Sub-Riemannian geometry
(Lecture 6)

Yuri Sachkov

yusachkov@gmail.com

«Geometric control theory, nonholonomic geometry, and their applications»

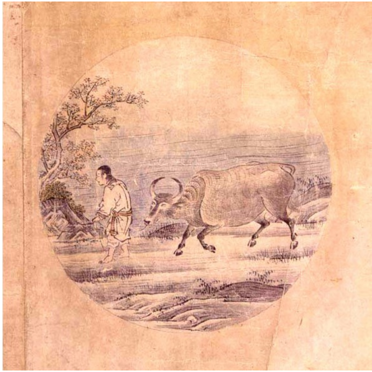
Lecture course in Dept. of Mathematics and Mechanics

Lomonosov Moscow State University

6 November 2024

5. Herding the Ox:

The boy is not to separate himself with his whip and tether,
Lest the animal should wander away into a world of defilements;
When the ox is properly tended to, he will grow pure and docile;
Without a chain, nothing binding, he will by himself follow the oxherd.
Pu-ming, "The Ten Oxherding Pictures"



Reminder: Plan of the previous lecture

1. Elements of symplectic geometry
2. Pontryagin maximum principle
3. Solution to examples of optimal control problems
4. Sub-Riemannian problems

Plan of this lecture

1. Sub-Riemannian problems
2. The sub-Riemannian problem on the Heisenberg group.

Sub-Riemannian optimal control problem

$$\dot{q} = \sum_{i=1}^{k} u_i f_i(q), \qquad q \in M, \quad u = (u_1, \dots, u_k) \in \mathbb{R}^k,$$

$$q(0) = q_0, \qquad q(t_1) = q_1,$$

$$l = \int_0^{t_1} \left(\sum_{i=1}^{k} u_i^2 \right)^{1/2} dt \to \min,$$

or, which is equivalent,

$$J = \frac{1}{2} \int_0^{t_1} \sum_{i=1}^{k} u_i^2 \, dt \to \min .$$

The Pontryagin maximum principle for SR problems

- Introduce the linear on fibers of T^*M Hamiltonians $h_i(\lambda) = \langle \lambda, f_i \rangle, \quad i = 1, \dots, k$. Then the Hamiltonian of PMP for SR problem takes the form

$$h_u^\nu(\lambda) = \sum_{i=1}^k u_i h_i(\lambda) + \frac{\nu}{2} \sum_{i=1}^k u_i^2.$$

- *The normal case: Let $\nu = -1$.*
- The maximality condition $\sum_{i=1}^k u_i h_i - \frac{1}{2} \sum_{i=1}^k u_i^2 \to \max\limits_{u_i \in \mathbb{R}}$ yields $u_i = h_i$, then the Hamiltonian takes the form

$$h_u^{-1}(\lambda) = \frac{1}{2} \sum_{i=1}^k h_i^2(\lambda) =: H(\lambda).$$

- The function $H(\lambda)$ is called the *normal maximized Hamiltonian*. Since it is smooth, in the normal case extremals satisfy the Hamiltonian system $\dot{\lambda} = \vec{H}(\lambda)$.

The abnormal case

- *Let* $\nu = 0$.
- The maximality condition
$$\sum_{i=1}^{k} u_i h_i \to \max_{u_i \in \mathbb{R}}$$
implies that $h_i(\lambda_t) \equiv 0, \quad i = 1, \dots, k$.
- Thus abnormal extremals satisfy the conditions:
$$\dot{\lambda}_t = \sum_{i=1}^{k} u_i(t) \vec{h}_i(\lambda_t),$$
$$h_1(\lambda_t) = \cdots = h_k(\lambda_t) \equiv 0.$$
- Normal length minimizers are projections of solutions to the smooth Hamiltonian system $\dot{\lambda} = \vec{H}(\lambda)$, thus they are smooth. An important *open question* of sub-Riemannian geometry is whether abnormal length minimizers are smooth.

Optimality of SR normal extremal trajectories

- *Let* $\nu = -1$.
- A horizontal curve $q(t)$ is called a *SR geodesic* if $g(\dot{q}, \dot{q}) \equiv \text{const}$ and short arcs of $q(t)$ are optimal.

Theorem 1 (Legendre)

Normal extremal trajectories are SR geodesics.

Proof.

