Exponential Mappings for Contact Sub-Riemannian Structures

A. A. Agrachev*

Abstract

On sub-Riemannian manifolds, any neighborhood of any point contains geodesics, which are not length minimizers; the closures of the cut and the conjugate loci of a point q contain q. We study this phenomenon in the case of a contact underlying distribution, essentially in the lowest possible dimension 3, where we extract differential invariants related to the singularities of the cut and the conjugate loci near q and give a generic classification of these singularities.

1 Introduction

1.1 Extremals

Let M be a smooth $(2n+1)$-dimensional manifold. A *contact sub-Riemannian structure* is a pair $\Delta, \langle\cdot|\cdot\rangle$, where $\Delta = \{\Delta_q\}_{q\in M}$, $\Delta_q \subset T_qM$, is a contact structure on M and $\langle\cdot|\cdot\rangle = \{\langle\cdot|\cdot\rangle_q\}_{q\in M}$ is a smooth in q family of Euclidean inner products

$$(v_1, v_2) \mapsto \langle v_1|v_2\rangle_q, \quad v_1, v_2 \in \Delta_q,$$

defined on Δ_q. A Lipschitzian curve $\xi : [0,1] \to M$ is called *admissible* for Δ if $\frac{d\xi(t)}{dt} \in \Delta_{\xi(t)}$ for almost all $t \in [0,1]$. The *length* of an admissble curve ξ is the integral $\int_0^1 |\frac{d\xi}{dt}|\, dt$, where $|v| = \sqrt{\langle v|v\rangle_q}$ $\forall v \in T_qM$. The problem

*Steklov Institute of Mathematics, ul. Vavilova 42, Moscow, 117966, Russia; e-mail: agrachev@mian.su; partially supported by the Russian Foundation for Fundamental Research, grant 95-01-00310 and by INTAS project 93-893

is to characterize length minimizers among all admissible curves connecting two fixed points in M and to extract differential invariants responsible for geometry of the length minimizers. The infimum of the lenghts of admissible curves connecting two points is the *Carnot–Caratheodory distance* between these points.

A nonconstant admissible curve ξ is a *(contact) sub-Riemannian geodesic,* if the restriction of ξ to any small enough subsegment in $[0,1]$ is a length minimizer and $\left|\frac{d\xi(t)}{dt}\right|$ does not depend on t. The first step is to characterize sub-Riemannian geodesics. It can be done with usual optimal control techniques, since we deal here with a rather simple optimal control problem. Take $q_0 \in M$; a standard existence theorem from optimal control theory implies that any point from a neighborhood of q_0 can be connected with q_0 by an admissible length minimizer. We thus should expect a $(2n+1)$-dimensional family of sub-Riemannian geodesics starting at q_0. On the other hand, each geodesic of such a family must be tangent to the $2n$-dimensional subspace Δ_{q_0}. This is the crucial difference of our problem with Riemannian geometry. For some technical reasons, it is convinient to replace the length functional by the functional $\mathcal{J}(\xi) = \frac{1}{2}\int\limits_0^1 |\frac{d\xi}{dt}|^2 dt$. It is easily shown that $\mathcal{J}$-minimizers are exactly the length minimizers parametrized in such a way that $\left|\frac{d\xi(t)}{dt}\right| = const$. In particular, nonconstant $\mathcal{J}$-minimizers are automatically sub-Riemannian geodesics and reparametrized by a homothety small pieces of a sub-Riemannian geodesic are $\mathcal{J}$-minimizers.

We now apply the Pontryagin maximum principle to the 2-point optimal control problem with the cost $\mathcal{J}(\xi)$ and constraints $\dot{\xi}(t) \in \Delta_{\xi(t)}$. The maximum principle gives a perfect characterization of contact sub-Riemannian geodesics since the optimal control problem under consideration admits only normal regular Pontryagin extremals. For $\lambda \in T_q^*M$, set

$$h(\lambda) = \frac{1}{2}(\max\{\langle\lambda, v\rangle : v \in \Delta_q,\ |v| = 1\})^2.$$

The function h is a *sub-Riemannian Hamiltonian.* The Hamiltonian h is a smooth function on the total space of the cotangent bundle T^*M. Set $h_q = h|_{T_q^*M;}$ then h_q is a rank 2n nonnegative quadratic form, $\ker h_q = \Delta_q^{\perp}$. We see that Δ and $\langle\cdot|\cdot\rangle$ can be reconstructed from h : the sub-Riemannian Hamiltonian contains all the information about the sub-Riemannian structure.

