Sub-Riemannian geodesics on the Heisenberg 3D nil-manifold

A. Glutsyuk*†‡, Yu. Sachkov§¶

June 17, 2025

Abstract

We study the projection of the left-invariant sub-Riemannian structure on the 3D Heisenberg group G to the Heisenberg 3D nil-manifold M — the compact homogeneous space of G by the discrete Heisenberg group.

First we describe dynamical properties of the geodesic flow for M: periodic and dense orbits, a dynamical characterization of the normal Hamiltonian flow of Pontryagin maximum principle and its integrability properties. We show that it is Liouville integrable on a nonzero level hypersurface Σ of the Hamiltonian outside an appropriate smaller proper hypersurface in Σ and has no nontrivial analytic integrals on all of Σ. Then we obtain sharp twoside bounds of sub-Riemannian balls and distance in G, and on this basis we estimate the cut time for sub-Riemannian geodesics in M.

Keywords: Sub-Riemannian geometry, Heisenberg group, Heisenberg 3D nil-manifold, geodesics, dynamics, Hamiltonian flow, sub-Riemannian balls and distance, optimality

MSC: 53C17, 37C10, 49K15

*UMPA, ENS de Lyon (UMR 5669 du CNRS), France. E-mail: aglutsyu@ens-lyon.fr

†HSE University, Moscow, Russia

‡Higher School of Modern Mathematics MIPT, 1 Klimentovskiy per., Moscow, Russia

§Program Systems Institute (Russian Academy of Sciences, Pereslavl-Zalessky, Russia), RUDN University (Moscow, Russia)

¶Corresponding author, yusachkov@gmail.com

Contents

1 Introduction

1.1 The goal and structure of the paper

The left-invariant sub-Riemannian structure on the 3D Heisenberg group G is a paradigmatic model of sub-Riemannian geometry [16, 1]. In this paper we study the projection of this sub-Riemannian structure to a compact homogeneous space of the group G — to the Heisenberg 3D nil-manifold M. The sub-Riemannian structure on M is locally isometric to the structure on G, thus these structures have local objects (geodesics and conjugate points) related by the projection. Although, the global issues as dynamical properties of geodesics and cut time are naturally different. We aim to study these global questions in some detail.

The structure of this work is as follows. In Subsection 1.2 we state the main results of the paper. Section 2 discusses a well known projection of Euclidean structure from $\mathbb{R}^n$ to the torus $\mathbb{T}^n$, which suggests a motivation of the subsequent study. In Sec. 3 we recall the construction of the Heisenberg 3D nil-manifold M. In Sec. 4 we present basic definitions of sub-Riemannian geometry, define the sub-Riemannian structures of G and M, and describe their geodesics; in particular, we recall the parametrization of two distinct classes of sub-Riemannian geodesics in G — lines and spirals. In Sec. 5 we show that sub-Riemannian geodesics-lines in M may be either closed or dense, and describe explicitly geodesics falling into these classes. In Sec. 6 we describe dynamical properties of geodesics-spirals in M: we show that such a geodesic is either closed or dense in a certain 2D torus, and distinguish geodesics of these classes. In Sec. 7 we describe dynamical properties of the restriction of the Hamiltonian vector field for geodesics to a compact invariant surface (common level surface of the Hamiltonian and the Casimir). We show that the flow of this restriction is conjugated to a p-standard flow.

It is well-known that the geodesic flow on the standard torus $\mathbb{T}^n$ is Liouville integrable. In contrast to this fact, we show that the restriction of our sub-Riemannian geodesic flow to a nonzero level hypersurface Σ of the Hamiltonian is Liouville integrable outside an appropriate hypersurface S_0 and has no nontrivial analytic integrals on the whole hypersurface Σ. Next in Sec. 8 we obtain sharp interior and exterior ellipsoidal bounds of sub-Riemannian balls in G, which improve the classical ball-box bounds. In Sec. 9 we estimate the cut time on geodesics in M and the diameter of M on the basis of above interior bounds of sub-Riemannian balls in G. Finally, in Sec. 10 we prove two-sided bounds of cut time on geodesics in M on the basis of above exterior bounds of sub-Riemannian balls in G.

Concerning related research on sub-Riemannian optimal control problems on compact homogeneous spaces, we are aware only on the works where left-invariant sub-Riemannian structures on SO(3) and SU(2) were studied [22, 9, 5, 6, 7, 10]; see also the work [23] on abnormal extremals in compact Lie groups. Moreover, sub-Riemannian structures on the lens spaces were studied in the paper [9]. As far as we know, our paper is the first one in the literature where dynamical properties of sub-Riemannian geodesic flow on a compact homogeneous space are investigated.

1.2 The main results

The Heisenberg group is the space $G = \{(a, b, c)\} \cong \mathbb{R}^3$ with the group operation $(a_1, b_1, c_1) \cdot (a_2, b_2, c_2) = (a_1 + a_2, b_1 + b_2, c_1 + c_2 + a_1 b_2)$. Consider the discrete subgroup $H = \mathbb{Z}^3$ and its quotient (the space of right cosets) $M := {}_H\backslash^G = \{Hg \mid g \in G\}$. Denote by $\pi : G \to M$ the canonical projection $g \mapsto Hg$.

Consider the sub-Riemannian structure on G with a left-invariant field of orthonormal frames $X_1 = \frac{\partial}{\partial x} - \frac{y}{2}\frac{\partial}{\partial z}$, $X_2 = \frac{\partial}{\partial y} + \frac{x}{2}\frac{\partial}{\partial z}$, where $a = x$, $b = y$, $c = z + \frac{xy}{2}$. The geodesics of this structure, issued from the identity element $\mathrm{Id} \in G$, are either straight lines in the plane $\{z = 0\}$, issued from the origin, or spirals of variable slope

$$(x(t), y(t), z(t)) = \left(\frac{\sin(\theta + \delta t) - \sin\theta}{\delta}, \frac{\cos\theta - \cos(\theta + \delta t)}{\delta}, \frac{\delta t - \sin\delta t}{2\delta^2}\right), \qquad \theta \in \mathbb{R},\ \delta \neq 0. \quad (1.1)$$

We also consider the sub-Riemannian structure on M induced by the canonical projection π. Geodesics on M have the form $\pi(g(t))$, where $g(t)$ are geodesics on G.

For $\lambda \in T^*N$, $N = G$ or M, we set $h_j(\lambda) := \langle \lambda, X_j \rangle$, $j = 1, 2, 3$, $X_3 := \frac{\partial}{\partial z} = [X_1, X_2]$. Sub-Riemannian geodesics on a manifold N (G or M) are projections of trajectories of the Hamiltonian flow onto T^*N with Hamiltonian function $F = (h_1^2 + h_2^2)/2$, $j = 1, 2, 3$, $\lambda \in T^*N$, which is the normal Hamiltonian of the Pontryagin maximum principle [1]. It has first integrals F and h_3. Their joint level set $S_\delta := \{F = \frac{1}{2},\ h_3 = \delta\}$ is naturally isomorphic to the product $S^1 \times N$, $S^1 = \mathbb{R}_\theta / 2\pi\mathbb{Z}$, where $\theta = \operatorname{arccot}\frac{h_1}{h_2}$. The restriction to S_δ of the Hamiltonian system is

$$(\dot\theta, \dot a, \dot b, \dot c) = (\delta, \cos\theta, \sin\theta, a\sin\theta). \quad (1.2)$$

The projection $G \to \mathbb{R} \times \mathbb{R}$, $(a, b, c) \mapsto (a, b)$, induces a projection $p : M \to \mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$.

1.2.1 Dynamics of sub-Riemannian geodesic flow on M

We show that the projection $p(\Gamma)$ of each geodesic-spiral Γ in M is a contractible closed curve $\gamma \subset \mathbb{T}^2$; Γ is closed for $\delta^2 \in \pi\mathbb{Q}$ and dense on the surface $p^{-1}(\gamma)$ for $\delta^2 \notin \pi\mathbb{Q}$. See Theorem 6.1.

Theorem 1.1 1) *The projection onto M of each geodesic line in G in the plane $z = 0$, originating from* Id *and having an irrational (rational) tangent of the slope in coordinates (x, y), is dense in the manifold M (respectively, periodic).*

2) *A flow* (1.2) *on $S^1 \times M$ with $\delta \neq 0$ is conjugate to the standard δ-flow given by the field $(\dot\theta, \dot a, \dot b, \dot c) = \left(\delta, 0, 0, \frac{1}{2\delta}\right)$, by a diffeomorphism isotopic to the identity.*

3) *The Hamiltonian flow on $T^*M \setminus \{F = 0\}$ is analytically Liouville integrable on the complement of the hypersurface $\{h_3 = 0\}$, but on the entire $T^*M \setminus \{F = 0\}$ this is not true.*

1.2.2 Estimates of the cut time on M

Denote by d and d' the sub-Riemannian distances on G and M, and by $t_{\text{cut}}(q(\cdot))$ the cut time on the geodesic $q(\cdot)$, i.e., the time after which the geodesic arc ceases to minimize the length. Let $q_0 = \pi(\text{Id}) \in M$. Denote the ball $B'_t = \{q \in M \mid d'(q_0, q) \leq t\}$, $t \geq 0$. Let $\bar t = \inf\{t > 0 \mid B'_t = M\}$. The following equalities hold:

$$\begin{aligned} \bar t &= \sup\{d'(q_0, q_1) \mid q_1 \in M\} \\ &= \sup\{t_{\text{cut}}(q(\cdot)) \mid q(\cdot) \subset M \text{ is a geodesic s.t. } q(0) = q_0\}, \end{aligned}$$

see Lemma 9.1. Let $\bar g = (\bar a, \bar b, \bar c)$, $\widetilde g = (\tilde a, \tilde b, \tilde c) \in G$ be such that $\bar a = \bar b = \bar c = \frac{1}{2}$ and $\tilde a = 1$, $\tilde b = \tilde c = 0$. Then $d(\widetilde g, \bar g) \approx 0.91$, this follows from the explicit expression for the sub-Riemannian distance on the Heisenberg group, see [19], Subsec. 3.3.6. There holds the following two-side bound of the number $\bar t$, see Theorems 9.3 and 9.4.

Theorem 1.2 *The following inequalities hold:*

$$d(\widetilde g, \bar g) \leq \bar t \leq \frac{1}{2}\sqrt{\frac{1}{2}\left(1 + \sqrt{1 + 1024\pi^2}\right)} \approx 3.56.$$

In order to prove Theorem 1.2 we first obtain the following estimate using formulas (1.1):

$$\sqrt{\frac{\sqrt{r^4 + 48z^2} + r^2}{2}} \leq d(\text{Id}, g) \leq \sqrt{\frac{\sqrt{r^4 + 64\pi^2 z^2} + r^2}{2}}, \qquad (1.3)$$

where $g = (x, y, z) \in G$, $r = \sqrt{x^2 + y^2}$, see Cor. 8.8.

Let $B_t(g) = \{q \in G \mid d(g, q) \leq t\}$, $t \geq 0$, $g \in G$. Denote $\widehat t = \sup\{t > 0 \mid B_t(h_1) \cap B_t(h_2) = \emptyset \ \ \forall h_1 \neq h_2 \in H\}$. The number $\widehat t$ is related to the cut time on the geodesic $g(\cdot)$ in G with origin Id and its projection $g'(\cdot)$ in M, see Cor. 10.5:

(1) $t_{\text{cut}}(g'(\cdot)) \leq t_{\text{cut}}(g(\cdot))$;

(2) If $t_{\text{cut}}(g(\cdot)) \geq \widehat{t}$, then $\widehat{t} \leq t_{\text{cut}}(g'(\cdot)) \leq t_{\text{cut}}(g(\cdot))$;

(3) If $t_{\text{cut}}(g(\cdot)) < \widehat{t}$, then $t_{\text{cut}}(g'(\cdot)) = t_{\text{cut}}(g(\cdot))$.

The lower bound in (1.3) implies the following statement, see Th. 10.7.

Corollary 1.3 *The equality* $\widehat{t} = \frac{1}{2}$ *holds.*

2 Motivating example

Geodesics in the Euclidean space $\mathbb{R}^n$ have trivial dynamics (they tend to infinity) and optimality properties (they are length minimizers).

The situation changes when we pass from $\mathbb{R}^n$ to its compact homogeneous space — the torus $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$. Consider the Riemannian structure on $\mathbb{T}^n$ obtained via the projection $\pi : \mathbb{R}^n \to \mathbb{T}^n$. Then the geodesics on $\mathbb{T}^n$ are orbits of the linear flows

$$t \mapsto (x_1^0 + \omega_1 t, \dots, x_n^0 + \omega_n t) (\mathrm{mod}\, 1), \qquad (2.1)$$
$$\omega_1^2 + \dots + \omega_n^2 \neq 0, \qquad (x_1^0, \dots, x_n^0) \in \mathbb{T}^n, \qquad t \in \mathbb{R}.$$

Kronecker's theorem [14] (Propos. 1.5.1) states that such a geodesic is dense in $\mathbb{T}^n$ if and only if the frequencies $\omega_1, \dots, \omega_n$ are linearly independent over $\mathbb{Q}$. In all other cases a geodesic is dense in a nontrivial k-dimensional torus in $\mathbb{T}^n$, $1 \leq k < n$, see Propos. 2.1 below; in particular, a geodesic is periodic if $k = 1$. See Figs. 1, 2 for $n = 2$ and Figs. 3–5 for $n = 3$.

The following well-known statement generalizes Kronecker's theorem.

Proposition 2.1 *(see [15, section 5.1.5]). Consider a geodesic* Γ *in* $\mathbb{T}^n$ *of the form* (2.1) *and the corresponding vector space*

$$R = \mathrm{span}_{\mathbb{Q}} \left\{ r = (r_1, \dots, r_n) \in \mathbb{Q}^n \mid \sum_{i=1}^{n} r_i \omega_i = 0 \right\}, \qquad \rho = \dim R.$$

Then Γ *is dense in a smooth manifold* $S \subset \mathbb{T}^n$ *diffeomorphic to a torus* $\mathbb{T}^{n-\rho}$*.*

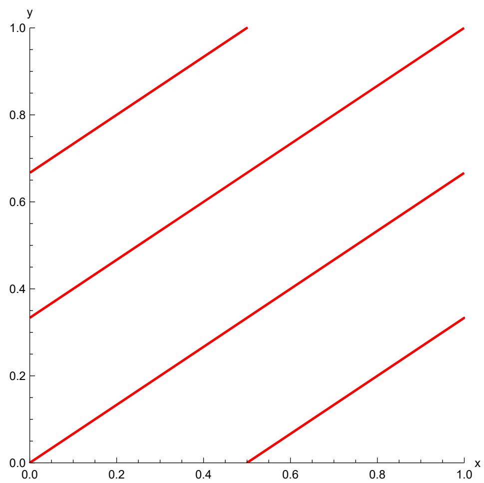

Figure 1: Geodesic (2.1) for $n = 2$, $(\omega_1, \omega_2) = (3, 2)$

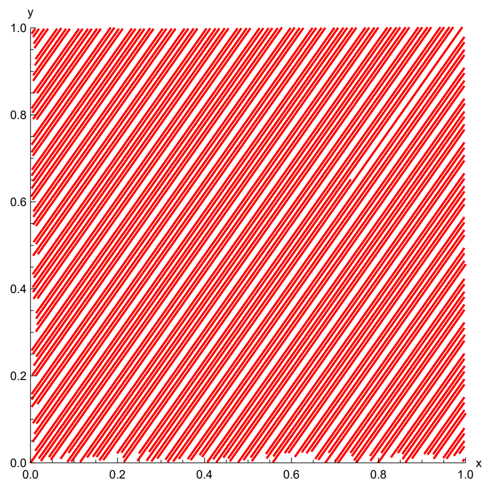

Figure 2: Geodesic (2.1) for $n = 2$, $(\omega_1, \omega_2) = (1, \sqrt{2})$

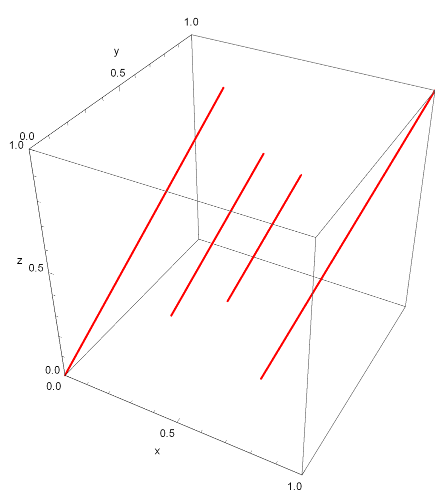

Figure 3: Geodesic (2.1) for $n = 3$, $(\omega_1, \omega_2, \omega_3) = (1, 2, 3)$

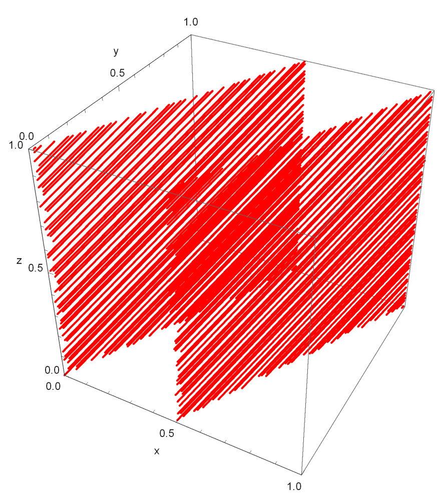

Figure 4: Geodesic (2.1) for $n = 3$, $(\omega_1, \omega_2, \omega_3) = (1, 2, \sqrt{2})$

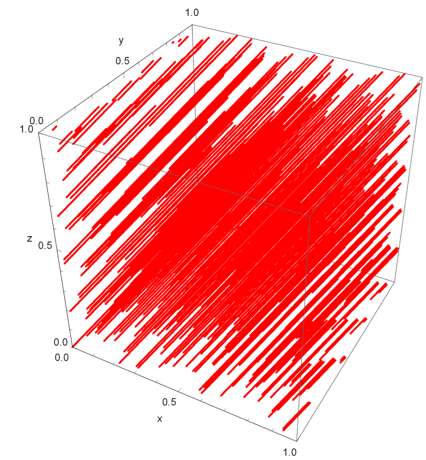

Figure 5: Geodesic (2.1) for $n = 3$, $(\omega_1, \omega_2, \omega_3) = (1, \sqrt{3}, \sqrt{2})$

Moreover, each geodesic (2.1) loses optimality at an instant

$$t_{\text{cut}} = \frac{1}{2 \max |\omega_i|} \in (0, +\infty). \tag{2.2}$$

More precisely, the *cut time* for a geodesic $x(t)$, $t > 0$, is defined as follows:

$$t_{\text{cut}}(x(\cdot)) = \sup\{t_1 > 0 \mid x(\cdot) \text{ is length minimizing on } [0, t_1]\}.$$

The reason for the loss of optimality is intersection with a symmetric geodesic starting from the same initial point in $\mathbb{T}^n$, see Fig. 6 for the case $n = 2$.