See A.A. Agrachev, Yu.L. Sachkov, *Control theory from the geometric viewpoint*,
А.А. Аграчев, Ю. Л. Сачков, *Геометрическая теория управления*. □

Example: Geodesics on S^2

- Consider the standard sphere $S^2 \subset \mathbb{R}^3$ with the Riemannian metric induced by the Euclidean metric of $\mathbb{R}^3$.
- Geodesics starting from the North pole $N \in S^2$ are great circles at the sphere passing through N (meridians). Such geodesics are optimal up to the South pole $S \in S^2$.
- Variation of geodesics passing through N yields the fixed point S, thus S is a *conjugate point* to N.
- On the other hand, S is the intersection point of different geodesics of the same length starting at N, thus S is a *Maxwell point*.
- In this example, a conjugate point coincides with a Maxwell point due to the one-parameter group of symmetries (rotations of S^2 around the line $NS \subset \mathbb{R}^3$). In order to distinguish these points, one should destroy the rotational symmetry as in the following example.

Example: Geodesics on an ellipsoid

- Consider a three-axes ellipsoid with the Riemannian metric induced by the Euclidean metric of the ambient $\mathbb{R}^3$.
- Construct the family of geodesics on the ellipsoid starting from a vertex N, and let us look at this family from the opposite vertex S.
- The family of geodesics has an envelope — an astroid centred at S. Each point of the astroid is a *conjugate point*. At such points the geodesics lose their local optimality.
- On the other hand, there is a segment joining a pair of opposite vertices of the astroid, where pairs of geodesics of the same length meet one another. This segment (except its endpoints) consists of *Maxwell points*. At such points geodesics on the ellipsoid lose their global optimality.

Example: Geodesics on an ellipsoid

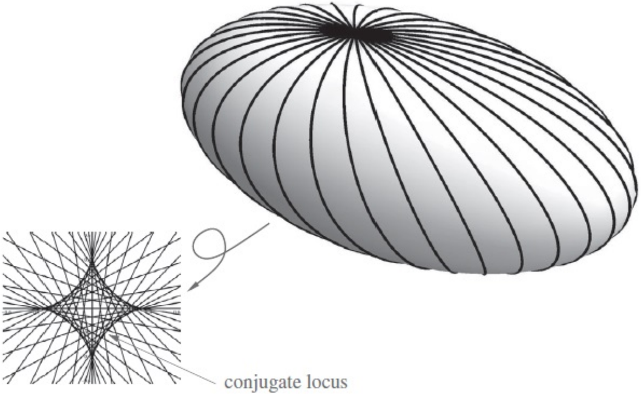


Figure from: Agrachev, D. Barilari, U. Boscain, A Comprehensive Introduction to sub-Riemannian Geometry from Hamiltonian viewpoint, Cambridge Studies in Advanced Mathematics, Cambridge Univ. Press, 2019

Sub-Riemannian exponential mapping

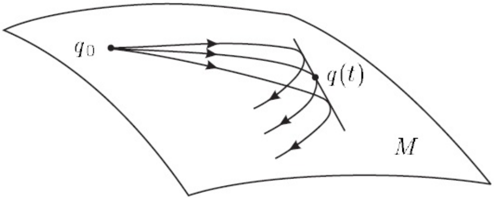
- Consider the normal Hamiltonian system of PMP $\dot{\lambda}_t = \vec{H}(\lambda_t)$.
- The Hamiltonian H is an integral of this system. We can assume that $H(\lambda_t) \equiv \frac{1}{2}$, this corresponds to the *arclength parametrization* of normal geodesics: $\|\dot{q}(t)\| \equiv 1$.
- Denote the cylinder $C = T^*_{q_0}M \cap \{H = \frac{1}{2}\}$ and define the sub-Riemannian *exponential mapping*

$$\mathrm{Exp} \;:\; C \times \mathbb{R}_+ \to M,$$

$$\mathrm{Exp}(\lambda_0, t) = \pi \circ e^{t\vec{H}}(\lambda_0) = q(t).$$

Conjugate points

- A point $\mathrm{Exp}(\lambda_0, t_1)$ is called a *conjugate point* along the geodesic $q(t) = \mathrm{Exp}(\lambda_0, t)$ if it is a critical value of Exp, i.e., $\mathrm{Exp}_{*(\lambda_0, t_1)}$ is degenerate.
- A point $\mathrm{Exp}(\lambda_0, t_1)$ is conjugate iff the Jacobian of the exponential mapping vanishes: $\det \left(\frac{\partial \, \mathrm{Exp}}{\partial(\lambda_0, t)} \right) \Big|_{t=t_1} = 0.$
- At a conjugate point $q(t)$ a geodesic $q(\cdot)$ is tangent to the envelope of the family of geodesics starting from the initial point q_0.



