UNIVERSITY OF JYVÄSKYLÄ
DEPARTMENT OF MATHEMATICS
AND STATISTICS

REPORT 163

UNIVERSITÄT JYVÄSKYLÄ
INSTITUT FÜR MATHEMATIK
UND STATISTIK

BERICHT 163

# TOPICS IN THE GEOMETRY OF NON-RIEMANNIAN LIE GROUPS

SEBASTIANO NICOLUSSI GOLO

JYVÄSKYLÄ
2017

UNIVERSITY OF JYVÄSKYLÄ
DEPARTMENT OF MATHEMATICS
AND STATISTICS

REPORT 163

UNIVERSITÄT JYVÄSKYLÄ
INSTITUT FÜR MATHEMATIK
UND STATISTIK

BERICHT 163

# TOPICS IN THE GEOMETRY OF NON-RIEMANNIAN LIE GROUPS

**SEBASTIANO NICOLUSSI GOLO**

To be presented, with permission of the Faculty of Mathematics and Science of the University of Jyväskylä, for public criticism in Auditorium MaA 211 on September 15th, 2017, at 13 o'clock noon.

JYVÄSKYLÄ
2017

Editor: Pekka Koskela
Department of Mathematics and Statistics
P.O. Box 35 (MaD)
FI–40014 University of Jyväskylä
Finland

ISBN 978-951-39-7148-9 (print)
ISBN 978-951-39-7149-6 (pdf)
ISSN 1457-8905



University Printing House
Jyväskylä 2017

## Acknowledgements

This dissertation is the final stage of a long, sometimes harsh, path from being a student to being a researcher. Whether I am a good or a bad researcher will be unveiled in the future, but the fruits I collected along the past years will remain my personal treasure. None of this could be possible without the involvement of the dozens of people I met. I thank them all for having been nice to me.

I wish to thank the staff of both the University of Jyväskylä and the University of Trento for helping me to fulfill my duties despite my desperate disorganization. I want to mention the financial support I received from the Department of Mathematics and Statistics of Jyväskylä and later from MAnET, a Marie Curie Initial Training Network. I thank also Raul P. Serapioni, who has been my reference person in the latter project.

I thank all my collaborators: Michael G. Cowling, Ville Kivijoja, Alessandro Ottazzi and Andrea Sambusetti. I hope we will keep working together in the future.

I wish to thank my advisor Francesco Serra Cassano for guiding me in the hard problems of the GMT on Carnot groups. His passion and his expertise have been fundamental for building up a big part of my research. I wish his counsel shall continue to be a precious part of my work.

This dissertation couldn't have been made without the work of my advisor Enrico Le Donne. His guidance has been immensely important for me. His patience, his thoughtful moderation, his brilliant attention to both the details and the greater picture, his courage to tackle the unknown, his ability to stay focused for endless hours, lay in front of me as a bright example of what I should aim at. I have tried and I will try to follow his example.

I am deeply grateful to my grandmother Nonna Sandra, to my parents Antonio and Ester, and to my siblings Dorotea and Giovanni.

I dedicate this dissertation to my grandparents, to my grand-grandparents, and to all past generations who suffered fatigue, famine, war and coercion.

Jyväskylä, September 2017

Sebastiano Nicolussi Golo.

## List of included articles

This dissertation consists of an introduction and the following articles:

[A] E. Le Donne and S. Nicolussi Golo. “Regularity properties of spheres in homogeneous groups”. In: *Transactions of the American Mathematical Society* (To appear). arXiv: 1509.03881.

[B] E. Le Donne, S. Nicolussi Golo, and A. Sambusetti. “Asymptotic behavior of the Riemannian Heisenberg group and its horoboundary”. In: *“Annali di Matematica Pura ed Applicata (1923 -)”* (2016), pp. 1–22. arXiv: 1509.00288.

[C] M. G. Cowling, V. Kivioja, E. Le Donne, S. Nicolussi Golo, and A. Ottazzi. “From homogeneous metric spaces to Lie groups”. In: *ArXiv: 1705.09648* (May 2017).

[D] S. Nicolussi Golo. “Some remarks on contact variations in the first Heisenberg group”. In: *ArXiv: 1611.07358* (Aug. 2017).

The author of this dissertation has actively taken part in the research of the joint papers.

# INTRODUCTION

Contents



## 1. Regularity of subFinsler distances

SubRiemannian geometry, as part of Geometry and Analysis on Metric Spaces, has several historical roots. Mostow's Rigidity Theorem ignited the interest in quasi-conformal mappings on metric spaces, [31, 32, 17, 35, 10, 20, 21, 39]. The characterization of groups with polynomial growth posed a bridge between Abstract Algebra and Metric Geometry on Lie groups, [42, 6, 19, 18, 34, 7, 8]. The study of hypoelliptic operators found deep connections with the geometry of subRiemannian metrics, [22, 41, 38, 33]. Finally, the point of view of Control Theory plays an important role in the study of the regularity of such distances and their geodesics, [1, 3].

Unlike on Riemannian manifolds, in subRiemannian geometry topological and Hausdorff dimensions may differ, the square of the distance function may be not smooth near the diagonal and small spheres may be not smooth. We are interested in such phenomena on subFinsler manifolds, the general notion of Carnot-Carathéodory spaces.

Let $G$ be a smooth manifold, $\Delta \subset TG$ a bracket-generating subbundle and $\|\cdot\| : \Delta \to [0,+\infty)$ a continuous function whose restriction to each fiber of $\Delta$ is a norm. An *horizontal curve* is an absolutely continuous curve $\gamma : [0,1] \to G$ such that $\gamma'(t) \in \Delta$ for almost all $t \in [0,1]$. The *length of a horizontal curve* $\gamma$ is $\ell(\gamma) := \int_0^1 \|\gamma'(t)\| \,\mathrm{d}t$. The *subFinsler distance*, or *Carnot-Carathéodory distance*, is defined for $p, q \in G$ as

$$d_{CC}(p,q) := \inf \{\ell(\gamma) : \ \gamma \text{ is a horizontal curve from } p \text{ to } q\} .$$

If the norm $\|\cdot\|$ is induced by a smooth scalar product on $G$, then $d_{CC}$ is said to be a *subRiemannian distance*.

The space of all Lipschitz curves $\gamma : [0,1] \to (G, d_{CC})$ has a natural structure of Banach manifold and the *End-point map* is the smooth map $\gamma \mapsto \gamma(1)$.

A curve is *regular*, respectively *singular*, if it is a regular, respectively singular, point of the End-point map. SubRiemannian regular length-minimizers behave like Riemannian geodesics. For instance, they are smooth. Moreover, if $p \in M$ and $\Sigma \subset M$ is the set of points $q \in M \setminus \{o\}$ such that there is a unique length minimizer from $p$ to $q$ and such curve is regular, then $\Sigma$ is open dense in $M$ and $q \mapsto d_{CC}(p, q)$ is smooth on $\Sigma$, see [2]. We extended the subRiemannian result to a very general class of subFinsler manifolds.

**Theorem 1.1** (with Le Donne, [A]). *Let $(G, d_{CC})$ be a subFinsler manifold of constant-type norm. Fix $p \in G$ and let $q \in G$ be such that all the length-minimizing curves from $p$ to $q$ are regular. Then there are a neighborhood $U$ of $q$ and a Riemannian distance $\rho$ on $U$ such that the map $x \mapsto d_{CC}(p, x)$ is Lipschitz on $U$ with respect to $\rho$.*

A subFinsler structure $(G, \Delta, \|\cdot\|)$ is *of constant-type norm* if there is a normed vector space $(\mathbb{E}, |\cdot|)$ and a bundle morphism $\mathtt{f} : G \times \mathbb{E} \to TG$ such that $\mathrm{Im}(\mathtt{f}) = \Delta$ and for every $v \in \Delta_p$ we have $\|v\| = \inf\{|e| : e \in \mathbb{E},\ \mathtt{f}(p, e) = v\}$.

In the subRiemannian case it is possible to prove a very strong result for spheres as well. Indeed, if all length minimizers are regular, then for almost every $r > 0$ the sphere centered at $p$ and of radius $r$ in $(G, d_{CC})$ is a Lipschitz hypersurface, see [36]. By *Lipschitz hypersurface* we mean a topological hypersurface $S \subset G$ such that for every $x \in S$ there is a neighborhood $U$ of $x$ so that $S \cap U$ is the graph of a Lipschitz function in some coordinates on $U$.

Such a property is a consequence of the semi-concavity of the subRiemannian distance in absence of singular geodesics. However, general Carnot-Carathéodory distances are not quasi-concave, as simple examples of norm distances on $\mathbb{R}^2$ show. Using the homogeneity of the distance, we proved Lipschitz regularity for spheres in Carnot groups. A subFinsler manifold $(G, d_{CC})$ is a *Carnot group* if $G$ is a simply connected Lie group whose Lie algebra $\mathfrak{g}$ admits a *stratification of step $s$* (i.e., a decomposition $\mathfrak{g} = \bigoplus_{i=1}^{s} V_i$ with $[V_1, V_i] = V_{i+1}$ for $i \in \{1, \dots, s-1\}$ and $[V_1, V_s] = \{0\}$), $\Delta$ is the left-invariant subbundle induced by the first stratum $V_1$ and $\|\cdot\|$ is left-invariant. It follows that $d_{CC}$ is left-invariant on $G$.

**Theorem 1.2** (with Le Donne, [A]). *Let $(G, d_{CC})$ be a Carnot group without non-constant singular length-minimizers. Fix a left-invariant Riemannian metric $\rho$ on $G$ and an open neighborhood $U$ of $0$. Then*

*(1) the step of $G$ is $1$ or $2$;*
*(2) the function $x \mapsto d_{CC}(0, x)$ is Lipschitz with respect to $\rho$ on $G \setminus U$;*
*(3) the function $x \mapsto d_{CC}(0, x)^2$ is Lipschitz with respect to $\rho$ on $U$;*
*(4) spheres are Lipschitz hypersurfaces.*

In this setting, we proved that on Carnot groups the absence of singular geodesics is equivalent to other three well-known properties. See Appendix in [A].

## 2. Self-similar Lie groups

In the study of Carnot groups, many propositions are proven thanks to the interaction between dilations and distance. Moreover, metric spaces endowed with dilations may arise as tangent cones or asymptotic cones. For these reasons we started to investigate self-similar Lie groups.

A *self-similar Lie group* is defined as a metric Lie group $(G, d)$ that admits a metric dilation that is an automorphism. More precisely, $(G, d)$ is a self-similar Lie group if $G$ is a Lie group, $d$ a left-invariant distance inducing the manifold topology and there are an automorphism $\delta : G \to G$ and a scalar $\lambda > 1$ such that $d(\delta p, \delta q) = \lambda d(p, q)$ for all $p, q \in G$.

In the paper [C], we gave a metric characterization of self-similar Lie groups. See also [26] for a similar characterization of Carnot groups.

**Theorem 2.1** (with Cowling, Kivioja, Le Donne, Ottazzi, [C])**.** *The self-similar Lie groups are the only metric spaces that are*

*(1) locally compact,*
*(2) connected,*
*(3) isometrically homogeneous,*
*(4) and admit a metric dilation.*

Self-similar Lie groups have a very rich structure, as we now recall. Following [27], a *graduable Lie group* is a simply connected Lie group $G$ whose Lie algebra $\mathfrak{g}$ admits a *grading*, i.e., a decomposition $\mathfrak{g} = \bigoplus_{i\in(0,+\infty)} V_i$ such that for all $i, j > 0$ it holds $[V_i, V_j] \subset V_{i+j}$. Graduable Lie groups are nilpotent and in particular $G$ is diffeomorphic to $\mathfrak{g}$ via the exponential map. If the grading is fixed, we say that $G$ is a *graded Lie group*. Stratified Lie groups are examples of graded Lie groups. Thanks to [40], self-similar Lie groups are graduable Lie groups with a grading induced by the metric dilation. A graded Lie group is endowed with *standard dilations* defined as the Lie group automorphisms $\delta_t : G \to G$ such that $(\delta_t)_* : \mathfrak{g} \to \mathfrak{g}$ is the map $(\delta_t)_*(v) = t^i v$, for all $v \in V_i$ and $t > 0$.

Together with my advisor, we extended some of the results obtained for Carnot groups to self-similar Lie groups.

**Theorem 2.2** (with Le Donne, [A])**.** *Let $G$ be a graded group. For $p \in G$, define $\bar\delta(p) := \frac{\mathrm{d}}{\mathrm{d}t}|_{t=1}\delta_t p \in T_pG$, where $\delta_t : G \to G$ are the standard dilations. Let $d$ be a left-invariant distance on $G$ such that $d(\delta_t q, \delta_t p) = td(q, p)$ for all $p, q \in G$ and all $t > 0$. Let $p \in G$ be such that*

$$\mathrm{d}L_p(V_1) + \mathrm{d}R_p(V_1) + \mathrm{span}\{\bar\delta(p)\} = T_pG. \tag{1}$$

*Then there is a neighborhood $U$ of $p$ such that $x \mapsto d(0, x)$ is Lipschitz on $U$ and the sphere of radius $d(0, p)$ is a Lipschitz hypersurface in $U$.*

In the case $G$ is a stratified group, condition (1) implies that all horizontal curves from 0 to $p$ are regular. Indeed, the left-hand side of (1) is contained in the image of the differential of the End-point map, because $\bar\delta$ is a contact vector field, see [28]. In this respect, Theorem 2.2 is a generalization of Theorem 1.2.

## 3. From homogeneous metric spaces to Lie groups

The metric characterization for self-similar Lie groups of Theorem 2.1 is part of a broader effort to characterize nilpotent metric Lie groups. If two nilpotent metric Lie groups are isometric, then they are isomorphic, see [24]. It still remains unknown if a similar result can be obtained with less regular maps, for example biLipschitz homeomorphisms or quasi-isometries. For instance, it is known that quasi-isometric Carnot groups are isomorphic as stratified Lie groups.

We recall that, given $L, C > 0$, a function $f : M \to N$ is a *(L,C)-quasi-isometry* between metric spaces $(M, d)$ and $(N, d)$ if $\frac{1}{L} d(x, y) - C \leq d(f(x), f(y)) \leq L d(x, y) + C$ for all $x, y \in M$, and every point in $N$ lies at most at distance $C$ from the image of $f$.

In the paper [C], we prove that solvable metric Lie groups represent, up to quasi-isometry, a large class of isometrically homogeneous metric spaces.

**Theorem 3.1.** *Let $(M, d)$ be a metric space that is*

*(1) locally compact,*
*(2) connected,*
*(3) isometrically homogeneous.*

*Then it is (1,C)-quasi-isometric to a simply connected solvable metric Lie group $(H, d_H)$, for some $C > 0$.*

As a consequence, any connected metric Lie group is quasi-isometric to a solvable metric Lie group.

## 4. Asymptotic estimates

Carnot groups have an important role as asymptotic cones of finitely generated groups with polynomial growth. Part of my research on Carnot groups is motivated by the aim of better understanding the asymptotic behaviour of such discrete groups.

If $\Gamma$ is a group generated by a finite and symmetric set $S$ containing the neutral element $e$, we say that $\Gamma$ has *polynomial growth* if it holds

$$\exists Q, V > 0, \ \forall k \in \mathbb{N} \qquad \#(S^k) \leq V k^Q. \tag{2}$$

Here $S^k = \{s_1 \cdots s_k : s_i \in S\}$ is the ball of radius $k$ and center $e$ for the left-invariant word distance $\rho_S$ induced by $S$. The property of being of polynomial growth and the exponent $Q$ do not depend on the generating set.

A celebrated theorem in Geometric Group Theory (see [42, 6, 18]), states that a finitely generated group has polynomial growth if and only if it has a nilpotent subgroup with finite index. The key point is that the asymptotic cone of such $(\Gamma, \rho_S)$ is a subFinsler Carnot group. The *asymptotic cone* of $(\Gamma, \rho_S)$ is the limit in Gromov-Hausdorff sense of the pointed metric spaces

$(\Gamma, \frac{1}{n}\rho_S, e)$, where $e$ is the neutral element of $\Gamma$ and $n \to \infty$. Similar results can be obtained in the class of compactly generated groups, see [19, 7].

A further analysis gives also an interpretation for the optimal constants $V$ and $Q$, see [34]. We will describe such interpretation in a subclass of finitely generated groups, see [8] and [7] for a general discussion. Let $G$ be a stratified Lie group and let $S \subset G$ be a finite symmetric set generating a cocompact subgroup $\Gamma \subset G$. Let $\|\cdot\|_\infty$ be the norm on the first stratum $V_1$ of $\mathfrak{g}$ whose unit ball is the convex hull of the projection of $\exp^{-1}(S)$ into $V_1$ along $\bigoplus_{i=2}^s V_i$. Let $d_\infty$ be the corresponding subFinsler distance on $G$. Then $(G, d_\infty)$ is a Carnot group and it is the asymptotic cone of $(\Gamma, \rho_S)$. Let $Q$ be the Hausdorff dimension of $(G, d_\infty)$, i.e., $Q = \sum_{i=1}^s i \dim(V_i)$, and $V$ the volume of the unit ball in $(G, d_\infty)$ with respect to the Haar measure on $G$ such that the lattice $\Gamma$ has co-volume 1. Then the property (2) improves into

$$\#(S^k) = Vk^Q + o(k^Q). \tag{3}$$

There are three kinds of open problems. First, one would like to have a stronger estimate for the error term in (3); for instance, one could ask whether $o(k^Q)$ can be replaced with $O(k^{Q-1})$. Second, one would like to approximate, quantitatively, the word distance $\rho_S$ with a subFinsler distance $d_S$ on $G$; usually, one consider the Stoll distance $d_S$, that is, the subFinsler distance on $G$ whose norm is induced by the convex hull of $S$. Third, one would like to estimate the difference $|d_1 - d_2|$ between two asymptotically equivalent left-invariant subFinsler distances on $G$; in particular, we are interested in $|d_S - d_\infty|$. In [8] it is shown that, if $d_1$ and $d_2$ are subFinsler distances on a stratified group $G$ of step $s$, then the following statements are equivalent:

(i) $\lim_{p\to\infty} \frac{d_1(0,p)}{d_2(0,p)} = 1$, i.e., $d_1$ and $d_2$ are asymptotically equivalent;

(ii) $|d_1(0,p) - d_2(0,p)| = O(d_1(0,p)^{\frac{1}{s}})$.

This equivalence is not always sharp. However, we obtained in [B] the following stronger estimate in the Heiseneberg group.

**Theorem 4.1** (with Le Donne, Sambusetti, [B]). *Let $d_1$ and $d_2$ be two left-invariant subRiemannian or Riemannian distances on the first Heisenberg group $\mathbb{H}^1$. Then, the following are equivalent:*

*(i)* $\lim_{p\to\infty} \frac{d_1(0,p)}{d_2(0,p)} = 1$;

*(ii)* $\exists C > 0,\ \forall p \in \mathbb{H}^1 \qquad |d_1(0,p) - d_2(0,p)| \le C.$

*Moreover, if $d_1$ is a left-invariant Riemannian distance on $\mathbb{H}^1$, then there exists a unique left-invariant subRiemannian (not Riemannian) distance $d_2$ such that*

$$\exists C > 0,\ \forall p \in \mathbb{H}^1 \qquad |d_1(0,p) - d_2(0,p)| \le \frac{C}{d_2(0,p)}. \tag{4}$$

Notice that (4) expresses a really strong asymptotic property. Indeed, we used it for the computation of the *horoboundary* of the Riemannian Heisenberg group.

**Corollary 4.2** (with Le Donne, Sambusetti, [B])**.** *The horoboundary of the Riemannian Heisenberg group coincides with the horoboundary of the sub-Riemannian Heisenberg group.*

For a precise description for the horoboundary of the subRiemannian Heisenberg group see [25].

## 5. Minimal surfaces in the Heisenberg group

Geometric Measure Theory grew enormously in the 1950s and 1960s thanks to De Giorgi, Federer and Fleming [11, 12, 29]. Since then, a great advance has been seen, with generalizations to metric measure spaces [4, 5] and to Carnot-Carathéodory spaces [9, 16, 13]. Notice that a purely metric approach to Carnot-Carathéodory spaces does not lead very far. For instance, the subRiemannian first Heisenberg group $\mathbb{H}^1$ is a totally non-rectifiable metric space. In particular, any topological surface in $\mathbb{H}^1$ has Hausdorff dimension larger or equal to three, whence the boundary of any open set is never contained in the image of any Lipschitz map $\mathbb{R}^2 \to \mathbb{H}^1$.

De Giorgi's fundamental work hinges on two results: a structure theorem for sets of finite perimeter, which describes the regularity of the boundary of such sets; and a regularity theorem for perimeter minimizers, which improves regularity of the boundary in the case of additional variational properties.

The first *Heisenberg group* $\mathbb{H}^1$ is the unique non-Abelian nilpotent simply connected Lie group of dimension three. It is a stratified group of step 2, in fact the smallest stratified group that is not Abelian. A smooth vector field on $\mathbb{H}^1$ is called *horizontal* if it takes values in the left-invariant subbundle generated by the first layer $V_1$ of the stratification. If we fix a scalar product $\langle\cdot,\cdot\rangle$ on the first layer $V_1$ of the stratification, we can define a subRiemannian distance and the intrinsic perimeter. The *intrinsic perimeter* of a measurable set $E \subset \mathbb{H}^1$ in an open set $\Omega$ is defined as

$$P(E;\Omega) := \sup \int_E \mathrm{div} V \,\mathrm{d}\mu,$$

where the supremum is taken among all horizontal vector fields $V \in \mathrm{Vec}(\mathbb{H}^1)$ with $\mathrm{spt}(V) \subset\subset \Omega$ and $\langle V, V\rangle \leq 1$, see [13, 14].

A structure theorem for the intrinsic perimeter in the Heisenberg group has been proven in [14]. Namely, the reduced boundary $\partial^* E$ of $E$, i.e., the support of the Radon measure induced by $\Omega \mapsto P(E;\Omega)$, is the countable union of compact subsets of $\mathscr{C}^1_{\mathbb{H}}$-hypersurfaces, up to a set of $\mathscr{S}^3_{CC}$-measure zero. Here $\mathscr{S}^3_{CC}$ is the spherical Hausdorff measure given by the Carnot-Carathéodory metric on $\mathbb{H}^1$. A $\mathscr{C}^1_{\mathbb{H}}$*-hypersurface* in $\mathbb{H}^1$ is locally the zero level of a function $\mathbb{H}^1 \to \mathbb{R}$ with continuous derivatives along horizontal vector fields. $\mathscr{C}^1_{\mathbb{H}}$-hypersurfaces may be fractals from the Euclidean point of view: in [23] it has been shown the existence of a $\mathscr{C}^1_{\mathbb{H}}$-hypersurface that has Hausdorff dimension larger that 2 with respect to any Riemannian metric on $\mathbb{H}^1$.

A measurable set $E$ is a *locally perimeter minimizer* if for every bounded open set $\Omega \subset \mathbb{H}^1$ and every measurable set $F \subset \mathbb{H}^1$ with symmetric difference compactly contained in $\Omega$,

we have
$$P(E;\Omega) \le P(F;\Omega).$$
In this case the reduced boundary $\partial^* E$ of $E$ is called *minimal surface.* There are no strong results of regularity for minimal surfaces in $\mathbb{H}^1$. In fact, there are examples of non-smooth minimal surfaces, see [37, 30].

A $\mathscr{C}^1_{\mathbb{H}}$-hypersurface is locally the intrinsic graph of a $\mathscr{C}^1_{\mathbb{W}}$-function. The notion of $\mathscr{C}^1_{\mathbb{W}}$-function is a little involved, but very well suited for the non-commutative geometry of the Heisenberg group. Let $\delta_t$ be the standard dilations on $\mathbb{H}^1$ induced by the stratification. Let $\mathbb{W}$ be a 2-dimensional normal $\delta_t$-homogeneous subgroup of $\mathbb{H}^1$ and $\mathbb{V}$ a 1-dimensional subgroup whose Lie algebra is a subspace of $V_1$ such that $\mathbb{W} \cap \mathbb{V} = \{e\}$. Then $\mathbb{H}^1 = \mathbb{W} \rtimes \mathbb{V}$, i.e., for every $p \in \mathbb{H}^1$ there are unique $p_{\mathbb{W}} \in \mathbb{W}$ and $p_{\mathbb{V}} \in \mathbb{V}$ such that $p = p_{\mathbb{W}} p_{\mathbb{V}}$. If $f : \mathbb{W} \to \mathbb{V}$ is a function, then the *intrinsic graph* of $f$ is
$$\Gamma_f := \{pf(p) : p \in \mathbb{W}\} \subset \mathbb{H}^1.$$
Left-translations and dilations of intrinsic graphs are again intrinsic graphs. A function $\mathrm{d}f_p : \mathbb{W} \to \mathbb{V}$ is the *intrinsic differential* of $f$ at $p \in \mathbb{W}$ if
$$\Gamma_{\mathrm{d}f_p} = \lim_{\epsilon \to 0} \delta_{\epsilon^{-1}} \left( (pf(p))^{-1} \Gamma_f \right),$$
where the limit is intended as a Kuratowski limit of sets. A function $f : \mathbb{W} \to \mathbb{V}$ is said to be *of class* $\mathscr{C}^1_{\mathbb{W}}$ if for every $p \in \mathbb{W}$ there is a linear intrinsic differential $\mathrm{d}f_p$ of $f$ at $p$ and the map $p \mapsto \mathrm{d}f_p$ is continuous.

If we adapt the *Bernstein's Problem* to the language of Geometric Measure Theory in $\mathbb{H}^1$, we obtain the following problem: If $f \in \mathscr{C}^1_{\mathbb{W}}$ is such that $\Gamma_f$ is a minimal surface, is then $\Gamma_f$ a vertical plane? This question has been already answered positively in the case $f \in \mathscr{C}^1(\mathbb{R}^2)$, see [15], and negatively if $f$ is not assumed to be in $\mathscr{C}^1_{\mathbb{W}}$, see [30].

The classical approach to Bernstein's Problem would involve a variational method along a one-parameter family of diffeomorphisms. However, there are two issues. On the one side, the only diffeomorphisms that keep finite the intrinsic perimeter are of contact type, see [D] for a proof. On the other side, the following result shows that diffeomorphisms of contact type are too poor.

**Theorem 5.1** ([D])**.** *There are functions $f \in \mathscr{C}^1_{\mathbb{W}}$ whose intrinsic graph $\Gamma_f$ is not a minimal surface, but it is a locally area minimizer along all variations by diffeomorphisms of contact type.*

The main difficulty arises from the non-linear behaviour of the operator $f \mapsto \mathrm{d}f_p$. In fact, the differential $\mathrm{d}f_p$ is completely determined by a number, so it is equivalent to a function $\nabla^f f : \mathbb{W} \to \mathbb{R}$, the *intrinsic gradient* of $f$, that is represented by $\nabla^f f = \partial_\eta f + f \partial_\tau f$ in a suitable choice of coordinates $(\eta, \tau)$ for $\mathbb{W}$.

The non-linearity that appears in the intrinsic differential brings to a new phenomenon: the space of all functions of class $\mathscr{C}^1_{\mathbb{W}}$ is not a vector space. Indeed, there is a $\mathscr{C}^1_{\mathbb{W}}$-function $f$ such that $\Gamma_{f+1}$ cannot be the boundary of a set of locally finite intrinsic perimeter (e.g., consider the example in [23]). In other words, $f+1$ is not even of bounded variation. This leads to the question whether $\mathscr{C}^1_{\mathbb{W}}$ is a Banach manifold. In particular, we could ask if the topological space $\mathscr{C}^1_{\mathbb{W}}$ is topologically homogeneous. More precisely, given $f, g \in \mathscr{C}^1_{\mathbb{W}}$, we could seek a continuous transformation $\Phi : \mathscr{C}^1_{\mathbb{W}} \to \mathscr{C}^1_{\mathbb{W}}$ such that $\Phi(f) = g$.

## References


[1] A. Agrachev, D. Barilari, and U. Boscain. *Introduction to Riemannian and Sub-Riemannian geometry (from Hamilton viewpoint).* Tech. rep. Sept. 2011.

[2] A. A. Agrachëv. "Any sub-Riemannian metric has points of smoothness". In: *Dokl. Akad. Nauk* 424.3 (2009), pp. 295–298.

[3] A. A. Agrachev and Y. L. Sachkov. *Control theory from the geometric viewpoint.* Vol. 87. Encyclopaedia of Mathematical Sciences. Berlin: Springer, 2004, pp. xiv+412.

[4] L. Ambrosio and B. Kirchheim. "Currents in metric spaces". In: *Acta Math.* 185.1 (2000), pp. 1–80.

[5] L. Ambrosio and B. Kirchheim. "Rectifiable sets in metric and Banach spaces". In: *Math. Ann.* 318.3 (2000), pp. 527–555.

[6] H. Bass. "The degree of polynomial growth of finitely generated nilpotent groups". In: *Proc. London Math. Soc. (3)* 25 (1972), pp. 603–614.

[7] E. Breuillard. "Geometry of locally compact groups of polynomial growth and shape of large balls". In: *Groups Geom. Dyn.* 8.3 (2014), pp. 669–732.

[8] E. Breuillard and E. Le Donne. "On the rate of convergence to the asymptotic cone for nilpotent groups and subFinsler geometry". In: *Proc. Natl. Acad. Sci. USA* 110.48 (2013), 19220–19226.

[9] L. Capogna, D. Danielli, and N. Garofalo. "The geometric Sobolev embedding for vector fields and the isoperimetric inequality". In: *Comm. Anal. Geom.* 2.2 (1994), pp. 203–215.

[10] J. Cheeger. "Differentiability of Lipschitz functions on metric measure spaces". In: *Geom. Funct. Anal.* 9.3 (1999), 428–517.

[11] E. De Giorgi. *Frontiere orientate di misura minima.* Seminario di Matematica della Scuola Normale Superiore di Pisa, 1960-61. Editrice Tecnico Scientifica, Pisa, 1961, p. 57.

[12] H. Federer. *Geometric measure theory.* Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969, pp. xiv+676.

[13] B. Franchi, R. Serapioni, and F. Serra Cassano. "Rectifiability and perimeter in the Heisenberg group". In: *Math. Ann.* 321.3 (2001), pp. 479–531.

[14] B. Franchi, R. Serapioni, and F. Serra Cassano. "Regular submanifolds, graphs and area formula in Heisenberg groups". In: *Adv. Math.* 211.1 (2007), pp. 152–203.

[15] M. Galli and M. Ritoré. "Area-stationary and stable surfaces of class $C^1$ in the sub-Riemannian Heisenberg group $\mathbb{H}^1$". In: *Adv. Math.* 285 (2015), pp. 737–765.

[16] N. Garofalo and D.-M. Nhieu. "Isoperimetric and Sobolev inequalities for Carnot-Carathéodory spaces and the existence of minimal surfaces". In: *Comm. Pure Appl. Math.* 49.10 (1996), pp. 1081–1144.

[17] M. Gromov. "Hyperbolic manifolds (according to Thurston and Jørgensen)". In: *Bourbaki Seminar, Vol. 1979/80.* Vol. 842. Lecture Notes in Math. Springer, Berlin-New York, 1981, 40–53.

[18] M. Gromov. "Groups of polynomial growth and expanding maps". In: *Inst. Hautes Études Sci. Publ. Math.* 53 (1981), pp. 53–73.

[19] Y. Guivarc'h. "Croissance polynomiale et périodes des fonctions harmoniques". In: *Bull. Soc. Math. France* 101 (1973), 333–379.

[20] P. Hajłasz and P. Koskela. "Sobolev met Poincaré". In: *Mem. Amer. Math. Soc.* 145.688 (2000), pp. x+101.

[21] J. Heinonen and P. Koskela. "Quasiconformal maps in metric spaces with controlled geometry". In: *Acta Math.* 181.1 (1998), pp. 1–61.

[22] L. Hörmander. "Hypoelliptic second order differential equations". In: *Acta Mathematica* 119.1 (1967), pp. 147–171.

[23] B. Kirchheim and F. Serra Cassano. "Rectifiability and parameterization of intrinsic regular surfaces in the Heisenberg group". In: *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* 3.4 (2004), 871–896.

[24] V. Kivioja and E. Le Donne. "Isometries of nilpotent metric groups". In: *ArXiv e-prints* (Jan. 2016).

[25] T. Klein and A. Nicas. "The horofunction boundary of the Heisenberg group: the Carnot-Carathéodory metric". In: *Conform. Geom. Dyn.* 14 (2010), pp. 269–295.

[26] E. Le Donne. "A metric characterization of Carnot groups". In: *Proc. Amer. Math. Soc.* 143.2 (2015), pp. 845–849.

[27] E. Le Donne and S. Rigot. *Besicovitch Covering Property on graded groups and applications to measure differentiation.* To appear in "Journal für die reine und angewandte Mathematik".

[28] E. Le Donne, R. Montgomery, A. Ottazzi, P. Pansu, and D. Vittone. "Sard property for the endpoint map on some Carnot groups". In: *Annales de l'Institut Henri Poincare (C) Non Linear Analysis* (2015).

[29] F. Maggi. *Sets of finite perimeter and geometric variational problems.* Vol. 135. Cambridge Studies in Advanced Mathematics. An introduction to geometric measure theory. Cambridge University Press, Cambridge, 2012, pp. xx+454.

[30] R. Monti, F. Serra Cassano, and D. Vittone. "A negative answer to the Bernstein problem for intrinsic graphs in the Heisenberg group". In: *Boll. Unione Mat. Ital. (9)* 1.3 (2008), pp. 709–727.

[31] G. D. Mostow. "Quasi-conformal mappings in $n$-space and the rigidity of hyperbolic space forms". In: *Inst. Hautes Études Sci. Publ. Math.* 34 (1968), pp. 53–104.

[32] G. D. Mostow. *Strong rigidity of locally symmetric spaces.* Annals of Mathematics Studies, No. 78. Princeton University Press, Princeton, N.J. and University of Tokyo Press, Tokyo, 1973, pp. v+195.

[33] A. Nagel, E. M. Stein, and S. Wainger. "Balls and metrics defined by vector fields. I. Basic properties". In: *Acta Math.* 155.1-2 (1985), pp. 103–147.

[34] P. Pansu. "Croissance des boules et des géodésiques fermées dans les nilvariétés". In: *Ergodic Theory Dynam. Systems* 3.3 (1983), pp. 415–445.

[35] P. Pansu. "Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un". In: *Ann. of Math. (2)* 129.1 (1989), pp. 1–60.

[36] L. Rifford. "À propos des sphères sous-riemanniennes". In: *Bull. Belg. Math. Soc. Simon Stevin* 13.3 (2006), pp. 521–526.

[37] M. Ritoré. "Examples of area-minimizing surfaces in the sub-Riemannian Heisenberg group $\mathbb{H}^1$ with low regularity". In: *Calc. Var. Partial Differential Equations* 34.2 (2009), pp. 179–192.

[38] L. P. Rothschild and E. M. Stein. "Hypoelliptic differential operators and nilpotent groups". In: *Acta Math.* 137.3-4 (1976), pp. 247–320.

[39] N. Shanmugalingam. "Newtonian spaces: an extension of Sobolev spaces to metric measure spaces". In: *Rev. Mat. Iberoamericana* 16.2 (2000), pp. 243–279.

[40] E. Siebert. "Contractive automorphisms on locally compact groups". In: *Math. Z.* 191.1 (1986), pp. 73–90.

[41] E. Stein. *Harmonic analysis : real-variable methods, orthogonality, and oscillatory integrals.* Princeton, N.J: Princeton University Press, 1993.

[42] J. A. Wolf. "Growth of finitely generated solvable groups and curvature of Riemanniann manifolds". In: *J. Differential Geometry* 2 (1968), pp. 421–446.

[A]

**Regularity properties of spheres in homogeneous groups**

E. Le Donne and S. Nicolussi Golo

First published in *Transactions of the AMS*
by the American Mathematical Society.

# REGULARITY PROPERTIES OF SPHERES IN HOMOGENEOUS GROUPS

ENRICO LE DONNE AND SEBASTIANO NICOLUSSI GOLO

Abstract. We study left-invariant distances on Lie groups for which there exists a one-parameter family of homothetic automorphisms. The main examples are Carnot groups, in particular the Heisenberg group with the standard dilations. We are interested in criteria implying that, locally and away from the diagonal, the distance is Euclidean Lipschitz and, consequently, that the metric spheres are boundaries of Lipschitz domains in the Euclidean sense. In the first part of the paper, we consider geodesic distances. In this case, we actually prove the regularity of the distance in the more general context of sub-Finsler manifolds with no abnormal geodesics. Secondly, for general groups we identify an algebraic criterium in terms of the dilating automorphisms, which for example makes us conclude the regularity of every homogeneous distance on the Heisenberg group. In such a group, we analyze in more details the geometry of metric spheres. We also provide examples of homogeneous groups where spheres present cusps.

Contents



2010 *Mathematics Subject Classification.* 28A75, 22E25, 53C60, 53C17, 26A16.

E.L.D. has been supported by the Academy of Finland project no. 288501. S.N.G. has been supported by the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n. 607643.

## 1. Introduction

The study of the asymptotic geometry of groups lead us to investigate spheres in homogeneous groups, examples of which are asymptotic cones of finitely generated nilpotent groups. A homogeneous group is a Lie group $G$ endowed with a family of Lie group automorphisms $\{\delta_\lambda\}_{\lambda>0}$ and a left-invariant distance $d$ for which each $\delta_\lambda$ multiplies the distance by $\lambda$, see Section 2.2. An algebraic characterization of these groups is known by [29]. In fact, the Lie algebra $\mathfrak{g}$ of $G$ admits a *grading*, i.e., a decomposition $\mathfrak{g} = \bigoplus_{i\geq 1} V_i$ such that $[V_i, V_j] \subset V_{i+j}$. For simplicity, we assume that the dilations are the ones induced by the grading. Namely, the *dilation of factor $\lambda$ relative to the grading* is the one such that $(\delta_\lambda)_*(v) = \lambda^i v$ for all $v \in V_i$. We denote by 0 the neutral element of $G$ and by $\mathbb{S}_d$ the unit sphere at 0 for a distance $d$ on $G$, i.e., $\mathbb{S}_d := \{p \in G : d(0,p) = 1\}$.

In this paper we want to exclude cusps in spheres since their presence in the asymptotic cone of a finitely generated nilpotent group may give a slower rate of convergence in the blow down, see [8]. We find criteria implying that the metric spheres are boundaries of Lipschitz domains and in fact that the distance function from a point is a locally Lipschitz function with respect to a Riemannian metric.

First, we address the case where the distance $d$ is a length distance. Thanks to a characterization of Carnot groups, see [18], the group $G$ is in this case a stratified group and $d$ is a sub-Finsler distance. Being a *stratified group* means that the grading of $\mathfrak{g}$ is such that the first layer $V_1$ generates $\mathfrak{g}$. Being a *sub-Finsler distance* means that there are a left-invariant subbundle $\Delta \subset TG$ and a left-invariant norm $\|\cdot\|$ on $\Delta$ such that the length induced by $d$ of an absolutely continuous curve $\gamma : [0,1] \to G$ is equal to $\int_0^1 \|\gamma'(t)\| \,\mathrm{d}t$, where $\|\gamma'(t)\| = +\infty$ if $\gamma'(t) \notin \Delta$. The left-invariant subbundle $\Delta$ is in fact the one generated by $V_1$.

In the sub-Finsler case, an obstruction to Lipschitz regularity of the sphere comes from the presence of length-minimizing curves (also called geodesics) that are not regular, in the sense that the first variation parallel to the subbundle $\Delta$ does not have maximal rank, see Definition 2.6.

**Theorem 1.1.** *Let $G$ be a stratified group endowed with a sub-Finsler metric $d$. Let $d_0 : G \to [0, +\infty)$, $p \mapsto d(0,p)$. Let $p \in G$ be such that all geodesics from 0 to $p$ are regular. Then for any Riemannian metric $\rho$ on $G$ the function $d_0$ is Lipschitz with respect to $\rho$ in some neighborhood of $p$.*

We will actually state and prove Theorem 1.1 in the more general setting of *sub-Finsler manifolds of constant-type norm*, see Section 2.1.

In case of homogeneity, the regularity of the distance implies also the regularity of the spheres. Hence, using Theorem 1.1 we easily get the second result for sub-Finsler homogeneous groups.

**Theorem 1.2.** *Let $G$ be a stratified group endowed with a sub-Finsler metric $d$. Let $p \in \mathbb{S}_d$ be such that all geodesics from $0$ to $p$ are regular. Then, in smooth coordinates, the set $\mathbb{S}_d$ is a Lipschitz graph in some neighborhood of $p$. In particular, if all non-constant geodesics are regular, then metric balls are Lipschitz domains.*

Notice that a ball may be a Lipschitz domain even if the distance from a point is not Lipschitz (we give an example in Remark 5.5). In Section 5 we also present examples of sub-Riemannian and sub-Finsler distances whose balls have a cusp.

At a second stage, we drop the hypothesis of $d$ being a length distance and we present a result similar to the previous Theorem 1.2 in the context of homogeneous groups. Hereafter we denote by $L_p$ and $R_p$ the left and the right translations on $G$, respectively, and by $\bar{\delta}(p)$ the vector $\frac{\mathrm{d}}{\mathrm{d}t}\delta_t(p)\big|_{t=1} \in T_pG$, where $\{\delta_t\}_{t>0}$ are the dilations relative to a grading.

**Theorem 1.3.** *Let $(G, d)$ be a homogeneous group with dilations relative to a grading, see Definition 2.12. Assume $p \in \mathbb{S}_d$ is such that*

$$\mathrm{d}L_p(V_1) + \mathrm{d}R_p(V_1) + \mathrm{span}\{\bar{\delta}(p)\} = T_pG. \tag{1.1}$$

*Then, in some neighborhood of $p$ we have that the sphere $\mathbb{S}_d$ is a Lipschitz graph and the distance $d_0$ from the identity is Lipschitz with respect to any Riemannian metric $\rho$.*

The similarity between Theorem 1.2 and Theorem 1.3 consists in the fact that, if $d$ is a sub-Finsler distance, then condition (1.1) implies that all geodesics from $0$ to $p$ are regular, see Remark 4.5.

The equality (1.1) or the absence of non-regular geodesics are actually quite strong conditions. However, in general we can give an upper bound for the Hausdorff dimension of spheres. In fact, if $d$ is a homogeneous distance on a graded group of maximal degree $s$, then

$$\dim_H^\rho(G) - 1 \leq \dim_H^\rho(\mathbb{S}_d) \leq \dim_H^\rho(G) - \frac{1}{s}, \tag{1.2}$$

where $\dim_H^\rho$ is the Hausdorff dimension with respect to some (therefore any) Riemannian metric $\rho$. We show with Proposition 5.1 that this estimate is sharp.

In the last part of the paper, we analyze in more details an important specific example: the Heisenberg group. In this graded group we consider all possible homogeneous distances and prove that in exponential coordinates

(i) the unit ball is a star-shaped Lipschitz domain (Proposition 6.1);
(ii) the unit sphere is a locally Lipschitz graph with respect to the direction of the center of the group (Proposition 6.2).

We also give a method to construct homogeneous distances in the Heisenberg group with arbitrary Lipschitz regularity of the sphere. Namely, the graph of each Lipschitz function defined on the unit disk, up to adding to it a constant, is the sphere of some homogeneous distance, see Proposition 6.3. The investigation of this class of examples is meaningful in connection to Besicovitch's covering property as studied in [21] and [22].

The paper is organized as follows. In section 2 we will present all preliminary notions needed in the paper. We introduce sub-Finsler manifolds of constant-type norm, graded and homogeneous groups and Carnot groups. Section 3 is devoted to the proof of Theorem 1.1; first in the setting of sub-Finsler manifolds, see Theorem 3.1 proved in Section 3.4, then with a more specific result for Carnot groups, see Proposition 3.3. In Section 4 we see metric spheres as graphs over smooth spheres. Hence, we show Theorem 1.2, the inequalities (1.2), and Theorem 1.3. In Section 5 we present six examples: three different grading of $\mathbb{R}^2$, the Heisenberg group, a sub-Finsler sphere with a cusp and a sub-Riemannian sphere with a cusp. In Section 6 we prove stronger properties for spheres of homogeneous distances on the Heisenberg group.

## 2. Preliminaries

2.1. **Sub-Finsler structures.** Let $M$ be a manifold of dimension $n$. We will write $TM$ for the tangent bundle and $\mathrm{Vec}(M)$ for the space of smooth vector fields on $M$.

**Definition 2.1** (Sub-Finsler structure)**.** A *sub-Finsler structure (of constant-type norm) of rank* $r$ on a manifold $M$ is a triple $(\mathbb{E}, \|\cdot\|, \mathtt{f})$, where $(\mathbb{E}, \|\cdot\|)$ is a normed vector space of dimension $r$ and $\mathtt{f} : M \times \mathbb{E} \to TM$ is a smooth bundle morphism with $\mathtt{f}(\{p\} \times \mathbb{E}) \subset T_pM$, for all $p \in M$.

We added the specification "of constant-type norm" because the norm $\|\cdot\|$ defined on the fibers of $M \times \mathbb{E}$ does not depend on the base point of each fiber.

**Definition 2.2** (Horizontal curve)**.** A curve $\gamma : [0,1] \to M$ is a *horizontal curve* if it is absolutely continuous and there is $u : [0,1] \to \mathbb{E}$ measurable, which is called *a control of* $\gamma$, such that

$$\gamma'(t) = \mathtt{f}(\gamma(t), u(t)) \qquad \text{for a.e. } t \in [0,1].$$

In this case $\gamma$ is called *integral curve* of $u$ and we write $\gamma_u$.

**Definition 2.3** (Space of controls)**.** The space of $\mathtt{L}^\infty$*-controls* is defined as[1]

$$\mathtt{L}^\infty([0,1]; \mathbb{E}) := \left\{ u : [0,1] \to \mathbb{E} \text{ measurable, } \operatorname*{ess\,sup}_{t \in [0,1]} \|u(t)\| < \infty \right\}.$$

This is a Banach space with norm $\|u\|_{\mathtt{L}^\infty} := \operatorname{ess\,sup}_{t\in[0,1]} \|u(t)\|$.

[1] Among the three norms $\mathtt{L}^1$, $\mathtt{L}^2$ and $\mathtt{L}^\infty$ for controls, we chose the latter because the unit ball in $\mathtt{L}^1([0,1];\mathbb{E})$ is not weakly compact and the $\mathtt{L}^2$-space is not a Hilbert space in our context.

Thanks to known results for ordinary differential equations, see [27], given a control $u \in \mathtt{L}^\infty([0,1];\mathbb{E})$ and a point $p \in M$ there is a unique solution $\gamma_{u,p}$ to the Cauchy problem

$$\begin{cases} \gamma_{u,p}(0) = p \\ \gamma_{u,p}'(t) = \mathtt{f}(\gamma_{u,p}(t), u(t)) \quad \text{for a.e. } t \text{ in a neighborhood of } 0. \end{cases}$$

*Remark* 2.4. We will always assume that every $u \in \mathtt{L}^\infty([0,1];\mathbb{E})$ and every $p \in M$ the curve $\gamma_{u,p}$ is defined on the interval $[0,1]$. This happens in many cases, for example for left-invariant sub-Finsler structures on Lie groups, in particular in Carnot groups.

**Definition 2.5** (End-point map)**.** Fix $o \in M$. Define the *End-point map* with *base point* $o$, $\mathtt{End}_o : \mathtt{L}^\infty([0,1];\mathbb{E}) \to M$, as

$$\mathtt{End}_o(u) = \gamma_{u,o}(1).$$

By standard result of ODE the map $\mathtt{End}_o$ is of class $\mathscr{C}^1$, see [27].

**Definition 2.6** (Regular curves)**.** Given $o \in M$, a control $u \in \mathtt{L}^\infty([0,1];\mathbb{E})$ is said to be *regular* if it is a regular point of $\mathtt{End}_o$, i.e., if $\mathrm{d}\mathtt{End}_o(u) : \mathtt{L}^\infty([0,1];\mathbb{E}) \to T_{\mathtt{End}_o(u)}M$ is surjective. A *singular control* is a control that is not regular.

**Definition 2.7** (Sub-Finsler distance)**.** The *sub-Finsler distance*, also called *Carnot-Carathéodory distance*, between two points $p, q \in M$ is

$$d(p,q) := \inf\left\{ \int_0^1 \|u(t)\| \,\mathrm{d}t : u \in \mathtt{L}^\infty([0,1];\mathbb{E}) \text{ with } \mathtt{End}_p(u) = q \right\}.$$

Clearly $(M,d)$ is a metric space, even though it might happen $d(p,q) = \infty$. Let $\ell_d(\gamma)$ be the length of a curve $\gamma$ with respect to $d$, see [4]. It can be proven that a curve $\gamma : [0,1] \to (M,d)$ is Lipschitz if and only if it is horizontal and it admits a control in $\mathtt{L}^\infty([0,1];\mathbb{E})$. Moreover, if $\gamma$ is Lipschitz, then

$$\ell_d(\gamma) = \inf\left\{ \int_0^1 \|u(t)\| \,\mathrm{d}t : u \in \mathtt{L}^\infty([0,1];\mathbb{E}) \text{ control of } \gamma \right\}.$$

We will use the term *geodesic* as a synonym of *length-minimizer.*

The distance can be expressed by using the $\mathtt{L}^\infty$-norm, i.e., for every $p, q \in M$

$$d(p,q) = \inf\left\{ \|u\|_{\mathtt{L}^\infty} : u \in \mathtt{L}^\infty([0,1];\mathbb{E}) \text{ with } \mathtt{End}_p(u) = q \right\}.$$

Moreover, if $u$ realizes the infimum above, then its integral curve $\gamma_u$ starting from $p$ is a length-minimizing curve parametrized by constant velocity, i.e.,

$$d(p,q) = \|u\|_{\mathtt{L}^\infty} = \ell_d(\gamma_u) = \|u(t)\|, \qquad \text{for a.e.} t \in [0,1].$$

Notice that the $\mathtt{L}^\infty$-norm plays a similar role here as the $\mathtt{L}^2$-energy in sub-Riemannian geometry.

**Definition 2.8** (Bracket-generating condition)**.** Let $\mathscr{A}$ be the Lie algebra generated by the set

$$\{p \mapsto \mathtt{f}(p, X(p)) \text{ with } X : M \to \mathbb{E} \text{ smooth}\} \subset \mathrm{Vec}(M).$$

We say that the sub-Finsler structure $(\mathbb{E}, \|\cdot\|, \mathtt{f})$ on $M$ satisfies the *bracket-generating condition* if for all $p \in M$

$$\{V(p) : V \in \mathscr{A}\} = T_pM.$$

As a consequence of the Orbit Theorem [17], we have the following basic well-known fact.

**Lemma 2.9.** *If $(\mathbb{E}, \|\cdot\|, \mathtt{f})$ satisfies the bracket-generating condition, then the distance $d$ induces the original topology of $M$ and $(M, d)$ is a locally compact and locally geodesic length space.*

By the Hopf-Rinow Theorem, see [9], the assumption in Remark 2.4 implies that $(M, d)$ is a complete, boundedly compact metric space.

2.2. **Graded groups.** All Lie algebras considered here are over $\mathbb{R}$ and finite-dimensional.

**Definition 2.10** (Graded group)**.** A Lie algebra $\mathfrak{g}$ is *graded* if it is equipped with a *grading*, i.e., with a vector-space decomposition $\mathfrak{g} = \bigoplus_{i>0} V_i$, where $i > 0$ means $i \in (0, \infty)$, such that for all $i, j > 0$ it holds $[V_i, V_j] \subset V_{i+j}$. A *graded Lie group* is a simply connected Lie group $G$ whose Lie algebra is graded. The *maximal degree* of a graded group $G$ is the maximum $i$ such that $V_i \neq \{0\}$.

Graded groups are nilpotent and the exponential map $\exp : \mathfrak{g} \to G$ is a global diffeomorphism. We will denote by 0 the neutral element of $G$ and identify $\mathfrak{g} = T_0 G$.

**Definition 2.11** (Dilations)**.** In a graded group for which the Lie algebra has the grading $\mathfrak{g} = \bigoplus_{i>0} V_i$, the *dilations* relative to the grading are the group homomorphisms $\delta_\lambda : G \to G$, for $\lambda \in (0, \infty)$, such that $(\delta_\lambda)_*(v) = \lambda^i v$ for all $v \in V_i$.

In the definition above, $\phi_*$ denotes the Lie algebra homomorphism associated to a Lie group homomorphism $\phi$, in particular, $\phi \circ \exp = \exp \circ \phi_*$. Since a graded group is simply connected, $\delta_\lambda$ is well defined. Notice that, for any $\lambda, \mu > 0$, $\delta_\lambda \circ \delta_\mu = \delta_{\lambda\mu}$.

**Definition 2.12** (Homogeneous distances)**.** Let $G$ be a graded group with a dilations $\{\delta_\lambda\}_{\lambda>0}$, relative to the grading. We say that a distance $d$ on $G$ is *homogeneous* if it is left-invariant, i.e., for every $g, x, y \in G$ we have $d(gx, gy) = d(x, y)$, and one-homogeneous with respect to the dilations, i.e., for all $\lambda > 0$ and all $x, y \in G$ we have $d(\delta_\lambda x, \delta_\lambda y) = \lambda d(x, y)$. If $d$ is one such a distance, then $(G, d)$ is called *homogeneous group* (with dilations relative to the grading).

*Remark* 2.13. A graded group admits a homogeneous distance if and only if for $i \in (0, 1)$ we have $V_i = \{0\}$, see [16].

Given a homogeneous distance $d$, the function $p \mapsto d_0(p) := d(0, p)$ is a homogeneous norm. Here with the term *homogeneous norm* we mean a function $N : G \to [0, +\infty)$ such that for all $p, q \in G$ and all $\lambda > 0$ it holds

1. $N(p) = 0 \;\Leftrightarrow\; p = 0$;
2. $N(pq) \le N(p) + N(q)$;
3. $N(p^{-1}) = N(p)$;
4. $N(\delta_\lambda p) = \lambda N(p)$.

In fact, homogeneous distances are in bijection with homogeneous norms on $G$ through the formula $d(p,q) = N(p^{-1}q)$.

Homogeneous distances induce the original topology of $G$, see [22]. Moreover, given two homogeneous distances $d_1, d_2$ on $G$, there is a constant $C > 0$ such that for all $p, q \in G$

$$\frac{1}{C} d_1(p,q) \leq d_2(p,q) \leq C d_1(p,q). \tag{2.1}$$

**Lemma 2.14.** *Let $G$ be a graded group and $0 < k_1 \leq k_2$ such that $V_i = \{0\}$ for all $i < k_1$ and all $i > k_2$. Let $d$ be a homogeneous distance and $\rho$ a left-invariant Riemannian metric on $G$. Then there are $C, \epsilon > 0$ such that for all $p, q \in G$ with $\rho(p,q) < \epsilon$ it holds*

$$\frac{1}{C} \rho(p,q)^{\frac{1}{k_1}} \leq d(p,q) \leq C \rho(p,q)^{\frac{1}{k_2}}. \tag{2.2}$$

*In particular, the homogeneous norm $d_0$ is locally $\frac{1}{k_2}$-Hölder.*

*Proof.* We identify $G = \mathfrak{g}$ via the exponential map. So, if $p \in G$, we denote by $p_i$ the $i$-th component in the decomposition $p = \sum_i p_i$ with $p_i \in V_i$. Fix a norm $|\cdot|$ on $\mathfrak{g}$. For any pair $(p,q) \in G \times G$ define

$$\eta(p,q) := \eta(0, p^{-1}q), \text{ where } \eta(0,p) := \max_i (|p_i|)^{\frac{1}{i}}.$$

The function $\eta$ is a so-called quasi-distance, see [22]. In particular, $\eta$ is continuous, left-invariant and one-homogeneous with respect to the dilations $\delta_\lambda$. Therefore, if $d$ is a homogeneous distance, then there is $C > 0$ such that

$$\frac{1}{C} \eta(p,q) \leq d(p,q) \leq C \eta(p,q).$$

So, we can prove (2.2) only for $\eta$.

Let $C, \epsilon > 0$ be with $C\epsilon < 1$ and such that, if $\rho(0,p) < \epsilon$, then

$$\frac{1}{C} \rho(0,p) \leq \max_i |p_i| \leq C \rho(0,p). \tag{2.3}$$

Therefore, if $\rho(p,q) < \epsilon$, then $|(p^{-1}q)_i| \leq C\rho(p,q) < 1$ for all $i$ and

$$\max_i |(p^{-1}q)_i|^{\frac{1}{k_1}} \leq \max_i (|(p^{-1}q)_i|)^{\frac{1}{i}} = \eta(p,q) \leq \max_i |(p^{-1}q)_i|^{\frac{1}{k_2}}, \tag{2.4}$$

thanks to the monotonicity of the function $x \mapsto a^x$ for $0 < a < 1$. The thesis follows immediately from (2.3) and (2.4) combined. □

Next lemma gives a characterization of sets that are the unit ball of a homogeneous distance. In this paper, we denote by $\mathrm{int}(B)$ the interior of a subset $B$.

**Lemma 2.15.** *Let $G$ be a graded group with dilations $\delta_\lambda$, $\lambda > 0$. A set $B \subset G$ is the unit ball with center $0$ of a homogeneous distance on $G$ if and only if $B$ is compact, $0 \in \mathrm{int}(B)$, $B = B^{-1}$ and*

$$\forall p, q \in B,\ \forall t \in [0,1] \quad \delta_t(p)\delta_{1-t}(q) \in B. \tag{2.5}$$

The proof of the latter fact is straightforward and hence omitted. One only needs to show that the function $N(p) := \inf\{t \geq 0 : \delta_{t^{-1}} p \in B\}$ is a homogeneous norm and $B = \{p : N(p) \leq 1\}$.

**Definition 2.16** (Stratified group)**.** A *stratified group* is a graded group $G$ such that its Lie algebra $\mathfrak{g}$ is generated by the layer $V_1$ of the grading of $\mathfrak{g}$.

Notice that in a stratified group $G$ the maximal degree $s$ of the grading equals the nilpotency step of $G$ and it holds $\mathfrak{g} = \bigoplus_{i=1}^{s} V_i$ with $[V_1, V_i] = V_{i+1}$ for all $i \in \{1, \dots, s\}$, with $V_{s+1} = \{0\}$. We also remark that all stratifications of a group $G$ are isomorphic to each other, i.e., if $\mathfrak{g} = \bigoplus_{i=1}^{s'} W_i$ is a second stratification, then there is a Lie group automorphism $\phi : G \to G$ such that $\phi_*(W_i) = V_i$ for all $i$, see [19].

In a stratified group, the map $\mathtt{f} : G \times V_1 \to TG$, $\mathtt{f}(g, v) := \mathrm{d}L_g(v)$, is a bundle morphism with $\mathtt{f}(g, v) \in T_gG$. So, if $\|\cdot\|$ is any norm on $V_1$, the triple $(V_1, \|\cdot\|, \mathtt{f})$ is a sub-Finsler structure on $G$. The stratified group $G$ endowed with the corresponding sub-Finsler distance $d$ is called *Carnot group*. Such a $d$ is an example of a homogeneous distance on $G$.

*Remark* 2.17. As already stated, singular curves play a central role in our analysis, because they disrupt the Lipschitz regularity of the distance function. We recall that every Carnot group of nilpotency step $s \geq 3$ has singular geodesics, see Appendix A. More precisely, there is $X \in V_1$ such that the curve $t \mapsto \exp(tX)$ is a singular geodesic. In particular, if all non-constant length-minimizing curves are regular, then the step of the group is necessarily at most 2.

## 3. Regularity of sub-Finsler distances

We will prove in this section that sub-Finsler distances are Lipschitz whenever all length-minimizing curves are regular, see Theorem 3.1. Theorem 1.1 expresses this result for Carnot groups.

It is important to remind what is known in the sub-Riemannian case. A *sub-Riemannian distance* is a sub-Finsler distance whose norm on the bundle $\mathbb{E}$ is induced by a scalar product. Rifford proved in [26] that, if there are no singular length-minimizers, for all $o \in M$, not only $d_o$ is locally Lipschitz, but also the spheres centered at $o$ are Lipschitz hypersurfaces for almost all radii. The key points of his proof are the tools of Clarke's non-smooth calculus (see [12]) and a version of Sard's Lemma for the distance function (see [25]). An exhaustive exposition of this topic can be found in [2].

In Rifford's version of Sard's Lemma, one uses the fact that the $\mathtt{L}^2$ norm in the Hilbert space $\mathtt{L}^2([0, 1]; \mathbb{E})$ is smooth away from the origin. If $\mathbb{E}$ is equipped with a generic norm, instead, the $\mathtt{L}^p$ norm on $\mathtt{L}^p([0, 1]; \mathbb{E})$ with $1 \leq p \leq \infty$ may be non-smooth, hence the proof does not work in the sub-Finsler case.

The non-smoothness of the norm can be seen in another dissimilarity between sub-Riemannian and sub-Finsler distances. Sub-Riemannian distances are proven to be locally semi-concave when there are no singular length-minimizing curves. We remind that a function $f : \mathbb{R}^n \to \mathbb{R}$ is *semi-concave* if for each $p \in \mathbb{R}^n$ there exists a $\mathscr{C}^2$ function $g : \mathbb{R}^n \to \mathbb{R}$ such that $f \leq g$ and $f(p) = g(p)$, see [27]. Semi-concavity is a stronger property than being Lipschitz. However, semi-concavity fails to hold in the sub-Finsler case. For example, the $\ell^1$-distance $d(0,(x,y)) := |x| + |y|$ on $\mathbb{R}^2$ is a sub-Finsler distance that is not semi-concave along the coordinate axis, although all curves are regular.[2]

We restrict our analysis to the Lipschitz regularity of the distance function, from which we deduce regularity properties of the spheres by means of the homogeneity of Carnot groups. With this aim in view, the core of the proof of Theorem 3.1 is the bound on the point-wise Lipschitz constant (see (3.5) at page 14), which already appeared in the sub-Riemannian context, see [1]. Our approach differs from the sub-Riemannian case for the fact that the set of optimal curves joining two points on a sub-Finsler manifold may not be compact in the $W^{1,\infty}$ topology. As an example, consider the set of all length-minimizers from $(0,0)$ to $(0,1)$ for the $\ell^\infty$-distance $d(0,(x,y)) := \max\{|x|,|y|\}$ on $\mathbb{R}^2$.[3] However, we are still able to obtain a bound on the pointwise Lipschitz constant, i.e., to prove (3.5), by use of the weak* topology on controls.