In this work we aim to generalize the above projection of Euclidean structure $\pi : \mathbb{R}^n \to \mathbb{T}^n$ to a projection of a left-invariant sub-Riemannian structure $\pi : G \to M$ from a Lie group G to its compact homogeneous space M. The simplest nontrivial case of such a projection is the case of the 3D Heisenberg group G and the 3D Heisenberg nil-manifold M, see Sec. 3.

Indeed, recall that a subgroup H of a Lie group G is called *uniform* (or *cocompact*) if the homogeneous space G/H is compact. The only connected and simply connected non-Abelian 2D Lie group $\mathbb{R} \rtimes \mathbb{R}_+$ does not contain uniform subgroups (see [21], Example 1.5). On the other hand, the 3D Heisenberg group G has a countable number of uniform subgroups, of which the subgroup (3.1) is the simplest one, see discussion in Remark 3.2 at the end of the next section.

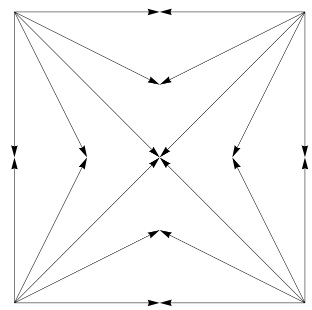

Figure 6: Optimal synthesis on $\mathbb{T}^2$

3 Heisenberg group and 3-dimensional nil-manifold

Recall that the Heisenberg group is

$$G = \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \mid (a, b, c) \in \mathbb{R}^3 \right\}.$$

Consider the following discrete subgroup and its quotient (the *right* cosets space):

$$H = \left\{ \begin{pmatrix} 1 & m & k \\ 0 & 1 & n \\ 0 & 0 & 1 \end{pmatrix} \mid (m, n, k) \in \mathbb{Z}^3 \right\}, \quad M := {H}\backslash{G} = \{Hg \mid g \in G\}. \tag{3.1}$$

The quotient M is a compact smooth manifold, which is called *Heisenberg 3-dimensional nil-manifold.*

Let $\pi : G \to M$ denote the canonical projection $g \mapsto Hg$. The functions

$$a' := \{a\}, \ b' := \{b\}, \ c' := \{c - [a]b\}$$

are coordinates on the homogeneous space M, and $\pi(a, b, c) = (a', b', c')$. Here and below $[x] = \max\{n \in \mathbb{Z} \mid n \leq x\}$ is the integer part of $x \in \mathbb{R}$, and $\{x\} = x - [x]$ is the fractional part of $x \in \mathbb{R}$.

Remark 3.1 The manifold M is not diffeomorphic to the 3-torus $\mathbb{T}^3 = \mathbb{R}^3_{a',b',c'}/\mathbb{Z}^3$, see [8, section 5]. Indeed, the first Betti number b_1 of the torus $\mathbb{T}^3$ is equal to 3. On the other hand, $b_1(M) = 2$. Indeed, the quotient projection $G \to M$ is a universal covering, since G is diffeomorphic to $\mathbb{R}^3_{a,b,c}$. Therefore, $\pi_1(M) = H$. The first homology of a path connected topological space is isomorphic to the quotient of the fundamental group by its commutant, by classical Poincaré Theorem [12, section 14.3, p.181]. Hence, $H_1(M, \mathbb{Z}) = H/[H, H]$. The commutant $[H, H]$ coincides with the subgroup of integer unipotent matrices that differ from the identity just by the upper-right corner element. Therefore, it is isomorphic to $\mathbb{Z}$. The map $H \mapsto (H_{12}, H_{23})$ is an isomorphism $H/[H, H] \to \mathbb{Z}^2$. Therefore, $b_1(M) = 2$.

The manifold M can be represented by a fundamental domain $D = \{(a, b, c) \mid 0 \le a, b, c < 1\}$ with identified facets $\{b = 0\} \leftrightarrow \{b = 1\}$, $\{c = 0\} \leftrightarrow \{c = 1\}$, while the facets $\{a = 0\}$ and $\{a = 1\}$ are identified by the rule $(0, b, c) \simeq (1, b, c + b)$, see Fig. 7.

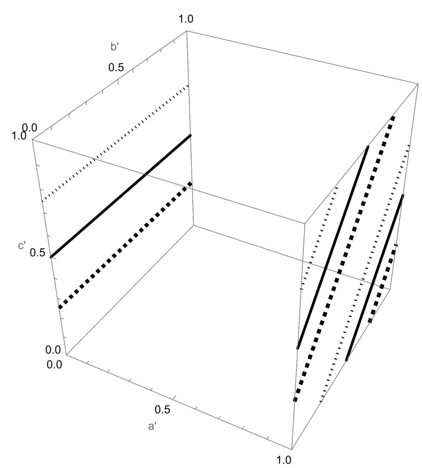

Figure 7: Heisenberg 3D nil-manifold

Remark 3.2 Any uniform subgroup of the Heisenberg group G is isomorphic to a subgroup

$$D(k) = \left\{ \begin{pmatrix} 1 & a & \frac{c}{k} \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \mid a, b, c \in \mathbb{Z} \right\}$$

for some $k \in \mathbb{N}$, see [21]; in particular, $H = D(1)$. Thus any compact 3D nil-manifold is a homogeneous space $G/D(k)$. Since G is simply connected, the fundamental group of such a space is $\pi_1(G/D(k)) = D(k)$. Hence

$$\pi_1(M) = \pi_1(G/D(1)) = D(1) \subset D(k) = \pi_1(G/G(k)), \qquad k \in \mathbb{N}.$$

Consequently, the Heisenberg nil-manifold M is the simplest compact 3D nil-manifold in the sense that it has the smallest fundamental group. This observation motivated us to study sub-Riemannian geometry on the Heisenberg nil-manifold M. This work may thus be seen as the first study of sub-Riemannian geometry of compact 3D nil-manifolds, starting from the simplest case of the Heisenberg nil-manifold M. We believe that our methods may be useful for the study of sub-Riemannian geometry on compact nil-manifolds of dimension 3 and greater than 3.

4 Sub-Riemannian structure on the Heisenberg group and its projection to the nil-manifold

4.1 Sub-Riemannian geometry

A sub-Riemannian structure [16, 1] on a smooth manifold M is a vector subspace distribution

$$\Delta = \{\Delta_q \subset T_qM \mid q \in M\} \subset TM, \qquad \dim \Delta_q \equiv \text{const},$$

endowed with an inner product

$$\langle \cdot, \cdot \rangle = \{\langle \cdot, \cdot \rangle_q \text{ inner product in } \Delta_q \mid q \in M\}.$$

A Lipschitzian curve $q : [0, t_1] \to M$ is called horizontal if $\dot{q}(t) \in \Delta_{q(t)}$ for almost all $t \in [0, t_1]$. The sub-Riemannian length of a horizontal curve $q(\cdot)$ is

$$l(q(\cdot)) = \int_0^{t_1} \langle \dot{q}(t), \dot{q}(t) \rangle^{1/2} \, dt.$$

The sub-Riemannian distance between points $q_0, q_1 \in M$ is

$$d(q_0, q_1) = \inf\{l(q(\cdot)) \mid q(\cdot) \text{ horiz. curve s.t. } q(0) = q_0, \ q(t_1) = q_1\}.$$

A horizontal curve is called a *sub-Riemannian length minimizer* if its sub-Riemannian length is equal to the sub-Riemannian distance between its endpoints. A *sub-Riemannian geodesic* is a horizontal curve whose sufficiently short arcs are length minimizers. Finally, a cut time along a sub-Riemannian geodesic $q(\cdot)$ is

$$t_{\text{cut}}(q(\cdot)) = \sup\{\tau > 0 \mid q(\cdot)|_{[0,\tau]} \text{ is a length minimizer}\}.$$

If a real-analytic distribution Δ is *completely nonholonomic* (*completely nonintegrable*), i.e., any points in M can be connected by a horizontal curve of Δ, then the sub-Riemannian distance d turns M into a metric space, and there are naturally defined a sub-Riemannian sphere of radius $R \geq 0$ centred at a point $q_0 \in M$:

$$S_R(q_0) = \{q \in M \mid d(q_0, q) = R\}$$

and the corresponding sub-Riemannian ball:

$$B_R(q_0) = \{q \in M \mid d(q_0, q) \leq R\}.$$

Let $X_1, \dots X_k$ be vector fields on M that form an orthonormal frame of a sub-Riemannian structure $(\Delta, \langle \cdot, \cdot \rangle)$:

$$\Delta_q = \operatorname{span}(X_1(q), \dots, X_k(q)), \quad \langle X_i(q), X_j(q) \rangle = \delta_{ij}, \qquad q \in M.$$

Then sub-Riemannian length minimizers that connect points $q_0, q_1 \in M$ are solutions to the optimal control problem

$$\dot{q} = \sum_{i=1}^{k} u_i(t) X_i(q), \qquad q \in M, \quad u = (u_1, \dots, u_k) \in \mathbb{R}^k,$$
$$q(0) = q_0, \qquad q(t_1) = q_1,$$
$$\int_0^{t_1} \left(\sum_{i=1}^{k} u_i^2(t) \right)^{1/2} dt \to \min .$$

4.2 Sub-Riemannian structures on G and M

The left-invariant sub-Riemannian problem on the Heisenberg group is stated as the following optimal control problem [1, 2, 19]:

$$\dot{g} = u_1(t) X_1(g) + u_2(t) X_2(g), \qquad g \in G, \quad u = (u_1, u_2) \in \mathbb{R}^2, \tag{4.1}$$
$$g(0) = g_0 = \mathrm{Id}, \qquad g(t_1) = g_1, \tag{4.2}$$
$$\int_0^{t_1} (u_1^2(t) + u_2^2(t))^{1/2} \, dt \to \min, \tag{4.3}$$
$$X_1 = \frac{\partial}{\partial x} - \frac{y}{2} \frac{\partial}{\partial z}, \quad X_2 = \frac{\partial}{\partial y} + \frac{x}{2} \frac{\partial}{\partial z}. \tag{4.4}$$

The fields X_1, X_2 are left-invariant vector fields on G.

Here and below we use coordinates (x, y, z) on the Heisenberg group G such that

$$a = x, \quad b = y, \quad c = z + \frac{xy}{2}. \tag{4.5}$$

The geodesics for this problem have the form:

$$x = t\cos\theta, \tag{4.6}$$
$$y = t\sin\theta, \tag{4.7}$$
$$z = 0 \tag{4.8}$$

for $\theta \in \mathbb{R}/(2\pi\mathbb{Z})$, and

$$x = (\sin(\theta + ht) - \sin\theta)/h, \tag{4.9}$$
$$y = (\cos\theta - \cos(\theta + ht))/h, \tag{4.10}$$
$$z = (ht - \sin ht)/(2h^2) \tag{4.11}$$

for $\theta \in \mathbb{R}/(2\pi\mathbb{Z})$ and $h \neq 0$.

Sub-Riemannian geodesics (4.6)–(4.8) are one-parametric subgroups in G, they are projected to the plane (x, y) into straight lines, thus we call them geodesics-lines in the sequel. Sub-Riemannian geodesics (4.9)–(4.11) are spirals of nonconstant slope in $\mathbb{R}^3_{x,y,z} \simeq G$, they are projected to the plane (x, y) into circles, and we call them geodesics-spirals in the sequel.

Sub-Riemannian problem (4.1)–(4.4) is left-invariant on the Heisenberg group G, thus its projection to the nil-manifold M is a well-defined sub-Riemannian problem on M:

$$\dot{g}' = u_1 X_1'(g') + u_2 X_2'(g'), \qquad g' \in M, \quad u = (u_1, u_2) \in \mathbb{R}^2, \tag{4.12}$$
$$g'(0) = g_0' = \pi(\mathrm{Id}), \qquad g'(t_1) = g_1', \tag{4.13}$$
$$\int_0^{t_1} (u_1^2 + u_2^2)^{1/2}\, dt \to \min, \tag{4.14}$$
$$X_i'(g') = \pi_*(X_i(g)), \qquad g' = \pi(g), \qquad i = 1, 2, \quad g \in G. \tag{4.15}$$

Geodesics of the projected problem (4.12)–(4.15) have the form $g'(t) = \pi(g(t))$, where $g(t)$ are geodesics of the initial problem (4.1)–(4.4).

Remark 4.1 The projection of sub-Riemannian structure from the Heisenberg group to the Heisenberg nil-manifold is a particular case of the following setting. Let G be a sub-Riemannian manifold, and let H be a discrete group that acts on G and preserves the sub-Riemannian structure. Let the action of H be free and proper. Then the quotient manifold $M = {}_H\backslash^G$ is endowed by a sub-Riemannian structure induced by the projection $\pi : G \to M$. We use this general setting in Section 10, Remark 10.6.

5 Projections of geodesics-lines to Heisenberg nil-manifold

For every $\theta \in \mathbb{R}/(2\pi\mathbb{Z})$ consider a geodesic-line (4.6)–(4.8) in G and its projection to M:

$$g(t) = (a(t), b(t), c(t)) = \left(t\cos\theta, t\sin\theta, \frac{t^2}{2}\sin\theta\cos\theta\right), \quad t \in \mathbb{R}.$$

$$g'(t) := \pi \circ g(t) = (a'(t), b'(t), c'(t)),$$

$$a'(t) = \{t\cos\theta\},\ b'(t) = \{t\sin\theta\},\ c'(t) = \left\{\frac{t^2}{2}\sin\theta\cos\theta - [t\cos\theta]t\sin\theta\right\}, \tag{5.1}$$

$$\Gamma := \{g'(t) \mid t \in \mathbb{R}\} \subset M.$$

Consider the projection $\rho : M \to \mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ and the image $\rho(\Gamma)$:

$$\rho : (a', b', c') \mapsto (a', b'), \quad \gamma := \rho(\Gamma) = \{(a'(t), b'(t)) \mid t \in \mathbb{R}\} \subset \mathbb{T}^2.$$

Remark 5.1 If $\theta = \frac{\pi n}{2}$, $n \in \mathbb{Z}$, then γ and Γ are both 1-periodic. If $\tan\theta \in \mathbb{Q} \setminus \{0\}$, then γ is periodic, and $\rho^{-1}(\gamma)$ is a two-dimensional torus.

Proposition 5.2 *If $\tan\theta = \frac{p}{q} \in \mathbb{Q} \setminus \{0\}$, then the curve γ is periodic with period $T = \frac{q}{\cos\theta}$; here $(p, q) = 1$. The curve Γ is periodic either with the same period, as γ, if some of the numbers p, q is even, or with twice bigger period otherwise.*

See Figs. 8, 9.

Proof One has

$$T\cos\theta = q \in \mathbb{Z}, \quad T\sin\theta = p \in \mathbb{Z}, \tag{5.2}$$

$$(a(t+T), b(t+T) = (a(t) + q, b(t) + p).$$

This implies T-periodicity of the curve γ. One has

$$c'(t+T) = \{c(t+T) - [a(t+T)]b(t+T)\}$$

$$= \left\{\frac{(t+T)^2}{2}\sin\theta\cos\theta - [(t+T)\cos\theta](t+T)\sin\theta\right\}$$

$$= \left\{\frac{t^2}{2}\sin\theta\cos\theta + tT\sin\theta\cos\theta + \frac{T^2}{2}\sin\theta\cos\theta - [(t+T)\cos\theta](t+T)\sin\theta\right\}.$$

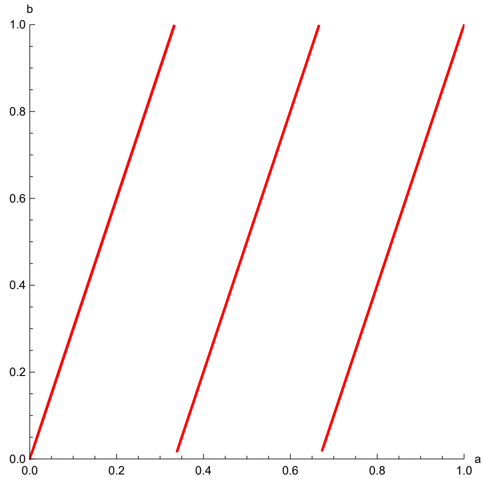

Figure 8: The curve γ for $\tan\theta = 3$

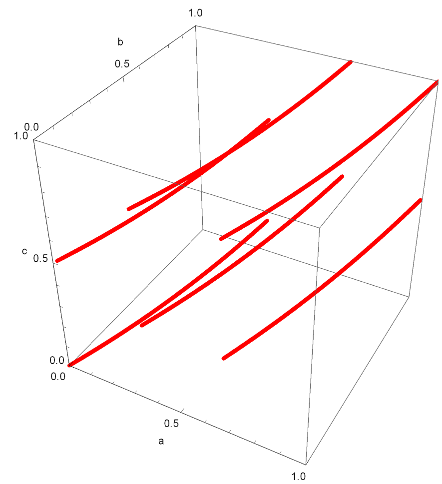

Figure 9: The curve Γ for $\tan\theta = 3$

Substituting (5.2) yields

$$c'(t+T) = \left\{ \frac{t^2}{2}\sin\theta\cos\theta + tq\sin\theta + \frac{pq}{2} - [t\cos\theta + q](t\sin\theta + p) \right\}$$

$$= \left\{ \frac{t^2}{2}\sin\theta\cos\theta + tq\sin\theta + \frac{pq}{2} - ([t\cos\theta] + q)(t\sin\theta + p) \right\}$$

$$= \left\{ \frac{t^2}{2}\sin\theta\cos\theta + tq\sin\theta + \frac{pq}{2} - [t\cos\theta]t\sin\theta - qt\sin\theta - p[t\cos\theta] - pq \right\}$$

$$= \left\{ \frac{t^2}{2}\sin\theta\cos\theta - [t\cos\theta]t\sin\theta - \frac{pq}{2} \right\}.$$

The latter expression is equal to either $c'(t)$, if pq is even, or $\{c'(t) + \frac{1}{2}\}$ otherwise. In the latter case replacing T by $2T$ and repeating the above argument with q, p replaced by $2q$, $2p$ yields $c'(t+2T) = c'(t)$. The proposition is proved. $\square$

Theorem 5.3 *The curve* Γ *is dense in* M *for every* $\theta \in \mathbb{R}$ *such that* $\tan\theta \notin \mathbb{Q} \cup \{\infty\}$. *In this case each its half* $\Gamma_\pm = \{g'(t) \mid t \in \mathbb{R}_\pm\}$ *is dense.*

See Figs. 10, 11.

As it is shown below, Theorem 5.3 is implied by the following theorem.

