HOW MANY GEODESICS JOIN TWO POINTS ON A CONTACT SUB-RIEMANNIAN MANIFOLD?

A. LERARIO AND L. RIZZI

Abstract. We investigate the structure and the topology of the set of geodesics (critical points for the *energy functional*) between two points on a contact Carnot group G (or, more generally, corank-one Carnot groups). Denoting by $(x, z) \in \mathbb{R}^{2n} \times \mathbb{R}$ exponential coordinates on G, we find constants $C_1, C_2 > 0$ and R_1, R_2 such that the number $\hat{\nu}(p)$ of geodesics joining the origin with a generic point $p = (x, z)$ satisfies:

$$C_1 \frac{|z|}{\|x\|^2} + R_1 \leq \hat{\nu}(p) \leq C_2 \frac{|z|}{\|x\|^2} + R_2. \qquad (1)$$

We give conditions for p to be joined by a unique geodesic and we specialize our computations to standard Heisenberg groups, where $C_1 = C_2 = \frac{8}{\pi}$.

The set of geodesics joining the origin with $p \neq p_0$, parametrized with their initial covector, is a topological space $\Gamma(p)$, that naturally splits as the disjoint union

$$\Gamma(p) = \Gamma_0(p) \cup \Gamma_\infty(p),$$

where $\Gamma_0(p)$ is a finite set of isolated geodesics, while $\Gamma_\infty(p)$ contains continuous families of non-isolated geodesics (critical *manifolds* for the energy). We prove an estimate similar to (1) for the "topology" (i.e. the total Betti number) of $\Gamma(p)$, with no restriction on p.

When G is the Heisenberg group, families appear if and only if p is a *vertical* nonzero point and each family is generated by the action of isometries on a given geodesic. Surprisingly, in more general cases, families of *non-isometrically equivalent* geodesics do appear.

If the Carnot group G is the *nilpotent approximation* of a contact sub-Riemannian manifold M at a point p_0, we prove that the number $\nu(p)$ of geodesics in M joining p_0 with p can be estimated from below with $\hat{\nu}(p)$. The number $\nu(p)$ estimates indeed geodesics whose image is contained in a coordinate chart around p_0 (we call these "affine" geodesics).

As a corollary we prove the existence of a sequence $\{p_n\}_{n\in\mathbb{N}}$ in M such that:

$$\lim_{n\to\infty} p_n = p_0 \qquad \text{and} \qquad \lim_{n\to\infty} \nu(p_n) = \infty,$$

i.e. the number of "affine" geodesics between two points can be arbitrarily large, in sharp contrast with the Riemannian case.

1. Introduction

If the topology of a Riemannian manifold M is "complicated enough" (for example if M is closed) a well known theorem of J-P. Serre [19] states that there are infinitely many geodesics[1] between any two points in M. These geodesics have the property of being "global", in the sense that their existence is guaranteed by the global topology of the manifold (see [17] for a recent quantitative proof of this result).

Somehow on the opposite, if the manifold M is a convex neighbourhood of a point in a Riemannian manifold, then between any two points in M there is only one geodesic: the local geometry of M is Euclidean and the structure of geodesics contained in a convex chart resembles the Euclidean one as well.

[1] In the spirit of Morse theory, we define Riemannian geodesics as locally length minimizing curves or, equivalently, critical points for the energy functional.

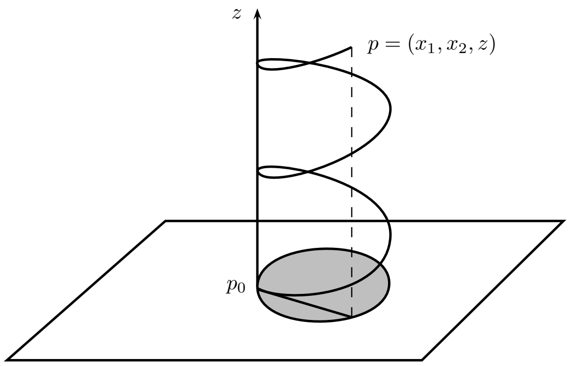

FIGURE 1. Geodesics in the Heisenberg group.

In the *subriemannian* case the situation dramatically changes. The study of geodesics that "loop" in the topology of the manifold was recently done by the first author and F. Boarotto in [7]: every two points on a compact sub-Riemannian contact manifold are joined by infinitely many geodesics (the result uses a weak homotopy equivalence between the space of all curves and the space of horizontal ones). Thus in a sense the *global* picture in the sub-Riemannian case is the same as the Riemannian. On the opposite, our main interest here will be in the set of "affine" geodesics, i.e. geodesics between two points whose image is contained in a coordinate chart: as we will see new interesting phenomena appear for their structure.

In this framework we consider a constant-rank distribution $\mathcal{D} \subset TM$ with the property that iterated brackets of vector fields on $\mathcal{D}$ generate the all tangent space (Hörmander's condition). This condition guarantees that any two points in M can be joined by a Lipschitz continuous curve whose velocity is a.e. in $\mathcal{D}$ (Chow-Rashevskii theorem).

The set of horizontal curves defined on the same interval I and joining two points p_0 and p is a smooth Hilbert manifold (for the generic choice of the two points). If a smooth scalar product is defined on $\mathcal{D}$, it makes sense to consider, for any horizontal curve γ, the norm of its velocity and the *energy* of this curve is defined by:

$$J(\gamma) = \frac{1}{2}\int_I \|\dot{\gamma}(t)\|^2 dt.$$

Sub-Riemannian *geodesics* between p_0 and p are critical points of J *constrained* to have endpoints p_0 and p. From now on the word geodesic will always mean sub-Riemannian geodesic.

Example 1 (Heisenberg). The *Heisenberg group* $\mathbb{H}_3$ is the smooth manifold $\mathbb{R}^3$ with coordinates (x_1, x_2, z) and the distribution:

$$\mathcal{D} = \operatorname{span}\left\{\frac{\partial}{\partial x_1} - \frac{x_2}{2}\frac{\partial}{\partial z}, \frac{\partial}{\partial x_2} + \frac{x_1}{2}\frac{\partial}{\partial z}\right\}.$$

The sub-Riemannian structure is given by declaring the above vector fields an orthonormal basis; higher dimensional Heisenberg groups $\mathbb{H}_{2n+1}$ are defined analogously.

Let $p_0 = (0, 0, 0)$ be the origin. Constrained critical points of the energy are curves whose projection on the (x_1, x_2)-plane is an arc of a circle (possibly with infinite radius, i.e. a segment on a straight line); the signed area swept out on the circle by this projection equals the z-coordinate of the final point p.

If p belongs to the (x_1, x_2)-plane there is only one geodesic joining it with the origin (this is precisely the segment trough p_0 and p); if p has both nonzero components in the (x_1, x_2) plane and the z axis, the number of geodesics is finite; finally, if p belongs to the z-axis there are infinitely many geodesics. In the latter case, given one geodesic, we obtain infinitely many others (a *continuous family*) by composing it with a rotation around the z-axis (see Figure 1).

The following fact is proved in [5, Proposition 7].

Proposition 1. *Let M be a step-two Carnot group such that $\mathrm{rk}(\mathcal{D}) > \frac{1}{2}\dim(M)$. Then for the generic choice of p_0 and p the number of geodesics between them is finite.*

The above proposition does not give a quantitative estimate and in general very little is known on the problem: a Sard's like argument for the sub-Riemannian exponential guarantees that for the generic choice of the two points geodesics are isolated, but finiteness is more delicate (this is related with the fact that the sub-Riemannian exponential might not be a proper map).

The goal of this paper is to make the above picture quantitative, at least in the case of contact[2] sub-Riemannian manifolds. We address the following question:

"How many geodesics join two points on a contact sub-Riemannian manifold?"

A contact sub-Riemannian manifold is the simplest example of nonholonomic geometry. From the point of view of differential geometry it consists of a $(2n + 1)$-dimensional, connected manifold M together with a distribution $\mathcal{D} \subset TM$ of hyperplanes locally defined as the kernel of a one-form α (the *contact form*) such that the restriction $d\alpha|_{\mathcal{D}}$ is non-degenerate. The sub-Riemannian structure is given by assigning a smooth metric on the hyperplane distribution. The non-degeneracy condition implies Hörmander's condition.

Example 2 (Heisenberg, continuation). The Heisenberg group is a contact manifold with contact form $\alpha = -dz + \frac{1}{2}(x_1 dx_2 - x_2 dx_1)$. As we will show later:

$$\#\{\text{geodesics between the origin and } p = (x_1, x_2, z)\} = \frac{8}{\pi}\frac{|z|}{\|x\|^2} + O(1). \tag{2}$$

In particular when p is "vertical", $p = (0, 0, z)$ the number of geodesics is infinite; otherwise it is finite and equals the r.h.s. (the $O(1)$ notation means "up to a bounded error").

For any point $p_0 \in M$ one can consider the so-called *nilpotent approximation* of the sub-Riemannian structure at p_0. The result of this construction (that depends only on the germ of the structure at p_0) is a sub-Riemannian manifold G_{p_0}, and is an example of a *Carnot group.*

One of our results (Theorem 7) states that the geodesic count on the Carnot group G_{p_0} controls the geodesic count on the original manifold M. For this reason, we start our analysis with the study of *contact Carnot groups*, namely Carnot groups arising as the nilpotent approximation of contact manifolds.

1.1. **Carnot groups.** A contact Carnot group is a connected, simply connected Lie group G, with $\dim G = 2n+1$, such that its Lie algebra $\mathfrak{g}$ of left-invariant vector fields admits a nilpotent stratification of step 2, namely:

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2, \qquad \mathfrak{g}_1, \mathfrak{g}_2 \neq \{0\},$$

with

$$[\mathfrak{g}_1, \mathfrak{g}_1] = \mathfrak{g}_2 \qquad \text{and} \qquad [\mathfrak{g}_1, \mathfrak{g}_2] = [\mathfrak{g}_2, \mathfrak{g}_2] = \{0\}.$$

[2]We stress here that all our results remain true with almost no modification for more general *corank-one* sub-Riemannian structure. For simplicity we restrict our exposition to the contact case.

A sub-Riemannian structure on G is obtained by defining a scalar product on $\mathfrak{g}_1$ or, equivalently, by declaring a set $f_1, \ldots, f_{2n} \in \mathfrak{g}_1$ a global orthonormal frame. The group exponential map:

$$\exp_G : \mathfrak{g} \to G,$$

associates with $v \in \mathfrak{g}$ the element $\gamma(1)$, where $\gamma : [0,1] \to G$ is the unique integral line of the vector field defined by v such that $\gamma(0) = 0$. Since G is simply connected and $\mathfrak{g}$ is nilpotent, $\exp_G$ is a smooth diffeomorphism. The choice of an orthonormal frame $f_1, \ldots, f_{2n} \in \mathfrak{g}_1$ and $f_0 \in \mathfrak{g}_2$ defines *exponential coordinates* $(x, z) \in \mathbb{R}^{2n} \times \mathbb{R}$ on G such that $p = (x, z)$ if and only if

$$p = \exp_G \left(\sum_{i=1}^{2n} x_i f_i + z f_0 \right).$$

For any such a choice there exists a skew-symmetric matrix $A \in \mathfrak{so}(2n)$ such that

$$[f_i, f_j] = A_{ij} f_0.$$

For contact Carnot groups A is non-degenerate. We denote by:

$$\alpha_1 < \cdots < \alpha_k \in \mathbb{R}_+$$

the distinct singular values of A and n_j their multiplicities. Let $x_j \in \mathbb{R}^{2n_j}$ be the projections of x on the invariant subspaces associated with α_j. Accordingly we write $p = (x_1, \ldots, x_k, z)$. The geodesic count for G can be made quite explicit in term of the exponential coordinates of p and the singular values of the matrix A. Define for this purpose the following "counting" function:

$$\hat{\nu}(p) = \#\{\text{geodesics in a Carnot group between the origin and } p\},$$

where, by convention, the "hat" stresses the fact that we refer to a Carnot group. We have the following estimates for $\hat{\nu}(p)$ (see Theorems 24-26). None of these bounds is trivial: the upper bound because the exponential map is not proper; the lower bound is in fact even more surprising, as the typical finiteness techniques from semialgebraic (semianalytic) geometry only produce upper bounds (we use indeed a kind of "ergodicity" property argument).

Theorem 2 (The "infinitesimal" bound)**.** *Given a contact Carnot group G, there exist constants $C_1, C_2 > 0$ and R_1, R_2 such that if $p = (x, z) \in G$ is a point with all components x_j different from zero, then:*

$$C_1 \frac{|z|}{\|x\|^2} + R_1 \leq \hat{\nu}(p) \leq C_2 \frac{|z|}{\|x\|^2} + R_2. \tag{3}$$

In fact C_1, C_2 (resp. R_1, R_2) are homogeneous of degree -1 (resp. 0) in the singular values $\alpha_1 < \cdots < \alpha_k$ of A and are given by:

$$C_1 = \frac{8}{\pi} \frac{\alpha_1}{\alpha_k^2} \sin\left(\frac{\delta\pi}{2}\right)^2 \qquad \text{with} \qquad \delta = \left(\sum_{j=1}^{k} \frac{\alpha_1}{\alpha_j} \left\lfloor \frac{\alpha_j}{\alpha_1} \right\rfloor \right)^{-1} \qquad \text{and} \qquad C_2 = \frac{8k}{\pi} \frac{\alpha_k}{\alpha_1^2}. \tag{4}$$

Remark 1. For any other choice of $f_1', \ldots, f_{2n}' \in \mathfrak{g}_1$ (orthonormal) and a complement $f_0' \in \mathfrak{g}_2$ there exists a matrix $M \in \mathrm{O}(2n)$ and a constant c such that:

$$f_i = \sum_{j=1}^{2n} M_{ij} f_j', \qquad f_0 = c f_0'.$$

Indeed this new choice defines a new skew-symmetric matrix A' and also new exponential coordinates (x', z'). One can easily check that:

$$A' = cM^* AM, \qquad x' = M^* x, \qquad z' = cz.$$

Since C_1, C_2 are homogeneous functions of degree -1 in the singular values of A, the upper and lower bounds (3) are invariant w.r.t. different choices of exponential coordinates.

Example 3 (Heisenberg, continuation). In the Heisenberg group $\mathbb{H}_{2n+1}$ there is only one singular value $\alpha = 1$, with multiplicity n. By using (3) and (4) one obtains:

$$C_1 = C_2 = \frac{8}{\pi},$$

recovering (2) (that holds true for any Heisenberg group, not just the three-dimensional one).

An interesting related question is to determine the set of points p such that $\hat{\nu}(p) = 1$ (as it happens for example if $p = (x, 0)$, i.e. p is horizontal). In the Heisenberg group:

$$\hat{\nu}(p) = 1 \iff \frac{|z|}{\|x\|^2} \leq \frac{\lambda_1}{4} \approx 1.12335,$$

where λ_1 is the first positive solution of $\tan\lambda = \lambda$.

For a general contact Carnot group the following proposition gives a quantitative way to estimate when there is only one geodesic (see Proposition 25).

Proposition 3. *Let G be a contact Carnot group and $p = (x, z)$ such that:*

$$|z| < \frac{\pi}{8}\left(\frac{2\alpha_1^2}{\alpha_k} - \alpha_k\right)\|x\|^2.$$

Then there is only one geodesic from p_0 to p. Moreover if $G = \mathbb{H}_{2n+1}$, let λ_1 be the first positive solution of $\lambda = \tan\lambda$. Then there is only one geodesic to p if and only if:

$$|z| \leq \frac{\lambda_1}{4}\|x\|^2.$$

1.2. **Critical manifolds.** It is interesting to discuss the structure of *all* geodesics ending at p, including the case when the point p belongs to a hyperplane coordinate space (i.e. $x_j = 0$ for some j), which was excluded from Theorem 2. We still exclude the case $p = p_0$, as for the case of Carnot groups there is only one geodesic: the trivial one $\gamma(t) \equiv p_0$.

Sub-Riemannian geodesics starting from p_0 are parametrized by their initial covector $\eta \in T^*_{p_0}M$. As such, the subset $\Gamma(p)$ of geodesics ending at p is naturally a topological space. We have the following characterization (see Theorem 23).

Theorem 4 (Topology of critical manifolds)**.** *Let G be a contact Carnot group. The set $\Gamma(p)$ of geodesics ending at $p \neq p_0$ can be decomposed into the disjoint union of two closed submanifolds:*

$$\Gamma(p) = \Gamma_0(p) \cup \Gamma_\infty(p).$$

The set $\Gamma_0(p)$ is finite and the set $\Gamma_\infty(p)$ is homeomorphic to a union of spheres. Moreover the energy function J is constant on each component of $\Gamma(p)$.

Remark 2. The structure of the sets of geodesics whose final point is vertical, in the general step-two Carnot group of type (k, n) is studied in [5]. Geodesics to p are critical points for the energy functional $J : \Omega_p \to \mathbb{R}$ (here Ω_p is the space of all admissible curves to p and J is defined as above); for the generic vertical p these geodesics appear in families, which are tori of finite dimension depending on the "multiplicity" of the Lagrange multiplier (in particular they are never isolated and J is a Morse-Bott function). A Morse theoretical study is performed and it is proved that:

$$\#\{\text{critical manifolds of } J \text{ with energy less then } c\} \leq O(c^{n-k}).$$

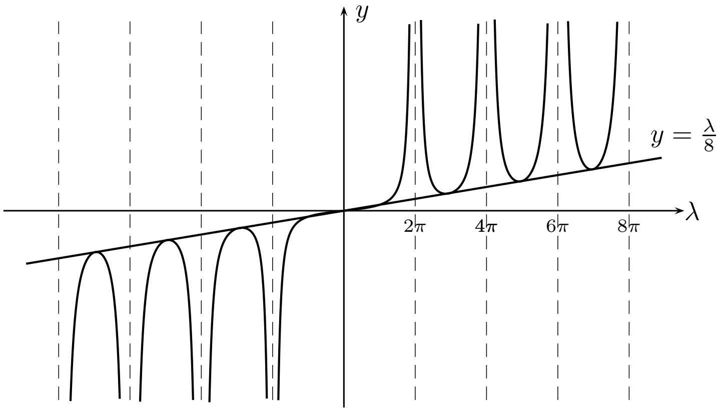

FIGURE 2. The graph of g.

On the other hand the "order of growth" of the topology[3] of $\Omega_p^c = \{\gamma \in \Omega_p \,|\, J(\gamma) \leq c\}$ (the sublevel set of the energy) turns out to be given by:

$$b(\Omega_p^c) \leq O(c^{n-k-1}),$$

an inequality which is stronger than the classical Morse-Bott prediction $b(\Omega_p^c) \leq O(c^{n-k})$.

In particular, since geodesics in $\Gamma_0(p)$ are always finite, the preimage of a regular value of $\hat{E}$ is finite. Geodesics in $\Gamma_\infty(p)$ appear in *families*. Since geodesics are critical points for the energy functional, we call each connected component of $\Gamma_\infty(p)$ a *critical family* (or *critical manifold*).

To be more precise on these continuous families, the set $\Gamma_\infty(p)$ admits the following description. Given $\alpha_1, \ldots, \alpha_k$ (the singular values of A) we define the function:

$$g(\lambda) = \frac{1}{8} \frac{\lambda - \sin\lambda}{\left(\sin\frac{\lambda}{2}\right)^2}.$$

and the sets:

$$\Lambda_j = \frac{2\pi}{\alpha_j}\mathbb{Z} \setminus \{0\}, \qquad \Lambda = \bigcup_{j=1}^{k} \Lambda_j \qquad \text{and} \qquad L(\lambda) = \{j \,|\, \lambda \in \Lambda_j\}.$$

Thus Λ_j consists of the poles of $\lambda \mapsto g(\lambda\alpha_j)$ and the set of indices $L(\lambda)$ tells how many of these poles occur at λ (see Figure 2). With these conventions we have:

$$\Gamma_\infty(p) \simeq \bigcup_{\lambda \in \Lambda_p} S^{2N(\lambda)-1}, \qquad N(\lambda) = \sum_{j \in L(\lambda)} n_j,$$

where:

$$\Lambda_p = \left\{ \lambda \in \Lambda \;\middle|\; \left(z - \sum_{x_j \neq 0} \alpha_j g(\lambda\alpha_j) \|x_j\|^2 \right) \lambda > 0 \right\}.$$

For the generic A all singular values are distinct ($k = n$) and non-commensurable, thus for every $\lambda \in \Lambda_p$ we have $\#L(\lambda) = 1$, $N(\lambda) = 1$ and all critical manifolds are homeomorphic to

[3]Here we measure the topological complexity of a space X with the sum $b(X)$ of its Betti numbers, the number of "holes" in X (see below).

	$\#\Gamma_0$	$\#\Gamma_\infty$	$\hat\nu$
all $x_j \neq 0$	$\frac{\lvert z\rvert}{\|x\|^2}$	0	$\frac{\lvert z\rvert}{\|x\|^2}$
some $x_j = 0$	$\frac{\lvert z\rvert}{\|x\|^2}$	∞	∞
$x = 0$	0	∞	∞

	$b(\Gamma_0)$	$b(\Gamma_\infty)$	$\hat\beta$
all $x_j \neq 0$	$\frac{\lvert z\rvert}{\|x\|^2}$	0	$\frac{\lvert z\rvert}{\|x\|^2}$
some $x_j = 0$	$\frac{\lvert z\rvert}{\|x\|^2}$	$\frac{\lvert z\rvert}{\|x\|^2}$	$\frac{\lvert z\rvert}{\|x\|^2}$
$x = 0$	0	∞	∞

FIGURE 3. The order of the contributions to $\hat\nu$ and $\hat\beta$ coming respectively from Γ_0 and Γ_∞ (it is assumed $p = (x, z) \neq (0, 0)$). The "topology" counting function $\hat\beta$ is more stable: it behaves as a rational function, whereas $\hat\nu$ has a "delta function" when some x_j is zero. Notice that isolated geodesics are always finite.

circles. If some of the singular values have multiplicities greater than one, but still are all pairwise non-commensurable, $\#L(\lambda) = 1$ but we can have critical manifolds of various dimensions.