**Theorem 3.1.** *Let $(\mathbb{E}, \|\cdot\|, \mathtt{f})$ be a sub-Finsler structure on $M$ with sub-Finsler distance $d$. Fix $o$ and $p$ in $M$. If all the length-minimizing curves from $o$ to $p$ are regular, then for every Riemannian metric $\rho$ on $M$ there are a neighborhood $U$ of $p$ and $L > 0$ such that*

$$\forall q_1, q_2 \in U \qquad d_o(q_1) - d_o(q_2) \le L\rho(q_1, q_2). \tag{3.1}$$

The proof is presented in Section 3.4.

*Remark* 3.2. Theorem 3.1 can be made more quantitative. Define

$$\tau_0 := \inf \left\{ \tau(\,\mathrm{d}\mathtt{End}_o(u)) : \mathtt{End}_o(u) = p \text{ and } \|u\|_{\mathtt{L}^\infty} = d(o,p) \right\},$$

where, for any linear operator $L$, $\tau(L)$ the minimal stretching, which we will recall in Definition 3.4. Then, for every $L > \frac{1}{\tau_0}$, there exists a neighborhood $U$ of $p$ such that (3.1) holds. The hypothesis of regularity of all length-minimizing curves from $o$ to $p$ is equivalent to $\tau_0 > 0$.

In the case of Carnot groups (of step 2, see Remark 2.17), we can obtain the following more global result.

**Proposition 3.3.** *Let $(G,d)$ be a Carnot group without non-constant singular geodesics. Then for every left-invariant Riemannian metric $\rho$ and every neighborhood $U$ of $0$ the function $d_0 : x \mapsto d(0,x)$ is Lipschitz on $G \setminus U$. Moreover, the function $d_0^2 : x \mapsto d(0,x)^2$ is Lipschitz in a neighborhood of $0$.*

*Proof.* Thanks to Theorem 3.1, one easily shows that there are $L > 0$ and an open neighborhood $\Omega$ of the unit sphere $\{p : d(0,p) = 1\}$ such that $d_0$ is $L$-Lipschitz on $\Omega$.

[2] We show that $d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is not locally semi-concave at the point $((0,0),(1,0))$. Suppose there is a function $\phi \in \mathscr{C}^2(\mathbb{R}^2 \times \mathbb{R}^2)$ with $\phi((0,0),(1,0)) = d((0,0),(1,0)) = 1$ and $\phi((x,y),(\bar{x},\bar{y})) \ge d((x,y),(\bar{x},\bar{y}))$ for $(x,y) \sim (0,0)$ and $(\bar{x},\bar{y}) \sim (1,0)$. Set $\psi(t) := \phi((0,0),(1,t))$. Then $\psi \in \mathscr{C}^2(\mathbb{R})$, $\psi(0) = 1$ and $\psi(t) \ge 1 + |t|$, which is impossible.

[3] If $f : [0,1] \to \mathbb{R}$ is a 1-Lipschitz map with $f(0) = 0$ and $f(1) = 0$, then $\gamma(t) := (t, f(t))$ is a length-minimizer from $(0,0)$ to $(0,1)$ for the $\ell^\infty$-distance on $\mathbb{R}^2$. Moreover, convergence in $W^{1,\infty}([0,1])$ and in $W^{1,\infty}([0,1];\mathbb{R}^2)$ are equivalent for such curves. Hence, the set of all length-minimizers from $(0,0)$ to $(0,1)$ contains as a closed subset the unit ball of $W^{1,\infty}([0,1])$, which is not compact.

Next, we claim that $d_0$ is locally $L$-Lipschitz on $G \setminus B_d(0,1)$. Indeed, let $r > 0$ be such that $B_\rho(x,r) \subset \Omega$ for all $x \in S_d(0,1)$. If $q_1, q_2 \in G \setminus B_d(0,1)$ are such that $\rho(q_1,q_2) < r$, then there is $o \in G$ such that $d(0,q_1) = d(0,o) + d(o,q_1)$ and $d(o,q_1) = 1$, therefore

$$d_0(q_2) - d_0(q_1) \le d(o,q_2) - d(o,q_1) \le L\rho(o^{-1}q_2, o^{-1}q_1) = L\rho(q_2,q_1).$$

In the second step of the proof, we prove that $d_0$ is $L$-Lipschitz on $G \setminus B_d(0,1)$. Let $p, q \in G \setminus B_d(0,1)$ and $\gamma : [0,1] \to G$ a $\rho$-length minimizing curve from $p$ to $q$. If $\Im\gamma \subset G \setminus B_d(0,1)$, then there are $0 = t_0 \le t_1 \le \cdots \le t_{k+1} = 1$ such that $d_0(\gamma(t_i)) - d_0(\gamma(t_{i+1})) \le L\rho(\gamma(t_i), \gamma(t_{i+1}))$ for all $i$. Hence

$$\begin{aligned} d_0(p) - d_0(q) &= \sum_{i=0}^{k} d_0(\gamma(t_i)) - d_0(\gamma(t_{i+1})) \\ &\le L \sum_{i=0}^{k} \rho(\gamma(t_i), \gamma(t_{i+1})) = L\rho(p,q). \end{aligned}$$

If instead $\Im\gamma \cap B_d(0,1) \ne \emptyset$, then there are $0 < s < t < 1$ such that $d_0(\gamma(s)) = d_0(\gamma(t)) = 1$ and $\gamma([0,s]) \subset G \setminus B_d(0,1)$ and $\gamma([t,1]) \subset G \setminus B_d(0,1)$. Then

$$\begin{aligned} d_0(p) - d_0(q) &= d_0(p) - d_0(\gamma(s)) + d_0(\gamma(t)) - d_0(q) \\ &\le L\left(\rho(p,\gamma(s)) + \rho(\gamma(t),q)\right) \le L\rho(p,q). \end{aligned}$$

Finally, let $p, q \in G \setminus B_d(0,r)$ for $0 < r < 1$. Then $\delta_{r^{-1}}p, \delta_{r^{-1}}q \in G \setminus B_d(0,1)$ and we have

(3.2)
$$d_0(p) - d_0(q) = r(d_0(\delta_{r^{-1}}p) - d_0(\delta_{r^{-1}}q)) \le Lr\rho(\delta_{r^{-1}}p, \delta_{r^{-1}}q) \le \frac{CL}{r}\rho(p,q),$$

where we used in the last step the fact that there exists $C > 0$ such that

$$\forall p, q \in G,\ \forall r \in (0,1) \qquad \rho(\delta_{r^{-1}}p, \delta_{r^{-1}}q) \le Cr^{-2}\rho(p,q).$$

Now, we need to prove that $d_0^2$ is Lipschitz on $B_d(0,1)$. We first claim that $d_0^2$ is locally $4L$-Lipschitz on $B_d(0,1) \setminus \{0\}$. Indeed, if $p, q \in B_d(0,1) \setminus \{0\}$ are such that

$$\frac{1}{2} \le \frac{d_0(p)}{d_0(q)} \le 2,$$

then

$$0 < d_0(p) + d_0(q) \le 4\min\{d_0(p), d_0(q)\}.$$

Therefore, using (3.2),

$$\begin{aligned} d_0(p)^2 - d_0(q)^2 &= (d_0(p) + d_0(q))(d_0(p) - d_0(q)) \\ &\le (d_0(p) + d_0(q)) \frac{CL}{\min\{d_0(p), d_0(q)\}} \rho(p,q) \\ &\le (d_0(p) + d_0(q)) \frac{4CL}{d_0(p) + d_0(q)} \rho(p,q) = 4CL\rho(p,q). \end{aligned}$$

Finally, using again the fact that $\rho$ is a geodesic distance, we get that $d_0^2$ is $4CL$ Lipschitz on $B_d(0,1) \setminus \{0\}$ and therefore on $B_d(0,1)$. $\square$

### 3.1. **About the minimal stretching.**

**Definition 3.4** (Minimal Stretching)**.** Let $(X, \|\cdot\|)$ and $(Y, \|\cdot\|)$ be normed vector spaces. We define for a continuous linear map $L : X \to Y$ the *minimal stretching*

$$\tau(L) := \inf\{\|y\| : y \in Y \setminus L(B_X(0,1))\}$$

where $B_X(p,r) = \{q \in X : \|q - p\| < r\}$.

It is easy to prove that $\tau : L(X;Y) \to [0,+\infty)$ is continuous, where $L(X;Y)$ is the space of continuous linear mappings $X \to Y$ endowed with the operator norm.

The next proposition applies this notion to smooth functions and it is a restatement of [15, Theorem 1].

**Proposition 3.5.** *Let $(X, \|\cdot\|)$ and $(Y, \|\cdot\|)$ be two Banach spaces and $F : \Omega \to Y$ a $\mathscr{C}^1$ map, where $\Omega \subset X$ is open. Fix $\hat{x} \in \Omega$ and let $\tau_0 := \tau(\mathrm{d}F(\hat{x})) > 0$. Then for every $C > 1$ there is $\hat{\epsilon} > 0$, such that for all $0 < \epsilon < \hat{\epsilon}$ it holds*

$$B_Y(F(\hat{x}), \epsilon) \subset F\left(B_X(\hat{x}, \frac{C}{\tau_0}\epsilon)\right).$$

### 3.2. **The End-point map is weakly* continuous.**

As before, let $(\mathbb{E}, \mathtt{f}, \|\cdot\|)$ be a sub-Finsler structure on a manifold $M$. We want to prove the following proposition.

**Proposition 3.6.** *Fix $o \in M$ and let $o_k \in M$ be a sequence converging to $o$. Let $u_k \in \mathtt{L}^\infty([0,1];\mathbb{E})$ be a sequence of controls weakly* converging to $u \in \mathtt{L}^\infty([0,1];\mathbb{E})$. Let $\gamma_k$ (resp. $\gamma$) be the curve with control $u_k$ (resp. $u$) and $\gamma_k(0) = o_k$ (resp. $\gamma(0) = o$). Then $\gamma_k$ uniformly converge to $\gamma$.*

In particular, it follows that the End-point map $\mathtt{End}_o : \mathtt{L}^\infty([0,1];\mathbb{E}) \to M$ is weakly* continuous.

*Proof.* Since the sequence $u_k$ is bounded in $\mathtt{L}^\infty([0,1];\mathbb{E})$ by the Banach-Steinhaus Theorem and the sequence $o_k$ is bounded in $(M,d)$, then there is a compact set $K \subset M$ such that $\gamma_k \subset K$ for all $k$. Let $R > 0$ be such that $\|u_k\|_{\mathtt{L}^\infty} \le R$ for all $k \in \mathbb{N}$.

Thanks to the Whitney Embedding Theorem, we can assume that $M$ is a submanifold of $\mathbb{R}^N$ for some $N \in \mathbb{N}$. Fix a basis $e_1, \dots, e_r$ of $\mathbb{E}$ and define the vector fields $X_i : M \to \mathbb{R}^N$ as

$$X_i(p) := \mathtt{f}(p, e_i).$$

Since they are smooth, they are $L$-Lipschitz on $K$ for some $L > 0$. We extend the vector fields $X_i : M \to \mathbb{R}^N$ to smooth functions $X_i : \mathbb{R}^N \to \mathbb{R}^N$.

Define $\eta_k : [0,1] \to \mathbb{R}^N$ as

$$\eta_k(t) := o_k + \int_0^t u_k^i(s) X_i(\gamma(s))\,\mathrm{d}s.$$

Since $t \mapsto X_i(\gamma(t)) \in \mathbb{R}^N$ are continuous, then $u_k^i X_i(\gamma) \overset{*}{\rightharpoonup} u^i X_i(\gamma)$, for all $i \in \{1,\dots,r\}$. In particular, $\eta_k(t) \to \gamma(t)$ for each $t \in [0,1]$. Moreover, since the $\eta_k$'s have uniformly bounded derivative, they are a pre-compact family of curves with respect to the topology of uniform convergence. This fact and the pointwise convergence imply that $\eta_k \to \gamma$ uniformly on $[0,1]$.

Set $\epsilon_k := \sup_{t\in[0,1]} |\eta_k(t) - \gamma(t)| + 2|o_k - o|$, so that $\epsilon_k \to 0$, where $|\cdot|$ is the usual norm in $\mathbb{R}^N$.

By Ascoli-Arzelá Theorem, the family of curves $\{\gamma_k\}_k$ is also precompact with respect to the uniform convergence. Hence, if we prove that the only accumulation curve of $\{\gamma_k\}_k$ is $\gamma$, then we obtain that $\gamma_k$ uniformly converges to $\gamma$. So, we can assume $\gamma_k \to \xi$ uniformly for some $\xi : [0,1] \to M$. Then we have (sums on $i$ are hidden)

$$\begin{aligned}
|\gamma_k(t) - \gamma(t)| &\leq |o_k - o| + \left| \int_0^t u_k^i(s) X_i(\gamma_k(s)) - u^i(s) X_i(\gamma(s)) \,\mathrm{d}s \right| \\
&\leq |o_k - o| + \int_0^t |u_k^i(s) X_i(\gamma_k(s)) - u_k^i(s) X_i(\gamma(s))| \,\mathrm{d}s + \\
&\qquad + \left| \int_0^t u_k^i(s) X_i(\gamma(s)) - u^i(s) X_i(\gamma(s)) \,\mathrm{d}s \right| \\
&\leq 2|o_k - o| + rRL \int_0^t |\gamma_k(s) - \gamma(s)| \,\mathrm{d}s + |\eta_k(t) - \gamma(t)| \\
&\leq rRL \int_0^t |\gamma_k(s) - \gamma(s)| \,\mathrm{d}s + \epsilon_k.
\end{aligned}$$

Passing to the limit $k \to \infty$, we get for all $t \in [0,1]$

$$|\xi(t) - \gamma(t)| \leq rRL \int_0^t |\xi(s) - \gamma(s)| \,\mathrm{d}s. \tag{3.3}$$

Starting with the fact that $\|\xi - \gamma\|_{\mathtt{L}^\infty} \leq C$ for some $C > 0$ and iterating the previous inequality, we claim that

$$|\xi(t) - \gamma(t)| \leq C \frac{(rRLt)^j}{j!} \qquad \forall j \in \mathbb{N},\ \forall t \in [0,1].$$

Indeed, by induction, from (3.3) we get

$$|\xi(t) - \gamma(t)| \leq rRL \int_0^t C \frac{(rRL)^j}{j!} t^j \,\mathrm{d}s = C \frac{(rRL)^{j+1}}{j!} \frac{t^{j+1}}{j+1}.$$

Finally, since $\lim_{j\to\infty} \frac{(rRLt)^j}{j!} = 0$, we have $|\xi(t) - \gamma(t)| = 0$ for all $t$. □

3.3. **The differential of the End-point map is an End-point map.** The End-point map behaves like the exponential map: its differential is again an End-point map. In order to make this statement precise, we consider the case $M = \mathbb{R}^n$. Notice that we don't need any bracket-generating condition. In Corollary 3.8 we will use the results on $\mathbb{R}^n$ to prove a statement for all manifolds.

Let $\mathtt{f} : \mathbb{R}^n \times \mathbb{E} \to \mathbb{R}^n$ be a smooth map. Given a basis $e_1, \dots, e_r$ of $\mathbb{E}$, we define the vector fields $X_i : \mathbb{R}^n \to \mathbb{R}^n$ as

$$X_i(p) := \mathtt{f}(p, e_i).$$

The differential of the End-point map with base point 0 is the map

$$\begin{array}{rccc}
\mathtt{dEnd}_0 : & \mathtt{L}^\infty([0,1];\mathbb{E}) \times \mathtt{L}^\infty([0,1];\mathbb{E}) & \to & \mathbb{R}^n \\
& (u, v) & \mapsto & \mathtt{dEnd}_0(u)[v].
\end{array}$$

Define $Y_i, Z_i : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ for $i = 1, \dots, r$ as

$$\begin{cases} Y_i(p,q) := (X_i(p),\, \mathrm{d}X_i(p)[q]) \\ Z_i(p,q) := (0, X_i(p)) \end{cases}$$

where $\mathrm{d}X_i(p) : \mathbb{R}^n \to \mathbb{R}^n$ is the differential of $X_i$ at $p$. These vector fields induce a new End-point map

$$\mathtt{End}_{00} : \mathtt{L}^\infty([0,1]; \mathbb{E} \times \mathbb{E}) \to \mathbb{R}^n \times \mathbb{R}^n$$

with starting point $(0,0) \in \mathbb{R}^n \times \mathbb{R}^n$.

**Proposition 3.7.** *For all $u, v \in \mathtt{L}^\infty([0,1]; \mathbb{E})$ it holds*

$$(\mathtt{End}_0(u),\, \mathrm{d}\mathtt{End}_0(u)[v]) = \mathtt{End}_{00}(u,v).$$

The proof is immediate, and hence omitted, once one has an explicit representation of the differential $\mathrm{d}\mathtt{End}_0(u)[v]$, see [23]. This result, together with Proposition 3.6, gives us the weakly* continuity of the differential of the End-point map. The next corollary is an application.

**Corollary 3.8.** *Let $(\mathbb{E}, \mathtt{f}, \|\cdot\|)$ be a sub-Finsler structure on a manifold $M$ and $o \in M$. Let $\rho$ be a Riemannian metric on $M$. Then the map $\mathtt{L}^\infty([0,1]; \mathbb{E}) \to [0, +\infty)$, $u \mapsto \tau(\mathrm{d}\mathtt{End}_o(u))$ is weakly* lower semi-continuous, where $\tau$ is the minimal stretching computed with respect to the norm given by $\rho$.*

*Proof.* Let $\{u_k\}_k \subset \mathtt{L}^\infty([0,1]; \mathbb{E})$ be a sequence with $u_k \overset{*}{\rightharpoonup} u \in \mathtt{L}^\infty([0,1]; \mathbb{E})$. Let $\gamma_u$ be the curve with control $u$ and starting point $o$. We can pull back the sub-Finsler structure from a neighborhood of $\gamma_u$ to an open subset of $\mathbb{R}^n$ via a covering map, so that we reduce the statement to the case $M = \mathbb{R}^n$.

We need to prove

$$\liminf_{k\to\infty} \tau(\mathrm{d}\mathtt{End}_o(u_k)) \geq \tau(\mathrm{d}\mathtt{End}_o(u)). \tag{3.4}$$

Set $\hat\tau := \tau(\mathrm{d}\mathtt{End}_0(u))$. If $\hat\tau = 0$, then (3.4) is fulfilled, so we assume $\hat\tau > 0$. Let $\hat\tau > \epsilon > 0$. Then there exists a finite-dimensional vector space $W \subset \mathtt{L}^\infty([0,1]; \mathbb{E})$ such that

$$B_{\mathtt{End}_0(u)}(0, \hat\tau - \frac{\epsilon}{2}) \subset\subset \mathrm{d}\mathtt{End}_0(u)[B_{\mathtt{L}^\infty}(0,1) \cap W],$$

where, for $p \in \mathbb{R}^n$, $B_p$ denotes a ball in $\mathbb{R}^n = T_p\mathbb{R}^n$ with respect to the norm given by $\rho$ at $p$, and $B_{\mathtt{L}^\infty}$ denotes a ball in $\mathtt{L}^\infty([0,1]; \mathbb{E})$ with respect to the $\mathtt{L}^\infty$-norm induced by $\|\cdot\|$. Since $\dim W < \infty$ and by Propositions 3.6 and 3.7, the maps $\mathrm{d}\mathtt{End}_0(u_k)|_W$ strongly converge to $\mathrm{d}\mathtt{End}_0(u)|_W$. Moreover, the norm on $\mathbb{R}^n = T_p\mathbb{R}^n$ given by $\rho$ continuously depends on $p \in \mathbb{R}^n$. Therefore, for $k$ large enough we have

$$B_{\mathtt{End}_0(u_k)}(0, \hat\tau - \epsilon) \subset \mathrm{d}\mathtt{End}_0(u_k)[B_{\mathtt{L}^\infty}(0,1) \cap W].$$

Hence

$$\liminf_{k\to\infty} \tau(\mathrm{d}\mathtt{End}_0(u_k)) \geq \hat\tau - \epsilon.$$

Since $\epsilon$ is arbitrary, (3.4) follows. $\square$

3.4. **The sub-Finsler distance is Lipschitz in absence of singular geodesics.** The proof of Theorem 3.1 is divided into the next two lemmas from which Theorem 3.1 follows.

**Lemma 3.9.** *Let $o, p \in M$ such that all $d$-minimizing curves from $o$ to $p$ are regular. Then there exist a compact neighborhood $K \subset M$ of $p$ and a weakly* compact set $\mathscr{K} \subset \mathtt{L}^\infty([0,1];\mathbb{E})$ such that:*

*(1) $\mathtt{End}_o : \mathscr{K} \to K$ is onto;*
*(2) $d_{CC}(o, \mathtt{End}_o(u)) = \|u\|_{\mathtt{L}^\infty}$ for all $u \in \mathscr{K}$;*
*(3) every $u \in \mathscr{K}$ is a regular point for $\mathtt{End}_o$.*

*Proof.* For any compact neighborhood $K$ of $p$, define the compact set

$$\mathscr{K}(K) := \{u \in \mathtt{L}^\infty([0,1];\mathbb{E}) : \mathtt{End}_o(u) \in K \text{ and } d(o, \mathtt{End}_o(u)) = \|u\|_{\mathtt{L}^\infty}\}.$$

Since the metric $d$ is geodesic, the End-point map $\mathtt{End}_o : \mathscr{K}(K) \to K$ is surjective, for all $K$. Moreover, the second requirement holds by definition. Finally, suppose that there exist a sequence $p_k \to p$ and a sequence $u_k \in \mathtt{L}^\infty([0,1];\mathbb{E})$ such that $\mathtt{End}_o(u_k) = p_k$, $d(0, p_k) = \|u_k\|_{\mathtt{L}^\infty}$ and $u_k$ is a singular point for $\mathtt{End}_o$, for all $k$. Since the sequence $u_k$ is bounded, thanks to the BanachAlaoglu Theorem there is $u \in \mathtt{L}^\infty([0,1];\mathbb{E})$ such that, up to a subsequence, $u_k \xrightharpoonup{*} u$. By the continuity of $\mathtt{End}_o$, we have $\mathtt{End}_o(u) = p$. By Corollary 3.8, we have $\tau(\,\mathrm{d}\mathtt{End}_o(u)) \le \liminf_k \tau(\,\mathrm{d}\mathtt{End}_o(u_k)) = 0$. Finally, by the lower-semicontinuity of the norm with respect to the weak* topology, we have

$$\|u\|_{\mathtt{L}^\infty} \le \liminf_{k\to\infty} \|u_k\|_{\mathtt{L}^\infty} = \liminf_{k\to\infty} d(o, p_k) = d(o,p) \le \|u\|_{\mathtt{L}^\infty}.$$

So, $u$ is the control of a singular length-minimizing curve from $o$ to $p$, against the assumption. Therefore, there exists a neighborhood $K$ of $p$ such that $\mathscr{K}(K)$ contains only regular points for $\mathtt{End}_o$. □

**Lemma 3.10.** *Let $o \in M$, and $K \subset M$ compact. Suppose there is a weakly* compact set $\mathscr{K} \subset \mathtt{L}^\infty([0,1];\mathbb{E})$ that satisfies all three properties listed in Lemma 3.9. Then for every Riemannian metric $\rho$ on $M$ there exists $L > 0$ such that the function $d_o : p \mapsto d(o,p)$ is locally $L$-Lipschitz on the interior of $K$.*

*Proof.* Let $\rho$ be a Riemannian metric on $M$. We first prove that

$$\begin{gathered}\exists L > 0,\ \forall q \in K,\ \exists \hat\epsilon_q > 0\ \forall q' \in K \\ \big[\rho(q,q') < \hat\epsilon_q \Rightarrow d_o(q') - d_o(q) \le L\rho(q,q')\big].\end{gathered} \tag{3.5}$$

Since $\mathscr{K}$ is a weakly* compact set of regular points for $\mathtt{End}_o$, then by Corollary 3.8 the function $u \mapsto \tau(\,\mathrm{d}\mathtt{End}_o(u))$ admits a minimum on $\mathscr{K}$ that is strictly positive. By Proposition 3.5, there is $L > 0$ such that for every $u \in \mathscr{K}$ there is $\hat\epsilon_u > 0$ such that

$$B_\rho(\mathtt{End}_o(u), \epsilon) \subset \mathtt{End}_o(B_{\mathtt{L}^\infty}(u, L\epsilon)) \quad \forall \epsilon < \hat\epsilon_u. \tag{3.6}$$

Let $q, q' \in K$ be such that $q = \mathtt{End}_o(u)$ with $u \in \mathscr{K}$ and $\epsilon := \rho(q,q') < \hat\epsilon_u$. Then, by the inclusion (3.6), there is $u' \in B_{\mathtt{L}^\infty}(u, L\epsilon)$ with $\mathtt{End}_o(u') = q'$. So

$$d_o(q') - d_o(q) \le \|u'\|_{\mathtt{L}^\infty} - \|u\|_{\mathtt{L}^\infty} \le \|u' - u\|_{\mathtt{L}^\infty} \le L\epsilon = L\rho(q,q').$$

This proves the claim (3.5).

Finally, if $p$ is an interior point of $K$, then there is a $\rho$-convex neighborhood $U$ of $p$ contained in $K$ (see [24]). So, if $q, q' \in U$, then there is a $\rho$-geodesic $\xi : [0,1] \to U$ joining $q$ to $q'$. Since the image of $\xi$ is compact, there are $0 = t_1 < s_1 < t_2 < s_2 < \cdots < s_{k-1} < t_k = 1$ such that $\rho(\xi(t_i), \xi(s_i)) < \hat{\epsilon}_{\xi(t_i)}$ and $\rho(\xi(s_i), \xi(t_{i+1})) < \hat{\epsilon}_{\xi(t_{i+1})}$. Therefore

$$\begin{aligned} d_o(q') - d_o(q) &\leq \sum_{i=1}^{k-1} d_o(\xi(t_{i+1})) - d_o(\xi(s_i)) + d_o(\xi(s_i)) - d_o(\xi(t_i)) \\ &\leq L \sum_{i=1}^{k-1} \rho(\xi(t_{i+1}), \xi(s_i)) + \rho(\xi(s_i), \xi(t_i)) = L\rho(q, q'). \end{aligned}$$

□

## 4. Regularity of spheres in graded groups

This section is devoted to the proof of Theorems 1.2 and 1.3 and of the inequalities (1.2).

### 4.1. **The sphere as a graph.**

Let $(G, d)$ be a homogeneous group and $\rho$ a Riemannian distance on $G$. Theorem 1.2 and the estimate (1.2) are both based on the following remark.

*Remark* 4.1. Let $\mathfrak{g} = \oplus_{i>0} V_i$ be a grading for the Lie algebra of $G$. Let $|\cdot|$ be the norm of a scalar product on $\mathfrak{g}$ that makes the layers orthogonal to each other and let $S = \exp(\{v : |v| = 1\}) \subset G$. The hypersurface $S$ is smooth and transversal to the dilations, i.e., for all $p \in S$ we have $\frac{\mathrm{d}}{\mathrm{d}t}|_{t=1} \delta_t p \notin T_p S$. Define

$$\begin{array}{rccc} \phi : & S \times (0, +\infty) & \to & G \setminus \{0\} \\ & (p, t) & \mapsto & \delta_{\frac{1}{t}} p. \end{array}$$

Since $S$ is transversal to the dilations, $\phi$ is a diffeomorphism. Moreover, if $\Gamma := \{(p, d_0(p)) : p \in S\} \subset S \times (0, +\infty)$ is the graph of the function $d_0$ restricted to $S$, then

$$\mathbb{S}_d = \phi(\Gamma).$$

Thanks to the last remark, the estimate (1.2) follows from the next lemma.

**Lemma 4.2.** *Let $\Omega \subset \mathbb{R}^n$ be an open set and let $f : \Omega \to \mathbb{R}$ be an $\alpha$-Hölder function, i.e., for all $x, y \in \Omega$ we have*

$$|f(x) - f(y)| \leq C|x - y|^\alpha,$$

*for some $C > 0$, where $\alpha \in (0, 1]$. Define the graph of $f$ as*

$$\Gamma_f := \{(x, f(x)) : x \in \Omega\} \subset \mathbb{R}^{n+1}.$$

*Then*

$$n \leq \dim_H \Gamma_f \leq n + 1 - \alpha,$$

*where* $\dim_H$ *is the Hausdorff dimension. Moreover, this estimate is sharp, i.e., there exists $f$ such that* $\dim_H \Gamma_f = n + 1 - \alpha$.

The proof is straightforward by use of a simple covering argument or by an estimate of the Minkowski content of the graph. The sharpness of this result has been shown in [5] for the case $n = 1$. The general case, as stated here, is a simple consequence. Indeed, if $g : (0, 1) \to \mathbb{R}$ is a $\alpha$-Hölder function such that $\dim_H(\Gamma_g) = 2-\alpha$, then the graph of the function $f(x_1, \dots, x_n) := g(x_1)$ is $\Gamma_f = \Gamma_g \times (0, 1)^{n-1}$. Therefore, $\dim_H(\Gamma_f) = n + 1 - \alpha$.

In the next easy-to-prove lemma we point out that a homogenous distance is locally Lipschitz if and only if the spheres are Lipschitz graphs in the directions of the dilations.

**Lemma 4.3.** *Let $d$ be a homogeneous distance on $G$. Let $S$ and $\mathbb{S}_d$ be as in Remark 4.1 and $p \in S$. Then the following conditions are equivalent:*

*(i) Setting $\hat{p} := \delta_{d_0(p)^{-1}}(p) \in \mathbb{S}_d$, the sphere $\mathbb{S}_d$ is a Lipschitz graph in the direction $\bar{\delta}(\hat{p})$ in some neighborhood of $p$;*
*(ii) $d_0|_S : S \to (0, +\infty)$ is Lipschitz in some neighborhood of $p$ in $S$;*
*(iii) $d_0$ is Lipschitz in some neighborhood of $\delta_\lambda p$ for one, hence all, $\lambda > 0$.*

Thanks to Lemma 4.3, Theorem 1.2 is a consequence of Theorem 1.1.

4.2. **An intrinsic approach.** In this section we will prove Theorem 1.3. We define a *cone* in $\mathbb{R}^n$ as

$$\texttt{Cone}(\alpha, h, v) := \{x \in \mathbb{R}^n : |x| \le h \text{ and } \angle(x, v) \le \alpha\} \subset \mathbb{R}^n,$$

where $\alpha \in [0, \pi]$, $h \in (0, +\infty]$, $v \in \mathbb{R}^n$ is the axis of the cone, and $\angle(x, v)$ is the angle between $x$ and $v$. The following lemma is a simple calculus exercise and it will be used later in the proof of Theorem 1.3. Roughly speaking, it states that a small smooth deformation of a cone still contains a cone with the same tip.

**Lemma 4.4.** *Let $m, k, n \in \mathbb{N}$, $p \in \mathbb{R}^m$ and $y_0 \in \mathbb{R}^k$. Let $\phi : \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}^n$ be a smooth map such that $\mathrm{d}(\phi_p)(y_0) : \mathbb{R}^k \to \mathbb{R}^n$ is surjective, where $\phi_x(y) := \phi(x, y)$. Let $C' \subset \mathbb{R}^k$ be a cone with axis $v' \in \mathbb{R}^k$. Then there exist a cone $C \subset \mathbb{R}^n$ with axis $\mathrm{d}(\phi_p)(y_0)v'$ and an open neighborhood $U \subset \mathbb{R}^m$ of $p$ such that for all $q \in U$*

$$\phi_q(y_0) + C \subset \phi_q(y_0 + C').$$

*Proof of Theorem 1.3.* In this proof, we consider the dilations $\delta_\lambda$ as defined for $\lambda \le 0$ too, with the same definition as for $\lambda > 0$. Notice that in this way the map $G \times \mathbb{R} \to G$, $(p, \lambda) \mapsto \delta_\lambda p$, is a smooth map.

Let $v_1, \dots, v_r$ be a basis for $V_1$ and set $p_i := \exp(v_i) \in G$. Up to a rescaling, we can assume $d_0(p_i) < 1$ for all $i$. For $p \in G$ define $\phi_p : \mathbb{R}^{2r+1} \to G$ as

$$\phi_p(u_1, \dots, u_r, s, v_1, \dots, v_r) = \delta_{u_1} p_1 \cdots \delta_{u_r} p_r \cdot \delta_s p \cdot \delta_{v_1} p_1 \cdots \delta_{v_r} p_r.$$

Let $\hat{x} \in \mathbb{R}^{2r+1}$ be the point with $u_i = 0$, $s = 1$ and $v_i = 0$, so that $\phi_p(\hat{x}) = p$. The differential of $\phi_p$ at $\hat{x}$ is given by the partial derivatives

$$\begin{aligned}
\frac{\partial \phi_p}{\partial u_i}(\hat{x}) &= \frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}\,(\delta_t p_i \cdot p) = \mathrm{d}R_p\left(\frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}(\delta_t p_i)\right) = \mathrm{d}R_p(v_i),\\
\frac{\partial \phi_p}{\partial s}(\hat{x}) &= \frac{\mathrm{d}}{\mathrm{d}t}|_{t=1}\,(\delta_t p) = \bar{\delta}(p),\\
\frac{\partial \phi_p}{\partial v_i}(\hat{x}) &= \frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}\,(p \cdot \delta_t p_i) = \mathrm{d}L_p\left(\frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}(\delta_t p_i)\right) = \mathrm{d}L_p(v_i).
\end{aligned}$$

Therefore, if $p \in \mathbb{S}_d$ is such that the condition (1.1) is true, then the differential $\mathrm{d}\phi_p$ has full rank at $\hat{x}$, hence in a neighborhood of $\hat{x}$.

Define

$$\Delta := \{(u_1, \dots, u_r, s, v_1, \dots, v_r) \in \mathbb{R}^{2r+1} : s + \sum_{i=1}^{r}(|u_i| + |v_i|) \leq 1\}.$$

We identify $G$ with $\mathbb{R}^n$ through an arbitrary diffeomorphism. By Lemma 4.4, there is a cone $C$ with axis $\bar{\delta}(p)$ and a neighborhood $U$ of $p$ such that for all $q \in U$

$$q + C \subset \phi_q(\Delta).$$

Up to restricting $U$, for all $q \in U$ there are cones $C_q$ with axis $\bar{\delta}(q)$, fixed amplitude and fixed height such that $q + C_q \subset q + C$. Notice that for all $q \in \mathbb{S}_d$ we have $\phi_q(\Delta) \subset \bar{B}_d(0,1)$ and $\phi_q(\Delta) \cap \mathbb{S}_d = \{q\}$. In particular, for all $q \in \mathbb{S}_d \cap U$, we have $q + C_q \subset \bar{B}_d(0,1)$ and $(q + C_q) \cap \mathbb{S}_d = \{q\}$, i.e., $\mathbb{S}_d \cap U$ is a Lipschitz graph in the direction of the dilations. Thanks to Lemma 4.3, we get that $d_0$ is Lipschitz in a neighborhood of $p$. □

Finally, some considerations on condition (1.1) are due.

*Remark* 4.5. If (1.1) holds at $p \in G$ and $u \in \mathtt{L}^\infty([0,1]; V_1)$ is a control such that $\mathtt{End}_0(u) = p$, then the differential $\mathrm{d}\mathtt{End}_0(u)$ is surjective, i.e., $p$ is a regular value of $\mathtt{End}$. Indeed, by [20] (see (2.6) and (2.11) there), we have

$$\mathrm{d}L_p(V_1) + \mathrm{d}R_p(V_1) + \mathrm{span}\{\bar{\delta}(p)\} \subset \Im(\mathrm{d}\mathtt{End}_0(u)),$$

because $q \mapsto \bar{\delta}(q)$ is a contact vector field of $G$.

**Proposition 4.6.** *Let $X \in V_1$. If* (1.1) *holds for $p = \exp(X)$, then*

$$\mathfrak{g} = V_1 + [X, V_1].$$

*Proof.* Let $X_1, \dots, X_r$ be a basis for $V_1$ and $Y_1, \dots, Y_\ell$ a basis for $[X, V_1]$. Let $\alpha^i_j \in \mathbb{R}$ be such that $[X, X_i] = \sum_{j=1}^{\ell} \alpha^i_j Y_j$. First, notice that

$$\begin{aligned}
T_0 G &= \mathrm{d}L_{\exp(-X)}\left(\mathrm{d}L_{\exp(X)}(V_1) + \mathrm{d}R_{\exp(X)}(V_1)\right)\\
&= V_1 + \mathrm{d}L_{\exp(-X)} \circ \mathrm{d}R_{\exp(X)}(V_1)\\
&= V_1 + \mathrm{Ad}_{\exp(X)}(V_1).
\end{aligned}$$

Then, using the formula $\mathrm{Ad}_{\exp(X)}(Y) = e^{\mathrm{ad}_X}(Y) = \sum_{k=0}^{\infty} \frac{1}{k!}\mathrm{ad}_X^k(Y)$, we have

$$\begin{aligned}
\mathrm{Ad}_{\exp(X)}(X_i) &= X_i + \left(\sum_{k=1}^{\infty} \frac{1}{k!}\mathrm{ad}_X^{k-1}([X,X_i])\right)\\
&= X_i + \left(\sum_{k=1}^{\infty} \mathrm{ad}_X^{k-1}(\sum_{j=1}^{\ell} \alpha_j^i Y_j)\right)\\
&= X_i + \sum_{j=1}^{\ell} \alpha_j^i \left(\sum_{k=1}^{\infty} \mathrm{ad}_X^{k-1}(Y_j)\right).
\end{aligned}$$

It follows that $\dim\left(V_1 + \mathrm{Ad}_{\exp(X)}V_1\right) \leq r + \ell$ and therefore $\dim \mathfrak{g} \leq r + \ell$, i.e., $\mathfrak{g} = V_1 + [X, V_1]$. $\square$

**Proposition 4.7.** *Let $Z \in V_k$, where $k > 0$ is such that $V_i = \{0\}$ for all $i > k$. If* (1.1) *holds for $p = \exp(Z)$, then*

$$\mathfrak{g} = V_1 + \mathrm{span}\{Z\}.$$

*Proof.* Since $[Z, \mathfrak{g}] = \{0\}$, we have $R_p = L_p$. Moreover, $\bar{\delta}(p) = \mathrm{d}L_p(kZ)$. So, condition (1.1) becomes $\mathrm{d}L_p(V_1) + \mathrm{d}L_p(\mathrm{span}\{Z\}) = T_pG$. $\square$

In particular, if (1.1) holds for all $p \in G \setminus \{0\}$, then $\mathfrak{g} = V_1 \oplus V_2$ with $\dim V_2 \leq 1$ and $[X, V_1] = V_2$ for all non-zero $X \in V_1$.

## 5. Examples

5.1. **Three gradings on $\mathbb{R}^2$.** We will present three examples of dilations on $\mathbb{R}^2$. In particular we want to illustrate two applications of Theorem 1.3 and show the sharpness of the dimension estimate (1.2). In Remark 5.5 we give an easy example of a homogeneous distance whose unit ball is a Lipschitz domain, but the distance is not locally Lipschitz away from the diagonal.

The first and the easiest is

$$\delta_\lambda(x, y) := (\lambda x, \lambda y),$$

which gives rise to the known structure of vector space. Here, homogeneous distances are given by norms and balls are convex, hence Lipschitz domains. It's trivial to see that condition (1.1) holds for all $p \in \mathbb{R}^2$.

The second example is given by the dilations

$$\delta_\lambda(x, y) := (\lambda x, \lambda^2 y).$$

In this case, $\mathbb{R}^2 = V_1 \oplus V_2$ with $V_1 = \mathbb{R} \times \{0\}$ and $V_2 = \{0\} \times \mathbb{R}$, and $\bar{\delta}(x, y) = (x, 2y)$. Condition (1.1) holds for all $(x, y) \in \mathbb{R}^2$ with $y \neq 0$. One can actually show that, for any homogeneous metric on $(\mathbb{R}^2, \delta_\lambda)$ with closed unit ball $B$ centered at 0, the set $I = \{x \in \mathbb{R} : (x, 0) \in B\}$ is a closed interval and there exists a function $f : I \to \mathbb{R}$ that is locally Lipschitz on the interior of $I$ such that

$$\mathbb{S}_d \cap \{(x, y) : y \geq 0\} = \{(x, f(x)) : x \in I\}.$$

We will prove a similar statement in the Heisenberg group with an argument that applies here too, see Section 6.

The third example is given by the dilations

$$\delta_\lambda(x, y) := (\lambda^2 x, \lambda^2 y), \tag{5.1}$$

and it is interesting because of the next proposition.

**Proposition 5.1.** *There exists a homogeneous (with respect to dilations* (5.1)*) distance $d$ on $\mathbb{R}^2$ whose unit sphere has Euclidean Hausdorff dimension $\frac{3}{2}$.*

Notice that $\frac{3}{2}$ is the maximal Hausdorff dimension that one gets by the estimate (1.2).

For proving Proposition 5.1, we need to find a set $B \subset \mathbb{R}^2$ that satisfies all four conditions listed in Lemma 2.15, in particular

$$\forall p, q \in B,\ \forall t \in [0, 1] \quad t^2 p + (1 - t)^2 q \in B. \tag{5.2}$$

One easily proves the following preliminary facts.

**Lemma 5.2.** *Let $p, q \in \mathbb{R}^2$ and $\gamma : [0, 1] \to \mathbb{R}^2$, $\gamma(t) := t^2 p + (1 - t)^2 q$.*

1. *The curve $\gamma$ is contained in the triangle of vertices $0, p, q$.*
2. *The curve $\gamma$ is an arc of the parabola passing through $p$ and $q$ and that is tangent to the lines $\mathrm{span}\{p\}$ and $\mathrm{span}\{q\}$.*
3. *If $B$ satisfies* (5.2) *and $A : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map, then $A(B)$ satisfies* (5.2) *as well.*

**Lemma 5.3.** *For $0 < C \le 1$, define*

$$Y_C := \{(x, y) : |x| \le 1,\ y \le 1 + C\sqrt{|x|}\}.$$

*Then $Y_C$ satisfies* (5.2).

*Proof.* Let $p, q \in Y_C$ and set $\gamma(t) = (\gamma_x(t), \gamma_y(t)) := t^2 p + (1 - t)^2 q$.

If both $p$ and $q$ stay on one side with respect to the vertical axis, then $\gamma(t) \in Y_C$ for all $t \in [0, 1]$ thanks to the first point of Lemma 5.2 and because the two sets $Y_C \cap \{x \ge 0\}$ and $Y_C \cap \{x \le 0\}$ are convex.

Therefore, we suppose that

$$p = (-p_x, p_y) \qquad q = (q_x, q_y)$$

with $p_x, q_x > 0$. Let $t_0 \in [0, 1]$ be the unique value such that $\gamma_x(t_0) = 0$. Then the curve $\gamma$ lies in the union of the two triangles with vertices $0, \gamma(0), \gamma(t_0)$ and $0, \gamma(1), \gamma(t_0)$, respectively. Therefore, $\gamma$ lies in $Y_C$ if and only if $\gamma_y(t_0) \le 1$. Solving the equation $\gamma_x(t_0) = t_0^2(q_x - p_x) - 2q_x t_0 + q_x = 0$, one gets

$$t_0 = \frac{\sqrt{q_x}}{\sqrt{q_x} + \sqrt{p_x}}, \qquad (1 - t_0) = \frac{\sqrt{p_x}}{\sqrt{q_x} + \sqrt{p_x}}.$$

From the expression of $\gamma_y(t_0) = t_0^2 p_y + (1 - t_0)^2 q_y$, we notice that, $p_x$ and $q_x$ fixed, the worst case is when $p_y$ and $q_y$ are maximal, i.e.,

$$p_y = 1 + C\sqrt{p_x}, \qquad q_y = 1 + C\sqrt{q_x}.$$

Finally

$$\begin{aligned}\gamma_y(t_0) &= t_0^2 p_y + (1-t_0)^2 q_y\\ &= \frac{q_x}{(\sqrt{q_x}+\sqrt{p_x})^2}(1+C\sqrt{p_x}) + \frac{p_x}{(\sqrt{q_x}+\sqrt{p_x})^2}(1+C\sqrt{q_x})\\ &= \frac{1}{(\sqrt{q_x}+\sqrt{p_x})^2}\left(q_x+p_x+Cq_x\sqrt{p_x}+Cp_x\sqrt{q_x}\right)\\ &= 1+\frac{-2\sqrt{p_xq_x}+Cq_x\sqrt{p_x}+Cp_x\sqrt{q_x}}{(\sqrt{q_x}+\sqrt{p_x})^2}\\ &= 1+\sqrt{p_xq_x}\frac{-2+C(\sqrt{q_x}+\sqrt{p_x})}{(\sqrt{q_x}+\sqrt{p_x})^2}.\end{aligned}$$

Since $-2+C(\sqrt{q_x}+\sqrt{p_x})\le 0$, then we have $\gamma_y(t_0)\le 1$, as desired. □

**Lemma 5.4.** *Let $\alpha,\beta>0$. For all $0<\epsilon\le\alpha$ and all $0<C\le\frac{\beta}{\sqrt{\alpha}}$, the set*

$$Y(\epsilon,\beta,C) := \{(x,y) : |x|\le\epsilon,\ y\le\beta+C\sqrt{|x|}$$

*satisfies* (5.2).

*Proof.* Define the linear map $A(x,y) := (\alpha x,\beta y)$ and set $C' := C\frac{\sqrt{\alpha}}{\beta}\le 1$. Then one just needs to check that

$$Y(\epsilon,\beta,C) = A(Y_{C'})\cap\{(x,y) : |x|\le\epsilon\},$$

where $Y_{C'}$ is defined as in the previous Lemma 5.3. □

*Proof of Proposition 5.1.* First of all, let $\theta_0>0$ be such that for all $|\theta|\le\theta_0$ it holds

$$\frac{|\theta|}{2}\le|\cos(\frac{\pi}{2}+\theta)| = |\sin\theta|\le 2|\theta|. \tag{5.3}$$

Moreover, let $L,m,M,C>0$ be such that

$$\frac{L\sqrt{2}}{\sqrt{m}}\le C\le\frac{m}{\sqrt{2M\theta_0}}.$$

Let $f:\mathbb{R}\to(0,+\infty)$ be a function such that

$$\forall s,t\in\mathbb{R}\quad |f(t)-f(s)|\le L\sqrt{|t-s|}, \tag{5.4}$$

$$\forall t\in\mathbb{R}\quad m\le f(t)\le M. \tag{5.5}$$

We claim that, for $|\theta|\le\theta_0$, we have

$$f\left(\frac{\pi}{2}+\theta\right)\cdot\left(\cos(\frac{\pi}{2}+\theta),\sin(\frac{\pi}{2}+\theta)\right)\in Y\left(2M\theta_0, f(\frac{\pi}{2}), C\right) \tag{5.6}$$

where $Y\left(2M\theta_0, f(\frac{\pi}{2}), C\right)$ is defined as in Lemma 5.4. Indeed, we have on one side

$$|x| := |f(\frac{\pi}{2}+\theta)\cos(\frac{\pi}{2}+\theta)|\le M2|\theta|\le 2M\theta_0.$$

On the other side,

$$\begin{aligned} y &:= f(\frac{\pi}{2}+\theta)\sin(\frac{\pi}{2}+\theta) \le f(\frac{\pi}{2}+\theta) \\ &\le f(\frac{\pi}{2}) + f(\frac{\pi}{2}+\theta) - f(\frac{\pi}{2}) \le f(\frac{\pi}{2}) + L\sqrt{|\theta|} \\ &\le f(\frac{\pi}{2}) + L\frac{\sqrt{2\cos(\frac{\pi}{2}+\theta)f(\frac{\pi}{2}+\theta)}}{\sqrt{f(\frac{\pi}{2}+\theta)}} \le f(\frac{\pi}{2}) + \frac{\sqrt{2}L}{\sqrt{m}}\sqrt{|x|} \\ &\le f(\frac{\pi}{2}) + C\sqrt{|x|}. \end{aligned}$$

So (5.6) is satisfied.

Since for $\alpha := 2M\theta_0$ and $\beta := f(\frac{\pi}{2})$ we have

$$\frac{\beta}{\sqrt{\alpha}} = \frac{f(\frac{\pi}{2})}{\sqrt{2M\theta_0}} \ge \frac{m}{\sqrt{2M\theta_0}} \ge C,$$

Lemma 5.4 applies and we get that $Y\left(2M\theta_0, f(\frac{\pi}{2}), C\right)$ satisfies (5.2).

For any $\theta$ we set $A_\theta$ to be the counterclockwise rotation of angle $\theta$:

$$A_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Define the curve $\phi(t) := f(t)(\cos t, \sin t)$. Notice that $A_\theta\phi(t) = f((t-\theta)+\theta)(\cos(t+\theta), \sin(t+\theta))$ and that the function $s \mapsto f(s+\theta)$ is still satisfying both (5.4) and (5.5). So we have that, for $|t|, |s| < \frac{\theta_0}{2}$,

$$\phi(\frac{\pi}{2}+t) \in A_s[Y(2M\theta_0, f(\frac{\pi}{2}+s), C)]$$

and the set $A_s[Y(2M\theta_0, f(\frac{\pi}{2}+s), C)]$ satisfies (5.2).

Set

$$B := \bigcap_{|s|<\frac{\theta_0}{2}} \left(A_s[Y(2M\theta_0, f(\frac{\pi}{2}+s), C)] \cap -A_s[Y(2M\theta_0, f(\frac{\pi}{2}+s), C)]\right).$$

The set $B$ satisfies all three conditions of Lemma 2.15, hence it is the unit ball of a homogeneous metric. Moreover,

$$\{\phi(\frac{\pi}{2}+t) : |t| < \frac{\theta_0}{2}\} \subset \partial B.$$

The statement of Proposition 5.1 follows because there are functions $f : \mathbb{R} \to [0, +\infty)$ that satisfy (5.4) and (5.5) and such that the image of the curve $\phi$ has Hausdorff dimension $\frac{3}{2}$. Indeed, the image of $\phi$ has the same Hausdorff dimension of the graph of $f$, and then one uses the sharpness of Lemma 4.2. $\square$

*Remark* 5.5. Using the same arguments as in the proof of Lemma 5.3, one easily shows that the set

$$B := \{(x,y) \in \mathbb{R}^2 : \ |x| \le 1, \ -f(-x) \le y \le f(x)\},$$

where

$$f(x) := \begin{cases} 1 & x \le 0 \\ 1+\sqrt{x} & x > 0, \end{cases}$$

is the unit ball of a homogeneous distance on $\mathbb{R}^2$ with dilations (5.1). Notice that such $B$ is a Lipschitz domain, but the associated homogeneous distance is not Lipschitz in any neighborhood of the point $(0,1)$, thanks to Lemma 4.3.

5.2. **The Heisenberg groups.** In the Heisenberg groups $\mathbb{H}^n$ (for an introduction see [10]) condition (1.1) holds at every non-zero point. Therefore, balls of any homogeneous metric on $\mathbb{H}^n$ are Lipschitz domains. We will treat more in detail the first Heisenberg group in Section 6.

5.3. **A sub-Finsler sphere with a cusp.** Let $\mathbb{H}$ be the first Heisenberg group (see Section 6 for the definition). The group $G = \mathbb{H} \times \mathbb{R}$ is a stratified group with grading $(V_1 \times \mathbb{R}) \oplus V_2$, where $V_1 \oplus V_2$ is a stratification for $\mathbb{H}$. The line $\{0_{\mathbb{H}}\} \times \mathbb{R}$ is a singular curve in $G$. Moreover, it has been shown in [8] that there exists a sub-Finsler distance on $G$ whose unit sphere $\mathbb{S}_d$ has a cusp in the intersection $\mathbb{S}_d \cap (\{0_{\mathbb{H}}\} \times \mathbb{R})$. However, for sub-Riemannian metrics we still have balls that are Lipschitz domains, as the following Proposition 5.7 shows. But let us first recall a simple fact:

**Lemma 5.6.** *Let $\mathbb{A}$ and $\mathbb{B}$ be two stratified groups with stratifications $\bigoplus V_i$ and $\bigoplus W_i$, respectively. Endow $V_1$ and $W_1$ with a scalar product each and let $d_A, d_B$ be the corresponding homogeneous sub-Riemannian distances.*

*Then $\mathbb{A} \times \mathbb{B}$ is a Carnot group with stratification $\bigoplus_i V_i \times W_i$ and metric*

$$d((a,b),(a',b')) := \sqrt{d_A(a,a')^2 + d_B(b,b')^2},$$

*which is the sub-Riemannian metric generated by the scalar product on $V_1 \times W_1$ induced by the scalar products on $V_1$ and $W_1$.*

One proves this lemma by using the fact that the energy of curves on $\mathbb{A} \times \mathbb{B}$ (i.e., the integral of the squared norm of the derivative) is the sum of the energies of the two components of the curve.

**Proposition 5.7.** *Any homogeneous sub-Riemannian metric on $\mathbb{H} \times \mathbb{R}$ is locally Lipschitz away from the diagonal.*

*Proof.* First of all, we show that, up to isometry, there is only one homogeneous sub-Riemannian distance on $\mathbb{H} \times \mathbb{R}$. Let $(X_1, Y_1, T_1)$ and $(X_2, Y_2, T_2)$ be two bases of $V_1 \times \mathbb{R}$ that are orthonormal for two sub-Riemannian structures, respectively. We may assume $T_1, T_2 \in \{0\} \times \mathbb{R}$. Notice that $[X_i, Y_i] \notin V_1 \times \mathbb{R}$. The linear map such that $X_1 \mapsto X_2$, $Y_1 \mapsto Y_2$, $T_1 \mapsto T_2$, $[X_1, Y_1] \mapsto [X_2, Y_2]$ is an automorphism of Lie algebras and induces an isometry between the two sub-Riemannian structures.

Denoting by $d_{\mathbb{H}}$ and $d_{\mathbb{R}}$ the standard metrics on $\mathbb{H}$ and $\mathbb{R}$, respectively, we prove the proposition for the product metric as in Lemma 5.6. Namely, we need to check that the function

$$(p,t) \mapsto d((0,0),(p,t)) = \sqrt{d_{\mathbb{H}}(0,p)^2 + t^2} \tag{5.7}$$

is locally Lipschitz at all $(\hat{p}, \hat{t}) \neq (0,0)$. This follows directly from Proposition 3.3. $\square$

### 5.4. **A sub-Riemannian sphere with a cusp.**

**Proposition 5.8.** *Let $G$ be a Carnot group of step 3 endowed with a sub-Riemannian distance $d_G$. Then the sub-Riemannian distance $d$ on $G \times \mathbb{R}$ given by*

$$d((p,s),(q,t)) = \sqrt{d_E(p,q)^2 + |t-s|^2}$$

*has a unit sphere with a cusp at $(0_G, 1)$.*

*Proof.* Let $\mathfrak{g} = \bigoplus_{i=1}^{3} V_i$ be the Lie algebra of $G$ and fix $Z \in V_3 \setminus \{0\}$. We identify $\mathfrak{g}$ with $G$ through the exponential map.

The intersection of the unit sphere in $(G\times\mathbb{R}, d)$ with the plane $\operatorname{span}\{Z\}\times\mathbb{R}$ is given by all points $(zZ, t)$ such that

$$d_G(0, zZ)^2 + t^2 = 1. \tag{5.8}$$

Since $d_G$ is homogeneous on $G$, there exists $C > 0$ such that for all $z \in \mathbb{R}$

$$d_G(0, zZ) = C|z|^{\frac{1}{3}}. \tag{5.9}$$

Putting together (5.8) and (5.9) we obtain that this intersection consists of all the points $(zZ, t)$ such that

$$|z| = \left(\frac{1+t}{C^2}\right)^{\frac{3}{2}} \cdot (1-t)^{\frac{3}{2}}.$$

One then easily sees that this set in $\mathbb{R}^2$ has a cusp at $(0, 1)$. $\square$

## 6. A closer look at the Heisenberg group

The Heisenberg group $\mathbb{H}$ is the easiest example of a stratified group that is not Abelian and for this reason it has been studied in large extend. The most common homogeneous metrics on $\mathbb{H}$ are the Korányi metric and the sub-Riemannian metric. Sub-Finsler metrics on $\mathbb{H}$ arise in the study of finitely-generated groups, see [7] and references therein. The geometry of sub-Finsler spheres has been studied in [21] and [14].

The Lie algebra $\mathfrak{h}$ of the Heisenberg group is a three dimensional vector space $\operatorname{span}\{X, Y, Z\}$ with a Lie bracket operation defined by the only nontrivial relation $[X, Y] = Z$.

We identify the Heisenberg group $\mathbb{H}$ again with $\operatorname{span}\{X, Y, Z\}$, where we define the group operation

$$p \cdot q := p + q + \frac{1}{2}[p, q] \qquad \forall p, q \in \mathbb{H}.$$

Hence $\mathfrak{h}$ is the Lie algebra of $\mathbb{H}$ and the exponential map $\mathfrak{h} \to \mathbb{H}$ is the identity map. Notice that the inverse of an element $p$ is $p^{-1} = -p$.

The Heisenberg Lie algebra admits the stratification $\mathfrak{h} = V_1 \oplus V_2$ with $V_1 = \operatorname{span}\{X, Y\}$ and $V_2 = \operatorname{span}\{Z\}$. Denote by $\pi$ the linear projection $\mathfrak{h} \to V_1$ along $V_2$. Notice that this map, regarded as $\pi : (\mathbb{H}, \cdot) \to (V_1, +)$, is a group morphism.

The dilations $\delta_\lambda : \mathbb{H} \to \mathbb{H}$ are explicitly expressed by

$$\delta_\lambda(xX + yY + zZ) = x\lambda X + y\lambda Y + z\lambda^2 Z, \qquad \forall \lambda > 0.$$

These are both Lie algebra automorphisms $\delta_\lambda : \mathfrak{h} \to \mathfrak{h}$ and Lie group automorphisms $\mathbb{H} \to \mathbb{H}$.

Three are the main results of this section.

**Proposition 6.1.** *Let $N : \mathbb{H} \to [0, +\infty)$ be a homogeneous norm. Then the unit ball*
$$B := \{p \in \mathbb{H} : N(p) \le 1\}$$
*is a star-like Lipschitz domain.*

*Proof.* One easily shows that condition (1.1) holds for all $p \in \mathbb{H} \setminus \{0\}$. In order to prove that $B$ is star-like, one first notice that if $p \in B$, then $-p \in B$, hence $\delta_t(p)\delta_{1-t}(-p) = (2t-1)p \in B$ for all $t \in [0, 1]$, and this is a straight line passing through zero. □

**Proposition 6.2.** *Let $N$ and $B$ as in Proposition 6.1. Set $K := \pi(B) \subset V_1$. Then $K$ is a compact, convex set with $K = -K$ and $K = \mathrm{cl}(\mathrm{int}(K))$, and there exists a function $f : K \to [0, +\infty)$, locally Lipschitz on $\mathrm{int}(K)$, such that*

$$B = \{v + zZ : \ v \in K, \ -f(-v) \le z \le f(v)\}. \tag{6.1}$$

The proof is postponed to Section 6.1.

We remark that homogeneous distances and sub-Finsler homogeneous distances on $\mathbb{H}$ have a precise relation. Indeed, if $d$ is a homogeneous distance on $\mathbb{H}$, then it is easy to show that the length distance generated by $d$ is exactly the sub-Finsler distance that has the norm on $V_1$ generated by the set $K$ defined in Proposition 6.2.

**Proposition 6.3.** *Let $K \subset V_1$ be a compact, convex set with $-K = K$ and $0 \in \mathrm{int}(K)$. Let $g : K \to \mathbb{R}$ be Lipschitz. Then there exists $b \in \mathbb{R}$ such that for $f := g + b$ the set $B$ as in (6.1) is the unit ball of a homogeneous norm.*

The proof will appear in Section 6.2.

As a consequence of Proposition 6.3, we get the existence of homogeneous distances on $\mathbb{H}$ that are not almost convex in the sense of [13]. Indeed, one can take the distance associated to $g(xX + yY) = |x|$ from Proposition 6.3.

## 6.1. **Proof of Proposition 6.2.**

**Lemma 6.4.** *Let $B \subset \mathbb{H}$ be an arbitrary closed set satisfying (2.5). If $p = v + zZ \in B$ with $v = \pi(p) \in V_1$, then $v + szZ \in B$, for all $s \in [0, 1]$. In particular,*

*(1) $\pi(B) = B \cap V_1$;*
*(2) $\pi(B) \subset V_1$ is convex.*

*Proof.* We have that for all $t \in [0, 1]$
$$B \ni \delta_t p \cdot \delta_{1-t} p = v + (t^2 + (1-t)^2) zZ.$$
Since the image of $[0, 1]$ through the map $t \mapsto (t^2 + (1-t)^2)$ is $[\frac{1}{2}, 1]$, then it follows $v + szZ \in B$ for all $s \in [\frac{1}{2}, 1]$. Iterating this process and using the closeness of $B$, we get $v + szZ \in B$ for all $s \in [0, 1]$. For the last statement, take $v, w \in \pi(B) \subset B$ and notice that $tv + (1-t)w = \pi(\delta_t v \cdot \delta_{1-t} w) \in \pi(B)$. □

Let $B = \{N \le 1\}$ be the unit ball of a homogeneous norm and set $K := \pi(B) \subset V_1$ and $\Omega := \mathrm{int}(K)$. First, we check that $\bar{\Omega} = K$. On the one hand, clearly we have $\bar{\Omega} \subset K$. On the other hand, if $v \in K$, then for any $t \in [0,1)$ we have $N(\delta_t v) = tN(v) < 1$, i.e., $\delta_t v = tv \in \mathrm{int}B \cap V \subset \Omega$. Hence $v \in \bar{\Omega}$.