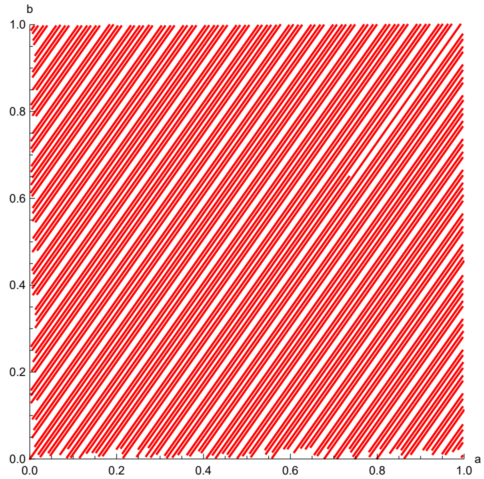

Figure 10: The curve γ for $\tan\theta = \sqrt{2}$

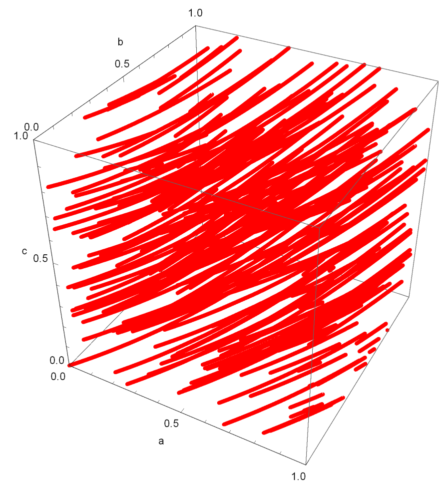

Figure 11: The curve Γ for $\tan\theta = \sqrt{2}$

Theorem 5.4 *The sequence* $\{(\{2rn\}, \{rn^2\}) \in \mathbb{T}^2 \mid n \in \mathbb{N}\}$ *is dense in* $\mathbb{T}^2$ *for every* $r \in \mathbb{R} \setminus \mathbb{Q}$*.*

Theorem 5.4 follows from a more general result due to H.Furstenberg, see [13, lemma 2.1], which yields unique ergodicity of the torus map (5.6). Below we present a proof of Theorem 5.4 for completeness of presentation.

Remark 5.5 It is known that for every real polynomial $P(n) = \alpha_0 n^m + \alpha_1 n^{m-1} + \cdots + \alpha_m$ with $\alpha_0 \notin \mathbb{Q}$ the values $P(n)$ are equidistributed (thus, dense) on the segment $[0,1]$ (Furstenberg's theorem, see [14, exercise 4.2.7]). This theorem also follows from the above-mentioned Furstenberg's result on unique ergodicity of map (5.6).

Proof of Theorem 5.3 modulo Theorem 5.4. We prove the statement of Theorem 5.3 for the half-curve Γ_+; the proof for Γ_- is analogous.

Let us do the above calculation with

$$T = \frac{q}{\cos\theta}, \quad q \in \mathbb{Z}_{\geq 0}:$$

$$T\cos\theta = q \in \mathbb{Z}_{\geq 0}, \quad T\sin\theta = q\tan\theta,$$

$$c'(t+T) = \{c(t+T) - [a(t+T)]b(t+T)\}$$

$$= \left\{\frac{(t+T)^2}{2}\sin\theta\cos\theta - [(t+T)\cos\theta](t+T)\sin\theta\right\}$$

$$= \left\{ \frac{t^2}{2} \sin\theta\cos\theta + tT\sin\theta\cos\theta + \frac{T^2}{2}\sin\theta\cos\theta - [(t+T)\cos\theta](t+T)\sin\theta \right\}$$

$$= \left\{ \frac{t^2}{2} \sin\theta\cos\theta + tq\sin\theta + \frac{q^2\tan\theta}{2} - ([t\cos\theta]+q)(t\sin\theta + q\tan\theta) \right\}$$

$$= c'_q(t) := \left\{ \frac{t^2}{2} \sin\theta\cos\theta - [t\cos\theta]t\sin\theta - [t\cos\theta]q\tan\theta - q^2\frac{\tan\theta}{2} \right\}. \tag{5.3}$$

Set

$$\widehat{\Gamma}_q := \{\Gamma(t + \frac{q}{\cos\theta}) \mid t \in [0, \frac{1}{\cos\theta})\} \quad \text{for every} \quad q \in \mathbb{Z}_{\geq 0}.$$

By definition,

$$\Gamma_+ \cup \{g'(0)\} = \cup_{q\in\mathbb{Z}_{\geq 0}} \widehat{\Gamma}_q. \tag{5.4}$$

Set

$$r = \frac{\tan\theta}{2}.$$

Each curve $\widehat{\Gamma}_q$ admits the coordinate representation

$$\widehat{\Gamma}_q(t) = (a'(t), \{b'(t) + 2qr\}, \{c'(t) - rq^2\}), \tag{5.5}$$

since for every $t \in [0, \frac{1}{\cos\theta})$ one has

$$c'_q(t) = \left\{ \frac{t^2}{2} \sin\theta\cos\theta - [t\cos\theta]t\sin\theta - q^2\frac{\tan\theta}{2} \right\} = \{c'(t) - rq^2\} :$$

$[t\cos\theta] = 0$, whenever $t \in [0, \frac{1}{\cos\theta})$. For every t the sequence $(\{b'(t) + 2qr\}, \{c'(t) - rq^2\})$ is dense in $[0,1)\times[0,1)$. To prove this, it suffices to show that the sequence $(\{2qr\}, \{-rq^2\})$ is dense in $[0,1)\times[0,1)$. Or equivalently, density of the projection to $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ of the sequence $(2qr, -rq^2)$. Indeed, the projection to $\mathbb{T}^2$ of the sequence $(2qr, rq^2)$ is dense, by Theorem 5.4. The sequence $(2qr, -rq^2)$ is obtained from the latter sequence with dense projection by the symmetry $(x, y) \mapsto (x, -y)$, which is the lifting to $\mathbb{R}^2$ of torus automorphism given by the same formula. Every torus automorphism sends any dense subset to a dense subset. The latter symmetry sends the projection of the sequence $(2qr, rq^2)$ to the projection of the sequence $(2qr, -rq^2)$. Therefore, the latter projection is dense. Hence, the sequence $(\{b'(t) + 2qr\}, \{c'(t) - rq^2\})$ is dense in $[0,1)\times[0,1)$ for every $t \in [0, \frac{1}{\cos\theta})$. This together with (5.5) and (5.4) implies density of the curve Γ_+ in M. □

For the proof of Theorem 5.4 let us introduce the torus map

$$T : \mathbb{T}^2 \to \mathbb{T}^2, \ T(x, y) = (x + \alpha, y + x + \beta). \tag{5.6}$$

Proposition 5.6 *Set $\alpha = 2r$, $\beta = r$. Then*

$$(\{2rn\}, \{rn^2\}) = T^n(0,0). \tag{5.7}$$

Proof Induction in n.

Induction base: for $n = 1$ one has $T(0,0) = (2r, r)(\operatorname{mod} \mathbb{Z}^2)$.

Induction step. Let $T^n(0,0) = (2rn, rn^2)(\operatorname{mod} \mathbb{Z}^2)$. Then modulo $\mathbb{Z}^2$,

$$T^{n+1}(0,0) = (2r(n+1), rn^2 + 2rn + r) = (2r(n+1), r(n+1)^2).$$

The induction step is done. The proposition is proved. □

Theorem 5.7 *For every $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and every $\beta \in \mathbb{R}$ the map T given by (5.6) is minimal: each its forward orbit is dense.*

Proof Suppose the contrary: there exists a point (x_0, y_0) with non-dense forward orbit. Let M denote the set of limit points of its orbit. (The orbit is non-periodic, as is the rotation $x \mapsto x + \alpha$.) The set M is a non-empty closed subset in $\mathbb{T}^2$ with a non-empty open complement

$$V := \mathbb{T}^2 \setminus M.$$

Remark 5.8 The set M is T- and T^{-1}-invariant, hence so is V.

Proposition 5.9 *The set V contains no fiber $z \times S^1$.*

Proof Suppose the contrary: V contains such a fiber. Then there exists an interval neighborhood $U = U(z) \subset S^1$ such that $U \times S^1 \subset V$. But the successive images $T^m(U \times S^1)$ cover all of $\mathbb{T}^2$: the images of the interval U by translations $x \mapsto x + m\alpha$ cover all of S^1, since α is irrational. Therefore, $V = \mathbb{T}^2$. The contradiction thus obtained proves the proposition. □

Fix $a, b, c, d \in [0,1)$, $a < b$, $c < d$, such that

$$\Pi := [a,b] \times [c,d] \subset V.$$

Set

$$h := d - c.$$

Proposition 5.10 *For every $k \in \mathbb{Z}_{\geq 0}$ there exists a fiber $S^1_k = z_k \times S^1$ such that $S^1_k \cap V$ contains an arc of length greater than $2^k h$. In the case, when $2^k h > 1$, this means that the whole fiber S^1_k lies in V.*

Proof The proof is based on area-preserving property of the map T and the fact that for every $N \in \mathbb{N}$ the iterate T^N lifted to $\mathbb{R}^2$ transforms horizontal lines to lines with the slope (the tangent of angle with the horizontal axis) equal to N. Thus, as $N \to \infty$, the images of horizontal lines tend to vertical lines. Therefore, the images of a rectangle become very long strips spiralling in nearly vertical direction.

Induction in k.

Induction base for $k = 0$. Each fiber $z \times S^1$, $z \in [a, b]$, intersects V by an arc strictly containing the arc $z \times [c, d]$ of length h, and thus, having a bigger length.

Induction step. Let there exist a $z_k \in S^1$ such that the intersection $(z_k \times S^1) \cap V$ contains a segment $z_k \times [c_k, d_k]$ of length $c_k - d_k > 2^k h$. Let us show that V contains a vertical circle arc of twice bigger length. To do this, fix an $\varepsilon > 0$ such that

$$\Pi_k := [z_k - 3\varepsilon, z_k + 3\varepsilon] \times [c_k, d_k] \subset V.$$

Fix a $\delta \in (0, \varepsilon)$ such that $2\delta < h$. For every $N \in \mathbb{N}$ there exists a point

$$(x_N, y_N) \in \Pi_{k,\delta} := [z_k - \varepsilon, z_k + \varepsilon] \times [c_k, c_k + \delta] \subset \Pi_k$$

such that

$$p_n := T^n(x_N, y_N) \in \Pi_{k,\delta} \text{ for some } n > N,$$

by area-preserving property and the Poincaré Recurrence Theorem [14, theorem 4.1.19].

Claim 1. *Let us choose* $N > \frac{d_k - c_k + 2\delta}{\varepsilon}$. *Let* (x_N, y_N) *and* n *be as above. Then the* T^n*-image of the lower horizontal side* $L := [z_k - 3\varepsilon, z_k + 3\varepsilon] \times c_k$ *of the rectangle* Π_k *intersects its upper side at some point* q_k. *See Fig.12.*

Proof The point (x_N, c_k) lies in the lower side L, $|x_N - z_k| \leq \varepsilon$, and $0 \leq y_N - c_k \leq \delta$. The y-coordinate of its image, $y_n := y(T^n(x_N, c_k))$ also differs from c_k by a quantity no greater than δ, since it is no greater than that $y'_n := y(T^n(x_N, y_N)) \in [c_k, c_k + \delta]$, $y'_n - y_n = y_N - c_k \in [0, \delta]$: the map T preserves the lengths of arcs of vertical fibers. The x-coordinate of the same image $T^n(x_N, c_k)$ lies in the segment $[z_k - \varepsilon, z_k + \varepsilon]$. This together with the inequality on N and the fact that the T^n-image of a horizontal segment has slope $n > N$ implies that $T^n(L)$ crosses the upper side of the rectangle Π_k. □

Let $q_k = (z_{k+1}, d_k) = T^n(p_k)$ be a point of the above crossing, $p_k = (s_k, c_k) \in L$. Then $s_k \times [c_k, d_k] \subset \Pi_k \subset V$, by assumption. Therefore, $z_{k+1} \times [d_k, d_k + (d_k - c_k)] = T^n(s_k \times [c_k, d_k]) \subset V$, by invariance of the set V.

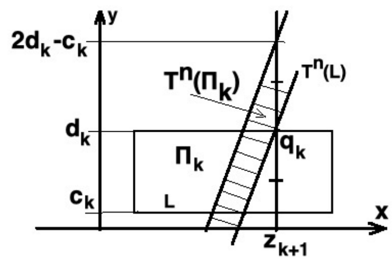

Figure 12: The rectangle Π_k, its image $T^n(\Pi_k)$ and the intersection point q_k.

Finally, the vertical circle arc $z_{k+1}\times[c_k, 2d_k-c_k]$ of length $2(d_k-c_k) > 2^{k+1}h$ lies in V. The induction step is done. Proposition 5.10 is proved. □

Proposition 5.10 applied to k large enough implies that V contains a vertical fiber $z \times S^1$. This contradicts Proposition 5.9. Theorem 5.7 is proved. □

Theorem 5.7 implies Theorem 5.4, and hence, Theorem 5.3.

6 Projections of geodesics-spirals to Heisenberg nil-manifold

For every $\theta \in \mathbb{R}/(2\pi\mathbb{Z})$, $h \neq 0$ consider a geodesic-spiral (4.9)–(4.11) in G and its projection to M:

$$a(t) = (\sin(\theta + ht) - \sin\theta)/h,$$
$$b(t) = (\cos\theta - \cos(\theta + ht))/h,$$
$$c(t) = (ht - \sin ht)/(2h^2) + a(t)b(t)/2,$$

$$g(t) = (a(t), b(t), c(t)), \quad t \in \mathbb{R}.$$
$$g'(t) := \pi \circ g(t) = (a'(t), b'(t), c'(t)),$$
$$\Gamma := \{g'(t) \mid t \in \mathbb{R}\} \subset M.$$

The projection $G \to \mathbb{R} \times \mathbb{R}$, $(a, b, c) \mapsto (a, b)$ passes to the quotient and induces the projection

$$p : M \to \mathbb{T}^2 = S^1 \times S^1, \quad S^1 = \mathbb{R}/\mathbb{Z}.$$

Theorem 6.1 *1) The projection p sends each geodesic-spiral Γ, see* (4.9)–(4.11)*, to a contractible closed curve $\gamma \subset \mathbb{T}^2$ that may have self-intersections.*

2) The geodesic-spiral Γ is

- either closed, which holds if and only if $h^2 \in \pi\mathbb{Q} \setminus \{0\}$;

- or dense in the preimage $p^{-1}(\gamma) \subset M$, if $h^2 \notin \pi\mathbb{Q}$.

See Figs. 13–16.

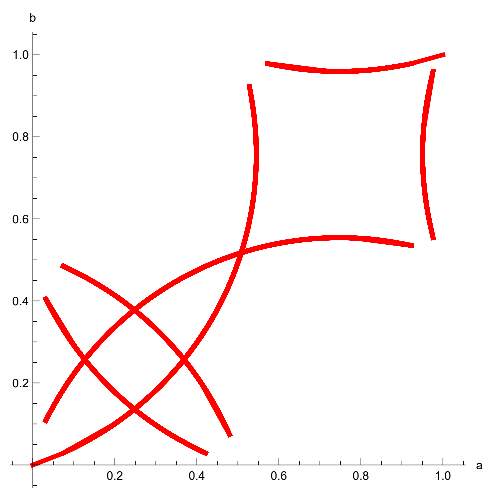

Figure 13: The curve γ for $h = \sqrt{\pi/2}$

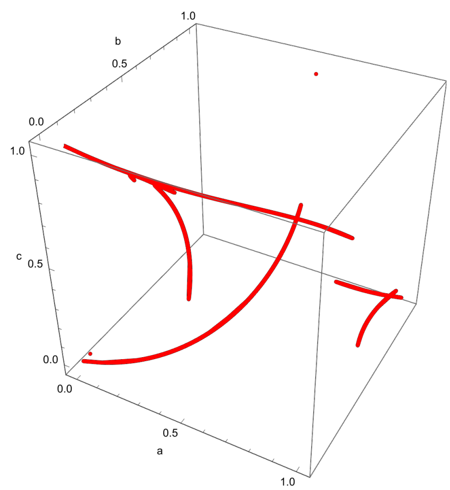

Figure 14: The curve Γ for $h = \sqrt{\pi/2}$

Proof Consider a spiral geodesic Γ given by (4.9)–(4.11) as a geodesic on the Heisenberg group $G \simeq \mathbb{R}^3_{x,y,z}$. Its projection to the (x, y)-plane is closed, being $\frac{2\pi}{h}$-periodic in t, see (4.9) and (4.10). It is clearly contractible in $\mathbb{R}^2_{x,y}$, as is every closed planar curve. Therefore, its projection $p(\Gamma)$ to $\mathbb{T}^2$ is also closed and contractible. Statement 1) is proved.

The coordinate $c' = \{c - [a]b\}$ of a point of the geodesic Γ is equal to

$$c'(t) = \left\{ \frac{t}{2h} - \frac{\sin ht}{2h^2} + \frac{x(t)y(t)}{2} - [x(t)]y(t) \right\}, \tag{6.1}$$

where $x(t)$, $y(t)$ are given by (4.9) and (4.10) respectively. All the terms in the right-hand side in (6.1) except for the first one are $\frac{2\pi}{h}$-periodic functions in t. Adding $\frac{2\pi}{h}$ to t results in adding $\frac{\pi}{h^2}$ to the first term $\frac{t}{2h}$. Therefore, closeness of the geodesic Γ is equivalent to commensurability of the numbers h^2 and π. If they are incommensurable, then the sequence of the numbers

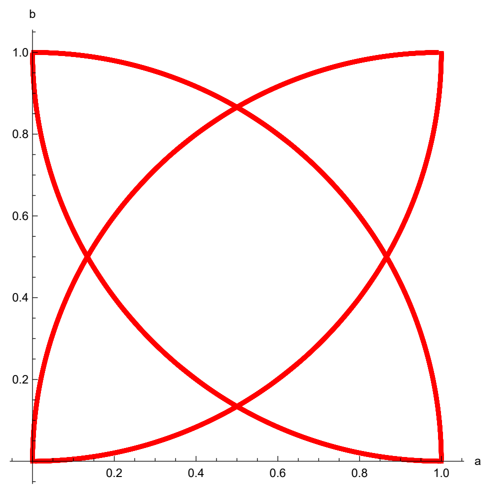

Figure 15: The curve γ for $h=1$

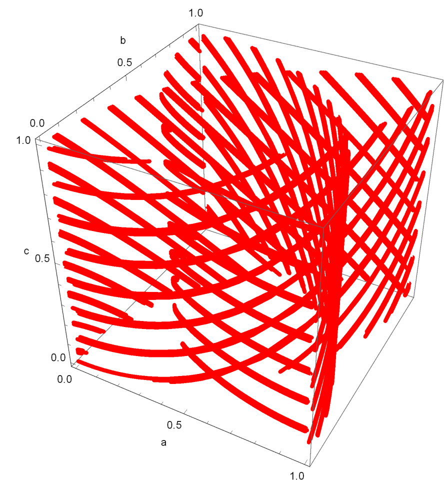

Figure 16: The curve Γ for $h=1$

$\{\frac{\pi n}{h^2}\}$ with $n \in \mathbb{N}$ is dense in $[0,1]$. This together with the above discussion implies that Γ is dense in $p^{-1}(\gamma)$. The theorem is proved. $\square$

7 Dynamics of the normal Hamiltonian flow on T^*M

Let T^*M be the cotangent bundle of the Heisenberg nil-manifold M. Introduce linear on fibers of T^*M Hamiltonians $h_i(\lambda) = \langle \lambda, X_i' \rangle$, $i = 1, 2, 3$, where $X_3' = [X_1', X_2'] = \frac{\partial}{\partial z'}$.