The cotangent bundle possesses the standard symplectic structure; hence a Hamiltonian vector field is associated to each smooth function on T^*M. By $\vec{h}$ we denote the Hamiltonian field on T^*M associated to h. Nonconstant trajectories of the Hamiltonian system $\dot{\lambda} = \vec{h}(\lambda)$, $\lambda \in T^*M$, are *(contact) sub-Riemannian extremals.*

Theorem 1.1 *Contact sub-Riemannian geodesics are exactly projections on M of contact sub-Riemannian extremals.*

The formulated theorem is well-known and I have know intention to prove it here, but some remarks could be useful. The theorem is obtained by a compilation of common facts from optimal control. The Pontryagin maximum principle is a necessary optimality condition, therefore any geodesic must satisfy this principle, i.e. must be the projection on M of a Pontryagin extremal in T^*M. In general, Pontryagin extremals can be normal or abnormal. Abnormal extremals appear in many sub-Riemannian structures but not in contact ones; normal Pontryagin extremals are exactly nonconstant trajectories of the Hamiltonian field $\vec{h}$. Moreover, any normal extremal for any sub-Riemannian problem is regular, i.e. satisfies the strengthened Legendre condition. The existence of a regular extremal over an admissible curve implies that small pieces of the curve are locally optimal in the integral $W_{1,1}$-topology (see, for instance, [13]). The global optimality of maybe smaller pieces follows from the fact that the written in coordinates length functional is, in fact, a $W_{1,1}$-norm (= L_1-norm for the velocities of curves).

1.2 Exponential mappings and minimality

Fix an initial point $q_0 \in M$ and consider all the extremals started at $T^*_{q_0}M$. The Hamiltonian field $\vec{h}$ vanishes on $\Delta^\perp$ and has no equilibriums in $T^*M \setminus \Delta^\perp$. Hence extremals started over q_0 are in a one–to–one correspondence with the points of $T^*_{q_0}M \setminus \Delta^\perp_{q_0}$. Let $t \mapsto \lambda(t;\lambda_0)$ be the extremal started at $\lambda_0 \in T^*_{q_0}M \setminus \Delta^\perp_{q_0}$ such that $\lambda(0;\lambda_0) = \lambda_0$. It can be easily checked that

$$\lambda(t;\tau\lambda_0) = \lambda(\tau t;\lambda_0) \quad \forall \tau > 0 \quad \text{and} \quad h(\lambda) = \frac{1}{2}|\pi_*\vec{h}_\lambda|^2 \quad \forall \lambda \in T^*M, \tag{1.1}$$

where $\pi : T^*M \to M$ is the standard projection. In particular, sub-Riemannian geodesics of length ℓ are projections of extremals containing in

$h^{-1}\left(\frac{\ell^2}{2}\right)$. We define an open subset $\mathcal{O}_{q_0}$ in $T^*_{q_0}M \setminus \Delta^{\perp}_{q_0}$ by the following property: $\lambda_0 \in \mathcal{O}_{q_0}$ iff $\lambda(t;\lambda_0)$ is defined for all $t \in [0,1]$. It is clear that $\mathcal{O}_{q_0} \cup \Delta^{\perp}_{q_0}$ is an open neighborhood of the equilibrium set $\Delta^{\perp}_{q_0}$ in $T^*_{q_0}M$.

Let us define the *exponential mapping* $\mathsf{F}_{q_0} : \mathcal{O}_{q_0} \to M$ by the rule $\mathsf{F}_{q_0}(\lambda_0) = \pi(\lambda(1;\lambda_0))$. Thus F_{q_0} takes λ_0 to the right endpoint of the correspondent geodesic. Note that $\mathcal{O}_{q_0}$ is a Lagrangian submanifold of the symplectic manifold T^*M, since it is an open subset of the fiber $T^*_{q_0}M$. It follows that $\lambda(1;\mathcal{O}_{q_0})$ is also a Lagrangian submanifold, since Hamiltonian flows preserve the symplectic structure. Hence the exponential mapping is the composition of a Lagrangian immersion and the standard projection π; we'll use this fact in sections **4**, **5**.

It follows from (1.1) that the curves

$$t \mapsto \mathsf{F}_{q_0}(t\lambda), \quad t \in [0,1] \quad \lambda \in \mathcal{O}_{q_0}, \tag{1.2}$$

are sub-Riemannian geodesics and any started at q_0 sub-Riemannian geodesic has a form (1.2). A number $t,\sqrt{2h(\lambda)} > 0$ is a *conjugate length* for the geodesic (1.2), if t,λ is a critical point of F_{q_0}.