If we define $f : K \to [0, +\infty)$ as $f(v) := \max\{z : v + zZ \in Q\}$, then we have (6.1). In order to prove that $f$ is locally Lipschitz on $\Omega$, we need to prove

$$\begin{gathered} \forall p \in \partial B \cap \{z \ge 0\} \cap \pi^{-1}(\Omega), \\ \exists U \ni p \text{ open }, \exists C \text{ vertical cone, s.t.} \\ \forall q \in U \cap \partial B \text{ it holds } q + C \subset B. \end{gathered} \tag{6.2}$$

Here a *vertical cone* is a Euclidean cone with axis $-Z$ and non-empty interior.

So, fix $p \in \partial Q \cap \{z \ge 0\}$ such that $\pi(p) \in \Omega$. Define for $\theta \in \mathbb{R}$ and $\epsilon > 0$

$$v_\theta := x_\theta X + y_\theta Y := \epsilon\cos(\theta)X + \epsilon\sin(\theta)Y.$$

For $\epsilon > 0$ small enough, $\pi(p) + v_\theta \in \Omega$ for all $\theta$. Define

$$\phi(t, \theta) := \delta_{(1-t)}p \cdot \delta_t(\pi(p) + v_\theta).$$

Clearly $\phi(t, \theta) \in B$ for $t \in [0,1]$ and $\theta \in \mathbb{R}$, and $\phi(0, \theta) = p$ for all $\theta$. Geometrically, $\phi([0,1] \times \mathbb{R})$ is a "tent" inside $B$ standing above the whole vertical segment from $\pi(p)$ to $p$. Notice that $p \ne \pi(p)$, indeed $N(p) = 1$ while $N(\pi(p)) < 1$, because $\pi(p) \in \Omega$.

We only need to prove that the curves $t \mapsto \phi(t, \theta)$ meet this vertical segment by an angle bounded away from 0. Some computations are needed: set $p = p_1X + p_2Y + p_3Z$, then

$$\phi(t, \theta) = \pi(p) + tv_\theta + \left(\frac{1}{2}t(1-t)(p_1y_\theta - p_2x_\theta) + (1-t)^2p_3\right)Z.$$

We take care only of the third coordinate. Set

$$\begin{aligned} g(t) :&= \frac{1}{2}t(1-t)(p_1y_\theta - p_2x_\theta) + (1-t)^2p_3 \\ &= t^2\ (-\frac{1}{2}(p_1y_\theta - p_2x_\theta) + p_3)\ +\ t\ (\frac{1}{2}(p_1y_\theta - p_2x_\theta) - 2p_3)\ +\ p_3. \end{aligned}$$

Saying that the angle between the curve $t \mapsto \phi(t, \theta)$ and the vertical segment at $p$ is uniformly grater than zero, is equivalent to give an upper bound to the derivative of $g$ at 0 for all $\theta$. Since

$$g'(0) = \frac{1}{2}(p_1y_\theta - p_2x_\theta) - 2p_3,$$

we are done.

Finally, since both $\epsilon$ and $g'(0)$ depend continuously on $p$, then (6.2) is satisfied. $\square$

6.2. **Proof of Proposition 6.3.** We consider the bilinear map $\omega : V_1 \times V_1 \to \mathbb{R}$ given by

$$\omega(v_1X + v_2Y, w_1X + w_2Y) := v_1w_2 - v_2w_1.$$

**Lemma 6.5.** *For any continuous function $f : K \to [0, +\infty)$, the set $B$ as in* (6.1) *is the unit ball of a homogeneous norm on $\mathbb{H}$ if and only if*

$$\begin{gathered} \forall v, w \in K \quad \forall t \in [0,1] \\ f(tv + (1-t)w) - t^2 f(v) - (1-t)^2 f(w) - \tfrac{t(1-t)}{2}\omega(v,w) \geq 0. \end{gathered} \tag{6.3}$$

*Proof.* One easily sees that $B = B^{-1}$. Notice that $B$ is the unit ball of a homogeneous norm if and only if it satisfies (2.5).

$\boxed{\Rightarrow}$ Assume that $B$ satisfies (2.5). Then for any $v, w \in K$ we have

$$\begin{aligned} B \ni\ & \delta_t(v + f(v)Z) \cdot \delta_{(1-t)}(w + f(w)Z) = \\ & = tv + (1-t)w + \left(t^2 f(v) + (1-t)^2 f(w) + \frac{1}{2}t(1-t)\omega(v,w)\right) Z, \end{aligned}$$

hence

$$t^2 f(v) + (1-t)^2 f(w) + \frac{1}{2}t(1-t)\omega(v,w) \leq f(tv + (1-t)w).$$

$\boxed{\Leftarrow}$ Suppose $f$ satisfies (6.3). Define

$$\begin{aligned} B^+ &:= \{v + zZ : v \in K \text{ and } z \leq f(v)\}, \\ B^- &:= \{v + zZ : v \in K \text{ and } -f(-v) \leq z\}. \end{aligned}$$

We will show that both $B^+$ and $B^-$ satisfy (2.5), from which it follows that $B = B^+ \cap B^-$ satisfies (2.5) as well.

So, let $v, w \in K$ and $z_1, z_2 \in \mathbb{R}$ such that $v + z_1 Z, w + z_2 Z \in B^+$. Then the third coordinate of $\delta_t(v + z_1 Z) \cdot \delta_{(1-t)}(w + z_2 Z)$ satisfies

$$\begin{aligned} & t^2 z_1 + (1-t)^2 z_2 + \frac{1}{2}t(1-t)\omega(v,w) \leq \\ & \qquad \leq t^2 f(v) + (1-t)^2 f(w) + \frac{1}{2}t(1-t)\omega(v,w) \leq f(tw + (1-t)v), \end{aligned}$$

therefore we have $\delta_t(v + z_1 Z) \cdot \delta_{(1-t)}(w + z_2 Z) \in B^+$ for all $t \in [0,1]$.

The calculation for $B^-$ is similar. □

The verification of the next lemma is simple and therefore it is omitted.

**Lemma 6.6.** *Suppose that $g : K \to \mathbb{R}$ is a continuous function such that there is a constant $A \in \mathbb{R}$ with*

$$\begin{gathered} \forall v, w \in K, \ \forall t \in [0,1] \\ g(tv + (1-t)w) - t^2 g(v) - (1-t)^2 g(w) \geq At(1-t). \end{gathered} \tag{6.4}$$

*Then $f := g + B$ satisfies* (6.3) *with*

$$B := \sup_{v,w \in K} \frac{1}{2}\left(\frac{1}{2}\omega(v,w) - A\right) = \frac{1}{4}\left(\sup_{v,w \in K} \omega(v,w)\right) - \frac{1}{2}A.$$

**Lemma 6.7.** *Let $g : K \to \mathbb{R}$ be $L$-Lipschitz. Then $g$ satisfies* (6.4) *for*

$$A := -2L\operatorname{diam}(K) - 4\sup_{p \in K} |g(p)|.$$

*Proof.* Notice that we need to show that (6.4) holds only for $t \in (0,1)$ and that (6.4) is symmetric in $t$ and $(1-t)$. So, it is enough to consider only the case $t \in (0, \frac{1}{2}]$:

$$\begin{aligned}
&\frac{g(tv+(1-t)w) - t^2 g(v) - (1-t)^2 g(w)}{t(1-t)} \\
&\qquad = \frac{g(w+t(v-w)) - g(w)}{t(1-t)} + \frac{(1-(1-t)^2)g(w)}{t(1-t)} - \frac{t}{1-t} g(v) \\
&\qquad \geq -\frac{L\|v-w\|}{1-t} + \frac{2-t}{1-t} g(w) - \frac{t}{1-t} g(v) \\
&\qquad\qquad \geq -2L\mathrm{diam}(K) - 4 \sup_{p \in K} |g(p)|.
\end{aligned}$$

□

Putting together Lemmas 6.7, 6.6, and 6.5, we get Proposition 6.3. □

## Appendix A. Equivalence of some definitions and existence of singular minimizers

We shall prove that on Carnot groups the absence of singular geodesics is equivalent to other three well-known properties. Consequently, we will prove that Carnot groups of step larger than 2 always have singular length minimizers, as we stated in Remark 2.17. Corollary A.2, and hence Theorem A.1, cannot be extended to the more general setting of sub-Finsler manifolds. Namely, it has been proven in [11] that a generic distribution of rank $m \geq 3$ on a manifold $M$ does not have singular curves. Note that if $\dim(M) \geq 2m$, then the step of all distributions of rank $m$ is larger than 2.

Before stating the theorem, we briefly introduce four classical properties present in literature. Let $G$ be a stratified group with Lie algebra $\mathfrak{g}$ and first layer $V_1$ of the stratification. The stratified Lie algebra $\mathfrak{g}$ is said to be *strongly bracket generating* if for all $X \in V_1 \setminus \{0\}$ it holds

$$\mathfrak{g} = V_1 + [X, V_1].$$

A stratified step-two Lie algebra $\mathfrak{g} = V_1 \oplus V_2$ is *of Métivier type* if there is a scalar product $\langle \cdot, \cdot \rangle$ on $\mathfrak{g}$ such that for all $Z \in V_2 \setminus \{0\}$ the map $J_Z : V_1 \to V_1$ defined by

$$\forall X, Y \in V_1 \qquad \langle J_Z X, Y \rangle = \langle Z, [X,Y] \rangle$$

is injective. The main examples of groups of Métivier type are those of $H$-type. See [6] for further reference.

We write $\Gamma(V_1)$ for the space of all vector fields of $G$ with values in the left-invariant tangent subbundle of $G$ generated by $V_1$. A stratified group $G$ is *fat* if for every vector field $X \in \Gamma(V_1)$ with $X(0) \neq 0$ it holds

$$\mathfrak{g} = V_1 + [X, \Gamma(V_1)]_0,$$

where $[X, \Gamma(V_1)]_0 = \mathrm{span}\{[X,Y](0) : Y \in \Gamma(V_1)\}$. See [27, 3] for reference.

A sub-Finsler manifold is said to be *ideal* if, except the constant curve, there are no singular length minimizers. The terminology is taken from [27].

**Theorem A.1.** *If $G$ is a Carnot group with stratified Lie algebra $\mathfrak{g}$, then the following properties are equivalent:*

*(i)* $\mathfrak{g}$ *is strongly bracket generating;*
*(ii)* $\mathfrak{g}$ *is of Métivier type;*
*(iii)* $G$ *is fat;*
*(iv)* $G$ *is an ideal sub-Finsler manifold.*

*Moreover, these properties imply that* $\mathfrak{g}$ *has step one or two.*

A direct consequence of the proof of the latter theorem is the following Corollary.

**Corollary A.2.** *In all sub-Finsler Carnot groups of step at least 3, there exists a one-parameter subgroup that is a singular non-constant length minimizer.*

*Proof of Theorem A.1.* $(i) \Rightarrow (ii)$. We give a proof by contraposition. If $\mathfrak{g}$ is not a of Métivier type, then there are a scalar product $\langle\cdot,\cdot\rangle$ on $\mathfrak{g}$ and $Z \in V_2\setminus\{0\}$ such that $J_Z : V_1 \to V_1$ is not injective. So, there is $X \in V_1\setminus\{0\}$ with $J_Z X = 0$. Therefore, for all $Y \in V_1$ we have $\langle Z, [X, Y]\rangle = \langle J_Z, Y\rangle = 0$, i.e., $Z \notin [X, V_1]$, so $\mathfrak{g}$ is not strongly bracket generating.

$(ii) \Rightarrow (i)$. We give a proof by contraposition. Suppose $\mathfrak{g}$ is not strongly bracket generating and let $\langle\cdot,\cdot\rangle$ be a scalar product on $\mathfrak{g}$. Then there are $X \in V_1 \setminus \{0\}$ and $Z \in V_2 \setminus \{0\}$ such that $Z$ is orthogonal to $[X, V_1]$. Hence, for all $Y \in V_1$ we have $\langle J_Z X, Y\rangle = \langle Z, [X, Y]\rangle = 0$, i.e., $\mathfrak{g}$ is not of Métivier type.

$(iii) \Rightarrow (i)$. The implication is trivial.

$(i) \Rightarrow (iii)$. Let $X_1, \dots, X_r$ be a basis for $V_1$ and $X = \sum_{i=1}^r a_i X_i \in \Gamma(V_1)$ with $a_i \in \mathscr{C}^\infty(G)$ with $X(0) \neq 0$, where $X_i$ are considered as left-invariant vector fields. Set $\tilde{X} := \sum_{i=1}^r a_i(0) X_i \in V_1 \setminus \{0\}$. Since $\mathfrak{g}$ is strongly bracket generating, $[\tilde{X}, V_1] = V_2$. Since $[X, X_j] = [\tilde{X}, X_j] + \sum_{i=i}^n (X_j a) X_i$, for $j \in \{1, \dots, r\}$, one easily sees that

$$V_1 + [X, \Gamma(V_1)]_0 = V_1 + \mathrm{span}\{[X, X_j]_0 : j = 1, \dots, r\} = \mathfrak{g}.$$

$(i) \Rightarrow (iv)$. This implication is well known. See for example [20, Remark 2.7].

$(iv) \Rightarrow (i)$. Before starting, recall that any horizontal one-parameter subgroup in a sub-Finsler Carnot group is length minimizer.

We begin by claiming that if $G$ is a Carnot group, $X \in V_1 \setminus \{0\}$, and $\gamma : [0, 1] \to G$, $\gamma(t) := \exp(tX)$, is a regular curve then

$$\mathrm{ad}_X : V_k \to V_{k+1} \text{ is surjective, for all } k \in \{1, \dots, s-1\}. \tag{A.1}$$

Indeed, for all $v \in V_1$ we have

$$\begin{aligned}\mathrm{Ad}_{\exp(tX)} v = e^{\mathrm{ad}_{tX}} v &= \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathrm{ad}_X^k v \\ &= v + t[X, v] + \frac{t^2}{2}[X, [X, v]] + \frac{t^3}{6}[X, [X, [X, v]]] + \dots\end{aligned}$$

Therefore,

$$\mathrm{span}\left\{\mathrm{Ad}_{\gamma(t)}[V_1] : t \in [0, 1]\right\} \subset \mathrm{span}\{\mathrm{ad}_X^k[V_1] : k \in \{0, \dots, s\}\}.$$

Thanks to [20, Proposition 2.3] the left hand side is $\mathrm{Lie}(G)$, since $\gamma$ is regular. Hence, since $\mathrm{ad}_X^k[V_1] \subset V_{k+1}$, then $\mathrm{ad}_X^k[V_1] = V_{k+1}$ and we get (A.1). To conclude the proof of the theorem, it is enough to show that if $G$ is a stratified group of step $s$ such that (A.1) holds for all $X \in V_1 \setminus \{0\}$, then $s \leq 2$. If $s > 3$, we can take the normal subgroup $H = \exp\left(\bigoplus_{i=4}^{s} V_i\right)$, so that the quotient $G/H$ is a stratified group of step 3. By taking a further quotient we may assume that the third layer $V_3$ has dimension 1. The quotient still satisfies (A.1) for all $X \in V_1 \setminus \{0\}$.

Therefore, we just need to show that there are no stratified groups of step 3 with $\dim(V_3) = 1$ that satisfy (A.1) for all $X \in V_1 \setminus \{0\}$. Let $r := \dim V_1$. Since for any $X \in V_1 \setminus \{0\}$ the map $\mathrm{ad}_X : V_1 \to V_2$ is surjective and has non-trivial kernel, then $m := \dim V_2 < r$. Let $Y_1, \ldots, Y_m$ be a basis of $V_2$. Since $V_3 \simeq \mathbb{R}$, we can interpret each $\mathrm{ad}_{Y_i}$ as an element of $(V_1)^*$. Since $m < r$, then $\mathrm{span}\{\mathrm{ad}_{Y_1}, \ldots, \mathrm{ad}_{Y_m}\}^\perp \neq \{0\}$, i.e., there exists $X \in V_1 \setminus \{0\}$ such that $\mathrm{ad}_{Y_i}(X) = 0$ for all $i \in \{1, \ldots, m\}$. We get now a contradiction with (A.1), because

$$\{0\} \neq V_3 = \mathrm{ad}_X(V_2) = \mathrm{span}\{[X, Y_i] : i \in \{1, \ldots, m\}\} = \{0\}.$$

□

**Acknowledgment.** This paper benefited from the help of A. Martini with discussion on groups of Métivier type. The anonymous referee and L. Rizzi suggested few improvements and in particular indicated that recently a similar proof of the implication $(iv) \Rightarrow (i)$ in Theorem A.1 appeared in [28]. The authors wish to thank all these people.

## References

[1] A. A. Agrachëv, *Any sub-Riemannian metric has points of smoothness*, Dokl. Akad. Nauk **424** (2009), no. 3, 295–298.

[2] A. A. Agrachëv, Davide Barilari, and Ugo Boscain, *Introduction to Riemannian and sub-Riemannian geometry*, July 2015.

[3] A. A. Agrachëv and Andrei V. Sarychev, *Sub-Riemannian metrics: minimality of abnormal geodesics versus subanalyticity*, ESAIM Control Optim. Calc. Var. **4** (1999), 377–403 (electronic).

[4] Luigi Ambrosio and Paolo Tilli, *Topics on analysis in metric spaces*, Oxford Lecture Series in Mathematics and its Applications, vol. 25, Oxford University Press, Oxford, 2004.

[5] A. S. Besicovitch and H. D. Ursell, *Sets of fractional dimensions (v): on dimensional numbers of some continuous curves*, J. London Math. Soc. **01** (1937), 18–25.

[6] A. Bonfiglioli, E. Lanconelli, and F. Uguzzoni, *Stratified Lie groups and potential theory for their sub-Laplacians*, Springer Monographs in Mathematics, Springer, Berlin, 2007.

[7] Emmanuel Breuillard, *Geometry of locally compact groups of polynomial growth and shape of large balls*, Groups Geom. Dyn. **8** (2014), no. 3, 669–732.

[8] Emmanuel Breuillard and Enrico Le Donne, *On the rate of convergence to the asymptotic cone for nilpotent groups and subFinsler geometry*, Proc. Natl. Acad. Sci. USA **110** (2013), no. 48, 19220–19226.

[9] Dmitri Burago, Yuri Burago, and Sergei Ivanov, *A course in metric geometry*, Graduate Studies in Mathematics, vol. 33, American Mathematical Society, Providence, RI, 2001.

[10] Luca Capogna, Donatella Danielli, Scott D. Pauls, and Jeremy T. Tyson, *An introduction to the Heisenberg group and the sub-Riemannian isoperimetric problem*, Progress in Mathematics, vol. 259, Birkhäuser Verlag, Basel, 2007.

[11] Y. Chitour, F. Jean, and E. Trélat, *Genericity results for singular curves*, J. Differential Geom. **73** (2006), no. 1, 45–73.

[12] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern, and P. R. Wolenski, *Nonsmooth analysis and control theory*, Graduate Texts in Mathematics, vol. 178, Springer-Verlag, New York, 1998.

[13] M. Duchin and M. Shapiro, *Rational growth in the Heisenberg group*, ArXiv e-prints (2014).

[14] Moon Duchin and Christopher Mooney, *Fine asymptotic geometry in the Heisenberg group*, Indiana Univ. Math. J. **63** (2014), no. 3, 885–916.

[15] Lawrence M. Graves, *Some mapping theorems*, Duke Math. J. **17** (1950), 111–114.

[16] Waldemar Hebisch and Adam Sikora, *A smooth subadditive homogeneous norm on a homogeneous group*, Studia Math. **96** (1990), no. 3, 231–236.

[17] Velimir Jurdjevic, *Geometric control theory*, Cambridge Studies in Advanced Mathematics, vol. 52, Cambridge University Press, Cambridge, 1997.

[18] Enrico Le Donne, *A metric characterization of Carnot groups*, Proc. Amer. Math. Soc. **143** (2015), no. 2, 845–849.

[19] ______, *A primer on Carnot groups: homogenous groups, CC spaces, and regularity of their isometries*, Manuscript, available on arxiv.org, 2016.

[20] Enrico Le Donne, Richard Montgomery, Alessandro Ottazzi, Pierre Pansu, and Davide Vittone, *Sard property for the endpoint map on some Carnot groups*, Annales de l'Institut Henri Poincare (C) Non Linear Analysis (2015).

[21] Enrico Le Donne and Séverine Rigot, *Besicovitch covering property for homogeneous distances on the Heisenberg groups*, Accepted in Journal of the European Mathematical Society.

[22] ______, *Remarks about the Besicovitch covering property in Carnot groups of step 3 and higher*, Proc. Amer. Math. Soc. **144** (2016), no. 5, 2003–2013.

[23] Richard Montgomery, *A tour of subriemannian geometries, their geodesics and applications*, Mathematical Surveys and Monographs, vol. 91, American Mathematical Society, Providence, RI, 2002.

[24] Barrett O'Neill, *Semi-Riemannian geometry: with applications to relativity*, Pure and applied mathematics, Academic Press, New York, 1983.

[25] Ludovic Rifford, *A Morse-Sard theorem for the distance function on Riemannian manifolds*, Manuscripta Math. **113** (2004), no. 2, 251–265.

[26] ______, *À propos des sphères sous-riemanniennes*, Bull. Belg. Math. Soc. Simon Stevin **13** (2006), no. 3, 521–526.

[27] ______, *Sub-Riemannian geometry and optimal transport*, Springer Briefs in Mathematics, Springer, Cham, 2014.

[28] Luca Rizzi, *Measure contraction properties of Carnot groups*, Calc. Var. Partial Differential Equations **55** (2016), no. 3, 55:60.

[29] Eberhard Siebert, *Contractive automorphisms on locally compact groups*, Math. Z. **191** (1986), no. 1, 73–90.

*E-mail address*: `enrico.ledonne@gmail.com`

*E-mail address*: `sebastiano2.72@gmail.com`

(Le Donne, Nicolussi Golo) Department of Mathematics and Statistics, University of Jyväskylä, 40014 Jyväskylä, Finland

[B]

**Asymptotic behavior of the Riemannian Heisenberg group and its horoboundary**

E. Le Donne, S. Nicolussi Golo, and A. Sambusetti

First published in *Annali di Matematica (2017) 196: 1251*
by Springer Berlin Heidelberg

# ASYMPTOTIC BEHAVIOR OF THE RIEMANNIAN HEISENBERG GROUP AND ITS HOROBOUNDARY

ENRICO LE DONNE, SEBASTIANO NICOLUSSI GOLO,
AND ANDREA SAMBUSETTI

Abstract. The paper is devoted to the large scale geometry of the Heisenberg group $\mathbb{H}$ equipped with left-invariant Riemannian metrics. We prove that two such metrics have bounded difference if and only if they are asymptotic, i.e., their ratio goes to one, at infinity. Moreover, we show that for every left-invariant Riemannian metric $d$ on $\mathbb{H}$ there is a unique subRiemanniann metric $d'$ for which $d - d'$ goes to zero at infinity, and we estimate the rate of convergence. As a first immediate consequence we get that the Riemannian Heisenberg group is at bounded distance from its asymptotic cone. The second consequence, which was our aim, is the explicit description of the horoboundary of the Riemannian Heisenberg group.

## Contents



## 1. Introduction

In large-scale geometry, various notions of space at infinity have received special interest for differently capturing the asymptotic geometric behavior. Two main examples of spaces at infinity are the asymptotic cone and the horoboundary. The description of the *asymptotic cone* for finitely generated groups is a crucial step in the algebraic characterization of groups of polynomial growth, [21, 30, 13, 34, 23, 3]. The notion of *horoboundary* has been formulated by Gromov [12], inspired by the seminal work of Busemann on the theory of parallels on geodesic spaces [6]. The horoboundary has a

2010 *Mathematics Subject Classification.* 20F69, 53C23, 53C17.

*Key words and phrases.* Heisenberg group, horoboundary, asymptotic cone, Riemannian geometry, sub-Riemannian geometry.

E.L.D. has been supported by the Academy of Finland project no. 288501. S.N.G. has been supported by the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n. 607643.

fully satisfying visual description in the framework of $CAT(0)$-spaces and of Gromov-hyperbolic spaces, [11, 2, 4]. It plays a major role in the study of dynamics and rigidity of negatively curved spaces, [15, 22, 11, 29, 24, 27]. The visual-boundary description breaks down for non-simply connected manifolds [9] and when the curvature has variable sign, as we will make evident for the Riemannian Heisenberg group.

This paper contributes to the study of the asymptotic geometry of the simplest non-Abelian nilpotent group: the Heisenberg group. The asymptotic cone of the Heisenberg group equipped with a left-invariant Riemannian metric $d_R$ is the Heisenberg group equipped with a Carnot-Carathéodory metric $d_{CC}$, see [23] and also [3]. Our contribution is a finer analysis of the asymptotic comparison of $d_R$ and $d_{CC}$. This leads to the explicit knowledge of the (Riemannian) horoboundary. We remark that the Heisenberg group is not hyperbolic, hence one does not consider its visual boundary.

We recall the definition of horoboundary. Let $(X, d)$ be a metric space. We consider the space of continuous real functions $\mathscr{C}(X)$ endowed with the topology of uniform convergence on compact sets. We denote by $\mathscr{C}(X)/\mathbb{R}$ the quotient with respect to the subspace of constant functions. The map $x \mapsto d(x, \cdot)$ induces an embedding $X \hookrightarrow \mathscr{C}(X)/\mathbb{R}$. The horoboundary of $X$ is defined as $\partial_h X := \bar{X} \setminus X \subset \mathscr{C}(X)/\mathbb{R}$. See Section 5 for a detailed exposition.

The horoboundary has been investigated for finite-dimensional normed vector spaces, [31], for Hilbert geometries, [32], and for infinite graphs [33]. For non-simply connected, negatively curved manifolds it has been studied in [9]. Nicas and Klein computed the horoboundary of the Heisenberg group when endowed with the Korany metric in [17], and with the metric $d_{CC}$ in [18].

We will show that the horoboundary of the Heisenberg group endowed with a left-invariant Riemannian metric $d_R$ coincides with the second case studied by Nicas and Klein, see Corollary 1.4. This will be an immediate consequence of our main result Theorem 1.3, which implies that the difference $d_R - d_{CC}$ converges to zero when evaluated on points $(p, q_n)$ with $q_n$ being a sequence that leaves every compact set.

**Remark 1.1.** Another term for Carnot-Carathéodory metric is subRiemannian metric. In this paper we should discuss subRiemannian metrics that may actually be Riemannian. Therefore we follow the convention that subRiemannian geometry includes as a particular case Riemannian geometry. This is in agreement with several established references in the field, see [1, 26, 16]. In the presence of a subRiemannian metric that is not Riemannian we shall use the term *strict subRiemannian.*

1.1. **Detailed results.** The Heisenberg group $\mathbb{H}$ is the simply connected Lie group whose Lie algebra $\mathfrak{h}$ is generated by three vectors $X, Y, Z$ with only non-zero relation $[X, Y] = Z$. A left-invariant Riemannian metric $d$ on $\mathbb{H}$ is determined by a scalar product $g$ on $\mathfrak{h}$; a left-invariant *strictly* subRiemannian metric $d$ is induced by a bracket generating plane $V \subset \mathfrak{h}$ and a scalar product $g$ on $V$ (see Section 2 for detailed exposition). In both cases we say that $d$ is subRiemannian with *horizontal space* $(V, g)$, where $\dim V$ is either 2 or 3.

We are interested in the asymptotic comparison between these metrics. Given two left-invariant subRiemannian metrics $d$ and $d'$ on $\mathbb{H}$, we deal with three asymptotic behaviors, in ascending order of strength, each of which defines an equivalence relation among subRiemannian metrics:

(i) $\lim_{d(p,q)\to\infty} \frac{d(p,q)}{d'(p,q)} = 1$;
(ii) There exists $c > 0$ such that $|d(p,q) - d'(p,q)| < c$, for all $p, q$;
(iii) $\lim_{d(p,q)\to\infty} |d(p,q) - d'(p,q)| = 0$.

A first example of the implication $(i) \Rightarrow (ii)$ was proved by Burago in [5] for $\mathbb{Z}^n$-invariant metrics $d$ on $\mathbb{R}^n$, by showing that $d$ and the associated stable norm stay at bounded distance from each other. This result has been extended quantitatively for $\mathbb{Z}^n$-invariant metrics on geodesic metric spaces in [8]. Gromov and Burago asked for other interesting cases where the same implication holds. Another well-known case where (i) is equivalent to (ii) is that of hyperbolic groups. Beyond Abelian and hyperbolic groups, Krat proved the equivalence for word metrics on the discrete Heisenberg group $\mathbb{H}(\mathbb{Z})$, [19]. For general subFinsler metrics on Carnot groups it has been proven in [3], following [28], that (i) is equivalent to the fact that the projections onto $\mathbb{H}/[\mathbb{H},\mathbb{H}]$ of the corresponding unit balls coincide, see (c) below. Our first result shows that this last condition is equivalent to each one of (i) and (ii) in the case of the Heisenberg group endowed with subRiemannian metrics.

**Theorem 1.2.** *Let $d$ and $d'$ be two left-invariant subRiemannian metrics on $\mathbb{H}$ whose horizontal spaces are $(V, g)$ and $(V', g')$ respectively. Let $\pi : \mathfrak{h} \to \mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ be the quotient projection and $\hat{\pi} : \mathbb{H} \to \mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ the corresponding group morphism.*

*Then the following assertions are equivalent:*

*(a) there exists $c > 0$ such that $|d(p,q) - d'(p,q)| < c$, for all $p, q$;*
*(b) $\frac{d(p,q)}{d'(p,q)} \to 1$ when $d(p,q) \to \infty$;*
*(c) $\hat{\pi}(\{p \in \mathbb{H} : \ d(0,p) \le R\}) = \hat{\pi}(\{p \in \mathbb{H} : \ d'(0,p) \le R\})$, for all $R > 0$, here $0$ denotes the neutral element of $\mathbb{H}$;*
*(d) $\pi(\{v \in V : \ g(v,v) \le 1\}) = \pi(\{v' \in V' : \ g'(v',v') \le 1\})$;*
*(e) there exists a scalar product $\bar{g}$ on $\mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ such that both*
$$\pi|_V : (V,g) \to (\mathfrak{h}/[\mathfrak{h},\mathfrak{h}], \bar{g}) \quad \text{and} \quad \pi|_{V'} : (V',g') \to (\mathfrak{h}/[\mathfrak{h},\mathfrak{h}], \bar{g})$$
*are submetries, i.e., they map balls to balls.*

Next, we prove that in every class of the equivalence relation (iii) there is exactly one strictly subRiemannian metric. To every left-invariant subRiemannian metric $d$ we define the *associated asymptotic metric* $d'$ as follows. If $d$ is Riemannian defined by a scalar product $g$ on $\mathfrak{h}$, then $d'$ is the strictly subRiemannian metric for which the horizontal space $V$ is $g$-orthogonal to $[\mathfrak{h},\mathfrak{h}]$ and the scalar product is $g|_V$. If $d$ is strictly subRiemannian, then $d' = d$.

**Theorem 1.3.** *Let $d$ and $d'$ be two left-invariant subRiemannian metrics on $\mathbb{H}$. Their associated asymptotic metrics are the same if and only if*

$$\lim_{d(p,q)\to\infty} |d(p,q) - d'(p,q)| = 0. \tag{1}$$

*Moreover, if* (1) *holds, then there is* $C > 0$ *such that*

$$|d(p,q) - d'(p,q)| \leq \frac{C}{d(p,q)}, \qquad \forall p, q \in \mathbb{H}. \tag{2}$$

We remark that the estimate (2) in Theorem 1.3 is sharp, as we will show in Remark 4.2.

The above result can be interpreted in terms of asymptotic cones. Namely, if $d$ is a left-invariant Riemannian metric on $\mathbb{H}$ and $d'$ is the associated asymptotic metric, then $(\mathbb{H}, d')$ is the asymptotic cone of $(\mathbb{H}, d)$. For the analogous result in arbitrary nilpotent groups see [23]. By Theorem 1.3, more is true: $(\mathbb{H}, d)$ is at bounded distance from $(\mathbb{H}, d')$. Notice that this consequence cannot be deduced by the similar results for discrete subgroups of the Heisenberg group in [19] and [10], because the word metric is only quasi-isometric to the Riemannian one. Moreover, we remark that there are examples of nilpotent groups of step two that are not at bounded distance from their asymptotic cone, see [3].

We now focus on the horoboundary. As a consequence of Theorem 1.3 and of the results of Klein-Nikas [18], we get:

**Corollary 1.4.** *If* $d_R$ *is a left-invariant Riemannian metric on* $\mathbb{H}$ *with associated asymptotic metric* $d_{CC}$*, then the horoboundary of* $(\mathbb{H}, d_R)$ *coincides with the horoboundary of* $(\mathbb{H}, d_{CC})$*; hence, it is homeomorphic to a* 2*-dimensional closed disk* $\bar{D}^2$.

More precisely, let $g$ be the scalar product of $d_R$ on $\mathfrak{h}$ and $W \subset \mathfrak{h}$ the orthogonal plane to $[\mathfrak{h}, \mathfrak{h}]$. Define the norm $\|w\| := \sqrt{g(w,w)}$ on $W$. Fix a orthonormal basis $(X, Y)$ for $W$ and set $Z := [X, Y] \in [\mathfrak{h}, \mathfrak{h}]$, so that $(X, Y, Z)$ is a basis of $\mathfrak{h}$. We identify $\mathfrak{h} \simeq \mathbb{H}$ via the exponential map, which is a global diffeomorphism. So, we write $p = w + zZ$ with $w \in W$ and $z \in \mathbb{R}$ for any point $p \in \mathbb{H}$. We say that a sequence of points $\{p_n\}_{n\in\mathbb{N}} \subset \mathbb{H}$ diverges if it leaves every compact set. Moreover, we shall use the following terminology for a diverging sequence of the form $p_n = w_n + z_n Z$:

(1) *vertical divergence*, if there exists $M < \infty$ such that $\|w_n\| < M$ for all $n$;
(2) *non-vertical divergence with quadratic rate* $\nu \in [-\infty, +\infty]$, if $w_n$ diverges and[1] $\lim_{n\to\infty} \frac{z_n}{4\|w_n\|^2} = -\nu$.

Then, according to [18] (see Corollary 5.6, 5.9 and 5.13 therein), we deduce the following description of the Riemannian horofunctions:

**(v):** a vertically diverging sequence $p_n = w_n + z_n Z$ converges to a horofunction $h$ if and only if $w_n \to w_\infty$, and in this case

$$h(w + zZ) = \|w_\infty\| - \|w_\infty - w\|;$$

**(nv):** a non-vertically diverging sequence $p_n = w_n + z_n Z$ with quadratic rate $\nu$ converges to a horofunction $h$ if and only if $\frac{w_n}{\|w_n\|} \to \hat{w}$, and then

$$h(w + zZ) = g(R_\vartheta(-\hat{w}), w)$$

[1] From the paper [18], there is an extra 4 and a change of sign due to our different choice of coordinates.

where $R_\vartheta$ is the anti-clockwise rotation in $W$ of angle $\vartheta = \mu^{-1}(\nu)$, and $\mu : [-\pi, \pi] \to \overline{\mathbb{R}}$ is the extended Gaveau function

$$\mu(\vartheta) := \frac{\vartheta - \sin\vartheta\cos\vartheta}{\sin^2(\vartheta)}.$$

Moreover, all the horofunctions of $(\mathbb{H}, d_R)$ are of type **(v)** or **(nv)**, by Theorem 5.16 in [18]; it is also clear that neither is of both types.

In section 5 we will also determine the Busemann points of $\partial_h(\mathbb{H}, d_R)$, that is those horofunctions obtained by points diverging along quasi-geodesics (see Definition 5.1). We obtain, as in the subRiemannian case:

**Corollary 1.5.** *The Busemann points of* $(\mathbb{H}, d_R)$ *are the horofunctions of type* ***(nv)*** *and can be identified to the boundary of the disk* $\bar{D}^2$.

The paper is organized as follows. In Section 2 we introduce the main objects and their basic properties. In Section 3 we estimate the difference between any two strictly subRiemannian left-invariant metrics on $\mathbb{H}$. In Section 4 we compare any Riemannian left-invariant metric on $\mathbb{H}$ and its associated asymptotic metric. At the end of the section we shall prove Theorems 1.2 and 1.3. In Section 5 we concentrate on the horofunctions and we prove Corollaries 1.4 and 1.5. Appendix A is devoted to the explicit description of subRiemannian geodesics.

*Acknowledgments.* The initial discussions for this work were done at the '2013 Workshop on Analytic and Geometric Group Theory' in Ventotene. We express our gratitude to the organizers: A. Iozzi, G. Kuhn and M. Sageev. We also thank the anonymous referee for several improving suggestions.

## 2. Preliminaries

2.1. **Definitions.** The *first Heisenberg group* $\mathbb{H}$ is the connected, simply connected Lie group associated to the Heisenberg Lie algebra $\mathfrak{h}$. The *Heisenberg Lie algebra* $\mathfrak{h}$ is the only three dimensional nilpotent Lie algebra that is not commutative. It can be proven that, for any two linearly independent vectors $X, Y \in \mathfrak{h} \setminus [\mathfrak{h}, \mathfrak{h}]$, the triple $(X, Y, [X, Y])$ is a basis of $\mathfrak{h}$ and $[X, [X, Y]] = [Y, [X, Y]] = 0$.

We denote by $\omega_{\mathbb{H}} : T\mathbb{H} \to \mathfrak{h}$ the left-invariant Maurer-Cartan form. Namely, denoting by 0 the neutral element of $\mathbb{H}$ and identifying $\mathfrak{h}$ with $T_0\mathbb{H}$, we have $\omega_{\mathbb{H}}(v) := \mathrm{d}L_p^{-1} v$ for $v \in T_p\mathbb{H}$, where $L_p$ is the left translation by $p$.

Let $\pi : \mathfrak{h} \to \mathfrak{h}/[\mathfrak{h}, \mathfrak{h}]$ be the quotient projection. Notice that $\mathfrak{h}/[\mathfrak{h}, \mathfrak{h}]$ is a commutative 2-dimensional Lie algebra. So the map $\pi$ induces a Lie group epimorphism $\hat{\pi} : \mathbb{H} \to \mathfrak{h}/[\mathfrak{h}, \mathfrak{h}] \simeq \mathbb{H}/[\mathbb{H}, \mathbb{H}]$.

2.2. **SubRiemannian metrics in $\mathbb{H}$.** Let $V \subset \mathfrak{h}$ be a bracket generating subspace. We have only two cases: either $V = \mathfrak{h}$ or $V$ is a plane and $\mathfrak{h} = V \oplus [\mathfrak{h}, \mathfrak{h}]$. In both cases the restriction of the projection $\pi|_V : V \to \mathfrak{h}/[\mathfrak{h}, \mathfrak{h}]$ is surjective. Let $g$ be a scalar product on $V$ and set the corresponding norm $\|v\| := \sqrt{g(v, v)}$ for $v \in V$.

An absolutely continuous curve $\gamma : [0,1] \to \mathbb{H}$ is said to be *horizontal* if $\omega_{\mathbb{H}}(\gamma'(t)) \in V$ for almost every $t$. For a horizontal curve we have the *length*

$$\ell(\gamma) := \int_0^1 \|\omega_{\mathbb{H}}(\gamma'(t))\| \,\mathrm{d}t.$$

A *subRiemannian metric* $d$ is hence defined as

$$d(p,q) := \inf\left\{\ell(\gamma) : \gamma \text{ horizontal curve from } p \text{ to } q\right\}.$$

SubRiemannian metrics on $\mathbb{H}$ are complete, geodesic, and left-invariant. They are either Riemannian, when $V = \mathfrak{h}$, or *strictly subRiemannian*, when $\dim V = 2$. The pair $(V, g)$ is called the *horizontal space of* $d$.

Since $\pi|_V : V \to \mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ is surjective, it induces a norm $\|\cdot\|$ on $\mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ such that $\pi : (V, \|\cdot\|) \to (\mathfrak{h}/[\mathfrak{h},\mathfrak{h}], \|\cdot\|)$ is an submetry, i.e., for all $w \in \mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ it holds $\|w\| = \inf\{\|v\| : \pi(v) = w\}$. Here we use the same notation for norms on $V$ and on $\mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$, because there will be no possibility of confusion. The norm on $\mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ is characterized by

$$\pi\left(\{v \in V : \|v\| \le R\}\right) = \{w \in \mathfrak{h}/[\mathfrak{h},\mathfrak{h}] : \|w\| \le R\}, \tag{3}$$

for all $R > 0$.

**Proposition 2.1.** *Let $d$ be subRiemannian metric on $\mathbb{H}$ with horizontal space $(V, g)$. Then for all $R > 0$*

$$\pi\left(\{v \in V : \|v\| \le R\}\right) = \hat{\pi}\left(\{p \in \mathbb{H} : d(0,p) \le R\}\right).$$

*In particular, $\hat{\pi} : (\mathbb{H}, d) \to (\mathfrak{h}/[\mathfrak{h},\mathfrak{h}], \|\cdot - \cdot\|)$ is a submetry, i.e., for all $v, w \in \mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$*

$$\|v - w\| = \inf\{d(p,q) : \hat{\pi}(p) = v,\ \hat{\pi}(q) = w\}.$$

*Proof.* $\boxed{\subset}$ Let $v \in V$ with $\|v\| \le R$. Set $\gamma(t) := \exp(tv)$. Then $\gamma : [0,1] \to \mathbb{H}$ is a horizontal curve with $d(0, \exp(v)) \le \ell(\gamma) = \|v\| \le R$. Since $\hat{\pi}(\exp(v)) = \pi(v)$, then we have proven this inclusion.

$\boxed{\supset}$ Let $p \in \mathbb{H}$ with $d(0,p) \le R$ and let $\gamma : [0,T] \to \mathbb{H}$ be a $d$-length-minimizing curve from $0$ to $p$ parametrized by arc-length, so $T = d(0,p)$. Then $\hat{\pi} \circ \gamma : [0,T] \to \mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ is a curve from $0$ to $\pi(p)$ and

$$\begin{aligned}
\|\hat{\pi}(p)\| &\le \int_0^T \|(\hat{\pi} \circ \gamma)'(t)\| \,\mathrm{d}t \\
&= \int_0^T \|\pi \circ \omega_{\mathbb{H}}(\gamma'(t))\| \,\mathrm{d}t \\
&\le \int_0^T \|\omega_{\mathbb{H}}(\gamma'(t))\| \,\mathrm{d}t \\
&= \ell(\gamma) = d(0,p).
\end{aligned}$$

In the first equality we used the fact that $\hat{\pi}$ is a morphism of Lie groups and its differential is $\pi$, i.e., $\omega_{\mathbb{H}/[\mathbb{H},\mathbb{H}]} \circ \mathrm{d}\hat{\pi} = \pi \circ \omega_{\mathbb{H}}$, where $\omega_{\mathbb{H}/[\mathbb{H},\mathbb{H}]}$ is the Mauer-Cartan form of $\mathbb{H}/[\mathbb{H},\mathbb{H}]$. $\square$

**Proposition 2.2.** *Let $d, d'$ be two subRiemannian metrics on $\mathbb{H}$ such that*

$$\lim_{p\to\infty} \frac{d(0,p)}{d'(0,p)} = 1.$$

*Then*

$$\hat{\pi}\left(\{p \in \mathbb{H} : d(0,p) \leq R\}\right) = \hat{\pi}\left(\{p \in \mathbb{H} : d'(0,p) \leq R\}\right). \tag{4}$$

*Proof.* Let $\|\cdot\|$ and $\|\cdot\|'$ be the norms on $\mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ induced by $d$ and $d'$, respectively. We will show that

$$\lim_{v\to\infty} \frac{\|v\|}{\|v\|'} = 1, \tag{5}$$

which easily implies $\|\cdot\| = \|\cdot\|'$, because for any fixed $v \in \mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ one has $1 = \lim_{t\to\infty} \frac{\|tv\|}{\|tv\|'} = \frac{\|v\|}{\|v\|'}$. Moreover, the equality (4) follows from Proposition 2.1 combined with (3) and (5).

Since both maps $\hat{\pi} : (\mathbb{H}, d) \to (\mathfrak{h}/[\mathfrak{h},\mathfrak{h}], \|\cdot\|)$ and $\hat{\pi} : (\mathbb{H}, d') \to (\mathfrak{h}/[\mathfrak{h},\mathfrak{h}], \|\cdot\|')$ are submetries, for every $v \in \mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ there are $p_v, p'_v \in \mathbb{H}$ such that $\hat{\pi}(p_v) = \hat{\pi}(p'_v) = v$, $\|v\| = d(0,p_v)$ and $\|v\|' = d'(0,p'_v)$.

Moreover it holds $\|v\|' \leq d'(0,p_v)$ and $\|v\| \leq d(0,p'_v)$, again because $\hat{\pi}$ is a submetry in both cases. Therefore

$$\frac{d(0,p_v)}{d'(0,p_v)} \leq \frac{\|v\|}{\|v\|'} \leq \frac{d(0,p'_v)}{d'(0,p'_v)}$$

Finally, if $v \to \infty$, then both $d(0,p_v)$ and $d(0,p'_v)$ go to infinity as well. The relation (5) is thus proven. □

2.3. **Balayage area and lifting of curves.** Let $V \subset \mathfrak{h}$ be a two-dimensional subspace with $V \cap [\mathfrak{h},\mathfrak{h}] = \{0\}$. Then $[\mathfrak{h},\mathfrak{h}] = [V,V]$, i.e., $V$ is bracket generating. Moreover, $\pi|_V : V \to \mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ is an isomorphism.

If $\rho : [0,T] \to \mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ is a curve with $\rho(0) = 0$, then there is a unique $\tilde{\rho} : [0,T] \to \mathbb{H}$ such that

$$\begin{cases} \tilde{\rho}(0) = 0, \\ \omega_{\mathbb{H}}(\tilde{\rho}'(t)) = \pi|_V^{-1}(\rho(t)'). \end{cases}$$

Since $(\pi \circ \tilde{\rho})' = \rho'$, then $\pi \circ \tilde{\rho} = \rho$. So, $\tilde{\rho}$ is called the *lift* of $\rho$.

The previous ODE system that defines $\tilde{\rho}$ can be easily integrated. Let $X, Y \in V$ be a basis, set $Z := [X,Y]$, so that $(X,Y,Z)$ is a basis of $\mathfrak{h}$. Let $(x,y,z) = \exp(xX + yY + zZ)$ be the exponential coordinates on $\mathbb{H}$ defined by $(X,Y,Z)$. Using the Backer-Campbell-Hausdorff formula, one shows that $X, Y, Z$ induce the following left-invariant vector fields on $\mathbb{H}$:

$$\hat{X} = \partial_x - \frac{y}{2}\partial_z, \qquad \hat{Y} = \partial_y + \frac{x}{2}\partial_z, \qquad \hat{Z} = \partial_z.$$

Thanks to these vector fields, we can describe the Maurer-Cartan form as

$$\omega_{\mathbb{H}}(a\hat{X} + b\hat{Y} + c\hat{Z}) = aX + bY + cZ.$$

The lift of $\rho$ is hence given by the ODE

$$\begin{cases} \tilde{\rho}'_1 = \rho'_1, \\ \tilde{\rho}'_2 = \rho'_2, \\ \tilde{\rho}'_3 = \frac{1}{2}\left(\rho_1\rho'_2 - \rho_2\rho'_1\right). \end{cases}$$

Take the coordinates $(x, y)$ on $\mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ given by the basis $(\pi(X), \pi(Y))$ and define the *balayage area* of a curve $\rho : [0, T] \to \mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ as

$$\mathscr{A}(\rho) = \frac{1}{2}\int_\rho (x\,\mathrm{d}y - y\,\mathrm{d}x). \tag{6}$$

If $\rho(0) = 0$, then the balayage area of $\rho$ corresponds to the signed area enclosed between the curve $\rho$ and the line passing through 0 and $\rho(T)$.

It follows that

$$\tilde{\rho}(t) = \left(\rho_1(t), \rho_2(t), \mathscr{A}(\rho|_0^t)\right).$$

In an implicit form we can write

$$\tilde{\rho}(t) = \exp\left((\pi|_V)^{-1}(\rho(t)) + \mathscr{A}(\rho|_0^t)Z\right). \tag{7}$$

Notice that the lift $\tilde{\rho}$ of a curve $\rho$ depends on the choice of $V$. Moreover, both the area and the Balayage area in $\mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ depend on the choice of the basis $(X, Y)$. Nevertheless, once a plane $V \subset \mathfrak{h}$ is fixed, the lift $\tilde{\rho}$ does not depend on the choice of the basis $X, Y$.

If $g$ is a scalar product on $V$ and $d$ is the corresponding strictly sub-Riemannian metric, the balayage area gives a characterization of $d$-length-minimizing curves. Let $\bar{g}$ be the scalar product on $\mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ induced by $g$. Then the $d$-length of a curve $\tilde{\rho} : [0, T] \to \mathbb{H}$ equals the length of $\rho = \pi \circ \tilde{\rho}$.

Therefore, given $p = (x, y, z) \in \mathbb{H}$, we have

$$d(0, p) = \inf\left\{\ell(\rho) : \rho : [0, 1] \to \mathfrak{h}/[\mathfrak{h},\mathfrak{h}], \rho(0) = 0, \rho(1) = \hat{\pi}(p), \mathscr{A}(\rho) = z\right\}.$$

This express the so-called *Dido's problem* in the plane, and the solutions are arc of circles. It degenerates into a line if $z = 0$. We can summarize the last discussion in the following result.

**Lemma 2.3.** *A curve $\tilde{\rho} : [0, 1] \to \mathbb{H}$ is $d$-length-minimizing from 0 to $p = (x, y, z)$ if and only if $\rho := \hat{\pi} \circ \tilde{\rho}$ is an arc of a circle from 0 to $\hat{\pi}(p)$ with $\mathscr{A}(\rho) = z$.*

## 3. Comparison between strictly subRiemannian metrics

The present section is devoted to comparing strictly subRiemannian metrics. For such metrics, Proposition 3.1 gives the only non-trivial implication in Theorem 1.2. The general case will follow from Proposition 4.1.

**Proposition 3.1.** *Let $d$ and $d'$ be two strictly subRiemannian metrics on $\mathbb{H}$ with horizontal spaces $(V, g)$ and $(V', g')$, respectively. Suppose that there exists a scalar product $\bar{g}$ on $\mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ such that both*

$$\pi|_V : (V, g) \to (\mathfrak{h}/[\mathfrak{h},\mathfrak{h}], \bar{g}) \quad \text{and} \quad \pi|_{V'} : (V', g') \to (\mathfrak{h}/[\mathfrak{h},\mathfrak{h}], \bar{g})$$

*are submetries.*

*Then*

$$\sup_{p\in\mathbb{H}} |d(0, p) - d'(0, p)| < \infty. \tag{8}$$

*Moreover, if $d \neq d'$, then*

$$\limsup_{p\to\infty} |d(0, p) - d'(0, p)| > 0. \tag{9}$$

In the proof we will give the exact value of the supremum in (8). Indeed, by (11) and (12), we get $\sup_{p\in\mathbb{H}} |d(0,p) - d'(0,p)| = 2|h|$, where $h$ is defined below. For (8), we will first prove that two of such subRiemannian metrics are isometric via a conjugation $x \mapsto gxg^{-1}$ for some $g \in \mathbb{H}$ and then we apply Lemma 3.2. For (9), we will give a sequence $p_n \to \infty$ and a constant $c > 0$ such that $|d(0,p_n) - d'(0,p_n)| > c$ for all $n \in \mathbb{N}$.

3.1. **Proof of** (8)**.** Since $\dim V = \dim V' = 2$, then $\pi|_V$ and $\pi|_{V'}$ are isomorphisms. Therefore by the assumption they are isometries onto $(\mathfrak{h}/[\mathfrak{h},\mathfrak{h}], \bar{g})$.

Let $X \in V \cap V'$ be with $g(X,X) = 1$. Then $g'(X,X) = 1$ as well.

Let $Y \in V$ be orthogonal to $X$ with $g(Y,Y) = 1$. Then $Z := [X,Y] \neq 0$ and $(X,Y,Z)$ is a basis of $\mathfrak{h}$.

Let $Y' := \pi|_{V'}^{-1}(\pi(Y)) \in V'$. Then $g'(Y',Y') = 1$ and $g'(X,Y') = 0$. Moreover, there is $h \in \mathbb{R}$ such that $Y' = Y + hZ$. In particular, $[X,Y'] = Z$.

Using the formula $\mathrm{Ad}_{\exp(hX)}(v) = e^{\mathrm{ad}_{hX}} v = v + h[X,v]$, we notice that

$$\begin{cases} \mathrm{Ad}_{\exp(hX)}(X) = X, \\ \mathrm{Ad}_{\exp(hX)}(Y) = Y', \\ \mathrm{Ad}_{\exp(hX)}(Z) = Z. \end{cases} \tag{10}$$

In particular $\mathrm{Ad}_{\exp(hX)}|_V : (V,g) \to (V',g')$ is an isometry.

Therefore, the conjugation

$$C_{\exp(hX)}(p) := \exp(hX) \cdot p \cdot \exp(hX)^{-1}$$

is an isometry $C_{\exp(hX)} : (\mathbb{H}, d) \to (\mathbb{H}, d')$.

We can now use the following Lemma 3.2 and get

$$\sup_{p\in\mathbb{H}} |d(0,p) - d'(0,p)| \leq 2|h|. \tag{11}$$

**Lemma 3.2.** *Let $G$ be a group with neutral element $e$ and let $d, d'$ be two left-invariant distances on $G$. If there is $g \in G$ such that for all $p \in G$*

$$d'(e,p) = d(e, gpg^{-1}),$$

*then for all $p \in G$*

$$|d(e,p) - d'(e,p)| \leq 2\min\{d(e,g), d'(e,g)\}.$$

*Proof.* Note that, since $d$ is left invariant, then for all $a, b \in G$ we have $d(e,ab) \leq d(e,a) + d(e,b)$ and $d(e,a) = d(e,a^{-1})$. On the one side, we have $d'(e,p) = d(e,gpg^{-1}) \leq d(e,p) + 2d(e,g)$. On the other side, we have $d(e,p) = d(e, g^{-1}gpg^{-1}g) \leq d(e,g^{-1}) + d(e,gpg^{-1}) + d(e,g) = 2d(e,g) + d'(e,p)$. Hence $|d(e,p) - d'(e,p)| \leq 2d(e,g)$. By symmetry, we have also $|d(e,p) - d'(e,p)| \leq 2d'(e,g)$. □

3.2. **Proof of** (9)**.** We keep the same notation of the previous subsection. Up to switching $V$ with $V'$, we can assume $h > 0$.

Let $(x,y,z)$ be the exponential coordinates on $\mathbb{H}$ induced by the basis $(X,Y,Z)$ of $\mathfrak{h}$, i.e., $(x,y,z) = \exp(xX + yY + zZ) \in \mathbb{H}$. Similary, on $\mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ we have coordinates $(x,y) = x\pi(X) + y\pi(Y)$.

For $R > 0$, define

$$p_R := \left(0, 2R, \frac{\pi R^2}{2} + 2hR\right).$$

We will show that

$$\lim_{R\to\infty} d(0, p_R) - d'(0, p_R) = 2h. \tag{12}$$

Fix $R > 0$. Let $\gamma : [0, T] \to \mathbb{H}$ be a $d'$-minimizing curve from 0 to $p_R$. Then $\hat{\pi} \circ \gamma : [0, T] \to \mathfrak{h}/[\mathfrak{h}, \mathfrak{h}]$ is half circle of center $(0, R)$ and radius $R$. The balayage area of $\hat{\pi} \circ \gamma$ is

$$\mathscr{A}(\hat{\pi} \circ \gamma) = \frac{\pi R^2}{2}.$$

Let $\eta : [0, T] \to \mathbb{H}$ be the $d$-length-minimizing curve from 0 to $p_R$. Then $\hat{\pi} \circ \eta : [0, T] \to \mathfrak{h}/[\mathfrak{h}, \mathfrak{h}]$ is an arc of a circle of radius $S_R$ whose balayage area is

$$\mathscr{A}(\hat{\pi} \circ \eta) = \frac{\pi R^2}{2} + 2Rh = \mathscr{A}(\hat{\pi} \circ \gamma) + 2Rh. \tag{13}$$

It is clear that $S_R > R$ and that the circle of $\hat{\pi} \circ \eta$ has center $(\mu_R, R)$ for some $\mu_R > 0$. So we have

$$S_R^2 = R^2 + \mu_R^2.$$

It is also clear from the picture that

$$\frac{\pi S_R^2}{2} + 2R\mu_R \leq \mathscr{A}(\hat{\pi} \circ \eta) \leq \frac{\pi S_R^2}{2} + 2S_R\mu_R. \tag{14}$$

Now, let's look at the lengths. First of all, notice that $\ell_{d'}(\gamma) = \ell(\hat{\pi} \circ \gamma)$ and $\ell_d(\eta) = \ell(\hat{\pi} \circ \eta)$. For one curve we have

$$\ell(\hat{\pi} \circ \gamma) = \pi R,$$

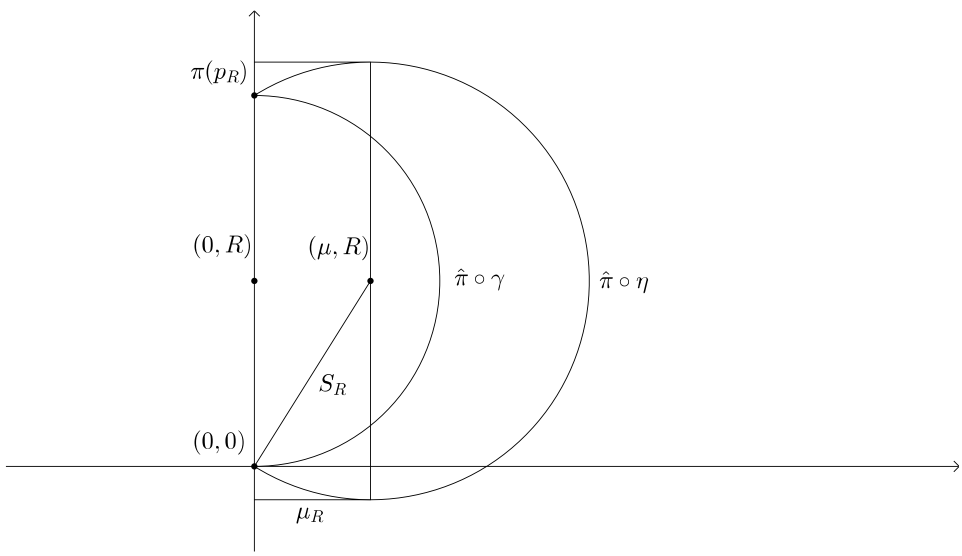


FIGURE 1. Curves in $\mathfrak{h}/[\mathfrak{h}, \mathfrak{h}]$ for the proof of (9).

for the other we have the estimate

$$\pi S_R + 2\mu_R \leq \ell(\hat{\pi} \circ \eta),$$

which is clear from the picture. Hence

$$\begin{aligned}\liminf_{R\to\infty} d(0,p) - d'(0,p) &= \liminf_{R\to\infty} \ell(\hat{\pi}\circ\eta) - \ell(\hat{\pi}\circ\gamma) \\ &\geq \lim_{R\to\infty} \pi S_R + 2\mu_R - \pi R \\ &= \lim_{R\to\infty} \pi(S_R - R) + 2\mu_R.\end{aligned}$$

We claim that

$$\lim_{R\to\infty} \pi(S_R - R) + 2\mu_R = 2h. \tag{15}$$

Let us start by checking that,

$$\mu_R < h. \tag{16}$$

Indeed, from the first inequality of (14) together with (13) it follows

$$\frac{\pi S_R^2}{2} + 2R\mu_R \leq \frac{\pi R^2}{2} + 2hR.$$

Since $S_R > R$, then

$$0 < \frac{\pi S_R^2}{2} - \frac{\pi R^2}{2} \leq 2R(h - \mu_R),$$

i.e., the inequality (16).

From the second inequality of (14) together with (13) we get

$$\frac{\pi R^2}{2} + 2Rh \leq \frac{\pi S_R^2}{2} + 2S_R\mu_R.$$

Using the facts $\mu_R \leq h$ and $S_R \leq R+\mu_R \leq R+h$, from the above inequality one gets

$$\begin{aligned}0 \leq 2(h-\mu_R) &\leq (S_R - R)\frac{1}{R}\left(\frac{\pi}{2}(S_R + R) + 2\mu_R\right) \\ &\leq (S_R - R)\left(\pi + \frac{h}{R}(\frac{\pi}{2} + 2)\right)\end{aligned} \tag{17}$$

Moreover, since $h^2 \geq \mu_R^2 = S_R^2 - R^2 = (S_R - R)(S_R + R)$, we also have

$$\lim_{R\to\infty} (S_R - R) = 0. \tag{18}$$

Finally, from (18) and (17) we obtain (15), as claimed. This completes the proof of (12) and of Proposition 3.1.

## 4. Comparison between Riemannian and strictly subRiemannian metrics

Let $d_R$ be a Riemannian metric on $\mathbb{H}$ with horizontal space $(\mathfrak{h}, g)$.

Let $V \subset \mathfrak{h}$ be the plane orthogonal to $[\mathfrak{h},\mathfrak{h}]$ and let $d_{CC}$ be the strictly subRiemannian metric on $\mathbb{H}$ with horizontal space $(V, g|_V)$.

Fix a basis $(X, Y, Z)$ for $\mathfrak{h}$ such that $(X, Y)$ is an orthonormal basis of $(V, g|_V)$ and $Z = [X, Y]$. The matrix representation of $g$ with respect to $(X, Y, Z)$ is

$$g = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \zeta^2 \end{pmatrix}$$

where $\zeta > 0$.

Let $d_{CC}$ be the strictly subRiemannian metric on $\mathbb{H}$ with horizontal space $(V, g|_V)$.

Our aim in this section is to prove the following proposition.

**Proposition 4.1.** *If $d_{CC}(0, p)$ is large enough, then:*

$$0 \le d_{CC}(0,p) - d_R(0,p) \le \frac{4\pi^2}{\zeta^2} \frac{1}{d_{CC}(0,p) - \frac{2^{3/2}\pi}{\zeta}}. \tag{19}$$

*In particular it holds*

$$\lim_{p\to\infty} |d_{CC}(0,p) - d_R(0,p)| = 0. \tag{20}$$

For the proof of this statement, we need to know length-minimizing curves for $d_R$ and $d_{CC}$, and a few properties of those, see the exposition in the Appendix A.