Let $H(\lambda) = (h_1^2(\lambda) + h_2^2(\lambda))/2$ be the normal Hamiltonian of the Pontryagin maximum principle [1, 19] for the sub-Riemannian problem (4.12)–(4.15), and let $\vec{H}$ be the corresponding Hamiltonian vector field on T^*M. Sub-Riemannian geodesics on M are projections of trajectories of the normal Hamiltonian system

$$\dot{\lambda} = \vec{H}(\lambda), \qquad \lambda \in T^*M, \tag{7.1}$$

in coordinates

$$\dot{h}_1 = -h_2 h_3, \quad \dot{h}_2 = h_1 h_3, \quad \dot{h}_3 = 0, \quad \dot{g}' = h_1 X_1' + h_2 X_2', \tag{7.2}$$

this follows from the classical coordinate expression of the Hamiltonian system for sub-Riemannian geodesics on the Heisenberg group [1, 19].

Each level surface

$$S_p = \{\lambda \in T^*M \mid H(\lambda) = 1/2,\ h_3(\lambda) = p\}, \qquad p \in \mathbb{R},$$

is invariant for the field $\vec{H}$. Denote on this level surface $h_1 = \cos\theta$, $h_2 = \sin\theta$. Denote also the restriction $V_p = \vec{H}\Big|_{S_p}$. The ODE

$$\dot{\lambda} = V_p(\lambda), \qquad \lambda \in S_p, \tag{7.3}$$

reads in coordinates as

$$\begin{cases} \dot{\theta} = p, \\ \dot{a}' = \cos\theta, \\ \dot{b}' = \sin\theta, \\ \dot{c}' = a'\sin\theta, \end{cases} \tag{7.4}$$

this follows immediately from ODEs (7.2) via the transformation formulas (4.5).

Set

$$S^1_\nu := \mathbb{R}_\nu/\mathbb{Z} \quad \text{for } \nu = a, b, c; \quad S^1_\theta = \mathbb{R}_\theta/2\pi\mathbb{Z}.$$

Let us introduce yet another flow on $S_p = S^1_\theta \times M$.

Definition 7.1 The *p-standard flow* F^t_p on $S^1_\theta \times M$ is given in the coordinates by the equation

$$\begin{cases} \dot{\theta} = p, \\ \dot{a}' = 0, \\ \dot{b}' = 0, \\ \dot{c}' = \frac{1}{2p}. \end{cases} \tag{7.5}$$

Remark 7.2 The projection of the vector field given by (7.5) to the 2-torus $\mathbb{T}^2_{\theta,c'}$ is the linear vector field $\dot{\theta} = p$, $\dot{c}' = \frac{1}{2p}$. Its flow map in time $\frac{2\pi}{p}$ fixes each circle $\{\theta\} \times S^1_{c'}$ and acts on it by translation (rotation)

$$c' \mapsto c' + \rho, \qquad \rho := \frac{\pi}{p^2}.$$

The number ρ is its rotation number, see the definition of rotation number in [3, p. 104].

Theorem 7.3 *Flow (7.3) is conjugated to the p-standard flow by a diffeomorphism of $S_p = S^1_\theta \times M$ preserving the θ-coordinate and isotopic to the identity in the class of diffeomorphisms preserving the θ-coordinate.*

Remark 7.4 Let $\pi : G \to M$ denote the quotient projection. Consider the pullbacks to $S^1_\theta \times G$ of vector fields (7.4), (7.5) under the induced projection $S^1 \times G \to S^1 \times M$. The pullbacks are written by the same formulas, as (7.4), (7.5), but in the coordinates (θ, a, b, c).

For the proof of Theorem 7.3 we first solve equations (7.4) and show that their solutions are all $\frac{2\pi}{p}$-periodic, except for the function $c(t)$, which is equal to $\frac{t}{2p}$ plus a $\frac{2\pi}{p}$-periodic function. Using formulas for solutions, we construct an explicit analytic family of diffeomorphisms $F_\nu : S^1_\theta \times G \to S^1_\theta \times G$ preserving the θ-coordinate, depending on the parameter $\nu \in [0,1]$, $F_0 = \mathrm{Id}$, such that F_1 transforms the lifted vector field (7.4) to the lifted field (7.5), see the above remark. We show that each F_ν is Γ-equivariant and thus, the family F_ν induces a family of diffeomorphisms $S^1 \times M \to S^1 \times M$ with F_1 sending (7.4) to (7.5). This will prove Theorem 7.3.

Set
$$\Sigma := \{H = \frac{1}{2}\} \subset T^*M.$$

Theorem 7.5 *1) The Hamiltonian flow on $T^*M^o := T^*M \setminus \{H = 0\}$ with Hamiltonian function H is integrable on the invariant domain $T^*M^{o,h_3} := T^*M^o \setminus \{h_3 = 0\}$: it has an additional integral $I(\lambda, a, b, c)$ analytic on T^*M^{o,h_3} that is in involution with the integrals H and h_3 for the canonical symplectic structure on T^*M. The latter integral I can be chosen to be any of the following functions:*

$$\cos(2\pi(a - \frac{\sin\theta}{h_3})),\ \sin(2\pi(a - \frac{\sin\theta}{h_3})),\ \cos(2\pi(b + \frac{\cos\theta}{h_3})),\ \sin(2\pi(b + \frac{\cos\theta}{h_3})). \tag{7.6}$$

2) For every integral I from (7.6) and $\nu \in (-1, 1)$, $p \neq 0$, the manifold

$$S_{p,\nu} := \{H = \frac{1}{2}\} \cap \{h_3 = p\} \cap \{I = \nu\} = S_p \cap \{I = \nu\}$$

is a transversal intersection, a disjoint union of two invariant 3-tori.

3) For $\nu = \pm 1$ (i.e., when ν is an extremum of the integral I) the latter tori coincide and the above intersection is one 3-torus.

4) The restriction of the Hamiltonian flow to any of the two latter tori is conjugated to a constant vector field on the standard 3-torus $\mathbb{R}^3/2\pi\mathbb{Z}^3$ with closed orbits for $p^2 \in \pi\mathbb{Q} \setminus \{0\}$ and each orbit dense in a 2-torus for $p^2 \notin \pi\mathbb{Q}$.

5) The flow restricted to the hypersurface $\Sigma = \{H = \frac{1}{2}\}$ has no non-trivial analytic integral: each analytic integral is a function of h_3.

Theorems 7.3, 7.5 are proved in Subsections 7.1 and 7.2 respectively.

7.1 Conjugacy with p-standard flow. Proof of Theorem 7.3

Proposition 7.6 *Each solution of the differential equation defined by the lifted vector field (7.4) (thus, written in the coordinates (θ, a, b, c)) with initial condition $(\theta_0, a_0, b_0, c_0)$ at $t = 0$ takes the form*

$$\begin{cases} \theta(t) = \theta_0 + pt \\ a(t) = \frac{1}{p}(\sin(\theta_0 + pt) - \sin\theta_0) + a_0 \\ b(t) = \frac{1}{p}(\cos\theta_0 - \cos(\theta_0 + pt)) + b_0 \\ c(t) = \frac{t}{2p} - \frac{1}{4p^2}(\sin 2(\theta_0 + pt) + \sin 2\theta_0) + \frac{1}{2p^2}(\sin(2\theta_0 + pt) - \sin pt) \\ -\frac{a_0}{p}(\cos(\theta_0 + pt) - \cos\theta_0) + c_0. \end{cases} \tag{7.7}$$

Proof The three first equations in (7.4) are solved by direct integration. The fourth equation is solved by taking the primitive:

$$c(t) = c_0 + \int_0^t a(t)\sin\theta(t)dt$$

$$= c_0 + \int_0^t (\frac{1}{p}(\sin(\theta_0 + pt) - \sin\theta_0) + a_0)\sin(\theta_0 + pt)dt. \tag{7.8}$$

The latter subintegral expression is equal to

$$\frac{1}{2p}(1 - \cos 2(\theta_0 + pt) + \cos(2\theta_0 + pt) - \cos pt) + a_0\sin(\theta_0 + pt).$$

Therefore, the integral is equal to

$$\frac{t}{2p} - \frac{1}{4p^2}(\sin 2(\theta_0 + pt) - \sin 2\theta_0 + 2\sin 2\theta_0 - 2\sin(2\theta_0 + pt) + 2\sin pt)$$

$$+\frac{a_0}{p}(\cos\theta_0 - \cos(\theta_0 + pt)).$$

This together with (7.8) yields (7.7). □

Proposition 7.7 *The phase curves of the lifted vector field (7.4) are graphs of vector functions $(a(\theta), b(\theta), c(\theta))$, where*

$$\begin{cases} a(\theta) = \frac{\sin\theta}{p} + a_0 \\ b(\theta) = \frac{1}{p}(1 - \cos\theta) + b_0 \\ c(\theta) = \frac{\theta}{2p^2} - \frac{1}{4p^2}\sin 2\theta - \frac{a_0}{p}(\cos\theta - 1) + c_0, \end{cases} \tag{7.9}$$

$$a_0 = a(0), b_0 = b(0), c_0 = c(0).$$

Proof Each phase curve intersects the fiber $\{\theta = 0\}$. Hence, it is the graph of a solution (7.7) with $\theta_0 = 0$. Substituting $\theta_0 = 0$ to (7.7) yields

$$\begin{cases} \theta(t) = pt \\ a(t) = \frac{\sin\theta}{p} + a_0 \\ b(t) = \frac{1}{p}(1 - \cos\theta) + b_0 \\ c(t) = \frac{t}{2p} - \frac{1}{4p^2}\sin 2\theta - \frac{a_0}{p}(\cos\theta - 1) + c_0, \end{cases}$$

which implies (7.9). □

Consider the following family of diffeomorphisms $F_\nu : S^1_\theta \times G \to S^1_\theta \times G$:

$$F_\nu(\theta, a, b, c) = (\theta,\ a - \frac{\nu}{p}\sin\theta,\ b - \frac{\nu}{p}(1 - \cos\theta),\ \widetilde{c}_\nu),$$

$$\widetilde{c}_\nu := c + \nu\left(\frac{1}{4p^2}\sin 2\theta + \frac{1}{p}(a - \frac{\sin\theta}{p})(\cos\theta - 1)\right),\ \nu \in [0, 1]. \qquad (7.10)$$

The action of the group Γ on G by left multiplication lifts to its action on $S^1_\theta \times G$:

$$\gamma(\theta, g) := (\theta, \gamma g) \text{ for every } \gamma \in \Gamma,\ g \in G.$$

Proposition 7.8 *1) For every $\nu \in \mathbb{R}$ the map F_ν is a diffeomorphism equivariant under left multiplications by elements of the group Γ:*

$$F_\nu(\theta, \gamma g) = \gamma F_\nu(\theta, g) \text{ for every } \gamma \in \Gamma,\ g \in G. \qquad (7.11)$$

2) One has $F_0 = id$, and F_1 transforms the lifting to $S^1 \times G$ of flow (7.4) to the lifting of the p-standard flow (7.5).

Proof Statement 2) follows from (7.10) and (7.9). Let us prove Statement 1). The group Γ has two generators:

$$\mathcal{A} := \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},\ \ \mathcal{B} := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Let us represent each $g \in G$ by its coordinates (a, b, c). The multiplication by $\mathcal{A}$ from the left acting on G lifts to the action on $S^1_\theta \times G$ by the formula $(\theta, a, b, c) \mapsto (\theta, a + 1, b, c + b)$. Therefore,

$$\mathcal{A}F_\nu(\theta, a, b, c) = (\theta, a + 1 - \frac{\nu}{p}\sin\theta, b - \frac{\nu}{p}(1 - \cos\theta), \widetilde{c}_{\nu,1}),$$

$$\widetilde{c}_{\nu,1} := c+\nu\left(\frac{1}{4p^2}\sin 2\theta + \frac{1}{p}(a - \frac{\sin\theta}{p})(\cos\theta - 1)\right)+b+\frac{\nu}{p}(\cos\theta-1), \quad (7.12)$$

$$F_\nu\circ\mathcal{A}(\theta,a,b,c) = F_\nu(\theta,a+1,b,c+b) = (\theta,a+1-\frac{\nu}{p}\sin\theta, b-\frac{\nu}{p}(1-\cos\theta),\widetilde{c}_{\nu,2}),$$

$$\widetilde{c}_{\nu,2} := c+\nu\left(\frac{1}{4p^2}\sin 2\theta + \frac{1}{p}(a+1 - \frac{\sin\theta}{p})(\cos\theta - 1)\right)+b.$$

This together with (7.12) implies that $c_{\nu,1} = c_{\nu,2}$ and proves (7.11) for $\gamma = \mathcal{A}$. Statement (7.11) for $\gamma = \mathcal{B}$ follows from (7.10) and the relation $\mathcal{B}(\theta,a,b,c) = (\theta,a,b+1,c)$. Proposition 7.8 is proved. □

The quotient of the diffeomorphism F_1 under the projection $\pi : S^1_\theta \times G \to S^1_\theta \times M$ is a well-defined diffeomorphism of the manifold $S_p = S^1_\theta \times M$ preserving the coordinate θ and isotopic to the identity in the class of diffeomorphisms preserving the coordinate θ. It transforms flow (7.4) to (7.5) by construction. This proves Theorem 7.3.

7.2 Integrability. Proof of Theorem 7.5

Let us prove Statement 1) of Theorem 7.5. The function $f = a - \frac{\sin\theta}{h_3}$ is defined on the cotangent bundle T^*G to the group G with the hypersurface $\{h_3 = 0\}$ deleted. It is a first integral of flow (7.4) lifted to T^*G. This follows immediately from the first and second equation in (7.4), since h_3 is a first integral. Similarly the function $g = b + \frac{\cos\theta}{h_3}$ is an integral. Each one of the two latter functions is automatically in involution with the Hamiltonian H, being an integral. Let us show that f is in involution with h_3, i.e.,

$$\omega(\vec{f},\vec{h}_3) = 0, \ \omega \ \text{ is the standard symplectic form on } \ T^*G. \quad (7.13)$$

Here by $\vec{\psi}$ we denote the skew gradient of a function ψ, which means by definition that $\omega(\vec{\psi}, v) = (d\psi)(v)$ for every $v \in TM$. Indeed, consider the coordinates (x,y,z) on G and the associated coordinates $(x,y,z,\lambda_1,\lambda_2,\lambda_3)$ on T^*G: λ_1, λ_2, λ_3 is the basis in $T^*_{(x,y,z)}G$ dual to the basis $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$, $\frac{\partial}{\partial z}$ in $T_{(x,y,z)}G$. The standard symplectic form on T^*G is

$$\omega = dx\wedge d\lambda_1 + dy\wedge d\lambda_2 + dz\wedge d\lambda_3.$$

One has

$$\vec{h}_3 = \vec{\lambda}_3 = (0,0,1,0,0,0). \quad (7.14)$$

To calculate $\vec{f}$, recall that

$$h_1 = \langle\lambda, X_1\rangle = \lambda_1 - \frac{y}{2}\lambda_3, \quad h_2 = \langle\lambda, X_2\rangle = \lambda_2 + \frac{x}{2}\lambda_3, \quad \theta = \arctan\frac{h_2}{h_1},$$

$$f = x - \frac{\sin\theta}{\lambda_3}.$$

The sixth, that is, λ_3-component of the skew gradient $\vec{f}$ is zero, since f is z-independent. Therefore, ω vanishes on the pair of vectors $\vec{h}_3$, $\vec{f}$, by (7.14). This proves (7.13). Analogously, the function g is in involution with h_3. This implies that each one of integrals in (7.6) is in involution with h_3 on T^*M^o and proves Statement 1).

Let us prove Statement 2). The submanifold $S_p = \{H = \frac{1}{2},\ h_3 = p\} = S^1 \times M \subset T^*M$ lifts to T^*G as a submanifold $\widetilde{S}_p = S^1 \times G$ covering S_p via the canonical projection $\widetilde{S}_p \to S_p$ induced by the quotient projection $G \to M$. The function $f = a - \frac{\sin\theta}{h_3} = a - \frac{\sin\theta}{p}$ is well-defined on $\widetilde{S}_p$ for $p \neq 0$. The surface $\widetilde{S}_p$ has natural coordinates (θ, a, b, c). The function $f|_{\widetilde{S}_p}$ has nowhere vanishing differential, since it has unit partial derivative in a. Therefore, the function $I = \cos(2\pi(a - \frac{\sin\theta}{h_3}))$ from (7.6) restricted to $S_p \subset T^*M^o$ also has nowhere vanishing differential, except for the points where the $|\cos|$ takes its maximal value 1. Hence, its level hypersurfaces $\{I = \nu\}$ with $\nu \in (-1, 1)$ are transversal to S_p. Writing

$$\nu = \cos 2\pi\alpha, \quad \alpha \in (0, \frac{1}{2}),$$

we get that

$$\{I = \nu\} \cap S_p = \cup_{\pm}\{(\theta, a, b, c) \mid a = \frac{\sin\theta}{p} \pm \alpha (\operatorname{mod} \mathbb{Z})\}. \tag{7.15}$$

The latter intersection is a union of two compact invariant 3-manifolds, each being the subset in the right-hand side with a given sign choice $\pm$. They are tori: this follows from Arnold–Liouville Theorem on integrable systems [4, chapter 10, section 49] and can be also proved directly. This together with analogous statements for the other integrals from (7.6) (proved similarly) proves Statement 2).

As $\nu = \pm 1$, one has $\alpha \in \{0, \frac{1}{2}\}$, and the two 3-tori in the union in (7.15) coincide, since in this case $\alpha \equiv -\alpha (\operatorname{mod} \mathbb{Z})$. This proves Statement 3).

Statement 4) follows from Theorem 6.1 and Arnold–Liouville Theorem.

In the proof of Statement 5) we use the following obvious corollaries of Theorem 5.3 on density of orbits in $S_0 = \Sigma \cap \{h_3 = 0\}$. To state them, let us introduce the following notations. For every $\theta_0 \in \mathbb{R}$, $p \in \mathbb{R}$, set

$$M_{\theta_0,p} := \Sigma \cap \{\theta = \theta_0\} \cap \{h_3 = p\} = \{\theta_0\} \times M \subset S_p = S^1 \times M.$$

Recall that for every θ_0 the fiber $M_{\theta_0,0}$ is an invariant manifold for the flow, since $\dot{\theta} = 0$.

Proposition 7.9 *1) For every θ with $\tan\theta \notin \mathbb{Q} \cup \{\infty\}$ the flow on $M_{\theta,0}$ is minimal: each orbit is dense.*

2) For every θ as above and $\varepsilon > 0$ there exists a $T = T_{\varepsilon,\theta} > 0$ such that each finite orbit of the flow on $M_{\theta,0}$ in times $t \in [0, T]$ is $\frac{\varepsilon}{4}$-dense in $M_{\theta,0}$: this means that it intersects the $\frac{\varepsilon}{4}$-neighborhood of each point in $M_{\theta,0}$.

