition 4.5.7 defines the boundary ∂X as a set. The standard compactopen topology on quasi-geodesics induces a topology on the boundary. We start with the following critical fact about quasi-isometries of hyperbolic spaces:

Theorem 4.5.9. Let $f : X \to Y$ be a quasi-isometric embedding of geodesic Gromov hyperbolic spaces. The induced boundary map $\partial f : \partial X \to \partial Y$ is then a continuous embedding. If f is furthermore a quasi-isometry, then ∂f is a homeomorphism.

Proof. The continuity is immediate from the use of the compact-open topology. The injectivity follows from the definition of equivalence: if two geodesic rays are not bounded distance in the original space, they cannot be bounded distance after a quasi-isometric embedding is applied to them. For quasi-isometries, a rough inverse provides existence and continuity of $(\partial f)^{-1}$.

The next lemma follows directly from the proof of Lemma 4.5.1:

Lemma 4.5.10. Topologically, the boundary at infinity of $\mathbb{H}^n_{\mathbb{R}}$ is \mathbb{S}^{n-1} , and the boundary at infinity of $\mathbb{H}^n_{\mathbb{C}}$ is \mathbb{S}^{2n-1} .

4.6 Quasi-isometries and quasi-symmetries

Theorem 4.5.9 states that every quasi-isometric embedding $f: X \to Y$ of Gromov hyperbolic geodesic spaces induces a continuous map on the boundary. Not every continuous map $F: \partial X \to \partial Y$ arises in this manner. To state this more precisely, we must put a family of metrics on the boundaries. It is convenient to simultaneously extend the class of Gromov hyperbolic spaces to include non-geodesic spaces.

Definition 4.6.1. Let X be a metric space (not necessarily geodesic). The Gromov product provides a quantitative measure of the excess of the triangle inequality:

$$(x,y)_z = \frac{1}{2}(|x-z| + |y-z| - |x-y|) \ge 0.$$

The space X is Gromov hyperbolic if the Gromov product satisfies a further inequality for some δ :

$$(x,y)_w \ge \min\{(x,z)_w, (z,y)_w\} - \delta_w$$

Remark 4.6.2. For geodesics spaces, the two definitions of hyperbolicity agree. While the specific values of δ for the definitions don't match up, each can be computed in terms of the other. For details, see [1]. Note, however, for non-geodesic spaces, hyperbolicity is not generally passed down through quasi-isometry (see Remark 4.5.4).

Let $\xi, \eta \in \partial X$, and $z \in X$. One extends the Gromov product by defining

$$(\xi,\eta)_z := \liminf(\xi(t),\eta(t))_z$$

where we make use of the identification of ξ and η with quasi-geodesic rays. A metric on ∂X is provided by the following lemma:

Lemma 4.6.3. Let X be a hyperbolic space and $z \in X$. Fix $\epsilon > 0$ and define

$$d_{\epsilon}(\xi,\eta) = e^{-\epsilon(\xi,\eta)_z}.$$

If X is hyperbolic, then there exists an ϵ_0 such that for $\epsilon \leq \epsilon_0$ the function d_{ϵ} is bi-Lipschitz to a metric.

The metrics on ∂X provided by Lemma 4.6.3 fall into the broader class of a *visual metrics*, for which a more robust equivalence is allowed.

Definition 4.6.4. Let X, Y be metric spaces (not necessarily hyperbolic). A topological embedding $f : X \to Y$ is a η -quasi-symmetric embedding, for an increasing bijection $\eta : \mathbb{R}^+ \to \mathbb{R}^+$, if one has for every $x, y, z \in X$ that

$$\frac{|fx - fy|}{|fx - fz|} \le \eta\left(\frac{x - y}{x - z}\right).$$

If f is furthermore a homeomorphism, we refer to it as a quasi-symmetry.

The visual metrics provided by Lemma 4.6.3 are all quasi-symmetrically equivalent. This leads to a broader collection of metrics on ∂X :

Definition 4.6.5. A metric on ∂X is in the *conformal gauge* if it is quasisymmetrically equivalent to a metric provided by Lemma 4.6.3.

In many spaces, including \mathbb{R}^n and Φ^n , quasi-conformal and quasi-symmetric maps coincide. More generally, quasi-conformality is a *local* version of quasisymmetry. In particular, quasi-symmetries must preserve the boundedness of a space, so that for example stereographic projection is 1-quasi-conformal but not quasi-symmetric.