Theorem 1.2 *There is at most finite number of conjugate lengths for any contact sub-Riemannian geodesic. Any free of conjugate lengths contact sub-Riemannian geodesic is a local length minimizer in C^0-topology. A geodesic is not a local length minimizer even in C^∞-topology, if it has a conjugate length that is strictly less than the length of the geodesic.*

The formulated theorem is an exact analog of what is known for classical regular variational problems. For contact sub-Riemannian structures it is also known, although its proof exploits a little bit more sophisticated techniques than the proof of theorem 1.1. There are various approuches and, in my opinion, a natural tool here is symplectic geometry (see [4] for an exposition and [6] for the detailed treatment). The following well-known fact indicates a point, where the analogy with classical Riemannian geometry fails. The reason is the noncompactness of the level sets $h_q^{-1}(l)$.

Proposition 1.1 *For any $l, \varepsilon > 0$ there exists a compact $K \subset\subset h_q^{-1}(l)$ such that the relation $\lambda_0 \in h_q^{-1}(l)\backslash K$ implies that the geodesic (1.2) has a conjugate length less than ε.* $\Box$

Let λ be a critical point of F_{q_0}; we say that the critical value $\mathsf{F}_{q_0}(\lambda)$ belongs to the *conjugate locus* Con_{q_0}, if $t\lambda$ is a regular point of Con_{q_0} $\forall t \in (0,1)$. Thus Con_{q_0} is a subset of M consisting of the points, where started at q_0 geodesics cease to be local length minimizers. It follows from Proposition 1.1 that $q_0 \in \overline{\mathsf{Con}_{q_0}}$.

A number $t_*\sqrt{2h(\lambda_0)}$ is called the *cut length* for the geodesic (1.2), if t_* is the maximal among $\tau > 0$ such that the geodesic $t \mapsto \mathsf{F}_{q_0}(t\tau\lambda_0)$, $t \in [0,1]$ is a global length minimizer. We say that $\mathsf{F}_{q_0}(\lambda)$ belongs to the *cut locus* Cut_{q_0}, if $\sqrt{2h(\lambda)}$ is the cut length for the geodesic $t \mapsto \mathsf{F}_{q_0}(t\lambda)$.

Proposition 1.2 *Suppose $t_*\sqrt{2h(\lambda_0)}$ is the cut length for a geodesic $t \mapsto \mathsf{F}_{q_0}(t\lambda)$; then $\mathsf{F}_{q_0}^{-1}(\mathsf{F}_{q_0}(t\lambda)) \cap h^{-1}(h(t\lambda)) = \{t\lambda\}$ $\forall t \in (0, t_*)$. If any point in a neighborhood of $\mathsf{F}_{q_0}(t_*\lambda)$ can be connected with q_0 by a (global) length minimizer, then either $t_*\sqrt{2h(\lambda)}$ is a conjugate length for the geodesic $t \mapsto \mathsf{F}_{q_0}(t\lambda)$ or it is the cut length for at least one more geodesic with endpoints q_0 and $\mathsf{F}_{q_0}(t_*\lambda)$.*

Proof. Suppose $\exists \tau < t_*, \lambda' \neq \lambda$ such that $h(\lambda') = h(\lambda)$, $\mathsf{F}_{q_0}(\tau\lambda') = \mathsf{F}_{q_0}(\tau\lambda)$;

$$\text{then the broken curve} \qquad t \mapsto \begin{cases} \mathsf{F}_{q_0}(t\lambda') & , \quad 0 \le t \le \tau \\ \mathsf{F}_{q_0}(t\lambda) & , \quad \tau \le t \le t_* \end{cases}$$

is a length minimizer. On the other hand, it can easily be checked that such a curve is not a contact sub-Riemannian geodesic and hence it cannot be a length minimizer. The contradiction proves the first statement of the proposition. Turn now to the second statement. Take $\tau > t_*$; if τ is close enough to t_* then there exists a length minimizer $t \mapsto \mathsf{F}_{q_0}(t\lambda_\tau)$ such that $\mathsf{F}_{q_0}(\tau\lambda_\tau) = \mathsf{F}_{q_0}(\tau\lambda)$, $h(\lambda_\tau) < h(\lambda)$. It follows from theorem 1.2 and proposition 1.1 that λ_τ remains in a compact as $\tau \longrightarrow t_*$; any limiting point of the family λ_τ as $\tau \longrightarrow t_*$ provides us with a required extra geodesic. □

1.3 The outline of the paper

It was already mentioned that for given $q_0 \in M$ there are arbitrary short nonminimizing sub-Riemannian geodesics started at q_0. It implies among other things a nonsmoothness of small Carnot–Caratheodory spheres. In this paper we study the local structure of the set of geodesics. When I say "local", I mean "near q_0": the being investigated local structure in M is determined by a nonlocal behavior of extremals in T^*M. Because of the local (in M) nature of our investigation we may assume that M is diffeomorphic to R^{2n+1} and that all vector fields on M we deal with, are complete vector fields. It is assumed without special mentioning, unless otherwise stated.