Proof Theorem 5.3 immediately implies density, which in its turn together with compactness implies Statement 2). □

Proposition 7.10 *For every $\varepsilon > 0$ there exists a $\delta > 0$ such that for every $p \in (0, \delta)$ each orbit of the restriction of the flow to S_p is ε-dense in S_p.*

Proof Choose a finite collection of numbers $\theta_1, \dots, \theta_N \in S^1 = \mathbb{R}/2\pi\mathbb{Z}$ with $\tan\theta_j \notin \mathbb{Q} \cup \{\infty\}$ that is $\frac{\varepsilon}{4}$-dense on S^1. There exists a $T > 0$ such that for every $j = 1, \dots, N$ each finite orbit of the flow on $M_{\theta_j,0}$ is $\frac{\varepsilon}{4}$-dense in $M_{\theta_j,0}$. The vector fields defining the flows on S_0 and S_p with $p \in (0, \delta)$ (both identified with the same product $S^1_\theta \times M$) are δ-close to each other. Therefore, as δ is small enough (depending on ε and T), for every $p \in (0, \delta)$ the finite orbit of the flow on $S_p = S^1 \times M$ in times $t \in [0, T]$ starting at each point (θ_j, x), $x \in M$, has $\frac{\varepsilon}{2}$-dense projection to M so that $\theta(t) \in [\theta_j, \theta_j + \frac{\varepsilon}{4}]$ for every $t \in [0, T]$. Along each full orbit in S_p the coordinate θ takes all values, including the above θ_j's. This together with the latter statement implies that it is ε-dense in S_p. The proposition is proved. □

Let us now prove Statement 5). Let the restriction of the Hamiltonian flow to Σ have a non-constant analytic integral I. Let us show that I is constant on each level hypersurface $S_p = \Sigma \cap \{h_3 = p\}$ of the function $h_3|_\Sigma$. This will imply that I is a function of h_3 and prove Statement 5).

The function $I_p := I|_{S_p}$ achieves its minimum on a compact invariant set for the flow in $S_p = S^1 \times M$, which will be denoted by $O_p \subset S^1 \times M$ (compactness). For arbitrarily small $\varepsilon > 0$ there exists a $\delta > 0$ such that for every $p \in (0, \delta)$ the invariant set O_p is ε-dense in $S^1 \times M$, since it consists of full orbits and by Proposition 7.10. Thus, it accumulates to all of S_0, as $p \to 0$. One has $dI_p = 0$ at all points in O_p, thus, at all points of an ε-dense subset accumulating to the whole hypersurface $S_0 \subset \Sigma$. This implies that $I \equiv const$ on S_0.

The manifold Σ is identified with the product $S^1_\theta \times M \times \mathbb{R}_{h_3}$. Suppose the contrary: there exists a $p \in \mathbb{R}$ such that I is non-constant along the hypersurface $S_p = S^1 \times M \times \{p\} \subset \Sigma$. Then some its first order partial derivative in local coordinates of the product $S^1 \times M$ is not identically zero. Let us denote the latter derivative by g. Fix a point $y \in S_0 = S^1 \times M \times \{0\}$

and consider the analytic extension of the function g to some its complex neighborhood $U = U(y)$ in the complexified manifold Σ. Then $g \not\equiv 0$ on U, by uniqueness of analytic extension and connectivity. On one hand, the zero locus $\{g = 0\}$ contains the sets $O_p \times \{p\} \subset \{dI_p = 0\}$. The latter sets, and hence, the intersections $\{g = 0\} \cap \{h_3 = p\}$ accumulate to all of $S_0 \cap U$, as do O_p.

On the other hand, the zero locus of a non-identically-zero analytic function g on U vanishing on S_0 (where $I = const$) is the union of the intersection $S_0 \cap U$ and another complex hypersurface that is a closed subset in U intersecting the complexified hypersurface S_0 by an analytic subset of complex codimension two. This follows from basic analytic set theory, see [11], which also implies that the latter codimension two analytic subset cannot contain all of the real part of the intersection $S_0 \cap U$. Thus, the set $\{g = 0\} \setminus S_0$ cannot accumulate to the real hypersurface $S_0 \cap U$. The contradiction thus obtained proves that $I_p \equiv const$ for every $p \in \mathbb{R}$ and finishes the proof of Statement 5) and hence, Theorem 7.5.

8 Two-sided bounds of the Heisenberg sub-Riemannian balls and distance

The sub-Riemannian sphere of radius $R > 0$ on the Heisenberg group G centred at the origin $g_0 = \mathrm{Id}$ is parameterized as follows [1, 20]:

$$x = R\frac{\sin p}{p}\cos\tau, \quad y = R\frac{\sin p}{p}\sin\tau, \quad z = R^2\frac{2p - \sin 2p}{8p^2},$$
$$p \in [-\pi, \pi], \quad \tau \in \mathbb{R}/(2\pi\mathbb{Z}),$$

denote it as S_R. Each sphere is a rotation surface around the z-axis, and spheres of different radii are transferred one into another by dilations

$$\delta_k(x, y, z) = (kx, ky, k^2 z), \qquad k > 0, \quad (x, y, z) \in G,$$

as follows:

$$\delta_k(S_R) = S_{kR}. \tag{8.1}$$

The unit sphere $S := S_1$ is a rotation surface around the z-axis of the curve

$$r = \frac{\sin p}{p}, \quad z = \frac{2p - \sin 2p}{8p^2}, \qquad p \in [-\pi, \pi], \tag{8.2}$$

where $r = \sqrt{x^2 + y^2}$, see this curve in Fig. 17. The curve (8.2) intersects the z-axis at the points $(r, z) = \left(0, \pm\frac{1}{4\pi}\right)$. The unit sphere S is shown below in coordinates (x, y, z) (Fig. 18) and in coordinates (a, b, c) (Fig. 19).

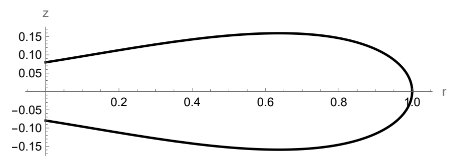

Figure 17: Section of the Heisenberg unit sphere

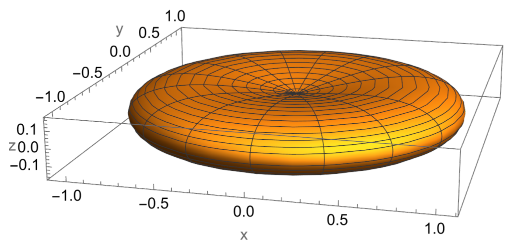

Figure 18: Sphere S in coordinates (x, y, z)

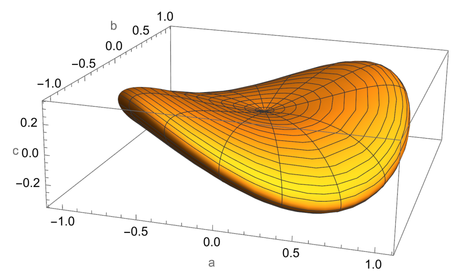

Figure 19: Sphere S in coordinates (a, b, c)

Consider the following domains bounded by ellipses in the plane $\mathbb{R}^2_{r,z}$:

$$e_1 \ : \ r^2 + 16\pi^2 z^2 \leq 1, \tag{8.3}$$

$$e_2 \ : \ r^2 + 12 z^2 \leq 1. \tag{8.4}$$

Lemma 8.1 *The ellipse ∂e_1 passes through the points $(r, z) = (1, 0)$ and $(r, z) = \left(0, \pm\frac{1}{4\pi}\right)$. The intersection $e_1 \cap \{r \geq 0\}$ is contained inside the curve* (8.2)*, see Fig.* 20*. Moreover, the curves $\partial e_1 \cap \{r \geq 0\}$ and* (8.2) *intersect only at the points $(r, z) = (0, \pm\frac{1}{4\pi})$ and $(r, z) = (1, 0)$.*

Proof First of all, it is obvious from (8.2) and (8.3) that the curves ∂e_1 and (8.2) intersect at the points $(r, z) = (0, \pm\frac{1}{4\pi})$ and $(r, z) = (1, 0)$.

Further, the ellipse ∂e_1 is the zero level curve of the function $f_1(r, z) = r^2 + 16\pi^2 z^2 - 1$. Evaluation of this function on the curve (8.2) is the function $\varphi_1(p) = \frac{\sin^2 p}{p^2} + \frac{\pi^2}{4p^4}(2p - \sin 2p)^2 - 1$. A standard calculus shows that $\varphi_1(0) = \varphi_1(\pm\pi) = 0$, and $\varphi_1(p) > 0$ for $0 < |p| < \pi$, see Fig. 22.

Indeed, we have $\varphi_1(p) = f_1(p) + f_2(p) - 1$, $f_1(p) = \frac{\sin^2 p}{p^2}$, $f_2(p) = \frac{\pi^2}{4p^4}(2p - \sin 2p)^2$. Notice that $f_2'(p) = \frac{2\pi^2}{p^5} \cos p (p \cos p - \sin p)(\sin 2p - 2p)$.

If $p \in (\pi/2, \pi)$, then $f_2'(p) < 0$, thus $f_2(p)$ decreases. Since $f_2(\pi) = 1$, then $f_2(p) > 1$, thus $\varphi_1(p) > 0$ for $p \in (\pi/2, \pi)$.

If $p \in (0, \pi/2)$, then $f_2'(p) > 0$, thus $f_2(p)$ increases. Since $f_2(\pi/6) = \frac{324}{\pi^2}\left(\frac{\pi}{3} - \frac{\sqrt{3}}{2}\right)^2 \approx 1.08 > 1$, then $f_2(p) > 1$ and $\varphi_1(p) > 0$ for $p \in [\pi/6, \pi/2]$.

In the proof below and in next lemmas we prove bounds of the form $g_1(p) < 0$, $g_1(0) = 0$, by comparing $g_1(p)$ with appropriate and more simple function $g_2(p)$, such that $(g_1(p)/g_2(p))' g_2^2(p) > 0$. We described this method and called it "divide et impera" in [18].

We have the following equalities:

$$\varphi_1'(p) = \frac{p\cos p - \sin p}{p^5} f_3(p),$$
$$f_3(p) = 2(p^2 + \pi^2 + \pi^2 \cos 2p)\sin p - 4p\pi^2 \cos p,$$
$$f_4(p) = \left(\frac{f_3(p)}{\sin p}\right)' \sin^2 p = 2p(1 + 2\pi^2 - \cos 2p) + \pi^2(\sin 4p - 4\sin 2p),$$
$$\left(\frac{f_4(p)}{\cos 2p - 2\pi^2 - 1}\right)' (\cos 2p - 2\pi^2 - 1)^2 = 4 f_5(p)\sin^2 p,$$
$$f_5(p) = -1 - 7\pi^2 + (1 + 2\pi^2 + 8\pi^4)\cos 2p - \pi^2 \cos 4p,$$
$$f_5'(p) = 4\cos p \sin p f_6(p),$$
$$f_6(p) = -1 - 2\pi^2 - 8\pi^4 + 4\pi^2 \cos 2p.$$

One has $f_6(p) \le -1 + 2\pi^2 - 8\pi^4 < 0$ for all p, since $\cos 2p \le 1$. Therefore the restriction to the semi-interval $(0, \pi/6]$ of the function $f_5(p)$ decreases, and hence, achieves its minimum at $p = \pi/6$. Its value there is equal to

$$-1 - 7\pi^2 + (1 + 2\pi^2 + 8\pi^4)/2 + \pi^2/2 = -1/2 - 5.5\pi^2 + 4\pi^4 > 0.$$

Therefore, $f_5(p) > 0$ on the above semi-interval. Hence, the function

$$\widetilde{f}_4(p) := \frac{f_4(p)}{\cos 2p - 2\pi^2 - 1}$$

increases there and thus, achieves there its minimum at $p = 0$. But $\widetilde{f}_4(0) = f_4(0) = 0$. Therefore, $\widetilde{f}_4(p) > 0$, hence $f_4(p) < 0$ for $p \in (0, \pi/6]$. Thus, the function

$$\widetilde{f}_3(p) := \frac{f_3(p)}{\sin p}$$

decreases on the same semi-interval. Hence, it achieves its supremum there at $p = 0$. But $\widetilde{f}_3(0) = 0$. Therefore, $\widetilde{f}_3(p) < 0$, hence $f_3(p) < 0$ on the semi-interval $(0, \pi/6]$. Thus, φ_1 increases there, by the above formula for its derivative and since $p \cos p - \sin p < 0$, i.e., $p < \tan p$, whenever $p \in (0, \pi/2)$. But $\varphi_1(0) = 0$. Hence, $\varphi_1 > 0$ there.

Summing up, if $p \in (0, \pi)$ then $\varphi_1(p) > 0$. Since $\varphi_1(p)$ is even, this inequality holds for $0 < |p| < \pi$. $\square$

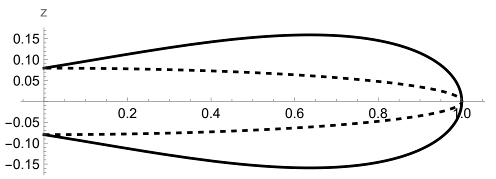

Figure 20: Ellipse ∂e_1 inside of section of sphere S

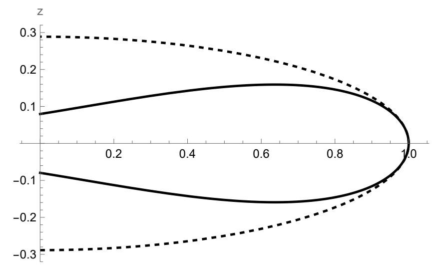

Figure 21: Ellipse ∂e_2 outside of section of sphere S

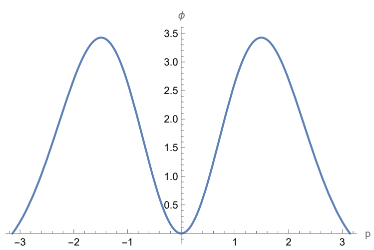

Figure 22: Plot of $\varphi_1(p)$

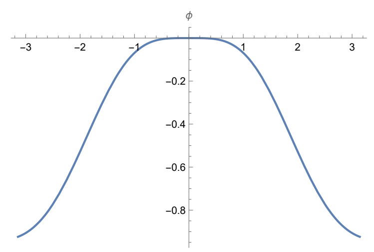

Figure 23: Plot of $\varphi_2(p)$

Remark 8.2 The ellipse ∂e_1 is the only ellipse in the plane (r, z), symmetric with respect to the z-axis, with the properties given in Lemma 8.1.

Lemma 8.3 *The curve ∂e_2 is tangent to the curve* (8.2) *with contact of order* 4*. The intersection $\partial e_2 \cap \{r \geq 0\}$ is contained outside of the curve* (8.2)*, see Fig.* 21*. Moreover, the curves $\partial e_2 \cap \{r \geq 0\}$ and* (8.2) *intersect only at the point $(r, z) = (1, 0)$.*

Proof The first statement is obtained by explicit differentiation. Indeed, for the ellipse ∂e_2 we have $r = \sqrt{1 - 12z^2} = 1 - 6z^2 + O(z^4)$, $z \to 0$. And for the curve (8.2) we have $r(0) = 1$, $z(0) = 0$. In a neighbourhood of the point $(r,z)=(1,0)$, the curves in question are graphs of even functions $r(z)$. Thus, it is sufficient to prove coincidence of their second derivatives at $z=0$. One has

$$\frac{dr}{dz} = \frac{dr/dp}{dz/dp} = -\frac{2p}{\cos p},$$
$$\frac{d^2r}{dz^2} = \frac{\frac{d}{dp}(-\frac{2p}{\cos p})}{dz/dp} = \frac{4p^3(1+p\tan p)}{\cos^3 p(p-\tan p)} \to -12, \quad p \to 0,$$

thus $r = 1 - 6z^2 + O(z^4)$, $z \to 0$.

The second statement follows since the function r^2+12z^2-1 whose zero level curve is the ellipse ∂e_2, when restricted to the curve (8.2), gives the function $\varphi_2(p) = \frac{\sin^2 p}{p^2} + \frac{3(2p-\sin 2p)^2}{16p^4} - 1$. A standard calculus shows that $\varphi_2(p) < 0$ for $0 < |p| \le \pi$, see Fig. 23.

Indeed, we have $\varphi_2(p) = f_1(p) + f_2(p) - 1$, $f_1(p) = \frac{\sin^2 p}{p^2}$, $f_2(p) = \frac{3(2p-\sin 2p)^2}{16p^4}$. Further,

$$f_3(p) = (16p^4\varphi_2(p))' = 40p - 64p^3 - 40p\cos 2p + 4(4p^2-3)\sin 2p + 6\sin 4p,$$
$$f_4(p) = f_3'(p) = 8(5 - 24p^2 + 4(p^2-2)\cos 2p + 3\cos 4p + 14p\sin 2p),$$
$$f_5(p) = f_4'(p) = -16(-18p\cos 2p + (4p^2-15)\sin 2p + 6(4p + \sin 4p)),$$
$$f_6(p) = f_5'(p) = -64(6 + 2(p^2-6)\cos 2p + 6\cos 4p + 11p\sin 2p),$$
$$f_7(p) = \left(\frac{f_6(p)}{\cos 2p}\right)' \cos^2 2p = 32(-48p - 4p\cos 4p + 12\sin 2p - 11\sin 4p + 12\sin 6p),$$
$$f_8(p) = f_7'(p) = 512 f_9(p) f_{10}(p),$$
$$f_9(p) = \sin p - \sin 3p, \qquad f_{10}(p) = -2p\cos p + 3\sin p + 9\sin 3p,$$
$$f_{11}(p) = \left(\frac{f_{10}(p)}{\cos p}\right)' \cos^2 p = 2 + 17\cos 2p + 9\cos 4p.$$

Let $p \in (0, \frac{\pi}{8}]$, then $f_{11}(p) > 0$, thus $\tilde{f}_{10}(p) = \frac{f_{10}(p)}{\cos p}$ increases. Since $\tilde{f}_{10}(0) = 0$, then $\tilde{f}_{10}(p) > 0$, so $f_{10}(p) > 0$.

Let $p \in (\frac{\pi}{8}, \frac{\pi}{4})$. Then $-2p\cos p > -\frac{\pi}{2}\cos\frac{\pi}{4} \approx -1.11 > -2$, $3\sin p > 3\sin\frac{\pi}{8} \approx 1.15 > 1$, $9\sin 3p > 9\sin\frac{3\pi}{4} \approx 6.36 > 6$, thus $f_{10}(p) > -2+1+6 > 0$.

Now let $p \in (0, \pi/4)$, we have proved that $f_{10}(p) > 0$. Since $f_9(p) < 0$, then $f_8(p) < 0$, thus $f_7(p)$ decreases. Since $f_7(0) = 0$, then $f_7(p) < 0$, thus $\tilde{f}_6(p) = \frac{f_6(p)}{\cos 2p}$ decreases. Since $\tilde{f}_6(0) = 0$ then $\tilde{f}_6(p) < 0$, thus $f_6(p) < 0$. Thus $f_5(p)$ decreases, and since $f_5(0) = 0$ then $f_5(p) < 0$. Thus

$f_4(p)$ decreases, and since $f_4(0) = 0$ then $f_4(p) < 0$. Thus $f_3(p)$ decreases, and since $f_3(0) = 0$ then $f_3(p) < 0$. Thus $p^4\varphi_2(p)$ decreases, and since $\lim_{p\to 0} p^4\varphi_2(p) = 0$ then $\varphi_2(p) < 0$ for $p \in (0, \pi/4]$.