As we will see, $\Gamma_\infty(p)$ is not empty only if some of the coordinates x_j vanish: the latter is a necessary condition for the occurrence of families. If $\Gamma_\infty(p)$ is not empty, each critical manifold is homeomorphic to a sphere; here the estimate (3) can be extended to *all* points $p \neq p_0$ if one adopts a "topological" viewpoint. Denoting by:

$$\hat\beta(p) = \{\text{sum of the Betti numbers of the set of geodesics from the origin to } p\},$$

we have the following generalization of Theorem 2 which bounds the number of spheres in $\Gamma_\infty(p)$ (see Theorems 24-30).

Theorem 5 (The "infinitesimal" bound for the topology)**.** *Let G be a contact Carnot group. There exist constants $C_1', C_2' > 0$ and R_1', R_2' such that for every $p = (x, z) \in G$, with $p \neq (0, 0)$:*

$$C_1' \frac{|z|}{\|x\|^2} + R_1' \leq \hat\beta(p) \leq C_2' \frac{|z|}{\|x\|^2} + R_2'.$$

As above, C_1', C_2' (resp. R_1', R_2') are homogeneous of degree -1 (resp. 0) in the singular values $\alpha_1 < \cdots < \alpha_k$ of A and are given by:

$$C_1' = \frac{8}{\pi}\frac{\alpha_1}{\alpha_k^2} \sin\left(\frac{\delta'\pi}{2}\right)^2 \qquad \text{with} \qquad \delta' = \left(\sum_{x_j \neq 0} \frac{\alpha_1}{\alpha_j}\left\lfloor \frac{\alpha_j}{\alpha_1}\right\rfloor\right)^{-1} \qquad \text{and} \qquad C_2' = \frac{8k}{\pi}\frac{\alpha_k}{\alpha_1^2};$$

and in particular again these upper bounds are invariant w.r.t. change of exponential coordinates.

Figure 3 compares the contribution to $\hat\nu$ and $\hat\beta$ coming respectively from Γ_0 and Γ_∞. In some sense, $\hat\beta(p)$ counts the geodesics "up to families". Thus if $x \neq 0$ (we say "p is not vertical") then geodesics might appear in families, but still the topology of these families is controlled, in particular the number of disjoint families is bounded.

Remark 3. On a contact Carnot group there is a well defined family of "non-homogeneous dilations" $\delta_\varepsilon(x, z) = (\varepsilon x, \varepsilon^2 z)$, where $\varepsilon > 0$ (see [2, 6]). These dilations have the property that if γ is a geodesic between the origin and p, then $\delta_\varepsilon\gamma$ is a geodesic between the origin and $\delta_\varepsilon(p)$ (the energies are though different, see Proposition 45 below). In particular both the counting function and the topology function are constants along the trajectories of δ_ε:

$$\hat\nu(\delta_\varepsilon(p)) = \hat\nu(p) \qquad \text{and} \qquad \hat\beta(\delta_\varepsilon(p)) = \hat\beta(p) \qquad \text{for all } \varepsilon > 0.$$

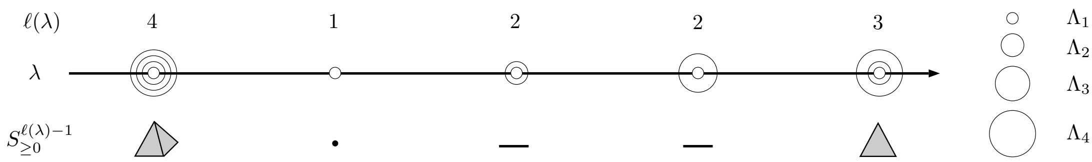

FIGURE 4. Equivalence classes of isometrically non-equivalent families for $k = 3$ commensurable eigenvalues $\alpha_i = 1/i$ for $i = 1, 2, 3, 4$. Thus $\Lambda_i = i\mathbb{Z}$.

1.3. **Families of geodesics.** A simple way to produce families of geodesics (critical manifolds) is to act on a geodesic γ with sub-Riemannian isometries fixing the endpoints of γ.

Example 4 (Heisenberg, continuation). Let us consider the Heisenberg group $\mathbb{H}_{2n+1}$. Thus $k = 1$ and $\alpha = 1$ (A is the canonical symplectic matrix). Let $p = (0, z)$ be a vertical point and γ a geodesic from the origin to p. The group of isometries fixing the origin is isomorphic to:

$$\mathrm{ISO}(\mathbb{H}_{2n+1}) \simeq U(n) \rtimes \mathbb{Z}_2.$$

Each isometry g in the connected component $\mathrm{U}(n)$ of the identity fixes the point $p = (0, z)$, thus $g\gamma$ is still a geodesic from the origin to p; such an isometry stabilizes the whole γ if it fixes the initial covector. Then, the stabiliser subgroup of the geodesic γ is $\mathrm{ISO}_\gamma(\mathbb{H}_{2n+1}) \simeq \mathrm{U}(n-1)$. In this way we produce a family:

$$X_\gamma = \mathrm{U}(n)/\mathrm{U}(n-1) \simeq S^{2n-1},$$

consisting of distinct geodesics isometrically equivalent to γ. In other words all geodesics in X_γ are obtained from γ by composition with an isometry (and they all have the same energy). In this case, it turns out that X_γ is a connected component of $\Gamma_\infty(p)$, i.e. a critical manifold.

Surprisingly this is not the case for more general Carnot groups. In fact, given a critical manifold $X \subset \Gamma_\infty(p)$ (one of the above spheres), this might not be obtained by acting with the stabilizer of p on a fixed geodesic. In other words, geodesics in X, although all with the same energy and endpoints, might be isometrically non-equivalent - they are "deformations" one of the other, but not via isometries.

We say that two geodesics with the same endpoints are *isometrically equivalent* if they are obtained one from the other by composition of an isometry of G. We denote by $\bar{\Gamma}_\infty(p)$ the set of equivalence classes of isometrically equivalent geodesics ending at p. For example, a family of isometrically equivalent geodesics corresponds to just a point in the quotient $\bar{\Gamma}_\infty(p)$.

The topology of this set (a quotient of $\Gamma_\infty(p)$) is related with the commensurability of the singular values of A (see Theorem 40).

Theorem 6 (Isometrically equivalent geodesics)**.** *Let G be a contact Carnot group. The set $\bar{\Gamma}_\infty(p)$ of equivalence classes of isometrically equivalent geodesics ending at $p \neq p_0$ is homeomorphic to the disjoint union of closed spheres quadrants:*

$$\bar{\Gamma}_\infty(p) \simeq \bigcup_{\lambda \in \Lambda_p} S^{\ell(\lambda)-1}_{\geq 0}, \qquad \ell(\lambda) := \#L(\lambda),$$

where $S^m_{\geq 0} = S^m \cap \mathbb{R}^{m+1}_{\geq 0}$ is the intersection of the m-sphere with the positive quadrant in $\mathbb{R}^{m+1}$.

See Figure 4. When A is generic, for every $\lambda \in \Lambda_p \subseteq \Lambda$ we have $\ell(\lambda) = 1$ and $\bar{\Gamma}_\infty$ is a discrete set of points, one for each $\lambda \in \Lambda_p$ (all the geodesics in a critical manifold $X \simeq S^1$ are isometrically equivalent to a given one). Nevertheless, non trivial manifolds of isometrically non-equivalent geodesics appear when there are *resonances*, i.e. some of the singular values of A are commensurable.

1.4. **A limiting procedure.** We discuss here the main ingredient of our study for contact sub-Riemannian manifolds: the *nilpotent approximation* of the structure at a point p_0. Because of the local nature of the problem, we can assume that $M = \mathbb{R}^{2n+1}$ and the point p_0 is the origin. Moreover, the distribution $\mathcal{D} \subset T\mathbb{R}^{2n+1}$ is given by:

$$\mathcal{D} = \text{span}\{f_1, \dots, f_{2n}\},$$

where $f_1, \dots, f_{2n}$ are bounded vector fields on $\mathbb{R}^{2n+1}$. The sub-Riemannian structure on $\mathcal{D}$ is obtained by declaring these vector fields to be orthonormal at each point.

We assume that the coordinates $(x, z) \in \mathbb{R}^{2n} \times \mathbb{R}$ are *adapted* to the distribution at the origin namely, $\mathcal{D}_{p_0} = \text{span}\{\partial_{x_1}, \dots, \partial_{x_{2n}}\}$ (for example we take canonical Darboux's coordinates). In the language of sub-Riemannian geometry these coordinates, at least in the contact (or step 2) case, are also called *privileged*.

Using these coordinates we define a family of "dilations" $\delta_\varepsilon : M \to M$ by:

$$\delta_\varepsilon(x, z) := (\varepsilon x, \varepsilon^2 z), \qquad \varepsilon > 0,$$

and the *nilpotent approximation at* p_0, another sub-Riemannian structure on the same base manifold M, given by declaring the following fields:

$$\hat{f}_i := \lim_{\varepsilon \to 0} \varepsilon \delta_{\frac{1}{\varepsilon}*} f_i, \qquad \forall i = 1, \dots, 2n,$$

a new orthonormal frame. Thus, the nilpotent approximation at a point p_0 is the "principal part" of the original sub-Riemannian structure in a neighbourhood of p_0 w.r.t. the non-homogeneous dilations δ_ε. Moreover, it turns out that the nilpotent approximation at any point p_0 of a contact sub-Riemannian manifold is a contact Carnot group.

The idea of the study of the asymptotic behaviour of geodesics in contact manifolds originates in [1]. In that case the author adopts the point of view that geodesics are rays through the exponential map and produces a kind of "blowup" of the initial covectors; as a result conjugate times for geodesics, at least asymptotically, are equal to the conjugate times for the nilpotent approximation [1, Theorem 3.1].

We introduce the following notation:

$$\nu(p) = \#\{\text{geodesics joining } p_0 \text{ and } p\}.$$

Thus $\nu(p)$ will denote the number of *affine* geodesics between p_0 and p, i.e. geodesics in M that are contained in a coordinate chart of p_0. Similarly $\hat{\nu}$ denoted the number of geodesics between the origin and p for the nilpotent approximation. The next theorem relates the geodesic count on the original structure and on the nilpotent Carnot group structure (see Theorem 47).

Theorem 7 (Counting in the limit)**.** *Let M be a contact sub-Riemannian manifold. For the generic $p \in M$ sufficiently close to p_0:*

$$\hat{\nu}(p) \leq \lim_{\varepsilon \to 0} \nu(\delta_\varepsilon(p)).$$

where δ_ε is the non-homogeneous dilation defined in some set of adapted coordinates in a neighbourhood of p_0.

Combining Theorem 7 and Theorem 2 we obtain an estimate for the order of growth of the number of "affine" geodesics between two close points on a contact manifold (see Theorem 48).

Theorem 8 (The local bound)**.** *Let M be a contact manifold. Denote by (x, z) Darboux's coordinates on a neighbourhood U of q. For every point $q \in M$ there exist constants $C(q), R(q)$ such that, for the generic $p = (x, z) \in U$:*

$$\lim_{\varepsilon \to 0} \nu(\delta_\varepsilon(p)) \geq C(q) \frac{|z|}{\|x\|^2} + R(q).$$

As we anticipated, a completely new phenomenon in the sub-Riemannian case is the existence of a sequence of points $q_n \to q$ with arbitrary large number of affine geodesics between the two (see Theorem 49).

Theorem 9 (Abundance of "affine" geodesics)**.** *Let M be a contact sub-Riemannian manifold and $q \in M$. Then there exists a sequence $\{q_n\}_{n\in\mathbb{N}}$ in M such that:*

$$\lim_{n\to\infty} q_n = q \qquad \textit{and} \qquad \lim_{n\to\infty} \nu(q_n) = \infty.$$

Notice that, in general, we cannot predict the existence of a point p with infinitely many affine geodesics between q and p.

1.5. **Structure of the paper.** In Section 2 we discuss some preliminary material, such as the definition of the main objects. Section 3 is devoted to the study of the exponential map for contact Carnot groups and terminates with the proof of the Theorem on the topology of the critical manifolds (Theorem 23). Theorem 2 above is split into two parts: the upper bound is discussed in Section 4 (Theorem 24) and the lower bound in Section 5 (Theorem 26 and Theorem 30); Proposition 25 is proved in section 4 (Proposition 25). In Sections 6 7 we study the isometry group of contact Carnot Groups. The theorem on equivalence classes of isometric geodesics is proved in Section 8 (Theorem 40). Section 9 is devoted to general contact manifolds: we first review their nilpotent approximation at a point and we prove some related properties; next we prove Theorem 7 (Theorem 47), Theorem 8 (Theorem 48) and Theorem 9 (Theorem 49).

Acknowledgements. The authors wish to thank A. A. Agrachev, D. Barilari and A. Gentile for stimulating comments. The first author was supported by the European Community's Seventh Framework Programme ([FP7/2007-2013] [FP7/2007-2011]) under grant agreement No. [258204].

2. Preliminaries

We recall some basic facts in sub-Riemannian geometry. We refer to [2, 18] for further details. Let M be a smooth, connected manifold of dimension $n \geq 3$. A sub-Riemannian structure on M is a pair $(\mathcal{D}, \langle\cdot|\cdot\rangle)$ where $\mathcal{D}$ is a smooth vector distribution of constant rank $k \leq n$ satisfying the *Hörmander condition* (i.e. $\mathrm{Lie}_x\mathcal{D} = T_xM$, $\forall x \in M$) and $\langle\cdot|\cdot\rangle$ is a smooth Riemannian metric on $\mathcal{D}$. A Lipschitz continuous curve $\gamma : [0,1] \to M$ is *admissible* (or *horizontal*) if $\dot\gamma(t) \in \mathcal{D}_{\gamma(t)}$ for a.e. $t \in [0,1]$. Given a horizontal curve $\gamma : [0,1] \to M$, the *energy* of γ is

$$J(\gamma) = \int_I \|\dot\gamma(t)\|^2 dt,$$

where $\|\cdot\|$ denotes the norm induced by $\langle\cdot|\cdot\rangle$.

The pair $(\mathcal{D}, \langle\cdot|\cdot\rangle)$ can be given, at least locally, by assigning a set of k smooth vector fields that span $\mathcal{D}$, orthonormal for $\langle\cdot|\cdot\rangle$. In this case, the set $\{f_1, \dots, f_k\}$ is called a *local orthonormal frame* for the sub-Riemannian structure.

In this paper we are mainly concerned with *contact* sub-Riemannian manifolds; in this framework admissible curves are also called *legendrian*.

Definition 10. A sub-Riemannian manifold is *contact* if locally there exists a one form α such that $\mathcal{D} = \ker\alpha$, and $d\alpha|_{\mathcal{D}}$ is non degenerate. Notice that, due to the non-degeneracy assumption, the rank of $\mathcal{D}$ must be even.

Definition 11. Let M be a contact manifold. A sub-Riemannian *geodesic* is a non-constant admissible curve $\gamma : [0,1] \to M$ that is locally energy minimizer. More precisely, for any $t \in [0,1]$ there exists a sufficiently small interval $I \subseteq [0,1]$, containing t, such that the restriction $\gamma|_I$ minimizes the energy between its endpoints.

Pontryagin maximum principle implies that any geodesic starting at p_0 can be lifted to a Lipschitz curve $\eta : [0,1] \to T^*M$ called *sub-Riemannian extremal*, as we discuss now. In general, sub-Riemannian extremals can be *normal* or *abnormal*. Abnormal extremals appear in many structures, and their regularity is one of the main open problems in sub-Riemannian geometry (see [13, 15] and references therein for recent progresses on this topic). Abnormal extremals do not appear in contact (or Riemannian) structures; for this reason, we now give an explicit characterization of *normal extremals*. We start with an important definition.

Definition 12. The *Hamiltonian function* $H \in C^\infty(T^*M)$ is

$$H(\eta) := \frac{1}{2}\sum_{i=1}^{k}\langle \eta, f_i\rangle^2, \qquad \forall \eta \in T^*M,$$

where $f_1, \dots, f_k$ is any local frame and $\langle \eta, \cdot\rangle$ denotes the action of the covector η on vectors.

Let σ be the canonical symplectic form on T^*M. With the symbol $\vec{a}$ we denote the Hamiltonian vector field on T^*M associated with a function $a \in C^\infty(T^*M)$. Indeed $\vec{a}$ is defined by the formula $da = \sigma(\cdot, \vec{a})$. Consider the *Hamiltonian vector field* $\vec{H} \in \mathrm{Vec}(T^*M)$.

Definition 13. Non-constant trajectories of the Hamiltonian system $\dot{\eta} = \vec{H}(\eta)$ are (normal) *sub-Riemannian extremals*.

Since normal sub-Riemannian extremals are non-constant trajectories of the Hamiltonian field $\vec{H}$, we obtain the following well known result.

Theorem 14. *Normal sub-Riemannian* geodesics *are exactly projections on M of normal sub-Riemannian* extremals*. In particular, all normal geodesics are smooth.*

We stress that in any structure where abnormal extremals do not exist (such as contact or Riemannian structures), Theorem 14 completely characterizes *all* geodesics. Moreover, any (normal) sub-Riemannian geodesic can be specified by its *initial covector*. In this setting, we introduce the sub-Riemannian generalization of the classical Riemannian exponential map. Let $\eta(t) = e^{t\vec{H}}(\eta_0)$ denote the integral line of $\vec{H}$, with initial condition $\eta(0) = \eta_0$.

Definition 15. The *sub-Riemannian exponential map* (with origin p_0) $E : T^*_{p_0}M \to M$ is

$$E(\eta_0) := \pi(e^{\vec{H}}(\eta_0)), \qquad \forall \eta_0 \in T^*_{p_0}M.$$

Thus all (normal) geodesics starting from p_0 are the image through E of the ray $t \mapsto t\eta$. We denote by $\Gamma(p)$ the set of geodesics from p_0 to $p \neq p_0$, with the topology induced by its inclusion:

$$\Gamma(p) = E^{-1}(p) \subset T^*_{p_0}M.$$

2.1. **Fibers of the exponential map and geodesics.** Notice that the correspondence:

$$\eta \mapsto \gamma_\eta, \qquad \gamma_\eta(t) = \pi(e^{t\vec{H}}(\eta))$$

defines a continuous map from $T^*_{p_0}M$ to the set of admissible curves. If we endow this set with the $W^{1,\infty}$-topology and we assume $p \neq p_0$, this map restricts to a homeomorphism between $\Gamma(p)$ and the set of geodesics to p: the topologies on $\Gamma(p)$ as a subset of $T^*_{p_0}M$ or as a subset of the space of admissible curves coincide and the point of view we adopted is not restrictive.

On the other hand, recall that extremals (resp. geodesics) are non-constant and for these reasons we will always make the assumption:

$$p \neq p_0.$$

In fact, most of our results are still true also for $p = p_0$, but then one should regard $\Gamma(p)$ simply as the fiber of E and not as the set of geodesics to p.

2.2. **Contact Carnot groups.** We start with the preliminary study of *contact Carnot groups*. As we will see, their analysis provides the main building block for the general theory on contact manifolds. Since we are mainly interested in contact structures, we restrict ourselves to step 2 Carnot groups. We refer to [11, 12, 16] for more details.

A Carnot group G of step 2 is a simply connected Lie group whose Lie algebra of left-invariant vector fields $\mathfrak{g}$ admits a nilpotent stratification of step 2, namely

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2, \qquad \mathfrak{g}_1, \mathfrak{g}_2 \neq \{0\},$$

with

$$[\mathfrak{g}_1, \mathfrak{g}_1] = \mathfrak{g}_2, \qquad \text{and} \qquad [\mathfrak{g}_1, \mathfrak{g}_2] = [\mathfrak{g}_2, \mathfrak{g}_2] = \{0\}.$$

A left-invariant sub-Riemannian structure on G is obtained by defining a scalar product on $\mathfrak{g}_1$ or, equivalently, by declaring a set $f_1, \dots, f_k \in \mathfrak{g}_1$ a global orthonormal frame. In particular, $\mathcal{D}|_x = \mathfrak{g}_1|_x$, for all $x \in G$. The group exponential map,

$$\exp_G : \mathfrak{g} \to G,$$

associates with $v \in \mathfrak{g}$ the element $\gamma(1)$, where $\gamma : [0,1] \to G$ is the unique integral line of the vector field v such that $\gamma(0) = 0$. Since G is simply connected and $\mathfrak{g}$ is nilpotent, $\exp_G$ is a smooth diffeomorphism. Thus we can identify $G \simeq \mathbb{R}^m$ with a polynomial product law.

Remark 4. In the literature, these structures are also referred to as *Carnot groups* of type (k, m), where $k = \dim \mathcal{D} = \operatorname{rank} \mathfrak{g}_1$ is the *rank* of the distribution and m is the dimension of G.

Definition 16. A contact Carnot group is a Carnot group of type $(2n, 2n+1)$ that admits a contact structure.