Local optimality of SR geodesics

A trajectory $\tilde{q}(t)$ of a control system with a control $\tilde{u}(t)$ and given boundary conditions q_0, q_1 is called *locally (strongly) optimal* if $J[\tilde{u}] \leq J[u]$ for any admissible control $u(t)$ such that the corresponding trajectory $q_u(t)$ satisfies the boundary conditions and is contained in some neighbourhood of $q_{\tilde{u}}(t)$ in the uniform topology of $C^0(M)$.

Theorem 2 (Jacobi)

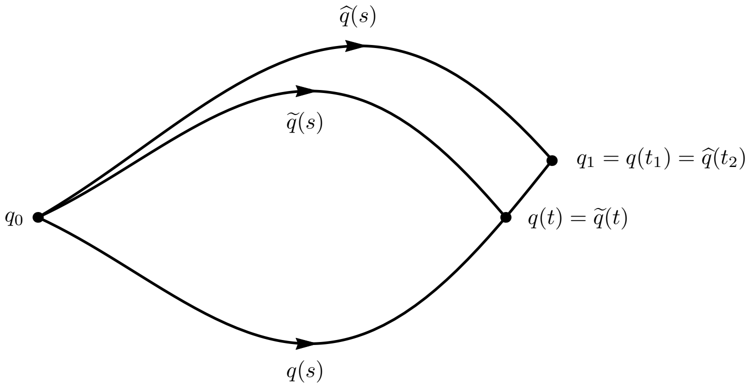
Let a normal geodesic $q(t)$ does not contain abnormal segments. Then $q(t)$ loses its local optimality at the first conjugate point.

Proof.

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 □

Maxwell points

- A point $q(t)$, $t > 0$, is called a *Maxwell point* along a geodesic $q(s) = \mathrm{Exp}(\lambda_0, s)$ if there exists another geodesic $\widetilde{q}(s) = \mathrm{Exp}(\widetilde{\lambda}_0, s) \not\equiv q(s)$ such that $q(t) = \widetilde{q}(t)$.
- See figure: there exists a geodesic $\widehat{q}(s)$ coming to the point $q_1 = q(t_1)$ earlier than $q(s)$.



Maxwell points and optimality

Lemma 3

If M and H are real-analytic, then a normal geodesic cannot be optimal after a Maxwell point.

Proof.

Let $q_1 = q(t_1)$ be a Maxwell point along a geodesic $q(t) = \text{Exp}(\lambda_0, t)$, and let $\tilde{q}(t) = \text{Exp}(\tilde{\lambda}_0, t) \not\equiv q(t)$ be another geodesic with $\tilde{q}(t_1) = q_1$. If $q(t)$, $t \in [0, t_1 + \varepsilon]$, $\varepsilon > 0$, is optimal, then the following curve is optimal as well:

$$\bar{q}(t) = \begin{cases} \tilde{q}(t), & t \in [0, t_1], \\ q(t), & t \in [t_1, t_1 + \varepsilon]. \end{cases}$$

The geodesics $q(t)$ and $\bar{q}(t)$ coincide at the segment $t \in [t_1, t_1 + \varepsilon]$. Since they are analytic, they should coincide at the whole domain $t \in [0, t_1 + \varepsilon]$. Thus $q(t) \equiv \tilde{q}(t)$, $t \in [0, t_1]$, a contradiction. □

Global optimality of SR geodesics

Theorem 4

- *Let M be a complete sub-Riemannian manifold (i.e., the topology on M defined by the sub-Riemannian distance, is complete).*
- *Let $q(t)$ be a normal geodesic that does not contain abnormal arcs.*

Then $q(t)$ loses its global optimality either at the first Maxwell point or at the first conjugate point (at the first one of these two points).

Proof.