If $p \in (\pi/4, 3\pi/8)$, then $-48p < -12\pi < -37.6$, $|4p\cos 4p| < 3\pi/2 < 4.72$, $|12\sin 6p| < 12$. One has $|12\sin 2p - 11\sin 4p| < 20.3$, since this is true at the endpoints $p = \pi/4, 3\pi/8$ and at the extremum point of the function under modulus in the interval in question. Indeed, its derivative in $x = 2p \in (\pi/2, 3\pi/4)$ is equal to $12\cos x - 22\cos 2x = 12u - 22(2u^2 - 1)$, $u = \cos x$. The latter derivative vanishes, if and only if $22u^2 - 6u - 11 = 0$. Solving the latter quadratic equation in negative $u = \cos x$ (which is indeed negative in the given interval) yields

$$u = \cos x = \frac{3 - \sqrt{251}}{22} \approx -0.58377, \ \sin x = \sqrt{1 - u^2} \approx \sqrt{0.65921} \approx 0.811,$$

$$|12\sin 2p - 11\sin 4p| = 2\sin x(6 - 11\cos x) \approx 20.169 < 20.3.$$

Thus, $\frac{f_7(p)}{32} < -37.6 + 4.72 + 20.3 + 12 < 0$. If $p \in [3\pi/8, \pi/2)$, then $-48p < -50$, $-4p\cos 4p \leq 0$, $|12\sin 2p - 11\sin 4p + 12\sin 6p| < 35$, thus $\frac{f_7(p)}{32} < -50 + 35 < 0$.

Thus for $p \in (\pi/4, \pi/2)$ we have $f_7(p) < 0$, so repeating the argument used two paragraphs above we get $f_i(p) < 0$ for $i = 3, \ldots, 6$, hence $\varphi_2(p) < 0$.

Finally, if $p \in [\pi/2, \pi)$, then $f_1(p) \leq \frac{4}{\pi^2} < \frac{1}{2}$. Since

$$f_2'(p) = -\frac{3}{p^5}\cos p(p\cos p - \sin p)(p - \cos p\sin p) \leq 0,$$

then $f_2(p)$ decreases, and since $f_2(\frac{\pi}{2}) = \frac{3}{\pi^2} \approx 0.3$ then $f_2(p) < \frac{1}{2}$. Thus $\varphi_2(p) < 0$.

If $p = \pi$ then $\varphi_2(p) = \frac{3 - 4\pi^2}{4\pi^2} < 0$.

We proved that $\varphi_2(p) < 0$ for $p \in (0, \pi]$. Since $\varphi_2(p)$ is even, this inequality holds for $0 < |p| \leq \pi$. □

Remark 8.4 The ellipse ∂e_2 is the smallest ellipse in the plane (r, z) among ellipses symmetric with respect to the z-axis, tangent to the curve (8.2) at the point $(r, z) = (1, 0)$ and encircling this curve.

Consider the projection

$$P : \mathbb{R}^3_{x,y,z} \to \{(r, z) \in \mathbb{R}^2 \mid r \geq 0\},$$
$$P(x, y, z) = (r, z) = (\sqrt{x^2 + y^2}, z)$$

and the corresponding ellipsoids $E_i = P^{-1}(e_i)$, $i = 1, 2$. Lemmas 8.1 and 8.3 plus equality (8.1) imply obviously the following two-sided ellipsoidal bounds of sub-Riemannian balls

$$B_R := \{g \in G \mid d(\mathrm{Id}, g) \leq R\}$$

on the Heisenberg group.

Corollary 8.5 *For any $R > 0$ we have*

$$\delta_R(E_1) \subset B_R \subset \delta_R(E_2). \tag{8.5}$$

Remark 8.6 Estimates (8.5) are sharp in the sense that the ellipsoids $\delta_R(E_1)$ and $\delta_R(E_2)$ are tangent to the sub-Riemannian ball B_R at its points in the plane $\{z = 0\}$. Moreover, the ellipsoid $\delta_R(E_1)$ intersects the sub-Riemannian ball B_R at its points in the line $\{x = y = 0\}$, see Figs. 20, 21.

Remark 8.7 In order to estimate precision of bounds (8.5), take the Euclidean volume $V = dx \wedge dy \wedge dz$ (in fact, Popp's volume [1]). Then

$$V(E_1) = \frac{1}{4} < V(B_1) = \frac{1}{12}\left(1 + 2\pi \int_0^{2\pi} \frac{\sin x}{x} dx\right) \approx 0.83$$
$$< V(E_2) = \frac{\pi}{2\sqrt{3}} \approx 0.91.$$

The above integral formula for the volume $V(B_1)$ was proved in [17], P. 587.

Corollary 8.8 *Let $g = (x, y, z) \in G$, and let $r = \sqrt{x^2 + y^2}$. Then*

$$\underline{d}(g) := \sqrt{\frac{\sqrt{r^4 + 48z^2} + r^2}{2}} \leq d(\mathrm{Id}, g) \leq \sqrt{\frac{\sqrt{r^4 + 64\pi^2 z^2} + r^2}{2}} =: \overline{d}(g). \tag{8.6}$$

Proof Since the statement holds trivially for $g = \mathrm{Id} = (0, 0, 0)$, we can assume that $g \neq \mathrm{Id}$, then $R := d(\mathrm{Id}, g) > 0$. Denote $g' = \delta_{\frac{1}{R}}(g)$, then $d(\mathrm{Id}, g') = 1$, and inclusions (8.5) imply that the functions $f_1(x, y, z) = r^2 + 16\pi^2 z^2$ and $f_2(x, y, z) = r^2 + 12z^2$, $r = \sqrt{x^2 + y^2}$, satisfy the inequalities

$$f_1(g') = f_1\left(\frac{x}{R}, \frac{y}{R}, \frac{z}{R^2}\right) \geq 1 = d(\mathrm{Id}, g') \geq f_2\left(\frac{x}{R}, \frac{y}{R}, \frac{z}{R^2}\right) = f_2(g').$$

Thus $\frac{r^2}{R^2} + 16\pi^2\frac{z^2}{R^4} \geq 1 \geq \frac{r^2}{R^2} + 12\frac{z^2}{R^4}$, i.e.,

$$16\pi^2 z^2\alpha^2 + r^2\alpha \geq 1 \geq 12z^2\alpha^2 + r^2\alpha, \qquad \alpha = \frac{1}{R^2}. \tag{8.7}$$

The second inequality in (8.7) solves to $0 < \alpha \leq \alpha_2 := \frac{2}{\sqrt{r^4+48z^2}+r^2}$, whence $R \geq \frac{1}{\sqrt{\alpha_2}} = \sqrt{\frac{\sqrt{r^4+48z^2}+r^2}{2}}$, which gives the first inequality in (8.6). Similarly, the first inequality in (8.7) solves to $\alpha \geq \alpha_1 := \frac{2}{\sqrt{r^4+64\pi^2z^2}+r^2}$, whence $R \leq \frac{1}{\sqrt{\alpha_1}} = \sqrt{\frac{\sqrt{r^4+64\pi^2z^2}+r^2}{2}}$, which gives the second inequality in (8.6). □

Remark 8.9 Estimates (8.6) are functional expressions of bounds (8.5):

$$\{g \in G \mid \underline{d}(g) \leq R\} = \delta_R(E_2), \qquad \{g \in G \mid \overline{d}(g) \leq R\} = \delta_R(E_1).$$

Remark 8.10 Estimates (8.6) are sharp in the following sense:

(1) in the case $z = 0$ these inequalities turn into equalities,

(2) in the case $r = 0$ the second inequality turns into equality corresponding to e_1.

In the case $rz \neq 0$ the both inequalities (8.6) are strict.

The second inclusion in (8.5) obviously implies the following inclusions.

Corollary 8.11 *For any $R > 0$ we have*

$$B_R \subset \left\{g = (x,y,z) \in G \mid \sqrt{x^2+y^2} \leq R\right\},$$
$$B_R \subset \left\{g = (x,y,z) \in G \mid |z| \leq \frac{R^2}{\sqrt{12}}\right\},$$

or, which is equivalent,

$$d(\mathrm{Id}, g) \geq \sqrt{x^2+y^2}, \tag{8.8}$$
$$d(\mathrm{Id}, g) \geq \sqrt[4]{12z^2}. \tag{8.9}$$

Notice that inequalities (8.8), (8.9) follow also from the first inequality in (8.6).

9 Bounds of cut time via lower bounds of sub-Riemannian balls

Fix a point $q_0 = g_0' \in M$. Denote the ball $B_t' = \{q \in M \mid d'(q_0, q) \leq t\}$, $t \geq 0$, where d' is the sub-Riemannian distance on M. Denote also

$$\bar{t} = \inf\{t > 0 \mid B_t' = M\}.$$

The following lemmas show the relevance of the number $\bar{t}$ for the sub-Riemannian manifold M.

Lemma 9.1 *We have the following:*

(1) $\bar{t} = \sup\{d'(q_0, q_1) \mid q_1 \in M\}$.

(2) $\bar{t} = \sup\{t_{\text{cut}}(q(\cdot)) \mid q(\cdot) \subset M$ *a geodesic s.t.* $q(0) = q_0\}$.

Proof (1) Denote $t_1 = \sup\{d'(q_0, q_1) \mid q_1 \in M\}$ and assume by contradiction that $\bar{t} \neq t_1$.

Let $t_1 < \bar{t}$. Then for every $t \in (t_1, \bar{t})$ and every $q_1 \in M$ we have $d'(q_0, q_1) < t$, i.e., $q_1 \in B_t'$. Since $t < \bar{t}$, this contradicts to definition of $\bar{t}$.

Let $t_1 > \bar{t}$. Then for every $t \in (\bar{t}, t_1)$ there exists $q_1 \in M$ such that $d'(q_0, q_1) > t$, i.e., $q_1 \notin B_t'$. Since $t > \bar{t}$, this contradicts to definition of $\bar{t}$ once more.

(2) Denote $t_2 = \sup\{t_{\text{cut}}(q(\cdot)) \mid q(\cdot) \subset M$ a geodesic s.t. $q(0) = q_0\}$ and assume by contradiction that $\bar{t} \neq t_2$.

Let $t_2 < \bar{t}$, take any $t \in (t_2, \bar{t})$. Then $B_t' \neq M$, thus there exists a point $q_1 \in M$ such that $d'(q_0, q_1) > t$. Take a sub-Riemannian length minimizer $q(\cdot)$ connecting q_0 and q_1. We have $t_{\text{cut}}(q(\cdot)) > t > t_2$, which contradicts the definition of t_2.

Let $\bar{t} < t_2$, take any $t \in (\bar{t}, t_2)$. We have $B_t' = M$, thus for every $q_1 \in M$ one has the inequality $d'(q_0, q_1) \leq t$. Then for every geodesic $q(\cdot) \subset M$ starting at q_0 we have $t_{\text{cut}}(q(\cdot)) \leq t < t_2$, which contradicts the definition of t_2. □

Remark 9.2 Consider the diameter of the sub-Riemannian manifold M:

$$\text{diam}(M) = \sup\{d'(q_1, q_2) \mid q_1, q_2 \in M\}.$$

By the triangle inequality, we have a bound $\text{diam}(M) \leq 2\bar{t}$.

Theorem 9.3 *We have* $\bar{t} \le t_* := \frac{1}{2}\sqrt{\frac{1}{2}\left(1+\sqrt{1+1024\pi^2}\right)} \approx 3.56$.

Proof We show that $B'_{t_*} = M$. By Corollary 8.5, we have $B_{t_*} \supset \delta_{t_*}(E_1) =: E_{1t_*}$, where the ellipsoid $E_{1t_*} \subset G$ is defined by the inequality $\frac{r^2}{t_*^2} + \frac{16\pi^2 z^2}{t_*^4} \le 1$. Thus $B'_{t_*} \supset E'_{1t_*} := \pi(E_{1t_*})$. We show that $E'_{1t_*} = M$.

The homogeneous space M can be represented by a fundamental domain $D = \{(a,b,c) \in G \mid 0 \le a,b,c < 1\}$, so that $\pi(D) = M$. We have $D \subset \cup_{i=1}^8 K_i$, where the cubes K_i are defined as follows:

$$K_1 \ : \ 0 \le a,b,c \le \frac{1}{2}, \qquad K_2 \ : \ 0 \le a,c \le \frac{1}{2} \le b \le 1,$$
$$K_3 \ : \ 0 \le a \le \frac{1}{2} \le b,c \le 1, \qquad K_4 \ : \ 0 \le a,b \le \frac{1}{2} \le c \le 1,$$
$$K_5 \ : \ 0 \le b,c \le \frac{1}{2} \le a \le 1, \qquad K_6 \ : \ 0 \le c \le \frac{1}{2} \le a,b \le 1,$$
$$K_7 \ : \ \frac{1}{2} \le a,b,c \le 1, \qquad K_8 \ : \ 0 \le b \le \frac{1}{2} \le a,c \le 1.$$

We show that $E'_{1t_*} \supset \pi(K_i)$, $i = 1, \ldots, 8$, which implies that $E'_{1t_*} = M$. To this end we define the following points $g_i \in H$ in the coordinates (a,b,c): $g_1 = (0,0,0)$, $g_2 = (0,1,0)$, $g_3 = (0,1,1)$, $g_4 = (0,0,1)$, $g_5 = (1,0,0)$, $g_6 = (1,1,0)$, $g_7 = (1,1,1)$, $g_8 = (1,0,1)$, and prove that $E_{1t_*} \supset \tilde{K}_i := g_i^{-1} K_i$, $i = 1, \ldots, 8$.

Let $L \subset \mathbb{R}^n$ be a convex compact set. We call a continuous function $f : L \to \mathbb{R}$ *quasiconvex* if $\max_L f = \max_{\partial L} f$. Since a convex function on a convex compact set attains maximum at points of the boundary of this set or at all points of this set, then a convex function on such a set is quasiconvex.

Now let $\Pi \subset \mathbb{R}^3$ be a parallelepiped whose all faces and edges are parallel to coordinate planes and axes, and let $\dim \Pi \in \{1,2,3\}$, i.e., Π is a 3D parallelepiped, a 2D rectangle, or a 1D segment. Let us study quasiconvexity of the function $f_t(a,b,c) = t^2(a^2+b^2) + 4\pi^2(2c-ab)^2 - t^4$ whose zero level is the ellipsoid ∂E_{1t}, $t > 0$. We have $\frac{\partial f_t}{\partial a} = 2t^2 a - 8\pi^2 b(2c-ab)$, $\frac{\partial f_t}{\partial b} = 2t^2 b - 8\pi^2 a(2c-ab)$, $\frac{\partial f_t}{\partial c} = 16\pi^2(2c-ab)$, thus f_t has only one critical point $(a,b,c) = (0,0,0)$, which is the minimum point. Thus if $\dim \Pi = 3$ then $f_t|_\Pi$ is quasiconvex.

If $\Pi \subset \{a = \text{const}\}$ or $\Pi \subset \{b = \text{const}\}$, then $f_t|_\Pi$ is convex, thus it is quasiconvex. Thus if $\dim \Pi = 3$ and the restriction of f_t to faces of Π parallel to the plane $\{c = 0\}$ is quasiconvex, then $f_t|_\Pi$ attains maximum at vertices of Π.

1) We prove that $E_{1t_*} \supset \tilde{K}_1 = K_1$. Since $\left.\frac{\partial f_t}{\partial a}\right|_{c=0} = 2a(4\pi^2 b^2 + t^2)$,

which is nonnegative and vanishes only for $a = 0$, then the function $f_{t_*}|_{K_1 \cap \{c=0\}}$ increases in a, thus $f_t|_{K_1 \cap \{c=0\}}$ is quasiconvex.

We have $p := \left(\frac{\partial f_{t_*}}{\partial a} + \frac{\partial f_{t_*}}{\partial b} \right)\Big|_{K_1 \cap \{c=1/2\}} = 2(a+b)(t_*^2 - 4(1-ab)\pi^2)$. Since in K_1 we have $ab \leq \frac{1}{4} < 1 - \frac{t_*^2}{4\pi^2} \approx 0.68$, then p is nonpositive and vanishes only at $(a,b) = (0,0)$, then $f_{t_*}|_{K_1 \cap \{c=1/2\}}$ is quasiconvex.

Thus $f_{t_*}|_{K_1}$ attains maximum at vertices of K_1. We have $f_{t_*}(0,0,0) \approx -161$, $f_{t_*}(0,0,1/2) \approx -122$, $f_{t_*}(0,1/2,0) = f_{t_*}(1/2,0,0) \approx -158$, $f_{t_*}(0,1/2,1/2) = f_{t_*}(1/2,0,1/2) \approx -118$, $f_{t_*}(1/2,1/2,0) \approx -152$, $f_{t_*}(1/2,1/2,1/2) \approx -133$, whence $f_{t_*}|_{K_1} < 0$, thus $E_{1t_*} \supset \tilde{K}_1 = K_1$.

2) We prove that $E_{1t_*} \supset \tilde{K}_2 = g_2^{-1} K_2$. Notice that for any elements (α,β,γ), (a,b,c) of G

$$(\alpha,\beta,\gamma)^{-1} \cdot (a,b,c) = (a-\alpha, b-\beta, c-\gamma+\alpha(\beta-b)).$$

Thus $\tilde{K}_2 = \{a,c \in [0,1/2],\ b \in [-1/2,0]\}$. By the argument of item 1), the function $f_{t_*}|_{\tilde{K}_2 \cap \{c=0\}}$ increases in a, thus $f_t|_{\tilde{K}_2 \cap \{c=0\}}$ is quasiconvex.

We have $\frac{\partial f_t}{\partial a}\Big|_{\tilde{K}_2 \cap \{c=1/2\}} = 8b(ab-1)\pi^2 + 2at^2$, which is nonnegative and vanishes only for $a = b = 0$, thus $f_t|_{\tilde{K}_2 \cap \{c=1/2\}}$ is quasiconvex.

So $f_{t_*}|_{\tilde{K}_2}$ attains maximum at vertices of $\tilde{K}_2$. In item 1) we showed that $f_{t_*} < 0$ at vertices of the square $[0,1/2]_a \times \{b=0\} \times [0,1/2]_c$. Further, we have $f_{t_*}(0,-1/2,0) \approx -158$, $f_{t_*}(0,-1/2,1/2) \approx -118$, $f_{t_*}(1/2,-1/2,0) \approx -152$, $f_{t_*}(1/2,-1/2,1/2) \approx -93$, then $f_{t_*}|_{\tilde{K}_2} < 0$, thus $E_{1t_*} \supset \tilde{K}_2$.