In section **2** we construct proper compactifications of the hypersurfaces $h^{-1}(l)$, $l > 0$, such that a $2n$-dimensional sphere is attached "at infinity" to each $h_q^{-1}(l)$, $q \in M$. The 1-foliation of $h^{-1}(l)$ generated by the trajectories of the Hamiltonian field $\vec{h}$ has a smooth extension to the compactified hypersurfaces; attached spheres are invariant for the extended foliation and are foliated by the trajectories of linear Hamiltonian systems with n degrees of freedom associated to positive definite quadratic Hamiltonians. Such a compactification plays a role of a resolution of singularities for the exponential mappings and, in principle, makes it possible to investigate singularities of these mappings. This possibility is realized in subsequent sections for the lowest dimensional case $n = 1$. The greater dimensions are still waiting for their investigators.

In section **3** we study the asymptotics of the conjugate lengths and loci and extract fundamental differential invariants (denoted by χ and κ), which appear as coefficients in the asymptotics. In section **4** we describe the form of the conjugate and cut loci near the initial point for the case $\chi \neq 0$ and finish by this the study of generic germs of sub-Riemannian structures under consideration.

Section **5** is devoted to the case χ vanishing at q_0. An extra differential invariant ρ is then responsible for the form of the conjugate and cut loci near the initial point. We describe the form of the loci in the case $\rho \neq 0$ and prove that generic contact sub-Riemannian structures on 3-dimensional manifolds admit only germs with nonvanishing χ or ρ. Obtained results are also applied to plane isoperimetric problems.

In section **6** we give an interpretation of the invariants in terms of the Carnot–Caratheodory distance. In other words, we define them as intrinsic invariants of the Carnot–Caratheodory metric space, without explicit using of the smooth structure on M. These approach is inspired by the paper [11], where sub-Riemannian structures are treated from the metric point of view. Our intrinsic definitions of the invariants have no Riemannian analogs:

they are completely based on the existence of arbitrary short nonminimizing geodesics started at given point.

We use operator notations and formulas of the chronological calculus for the calculation of asymptotic expansions. I realize that these notations are not common and try to use them only in proves, where they are perhaps inevitable, but not in the formulations of the results. The necessary notations and formulas are listed in the Appendix, see [3, 5] for details.

The form of the conjugate locus near the initial point in the case $\chi \neq 0$ was presented for the first time in May 1994 at the seminar on nonholonomic geometry in ENS, Paris, and then at ICM-94 in Zürich (see [1]); the form of the cut locus in this case was understood soon. Visiting INSA, Rouen, in July 1995 I found out that El Alaoui, Gauthier, and Kupka obtained similar results by other methods. Moreover, at that time they already had a preliminary picture for the conjugate locus in the degenerate case, made with the help of "Mathematica". Further investigation of the degenerate case was made in the intensive exchange by ideas; results were announced in [2]. I am grateful to these colleagues and would be glad to continue the cooperation. The most part of the paper was written, when I was visiting the Mathematics department at the University Aveiro, Portugal. My warmest thanks to the mathematicians of the department for the stimulating friendly atmosphere. I also thanks R. V. Gamkrelidze and A. V. Sarychev for constant support and numerous valuable discussions.

2 Compactification of the space of extremals

2.1 Moving frames

Suppose that an orientation of the contact distribution Δ is fixed. Thus each Δ_q is a $2n$-dimensional oriented Euclidean space.

Lemma 2.1 *There exists a unique contact form ω on M such that $\omega|_{\Delta_q = 0}$ and the form $(\bigwedge^n d\omega)|_{\Delta_q}$ coincides with the volume form on the oriented Euclidean space Δ_q, $\forall q \in M$.*

Proof. The condition $\omega|_{\Delta_q} = 0$ defines a contact form ω up to a nonvanishing multiplier $a \in C^\infty(M)$. We have

$$\bigwedge^n d(a\omega) = a^n \bigwedge^n d\omega + na^{n-1}(\bigwedge^n da) \wedge \omega.$$

Hence $\bigwedge^n d(a\omega)|_{\Delta_q} = a^n \bigwedge^n d\omega|_{\Delta_q}$. $\square$

The form $\Omega_q = d\omega|_{\Delta_q}$ is a symplectic form on the Euclidean space Δ_q. Then

$$\langle \Omega_q, v_1 \wedge v_2 \rangle = \langle \bar{\Omega}_q v_1 | v_2 \rangle_q, \quad \forall v_1, v_2 \in \Delta_q,$$

where $\bar{\Omega}$ is a nondegenerate anti-symmetric operator on Δ_q. Let $\pm ib_1(q), \ldots, \pm ib_n(q)$ be eigenvalues of $\bar{\Omega}_q$, where $b_1(q) \geq \cdots \geq b_n(q) > 0$. The numbers $b_1(q), \ldots, b_n(q)$ are called *fundamental frequences* of the sub-Riemannian structure; they are continuous in q (and smooth, if simple). The normalization condition (Lemma 2.1) implies $\prod\limits_{j=1}^{n} b_j(q) = 1$. So there are only $n-1$ independent frequencies.