Notice that the only non-trivial request is the non-degeneracy of the contact form. In fact, let G be a contact Carnot group. Let $f_1, \dots, f_{2n} \in \mathfrak{g}_1$ be a global orthonormal frame of left-invariant vector fields, and $f_0 \in \mathfrak{g}_2$ a generator for the second layer. Indeed:

$$[f_i, f_j] = A_{ij} f_0, \qquad \forall i, j = 1, \dots, 2n,$$

for some constant matrix $A \in \mathfrak{so}(2n)$. Observe that there exists a unique never-vanishing left invariant one-form α (up to multiplication by a constant) such that $\mathcal{D} = \ker \alpha$. Using the identity $d\alpha(X, Y) = X\alpha(Y) - Y\alpha(X) - \alpha([X, Y])$ we obtain:

$$d\alpha(f_i, f_j) = -\alpha([f_i, f_j]) = A_{ji}\alpha(f_0). \tag{5}$$

Since $\alpha(f_0) \neq 0$, a Carnot group G of type $(2n, 2n+1)$ is a contact Carnot group if and only if the matrix A is non-degenerate.

2.3. **Normal form of contact Carnot groups.** By acting on $\mathfrak{g}_1$ with an orthogonal transformation, it is always possible to put A in its canonical form. Such a transformation can be trivially extended to an automorphism of $\mathfrak{g}$, and thus lifts to a group automorphism of G that preserves the scalar product. Therefore, isometry classes of contact Carnot groups are parametrised by the possible singular values of non-degenerate matrices $A \in \mathfrak{so}(2n)$. In the following we describe the possible normal forms of contact Carnot groups. Consider the triple $(k, \vec{n}, \vec{\alpha})$, where:

(i) $k \in \mathbb{N}$, with $1 \leq k \leq n$,
(ii) $\vec{n} = (n_1, \dots, n_k)$ is a partition of n, namely $n_j \in \mathbb{N}$ and $\sum_{j=1}^k n_j = n$,
(iii) $\vec{\alpha} = (\alpha_1, \dots, \alpha_k)$ with $0 < \alpha_1 < \dots < \alpha_k$.

For a fixed choice of $(k, \vec{n}, \vec{\alpha})$, let:

$$A := \operatorname{diag}(\alpha_1 J_{n_1}, \dots, \alpha_k J_{n_k}) \in \mathfrak{so}(2n), \qquad \text{with} \qquad J_m = \begin{pmatrix} 0 & \mathbb{1}_m \\ -\mathbb{1}_m & 0 \end{pmatrix}. \tag{6}$$

In other words, A has k distinct singular values $0 < \alpha_1 < \ldots < \alpha_k$, with multiplicities $n_1, \ldots, n_k$ (half the dimension of the corresponding invariant subspaces). This gives the normal form of the $(2n, 2n+1)$ graded Lie algebra with parameters $(k, \vec{n}, \vec{\alpha})$. As an abstract algebra is given by:

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2, \qquad \mathfrak{g}_1 = \operatorname{span}\{f_1, \ldots, f_{2n}\}, \quad \mathfrak{g}_2 = \operatorname{span}\{f_0\},$$

with:

$$[f_i, f_j] = A_{ij} f_0, \qquad i, j = 1, \ldots, 2n.$$

Let G be the unique connected, simply connected Lie group such that $\mathfrak{g}$ is its Lie algebra. We define a scalar product on $\mathfrak{g}_1$ in such a way that $f_1, \ldots, f_{2n}$ is an orthonormal frame. This gives G the structure of a contact Carnot group associated with the parameters $(k, \vec{\alpha}, \vec{n})$. Any contact Carnot group is isomorphic to one of these structures, for a choice of $(k, \vec{\alpha}, \vec{n})$. Notice that the normal form is determined only up to rescaling of the eigenvalues $\vec{\alpha}$ (see [5, Remark 1] for more details).

2.4. **Exponential coordinates.** The orthonormal basis $f_1, \ldots, f_{2n}$ and f_0 realize the splitting

$$\mathfrak{g} = \mathfrak{g}_1^{\alpha_1} \oplus \ldots \oplus \mathfrak{g}_1^{\alpha_k} \oplus \mathfrak{g}_2,$$

according to the generalized eigenspaces of A. Accordingly, we identify:

$$G \simeq \mathbb{R}^{2n_1} \oplus \ldots \oplus \mathbb{R}^{2n_k} \oplus \mathbb{R}$$

through the group exponential map $\exp_G : \mathfrak{g} \to G$, in such a way that $p \in G$ has exponential coordinates $(x_1, \ldots, x_k, z)$ with $x_i \in \mathbb{R}^{2n_i}$ for $i = 1, \ldots, k$ and $z \in \mathbb{R}$.

2.5. **An explicit representation.** An explicit representation of the contact Carnot group with parameters $(k, \vec{\alpha}, \vec{n})$ is given by the sub-Riemannian structure induced by the following vector fields on $\mathbb{R}^{2n+1}$, with coordinates $(x, z) \in \mathbb{R}^{2n} \times \mathbb{R}$:

$$f_i := \frac{\partial}{\partial x_i} - \frac{1}{2} \sum_{j=1}^{2n} A_{ij} x_j \frac{\partial}{\partial z}, \qquad f_0 := \frac{\partial}{\partial z}, \qquad i = 1, \ldots, 2n.$$

where A is the matrix of Eq. (6) with k singular values $\vec{\alpha}$ and multiplicities $\vec{n}$.

The next lemma clarifies the relation between "tautological" coordinates (x, z) and exponential coordinates. In fact, there is no difference.

Lemma 17. *The coordinates (x, z) are the exponential coordinates induced by $f_1, \ldots, f_{2n}, f_0$.*

Proof. Assume that $p = (x, z)$ has exponential coordinates (θ, ρ), namely

$$(x, z) = \exp_G \left(\sum_{i=1}^{2n} \theta_i f_i + \rho f_0 \right).$$

This means that $(x, z) = \gamma(1)$, where $\gamma(t) = (x(t), z(t))$ is the solution of the Cauchy problem

$$\dot{x}_i(t) = \theta_i, \qquad \dot{z}(t) = \rho + \frac{1}{2} \sum_{i,j=1}^{2n} x_i A_{ij} \theta_j, \qquad \gamma(0) = (0, 0),$$

By the skew-symmetry of A, the solution is $x(t) = \theta t$ and $z(t) = \rho t$. Then $(x, z) = (\theta, \rho)$. □

By the Campbell-Baker-Hausdorff formula, the product on G, in exponential coordinates, is:

$$(x, z) \cdot (x', z') = \left(x + x', z + z' + \frac{1}{2} x^* A x \right).$$

Example 5. A classical example is the $(2n+1)$-dimensional contact Carnot group $\mathbb{H}_{2n+1}$. This is the case with $k=1$, i.e. a unique singular value $\alpha_1 = 1$ with multiplicity n. In this case

$$f_i := \frac{\partial}{\partial x_i} - \frac{1}{2}x_{i+n}\frac{\partial}{\partial z}, \qquad f_{n+i} := \frac{\partial}{\partial x_{n+i}} + \frac{1}{2}x_i\frac{\partial}{\partial z}, \qquad f_0 := \frac{\partial}{\partial z}, \qquad i = 1,\dots,n,$$

and A is the standard symplectic matrix $J = \left(\begin{smallmatrix} 0 & \mathbb{1}_n \\ -\mathbb{1}_n & 0 \end{smallmatrix}\right)$.

Remark 5. We introduced contact Carnot groups as special contact sub-Riemannian structures. We stress that the following construction can be generalized, with minor modifications, to any corank-one Carnot group, namely structures of type $(k, k+1)$ where A may also have odd dimension and some vanishing singular values.

3. The fibers of the exponential map for contact Carnot groups

Let $\hat{E} : T_0^*G \to G$ be the exponential map for the contact Carnot group whose (nonzero) structure constants for its Lie algebra are given by:

$$A = \operatorname{diag}(\alpha_1 J_{n_1}, \dots, \alpha_k J_{n_k}), \qquad \text{where} \qquad J_m = \begin{pmatrix} 0 & \mathbb{1}_m \\ -\mathbb{1}_m & 0 \end{pmatrix}.$$

In the following, we write $p \in G$ in exponential coordinates as $p = (x_1, \dots, x_k, z)$, with $x_j \in \mathbb{R}^{2n_j}$ and analogously, for $\eta \in T_0^*G$, we write $\eta = (u_1, \dots, u_k, \lambda)$, with $u_j \in \mathbb{R}^{2n_j}$. Thus:

$$\hat{E}(u_1, \dots, u_k, \lambda) = (x_1, \dots, x_k, z) \qquad \text{with} \qquad x_j, u_j \in \mathbb{R}^{2n_j}, \quad j = 1, \dots, k.$$

When convenient, we adopt the compact notation $p = (x, z)$ and $\eta = (u, \lambda)$, with $x, u \in \mathbb{R}^{2n}$ and $n = \sum_{j=1}^k n_j$.

Proposition 18. *With the above notation we have for every $j = 1, \dots, k$:*

$$x_j = \left(\frac{\sin(\lambda\alpha_j)}{\lambda\alpha_j}\mathbb{1} + \frac{\cos(\lambda\alpha_j) - 1}{\lambda\alpha_j}J\right)u_j \qquad \text{and} \qquad z = \sum_{j=1}^k \left(\frac{\lambda\alpha_j - \sin(\lambda\alpha_j)}{2\lambda^2\alpha_j}\right)\|u_j\|^2, \tag{7}$$

where we understand that, if $\lambda = 0$, then $x_j = u_j$ for $j = 1, \dots, k$ and $z = 0$, i.e. $\hat{E}(u, 0) = (u, 0)$.

Proof. We recall that the sub-Riemannian exponential map is given explicitly by [3]:

$$(u, \lambda) \mapsto \left(\int_0^1 e^{-\lambda A t} u dt, -\frac{1}{2}\int_0^1 \left\langle e^{-\lambda A t}u, A\int_0^t e^{-\lambda A s} u ds \right\rangle dt\right).$$

We start by considering the horizontal components (we omit the subscript for $J = J_{n_j}$):

$$x_j = \int_0^1 e^{-\lambda\alpha_j J t} u_j dt.$$

If $\lambda = 0$, then $e^{-\lambda\alpha_j J t} = \mathbb{1}$ and $x_j = u_j$; otherwise the expression for x_j follows immediately from writing the integrand matrix as:

$$e^{-\lambda\alpha_j J t} = \cos(\lambda\alpha_j t)\mathbb{1} - \sin(\lambda\alpha_j t) J. \tag{8}$$

In fact using (8) we can also evaluate the matrix integral:

$$\int_0^t e^{-\lambda\alpha_j J t} dt = \frac{\sin(\lambda\alpha_j t)}{\lambda\alpha_j}\mathbb{1} + \frac{\cos(\lambda\alpha_j t) - 1}{\lambda\alpha_j}J = a(t)\mathbb{1} + b(t)J. \tag{9}$$

For the z component, we notice that it can be rewritten as $z = u^* S u$, where S is the matrix:

$$S = -\frac{1}{2}\int_0^1 \int_0^t e^{\lambda A t} A e^{-\lambda A s} ds dt,$$

and since A is assumed to be block-diagonal, we obtain:

$$z=\sum_{j=1}^{k} u_j^* S_j u_j \qquad \text{with} \qquad S_j=-\frac{1}{2}\int_0^1 e^{\lambda\alpha_j J t}\alpha_j J\int_0^t e^{-\lambda\alpha_j J s}ds dt,$$

where once again we removed the subscript from the matrix $J=J_{n_j}$ in the expression for S_j. Notice that if $\lambda=0$ then $S=-\frac{1}{4}A$ and, being skew-symmetric, $z=u^*Su=0$. If $\lambda\neq 0$ the integrand matrix in S_j equals, using (8):

$$\begin{aligned}(10)\quad e^{\lambda\alpha_j J t}\alpha_j J\int_0^t e^{-\lambda\alpha_j J s}ds &= \left(\alpha_j\cos(\lambda\alpha_j t)J-\alpha_j\sin(\lambda\alpha_j t)\mathbb{1}\right)\left(a(t)\mathbb{1}+b(t)J\right)=\\ &=\left(c(t)\mathbb{1}+d(t)J\right)\left(a(t)\mathbb{1}+b(t)J\right)=(ac-bd)(t)\mathbb{1}+(ad+bc)(t)J.\end{aligned}$$

where we defined $c(t)=\alpha_j\cos(\lambda\alpha_j t)$ and $d(t)=-\alpha_j\sin(\lambda\alpha_j t)$. Since $\int(ad+bc)J$ will be a skew-symmetric matrix, then:

$$u_j^*S_ju_j=u_j^*\mathbb{1}\left(-\frac{1}{2}\int_0^1(ac-bd)(t)dt\right)u=-\|u_j\|^2\frac{1}{2}\int_0^1(ac-bd)(t)dt.$$

Using the explicit expression of a,b,c,d (given by (9) and (10)), we obtain $(ac-bd)(t)=\frac{\cos(\lambda\alpha_j t)-1}{\lambda}$, whose integral equals:

$$\int_0^1\frac{\cos(\lambda\alpha_j t)-1}{\lambda}dt=\frac{\sin(\lambda\alpha_j)-\lambda\alpha_j}{\lambda^2\alpha_j}.$$

Substituting this into the above formula for $u_j^*Su_j$ concludes the proof. □

For all $j=1,\dots,k$, we define the $2n_j\times 2n_j$ matrix:

$$I(\lambda\alpha_j)=\frac{\sin(\lambda\alpha_j)}{\lambda\alpha_j}\mathbb{1}+\frac{\cos(\lambda\alpha_j)-1}{\lambda\alpha_j}J,$$

where we understand that $I(0)=\mathbb{1}$. I this way, equation (7) reads $x_j=I(\alpha_j\lambda)u_j$.

Proposition 19. *Assume $\lambda\alpha_j\notin 2\pi\mathbb{Z}\setminus\{0\}$. Then $I(\lambda\alpha_j)$ is invertible with inverse:*

$$I(\lambda\alpha_j)^{-1}=\frac{\lambda\alpha_j}{2}\cot\left(\frac{\lambda\alpha_j}{2}\right)\mathbb{1}+\frac{\lambda\alpha_j}{2}J,$$

where if $\lambda\alpha_j=0$ we have $I(0)^{-1}=\mathbb{1}$. In particular if $x_j=I(\lambda\alpha_j)u_j$, then:

$$\frac{\lambda\alpha_j-\sin(\lambda\alpha_j)}{2\lambda^2\alpha_j}\|u_j\|^2=\frac{\alpha_j}{8}\frac{\lambda\alpha_j-\sin(\lambda\alpha_j)}{\sin\left(\frac{\lambda\alpha_j}{2}\right)}\|x_j\|^2.$$

Proof. The determinant if $I(\lambda\alpha_j)$ is:

$$\det I(\lambda\alpha_j)=2\left(\frac{1-\cos(\lambda\alpha_j)}{\lambda^2\alpha_j^2}\right),$$

and is nonzero if and only if $\frac{\lambda\alpha_j}{2\pi}\notin\mathbb{Z}\setminus 0$; in this case the matrix $I(\lambda\alpha_j)^{-1}$ as above is well defined and is the inverse of $I(\lambda\alpha_j)$.

For the second part of the statement we write $I(\lambda\alpha_j)^{-1}=c_1\mathbb{1}+c_2J$, where $c_1=\frac{\lambda\alpha_j}{2}\cot\left(\frac{\lambda\alpha_j}{2}\right)$ and $c_2=\frac{\lambda\alpha_j}{2}$. Then, $u_j=c_1x_j+c_2Jx_j$ and since x_j and Jx_j are orthogonal we obtain:

$$\|u_j\|^2=c_1^2\|x_j\|^2+c_2^2\|Jx_j\|^2=(c_1^2+c_2^2)\|x_j\|^2.$$

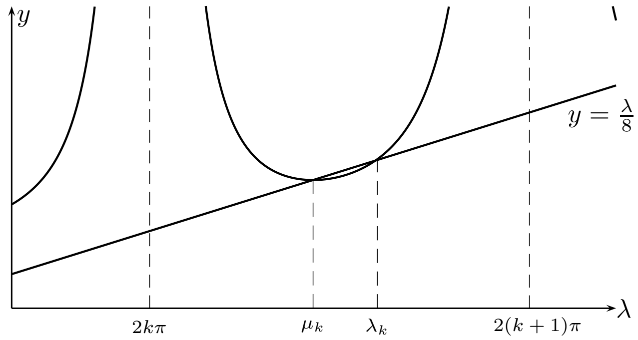

FIGURE 5. Detail of the function $g(\lambda)$ in the interval $I_k = (2k\pi, 2k\pi + 2\pi)$.

Computing $c_1^2 + c_2^2 = (\frac{\lambda\alpha_j}{2}\frac{1}{\sin(\lambda\alpha_j/2)})^2$, and setting $y = \lambda\alpha_j$ we finally obtain:

$$\frac{y - \sin y}{2y^2/\alpha_j}\|u_j\|^2 = \frac{y - \sin y}{2y^2/\alpha_j}\left(\frac{y}{2}\frac{1}{\sin(y/2)}\right)^2\|x_j\|^2 = \frac{\alpha_j}{8}\frac{y - \sin y}{(\sin\frac{y}{2})^2}. \qquad \square$$

Lemma 20. *Assume $\lambda\alpha_j \in 2\pi\mathbb{Z}\setminus\{0\}$. Then $x_j = 0$.*

Proof. It follows immediately from (7). $\square$

3.1. **A relevant function.** We introduce the function $g : \mathbb{R} \to \mathbb{R}\cup\{\infty\}$ defined by:

$$g(\lambda) = \frac{1}{8}\frac{\lambda - \sin\lambda}{\left(\sin\frac{\lambda}{2}\right)^2}.$$

Each pole of g is of order two and lies on $2\pi\mathbb{Z}\setminus\{0\}$ (see Figure 2 in Section 1 and Figure 5 above).

Proposition 21. *Let $k \in \mathbb{Z}$ and $I_k = (2k\pi, 2k\pi + 2\pi)$. Then:*

1. *$g(\lambda) = -g(-\lambda)$ and $g(\lambda) > 0$ if $\lambda > 0$;*
2. *$|g|$ is strictly convex on each interval I_k;*
3. *if μ_k is the point of minimum of $|g|$ on I_k, we have $g(\mu_k) = \frac{\mu_k}{8} < \frac{(2k+1)\pi}{8}$;*
4. *$g(|\lambda|) > \frac{|\lambda|}{8} - \frac{\pi}{8}$ for every λ.*

Proof. The first property is clear, since g is the sum of two odd functions. To derive the other properties, we compute the first two derivatives of g:

$$g'(\lambda) = -\frac{1}{8}\frac{\lambda\cot\left(\frac{\lambda}{2}\right) - 2}{\left(\sin\frac{\lambda}{2}\right)^2} \quad \text{and} \quad g''(\lambda) = \frac{1}{16}\frac{2\lambda + \lambda\cos\lambda - 3\sin\lambda}{\left(\sin\frac{\lambda}{2}\right)^4}.$$

For $\lambda > 0$, a standard check shows that $g''(\lambda) > 0$, which proves g is strictly convex for $\lambda > 0$; being an odd function g is strictly concave for $\lambda < 0$ and the statement for $|g|$ follows.