*Proof.* Let $(x, y, z)$ be the exponential coordinates on $\mathbb{H}$ induced by the basis $(X, Y, Z)$ of $\mathfrak{h}$, i.e., $(x, y, z) = \exp(xX + yY + zZ) \in \mathbb{H}$. Fix $p = (p_1, p_2, p_3) \in \mathbb{H}$.

Notice that both $d_R$ and $d_{CC}$ are generated as length metrics using the same length measure $\ell$, with the difference that $d_R$ minimizes the length among all the curves, while $d_{CC}$ takes into account only the curves tangent to $V$. This implies that

$$\forall p, q \in \mathbb{H} \qquad d_{CC}(p,q) \ge d_R(p,q),$$

therefore we get the first inequality in (19). We need to prove the second inequality of (19).

If $p \in \{z = 0\}$, then $d_{CC}(0, p) = d_R(0, p)$ by Corollary A.7, and the claim is true.

Suppose $p \notin \{z = 0\}$ and let $\gamma : [0, T] \to \mathbb{H}$ be a $d_R$-length minimizing curve from $0 = \gamma(0)$ to $p = \gamma(T)$. Since $p \notin \{z = 0\}$ and since we supposed that $d_{CC}(0, p)$ is large enough, then by Corollary A.5 we can parametrize $\gamma$ in such a way that $\gamma$ is exactly in the form expressed in Type II in Proposition A.2 for some $k > 0$ and $\theta \in \mathbb{R}$.

By Corollary A.4 it holds

$$kT \le 2\pi. \tag{21}$$

Moreover, by Corollary A.8

$$d_R(0,p) = \|\omega_{\mathbb{H}}(\gamma')\| \cdot T = \sqrt{1 + \frac{k^2}{\zeta^2}} \cdot T. \tag{22}$$

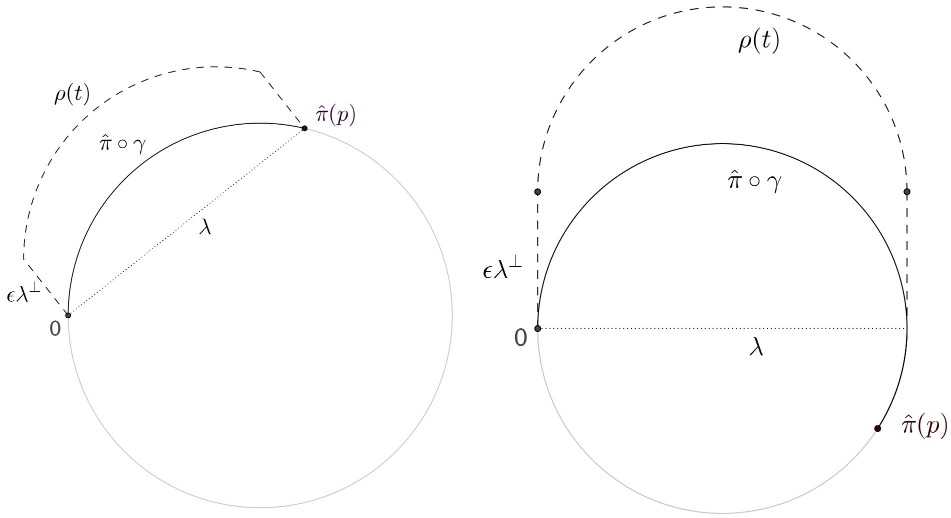


FIGURE 2. Curves for Case 1 and Case 2.

Let $\eta : [0, T] \to \mathbb{H}$ be the $d_{CC}$-length-minimizing curve corresponding to $\gamma$ as shown in Corollary A.8. Then we know that $d_{CC}(0, \eta(T)) = \ell(\eta) = T$, and

$$p = \gamma(T) = \eta(T) + (0, 0, \frac{kT}{\zeta^2}). \tag{23}$$

Hence by Corollary A.8 and (21)

$$d_{CC}(0, p) \leq d_{CC}(0, \eta(T)) + d_{CC}(\eta(T), \gamma(T)) \leq T + \frac{2^{3/2}\pi}{\zeta},$$

i.e.,

$$\frac{1}{T} \leq \frac{1}{d_{CC}(0, p) - \frac{2^{3/2}\pi}{\zeta}}. \tag{24}$$

Since $\eta$ is a $d_{CC}$-rectifiable curve, then $\eta(T)_3 = \mathscr{A}(\hat{\pi} \circ \eta)$, where $\eta(T)_3$ is the third coordinate of the point in the exponential coordinates. Since $\hat{\pi} \circ \gamma = \hat{\pi} \circ \eta$, then we have by (23)

$$p_3 = \mathscr{A}(\hat{\pi} \circ \gamma) + \frac{kT}{\zeta^2}. \tag{25}$$

Notice that $\hat{\pi} \circ \gamma$ is an arc of a circle in $\mathfrak{h}/[\mathfrak{h}, \mathfrak{h}]$ of radius $\frac{1}{k}$, see Proposition A.2.

Now we want to define a horizontal curve $\tilde{\rho} : [-\epsilon, T + \epsilon] \to \mathbb{H}$, where $\epsilon > 0$ has to be chosen, such that $\tilde{\rho}(-\epsilon) = 0$ and $\tilde{\rho}(T + \epsilon) = p$. We first define a curve $\rho : [-\epsilon, T + \epsilon] \to \mathfrak{h}/[\mathfrak{h}, \mathfrak{h}]$ and then take its lift to $\mathbb{H}$.

For the definition of $\rho$ we follow two different strategies for two different cases:

CASE 1. Suppose that $\hat{\pi} \circ \gamma$ doesn't cover the half of the circle, i.e., $T \leq \frac{\pi}{k}$. Set $\lambda = \hat{\pi}(p) \in \mathfrak{h}/[\mathfrak{h}, \mathfrak{h}]$. Then $T$ is smaller than the circle of diameter $\|\lambda\|$, i.e.,

$$\|\lambda\| \geq \frac{T}{\pi}. \tag{26}$$

Let $\lambda^\perp \in \mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ be the unit vector perpendicular to $\lambda$ and forming an angle smaller than $\pi/2$ with the arc $\hat{\pi}\circ\gamma$. Let $\epsilon > 0$ such that

$$\epsilon \cdot \|\lambda\| = \frac{kT}{\zeta^2}. \tag{27}$$

Now, define $\rho : [-\epsilon, T+\epsilon] \to \mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ as

$$\rho(t) = \begin{cases} (t+\epsilon)\lambda^\perp & \text{for } -\epsilon \leq t \leq 0 \\ \epsilon\lambda^\perp + \hat{\pi}\circ\gamma(t) & \text{for } 0 \leq t \leq T \\ \epsilon\lambda^\perp + \hat{\pi}\circ\gamma(T) - (t-T)\lambda^\perp & \text{for } T \leq t \leq T+\epsilon \end{cases}$$

Notice that

$$\mathscr{A}(\rho) = \mathscr{A}(\hat{\pi}\circ\gamma) + \epsilon\cdot\|\lambda\| = \mathscr{A}(\hat{\pi}\circ\gamma) + \frac{kT}{\zeta^2} \overset{(25)}{=} p_3$$

and that $\rho(T+\epsilon) = \hat{\pi}\circ\gamma(T) = \hat{\pi}(p)$. Then the horizontal lift $\tilde{\rho} : [-\epsilon, T+\epsilon] \to \mathbb{H}$ of $\rho$ is a $d_{CC}$-rectifiable curve from 0 to $p$.

Case 2. Suppose that $\hat{\pi}\circ\gamma$ covers more than half of the circle. Let $\lambda \in \mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ be the diameter of the circle that contains 0. Since $T$ is shorter than the whole circle, then

$$\|\lambda\| \geq \frac{T}{\pi}. \tag{28}$$

Let $\lambda^\perp$ be the unit vector perpendicular to $\lambda$ and forming an angle smaller than $\pi/2$ with the arc $\hat{\pi}\circ\gamma$. Let $\epsilon > 0$ be such that

$$\epsilon \cdot \|\lambda\| = \frac{kT}{\zeta^2}. \tag{29}$$

Now, define $\rho : [-\epsilon, T+\epsilon] \to \mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ as

$$\rho(t) = \begin{cases} (t+\epsilon)\lambda^\perp & \text{for } -\epsilon \leq t \leq 0 \\ \epsilon\lambda^\perp + \hat{\pi}\circ\gamma(t) & \text{for } 0 \leq t \leq \frac{\pi\|\lambda\|}{2} \\ \epsilon\lambda^\perp + \lambda - (t - \frac{\pi\|\lambda\|}{2})\lambda^\perp & \text{for } \frac{\pi\|\lambda\|}{2} \leq t \leq \frac{\pi\|\lambda\|}{2} + \epsilon \\ \hat{\pi}\circ\gamma(t-\epsilon) & \text{for } \frac{\pi\|\lambda\|}{2} + \epsilon \leq t \leq T+\epsilon \end{cases}$$

where we used the fact $\lambda = \hat{\pi}\circ\gamma(\frac{\pi\|\lambda\|}{2})$. Notice that

$$\mathscr{A}(\rho) = \mathscr{A}(\hat{\pi}\circ\gamma) + \epsilon\cdot\|\lambda\| = \mathscr{A}(\hat{\pi}\circ\gamma) + \frac{kT}{\zeta^2} = p_3.$$

Then the horizontal lift $\tilde{\rho} : [-\epsilon, T+\epsilon] \to \mathbb{H}$ of $\rho$ is a $d_{CC}$-rectifiable curve from 0 to $p$.

In both cases $\tilde{\rho}$ is a horizontal curve from 0 to $p$ of length

$$\ell(\tilde{\rho}) = T + 2\epsilon \tag{30}$$

Moreover, from (26) and (27) (respectively (28) and (29)) we get

$$\epsilon = \frac{kT}{\zeta^2\|\lambda\|} \leq \frac{kT}{\zeta^2}\frac{\pi}{T} \overset{(21)}{\leq} \frac{2\pi}{\zeta^2}\frac{\pi}{T} = \frac{2\pi^2}{\zeta^2 T} \tag{31}$$

Finally using in order (22), (30), (31), (24)

$$\begin{aligned} d_{CC}(0,p) - d_R(0,p) &\leq \ell(\tilde{\rho}) - \sqrt{1+\frac{k^2}{\zeta^2}}\cdot T \leq \\ &\leq T + 2\epsilon - \sqrt{1+\frac{k^2}{\zeta^2}}\cdot T \leq 2\epsilon \leq 2\frac{2\pi^2}{\zeta^2 T} \leq \frac{4\pi^2}{\zeta^2}\frac{1}{d_{CC}(0,p)-\frac{2^{3/2}\pi}{\zeta}}. \end{aligned}$$

□

**Remark 4.2.** The inequality (2) is sharp. Indeed, for $z\to\infty$, we have the asymptotic equivalence

$$d_{CC}(0,(0,0,z)) - d_R(0,(0,0,z)) \sim \frac{4\pi^2}{\zeta^2}\frac{1}{d_{CC}(0,(0,0,z))}. \tag{32}$$

*Proof of* (32). We claim that, for $z>0$ large enough,

$$d_R(0,(0,0,z)) = 2\sqrt{\pi}\sqrt{z-\frac{\pi}{\zeta^2}}. \tag{33}$$

Let $\gamma:[0,T]\to\mathbb{H}$ be a $d_R$-length-minimizing curve from 0 to $(0,0,z)$. Since $z$ is large, we assume that $\gamma$ is of (Type II), see Proposition A.2, for some $k>0$ and $\theta=0$. Since the end point is on the $Z$ axis, we have

$$kT = 2\pi \tag{34}$$

and $z=\frac{T}{2k}+\frac{kT}{\zeta^2}$, from which follows

$$T^2 = 4\pi\left(z-\frac{2\pi}{\zeta^2}\right). \tag{35}$$

We know also the length of $\gamma$ (see Corollary A.8) and so we get

$$\begin{aligned} d_R(0,(0,0,z)) = \ell(\gamma) = T\|\omega_{\mathbb{H}}(\gamma')\| = T\sqrt{1+\frac{k^2}{\zeta^2}} &= \sqrt{T^2+\frac{4\pi^2}{\zeta^2}} \\ &= \sqrt{4\pi\left(z-\frac{2\pi}{\zeta^2}\right)+\frac{4\pi^2}{\zeta^2}} = 2\sqrt{\pi}\sqrt{z-\frac{\pi}{\zeta^2}}. \end{aligned}$$

Claim (33) is proved. From Corollary A.6 we get $d_{CC}(0,(0,0,z)) = 2\sqrt{\pi}\sqrt{z}$ and

$$\begin{aligned} d_{CC}(0,(0,0,z)) - d_R(0,(0,0,z)) &= 2\sqrt{\pi}\left(\sqrt{z}-\sqrt{z-\frac{\pi}{\zeta^2}}\right) \\ &= \frac{2\sqrt{\pi}}{\sqrt{z}}\frac{\frac{\pi}{\zeta^2}}{1+\sqrt{1-\frac{\pi}{\zeta^2 z}}} = \frac{1}{2\sqrt{\pi}\sqrt{z}}\frac{4\pi^2}{\zeta^2}\frac{1}{1+\sqrt{1-\frac{\pi}{\zeta^2 z}}}. \end{aligned}$$

□

We are now ready to give the proof of the main theorems:

*Proof of Theorem 1.2.* The implication $(a) \Rightarrow (b)$ is trivial. The implication $(b) \Rightarrow (c)$ is proven in Proposition 2.2. The equivalence $(c) \Leftrightarrow (d)$ follows from Proposition 2.1. The assertion $(e)$ is a restatement of $(d)$. For $(d) \Rightarrow (a)$ one uses Proposition 4.1 in order to reduce to the case when both $d$ and $d'$ are strictly subRiemannian and then one applies Proposition 3.1. □

*Proof of Theorem 1.3.* This is a consequence of Propositions 4.1 and of the sharpness result (9) of Proposition 3.1. □

## 5. The horoboundary

Let $(X, d)$ be a geodesic space and $\mathscr{C}(X)$ the space of continuous functions $X \to \mathbb{R}$ endowed with the topology of the uniform convergence on compact sets. The map $\iota : X \hookrightarrow \mathscr{C}(X)$, $(\iota(x))(y) := d(x, y)$, is an embedding, i.e., a homeomorphism onto its image.

Let $\mathscr{C}(X)/\mathbb{R}$ be the topological quotient of $\mathscr{C}(X)$ with kernel the constant functions, i.e., for every $f, g \in \mathscr{C}(X)$ we set the equivalence relation $f \sim g \Leftrightarrow f - g$ is constant.

Then the map $\hat{\iota} : X \hookrightarrow \mathscr{C}(X)/\mathbb{R}$ is still an embedding. Indeed, since the map $\mathscr{C}(X) \to \mathscr{C}(X)/\mathbb{R}$ is continuous and open, we only need to show that $\hat{\iota}$ is injective: if $x, x' \in X$ are such that $\iota(x) - \iota(x')$ is constant, then one takes $z \in Z$ such that $d(x, z) = d(x', z)$, which exists because $(X, d)$ is a geodesic space, and checks that

$$d(x, x') = \iota(x)(x') - \iota(x')(x') = \iota(x)(z) - \iota(x')(z) = 0.$$

Define the *horoboundary of* $(X, d)$ as

$$\partial_h X := cl(\hat{\iota}(X)) \setminus \hat{\iota}(X) \subset \mathscr{C}(X)/\mathbb{R},$$

where $cl(\hat{\iota}(X))$ is the topological closure.

Another description of the horoboundary is possible. Fix $o \in X$ and set

$$\mathscr{C}(X)_o := \{f \in \mathscr{C}(X) : f(o) = 0\}.$$

Then the restriction of the quotient projection $\mathscr{C}(X)_o \to \mathscr{C}(X)/\mathbb{R}$ is an isomorphism of topological vector spaces. Indeed, one easily checks that it is both injective and surjective, and that its inverse map is $[f] \mapsto f - f(o)$, where $[f] \in \mathscr{C}(X)/\mathbb{R}$ is the class of equivalence of $f \in \mathscr{C}(X)$.

Hence, we can identify $\partial_h X$ with a subset of $\mathscr{C}(X)_o$. More explicitly: $f \in \mathscr{C}(X)_o$ belongs to $\partial_h X$ if and only if there is a sequence $p_n \in X$ such that $p_n \to \infty$ (i.e., for every compact $K \subset X$ there is $N \in \mathbb{N}$ such that $p_n \notin K$ for all $n > N$) and the sequence of functions $f_n \in \mathscr{C}(X)_o$,

$$f_n(x) := d(p_n, x) - d(p_n, o), \tag{36}$$

converge uniformly on compact sets to $f$.

*Proof of Corollary 1.4.* Let us first remark that if $d, d'$ are two geodesic distances on $X$ and

$$\lim_{d(p,q)+d'(p,q)\to\infty} |d'(p, q) - d(p, q)| = 0. \tag{37}$$

then

$$\partial_h(X, d') = \partial_h(X, d).$$

Indeed, first of all the space $\mathscr{C}(X)_o$ depends only on the topology of $X$. Moreover, if $f \in \partial_h(X,d)$, let $p_n \in X$ be a sequence as in (36) and set $f'_n(x) := d'(p_n,x) - d'(p_n,o)$. Then

$$|f'_n(x) - f_n(x)| \leq |d'(p_n,x) - d(p_n,x)| + |d'(p_n,o) - d(p_n,o)|,$$

and as a consequence of (37) we get $f'_n \to f$ uniformly on compact sets. This shows $\partial_h(X,d) \subset \partial_h(X,d')$. The other inclusion follows by the simmetry of (37) in $d$ and $d'$.
Now, if $d_R$ and $d_{CC}$ are metrics on $\mathbb{H}$ like in Corollary 1.4, then (37) is easily satisfied thanks to Theorem 1.3, and therefore $\partial_h(\mathbb{H},d_R) = \partial_h(\mathbb{H},d_{CC})$ if the Riemannian metric $d_R$ and the subRiemannian metric $d_{CC}$ are compatible. The conclusion follows from [18]. □

The Busemann points in the boundary $\partial_h(X,d)$ are usually defined as the horofunctions associated to sequences of points $(p_n)$ diverging to infinity along rays or "almost geodesic rays". However, in literature there are different definitions of almost geodesic rays, according to the generality of the metric space $(X,d)$ under consideration ([14], [25], [9]). A map $\gamma : I = [0,+\infty) \to (X,d)$ into a complete length space is called

- a *quasi-ray*, if the length excess
$$\Delta_N(\gamma) = \sup_{t,s \in [N,+\infty)} \ell(\gamma;t,s) - d(\gamma(t),\gamma(s))$$
tends to zero for $N \to +\infty$;
- an *almost geodesic ray*, if
$$\Theta_N(\gamma) = \sup_{t,s \in [N,+\infty)} d(\gamma(t),\gamma(s)) + d(\gamma(s),\gamma(0)) - t$$
tends to zero for $N \to +\infty$.

(Notice that the second definition depends on the parametrization, while the first one is intrinsic). We will use here a notion of Busemann points which is more general than both of them:

**Definition 5.1.** A sequence of points $(p_n)$ diverging to infinity in $(X,d)$ is said to diverge *almost straightly* if for all $\epsilon > 0$ there exists $L$ such that for every $n \geq m \geq L$ we have

$$d(p_L,p_m) + d(p_m,p_n) - d(p_L,p_n) < \epsilon \tag{38}$$

It is easy to verify that points diverging along a quasi-ray or along an almost-geodesic ray diverge almost straightly. We then define a *Busemann point* as a horofunction $f$ which is the limit of a sequence $f_n(x) = d(p_n,x) - d(p_n,o)$, for points $(p_n)$ diverging to infinity almost straightly.

To prove Corollary 1.5, we need the following lemma. We remind that a metric space is *boundedly compact* if closed balls are compact.

**Lemma 5.2.** *Let $(X,d)$ be a boundedly compact geodesic space, $o \in X$ and $\{p_n\}_{n\in\mathbb{N}} \subset X$ a sequence of points diverging almost straightly. Then:*

*(i) the sequence $f_n(x) = d(p_n,x) - d(p_n,o)$ converges uniformly on compacts to a horofunction $f$;*
*(ii)* $\lim_{n\to\infty} f(p_n) + d(o,p_n) = 0$.

*Proof.* Since the 1-Lipschitz functions $f_n$ are uniformly bounded on compact sets and $(X, d)$ is boundedly compact, then the family $\{f_n\}_{n\in\mathbb{N}}$ is precompact with respect to the uniform convergence on compact sets. Hence, if we prove that there is a unique accumulation point, then we obtain that the whole sequence $\{f_n\}_{n\in\mathbb{N}}$ converges.

So, let $g, g' \in \mathscr{C}(X)$ and let $\{f_{n_k}\}_{k\in\mathbb{N}}$ and $\{f_{n'_k}\}_{k\in\mathbb{N}}$ be two subsequences of $\{f_n\}_{n\in\mathbb{N}}$ such that $f_{n_k} \to g$ and $f_{n'_k} \to g'$ uniformly on compact sets. We claim

$$\forall\epsilon > 0\ \exists R_\epsilon \in \mathbb{R}\ \forall x \in X \quad |g'(x) + R_\epsilon - g(x)| \leq \epsilon. \tag{39}$$

Let $\epsilon > 0$. Let $L \in \mathbb{N}$ be such that (38) holds for every $n \geq m \geq L$. Define for $x \in X$

$$\begin{aligned} g_L(x) &:= \lim_{k\to\infty} d(p_{n_k}, x) - d(p_{n_k}, p_L) = g(x) - g(p_L) \\ g'_L(x) &:= \lim_{k\to\infty} d(p_{n'_k}, x) - d(p_{n'_k}, p_L) = g'(x) - g'(p_L). \end{aligned}$$

Then for $n_i \geq n'_j \geq L$, we get for all $x \in X$

$$\begin{aligned} d(p_{n_i}, x) - d(p_{n_i}, p_L) - d(p_{n'_j}, x) + d(p_{n'_j}, p_L) \\ \leq d(p_{n_i}, p_{n'_j}) - d(p_{n_i}, p_L) + d(p_{n'_j}, p_L) \leq \epsilon. \end{aligned}$$

By taking the limit $i \to \infty$ and $j \to \infty$, we obtain for all $x \in X$

$$g_L(x) - g'_L(x) \leq \epsilon.$$

By the symmetry of the argument, also $g'_L(x) - g_L(x) \leq \epsilon$ holds. Therefore for all $x \in X$

$$\epsilon \geq |g'_L(x) - g_L(x)| = |g'(x) - g(x) + g(p_L) - g'(p_L)|.$$

Setting $R_\epsilon = g(p_L) - g'(p_L)$, we conclude the proof of claim (39).

It is now easy to conclude $(i)$ from (39). Indeed, taking $x = o$, we have $|R_\epsilon| \leq \epsilon$, therefore for all $\epsilon > 0$ and for all $x \in X$ $|g(x) - g'(x)| \leq 2\epsilon$, i.e., $g = g'$. This completes the proof of $(i)$.

To prove assertion (ii), fix $\epsilon > 0$ and let $L \in \mathbb{N}$ be as above. Then we have for all $n \geq m \geq L$

$$\begin{aligned} 0 &\leq d(p_m, p_n) - d(p_n, o) + d(p_m, o) \\ &= d(p_m, p_n) + d(p_L, p_m) - d(p_L, p_n) + \\ &\qquad + d(p_L, p_n) - d(p_n, o) - d(p_L, p_m) + d(p_m, o) \\ &\leq \epsilon + d(p_L, p_n) - d(p_n, o) - d(p_L, p_m) + d(p_m, o). \end{aligned}$$

Taking first the limit $n \to \infty$ and then $m \to \infty$ in the above lines, we obtain the estimate

$$\begin{aligned} 0 &\leq \liminf_{m\to\infty} f(p_m) + d(p_m, o) \\ &\leq \limsup_{m\to\infty} f(p_m) + d(p_m, o) \leq \epsilon + f(p_L) - f(p_L) = \epsilon. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, then $(ii)$ holds true. □

Then, the proof of Corollary 1.5 runs similarly to Theorem 6.5 of [18].

*Proof of Corollary 1.5.* The horofunctions of type **(nv)** clearly are Busemann points, as they are limits, in particular, of the Riemannian geodesic rays which are the horizontal half-lines issued from the origin and which are always minimizing, see Proposition A.2 and Corollary A.3 in the Appendix. On the other hand, consider a horofunction of type **(v)**, $h_u = (v, z) = |u| - |u - v|$, for $u \in W$. Assume that there exists an almost straightly diverging sequence of points $p_n = v_n + z_n Z$ converging to $h_u$. By Lemma 5.2 (ii), we deduce that

$$\lim_{n\to\infty} f_u(p_n) + d_R(o, p_n) = \lim_{n\to\infty} |u| - |u - v_n| + d_R(o, p_n) = 0,$$

hence $\{v_n\}_{n\in\mathbb{N}}$ is necessarily an unbounded sequence. By Corollary 1.4 and the following description of horofunctions, it follows that $h_u$ should be of type **(nv)**, a contradiction. $\square$

5.0.1. *Concluding remarks.* The Riemannian Heisenberg group shows a number of counterintuitive features which is worth to stress:

(i) in view of Corollary 1.4, all Riemannian metrics on $\mathbb{H}$ with the same associated asymptotic metric have the same Busemann functions, though they are not necessarily isometric (in contrast, notice that all strictly subRiemannian metrics on $\mathbb{H}$ are isometric). However, this is not surprising, because all left-invariant Riemannian metrics on $\mathbb{H}$ are homothetic.

(ii) there exist diverging sequences of points $\{p_n\}_{n\in\mathbb{N}}$ that *visually converge* to a limit direction $v$ (that is, the minimizing geodesics $\gamma_n$ from $o$ to $p_n$ tend to a limit, minimizing geodesic $\gamma_v$ with initial direction $v$), but whose associated limit point $h_{\{p_n\}}$ is not given by the limit point $\gamma_v(+\infty)$ of $\gamma_v$. This happens for all vertically divergent sequences $\{p_n\}_{n\in\mathbb{N}}$, as the limit geodesic $\gamma_v$ is horizontal in this case (see Proposition A.2 and Corollary A.4 in the Appendix).

(iii) there exist diverging trajectories $\{p_n\}_{n\in\mathbb{N}}$, $\{q_n\}_{n\in\mathbb{N}}$ staying at bounded distance from each other, but defining different limit points (e.g., vertically diverging sequences of points with different limit horofunctions).

(iv) it is not true that, for a cocompact group of isometries $G$ of $(\mathbb{H}, d_R)$, the limit set of $G$ (which is the set of accumulation points of an orbit $Gx_0$ in $\partial(\mathbb{H}, d_R)$) equals the whole Gromov boundary; for instance, the discrete Heisenberg group $G = \mathbb{H}(\mathbb{Z})$, has a limit set equal to the set of all Busemann points, plus a discrete subset of the interior of the disk boundary $\bar{D}^2$. Also, the limit set may depend on the choice of the base point $x_0 \in \mathbb{H}$.

(v) The functions appearing in **(nv)** coincide with the Busemann functions of a Euclidean plane in the direction $R_\vartheta(\hat{v}_\infty)$; that is, the horofunction $h(v, z)$ associated to a diverging sequence $P_n = (v_n, z_n)$ of $(\mathbb{H}, d_R)$ is obtained just by dropping the vertical component $z$ of the argument, and then applying to $v$ the usual Euclidean Busemann function in the direction which is opposite to the limit direction of the $v_n$'s, rotated by an angle $\vartheta$ depending on the quadratic rate of divergence of the sequence ($\vartheta$ is zero for points diverging sub-quadratically, and $\vartheta = \pm\pi$ when the divergence is sup-quadratical).

These properties mark a remarkable difference to the theory of nonpositively curved, simply connected spaces.

## Appendix A. Length-minimizing curves for $d_{CC}$ and $d_R$

In the Heisenberg group, locally length-minimizing curves are smooth solutions of an Hamiltonian system both in the Riemannian and in the subRiemannian case. Locally length-minimizing curves are also called *geodesics*.

Let $d_R$ be a Riemannian metric on $\mathbb{H}$ with horizontal space $(\mathfrak{h}, g)$. Let $V \subset \mathfrak{h}$ be the plane orthogonal to $[\mathfrak{h}, \mathfrak{h}]$ and let $d_{CC}$ be the strictly subRiemannian metric on $\mathbb{H}$ with horizontal space $(V, g|_V)$.

Fix a basis $(X, Y, Z)$ for $\mathfrak{h}$ such that $(X, Y)$ is an orthonormal basis of $(V, g|_V)$ and $Z = [X, Y]$. Set $\zeta = \sqrt{g(Z, Z)}$. The basis $(X, Y, Z)$ induces the exponential coordinates $(x, y, z)$ on $\mathbb{H}$, i.e., $(x, y, z) = \exp(xX + yY + zZ)$. We will work in this coordinate system.

The Riemannian and subRiemannian length-minimizing curves are known and we recall their parametrization in the following two propositions. SubRiemannian geodesics can be found with different notation in [7]. Riemannian geodesics are found in [20], in different coordinates and with parametrization by arc-length.

**Proposition A.1** (subRiemannian geodesics)**.** *All the non-constant locally length-minimizing curves of $d_{CC}$ starting from $0$ and parametrized by arc-length are the following: given $k \in \mathbb{R} \setminus \{0\}$ and $\theta \in \mathbb{R}$*

(Type I) *The horizontal lines $t \mapsto (t\cos\theta, t\sin\theta, 0)$;*

(Type II) *The curves $t \mapsto (x(t), y(t), z(t))$ given by*

$$\begin{cases} x(t) = \dfrac{1}{k}\left(\cos\theta(\cos(kt) - 1) - \sin\theta\sin(kt)\right) \\ y(t) = \dfrac{1}{k}\left(\sin\theta(\cos(kt) - 1) + \cos\theta\sin(kt)\right) \\ z(t) = \dfrac{1}{2k}t - \dfrac{1}{2k^2}\sin(kt). \end{cases}$$

*Here the derivative at $t = 0$ is $(-\sin\theta, \cos\theta, 0)$.*

**Proposition A.2** (Riemannian geodesics)**.** *All non-constant locally length-minimizing curves of $d_R$ parametrized by a multiple of arc-length and starting from $0$ are the following: given $k \in \mathbb{R} \setminus \{0\}$ and $\theta \in \mathbb{R}$*

(Type 0) *The vertical line $t \mapsto (0, 0, t)$;*

(Type I) *The horizontal lines $t \mapsto (t\cos\theta, t\sin\theta, 0)$;*

(Type II) *The curves $t \mapsto (x(t), y(t), z(t))$ given by*

$$\begin{cases} x(t) = \dfrac{1}{k}\left(\cos\theta(\cos(kt) - 1) - \sin\theta\sin(kt)\right) \\ y(t) = \dfrac{1}{k}\left(\sin\theta(\cos(kt) - 1) + \cos\theta\sin(kt)\right) \\ z(t) = \dfrac{1}{2k}t - \dfrac{1}{2k^2}\sin(kt) + \dfrac{k}{\zeta^2}t. \end{cases}$$

*Here the derivative at $t = 0$ is $(-\sin\theta, \cos\theta, \frac{k}{\zeta^2})$, which has Riemannian length $\sqrt{1 + \frac{k^2}{\zeta^2}}$.*

The expression of geodesics helps us to prove the following facts.

**Corollary A.3.** *The horizontal lines of* Type I *are globally $d_R$- and $d_{CC}$-length-minimizing curves.*

**Corollary A.4.** *Both $d_R$- and locally $d_{CC}$-length-minimizing curves $\gamma$ of* Type II *are not minimizing from* $0$ *to $\gamma(t)$ if $|t| > \frac{2\pi}{k}$.*

*Proof.* This statement depends on the fact that in both cases, if we fix $k \in \mathbb{R} \setminus \{0\}$, then for all $\theta$ the corresponding length-minimizing curves $\gamma_{k,\theta}$ of Type II meet each other at the point $\gamma_{k,\theta}(2\pi/k)$. ☐

**Corollary A.5.** *The locally $d_R$-length-minimizing curve $\gamma$ of* Type 0*, $t \mapsto (0, 0, t)$, is not minimizing from* $0$ *to $\gamma(t)$ for $|t| > \frac{2\pi}{\zeta^2}$.*

*Proof.* For $k > 0$ let $\gamma_k$ be the $d_R$-length-minimizing curve of Type II with this $k$ and $\theta = 0$. Then $(\gamma_k)_3(\frac{2\pi}{k}) = \frac{\pi}{k^2} + \frac{2\pi}{\zeta^2}$. Letting $k \to \infty$ we obtain $\hat{z} := \frac{2\pi}{\zeta^2}$. This means that for every $\epsilon > 0$ there is $z \leq \hat{z} + \epsilon$ and $k > 0$ such that $\gamma_k(\frac{2\pi}{k}) = (0, 0, z)$. Therefore $t \mapsto (0, 0, t)$ cannot be minimizing after $z$, and therefore after $\hat{z}$. ☐

**Corollary A.6.** *If $p = (x, y, p_3)$ and $q = (x, y, q_3)$, then*

$$d_{CC}(p, q) = 2\sqrt{\pi} \cdot \sqrt{|p_3 - q_3|}.$$

*Proof.* First suppose $p = 0$: we have to prove that $d_{CC}(0, (0, 0, z)) = 2\sqrt{\pi}\sqrt{|z|}$. This is done by looking at the length-minimizing curves: they come from complete circle of perimeter $2\pi R = d$ and area $\pi R^2 = |z|$, so that we have $d_{CC}(0, (0, 0, z)) = 2\pi\sqrt{\frac{|z|}{\pi}} = 2\sqrt{\pi}\sqrt{|z|}$. The general case follows from the left-invariance of $d_{CC}$:

$$\begin{aligned} d_{CC}((x, y, p_3), (x, y, q_3)) &= d_{CC}(0, (x, y, p_3)^{-1}(x, y, q_3)) \\ &= d_{CC}(0, (0, 0, q_3 - p_3)). \end{aligned}$$

☐

**Corollary A.7.** *If $p \in \{z = 0\}$, then*

$$d_{CC}(0, p) = d_R(0, p)$$

**Corollary A.8.** *$d_R$- and $d_{CC}$-length-minimizing curves of* Type II *are in bijection via the following rule: If $\eta : [0, T] \to \mathbb{H}$ is a $d_{CC}$-length-minimizing curve of* Type II*, then*

$$\gamma(t) = \eta(t) + \left(0, 0, \frac{kt}{\zeta^2}\right)$$

*is $d_R$-length-minimizing of* Type II*, where $k \in \mathbb{R}$ is given by $\eta$. Moreover, it holds*

$$\|\omega_{\mathbb{H}}(\gamma')\|^2 = 1 + \frac{k^2}{\zeta^4}$$

*and*

$$d_{CC}(\gamma(t), \eta(t)) = 2\sqrt{\pi}\sqrt{\frac{kt}{\zeta^2}}$$

*Proof.* All the statements come directly from the expression of the geodesics. Notice that a $d_{CC}$-length-minimizing curve $\eta$ of TYPE II is parametrized by arc-length, i.e., $\|\omega_{\mathbb{H}}(\eta')\| \equiv 1$. On the other hand, the corresponding $d_R$-length-minimizing curve $\gamma$ has derivative $\omega_{\mathbb{H}}(\gamma') = \omega_{\mathbb{H}}(\eta') + \frac{k}{\zeta^2}Z$, where $\omega_{\mathbb{H}}(\eta')$ is orthogonal to $Z$. Hence $\|\omega_{\mathbb{H}}(\gamma')\|^2 = 1 + \frac{k^2}{\zeta^4}$ □

## References

[1] A. A. Agrachëv, D. Barilari, and U. Boscain. *Introduction to Riemannian and sub-Riemannian geometry.* 2015.

[2] W. Ballmann. *Lectures on spaces of nonpositive curvature.* Vol. 25. DMV Seminar. With an appendix by Misha Brin. Birkhäuser Verlag, Basel, 1995, pp. viii+112. ISBN: 3-7643-5242-6.

[3] E. Breuillard and E. Le Donne. "On the rate of convergence to the asymptotic cone for nilpotent groups and subFinsler geometry". In: *Proc. Natl. Acad. Sci. USA* 110.48 (2013), pp. 19220–19226. ISSN: 1091-6490.

[4] M. R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature.* Vol. 319. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1999, pp. xxii+643. ISBN: 3-540-64324-9.

[5] D. Burago. "Periodic metrics". In: *Seminar on Dynamical Systems (St. Petersburg, 1991).* Vol. 12. Progr. Nonlinear Differential Equations Appl. Birkhäuser, Basel, 1994, pp. 90–95.

[6] H. Busemann. *The geometry of geodesics.* Academic Press Inc., New York, N. Y., 1955, pp. x+422.

[7] L. Capogna et al. *An introduction to the Heisenberg group and the sub-Riemannian isoperimetric problem.* Vol. 259. Progress in Mathematics. Birkhäuser Verlag, Basel, 2007, pp. xvi+223. ISBN: 978-3-7643-8132-5; 3-7643-8132-9.

[8] F. Cerocchi and A. Sambusetti. "Quantitative Bounded Distance Theorem and Margulis' Lemma for Z^n actions with applications to homology". In: *ArXiv e-prints* (Dec. 2014). arXiv: 1412.6516.

[9] F. Dal'bo, M. Peigné, and A. Sambusetti. "On the horoboundary and the geometry of rays of negatively curved manifolds". In: *Pacific J. Math.* 259.1 (2012), pp. 55–100. ISSN: 0030-8730.

[10] M. Duchin and M. Shapiro. "Rational growth in the Heisenberg group". In: *ArXiv e-prints* (Nov. 2014). arXiv: 1411.4201.

[11] M. Gromov. "Hyperbolic groups". In: *Essays in group theory.* Vol. 8. Math. Sci. Res. Inst. Publ. Springer, New York, 1987, pp. 75–263.

[12] M. Gromov. "Hyperbolic manifolds, groups and actions". In: *Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference (State Univ. New York, Stony Brook, N.Y., 1978).* Vol. 97. Ann. of Math. Stud. Princeton Univ. Press, Princeton, N.J., 1981, pp. 183–213.

[13] M. Gromov. "Groups of polynomial growth and expanding maps". In: *Inst. Hautes Études Sci. Publ. Math.* 53 (1981), pp. 53–73. ISSN: 0073-8301.

[14] A. Haas. "Dirichlet points, Garnett points, and infinite ends of hyperbolic surfaces. I". In: *Ann. Acad. Sci. Fenn. Math.* 21.1 (1996), pp. 3–29. ISSN: 0066-1953.

[15] J. Hadamard. "Les surfaces à courbures opposées et leurs lignes géodésiques". In: *Jorn. de Math.* 4.5 (1898), pp. 27–74.

[16] F. Jean. *Control of nonholonomic systems: from sub-Riemannian geometry to motion planning.* Springer Briefs in Mathematics. Springer, Cham, 2014, pp. x+104. ISBN: 978-3-319-08689-7; 978-3-319-08690-3.

[17] T. Klein and A. Nicas. "The horofunction boundary of the Heisenberg group". In: *Pacific J. Math.* 242.2 (2009), pp. 299–310. ISSN: 0030-8730.

[18] T. Klein and A. Nicas. "The horofunction boundary of the Heisenberg group: the Carnot-Carathéodory metric". In: *Conform. Geom. Dyn.* 14 (2010), pp. 269–295. ISSN: 1088-4173.

[19] S. A. Krat. "Asymptotic properties of the Heisenberg group". In: *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* 261. Geom. i Topol. 4 (1999), pp. 125–154, 268. ISSN: 0373-2703.

[20] V. Marenich. "Geodesics in Heisenberg groups". In: *Geom. Dedicata* 66.2 (1997), pp. 175–185. ISSN: 0046-5755.

[21] J. Milnor. "A note on curvature and fundamental group". In: *J. Differential Geometry* 2 (1968), pp. 1–7. ISSN: 0022-040X.

[22] G. D. Mostow. *Strong rigidity of locally symmetric spaces.* Annals of Mathematics Studies, No. 78. Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1973, pp. v+195.

[23] P. Pansu. "Croissance des boules et des géodésiques fermées dans les nilvariétés". In: *Ergodic Theory Dynam. Systems* 3.3 (1983), pp. 415–445. ISSN: 0143-3857.

[24] S. J. Patterson. "The limit set of a Fuchsian group". In: *Acta Math.* 136.3-4 (1976), pp. 241–273. ISSN: 0001-5962.

[25] M. A. Rieffel. "Group $C^*$-algebras as compact quantum metric spaces". In: *Doc. Math.* 7 (2002), 605–651 (electronic). ISSN: 1431-0635.

[26] L. Rifford. *Sub-Riemannian geometry and optimal transport.* Springer Briefs in Mathematics. Springer, Cham, 2014, pp. viii+140. ISBN: 978-3-319-04803-1; 978-3-319-04804-8.

[27] T. Roblin. "Ergodicité et équidistribution en courbure négative". In: *Mém. Soc. Math. Fr. (N.S.)* 95 (2003), pp. vi+96. ISSN: 0249-633X.

[28] M. Stoll. "On the asymptotics of the growth of 2-step nilpotent groups". In: *J. London Math. Soc. (2)* 58.1 (1998), pp. 38–48. ISSN: 0024-6107.

[29] D. Sullivan. "The density at infinity of a discrete group of hyperbolic motions". In: *Inst. Hautes Études Sci. Publ. Math.* 50 (1979), pp. 171–202. ISSN: 0073-8301.

[30] J. Tits. "Free subgroups in linear groups". In: *J. Algebra* 20 (1972), pp. 250–270. ISSN: 0021-8693.

[31] C. Walsh. "The horofunction boundary of finite-dimensional normed spaces". In: *Math. Proc. Cambridge Philos. Soc.* 142.3 (2007), pp. 497–507. ISSN: 0305-0041.

[32] C. Walsh. "The horofunction boundary of the Hilbert geometry". In: *Adv. Geom.* 8.4 (2008), pp. 503–529. ISSN: 1615-715X.

[33] C. Webster and A. Winchester. "Busemann points of infinite graphs". In: *Trans. Amer. Math. Soc.* 358.9 (2006), 4209–4224 (electronic). ISSN: 0002-9947.

[34] J. A. Wolf. "Growth of finitely generated solvable groups and curvature of Riemanniann manifolds". In: *J. Differential Geometry* 2 (1968), pp. 421–446. ISSN: 0022-040X.

(Le Donne, Nicolussi Golo) University of Jyvaskyla, Finland

(Sambusetti) Sapienza Università di Roma, Italy

[C]

**From homogeneous metric spaces to Lie groups**

M. G. Cowling, V. Kivioja, E. Le Donne, S. Nicolussi Golo, and A. Ottazzi

ArXiv 1705.09648

# FROM HOMOGENEOUS METRIC SPACES TO LIE GROUPS

MICHAEL G. COWLING, VILLE KIVIOJA, ENRICO LE DONNE, SEBASTIANO NICOLUSSI GOLO, AND ALESSANDRO OTTAZZI

Abstract. We study connected, locally compact metric spaces with transitive isometry groups. For all $\varepsilon \in \mathbb{R}^+$, each such space is $(1,\varepsilon)$-quasi-isometric to a Lie group equipped with a left-invariant metric. Further, every metric Lie group is $(1,C)$-quasi-isometric to a solvable Lie group, and every simply connected metric Lie group is $(1,C)$-quasi-isometrically homeomorphic to a solvable-by-compact metric Lie group. While any contractible Lie group may be made isometric to a solvable group, only those that are solvable and of type (R) may be made isometric to a nilpotent Lie group, in which case the nilpotent group is the nilshadow of the group. Finally, we give a complete metric characterisation of metric Lie groups for which there exists an automorphic dilation. These coincide with the metric spaces that are locally compact, connected, homogeneous, and admit a metric dilation.

## Contents



*Date*: August 24, 2017.
2010 *Mathematics Subject Classification.* 53C30, 22F30, 20F69, 22E25.

## 1. Introduction

1.1. **Overview.** This paper presents some links between Lie theory and metric geometry. We study connected locally compact metric spaces with transitive isometry groups. Prototypical examples are Lie groups equipped with a left-invariant metrics. We assume that distance functions are compatible with the topology but *not* that they are Riemannian, or even geodesic. This permits us to consider a very broad setting including sub-Riemannian groups and their subgroups, as well as homogeneous groups in the sense of Folland and Stein. Nilpotent and solvable Lie groups play a special role in our analysis. We recall a number of developments that underlie our work.

First, in the 1960s, Harish-Chandra, Helgason, and many others developed the theory of semisimple Lie groups and Riemannian symmetric spaces; see [28] for an overview of the geometric aspects of this work. Then Milnor [43], Wolf [59], Gordon and Wilson [23, 24], and Wilson [58], amongst others, made important contributions to the theory of Riemannian Lie groups. In a parallel development, following Hörmander [29], in the 1970s Folland and Stein [21] and Rothschild and Stein [50] showed that nilpotent Lie groups are good model spaces for the study of subelliptic operators much as Euclidean space is a model for the study of elliptic operators. These operators lead naturally to distance functions on the group that are not Riemannian. They may be Carnot–Carathéodory distance functions, or be given by homogeneous norms, which in general are not geodesic. For more on analysis on nilpotent groups and on groups of polynomial growth, we refer to the monographs of Dungey, ter Elst and Robinson [18], of Corwin and Greenleaf [16] and of Goodman [22]. At about the same time, Gromov [25] combined abstract metric space ideas with group theory to prove his celebrated theorem that finitely generated groups of polynomial growth are virtually nilpotent. This is the discrete version of the Lie group theorem proved by Guivarc'h [26] and Jenkins [32]. Subsequently, Pansu shed light on the growth of nilpotent groups by showing that the asymptotic cones of nilpotent groups are Carnot groups [47]. Further, analysis on Carnot groups was used in complex geometry, to study CR manifolds by Korányi and Reimann [36], and to reprove some of Mostow's rigidity results by Pansu [48]. Finally, many authors, including Bellaïche [4], Hamenstädt [27], Montgomery [45], and Strichartz [53], developed the links between sub-Riemannian geometry on Lie groups and nonholonomic mechanics.

Today, the study of Lie groups equipped with general metrics is a thriving field of research, as evidenced by the work of Breuillard [7], Cornulier [17], Cornulier and de la Harpe [14], Cornulier and Tessera [15], Pauls [49], Stoll [52], and Tessera [56]. The connections with analysis on metric spaces and sub-Riemannian geometry became stronger after the work of Cheeger and Kleiner [10, 11] and Lee and Naor [42]. There are also developments in geometric measure theory on homogeneous groups; see, for example, Le Donne and Rigot [38, 39].

In this paper, we prove that Lie groups are models for connected locally compact homogeneous metric spaces up to quasi-isometry. More precisely, in Theorem 1.1, we show that for all $\varepsilon \in \mathbb{R}^+$, each such space is $(1,\varepsilon)$-quasi-isometric to a Lie group equipped with left-invariant metric, and hence that any homogenous metric space is $(1,C)$-quasi-isometric to a solvable Lie group. In Theorem 1.2, we prove that every simply connected metric Lie group is $(1,C)$-quasi-isometrically homeomorphic to a solvable-by-compact metric Lie group. We observe that any contractible Lie group may be made isometric to a solvable group (Remark 3.10). However, in Theorem 1.3, we see that only those that are solvable and of type (R) may be made isometric to a nilpotent Lie group, in which case the nilpotent group is the nilshadow of the group, in the sense of Auslander and Green [2]. Finally, we give a complete metric characterisation of those metric Lie groups that admit an automorphic dilation: according to Theorem 1.4, these coincide with the metric spaces that are connected, locally compact and homogeneous and admit a metric dilation.

1.2. **Statements of the results.** In this paper, *metric spaces are always assumed to be connected and locally compact, unless explicitly stated otherwise.* Some of our results may be proved in greater generality, but this assumption will save space. The main additional assumption is that the isometry group acts transitively, in which case we talk of a *homogeneous metric space*. The prototypical examples are connected locally compact groups with left-invariant metrics, such as Riemannian and sub-Riemannian Lie groups. Starting with these, one may obtain new examples by considering $\ell^p$ products, passing to subgroups, and composing the distance function with concave functions, as in the snowflake construction. We consider locally compact groups and Lie groups equipped with admissible left-invariant distance functions, which we call *metric groups* and *metric Lie groups*; by admissible we mean that the distance function induces the manifold topology. We stress that we do not restrict to quasigeodesic nor proper spaces.

Using the Gleason–Yamabe–Montgomery–Zippin structure theory of locally compact groups (see [55]), we reduce the study of homogeneous metric spaces to the study of metric Lie groups, up to quasi-isometry. Moreover, using the Levi decomposition and Iwasawa decompositions, we reduce further to the consideration of simply connected solvable groups. Before we state our main results, we state our convention on constants: these are always nonnegative real numbers, possibly with additional restrictions, and may vary from one occurrence to the next. These are often denoted by $C$, $L$, $Q$ or $\varepsilon$; we do not specify that these letters denote constants when they occur. As usual, we use $\varepsilon$ for a positive constant that may be chosen to be arbitrarily small.

**Theorem 1.1.** *Let $M$ be a homogeneous metric space. Then $M$ is*

*(a) $(1,\varepsilon)$-quasi-isometric to a connected metric Lie group, and*
*(b) $(1,C)$-quasi-isometric to a simply connected solvable metric Lie group.*

Part (a) of this theorem is related to the following result of Montgomery and Zippin [44, p. 243]: *a homogeneous space that is locally compact, locally connected and has finite topological dimension may be identified with a quotient of a Lie group by a compact subgroup.* Part (b) is known for geodesic distance functions; see for example [7, Proposition 1.3].

One of our aims is to study the following relation between metric groups. Given two topological groups $G$ and $H$, we say that $G$ *may be made isometric to* $H$ if there exist admissible left-invariant distance functions $d_G$ and $d_H$ such that the metric spaces $(G, d_G)$ and $(H, d_H)$ are isometric. Moreover, if $G$ is already a metric group, then we may impose the extra condition that the new distance function is $(1, C)$-quasi-isometric to the initial one; in this case, the Gromov–Hausdorff distance of the new metric space from the original one is bounded.

As a consequence of our next theorem, every simply connected Lie group may be made isometric to a direct product of a solvable and a compact Lie group.

**Theorem 1.2.** *Let $(G, d_G)$ be a simply connected metric Lie group. Then there are a solvable Lie group $S$, a compact Lie group $K$, and admissible left-invariant distance functions $d'_G$ and $d_{S\times K}$ such that*

*(i) the spaces $(G, d'_G)$ and $(S \times K, d_{S\times K})$ are isometric, and*
*(ii) the identity map on $G$ is a $(1, C)$-quasi-isometry from $d_G$ to $d'_G$.*

In this theorem, $S$ and $K$ are constructed explicitly: if $R$ is the radical of $G$, $L$ is a Levi subgroup of $G$, and $K_1AN$ is the Iwasawa decomposition of $L$, then we may decompose $K_1$ as $V \times K$, where $V$ is a vector group and $K$ is compact; we take $S$ to be $(R \rtimes AN) \times V$.

The theorem still holds if we assume that $R \cap L$ is trivial instead of assuming that $G$ is simply connected. This is the case if $G$ is semisimple; see Corollary 3.9 and Corollary 3.11.

The next step in our analysis is to consider metric Lie groups of polynomial volume growth. A compactly generated locally compact group $G$, with Haar measure $\mu$, is said to be *of polynomial growth* if there is a compact generating neighbourhood $U$ of the identity in $G$

$$\mu(U^n) \leq Cn^Q \qquad \forall n \in \mathbb{Z}^+. \tag{1.1}$$

We recall that a Lie group is of polynomial growth if and only if its Lie algebra is of type (R); see [26, 31].

It is known that groups of polynomial growth with quasigeodesic distance functions are quasi-isometric to nilpotent groups; see [7]. We generalise this to quasigeodesic homogeneous spaces in Corollary 4.17. It is not clear whether this generalisation holds for all admissible metrics.

We refine Theorem 1.2, and study when a Lie group may be made isometric to a nilpotent group. This question is tackled in Section 4.4; the main tools are the modifications of Gordon and Wilson [24].

**Theorem 1.3.** *Let $H$ and $N$ be connected simply connected Lie groups and assume that $N$ is nilpotent. The following are equivalent:*

*(i) $H$ may be made isometric to the nilpotent group $N$;*

*(ii) $H$ is a modification of $N$;*
*(iii) $H$ is solvable and of polynomial growth, and $N$ is its nilshadow.*

The nilshadow of a Lie group is uniquely defined up to isomorphism; see Section 4.3 for the definition following [18]. Hence for every solvable simply connected Lie group $G$ of polynomial growth, there exists exactly one nilpotent Lie group $N$, its nilshadow, with the property that $G$ and $N$ are isometric when these groups are appropriately metrised.

It is easy to construct groups that are not nilpotent but may be made isometric to nilpotent groups. For example, take a nilpotent group $N$ with a one-parameter isometry group of automorphisms, such as a Euclidean space, a generalised Heisenberg group, or a free nilpotent Lie group. Then $\mathbb{R}$ acts by isometries on $N$, and the direct product $N \times \mathbb{R}$ is a nilpotent group isometric to the semidirect product $N \rtimes \mathbb{R}$, which is not nilpotent. Moreover, $N \times \mathbb{R}$ is a Carnot group when $N$ is a Carnot group.

Further, if $H$ admits a quasigeodesic distance function $d$ making it isometric to $(N_1, d_1)$ and another quasigeodesic distance function $d'$ making it isometric to $(N_2, d_2)$, then necessarily $(N_1, d_1)$ and $(N_2, d_2)$ are quasi-isometric. However, the classification of nilpotent groups up to quasi-isometry is an important unsolved problem. Still, our theorem implies that $N_1$ and $N_2$ are isomorphic.

Parts of Theorem 1.3 were proved by Breuillard [7] and Gordon and Wilson [24], see also [12]; however our proof is different and more direct.

A map $\delta : X \to Y$ between metric spaces is called a *metric dilation* if $\delta$ is bijective and $d(\delta(x), \delta(x')) = \lambda d(x, x')$ for all $x, x' \in X$, for some $\lambda \in (1, \infty)$, and a *self-similar group* is a metric group $(G, d)$ that admits a map $\delta : G \to G$ that is both a metric dilation and an automorphism. Finite dimensional normed spaces and Carnot groups are self-similar groups; the homogeneous groups of Folland and Stein [21], equipped with Hebisch–Sikora distance functions [39], are more general examples.

**Theorem 1.4.** *If a metric space is locally compact, connected, isometrically homogeneous, and it admits a metric dilation, then it is isometric to self-similar Lie group. Moreover, all metric dilations of a self-similar Lie group are automorphisms.*

As a consequence of [51, Proposition 2.2] and [34], if a metric space $M$ is isometric to a self-similar Lie group $(G, d')$, then $G$ is a gradable, connected simply connected nilpotent Lie group isomorphic to the nilradical of $\mathrm{Iso}(M)$. However, $M$ may also be isometric to a Lie group that is not nilpotent. As discussed after Theorem 1.3, there are metric groups that are not nilpotent but which are isometric to self-similar metric Lie groups; it follows from Theorem 1.4 that if $M$ is a metric Lie group and $\delta$ is a metric dilation, then $\delta$ is an automorphism if and only if $M$ is nilpotent.

Theorem 1.4 generalises a result of [40], where it is shown that a space is a sub-Finsler Carnot group if and only if the conditions in Theorem 1.4 hold and moreover the distance function is geodesic.

The scheme of the proof of Theorem 1.4 is the following. We show that a metric space satisfying the hypotheses of the theorem is doubling. Then we show that its isometry group $G$ is a Lie group of polynomial growth, whence every Levi subgroup of $G$ is compact. However, the metric space is contractible, so the stabiliser $K$ of a point is a maximal compact subgroup containing a Levi subgroup. This allows us to find a subgroup $S$ of $G$ that is transverse to $K$: namely, the orthogonal complement of $K$ with respect to the Killing form. This subgroup $S$ induces the group structure on the metric space.

To link the doubling property of a metric space with the polynomial growth of its isometry group, we introduce a notion of polynomial growth for homogeneous metric spaces. Consider a Radon measure $m$ on a homogeneous metric space $M$ that is invariant under isometries, which exists and is unique up to a multiplicative constant. We say that $M$ is *of polynomial growth* if for one point, and hence for all points $o \in M$,

$$m(B(o,r)) \le Cr^Q \tag{1.2}$$

for all sufficiently large $r$. At this point, for a metric Lie group we have two notions of polynomial growth, which in general are not equivalent. For instance, $\mathbb{R}$ is a group of polynomial growth, but if we define the metric $d$ on $\mathbb{R}$ by

$$d(x,y) := \log(|x-y|+1) \qquad \forall x,y \in \mathbb{R},$$

then $(\mathbb{R}, d)$ is not of polynomial growth. Nonetheless, if a homogeneous metric space $M$ is of polynomial growth as in (1.2), then its isometry group $\mathrm{Iso}(M)$ is of polynomial growth in the sense of (1.1); see Lemma 2.21. In particular, a metric Lie group that is of polynomial growth as a metric space is also of polynomial growth as a group.

Let $(M, d)$ be a connected locally compact homogeneous metric space of polynomial growth. If $d$ is a quasigeodesic distance function, then $(M, d)$ is quasi-isometric to a simply connected nilpotent Riemannian Lie group; see Corollary 4.17. If, in addition, $M$ is contractible, then the quasi-isometry may be chosen to be a homeomorphism; see Corollary 4.16.

Polynomial growth is often linked with the property of being doubling at large scale. We observe that these two notions are not equivalent in our setting. More precisely, if a metric space $M$ is doubling at large scale, it may fail to be of polynomial growth; for instance, the space $\mathbb{R}$ with the distance function $d$ given by $d(x,y) = \min\{|x-y|, 1\}$ is trivially doubling at large scale, but is evidently not of polynomial growth. However, if $M$ is doubling at large scale and proper, then it is of polynomial growth; see Remark 2.18. Conversely, if $M$ is of polynomial growth, then it is proper, but it does not need to be doubling at large scale; see Remarks 2.19 and 2.20. Finally, if $M$ is proper and quasigeodesic, then it is of polynomial growth if and only if it is doubling at large scale; see, for instance, [13]. This paradoxical behaviour reflects the fact that polynomial growth and properness are not quasi-isometric invariants.

1.3. **Structure of the paper.** This paper is organised as follows. Section 2 contains several useful preliminary results. In particular, in Section 2.2, we consider homogeneous metric spaces, and in Section 2.3, we discuss contractibility in locally compact groups. In Section 2.4, we establish some Lie theory, and in Section 2.5, we deal with polynomial growth. While some of the results in Section 2 may be familiar to the expert, we decided to include proofs if we could not find an explicit proof in the literature or if we could give an easier one. In Section 3, we prove Theorems 1.1 and 1.2 and consider some of their consequences. Section 4 contains the proof of Theorem 1.3. In particular, we establish the preliminary results on modifications and nilshadows that are important for this proof in Sections 4.2 and 4.3. In Section 4.5, we prove a stronger version of Theorem 1.2 for homogeneous spaces of polynomial growth and quasigeodesic distance functions. In Section 5, we prove Theorem 1.4.

## 2. Preliminaries

In this section, we recall some more or less familiar facts. First, we discuss homogeneous metric spaces, then contractibility. Third, we bring in some Lie theory, and finally, we discuss polynomial growth.

2.1. **Notation.** If $(M, d)$ is a metric space, we sometimes write just $M$, leaving the metric $d$ implicit. We denote by $B(x, r)$ or $B_d(x, r)$ the open ball $\{y \in M : d(x, y) < r\}$, and by $\bar{B}(x, r)$ or $\bar{B}_d(x, r)$ the closed ball $\{y \in M : d(x, y) \leq r\}$.

A function $f : (M_1, d_1) \to (M_2, d_2)$ is an $(L, C)$-*quasi-isometry* if

$$L^{-1} d_1(x, y) - C \leq d_2(f(x), f(y)) \leq L d_1(x, y) + C$$

for all $x, y \in M_1$, and for every $z \in M_2$ there is $x \in M_1$ such that $d_2(f(x), z) \leq C$. If such a function exists between two metric spaces, then we say that they are $(L, C)$-*quasi-isometric*.

We denote by $e_G$, or more simply $e$, the identity element of a group $G$. We denote the Lie algebra of a Lie group $G$ by the corresponding fraktur letter $\mathfrak{g}$ or by $\mathrm{Lie}(G)$.

2.2. **Homogeneous metric spaces.** We define an *isometry* of a metric space $(M, d)$ to be a surjective map $f$ on $M$ such that

$$d(f(x), f(y)) = d(x, y) \qquad \forall x, y \in M.$$

We denote by $\mathrm{Iso}(M, d)$ the group of all isometries of $(M, d)$, where the group law is composition. A metric space $(M, d)$ is said to be *homogeneous* if its isometry group acts transitively.

We recall our convention that metric spaces are connected and locally compact unless explicitly stated. We prove that $\mathrm{Iso}(M, d)$ is a topological group (Lemma 2.1), that is metrisable (Proposition 2.7), locally compact and $\sigma$-compact (Proposition 2.11), and whose identity component acts transitively (Proposition 2.13) with compact stabiliser (Lemma 2.9). The main result of this section, Proposition 2.7, is that, for every $\varepsilon > 0$, $\mathrm{Iso}(M, d)$ may be metrised so that the identity component is $(1, \varepsilon)$-quasi-isometric to $(M, d)$.

**Lemma 2.1.** *Let $(M, d)$ be a metric space, not necessarily connected or locally compact. The group $\mathrm{Iso}(M, d)$, endowed with the topology of pointwise convergence, is a topological group, and its action on $M$ is a topological action.*

*Proof.* First, we show that the map $(f,g) \mapsto f^{-1} \circ g$ is continuous from $\mathrm{Iso}(M,d) \times \mathrm{Iso}(M,d)$ to $\mathrm{Iso}(M,d)$. Let $\{f_\nu\}_{\nu\in\mathrm{N}}$ and $\{g_\nu\}_{\nu\in\mathrm{N}}$ be nets in $\mathrm{Iso}(M,d)$ that converge to $f$ and $g$. For each $p \in M$,

$$\begin{aligned}
&d(f_\nu^{-1}(g_\nu(p)), f^{-1}(g(p))) \\
&\quad \le d(f_\nu^{-1}(g_\nu(p)), f_\nu^{-1}(g(p))) + d(f_\nu^{-1}(g(p)), f^{-1}(g(p))) \\
&\quad = d(g_\nu(p), g(p)) + d(g(p), f_\nu(f^{-1} \circ g(p))) \longrightarrow 0,
\end{aligned}$$

that is, the net $\{f_\nu^{-1} \circ g_\nu\}_{\nu\in\mathrm{N}}$ converges to $f^{-1} \circ g$.