Let us give a dual definition of the same frequencies. As any bilinear form on Δ_q, the form Ω can be identified with a skew-adjoint mapping $v \mapsto v \rfloor \Omega$ from Δ_q into the Δ_q^*. The inverse mapping from Δ_q^* into Δ_q defines a symplectic form Ω^{-1} on $\Delta_q^* \cong T_q^*M/\Delta_q^\perp$. Recall that $\Delta_q^\perp$ is the kernel of the quadratic form $h_q = h|_{T_q^*M}$, where h is the sub-Riemannian Hamiltonian. Hence h_q is a correctly defined positively definite quadratic form on the space Δ_q^* endowed with the symplectic structure Ω_q^{-1}. Let $\vec{h}_q^\infty$ be the the linear Hamiltonian field on Δ_q^* associated to the Hamiltonian function h_q; then $b_1(q), \ldots, b_n(q)$ are just fundamental frequencies of $\vec{h}_q^\infty$. In other words, there exists a basis $v_q^1, \ldots, v_q^{2n}$ of Δ_q^* such that

$$\Omega_q = \sum_{i=1}^{n} b_i v_q^i \wedge v_q^{n+i}, \quad h(\sum_{j=1}^{2n} u_j v_q^j) = \frac{1}{2} \sum_{i=1}^{n} (u_i^2 + u_{n+i}^2).$$

Then

$$\vec{h}_q^\infty = \sum_{i=1}^{n} b_i (u_{n+i} \partial_{u_i} - u_i \partial_{u_{n+i}}).$$

The *Stokes field* e associated with the contact form ω is a smooth vector field on M uniquely defined by the relations $e \rfloor d\omega = 0,\ e \rfloor \omega = 1$. The Stokes field is transversal to Δ. Hence $T^*M = \Delta^\perp \oplus e^\perp$ and there are canonical identifications $e_q^\perp \cong \Delta_q^*,\ T_q^*M \cong \Delta_q^\perp \oplus \Delta_q^*$. We shall use these identifications without special mentioning. In particular, $\omega_q, v_q^1, \ldots, v_q^{2n}$ can be chosen smooth in q such that the 1-forms $\omega, v^1, \ldots, v^{2n}$ define a trivialization of the bundle T^*M.

The "tautological" 1-form ς on T^*M has the expression

$$\varsigma = \nu\omega + \sum_{j=1}^{2n} u_j \upsilon^j,$$

where $(\nu, u_1, \ldots, u_{2n}) = (\nu, u)$ are coordinates on the fibers of T^*M defined by the trivialization. Further,

$$d\omega = \sum_{i=1}^{n} b_i \upsilon^i \wedge \upsilon^{n+i}, \tag{2.1}$$

$$d\upsilon^j = \sum_{k=1}^{n} c_{0k}^j \omega \wedge \upsilon^k + \sum_{1\leq k<l\leq n} c_{kl}^j \upsilon^k \wedge \upsilon^l, \tag{2.2}$$

where smooth functions $c_{0k}^j,\ c_{kl}^j$ are *structural constants* associated to the frame $\omega, \upsilon^1, \ldots, \upsilon^{2n}$. Differentiation of the equations (2.1), (2.2) gives *structural equations* for the structural constants. In particular, we derive from (2.1)

$$\begin{gathered} c_{0j}^{n+i} - c_{0i}^{n+j} = c_{0n+j}^{i} - c_{0n+i}^{j} = 0,\ c_{0j}^{i} + c_{0n+i}^{n+j} = 0 \text{ for } i \neq j, \\ c_{0i}^{i} + c_{0n+i}^{n+i} = -eb_i, \quad i, j = 1, \ldots, n, \end{gathered} \tag{2.3}$$

and similar relations for other structural constants.