To estimate the value of the minimum of $|g|$ on I_k, we evaluate g and its derivative at the middle point $\lambda_k = (2k+1)\pi$ for $k > 0$ (the situation being symmetric for $k < 0$):

$$g(\lambda_k) = \frac{(2k+1)\pi}{8} \qquad \text{and} \qquad g'(\lambda_k) = \frac{1}{4}.$$

Thus g is increasing at λ_k, its point of minimum μ_k must be in the interval $(2k\pi, \lambda_k)$ and its value must be smaller than the value $g(\lambda_k) = \frac{(2k+1)\pi}{8}$ (see Figure 5). Let's prove $g(\mu_k) = \frac{\mu_k}{8}$. If

μ_k is a minimum, then $g'(\mu_k) = 0$. Using the above expression for the derivative we must have:

$$\mu_k \cot\left(\frac{\mu_k}{2}\right) - 2 = 0 \qquad \text{i.e.} \qquad \sin\frac{\mu_k}{2} = \frac{\mu_k}{2}\cos\frac{\mu_k}{2}.$$

Substituting this into the definition of g we have:

$$\begin{aligned} g(\mu_k) &= \frac{1}{8}\frac{\mu_k - \sin\mu_k}{\left(\sin\frac{\mu_k}{2}\right)^2} = \frac{1}{8}\frac{\mu_k - 2\left(\sin\frac{\mu_k}{2}\right)\left(\cos\frac{\mu_k}{2}\right)}{\left(\sin\frac{\mu_k}{2}\right)^2} = \\ &= \frac{1}{8}\frac{\mu_k - \mu_k\left(\cos\frac{\mu_k}{2}\right)^2}{1-\left(\cos\frac{\mu_k}{2}\right)^2} = \frac{\mu_k}{8}. \end{aligned}$$

For what concerns the last property, it is sufficient to prove it for $\lambda > 0$. In the (λ, y)-plane, the equation for the tangent line to the graph of g at the point $\lambda_k > 0$ is:

$$y = \frac{\lambda}{4} - \frac{(2k+1)\pi}{8}.$$

Since the function is strictly convex we must have $g(\lambda) \geq \frac{\lambda}{4} - \frac{(2k+1)\pi}{8}$ for all $\lambda \in I_k$. On this interval $\lambda > 2k\pi$ and thus:

$$g(\lambda) \geq \frac{\lambda}{4} - \frac{(2k+1)\pi}{8} = \frac{\lambda}{8} + \frac{\lambda}{8} - \frac{2k\pi}{8} - \frac{\pi}{8} > \frac{\lambda}{8} - \frac{\pi}{8}.$$

□

3.2. **Decomposition of the fiber.** In this section we give a description of the fibers of the exponential map for the contact Carnot group G. We introduce the notation:

$$\Gamma(p) = \hat{E}^{-1}(x,z), \qquad p = (x,z).$$

Recall that we are always assuming $p \neq p_0$, so that $\Gamma(p)$ consists of all geodesics ending at p. Given $\alpha_1, \ldots, \alpha_k$ we define:

$$\Lambda_j = \{\text{poles of } \lambda \mapsto g(\lambda\alpha_j)\}, \qquad \Lambda = \bigcup_{j=1}^{k}\Lambda_j \qquad \text{and} \qquad I_0 = \{j \,|\, x_j = 0\}.$$

Notice that Lemma 20 implies that, if $(u,\lambda) \in \Gamma(p)$, then:

$$L(\lambda) := \{j \mid \lambda \in \Lambda_j\} \subseteq I_0. \tag{11}$$

Proposition 22 (Characterization of the fiber)**.** *Let $p = (x,z) \in G$, $p \neq (0,0)$. The set $\Gamma(p)$ consists of the points (u,λ) such that $x_j = I(\lambda\alpha_j)u_j$ for every $j = 1,\ldots,k$ and:*

$$z = \sum_{j\notin I_0}\alpha_j g(\lambda\alpha_j)\|x_j\|^2 + \frac{1}{2\lambda}\sum_{j\in I_0}\|u_j\|^2. \tag{12}$$

Proof. The condition on the x_j's is given by Proposition 18 and it remains to understand the equation for z in (7). Now we can decompose the summation in the terms defining z as:

$$z = \sum_{j\notin I_0}\left(\frac{\lambda\alpha_j - \sin(\lambda\alpha_j)}{2\lambda^2\alpha_j}\right)\|u_j\|^2 + \sum_{j\in I_0}\left(\frac{\lambda\alpha_j - \sin(\lambda\alpha_j)}{2\lambda^2\alpha_j}\right)\|u_j\|^2. \tag{13}$$

If $j \notin I_0$ then $j \notin L(\lambda)$ by (11) and Proposition 19 allows to write:

$$\left(\frac{\lambda\alpha_j - \sin(\lambda\alpha_j)}{2\lambda^2\alpha_j}\right)\|u_j\|^2 = \frac{\alpha_j}{8}\frac{\lambda\alpha_j - \sin(\lambda\alpha_j)}{\sin\left(\frac{\lambda\alpha_j}{2}\right)}\|x_j\|^2.$$

On the other hand the sum $\sum_{j\in I_0}\left(\frac{\lambda\alpha_j-\sin(\lambda\alpha_j)}{2\lambda^2\alpha_j}\right)\|u_j\|^2$ can be split as:

$$\sum_{j\in I_0\cap L(\lambda)}\left(\frac{\lambda\alpha_j-\sin(\lambda\alpha_j)}{2\lambda^2\alpha_j}\right)\|u_j\|^2+\sum_{j\in I_0\cap L(\lambda)^c}\left(\frac{\lambda\alpha_j-\sin(\lambda\alpha_j)}{2\lambda^2\alpha_j}\right)\|u_j\|^2.$$

The second summation is zero, because for a $j\notin L(\lambda)$ the matrix $I(\lambda\alpha_j)$ is invertible and $u_j=I(\lambda\alpha_j)x_j=0$. By (11), the index set for the first summation equals $L(\lambda)$ itself. Moreover, for each term $j\in L(\lambda)$ we have $\lambda\alpha_j\in 2\pi\mathbb{Z}\setminus\{0\}$ and, for some $k_j\in\mathbb{Z}\setminus\{0\}$:

$$\frac{\lambda\alpha_j-\sin(\lambda\alpha_j)}{2\lambda^2\alpha_j}=\frac{2\pi k_j-\sin(2\pi k_j)}{2\lambda(2\pi k_j)}=\frac{1}{2\lambda}.$$

Subsituting what we got into (13) we finally obtain:

$$z=\sum_{j\notin I_0}\alpha_j g(\lambda\alpha_j)\|x_j\|^2+\frac{1}{2\lambda}\sum_{j\in I_0}\|u_j\|^2. \qquad \square$$

We finally introduce a decomposition of the set $\Gamma(p)$ into two closed disjoint subsets, reflecting its "discrete" and "continuous" part. We set indeed $\Gamma(p)=\Gamma_0(p)\cup\Gamma_\infty(p)$ where:

$$\Gamma_0(p)=\left\{(u,\lambda)\in\Gamma(p)\,\middle|\,\sum_{j\in I_0}\|u_j\|^2=0\right\}\qquad\text{and}\qquad\Gamma_\infty(p)=\Gamma_0(p)^c.$$

The next theorem clarifies the subscripts and the terminology "discrete" and "continuous" part.

Theorem 23. *If $p\neq p_0$, the set $\Gamma_0(p)$ is finite and $\Gamma_\infty(p)$ is a closed set homeomorphic to:*

$$\Gamma_\infty(p)\simeq\bigcup_{\lambda\in\Lambda_p}S^{2N(\lambda)-1},\qquad N(\lambda)=\sum_{j\in L(\lambda)}n_j,$$

where:

$$\Lambda_p=\left\{\lambda\in\Lambda\,\middle|\,\left(z-\sum_{j\notin I_0}\alpha_j g(\lambda\alpha_j)\|x_j\|^2\right)\lambda>0\right\}. \tag{14}$$

Moreover the energy function J is constant on each component of $\Gamma(p)$.

Remark 6. By definition, $\Gamma_\infty(p)\neq\emptyset$ implies $I_0\neq\emptyset$. Thus, a necessary condition for occurrence of families of geodesics ending at $p=(x,z)$ is that some of the components x_j must vanish.

Proof. We start noticing that if $(u,\lambda)\in\Gamma_0(p)$ then all the u_j's are determined. In fact if $j\notin I_0$ then, by 11, $j\notin L(\lambda)$, $I(\alpha_j\lambda)$ is invertible and $u_j=I(\alpha_j\lambda)^{-1}x_j$; if $j\in I_0$, then the condition $\sum_{j\in I_0}\|u_j\|^2=0$ implies $u_j=0$ (notice that this condition is automatically satisfied if $I_0=\emptyset$).

Consider now the projection q onto the λ-axis:

$$q:T_0^*G\to\mathbb{R},\qquad(u,\lambda)\mapsto\lambda.$$

By the above discussion $q|_{\Gamma_0(p)}$ is one-to-one onto its image $q(\Gamma_0(p))$ and it is enough to show that this last set is discrete. To this end we notice that by Proposition 22 if $(u,\lambda)\in\Gamma_0(p)$ then:

$$z=\sum_{j\notin I_0}\alpha_j g(\alpha_j\lambda)\|x_j\|^2. \tag{15}$$

The set of solutions in λ of this equation coincides with $q(\Gamma_0(p))$ and is discrete: remember that (x,z) is fixed, the function g is strictly convex (by Proposition 21) and a linear combination of strictly convex functions is still strictly convex (on the domains of definition). Moreover since the set of solutions of (15) has no accumulation points, $q(\Gamma_0(p))$ is closed and $\Gamma_0(p) = q^{-1}(q(\Gamma_0(p)))$ is closed as well.

We prove that $\Gamma_0(p)$ is finite. If $x \neq 0$ the cardinality of $\Gamma_0(p)$ is bounded by Theorem 24 below; if $x = 0$ then equation (12) reduces to $z = \frac{1}{2\lambda}\sum_{j\in I_0}\|u_j\|^2$ and since Γ_0 is defined by $\sum_{j\in I_0}\|u_j\|^2 = 0$, it implies $z = 0$ as well, contradicting the assumption $p \neq p_0$.

Now we turn to $\Gamma_\infty(p)$. For each fixed $\lambda \in q(\Gamma_\infty(p))$ consider the fiber of the projection (the set of pairs $(u,\lambda) \in \Gamma_\infty(p)$). We show that $\lambda \in \Lambda_p$ and that the fiber is a sphere. By Proposition 22, this is the set of $u \in \mathbb{R}^{2n}$ such that $x_j = I(\lambda\alpha_j)u_j$ for every $j = 1,\dots,k$ and:

$$\frac{1}{2\lambda}\sum_{j\in I_0}\|u_j\|^2 = z - \sum_{j\notin I_0}\alpha_j g(\lambda\alpha_j)\|x_j\|^2. \tag{16}$$

Now, if $j \notin L(\lambda)$, then u_j is fixed by the value of x_j (since $I(\alpha_j\lambda)$ is invertible). For the remaining ones the only constraint comes from Eq. (16). Consider the summation in the l.h.s. Notice that $L(\lambda) \subseteq I_0$, but if $j \in I_0 \cap L(\lambda)^c$ then $u_j = 0$. Therefore:

$$\sum_{j\in I_0}\|u_j\|^2 = \sum_{j\in L(\lambda)}\|u_j\|^2. \tag{17}$$

In particular, since $(u,\lambda) \in \Gamma_\infty(p)$ this implies that $L(\lambda)$ must be non-empty, namely $\lambda \in \Lambda$. Moreover Eq. (16) reduces to:

$$\frac{1}{2\lambda}\sum_{j\in L(\lambda)}\|u_j\|^2 = z - \sum_{j\notin I_0}\alpha_j g(\lambda\alpha_j)\|x_j\|^2. \tag{18}$$

The r.h.s. of the above equation has the same sign of λ. Thus $\lambda \in \Lambda_p$ and $q^{-1}(\lambda)$ is a sphere of dimension $2N(\lambda) - 1$.

Finally q is surjective over Λ_p. In fact, for any $\lambda \in \Lambda_p$, we choose for $j \in L(\lambda)$, u_j that satisfies (18), and for $j \notin L(\lambda)$ we set $u_j = I(\alpha_j\lambda)^{-1}x_j$. The point $(u,\lambda) \in \Gamma_\infty(p)$ by construction.

The image $q(\Gamma_\infty(p))$ of this projection is certainly discrete, as it is contained into Λ. Also, it has no accumulation points, since Λ itself has no accumulation points. Thus $q(\Gamma_\infty(p))$ is closed and $\Gamma_\infty(p)$ is closed as well.

Since the energy of a geodesic (u,λ) is given by $\|u\|^2/2$, it is constant on each component. □

Remark 7. Rephrasing the above theorem, we have essentially proven that $\Gamma(p)$ is a smooth submanifold of T_0^*M; this manifold can be decomposed into the union its zero-dimensional components, $\Gamma_0(p)$, and the union of the components of dimension at least one. Since the energy function J is constant on each component, these is really a union of "critical manifolds".

4. Upper bounds

Let us introduce the following "counting" functions $\hat\nu, \hat\beta : G \to \mathbb{R}\cup\{\infty\}$:

$$\hat\nu(p) = \#\Gamma(p) \quad \text{and} \quad \hat\beta(p) = b\left(\Gamma(p)\right),$$

where $b(X)$ denotes the sum of the Betti numbers of X (which might as well be infinite a priori).

Remark 8. The Betti numbers $b_i(X)$ of a topological space X are the ranks of $H_i(X,\mathbb{Z})$ (the homology groups of X) and they measure the number of "holes" of X, see [10]. For example for a point or a line all b_i are zeroes except $b_0 = 1$; for a sphere S^k they are all zero except $b_0, b_k = 1$ (here $k > 1$). The sum of the Betti numbers $b(X)$ is sometimes called the *homological complexity* and measure how complicated X is from the topological viewpoint; for example $b(S^k) = 2$.

If $\hat{E}^{-1}(p)$ is finite, then $\hat{\nu}(p) = \hat{\beta}(p)$; on the other hand if a point p has infinitely many geodesics arriving on it $\hat{\nu}(p) = \infty$ and it could either be that they are "genuinely" infinite, i.e. also $\hat{\beta}(p) = \infty$, or they arrange in finitely many families with controlled topology, i.e. $\hat{\beta}(p) < \infty$.

Theorem 24. *Let G be a contact Carnot group. Then there exists a constant R_2 such that, for every point $p = (x, z)$, with $p \neq p_0$:*

$$\hat{\beta}(p) \leq \left(\frac{8k}{\pi}\frac{\alpha_k}{\alpha_1^2}\right)\frac{|z|}{\|x\|^2} + R_2,$$

R_2 is homogeneous of degree 0 in the singular values $\alpha_1 < \ldots < \alpha_k$ of A. In particular, if $x = (x_1, \ldots, x_k)$ has all components different from zero, then $\Gamma(p) = \Gamma_0(p)$ and:

$$\hat{\nu}(p) \leq \left(\frac{8k}{\pi}\frac{\alpha_k}{\alpha_1^2}\right)\frac{|z|}{\|x\|^2} + R_2.$$

Remark 9. Thus, whenever at least one x_j is not zero, the topology of $\Gamma(p)$ is finite; if $z \neq 0$ and $x = 0$, then the above formulas are meaningful in the sense that $\frac{|z|}{0} = \infty$.

Proof. The decomposition of Theorem 23 implies:

$$b\left(\Gamma(p)\right) = b\left(\Gamma_0(p)\right) + b\left(\Gamma_\infty(p)\right).$$

Let us start with the computation of $b(\Gamma_0(p))$. Notice that, since $\Gamma_0(p)$ consists of points, then $b(\Gamma_0(p)) = \#\Gamma_0(p)$ and:

$$\#\left(\Gamma_0(p)\right) = \#\left\{\lambda \,\middle|\, z = \sum_{j\notin I_0}\alpha_j g(\lambda\alpha_j)\|x_j\|^2\right\}. \tag{19}$$

We recall that $I_0 = \{j \mid x_j = 0\}$ and distinguish two cases.

1. If $I_0 = \{1, \ldots, k\}$ (i.e. $x = 0$), then $\Gamma_0(p)$ is empty: in fact from 19 we obtain that also $z = 0$, contradicting the assumption $p \neq p_0$.
2. If $I_0 \subsetneq \{1, \ldots, k\}$ (there is at least one $x_j \neq 0$), then property 4 of Proposition 21 implies that:

$$|z| = \left|\sum_{j\notin I_0}\alpha_j g(\lambda\alpha_j)\|x_j\|^2\right| > \frac{|\lambda|}{8}\sum_{j\notin I_0}\alpha_j^2\|x_j\|^2 - \frac{\pi}{8}\sum_{j\notin I_0}\alpha_j\|x_j\|^2,$$

which tells equivalently that:

$$|\lambda| < \frac{8|z|}{\sum_{j\notin I_0}\alpha_j^2\|x_j\|^2} + \frac{\pi\sum_{j\notin I_0}\alpha_j\|x_j\|^2}{\sum_{j\notin I_0}\alpha_j^2\|x_j\|^2} \leq \frac{8|z|}{\alpha_1^2\|x\|^2} + \frac{\pi\alpha_k}{\alpha_1^2} =: \rho, \tag{20}$$

where in the last inequality we have used the fact that $\|x\|^2 = \sum_{j\notin I_0}\|x_j\|^2$. The number of solutions of (19) is the number of intersections of the horizontal line $w = z$ with the graph of:

$$G_0(\lambda) = \sum_{j\notin I_0}\alpha_j g(\lambda\alpha_j)\|x_j\|^2,$$

in the (λ, w)-plane, with the restriction $|\lambda| < \rho$ we found in (20). The function G_0, being a sum of strictly convex functions, is itself strictly convex (on each interval where it is defined) and the number of points of intersections of $w = z$ with its graph is:

$$b\left(\Gamma_0(p)\right) \leq 2\#\{\text{poles of } G_0 \text{ on the interval } (0, \rho)\} + 1.$$

Since the function G_0 has poles exactly on the sets $\Lambda_j = \{\lambda \neq 0 \,|\, \lambda\alpha_j \in 2\pi\mathbb{Z},\, j \notin I_0\}$, we obtain:

$$
\begin{aligned}
b(\Gamma_0(p)) &\leq 2\sum_{j\notin I_0} \left\lfloor \frac{\rho\alpha_j}{2\pi} \right\rfloor + 1 \leq 2\sum_{j\notin I_0} \left\lfloor \frac{4\alpha_j|z|}{\pi\alpha_1^2\|x\|^2} + \frac{\alpha_k\alpha_j}{2\alpha_1^2} \right\rfloor + 1 \leq \\
&\leq (k - \#I_0)\frac{8}{\pi}\frac{\alpha_k}{\alpha_1^2}\frac{|z|}{\|x\|^2} + r_0,
\end{aligned} \tag{21}
$$

where r_0 is a bounded remainder (homogeneous of degree 0 in the singular values) given by:

$$
r_0 = (k - \#I_0)\frac{\alpha_k^2}{\alpha_1^2} + 1.
$$

Let us consider now $b(\Gamma_\infty(p))$. By Theorem 23, $\Gamma_\infty(p)$ is a disjoint union of spheres, one sphere for each point $\lambda \in \Lambda_p$, where:

$$
\Lambda_p = \{\lambda \in \Lambda \mid (z - G_0(\lambda))\lambda > 0\}.
$$

Since the total Betti number of sphere is 2 (independently on the dimension), we have:

$$
b(\Gamma_\infty(p)) = b\left(\bigcup_{\lambda\in\Lambda_p} S^{2N(\lambda)-1}\right) = 2\#\Lambda_p. \tag{22}
$$

We assume $z \geq 0$ for simplicity. This implies $\lambda > 0$. Moreover, if $\lambda \in \Lambda_p \subseteq \Lambda$, then λ must belong to the complement of the set of poles of the function G_0, namely

$$
\lambda \in \Lambda_0 := \bigcup_{j\in I_0} \Lambda_j = \bigcup_{j\in I_0} \frac{2\pi}{\alpha_j}\mathbb{Z} \setminus \{0\} \subseteq \Lambda.
$$

Thus we finally rewrite:

$$
\Lambda_p = \{\lambda \in \Lambda_0 \mid \lambda > 0, \quad z > G_0(\lambda)\}.
$$

It only remains to estimate the cardinality of Λ_p. We distinguish again two cases.

1. $I_0 = \{1, \dots, k\}$ (i.e. $x = 0$). By our assumption $p \neq p_0$ it follows that $z > 0$. Moreover, in this case $G_0(\lambda) \equiv 0$ and $\Lambda_0 = \Lambda$. Therefore $\Lambda_p = \Lambda$ is infinite and $\Gamma_\infty(p)$ consists of infinitely many spheres, thus $b(\Gamma_\infty(p)) = \infty$.

2. $I_0 \subsetneq \{1, \dots, k\}$. In this case we have to count the $\bar\lambda > 0$, such that:

$$
z > \sum_{j\notin I_0} \alpha_j g(\bar\lambda\alpha_j)\|x_j\|^2, \qquad \text{with } \bar\lambda \in \Lambda_0. \tag{23}
$$

Arguing exactly as in (20) we obtain that:

$$
|\bar\lambda| < \frac{8|z|}{\alpha_1^2\|x\|^2} + \frac{\pi\alpha_k}{\alpha_1^2} := \rho.
$$

Thus the number of $\bar\lambda$ satisfying (23) is bounded by the (finite) number of elements $\bar\lambda \in \Lambda_0$ in the interval $(0, \rho)$ (arguing as in (21)):

$$
\sum_{j\in I_0} \left\lfloor \frac{\rho\alpha_j}{2\pi} \right\rfloor \leq \#I_0\frac{4}{\pi}\frac{\alpha_k}{\alpha_1^2}\frac{|z|}{\|x\|^2} + \frac{\#I_0}{2}\frac{\alpha_k^2}{\alpha_1^2},
$$

Combining this with (22) we get:

$$
b\left(\Gamma_\infty(p)\right) \leq \#I_0\frac{8}{\pi}\frac{\alpha_k}{\alpha_1^2}\frac{|z|}{\|x\|^2} + r_\infty,
$$

where r_∞ is a bounded remainder (homogeneous of degree 0 in the singular values) given by:

$$r_\infty = \# I_0 \frac{\alpha_k^2}{\alpha_1^2}.$$

Finally, since the union $\Gamma_0(p) \cup \Gamma_\infty(p)$ is disjoint and closed, we obtain:

$$\begin{aligned} b\left(\Gamma(p)\right) &= b\left(\Gamma_0(p)\right) + b\left(\Gamma_\infty(p)\right) \leq \\ &\leq (k - \# I_0)\frac{8}{\pi}\frac{\alpha_k}{\alpha_1^2}\frac{|z|}{\|x\|^2} + \# I_0 \frac{8}{\pi}\frac{\alpha_k}{\alpha_1^2}\frac{|z|}{\|x\|^2} + r_0 + r_\infty = \qquad (24) \\ &= \left(k\frac{8}{\pi}\frac{\alpha_k}{\alpha_1^2}\right)\frac{|z|}{\|x\|^2} + R_2, \end{aligned}$$

where R_2 is a bounded remainder (homogeneous of degree 0 in the singular values) given by:

$$R_2 = r_0 + r_\infty = k\frac{\alpha_k^2}{\alpha_1^2}.$$

Notice that if all $x_j \neq 0$, then $I_0 = \emptyset$ and $\Gamma(p)$ reduces simply to $\Gamma_0(p)$, which is necessarily finite by the first part of the statement. □

Remark 10. It is interesting to notice that equation (24) splits clearly the contribution to the topology into two pieces:

$$b\left(\Gamma(p)\right) \leq \underbrace{(k - \# I_0)\frac{8}{\pi}\frac{\alpha_k}{\alpha_1^2}\frac{|z|}{\|x\|^2}}_{\geq b(\Gamma_0(p))} + \underbrace{\# I_0 \frac{8}{\pi}\frac{\alpha_k}{\alpha_1^2}\frac{|z|}{\|x\|^2}}_{\geq b(\Gamma_\infty(p))} + R_2.$$

If we ignore the error term and interpret the r.h.s. with the convention of Remark 9, this inequality is correct (for example if $\# I_0 = 0$ then $b(\Gamma_\infty(p)) = 0$ and if $\# I_0 = k$ there is no discrete part).