Next, we show that the map $(f,p) \mapsto f(p)$ from $\mathrm{Iso}(M,d) \times M$ to $M$ is jointly continuous. Let $\{f_\nu\}_{\nu\in\mathrm{N}}$ and $\{p_\nu\}_{\nu\in\mathrm{N}}$ be nets in $\mathrm{Iso}(M,d)$ and $M$ that converge to $f$ and $p$. Then

$$\begin{aligned}
d(f_\nu(p_\nu), f(p)) &\le d(f_\nu(p_\nu), f_\nu(p)) + d(f_\nu(p), f(p)) \\
&= d(p_\nu, p) + d(f_\nu(p), f(p)) \to 0,
\end{aligned}$$

that is, $\{f_\nu(p_\nu)\}_{\nu\in\mathrm{N}}$ converges to $f(p)$. □

*Remark* 2.2. The topology of uniform convergence on compacta and the topology of pointwise convergence agree on $\mathrm{Iso}(M,d)$, since $\mathrm{Iso}(M,d)$ is an equicontinuous family of maps; see [33, p. 232].

To pass from local to global statements, we introduce the following notation. For $\ell \in \mathbb{R}^+$ and a subset $A$ of $M$, define the sets $V_n(A,\ell)$ by iteration on $n \in \mathbb{N}$: first, $V_0(A,\ell) := A$, and then

$$V_n(A,\ell) := \bigcup_{y\in V_{n-1}(A,\ell)} \bar{B}(y,\ell) \tag{2.1}$$

when $n \in \mathbb{Z}^+$. We usually write $V_n(p,\ell)$ rather than $V_n(\{p\},\ell)$.

**Lemma 2.3.** *Let $A$ be a nonempty subset of a homogeneous metric space $(M,d)$. Then $M = \bigcup_{n\in\mathbb{N}} V_n(A,\ell)$ for all $\ell \in \mathbb{R}^+$. If moreover $A$ is compact, then there exists $\ell \in \mathbb{R}^+$ such that $V_n(A,\ell)$ is compact for all $n \in \mathbb{N}$. Consequently, $(M,d)$ is $\sigma$-compact.*

*Proof.* It is easy to see that $\bigcup_{n\in\mathbb{N}} V_n(p,\ell)$ is a nonempty open and closed set in $M$, so it coincides with the connected set $M$.

Since $(M,d)$ is homogeneous, all closed balls with the same radius are homeomorphic. Take $\ell \in \mathbb{R}^+$ such that the closed balls of radius $2\ell$ are compact, and a nonempty compact subset $A$ of $M$. We prove by induction that $V_n(A,\ell)$ is compact for all $n \in \mathbb{N}$. By definition, $V_0(A,\ell)$ is compact. Further, if $V_n(A,\ell)$ is compact, then there are finitely many balls $B(x_i,\ell)$ such that $V_n(A,\ell) \subseteq \bigcup_i B(x_i,\ell)$; it follows that $V_{n+1}(A,\ell)$ is contained in the finite union of compact balls $\bigcup_i \bar{B}(x_i,2\ell)$, and hence is compact. □

**Lemma 2.4.** *Let $(M,d)$ be a homogeneous metric space. Then every distance-preserving map is surjective. Consequently, $\mathrm{Iso}(M,d)$ is closed in the space of all maps on $M$ equipped with the pointwise topology.*

*Proof.* Take a distance-preserving map $f$ on $M$; we must show that $f$ is surjective. By homogeneity, we may assume without loss of generality that $f$ fixes a point $o$. Take $\ell \in \mathbb{R}^+$ such that the sets $V_n(o,\ell)$ are compact, as in Lemma 2.3. Now $f$ is a distance-preserving map from $V_n(o,\ell)$ into $V_n(o,\ell)$

for all $n \in \mathbb{N}$. By [9, Theorem 1.6.14], a distance-preserving map from a compact metric space into itself is surjective. Hence $f(V_n(o,\ell)) = V_n(o,\ell)$ for all $n \in \mathbb{N}$. By Lemma 2.3, $f$ is surjective. Finally, if $f$ is the pointwise limit of a net of isometries, then $f$ is distance-preserving, and hence an isometry. Thus $\operatorname{Iso}(M,d)$ is closed in the space of all functions on $M$. □

The hypothesis of homogeneity is important: the set $\mathbb{R}^+$ with the metric $d$ given by $d(x,y) = |x-y|$ is not homogeneous, and the map $x \mapsto x+1$ is distance-preserving but not surjective.

To metrise the isometry group, we introduce more terminology.

**Definition 2.5.** Let $(M,d)$ be a homogeneous metric space and fix $o \in M$. A *Busemann gauge on* $(M,d)$ *with base point* $o$ is a function $\rho : M \to [0,+\infty)$ such that

(1) $\rho(o) = 0$ and $d(o,p) \le \rho(p)$ for all $p \in M$,
(2) a subset $A$ of $M$ is precompact if and only if $\sup_{p\in A} \rho(p) < \infty$.

*Remark* 2.6. Every homogeneous metric space admits a Busemann gauge, for instance,

$$\rho(p) \coloneqq \ell \min\{n \in \mathbb{N} : p \in V_n(o,\ell)\}, \tag{2.2}$$

where $\ell$ is such that $V_n(o,\ell)$ is compact for all $n \in \mathbb{N}$, as in Lemma 2.3. Indeed, define $\rho$ as in (2.2). Clearly $\rho(o) = 0$. If $\rho(p) = \ell n$, then there are points $p_0, p_1, \dots, p_n$ in $M$ such that $d(p_{i-1}, p_i) \le \ell$ for all $i = 1, \dots, n$ and $p_0 = o$ while $p_n = p$; hence $d(o,p) \le n\ell = \rho(p)$. Next, if $r \ge 0$, then $\{p : \rho(p) \le \ell r\} = V_{\lfloor r \rfloor}(o,\ell)$. Thus, if $\sup_{p\in A} \rho(p) \le \ell r$, then $A$ is precompact. Conversely, if $A \subseteq M$, then $\bar{A}$ may be covered by the sets $V_n(o,\ell)$ as $n$ increases; notice that the interior of $V_n(o,\ell)$ contains $V_{n-1}(o,\ell)$; if $A$ is precompact, then there exists $n$ such that $A \subseteq V_n(o,\ell)$.

**Proposition 2.7.** *Let* $(M,d)$ *be a homogeneous metric space and* $G$ *be a subgroup of* $\operatorname{Iso}(M,d)$ *that acts transitively on* $M$*. Take* $o \in M$ *and* $\varepsilon \in \mathbb{R}^+$*, and fix a Busemann gauge* $\rho$ *with base point* $o$*. Then the Busemann distance function* $d_G$ *on* $G$*, defined by*

$$d_G(g,h) \coloneqq \sup\{d(gp,hp)\mathrm{e}^{-\rho(p)/\varepsilon} \ : \ p \in M\},$$

*is an admissible left-invariant distance function on* $G$ *and the map* $\pi : g \mapsto g(o)$ *from* $(G,d_G)$ *to* $(M,d)$ *is 1-Lipschitz and a* $(1, 2\varepsilon/\mathrm{e})$*-quasi-isometry. In particular,* $\operatorname{Iso}(M,d)$ *is metrisable.*

*Proof.* Remark 2.6 exhibits an explicit Busemann gauge. The Busemann distance function $d_G$ is clearly left-invariant; we need to show that it is admissible. Let $\{g_\nu\}_{\nu\in\mathrm{N}}$ be a net in $G$.

On the one hand, if $g_\nu \to g$ in $(G,d_G)$, then

$$d(g_\nu(p), g(p)) \le \mathrm{e}^{\rho(p)/\varepsilon} d_G(g_\nu, g),$$

for all $p \in M$, and hence $g_\nu$ converges to $g$ pointwise, and so in $G$.

On the other hand, if $g_\nu \to g$ in $G$, then the convergence is uniform on compacta, by Remark 2.2. Fix $\eta \in (0,1)$. Then there is $R \in \mathbb{R}^+$ such that $t\mathrm{e}^{-t/\varepsilon} < \eta$ whenever $t > R$. Define $A$ to be the closure of $\{p \in M : \rho(p) \le R\}$. Then $A$ contains $o$ and is compact in $M$ by the definition of a Busemann gauge. Hence there is $\nu_0 \in \mathrm{N}$ such that $d(g_\nu(p), g(p)) \leq \eta$ for all $p \in A$ and all $\nu \geq \nu_0$. Therefore

$$d(g_\nu(p), g(p))\mathrm{e}^{-\rho(p)/\varepsilon} \leq \eta,$$

if $\nu \geq \nu_0$ and $p \in A$, while if $\nu \geq \nu_0$ and $p \notin A$, then

$$\begin{aligned} &d(g_\nu(p), g(p))\mathrm{e}^{-\rho(p)/\varepsilon} \\ &\quad\leq \left(d(g_\nu(p), g_\nu(o)) + d(g_\nu(o), g(o)) + d(g(o), g(p))\right)\mathrm{e}^{-\rho(p)/\varepsilon} \\ &\quad\leq \left(2d(o,p) + \eta\right)\mathrm{e}^{-\rho(p)/\varepsilon} \\ &\quad\leq \left(2\rho(p) + \eta\right)\mathrm{e}^{-\rho(p)/\varepsilon} \\ &\quad\leq 2\eta + \eta = 3\eta. \end{aligned}$$

We conclude that $d_G(g_\nu, g) \leq 3\eta$ for all $\nu \geq \nu_0$. As $\eta$ may be arbitrarily small, $g_\nu \to g$ in $(G, d_G)$.

By definition, $d(\pi(g), \pi(h)) = d(go, ho) \leq d_G(g, h)$ for all $g, h \in G$, so $\pi$ is 1-Lipschitz. Moreover, $\pi$ is surjective by assumption, and

$$\begin{aligned} d_G(g,h) &\leq \sup\{(d(gp, go) + d(go, ho) + d(ho, hp))\mathrm{e}^{-\rho(p)/\varepsilon} : p \in M\} \\ &\leq d(go, ho)\sup\{\mathrm{e}^{-\rho(p)/\varepsilon} : p \in M\} \\ &\qquad + 2\sup\{d(o,p)\mathrm{e}^{-\rho(p)/\varepsilon} : p \in M\} \\ &\leq d(\pi(g), \pi(h)) + 2\varepsilon/\mathrm{e} \end{aligned}$$

for all $g, h \in G$, whence $\pi$ is a $(1, 2\varepsilon/\mathrm{e})$-quasi-isometry. $\square$

**Lemma 2.8.** *Let $(M,d)$ be a homogeneous metric space and $G$ be a subgroup of $\mathrm{Iso}(M,d)$ that acts transitively on $M$. Take $\ell \in \mathbb{R}^+$ and $o \in M$, and set $U \coloneqq \{f \in G : f(o) \in \bar{B}(o,\ell)\}$. Then for all $n \in \mathbb{N}$,*

$$U^n = \{f \in G : f(o) \in V_n(o,\ell)\}. \tag{2.3}$$

*Proof.* If $n = 1$, then (2.3) holds by definition. Assume that (2.3) holds when $n = k$. On the one hand, if $f \in U^{k+1}$, then $f = gh$ where $g \in U^k$ and $h \in U$, so $f(o) \in g(\bar{B}(o,\ell)) = \bar{B}(g(o),\ell) \subseteq V_{k+1}(o,\ell)$. On the other hand, suppose that $f(o) \in V_{k+1}(o,\ell)$. Since $G$ acts transitively on $M$, there is $g \in G$ such that $g(o) \in V_k(o,\ell)$ and $f(o) \in \bar{B}(g(o),\ell)$. First, $g \in U^k$ by assumption. Second, $g^{-1}f(o) \in \bar{B}(o,\ell)$, that is, $g^{-1}f \in U$, since $\bar{B}(g(o),\ell) = g(\bar{B}(o,\ell))$. We conclude that $f \in U^{k+1}$. By induction, (2.3) holds for all $n$. $\square$

**Lemma 2.9.** *Let $(M,d)$ be a homogeneous metric space. If $A, B$ are compact subsets of $M$, then the set $U(A,B)$, given by*

$$U(A,B) \coloneqq \{f \in \mathrm{Iso}(M,d) : f(A) \subseteq B\},$$

*is compact. In particular, the stabiliser of a point is compact.*

*Proof.* Fix compacta $A, B$ in $M$. By Lemma 2.3, there is $\ell \in \mathbb{R}^+$ such that the sets $V_n(A,\ell)$ are compact. Note that $f(V_n(A,\ell)) = V_n(f(A),\ell)$ for all $f \in \mathrm{Iso}(M,d)$.

By Remark 2.2 and the Ascoli–Arzelà theorem (see [33, p. 233]), we need to show that

(a) $U(A,B)$ is closed in the space of continuous functions on $M$ in the topology of uniform convergence on compacta,

(b) $\{f(p) : f \in U(A,B)\}$ has compact closure for every $p \in M$,

(c) the family $U(A,B)$ is equicontinuous.

First, $U(A,B)$ is clearly closed in $\mathrm{Iso}(M,d)$, which is closed in the space of all continuous functions on $M$ by Lemma 2.4. Second, for all $p \in M$, the set $\{f(p) : f \in U(A,B)\}$ has compact closure in $M$: indeed, for each $p \in M$, there is $n \in \mathbb{N}$ such that $p \in V_n(A,\ell)$ and thus if $f \in U(A,B)$, then $f(p) \in f(V_n(A,\ell)) \subseteq V_n(B,\ell)$, that is, $\{f(p) : f \in U(A,B)\} \subseteq V_n(B,\ell)$. Finally, the family of isometries $U(A,B)$ is equicontinuous because $\mathrm{Iso}(M,d)$ is. By the Ascoli–Arzelà theorem, $U(A,B)$ is compact. □

*Remark* 2.10. If $M = \mathbb{Z}$ and $d(m,n) = 0$ if $m = n$ and 1 otherwise, then the metric space $(M,d)$ is locally compact and homogeneous but *not connected*, and the stabiliser of 0 is not compact. In this space, distance-preserving mappings need not be surjective.

**Proposition 2.11.** *Let $(M,d)$ be a homogeneous metric space. Then the group $\mathrm{Iso}(M,d)$ is locally compact, $\sigma$-compact and second countable. Hence if $G$ is a closed subgroup of $\mathrm{Iso}(M,d)$ that acts transitively on $M$ and $S$ is the stabiliser in $G$ of a point $o$ in $M$, then the map $gS \mapsto go$ is a homeomorphism from $G/S$ to $M$.*

*Proof.* Fix $\ell \in \mathbb{R}^+$ such that $\bar{B}(o,2\ell)$ is compact. Define

$$U := \{f \in \mathrm{Iso}(M,d) : f(o) \in \bar{B}(o,\ell)\};$$

then $U$ is a neighbourhood of the identity element in $\mathrm{Iso}(M,d)$. By Lemmas 2.8 and 2.9, the set $U^n$ is compact for all $n \in \mathbb{N}$, hence $\mathrm{Iso}(M,d)$ is locally compact and $\sigma$-compact. Since $\mathrm{Iso}(M,d)$ is also metrisable by Proposition 2.7, it is second countable. The last part of the proposition follows from [28, Theorem 3.2, p. 121]. □

**Lemma 2.12.** *Let $(M,d)$ be a homogeneous metric space and $G$ be a group of isometries of $(M,d)$. If there are $\ell \in \mathbb{R}^+$ and $o \in M$ such that $\bar{B}(o,\ell) \subseteq Go$, then $G$ acts transitively on $M$. In particular, every open subgroup of $\mathrm{Iso}(M,d)$ acts transitively.*

*Proof.* We show by induction on $n$ that $V_n(o,\ell) \subseteq Go$ for all $n \in \mathbb{N}$. If $n = 0$, then there is nothing to prove. Assume that $V_n(o,\ell) \subseteq Go$ and take $x \in V_{n+1}(o,\ell)$. Then there is $y \in V_n(o,\ell)$ such that $x \in \bar{B}(y,\ell)$. Since $V_n(o,\ell) \subseteq Go$, there is $g \in G$ such that $go = y$, hence $d(x,go) \leq \ell$, that is, $g^{-1}x \in \bar{B}(o,\ell)$. Since $\bar{B}(o,\ell) \subseteq Go$ by hypothesis, there is $f \in G$ such that $fo = g^{-1}x$ and thus $x = gfo \in Go$. This implies that $V_{n+1}(o,\ell) \subseteq Go$ and the inductive step is proved. It follows that $Go = M$ by Lemma 2.3.

Finally, suppose that $G$ is a open subgroup of $\mathrm{Iso}(M,d)$. By Proposition 2.11, the map $f \mapsto fo$ from $\mathrm{Iso}(M,d)$ to $M$ is open. Hence there is $\ell \in \mathbb{R}^+$ such that $\bar{B}(o,\ell)$ is compact and is a subset of $Go$. Therefore $G$ acts transitively on $M$, by the first part of the lemma. □

**Proposition 2.13.** *Let $(M,d)$ be a homogeneous metric space. The connected component $G$ of $\mathrm{Iso}(M,d)$ acts transitively on $M$.*

*Proof.* The totally disconnected locally compact group $\mathrm{Iso}(M,d)/G$ has a neighbourhood base N of the identity consisting of open and closed subgroups, ordered by reverse inclusion; see [54, Proposition 4.13]. For each

$\nu \in \mathrm{N}$, let $G_\nu$ be the preimage of $\nu$ in $\mathrm{Iso}(M,d)$. Then $\{G_\nu\}_{\nu\in\mathrm{N}}$ is a net of open and closed subgroups of $\mathrm{Iso}(M,d)$ such that $G = \bigcap_{\nu\in\mathrm{N}} G_\nu$, and $G_\nu$ acts transitively on $M$ for every $\nu \in \mathrm{N}$ by Lemma 2.12.

Take $o, p \in M$. For each $\nu \in \mathrm{N}$, there is $g_\nu \in G_\nu$ such that $g_\nu(o) = p$. By Lemma 2.9, $U(\{o\},\{p\})$ is compact; since $g_\nu \in U(\{o\},\{p\})$, we may assume that $g_\nu$ converges to $g \in U(\{o\},\{p\})$ by passing to a subnet if necessary. For each $\nu \in \mathrm{N}$, $g_{\nu'} \in G_\nu$ when $\nu' \geq \nu$, and hence $g \in G_\nu$. In conclusion, $g \in \bigcap_{\nu\in\mathrm{N}} G_\nu = G$ and $go = p$. □

2.3. **Contractibility.** We will need some information about maximal compact subgroups of locally compact groups. The following result is almost standard and may be extended (see [1]); compact contractibility is the only new ingredient. We say that a topological space $M$ is *compactly contractible* if, for each compact subset $S$ of $M$, there are $x \in M$ and a continuous map $F : [0,1] \times S \to M$ such that $F(0,s) = s$ and $F(1,s) = x$ for all $s \in S$.

**Lemma 2.14.** *If $K$ is a compact subgroup of a connected locally compact group $G$, then the following are equivalent:*

*(i) $K$ is a maximal compact subgroup of $G$;*
*(ii) $G/K$ is homeomorphic to a Euclidean space;*
*(iii) $G/K$ is contractible;*
*(iv) $G/K$ is compactly contractible.*

*Proof.* By [44, p. 188], (i) implies (ii). It is trivial that (ii) implies (iii) and (iii) implies (iv). We prove that (iv) implies (i) by modifying the argument of [1, Theorem 1.3].

Suppose that (iv) holds. By [3], there is a maximal compact subgroup $K_0$ of $G$ that contains $K$, and then by [44, p. 188], there is a map $\Phi : \mathbb{R}^n \to G$ such that the map $(x,y) \mapsto \Phi(x)y$ is a homeomorphism from $\mathbb{R}^n \times K_0$ to $G$. Hence $G/K$ is homeomorphic to $\mathbb{R}^n \times K_0/K$. The contraction of the compact set $K_0/K$ in $G/K$ composed with the projection onto $K_0/K$ is a contraction of $K_0/K$. From Antonyan [1], $K_0/K$ is contractible if and only if $K = K_0$, so $K$ is maximal. □

2.4. **Lie theory.** The main result of this section, Proposition 2.17, is an algebraic criterion for the existence of closed subgroups of the isometry group of a homogeneous metric space that act simply transitively.

Recall that if $G$ is a Lie group with Lie algebra $\mathfrak{g}$ and $\mathfrak{h}$ is a subalgebra of $\mathfrak{g}$, then there is a Lie subgroup $H$ of $G$ whose Lie algebra is $\mathfrak{h}$, but $H$ need not be closed. Moreover, if $H$ is a Lie subgroup of $G$, then $H$ with its own Lie structure is analytically immersed, but not necessarily embedded, in $G$. Recall also that if $H$ and $K$ are subgroups of a group $G$, then $HK$ denotes the subset $\{hk : h \in H, k \in K\}$ of $G$. The next lemma gives a criterion for $H$ to be closed.

**Lemma 2.15.** *Suppose that $K$ is a closed subgroup of a connected Lie group $G$ and denote by $\pi$ the quotient map from $G$ to $G/K$. Let $H$ be a Lie subgroup of $G$ such that $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{k}$ as vector spaces. Then*

*(i) $G = HK$,*
*(ii) the map $\pi|_H : H \to G/K$ is a covering map,*
*(iii) $H$ is closed in $G$ if and only if $H \cap K$ is discrete in $G$, and*

*(iv) if $G/K$ is simply connected, then $H \cap K = \{e\}$ and $H$ is closed.*

*Proof.* Denote by $M$ the quotient space $G/K$, which is a connected manifold, and by $o$ the point $K$ in $G/K$. The restriction to $H$ of the action of $G$ on $M$ is analytic. Since the map $\pi|_H : H \to M$ is smooth and its differential at $e_H$ is a linear isomorphism, there are an open neighbourhood $U$ of $e_H$ in $H$ and an open neighbourhood $V$ of $o$ in $M$ such that $\pi|_H : U \to V$ is a homeomorphism.

By introducing an auxiliary $G$-invariant metric on $M$ and by Lemma 2.12, we deduce that $Ho = M$; it follows immediately that $G = HK$. Indeed, if $g \in G$, then there is $h \in H$ such that $h^{-1}go = o$, that is, $h^{-1}g \in K$, and (i) is proved.

Since $H$ acts continuously on $M$, the stabiliser of $o$ in $H$ is closed. Since $\mathfrak{h} \cap \mathfrak{k} = \{0\}$, the intersection $H \cap K$ is discrete in $H$. Therefore, after shrinking the set $U$ that we produced above if necessary, we may assume that $Uk \cap Uk' = \emptyset$ when $k, k' \in H \cap K$ and $k \neq k'$. If $p \in M$, then $p = ho$ for some $h \in H$, so $\pi|_H^{-1}(hV)$ is equal to $\bigcup_{k \in H \cap K} hUk$, a disjoint union of open sets on each of which $\pi|_H$ is a homeomorphism onto $hV$. Thus $\pi|_H : H \to M$ is a covering map, which proves (ii).

If $H$ is closed, then $H \cap K$ is a closed zero-dimensional subgroup of $G$, and hence is discrete. Conversely, if $H \cap K$ is discrete, then there is an open subset $\Omega$ of $G$ such that $\Omega \cap H \cap K = \{e_G\}$. By shrinking the set $U$ produced above if necessary, we may assume that $U^{-1}U \subseteq \Omega$. The map $\varphi : (h, k) \mapsto hk$ from $U \times K$ into $G$ is trivially continuous; we claim that it is also injective. Indeed, assume that $h_1, h_2 \in U$ and $k_1, k_2 \in K$. If $h_1k_1 = h_2k_2$, then

$$h_2^{-1}h_1 = k_2k_1^{-1} \in U^{-1}U \cap H \cap K,$$

whence $h_1 = h_2$ and $k_1 = k_2$. Again by invariance of domain, $\varphi$ is a homeomorphism from $U \times K$ onto its image, and $U$ is closed in the open subset $UK$ of $G$. Hence $H$ is closed in $G$, and (iii) holds.

Finally, if $M$ is simply connected, then $\pi|_H$ is a homeomorphism, whence $H \cap K = \{e\}$. From part (iii), $H$ is closed in $G$. □

*Remark* 2.16. We recall an elementary fact that will be useful. If $G$ acts transitively on a set $M$, then the stabilizers of two points in $M$ are conjugated with each other. Hence, if a normal subgroup of $G$ is contained in one of the stabilizers, then it is contained in all stabilizers, i.e., it fixes all points. In particular, if $G$ acts faithfully and transitively on a set, then no normal subgroups of $G$ are contained in a stabilizer.

To state our next result, we introduce more notation. We denote by $\mathfrak{r}$ and $\mathfrak{n}$ the radical and nilradical of a Lie algebra $\mathfrak{g}$, and by $B$ its Killing form. We define the annihilator $\mathfrak{s}^B$ of a subspace $\mathfrak{s}$ of $\mathfrak{g}$ by

$$\mathfrak{s}^B = \{X \in \mathfrak{g} : B(X, \mathfrak{s}) = \{0\}\}.$$

The following result is close to and inspired by [24, Lemma 1.8].

**Proposition 2.17.** *Let $K$ be a compact subgroup of a connected Lie group $G$ with Lie algebra $\mathfrak{k}$, let $\mathfrak{h}$ be $\mathfrak{k}^B$, and let $H$ be the connected Lie subgroup of $G$ with Lie algebra $\mathfrak{h}$. Suppose that $K$ contains a Levi subgroup of $G$ and acts effectively on $G/K$. Then*

*(i)* $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{k}$ *as vector spaces, and*
*(ii)* $\mathfrak{n} \subseteq \mathfrak{h} \subseteq \mathfrak{r}$*, and* $\mathfrak{h}$ *is a solvable ideal of* $\mathfrak{g}$*.*

*Moreover, if $G/K$ is simply connected, then $H$ is closed, $G = HK$ and the map $h \mapsto hK$ from $H$ to $G/K$ is a diffeomorphism.*

*Proof.* First we show that $-B$ is positive definite on $\mathfrak{k}$. Since $K$ is compact, if $X \in \mathfrak{k}$, then $\operatorname{ad} X$ is semisimple and has eigenvalues $i\lambda_1, \ldots, i\lambda_n$, where each $\lambda_i \in \mathbb{R}$. Hence $B(X, X) = -(\lambda_1^2 + \cdots + \lambda_n^2)$, and so $B(X, X) = 0$ implies that $\operatorname{ad} X = 0$. As $G$ is connected, the one-parameter subgroup $\{\exp(tX) : t \in \mathbb{R}\} \subset K$ is central in $G$, whence $X = 0$ by Remark 2.16.

Since $K$ acts effectively on $G/K$, so does $\{\exp(tX) : t \in \mathbb{R}\}$, whence $X = 0$ by Lemma **??**.

It follows that $\mathfrak{h} \cap \mathfrak{k} = \{0\}$. Note that $\mathfrak{h}$ is the kernel of the map $X \mapsto B(X, \cdot)$ from $\mathfrak{g}$ to the dual $\mathfrak{k}^*$. Therefore

$$\dim(\mathfrak{g}) \leq \dim(\mathfrak{h}) + \dim(\mathfrak{k}) \leq \dim(\mathfrak{g}),$$

from which it follows that $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{k}$, and (i) holds.

We now prove (ii). Let $\mathfrak{l}$ be a Levi subgroup of $\mathfrak{g}$ contained in $\mathfrak{k}$. Since $\mathfrak{l} = [\mathfrak{l}, \mathfrak{l}] \subseteq [\mathfrak{g}, \mathfrak{g}]$, and $B(\mathfrak{r}, [\mathfrak{g}, \mathfrak{g}]) = \{0\}$ by [30, Theorem 5, Chapter III], it follows that $B(\mathfrak{r}, \mathfrak{l}) = \{0\}$. Now if $Z \in \mathfrak{h}$ and $Z = X + Y$, where $X \in \mathfrak{r}$ and $Y \in \mathfrak{l}$, then

$$0 = B(Z, Y) = B(X, Y) + B(Y, Y) = B(Y, Y),$$

so $Y = 0$ and $Z \in \mathfrak{r}$. Thus $\mathfrak{h} \subseteq \mathfrak{r}$.

If $X \in \mathfrak{n}$ and $Y \in \mathfrak{g}$, then $B(X, Y) = 0$; see [6, Chapter I, Section 4, Proposition 6]. It follows that $\mathfrak{n} \subseteq \mathfrak{h}$. Moreover, $\mathfrak{h}$ is an ideal since

$$[\mathfrak{g}, \mathfrak{h}] \subseteq [\mathfrak{g}, \mathfrak{r}] \subseteq \mathfrak{n} \subseteq \mathfrak{h}.$$

Finally, the last statement follows from Lemma 2.15. □

2.5. **Polynomial growth.** Let $G$ be a locally compact group, equipped with a left-invariant Haar measure $\mu$. If $K$ is a compact subgroup of $G$ and $\pi : G \to G/K$ is the quotient map, then there is a unique $G$-invariant Radon measure $m$ on $G/K$ such that

$$m(U) = \mu(\pi^{-1}(U)) \tag{2.4}$$

for all Borel subsets $U$ of $G/K$; see [20] or [46]. From Proposition 2.11, if $(M, d)$ is a homogeneous metric space and $G$ is the identity component of $\operatorname{Iso}(M, d)$, then $M$ may be identified with $G/K$ for some compact subgroup $K$ of $G$.

We now recall some standard terminology. First, a metric space $M$ is said to be *proper* if bounded sets are relatively compact, or equivalently, a subset is compact if and only if it is closed and bounded. Next, $M$ is said to be *doubling* if there is a constant $N$ such that each ball of radius $2r$ may be covered by at most $N$ balls of radius $r$ for all $r \in \mathbb{R}^+$. Finally, $M$ is $(L, C)$-*quasigeodesic* if for every $x, y \in M$ there are $n \in \mathbb{N}$ and $x_0, x_1, \ldots, x_n \in M$ such that $d(x_{j-1}, x_j) < C$ when $j \in \{1, \ldots, n\}$ and $\sum_{j=1}^n d(x_{j-1}, x_j) \leq Ld(x, y) + C$.

*Remark* 2.18. If a homogeneous metric space is proper and doubling, then it is of polynomial growth. Indeed, if every ball of radius $2r$ may be covered by $N$ balls of radius $r$, then one may check that

$$m(B(o,r)) \leq N m(B(o,1)) r^{\log_2(N)}$$

when $r > 1$.

*Remark* 2.19. A space of polynomial growth need not be doubling. The next example shows that having polynomial growth does not even imply being doubling at large scale.

Consider the piecewise linear function $D : [0,+\infty) \to [0,+\infty)$ with nodes at $(0,0)$, $(1,1)$, and $(x_n, y_n)$, where $n \in \mathbb{N}$, given by $x_n = 2^{2^{n+1}}$ and $y_n = 2^{2^n}$. The nodes all lie on the graph $y = x^{1/2}$, so $D$ is evidently increasing and concave. Hence $d(x,y) := D(|x-y|)$ is a translation-invariant metric on $\mathbb{R}$, and $|B(x_0,r)| = 2D^{-1}(r)$ for all $r \in [0,+\infty)$.

Take $r = y_n$, and consider the ratio

$$\frac{|B(0,2r)|}{|B(0,r)|} = \frac{D^{-1}(2y_n)}{D^{-1}(y_n)} = \frac{D^{-1}(2y_n)}{x_n} .$$

We will now show that the right hand fraction is unbounded in $n$, which shows that $d$ is not a doubling metric.

If $(x,y)$ lies on the line segment between $(x_n,y_n)$ and $(x_{n+1},y_{n+1})$, then

$$\frac{y-y_n}{x-x_n} = \frac{y_{n+1}-y_n}{x_{n+1}-x_n} = \frac{y_n^2-y_n}{y_n^4-y_n^2} = \frac{1}{y_n(y_n+1)},$$

so

$$x = x_n + y_n(y_n+1)(y-y_n).$$

Since $2y_n \leq y_{n+1}$, if $D(x) = 2y_n$, then $(x, 2y_n)$ lies on the line segment, and so $x = x_n + x_n(y_n+1)$ and

$$\frac{D^{-1}(2y_n)}{x_n} = \frac{x}{x_n} = y_n + 2,$$

which tends to infinity as $n$ increases.

The same argument also shows that if $(x,y)$ lies on this line segment, then

$$\begin{aligned} |B(0,y)| = 2x &= 2x_n + 2y_n(y_n+1)(y-y_n) \\ &\leq 2y_n^2 + 2y_n y(y_n+1) \leq 2y^2 + 2y^2(y+1), \end{aligned}$$

and it follows that $d$ is of polynomial growth.

*Remark* 2.20. If $(M,d)$ is a homogeneous metric space of polynomial growth, then it is proper. Indeed, if there were a noncompact closed ball $\bar{B}(p,r)$, then there would be $\varepsilon \in \mathbb{R}^+$ and points $x_i$ in $\bar{B}(p,r)$, where $i \in \mathbb{N}$, such that $d(x_i,x_j) > 2\varepsilon$ if $i \neq j$. But then it would follow that

$$C(r+\varepsilon)^Q \geq m(\bar{B}(p,r+\varepsilon)) \geq \sum_{i\in\mathbb{N}} m(B(x_i,\varepsilon)) = \infty,$$

which would be a contradiction.

A quasigeodesic homogeneous metric space is of polynomial growth if and only if its isometry group if of polynomial growth. For general metric spaces, the following implication may be proved.

**Lemma 2.21.** *If $M$ is a homogeneous metric space of polynomial growth, then* $\mathrm{Iso}(M)$ *and its identity component are of polynomial growth.*

*Proof.* Let $G$ be either $\mathrm{Iso}(M)$ or its identity component. By Lemma 2.3, we may fix $o \in M$ and $\ell \in \mathbb{R}^+$ such that the sets $V_n(o,\ell)$ are compact for all $n \in \mathbb{N}$. By Lemma 2.9, the set $U := \{f \in G : f(o) \in \bar{B}(o,\ell)\}$ is a compact neighbourhood of the identity element in $G$. By Lemma 2.8 and Proposition 2.13,

$$U^n = \{f \in G : f(o) \in V_n(o,\ell)\}.$$

Let $\mu$ be a Haar measure on $G$ and $m$ be an invariant measure on $M$ such that (2.4) holds, as discussed at the beginning this section, and suppose that $m(B(o,r)) \leq Cr^Q$ for all sufficiently large $r$. Then

$$\mu(U^n) = m(V_n(o,\ell)) \leq C\ell^Q(n+1)^Q$$

since $V_n(o,\ell) \subseteq B(o,(n+1)\ell)$. □

If $G$ is a connected Lie group, then it is of polynomial growth if and only if its Lie algebra $\mathfrak{g}$ is of type (R), that is, the eigenvalues of $\mathrm{ad}\, X$ are purely imaginary for each $X \in \mathfrak{g}$. For instance, nilpotent Lie groups are of polynomial growth. For more on this, see [32, 26].

**Lemma 2.22.** *Let $G$ be a connected Lie group of polynomial growth. Then each Levi subgroup of $G$ is compact. If moreover $G$ is simply connected, then $G = R \rtimes L$, where $R$ is the radical and $L$ is semisimple and compact. If $G$ is also contractible, then $G$ is solvable.*

*Proof.* Let $\tilde{G}$ be the universal cover of $G$. Since $G$ is of polynomial growth, the Lie algebra of $G$, and of $\tilde{G}$, is of type (R), hence $\tilde{G}$ is of polynomial growth. Since $\tilde{G}$ is connected and simply connected, $\tilde{G} = \tilde{R} \rtimes \tilde{L}$, where $\tilde{R}$ is the radical of $\tilde{G}$ and $\tilde{L}$ is a semisimple Lie subgroup of $\tilde{G}$; this is the Levi decomposition; see, for example, [57, Theorem 3.18.13]. Since $\tilde{G}$ is of polynomial growth, $\tilde{L}$ is compact; see, for example, [18, Theorem II.4.8].

Let $\pi : \tilde{G} \to G$ be the quotient projection, and write $R$ and $L$ for $\pi(\tilde{R})$ and $\pi(\tilde{L})$. The subgroup $R$ is the radical of $G$, and $L$ is a Levi subgroup of $G$, which is compact as $\tilde{L}$ is. Since all Levi subgroups of $G$ are conjugate to each other, all Levi subgroups of $G$ are compact.

If $G$ is contractible, then $G$ is the topological product of $R$ and $L$ and thus $L$ is contractible. A contractible compact Lie group is trivial, by Lemma 2.14, so $G$ coincides with $R$ and is solvable. □

## 3. From homogeneous spaces to solvable Lie groups

In this section, we first discuss some modifications of a metric space that do not change its quasi-isometry class. Next, we prove a key technical result; finally, we prove Theorems 1.1 and 1.2.

3.1. **From spaces to groups.** In this subsection, we first treat distance functions on quotients, in Lemma 3.1, and then we show how to enlarge isometry groups in Lemma 3.3. Finally, in Corollary 3.7, we use the solution to Hilbert's fifth problem to relate homogeneous metric spaces to Lie groups.

We will often deal with metric groups. Each element $g$ of a metric group $(M, d)$ is associated to a left translation map $L_g : p \mapsto gp$ and a right translation map $R_g : p \mapsto pg$; left translations are isometries of $M$ by definition, while right translations need not be.

**Lemma 3.1.** *Suppose that $K$ is a compact group of isometries of a metric space $(M, d)$ such that*

$$C := \sup\{d(kp, k'p) : k, k' \in K, p \in M\} < \infty,$$

*and define the function $d'$ on the orbit space $K\backslash M$ by*

$$d'(Kx, Ky) := \min\{d(fx, f'y) : f, f' \in K\} \qquad \forall x, y \in M.$$

*Then $d'$ is an admissible distance function, and*

$$d(x, y) - C \le d'(Kx, Ky) \le d(x, y) \qquad \forall x, y \in M, \tag{3.1}$$

*that is, the quotient map $\pi : p \mapsto Kp$ from $(M, d)$ to $(K\backslash M, d')$ is 1-Lipschitz and $(1, C)$-quasi-isometric.*

*Proof.* Since $d'(Kx, Ky) = \min\{d(x, ky) : k \in K\}$ for all $x, y \in M$, it is clear that $(K\backslash M, d')$ is a metric space. Moreover, if $x, y \in M$, then

$$d(x, y) \le d(x, ky) + d(ky, y) \le d(x, ky) + C$$

for all $k \in K$, from which the first inequality of (3.1) follows. The second inequality of (3.1) follows straight from the definition of $d'$.

Now we need to show that $d'$ is admissible, that is, that $d'$ induces the quotient topology on $K\backslash M$. We recall that a subset $U$ of $K\backslash M$ is open if and only if $\pi^{-1}(U)$ is open in $M$. On the one hand,

$$\pi^{-1}(B_{d'}(Kx, r)) = \bigcup_{y \in Kx} B_d(y, r);$$

this right hand side is clearly open in $M$ for all $x \in M$ and $r \in \mathbb{R}^+$; consequently, $B_{d'}(Kx, r)$ is open in $K\backslash M$. On the other hand, suppose that $U$ is an open subset of $K\backslash M$ and fix a point $x \in M$ such that $Kx \in U$. Define $\rho : K \to [0, +\infty]$ by

$$\rho(k) := \inf\{d(kx, y) : y \in M, \pi(y) \notin U\}.$$

The function $\rho$ is clearly lower semicontinuous and strictly positive. Since $K$ is compact, $r_0$, the minimum of $\rho$ on $K$, is strictly positive. Therefore $B_d(kx, r_0) \subseteq \pi^{-1}(U)$ for all $k \in K$ and $B_{d'}(Kx, r_0) \subseteq U$. We conclude that $U$ is open with respect to $d'$. □

**Corollary 3.2.** *Let $(M, d)$ be a metric group and $K$ be a compact normal subgroup of $M$. Then there is a distance function $d'$ on the quotient group $M/K$ such that the quotient map $\pi : x \mapsto xK$ from $(M, d)$ to $(M/K, d')$ is 1-Lipschitz and a $(1, \operatorname{diam}(K))$-quasi-isometry.*

*Proof.* Since $K$ is normal, left and right cosets coincide and $\operatorname{diam}(Kp) = \operatorname{diam}(K)$ for all $p \in M$. Lemma 3.1 may now be applied. □

**Lemma 3.3.** *Let $(M, d)$ be a locally compact metric space. Let $A$ be a group of homeomorphisms of $(M, d)$ that is compact in the topology of uniform* *convergence on compacta and that normalises a group $J$ of isometries of $M$. Define*

$$d_A(x, y) := \max\{d(ax, ay) : a \in A\} \qquad \forall x, y \in M.$$

*Then $d_A$ is a $JA$-invariant admissible metric on $M$, that is,*

$$d_A(gax, gay) = d_A(x, y) \qquad \forall x, y \in M, \quad \forall g \in J, \quad \forall a \in A.$$

*If all the maps in $A$ are $(L, C)$-quasi-isometries, then the identity map on $M$ is an $(L, C)$-quasi-isometry from $d$ to $d_A$.*

*Proof.* Since $A$ is compact and acts continuously on $M$, $d_A$ is finite-valued. Clearly $d_A$ is a metric and $d(x, y) \leq d_A(x, y)$ for all $x, y \in M$.

Now we show that $d$ and $d_A$ induce the same topology. Fix $x \in M$ and $\varepsilon \in \mathbb{R}^+$. On the one hand, if $d_A(x, y) < \varepsilon$, then $d(x, y) < \varepsilon$. On the other hand, by the Ascoli–Arzelà theorem, $A$ is an equicontinuous family of functions. Hence there is $\eta \in \mathbb{R}^+$ such that $d(kx, ky) < \varepsilon$ for all $k \in A$ and all $y \in B_d(x, \eta)$. Thus $d_A(x, y) < \varepsilon$ if $d(x, y) < \eta$.

If $x, y \in M$, $g \in J$ and $k \in A$, then

$$\begin{aligned} d_A(gkx, gky) &= \max\{d(k'gkx, k'gky) : k' \in A\} \\ &= \max\{d((k'gk'^{-1})k'kx, (k'gk'^{-1})k'ky) : k' \in A\} \\ &= \max\{d(k''x, k''y) : k'' \in A\} = d_A(x, y), \end{aligned}$$

since $A$ normalises $J$. Hence $d_A$ is $JA$-invariant.

The last statement is trivially true. □

**Corollary 3.4.** *Let $(M, d)$ be a locally compact metric group and $K$ be a subgroup of $M$; write $\alpha(k)$ for the inner automorphism $x \mapsto kxk^{-1}$ of $M$. Suppose that $\alpha(K)$ is a compact group of automorphisms of $M$. Then there is an $M$-left-invariant, $K$-right-invariant admissible distance function $d_K$ on $M$ such that the identity map from $(M, d)$ to $(M, d_K)$ is a $(1, C)$-quasi-isometry. If $K$ is itself compact, then there is an admissible distance function $d'_K$ on $M/K$ such that the quotient map $p \mapsto pK$ from $(M, d)$ to $(M/K, d'_K)$ is a $(1, C')$-quasi-isometry.*

*Proof.* Set $A := \alpha(K)$. By assumption, $A$ is a compact group of homeomorphisms of $M$ that normalises the group $J$ of left translations of $M$. Indeed, if $p, x \in M$ and $k \in K$, then

$$\alpha(k) \circ L_p \circ \alpha(k)^{-1}(x) = kp(k^{-1}xk)k^{-1} = (kpk^{-1})x.$$

Lemma 3.3 above constructs a $JA$-invariant admissible distance function $d_K$ on $M$. Since $R_k = L_k \circ \alpha(k^{-1})$ for all $k \in K$, the distance function $d_K$ is also $K$-right-invariant.

On the one hand, one easily shows that

$$d(\alpha(k)x, \alpha(k)y) \leq d(x, y) + 2d(e, k)$$

for all $k \in K$ and $x, y \in M$. On the other hand, since $\alpha(k)^{-1} = \alpha(k^{-1})$ and $d(e, k) = d(e, k^{-1})$,

$$d(x, y) = d(\alpha(k^{-1})\alpha(k)x, \alpha(k^{-1})\alpha(k)y) \leq d(\alpha(k)x, \alpha(k)y) + 2d(e, k)$$

for all $k \in K$ and $x, y \in M$. Since $\alpha$ is an open map onto $A$ and $A$ is compact, there is a constant $r$ such that $A = \alpha(B_d(e, r))$. Therefore $\alpha(k)$ is a $(1, 2r)$-quasi-isometry and by Lemma 3.3, the identity map from $(M, d)$

to $(M, d_K)$ is a $(1, 2r)$-quasi-isometry. If $K$ is compact, then one may take $r$ equal to $\mathrm{diam}_d(K)$.

Assume that $K$ is compact. Observe that $d_K(p, pk) = d_K(e, k)$ for all $p \in M$ and $k \in K$, so, with respect to $d_K$, the diameter of each orbit $pK$ is equal to the diameter of $K$. Therefore by Lemma 3.1, applied to the group of right translations by $K$, the composition of the identity map on $M$ with a change of metric from $d$ to $d_K$ and the quotient map form $(M, d_K)$ to $(M/K, d'_K)$ is a $(1, C)$-quasi-isometry, where $C = \mathrm{diam}_d(K) + \mathrm{diam}_{d_K}(K)$. □

The next lemma restates the solution to Hilbert's fifth problem by Gleason, Yamabe, Montgomery and Zippin. We quote [55].

**Lemma 3.5.** *Let $G$ be a locally compact group. There is an open subgroup $G'$ of $G$ with the property that every neighbourhood $U$ of the identity element of $G'$ contains a normal compact subgroup $K$ of $G'$ such that $G'/K$ is a Lie group.*

**Proposition 3.6.** *Let $(M, d)$ be a homogeneous metric space and $G$ be the connected component of the identity in $\mathrm{Iso}(M, d)$. For each $\varepsilon \in \mathbb{R}^+$, there is a compact normal subgroup $K_\varepsilon$ of $G$ such that $G/K_\varepsilon$ is a Lie group and the orbit space $K_\varepsilon\backslash M$ is an analytic manifold. Moreover, there is a distance function $d_\varepsilon$ on $K_\varepsilon\backslash M$ such that $G/K_\varepsilon$ acts transitively and effectively by isometries on $(K_\varepsilon\backslash M, d_\varepsilon)$. The quotient map from $(M, d)$ to $(K_\varepsilon\backslash M, d_\varepsilon)$ is 1-Lipschitz and a $(1, \varepsilon)$-quasi-isometry.*

*Proof.* We apply Lemma 3.5 to $G$, which is locally compact by Proposition 2.11. The open subgroup $G'$ of the lemma above coincides with $G$, because $G$ is connected. Fix $o \in M$ and $\varepsilon \in \mathbb{R}^+$, and let $B$ be a compact ball with center $o$ and radius less than $\varepsilon$. The set $U = \{f \in G : f(o) \in B\}$ is a neighbourhood of the identity element in $G$. By Lemma 3.5, there is a compact normal subgroup $K_\varepsilon$ of $G$, contained in $U$, such that $G/K_\varepsilon$ is a Lie group. Let $S$ be the stabiliser of $o$ in $G$. The stabiliser of $K_\varepsilon o$ in $G/K_\varepsilon$ is $(SK_\varepsilon)/K_\varepsilon$, which is a compact subgroup of the Lie group $G/K_\varepsilon$. Hence the orbit space $K_\varepsilon\backslash M$ is homeomorphic to $G/(SK_\varepsilon)$ and is an analytic manifold.

If $p \in M$ and $f \in K_\varepsilon$, then there are $g \in G$ with $g(o) = p$ and $f' \in K_\varepsilon$ such that $fg = gf'$. Thus

$$d(f(p), p) = d(fg(o), g(o)) = d(gf'(o), g(o)) \le \varepsilon, \tag{3.2}$$

that is, the diameter of $K_\varepsilon p$ is no greater than $\varepsilon$ for all $p \in M$.

The proposition now follows from Lemma 3.1. □

**Corollary 3.7.** *Let $(M, d)$ be a homogeneous metric space. For all $\varepsilon \in \mathbb{R}^+$, there is a connected metric Lie group $(G_\varepsilon, d_\varepsilon)$ that is $(1, \varepsilon)$-quasi-isometric to $(M, d)$.*

*Proof.* Let $G$ be the connected component of the identity in $\mathrm{Iso}(M, d)$. Proposition 3.6 guarantees the existence of a subgroup $K_\varepsilon$ of $G$ such that $G_\varepsilon := G/K_\varepsilon$ is a Lie group, $M_\varepsilon := K_\varepsilon\backslash M$ is an analytic manifold endowed with a distance function $d'_\varepsilon$ so that $G_\varepsilon$ acts transitively and effectively by isometries on $(M_\varepsilon, d'_\varepsilon)$, and the projection map from $(M, d)$ to $(M_\varepsilon, d'_\varepsilon)$ is a $(1, \varepsilon/2)$-quasi-isometry.

Now $G_\varepsilon$ acts transitively and effectively by isometries on $M_\varepsilon$, so from Proposition 2.7 we deduce that there is an admissible left-invariant distance function $d_\varepsilon$ on $G_\varepsilon$ such that the projection from $(G_\varepsilon, d_\varepsilon)$ to $(M_\varepsilon, d'_\varepsilon)$ is a $(1, \varepsilon/2)$-quasi-isometry.

Therefore $(M, d)$ is $(1, \varepsilon)$-quasi-isometric to $(G_\varepsilon, d_\varepsilon)$. □

3.2. **From groups to solvable groups.** The aim of this section is to prove Theorem 3.8. More precisely, given a connected Lie group $G$, we construct a solvable group that is a model space for $G$.

We will use several well-known facts about semisimple Lie groups, for which see [28] or [35]. Let $L$ be a connected semisimple Lie group, with Iwasawa decomposition $ANK$, where $A$, $N$ and $K$ are closed Lie subgroups, $A$ is a vector group, $N$ is nilpotent and simply connected, and the map $(a, n, k) \mapsto ank$ from $A \times N \times K$ to $G$ is a diffeomorphism. Then $AN$ is a solvable Lie subgroup of $L$, the center $Z(L)$ of $L$ is discrete and contained in $K$, and $K/Z(L)$ is compact. We denote by $Z_F$ the intersection of the kernels of all finite-dimensional representations of $L$. Then $Z_F < Z(L)$; further, $Z(L)/Z_F$ is finite and $K/Z_F$ is compact. We may further decompose $K$ as $V \times K_0$, where $V$ is a closed vector subgroup of $K$ and $K_0$ is a maximal compact subgroup of $L$. The center $Z(L)$ is thus the direct product $Z_V \times Z_0$, where $Z_V$ is a lattice in $V$ and $Z_0$ is a finite subgroup of $K_0$.

**Theorem 3.8.** *If $G$ is a connected Lie group, then there exists a connected Lie group $H$ with the following properties.*

*(1) $H = H_0 \times K_0$, where $H_0$ is solvable and $K_0$ is compact.*
*(2) $H$ acts analytically and transitively on $G$, with finite stabiliser. In particular, the analytic map $h \mapsto h \,.\, e_G$ from $H$ to $G$ is a finite covering map, whose degree is bounded by the cardinality of $R \cap L$, where $R$ is the radical of $G$ and $L$ is a Levi subgroup.*
*(3) There is a connected solvable subgroup $S$ of $H_0$ whose action on $G$ is simple and cocompact.*
*(4) If $d$ is an admissible left-invariant distance function on $G$, then there exists an admissible left-invariant distance function $d_G$ on $G$ such that the action of $H$ on $(G, d_G)$ is by isometries and the identity map $(G, d) \to (G, d_G)$ is a $(1, C)$-quasi-isometry.*
*(5) There is an admissible left-invariant distance function $d_S$ on $S$ such that the map $s \mapsto s \,.\, e_G$ is an isometric embedding of $(S, d_S)$ into $(G, d_G)$ and a $(1, C)$-quasi-isometry.*

*Proof.* Let $G = RL$ be a Levi decomposition of $G$, where $R$ is the radical of $G$, and $L$ is a connected semisimple Lie subgroup. Using the notation introduced at the beginning of this section, we fix an Iwasawa decomposition $ANK$ of $L$, and we further decompose $K$ as $V \times K_0$.

Define $\Gamma$ to be $R \cap L$. Note that $L$ does not need to be closed in $G$, but since $R$ is normal and closed, $\Gamma$ is a closed normal zero-dimensional subgroup of $L$, so it is central and discrete in $L$.

Note that $Z_F \subseteq Z(L) \cap Z(G)$, where $Z(G)$ is the center of $G$. Write $\Delta$ for $\Gamma \cap Z_F \cap Z_V$. We claim that $\Gamma/\Delta$ is finite. Indeed, algebraically,

$$\Gamma/\Delta \simeq \Gamma(Z_F \cap Z_V)/(Z_F \cap Z_V) < Z(L)/(Z_F \cap Z_V).$$

Second, since $Z_F \cap Z_V$ is the kernel of the restriction to $Z_F$ of the projection from $Z_V \times Z_0$ onto the second factor $Z_0$ and $Z_0$ is finite, $Z_F/(Z_F \cap Z_V)$ is finite. Third, $Z(L)/Z_F$ is finite, and

$$Z(L)/Z_F \simeq \big(Z(L)/(Z_F \cap Z_V)\big) \; / \; \big(Z_F/(Z_F \cap Z_V)\big).$$

Therefore $Z(L)/(Z_F \cap Z_V)$ is finite and thus $\Gamma/\Delta$ is finite too, and the claim is proved.

Define $H'$ to be $(R \rtimes AN) \times K$ and $\Psi : H' \times G \to G$ by

$$\Psi\big((x, y), g\big) := L_x \circ R_{y^{-1}}(g) = xgy^{-1},$$

for all $(x, y) \in H'$ and all $g \in G$. The analytic map $\Psi$ defines a left action of $H'$ on $G$. We write $(x, y) \,.\, g$ for $\Psi\big((x, y), g\big)$.

The action is transitive, because if $g \in G$ then there are $r \in R$, $a \in A$, $n \in N$, and $k \in K$ such that $rank^{-1} = g$, that is, $g = (ran, k) \,.\, e_G$. Consequently, all stabilisers are conjugate to the stabiliser $\mathrm{Stab}_{H'}(e_G)$ of $e_G$ in $H'$, which is $\{(x, x) : x \in \Gamma\}$. The kernel $\ker \Psi$ of the action is $\{(x, x) : x \in Z(G) \cap \Gamma\}$. Indeed, on the one hand, if $x \in Z(G) \cap \Gamma$ then $(x, x) \,.\, g = xgx^{-1} = g$ for all $g \in G$. On the other hand, if $(x, y) \,.\, g = g$ for all $g \in G$, then $xy^{-1} = e$, that is, $x = y \in \Gamma$, and $x \in Z(G)$.

Define $\tilde{\Delta}$ to be $\{(x, x) : x \in \Delta\}$, and the groups $H_0$ and $H$ by

$$H_0 := ((R \rtimes AN) \times V)/\tilde{\Delta} \quad \text{and} \quad H := H_0 \times K_0.$$

Note that $H$ is equal to $H'/\tilde{\Delta}$, since $\Delta \cap K_0 = \{e\}$. Now $\Delta \subseteq Z(G)$, so $\tilde{\Delta}$ is a central subgroup of $(R \rtimes AN) \times V$ and therefore $H_0$ is a solvable Lie group. Since $\tilde{\Delta}$ is contained in the kernel of the action of $H'$, the group $H$ still acts transitively on $G$. Moreover, $\mathrm{Stab}_H(e_G)$, the stabiliser of $e_G$ in $H$, is $\mathrm{Stab}_{H'}(e_G)/\tilde{\Delta}$, which is isomorphic to the finite group $\Gamma/\Delta$.

Parts (1) and (2) of the theorem are proved. Now we will prove (3). Let $V_2$ be the linear span in $V$ of the set $\Gamma \cap V$. We claim that

$$\Gamma \subseteq V_2 \times K_0. \tag{3.3}$$

Indeed, $Z(L) = Z_V \times Z_0$, where $Z_0$ is a finite subgroup of $K_0$ and $Z_V$ is a subgroup of $V$, so if $\gamma \in \Gamma$ is written as $(z_V, z_0) \in Z_V \times Z_0$, then $\gamma^n = (z_V^n, e)$, where $n$ is the order of $Z_0$. Hence $z_V \in V_2$ and (3.3) is proved.

Let $V_1$ be a subspace of $V$ complementary to $V_2$, so $V = V_1 \times V_2$, and take $S$ to be $(R \rtimes AN) \times V_1$, which is a connected solvable subgroup of $H'$, and so acts on $G$. Now $S \cap \mathrm{Stab}_{H'}(e_G) = \{e\}$ since $\Gamma \subseteq V_2 \times K_0$, and thus $S$ acts simply on $G$ and is a subgroup of $H$.

In general, $S$ does not act transitively on $G$; however, the orbit space $S\backslash G$ is compact. Indeed, topologically,

$$S\backslash G \simeq S\backslash \big(H'/\, \mathrm{Stab}_{H'}(e_G)\big) \simeq \big(S\backslash H'\big) \,/\, \mathrm{Stab}_{H'}(e_G) \simeq (V_2 \times K_0)/\Gamma.$$

Moreover, $(V_2 \times K_0)/\Gamma$ is compact. Indeed, write $\Gamma_0$ for $\Gamma \cap K_0$ and $\Gamma_2$ for $\Gamma \cap V_2$; then $\Gamma_2 \times \Gamma_0 \subseteq \Gamma$ and thus

$$\begin{aligned}(V_2 \times K_0)/\Gamma &\simeq \big((V_2 \times K_0)/(\Gamma_2 \times \Gamma_0)\big) \; / \; \big(\Gamma/(\Gamma_2 \times \Gamma_0)\big) \\ &\simeq \big((V_2/\Gamma_2) \times (K_0/\Gamma_0)\big) \; / \; \big(\Gamma/(\Gamma_2 \times \Gamma_0)\big),\end{aligned}$$

where $(V_2/\Gamma_2) \times (K_0/\Gamma_0)$ is compact. This completes the proof of (3).

To prove (4), we define the analytic map $\psi : s \mapsto s\,.\,e_G$ from $S$ to $G$, that is,

$$\psi(ran, v_1) = ranv_1^{-1}.$$

Since the action of $S$ is simple, the map $\psi$ is injective. We prove that $\psi$ is a topological embedding, that is, that the inverse $\psi^{-1}$ is continuous from $\psi(S)$ to $S$. Let $p_i = (r_i a_i n_i, v_i) \in S$ be a sequence such that $\lim_{i\to\infty} \psi(p_i) = \psi(p)$ for some $p = (ran, v) \in S$. We need to show that $p_i \to p$. Consider the quotient $R\backslash G$, which is a connected semisimple Lie group isomorphic to $L/(L \cap R) = (ANV_1) \cdot ((V_2K_0)/\Gamma)$. Consider also the quotient map $\tilde{\pi} : G \to R\backslash G$, and the standard isomorphism $\tau : R\backslash G \to L/(L \cap R)$. Then $\lim_{i\to\infty} \tau \circ \tilde{\pi} \circ \psi(p_i) = \tau \circ \tilde{\pi} \circ \psi(p)$ by continuity. Since $\tau \circ \tilde{\pi} \circ \psi(p_i) = a_i n_i v_i^{-1}$ and $ANV_1$ is the topological product of $A$, $N$ and $V_1$, it follows that $\lim_{i\to\infty} a_i = a$, $\lim_{i\to\infty} n_i = n$ and $\lim_{i\to\infty} v_i = v$. Therefore $\lim_{i\to\infty} r_i = \lim_{i\to\infty} r_i a_i n_i v_i^{-1}(anv^{-1})^{-1} = r$, and we conclude that $p_i \to p$.

Suppose that $d$ is an admissible left-invariant distance function on $G$. Note that $K$ need not be compact (for instance, if $L$ is the universal covering group of $\mathrm{SL}(2,\mathbb{R})$); however, it is still true that $K/Z(G)$ is compact, because $Z_F \subseteq Z(G)$. By Corollary 3.4, there is an admissible distance function $d_G$ on $G$ that is $G$-left-invariant and $K$-right-invariant and such that the identity map on $G$ is a $(1, C)$-quasi-isometry from $d$ to $d_G$. Therefore $H'$ acts by isometries on $(G, d_G)$, and thus both $H$ and $S$ also act by isometries. This proves (4).

Define $d_S$ on $S$ by

$$d_S(p, q) := d_G(\psi(p), \psi(q)).$$

Since $\psi : S \to \psi(S)$ is a homeomorphism, $d_S$ is an admissible distance function on $S$. Further, $d_S$ is left-invariant on $S$. Indeed, if $\bar{p}, p, p' \in S$, then

$$\begin{aligned} d_S(\bar{p}p, \bar{p}p') &= d_G((\bar{p}p).e_G, (\bar{p}p').e_G) \\ &= d_G(\bar{p}.(p.e_G), \bar{p}.(p'.e_G)) \\ &= d_G(p.e_G, p'.e_G) \\ &= d_S(p, p'). \end{aligned}$$

Finally, we show that $\psi : (S, d_S) \to (G, d_G)$ is a $(1, C)$-quasi-isometry or, equivalently, a $C$-neighbourhood of $\psi(S)$ with respect to $d_G$ covers $G$. If $\{U_n\}_{n\in\mathbb{N}}$ is a nested sequence of precompact open sets in $G$ such that $e_G \in U_n$ for all $n$ and $G = \bigcup_n U_n$, then $\pi(U_k) = S\backslash G$ for some $k$ because $S\backslash G$ is compact. Set

$$C := \max\{d_G(e_G, x), x \in \bar{U}_k\},$$

and observe that if $y \in G$, then there are $x \in U_k$ and $\bar{s} \in S$ such that $\varphi(\bar{s})x = y$, whence

$$\begin{aligned} d_G(y, \psi(S)) &= \inf\{d_G(y, \varphi(s)e_G) : s \in S\} \\ &\leq d_G(\varphi(\bar{s})x, \varphi(\bar{s})e_G) = d_G(x, e_G) \leq C. \end{aligned}$$

The proof of (5) is now complete. □

3.3. **Proof of Theorem 1.1.** The first part of Theorem 1.1 is the content of Corollary 3.7. We recall here the statement of the second part for the reader's convenience.