The following is a classical result:

Theorem 4.6.6. Let $f : X \to Y$ be a quasi-isometric embedding. The $\partial f : \partial X \to \partial Y$ is a quasi-symmetric embedding, with respect to the conformal gauge. Furthermore, if f is a quasi-isometry, ∂f is a quasi-symmetry.

The converse is provided by multiple constructions, e.g. Bonk–Schramm [3].

Theorem 4.6.7. Let $F : \partial X \to \partial Y$ be a quasi-symmetry. Then there exists a quasi-isometry $f : X \to Y$ such that $\partial f = F$.

A related question is whether every space serves as the boundary of a hyperbolic space. Bonk–Schramm answer this affirmatively for complete bounded spaces via the following construction (for unbounded complete spaces serving as *parabolic* boundaries, see the more intricate construction in [6]).

Theorem 4.6.8. Let X be a complete metric space of diameter $D < \infty$. Then $X = \partial Y$ for a hyperbolic space $Y = X \times (0, D)$ with metric

$$|(x_1, t_1) - (x_2, t_2)| = 2\log\left(\frac{|x_1 - x_2| + \max\{t_1, t_2\}}{\sqrt{t_1 t_2}}\right)$$

We thus have that, up to a quasi-isometry, every hyperbolic space is given by the construction in Theorem 4.6.8. One proves Theorem 4.6.7 by interpreting a point $(x, t) \in Y$ as the ball B(x, t) in X having center x and radius t.

4.7 Quasi-isometric rigidity and quasi-conformal extension

Given an isometry f between two metric spaces, it is easy enough to perturb f on a bounded region to turn it into a (L, C)-quasi-isometry with $L \neq 1$ and $C \neq 0$. However, this tells us nothing about the structure of the spaces. Under what conditions on the metric spaces do "real" quasi-isometries not exist between them? In other words, when can one perturb every quasi-isometry to force L = 1 or C = 0?

Theorem 4.7.1 (Pansu [38]). Let X be a quaternionic hyperbolic space or the octonionic hyperbolic plane. Then every quasi-isometry $f : X \to X$ is bounded distance from an isometry.

Sketch of proof. The result follows from Pansu's differentiability theorem 3.6.4 and its Corollary 3.6.6. Namely, Pansu extends the quasi-isometry f to a quasiconformal map $\partial f : \partial X \to \partial X$, and shows that every quasi-conformal map on the boundary is in fact conformal. Since every conformal map is a Möbius transformation, we immediately obtain an isometry $F : X \to X$ satisfying $\partial F = \partial f$. It then follows from Lemma 4.5.6 that F and f are some bounded distance K apart, that is for every $x \in X$ we have $|f(x) - F(x)| < K < \infty$. \Box

In $\mathbb{H}^n_{\mathbb{R}}$ and $\mathbb{H}^n_{\mathbb{C}}$, a similar argument is unavailable. Indeed, there are many quasi-conformal self-mappings of \mathbb{R}^{n-1} and Φ^{n-1} (see Theorem 3.4.10), and therefore many non-trivial quasi-isometries of $\mathbb{H}^n_{\mathbb{R}}$ and $\mathbb{H}^n_{\mathbb{C}}$. One method for generating quasi-conformal mappings of \mathbb{R}^n was provided by Tukia–Väisälä:

Theorem 4.7.2 (Tukia–Väisälä [49]). Let $f : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ be a quasiconformal mapping (or quasi-symmetric if n = 2). Then there exists a quasiconformal mapping $F : \mathbb{R}^n \to \mathbb{R}^n$ such that $F|_{\mathbb{R}^{n-1}} = f$.

The essential step in the proof of Theorem 4.7.2 can be rephrased as:

Theorem 4.7.3 (Tukia–Väisälä [49]). Let $f : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ be a quasiconformal mapping (or quasi-symmetric if n = 2). Then there exists a bi-Lipschitz mapping $F : \mathbb{H}^n_{\mathbb{R}} \to \mathbb{H}^n_{\mathbb{R}}$ such that $\partial F = f$ (in the upper half-plane model).