The following proposition gives a quantitative way to estimate when there is only one geodesic.

Proposition 25. *Let G be a contact Carnot group and $p = (x, z)$ such that:*

$$|z| < \frac{\pi}{8}\left(\frac{2\alpha_1^2}{\alpha_k} - \alpha_k\right)\|x\|^2. \qquad (25)$$

Then there is only one geodesic from p_0 to p. Moreover if $G = \mathbb{H}_{2n+1}$, let λ_1 be the first positive solution of $\lambda = \tan\lambda$. Then there is only one geodesic to p if and only if:

$$|z| \leq \frac{\lambda_1}{4}\|x\|^2. \qquad (26)$$

Proof. We give the proof only for the Heisenberg case; we leave to the reader the verification that condition (25) guarantees that $\Gamma_\infty(p) = \emptyset$ and that $\hat{\nu}(p) = 1$.

In the case of the Heisenberg group, condition (26) implies automatically that p is not a vertical point and by Theorem 23 $\Gamma_\infty(p) = \emptyset$. In particular, using (19), we can write:

$$\hat{\nu}(p) = \#\left\{\lambda \,\middle|\, z = g(\lambda)\|x\|^2\right\}. \qquad (27)$$

Let $2\lambda_1$ be the first positive local minimum of g; $2\lambda_1$ is also a zero of g' and by property 3. of Proposition 21 we have $g(2\lambda_1) = \frac{\lambda_1}{4}$. Computing the derivative of g we obtain:

$$g'(\lambda) = \frac{2 - \lambda\cot\left(\frac{\lambda}{2}\right)}{\left(\sin\frac{\lambda}{2}\right)^2}.$$

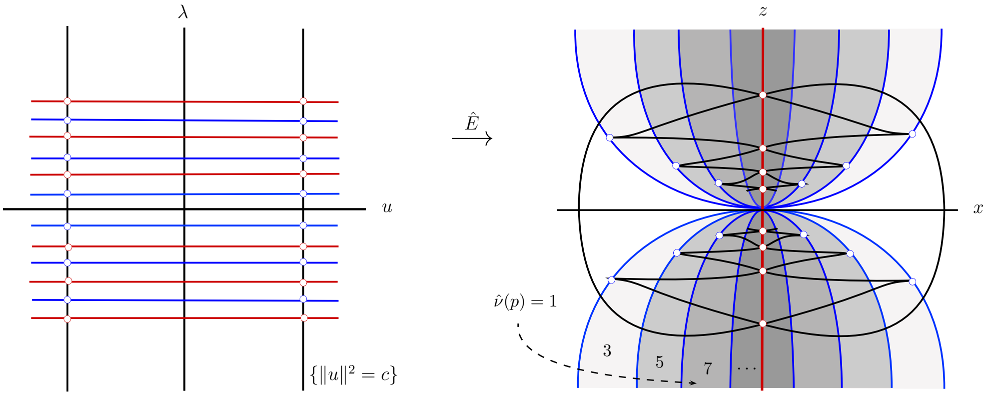

FIGURE 6. A two-dimensional qualitative picture of the exponential map for the Heisenberg group. The critical points are the λ-axes, the set R (in red) and the set B (in blue). The broken curve is the section in the (x, z)-plane of the image of the cylinder $\{\|u\|^2 = c\}$. The number of geodesics to p is constant on each shaded region (the white one is where $\hat{\nu}(p) = 1$). When c varies the blue dots on the right figure (the images of $B \cap \{\|u\|^2 = c\}$) "span" all the paraboloids $|z| = g(\lambda_k)\|x\|^2$.

Thus λ_1 is the first positive solution of $\tan\lambda = \lambda$ and:

$$g(\lambda_1)\|x\|^2 = \frac{\lambda_1}{4}\|x\|^2. \tag{28}$$

By (27) $\hat{\nu}(p) = 1$ as long as $|z| < G_0(\lambda_1)$: condition (26) and (28) imply the statement. □

Example 6 (Heisenberg, conclusion). The Jacobian of the exponential map in the 3-dimensional standard Heisenberg case can be computed explicitly using (7):

$$\det\left(d_{(u,\lambda)}\hat{E}\right) = -\frac{\|u\|^2(\lambda\sin\lambda + 2\cos\lambda - 2)}{\lambda^4}.$$

Setting to zero the previous equation we find critical points of $\hat{E}$:

$$\{\text{critical points of } \hat{E}\} = \underbrace{\left\{\|u\|^2 = 0\right\}}_{A} \cup \underbrace{\left\{\lambda = 2k\pi, k \neq 0\right\}}_{R} \cup \underbrace{\left\{\lambda \mid \tfrac{\lambda}{2} = \tan\tfrac{\lambda}{2}\right\}}_{B}.$$

The critical values are the images of these sets. For convenience of notations we label λ_k the solutions of $\frac{\lambda}{2} = \tan\frac{\lambda}{2}$: these numbers, in the case of the Heisenberg group, coincide with the minima of the function g and are of the form $\lambda_k = (2k+1)\pi + \varepsilon_k$. The critical values of $\hat{E}$ decompose into the union of the three sets:

$$\hat{E}(A) = \text{origin}, \qquad \hat{E}(R) = z\text{-axis}, \qquad \hat{E}(B) = \left\{z = \|x\|^2 g(\lambda_k) \mid k \in \mathbb{Z}\right\}.$$

In particular, $\hat{E}(B)$ is an hypersurface (a paraboloid), and has the following interesting characterization: for $x \neq 0$, we have:

$$\hat{\nu}(p) = \#\{\lambda \mid z = g(\lambda)\|x\|^2\}.$$

By the properties of the function g (Proposition 21), and assuming $z \geq 0$, two new contributions to $\hat{\nu}(p)$ appear (or disappear) every time the ratio $|z|/\|x\|^2$ crosses the values $g(\lambda_k)$, for $k \geq 0$ (resp. for $k \leq 0$ assuming ≤ 0). Thus the function $\hat{\nu}(p)$ "jumps" by two every time p crosses $\hat{E}(B)$ transversally (see Figure 6).

In the general case, the computation of the Jacobian of the exponential map for corank-one Carnot groups is done in [3, Lemma 38].

5. Lower bounds

In this section we prove the lower bounds of Theorems 2 and 5 for the number of geodesics $\hat{\nu}(p) = \#\Gamma(p)$ and the topology $\hat{\beta}(p) = b(\Gamma(p))$ in a contact Carnot group G, with constants $(k, \vec{n}, \vec{\alpha})$. According to the decomposition of Section 3, for $p \neq p_0$ we have the following splitting:

$$\Gamma(p) = \Gamma_0(p) \cup \Gamma_\infty(p),$$

where $\Gamma_0(p)$ is a finite set and $\Gamma_\infty(p)$ is homeomorphic to a disjoint union of spheres. According to Remark 6, if $p = (x, z)$ is a point with all components $x_j \neq 0$, then $\Gamma_\infty(p) = \emptyset$ (in particular this is the case for a generic point p). In this setting we prove the next theorem.

Theorem 26 (The "infinitesimal" lower bound)**.** *Given a contact Carnot group G, there exist constants C_1, R_1 such that if $p = (x, z) \in G$ has all components x_j different from zero, then:*

$$C_1 \frac{|z|}{\|x\|^2} + R_1 \leq \hat{\nu}(p).$$

In particular, denoting by α_1 and α_k the smallest and the largest singular values of A:

$$C_1 = \frac{8}{\pi}\frac{\alpha_1}{\alpha_k^2}\sin\left(\frac{\delta\pi}{2}\right)^2 \quad \textit{with} \quad \delta = \left(\sum_{j=1}^k \frac{\alpha_1}{\alpha_j}\left\lfloor\frac{\alpha_j}{\alpha_1}\right\rfloor\right)^{-1}.$$

Moreover, R_1 (resp. C_1) is homogeneous of degree 0 (resp. -1) in the singular values $\alpha_1, \dots, \alpha_k$.

Proof. When all the $x_j \neq 0$, then $\Gamma(p) = \Gamma_0(p)$. According to Proposition 22, and recalling that $I_0 = \emptyset$, the number of geodesics ending at $p = (x, z)$ is computed by:

$$\hat{\nu}(p) = \#\{\lambda \mid z = G(\lambda)\}, \qquad G(\lambda) := \sum_{j=1}^k \alpha_j \|x_j\|^2 g(\alpha_j \lambda).$$

The idea of the proof is to build an increasing sequence of values $\hat{\lambda}_n$, with "spacing" growing linearly with n, such that $G(\hat{\lambda}_n) \leq cn + d$ for some constants c and d. By the strict convexity of $G(\lambda)$, we have at least one contribution to $\hat{\nu}(p)$ for any point $\hat{\lambda}_n$ of the sequence such that $G(\hat{\lambda}_n) < z$. Then it is sufficient to count the number of points $\hat{\lambda}_n$ of the sequence such that $G(\hat{\lambda}_n) < z$. We now proceed with the proof.

Without loss of generality, we assume $z \geq 0$ and then $\lambda \geq 0$. For fixed $0 < \delta \leq 1$ and every $j = 1, \dots, k$ define the intervals:

$$I_{n,j} := \left[\frac{2n\pi}{\alpha_j}, \frac{2(n+1)\pi}{\alpha_j}\right] \qquad \text{and} \qquad \hat{I}_{n,j} := \left[\frac{2n\pi}{\alpha_j} + \frac{\delta\pi}{\alpha_j}, \frac{2(n+1)\pi}{\alpha_j} - \frac{\delta\pi}{\alpha_j}\right].$$

Each interval $\hat{I}_{n,j}$ is contained in $I_{n,j}$ and the lengths of these two intervals are (see Figure 7):

$$|I_{n,j}| = \frac{2\pi}{\alpha_j} =: a_j \qquad \text{and} \qquad |\hat{I}_{n,j}| = a_j(1-\delta).$$

The singular values $0 < \alpha_1 < \ldots < \alpha_k$ are ordered, then the intervals $I_{n,1}$, for $n \in \mathbb{N}$ are the largest. We also define $y_n := (2n - \delta)\pi$. Notice that $\frac{y_{n+1}}{\alpha_j}$ is the maximum of the interval $\hat{I}_{n,j}$, and will play an important role in the proof.

Each function $\lambda \mapsto g(\alpha_j \lambda)$ is unbounded in the intervals $I_{n,j}$ (it has poles at the extrema), but it is controlled on all the smaller intervals $\hat{I}_{n,j}$, as stated by the next lemma.

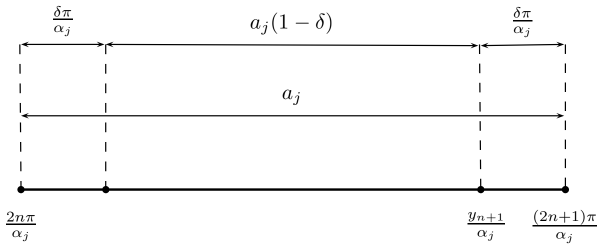

FIGURE 7. The intervals $\hat{I}_{n,j} \subset I_{n,j}$.

Lemma 27. *There exist constants $c_1(\delta)$, $d_1(\delta)$ such that, for all $j = 1, \ldots, k$:*

$$g(\alpha_j \lambda) \leq c_1(\delta)n + d_1(\delta), \qquad \forall \lambda \in \hat{I}_{n,j}.$$

Proof. By Proposition 21, for all $j = 1, \ldots, k$ the functions $\lambda \mapsto g(\alpha_j \lambda)$ are strictly convex on the intervals $\hat{I}_{n,j} \subset I_{n,j}$. Each function is clearly unbounded on $I_{n,j}$ but, when restricted on $\hat{I}_{n,j}$, it achieves its maximum value at the point $\frac{y_{n+1}}{\alpha_j}$ (i.e. the maximum of the interval $\hat{I}_{n,j}$). Therefore, by explicit evaluation, for all $\lambda \in \hat{I}_{n,j}$ we have:

$$g(\alpha_j \lambda) \leq g(y_{n+1}) = \frac{2\pi}{8\sin(\delta\pi/2)^2} n + \frac{2\pi - \delta\pi + \sin(\delta\pi)}{8\sin(\delta\pi/2)^2} = c_1(\delta)n + d_1(\delta). \qquad \square$$

The next lemma implies that, for each $n \geq 0$, the large interval $I_{n,1}$ contains at least one point that belongs to all the smaller intervals $\hat{I}_{m_1,1}, \ldots, \hat{I}_{m_k,k}$, for some $m_1, \ldots, m_k$.

Lemma 28. *Let $\hat{I}_j = \bigcup_{m \geq 0} \hat{I}_{m,j}$ for all $j = 1, \ldots, k$. If $0 < \delta \leq 1$ is small enough then:*

$$\forall n \geq 0 \qquad I_{n,1} \cap \bigcap_{j=1}^{k} \hat{I}_j \neq \emptyset.$$

Proof. We argue by contradiction. Assume there exists $n \geq 0$ such that for all $\lambda \in I_{n,1}$ we can find $j \in \{1, \ldots, k\}$ with $\lambda \notin \hat{I}_j$. This implies:

$$s_n := \int_{I_{n,1}} \# \left\{ j \mid \lambda \in \hat{I}_j \right\} dz \leq (k-1)a_1. \tag{29}$$

On the other hand the above integral equals:

$$s_n = \sum_{j=1}^{k} |\hat{I}_j \cap I_{n,1}| \geq k a_1 - \sum_{j=1}^{k} \frac{2\delta\pi}{\alpha_j} \left\lfloor \frac{\alpha_j}{\alpha_1} \right\rfloor \geq (k-1)a_1 + \left(a_1 - \delta \sum_{j=1}^{k} \frac{2\pi}{\alpha_j} \left\lfloor \frac{\alpha_j}{\alpha_1} \right\rfloor \right).$$

Recalling that $a_1 = \frac{2\pi}{\alpha_1}$, if we choose

$$0 < \delta < \left(\sum_{j=1}^{k} \frac{\alpha_1}{\alpha_j} \left\lfloor \frac{\alpha_j}{\alpha_1} \right\rfloor \right)^{-1}, \tag{30}$$

we obtain $s_n > (k-1)a_1$, contradicting (29). $\square$

The next lemma builds a sequence $\hat{\lambda}_n$ where the behaviour of $G(\lambda)$ is controlled.

Lemma 29. *There exists an unbounded, increasing sequence $\{\hat{\lambda}_n \in I_n\}_{n\in\mathbb{N}}$ and constants $c_k(\delta)$, $d_k(\delta)$ such that:*

$$\sum_{j=1}^{k} g(\hat{\lambda}_n \alpha_j) \leq c_k(\delta) n + d_k(\delta).$$

Proof. By Lemma (28), for all $n \geq 0$ there is a point $\hat{\lambda}_n \in I_{n,1} \cap \hat{I}_{m_1,1} \cap \hat{I}_{m_2,2} \cap \ldots \cap \hat{I}_{m_k,k}$, for some $m_1, \ldots, m_k$. This sequence is unbounded and increasing by construction. Notice that, by construction, $m_1 = n$ and

$$m_j \leq \left\lfloor \frac{(n+1)\alpha_j}{\alpha_1} \right\rfloor \leq \frac{\alpha_k}{\alpha_1} n + 2\frac{\alpha_k}{\alpha_1}, \qquad j = 2, \ldots, k.$$

By the estimates of Lemma 27, we have

$$\sum_{j=1}^{k} g(\hat{\lambda}_n \alpha_j) \leq \sum_{j=1}^{k} (c_1(\delta) m_j + d_1(\delta)) \leq \underbrace{\left[c_1(\delta) k \frac{\alpha_k}{\alpha_1}\right]}_{c_k(\delta)} n + \underbrace{\left[2c_1(\delta)(k-1)\frac{\alpha_k}{\alpha_1} + k d_1(\delta)\right]}_{d_k(\delta)}. \qquad \square$$

We are now ready for the computation of the lower bound for $\hat{\nu}(p)$. Indeed

$$\hat{\nu}(p) = \#\{\lambda \mid z = G(\lambda)\}, \qquad G(\lambda) := \sum_{j=1}^{k} \alpha_j g(\lambda \alpha_j) \|x_j\|^2. \tag{31}$$

By Proposition 21, each function $\lambda \mapsto g(\alpha_j \lambda)$ is strictly convex in the intervals $I_{n,j}$, for $n \in \mathbb{N}$, and has poles at the extrema of $I_{n,j}$ (excluding the point $\lambda = 0$), i.e. the discrete set Λ_j. Then also $G(\lambda)$ is a strictly convex function in each interval in which it is defined, with poles at the points of $\Lambda = \cup_{j=1}^{k} \Lambda_j$.

Consider the sequence $\hat{\lambda}_n$ of Lemma 29. There are at least 2 solutions contributing to Eq. (31) for any value $\hat{\lambda}_n$ such that $G(\hat{\lambda}_n) < z$. This follows by strict convexity of G in the interval between two successive poles containing $\hat{\lambda}_n$. The only exception to this rule is when $\hat{\lambda}_n$ belongs to $I_{0,k}$: in this case there is only 1 solution (there is no pole at $\lambda = 0$). We have:

$$G(\hat{\lambda}_n) \leq \alpha_k \|x\|^2 \sum_{j=1}^{k} g(\alpha_j \hat{\lambda}_n) \leq \alpha_k \|x\|^2 \left[c_k(\delta) n + d_k(\delta)\right].$$

According to the discussion above, $\hat{\lambda}_0$ gives a contribution of 1 to $\hat{\nu}(p)$, while each point $\hat{\lambda}_n$, with $n \geq 1$ of the sequence, such that $G(\hat{\lambda}_n) < z$, give a contribution of 2 to $\hat{\nu}(p)$. Therefore, taking in account all the contributions, we obtain:

$$\hat{\nu}(p) \geq 2\left\lfloor \frac{1}{c_k(\delta)} \frac{|z|}{\alpha_k \|x\|^2} - \frac{d_k(\delta)}{c_k(\delta)} \right\rfloor + 1 \geq \frac{2}{\alpha_k c_k(\delta)} \frac{|z|}{\|x\|^2} - 2\frac{d_k(\delta)}{c_k(\delta)} - 1.$$

Plugging in the constants obtained above, we obtain:

$$\hat{\nu}(p) \geq C(\delta) \frac{|z|}{\|x\|^2} + R(\delta),$$

with:

$$C(\delta) := \frac{8}{\pi} \frac{\alpha_1}{\alpha_k^2} \sin\left(\frac{\delta\pi}{2}\right)^2, \qquad R(\delta) := 4\frac{1-k}{k} + \frac{\alpha_1}{\alpha_k} \frac{\delta\pi - \sin(\delta\pi) - 2\pi}{\pi} - 1.$$

Both $C(\delta)$ and $R(\delta)$ are non-decreasing functions of δ, for $0 < \delta \leq 1$, thus the best estimate is given by the values at the largest δ. According to (30) this value is:

$$\delta_M := \left(\sum_{j=1}^{k} \frac{\alpha_1}{\alpha_j} \left\lfloor \frac{\alpha_j}{\alpha_1} \right\rfloor \right)^{-1}.$$

Notice that $C_1 := C(\delta_M)$ is homogeneous of degree -1 w.r.t. the singular values $\alpha_1, \dots, \alpha_k$, while $R_1 := R(\delta_M)$ is homogeneous of degree 0. □

The previous theorem holds if all the x_j are different from zero. When some of the $x_j = 0$, continuous families might appear. Nevertheless, the topology of these families is controlled. We obtain the following lower bound.

Theorem 30 (The "infinitesimal" lower bound for the topology)**.** *Let G be a contact Carnot group. There exist constants R_1', C_1' such that for every $p = (x, z) \in G$ with $p \neq p_0$:*

$$C_1' \frac{|z|}{\|x\|^2} + R_1' \leq \hat{\beta}(p).$$

In particular, denoting by α_1 and α_k the smallest and the largest singular values of A:

$$C_1' = \frac{8}{\pi} \frac{\alpha_1}{\alpha_k^2} \sin\left(\frac{\delta' \pi}{2}\right)^2 \quad \text{with} \quad \delta' = \left(\sum_{j \notin I_0} \frac{\alpha_1}{\alpha_j} \left\lfloor \frac{\alpha_j}{\alpha_1} \right\rfloor \right)^{-1}.$$

Moreover, R_1' (resp. C_1') is homogeneous of degree 0 (resp. -1) in the singular values $\alpha_1, \dots, \alpha_k$.

Proof. Recall that $I_0 = \{j \mid x_j = 0\}$. If $I_0 = \emptyset$ the statement reduces to Theorem 30 since $\Gamma(p) = \Gamma_0(p)$ is finite and $\hat{\nu}(p) = \#\Gamma(p) = b(\Gamma(p)) = \hat{\beta}(p)$. Then assume $I_0 \neq \emptyset$. By Theorem 22, $\Gamma(p) = \Gamma_0(p) \cup \Gamma_\infty(p)$ and:

$$\hat{\beta}(p) = b(\Gamma(p)) \geq b(\Gamma_0(p)) = \#\Gamma_0(p).$$

In particular $\Gamma_0(p)$ is in one-to-one correspondence with its projection on the λ component, since all the u_j are uniquely determined by the point $p = (x, z)$ once λ is known. Therefore

$$\#\Gamma_0(p) = \#\{\lambda \mid z = G_0(\lambda)\}, \qquad G_0(\lambda) := \sum_{j \notin I_0} \alpha_j g(\alpha_j \lambda) \|x_j\|^2.$$

Now we only have to bound from below the number of solutions of $z = G(\lambda)$. The proof is analogous to the one of Theorem 26, where only the indices $j \notin I_0$ appear. □

6. Isometries of the Heisenberg group

Sub-Riemannian isometries are distance-preserving transformations. According to the result in [8], isometries of Carnot groups are smooth. Moreover the set of all sub-Riemannian isometries $\mathrm{ISO}(G)$ of a Carnot group is a Lie group, and any isometry is the composition of a group automorphism and a group translation (see [9, 14]).