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 □

A symmetry method for construction of optimal synthesis

- A general method for construction of optimal synthesis for sub-Riemannian problems with a big group of symmetries (e.g. for left-invariant SR problems on Lie groups)
- Assume that M is a complete sub-Riemannian manifold. Then for any $q_1 \in M$ there exists a length minimizer $q(t)$ that connects q_0 and q_1.
- Moreover, suppose for simplicity that all abnormal geodesics are simultaneously normal. Thus all geodesics are parametrised by the normal exponential mapping

$$\operatorname{Exp} : \; N \to M, \qquad N = C \times \mathbb{R}_+, \quad C = T^*_{q_0} M \cap \left\{ H = \frac{1}{2} \right\}.$$

- If this mapping is bijective onto $M \setminus \{q_0\}$, then any point $q_1 \in M$ is connected with q_0 by a unique geodesic $q(t)$, and by virtue of existence of length minimizers this geodesic is optimal.

A symmetry method for construction of optimal synthesis

- But typically the exponential mapping is not bijective due to Maxwell points.
- Denote by $t^1_{\mathrm{Max}}(\lambda) \in (0, +\infty]$ the first Maxwell time along a geodesic $\mathrm{Exp}(\lambda, t)$, $\lambda \in C$. Consider the Maxwell set in the image of the exponential mapping

$$\mathrm{Max} = \left\{ \mathrm{Exp}(\lambda, t^1_{\mathrm{Max}}(\lambda)) \mid \lambda \in C \right\}.$$

- Introduce the restricted exponential mapping

$$\mathrm{Exp} : \ \widetilde{N} \to \widetilde{M},$$
$$\widetilde{N} = \left\{ (\lambda, t) \in N \mid t < t^1_{\mathrm{Max}}(\lambda) \right\},$$
$$\widetilde{M} = M \setminus \mathrm{cl}(\mathrm{Max}).$$

- This mapping may well be bijective, and if this is the case, then any point $q_1 \in \widetilde{M}$ is connected with q_0 by a unique candidate optimal geodesic; by virtue of existence, this geodesic is optimal.
- The bijective property of the restricted exponential mapping can often be proved via the following classic theorem due to Hadamard.

Theorem 5 (Hadamard)

Let $F : X \to Y$ be a smooth mapping between smooth manifolds for which the following conditions hold:

(1) $\dim X = \dim Y$

(2) *X, Y are connected, and Y is simply connected*

(3) *F is nondegenerate*

(4) *F is* proper *(preimage of a compact set is compact).*

Then F is a diffeomorphism, thus a bijection.

A symmetry method for construction of optimal synthesis

- Usually it is difficult to describe all Maxwell points (and respectively to describe the first of them), but one can often do this for a group of symmetries G of the exponential mapping.
- Suppose that we have a mapping ε acting both in the preimage and image of the exponential mapping: $\varepsilon : N \to N, \qquad \varepsilon : M \to M$. This mapping is called a *symmetry of the exponential mapping* if it commutes with this mapping: $\varepsilon \circ \mathrm{Exp} = \mathrm{Exp} \circ \varepsilon$ and if it preserves time: $\varepsilon(\lambda, t) = (* , t), \qquad (\lambda, t) \in N$.
- Suppose that there is a group G of symmetries of the exponential mapping. If

  $$\varepsilon(\lambda, t) \neq (\lambda, t) \text{ and } \mathrm{Exp} \circ \varepsilon(\lambda, t) = \mathrm{Exp}(\lambda, t) = q_1, \qquad \varepsilon \in G, \quad (\lambda, t) \in N,$$

  then q_1 is a Maxwell point.
- In such a way, one can describe the Maxwell points corresponding to the group of symmetries G, and consequently describe *the first Maxwell time corresponding to the group G*: $t^G_{\mathrm{Max}} : C \to (0, +\infty]$.
- Then one can apply the above procedure with the restricted exponential mapping. Thus one can often construct optimal synthesis.

Examples of successful application of the symmetry method

- Dido's problem (the sub-Riemannian problem on the Heisenberg group)
- the sub-Riemannian problem in the flat Martinet case
- axisymmetric sub-Riemannian problems on the Lie groups SO(3), SU(2), SL(2)
- a general left-invariant sub-Riemannian problem on the Lie group SO(3)
- the sub-Riemannian problem with the growth vector $(3, 6)$
- the two-step sub-Riemannian problems of coranks 1 and 2
- the sub-Riemannian problem on the group of Euclidean motions of the plane
- the sub-Riemannian problem on the group of hyperbolic motions of the plane
- Euler's elastic problem
- the problem on optimal rolling of a sphere on a plane without slipping, with twisting
- the plate-ball problem
- sub-Riemannian problem on the Engel group
- sub-Riemannian problem on the Cartan group
- axisymmetric Riemannian problems on the Lie groups SO(3), SU(2), SL(2), PSL(2).