Rephrasing further, Theorem 4.7.3 states that every quasi-isometry of $\mathbb{H}^n_{\mathbb{R}}$ is bounded distance from a bi-Lipschitz mapping. In this interpretation, Theorem 4.7.3 was extended to two classes of spaces including $\mathbb{H}^n_{\mathbb{C}}$ in [51] and [30]. We state our phrasing of the result for $\mathbb{H}^n_{\mathbb{C}}$:

Theorem 4.7.4 (Lukyanenko [30], cf. Xie [51]). Let $f : \Phi^{n-1} \to \Phi^{n-1}$ be quasiconformal, for $n \neq 2$. Then there exists a bi-Lipschitz map $F : \mathbb{H}^n_{\mathbb{C}} \to \mathbb{H}^n_{\mathbb{C}}$ such that in the horospherical model one has $\partial F = f$.

Sketch of proof. Recall that in horospherical coordinates we have $\mathbb{H}^n_{\mathbb{C}} = \Phi^{n-1} \times \mathbb{R}^+$. We say that a *lifting method* is a way of assigning to each homeomorphism $f : \Phi^{n-1} \to \Phi^{n-1}$ a homeomorphism $\widehat{f} : \mathbb{H}^n_{\mathbb{C}} \to \mathbb{H}^n_{\mathbb{C}}$ satisfying the additional properties:

- 1. Extension: \widehat{f} extends continuously to $\Phi^{n-1} \times \{0\}$ as f,
- 2. Continuity: $f \mapsto \hat{f}$ is continuous in the compact-open topology,
- 3. Equivariance: for any homotheties g_1, g_2 of Φ^{n-1} , we have

$$\widehat{g_1 \circ f \circ g_2} = g_1 \circ \widehat{f} \circ g_2,$$

where on the right the maps g_1 and g_2 are viewed as isometries of $\mathbb{H}^n_{\mathbb{C}}$.

Example 4.7.5. The Tukia–Väisälä lifting method views a point $(p, r) \in \Phi^{n-1} \times \mathbb{R}^+$ as a ball B(p, r) of radius r in Φ^{n-1} centered at p. One then defines $\widehat{f}(p, r) = (f(p), r')$, where r' is the inner radius of fB(p, r).

It will follow from the proof that any lifting method sends quasi-symmetries to quasi-isometries, largely due to the following lemma:

Lemma 4.7.6 (Corollary 10.30 of [18]). Fix a control function η and let \mathcal{F} be a family of η -quasi-symmetric embeddings of X in Y. Assume that Y is proper and there is $x_0 \in X$ and $y_0 \in Y$ such that $f(x_0) = y_0$ for each $f \in \mathcal{F}$. If

$$C^{-1} \le |f(a) - f(b)| \le C$$

for some $C \ge 1$, for some pair of points $a, b \in X$ and for all $f \in \mathcal{F}$, then \mathcal{F} is sequentially compact: every sequence in \mathcal{F} subconverges to an element of \mathcal{F} .

Returning to the theorem, we have a quasi-conformal map f and a lift \hat{f} , which is now a homeomorphism of $\mathbb{H}^n_{\mathbb{C}}$, but might not be bi-Lipschitz, perhaps not even locally. To iron out the wrinkles in \hat{f} , we invoke

Theorem 4.7.7 (PL approximation [40]). For $n \neq 4$, every homeomorphism of \mathbb{R}^n to itself is arbitrarily closely approximated by a piecewise-linear (hence locally bi-Lipschitz) homeomorphism.

Remark 4.7.8. Counter-examples exist in dimension 4.

We are thus able to adjust \hat{f} to a bi-Lipschitz mapping on any compact piece of $\mathbb{H}^n_{\mathbb{C}}$. We start by breaking $\mathbb{H}^n_{\mathbb{C}}$ up into pieces. More precisely, we create a Dyadic-style decomposition of $\mathbb{H}^n_{\mathbb{C}}$ by starting with a Strichartz tile K_S (cf. Theorem 2.1.7 and Figures 2.6 and 4.7.4), and turning it into a higherdimensional tile

$$K = K_S \times [1, 2] \subset \Phi^{n-1} \times \mathbb{R}^+ = \mathbb{H}^n_{\mathbb{C}}.$$

Let Γ_{α} be a set of isometries of $\mathbb{H}^{n}_{\mathbb{C}}$ such that the tiles $\Gamma_{\alpha} \cdot K$ tile $\mathbb{H}^{n}_{\mathbb{C}}$ as in Figure 4.7.4. By abuse of notation, we also think of Γ_{α} as a set of homotheties of Φ^{n-1} . Note that Γ_{α} is not a group.



Figure 4.6: Dyadic decomposition of $\mathbb{H}^2_{\mathbb{R}}$. An analogous construction is available in $\mathbb{H}^n_{\mathbb{C}}$.