Remark 11. For our purposes it is sufficient to consider the subgroup $\mathrm{ISO}_0(G)$ of isometries that fix the identity. With no risk of confusion, we denote this subgroup simply $\mathrm{ISO}(G)$.

In this section we focus on the case $G = \mathbb{H}_{2n+1}$ (we refer to the notation of Section 2, and Example 5). Then we use this result to study the isometry group of any contact Carnot group.

Lemma 31. *The isometry group of* $\mathbb{H}_{2n+1}$ *is:*

$$\mathrm{ISO}(\mathbb{H}_{2n+1}) = \{(M,\theta) \mid \theta \in \{-1,+1\},\ MM^* = \mathbb{1}_{2n},\ MJM^* = \theta J\},$$

with the action of $\mathrm{ISO}(\mathbb{H}_{2n+1})$ *on* $\mathbb{H}_{2n+1}$ *given by:*

$$(M,\theta)\cdot(x,z) = (Mx, \theta z).$$

Moreover:

$$\mathrm{ISO}(\mathbb{H}_{2n+1}) \simeq \mathrm{O}(2n)\cap \mathrm{Sp}(2n) \rtimes \mathbb{Z}_2 \simeq \mathrm{U}(n)\rtimes \mathbb{Z}_2.$$

Proof. A diffeomorphism is an isometry of Carnot groups fixing the identity if and only if is a Lie group isomorphism. In particular, it is induced by Lie algebra isomorphisms $\phi : \mathfrak{h}_{2n+1} \to \mathfrak{h}_{2n+1}$ that are orthogonal transformations on the first layer. Since ϕ is a Lie algebra isomorphism, it preserves the stratification. Then we can write $\phi = (M,\theta) \in \mathrm{O}(2n)\times\mathbb{R}$, such that

$$\phi(f_i) = \sum_{j=1}^{2n} M_{ji} f_j, \qquad \phi(f_0) = \theta f_0.$$

The isomorphism condition $[\phi(f_i),\phi(f_j)] = J_{ij}\phi(f_0)$ implies:

$$MJM^* = \theta J.$$

It follows that $\theta^2 = 1$. Then:

$$\mathrm{ISO}(\mathbb{H}_{2n+1}) = \{(M,\theta) \mid \theta \in \{-1,+1\},\ MM^* = \mathbb{1}_{2n},\ MJM^* = \theta J\}.$$

This Lie algebra isomorphism generates a Lie group isomorphism that, in exponential coordinates, reads $(M,\theta)\cdot(x,z) = (Mx,\theta z)$. Let $\mathrm{ISO}(\mathbb{H}_{2n+1})_+ \lhd \mathrm{ISO}(\mathbb{H}_{2n+1})$ be the normal subgroup:

$$\mathrm{ISO}(\mathbb{H}_{2n+1})_+ := \{(M,1) \mid MM^* = \mathbb{1},\ MJM^* = J\} \simeq \mathrm{O}(2n)\cap\mathrm{Sp}(2n).$$

Moreover, let K be any matrix such that $KJK^* = -J$. Then, let :

$$H := \{(\mathbb{1},1),(K,-1)\} \simeq \mathbb{Z}_2$$

be another subgroup of $\mathrm{ISO}(\mathbb{H}_{2n+1})$. Any element of $\mathrm{ISO}(\mathbb{H}_{2n+1})$ can be written uniquely as the product mh of an element of $m \in \mathrm{ISO}(\mathbb{H}_{2n+1})_+$ and an element of $h \in H$. Thus the map $mh \mapsto (m,h)$ is a group isomorphism:

$$\mathrm{ISO}(\mathbb{H}_{2n+1}) = \mathrm{ISO}(\mathbb{H}_{2n+1})_+ \rtimes H,$$

where H acts on $\mathrm{ISO}(\mathbb{H}_{2n+1})_+$ with the adjoint action. As we observed $\mathrm{ISO}(\mathbb{H}_{2n+1})_+ \simeq \mathrm{O}(2n)\cap \mathrm{Sp}(2n)$ and $H \simeq \mathbb{Z}_2$, thus

$$\mathrm{ISO}(\mathbb{H}_{2n+1}) \simeq \mathrm{O}(2n)\cap\mathrm{Sp}(2n)\rtimes\mathbb{Z}_2.$$

Remark 12. With this identification, the action of $\varphi : \mathbb{Z}_2 \to \mathrm{Aut}(\mathrm{O}(2n)\cap\mathrm{Sp}(2n))$ is:

$$\varphi(1)M = M, \qquad \varphi(-1)M = KMK^*,$$

the product on $\mathrm{O}(2n)\cap\mathrm{Sp}(2n)\rtimes\mathbb{Z}_2$ reads:

$$(M,\theta)(M',\theta') = (M\varphi(\theta)M', \theta\theta'),$$

and the action of $\mathrm{O}(2n)\cap\mathrm{Sp}(2n)\rtimes\mathbb{Z}_2$ on $\mathbb{H}_{2n+1}$ is:

$$(M,\theta)\cdot(x,z) = \begin{cases} (Mx,z) & \theta = 1, \\ (MKx,-z) & \theta = -1. \end{cases}$$

Finally, to see that $\mathrm{O}(2n) \cap \mathrm{Sp}(2n) \simeq \mathrm{U}(n)$, write $M \in \mathrm{GL}(2n, \mathbb{R})$ as $M = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$. Then $M \in \mathrm{O}(2n) \cap \mathrm{Sp}(2n)$ if and only if:

$$M = \begin{pmatrix} A & B \\ -B & A \end{pmatrix}, \qquad AA^* + BB^* = \mathbb{1}_n, \quad AB^* - BA^* = 0.$$

Thus the map $M \mapsto A + iB$ is the group isomorphism $\mathrm{O}(2n) \cap \mathrm{Sp}(2n) \simeq \mathrm{U}(n)$. □

6.1. **Stabilizers of points.** Let $p \in \mathbb{H}_{2n+1}$. From now on we restrict our attention to the connected component $\mathrm{ISO}(\mathbb{H}_{2n+1})_+$ that contains the identity. As in the proof of Lemma 31, we identify:

$$\mathrm{ISO}(\mathbb{H}_{2n+1})_+ = \mathrm{U}(n).$$

With this identification, the action $\rho : \mathrm{U}(n) \times \mathbb{H}_{2n+1} \to \mathbb{H}_{2n+1}$ is

$$\rho(A + iB, (x, z)) = (Mx, z), \qquad M = \begin{pmatrix} A & B \\ -B & A \end{pmatrix}.$$

What is the stabilizer subgroup $\mathrm{ISO}_p(\mathbb{H}_{2n+1}) \subseteq \mathrm{ISO}(\mathbb{H}_{2n+1})_+$ that fixes $p \in \mathbb{H}_{2n+1}$?

Lemma 32. *Let $p = (x, z) \in \mathbb{H}_{2n+1}$. Then*

$$\mathrm{ISO}_p(\mathbb{H}_{2n+1}) \simeq \begin{cases} \mathrm{U}(n) & x = 0, \\ \mathrm{U}(n-1) & x \neq 0. \end{cases}$$

Proof. Let $A + iB \in \mathrm{U}(n)$. Let $p = (x, z) \in \mathbb{H}_{2n+1}$, with $x \neq 0$ and write $x = (v, w)$ with $v, w \in \mathbb{R}^n$. Then

$$\rho(A + iB, p) = p \Longleftrightarrow Mx = x \Longleftrightarrow (A + iB)(v - iw) = v - iw.$$

This means that $A + iB$ must be a unitary matrix with a prescribed eigenvector $v - iw$ with eigenvalue 1. This identifies a copy of $\mathrm{U}(n-1) \subset \mathrm{U}(n)$ that fixes p. On the other hand, if $x = 0$, the point $p = (0, z)$ is fixed for any element of $\mathrm{ISO}(\mathbb{H}_{2n+1})_+$. □

6.2. **Stabilizers of geodesics.** Let $(u, \lambda) \in T_0^*\mathbb{H}_{2n+1}$. Let $\gamma(t)$ the associated normal geodesic. What is the subgroup $\mathrm{ISO}_\gamma(\mathbb{H}_{2n+1}) \subset \mathrm{ISO}(\mathbb{H}_{2n+1})_+$ that fixes the whole geodesic? Recall that

$$\gamma(t) = \left(\int_0^t e^{-\lambda J \tau} u d\tau, -\frac{1}{2} \int_0^t \langle e^{-\lambda J \tau} u, J \int_0^\tau e^{-\lambda J s} u ds \rangle d\tau \right).$$

Lemma 33. *Let $(u, \lambda) \in T_0^*\mathbb{H}_{2n+1}$ be the initial covector of the geodesic γ. Then*

$$\mathrm{ISO}_\gamma(\mathbb{H}_{2n+1}) \simeq \begin{cases} \mathrm{U}(n) & u = 0, \\ \mathrm{U}(n-1) & u \neq 0. \end{cases}$$

Proof. Let $A + iB \in \mathrm{U}(n)$ be an isometry and $\gamma(t) = (x(t), z(t))$. Then $\rho(A + iB, (x(t), z(t))) = (x(t), z(t))$ if and only if $Mx(t) = x(t)$ for all t, namely to stabilize the geodesic is equivalent to stabilize its "horizontal" component. For $u \neq 0$, take one derivative w.r.t. t at $t = 0$; we obtain:

$$Mu = u,$$

as in the proof of Lemma 32, this identifies a subgroup $\mathrm{U}(n-1) \subset \mathrm{U}(n)$. Notice that the condition $Mu = u$ implies also $Mx(t) = x(t)$. In fact:

$$Mx(t) = M \int_0^t e^{-\tau \lambda J} u = \int_0^t e^{-\tau \lambda J} Mu = x(t),$$

where we used the fact that, being an isometry, $MJ = JM$. Thus, in this case, $\mathrm{ISO}_\gamma(\mathbb{H}_{2n+1}) = \mathrm{U}(n-1)$. When $u = 0$ the geodesic is the trivial one, and is stabilized by the whole $\mathrm{U}(n)$. □

Remark 13. Notice that in this case $u = 0$ if and only the geodesic is trivial $\gamma(t) \equiv 0$. When $u \neq 0$ two possibilities can occur: 1) $x \neq 0$, in which case $\mathrm{ISO}_\gamma(\mathbb{H}_{2n+1}) = \mathrm{ISO}_p(\mathbb{H}_{2n+1}) \simeq U(n-1)$; 2) $x = 0$ and the subgroup $\mathrm{ISO}_\gamma(\mathbb{H}_{2n+1}) \simeq U(n-1)$ is properly contained in $\mathrm{ISO}_p(\mathbb{H}_{2n+1}) \simeq U(n)$.

6.3. **Isometrically equivalent geodesics.** We introduce the following equivalence relation on geodesics with the same endpoint in a contact Carnot group G.

Definition 34 (Isometrically equivalent geodesics)**.**

$$\gamma_1 \sim \gamma_2 \iff \text{there exists } g \in \mathrm{ISO}(G) \text{ such that } \gamma_1 = g\gamma_2. \tag{32}$$

Let $p \in \mathbb{H}_{2n+1}$, and γ be a normal geodesic such that $\gamma(0) = 0$ and $\gamma(1) = p$. By acting with $\mathrm{ISO}_p(\mathbb{H}_{2n+1})$ we obtain families of isometrically equivalent by construction. Still, since $\mathrm{ISO}_\gamma(\mathbb{H}_{2n+1}) \subseteq \mathrm{ISO}_p(\mathbb{H}_{2n+1})$, we may obtain in this way non-distinct geodesics. To avoid duplicates, we have to take the quotient w.r.t. the subgroup $\mathrm{ISO}_\gamma(\mathbb{H}_{2n+1})$.

Let X_γ be the set of geodesics isometrically equivalent to a given one γ. This is a homogeneous space w.r.t. the action of $\mathrm{ISO}_p(\mathbb{H}_{2n+1})$. From Lemma 32 and 33 we obtain the structure of X_γ.

Proposition 35. *Let γ be a geodesic such that $\gamma(0) = 0$ and $\gamma(1) = p$, with initial covector $(u, \lambda) \in T_0^*\mathbb{H}_{2n+1}$. Then:*

$$X_\gamma = \mathrm{ISO}_p(\mathbb{H}_{2n+1}) / \mathrm{ISO}_\gamma(\mathbb{H}_{2n+1}) \simeq \begin{cases} S^{2n-1} & u \neq 0, \lambda = 2m\pi,\ m \in \mathbb{Z} \setminus \{0\}, \\ 1 & \textit{otherwise}. \end{cases}$$

Proof. Observe that if $u = 0$, then $\gamma(t) = 0$ is the trivial geodesic. In this case X_0 is just a point (the trivial geodesic). Then we may assume $u \neq 0$ in the following. Let $p = (0, z)$. Then an explicit computation leads to

$$0 = \int_0^1 e^{-\tau\lambda J} u \iff \lambda = 2m\pi, \quad m \in \mathbb{Z} \setminus \{0\}.$$

Then, when $\lambda = 2m\pi$ (and $u \neq 0$), according to Lemma 32 and 33 we have:

$$\mathrm{ISO}_p(\mathbb{H}_{2n+1}) / \mathrm{ISO}_\gamma(\mathbb{H}_{2n+1}) = \mathrm{U}(n)/\mathrm{U}(n-1) \simeq S^{2n-1}.$$

If $\lambda \neq 2m\pi$, then $p = (x, z)$ with $x \neq 0$. According to Lemma 32 and 33 (see also Remark 13) we have $\mathrm{ISO}_p(\mathbb{H}_{2n+1}) = \mathrm{ISO}_\gamma(\mathbb{H}_{2n+1}) = \mathrm{U}(n-1)$. Thus their quotient is the trivial Lie group. □

Remark 14. In fact, in terms of the endpoint, the only possibility for having a family of isometrically equivalent geodesics ending at p is that $x = 0$ zero. In fact, $\lambda = 2m\pi$ and $u \neq 0$ if and only if $p = (0, z)$ with $z \neq 0$. This means that for *non-vertical* points p, all the geodesics connecting p with the origin are not isometrically equivalent, while if $p = (0, z)$ is *vertical*, for any geodesic γ connecting p with the origin, we have a family of distinct geodesics (all with the same energy) diffeomorphic to S^{2n-1}.

7. Isometries of contact Carnot groups

We now pass to the study of the isometry group of a contact Carnot group G. By "isometry" we still mean distance preserving transformations that fix the origin.

Lemma 36. *The isometry group of the contact Carnot group G with parameters $(k, \vec{n}, \vec{\alpha})$ is:*

$$\mathrm{ISO}(G) = \{(M_1, \dots, M_k, \theta) \mid \theta \in \{-1, +1\},\ M_i M_i^* = \mathbb{1}_{2n_i},\ M_i J_{n_i} M_i^* = \theta J_{n_i}\},$$

with the action of $\mathrm{ISO}(G)$ *on G given by:*

$$(M_1, \dots, M_k, \theta) \cdot (x_1, \dots, x_k, z) = (M_1 x_1, \dots, M_k x_k, \theta z).$$

Moreover this group is isomorphic to:

$$\mathrm{ISO}(G) \simeq \mathrm{O}(2n_1) \cap \mathrm{Sp}(2n_1) \times \ldots \times \mathrm{O}(2n_k) \cap \mathrm{Sp}(2n_k) \rtimes \mathbb{Z}_2 \simeq \mathrm{U}(n_1) \times \ldots \times \mathrm{U}(n_k) \rtimes \mathbb{Z}_2.$$

Proof. The proof is analogous to the proof of Lemma 31, after splitting the equations in the real eigenspaces associated with the eigenvalues of A. □

Remark 15. Once again, we restrict our attention to the connected component $\mathrm{ISO}(G)_+$ containing the identity. We identify:

$$\mathrm{ISO}(G)_+ = \mathrm{U}(n_1) \times \ldots \times \mathrm{U}(n_k).$$

With this identification, the action $\rho : \mathrm{ISO}(G)_+ \times G \to G$ is

$$\rho(A_1 + iB_1, \ldots, A_k + iB_k, (x_1, \ldots, x_k, z)) = (M_1x_1, \ldots, M_kx_k, z),$$

where $A_j + iB_j \in \mathrm{U}(n_j)$ for all $j = 1, \ldots, k$ and $M_j := \begin{pmatrix} A_j & B_j \\ -B_j & A_j \end{pmatrix}$.

7.1. **Stabilizers of points.** Which is the subgroup $\mathrm{ISO}_p(G) \subseteq \mathrm{ISO}(G)_+$ that fixes $p \in G$?

Lemma 37. *Let $p = (x_1, \ldots, x_k, z) \in G$. Then:*

$$\mathrm{ISO}_p(G) = \begin{cases} \mathrm{U}(n_1) & x_1 = 0 \\ \mathrm{U}(n_1 - 1) & x_1 \neq 0 \end{cases} \times \ldots \times \begin{cases} \mathrm{U}(n_k) & x_k = 0 \\ \mathrm{U}(n_k - 1) & x_k \neq 0 \end{cases}.$$

Proof. Following Remark 15, an isometry $(A_1 + iB_1, \ldots, A_k + iB_k) \in \mathrm{ISO}(G)_+$ fixes $p = (x_1, \ldots, x_k, z)$ if and only if $(A_j + iB_j)x_j = x_j$ for all $j = 1, \ldots, k$. This means that $A_j + iB_j \in \mathrm{ISO}(\mathbb{H}_{2n_j+1})_+$ fixes the point $p_j := (x_j, z) \in \mathbb{H}_{2n_j+1}$, for all $j = 1, \ldots, k$. Then :

$$\mathrm{ISO}_p(G) = \mathrm{ISO}_{p_1}(\mathbb{H}_{2n_1+1}) \times \ldots \times \mathrm{ISO}_{p_k}(\mathbb{H}_{2n_k+1}),$$

and the result follows from Lemma 32. □

7.2. **Stabilizers of geodesics.** Let $(u, \lambda) \in T_0^*G$. Let γ the associated geodesic, such that $\gamma(0) = 0$ and $p = \gamma(1)$. What is the stabilizer subgroup of the geodesic $\mathrm{ISO}_\gamma(G) \subseteq \mathrm{ISO}_p(G)$? As usual, we write $u = (u_1, \ldots, u_k)$, with $u_i \in \mathbb{R}^{2n_i}$. Accordingly $\gamma(t) = (x_1(t), \ldots, x_k(t), z(t))$, with $x_i(t) \in \mathbb{R}^{2n_i}$. In particular:

$$x_i(t) = \int_0^t e^{-\tau\lambda\alpha_i J} u_i d\tau, \qquad z(t) = -\frac{1}{2}\sum_{i=1}^k \int_0^t \langle e^{-\tau\lambda\alpha_i J} u_i, \alpha_i J \int_0^\tau e^{-s\lambda\alpha_i J} u_i ds\rangle d\tau,$$

where we suppressed the explicit mention of the dimension of the matrices J_{n_i}. Notice that $u = 0$ if and only if the geodesic is the trivial one $\gamma(t) \equiv 0$.

Lemma 38. *Let $(u_1, \ldots, u_k, \lambda) \in T_0^*G$ the initial covector of the geodesic γ. Then:*

$$\mathrm{ISO}_\gamma(G) = \begin{cases} \mathrm{U}(n) & u_1 = 0 \\ \mathrm{U}(n-1) & u_1 \neq 0 \end{cases} \times \ldots \times \begin{cases} \mathrm{U}(n) & u_k = 0 \\ \mathrm{U}(n-1) & u_k \neq 0 \end{cases}.$$

Proof. Let $(A_1 + iB_1, \ldots, A_k + iB_k) \in \mathrm{ISO}(G)_+$. Then, according to Remark 15, this isometry fixes the geodesic $(x_1(t), \ldots, x_k(t), z(t))$ if and only if

$$M_jx_j(t) = x_j(t), \qquad M_j = \begin{pmatrix} A_j & B_j \\ -B_j & A_j \end{pmatrix}, \qquad \forall j = 1, \ldots, k.$$

This implies that $A_j + iB_j \in \mathrm{ISO}(\mathbb{H}_{2n_j+1})_+$ fixes the geodesic γ_j of $\mathbb{H}_{2n_j+1}$ associated with the initial covector $(u_j, \alpha_j\lambda)$. Then:

$$\mathrm{ISO}_\gamma(G) = \mathrm{ISO}_{\gamma_1}(\mathbb{H}_{2n_1+1}) \times \ldots \times \mathrm{ISO}_{\gamma_k}(\mathbb{H}_{2n_k+1}),$$

and the result follows from Lemma 33. □

Remark 16. In the proof of Lemma 38, starting from a geodesic γ in G with initial covector $(u_1,\dots,u_k,\lambda)$, we built "auxiliary" geodesics γ_i in $\mathbb{H}_{2n_i+1}$ associated with initial covector $(u_i,\alpha_i\lambda)$. Notice that the points $p_i := (x_i,z)$ are *not* the final points of the geodesics γ_i. In fact, if we write $\gamma(t)=(x_1(t),\dots,x_k(t),z(t))$, we have that $\gamma_i(t)=(x_i(t),z_i(t))$, with

$$z(t) = -\frac{1}{2}\sum_{i=1}^k \int_0^t \langle e^{-\tau\lambda\alpha_i J}u_i, \alpha_i J\int_0^\tau e^{-s\lambda\alpha_i J}u_i ds\rangle d\tau = \sum_{i=1}^k \alpha_i z_i(t).$$

7.3. **Isometrically equivalent geodesics.** Let γ be a geodesic connecting the origin with a point $p\in G$. Let $(u_1,\dots,u_k,\lambda)$ be the initial covector of the geodesic, and let $p=(x_1,\dots,x_k,z)$ its endpoint. Let X_γ be the set of geodesic isometrically equivalent to the given one. This is an homogeneous space w.r.t. the action of $\mathrm{ISO}_p(G)$.