Standard symplectic structure $\sigma = d\varsigma$ on T^*M and the sub-Rimannian Hamiltonian function h have a form

$$\sigma = d\nu \wedge \omega = \sum_{j=1}^{2n} du_j \wedge \upsilon^j + \nu d\omega + \sum_{j=1}^{2n} u_j d\upsilon^j; \quad h = \frac{1}{2}\sum_{j=1}^{2n} u_j^2.$$

We shall consider one more Hamiltonian function

$$e^* : \lambda \mapsto \langle \lambda, e_q \rangle, \quad \forall \lambda \in T_q^*M,\ q \in M;$$

then $e^* = u_0$. Let $\{h, e^*\} = \vec{h}e^* = -\vec{e^*}h$ be the Poisson brackets of Hamiltonians h, e^*. Equations (2.1), (2.2) and the relation $\vec{h}\rfloor = -dh$ imply

$$\{h, e^*\} = \sum_{j,k=1}^{2n} u_j u_k c_{0k}^j, \tag{2.4}$$

$$\vec{h} = f(u) + \{h, e^*\}\partial_\nu + \nu \vec{h}_q^\infty - \sum_{j=1}^{2n} u_j \sum_{1\le k<l\le n} c_{kl}^j \partial_{\theta_{kl}}, \tag{2.5}$$

where $f(u) = \sum\limits_{j=1}^{2n} u_j f^j$ and vector fields $f^1, \dots, f^{2n}$ form the basis of Δ dual to $\upsilon^1, \dots, \upsilon^{2n}$ (i.e. $\langle \upsilon^i, f^j \rangle = \delta_{ij}$); $\partial_{\theta_{kl}} = u_k \partial_{u_l} - u_l \partial_{u_k}$ are vector fields generating rotations in the planes $span\{\upsilon^k, \upsilon^l\}$.

Note that the fields $e, f^1, \dots, f^{2n}$ satisfy structural equations dual to (2.1), (2.2):

$$[f^j, e] = \sum_{k=1}^{2n} c_{0j}^k f^k \tag{2.6}$$

$$[f^j, f^i] = \sum_{k=1}^{2n} c_{ij}^k f^k + b_{ij} e \tag{2.7}$$

for $1 \le j < i \le 2n$, where we put $b_{ij} = 0$ for $j \ne n+i$, $b_{in+i} = b_i$.

2.2 Principle bundles

Classical Nöther theorem implies that the sub-Riemannian structure is invariant under the action of 1-parametric group generated by e iff $\{e^*, h\} = 0$. Let $\mathcal{E}$ be the 1-foliation in M generated by the vector field e. Assume that $M/\mathcal{E}$ is a smooth manifold and the canonical projection $M \to M/\mathcal{E}$ defines a principle bundle with a 1-dimensional structure group generated by e. If $\{e^*, h\} = 0$, then:

i) Δ is a connection on this bundle;

ii) the sub-Riemannian structure on M defines a Riemannian structure on $M/\mathcal{E}$;

iii) the n-th exterior power of the curvature form associated to the connection Δ coincides with a volume form on $M/\mathcal{E}$ defined by given Riemannian structure.

Conversely, we can start with a connection on a line or circle principle bundle over an arbitrary $2n$-dimensional Riemannian manifold. The connection is a distribution that is transversal to fibers and invariant under the action of the structure group. It is a contact distribution, iff the curvature 2-form associated to the connection is nondegenerate. The Riemannian structure on M induces a sub-Riemannian structure on this distribution. The identity $\{e^*, h\} = 0$ is valid for this sub-Riemannian structure, iff the curvature form R satisfies condition iii). In the last case the structure group coincides with the group generated by e.

2.3 Behavior at infinity

Let $S_q \subset \Delta_q^*$ be the unit sphere with respect to the norm in Δ_q^* induced by the Euclidean norm in Δ_q; then $h_q^{-1}(\frac{1}{2}) \cong S_q \times R\omega_q \cong S_q \times R$. This means that $h^{-1}(\frac{1}{2})$ is a fiber bundle over M with fibers naturally diffeomorphic to $S_q \times R$. The standard projective compactification $R \hookrightarrow RP^1$ results in a compactification $S_q \times R \hookrightarrow S_q \times RP^1$ and, finally, in a compactification $h^{-1}(\frac{1}{2}) \hookrightarrow G$, where G is a fiber bundle over M with fibers G_q naturally diffeomorphic to $S_q \times RP^1 = (S_q \times R) \cup (S_q \times \infty)$.

Trajectories of the Hamiltonian field $\vec{h}$ that are contained in $h^{-1}(\frac{1}{2})$ are exactly sub-Riemannian extremals parametrized by the length of their projections in M. Neglecting the parametrization we obtain a 1-foliation $\mathcal{H}$ of $h^{-1}(\frac{1}{2})$, where leaves of $\mathcal{H}$ are unparametrized extremals.