We would now like to iron out the lifting \hat{f} on each tile γK , for $\gamma \in \Gamma_{\alpha}$. However, if we invoke Theorem 4.7.7 on two adjacent tiles, the resulting straightenings need not agree on the border. This problem is resolved by invoking Sullivan's deep Bi-Lipschitz Extension Theorem, which essentially states that at some cost it is possible to reconcile two bi-Lipschitz functions defined on overlapping domains (see [45, 48, 49]).

We are thus able to take our lifting \hat{f} and replace it, tile by tile, with a bi-Lipschitz mapping. Unfortunately, as we iron \hat{f} on infinitely many tiles, we lose control of the bi-Lipschitz constant. It is at this point that we invoke both the combinatorics of our tiling and the fact that our initial map f was a quasi-symmetry.

Note that instead of correcting \widehat{f} on each tile γK , we can correct $f \circ \gamma^{-1}$ on the tile K. Indeed, Lemma 4.7.6 implies that the mappings $\{f \circ \gamma^{-1}|_K : \gamma \in \Gamma_{\alpha}\}$ form a relatively compact family. That is, for each ϵ there is a finite collection of mappings f_1, \ldots, f_N such that we can "model" any mapping $f \circ \gamma^{-1}|_K$ by one of the f_1, \ldots, f_N . We apply Theorem 4.7.7 only to these mappings, obtaining corresponding mappings f'_1, \ldots, f'_N .

Consider now a maximal mutually disjoint collection of tiles γK . For each γK , consider $f \circ \gamma^{-1}|_K$ and pick the closest approximation $f_j \in \{f_1, \ldots, f_N\}$. Approximate $f|_{\gamma K}$ by $f_j \circ \gamma$. We have thus straightened f out on infinitely many tiles while using Theorem 4.7.7 only finitely many times.



Figure 4.7: A coloring of the dyadic decomposition, similar to the coloring of $\mathbb{H}^n_{\mathbb{C}}$ used in the proof of Theorem 4.7.4.

Consider next another mutually disjoint collection of tiles γK , also disjoint from the previous collection. We have to approximate f on each tile γK and also reconcile the approximation with the previous paragraph. Again, up to some error, we see only finitely many configurations and have to invoke PL approximation and Sullivan Extension only finitely many times.

One shows that the adjacency graph of the dyadic tiling is finitely colorable (Figure 4.7) so that a finite number of iterations of the previous paragraph completes the approximation of f on the full space $\mathbb{H}^n_{\mathbb{C}}$.

We thus have a uniformly-locally-bi-Lipschitz map F that was obtained from \hat{f} by arbitrarily small perturbations. Because \hat{f} is a homeomorphism, a sufficiently close approximate F must, in fact, be bi-Lipschitz.

4.8 Co-Hopficity of co-compact lattices

We now return to geometric group theory. Let Γ be a lattice in Isom(H) for a non-compact rank one symmetric space, say $\mathbb{H}^2_{\mathbb{R}}$. We saw in §4.2 that Γ must then be finitely generated and quasi-isometric to either all of H if $\Gamma \setminus H$ is compact, or to a truncated space $H \setminus \cup \mathcal{B}$ if Γ is non-uniform.

Definition 4.8.1. Let Γ be a group. One says that Γ is *co-Hopf* if every injective homomorphism $f : \Gamma \hookrightarrow \Gamma$ is in fact an isomorphism. Alternately, Γ is *almost co-Hopf* if every injective homomorphism $f : \Gamma \hookrightarrow \Gamma$ satisfies $[\Gamma : f(\Gamma)] < \infty$.

Example 4.8.2. We provide some examples of co-Hopficity:

- Every finite group is co-Hopf.
- The group (\mathbb{R} , +) is co-Hopf. Indeed, every homomorphism is either an isomorphism or the trivial map.
- The group $(\mathbb{Z}, +)$ is not co-Hopf, as one has the embedding f(x) = 2x. However, it is almost co-Hopf: for any non-trivial $f : \mathbb{Z} \hookrightarrow \mathbb{Z}$, we have $[\mathbb{Z} : f(\mathbb{Z})] = |f(1)|$.
- More generally, the free group \mathbb{F}_n is *not* co-Hopf. A generic set of elements $g_1, g_2, \ldots, g_k \in \mathbb{F}_n$ generates a proper free subgroup of \mathbb{F}_n isomorphic to \mathbb{F}_k . For $n \geq 2$, \mathbb{F}_n is not even almost co-Hopf: the map $f : \mathbb{F}_n \hookrightarrow \mathbb{F}_n$ that doubles the generators has infinite co-dimension.