**Theorem.** *If $(M, d)$ is a homogeneous metric space, then it is (1, C)-quasi-isometric to a simply connected solvable metric Lie group.*

*Proof.* By part A of Theorem 1.1, $(M, d)$ is $(1, C)$-quasi-isometric to a metric Lie group $(G, d_G)$. By Theorem 3.8, there is a connected solvable metric Lie group $(H, d_H)$ that is $(1, C)$-quasi-isometric to $(G, d_G)$. We will prove that there is a simply connected solvable metric Lie group $(J, d_J)$ that is $(1, C)$-quasi-isometric to $(H, d_H)$.

Let $K$ be a maximal compact subgroup of $H$. We may assume that $K$ acts effectively on $H/K$, by taking the quotient of $H$ by the kernel $K'$ of the action of $H$ on $H/K$ otherwise. Indeed, $K'$ is a compact normal subgroup of $H$, and Corollary 3.2 applies.

Note that the Levi subgroup of $H$ is trivial because $H$ is solvable. The quotient space $H/K$ is simply connected by Lemma 2.14. Now we may apply Proposition 2.17 to obtain a simply connected closed normal solvable subgroup $J$ of $H$ such that the restricted quotient map from $J$ to $H/K$ is a homeomorphism. Moreover, $H/J \simeq K$ is compact. Therefore, $(J, d_H)$ is a metric Lie group (1, C)-quasi-isometric to $(H, d_H)$.

Finally, $(M, d)$ is (1, C)-quasi-isometric to the simply connected solvable metric Lie group $(J, d_J)$. □

3.4. **Proof of Theorem 1.2.** Let $G$ be a connected Lie group with radical $R$ and Levi subgroup $L$. Let $H$ be the group constructed in Theorem 3.8. If $R \cap L = \{e\}$, then the stabiliser of $e_G$ in $H$ is trivial. There are two simple cases in which this happens: if $G$ is simply connected, because then $G = R \rtimes L$, and if $L$ is semisimple, because then $R = \{e\}$. If the stabiliser of $e_G$ in $H$ is trivial, then the covering map $h \mapsto h.e_G$ described in Theorem 3.8.(2) is a homeomorphism and we can pull back from $G$ to $H$ the distance $d_G$ given by Theorem 3.8.(4). We denote by $d_H$ the new distance on $H$. Since the action of $H$ on $(G, d_G)$ is by isometry, then $d_H$ is left-invariant. Indeed, if $h, h_1, h_2 \in H$, then

$$d_H(hh_1, hh_2) = d_G(hh_1.e_G, hh_2.e_G) = d_G(h_1.e_G, h_2.e_G) = d_H(h_1, h_2).$$

Therefore we obtain the following results, which contain a restatement of Theorem 1.2.

**Corollary 3.9.** *Suppose that $(G, d)$ is either a simply connected metric Lie group or a connected semisimple metric Lie group. Then there exist a connected Lie group $H$ that is the product of a solvable and a compact Lie group, and admissible left-invariant distance functions $d_G$ and $d_H$ such that $(G, d_G)$ and $(H, d_H)$ are isometric and the identity map on $G$ is a $(1, C)$-quasi-isometry from $d$ to $d_G$.*

*Remark* 3.10. If $G$ is a contractible Lie group, then the group $H$ given by Corollary 3.9 has no nontrivial compact subgroup, by Lemma 2.14, whence $H$ is solvable and $G$ may be made isometric to a solvable Lie group.

**Corollary 3.11.** *Let $G$ be a connected semisimple Lie group with Iwasawa decomposition $ANK$. Write $K$ as $V \times K'$, where $V$ is a vector group and $K'$ is compact. Then $G$ may be made isometric to the direct product $AN \times V \times K'$.*

## 4. Nilpotent groups and polynomial growth

The aim of this section is to prove Theorem 1.3 and discuss when a homogeneous metric space is quasi-isometric to a simply connected nilpotent Lie group. We first recall some definitions and results, on modifications of nilpotent Lie algebras and groups in Section 4.2, and on nilshadows of solvable Lie groups in Section 4.3. These notions are then used to prove Theorem 1.3 in Section 4.4. In Section 4.5, we deduce that quasigeodesic homogeneous spaces of polynomial growth are quasi-isometric to nilpotent groups.

4.1. **Notation.** We write $\mathrm{Aut}(G)$ for the group of automorphisms of a Lie group $G$, and $\mathrm{Aff}(G)$ for the Lie group of affine transformations of $G$, which may be identified with $G \rtimes \mathrm{Aut}(G)$. Given a Lie algebra $\mathfrak{g}$, we write $\mathrm{nil}(\mathfrak{g})$ for the nilradical of $\mathfrak{g}$, and $\mathrm{der}(\mathfrak{g})$ and $\mathrm{Aut}(\mathfrak{g})$ for the Lie algebra of derivations of $\mathfrak{g}$ and the group of Lie algebra automorphisms of $\mathfrak{g}$. The Lie algebra of $\mathrm{Aut}(\mathfrak{g})$ coincides with $\mathrm{der}(\mathfrak{g})$. For $A \in \mathrm{Aut}(G)$, we denote by $A_*$ the corresponding Lie algebra morphism of $\mathfrak{g}$. In general, the map $A \mapsto A_*$ is a homomorphism from $\mathrm{Aut}(G)$ to $\mathrm{Aut}(\mathfrak{g})$; it is an isomorphism if $G$ is connected and simply connected.

4.2. **Modifications of algebras and groups.** We define modifications of nilpotent Lie algebras. A modification map according to our definition coincides with what Gordon and Wilson [24, (2.2) and (2.4)] call a normal modification map. More precisely, they give a more general definition of modification map for solvable Lie algebras, and then prove that modification maps on nilpotent Lie algebras are normal in [24, (2.5)].

In this section, $\mathfrak{n}$ denotes a nilpotent Lie algebra.

Consider the semidirect sum of Lie algebras $\mathfrak{n} \oplus \mathrm{der}(\mathfrak{n})$ whose Lie product is defined by

$$[X + D, X' + D'] = [X, X'] + DX' - D'X + DD' - D'D$$

for all $X, X' \in \mathfrak{n}$ and all $D, D' \in \mathrm{der}(\mathfrak{n})$, From the definition, $\mathrm{ad}\, D|_{\mathfrak{n}}$ coincides with $D$.

**Definition 4.1.** A linear map $\sigma : \mathfrak{n} \to \mathrm{der}(\mathfrak{n})$ is called a *modification map* of the nilpotent Lie algebra $\mathfrak{n}$ if

($\mathfrak{m}1$) $\sigma$ is a Lie algebra homomorphism,
($\mathfrak{m}2$) $\exp(\sigma(\mathfrak{n}))$ is precompact in $\mathrm{Aut}(\mathfrak{n})$, and
($\mathfrak{m}3$) $[\sigma(\mathfrak{n}), \mathfrak{n}] \subseteq \ker(\sigma)$.

*Remark* 4.2. From ($\mathfrak{m}1$) and ($\mathfrak{m}2$), the closure of $\exp(\sigma(\mathfrak{n}))$ is a compact nilpotent Lie group, hence $\sigma(\mathfrak{n})$ is abelian and $[\mathfrak{n}, \mathfrak{n}] \subseteq \ker(\sigma)$.

*Remark* 4.3. From ($\mathfrak{m}3$) and Remark 4.2, $\mathrm{Gr}(\sigma)$, the graph of $\sigma$, that is, $\{X + \sigma(X) : X \in \mathfrak{n}\}$, is a Lie subalgebra of $\mathfrak{n} \oplus \mathrm{der}(\mathfrak{n})$. Moreover, $[\mathrm{Gr}(\sigma), \mathrm{Gr}(\sigma)] \subseteq \mathfrak{n}$, so $\mathrm{Gr}(\sigma)$ is solvable.

*Remark* 4.4. From ($\mathfrak{m}2$) and Remark 4.2, there exists a scalar product on $\mathfrak{n}$ such that each element in $\sigma(\mathfrak{n})$ is a skew-symmetric transformation of $\mathfrak{n}$. Fix such a scalar product and denote the orthogonal complement of $\ker(\sigma)$ in $\mathfrak{n}$ by $\mathfrak{w}$. Using ($\mathfrak{m}3$), one may easily show that

$$\mathfrak{n} = \ker(\sigma) \oplus \mathfrak{w} \quad \text{and} \quad [\sigma(\mathfrak{n}), \mathfrak{w}] = \{0\}. \tag{4.1}$$

**Lemma 4.5.** *Let $\sigma : \mathfrak{n} \to \mathrm{der}(\mathfrak{n})$ be a linear map. Assume that $\sigma$ has property ($\mathfrak{m}2$), that $\sigma(\mathfrak{n})$ is abelian and that $\mathrm{Gr}(\sigma)$ is a Lie subalgebra of $\mathfrak{n} \oplus \mathrm{der}(\mathfrak{n})$. Then $\sigma$ is a modification map.*

The proof is postponed to the end of this subsection.

**Definition 4.6.** Let $N$ be a connected Lie group. A Lie group homomorphism $\varphi : N \to \mathrm{Aut}(N)$ is called a *modification map* if

- (M1) $\varphi_g(\ker(\varphi)) \subseteq \ker(\varphi)$ for all $g \in N$,
- (M2) $\varphi(N)$ is precompact in $\mathrm{Aut}(N)$,
- (M3) there is a submanifold $P$ of $N$ containing $e$ and transverse to $\ker(\varphi)$ such that $\varphi_g(p) = p$ for all $p \in P$ and $g \in N$.

*Remark* 4.7. It follows immediately from the definition that $\ker(\varphi)P$ is a neighbourhood of the identity element. Hence

$$\varphi(x) = (\varphi \circ \varphi_g)(x) \quad \forall g, x \in N. \tag{4.2}$$

**Lemma 4.8.** *Assume that $N$ is a simply connected nilpotent Lie group with Lie algebra $\mathfrak{n}$. Let $\varphi : N \to \mathrm{Aut}(N)$ be a Lie group homomorphism with induced Lie algebra homomorphism $\sigma : \mathfrak{n} \to \mathrm{der}(\mathfrak{n})$. Then the following are equivalent:*

- *(i) $\sigma$ is a modification map, and*
- *(ii) $\varphi$ is a modification map.*

*Proof.* We may identify $\ker(\sigma)$ with the tangent space $T_e \ker(\varphi)$. Suppose that $\mathfrak{w}$ is a subspace of $T_eN$ complementary to $\ker(\sigma)$; define

$$\begin{aligned} L &:= \{A \in \mathrm{Aut}(N) : A_*(\ker(\sigma)) \subseteq \ker(\sigma),\ A_*|_{\mathfrak{w}} = \mathrm{Id}\} \\ &= \{A \in \mathrm{Aut}(N) : A(\ker(\varphi)) \subseteq \ker(\varphi),\ A|_{\exp(\mathfrak{w})} = \mathrm{Id}\}, \end{aligned}$$

which is a closed subgroup of $\mathrm{Aut}(N)$, and hence a Lie group, whose Lie algebra we denote by $\mathfrak{l}$. One may check directly that

$$\mathfrak{l} := \{D \in \mathrm{der}(\mathfrak{n}) : D(\ker(\sigma)) \subseteq \ker(\sigma),\ D(\mathfrak{w}) = \{0\}\}.$$

We show that (i) implies (ii). Since $\varphi(N) = \exp(\sigma(\mathfrak{n}))$, (M2) follows from ($\mathfrak{m}2$). Let $\mathfrak{w}$ be as in Remark 4.4 and $U$ be an open neighbourhood of 0 in $\mathfrak{n}$ on which exp is a diffeomorphism. Then $P := \exp(U \cap \mathfrak{w})$ is a submanifold of $N$ that contains $e$. Now $\sigma(\mathfrak{n}) \subseteq \mathfrak{l}$, so $\varphi(N) \subseteq L$, that is, $\varphi$ satisfies both (M1) and (M3).

Now we prove that (ii) implies (i). Property ($\mathfrak{m}1$) holds by assumption. Next, ($\mathfrak{m}2$) holds since $\varphi(N) = \exp(\sigma(\mathfrak{n}))$. We take $\mathfrak{w}$ to be $T_eP$, and see that $\varphi(N) \subseteq L$, whence $\sigma(\mathfrak{n}) \subseteq \mathfrak{l}$, and ($\mathfrak{m}3$) holds. ☐

**Lemma 4.9.** *Let $\varphi$ be a modification map on a simply connected nilpotent Lie group $N$. Then the graph of $\varphi$, that is,*

$$\mathrm{Gr}(\varphi) := \{L_n \circ \varphi_n : n \in N\} \subseteq \mathrm{Aff}(N),$$

*is a closed Lie subgroup of* $\mathrm{Aff}(N)$ *homeomorphic to* $N$*, with Lie algebra* $\mathrm{Gr}(\varphi_*)$.

*Proof.* To show that $\mathrm{Gr}(\varphi)$ is a subgroup of $\mathrm{Aff}(N)$, we take $g_1, g_2 \in N$, choose $g \coloneqq g_1\varphi_{g_1}(g_2)$, and prove that $L_g \circ \varphi_g = L_{g_1} \circ \varphi_{g_1} \circ L_{g_2} \circ \varphi_{g_2}$. Since $\varphi$ is a homomorphism,

$$\varphi_g = \varphi(g_1\varphi_{g_1}(g_2)) = \varphi(g_1)\varphi(\varphi_{g_1}(g_2)) = \varphi_{g_1} \circ \varphi_{g_2},$$

by (4.2). Since $\varphi_{g_1}$ is an automorphism,

$$L_g(x) = g_1\varphi_{g_1}(g_2\varphi_{g_1}^{-1}(x)) = \left(L_{g_1} \circ \varphi_{g_1} \circ L_{g_2} \circ \varphi_{g_1}^{-1}\right)(x)$$

for all $x \in N$. Therefore

$$L_g \circ \varphi_g = \left(L_{g_1} \circ \varphi_{g_1} \circ L_{g_2} \circ \varphi_{g_1}^{-1}\right) \circ \varphi_{g_1} \circ \varphi_{g_2} = L_{g_1} \circ \varphi_{g_1} \circ L_{g_2} \circ \varphi_{g_2}.$$

This shows that $\mathrm{Gr}(\varphi)$ is closed under composition. Similarly, one may prove that if $g' \coloneqq \varphi_g^{-1}(g^{-1})$, then $L_{g'} \circ \varphi_{g'} = (L_g \circ \varphi_g)^{-1}$, hence $\mathrm{Gr}(\varphi)$ is also closed under inversion.

Now we prove that $\mathrm{Gr}(\varphi)$ is closed in $\mathrm{Aff}(N)$. Take $g_i \in N$ such that $L_{g_i} \circ \varphi_{g_i} \to f$, where $f \in \mathrm{Aff}(N)$. We define $g$ to be $f(e)$ and observe that $g_i = L_{g_i} \circ \varphi_{g_i}(e) \to g$. Since $\varphi$ is continuous,

$$f = \lim_{i\to\infty} L_{g_i} \circ \varphi_{g_i} = L_g \circ \varphi_g \in \mathrm{Gr}(\varphi),$$

and $\mathrm{Gr}(\varphi)$ is closed. Since $\mathrm{Aff}(N)$ is a Lie group, $\mathrm{Gr}(\varphi)$ is a closed Lie subgroup thereof.

Finally, $\psi : g \mapsto L_g \circ \varphi_g$ from $N$ to $\mathrm{Aff}(N)$ is an analytic homeomorphism onto $\mathrm{Gr}(\varphi)$, and $\mathrm{d}\psi(e)(T_eN) = T_e\,\mathrm{Gr}(\varphi)$. By direct computation, $\mathrm{d}\psi(e)(v) = v + \varphi_*(v)$ for all $v \in \mathfrak{n}$. In particular,

$$\mathrm{Lie}(\mathrm{Gr}(\varphi)) = T_e\,\mathrm{Gr}(\varphi) = \mathrm{d}\psi(e)(T_eN) = \mathrm{Gr}(\varphi_*),$$

and we are done. □

**Definition 4.10.** Let $N$ be a simply connected nilpotent Lie group. A Lie group $G$ is called a *modification of* $N$ if there is a modification map $\varphi$ on $N$ such that $G$ is isomorphic to $\mathrm{Gr}(\varphi)$.

In light of Lemmas 4.8 and 4.9, a Lie group $G$ is a modification of $N$ if and only if $G$ is simply connected and there exists a modification map $\sigma$ on $\mathfrak{n}$ such that $\mathfrak{g}$ is isomorphic to $\mathrm{Gr}(\sigma)$.

4.2.1. *Proof of Lemma 4.5.* Let $\mathfrak{n}$ be a nilpotent Lie algebra and $\sigma$ be as in the hypotheses of Lemma 4.5. Since $\mathrm{Gr}(\sigma)$ is a Lie algebra and $\sigma(\mathfrak{n})$ is abelian,

$$[X + \sigma X, Y + \sigma Y] = [X, Y] + [X, \sigma Y] + [\sigma X, Y] \in \mathrm{Gr}(\sigma) \cap \mathfrak{n},$$

and hence

$$\sigma[X, Y] = \sigma[\sigma Y, X] - \sigma[\sigma X, Y] \tag{4.3}$$

for all $X, Y \in \mathfrak{n}$. If moreover $Z \in \mathfrak{n}$, then $\sigma Z \in \mathrm{der}(\mathfrak{n})$, and

$$[\sigma Z, [X, Y]] = [[\sigma Z, X], Y] + [X, [\sigma Z, Y]]. \tag{4.4}$$

These two formulae imply the following, which will be used extensively: for all $X_1, X_2, X_3 \in \mathfrak{n}$,

$$
\begin{aligned}
(4.5)\quad & \sigma[[X_1, X_2], X_3] \\
&= \sigma\big([\sigma[\sigma X_1, X_2], X_3] - [\sigma[\sigma X_2, X_1], X_3] - [\sigma[\sigma X_3, X_1], X_2] \\
&\qquad + [\sigma[\sigma X_3, X_2], X_1] + [\sigma X_2, [\sigma X_3, X_1]] - [\sigma X_1, [\sigma X_3, X_2]]\big).
\end{aligned}
$$

Let $\{\mathfrak{n}_k\}_{k=1}^s$ be the ascending central sequence of $\mathfrak{n}$. Since the abelian algebra $\sigma(\mathfrak{n})$ acts on $\mathfrak{n}$ by skew-symmetric maps, we may write $\mathfrak{n}$ as a direct sum $\bigoplus_{j=1}^J \mathfrak{w}_j$, where each $\mathfrak{w}_j$ is a minimal irreducible subspace of $\mathfrak{n}$ for $\sigma(\mathfrak{n})$. In particular, $\dim(\mathfrak{w}_j)$ is either 1 or 2. Moreover, we may assume that for each $k$ there exists $J_k$ such that $\mathfrak{n}_k = \bigoplus_{j=1}^{J_k} \mathfrak{w}_j$.

*Claim* 4.11. If $\sigma(\mathfrak{n}_k) = \{0\}$, then $\sigma[\sigma X_i, X_j] = 0$ for all $X_i \in \mathfrak{w}_i$ and $X_j \in \mathfrak{w}_j$ such that $[X_i, X_j] \in \mathfrak{n}_{k+1}$.

To prove Claim 4.11, we may assume that $X_i$ and $X_j$ have norm one. In the case where $\dim(\mathfrak{w}_i) = 2$, we define $X_{i*}$ to be a unit vector in $\mathfrak{w}_i$ orthogonal to $X_i$. We define $X_{j*}$ similarly. We consider four cases:

(a) $i = j$ and $\dim(\mathfrak{w}_i) = 2$;
(b) $i \neq j$, $\dim(\mathfrak{w}_i) = 2$ and $\dim(\mathfrak{w}_j) = 2$;
(c) $i \neq j$ and $\dim(\mathfrak{w}_i) + \dim(\mathfrak{w}_j) = 3$;
(d) $i \neq j$, $\dim(\mathfrak{w}_i) = 1$ and $\dim(\mathfrak{w}_j) = 1$.

In case (a), we need to show that $\sigma[\sigma X_i, X_{i*}] = \sigma[\sigma X_i, X_i] = 0$. By assumption, there are $a_i, a_{i*} \in \mathbb{R}$ such that

$$
\begin{aligned}
[\sigma X_i, X_i] &= -a_i X_{i*}, & [\sigma X_i, X_{i*}] &= a_i X_i, \\
[\sigma X_{i*}, X_i] &= -a_{i*} X_{i*}, & [\sigma X_{i*}, X_{i*}] &= a_{i*} X_i.
\end{aligned}
$$

By hypothesis and (4.5),

$$
\begin{aligned}
0 &= \sigma[[X_i, X_{i*}], X_i] = -(a_i^2 + a_{i*}^2)\sigma X_{i*}, \\
0 &= \sigma[[X_i, X_{i*}], X_{i*}] = (a_i^2 + a_{i*}^2)\sigma X_i.
\end{aligned}
$$

Hence $a_i = a_{i*} = 0$ or $\sigma X_i = \sigma X_{i*} = 0$. In both cases,

$$
\sigma[\sigma X_i, X_{i*}] = a_i \sigma X_i = 0 \qquad \text{and} \qquad \sigma[\sigma X_i, X_i] = -a_i \sigma X_{i*} = 0.
$$

To treat case (b), we may assume that $[\mathfrak{w}_i, \mathfrak{w}_i] \subseteq \mathfrak{n}^{(k)}$. By case (a), $[\sigma U, V] = 0$ for all $U, V \in \mathfrak{w}_i$. Thus there are $a_j, a_{j*}, b_i, b_{i*}, b_j, b_{j*} \in \mathbb{R}$ such that

$$
\begin{aligned}
[\sigma X_i, X_i] &= 0 & [\sigma X_i, X_{i*}] &= 0 \\
[\sigma X_{i*}, X_i] &= 0 & [\sigma X_{i*}, X_{i*}] &= 0 \\
[\sigma X_j, X_j] &= -a_j X_{j*} & [\sigma X_j, X_{j*}] &= a_j X_j \\
[\sigma X_{j*}, X_j] &= -a_{j*} X_{j*} & [\sigma X_{j*}, X_{j*}] &= a_{j*} X_j \\
[\sigma X_i, X_j] &= -b_i X_{j*} & [\sigma X_i, X_{j*}] &= b_i X_j \\
[\sigma X_{i*}, X_j] &= -b_{i*} X_{j*} & [\sigma X_{i*}, X_{j*}] &= b_{i*} X_j \\
[\sigma X_j, X_i] &= -b_j X_{i*} & [\sigma X_j, X_{i*}] &= b_j X_i \\
[\sigma X_{j*}, X_i] &= -b_{j*} X_{i*} & [\sigma X_{j*}, X_{i*}] &= b_{j*} X_i.
\end{aligned}
$$

By hypothesis and (4.5),

$$0 = \sigma[[X_j, X_i], X_i] = -2b_i b_{j*}\sigma(X_{i*}) - b_i^2\sigma(X_j) \tag{4.6}$$

$$0 = \sigma[[X_{j*}, X_i], X_i] = -2b_i b_j\sigma(X_{i*}) + b_i^2\sigma(X_{j*}) \tag{4.7}$$

$$0 = \sigma[[X_j, X_{i*}], X_{i*}] = (b_{j*}^2 + b_{i*}b_{j*})\sigma(X_i) - b_{i*}^2\sigma(X_j) \tag{4.8}$$

$$\begin{aligned} 0 = \sigma[[X_{i*}, X_j], X_j] &= (b_{i*}a_{j*} + 2b_i b_j)\sigma(X_{j*}) - a_j b_{j*}\sigma(X_i) \\ &\quad - b_j^2\sigma(X_{i*}) + a_j b_{i*}\sigma(X_j) \end{aligned} \tag{4.9}$$

We will show that

$$\sigma[\sigma X_i, X_j] = -b_i\sigma X_{j*} = 0 \qquad \text{and} \qquad \sigma[\sigma X_j, X_i] = -b_j\sigma X_{i*} = 0.$$

We apply (4.6) to $X_{i*}$ and deduce that $b_i b_j = 0$. Hence (4.7) reduces to $b_i\sigma(X_{j*}) = 0$. If $b_j = 0$, then $b_j\sigma(X_{i*}) = 0$. Otherwise, $b_j \neq 0$, and by applying (4.8) to $X_i$, we deduce that $b_{i*}b_j = 0$. Hence $b_{i*} = 0$ and (4.8) reduces to $b_{j*}\sigma(X_i) = 0$. Finally, (4.9) simplifies to $b_j\sigma(X_{i*}) = 0$.

In case (c), if $\dim(\mathfrak{w}_j) = 1$, then $\sigma[\sigma X_i, X_j] = 0$ trivially. So we show that $\sigma[\sigma X_i, X_j] = 0$ when $\dim(\mathfrak{w}_i) = 1$ and $\dim(\mathfrak{w}_j) = 2$. Fix $X_i \in \mathfrak{w}_i \setminus \{0\}$, and take $b_i \in \mathbb{R}$ such that

$$[\sigma X_i, X_j] = -b_i X_{j*} \quad \text{and} \quad [\sigma X_i, X_{j*}] = b_i X_j.$$

By hypothesis and (4.5),

$$0 = \sigma[[X_{j*}, X_i], X_i] = -b_i^2\sigma(X_{j*}).$$

In case (d), $\sigma[\sigma X_i, X_j] = 0$ trivially.

Claim 4.11 is now proved. To finish the proof of Lemma 4.5, we need to show that $\sigma$ is a homomorphism and $[\sigma(\mathfrak{n}), \mathfrak{n}] \subseteq \ker(\sigma)$.

Since $\sigma(\mathfrak{n})$ is abelian, $\sigma$ is a Lie algebra homomorphism if and only if $[\mathfrak{n}, \mathfrak{n}] \subseteq \ker(\sigma)$. By (4.3), we only have to show ($\mathfrak{m}3$).

By linearity, ($\mathfrak{m}3$) is equivalent to the condition that

$$\sigma[\sigma X_i, X_j] = 0 \qquad \forall X_i \in \mathfrak{w}_i \quad \forall X_j \in \mathfrak{w}_j. \tag{4.10}$$

But Claim 4.11 implies (4.10) for all $X_i, X_j$ by induction on $k$. Indeed, if $[X_i, X_j] \in \mathfrak{n}_1$, then (4.10) follows directly from the fact that $\mathfrak{n}_0 = \{0\}$ and Claim 4.11. If (4.10) holds for all $X_i, X_j$ with $[X_i, X_j] \in \mathfrak{n}_k$, then $\sigma(\mathfrak{n}_k) = \{0\}$ by (4.3) and because $\mathfrak{n}_k$ is spanned by elements of the type $[X_i, X_j]$. Thus (4.10) holds also for all $X_i, X_j$ with $[X_i, X_j] \in \mathfrak{n}_{k+1}$ by Claim 4.11. □

4.3. **Nilshadows of solvable groups of polynomial growth.** In this section, we follow [18] and [7].

For each element $X$ of a Lie algebra, the linear map $\operatorname{ad} X$ admits a unique Jordan decomposition as a sum of a semisimple map, denoted by $\operatorname{ad_s}(X)$, and a nilpotent map.

Let $\mathfrak{g}$ be a solvable Lie algebra of type (R). Let $\mathfrak{v}$ be a subspace of $\mathfrak{g}$ such that

$$\mathfrak{g} = \operatorname{nil}(\mathfrak{g}) \oplus \mathfrak{v} \quad \text{and} \quad \operatorname{ad_s}(\mathfrak{v})\mathfrak{v} = \{0\}, \tag{4.11}$$

which exists by [7, p. 689]. Let $\pi_{\mathfrak{v}} : \mathfrak{g} \to \mathfrak{v}$ be the projection with kernel $\operatorname{nil}(\mathfrak{g})$. On the vector space $\mathfrak{g}$, define the new Lie product $[X, Y]_{\mathrm{nil}}$ by

$$[X, Y]_{\mathrm{nil}} \coloneqq [X, Y] - \operatorname{ad_s}(\pi_{\mathfrak{v}}(X))Y + \operatorname{ad_s}(\pi_{\mathfrak{v}}(Y))X.$$

The Lie algebra $(\mathfrak{g}, [\cdot,\cdot]_{\mathrm{nil}})$ is the *nilshadow* of $\mathfrak{g}$, which is nilpotent and unique up to isomorphism; see, for example, [18].

We show now that the modifications of a nilpotent Lie algebra $\mathfrak{n}$ are exactly the Lie algebras whose nilshadow is $\mathfrak{n}$.

**Proposition 4.12.** *If $\mathfrak{n}$ is a nilpotent Lie algebra and $\sigma$ is a modification map on $\mathfrak{n}$, then $\ker(\sigma) = \mathrm{nil}(\mathrm{Gr}(\sigma))$ and the nilshadow of $\mathrm{Gr}(\sigma)$ is isomorphic to $\mathfrak{n}$.*

*Proof.* Set $\mathfrak{k} := \mathfrak{n} \oplus \sigma(\mathfrak{n})$. Take $X \in \mathfrak{n}$. Since $\mathrm{ad}\, X|_{\mathfrak{n}}^{k} = 0$ for some $k \in \mathbb{N}$, we see that $\mathrm{ad}\, X|_{\mathfrak{n}\oplus\mathrm{der}(\mathfrak{n})}^{k+1} = 0$. Thus

$$\mathrm{ad}_{\mathrm{s}}(X + \sigma(X))|_{\mathfrak{k}} = \mathrm{ad}\,\sigma(X)|_{\mathfrak{k}} \qquad \forall X \in \mathfrak{n}, \tag{4.12}$$

since $\sigma(\mathfrak{n})$ is commutative, $\sigma(\mathfrak{n}) \subseteq \ker(\mathrm{ad}\,\sigma(X))$ and $\mathrm{ad}\,\sigma(X)$ is semisimple on $\mathfrak{k}$ by Remark 4.4.

Now we claim that $\ker(\sigma) = \mathrm{nil}(\mathrm{Gr}(\sigma))$. Since $\mathrm{Gr}(\sigma)$ is solvable by Remark 4.3, we only need to show that the nilpotent elements of $\mathrm{Gr}(\sigma)$ are those in $\ker(\sigma)$. On the one hand, if $\sigma(X) = 0$, then $\mathrm{ad}\, X + \sigma(X)$ is nilpotent on $\mathfrak{n}\oplus\mathrm{der}(\mathfrak{n})$, and in particular on $\mathrm{Gr}(\sigma)$. On the other hand, if $\mathrm{ad}\, X + \sigma(X)$ is nilpotent on $\mathrm{Gr}(\sigma)$, then $\mathrm{ad}\,\sigma(X)|_{\mathrm{Gr}(\sigma)} = 0$, by (4.12), which implies that

$$0 = \mathrm{ad}\,\sigma(X)(Y + \sigma(Y)) = \sigma(X)(Y)$$

for all $Y \in \mathfrak{n}$, and thus $\sigma(X) = 0$.

Let $\mathfrak{w}$ be the subspace of $\mathfrak{n}$ defined in Remark 4.4, and set

$$\mathfrak{v} := \{X + \sigma(X) : X \in \mathfrak{w}\}.$$

Clearly $\mathrm{Gr}(\sigma) = \mathrm{nil}(\mathrm{Gr}(\sigma)) \oplus \mathfrak{v}$ and $\mathrm{ad}_{\mathrm{s}}(\mathfrak{v})\mathfrak{v} = \{0\}$, by (4.12) and (4.1). From (4.12), we also see that $\mathrm{ad}_{\mathrm{s}}(\pi_{\mathfrak{v}}(X + \sigma(X))) = \mathrm{ad}\,\sigma(X)$. So

$$[X + \sigma(X), Y + \sigma(Y)]_{\mathrm{nil}} = [X, Y] \qquad \forall X, Y \in \mathfrak{n}.$$

This shows that the map $X \mapsto X + \sigma(X)$ is an isomorphism from $\mathfrak{n}$ to the nilshadow of $\mathrm{Gr}(\sigma)$. □

The converse of the previous proposition also holds: every simply connected solvable group of polynomial growth is a modification of its nilshadow. We will not use this, but see Remark 4.14 for more.

**Proposition 4.13.** *If $\mathfrak{g}$ is a solvable Lie algebra of type (R) and $\mathfrak{v}$ is chosen such that (4.11) holds, then the map $\sigma : X \mapsto \mathrm{ad}_{\mathrm{s}}(\pi_{\mathfrak{v}}(X))$ from $\mathfrak{g}$ to $\mathfrak{gl}(\mathfrak{g})$ is a modification map of $(\mathfrak{g}, [\cdot,\cdot]_{\mathrm{nil}})$ and $\mathrm{Gr}(\sigma)$ is isomorphic to $\mathfrak{g}$.*

4.4. **Proof of Theorem 1.3.** Let $(N, d)$ be a connected simply connected nilpotent metric Lie group and $G$ be the connected component of the identity of $\mathrm{Iso}(N, d)$. We aim to characterise the Lie groups $H$ that may be equipped with a metric $d_H$ so that $(H, d_H)$ is isometric to $(N, d)$.

As $G$ is of polynomial growth, so is $H$, and $H$ is contractible since it is isometric to $N$. By Lemma 2.22, $H$ is solvable. It is reasonable to expect that there are similarities between $H$ and $N$; in fact we will see that $H$ is a modification of $N$ and $N$ is the nilshadow of $H$.

We restate Theorem 1.3 for the reader's convenience.

**Theorem.** *Let $N$ and $H$ be simply connected Lie groups and assume that $N$ is nilpotent. The following are equivalent:*

*(i) $H$ and $N$ may be made isometric;*
*(ii) $H$ is a modification of $N$;*
*(iii) $H$ is solvable and of polynomial growth and $N$ is its nilshadow.*

We shall prove the claim by establishing that (i) implies (ii), that (ii) implies (iii), and that (iii) implies (i); we discuss other implications after the proof. We point out that Breuillard [7] also showed that (iii) implies (i), and Gordon and Wilson [24] showed essentially that (i) and (ii) are equivalent.

*Proof.* We start by showing that (i) implies (ii). Let $d_N$ and $d_H$ be admissible left-invariant distance functions on $N$ and $H$ and let $F : (N, d_N) \to (H, d_H)$ be an isometry. As the distance functions are left-invariant, we may assume that $F(e_N) = e_G$ without loss of generality. The groups $\mathrm{Iso}(H, d_H)$ and $\mathrm{Iso}(N, d_N)$ are naturally endowed with Lie group structures; see, for example, [34, Section 2.1]. Define $\Psi : \mathrm{Iso}(H, d_H) \to \mathrm{Iso}(N, d_N)$ by $\Psi(f) := F^{-1} \circ f \circ F$. The map $\Psi$ is a continuous group isomorphism, hence a diffeomorphism. In particular, $F$ is also smooth, because it is a composition of smooth maps: indeed, $F(x) = \Psi^{-1}(L_x)(e_N)$ for all $x \in N$. Since $N$ is nilpotent, the stabiliser $\mathrm{Stab}(e_N)$ is a subgroup of $\mathrm{Aut}(N)$ and $\mathrm{Iso}(N, d_N)$ is a closed subgroup of $\mathrm{Aff}(N)$; see [34]. Therefore $\Psi$ is a smooth embedding of $\mathrm{Iso}(H, d_H)$ in $\mathrm{Aff}(N)$. Define the linear map $\sigma : \mathfrak{n} \to \mathrm{der}(\mathfrak{n})$ by

$$\sigma := \pi_* \circ \Psi_* \circ (\mathrm{d}F)_{e_N}, \tag{4.13}$$

where $\pi : \mathrm{Aff}(N) = N \rtimes \mathrm{Aut}(N) \to \mathrm{Aut}(N)$ is the quotient map.

We first prove that

$$\mathrm{Gr}(\sigma) = \Psi_*(\mathfrak{h}), \tag{4.14}$$

that is, $\mathrm{Gr}(\sigma)$ is isomorphic to $\mathfrak{h}$. Recall that $\mathrm{Lie}(\mathrm{Iso}(N, d_N))$ may be represented as a Lie algebra of smooth vector fields on $N$ whose flows are one-parameter groups of isometries. In this representation, a vector $X \in \mathfrak{n}$ corresponds to the right-invariant vector field $X^\dagger$ on $N$ such that $X^\dagger(e_N) = X$, and the Lie algebra of the stabiliser of $e_N$ corresponds to the space of vector fields that vanish at $e_N$. Moreover, if $Y \in \mathfrak{h}$, then $\Psi_*(Y) \in \mathrm{Lie}(\mathrm{Iso}(N, d_N))$ corresponds to the vector field $F^*Y^\dagger$ on $N$. So take $X \in \mathfrak{n}$ and set $Y := (\mathrm{d}F)_{e_N}(X) \in T_{e_H}H = \mathfrak{h}$. Then $(F^*Y^\dagger - X^\dagger)(e_N) = 0$, that is, $\Psi_* \circ (\mathrm{d}F)_{e_N}(X) - X \in \mathrm{der}(\mathfrak{n})$. It follows that $\sigma(X) = \Psi_* \circ (\mathrm{d}F)_{e_N}(X) - X$ and thus

$$X + \sigma(X) = \Psi_* \circ (\mathrm{d}F)_{e_N}(X) \in \Psi_*(\mathfrak{h}).$$

This shows that $\mathrm{Gr}(\sigma) \subseteq \Psi_*(\mathfrak{h})$. Since $\mathrm{Gr}(\sigma)$ has the same dimension as $\mathfrak{n}$ and thus as $\Psi_*(\mathfrak{h})$, we conclude that (4.14) holds.

We need to show that $\sigma$ is a modification map. Since $N$ is simply connected and nilpotent, it is contractible and so $H$ is also contractible. Moreover, $H$ is of polynomial growth, because $N$ is. By Lemma 2.22, $H$ is solvable. Observe that $\sigma(\mathfrak{n}) = \pi_*(\Psi_*(\mathfrak{h}))$, where $\Psi_*(\mathfrak{h})$ is a solvable Lie algebra and $\pi_*$ is a Lie algebra homomorphism. Hence $\sigma(\mathfrak{n})$ is a solvable subalgebra of $\mathrm{Lie}(\mathrm{Stab}(e_N))$, which is a compact Lie algebra, and thus $\sigma(\mathfrak{n})$ is abelian. Property ($\mathfrak{m}2$) is easily checked because $\sigma(\mathfrak{n}) \subseteq \mathrm{Lie}(\mathrm{Stab}(e_N))$. Lemma 4.5 yields that $\sigma$ is a modification map. Now Lemma 4.8 and Lemma 4.9 imply (ii).

Remark 4.3 and Proposition 4.12 show that (ii) implies (iii).

Finally, from [7, Lemma 3.11], on each simply connected solvable group of polynomial growth there exists a Riemannian metric that is left-invariant for both the original Lie structure and for the nilshadow Lie structure, so (iii) implies (i). ☐

*Remark* 4.14. Note that if (iii) holds, then, as already stated, the natural map from $N$ to $H$ is an isometry for some left-invariant distance functions. One may then show that the modification map on $\mathfrak{n}$ constructed in showing that (i) implies (ii) is the differential of a modification map $\varphi$ on $N$ that satisfies $L_p^{(N)} \circ \varphi_p = L_p^{(H)}$, where the superscript indicates the group law for the left translation. Using the formula for the nilshadow product [7, p. 690], one deduces that the modification map $\varphi$ is the group homomorphism $T$ with differential $X \mapsto \mathrm{ad}_{\mathrm{s}}(\pi_{\mathfrak{v}}(X))$. This last observation motivates Proposition 4.13.

*Remark* 4.15. To see that (ii) implies (i), note that there is a left-invariant Riemannian distance function $d_N$ on $N$ that is also $\varphi(N)$-invariant, since $\varphi(N)$ is precompact in $\mathrm{Aut}(N)$. Hence we define

$$d_{\mathrm{Gr}(\varphi)}(L_x \circ \varphi_x, L_y \circ \varphi_y) \coloneqq d_N(x, y)$$

for all $x, y \in N$; it is easy to check that the map $x \mapsto L_x \circ \varphi_x$ is an isometry from the metric Lie group $(\mathrm{Gr}(\varphi), d_{\mathrm{Gr}(\varphi)})$ to $(N, d_N)$ (recall Proposition 4.12).

Here are more observations about Theorem 1.3. We may change a metric on an isometrically homogeneous metric space and change the isometry group by doing so. For instance, we may equip $\mathbb{R}^2$ with any one of the biLipschitz equivalent translation-invariant metrics

$$d((x_1, y_1), (x_2, y_2)) = (|x_1 - x_2|^p + a|y_1 - y_2|^p)^{1/p},$$

where $1 \le p < \infty$ and $0 < a < \infty$. When $p = 2$, the isometry group includes rotations, but otherwise it does not. And when $p = 2$, the rotation group depends on the parameter $a$. However, each of the isometry groups act by bi-Lipschitz transformations with respect to all the other metrics.

However, we may equip a simply connected solvable group $G$ with a left-invariant distance function $d$ so that $(G, d)$ cannot be bi-Lipschitz equivalent to $N$ endowed with a left-invariant distance function.

The universal covering group $H$ of the group $\mathbb{R}^2 \rtimes \mathrm{SO}(2)$ of orientation-preserving rigid motions of $\mathbb{R}^2$ is a simply connected three-dimensional solvable Lie group that admits a left-invariant distance function $d$ such that $(H, d)$ is not bi-Lipschitz equivalent to any nilpotent group. Indeed, the two simply connected three-dimensional nilpotent Lie groups are the abelian group $\mathbb{R}^3$, which is the nilshadow of $H$, and the nonabelian Heisenberg group $\mathbb{H}$. However, if $d$ is a suitable left-invariant sub-Riemannian distance function on $H$, then $(H, d)$ is not even quasiconformally equivalent to either $\mathbb{R}^3$ or $\mathbb{H}$; see [19]. Nevertheless, $(H, d)$ is locally bi-Lipschitz to $\mathbb{H}$ with the standard sub-Riemannian distance function.

4.5. **Metric spaces of polynomial growth.** We now derive some consequences of Theorem 1.3. These results are not surprising since we consider distance functions that are proper and quasigeodesic.

**Corollary 4.16.** *Let $(M, d)$ be a homogeneous metric space of polynomial growth. Suppose that $M$ is a contractible manifold and that $d$ is quasigeodesic. Then $(M, d)$ is quasi-isometrically homeomorphic to a simply connected nilpotent Riemannian Lie group.*

*Proof.* Let $G$ be the connected component of $\mathrm{Iso}(M, d)$ that contains the identity and $K$ be the stabiliser in $G$ of a point $o \in M$. By Proposition 2.11, $M$ is homeomorphic to $G/K$. By Lemma 2.14, $K$ is a maximal compact subgroup of $G$. Moreover, $G$ is a Lie group; see [44, p. 243] or the statement after Theorem 1.1.

Let $L$ be a Levi subgroup of $G$. By Lemma 2.21, $G$ is of polynomial growth, so $L$ is compact by Lemma 2.22. Therefore, after a conjugation if necessary, $L$ is contained in $K$. By Propositions 2.17 and 2.11, there is a solvable Lie subgroup $H$ in $G$ such that the map $f \mapsto f(o)$ is a homeomorphism from $H$ to $M$. Let $d_H$ be the distance function on $H$ pulled back from that on $M$, which is left-invariant and admissible since $H$ acts by isometries on $(M, d)$. Hence $(M, d)$ is isometric to the simply connected solvable metric Lie group $(H, d_H)$.

Let $N$ be the nilshadow of $H$. By Theorem 1.3, there are distance functions $d'_H$ and $d'_N$ on $H$ and $N$ such that $(H, d'_H)$ and $(N, d'_N)$ are isometric. We may assume that $d'_H$ and $d'_N$ are Riemannian, by [34, Section 2.3]. Since $d$ is assumed to be of polynomial growth, $(H, d_H)$ is proper by Remark 2.20. Finally, admissible proper left-invariant quasigeodesic distance functions on a Lie group are quasi-isometric (see [8]), and $d$ is assumed to be quasigeodesic, so the identity map on $H$ is a quasi-isometry from $d_H$ to $d'_H$. □

**Corollary 4.17.** *Let $(M, d)$ be a homogeneous metric space of polynomial growth. Suppose that the distance function $d$ is quasigeodesic. Then $(M, d)$ is quasi-isometric to a simply connected nilpotent Riemannian Lie group.*

*Proof.* Let $G$ be the connected component of the identity in the group of isometries of $(M, d)$ and $d_G$ be a Busemann distance function on $G$, as defined in Proposition 2.7 (using the transitivity of $G$ established in Proposition 2.13). Let $\mu$ be a Haar measure on $G$ and $m$ be a Radon measure on $M$ such that (2.4) holds. Using the fact that the quotient map from $(G, d_G)$ to $(M, d)$ is a $(1, C)$-quasi-isometry and the relation (2.4) between the measures, one may easily show that the metric space $(G, d_G)$ is of polynomial growth.

Let $K_0$ be a maximal compact subgroup of $G$ and define $M'$ to be $G/K_0$. By Lemma 2.14, $M'$ is a contractible manifold. By Corollary 3.4, there is an admissible $G$-invariant distance function $d'$ on $M'$ such that the quotient map from $(G, d_G)$ to $(M', d')$ is a $(1, C)$-quasi-isometry. Let $m'$ be a $G$-invariant Radon measure on $M'$ such that (2.4) holds. Using the relation (2.4) between the measures and the fact that the quotient map from $(G, d_G)$ to $(M', d')$ is a $(1, C)$-quasi-isometry, we may now prove that $(M', d')$ is of polynomial growth.

Since $(M', d')$ is quasi-isometric to $(M, d)$, the metric $d'$ is quasigeodesic. We conclude by applying Corollary 4.16 to $(M', d')$. □

## 5. Characterisation of self-similar Lie groups

5.1. **Basic properties of self-similar Lie groups.** We recall the definition of self-similar Lie group and we present some examples and properties.

**Definition 5.1.** A *self-similar Lie group* is given by $(G, d, \delta)$ where $G$ is a connected Lie group, $d$ is a left-invariant distance on $G$ inducing the manifold topology and $\delta : G \to G$ is an automorphism such that $d(\delta x, \delta y) = \lambda d(x, y)$ for some $\lambda \neq 1$.

The basic examples of self-similar Lie groups are normed vector spaces of finite dimension with a dilation $\delta v = \lambda v$. Several other examples are already available using $G = \mathbb{R}^2$.

If $\alpha, \beta \geq 1$, the automorphisms $\delta_\lambda = \begin{pmatrix} \lambda^\alpha & 0 \\ 0 & \lambda^\beta \end{pmatrix}$ are all dilations of factor $\lambda$ for several distances such as $d((x, y), (x', y')) = \max\{|x-x'|^{1/\alpha}, |y-y'|^{1/\beta}\}$ or, if $\alpha = \beta$, $d(x, y) = \|x - y\|^{1/\alpha}$ where $\|\cdot\|$ is a norm on $\mathbb{R}^2$. It has been shown in [37] that, for $\alpha = 2$, there exists a homogeneous distance $d$ whose spheres are fractals in $\mathbb{R}^2$.

The automorphisms $\delta_\lambda = \lambda^\alpha \begin{pmatrix} \cos(\log\lambda) & -\sin(\log\lambda) \\ \sin(\log\lambda) & \cos(\log\lambda) \end{pmatrix}$ are dilations of factor $\lambda$ for the distance $d(x, y) = \|x - y\|^{\frac{1}{\alpha}}$, where $\|\cdot\|$ is the Euclidean norm and $\alpha \geq 1$.

If $\alpha > 1$, then there is a left-invariant distance $d$ on $\mathbb{R}^2$ for which the automorphisms $\delta_\lambda = \begin{pmatrix} \lambda^\alpha & \lambda^\alpha \log(\lambda^\alpha) \\ 0 & \lambda^\alpha \end{pmatrix}$ are dilations of factor $\lambda$. These dilations appear in [5] in the study of visual boundaries of Gromov hyperbolic spaces. See also [60] for furter results and examples in $\mathbb{R}^n$. In [39] the authors have studied those self-similar Lie groups that admit a Besicovitch covering property. See also [41] for further references.

**Definition 5.2.** A *(positive) grading* of a Lie algebra $\mathfrak{g}$ is a splitting $\mathfrak{g} = \bigoplus_{t>0} V_t$ so that $[V_s, V_t] \subset V_{s+t}$ for all $s, t > 0$. A Lie group $G$ is *graduable* if it is simply connected and its Lie algebra admits a grading.

Notice that only a finite number of $V_t$'s are not $\{0\}$, because $\mathfrak{g}$ has finite dimension. Moreover, a graduable group is necessarily nilpotent. If $G$ is a graduable Lie group with grading $\mathfrak{g} = \bigoplus_{t>0} V_t$, we may define the *standard dilations* $\delta_\lambda : G \to G$ by imposing $(\delta_\lambda)_* v = \lambda^t v$ for all $v \in V_t$. It is known that a distance $d$ exists on $G$ so that $(G, d, \delta_\lambda)$ is a self-similar group if and only if $V_t = \{0\}$ for all $t < 1$, see [39] for references.

Graduable groups are in fact the only Lie groups that support a dilation by the following theorem due to Siebert [51].

**Theorem 5.3** (Siebert)**.** *Let $G$ be a connected Lie group and let $\delta : G \to G$ be a Lie group automorphism such that for all $g \in G$*

$$\lim_{n\to\infty} \delta^n g = e_G.$$

*Then $G$ is graduable, nilpotent and simply connected.*

The proof constructs a grading for $G$ as follows. One denotes by $\mathfrak{g}_{\mathbb{C}}$ the complexified Lie algebra and by $W_\alpha$ the generalized eigenspace of $(\delta_*)_{\mathbb{C}} : \mathfrak{g}_{\mathbb{C}} \to \mathfrak{g}_{\mathbb{C}}$ with respect to $\alpha \in \mathbb{C}$, that is,

$$W_\alpha = \{v \in \mathfrak{g}_{\mathbb{C}} : \exists n \in \mathbb{N}\ ((\delta_*)_{\mathbb{C}} - \alpha \mathrm{Id})^n v = 0\}.$$

It can be proven that $[W_\alpha, W_\beta] = W_{\alpha\beta}$ and that $(\delta_*)_{\mathbb{C}} W_\alpha = W_\alpha$, for all $\alpha, \beta \in \mathbb{C}$. Thus, one has that $V_t = \mathfrak{g} \cap \bigoplus_{\log|\alpha| = -t} W_\alpha$ defines the layers of a grading for $\mathfrak{g}$.

**Corollary 5.4.** *If $(G, d, \delta)$ is a self-similar Lie group, then $G$ is graduable, nilpotent and simply connected. Moreover, all metric dilations on $(G, d)$ are Lie group automorphisms of $G$.*

*Proof.* Since on a self-similar Lie group one has a contractive automorphism, the first statement follows from Theorem 5.3. Recall that a metric dilation on a metric space $(G, d)$ is a bijection $f : G \to G$ such that $d(f(x), f(y)) = \mu d(x, y)$ for all $x, y \in G$ and some $\mu \neq 1$. Notice that such a map is also an isometry from $(G, \mu d)$ to $(G, d)$. By [34], isometries between connected nilpotent Lie groups are Lie group isomorphisms. $\square$

5.2. **Proof of Theorem 1.4.** The last sentence in Theorem 1.4 has been proven in Corollary 5.4. We restate the first part of Theorem 1.4 for the reader's convenience.

**Theorem.** *If a metric space is locally compact, connected, isometrically homogeneous, and it admits a metric dilation, then it is isometric to self-similar Lie group.*

The converse part of the theorem is obvious. Hence we focus on metric spaces with a dilation. Throughout this section, we assume that $(M, d)$ is a homogeneous metric space, $\lambda > 1$, and $\delta$ is a bijection of $M$ such that $d(\delta x, \delta y) = \lambda d(x, y)$ for all $x, y \in M$. Since $M$ is locally compact and isometrically homogeneous, it is complete, and the Banach fixed point theorem shows that $\delta$ has a unique fixed point, $o$ say. As usual, $G$ denotes the connected component of the identity in $\mathrm{Iso}(M)$. We prove a few preliminary results.

**Lemma 5.5.** *The metric space $(M, d)$ is proper and doubling.*

*Proof.* The ball $B(o, r)$ is relatively compact for all sufficiently small $r$; using the dilation we see that this holds for all $r \in \mathbb{R}$.

We now show that $(M, d)$ is a doubling metric space. Since $\bar{B}(o, \lambda)$ is compact, there are points $x_1, \dots, x_k \in \bar{B}(o, \lambda)$ such that

$$\bar{B}(o, \lambda) \subseteq \bigcup_{i=1}^{k} B(x_i, 1/2).$$

Take $R \in \mathbb{R}^+$, and let $n \coloneqq \lfloor \log_\lambda R \rfloor$, so that $1 \leq \lambda^{-n} R < \lambda$. Then

$$\delta^n B(x_i, 1/2) \subseteq \delta^n B(x_i, \lambda^{-n} R/2) = B(\delta^n x_i, R/2),$$

and so

$$B(o, R) = \delta^n (B(o, \lambda^{-n} R)) \subseteq \delta^n (B(o, \lambda)) \subseteq \bigcup_{i=1}^{k} B(\delta^n x_i, R/2).$$

Since $(M, d)$ is isometrically homogeneous, $(M, d)$ is doubling. $\square$

**Lemma 5.6.** *The space $M$ is an analytic contractible manifold and $G$ is a Lie group that acts on $M$ analytically and transitively. Moreover $G$ is of polynomial growth.*

*Proof.* Let $\pi : f \mapsto fo$ be the map from $G$ to $M$. Define $T : G \to G$ by $Tf = \delta \circ f \circ \delta^{-1}$; then $\pi \circ T = \delta \circ \pi$. Let $K$ be the maximal compact normal subgroup of $G$. Note that $T(K) = K$, since $T$ is an automorphism of $G$. Then $\pi(K)$ is a compact subset of $M$: let $r \coloneqq \max\{d(o,p) : p \in \pi(K)\}$. Then

$$\pi(K) = \pi T^{-1}(K) = \delta^{-1}\pi(K) \subseteq B(o, \lambda^{-1} r),$$

that is, $r = 0$. Therefore $\pi(K) = \{o\}$, and $K$ is contained in the stabiliser of $o$ in $G$. Since $G$ acts transitively by Proposition 2.13, $K = \{e_G\}$. By Proposition 2.11 and Lemma 3.5, $G$ is a Lie group, $M$ is a manifold and the action of $G$ on $M$ is analytic.

Since $M$ is a manifold and admits a metric dilation, it is compactly contractible, and hence contractible by Lemma 2.14. Since moreover $M$ is doubling and proper, it is of polynomial growth by Remark 2.18. By Lemma 2.21, $G$ is a group of polynomial growth. $\square$

*Proof of Theorem 1.4.* Let $(M, d)$ be a homogeneous metric space. Let $\delta$ be a metric dilation of factor $\lambda \in (1, \infty)$ and with fixed point $o$. Let $G$ denote the connected component of the identity in $\mathrm{Iso}(M)$. By Lemma 5.6, $G$ is a Lie group of polynomial growth and $M$ may be identified with $G/K$, where $K$ is the stabiliser of $o$ in $G$.

We will apply Proposition 2.17. Since $G$ is of polynomial growth, each Levi subgroup of $G$ is compact, by Lemma 2.22. Since $G/K$ is contractible by Lemma 5.6, $K$ is a maximal compact subgroup by Lemma 2.14, and therefore $K$ contains a Levi subgroup.

From Proposition 2.17, there exists a connected Lie subgroup $H$ of $G$ such that the restricted quotient map $h \mapsto ho$ from $H$ to $M$ is a homeomorphism. We use this homeomorphism to make $H$ into a self-similar Lie group isometric to $(M, d)$.

Define the distance function $d_H$ on $H$ by $d_H(h, h') = d(ho, h'o)$. It is clear that this is an admissible metric, and it is left-invariant because

$$d_H(hh', hh'') = d(h(h'(o)), h(h''(o))) = d(h'o, h''o) = d_H(h', h'')$$

for all $h, h', h'' \in H$. Further, define the map $T$ on $G$ by

$$Tg \coloneqq \delta \circ g \circ \delta^{-1}.$$

Then $T$ is a Lie group automorphism of $G$. Since $TK = K$ and the Killing form is invariant under automorphisms, $TH = H$. Thus $T|_H$ is a Lie group automorphism of $H$.

We note that after the identification of $H$ with $M$, the map $T|_H$ coincides with $\delta$. Indeed,

$$(Th)(o) = (\delta h \delta^{-1})(o) = \delta(ho),$$

and the proof is complete. $\square$

## References


[1] S. A. Antonyan, 'Characterizing maximal compact subgroups', *Arch. Math. (Basel)* **98** (6) (2012), 555–560.

[2] L. Auslander and L. W. Green, '$G$-induced flows', *Amer. J. Math.* **88** (1966), 43–60.

[3] R. W. Bagley and M. R. Peyrovian 'A note on compact subgroups of topological groups' *Bull. Austral. Math. Soc.* **33** (1986), 273–278.

[4] A. Bellaïche, 'The tangent space in sub-Riemannian geometry'. In: *Sub-Riemannian Geometry.* Progress in Math., Vol. 144, Birkhäuser, Basel, 1996, pp. 1–78.

[5] M. Bonk and B. Kleiner, 'Conformal dimension and Gromov hyperbolic groups with 2-sphere boundary', *Geom. Topol.* **9** (2005), 219–246.

[6] N. Bourbaki, *Lie Groups and Lie Algebras. Chapters 1–3.* Elements of Mathematics. Springer-Verlag, Berlin, 1989. Translated from the French, reprint of the 1975 edition.

[7] E. Breuillard, 'Geometry of locally compact groups of polynomial growth and shape of large balls', *Groups Geom. Dyn.* **8** (3) (2014), 669–732.

[8] E. Breuillard and E. Le Donne, 'On the rate of convergence to the asymptotic cone for nilpotent groups and subFinsler geometry', *Proc. Nat. Acad. Sci. USA* **110** (48) (2013), 19220–19226.

[9] D. Burago, Y. Burago, and S. Ivanov, *A Course in Metric Geometry.* Graduate Studies in Math., Vol. 33. American Mathematical Society, Providence, 2001.

[10] J. Cheeger and B. Kleiner, 'Differentiating maps into $L^1$, and the geometry of BV functions', *Ann. of Math. (2)* **171** (2) (2010), 1347–1385.

[11] J. Cheeger and B. Kleiner, 'Metric differentiation, monotonicity and maps to $L^1$', *Invent. Math.* **182** (2) (2010), 335–370.

[12] Y. Cornulier, 'Commability and focal locally compact groups', *Indiana Univ. Math. J.* **64** (1) (2015), 115–150.

[13] Y. Cornulier, 'On sublinear bilipschitz equivalence of groups', *ArXiv 170206618C.*

[14] Y. Cornulier and P. de la Harpe, 'Metric geometry of locally compact groups, *ArXiv 1403.3796C.*

[15] Y. Cornulier and R. Tessera, 'Geometric presentations of Lie groups and their Dehn functions', *ArXiv 1310.5373C.*

[16] L. J. Corwin and F. P. Greenleaf, *Representations of Nilpotent Lie Groups and their Applications. Part I. Basic Theory and Examples.* Cambridge Studies in Advanced Math., Vol. 18. Cambridge University Press, Cambridge, 1990.

[17] Y. de Cornulier, 'Dimension of asymptotic cones of Lie groups', *J. Topol.* **1** (2) (2008), 342–361.

[18] N. Dungey, A. F. M. ter Elst, and D. W. Robinson, *Analysis on Lie Groups with Polynomial Growth.* Progress in Math., Vol. 214. Birkhäuser, Boston, 2003.

[19] K. Fässler, P. Koskela, and E. Le Donne, 'Nonexistence of quasiconformal maps between certain metric measure spaces', *Int. Math. Res. Not. IMRN* (16) (2015), 6968–6987.

[20] H. Federer, *Geometric Measure Theory.* Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag, New York, 1969.

[21] G. B. Folland and E. M. Stein, *Hardy Spaces on Homogeneous Groups.* Mathematical Notes, Vol. 28. Princeton University Press, Princeton; University of Tokyo Press, Tokyo, 1982.

[22] R. W. Goodman, *Nilpotent Lie groups: Structure and Applications to Analysis.* Lecture Notes in Math., Vol. 562. Springer-Verlag, Berlin–New York, 1976.

[23] C. S. Gordon and E. N. Wilson, 'The fine structure of transitive Riemannian isometry groups. I', *Trans. Amer. Math. Soc.* **289** (1) (1985), 367–380.

[24] C. S. Gordon and E. N. Wilson, 'Isometry groups of Riemannian solvmanifolds', *Trans. Amer. Math. Soc.* **307** (1) (1988), 245–269.

[25] M. Gromov, 'Groups of polynomial growth and expanding maps', *Inst. Hautes Études Sci. Publ. Math.* **53** (1981), 53–73.

[26] Y. Guivarc'h, 'Croissance polynomiale et périodes des fonctions harmoniques', *Bull. Soc. Math. France* **101** (1973), 333–379.

[27] U. Hamenstädt, 'Some regularity theorems for Carnot-Carathéodory metrics', *J. Differential Geom.* **32** (3) (1990), 819–850.

[28] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces.* Graduate Studies in Math., Vol. 34. American Mathematical Society, Providence, 2001. Corrected reprint of the 1978 original.

[29] L. Hörmander, 'Hypoelliptic second order differential equations', *Acta Math.* **119** (1967), 147–171.

[30] N. Jacobson, *Lie Algebras.* Dover, New York, 1979. Republication of the 1962 original.

[31] J. W. Jenkins, 'A characterization of growth in locally compact groups', *Bull. Amer. Math. Soc.* **79** (1973), 103–106.

[32] J. W. Jenkins, 'Growth of connected locally compact groups', *J. Funct. Anal.* **12** (1973), 113–127.

[33] J. L. Kelley. *General Topology.* Graduate Texts in Math., Vol. 27. Springer-Verlag, New York-Berlin, 1975. Reprint of the 1955 edition.

[34] V. Kivioja and E. Le Donne, 'Isometries of nilpotent metric groups', *J.École Polytechnique – Math.* **4** (2017), 473–482.