Proposition 39. *Let G a contact Carnot group with parameters $(k,\vec{n},\vec{\alpha})$. Let γ be a geodesic in G with initial covector $(u_1,\dots,u_k,\lambda)$, such that $\gamma(0)=0$ and $\gamma(1)=p$. Then:*

$$X_\gamma = \mathrm{ISO}_p(G)/\,\mathrm{ISO}_\gamma(G) \simeq X_{\gamma_1}\times\dots\times X_{\gamma_k}, \tag{33}$$

where:

$$X_{\gamma_i} := \begin{cases} S^{2n_i-1} & u_i\neq 0, \quad \alpha_i\lambda = 2m_i\pi, \\ 1 & \textit{otherwise}, \end{cases} \qquad m_i\in\mathbb{Z}\setminus\{0\}.$$

Proof. By the proofs of Lemma 38 and 37 we have

$$\mathrm{ISO}_p(G) = \mathrm{ISO}_{p_1}(\mathbb{H}_{2n_1+1})\times\dots\times\mathrm{ISO}_{p_k}(\mathbb{H}_{2n_k+1}),$$

$$\mathrm{ISO}_\gamma(G) = \mathrm{ISO}_{\gamma_1}(\mathbb{H}_{2n_1+1})\times\dots\times\mathrm{ISO}_{\gamma_k}(\mathbb{H}_{2n_k+1}),$$

where $p_i=(x_i,z)\in\mathbb{H}_{2n_i+1}$ and γ_i is the normal geodesic in $\mathbb{H}_{2n_i+1}$ with initial covector $(u_i,\alpha_i\lambda)\in T_0^*\mathbb{H}_{2n_i+1}$, for all $i=1,\dots,k$. Since each factor $\mathrm{ISO}_{\gamma_i}(\mathbb{H}_{2n_i+1})$ is a subgroup of the corresponding $\mathrm{ISO}_{p_i}(\mathbb{H}_{2n_i+1})$, the quotient of the direct product of Lie groups factors in the direct product of the quotients:

$$\mathop{\times}_{i=1}^k \mathrm{ISO}_{p_i}(\mathbb{H}_{2n_i+1})/\mathop{\times}_{i=1}^k \mathrm{ISO}_{\gamma_i}(\mathbb{H}_{2n_i+1}) = \mathop{\times}_{i=1}^k \mathrm{ISO}_{p_i}(\mathbb{H}_{2n_i+1})/\,\mathrm{ISO}_{\gamma_i}(\mathbb{H}_{2n_i+1}).$$

Then:

$$X_\gamma = \mathop{\times}_{i=1}^k X_{\gamma_i}, \qquad X_{\gamma_i} = \mathrm{ISO}_{p_i}(\mathbb{H}_{2n_i+1})/\,\mathrm{ISO}_{\gamma_i}(\mathbb{H}_{2n_i+1}).$$

Recall that the geodesic γ_i of $\mathbb{H}_{2n_i+1}$ is associated with initial covector $(u_i,\alpha_i\lambda)$ by construction. Thus for each factor X_{γ_i} we proceed as in the proof of Proposition 35 and we obtain the result. □

Example 7. Proposition 39 implies that the for generic geodesic (i.e. with generic initial covector), the manifold X_γ of distinct isometrically equivalent geodesics is trivial.

Example 8. When $k=1$, G is isometric to $\mathbb{H}_{2n+1}$. Then Proposition 39 recovers Proposition 35.

Example 9. Consider the generic Carnot group G, associated with the generic choice of $A\in\mathfrak{so}(2n)$. In this case $n=k$, $n_1=\dots=n_k=1$ and all the α_i are not commensurable. Thus the only geodesics γ admitting a non-trivial manifold X_γ of distinct isometrically equivalent geodesics are those with initial covector (u,λ), such that $\lambda=2m\pi/\alpha_i$ for a unique $i\in\{1,\dots,n\}$ and $m\in\mathbb{Z}\setminus\{0\}$. In this case:

$$X_\gamma \simeq S^1.$$

In fact $\alpha_j\lambda \neq 2m_j\pi$ for all $j \neq i$ otherwise some α_j would be commensurable with α_i. Then there is only one factor in Eq. (33). Notice that these geodesics have endpoint (x, z), with $z \neq 0$, $x_i = 0$.

8. Families of isometrically non-equivalent geodesics

We ended the previous section discussing families X_γ of *isometrically equivalent* geodesics connecting two points. These families arose as homogeneous space w.r.t. the stabilizer $\mathrm{ISO}_p(G)$ of the final point $p = \gamma(1)$ of a fixed geodesic γ. In this section we adopt a different point of view, and we investigate how many *isometrically non-equivalent* geodesics join two points in G.

It may well be that some of the families of Theorem 23 contain geodesics that are isometrically equivalent, as defined in (32). This is the case in the Heisenberg groups $\mathbb{H}_{2n+1}$, where all the families are S^1 of equivalent geodesics. Is this the correct picture for any contact Carnot group? In other words, are the spheres appearing in $\Gamma_\infty(p)$ families of isometrically equivalent geodesics? In general the answer is no, and the picture is more complicated as shown in the next theorem.

Theorem 40. *Let G be a contact Carnot group. The set $\bar{\Gamma}_\infty(p)$ of equivalence classes of isometrically equivalent geodesics ending at $p \neq p_0$ is homeomorphic to the disjoint union of closed spheres quadrants:*

$$\bar{\Gamma}_\infty(p) \simeq \bigcup_{\lambda\in\Lambda_p} S^{\ell(\lambda)-1}_{\geq 0} \qquad \ell(\lambda) := \#L(\lambda),$$

where $S^n_{\geq 0} = S^n \cap \mathbb{R}^{n+1}_{\geq 0}$ is the intersection of the n-sphere with the positive quadrant in $\mathbb{R}^{n+1}$ and Λ_p is defined in (14).

Remark 17. Thus a family $X \subset \Gamma_\infty(p)$ is made of equivalent geodesics if under the above equivalence relation it corresponds to just a point in $\bar{\Gamma}_\infty(p)$.

When all the $\alpha_1, \dots, \alpha_k$ are pair-wise non-commensurable (e.g. the Heisenberg groups), then $\#L(\lambda) = 1$ for all $\lambda \in \Lambda_p \subseteq \Lambda$ and $N(\lambda) = 1$. Thus all the "continuous" families in $\Gamma_\infty(p)$ are topologically S^1 of isometrically equivalent geodesics. Nevertheless, for *resonant structures* (i.e. when some of the α_i are commensurable) there exist continuous families of non-isometrically equivalent geodesics.

Proof. Fix $\bar{\lambda} \in \Lambda_p$. Without loss of generality, we can assume that $L(\bar{\lambda}) = \{1, \dots, \ell\}$ for $\ell = \#L(\bar{\lambda})$. Notice that this implies $x_1 = \dots = x_\ell = 0$ by Lemma 20. From Lemma 37:

$$\mathrm{ISO}_p(G) = \mathrm{U}(n_1) \times \ldots \times \mathrm{U}(n_\ell) \times \mathrm{U}(n_{\ell+1} - 1) \times \ldots \times \mathrm{U}(n_k - 1),$$

and the action $\rho : \mathrm{ISO}_p(G) \times G \to G$ is:

$$\rho(A_1 + iB_1, \dots, A_k + iB_k, (x_1, \dots, x_k, z)) = (M_1x_1, \dots M_kx_k, z), \qquad M_i = \begin{pmatrix} A_j & B_j \\ -B_j & A_j \end{pmatrix}.$$

In particular $\mathrm{ISO}_p(G)$ is the subgroup that fixes all the components $x_{\ell+1}, \dots, x_k$ (with no other restriction on the other components). It is easy to check that the action on the initial covector $(u_1, \dots, u_k, \lambda)$ is exactly the same. In particular, $\mathrm{ISO}_p(G)$ is the subgroup that fixes all the components $u_{\ell+1}, \dots, u_k$ with no other restriction on the other components.

According to Proposition 22 and the definition of $\Gamma_\infty(p)$:

$$\Gamma_\infty(p) \cap \{\lambda = \bar{\lambda}\} \simeq \left\{ (u_1, \dots, u_{2n}) \,\middle|\, \sum_{j\in I_0} \|u_j\|^2 = c(\bar{\lambda}) \right\},$$

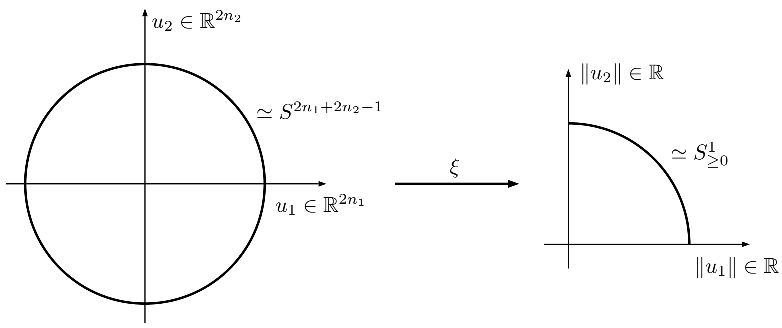

FIGURE 8. An example of the quotient under the action of $\mathrm{ISO}_p(G)$, for the case $L(\bar\lambda) = \{1, 2\}$.

where $c(\bar\lambda) > 0$. On the other hand, all the u_j with $j \notin L(\bar\lambda)$ are fixed by the equation $u_j = I(\alpha_j\lambda)^{-1}x_j$, thus we can rewrite, as in (17) the above formula as:

$$\Gamma_\infty(p) \cap \{\lambda = \bar\lambda\} \simeq \left\{ (u_1, \dots, u_\ell) \,\middle|\, \sum_{j \in L(\lambda)} \|u_j\|^2 = c(\bar\lambda) \right\} \simeq S^{2N(\bar\lambda)-1}.$$

The action of $\mathrm{ISO}_p(G)$ on $S^{2N(\bar\lambda)-1}$ is the action of $\mathrm{U}(n_1) \times \ldots \times \mathrm{U}(n_\ell)$, namely one copy of $\mathrm{U}(n_j)$ on each component u_j with $j \in L(\bar\lambda)$. Thus consider the map:

$$\xi : S^{2N(\bar\lambda)-1} \to S^{\ell-1}_{\geq 0} \qquad \xi(u_1, \dots, u_\ell) := (\|u_1\|, \dots, \|u_\ell\|).$$

This map indeed descends to a continuous map on the quotient (see Figure 8):

$$\tilde\xi : S^{2N(\bar\lambda)-1}/\mathrm{U}(n_1) \times \ldots \times \mathrm{U}(n_\ell) \to S^{\ell-1}_{\geq 0}.$$

It is indeed surjective and injective (recall that $u_j \in \mathbb{R}^{2n_j}$ and the action of $\mathrm{U}(n_j)$ on $\mathbb{R}^{2n_j}$ is exactly the classical action of $\mathrm{U}(n_j)$ on $\mathbb{C}^{n_j}$, which is transitive on spheres with the same radius). Being a continuous map from a compact space to a Hausdorff space, $\bar\xi$ is closed, then is open, thus it is an homemorphism. □

9. Contact sub-Riemannian manifolds

In this section we define the nilpotent approximation of a contact sub-Riemannian structure at a point p_0, and we review some of its basic properties. Then we relate the local geodesic count on the original manifold with the geodesic count on the nilpotent structure.

9.1. **The nilpotent approximation.**

Let M be a contact sub-Riemannian manifold and let $p_0 \in M$. All our considerations being local, up to restriction to a coordinate neighbourhood U of p_0, we assume that $M = \mathbb{R}^{2n+1}$ and the sub-Riemannian structure $(\mathcal{D}, \langle\cdot|\cdot\rangle)$ on M is defined by a set $f_1, \dots, f_{2n}$ of global orthonormal vector fields. Namely

$$\mathcal{D} = \mathrm{span}\{f_1, \dots, f_{2n}\}, \qquad \text{and} \qquad \langle f_i | f_j \rangle = \delta_{ij}.$$

The vector fields f are assumed to be bounded with all derivatives as well. This will certainly be true if they are the coordinate representation of local orthonormal fields on a neighbourhood U of p_0 of a larger sub-Riemannian manifold. Due to this assumption, it is easy to verify that *any* convergence below is uniform in all derivatives.

Definition 41. Coordinates $(x, z) \in \mathbb{R}^{2n} \times \mathbb{R}$ are *adapted* at p_0 if they are centred at p_0 and

$$\mathcal{D}_{p_0} = \operatorname{span}\left\{\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_{2n}}\right\}.$$

Example 10. *Darboux's coordinates* on a contact manifolds are local coordinates $(x, z) \in \mathbb{R}^{2n} \times \mathbb{R}$ such that the contact form has the following form:

$$\alpha = -dz + \frac{1}{2}\sum_{i,j=1}^{2n} J_{ij} x_i dx_j, \qquad \text{where} \qquad J = \begin{pmatrix} 0 & \mathbb{1}_n \\ -\mathbb{1}_n & 0 \end{pmatrix}.$$

In particular, in these coordinates $d\alpha = \sum_{i<j} J_{ij} dx_i \wedge dx_j$. The classical Darboux's theorem states that Darboux's coordinates always exist in a neighbourhood of any point p_0. Since $\mathcal{D}_{p_0} = \ker \alpha|_{p_0} = \operatorname{span}\{\partial_{x_1}, \dots, \partial_{x_{2n}}\}$, Darboux's coordinates are indeed adapted at p_0.

Remark 18. In the language of sub-Riemannian geometry, and strictly speaking of contact structures, adapted coordinate are called *privileged*. For more general sub-Riemannian manifolds some more properties are required (see [2, 6]).

In these coordinates we define "non-homogeneous dilations" $\delta_\varepsilon : M \to M$ by:

$$\delta_\varepsilon(x, z) = (\varepsilon x, \varepsilon^2 z), \qquad \varepsilon > 0,$$

and the following family of vector fields:

$$f_i^\varepsilon := \varepsilon \delta_{\frac{1}{\varepsilon}*} f_i = \hat{f}_i + \varepsilon W_i^\varepsilon, \qquad \varepsilon > 0.$$

The fields f_i^ε represent the "blowup" of the original structure in a neighbourhood of p_0 through the dilations δ_ε. The nilpotent approximation is the "principal part" of the original structure w.r.t. this non-homogeneous blowup.

Definition 42. For all $\varepsilon > 0$, the ε-*blowup* is the sub-Riemannian structure (M, f^ε) on M defined by declaring $f_1^\varepsilon, \dots, f_{2n}^\varepsilon$ a set of global orthonormal fields. Likewise, the *nilpotent approximation* (at p_0) is the sub-Riemannian structure $(M, \hat{f})$ on M defined by declaring $\hat{f}_1, \dots, \hat{f}_{2n}$ a set of global orthonormal fields.

We call $\mathcal{D}^\varepsilon$ (resp. $\hat{\mathcal{D}}$) the distribution of the ε-blowup (resp. of the nilpotent structure). The next proposition describes the structure of the nilpotent approximation of a contact sub-Riemannian manifold.

Proposition 43. *The nilpotent approximation $(M, \hat{f})$ at p_0 of a contact manifold is a contact Carnot group, with contact form given by*

$$\hat{\alpha} = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \delta_\varepsilon^* \alpha.$$

Let f_0 be a vector field transversal to $\mathcal{D}$ (in the original structure), and let

$$\hat{f}_0 := \lim_{\varepsilon \to 0} \varepsilon^2 \delta_{\frac{1}{\varepsilon}*} f_0.$$

Then the Lie algebra $\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2$ of the contact Carnot group $G = (M, \hat{f})$ is

$$\mathfrak{g}_1 = \operatorname{span}\{\hat{f}_1, \dots, \hat{f}_{2n}\}, \qquad \mathfrak{g}_2 = \operatorname{span}\{\hat{f}_0\}.$$

with structural constants given by $A \in \mathfrak{so}(2n)$ such that:

$$[\hat{f}_i, \hat{f}_j] = A_{ij} \hat{f}_0, \qquad A_{ij} = -\left.\frac{d\alpha(f_i, f_j)}{\alpha(f_0)}\right|_{p_0},$$

in terms of the given frame.

Proof. We first prove that the nilpotent structure is contact. For $\varepsilon > 0$ let $\alpha^\varepsilon := \frac{1}{\varepsilon^2}\delta_\varepsilon^*\alpha$. Indeed $\mathcal{D}^\varepsilon = \ker\alpha^\varepsilon$. Let (x, z) be the set of adapted coordinates that define the dilation δ_ε. Then

$$\alpha = \sum_{i=1}^{2n} \xi_i dx_i + w dz,$$

for some smooth functions $\xi_i, w : \mathbb{R}^{2n+1} \to \mathbb{R}$, bounded with all their derivatives. Since $\mathcal{D}_{p_0} = \ker \alpha|_{p_0} = \text{span}\{\partial_{x_1}, \dots, \partial_{x_{2n}}\}$ in adapted coordinates we have the following Taylor expansions

$$\xi_i(x, z) = \sum_{j=1}^{2n} a_{ij} x_j + bz + R_i(x, z), \qquad w(x, z) = w_0 + R_0(x, z). \tag{34}$$

where the remainder terms $R_i(x, z)$ (resp. $R_0(x, z)$) are actually bounded by polynomials of degree ≥ 2 (resp. ≥ 1) in (x, z). Moreover a_{ij} is non-degenerate since $d\alpha|_{\mathcal{D}}$ is non-degenerate and $w_0 \neq 0$. A straightforward calculation using the definition of δ_ε^* gives

$$\alpha^\varepsilon = \sum_{i=1}^{2n} \frac{1}{\varepsilon}\xi_i(\varepsilon x, \varepsilon^2 z) dx_i + w(\varepsilon x, \varepsilon^2 z) dz.$$

In particular, using Eq. (34), we notice that α^ε converges uniformly to $\hat{\alpha}$:

$$\hat{\alpha} = \lim_{\varepsilon \to 0} \alpha^\varepsilon = \sum_{i,j=1}^{2n} a_{ij} x_j dx_i + w_0 dz.$$

Indeed $\alpha \wedge (d\alpha)^n = w_0 \det(a) \neq 0$, which implies non-degeneracy of the contact form. Moreover, $\ker \hat{\alpha} = \text{span}\{\hat{f}_1, \dots, \hat{f}_{2n}\}$. In fact, for all $i = 1, \dots, 2n$, we have

$$\hat{\alpha}(\hat{f}_i) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2}\delta_\varepsilon^* \alpha(\varepsilon \delta_{\frac{1}{\varepsilon}*} f_i) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\alpha(f_i) = 0.$$

Now we show that the nilpotent approximation $(M, \hat{f})$ is a Carnot group. Consider the fields $f_1, \dots, f_{2n}$ defining the original structure, and any field f_0 transversal to $\mathcal{D}$. Then

$$[f_i, f_j] = \sum_{k=1}^{2n} c_{ij}^k f_k + c_{ij}^0 f_0, \qquad \forall i, j = 1, \dots, 2n, \tag{35}$$

for some family of smooth functions c_{ij}^0 and c_{ij}^k. Now consider the blowup of Eq. (35), namely we act on both sides with $\varepsilon^2\delta_{1/\varepsilon *}$, and we take the limit for $\varepsilon \to 0$. The first term on the r.h.s. vanishes in the limit (due to the factor ε^2), and we obtain

$$[\hat{f}_i, \hat{f}_j] = A_{ij}\hat{f}_0.$$

where $A_{ij} := c_{ij}^0(p_0)$ is a constant skew-symmetric matrix. Analogously, one can check that

$$[\hat{f}_i, \hat{f}_0] = [\hat{f}_0, \hat{f}_0] = 0, \qquad \forall i = 1, \dots, 2n.$$

Thus the fields $\hat{f}_1, \dots, \hat{f}_{2n}$ and $\hat{f}_0$ define a graduated, nilpotent Lie algebra $\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2$ with

$$\mathfrak{g}_1 := \text{span}\{\hat{f}_1, \dots, \hat{f}_{2n}\}, \qquad \mathfrak{g}_2 = \text{span}\{\hat{f}_0\}.$$

Since $M = \mathbb{R}^{2n+1}$ is simply connected and the Lie algebra of vector fields $\mathfrak{g}$ is nilpotent, there exists a unique group structure on M such that $\mathfrak{g}$ is its Lie algebra of left-invariant vector fields. The definition of the product law can be written explicitly in exponential coordinates on G

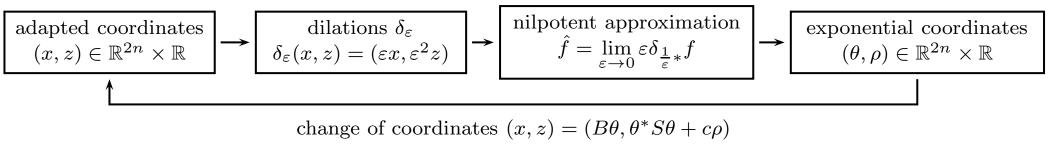

FIGURE 9. Adapted coordinates on M define the dilation map δ_ε that, in turn, defines the nilpotent approximation $(M, \hat{f})$. The latter is a contact Carnot group, endowed with exponential coordinates.

induced by the fields $\hat{f}_1, \ldots, \hat{f}_{2n}, \hat{f}_0$ through the Backer-Campbell-Hausdorff formula and is left to the reader. Thus $G := (M, \hat{f})$ has the structure of a contact Carnot group. Finally, see (5):

$$d\alpha(f_i, f_j) = f_i(\alpha(f_j)) - f_j(\alpha(f_i)) - \alpha([f_i, f_j]) = -c_{ij}^0 \alpha(f_0).$$

Using the relation $A_{ij} = c_{ij}^0(p_0)$, it is sufficient to evaluate the above formula at p_0 to obtain

$$A_{ij} = -\left.\frac{d\alpha(f_i, f_j)}{\alpha(f_0)}\right|_{p_0}.$$

Indeed A is not degenerate, as a consequence of the non-degeneracy assumption on $d\alpha|_{\mathcal{D}}$. □

9.2. **Adapted vs exponential coordinates.** Recall that, at the beginning of this section we put adapted coordinates $(x, z) \in \mathbb{R}^{2n} \times \mathbb{R}$ on M (e.g. Darboux's coordinates). This choice defined the family of non-homogeneous dilations δ_ε that, in turn defined the nilpotent approximation $(M, \hat{f})$ as the "limit" of the ε-blowup structures. Any choice of a global orthonormal frame f_i and f_0 transversal to $\mathcal{D}$ for the original structure induces a global orthonormal frame $\hat{f}_i$ and $\hat{f}_0$ (transversal to $\hat{\mathcal{D}}$) for the nilpotent approximation, where

$$\hat{f}_i = \lim_{\varepsilon\to 0} \varepsilon\delta_{\frac{1}{\varepsilon}*} f_i, \qquad \hat{f}_0 = \lim_{\varepsilon\to 0} \varepsilon^2\delta_{\frac{1}{\varepsilon}*} f_0.$$

Since $G = (M, \hat{f})$ is a contact Carnot group, the fields $\hat{f}_1, \ldots, \hat{f}_{2n}$ and $\hat{f}_0$ induce exponential coordinates $(\theta, \rho) \in \mathbb{R}^{2n} \times \mathbb{R}$. Namely a point has coordinates (θ, ρ) if and only if

$$(x, z) = \exp_G\left(\sum_{i=1}^{2n} \theta_i \hat{f}_i + \rho \hat{f}_0\right),$$

See Figure 9 for a diagram of logical implications. The next lemma clarifies the relation between adapted coordinates (x, z) and exponential coordinates (θ, ρ) on the same base space $M = \mathbb{R}^{2n+1}$.