We would now like to prove:

Theorem 4.8.3. Let H be a non-compact rank one symmetric space and $\Gamma \subset Isom(H)$ a lattice. If Γ is co-compact or if $H \neq \mathbb{H}^2_{\mathbb{R}}$, then Γ is almost co-Hopf.

In this section, we provide a proof of the co-compact case. In §4.9 we sketch the proof of the non-uniform case. Figure 4.8 shows that \mathbb{F}_2 is a non-uniform lattice in $\text{Isom}(\mathbb{H}^2_{\mathbb{R}})$ justifying the constraints in Theorem 4.8.3.



Figure 4.8: The free group on two generators is a lattice in $\text{Isom}(\mathbb{H}^2_{\mathbb{R}})$.

Before proving the co-compact case of Theorem 4.8.3, we rephrase the assertion in terms of geometry.

Definition 4.8.4. A space X is quasi-isometrically co-Hopf if every quasiisometric embedding $f: X \hookrightarrow X$ is, in fact, a quasi-isometry.

Lemma 4.8.5. Let Γ be a finitely generated group. If Γ is quasi-isometrically co-Hopf, then it is almost co-Hopf.

Proof. Let $f : \Gamma \hookrightarrow \Gamma$ be an injective homomorphism. Note that distances in Γ are measured along the edges of the Cayley graph for some generating set. Now,

for each generator γ , the image $f(\gamma)$ becomes a finite word in the generators, i.e. the distance from $f(\gamma)$ to the identity is bounded. Thus, *any* path in the Cayley graph of Γ is distorted by only a bounded amount. That is, f is a quasi-isometric embedding. By the assumed quasi-isometric co-Hopficity, f is furthermore a quasi-isometry and every point in Γ is some bounded distance Rfrom $f(\Gamma)$.

Consider now the coset space $\Gamma/f(\Gamma)$. For each coset, we can choose a representative that is within R of some point of $\gamma \in f(\Gamma)$ and then translate it by γ^{-1} so that it is within R of the identity. The resulting points have to be disjoint, and there are finitely many of them since the ball of radius R in the Cayley graph of Γ contains a finite number of vertices. Thus, Γ is almost co-Hopf. Note that if we are able to take R = 0, then Γ is co-Hopf. \Box

We now provide a proof of the following classical result:

Theorem 4.8.6. Let Γ be a co-compact lattice in Isom(H), with H a noncompact rank one symmetric space. Then Γ is almost co-Hopf.

Proof. By Lemma 4.8.5 it suffices to prove that Γ is quasi-isometrically co-Hopf. Note that quasi-isometric co-Hopficity is a quasi-isometry invariant, so that by Theorem 4.2.2 it suffices to prove that H is quasi-isometrically co-Hopf.

Consider now a quasi-isometric embedding $f : H \to H$ and let $n+1 = \dim H$. We have an induced topological embedding $\partial f : \partial H \to \partial H$. That is, we have an injective continuous map of \mathbb{S}^n to itself. If f is not a homeomorphism, then it misses a point, so that we would have an injective continuous map of \mathbb{S}^n into \mathbb{R}^n , but this is impossible.

We thus have that $\partial f : \partial H \to \partial H$ is a homeomorphism, indeed a quasisymmetry by Theorem 4.6.6. Let F^{-1} be a quasi-isometry of H satisfying $\partial(F^{-1}) = (\partial f)^{-1}$. We then have that $F^{-1} \circ f$ is a quasi-isometric embedding whose boundary map is the identity. But then $F^{-1} \circ f$ is bounded distance from the identity, and in particular itself a quasi-isometry. Therefore, f was a quasi-isometry in the first place.

4.9 Co-Hopficity of non-uniform lattices

In this section, we sketch a proof of the non-uniform case of Theorem 4.8.3. The co-compact case is given by Theorem 4.8.6. Analogously to Theorem 4.8.6, we phrase the non-uniform case as a geometric statement. For full details, see [22].