Proposition 2.1 *There exists a unique smooth 1-foliation $\overline{\mathcal{H}}$ of G such that $\overline{\mathcal{H}}|_{h^{-1}(\frac{1}{2})} = \mathcal{H}$. Spheres at infinity $S_q \times \infty$ are invariant submanifolds for $\overline{\mathcal{H}}$. The foliation $\overline{\mathcal{H}}|_{S_q \times \infty}$ is generated by the field $\vec{h}_q^\infty$ restricted to the sphere S_q.*

Proof. Set $w = -\frac{1}{\nu}$; then the triple $(w, u, q),\ w \in R,\ u \in S^{2n-1},\ q \in M$, defines "coordinates" on a neighborhood of the submanifold $G \setminus h_q^{-1}(\frac{1}{2})$ of G (i.e. on a neighborhood of the infinity). The infinity $G \setminus h^{-1}(\frac{1}{2})$ is defined by the equation $w = 0$. We have

$$\vec{h} = f(u) + w^2\{h, e^*\}\partial_w - \frac{1}{w}\vec{h}_q^\infty - \sum_{j=1}^{2n} u_j \sum_{1 \le k < l \le n} c_{kl}^j \partial_{\theta_{kl}}. \tag{2.8}$$

The vector field $w\vec{h} = \vec{h}_q^\infty + O(w)$ (as $w \longrightarrow 0$) generates the restiction of the foliation $\mathcal{H}$ to the neighborhood of the infinity. The statement of the proposition is an immediate consiquence of the fact that $\vec{h}_q^\infty$ has no equilibriums.

3 Three-dimensional structures

3.1 Rescaling

Let us come back to the mapping $\mathsf{F}_{q_0} : \mathcal{O}_{q_0} \to M$, $\mathcal{O}_{q_0} \subseteq T^*_{q_0}M \setminus \Delta^{\perp}_{q_0}$. Any vector $\lambda \in T^*_{q_0}M \setminus (\Delta^{\perp}_{q_0} \cup e^{\perp}_{q_0})$ has a form

$$\lambda = \sqrt{2h(\lambda)}(\frac{1}{w_\lambda}\omega + \upsilon_\lambda), \quad \text{where } w_\lambda \in R \setminus 0, \upsilon_\lambda \in S_{q_0}.$$

Proposition 2.1 implies that $\mathsf{F}_{q_0}(\lambda)$ tends to q_0 as $w_\lambda \longrightarrow 0$, if $h(\lambda)$ remains uniformly bounded. The mapping $\lambda \mapsto (h(\lambda), \upsilon_\lambda, w_\lambda)$ gives an identification

$$T^*_{q_0}M \setminus (\Delta^{\perp}_{q_0} \cup e^{\perp}_{q_0}) \cong R_+ \times S_{q_0} \times (R \setminus 0). \tag{3.1}$$

It follows from (2.8) that the mapping $\lambda \mapsto \mathsf{F}_{q_0}(w_\lambda \lambda)$ originally defined on a domain in (3.1), has a smooth extension to a bigger domain includin $S_{q_0} \times \infty$. Moreover, it follows that the above mapping admits an explicit asymptotic expansion in a power series in w as $w \longrightarrow 0$. Such an expansion of the exponential mapping gives a base for the investigating of the singularity of the mapping at q_0, although explicit expressions are very involved. We focus our attention on a lowest dimensional case $n = 1$, $\dim M = 3$ and use coordinate free notations and formulas from the chronological calculus (see Appendix) in order to simplify and clarify calculations.

First of all, for $n = 1$ identities (2.6), (2.7), (2.3) take a form

$$[f^j, e] = c^1_{0j} f^1 + c^2_{0j} f^2 \tag{3.2}$$

$$[f^2, f^1] = e + c^1_{12} f^1 + c^2_{12} f^2 \tag{3.3}$$

$$c^1_{01} + c^2_{02} = 0. \tag{3.4}$$

Let θ be an angle coordinate on the oriented circle S^1; then $\partial_\theta = \partial_{\theta_{12}} = -\vec{h}^{\infty}_q$ (see subsection 2.1). Thus

$$\vec{h} = f(u) + w^2\{h, e^*\}\partial_w + \frac{1}{w}(1 - w(c^1_{12}u_1 + c^2_{12}u_2))\partial_\theta.$$

Let us consider the vector field

$$\frac{w}{1 - w(c^1_{12}u_1(\theta) + c^2_{12}u_2(\theta))}(wf(u(\theta)) + w^3\{h, e^*\}\partial_w) + \partial_\theta. \tag{3.5}$$

Trajectories of the field (3.5) are just reparametrizations of the extremals.

Suppose $t \mapsto (q(t;\theta,\eta), w(t;\theta,\eta))$ is the solution of the system

$$\left\{ \begin{array}{rcl} \dot{q} & = & \frac{w}{1-w(c_{12}^1 u_1 + c_{12}^2 u_2)} f(u(t)) \\ \dot{w} & = & \frac{w^3}{1-w(c_{12}^1 u_1 + c_{12}^2 u_2)} \{h, e^*\} \end{array} \right. \quad q(\theta;\theta,\eta) = q_0,\ w(\theta;\theta,\eta) = \eta, \tag{3.6}$$

where $u_1(t) = \cos t, u_2(t) = \sin t$.