Lemma 44. *Let $(x, z) \in \mathbb{R}^{2n} \times \mathbb{R}$ adapted coordinates for the contact structure $(\mathbb{R}^{2n+1}, f)$, and let $(\theta, \rho) \in \mathbb{R}^{2n} \times \mathbb{R}$ exponential coordinates for the Carnot structure $(\mathbb{R}^{2n+1}, \hat{f})$, induced by some choice f_i, f_0 (and consequently $\hat{f}_i, \hat{f}_0$). Then the two sets of coordinates are related by the following transformation*

$$\begin{cases} x = B\theta \\ z = \theta^* S\theta + c\rho \end{cases}$$

where

$$B \in \mathrm{GL}(2n), \qquad c \in \mathbb{R} \setminus \{0\}, \qquad S \in \mathrm{Mat}(2n)$$

Proof. For $i = 1, \dots, 2n$ we have, in adapted coordinates:

$$f_i = \sum_{j=1}^{2n} B_{ji}(x,z)\frac{\partial}{\partial x_j} + b_i(x,z)\frac{\partial}{\partial z}, \qquad f_0 = \sum_{j=1}^{2n} C_j(x,z)\frac{\partial}{\partial x_j} + c(x,z)\frac{\partial}{\partial z},$$

for some smooth functions $B_{ij}, b_i, C_j, c : \mathbb{R}^{2n+1} \to \mathbb{R}$ that satisfy:

$$b_i(0,0) = 0, \qquad \det B_{ij}(0,0) \neq 0, \qquad c(0,0) \neq 0. \tag{36}$$

By explicit computation we obtain

$$\hat{f}_i = \sum_{j=1}^{2n} B_{ji}(0,0)\frac{\partial}{\partial x_j} + \sum_{j=1}^{2n} \frac{\partial b_i}{\partial x_j}(0,0) x_j \frac{\partial}{\partial z}, \qquad \hat{f}_0 = c(0,0)\frac{\partial}{\partial z}.$$

By definition of exponential coordinates (see the proof of Lemma 17) we obtain that

$$x = B\theta, \qquad \text{and} \qquad z = \theta^* S\theta + c\rho,$$

where B is the matrix with components $B_{ij}(0,0)$, $c = c(0,0)$ and the matrix S has components

$$S_{ij} = \frac{1}{2}\sum_{\ell=1}^{2n} \frac{\partial b_i}{\partial x_\ell}(0,0) B_{\ell j}(0,0).$$

Indeed $c \neq 0$ and $B \in \mathrm{GL}(2n)$ by (36). □

The following proposition compares the geometry of the original structure with the ε-blowup.

Proposition 45. *The composition $\gamma \mapsto \gamma_\varepsilon = \delta_{\frac{1}{\varepsilon}}\gamma$ gives a homeomorphism between the set of admissible curves for (M, f) and admissible curves for (M, f^ε). If $\gamma(0) = 0, \gamma(1) = p$ and γ is a geodesic for (M, f), then γ_ε is a geodesic for (M, f^ε) with $\gamma_\varepsilon(0) = 0, \gamma_\varepsilon(1) = \delta_{\frac{1}{\varepsilon}}(p)$; the Energies of these curves are related by $J_\varepsilon(\gamma_\varepsilon) = \varepsilon^{-2}J(\gamma)$.*

Proof. Since $\gamma : [0,1] \to M$ is admissible, there exists $u_1, \dots, u_{2n} \in L^1(I)$ such that $\dot\gamma(t) = \sum u_i(t) f_i(\gamma(t))$ a.e. and:

$$\dot\gamma_\varepsilon(t) = \sum_{i=1}^{2n} u_i(t)\delta_{\frac{1}{\varepsilon}*} f_i(\gamma_\varepsilon(t)) = \sum_{i=1}^{2n} \frac{u_i(t)}{\varepsilon} f_i^\varepsilon(\gamma_\varepsilon(t)) = \sum_{i=1}^{2n} u_i^\varepsilon(t) f_i^\varepsilon(\gamma_\varepsilon(t)),$$

where we have set $u_i^\varepsilon = \varepsilon^{-1}u_i$. In particular γ_ε is admissible, and its energy is obtained dividing by ε^2 the energy of γ. Thus, a critical point for J, i.e. a geodesics for (M, f), is mapped to a critical point for $J_\varepsilon = \varepsilon^{-2}J$, and all critical points arise in this way, as $\gamma \mapsto \gamma_\varepsilon$ is invertible. □

We denote by E_ε and $\hat{E}$ respectively the sub-Riemannian exponential map at the origin for (M, f^ε) and for $(M, \hat{f})$:

$$E_\varepsilon, \hat{E} : T^*_{p_0}M \to M.$$

The next proposition states that the ε-blowups "converge" to the nilpotent approximation.

Proposition 46. *The family of maps $\{E_\varepsilon\}_{\varepsilon>0}$ converges uniformly with all derivatives on compact sets to $\hat{E}$.*

Proof. See [4, Proposition 5.2]. □

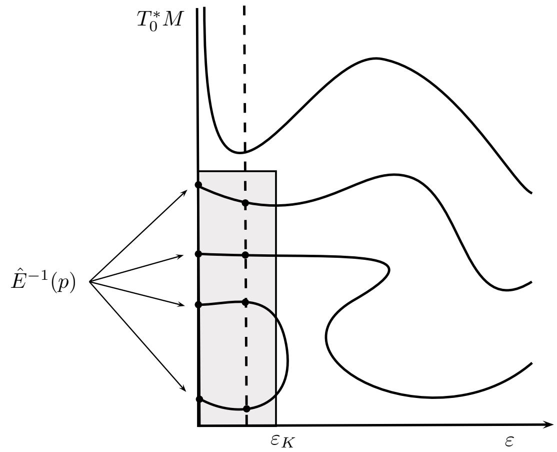

FIGURE 10. Picture of $\bigcup_{\varepsilon \in I} E_\varepsilon^{-1}(p) \subset I \times T_0^*M$. Even if ε_K is small, some geodesics can still "escape" out of K. The shaded region denotes $[0, \varepsilon_K] \times K$.

9.3. **Semicontinuity of the counting function.** We define now the *counting* functions $\nu_\varepsilon, \hat{\nu} : M \to (0, \infty]$ as:

$$\nu_\varepsilon(p) = \#E_\varepsilon^{-1}(p) \qquad \text{and} \qquad \hat{\nu}(p) = \#\hat{E}^{-1}(p).$$

In other words, $\nu_\varepsilon(p)$ counts the number of geodesics between 0 and p for the ε-blowup and $\hat{\nu}(p)$ for the limit Carnot group. Setting $\nu = \nu_1$ (the counting function for the original structure (M, f)), we notice that Proposition 45 implies indeed:

$$\nu_\varepsilon(p) = \nu(\delta_\varepsilon(p)).$$

In fact given a geodesic $\gamma : I \to M$ for (M, f) between 0 and $\delta_\varepsilon(p)$, then $\delta_{\frac{1}{\varepsilon}}\gamma$ is a geodesic for (M, f^ε) with final point $\delta_{\frac{1}{\varepsilon}}(\gamma(1)) = \delta_{\frac{1}{\varepsilon}}(\delta_\varepsilon(p)) = p$ (and vice-versa). The next theorem compares the asymptotic of $\nu(\delta_\varepsilon(p))$ with $\hat{\nu}(p)$.

Theorem 47 (Counting in the limit)**.** *Let M be a contact sub-Riemannian manifold. For the generic $p \in M$ sufficiently close to p_0:*

$$\hat{\nu}(p) \leq \varliminf_{\varepsilon \to 0} \nu(\delta_\varepsilon(p)).$$

where δ_ε is the non-homogeneous dilation defined in some set of adapted coordinates in a neighbourhood of p_0.

Proof. If p is a regular value of $\hat{E}$, then the fiber if $\hat{E}^{-1}(p)$ is discrete, hence $\hat{\nu}(p)$ is finite by Theorem 23. Consider an open bounded set $U \subset T_0^*M$, where bounded means that it is contained in a compact set K, such that:

$$\hat{E}^{-1}(p) \subset U \subset K.$$

We claim that there exists $\varepsilon_K > 0$ such that p is a regular value of $E_\varepsilon|_U$ for every $\varepsilon < \varepsilon_K$. If this was not true, then we can find a sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$ converging to zero and a sequence $\{\lambda_n\}_{n\in\mathbb{N}} \subset K$ such that $E_{\varepsilon_n}(\lambda_n) = p$ and $\mathrm{rank}(d_{\lambda_n}E_{\varepsilon_n}) < \dim(M)$. Then, by compactness of K, up to subsequences we can assume $\lambda_n \to \hat{\lambda}$ with $\hat{E}(\hat{\lambda}) = p$, by convergence of $E_{\varepsilon_n}|_K$ to $\hat{E}|_K$ as guaranteed by Proposition 46; moreover, by the same Proposition, $d_{\lambda_n}E_{\varepsilon_n} \to d_{\hat{\lambda}}\hat{E}$

and since the set of points where the rank of $d\hat{E}$ is not maximal is a closed set, we also have $\mathrm{rank}(d_{\hat{\lambda}}\hat{E}) < \dim(M)$, which contradicts the fact that p was a regular value of $\hat{E}$.

Consider now the function $\bar{E} : \overline{U} \to M$ (where $\overline{U} = U \times [0, \varepsilon_K])$ given by $(u, \varepsilon) \mapsto E_\varepsilon(u)$ (where we have set $E_0 = \hat{E}$); the uniform convergence of E_ε with all derivatives on compact sets implies $\bar{E}$ is smooth (in fact C^1 is enough for us). By the above observation $\bar{X} = \bar{E}^{-1}(p)$ is a smooth submanifold of $\overline{U}$ and its dimension is one. In fact:

$$(d_{(u,\varepsilon)}\bar{E})(\dot{u}, \dot{\varepsilon}) = (d_u E_\varepsilon)\dot{u} + \frac{\partial \bar{E}}{\partial \varepsilon}(u, \varepsilon)\dot{\varepsilon}, \qquad (\dot{u}, \dot{\varepsilon}) \in T_{(u,\varepsilon)}\overline{U}.$$

Since p is a regular value of E_ε for all $\varepsilon \in [0, \varepsilon_K]$, the image of $d_u E_\varepsilon$ is enough to generate $T_p M$.

On the other hand *zero* is a regular value for the the projection $\pi : \bar{X} \to [0, \varepsilon_K]$ on the second factor. The tangent space to $\bar{X}$ at $(u, 0)$ equals:

$$T_{(u,0)}\bar{X} = \left\{ (\dot{u}, \dot{\varepsilon}) \,\Big|\, (d_u\hat{E})\dot{u} + \frac{\partial \bar{E}}{\partial \varepsilon}(u, 0)\dot{\varepsilon} = 0 \right\}$$

and since $\hat{E}$ a submersion at p:

$$T_{(u,0)}\bar{X} \cap \ker d\pi \simeq T_u\hat{E}^{-1}(p) = \{0\}.$$

Thus $T_{(u,0)}\bar{X}$ must contain some vector $(\dot{u}, \dot{\varepsilon})$ with $\dot{\varepsilon} \neq 0$, i.e. zero is not critical for π. Then $\varepsilon' > 0$ small enough also is noncritical for π; in particular, by Ehresmann's theorem, $\pi|_{\pi^{-1}[0,\varepsilon']}$ is a fibration (U is contained in a compact set) and:

$$\forall \varepsilon < \varepsilon' : \quad E_\varepsilon|_U^{-1}(p) \simeq \hat{E}|_U^{-1}(p).$$

Since $\nu_\varepsilon(p) \geq \# E_\varepsilon|_U^{-1}(p)$ the conclusion for this case follows (see Figure 10). □

Remark 19. The definition of ε-blowup and nilpotent approximation remains unchanged for a general sub-Riemannian manifold provided that i) the point p_0 is *equiregular* (this is an assumption on the dimension of higher order distributions, true for the generic choice of the point p_0) and ii) we choose *privileged coordinates.* Since we are mainly interested in contact structures, we do not dwell into further details. Propositions 45-46 and 47 remain true with no change whatsoever in their proof in this more general setting.

Everything is now ready for the next theorem.

Theorem 48. *Let M be a contact manifold. Denote by (x, z) Darboux's coordinates on a neighbourhood U of q. For every point $q \in M$ there exist constants $C(q), R(q)$ such that, for the generic $p = (x, z) \in U$:*

$$\lim_{\varepsilon \to 0} \nu(\delta_\varepsilon(p)) \geq C(q)\frac{|z|}{\|x\|^2} + R(q).$$

Proof. Darboux's coordinates are adapted (see 10). We consider on U the original structure (U, f) and the nilpotent structure $(U, \hat{f})$ defined in terms of these adapted coordinates (see Figure 9). Classical Sard theorem implies that the generic $p \in U$ is a regular value for $\hat{E} : T_q^*U \to U$. Then, by Theorem 47,

$$\lim_{\varepsilon \to 0} \nu(\delta_\varepsilon(p)) \geq \hat{\nu}(p).$$

Now choose some orthogonal local frame $f_1, \ldots, f_{2n}$ and f_0 transversal to $\mathcal{D}$ for the original structure. This induces exponential coordinates (θ, ρ) on U (see Section 9.2). By Proposition 43, the nilpotent structure $(U, \hat{f})$ is a contact Carnot group such that

$$[\hat{f}_i, \hat{f}_j] = A_{ij}\hat{f}_0, \qquad A_{ij} = \left.\frac{d\alpha(f_j, f_i)}{\alpha(f_0)}\right|_q.$$

The generic point p has exponential coordinates (θ, ρ) with all $\theta_j \neq 0$. Theorem 26 gives

$$\hat{\nu}(p) \geq C_1 \frac{|\rho|}{\|\theta\|^2} + R_1$$

where $C_1 = C_1(q)$ and $R_1 = R_1(q)$ are computed in the proof of Theorem 26 in terms of the singular values of A. Indeed they depend on the point q at which we consider the nilpotentization. By Lemma 44, Darboux's (adapted) coordinates (x, z) and exponential coordinates (θ, ρ) are related by a quadratic change of coordinates, namely $(x, z) = (B\theta, \theta^* S\theta + c\rho)$ for some $B \in \mathrm{GL}(2n)$, $S \in \mathrm{Mat}(2n)$ and $c \in \mathbb{R} \setminus \{0\}$ and we obtain the result. □

We almost immediately obtain the following corollary.

Theorem 49. *Let M be a contact sub-Riemannian manifold and $q \in M$. Then there exists a sequence $\{q_m\}_{m\in\mathbb{N}}$ in M such that:*

$$\lim_{n\to\infty} q_m = q \qquad \textit{and} \qquad \lim_{m\to\infty} \nu(q_m) = \infty.$$

Proof. In Darboux's coordinates in a neighbourhood U of q, for every $m \in \mathbb{N}$ we pick a point $p_m = (x_m, z_m)$ such that: 1) p_m is a regular value of $\hat{E}$ and 2) $\frac{|z_m|}{\|x_m\|^2} \geq m$. The existence of such p_m is guaranteed simply by Sard's Lemma.

Consider now $\delta_\varepsilon(p_m)$. Using the fact that p_m is a regular value for $\hat{E}$, by Theorem 48 (its proof indeed implies that "generic" means "regular value of $\hat{E}$") we have:

$$\lim_{\varepsilon\to 0} \nu(\delta_\varepsilon(p_m)) \geq C(q)m + R(q).$$

For each such m let ε_m be such that:

$$\nu(\delta_{\varepsilon_m}(p_m)) \geq mC(q) + R(q) \qquad \forall \varepsilon \leq \varepsilon_m.$$

Notice that we can assume $\lim_{m\to\infty} \varepsilon_m = 0$. Setting $q_m = \delta_{\varepsilon_m}(p_m)$ we obtain the statement. □

References

[1] A. A. Agrachev. Exponential mappings for contact sub-Riemannian structures. *J. Dynam. Control Systems*, 2(3):321–358, 1996.

[2] A. A. Agrachev, D. Barilari, and U. Boscain. Introduction to Riemannian and sub-Riemannian geometry (Lecture Notes), http://people.sissa.it/agrachev/agrachev_files/notes.html. 2012.

[3] A. A. Agrachev, D. Barilari, and U. Boscain. On the hausdorff volume in sub-riemannian geometry. *Calculus of Variations and Partial Differential Equations*, 43(3-4):355–388, 2012.

[4] A. A. Agrachev, D. Barilari, and L. Rizzi. The curvature: a variational approach. *ArXiv e-prints*, June 2013.

[5] A. A. Agrachev, A. Gentile, and A. Lerario. Geodesics and admissible-path spaces in Carnot Groups. *to appear on Geometry & Topology*, Nov. 2013.

[6] A. Bellaïche. The tangent space in sub-Riemannian geometry. In *Sub-Riemannian geometry*, volume 144 of *Progr. Math.*, pages 1–78. Birkhäuser, Basel, 1996.

[7] F. Boarotto and A. Lerario. Homotopy properties of endpoint maps and a theorem of serre in subriemannian geometry. *ArXiv e-prints*, 2015.

[8] L. Capogna and E. Le Donne. Smoothness of subRiemannian isometries. *ArXiv e-prints*, May 2013.

[9] U. Hamenstädt. Some regularity theorems for Carnot-Caratheodory metrics. *J. Differ. Geom.*, 32(3):819–850, 1990.

[10] A. Hatcher. *Algebraic topology.* Cambridge University Press, Cambridge, 2002.

[11] F. Jean. *Control of Nonholonomic Systems: from Sub-Riemannian Geometry to Motion Planning.* to appear.

[12] E. Le Donne. *Lecture notes on sub-Riemannian geometry.* Lecture Notes, 2014.

[13] E. Le Donne, G. P. Leonardi, R. Monti, and D. Vittone. Extremal curves in nilpotent Lie groups. *Geom. Funct. Anal.*, 23(4):1371–1401, 2013.

[14] E. Le Donne and A. Ottazzi. Isometries of Carnot groups and subFinsler homogeneous manifolds. *ArXiv e-prints*, Oct. 2012.

[15] G. P. Leonardi and R. Monti. End-point equations and regularity of sub-Riemannian geodesics. *Geom. Funct. Anal.*, 18(2):552–582, 2008.

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

[17] A. Nabutovsky and R. Rotman. Length of geodesics and quantitative Morse theory on loop spaces. *Geom. Funct. Anal.*, 23(1):367–414, 2013.

[18] L. Rifford. *Sub-Riemannian Geometry and Optimal Transport*. Lecture Notes, 146 pp., 2012.

[19] J.-P. Serre. Homologie singulière des espaces fibrés. Applications. *Ann. of Math. (2)*, 54:425–505, 1951.

(A. Lerario) Institut Camille Jordan, Universitè Claude Bernard Lyon 1, 43 boulevard du 11 novembre 1918, 69622 Villeurbanne cedex France

E-mail address: lerario@math.univ-lyon1.fr

(L. Rizzi) CNRS, CMAP École Polytechnique and Équipe INRIA GECO Saclay Île-de-France, Paris, France.

E-mail address: luca.rizzi@cmap.polytechnique.fr