3) We prove that $E_{1t_*} \supset \tilde{K}_3 = g_3^{-1} K_3 = \{a \in [0,1/2],\ b,c \in [-1/2,0]\}$. Consider the involution $i : (a,b,c) \mapsto (a,-b,-c)$. Then $i(\tilde{K}_1) = \tilde{K}_3$ and $f_t \circ i = f_t$, thus $f_{t_*}|_{\tilde{K}_3} < 0$ since $f_{t_*} \circ i|_{\tilde{K}_1} = f_{t_*}|_{\tilde{K}_1} < 0$. Thus $E_{1t_*} \supset \tilde{K}_3$.

4) We prove that $E_{1t_*} \supset \tilde{K}_4 = g_4^{-1} K_4 = \{a,b \in [0,1/2],\ c \in [-1/2,0]\}$. Consider the involution $i : (a,b,c) \mapsto (a,-b,-c)$. Then $i(\tilde{K}_2) = \tilde{K}_4$ and $f_t \circ i = f_t$, thus $f_{t_*}|_{\tilde{K}_4} < 0$ by virtue of $f_{t_*}|_{\tilde{K}_2} < 0$ similarly to item 3).

5) We prove that $E_{1t_*} \supset \tilde{K}_5 = g_5^{-1} K_5$. We have $\tilde{K}_5 \subset \hat{K}_5 = \{a \in [-1/2,0],\ b \in [0,1/2],\ c \in [-1/2,1/2]\}$. Notice that $\frac{\partial f_t}{\partial a}\Big|_{c=-1/2} = 2a(t^2 + 4\pi^2 b^2) + 8\pi^2 b = 0$ only if $a = \bar{a} := -\frac{4\pi^2 b}{t^2 + 4\pi^2 b^2}$, and

$$\frac{\partial f_{t_*}}{\partial b}\Big|_{a=\bar{a},\ b\in[0,1/2],\ c=-1/2} = \frac{2bt_*^2(16\pi^4(b^4-1) + 8b^2\pi^2 t_*^2 + t_*^4)}{(4b^2\pi^2 + t_*^2)^2},$$

which is nonpositive (since the quartic polynomial in brackets in numerator is negative for $b \in [0,1/2]$) and vanishes only for $b = 0$. Thus $f_{t_*}|_{\hat{K}^5 \cap \{c=-1/2\}}$ has no interior critical points, so this function is quasiconvex.

We have $\frac{\partial f_t}{\partial a}\Big|_{\hat{K}^5\cap\{c=1/2\}} = 2a(t^2+4\pi^2b^2) - 8\pi^2 b = 0$ only if $a = \bar{a} := \frac{4\pi^2 b}{t^2+4\pi^2b^2}$, and $\frac{\partial f_{t_*}}{\partial b}\Big|_{a=\bar{a},\ b\in[0,1/2],\ c=1/2} = \frac{2bt_*^2(16\pi^4(b^4-1)+8b^2\pi^2t_*^2+t_*^4)}{(4b^2\pi^2+t_*^2)^2}$, which is nonpositive and vanishes only for $b=0$ by the previous paragraph. Thus $f_{t_*}|_{\hat{K}_5\cap\{c=1/2\}}$ has no interior critical points, so this function is quasiconvex.

So $f_{t_*}|_{\hat{K}_5}$ attains maximum at vertices of $\hat{K}_5$. Since $f_{t_*}(-1/2,0,-1/2) = f_{t_*}(-1/2,0,1/2) = f_{t_*}(1/2,0,1/2) = f_{t_*}(0,1/2,1/2) = f_{t_*}(0,1/2,-1/2) \approx -118$, $f_{t_*}(-1/2,1/2,-1/2) = f_{t_*}(1/2,1/2,1/2) \approx -133$, $f_{t_*}(-1/2,1/2,1/2) = f_{t_*}(1/2,-1/2,1/2) \approx -93$, $f_{t_*}(0,0,-1/2) = f_{t_*}(0,0,1/2) \approx -121$, see items 1)–3) above, then $f_{t_*}|_{\hat{K}_5} \leq 0$, thus $f_{t_*}|_{\tilde{K}_5} \leq 0$, and $E_{1t_*} \supset \tilde{K}_5$.

6) We prove that $E_{1t_*} \supset \tilde{K}_6 = g_6^{-1}K_6$. We have $\tilde{K}_6 \subset \hat{K}_6 = \{a,b \in [-1/2,0],\ c\in[0,1]\}$. Since $\frac{\partial f_t}{\partial a}\Big|_{\hat{K}_6\cap\{c=0\}} = 2a(4b^2\pi^2+t^2)$, which is nonpositive and vanishes only for $a=0$, then $f_t|_{\hat{K}_6\cap\{c=0\}}$ is quasiconvex.

We have $\delta := \left(\frac{\partial f_t}{\partial a} + \frac{\partial f_t}{\partial b}\right)\Big|_{\hat{K}_6\cap\{c=1\}} = 2(a+b)(t^2-4\pi^2(2-ab))$. Since $\frac{t_*^2}{4\pi^2} \approx 0.32 < \frac{7}{4} \leq 2-ab$, then δ is nonnegative and vanishes only for $a=b=0$. Thus $f_t|_{\hat{K}_6\cap\{c=1\}}$ is quasiconvex. So $f_{t_*}|_{\tilde{K}_6}$ attains maximum at vertices of $\tilde{K}_6$.

Since $f_{t_*}(-1/2,-1/2,0) = f_{t_*}(1/2,1/2,0) \approx -152$, see item 1),
$f_{t_*}(-1/2,-1/2,1) \approx -34$,
$f_{t_*}(-1/2,0,0) = f_{t_*}(0,-1/2,0) \approx -158$, see item 1), $f_{t_*}(-1/2,0,1) = f_{t_*}(0,-1/2,1) = 0$, $f_{t_*}(0,0,0) \approx -161$, see item 1), $f_{t_*}(0,0,1) \approx -3$, then $f_{t_*}|_{\hat{K}^6} \leq 0$, thus $f_{t_*}|_{\tilde{K}^6} \leq 0$, and $E_{1t_*} \supset \tilde{K}_6$.

7) We prove that $E_{1t_*} \supset \tilde{K}_7 = g_7^{-1}K_7$. We have $\tilde{K}_7 \subset \hat{K}_7 = \{a,b \in [-1/2,0],\ c\in[-1/2,1/2]\}$. Consider the involution $i : (a,b,c) \to (a,-b,-c)$. Then $i(\hat{K}_5) = \hat{K}_7$ and $f_t \circ i = f_t$. Since $f_{t_*}|_{\tilde{K}^5} \leq 0$, then $f_{t_*}|_{\tilde{K}^7} \leq 0$, and $E_{1t_*} \supset \tilde{K}_7$.

8) Finally, we prove that $E_{1t_*} \supset \tilde{K}_8 = g_8^{-1}K_8$. We have $\tilde{K}_8 \subset \hat{K}_8 = \{a \in [-1/2,0],\ b\in[0,1/2],\ c\in[-1,0]\}$. Consider the involution $i : (a,b,c) \to (a,-b,-c)$. Then $i(\hat{K}_6) = \hat{K}_8$ and $f_t \circ i = f_t$. Since $f_{t_*}|_{\tilde{K}^6} \leq 0$, then $f_{t_*}|_{\tilde{K}^8} \leq 0$, and $E_{1t_*} \supset \tilde{K}_8$.

Summing up, we proved that $E_{1t_*} \supset \cup_{i=1}^8 g_i^{-1}K_i$. Thus

$$E'_{1t_*} \supset \cup_{i=1}^8 \pi(g_i^{-1}(K_i)) \supset \pi(D) = M,$$

so the required inclusion $B'_{t_*} = M$ follows. □

We plot a union of sub-Riemannian balls $B_{t_*}(h)$ for some $h\in H$ covering the fundamental domain D in Fig. 24.

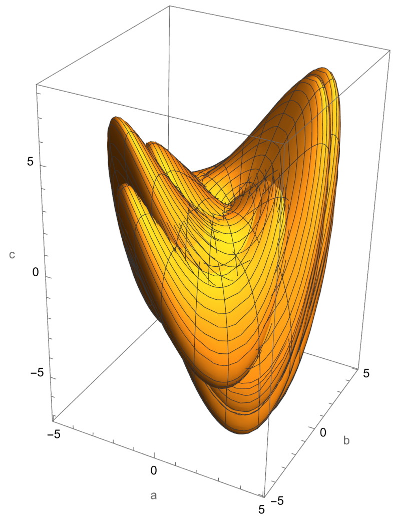

Figure 24: Union of sub-Riemannian Heisenberg balls $B_t(h)$ covering the fundamental domain D

Now we provide a lower bound of the number $\bar{t}$. Denote the points $\bar{g} = (\bar{a}, \bar{b}, \bar{c}), \tilde{g} = (\tilde{a}, \tilde{b}, \tilde{c}) \in G$ such that $\bar{a} = \bar{b} = \bar{c} = \frac{1}{2}$ and $\tilde{a} = 1$, $\tilde{b} = \tilde{c} = 0$.

Theorem 9.4 *We have* $\bar{t} \geq d(\tilde{g}, \bar{g})$.

Proof By left-invariance of the metric d, for any $g_1, g_2 \in G$ we have $d(g_1, g_2) = d(\mathrm{Id}, g_0)$, $g_0 = g_1^{-1} g_2$. The distance $d(\mathrm{Id}, g_0)$, $g_0 = (x, y, z)$, is computed explicitly [20]: if $(x, y) \neq (0, 0)$ then

$$d(\mathrm{Id}, g_0) = \frac{p}{\sin p} \sqrt{x^2 + y^2}, \tag{9.1}$$

$$\frac{2p - \sin 2p}{8 \sin^2 p} = \frac{|z|}{x^2 + y^2}, \qquad p \in [0, \pi], \tag{9.2}$$

and if $(x, y) = (0, 0)$ then $d(\mathrm{Id}, g_0) = 2\sqrt{\pi |z|}$.

Now we show that $d(\tilde{g}, \bar{g}) < 1$. We have $d(\tilde{g}, \bar{g}) = d(\mathrm{Id}, g)$, where $g = \tilde{g}^{-1}\bar{g} = (x, y, z) = (-1/2, 1/2, 1/8)$. The function $\psi(p) := \frac{2p - \sin 2p}{8 \sin^2 p}$ that appears in (9.2) increases as $p \in (0, \pi)$ from 0 to $+\infty$. Indeed, changing p to $u = 2p$, we get

$$4(1 - \cos u)^2 \frac{d\psi}{du} = (1 - \cos u)^2 - (u - \sin u) \sin u = 2(1 - \cos u) - u \sin u$$

$$= 4\sin p(\sin p - p\cos p) > 0 \quad \text{for} \quad p \in (0, \pi).$$

Since $\psi(1.25) \approx 0.26$ then $\psi(1.25) > \frac{1}{4}$. Let $p_* \in (0, \pi)$ be the root of the equation $\psi(p_*) = \frac{1}{4} = \frac{|z|}{x^2+y^2}$. Then $p_* = \psi^{-1}(\frac{1}{4}) < 1.25$, thus by equalities (9.1), (9.2)

$$d(\mathrm{Id}, \tilde{g}^{-1}\bar{g}) = \frac{p_*}{\sin p_*}\sqrt{\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2} < \frac{1.25}{\sin 1.25}\frac{1}{\sqrt{2}} \approx 0.93 < 1.$$

Here we use increasing of the function $\frac{p}{\sin p}$ in $p \in (0, \pi)$. Indeed, its derivative multiplied by $\sin^2 p$ is equal to $\sin p - p\cos p > 0$. So $d(\tilde{g}, \bar{g}) < 1$.

Take any number $\tilde{t} \in (0, d(\tilde{g}, \bar{g}))$. We show that $B'_{\tilde{t}} \neq M$. To this end we show that $\pi(\bar{g}) \notin B'_{\tilde{t}}$, i.e., that $\bar{g} \notin HB_{\tilde{t}}$.

We take any $h = (a, b, c) \in H$ and show that $\bar{g} \notin hB_{\tilde{t}}$. The following cases are possible:

1) $0 \leq a, b, c \leq 1$,

2) $(a \leq -1) \quad \vee \quad (a \geq 2) \quad \vee \quad (b \leq -1) \quad \vee \quad (b \geq 2)$,

3) $(0 \leq a, b \leq 1) \quad \wedge \quad ((c \leq -1) \quad \vee \quad (c \geq 2))$.

1) Let $0 \leq a, b, c \leq 1$.

1.1) Let $h = (0, 0, 0)$ in coordinates a, b, c, we denote this as $h = (0, 0, 0)_{abc}$. Then $g_0 = h^{-1}\bar{g} = \bar{g} = (1/2, 1/2, 1/2)_{abc} = (1/2, 1/2, 3/8)_{xyz}$, i.e., the point $\bar{g}$ has coordinates $(x, y, z) = (1/2, 1/2, 3/8)$. Thus by Corollary 8.8 $d(\mathrm{Id}, g_0) \geq \underline{d}(g_0) = \sqrt{\frac{1}{2}\left(\frac{1}{2} + \sqrt{7}\right)} \approx 1.25 > 1 > d(\tilde{g}, \bar{g}) > \tilde{t}$.

1.2) Let $h = (0, 0, 1)_{abc}$. Then $g_0 = h^{-1}\bar{g} = (1/2, 1/2, -1/2)_{abc} = (1/2, 1/2, -5/8)_{xyz}$, thus by Corollary 8.8 $d(\mathrm{Id}, g_0) \geq \underline{d}(g_0) = \sqrt{\frac{1}{2}\left(\frac{1}{2} + \sqrt{19}\right)} \approx 1.56 > 1 > d(\tilde{g}, \bar{g}) > \tilde{t}$.

1.3) Let $h = (0, 1, 0)_{abc}$.
Then $g_0 = h^{-1}\bar{g} = (1/2, -1/2, 1/2)_{abc} = (1/2, -1/2, 5/8)_{xyz}$, thus by Corollary 8.8
$d(\mathrm{Id}, g_0) \geq \underline{d}(g_0) = \sqrt{\frac{1}{2}\left(\frac{1}{2} + \sqrt{19}\right)} \approx 1.56 > 1 > d(\tilde{g}, \bar{g}) > \tilde{t}$.

1.4) Let $h = (0, 1, 1)_{abc}$. Then $g_0 = h^{-1}\bar{g} = (1/2, -1/2, -1/2)_{abc} = (1/2, -1/2, -3/8)_{xyz}$, thus by Corollary 8.8 $d(\mathrm{Id}, g_0) \geq \underline{d}(g_0) = \sqrt{\frac{1}{2}\left(\frac{1}{2} + \sqrt{7}\right)} \approx 1.25 > 1 > d(\tilde{g}, \bar{g}) > \tilde{t}$.

1.5) Let $h = (1, 0, 0)_{abc}$. Then $h = \tilde{g}$, thus $d(h, \bar{g}) = d(\tilde{g}, \bar{g}) > \tilde{t}$.

1.6) Let $h = (1, 0, 1)_{abc}$. Then $g_0 = h^{-1}\bar{g} = (-1/2, 1/2, -1)_{abc} = (-1/2, 1/2, -7/8)_{xyz}$, thus by Corollary 8.8 $d(\mathrm{Id}, g_0) \geq \underline{d}(g_0) = \sqrt{\frac{1}{2}\left(\frac{1}{2} + \sqrt{37}\right)} \approx 1.81 > 1 > d(\tilde{g}, \bar{g}) > \tilde{t}$.

1.7) Let $h = (1,1,0)_{abc}$. Then $g_0 = h^{-1}\bar{g} = (-1/2,-1/2,1)_{abc} = (-1/2,-1/2,7/8)_{xyz}$, thus by Corollary 8.8
$d(\mathrm{Id}, g_0) \geq \underline{d}(g_0) = \sqrt{\frac{1}{2}\left(\frac{1}{2}+\sqrt{37}\right)} \approx 1.81 > 1 > d(\tilde{g},\bar{g}) > \tilde{t}$.

1.8) Let $h = (1,1,1)_{abc}$. Then $g_0 = h^{-1}\bar{g} = (-1/2,-1/2,0)_{abc} = (-1/2,-1/2,-1/8)_{xyz}$. Consider the involution $i : (x,y,z) \to (x,-y,-z)$, then $d(\mathrm{Id}, i(g)) = d(\mathrm{Id}, g)$. Since $i(g_0) = \tilde{g}^{-1}\bar{g}$, then $d(\mathrm{Id}, g_0) = d(\mathrm{Id}, \tilde{g}^{-1}\bar{g}) > \tilde{t}$.

2) Let $(a \leq -1) \quad \vee \quad (a \geq 2) \quad \vee \quad (b \leq -1) \quad \vee \quad (b \geq 2)$. Since $d(h,\bar{g}) = d(\mathrm{Id}, h^{-1}\bar{g})$ and $h^{-1}\bar{g} = (\frac{1}{2}-a, \frac{1}{2}-b, *)$, then by inequality (8.8)

$$d(h,\bar{g}) \geq \sqrt{\left(\frac{1}{2}-a\right)^2 + \left(\frac{1}{2}-b\right)^2} \geq \sqrt{\left(\frac{3}{2}\right)^2 + \left(\frac{1}{2}\right)^2} = \frac{\sqrt{10}}{2} \approx 1.58.$$

Thus $d(h,\bar{g}) > d(\tilde{g},\bar{g}) > \tilde{t}$.

3) Let $(0 \leq a, b \leq 1) \quad \wedge \quad ((c \leq -1) \quad \vee \quad (c \geq 2))$. We have $d(h,\bar{g}) = d(\mathrm{Id}, g_0)$,

$$g_0 = h^{-1}\bar{g} = (x_0, y_0, z_0) = \left(\frac{1}{2}-a, \frac{1}{2}-b, \frac{ab}{2} + \frac{3}{8} + \frac{b-a}{4} - c\right).$$

If $c \geq 2$, then $|z_0| \geq \frac{7}{8}$. And if $c \leq -1$, then $|z_0| \geq \frac{9}{8}$. In both cases inequality (8.9) implies that $d(h,\bar{g}) = d(\mathrm{Id}, h^{-1}\bar{g}) \geq \sqrt[4]{12 z_0^2}\Big|_{|z_0|=\frac{7}{8}} = \sqrt[4]{\frac{147}{16}} > 1 > \tilde{t}$.

Summing up, we proved that $\bar{g} \notin HB_{\tilde{t}}$, and the statement of this theorem follows. $\square$

Remark 9.5 Numerical computations on the basis of equalities (9.1), (9.2) imply that $d(\widetilde{g},\bar{g}) \in (0.91, 0.92)$.

Remark 9.6 For comparison, consider the standard Euclidean metric on $\mathbb{R}^3$ and its quotient on the torus $\mathbb{T}^3 = \mathbb{R}^3/\mathbb{Z}^3$ (see Sec. 2). Then formula (2.2) yields

$$\sup\{t_{\mathrm{cut}}(q(\cdot)) \mid q(\cdot) \text{ geodesic on } \mathbb{T}^3\} = \frac{\sqrt{3}}{2} \approx 0.87.$$

This value is essentially less than our bound $\bar{t} \leq t_* \approx 3.56$ since in the Heisenberg group the sub-Riemannian distance grows slowly near the origin in the direction of the vector field X_3, see Fig. 24 and estimates (8.6).