The sub-Riemannian problem on the Heisenberg group

Dido's problem is stated as the following optimal control problem:

$$\dot{q} = u_1 f_1(q) + u_2 f_2(q), \qquad q \in M = \mathbb{R}^3_{x,y,z}, \quad u = (u_1, u_2) \in \mathbb{R}^2,$$

$$q(0) = q_0 = (0,0,0), \qquad q(t_1) = q_1,$$

$$J = \frac{1}{2} \int_0^{t_1} (u_1^2 + u_2^2)\, dt \to \min,$$

$$f_1 = \frac{\partial}{\partial x} - \frac{y}{2} \frac{\partial}{\partial z}, \quad f_2 = \frac{\partial}{\partial y} + \frac{x}{2} \frac{\partial}{\partial z}.$$

- *Existence of solutions*.
- We have $[f_1, f_2] = f_3 = \frac{\partial}{\partial z}$. The system is symmetric and full-rank, thus it is completely controllable.
- The right-hand side satisfies the bound

$$|u_1 f_1(q) + u_2 f_2(q)| \leq C(1 + |q|), \qquad q \in M, \quad u_1^2 + u_2^2 \leq 1.$$

Thus the Filippov theorem gives existence of optimal controls.

The sub-Riemannian problem on the Heisenberg group

- *Extremals*.
- Introduce linear on fibers of T^*M Hamiltonians:

$$h_i(\lambda) = \langle \lambda, f_i \rangle, \qquad i = 1, 2, 3, \quad \lambda \in T^*M.$$

- *Abnormal extremals* satisfy the Hamiltonian system $\dot{\lambda} = u_1 \vec{h}_1(\lambda) + u_2 \vec{h}_2(\lambda)$, in coordinates:

$$\begin{aligned} \dot{h}_1 &= -u_2 h_3, \\ \dot{h}_2 &= u_1 h_3, \\ \dot{h}_3 &= 0, \\ \dot{q} &= u_1 f_1 + u_2 f_2, \end{aligned}$$

plus the identities

$$h_1(\lambda_t) = h_2(\lambda_t) \equiv 0.$$

Thus $h_3(\lambda_t) \neq 0$, and the first two equations of the Hamiltonian system yield $u_1(t) = u_2(t) \equiv 0$. So abnormal trajectories are constant.

The sub-Riemannian problem on the Heisenberg group

- *Normal extremals* satisfy the Hamiltonian system $\dot{\lambda} = \vec{H}(\lambda)$ with the Hamiltonian $H = \frac{1}{2}(h_1^2 + h_2^2)$, in coordinates:

$$\dot{h}_1 = -h_2 h_3, \tag{1}$$

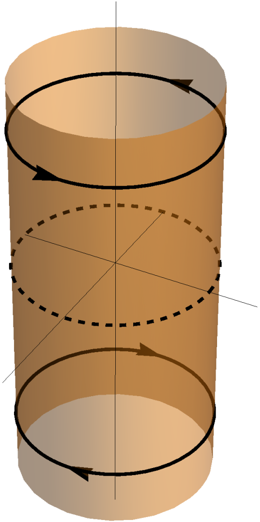
$$\dot{h}_2 = h_1 h_3, \tag{2}$$

$$\dot{h}_3 = 0, \tag{3}$$

$$\dot{q} = h_1 f_1 + h_2 f_2. \tag{4}$$

- The subsystem of the Hamiltonian system for the adjoint variables h_1, h_2, h_3 (the *vertical subsystem*) (1)–(3) has integrals H and h_3. Moreover, in the plane $\{h_3 = 0\}$ the vertical subsystem stays fixed. Thus at the level surface $\{H = 1/2\}$ it has the flow shown in the next slide: rotations in the circles $\{H = 1/2, \ h_3 = \text{const} \neq 0\}$ and fixed points in the circle $\{H = 1/2, \ h_3 = 0\}$.