Theorem 4.9.1 (Kapovich–Lukyanenko [22]). Let Γ be a non-uniform lattice in Isom(H) for $H \neq \mathbb{H}^2_{\mathbb{R}}$ a non-compact rank-one symmetric space. Then every quasi-isometric embedding $f : \Gamma \hookrightarrow \Gamma$ is in fact a quasi-isometry. In particular, as a group Γ is almost co-Hopf. The proof of Theorem 4.9.1 starts by transitioning to a statement about truncated spaces. Recall from Theorem 4.2.3 that every non-uniform lattice is quasi-isometric to a truncated space $H \setminus \bigcup \mathcal{B}$, where \mathcal{B} is a family of pairwise disjoint horoballs invariant under Γ . Thus, Theorem 4.9.1 is equivalent to saying that a quasi-isometric embedding of a symmetric truncated space to itself is always a quasi-isometry (note that we don't claim this for all truncated spaces).

Because the large-scale geometry of truncated spaces is not well-understood (in particular, there is no meaningful notion of boundary at infinity), we make use of the large-scale geometry of H instead. Namely, we use the following lemma of Richard Schwartz:

Lemma 4.9.2 (R. Schwartz [41]). Let X be a horosphere in a rank-one symmetric space $H \neq \mathbb{H}^2_{\mathbb{R}}$ with its Riemannian metric (not the subspace metric), and $f: X \to H \setminus \cup \mathcal{B}$ a quasi-isometric embedding into a truncated space. Then there is a constant C depending only on H such that f(X) is within distance C of a boundary horosphere of $H \setminus \cup \mathcal{B}$.

Modifying the map f by at worst the constant C from Lemma 4.9.2, we have that the boundary horospheres of $H \setminus \bigcup \mathcal{B}$ get mapped by quasi-isometric embeddings to other boundary horospheres of the truncated space. We now extend f to a mapping F of all of H as follows.

Recall that a horoball B is foliated by a family of geodesic rays, each one starting from a point on the horosphere ∂B and terminating in the same point $\xi \in \partial H$. Now, for each $B_1 \in \mathcal{B}$, f sends ∂B to another horosphere ∂B_2 , so we have a way of matching up the geodesic rays foliating B_1 with the geodesic rays foliating B_2 . Define the extension F on B via this matching up of geodesic rays.

We now make use of another lemma of Richard Schwartz:

Lemma 4.9.3 (R. Schwartz [41]). Let $f : \partial B_1 \to \partial B_2$ be a quasi-isometry of Riemannian horospheres. Then $F : B_1 \to B_2$ defined by extending along geodesic rays is a quasi-isometry.

Remark 4.9.4. Lemma 4.9.3 only applies to quasi-isometries, and we only have a quasi-isometric embedding. This is resolved by Lemma 4.9.6 below.

Modulo Remark 4.9.4, we thus have transitioned from a quasi-isometric embedding $f: \Gamma \to \Gamma$ to a mapping $F: H \to H$ that is a quasi-isometric embedding on the truncated space $H \setminus \cup \mathcal{B}$ and a quasi-isometry from each $B \in \mathcal{B}$ onto its image horoball. We would like to combine these facts to say that $F: H \to H$ is a quasi-isometric embedding, but we only obtain:

Lemma 4.9.5. The extension $F : H \to H$ is a coarse embedding. That is, there exist control functions $\alpha(d)$ and $\omega(d)$ such that we have

$$\alpha(|x-y|) \le |fx - fy| \le \omega(|x-y|)$$

with $\alpha, \beta : \mathbb{R} \to \mathbb{R}$ increasing proper functions.

The proof of Theorem 4.9.1 now comes down to the following key lemma (which also resolves Remark 4.9.4:

Lemma 4.9.6 (Theorem 3.8 of [22]). Let X be a uniformly contractible uniformly acyclic metric space space homeomorphic to a disk (such as \mathbb{R}^n, Φ^n , or non-compact rank one symmetric spaces with the standard Riemannian metrics). Let $f: X \to X$ be a (α, ω) -coarse embedding. Then f is roughly onto, with the constant depending quantitatively on the space, α , and ω .

Sketch of proof. Using the uniform acyclicity of X, f can be replaced by a continuous map (consider a mesh in X and replace f by a piecewise-linear map defined by interpolating on the mesh). By uniform contractibility, the continuous map must then induce an isomorphism on the compactly supported cohomology. In particular, it must be surjective.

We thus have that the extended map $F: H \to H$ is roughly onto. It follows that the map $f: H \setminus \cup \mathcal{B} \to H \setminus \cup \mathcal{B}$ is roughly onto, and therefore that $f: \Gamma \to \Gamma$ is roughly onto.