[35] A. W. Knapp, *Lie Groups Beyond an Introduction.* Progress in Math., Vol. 140. Birkhäuser, Boston, 2002. 2nd edn.

[36] A. Korányi and H. M. Reimann, 'Foundations for the theory of quasiconformal mappings on the Heisenberg group', *Adv. Math.* **111** (1) (1995), 1–87.

[37] E. Le Donne and S. Nicolussi Golo, 'Regularity properties of spheres in homogeneous groups', *Trans. of the AMS*, to appear; arXiv: 1509.03881

[38] E. Le Donne and S. Rigot, 'Besicovitch Covering Property for homogeneous distances in the Heisenberg groups', *ArXiv 1406.1484L.*

[39] E. Le Donne and S. Rigot, 'Besicovitch Covering Property on graded groups and applications to measure differentiation', *J reine angew. Math.*, to appear; arXiv: 1512.04936.

[40] E. Le Donne, 'A metric characterization of Carnot groups', *Proc. Amer. Math. Soc.* **143** (2) (2015), 845–849.

[41] E. Le Donne, 'A primer on Carnot groups: homogenous groups, CC spaces, and regularity of their isometries', *ArXiv 1604.08579* (2016).

[42] J. Lee and A. Naor, '$L^p$' metrics on the Heisenberg group and the Goemans-Linial conjecture', *54th Annual Symposium on Foundations of Computer Science (FOCS'06)* (2006), 99–108.

[43] J. Milnor, 'A note on curvature and fundamental group', *J. Differential Geom.* **2** (1968), 1–7.

[44] D. Montgomery and L. Zippin, *Topological Transformation Groups.* Krieger, Huntington, N.Y., 1974. Reprint of the 1955 original.

[45] R. Montgomery, *A Tour of Subriemannian Geometries, Their Geodesics and Applications.* Mathematical Surveys and Monographs, Vol. 91. American Mathematical Society, Providence, 2002.

[46] L. Nachbin, *The Haar Integral.* Van Nostrand, Princeton–Toronto–London, 1965.

[47] P. Pansu, 'Croissance des boules et des géodésiques fermées dans les nilvariétés', *Ergodic Theory Dynam. Systems* **3** (3) (1983), 415–445.

[48] P. Pansu, 'Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un', *Ann. of Math. (2)* **129** (1) (1989), 1–60.

[49] S. D. Pauls, 'The large scale geometry of nilpotent Lie groups', *Comm. Anal. Geom.* **9** (5) (2001), 951–982.

[50] L. P. Rothschild and E. M. Stein, 'Hypoelliptic differential operators and nilpotent groups', *Acta Math.* **137** (3-4) (1976), 247–320.

[51] E. Siebert, 'Contractive automorphisms on locally compact groups', *Math. Z.* **191** (1) (1986), 73–90.

[52] M. Stoll, 'On the asymptotics of the growth of 2-step nilpotent groups', *J. London Math. Soc. (2)* **58** (1) (1998), 38–48.

[53] R. S. Strichartz, 'Sub-Riemannian geometry', *J. Differential Geom.* **24** (2) (1986), 221–263.

[54] M. Stroppel, *Locally Compact Groups.* EMS Textbooks in Math. European Mathematical Society, Zürich, 2006.
[55] T. Tao, *Hilbert's Fifth Problem and Related Topics.* Graduate Studies in Math., Vol. 153. American Mathematical Society, Providence, 2014.
[56] R. Tessera, 'Volume of spheres in doubling metric measured spaces and in groups of polynomial growth', *Bull. Soc. Math. France* **135** (1) (2007), 47–64.
[57] V. S. Varadarajan, *Lie Groups, Lie Algebras, and Their Representations.* Graduate Texts in Math., Vol. 102. Springer-Verlag, New York, 1984. Reprint of the 1974 edition.
[58] E. N. Wilson, 'Isometry groups on homogeneous nilmanifolds', *Geom. Dedicata* **12** (3) (1982), 337–346.
[59] J. A. Wolf, 'Growth of finitely generated solvable groups and curvature of Riemannian manifolds', *J. Differential Geom.* **2** (1968), 421–446.
[60] X. Xie, 'Large scale geometry of negatively curved $\mathbb{R}^n \rtimes \mathbb{R}$' *Geom. Topol.* **18** (2015), 831–872.

School of Mathematics and Statistics, University of New South Wales, UNSW Sydney 2052, Australia

Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä FI-40014 Finland

Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä FI-40014 Finland

Department of Mathematics, University of Trento, via Calepina 14, I-38122 Trento, Italy

School of Mathematics and Statistics, University of New South Wales, UNSW Sydney 2052, Australia

[D]

**Some remarks on contact variations in the first Heisenberg group**
S. Nicolussi Golo
ArXiv 1611.07358

# SOME REMARKS ON CONTACT VARIATIONS IN THE FIRST HEISENBERG GROUP

SEBASTIANO GOLO

Abstract. We show that in the first sub-Riemannian Heisenberg group there are intrinsic graphs of smooth functions that are both critical and stable points of the sub-Riemannian perimeter under compactly supported variations of contact diffeomorphisms, despite the fact that they are not area-minimizing surfaces. In particular, we show that if $f : \mathbb{R}^2 \to \mathbb{R}$ is a $\mathscr{C}^1$-intrinsic function, and $\nabla^f \nabla^f f = 0$, then the first contact variation of the sub-Riemannian area of its intrinsic graph is zero and the second contact variation is positive.

Contents



## 1. Introduction

We want to address some new features of the sub-Riemannian perimeter in the Heisenberg group. The notion of sub-Riemannian perimeter in the Heisenberg group, the so-called *intrinsic perimeter*, has been enstablished as a direct and natural extension from the Euclidean perimeter in $\mathbb{R}^n$. However, in many aspects, there are fundamental differences that lead to new open questions [6, 7, 10, 18, 16].

Before a detailed explanation, let us introduce some basic notions and notations we need in this introduction. The (first) Heisenberg group $\mathbb{H}$ is a

*Date*: August 24, 2017.

2010 *Mathematics Subject Classification.* 53C17,49Q20.

*Key words and phrases.* Sub-Riemannian Geometry, Sub-Riemannian perimeter, Heisenberg Group, Bernstein's Problem, Contact diffeomorphisms.

The author has been supported by the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n. 607643.

three dimensional Lie group diffeomorphic to $\mathbb{R}^3$. However, when endowed with a left-invariant sub-Riemannian distance, it becomes a metric space with Hausdorff dimension equal to four; see [3].

By standard methods of Geometric Measure Theory, one defines the *intrinsic perimeter* $P(E;\Omega)$ of a measurable set $E \subset \mathbb{H}$ in an open set $\Omega \subset \mathbb{H}$. We will denote it also by $\mathscr{A}(\partial E \cap \Omega)$.

*Regular surfaces* are topological surfaces in $\mathbb{H}$ that admit a continuously varying tangent plane and they play an important role in the theory of sets with finite intrinsic perimeter. They are the sub-Riemannian counterpart of smooth hypersurfaces in $\mathbb{R}^n$. Regular surfaces are locally graphs of so-called $\mathscr{C}^1$*-intrinsic functions* $\mathbb{R}^2 \to \mathbb{R}$. We will focus on these functions and their graphs.

The space of all $\mathscr{C}^1$-intrinsic functions will be denoted by $\mathscr{C}^1_{\mathbb{W}}$ and the graph of $f : \mathbb{R}^2 \to \mathbb{R}$ by $\Gamma_f \subset \mathbb{H}$. It is well known that $f \in \mathscr{C}^1_{\mathbb{W}}$ if and only if $f \in \mathscr{C}^0(\mathbb{R}^2)$ and the distributional derivative

$$\nabla^f f = \partial_\eta f + \frac{1}{2}\partial_\tau(f^2)$$

is continuous, where we denote by $(\eta, \tau)$ the coordinates on $\mathbb{R}^2$; see [1, 18]. If $\omega \subset \mathbb{R}^2$, the intrinsic area of $\Gamma_f$ above $\omega$ is

$$\mathscr{A}(\Gamma_f \cap \Omega_\omega) = \int_\omega \sqrt{1 + (\nabla^f f)^2}\, \mathrm{d}\eta\, \mathrm{d}\tau,$$

where $\Omega_\omega = \{(0, \eta, \tau) * (\xi, 0, 0) : (\eta, \tau) \in \omega,\ \xi \in \mathbb{R}\}$, with $*$ denoting the group operation in $\mathbb{H}$.

An important open problem concerning $\mathscr{C}^1_{\mathbb{W}}$ is *Bernstein's problem*: If the graph $\Gamma_f$ of $f \in \mathscr{C}^1_{\mathbb{W}}$ is a locally minimizer of the intrinsic area, is $\Gamma_f$ a plane? See Section 2.4 for a precise statement and [2, 4, 17, 9] for further reading.

In the study of perimeter minimizers in $\mathbb{H}$, we identify three main issues that mark the gap from the Euclidean theory. First, the map $f \mapsto \nabla^f f$ is a nonlinear operator. Such non-linearity reflects on the fact that basic function spaces like $\mathscr{C}^1_{\mathbb{W}}$ itself, or the space of functions with bounded intrinsic variation, are not vector spaces. See Remark 2.3 for details.

Second, the area functional is not convex (say on $\mathscr{C}^1(\mathbb{R}^2)$). In particular, there are critical points that are not extremals, see [4]. In other words, a first variation condition

$$\left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}\right|_{\epsilon=0} \mathscr{A}(\Gamma_{f+\epsilon\phi} \cap (\Omega_\omega)) = 0 \qquad \forall \phi \in \mathscr{C}^\infty_c(\omega) \tag{1}$$

does not characterize minimizers. However, if $f \in \mathscr{C}^1(\mathbb{R}^2)$, a second variation condition $\left.\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2}\right|_{\epsilon=0} \mathscr{A}(\Gamma_{f+\epsilon\phi} \cap (\Omega_\omega)) \geq 0$ does, see [9].

Third, there are objects among sets of finite intrinsic perimeter with very low regularity, see Remark 2.3. The standard variational approach as in (1) fails when applied to these objects. More precisely, if $f \in \mathscr{C}^1_{\mathbb{W}}$, then $\mathscr{A}(\Gamma_{f+\epsilon\phi} \cap (\Omega_\omega))$ may be $+\infty$ for all $\epsilon \neq 0$, all $\omega \subset \mathbb{R}^2$ open and all $\phi \in \mathscr{C}^\infty_c(\omega) \setminus \{0\}$. In another approach, one can consider smooth one-parameter families of diffeomorphisms $\Phi_\epsilon : \mathbb{H} \to \mathbb{H}$ with $\Phi_0 = \mathrm{Id}$ and $\{\Phi_\epsilon \neq \mathrm{Id}\} \subset\subset \Omega$, and take variations of $\mathscr{A}(\Phi_\epsilon(\Gamma_f) \cap \Omega)$. However, it may happen again that $\mathscr{A}(\Phi_\epsilon(\Gamma_f) \cap \Omega) = +\infty$ for all $\epsilon \neq 0$.

After further considerations, one understands that we need to restrict the choice of $\Phi_\epsilon$ to *contact diffeomorphisms*, see Proposition 5.1. See also [3] and [12] for references on contact diffeomorphisms. In this setting, we address the question whether, despite this restriction, conditions on the first and second variations with contact diffeomorphisms can single out minimal graphs. Our answer is no:

**Theorem 1.1.** *There is $f \in \mathscr{C}^1_{\mathbb{W}}$ such that, for all $\Omega \subset \mathbb{H}$ open and all smooth one-parameter families of contact diffeomorphisms $\Phi_\epsilon : \mathbb{H} \to \mathbb{H}$ with $\Phi_0 = \mathrm{Id}$ and $\{\Phi_\epsilon = \mathrm{Id}\} \subset\subset \Omega$, it holds*

$$\left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}\right|_{\epsilon=0} \mathscr{A}(\Phi_\epsilon(\Gamma_f) \cap \Omega) = 0 \quad \text{and} \quad \left.\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2}\right|_{\epsilon=0} \mathscr{A}(\Phi_\epsilon(\Gamma_f) \cap \Omega) \geq 0,$$

*but $\Gamma_f$ is not an area-minimizing surface.*

In fact, all smooth solutions to the equation $\nabla^f(\nabla^f f) = 0$ are examples of the functions appearing in the theorem.

The proof of Theorem 1.1 is based on a "Lagrangian" approach to $\mathscr{C}^1_{\mathbb{W}}$. Indeed, a function $f \in \mathscr{C}^1_{\mathbb{W}}$ is uniquely characterized by the integral curves of the planar vector field $\nabla^f = \partial_\eta + f\partial_\tau$. We will thus take variations of $f$ via smooth one-parameter families of diffeomorphisms $\phi_\epsilon : \mathbb{R}^2 \to \mathbb{R}^2$, i.e., by smoothly varying the integral curves of $\nabla^f$; see Section 4. We will then prove that this approach is equivalent to the use of contact diffeomorphisms $\Phi_\epsilon : \mathbb{H} \to \mathbb{H}$; see Section 5.

Finally, we will consider functions $f \in \mathscr{C}^1_{\mathbb{W}}$ that are solutions to the equation $\nabla^f \nabla^f f = 0$ in a Lagrangian sense, that is, functions such that $\nabla^f f$ is constant along the integral curves of $\nabla^f$. We will characterize such functions as the ones for which the integral curves of $\nabla^f$ are parabolas, or, equivalentely, as the ones whose graph $\Gamma_f$ is ruled by horizontal straight lines. These functions are the ones appearing in Theorem 1.1.

The paper is organized as follows. Section 2 is devoted to the presentation of all main definitions. In the next Section 3, we study solutions to the equation $\nabla^f \nabla^f f = 0$. We construct a Lagrangian variation of a function $f \in \mathscr{C}^1_{\mathbb{W}}$ in Section 4. In Section 5, we prove some basic properties of contact diffeomorphisms. Section 6 is devoted to the first contact variation and Section 7 to the second contact variation for functions $f \in \mathscr{C}^1_{\mathbb{W}}$. Finally, in Section 8 we prove our main theorem. An Appendix is added as a reference for a few equalities that are applied all over the paper.

The author thanks his advisor Francesco Serra Cassano for many fruitful discussions and Katrin Fässler for her comments and corrections on a draft of this paper.

## 2. Preliminaries

2.1. **The Heisenberg group.** The *first Heisenberg group* $\mathbb{H}$ is the connected, simply connected Lie group associated to the Heisenberg Lie algebra $\mathfrak{h}$. The *Heisenberg Lie algebra* $\mathfrak{h}$ is the only three-dimensional nilpotent Lie algebra that is not commutative. It can be proven that, for any two linearly independent vectors $A, B \in \mathfrak{h} \setminus [\mathfrak{h}, \mathfrak{h}]$, the triple $(A, B, [A, B])$ is a basis of $\mathfrak{h}$ and $[A, [A, B]] = [B, [A, B]] = 0$. The Heisenberg group has the structure of a *stratified Lie group*, i.e., $\mathfrak{h} = \operatorname{span}\{A, B\} \oplus \operatorname{span}\{[A, B]\}$, see [13, 14].

We then identify $\mathbb{H} = (\operatorname{span}\{A, B, [A, B]\}, *)$, where

$$p * q := p + q + \frac{1}{2}[p, q].$$

In the coordinates $(x, y, z) = xA + yB + z[A, B]$, which are the exponential coordinates of first kind, we have

$$(a, b, c) * (x, y, z) = (a + x, b + y, c + z + \frac{1}{2}(ay - bx)).$$

The inverse is $(x, y, z)^{-1} = (-x, -y, -z)$.

The elements $A, B, [A, B] \in \mathfrak{h}$ induce a frame of left-invariant vector fields on $\mathbb{H}$:

$$X := \partial_x - \frac{1}{2} y \partial_z, \qquad Y := \partial_y + \frac{1}{2} x \partial_z, \qquad Z := \partial_z.$$

The *horizontal subbundle* is the vector bundle

$$H := \bigsqcup_{p \in \mathbb{H}} \operatorname{span}\{X(p), Y(p)\} \subset T\mathbb{H}.$$

The maps $\delta_\lambda(x, y, z) := (\lambda x, \lambda y, \lambda^2 z)$, $\lambda > 0$, are called *dilations*. They are group automorphisms of $\mathbb{H}$ and for all $\lambda, \mu > 0$ it holds $\delta_\lambda \circ \delta_\mu = \delta_{\lambda\mu}$.

2.2. **Intrinsic graphs and intrinsic differentials.** A *vertical plane* is a plane containing the $z$-axis. Explicitly, for $\theta \in \mathbb{R}$,

$$\mathbb{W}_\theta := \{(\eta \sin\theta, \eta \cos\theta, \tau) : \eta, \tau \in \mathbb{R}\} \subset \mathbb{H}.$$

Vertical planes are the only 2-dimensional subgroups of $\mathbb{H}$ that are $\delta_\lambda$-homogeneous, i.e., $\delta_\lambda(\mathbb{W}_\theta) = \mathbb{W}_\theta$ for all $\lambda > 0$.

The *intrinsic $X$-graph* (or simply *intrinsic graph*) of a function $f : \mathbb{R}^2 \to \mathbb{R}$ is the set[1]

$$\begin{aligned} \Gamma_f &:= \left\{(0, \eta, \tau) * (f(\eta, \tau), 0, 0) : \eta, \tau \in \mathbb{R}^2\right\} \\ &= \left\{(f(\eta, \tau), \eta, \tau - \frac{1}{2}\eta f(\eta, \tau)) : \eta, \tau \in \mathbb{R}^2\right\}. \end{aligned}$$

If one look at $f$ as a function $\mathbb{W}_0 \to \operatorname{span}\{A\}$, then $\Gamma_f = \{p * f(p) : p \in \mathbb{W}_0\}$. Left translations and dilations of an intrinsic graph are also intrinsic graphs. For $\alpha \in \mathbb{R}$, the vertical plane $\mathbb{W}_{\arctan(\alpha)}$ is the intrinsic graph of the function $f(\eta, \tau) = \alpha\eta$. We will use the map $\pi_X : \mathbb{H} \to \mathbb{R}^2$, $\pi_X(x, y, z) = (y, z + \frac{1}{2}xy)$. Note that $\pi_X(p * f(p)) = p$.

For $(\eta_0, \tau_0) \in \mathbb{R}^2$ and $f : \mathbb{R}^2 \to \mathbb{R}$ continuous, set $f_0 := f(\eta_0, \tau_0)$ and $p_0 := (0, \eta_0, \tau_0) * (f_0, 0, 0) = (f_0, \eta_0, \tau_0 - \frac{1}{2}\eta_0 f_0)$. We say that $f$ is *intrinsically $\mathscr{C}^1$*, or *belonging to $\mathscr{C}^1_{\mathbb{W}}$*, with differential $\psi : \mathbb{R}^2 \to \mathbb{R}$, if $\delta_\lambda(p_0^{-1}\Gamma_f)$ converge to $\mathbb{W}_{\arctan(\psi(\eta_0, \tau_0))}$ in the sense of the local Hausdorff convergence of sets as $\lambda \to \infty$, and the convergence is uniform on compact sets in $(\eta_0, \tau_0)$.

[1] In a different choice of coordinates in $\mathbb{H}$, we can have $(0, \eta, \tau) * (f(\eta, \tau), 0, 0) = (f(\eta, \tau), \eta, \tau)$. For instance, we will use these coordinates in Section 5.2

Notice that $\delta_\lambda(p_0^{-1}\Gamma_f) = \Gamma_{f_{(\eta_0,\tau_0);\lambda}}$, where

$$f_{(\eta_0,\tau_0);\lambda}(\eta,\tau) = \lambda\left(-f_0 + f(\eta_0 + \frac{\eta}{\lambda}, \tau_0 + f_0\frac{\eta}{\lambda} + \frac{\tau}{\lambda^2})\right).$$

Therefore, $f$ belongs to $\mathscr{C}^1_{\mathbb{W}}$ with differential $\psi$ if and only if $f_{(\eta_0,\tau_0);\lambda}$ converge uniformly on compact sets to the function $(\eta,\tau)\mapsto\psi(\eta_0,\tau_0)\eta$, as $\lambda\to+\infty$, and the convergence is uniform on compact sets in $(\eta_0,\tau_0)$. Notice that $\psi$ has to be continuous.

The *intrinsic gradient* of a function $f:\mathbb{R}^2\to\mathbb{R}$ is the vector field on $\mathbb{R}^2$ defined as

$$\nabla^f := \partial_\eta + f\partial_\tau.$$

We can express the intrinsic differentiability in terms of the differentiability of $f$ along the integral curves of $\nabla^f$: from [18, Theorem 4.95] we obtain the following characterisation, which justify the notation $\nabla^f f$ for the differential $\psi$ of $f\in\mathscr{C}^1_{\mathbb{W}}$.

**Lemma 2.1.** *A continuous function $f:\mathbb{R}^2\to\mathbb{R}$ is in $\mathscr{C}^1_{\mathbb{W}}$ with differential $\psi$ if and only if for every $p\in\mathbb{R}^2$ there exists a $\mathscr{C}^2$-function $g_p: I\to\mathbb{R}$, where $I\subset\mathbb{R}$ is a neighbourhood of $0$, such that*

$$\begin{cases} g_p(0) = 0, \\ g_p'(t) = f(p+(t,g_p(t))) & \forall t\in I, \\ g_p''(t) = \psi(p+(t,g_p(t))) & \forall t\in I. \end{cases}$$

Note that $t\mapsto p+(t,g_p(t))$ is an integral curve of $\nabla^f$ and that $g_p$ is not unique in general. Another interpretation of these curves will be useful:

**Lemma 2.2.** *Let $f\in\mathscr{C}^1_{\mathbb{W}}$. A curve $\gamma: I\to\mathbb{R}^2$ of class $\mathscr{C}^1$, where $I\subset\mathbb{R}$ is an interval, is an integral curve of $\nabla^f$ if and only if the curve $t\mapsto\gamma(t)*f(\gamma(t))\in\Gamma_f$ is a curve of class $\mathscr{C}^1$ tangent to the horizontal bundle $H$.*

**Remark 2.3.** In [11] it has been shown that there exists $f\in\mathscr{C}^1_{\mathbb{W}}$ whose intrinsic graph $\Gamma_f$ has Euclidean Hausdorff dimension (seen as a subset of the Euclidean $\mathbb{R}^3$) strictly larger than two. It is possible to prove, for example using Lemma 5.4, that $\Gamma_{f+1}$ does not have locally finite intrinsic perimeter and in particular $f+1\notin\mathscr{C}^1_{\mathbb{W}}$. This shows that $\mathscr{C}^1_{\mathbb{W}}$ is not a vector space.

2.3. **Smooth approximation.** A sequence $\{f_k\}_{k\in\mathbb{N}}\subset\mathscr{C}^1_{\mathbb{W}}$ converges to $f$ in $\mathscr{C}^1_{\mathbb{W}}$ if $f_k$ and $\nabla^{f_k}f_k$ converge to $f$ and $\nabla^f f$ uniformly on compact sets. The following lemma has been proven in [1].

**Lemma 2.4.** *If $f\in\mathscr{C}^1_{\mathbb{W}}$ then there is a sequence of functions $\{f_k\}_{k\in\mathbb{N}}\subset\mathscr{C}^\infty(\mathbb{R}^2)$ that converges to $f$ in $\mathscr{C}^1_{\mathbb{W}}$.*

2.4. **Perimeter and Bernstein's Problem.** The Lebesgue measure $\mathscr{L}^3$ in $\mathbb{R}^3$ is a Haar measure on $\mathbb{H}$ in the exponential coordinates introduced Section 2.1. Notice that for any measurable set $E\subset\mathbb{H}^1$ and any $\lambda>0$ it holds $\mathscr{L}^3(\delta_\lambda(E)) = \lambda^4\mathscr{L}^3(E)$.

Let $\langle\cdot,\cdot\rangle$ be the left-invariant scalar product on the subbundle $H$ such that $(X,Y)$ is an orthonormal frame and set $\|v\| := \sqrt{\langle v,v\rangle}$ for $v\in H$. The *sub-Riemannian perimeter* of a measurable set $E\subset\mathbb{H}^1$ in an open set $\Omega$ is

$$P(E;\Omega) := \sup\left\{\int_E \mathrm{div}V\,\mathrm{d}\mathscr{L}^3 :\ V\in\Gamma(H),\ \mathrm{spt}(V)\subset\subset\Omega,\ \|V\|\le 1\right\},$$

where $\Gamma(H)$ contains all the smooth sections of the horizontal subbundle and $\mathrm{div}V$ is the divergence of vector fields on $\mathbb{R}^3$. One can show that, for every $V_1, V_2 \in \mathscr{C}^\infty(\mathbb{R}^3)$,

$$\mathrm{div}(V_1 X + V_2 Y) = XV_1 + YV_2.$$

A set $E$ has *locally finite perimeter* if $P(E;\Omega) < \infty$ for all $\Omega \subset \mathbb{H}$ open and bounded. If $E$ has locally finite perimeter, the function $\Omega \mapsto P(E;\Omega)$ induces a Radon measure $|\partial E|$ on $\mathbb{H}^1$, which is concentrated on the so-called *reduced boundary* $\partial^* E \subset \partial E$. Moreover, up to a set of $|\partial E|$-measure zero and a rotation around the $z$-axis, $\partial^* E$ is the countable union of intrinsic graphs of $\mathscr{C}^1_{\mathbb{W}}$ functions. See [6] and [7] for further reading.

A measurable set $E$ has *minimal perimeter* if, for every bounded open set $\Omega \subset \mathbb{H}^1$ and every measurable set $F \subset \mathbb{H}^1$ with symmetric difference $E \Delta F \subset\subset \Omega$, we have

$$P(E;\Omega) \le P(F;\Omega).$$

In this case, the reduced boundary $\partial^* E$ of $E$ is called *area-minimizing surface*. We are interested in area minimizers that are global intrinsic graphs.

**Conjecture 2.5** (Bernstein's Problem)**.** *If $f \in \mathscr{C}^1_{\mathbb{W}}$ is such that $\Gamma_f$ is an area-minimizing surface, then $\Gamma_f$ is a vertical plane up to left-translations.*

Such conjecture has been proven in the case $f \in \mathscr{C}^1(\mathbb{R}^2)$ in [9], while it has been presented a counterexample in [17] with $f \in \mathscr{C}^0(\mathbb{R}^2) \setminus \mathscr{C}^1_{\mathbb{W}}$.

For an open domain $\omega \subset \mathbb{R}^2$, set

$$\Omega_\omega := \{(0,\eta,\tau) * (\xi,0,0) : (\eta,\tau) \in \omega,\ \xi \in \mathbb{R}\}.$$

If $f \in \mathscr{C}^1_{\mathbb{W}}$ and $E_f = \{(0,\eta,\tau) * (\xi,0,0) \in \mathbb{R}^2,\ \xi \le f(\eta,\tau)\}$, then

$$P(E_f;\Omega_\omega) = \int_\omega \sqrt{1 + (\nabla^f f)^2}\, \mathrm{d}\eta\, \mathrm{d}\tau.$$

If $E_f$ has minimal perimeter, then, for every $g \in \mathscr{C}^1_{\mathbb{W}}$ with $\{f \ne g\} \subset\subset \omega$, it holds

$$\int_\omega \sqrt{1 + (\nabla^f f)^2}\, \mathrm{d}\eta\, \mathrm{d}\tau \le \int_\omega \sqrt{1 + (\nabla^g g)^2}\, \mathrm{d}\eta\, \mathrm{d}\tau.$$

It is not known whether the converse implication holds.

## 3. Lagrangian solutions to $\Delta^f f = 0$

For $f \in \mathscr{C}^1_{\mathbb{W}}$ and $v \in \mathscr{C}^2(\mathbb{R}^2)$, we define the differential operator

$$\Delta^f v := \partial_\eta^2 v + 2f\partial_\eta\partial_\tau v + f^2\partial_\tau^2 v + \nabla^f f \partial_\tau v. \tag{2}$$

Notice that, if $f \in \mathscr{C}^2(\mathbb{R}^2)$, then

$$\Delta^f v = \nabla^f(\nabla^f v).$$

The next lemma will be a fundamental tool for extending some results beyond the smooth case via approximation. The proof trivially follows from the explicit expressions of the differential operators $\nabla^f$ and $\Delta^f$.

**Lemma 3.1.** *If $\{f_k\}_{k\in\mathbb{N}} \subset \mathscr{C}^1_{\mathbb{W}}$ and $\{v_k\}_{k\in\mathbb{N}} \subset \mathscr{C}^2(\mathbb{R}^2)$ are sequences converging to $f$ and $v$ in their respective spaces, then the sequences $\{\nabla^{f_k} v_k\}_{k\in\mathbb{N}}$ and $\{\Delta^{f_k} v_k\}_{k\in\mathbb{N}}$ converge to $\nabla^f v$ and $\Delta^f v$ uniformly on compact sets.*

If $f \in \mathscr{C}^2(\mathbb{R}^2)$ is such that $\Gamma_f$ is a minimal surface in $\mathbb{H}$, then one shows that $f$ satisfies the differential equation (see [2])

$$\nabla^f \left( \frac{\nabla^f f}{\sqrt{1 + (\nabla^f f)^2}} \right) = 0. \tag{3}$$

Equation (3) is equivalent, for $f \in \mathscr{C}^2(\mathbb{R}^2)$, to

$$\Delta^f f = 0. \tag{4}$$

For a generic $f \in \mathscr{C}^1_{\mathbb{W}}$, equation (4) has not the classical interpretation (2). However, using a "Lagrangian interpretation" of $\nabla^f(\nabla^f f) = 0$, we give the following definition:

**Definition 3.2.** A function $f \in \mathscr{C}^1_{\mathbb{W}}$ satisfies $\Delta^f f = 0$ *in weak Lagrangian sense*, if for every $p \in \mathbb{R}^2$ there is an integral curve $\gamma$ of $\nabla^f$ passing through $p$ such that $\nabla^f f$ is constant along $\gamma$.

If $f \in \mathscr{C}^2(\mathbb{R}^2)$ then $\Delta^f f = \nabla^f(\nabla^f f) = 0$ if and only if $\nabla^f f$ is constant along *all* integral curves of $\nabla^f$, i.e., $\Delta^f f = 0$ holds in a *strong* Lagrangian sense, see Remark 3.7.

Lemma 3.5 will characterize such functions by the integral curves of $\nabla^f$.

**Lemma 3.3.** *Let $A, B \in \mathscr{C}^0(\mathbb{R})$. The map $\mathbb{R}^2 \to \mathbb{R}^2$ given by*

$$G : (t, \zeta) \mapsto \left( t, \frac{A(\zeta)}{2} t^2 + B(\zeta) t + \zeta \right)$$

*is a homeomorphism if and only if*

(1) *For all $\zeta, \zeta' \in \mathbb{R}$*
 (1a) *either $A(\zeta) = A(\zeta')$ and $B(\zeta) = B(\zeta')$,*
 (1b) *or $2\big(A(\zeta) - A(\zeta')\big)(\zeta - \zeta') > \big(B(\zeta) - B(\zeta')\big)^2$.*

(2) *If there exists $\zeta_0 \in \mathbb{R}$ such that $A(\zeta_0) > 0$, then*

$$\limsup_{\zeta \to \infty} \left( \zeta - \frac{B(\zeta)^2}{2A(\zeta)} \right) = +\infty.$$

(3) *If there exists $\zeta_0 \in \mathbb{R}$ such that $A(\zeta_0) < 0$, then*

$$\liminf_{\zeta \to -\infty} \left( \zeta - \frac{B(\zeta)^2}{2A(\zeta)} \right) = -\infty.$$

*Proof.* Define $g(t, \zeta) = \frac{A(\zeta)}{2} t^2 + B(\zeta) t + \zeta$, so that $G(t, \zeta) = (t, g(t, \zeta))$. We first show that $G$ is injective if and only if property (1) holds. For $\zeta' > \zeta$, define the quadratic polynomial

$$Q_{\zeta',\zeta}(t) = g(t, \zeta') - g(t, \zeta) = \frac{A(\zeta') - A(\zeta)}{2} t^2 + (B(\zeta') - B(\zeta)) t + (\zeta' - \zeta).$$

The map $G$ is injective if and only if for all $\zeta', \zeta \in \mathbb{R}$ with $\zeta' > \zeta$ the polynomial $Q_{\zeta',\zeta}$ has no zeros. If $A(\zeta') = A(\zeta)$, then $Q_{\zeta',\zeta}$ is in fact linear, thus it has no zeros if and only if $B(\zeta') = B(\zeta)$ and we obtain property (1a). If $A(\zeta') \neq A(\zeta)$, then $Q_{\zeta',\zeta}$ has no zeros if and only if its discriminant is strictly negative, i.e., property (1b) holds.

Next, we assume that $G$ is injective, i.e., that property (1) holds, and we will show that $G$ is surjective if and only if properties (2) and (3) hold. By the Invariance of Domain Theorem, the fact that $G$ is surjective is equivalent to $G$ being a homeomorphism. Notice that, since $Q_{\zeta',\zeta}(0) = \zeta' - \zeta > 0$ for all $\zeta' > \zeta$, we have

$$\zeta' > \zeta \quad \Rightarrow \quad \forall t \in \mathbb{R} \quad g(t,\zeta') > g(t,\zeta). \tag{5}$$

Suppose that $G$ is surjective, hence a homeomorphism. Suppose $\zeta_0 \in \mathbb{R}$ is such that $A(\zeta_0) > 0$. By (1) we have that $A$ is monotone increasing, therefore $A(\zeta) > 0$ for all $\zeta \geq \zeta_0$. It follows that if $\zeta \geq \zeta_0$ then

$$\zeta - \frac{B(\zeta)^2}{2A(\zeta)} = \inf_{t\in\mathbb{R}} g(t,\zeta).$$

For $M \in \mathbb{R}$ define $K_M = \{(\eta,\tau) \in \mathbb{R}^2 : g(\eta,\zeta_0) \leq \tau \leq M\}$. Since $A(\zeta_0) > 0$, the set $K_M$ is compact (possibly empty) for all $M \in \mathbb{R}$. Next, for $\zeta \in \mathbb{R}$ define $U_\zeta = G(\mathbb{R} \times (-\infty,\zeta)) = \{(\eta,\tau) : \tau < g(\eta,\zeta)\}$. Since $G$ is surjective, the open sets $U_\zeta$ cover $\mathbb{R}^2$. Hence, there is $\zeta_1 \geq \zeta_0$ such that $K_M \subset U_{\zeta_1}$. Using (5), we obtain

$$\forall \zeta \geq \zeta_1 \quad \inf_{t\in\mathbb{R}} g(t,\zeta) \geq M.$$

Since $M$ is arbitrary, we have proven (2). Property (3) is proven with a similar argument.

Now we prove the converse implication. Suppose that $A$ and $B$ satisfy properties (2) and (3). In order to prove that $G$ is surjective, we need only to prove that $\lim_{\zeta\to\infty} g(t,\zeta) = +\infty$ and $\lim_{\zeta\to-\infty} g(t,\zeta) = -\infty$, for every $t \in \mathbb{R}$.

If $A(\zeta) = A(0)$ for all $\zeta \geq 0$, then $g(t,\zeta) = g(t,0) + \zeta$ and therefore $\lim_{\zeta\to\infty} g(t,\zeta) = +\infty$. If $A(\zeta) \leq 0$ for all $\zeta \in \mathbb{R}$, then there is $C > 0$ such that $0 \leq A(\zeta) - A(0) \leq C$ for all $\zeta > 0$. We may suppose $A(\zeta) > A(0)$ for $\zeta$ large enough. Thus, using (1$b$),

$$\begin{aligned}
g(t,\zeta) &\geq \frac{A(0)}{2}t^2 + B(0)t + \zeta + (B(\zeta) - B(0))t \\
&\geq \frac{A(0)}{2}t^2 + B(0)t + \zeta - |t|\sqrt{2(A(\zeta) - A(0))\zeta} \\
&\geq \frac{A(0)}{2}t^2 + B(0)t + \zeta - |t|\sqrt{2C}\sqrt{\zeta}.
\end{aligned}$$

The limit $\lim_{\zeta\to\infty} g(t,\zeta) = +\infty$ follows. Finally, if $A(\zeta_0) > 0$ for some $\zeta_0 \in \mathbb{R}$, then for all $\zeta \geq \zeta_0$ we have $\inf_{t\in\mathbb{R}} g(t,\zeta) = \zeta - \frac{B(\zeta)^2}{2A(\zeta)}$. Property (2) implies that $\lim_{\zeta\to\infty} g(t,\zeta) = +\infty$.

The limit $\lim_{\zeta\to-\infty} g(t,\zeta) = -\infty$ is deduced similarly from (3). □

**Remark 3.4.** If $A, B \in \mathscr{C}(\mathbb{R})$ satisfy properties (1), (2) and (3) of the previous Lemma 3.3, then the function $f$ defined by $f(G(t,\zeta)) = \partial_t g(t,\zeta) = A(\zeta)t + B(\zeta)$ belongs to $\mathscr{C}^1_{\mathbb{W}}$ by Lemma 2.1. Moreover, the curves $t \mapsto g(t,\zeta)$ are integral curves of $\nabla^f$ along which $\nabla^f f(G(t,\zeta)) = \partial_t^2 g(t,\zeta) = A(\zeta)$ is constant. So, $\Delta^f f = 0$ in weak Lagrangian sense. The graphs of these functions are examples of "graphical strips" as introduced in [4]. For example, for any $A \in \mathscr{C}^0(\mathbb{R})$ non-decreasing, we can define $g(t,\zeta) := A(\zeta)t^2 + \zeta$ and we obtain a well defined $f \in \mathscr{C}^1_{\mathbb{W}}$ with $\Delta^f f = 0$ given by

$$f(t, A(\zeta)t^2 + \zeta) = 2A(\zeta)t.$$

The converse also holds, as the next lemma shows.

**Lemma 3.5.** *Let $f \in \mathscr{C}^1_{\mathbb{W}}$ satisfying $\Delta^f f = 0$ in weak Lagrangian sense. Then the curves $t \mapsto (t, g(t,\zeta))$, where $\zeta \in \mathbb{R}$ and*

$$g(t,\zeta) = \frac{\nabla^f f(0,\zeta)}{2} t^2 + f(0,\zeta)t + \zeta, \tag{6}$$

*are the integral curves of $\nabla^f$ along which $\nabla^f f$ is constant. Moreover, the functions $\zeta \mapsto \nabla^f f(0,\zeta)$ and $\zeta \mapsto f(0,\zeta)$ satisfy the conditions (1), (2) and (3) in Lemma 3.3. In particular, $\tau \mapsto \nabla^f f(\eta,\tau)$ is non-decreasing, for all $\eta \in \mathbb{R}$.*

*Proof of Lemma 3.5.* Given a function $g_p : I \to \mathbb{R}$ like in Lemma 2.1 along which $\nabla^f f$ is constant, we have $g_p''(t) = \nabla^f f(p)$ for all $t \in I$, i.e., $g_p$ is a polynomial of second degree. Moreover, such a $g_p$ is unique for every $p$, because it is completely determined by $f(p)$ and $\nabla^f f(p)$.

It follows that $g_p$ is defined on $\mathbb{R}$. Indeed, suppose $I = (a,b)$ and set $q = \lim_{t\to b} p + (t, g_p(t))$, which exists because $g_p$ is a polynomial. If $g_q : J \to \mathbb{R}$ is a function like in Lemma 2.1 along which $\nabla^f f$ is constant, then $g_q$ is uniquely determined by $f(q)$ and $\nabla^f f(q)$, where

$$f(q) = \lim_{t\to b} f(p + (t, g_p(t))) = \lim_{t\to b} g_p'(t),$$
$$\nabla^f f(q) = \lim_{t\to b} \nabla^f f(p + (t, g_p(t))) = \lim_{t\to b} g_p''(t).$$

Hence, $g_q(t) = g_p(b+t)$ for $t < 0$ and so $g_p$ can be extended beyond $b$. Similarly, we can extend $g_p$ to values below $a$.

If we consider $p = (0,\zeta)$, then $g_p(t) = g(t,\zeta)$, where $g(t,\zeta)$ is given in (6). If $p \in \mathbb{R}^2$, then the curve $t \mapsto p + (t, g_p(t))$ intersects the axis $\{0\} \times \mathbb{R}$ at some point, and thus $g_p$ is of the form described in (6) up to a change of variables in $t$. We conclude that the map $(t,\zeta) \mapsto (t, g(t,\zeta))$ is a homeomorphism. Therefore, the conditions stated in Lemma 3.3 hold true.

Finally, since $(f(0,\zeta) - f(0,\zeta'))^2 \geq 0$, then $\zeta \mapsto \nabla^f f(0,\zeta)$ is non-decreasing. Since $\nabla^f f(t, g(t,\zeta)) = \nabla^f f(0,\zeta)$ and since, for $t \in \mathbb{R}$ fixed, the map $\zeta \mapsto g(t,\zeta)$ is a ordering-preserving homeomorphism $\mathbb{R} \to \mathbb{R}$, then the map $\tau \mapsto \nabla^f f(\eta,\tau)$ is non-decreasing as well, for all $\eta \in \mathbb{R}$. □

**Remark 3.6.** Lemma 3.5 states in particular that, if $\Delta^f f = 0$ in weak Lagrangian sense then $\Gamma_f$ is foliated by horizontal straight lines. Indeed, notice that any parabola $t \mapsto g(t,\zeta)$ in $\mathbb{R}^2$ lifts to a straight line in $\Gamma_f$. In [9, Theorem 3.5] Galli and Ritoré are able to prove that, if $f \in \mathscr{C}^1(\mathbb{R}^2)$ and if $\Gamma_f$ is a minimal surface in $\mathbb{H}$, then $\Gamma_f$ is foliated by horizontal straight lines, i.e., $\Delta^f f = 0$ holds in weak Lagrangian sense.

**Remark 3.7.** One may wonder wether Definition 3.2 for weak Lagrangian solutions to $\Delta^f f = 0$ is equivalent to a stronger condition, namely that $\nabla^f f$ is constant along *all* integral curves of $\nabla^f$. This is the case when $f \in C^1(\mathbb{R}^2)$, because integral curves are unique at each point. The following example shows that strong and weak conditions are not equivalent. Indeed, there are functions for which the curves $t \mapsto (t, g(t,\zeta))$ described in Lemma 3.5 do not exhaust all the integral curves of $\nabla^f$.

Let $h \in \mathscr{C}^2(\mathbb{R})$ and define $k : \mathbb{R}^2 \to \mathbb{R}$ by requiring that for each $s \in \mathbb{R}$ the function $t \mapsto k(t,s)$ is the unique polynomial of second degree with $k(s,s) = h(s)$, $\partial_t k(s,s) = h'(s)$ and $\partial_t^2 k(s,s) = h''(s)$. Explicitly, we have

$$k(t,s) = \frac{h''(s)}{2}t^2 + (h'(s) - h''(s)s)t + h(s) - h'(s)s + \frac{h''(s)}{2}s^2.$$

If the map $K(t,s) = (t, k(t,s))$ is a homeomorphism $\mathbb{R}^2 \to \mathbb{R}^2$, then we may define a function $f \in \mathscr{C}^1_{\mathbb{W}}$ by $f(K(t,s)) = \partial_t k(t,s)$, as we did in Remark 3.4. Then $t \mapsto K(t,s)$ are integral curves of $\nabla^f$ and $\nabla^f f(K(t,s)) = \partial_t^2 k(t,s) = h''(s)$. It follows that $\Delta^f f = 0$ holds in weak Lagrangian sense. However, $s \mapsto K(s,s) = (s, h(s))$ is an integral curve of $\nabla^f$, because $f(K(s,s)) = h'(s)$. Since $\nabla^f f(K(s,s)) = h''(s)$, there is no need for $\nabla^f f$ to be constant along this curve.

As an example, consider $h(s) = s^3$, for which we have $k(t,s) = 3st^2 - 3s^2t + s^3$. We show that the map $K$ is in this case a homeomorphism. Define $\zeta(s) = s^3$, $A(\zeta(s)) = 6s = 6\zeta^{1/3}$, $B(\zeta(s)) = -3s^2 = -3\zeta^{2/3}$ and the functions $g(t,\zeta)$ and $G(t,\zeta)$ as in Lemma 3.3. Since $K(t,s) = G(t,\zeta(s))$ and since $\zeta(\cdot)$ is a homeomorphism $\mathbb{R} \to \mathbb{R}$, we need only to show that $G$ is a homeomorphism $\mathbb{R}^2 \to \mathbb{R}^2$, i.e., that the functions $A$ and $B$ satisfy all conditions of Lemma 3.3:

(1): Let $\zeta = \zeta(s), \zeta' = \zeta(s') \in \mathbb{R}$. If $A(\zeta) = A(\zeta')$, then $s = s'$ and thus $B(\zeta) = B(\zeta')$. If instead $A(\zeta) \neq A(\zeta')$, then $s \neq s'$ and thus

$$\begin{aligned} &2(A(\zeta) - A(\zeta'))(\zeta - \zeta') - (B(\zeta) - B(\zeta'))^2 \\ &\qquad = 2(6s - 6s')(s^3 - s'^3) - 9(s'^2 - s^2)^2 = 3(s - s')^4 > 0. \end{aligned}$$

(2)&(3): Since $\zeta - \frac{B(\zeta)^2}{2A(\zeta)} = \frac{1}{4}s^3$ and since $\zeta \to \pm\infty$ if and only if $s \to \pm\infty$, then $\lim_{\zeta\to+\infty} \zeta - \frac{B(\zeta)^2}{2A(\zeta)} = \lim_{s\to+\infty} \frac{1}{4}s^3 = +\infty$ and $\lim_{\zeta\to-\infty} \zeta - \frac{B(\zeta)^2}{2A(\zeta)} = \lim_{s\to-\infty} \frac{1}{4}s^3 = -\infty$.

The function $f$ can be explicitly computed as $f(\eta,\tau) = 3\eta^2 - 3(\tau - \eta^3)^{2/3}$. Finally, as we noticed before, $s \mapsto (s, s^3)$ is an integral curve of $\nabla^f$ and $\nabla^f f(s, s^3) = 6s$ is not constant.

**Lemma 3.8.** *Let $f \in \mathscr{C}^1_{\mathbb{W}}$. If $\Delta^f f = 0$ in weak Lagrangian sense, then there is a sequence $\{f_k\}_{k\in\mathbb{N}} \subset \mathscr{C}^\infty(\mathbb{R}^2)$ converging to $f$ in $\mathscr{C}^1_{\mathbb{W}}$ such that $\Delta^{f_k} f_k = 0$ for all $k \in \mathbb{N}$.*

*Proof.* Let $\{\rho_\epsilon\}_{\epsilon>0} \subset \mathscr{C}^\infty(\mathbb{R})$ be a family of mollifiers with $\mathrm{spt}(\rho_\epsilon) \subset [-\epsilon, \epsilon]$, $\rho_\epsilon \geq 0$, $\rho_\epsilon(0) > 0$ and $\int_{\mathbb{R}} \rho_\epsilon(r)\,\mathrm{d}r = 1$. Fix $f \in \mathscr{C}^1_{\mathbb{W}}$ with $\Delta^f f = 0$. Set $A(\zeta) := \nabla^f f(0,\zeta)$ and $B(\zeta) := f(0,\zeta)$. Define

$$\begin{aligned} A_\epsilon(\zeta) &:= \int_{\mathbb{R}} \nabla^f f(0, \zeta - r)\rho_\epsilon(r)\,\mathrm{d}r, \\ B_\epsilon(\zeta) &:= \int_{\mathbb{R}} f(0, \zeta - r)\rho_\epsilon(r)\,\mathrm{d}r, \\ g_\epsilon(t,\zeta) &:= \frac{A_\epsilon(\zeta)}{2}t^2 + B_\epsilon(\zeta)t + \zeta. \end{aligned}$$

We claim that, for all $\epsilon > 0$, all conditions stated in Lemma 3.3 hold for $A_\epsilon$ and $B_\epsilon$. Let $\zeta, \zeta' \in \mathbb{R}$ with $\zeta < \zeta'$. First, suppose that $A_\epsilon(\zeta) = A_\epsilon(\zeta')$.

Notice that $A(\zeta-r)-A(\zeta'-r)\leq 0$ for all $r\in\mathbb{R}$, because $A$ is non-decreasing. Thus, we deduce from

$$0 = A_\epsilon(\zeta) - A_\epsilon(\zeta') = \int_{\mathbb{R}} (A(\zeta-r) - A(\zeta'-r))\rho_\epsilon(r)\,\mathrm{d}r$$

that $(B(\zeta-r)-B(\zeta'-r))\rho_\epsilon(r)=0$ for all $r\in\mathbb{R}$ and therefore that $B_\epsilon(\zeta)=B_\epsilon(\zeta')$, i.e., (1a) holds

Second, suppose that $A_\epsilon(\zeta)\neq A_\epsilon(\zeta')$. Using Jensen's inequality, we have

$$\begin{aligned}
&2\big(A_\epsilon(\zeta) - A_\epsilon(\zeta')\big)(\zeta-\zeta')\\
&\qquad = \int_{\mathbb{R}} 2\big(A(\zeta-r)-A(\zeta'-r)\big)\big((\zeta-r)-(\zeta'-r)\big)\rho_\epsilon(r)\,\mathrm{d}r\\
&> \int_{\mathbb{R}} \big(B(\zeta-r)-B(\zeta'-r)\big)^2\rho_\epsilon(r)\,\mathrm{d}r \geq \left(\int_{\mathbb{R}} (B(\zeta-r)-B(\zeta'-r))\rho_\epsilon(r)\,\mathrm{d}r\right)^2.
\end{aligned}$$

So, condition (1b) is also verified.

Suppose that $A_\epsilon(\zeta_0)>0$ for some $\zeta_0\in\mathbb{R}$. By the monotonicity of $A$ and the positivity of $\rho_\epsilon$, we may assume $A(\zeta_0)>0$. Let $M>0$. Since property (2) of Lemma 3.3 holds for $A$, there is $\zeta_1>\zeta_0$ so that for all $\zeta>\zeta_1$

$$M < \zeta - \frac{B(\zeta)^2}{2A(\zeta)} = \frac{2A(\zeta)\zeta - B(\zeta)^2}{2A(\zeta)}.$$

Using Jensen inequality, we have for all $\zeta>\zeta_1+\epsilon$

$$\begin{aligned}
2A_\epsilon(\zeta)\zeta - B_\epsilon(\zeta)^2 &\geq \int_{\mathbb{R}} \left(2\zeta A(\zeta-r) - B(\zeta-r)^2\right)\rho_\epsilon(r)\,\mathrm{d}r\\
&= 2\int_{\mathbb{R}} A(\zeta-r) r\rho_\epsilon(r)\,\mathrm{d}r + \int_{\mathbb{R}} \left(2(\zeta-r)A(\zeta-r) - B(\zeta-r)^2\right)\rho_\epsilon(r)\,\mathrm{d}r\\
&\geq -2\epsilon\int_{\mathbb{R}} A(\zeta-r)\rho_\epsilon(r)\,\mathrm{d}r + 2M\int_{\mathbb{R}} A(\zeta-r)\rho_\epsilon(r)\,\mathrm{d}r = 2A_\epsilon(\zeta)(M-\epsilon).
\end{aligned}$$

Thus, $M-\epsilon<\zeta-\frac{B_\epsilon(\zeta)^2}{2A_\epsilon(\zeta)}$ for all $\zeta>\zeta_1+\epsilon$. Since $M$ was arbitrary, we obtain property (2) of Lemma 3.3. Property (3) can be similarly obtained.

The functions $G_\epsilon:\mathbb{R}^2\to\mathbb{R}^2$, $G_\epsilon(t,\zeta):=(t,g_\epsilon(t,\zeta))$, are homeomorphisms and, as $\epsilon\to 0$, they converge to $G_0$ uniformly on compact sets. It follows that $G_\epsilon^{-1}$ also converge to $G_0^{-1}$, as $\epsilon\to 0$.

For $\epsilon>0$, define $f_\epsilon\in\mathscr{C}^\infty(\mathbb{R}^2)$ via

$$f_\epsilon(t,g_\epsilon(t,\zeta)) = A_\epsilon(\zeta)t + B_\epsilon(\zeta).$$

By the continuity of $G_\epsilon$ and $G_\epsilon^{-1}$ in $\epsilon$, $f_\epsilon$ and $\nabla^{f_\epsilon}f_\epsilon$ converge to $f_0$ and $\nabla^{f_0}f_0$ uniformly on compact sets. Finally, $\Delta^{f_\epsilon}f_\epsilon=0$ by construction. □

## 4. A Lagrangian approach to contact variations

**Proposition 4.1.** *Let $\phi=(\phi_1,\phi_2):\mathbb{R}^2\to\mathbb{R}^2$ be a $\mathscr{C}^\infty$-diffeomorphism. Let $f\in\mathscr{C}^1_{\mathbb{W}}$ and assume*

$$\nabla^f\phi_1(p)\neq 0 \qquad \forall p\in\mathbb{R}^2. \tag{7}$$

*Define $\bar f:\mathbb{R}^2\to\mathbb{R}$ as*

$$\bar f\circ\phi = \frac{\nabla^f\phi_2}{\nabla^f\phi_1}.$$

*Then* $\bar f \in \mathscr{C}^1_{\mathbb{W}}$ *and*

$$\nabla^{\bar f}\bar f\circ\phi = \frac{\Delta^f\phi_2}{(\nabla^f\phi_1)^2} - \frac{\nabla^f\phi_2}{(\nabla^f\phi_1)^3}\Delta^f\phi_1. \tag{8}$$

Notice that, if $f\in\mathscr{C}^1(\mathbb{R}^2)$, then $\bar f\in\mathscr{C}^1(\mathbb{R}^2)$ as well.

**Remark 4.2.** If $\{\phi^\epsilon\}_{\epsilon>0}$ is a smooth one-parameter family of diffeomorphisms $\phi^\epsilon:\mathbb{R}^2\to\mathbb{R}^2$ with $\phi^0=\mathrm{Id}$, then, for $\epsilon>0$ small enough, the functions $f_\epsilon$ defined by

$$f_\epsilon\circ\phi^\epsilon = \frac{\nabla^f\phi_2^\epsilon}{\nabla^f\phi_1^\epsilon}.$$

belong to $\mathscr{C}^1_{\mathbb{W}}$ and converge to $f$ in $\mathscr{C}^1_{\mathbb{W}}$.

*Proof.* The idea is to transform via $\phi$ the integral curves of $\nabla^f$ into the ones of $\nabla^{\bar f}$. Fix $p=(\eta,\tau)$, let $q:=(\bar\eta,\bar\tau):=\phi(p)$ and let $g_p:I\to\mathbb{R}$ be like in Lemma 2.1. Thanks to the condition $\nabla^f\phi_1\neq 0$ and the Implicit Function Theorem, there exist two $\mathscr{C}^2$-function $s:I\to\mathbb{R}$ and $\bar g_q:s(I)\to\mathbb{R}$, such that

$$q+(s,\bar g_q(s)) = \phi(p+(t,g_p(t))),\qquad \forall t\in I.$$

Therefore

$$\begin{cases} s(t)=\phi_1(\eta+t,\tau+g_p(t))-\bar\eta\\ \bar g_q(s(t))=\phi_2(\eta+t,\tau+g_p(t))-\bar\tau.\end{cases}$$

We define

$$\bar f(q):=\bar g_q'(0).$$

Notice that this value does not depend on the choice of $g_p$, as far as $t\mapsto(t,g_p(t))$ is an integral curve of $\nabla^f$.

We want to write $\bar g_q'(0)$. Set

$$p_t:=(\eta+t,\tau+g_p(t)).$$

First

$$\frac{\mathrm{d}}{\mathrm{d}t}s(t)=\partial_\eta\phi_1(p_t)+\partial_\tau\phi_1(p_t)g_p'(t)=\nabla^f\phi_1(p_t),$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\bar g_q(s(t))=\partial_\eta\phi_2(p_t)+\partial_\tau\phi_2(p_t)g_p'(t)=\nabla^f\phi_2(p_t).$$

Since

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar g_q(s(t))=\bar g_q'(s(t))\cdot\frac{\mathrm{d}}{\mathrm{d}t}s(t),$$

we have for $s=0=t$

$$\bar f(q)=\frac{\nabla^f\phi_2(p)}{\nabla^f\phi_1(p)}.$$

$\nabla^{\bar f}\bar f(q)$ is the derivative of $\bar f$ along the curve $q+(s,\bar g_q(s))$ at $s=0$, i.e.,

$$\nabla^{\bar f}\bar f(q)=\bar g_q''(0).$$

As above, we want to write down $\bar g_q''(0)$ in a more explicit way.

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}s(t)|_{t=0}=\partial_\eta^2\phi_1(p)+\partial_\tau\partial_\eta\phi_1(p)f(p)+ \\ +\partial_\eta\partial_\tau\phi_1(p)f(p)+\partial_\tau^2\phi_1(p)(f(p))^2+\partial_\tau\phi_1(p)\nabla^f f(p)=\Delta^f\phi_1(p).$$

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\bar g_q(s(t))|_{t=0} = \partial_\eta^2\phi_2(p) + \partial_\tau\partial_\eta\phi_2(p)f(p) + \\ + \partial_\eta\partial_\tau\phi_2(p)f(p) + \partial_\tau^2\phi_2(p)(f(p))^2 + \partial_\tau\phi_2(p)\nabla^f f(p) = \Delta^f\phi_2(p)$$

Since

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\bar g_q(s(t)) = \bar g_q''(s(t))\cdot\left(\frac{\mathrm{d}}{\mathrm{d}t}s(t)\right)^2 + \bar g_q'(s(t))\cdot\frac{\mathrm{d}^2}{\mathrm{d}t^2}s(t),$$

we have

$$\nabla^{\bar f}\bar f(q) = \bar g_q''(0) = \frac{\frac{\mathrm{d}^2}{\mathrm{d}t^2}\bar g_q|_{t=0} - \bar g_q'(0)\cdot\frac{\mathrm{d}^2}{\mathrm{d}t^2}s|_{t=0}}{(\frac{\mathrm{d}}{\mathrm{d}t}s|_{t=0})^2} \\ = \frac{1}{(\nabla^f\phi_1(p))^2}\cdot\left(\Delta^f\phi_2(p) - \frac{\nabla^f\phi_2(p)}{\nabla^f\phi_1(p)}\cdot\Delta^f\phi_1(p)\right).$$

By Lemma 2.1, the function $\bar f$ belongs to $\mathscr{C}^1_{\mathbb{W}}$. □

## 5. Contact transformations

A diffeomorphism $\Phi:\mathbb{H}\to\mathbb{H}$ is a *contact diffeomorphism* if $\mathrm{d}\Phi(H)\subset H$, see [3, 12]. Contact diffeomorphisms are the only diffeomorphisms that preserve the sub-Riemannian perimeter.

**Proposition 5.1.** *Let $\Phi:\mathbb{H}\to\mathbb{H}$ be a diffeomorphism of class $\mathscr{C}^2$. If, for all $E\subset\mathbb{H}$ measurable and all $\Omega\subset\mathbb{H}$ open, it holds*

$$P(E;\Omega)<\infty \qquad\Rightarrow\qquad P(\Phi(E);\Phi(\Omega))<\infty, \tag{9}$$

*then $\Phi$ is contact.*

We will show in this section that any variation of an intrinsic graph $\Gamma_f$ via contact diffeomorphisms is equivalent to a variation of $f$ via the transformations of Proposition 4.1 and Remark 4.2.

**Proposition 5.2.** *Let $\phi:\mathbb{R}^2\to\mathbb{R}^2$ be a $\mathscr{C}^\infty$-diffeomorphism and $f,\bar f\in\mathscr{C}^1_{\mathbb{W}}$ as in Proposition 4.1. Then there is a contact diffeomorphism $\Phi:\Omega\to\Phi(\Omega)$, where $\Omega$ and $\Phi(\Omega)$ are open subsets of $\mathbb{H}$ with $\Gamma_f\subset\Omega$, such that $\Phi(\Gamma_f)=\Gamma_{\bar f}$.*

**Proposition 5.3.** *Let $\Phi^\epsilon:\mathbb{H}\to\mathbb{H}$, $\epsilon\in\mathbb{R}$, be a smooth one-parameter family of contact diffeomorphisms such that there is a compact set $K\subset\mathbb{H}$ with $\Phi^\epsilon|_{\mathbb{H}\setminus K}=\mathrm{Id}$ for all $\epsilon$ and $\Phi^0=\mathrm{Id}$. Let $f\in\mathscr{C}^\infty(\mathbb{R}^2)$. Then there is $\epsilon_0>0$ such that for all $\epsilon$ with $|\epsilon|<\epsilon_0$, the maps $\phi^\epsilon:\mathbb{R}^2\to\mathbb{R}^2$,*

$$\phi^\epsilon(p) := \pi_X\circ\Phi^\epsilon(p*f(p)),$$

*form a smooth family of $\mathscr{C}^\infty$-diffeomorphism of $\mathbb{R}^2$.*

*Moreover, if $f^\epsilon$ is the function defined via $f$ and $\phi^\epsilon$ as in Proposition 4.1, then*

$$\Phi^\epsilon(\Gamma_f)=\Gamma_{f^\epsilon}.$$

5.1. **Proof of Proposition 5.1.** We use an argument by contradiction. Assume that $\Phi$ is not a contact diffeomorphism. Then there is an open and bounded set $\Omega \subset \mathbb{H}$ such that for all $p \in \Omega$ it holds $\mathrm{d}\Phi(H_p) \not\subset H_{\Phi(p)}$. Thanks to the following lemma and Remark 2.3, we get a contradiction with the property (9).