The sub-Riemannian problem on the Heisenberg group: The flow of the vertical subsystem of the Hamiltonian system of PMP



The sub-Riemannian problem on the Heisenberg group

- On the level surface $\{H = \frac{1}{2}\}$, we introduce the polar coordinate θ:

$$h_1 = \cos\theta, \quad h_2 = \sin\theta.$$

Arclength parametrized minimizers satisfy the normal Hamiltonian system

$$\begin{aligned}
\dot\theta &= h_3,\\
\dot h_3 &= 0,\\
\dot x &= \cos\theta,\\
\dot y &= \sin\theta,\\
\dot z &= -\frac{y}{2}\cos\theta + \frac{x}{2}\sin\theta,\\
(x, y, z)(0) &= (0, 0, 0).
\end{aligned}$$

The sub-Riemannian problem on the Heisenberg group: Geodesics

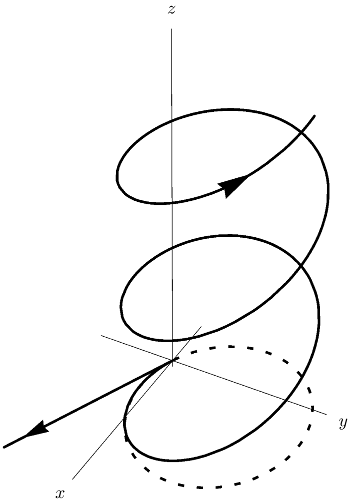
1. If $h_3 = 0$, then lines in the plane $\{z = 0\}$:

$$\begin{aligned} \theta &\equiv \theta_0, \\ x &= t\cos\theta_0, \\ y &= t\sin\theta_0, \\ z &= 0. \end{aligned}$$

2. If $h_3 \neq 0$, then helices of nonconstant slope:

$$\begin{aligned} \theta &= \theta_0 + h_3 t, \\ x &= (\sin(\theta_0 + h_3 t) - \sin\theta_0)/h_3, \\ y &= (\cos\theta_0 - \cos(\theta_0 + h_3 t))/h_3, \\ z &= (h_3 t - \sin h_3 t)/(2h_3^2). \end{aligned}$$

The sub-Riemannian problem on the Heisenberg group: Geodesics

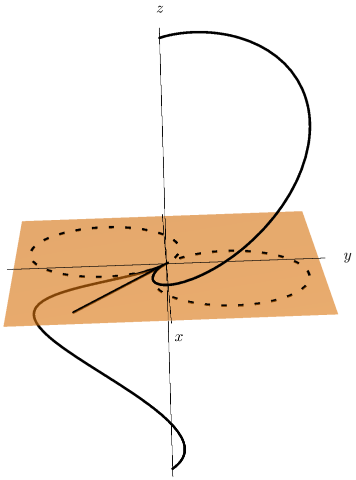


The sub-Riemannian problem on the Heisenberg group: Optimality of geodesics

- Straight lines (case $h_3 = 0$) minimize the Euclidean distance in $\mathbb{R}^2_{x,y}$, thus they are optimal on any segment $t \in [0, t_1]$, $t_1 > 0$.
- Helices (case $h_3 \neq 0$) are not optimal after the first intersection with the z-axis at $t = \frac{2\pi}{|h_3|}$ since these intersections are Maxwell points.
- If $t_1 = \frac{2\pi}{|h_3|}$, then there is a continuum of helices $q(t)$, $t \in [0, t_1]$, coming to the same point $q(t_1)$ at the z-axis; they are obtained one from another by rotations around this axis, thus they all are optimal.
- A part of an optimal arc is optimal, thus the helices are optimal also for $t \in [0, t_1]$, $t_1 \in (0, \frac{2\pi}{|h_3|})$.
- Summing up, the cut time along a geodesic $\mathrm{Exp}(\lambda, t)$ is

$$t_{\mathrm{cut}}(\lambda) = \begin{cases} \frac{2\pi}{|h_3|} & \text{for } h_3 \neq 0, \\ +\infty & \text{for } h_3 = 0. \end{cases} \tag{5}$$