This concludes the sketch of Theorem 4.9.1 and therefore of Theorem 4.8.3.

References

- J. M. Alonso, T. Brady, D. Cooper, V. Ferlini, M. Lustig, M. Mihalik, M. Shapiro, and H. Short. Notes on word hyperbolic groups. In *Group* theory from a geometrical viewpoint (Trieste, 1990), pages 3–63. World Sci. Publ., River Edge, NJ, 1991. Edited by Short.
- [2] R. Benedetti and C. Petronio. Lectures on hyperbolic geometry. Springer, 1992.
- [3] M. Bonk and O. Schramm. Embeddings of Gromov hyperbolic spaces. Geom. Funct. Anal., 10(2):266–306, 2000.
- [4] M. R. Bridson and A. Haefliger. Metric spaces of non-positive curvature, volume 319 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1999.
- [5] D. Burago, Y. Burago, and S. Ivanov. A course in metric geometry, volume 33 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2001.
- [6] S. Buyalo and V. Schroeder. Spaces of curvature bounded above. In Surveys in differential geometry. Vol. XI, volume 11 of Surv. Differ. Geom., pages 295–327. Int. Press, Somerville, MA, 2007.
- [7] L. Capogna. Regularity of quasi-linear equations in the Heisenberg group. Comm. Pure Appl. Math., 50(9):867–889, 1997.
- [8] L. Capogna, D. Danielli, S. Pauls, and J. T. Tyson. An introduction to the Heisenberg group and the sub-Riemannian isoperimetric problem, volume 259 of Progress in Mathematics. Birkhäuser Verlag, Basel, 2007.
- [9] L. Capogna and E. Le Donne. Smoothness of subriemannian isometries. arXiv preprint arXiv:1305.5286, 2013.
- [10] L. Capogna and P. Tang. Uniform domains and quasiconformal mappings on the Heisenberg group. *Manuscripta Math.*, 86(3):267–281, 1995.
- [11] P. Deligne and D. Sullivan. Fibrés vectoriels complexes à groupe structural discret. C. R. Acad. Sci. Paris Sér. A-B, 281(24):Ai, A1081–A1083, 1975.
- [12] E. Falbel, G. Francsics, P. Lax, and J. Parker. Generators of a Picard modular group in two complex dimensions. *Proc. Amer. Math. Soc.*, 139(7):2439– 2447, 2011.
- [13] K. Fässler, A. Lukyanenko, and K. Peltonen. Quasiregular mappings on sub-riemannian manifolds. *preprint arXiv:1312.0271*, 2013.

- [14] J. Ferrand. The action of conformal transformations on a riemannian manifold. Mathematische Annalen, 304(1):277–291, 1996.
- [15] F. Gehring. Rings and quasiconformal mappings in space. Transactions of the American Mathematical Society, pages 353–393, 1962.
- [16] W. Goldman. Complex hyperbolic geometry. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 1999. Oxford Science Publications.
- [17] U. Hamenstädt. Some regularity theorems for Carnot-Carathéodory metrics. J. Differential Geom., 32(3):819–850, 1990.
- [18] J. Heinonen. Lectures on Analysis on Metric Spaces. Springer, New York, 2000.
- [19] J. Heinonen. Lectures on Lipschitz analysis, volume 100 of Report. University of Jyväskylä Department of Mathematics and Statistics. University of Jyväskylä, Jyväskylä, 2005.
- [20] J. Heinonen and S. Semmes. Thirty-three yes or no questions about mappings, measures, and metrics. *Conform. Geom. Dyn.*, 1:1–12 (electronic), 1997.
- [21] T. Iwaniec and G. Martin. Quasiregular semigroups. Ann. Acad. Sci. Fenn. Math., 21(2):241–254, 1996.
- [22] I. Kapovich and A. Lukyanenko. Quasi-isometric co-Hopficity of nonuniform lattices in rank-one semi-simple Lie groups. *Conform. Geom. Dyn.*, 16:269–282, 2012.
- [23] M. Kapovich and C. Druţu. Lectures on the Geometric Group Theory. 2013.
- [24] A. Korányi and H. Reimann. Quasiconformal mappings on the Heisenberg group. *Invent. Math.*, 80(2):309–338, 1985.
- [25] A. Korányi and H. Reimann. Foundations for the theory of quasiconformal mappings on the Heisenberg group. Adv. Math., 111(1):1–87, 1995.
- [26] E. Le Donne. A metric characterization of Carnot groups. arXiv preprint arXiv:1304.7493, 2013.
- [27] E. Le Donne, A. Ottazzi, and B. Warhurst. Ultrarigid tangents of sub-Riemannian nilpotent groups. arXiv preprint arXiv:1104.3071v2, 2011.
- [28] J. Lelong-Ferrand. Geometrical interpretations of scalar curvature and regularity of conformal homeomorphisms. In *Differential geometry and relativity*, pages 91–105. Springer, 1976.
- [29] P. Libermann. Sur les automorphismes infinitésimaux des structures symplectiques et des structures de contact. In *Colloque Géom. Diff. Globale* (*Bruxelles, 1958*), pages 37–59. Centre Belge Rech. Math., Louvain, 1959.
- [30] A. Lukyanenko. Bi-Lipschitz extension from boundaries of certain hyperbolic spaces. *Geom. Dedicata*, 164:47–71, 2013.
- [31] A. Lukyanenko and J. Vandehey. Continued fractions on the Heisenberg group. arXiv preprint arXiv:1305.5286, 2013.