**Lemma 5.4.** *Let $\Phi : \mathbb{H} \to \mathbb{H}$ be a diffeomorphism of class $\mathscr{C}^2$. Let $\Omega \subset \mathbb{H}$ be an open and bounded set such that for all $p \in \Omega$*

$$\mathrm{d}\Phi(H_p) \not\subset H_{\Phi(p)}.$$

*Let $E \subset \mathbb{H}$ be measurable. If $P(E;\Omega) < \infty$ and $P(\Phi(E);\Phi(\Omega)) < \infty$, then $E$ has finite Riemannian perimeter in $\Omega$.*

*Proof.* We extend the scalar product $\langle \cdot, \cdot \rangle$ to the whole $T\mathbb{H}$ in such a way that $(X, Y, Z)$ is an orthonormal frame. The Riemannian perimeter is defined as

$$P_{\mathscr{R}}(E;\Omega) := \sup \left\{ \int_E \mathrm{div} U \,\mathrm{d}\mathscr{L}^3 : \ U \in \mathrm{Vec}(T\mathbb{H}),\ \mathrm{spt} U \subset\subset \Omega,\ \|U\| \le 1 \right\}.$$

Let $U \in \mathrm{Vec}(T\mathbb{H})$ with $\mathrm{spt}(U) \subset\subset \Omega$ and $\|U\| \le 1$. Then there are $V, W \in \mathrm{Vec}(T\mathbb{H})$ with $\mathrm{spt}(V) \cup \mathrm{spt}(W) = \mathrm{spt}(U)$, $V + W = U$, $V(p) \in H_p$ for all $p$, $\|V\| \le K$ and $\|W\| \le K$, and $\Phi_* W(p) \in H_p$ for all $p$, where $K \ge 0$ depends on $\Phi$ and $\Omega$, but not on $U$.

Remind that, if $W$ is a smooth vector field on $\mathbb{H}$, then[2]

$$\mathrm{div}(\Phi_* W) = \mathrm{div}(W) \circ \Phi^{-1} \cdot J(\Phi^{-1}).$$

Hence,$\int_E \mathrm{div} W \,\mathrm{d}\mathscr{L}^3 = \int_{\Phi(E)} (\mathrm{div} W) \circ \Phi^{-1} J\Phi^{-1} \,\mathrm{d}\mathscr{L}^3 = \int_{\Phi(E)} \mathrm{div}(\Phi_* W) \,\mathrm{d}\mathscr{L}^3$. Moreover, since $\Omega$ is bounded, we can assume $\|\mathrm{d}\Phi(v)\| \le K\|v\|$ for all $v \in T\Omega$, where $K \ge 0$ is the same constant as above. Therefore

$$\begin{aligned}
\int_E \mathrm{div} U \,\mathrm{d}\mathscr{L}^3 &= \int_E \mathrm{div} V \,\mathrm{d}\mathscr{L}^3 + \int_E \mathrm{div} W \,\mathrm{d}\mathscr{L}^3 \\
&= \int_E \mathrm{div} V \,\mathrm{d}\mathscr{L}^3 + \int_{\Phi(E)} \mathrm{div}(\Phi_* W) \,\mathrm{d}\mathscr{L}^3 \\
&\le K P(E;\Omega) + K^2 P(\Phi(E);\Phi(\Omega)).
\end{aligned}$$

This implies that $P_{\mathscr{R}}(E;\Omega) \le K P(E;\Omega) + K^2 P(\Phi(E);\Phi(\Omega)) < \infty$. ☐

5.2. **Proof of Proposition 5.2.** In this case our choice of coordinates is not helpful. So, we consider the exponential coordinates of second kind $(\xi, \eta, \tau) \mapsto \exp(\eta B + \tau C) * \exp(\xi A)$, using the notation of Section 2.1.

We define the map $\Phi$ as

$$\Phi(\xi, \eta, \tau) := \left( \frac{\nabla^\xi \phi_2}{\nabla^\xi \phi_1}(\eta, \tau), \phi_1(\eta, \tau), \phi_2(\eta, \tau) \right)$$

Clearly, $\Phi$ is well defined and smooth on the open set

$$\Omega := \{(\xi, \eta, \tau) : \nabla^\xi \phi_1(\eta, \tau) \neq 0\},$$

[2]A sketch of the proof of this formula: it is clearer to show the dual formula $\mathrm{div}(\Phi^* W) = \mathrm{div}(W) \circ \Phi \cdot J(\Phi)$; consider $W$ as a 2-form and the divergence as the exterior derivative d; remind that $\mathrm{d}\Phi^* = \Phi^* \mathrm{d}$; the formula follows.

$\Gamma_f \subset \Omega$ by the hypothesis of Proposition 4.1 and $\Phi(\Gamma_f) = \Gamma_{\bar{f}}$. In these coordinates, the differential of $\Phi$ is

$$\mathrm{d}\Phi(\xi,\eta,\tau) = \begin{pmatrix} \partial_\xi\left(\frac{\nabla^\xi\phi_2}{\nabla^\xi\phi_1}\right) & \partial_\eta\left(\frac{\nabla^\xi\phi_2}{\nabla^\xi\phi_1}\right) & \partial_\tau\left(\frac{\nabla^\xi\phi_2}{\nabla^\xi\phi_1}\right) \\ 0 & \partial_\eta\phi_1 & \partial_\tau\phi_1 \\ 0 & \partial_\eta\phi_2 & \partial_\tau\phi_2 \end{pmatrix}$$

Since $\phi$ is a diffeomorphism, $\Phi$ is a diffeomorphism if and only if $\partial_\xi\left(\frac{\nabla^\xi\phi_2}{\nabla^\xi\phi_1}\right) \neq 0$. A short computation shows that

$$\partial_\xi\left(\frac{\nabla^\xi\phi_2}{\nabla^\xi\phi_1}\right) = \frac{\det(\mathrm{d}\phi)}{(\nabla^\xi\phi_1)^2},$$

which is non-zero.

Now, we need to show that $\Phi$ is a contact diffeomorphism. In this system of coordinates, the left-invariant vector fields $X, Y, Z$ are written as

$$\tilde{X}(\xi,\eta,\tau) = \partial_\xi, \qquad \tilde{Y}(\xi,\eta,\tau) = \partial_\eta + \xi\partial_\tau, \qquad \tilde{Z}(\xi,\eta,\tau) = \partial_\tau.$$

We have

$$\mathrm{d}\Phi\left(\tilde{X}(\xi,\eta,\tau)\right) = \partial_\xi\left(\frac{\nabla^\xi\phi_2}{\nabla^\xi\phi_1}\right)\tilde{X}(\Phi(\xi,\eta,\tau)),$$

$$\mathrm{d}\Phi\left(\tilde{Y}(\xi,\eta,\tau)\right) = \nabla^\xi\left(\frac{\nabla^\xi\phi_2}{\nabla^\xi\phi_1}\right)\tilde{X}(\Phi(\xi,\eta,\tau)) + \nabla^\xi\phi_1\tilde{Y}(\Phi(\xi,\eta,\tau)).$$

Therefore, $\mathrm{d}\Phi(H) \subset H$. □

5.3. **Proof of Proposition 5.3.** The functions $\phi^\epsilon : \mathbb{R}^2 \to \mathbb{R}^2$ are well defined and smooth for all $\epsilon \in \mathbb{R}$. Since $\Phi^\epsilon$ and all its derivative converge to Id uniformly on $\mathbb{H}$, there exists $\epsilon_0 > 0$ such that for all $\epsilon$ with $|\epsilon| < \epsilon_0$, the vector field $X$ is not tangent to $\Phi^\epsilon(\Gamma_f)$ at any point. Therefore, $\det(\mathrm{d}\phi^\epsilon) \neq 0$ for all such $\epsilon$. Since $\phi^\epsilon|_{\pi_X(K)} = \mathrm{Id}$, $\phi^\epsilon$ is a covering map and therefore it is a smooth diffeomorphism.

The last statement is a direct consequence Lemma 2.2. □

## 6. First Contact Variation

Similar formulas for the first and the second variation for the sub-Riemannian perimeter in the Heisenberg group can be found in [4, 5, 8, 15].

In all the formulas below, we set $\psi := \nabla^f f$.

**Proposition 6.1.** *Let $f \in \mathscr{C}^1_{\mathbb{W}}$ be such that $\Gamma_f$ is an area-minimizing surface. Then for all $V_1, V_2 \in \mathscr{C}^\infty_c(\mathbb{R}^2)$ it holds*

$$0 = \int_{\mathbb{R}^2}\left[\frac{\psi}{\sqrt{1+\psi^2}}\left(-2\psi\cdot\nabla^f V_1 - f\cdot\Delta^f V_1\right) + \sqrt{1+\psi^2}\partial_\eta V_1\right]\mathrm{d}\eta\,\mathrm{d}\tau. \tag{10}$$

*and*

$$0 = \int_{\mathbb{R}^2}\left[\frac{\psi}{\sqrt{1+\psi^2}}\Delta^f V_2 + \sqrt{1+\psi^2}\partial_\tau V_2\right]\mathrm{d}\eta\,\mathrm{d}\tau. \tag{11}$$

**Proposition 6.2.** *Let $f \in \mathscr{C}^\infty(\mathbb{R}^2)$ be such that for all $V_2 \in \mathscr{C}^\infty_c(\mathbb{R}^2)$ the equation* (11) *holds. Then* (10) *holds as well for all $V_1 \in \mathscr{C}^\infty_c(\mathbb{R}^2)$.*

**Proposition 6.3.** *A function $f \in \mathscr{C}^\infty(\mathbb{R}^2)$ satisfies* (11) *for all $V_2 \in \mathscr{C}^\infty_c(\mathbb{R}^2)$ if and only if*

$$(\nabla^f + 2\partial_\tau f)\nabla^f \left( \frac{\psi}{\sqrt{1+\psi^2}} \right) = 0. \tag{12}$$

6.1. **Proof of Proposition 6.1.** Let $f \in \mathscr{C}^1_{\mathbb{W}}$, $\omega \subset \mathbb{R}^2$ an open and bounded set and $V = (V_1, V_2) : \mathbb{R}^2 \to \mathbb{R}^2$ a smooth vector field with $\mathrm{spt} V \subset\subset \omega$. Let $\phi^\epsilon : \mathbb{R}^2 \to \mathbb{R}^2$ be a smooth one-parameter family of diffeomorphism such that $\{\phi^\epsilon \neq \mathrm{Id}\} \subset \mathrm{spt} V$ for all $\epsilon > 0$ and, for all $p \in \mathbb{R}^2$,

$$\begin{cases} \phi^0(p) = p \\ \partial_\epsilon \phi^\epsilon(p)|_{\epsilon=0} = V(p). \end{cases}$$

Notice that $\nabla^f \phi_1^\epsilon = \partial_\eta \phi_1^\epsilon + f \partial_\tau \phi_1^\epsilon$ is not zero for $\epsilon$ small enough, because $\nabla^f \phi_1^\epsilon$ converges to 1 uniformly as $\epsilon \to 0$. Hence, by Proposition 4.1, there is an interval $I = (-\hat\epsilon, \hat\epsilon)$ such that the function given by

$$f_\epsilon \circ \phi^\epsilon = \frac{\nabla^f \phi_2^\epsilon}{\nabla^f \phi_1^\epsilon} \tag{13}$$

is well defined for all $\epsilon \in I$. Define $\gamma : I \to \mathbb{R}$ as

$$\begin{aligned} \gamma(\epsilon) &:= \int_\omega \sqrt{1 + (\nabla^{f_\epsilon} f_\epsilon)^2}\, \mathrm{d}\eta \, \mathrm{d}\tau \\ &= \int_\omega \sqrt{1 + ((\nabla^{f_\epsilon} f_\epsilon) \circ \phi^\epsilon)^2} J_{\phi^\epsilon} \, \mathrm{d}\eta \, \mathrm{d}\tau, \end{aligned}$$

where we performed a change of coordinates via $\phi^\epsilon$ and

$$J_{\phi^\epsilon} = \partial_\eta \phi_1^\epsilon \partial_\tau \phi_2^\epsilon - \partial_\tau \phi_1^\epsilon \partial_\eta \phi_2^\epsilon$$

is the Jacobian of $\phi^\epsilon$. Using equality (8) and Lemma 3.1, it is immediate to see that $\gamma$ is continuous.

**Lemma 6.4.** *The function $\gamma : I \to \mathbb{R}$ is continuously differentiable and*

$$\begin{aligned} \gamma'(\epsilon) = \int_\omega \Bigg[ &\frac{((\nabla^{f_\epsilon} f_\epsilon) \circ \phi^\epsilon)}{\sqrt{1 + ((\nabla^{f_\epsilon} f_\epsilon) \circ \phi^\epsilon)^2}} A_f(\epsilon) J_{\phi^\epsilon} + \\ &+ \sqrt{1 + ((\nabla^{f_\epsilon} f_\epsilon) \circ \phi^\epsilon)^2} \partial_\epsilon J_{\phi^\epsilon} \Bigg] \mathrm{d}\eta \, \mathrm{d}\tau, \end{aligned} \tag{14}$$

*where*

$$\begin{aligned} A_f(\epsilon) := &\frac{\Delta^f \partial_\epsilon \phi_2^\epsilon}{(\nabla^f \phi_1^\epsilon)^2} - 2 \frac{\Delta^f \phi_2^\epsilon}{(\nabla^f \phi_1^\epsilon)^3} \nabla^f \partial_\epsilon \phi_1^\epsilon + \\ &- \frac{\nabla^f \partial_\epsilon \phi_2^\epsilon}{(\nabla^f \phi_1^\epsilon)^3} \Delta^f \phi_1^\epsilon + 3 \frac{\nabla^f \phi_2^\epsilon}{(\nabla^f \phi_1^\epsilon)^4} \nabla^f \partial_\epsilon \phi_1^\epsilon \cdot \Delta^f \phi_1^\epsilon - \frac{\nabla^f \phi_2^\epsilon}{(\nabla^f \phi_1^\epsilon)^3} \Delta^f \partial_\epsilon \phi_1^\epsilon. \end{aligned} \tag{15}$$

*Proof of Lemma 6.4.* First, suppose $f \in \mathscr{C}^\infty(\mathbb{R}^2)$. Then $\gamma \in \mathscr{C}^\infty(I)$ and

$$\begin{aligned} \gamma'(\epsilon) = \int_\omega \Bigg[ &\frac{((\nabla^{f_\epsilon} f_\epsilon) \circ \phi^\epsilon)}{\sqrt{1 + ((\nabla^{f_\epsilon} f_\epsilon) \circ \phi^\epsilon)^2}} \partial_\epsilon ((\nabla^{f_\epsilon} f_\epsilon) \circ \phi^\epsilon) J_{\phi^\epsilon} + \\ &+ \sqrt{1 + ((\nabla^{f_\epsilon} f_\epsilon) \circ \phi^\epsilon)^2} \partial_\epsilon J_{\phi^\epsilon} \Bigg] \mathrm{d}\eta \, \mathrm{d}\tau. \end{aligned}$$

Applying the formula in Proposition 4.1 and the identity $\nabla^f\partial_\epsilon = \partial_\epsilon\nabla^f$, one obtains

$$\partial_\epsilon((\nabla^{f_\epsilon} f_\epsilon)\circ\phi^\epsilon) = A_f(\epsilon)$$

and thus formula (14) holds in the smooth case.

Next, suppose $f = f_\infty$ is the limit in $\mathscr{C}^1_{\mathbb{W}}$ of a sequence $f_k \in \mathscr{C}^\infty(\mathbb{R}^2)$, as in Lemma 2.4. Notice that $\nabla^{f_k}\phi_1^\epsilon$ is not zero for $\epsilon$ small enough and $k$ large enough. Indeed, $|\nabla^{f_k}\phi_1^\epsilon - \nabla^{f_\infty}\phi_1^\epsilon| \le \|f_k - f\|_{\mathscr{L}^\infty(\mathrm{spt}V)}\|\partial_\tau\phi_1^\epsilon\|_{\mathscr{L}^\infty(\mathrm{spt}V)}$ and $\nabla^{f_\infty}\phi_1^\epsilon$ converges to one uniformly on $\mathbb{R}^2$ as $\epsilon\to 0$. Hence, there is an interval $I\subset\mathbb{R}$ centered at zero such that the functions $f_{k,\epsilon}$ as in Proposition 4.1 are well defined for $\epsilon\in I$ and, without loss of generality, for all $k\in\mathbb{N}\cup\{\infty\}$. For $k\in\mathbb{N}\cup\{\infty\}$, define $\gamma_k : I\to\mathbb{R}$ as

$$\gamma_k(\epsilon) := \int_\omega \sqrt{1+(\nabla^{f_{k,\epsilon}} f_{k,\epsilon})^2}\,\mathrm{d}\eta\,\mathrm{d}\tau$$

Define also the function $\eta : I\to\mathbb{R}$ as the right-hand side of (14). From Lemma 3.1, it follows that $\{A_{f_k}\}_{k\in\mathbb{N}}$ converges to $A_f$ uniformly on $I$. Therefore, we have that $\{\gamma_k\}_{k\in\mathbb{N}}$ and $\{\gamma_k'\}_{k\in\mathbb{N}}$ converge to $\gamma$ and $\eta$ uniformly on $I$. We conclude that $\gamma\in\mathscr{C}^1(I)$ and $\gamma'=\eta$. □

In order to evaluate $\gamma'(0)$, notice that

$$\begin{aligned}
\nabla^f\phi_1^0 &= 1 & \nabla^f\phi_2^0 &= f\\
\nabla^f\partial_\epsilon\phi_1^\epsilon|_{\epsilon=0} &= \nabla^f V_1 & \nabla^f\partial_\epsilon\phi_2^\epsilon|_{\epsilon=0} &= \nabla^f V_2\\
\Delta^f\phi_1^0 &= 0 & \Delta^f\phi_2^0 &= \psi\\
\Delta^f\partial_\epsilon\phi_1^\epsilon|_{\epsilon=0} &= \Delta^f V_1 & \Delta^f\partial_\epsilon\phi_2^\epsilon|_{\epsilon=0} &= \Delta^f V_2.
\end{aligned}$$

Therefore

$$A_f(0) = \Delta^f V_2 - 2\psi\nabla^f V_1 - f\Delta^f V_1.$$

Moreover, using the facts $\partial_\tau\phi_1^0 = \partial_\eta\phi_2^0 = 0$ and $\partial_\eta\phi_1^0 = \partial_\tau\phi_2^0 = 1$ and that the derivatives $\partial_\epsilon$, $\partial_\eta$ and $\partial_\tau$ commute, we have

$$\partial_\epsilon J_{\phi^\epsilon}|_{\epsilon=0} = \partial_\eta V_1 + \partial_\tau V_2.$$

Putting all together, we obtain

$$\begin{aligned}
\gamma'(0) = \int_\omega \Bigg[&\frac{\psi}{\sqrt{1+\psi^2}}\left(\Delta^f V_2 - 2\psi\cdot\nabla^f V_1 - f\cdot\Delta^f V_1\right) + \\
&+\sqrt{1+\psi^2}(\partial_\eta V_1 + \partial_\tau V_2)\Bigg]\,\mathrm{d}\eta\,\mathrm{d}\tau.
\end{aligned}$$

Since $\Gamma_f$ is an area-minimizing surface, then $\gamma'(0) = 0$ for all $V_1, V_2 \in \mathscr{C}^\infty_c(\mathbb{R}^2)$. Since this expression is linear in $V$, then we obtain both conditions (10) and (11). □

6.2. **Proof of Proposition 6.2.** Let $V_1\in\mathscr{C}^\infty(\mathbb{R}^2)$ and set $V_2 := fV_1 \in \mathscr{C}^\infty_c(\mathbb{R}^2)$. Then

$$\begin{aligned}
0 &= \int_{\mathbb{R}^2}\left[\frac{\psi}{\sqrt{1+\psi^2}}\Delta^f V_2 + \sqrt{1+\psi^2}\partial_\tau V_2\right]\mathrm{d}\eta\,\mathrm{d}\tau\\
&= \int_{\mathbb{R}^2}\Bigg[\frac{\psi}{\sqrt{1+\psi^2}}(\nabla^f\psi V_1 + 2\psi\nabla^f V_1 + f\Delta^f V_1) +
\end{aligned}$$

$$
\begin{aligned}
&\qquad\qquad\qquad\qquad\qquad + \sqrt{1+\psi^2}(\partial_\tau f V_1 + f\partial_\tau V_1)\Big]\,\mathrm{d}\eta\,\mathrm{d}\tau \\
&= \int_{\mathbb{R}^2}\Bigg[\frac{\psi}{\sqrt{1+\psi^2}}(2\psi\nabla^f V_1 + f\Delta^f V_1)+ \\
&\qquad\qquad + \left(\frac{\psi\nabla^f\psi}{\sqrt{1+\psi^2}} + \sqrt{1+\psi^2}\partial_\tau f\right)V_1+ \\
&\qquad\qquad\qquad\qquad\qquad + \sqrt{1+\psi^2}(\nabla^f V_1 - \partial_\eta V_1)\Bigg]\,\mathrm{d}\eta\,\mathrm{d}\tau \\
&= \int_{\mathbb{R}^2}\Bigg[\frac{\psi}{\sqrt{1+\psi^2}}(2\psi\nabla^f V_1 + f\Delta^f V_1) - \sqrt{1+\psi^2}\nabla^f V_1+ \\
&\qquad\qquad\qquad\qquad\qquad + \sqrt{1+\psi^2}(\nabla^f V_1 - \partial_\eta V_1)\Bigg]\,\mathrm{d}\eta\,\mathrm{d}\tau \\
&= \int_{\mathbb{R}^2}\left[\frac{\psi}{\sqrt{1+\psi^2}}(2\psi\nabla^f V_1 + f\Delta^f V_1) - \sqrt{1+\psi^2}\partial_\eta V_1)\right]\mathrm{d}\eta\,\mathrm{d}\tau.
\end{aligned}
$$

Hence (10) holds true for $V_1$ as well. □

6.3. **Proof of Proposition 6.3.** We have for all $V_2 \in \mathscr{C}_c^\infty(\mathbb{R}^2)$

$$
\begin{aligned}
&\int_{\mathbb{R}^2}\left[\frac{\psi}{\sqrt{1+\psi^2}}\nabla^f\nabla^f V_2 + \sqrt{1+\psi^2}\partial_\tau V_2\right]\mathrm{d}\eta\,\mathrm{d}\tau \\
&= -\int_{\mathbb{R}^2}\left[\nabla^f\left(\frac{\psi}{\sqrt{1+\psi^2}}\right)\nabla^f V_2 + \frac{\partial_\tau f\psi}{\sqrt{1+\psi^2}}\nabla^f V_2 + \partial_\tau(\sqrt{1+\psi^2})V_2\right]\mathrm{d}\eta\,\mathrm{d}\tau \\
&= \int_{\mathbb{R}^2}\Bigg[\nabla^f\nabla^f\left(\frac{\psi}{\sqrt{1+\psi^2}}\right)V_2 + \partial_\tau f\nabla^f\left(\frac{\psi}{\sqrt{1+\psi^2}}\right)V_2+ \\
&\quad + \nabla^f(\partial_\tau f)\frac{\psi}{\sqrt{1+\psi^2}}V_2 + \partial_\tau f\nabla^f\left(\frac{\psi}{\sqrt{1+\psi^2}}\right)V_2 + (\partial_\tau f)^2\frac{\psi}{\sqrt{1+\psi^2}}V_2+ \\
&\qquad\qquad\qquad\qquad\qquad - \frac{\psi}{\sqrt{1+\psi^2}}\partial_\tau\psi V_2\Bigg]\,\mathrm{d}\eta\,\mathrm{d}\tau.
\end{aligned}
$$

Therefore, using the fact that $\partial_\tau\psi = \nabla^f(\partial_\tau f) + (\partial_\tau f)^2$, we get that (11) is equivalent to

$$
\nabla^f\nabla^f\left(\frac{\psi}{\sqrt{1+\psi^2}}\right) + 2\partial_\tau f\cdot\nabla^f\left(\frac{\psi}{\sqrt{1+\psi^2}}\right) = 0.
$$

□

## 7. Second Contact Variation

Similarly to the previous sections, we set $\psi := \nabla^f f$.

**Proposition 7.1.** *If the intrinsic graph of $f \in \mathscr{C}^1_{\mathbb{W}}$ is an area-minimizing surface, then, for all $V_1, V_2 \in \mathscr{C}_c^\infty(\mathbb{R}^2)$, we have:*

$$(16)\quad 0 \le II_f(V_1,V_2) := \int_{\mathbb{R}^2} \left[ \frac{(\Delta^f V_2 - 2\psi\nabla^f V_1 - f\Delta^f V_1)^2}{(1+\psi^2)^{\frac{3}{2}}} + \right.$$
$$+ \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \Big( -4\Delta^f V_2 \cdot \nabla^f V_1 - 2\nabla^f V_2 \cdot \Delta^f V_1 +$$
$$+6f \cdot \nabla^f V_1 \cdot \Delta^f V_1 + 6\psi \cdot (\nabla^f V_1)^2 \Big) +$$
$$+ 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} (\Delta^f V_2 - 2\psi\nabla^f V_1 - f\Delta^f V_1)(\partial_\eta V_1 + \partial_\tau V_2) +$$
$$\left. + 2(1+\psi^2)^{\frac{1}{2}} (\partial_\eta V_1 \partial_\tau V_2 - \partial_\tau V_1 \partial_\eta V_2) \right] \mathrm{d}\eta \, \mathrm{d}\tau.$$

7.1. **Proof of Proposition 7.1.** Let $\omega \subset \mathbb{R}^2$ be an open and bounded set and $V = (V_1, V_2) : \mathbb{R}^2 \to \mathbb{R}^2$ a smooth vector field with $\mathrm{spt} V \subset\subset \omega$. Let $\phi^\epsilon = (\phi_1^\epsilon, \phi_2^\epsilon) : \mathbb{R}^2 \to \mathbb{R}^2$ be a smooth one-parameter family of diffeomorphism such that $\{\phi^\epsilon \neq \mathrm{Id}\} \subset \mathrm{spt} V$ for all $\epsilon > 0$ and, for all $p \in \mathbb{R}^2$,

$$\begin{cases} \phi^0(p) = p \\ \partial_\epsilon \phi^\epsilon(p)|_{\epsilon=0} = V(p). \end{cases}$$

Define $W_i(p) := \partial_\epsilon^2 \phi_i^\epsilon(p)|_{\epsilon=0}$. Then $W = (W_1, W_2) : \mathbb{R}^2 \to \mathbb{R}^2$ is a smooth vector field with $\mathrm{spt} W \subset\subset \omega$.

As for the first variation, see Section 6.1, define

$$\gamma(\epsilon) := \int_\omega \sqrt{1 + (\nabla^{f_\epsilon} f_\epsilon)^2} \, \mathrm{d}\eta \, \mathrm{d}\tau.$$

**Lemma 7.2.** *The function $\gamma : I \to \mathbb{R}$ is twice continuously differentiable and*

$$(17)\quad \gamma''(\epsilon) = \int_\omega \left[ \frac{A_f(\epsilon)^2}{(1+(\nabla^{f_\epsilon} f_\epsilon \circ \phi^\epsilon)^2)^{\frac{3}{2}}} J_{\phi^\epsilon} + \frac{(\nabla^{f_\epsilon} f_\epsilon \circ \phi^\epsilon) B_f(\epsilon)}{(1+(\nabla^{f_\epsilon} f_\epsilon \circ \phi^\epsilon)^2)^{\frac{1}{2}}} J_{\phi^\epsilon} + \right.$$
$$\left. + 2\frac{(\nabla^{f_\epsilon} f_\epsilon \circ \phi^\epsilon) A_f(\epsilon)}{(1+(\nabla^{f_\epsilon} f_\epsilon \circ \phi^\epsilon)^2)^{\frac{1}{2}}} \partial_\epsilon J_{\phi^\epsilon} + (1+(\nabla^{f_\epsilon} f_\epsilon \circ \phi^\epsilon)^2)^{\frac{1}{2}} \partial_\epsilon^2 J_{\phi^\epsilon} \right] \mathrm{d}y \, \mathrm{d}z,$$

*where $A_f(\epsilon)$ is defined as in* (15) *and*

$$B_f(\epsilon) := \frac{\Delta^f \partial_\epsilon^2 \phi_2^\epsilon}{(\nabla^f \phi_1^\epsilon)^2} - 2\frac{\Delta^f \partial_\epsilon \phi_2^\epsilon \cdot \nabla^f \partial_\epsilon \phi_1^\epsilon}{(\nabla^f \phi_1^\epsilon)^3} +$$
$$- 2\frac{\Delta^f \partial_\epsilon \phi_2^\epsilon \cdot \nabla^f \partial_\epsilon \phi_1^\epsilon}{(\nabla^f \phi_1^\epsilon)^3} - 2\frac{\Delta^f \phi_2^\epsilon \cdot \nabla^f \partial_\epsilon^2 \phi_1^\epsilon}{(\nabla^f \phi_1^\epsilon)^3} + 6\frac{\Delta^f \phi_2^\epsilon \cdot (\nabla^f \partial_\epsilon \phi_1^\epsilon)^2}{(\nabla^f \phi_1^\epsilon)^4} +$$
$$- \frac{\nabla^f \partial_\epsilon^2 \phi_2^\epsilon \cdot \Delta^f \phi_1^\epsilon}{(\nabla^f \phi_1^\epsilon)^3} - \frac{\nabla^f \partial_\epsilon \phi_2^\epsilon \cdot \Delta^f \partial_\epsilon \phi_1^\epsilon}{(\nabla^f \phi_1^\epsilon)^3} + 3\frac{\nabla^f \partial_\epsilon \phi_2^\epsilon \cdot \Delta^f \phi_1^\epsilon \cdot \nabla^f \partial_\epsilon \phi_1^\epsilon}{(\nabla^f \phi_1^\epsilon)^4} +$$
$$+ 3\frac{\nabla^f \partial_\epsilon \phi_2^\epsilon \cdot \nabla^f \partial_\epsilon \phi_1^\epsilon \cdot \Delta^f \phi_1^\epsilon}{(\nabla^f \phi_1^\epsilon)^4} + 3\frac{\nabla^f \phi_2^\epsilon \cdot \nabla^f \partial_\epsilon^2 \phi_1^\epsilon \cdot \Delta^f \phi_1^\epsilon}{(\nabla^f \phi_1^\epsilon)^4} +$$
$$+ 3\frac{\nabla^f \phi_2^\epsilon \cdot \nabla^f \partial_\epsilon \phi_1^\epsilon \cdot \Delta^f \partial_\epsilon \phi_1^\epsilon}{(\nabla^f \phi_1^\epsilon)^4} - 12\frac{\nabla^f \phi_2^\epsilon \cdot (\nabla^f \partial_\epsilon \phi_1^\epsilon)^2 \cdot \Delta^f \phi_1^\epsilon}{(\nabla^f \phi_1^\epsilon)^5} +$$
$$- \frac{\nabla^f \partial_\epsilon \phi_2^\epsilon \cdot \Delta^f \partial_\epsilon \phi_1^\epsilon}{(\nabla^f \phi_1^\epsilon)^3} - \frac{\nabla^f \phi_2^\epsilon \cdot \Delta^f \partial_\epsilon^2 \phi_1^\epsilon}{(\nabla^f \phi_1^\epsilon)^3} + 3\frac{\nabla^f \phi_2^\epsilon \cdot \Delta^f \partial_\epsilon \phi_1^\epsilon \cdot \nabla^f \partial_\epsilon \phi_1^\epsilon}{(\nabla^f \phi_1^\epsilon)^4}.$$

*Proof of Lemma 7.2.* This lemma is a continuation of Lemma 6.4.

First, suppose $f \in \mathscr{C}^\infty(\mathbb{R}^2)$. Then, the function $\gamma$ is smooth and its second derivative is

$$\begin{aligned}\gamma''(\epsilon) = \int_\omega \Bigg[ & \frac{(\partial_\epsilon(\nabla^{f_\epsilon} f_\epsilon \circ \phi^\epsilon))^2}{(1+(\nabla^{f_\epsilon} f_\epsilon \circ \phi^\epsilon)^2)^{\frac{3}{2}}} J_{\phi^\epsilon} + \\ & + \frac{(\nabla^{f_\epsilon} f_\epsilon \circ \phi^\epsilon)\partial_\epsilon^2(\nabla^{f_\epsilon} f_\epsilon \circ \phi^\epsilon)}{(1+(\nabla^{f_\epsilon} f_\epsilon \circ \phi^\epsilon)^2)^{\frac{1}{2}}} J_{\phi^\epsilon} + \\ & + 2\frac{(\nabla^{f_\epsilon} f_\epsilon \circ \phi^\epsilon)\partial_\epsilon(\nabla^{f_\epsilon} f_\epsilon \circ \phi^\epsilon)}{(1+(\nabla^{f_\epsilon} f_\epsilon \circ \phi^\epsilon)^2)^{\frac{1}{2}}} \partial_\epsilon J_{\phi^\epsilon} + \\ & + (1+(\nabla^{f_\epsilon} f_\epsilon \circ \phi^\epsilon)^2)^{\frac{1}{2}} \partial_\epsilon^2 J_{\phi^\epsilon} \Bigg] \mathrm{d}y\,\mathrm{d}z.\end{aligned}$$

One checks by direct computation that

$$\begin{aligned}\partial_\epsilon(\nabla^{f_\epsilon} f_\epsilon \circ \phi^\epsilon) &= A_f(\epsilon), \\ \partial_\epsilon^2(\nabla^{f_\epsilon} f_\epsilon \circ \phi^\epsilon) &= B_f(\epsilon),\end{aligned}$$

thus (17) is proven in the smooth case.

Next, suppose $f = f_\infty$ is the limit in $\mathscr{C}^1_{\mathbb{W}}$ of a sequence $f_k \in \mathscr{C}^\infty(\mathbb{R}^2)$, as in Lemma 2.4. Define $f_{k,\epsilon}$ and $I \subset \mathbb{R}$ and $\gamma_k : I \to \mathbb{R}$ as in the proof of Lemma 6.4. Define also $\eta : I \to \mathbb{R}$ as the right-hand side of (17). By Lemma 3.1, $\{A_{f_k}\}_{k\in\mathbb{N}}$ and $\{B_{f_k}\}_{k\in\mathbb{N}}$ converge to $A_f$ and $B_f$ uniformly on $I$. Therefore, we have that the convergences $\gamma_k \to \gamma$ and $\gamma_k' \to \gamma'$ and $\gamma_k'' \to \eta$ are uniform on $I$. We conclude that $\gamma \in \mathscr{C}^2(I)$ and $\gamma'' = \eta$. □

Next, one can directly check that

$$\begin{aligned}\gamma''(0) = \int_\omega \Bigg[ & \frac{(\Delta^f V_2 - 2\psi\nabla^f V_1 - f\Delta^f V_1)^2}{(1+\psi^2)^{\frac{3}{2}}} + \\ & + \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}(\Delta^f W_2 - f\cdot\Delta^f W_1 - 2\psi\cdot\nabla^f W_1 - 4\Delta^f V_2\cdot\nabla^f V_1 - 2\nabla^f V_2\cdot\Delta^f V_1 + \\ & + 6f\cdot\nabla^f V_1\cdot\Delta^f V_1 + 6\psi\cdot(\nabla^f V_1)^2) + \\ & + 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}(\Delta^f V_2 - 2\psi\nabla^f V_1 - f\Delta^f V_1)(\partial_\eta V_1 + \partial_\tau V_2) + \\ & + (1+\psi^2)^{\frac{1}{2}}(\partial_\eta W_1 + \partial_\tau W_2 + 2(\partial_\eta V_1\partial_\tau V_2 - \partial_\tau V_1\partial_\eta V_2))\Bigg] \mathrm{d}\eta\,\mathrm{d}\tau.\end{aligned}$$

Finally, if $\Gamma_f$ is an area-minimizing surface, then $\gamma'(0) = 0$ and $\gamma''(0) \geq 0$. Notice that the terms containing $W_1$ and $W_2$ in the expression of $\gamma''(0)$ are zero because $\gamma'(0) = 0$. So, the second variation formula (16) is proven. □

## 8. Contact variations in the case $\Delta^f f = 0$

In this final section we prove our main result. We show that there is a quite large class of functions in $\mathscr{C}^1_{\mathbb{W}}$ that satisfy both conditions on the first and second contact variation. Since we know that the only intrinsic graphs of smooth functions that are area minimizers are the vertical planes, our result shows that variations along contact diffeomorphisms are not selective enough.

As usual, we set $\psi := \nabla^f f$.

**Lemma 8.1.** *Let $f \in \mathscr{C}^\infty(\mathbb{R}^2)$ be such that $\Delta^f f = 0$. Then*

$$II_f(V_1, V_2) = \int_{\mathbb{R}^2} \left[ \frac{(\Delta^f V_2 - 2\psi \nabla^f V_1 - f \Delta^f V_1)^2}{(1+\psi^2)^{\frac{3}{2}}} + \right.$$
$$\left. + \partial_\tau \left( \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \right) \Big( \nabla^f V_2 - \nabla^f (f V_1) \Big)^2 \right] \mathrm{d}\eta\, \mathrm{d}\tau.$$

The proof is very technical and it is postponed to the last section below.

**Theorem 8.2.** *Let $f \in \mathscr{C}^1_{\mathbb{W}}$ be such that $\Delta^f f = 0$ in weak Lagrangian sense. Then both equalities* (10) *and* (11) *and also the inequality* (16) *are satisfied for all $V_1, V_2 \in \mathscr{C}^\infty_c(\mathbb{R}^2)$.*

*Proof.* We first prove that both equalities (10) and (11) are satisfied. Let $\{f_k\}_{k\in\mathbb{N}} \subset \mathscr{C}^\infty(\mathbb{R}^2)$ be a sequence converging to $f$ in $\mathscr{C}^1_{\mathbb{W}}$ and such that $\Delta^{f_k} f_k = 0$, as in Lemma 3.8. Fix $V_1, V_2 \in \mathscr{C}^\infty_c(\mathbb{R}^2)$. Then (12) and (10) are satisfied by all $f_k$ thanks to Propositions 6.2 and 6.3. Passing to the limit $k \to \infty$, we prove that $f$ satisfies them too.

Now, we prove that the inequality (16) holds true. If $f \in \mathscr{C}^\infty(\mathbb{R}^2)$, then we can apply Lemma 8.1, where $\partial_\tau \left( \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \right) = \frac{\partial_\tau \psi}{(1+\psi^2)^{\frac{3}{2}}} \geq 0$ because of Lemma 3.5. So, (16) is proven for $f$ smooth. For $f \in \mathscr{C}^1_{\mathbb{W}}$, let $\{f_k\}_{k\in\mathbb{N}} \subset \mathscr{C}^\infty(\mathbb{R}^2)$ as in Lemma 3.8. From Lemma 3.1 follows that, for fixed $V_1, V_2 \in \mathscr{C}^\infty_c(\mathbb{R}^2)$, it holds

$$\lim_{k\to\infty} II_{f_k}(V_1, V_2) = II_f(V_1, V_2),$$

thus $II_f(V_1, V_2) \geq 0$. □

8.1. **Proof of Lemma 8.1.** The proof of this lemma is just a computation, but quite elaborate. For making the formulas more readable, we decided to drop the sign of integral along the proof. In other words, all equalities in this section are meant as equalities of integrals on $\mathbb{R}^2$. We will constantly use the formulas listed in Appendix A together with $\nabla^f \psi = 0$.

Before of all, we reorganise the integral in (16):

$$\text{(ⓐ)} \qquad \frac{(\Delta^f V_2 - 2\psi \nabla^f V_1 - f \Delta^f V_1)^2}{(1+\psi^2)^{\frac{3}{2}}} +$$

$$\text{(ⓑ)} \qquad + \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \Big( +6f \cdot \nabla^f V_1 \cdot \Delta^f V_1 + 6\psi \cdot (\nabla^f V_1)^2 \Big)$$

$$\text{(ⓒ)} \qquad + 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \Big( -2\psi \nabla^f V_1 - f \Delta^f V_1 \Big) \partial_\eta V_1$$

$$\text{(ⓓ)} \qquad + 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \Delta^f V_2 \partial_\tau V_2$$

$$\text{(ⓔ)} \qquad + \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \Big( -4\Delta^f V_2 \cdot \nabla^f V_1 - 2\nabla^f V_2 \cdot \Delta^f V_1 \Big)$$

$$(\text{f}) \qquad + 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}\Big(\Delta^f V_2 \partial_\eta V_1 + (-2\psi\nabla^f V_1 - f\Delta^f V_1)\partial_\tau V_2\Big)$$

$$(\text{g}) \qquad + 2(1+\psi^2)^{\frac{1}{2}}(\partial_\eta V_1 \partial_\tau V_2 - \partial_\tau V_1 \partial_\eta V_2).$$

In the following lemmas we will study b+c, d and e + f + g separately in order to obtain the expansion of the square in the second term of the integral in Lemma 8.1.

**Lemma 8.3.**

$$\text{b} + \text{c} = \partial_\tau\left(\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}\right)\Big(\nabla^f(fV_1)\Big)^2.$$

*Proof of Lemma 8.3.*

$$\begin{aligned}
\text{b} &= \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}\Big(6f\nabla^f V_1 \Delta^f V_1 + 6\psi(\nabla^f V_1)^2\Big) \\
&= \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}\Big(3f\nabla^f(\nabla^f V_1)^2 + 6\psi(\nabla^f V_1)^2\Big) \\
&= 3\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}\Big(\nabla^f(f(\nabla^f V_1)^2) + \psi(\nabla^f V_1)^2\Big) \\
&= -3\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}f\partial_\tau f(\nabla^f V_1)^2 + 3\frac{\psi^2}{(1+\psi^2)^{\frac{1}{2}}}(\nabla^f V_1)^2 \\
&= -\frac{3}{2}\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}\partial_\tau(f^2)(\nabla^f V_1)^2 + 3\frac{\psi^2}{(1+\psi^2)^{\frac{1}{2}}}(\nabla^f V_1)^2.
\end{aligned}$$

$$\begin{aligned}
\text{c} &= -2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}(2\psi\nabla^f V_1 + f\Delta^f V_1)\partial_\eta V_1 \\
&= -2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}(2\psi\nabla^f V_1 + f\Delta^f V_1)(\nabla^f V_1 - f\partial_\tau V_1) \\
&= -4\frac{\psi^2}{(1+\psi^2)^{\frac{1}{2}}}(\nabla^f V_1)^2 + 4\frac{\psi^2}{(1+\psi^2)^{\frac{1}{2}}}f\nabla^f V_1 \partial_\tau V_1 + \\
&\qquad - 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}f\Delta^f V_1 \nabla^f V_1 + 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}f^2\Delta^f V_1 \partial_\tau V_1.
\end{aligned}$$

We have two particular terms in this expression:

$$\begin{aligned}
\text{@} &:= 4\frac{\psi^2}{(1+\psi^2)^{\frac{1}{2}}}f\nabla^f V_1 \partial_\tau V_1 + 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}f^2\Delta^f V_1 \partial_\tau V_1 \\
&= 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}\nabla^f(f^2\nabla^f V_1)\partial_\tau V_1 \\
&= -2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}(f^2\nabla^f V_1)\nabla^f\partial_\tau V_1 - 2\partial_\tau f\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}(f^2\nabla^f V_1)\partial_\tau V_1 \\
&= -2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}f^2\nabla^f V_1(\nabla^f\partial_\tau V_1 + \partial_\tau f\partial_\tau V_1)
\end{aligned}$$

$$
\begin{aligned}
&= -2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} f^2 \nabla^f V_1 \partial_\tau \nabla^f V_1 \\
&= -\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} f^2 \partial_\tau (\nabla^f V_1)^2 \\
&= \partial_\tau \left( \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \right) f^2 (\nabla^f V_1)^2 + \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \partial_\tau (f^2) (\nabla^f V_1)^2
\end{aligned}
$$

and

$$
\begin{aligned}
\beta := &-2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} f \Delta^f V_1 \nabla^f V_1 \\
&= -\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} f \nabla^f (\nabla^f V_1)^2 \\
&= \frac{\psi^2}{(1+\psi^2)^{\frac{1}{2}}} (\nabla^f V_1)^2 + \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} f \partial_\tau f (\nabla^f V_1)^2 \\
&= \frac{\psi^2}{(1+\psi^2)^{\frac{1}{2}}} (\nabla^f V_1)^2 + \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \frac{\partial_\tau (f^2)}{2} (\nabla^f V_1)^2 .
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
c = &-4\frac{\psi^2}{(1+\psi^2)^{\frac{1}{2}}} (\nabla^f V_1)^2 + \alpha + \beta \\
= -3\frac{\psi^2}{(1+\psi^2)^{\frac{1}{2}}} (\nabla^f V_1)^2 &+ \partial_\tau \left( \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \right) f^2 (\nabla^f V_1)^2 + \\
&+ \frac{3}{2} \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \partial_\tau (f^2) (\nabla^f V_1)^2 .
\end{aligned}
$$

Putting this together,

$$
\begin{aligned}
b + c &= \partial_\tau \left( \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \right) f^2 (\nabla^f V_1)^2 \\
&= \partial_\tau \left( \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \right) (\nabla^f (f V_1) - \psi V_1)^2 \\
&= \partial_\tau \left( \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \right) ((\nabla^f (f V_1))^2 + (\psi V_1)^2 - 2\nabla^f (f V_1) \psi V_1) \\
&= \partial_\tau \left( \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \right) ((\nabla^f (f V_1))^2 - (\psi V_1)^2 - f \nabla^f (\psi V_1^2)) \\
&\overset{(*)}{=} \partial_\tau \left( \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \right) ((\nabla^f (f V_1))^2 - (\psi V_1)^2 + \nabla^f f \psi V_1^2) \\
&= \partial_\tau \left( \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \right) (\nabla^f (f V_1))^2 .
\end{aligned}
$$

In $(*)$ we used formula (18). $\square$

**Lemma 8.4.**

$$ⓓ = \partial_\tau \left( \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \right) (\nabla^f V_2)^2.$$

*Proof of Lemma 8.4.*

$$\begin{aligned} ⓓ &= 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \Delta^f V_2 \partial_\tau V_2 \\ &\overset{(*)}{=} -2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \nabla^f V_2 \partial_\tau \nabla^f V_2 \\ &= -\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \partial_\tau (\nabla^f V_2)^2 \\ &= \partial_\tau \left( \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \right) (\nabla^f V_2)^2. \end{aligned}$$

In $(*)$ we used formula (18). □

**Lemma 8.5.**

$$ⓔ + ⓕ + ⓖ = -2\partial_\tau \left( \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \right) \nabla^f (f V_1) \nabla^f V_2.$$

*Proof.*

$$\begin{aligned} ⓖ &= 2(1+\psi^2)^{\frac{1}{2}} (\partial_\eta V_1 \partial_\tau V_2 - \partial_\tau V_1 \partial_\eta V_2) \\ &= 2(1+\psi^2)^{\frac{1}{2}} ((\nabla^f V_1 - f\partial_\tau V_1)\partial_\tau V_2 - \partial_\tau V_1 (\nabla^f V_2 - f \partial_\tau V_2)) \\ &= 2(1+\psi^2)^{\frac{1}{2}} (\nabla^f V_1 \partial_\tau V_2 - \partial_\tau V_1 \nabla^f V_2). \end{aligned}$$

$$\begin{aligned} ⓕ &= 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \Big( \Delta^f V_2 \partial_\eta V_1 + (-2\psi \nabla^f V_1 - f \Delta^f V_1) \partial_\tau V_2 \Big) \\ &= 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \Big( \Delta^f V_2 (\nabla^f V_1 - f\partial_\tau V_1) - 2\psi \nabla^f V_1 \partial_\tau V_2 - f \Delta^f V_1 \partial_\tau V_2 \Big) \\ &= 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \Big( -f\partial_\tau V_1 \Delta^f V_2 - f \partial_\tau V_2 \Delta^f V_1 + \Delta^f V_2 \nabla^f V_1 - 2\psi \nabla^f V_1 \partial_\tau V_2 \Big) \\ &\overset{(*)}{=} 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \Big( \psi \partial_\tau V_1 \nabla^f V_2 + f \partial_\tau (\nabla^f V_1) \nabla^f V_2 + \psi \partial_\tau V_2 \nabla^f V_1 + \\ &\qquad + f\partial_\tau (\nabla^f V_2) \nabla^f V_1 + \Delta^f V_2 \nabla^f V_1 - 2\psi \nabla^f V_1 \partial_\tau V_2 \Big) \\ &= 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \Big( \psi \partial_\tau V_1 \nabla^f V_2 + f \partial_\tau (\nabla^f V_1) \nabla^f V_2 + \\ &\qquad + f\partial_\tau (\nabla^f V_2) \nabla^f V_1 + \Delta^f V_2 \nabla^f V_1 - \psi \nabla^f V_1 \partial_\tau V_2 \Big). \end{aligned}$$

In $(*)$ we used formula (18).

$$ⓕ + ⓖ = 2\frac{(1+\psi^2)}{(1+\psi^2)^{\frac{1}{2}}} (\nabla^f V_1 \partial_\tau V_2 - \partial_\tau V_1 \nabla^f V_2) +$$

$$+2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}\Big(\psi\partial_\tau V_1\nabla^f V_2 + f\partial_\tau(\nabla^f V_1)\nabla^f V_2 + $$
$$+ f\partial_\tau(\nabla^f V_2)\nabla^f V_1 + \Delta^f V_2\nabla^f V_1 - \psi\nabla^f V_1\partial_\tau V_2\Big)$$
$$= 2\frac{1}{(1+\psi^2)^{\frac{1}{2}}}(\nabla^f V_1\partial_\tau V_2 - \partial_\tau V_1\nabla^f V_2) +$$
$$+2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}\Big(f\partial_\tau(\nabla^f V_1\nabla^f V_2) + \Delta^f V_2\nabla^f V_1\Big).$$

$$ⓔ + ⓕ + ⓖ = \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}\Big(-4\Delta^f V_2\nabla^f V_1 - 2\nabla^f V_2\Delta^f V_1\Big) + ⓕ + ⓖ$$
$$= 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}\Big(-\Delta^f V_2\nabla^f V_1 - \nabla^f V_2\Delta^f V_1 + f\partial_\tau(\nabla^f V_1\nabla^f V_2)\Big) +$$
$$+2\frac{1}{(1+\psi^2)^{\frac{1}{2}}}(\nabla^f V_1\partial_\tau V_2 - \partial_\tau V_1\nabla^f V_2)$$
$$= 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}\Big(-\nabla^f(\nabla^f V_2\nabla^f V_1) + f\partial_\tau(\nabla^f V_1\nabla^f V_2)\Big) +$$
$$+2\frac{1}{(1+\psi^2)^{\frac{1}{2}}}(\nabla^f V_1\partial_\tau V_2 - \partial_\tau V_1\nabla^f V_2)$$
$$= 2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}\partial_\tau f\nabla^f V_2\nabla^f V_1 - 2\partial_\tau\left(\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}\right) f\nabla^f V_1\nabla^f V_2 +$$
$$-2\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}\partial_\tau f\nabla^f V_1\nabla^f V_2 + 2\frac{1}{(1+\psi^2)^{\frac{1}{2}}}(\nabla^f V_1\partial_\tau V_2 - \partial_\tau V_1\nabla^f V_2)$$
$$= -2\partial_\tau\left(\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}\right) f\nabla^f V_1\nabla^f V_2 + 2\frac{1}{(1+\psi^2)^{\frac{1}{2}}}(\nabla^f V_1\partial_\tau V_2 - \partial_\tau V_1\nabla^f V_2).$$

In particular, we have

$$\frac{1}{(1+\psi^2)^{\frac{1}{2}}}(\nabla^f V_1\partial_\tau V_2 - \partial_\tau V_1\nabla^f V_2) = -\frac{1}{(1+\psi^2)^{\frac{1}{2}}}\partial_\tau\nabla^f V_2 V_1 +$$
$$+\partial_\tau\left(\frac{1}{(1+\psi^2)^{\frac{1}{2}}}\right)\nabla^f V_2 V_1 + \frac{1}{(1+\psi^2)^{\frac{1}{2}}}\partial_\tau\nabla^f V_2 V_1$$
$$= \partial_\tau\left(\frac{1}{(1+\psi^2)^{\frac{1}{2}}}\right)\nabla^f V_2 V_1$$

and

$$\partial_\tau\left(\frac{1}{(1+\psi^2)^{\frac{1}{2}}}\right) = -\frac{1}{(1+\psi^2)^{\frac{3}{2}}}\psi\partial_\tau\psi = -\psi\partial_\tau\left(\frac{\psi}{(1+\psi^2)^{\frac{1}{2}}}\right).$$

Therefore

$$ⓔ + ⓕ + ⓖ =$$

$$= -2\partial_\tau \left( \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \right) f \nabla^f V_1 \nabla^f V_2 + 2 \frac{1}{(1+\psi^2)^{\frac{1}{2}}} (\nabla^f V_1 \partial_\tau V_2 - \partial_\tau V_1 \nabla^f V_2)$$

$$= -2\partial_\tau \left( \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \right) f \nabla^f V_1 \nabla^f V_2 - 2\psi \partial_\tau \left( \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \right) \nabla^f V_2 V_1$$

$$= -2\partial_\tau \left( \frac{\psi}{(1+\psi^2)^{\frac{1}{2}}} \right) \nabla^f (f V_1) \nabla^f V_2.$$

□

## Appendix A. Useful formulas

In the case $f \in \mathscr{C}^\infty(\mathbb{R}^2)$, the adjoint operator of $\nabla^f$ is

$$(\nabla^f)^* = -\nabla^f - \partial_\tau f,$$

i.e., if $A, B \in \mathscr{C}^\infty(\mathbb{R}^2)$ and one of them has compact support, then

$$\int_{\mathbb{R}^2} A \cdot \nabla^f B \,\mathrm{d}\eta \,\mathrm{d}\tau = -\int_{\mathbb{R}^2} \left[ \nabla^f A \cdot B + \partial_\tau f \cdot A \cdot B \right] \mathrm{d}\eta \,\mathrm{d}\tau.$$

Notice that, if $f$ is smooth, the following holds:

$$\partial_\eta = \nabla^f - f\partial_\tau,$$

$$\partial_\tau \nabla^f = \nabla^f \partial_\tau + \partial_\tau f \partial_\tau.$$

If $A, B, C \in \mathscr{C}^\infty(\mathbb{R}^2)$ and one of them has compact support, then

(18)

$$\int_{\mathbb{R}^2} A \cdot \partial_\tau B \cdot \nabla^f C \,\mathrm{d}\eta \,\mathrm{d}\tau = -\int_{\mathbb{R}^2} \left( \nabla^f A \cdot \partial_\tau B \cdot C + A \cdot \partial_\tau \nabla^f B \cdot C \right) \mathrm{d}\eta \,\mathrm{d}\tau.$$

## References

[1] L. Ambrosio, F. Serra Cassano, and D. Vittone. "Intrinsic regular hypersurfaces in Heisenberg groups". In: *J. Geom. Anal.* 16.2 (2006), pp. 187–232.

[2] V. Barone Adesi, F. Serra Cassano, and D. Vittone. "The Bernstein problem for intrinsic graphs in Heisenberg groups and calibrations". In: *Calc. Var. Partial Differential Equations* 30.1 (2007), pp. 17–49.

[3] L. Capogna, D. Danielli, S. D. Pauls, and J. T. Tyson. *An introduction to the Heisenberg group and the sub-Riemannian isoperimetric problem.* Vol. 259. Progress in Mathematics. Birkhäuser Verlag, Basel, 2007, pp. xvi+223.

[4] D. Danielli, N. Garofalo, D. M. Nhieu, and S. D. Pauls. "Instability of graphical strips and a positive answer to the Bernstein problem in the Heisenberg group $\mathbb{H}^1$". In: *J. Differential Geom.* 81.2 (2009), pp. 251–295.

[5] M. Fogagnolo, R. Monti, and D. Vittone. "Variation formulas for $H$-rectifiable sets". In: *Ann. Acad. Sci. Fenn. Math.* 42.1 (2017), pp. 239–256.

[6] B. Franchi, R. Serapioni, and F. Serra Cassano. "Rectifiability and perimeter in the Heisenberg group". In: *Math. Ann.* 321.3 (2001), pp. 479–531.

[7] B. Franchi, R. Serapioni, and F. Serra Cassano. "Regular submanifolds, graphs and area formula in Heisenberg groups". In: *Adv. Math.* 211.1 (2007), pp. 152–203.

[8] M. Galli. "First and second variation formulae for the sub-Riemannian area in three-dimensional pseudo-Hermitian manifolds". In: *Calc. Var. Partial Differential Equations* 47.1-2 (2013), pp. 117–157.

[9] M. Galli and M. Ritoré. "Area-stationary and stable surfaces of class $C^1$ in the sub-Riemannian Heisenberg group $\mathbb{H}^1$". In: *Adv. Math.* 285 (2015), pp. 737–765.

[10] N. Garofalo and D.-M. Nhieu. "Isoperimetric and Sobolev inequalities for Carnot-Carathéodory spaces and the existence of minimal surfaces". In: *Comm. Pure Appl. Math.* 49.10 (1996), pp. 1081–1144.

[11] B. Kirchheim and F. Serra Cassano. "Rectifiability and parameterization of intrinsic regular surfaces in the Heisenberg group". In: *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* 3.4 (2004), pp. 871–896.

[12] A. Korányi and H. M. Reimann. "Foundations for the theory of quasiconformal mappings on the Heisenberg group". In: *Adv. Math.* 111.1 (1995), pp. 1–87.

[13] E. Le Donne and S. Nicolussi Golo. *Regularity properties of spheres in homogeneous groups.* In press on *Trans. Amer. Math. Soc.* Preprint available on arXiv: 1509.03881.

[14] E. Le Donne. "A primer on Carnot groups: homogenous groups, CC spaces, and regularity of their isometries". In: *ArXiv e-prints* (Apr. 2016). arXiv: 1604.08579.

[15] F. Montefalcone and F. Serra Cassano. "Intrinsic variational formulae for sets of finite H -perimeter in 2-step Carnot Groups". Work in progress. 2015.

[16] R. Monti. "Isoperimetric problem and minimal surfaces in the Heisenberg group". In: *Geometric measure theory and real analysis.* Vol. 17. CRM Series. Ed. Norm., Pisa, 2014, pp. 57–129.

[17] R. Monti, F. Serra Cassano, and D. Vittone. "A negative answer to the Bernstein problem for intrinsic graphs in the Heisenberg group". In: *Boll. Unione Mat. Ital. (9)* 1.3 (2008), pp. 709–727.

[18] F. Serra Cassano. "Some topics of geometric measure theory in Carnot groups". In: *Geometry, analysis and dynamics on sub-Riemannian manifolds. Vol. 1.* EMS Ser. Lect. Math. Eur. Math. Soc., Zürich, 2016, pp. 1–121.


(Golo) University of Jyvaskyla, Finland; University of Trento, Italy

130. KEISALA, JUKKA, Existence and uniqueness of $p(x)$-harmonic functions for bounded and unbounded $p(x)$. (56 pp.) 2011
131. SEPPÄLÄ, HEIKKI, Interpolation spaces with parameter functions and $L_2$-approximations of stochastic integrals. (18 pp.) 2011
132. TUHOLA-KUJANPÄÄ, ANNA, On superharmonic functions and applications to Riccati type equations. (17 pp.) 2012
133. JIANG, RENJIN, Optimal regularity of solutions to Poisson equations on metric measure spaces and an application. (13 pp.) 2012
134. TÖRMÄKANGAS, TIMO, Simulation study on the properties of quantitative trait model estimation in twin study design of normally distributed and discrete event-time phenotype variables. (417 pp.) 2012
135. ZHANG, GUO, Liouville theorems for stationary flows of generalized Newtonian fluids. (14 pp.) 2012
136. RAJALA, TUOMAS, Use of secondary structures in the analysis of spatial point patterns. (27 pp.) 2012
137. LAUKKARINEN, EIJA, On Malliavin calculus and approximation of stochastic integrals for Lévy processes. (21 pp.) 2012
138. GUO, CHANGYU, Generalized quasidisks and the associated John domains. (17 pp.) 2013
139. ÄKKINEN, TUOMO, Mappings of finite distortion: Radial limits and boundary behavior. (14 pp.) 2014
140. ILMAVIRTA, JOONAS, On the broken ray transform. (37 pp.) 2014
141. MIETTINEN, JARI, On statistical properties of blind source separation methods based on joint diagonalization. (37 pp.) 2014
142. TENGVALL, VILLE, Mappings of finite distortion: Mappings in the Sobolev space $W^{1,n-1}$ with integrable inner distortion. (22 pp.) 2014
143. BENEDICT, SITA, Hardy-Orlicz spaces of quasiconformal mappings and conformal densities. (16 pp.) 2014
144. OJALA, TUOMO, Thin and fat sets: Geometry of doubling measures in metric spaces. (19 pp.) 2014
145. KARAK, NIJJWAL, Applications of chaining, Poincaré and pointwise decay of measures. (14 pp.) 2014
146. JYLHÄ, HEIKKI, On generalizations of Evans and Gangbo's approximation method and $L^\infty$ transport. (20 pp.) 2014
147. KAURANEN, AAPO, Space-filling, energy and moduli of continuity. (16 pp.) 2015
148. YLINEN, JUHA, Decoupling on the Wiener space and variational estimates for BSDEs. (45 pp.) 2015
149. KIRSILÄ, VILLE, Mappings of finite distortion on generalized manifolds. (14 pp.) 2015
150. XIANG, CHANG-LIN, Asymptotic behaviors of solutions to quasilinear elliptic equations with Hardy potential. (20 pp.) 2015
151. ROSSI, EINO, Local structure of fractal sets: tangents and dimension. (16 pp.) 2015
152. HELSKE, JOUNI, Prediction and interpolation of time series by state space models. (28 pp.) 2015
153. REINIKAINEN, JAAKKO, Efficient design and modeling strategies for follow-up studies with time-varying covariates. (36 pp.) 2015
154. NUUTINEN, JUHO, Maximal operators and capacities in metric spaces. (22 pp.) 2016
155. BRANDER, TOMMI, Calderón's problem for $p$-Laplace type equations. (21 pp.) 2016
156. ÄRJE, JOHANNA, Improving statistical classification methods and ecological status assessment for river macroinvertebrates. (30 pp.) 2016
157. HELSKE, SATU, Statistical analysis of life sequence data. (40 pp.) 2016
158. RUOSTEENOJA, EERO, Regularity properties of tug-of-war games and normalized equations. (16 pp.) 2017
159. ZHANG, YI, Planar Sobolev extension domains. (13 pp.) 2017
160. YLITALO, ANNA-KAISA, Statistical inference for eye movement sequences using spatial and spatio-temporal point processes. (45 pp.) 2017
161. LINDQVIST, PETER, Notes on the p-Laplace equation. (106 pp.) 2017
162. LEHTONEN, JERE, Injectivity results for the geodesic ray transform. (15 pp.) 2017

ISBN 978-951-39-7148-9 (print)
ISBN 978-951-39-7149-6 (pdf)
ISSN 1457-8905