The sub-Riemannian problem on the Heisenberg group: Optimal geodesics



The sub-Riemannian problem on the Heisenberg group: Cut locus and caustic

In Dido's problem the *cut locus*

$$\mathrm{Cut} = \{\mathrm{Exp}(\lambda, t_{\mathrm{cut}}(\lambda)) \mid \lambda \in C\}$$

and the first *caustic*

$$\mathrm{Conj}^1 = \left\{\mathrm{Exp}(\lambda, t^1_{\mathrm{conj}}(\lambda)) \mid \lambda \in C\right\}$$

coincide one with another:

$$\mathrm{Cut} = \mathrm{Conj}^1 = \{(0, 0, z) \in \mathbb{R}^3 \mid z \neq 0\}.$$

The sub-Riemannian problem on the Heisenberg group: Sub-Riemannian distance

Let us describe the *SR distance* $d_0(q) = d(q_0, q)$, $q = (x, y, z) \in \mathbb{R}^3$:

- if $z = 0$, then $d_0(q) = \sqrt{x^2 + y^2}$,
- if $z \neq 0$, $x^2 + y^2 = 0$, then $d_0(q) = \sqrt{2\pi|z|}$,
- if $z \neq 0$, $x^2 + y^2 \neq 0$, then the distance is determined by the conditions

$$d_0(q) = \frac{p}{\sin p}\sqrt{x^2 + y^2},$$

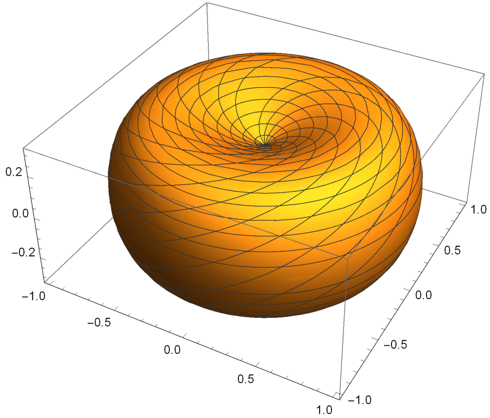
$$\frac{2p - \sin 2p}{4\sin^2 p} = \frac{z}{x^2 + y^2}.$$

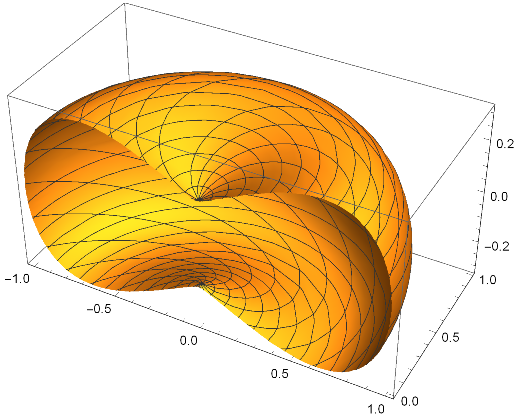
The sub-Riemannian problem on the Heisenberg group: Sub-Riemannian spheres

- The unit *sub-Riemannian sphere* $S = \{q \in \mathbb{R}^3 \mid d_0(q) = 1\}$ is a surface of revolution around the axis z in the form of an apple, see figures at the next slide.
- It has two singular conical points $z = \pm\frac{1}{4\pi}$, $x^2 + y^2 = 0$.
- The remaining spheres $S_R = \{q \in \mathbb{R}^3 \mid d_0(q) = R\}$ are obtained from S by virtue of *dilations*:

$$\delta_s : (x, y, z) \mapsto (e^s x, e^s y, e^{2s} z), \qquad s \in \mathbb{R},$$
$$S_R = \delta_s(S), \qquad s = \ln R.$$

The sub-Riemannian problem on the Heisenberg group: Sub-Riemannian spheres





Exercises

1. Prove that the product

$$(x_1, y_1, z_1) \cdot (x_2, y_2, z_2) = (x_1 + x_2, y_1 + y_2, z_1 + z_2 + (x_1 y_2 - x_2 y_1)/2),$$
$$(x_i, y_i, z_i) \in \mathbb{R}^3, \qquad i = 1, 2,$$

   turns $\mathbb{R}^3$ into a Lie group called the *Heisenberg group*. Show that Dido's problem is left-invariant on this Lie group.

2. Find all conjugate points in Dido's problem.
3. Show that in Dido's problem $d_0 \in C(\mathbb{R}^3)$, but $d_0 \notin C^1(q)$ for any $q = (0, 0, z)$, $z \in \mathbb{R}$.
4. Prove that the sub-Riemannian spheres in Dido's problem are semianalytic (thus subanalytic).