- [32] V. Magnani. Towards differential calculus in stratified groups. J. Aust. Math. Soc., 95(1):76–128, 2013.
- [33] G. Margulis and G. Mostow. The differential of a quasi-conformal mapping of a Carnot-Carathéodory space. *Geom. Funct. Anal.*, 5(2):402–433, 1995.
- [34] J. Mitchell. On Carnot-Caratheodory metrics. J. Differential Geom., 21(1):35–45, 1985.
- [35] G. Mostow. Strong rigidity of locally symmetric spaces. Princeton University Press, Princeton, N.J., 1973. Annals of Mathematics Studies, No. 78.
- [36] S. Myers and N. Steenrod. The group of isometries of a riemannian manifold. Annals of Mathematics, pages 400–416, 1939.
- [37] P. Pansu. An isoperimetric inequality on the Heisenberg group. Rend. Sem. Mat. Univ. Politec. Torino, (Special Issue):159–174 (1984), 1983. Conference on differential geometry on homogeneous spaces (Turin, 1983).
- [38] P. Pansu. Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un. Ann. of Math. (2), 129(1):1–60, 1989.
- [39] S. Rickman. Quasiregular mappings, volume 26 of Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]. Springer-Verlag, Berlin, 1993.
- [40] T. B. Rushing. *Topological embeddings*. Academic Press, New York, 1973. Pure and Applied Mathematics, Vol. 52.
- [41] R. E. Schwartz. The quasi-isometry classification of rank one lattices. Inst. Hautes Études Sci. Publ. Math., (82):133–168 (1996), 1995.
- [42] F. Schweiger. Multidimensional continued fractions. Oxford Science Publications. Oxford University Press, Oxford, 2000.
- [43] A. Selberg. On discontinuous groups in higher-dimensional symmetric spaces. In Contributions to function theory (internat. Colloq. Function Theory, Bombay, 1960), pages 147–164. Tata Institute of Fundamental Research, Bombay, 1960.
- [44] R. Strichartz. Self-similarity on nilpotent Lie groups. In *Geometric analysis* (*Philadelphia, PA, 1991*), volume 140 of *Contemp. Math.*, pages 123–157. Amer. Math. Soc., Providence, RI, 1992.
- [45] D. Sullivan. Hyperbolic geometry and homeomorphisms. In Geometric topology (Proc. Georgia Topology Conf., Athens, Ga., 1977), pages 543– 555, 1979.
- [46] P. Tukia. On quasiconformal groups. J. Analyse Math., 46:318–346, 1986.
- [47] P. Tukia and J. Väisälä. Lipschitz and quasiconformal approximation and extension. Ann. Acad. Sci. Fenn. Ser. A I Math., 6(2):303–342 (1982), 1981.
- [48] P. Tukia and J. Väisälä. Lipschitz and quasiconformal approximation and extension. Ann. Acad. Sci. Fenn. Ser. A I Math., 6(2):303–342 (1982), 1981.
- [49] P. Tukia and J. Väisälä. Quasiconformal extension from dimension n to n + 1. Ann. of Math. (2), 115(2):331–348, 1982.
- [50] A. Varčenko. Obstructions to local equivalence of distributions. Mat. Zametki, 29(6):939–947, 957, 1981.
- [51] X. Xie. Quasiisometries between negatively curved Hadamard manifolds. J. Lond. Math. Soc. (2), 79(1):15–32, 2009.