10 Bounds of cut time via upper bounds of sub-Riemannian balls

Recall that $B_t(g) \subset G$ is the closed sub-Riemannian ball of radius $t \geq 0$ centred at a point $g \in G$, and $B_t := B_t(\mathrm{Id})$. Denote

$$\widehat{t} = \sup\{t > 0 \mid B_t(h_1) \cap B_t(h_2) = \emptyset \ \ \forall h_1 \neq h_2 \in H\}.$$

Since $B_t(h_i) = h_i B_t$, then

$$\widehat{t} = \sup\{t > 0 \mid B_t \cap B_t(h) = \emptyset \ \ \forall h \neq \mathrm{Id} \in H\}.$$

Recall that $g_0 = \mathrm{Id} \in G$, and for an element $g \in G$ we denote its projection to M as $g' := \pi(g)$.

Lemma 10.1 *If $d(g_0, g_1) < \widehat{t}$ for an element $g_1 \in G$, then $d'(g_0', g_1') = d(g_0, g_1)$.*

Proof Let $d(g_0, g_1) < \widehat{t}$. Notice that

$$d'(g_0', g_1') = \min\{d(\bar{g}_0, \bar{g}_1) \mid \pi(\bar{g}_i) = g_i'\} = d(\widetilde{g}_0, \widetilde{g}_1) \tag{10.1}$$

for some $\widetilde{g}_i \in \pi^{-1}(g_i')$. We have $\widetilde{g}_i = h_i g_i$, $h_i \in H$.

If $h_0 = h_1$, then $d'(g_0', g_1') = d(h_0 g_0, h_0 g_1) = d(g_0, g_1)$, and the claim follows.

Let $h_0 \neq h_1$. Then $d(\widetilde{g}_0, \widetilde{g}_1) = d(h_0 g_0, h_1 g_1) = d(g_0, h_0^{-1} h_1 g_1)$. Moreover, $h_0^{-1} h_1 g_1 \in h_0^{-1} h_1 B_t = B_t(h_0^{-1} h_1)$, $t = d(g_0, g_1) < \widehat{t}$. Since $h_0^{-1} h_1 \neq \mathrm{Id}$, then $B_t \cap B_t(h_0^{-1} h_1) = \emptyset$, thus $h_0^{-1} h_1 g_1 \notin B_t$. So $d(g_0, h_0^{-1} h_1 g_1) > t$, i.e., $d(g_0, h_0^{-1} h_1 g_1) > d(g_0, g_1)$, which contradicts to (10.1). Thus $h_0 = h_1$, and the claim follows by the previous paragraph. □

Lemma 10.2 *Let $g(\cdot)$ be a sub-Riemannian geodesic in G starting at $g_0 = \mathrm{Id}$ such that $t_{\mathrm{cut}}(g(\cdot)) \geq \widehat{t}$. Then $t_{\mathrm{cut}}(g'(\cdot)) \geq \widehat{t}$ as well.*

Proof Let $t_{\mathrm{cut}}(g(\cdot)) \geq \widehat{t}$. Take any $t \in (0, \widehat{t})$. The geodesic $g(\cdot)|_{[0,\widehat{t}]}$ is optimal, thus $d(g_0, g(t)) = t < \widehat{t}$. By Lemma 10.1 we have $d'(g_0', g'(t)) = d(g_0, g(t)) = t$, i.e., the geodesic $g'(\cdot)|_{[0,t]}$ is optimal. Thus $t_{\mathrm{cut}}(g'(\cdot)) \geq t$. Taking t arbitrarily close to $\widehat{t}$, we get the required bound $t_{\mathrm{cut}}(g'(\cdot)) \geq \widehat{t}$. □

Lemma 10.3 *Let $g(\cdot)$ be a sub-Riemannian geodesic in G starting at $g_0 = \mathrm{Id}$ such that $t_{\mathrm{cut}}(g(\cdot)) < \widehat{t}$. Then $t_{\mathrm{cut}}(g'(\cdot)) = t_{\mathrm{cut}}(g(\cdot))$.*

Proof Let $t_{\mathrm{cut}}(g(\cdot)) < \widehat{t}$. Take any $t \in (0, t_{\mathrm{cut}}(g(\cdot)))$. The geodesic $g(\cdot)|_{[0,t]}$ is optimal, thus $d(g_0, g(t)) = t < \widehat{t}$. By Lemma 10.1 we have $d'(g_0', g'(t)) = d(g_0, g(t)) = t$, i.e., the geodesic $g'(\cdot)|_{[0,t]}$ is optimal, so $t_{\mathrm{cut}}(g'(\cdot)) \geq t$. Taking t arbitrarily close to $t_{\mathrm{cut}}(g(\cdot))$, we get $t_{\mathrm{cut}}(g'(\cdot)) \geq t_{\mathrm{cut}}(g(\cdot))$.

Take any $\tau \in (t_{\mathrm{cut}}(g(\cdot)), \widehat{t})$. The geodesic $g(\cdot)|_{[0,\tau]}$ is not optimal, so $d(g_0, g(\tau)) < \tau$, thus $d(g_0, g(\tau)) < \widehat{t}$. By Lemma 10.1 we have $d'(g_0', g'(\tau)) = d(g_0, g(\tau)) < \tau$, i.e., the geodesic $g'(\cdot)|_{[0,\tau]}$ is not optimal. Thus $t_{\mathrm{cut}}(g'(\cdot)) = t_{\mathrm{cut}}(g(\cdot))$. □

Lemma 10.4 *Let $g(\cdot)$ be a sub-Riemannian geodesic in G starting at $g_0 = \mathrm{Id}$. Then $t_{\mathrm{cut}}(g'(\cdot)) \leq t_{\mathrm{cut}}(g(\cdot))$.*

Proof By contradiction, assume that $t_{\mathrm{cut}}(g'(\cdot)) > t_{\mathrm{cut}}(g(\cdot))$. Take any $t \in (t_{\mathrm{cut}}(g(\cdot)), t_{\mathrm{cut}}(g'(\cdot)))$. Then the geodesic $g'(\cdot)|_{[0,t]}$ is optimal, thus its length is equal to t: $l\left(g'(\cdot)|_{[0,t]}\right) = t$. But $l\left(g(\cdot)|_{[0,t]}\right) = l\left(g'(\cdot)|_{[0,t]}\right)$, and the geodesic $g(\cdot)|_{[0,t]}$ is not optimal, since $t > t_{\mathrm{cut}}(g(\cdot))$. Therefore, there exists another geodesic from $g(0)$ to $g(t)$ of length less than t. Its projection to M is a geodesic connecting $g'(0)$ and $g'(t)$ of the same length less than t. Therefore, the geodesic $g'(\cdot)|_{[0,t]}$ is not optimal, while $t < t_{\mathrm{cut}}(g'(\cdot))$. The contradiction thus obtained proves the lemma. □

Summing up, we have the following bounds of the cut time in M.

Corollary 10.5 *Let $g(\cdot)$ be a sub-Riemannian geodesic in G starting at $g_0 = \mathrm{Id}$. Then the following bounds hold:*

(1) $t_{\mathrm{cut}}(g'(\cdot)) \leq t_{\mathrm{cut}}(g(\cdot))$.

(2) *If* $t_{\mathrm{cut}}(g(\cdot)) \geq \widehat{t}$, *then* $\widehat{t} \leq t_{\mathrm{cut}}(g'(\cdot)) \leq t_{\mathrm{cut}}(g(\cdot))$.

(3) *If* $t_{\mathrm{cut}}(g(\cdot)) < \widehat{t}$, *then* $t_{\mathrm{cut}}(g'(\cdot)) = t_{\mathrm{cut}}(g(\cdot))$.

Remark 10.6 Corollary 10.5 holds in the general setting of Remark 4.1.

Now we compute the number $\widehat{t}$.

Theorem 10.7 *We have* $\widehat{t} = \frac{1}{2}$.

Proof In this proof we compute in coordinates (a, b, c). Denote the intersection $S_h^t = B_t \cap B_t(h)$, $h \in H$, $t > 0$. Then $S_{(\pm 1,0,0)}^{1/2} = \{(\pm\frac{1}{2}, 0, 0)\}$,

$S^{1/2}_{(0,\pm 1,0)} = \{(0, \pm\frac{1}{2}, 0)\}$, and all the rest of the sets $S^{1/2}_h$ are empty. Thus $\widehat{t} \leq \frac{1}{2}$.

Take any $t \in (0, \frac{1}{2})$, then for any $h \in H \setminus \{\mathrm{Id}\}$ we have $d(\mathrm{Id}, S^t_h) > \frac{1}{2}$, by the above statements and since the points $(\pm\frac{1}{2}, 0, 0)$, $(0, \pm\frac{1}{2}, 0)$ are on distance $\frac{1}{2}$ from the identity $\mathrm{Id} = (0, 0, 0)$. Thus $S^t_h = \emptyset$. So $\widehat{t} = \frac{1}{2}$. □

We plot the balls $B_{1/2}(h) \subset G$ for $h \in \{(0, 0, 0), (\pm 1, 0, 0), (0, \pm 1, 0)\}$ in Fig. 25, and the ball $B'_{1/2} \subset M$ in Fig. 26.

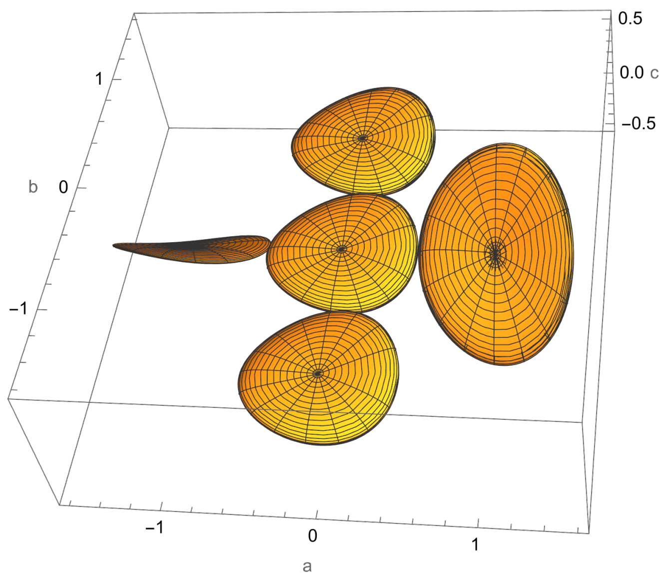

Figure 25: Sub-Riemannian balls $B_{1/2}(h) \subset G$ touching one another

Remark 10.8 For the quotient of the Euclidean metric from $\mathbb{R}^n$ to $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ (see Sec. 2) we have $\widehat{t} = \frac{1}{2}$ as well.

Theorem 10.9 *We have*

$$\sup\{t > 0 \mid \delta_t(E_2) \cap h\delta_t(E_2) = \emptyset \ \forall h \neq \mathrm{Id} \in H\} = \frac{1}{2}.$$

Proof Similarly to Theorem 10.7 since the ellipsoid $\delta_t(E_2)$ is tangent to the sub-Riemannian sphere B_t along the equator $B_t \cap \{z = 0\}$. □

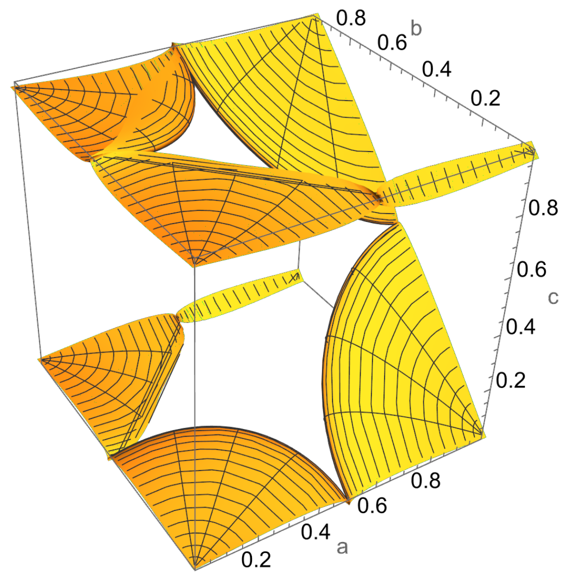

Figure 26: Sub-Riemannian ball $B'_{1/2} \subset M$ touching itself

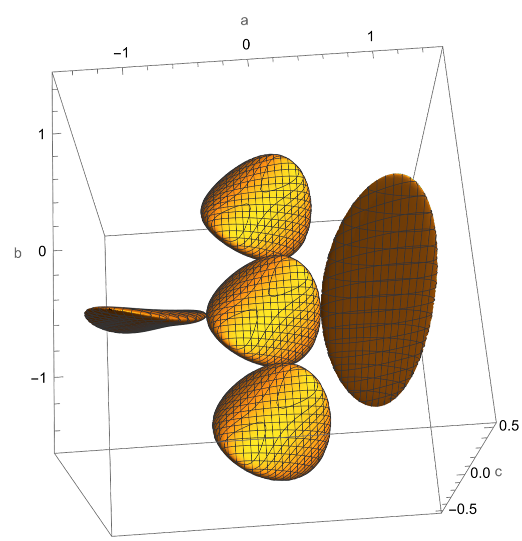

Figure 27: Sets $h\delta_{1/2}(E_2)$ touching one another

We plot the sets $h\delta_{1/2}(E_2)$ for $h \in \{(0,0,0), (\pm 1,0,0), (0,\pm 1,0)\}$ in Fig. 27.

Remark 10.10 Let $q(t)$ be a periodic sub-Riemannian geodesic in M of period T. Then it is obvious that $t_{\mathrm{cut}}(q(\cdot)) \leq \frac{T}{2}$ since $q(T/2)$ is a Maxwell point [19], i.e., an intersection point of two symmetric geodesics.

In a special case this bound turns into equality. Consider a geodesic $q(t)$ of the form (5.1) with $\theta = \frac{\pi n}{2}$, $n \in \mathbb{Z}$. Then it is easy to see that $t_{\mathrm{cut}}(q(\cdot)) = \frac{T}{2} = \frac{1}{2}$.

11 Conclusion

This work seems to be the first study of a projection of a left-invariant sub-Riemannian structure on a Lie group to a compact homogeneous space. It reveals essential difference between the initial structure and its projection despite their local isometry.

For example, dynamical behaviour of sub-Riemannian geodesics on the Heisenberg group G is trivial: all geodesics tend to infinity. Dynamics of sub-Riemannian geodesics on the Heisenberg 3D nil-manifold M includes closed geodesics, dense in M geodesics, and geodesics dense in a 2D torus.

Further, optimality of sub-Riemannian geodesics in G is very well understood; the corresponding cut time arises due to continuous symmetries of the sub-Riemannian structure on G. Description of optimality and cut time on M is much delicate since there are no continuous symmetries; and discrete symmetries which seem to generate the cut locus are hidden since they do not respect the projection mapping $\pi : G \to M$. Although, some two-sided bounds of the cut time in M are possible due to two-sided bounds of sub-Riemannian balls and distance in G, which may be of independent interest.

Acknowledgement

The authors thank V.M. Buchstaber, I.A. Taimanov, and A.V. Podobryaev for helpful discussions of certain aspects of this work.

References

[1] A. Agrachev, D. Barilari, U. Boscain, *A Comprehensive Introduction to sub-Riemannian Geometry from Hamiltonian viewpoint*, Cambridge Studies in Advanced Mathematics, Cambridge Univ. Press, 2019

[2] Agrachev A.A., Sachkov Yu.L., *Control theory from the geometric viewpoint*, Springer, 2004.

[3] Arnold, V. I. *Geometrical Methods in the Theory of Ordinary Differential Equations.* Second edition. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 250. Springer-Verlag, New York, 1988.

[4] Arnold V.I. *Mathematical methods of classical mechanics.* 2nd ed. Graduate texts in mathematics, 60. Springer-Verlag, New York, 1989.

[5] Berestovskii, V.N., Zubareva, I.A. Shapes of Spheres of Special Nonholonomic Left-Invariant Intrinsic Metrics on Some Lie Groups. Siberian Mathematical Journal 42, 613–628 (2001)

[6] Berestovskii, V.N., Zubareva, I.A. Geodesics and shortest arcs of a special sub-Riemannian metric on the Lie group SO(3). Sib Math J 56, 601–611 (2015).

[7] Berestovskii, V.N., Zubareva, I.A. Sub-Riemannian distance in the Lie groups SU(2) and SO(3). Sib. Adv. Math. 26, 77–89 (2016).

[8] Bock, C. *On low-dimensional solvmanifolds.* Asian J. Math. **20** (2016), No. 2, 199–262.

[9] U. Boscain, F. Rossi. "Invariant Carnot–Caratheodory metrics on S^3, SO(3), SL(2) and lens spaces", SIAM J. Control Optim., 47 (2008), pp. 1851–1878

[10] D.Ch. Chang, I. Markina, A. Vasiliev, Sub-riemannian geodesics on the 3-sphere, *Compl. anal. oper. theory*, 3 (2009), 361–377.

[11] Chirka, E.M., *Complex analytic sets,* Moscow, Nauka, 1985.

[12] Fomenko, A.; Fuchs, D. *Homotopical topology.* Second edition. Graduate Texts in Mathematics **273**, Springer, 2016.

[13] Furstenberg, H. *Strict Ergodicity and Transformation of the Torus.* Amer. J. Math. **83** (1961), No. 4, 573–601.

[14] Katok, A.; Hasselblatt, B. *Introduction to the modern theory of dynamical systems.* Cambridge University Press, 1995.

[15] Katok, A.; Hasselblatt, B. *A first course in dynamics.* Cambridge University Press, 2003.

[16] R. Montgomery, *A tour of subriemannnian geometries, their geodesics and applications*, Amer. Math. Soc., 2002

[17] E. F. Sachkova, Sub-Riemannian balls on the Heisenberg group: an invariant volume, *Journal of Mathematical Sciences*, 199:5 (2014), 583–587

[18] Yu.Sachkov, Conjugate Time in the Sub-Riemannian Problem on the Cartan Group. *J Dyn Control Syst* 27, 709–751 (2021).

[19] Yu.Sachkov, *Introduction to geometric control*, Springer, 2022.

[20] Yu.Sachkov, Left-invariant optimal control problems on Lie groups: classifications and problems integrable by elementary functions, Russian math. surveys, **77:1** (463) (2022), 109–176

[21] Vinberg, È. B.; Gorbatsevich, V. V.; Shvartsman, O. V. Discrete subgroups of Lie groups. Current problems in mathematics. Fundamental directions, Vol. 21 (Russian), 5–120, 215 Itogi Nauki i Tekhniki [Progress in Science and Technology] Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow, 1988

[22] Vershik, A.M., Gershkovich, V.Y. Nonholonomic problems and the theory of distributions. Acta Appl Math 12, 181–209 (1988)

[23] Montgomery, R. *Singular extremals on Lie groups*, Math. Control Signals Systems, Mathematics of Control, Signals, and Systems, 7, 1994, 3, 217–234