In further calculations we omit arguments $\theta,\ \eta$ that are situated after ";", if it does not lead to collisions. Let us introduce simplified notations

$$a(t) = \{h, e^*\}|_{u(t)} = \sum_{j,k=1}^{2} c_{0k}^j u_j(t) u_k(t), \quad b(t) = c_{12}^1 u_1 + c_{12}^2 u_2;$$

thus $a(t),\ b(t)$ are smooth functions on M, if t is fixed.

Set $F(t,\theta,\eta) = q(t+\theta;\theta,\eta)$,

$$\begin{array}{rcl} F(t,\theta,\eta) & = & \mathsf{F}_{q_0}\left(\int\limits_{\theta}^{\theta+t} \frac{w(\tau)}{1-w(\tau)b(\tau)}\, d\tau f(u(\theta)) - \frac{1}{\eta}\omega \right) \\ & = & \mathsf{F}_{q_0}(\eta(t+O(\eta))(f(u(\theta)) - \frac{1}{\eta}\omega)) \end{array} \tag{3.7}$$

as $\eta \longrightarrow 0$, if $t \geq 0$ remains uniformly bounded.

3.2 Conjugate lengths

For given $q_0 \in M,\ \upsilon \in S_{q_0}$, consider the family of geodesics

$$\gamma_\nu : \tau \mapsto \mathsf{F}_{q_0}(\tau(\upsilon + \nu\omega_{q_0})).$$

Recall that a number $\tau > 0$ is a conjugate length for γ_ν, if $\tau(\upsilon + \nu\omega_{q_0})$ is a critical point of F_{q_0}. Conjugate lengths for given geodesic are isolated in R and thus linearly ordered. We denote by $\ell_m(\nu)$ the m-th conjugate length for γ_ν.

Theorem 3.1 *For any $\varepsilon > 0$, the number of conjugate lengths for γ_ν that are contained in $(0,\varepsilon)$ tends to infinity with $|\nu|$ and the following asymptotic relations hold as $\nu \longrightarrow \pm\infty$:*

$$|\nu|\ell_m(\nu) = \tau_m + O(\frac{1}{\nu}),$$

where τ_m is the m-th positive root of the equation $\tau \sin\tau + 2\cos\tau = 2$, $\tau_{2k-1} = 2\pi k$;

$$\gamma_\nu(\ell_m(\nu)) = q_0 + \frac{\sqrt{\tau_m \sin\tau_m}}{|\nu|} v_m^\pm + O(\frac{1}{\nu^2}),$$

where $v_m^\pm$ is the result of the rotation of the vector $\dot\gamma_\nu(0)$ on the angle $\pm \arccos\sqrt{\frac{\sin\tau_m}{\tau_m}}$ in the oriented Euclidean plane Δ_{q_0};

$$|\nu|\ell_{2k-1}(\nu) = 2\pi k + O(\frac{1}{\nu^2}), \quad \gamma_\nu(\ell_{2k-1}(\nu)) = q_0 \pm \frac{k\pi}{\nu^2} e_{q_0} + O(\frac{1}{\nu^3}).$$

Proof. The first statement of the theorem is a corollary of (3.7) and the following lemma. Starting with this lemma we shall make calculations only for positive $\eta = -\frac{1}{\nu}$. One can reduce the case of negative η to the positive one just replacing e with $-e$ and changing the orientation of Δ_q.

Lemma 3.1 *Let us consider the equation in t*

$$\frac{\partial F}{\partial t}(t,\theta,\eta) \wedge \frac{\partial F}{\partial \theta}(t,\theta,\eta) \wedge \frac{\partial F}{\partial \eta}(t,\theta,\eta) = 0, \tag{3.9}$$

where $\theta \in S^1$, $\eta \in R \setminus 0$ are parameters. For given θ and integral number $m > 0$ denote by $t_m(\eta)$ the m-th positive root of the equation (3.9), if such a root exist. Then $t_m(\eta)$ exists for all small enough η and

$$t_m(\eta) = \tau_m + O(\eta) \qquad (\eta \longrightarrow 0).$$

Proof. Consider the nonstationary vector field on M

$$\frac{w(t)}{1 - w(t)b(t)} f(u(t)) = (\eta + \eta^2 c) f(u(t)), \tag{3.10}$$

where c is a smooth scalar function of (t, η, q).

Let $Q_\theta^t(\eta) = \overrightarrow{\exp} \int\limits_\theta^{\theta+t} \frac{w(\tau)}{1-w(\tau)b(\tau)} f(u(\tau))\, d\tau$ be the flow in M generated by the field (3.10); then $F(t,\theta,\eta)$ is the diffeomorphism $Q_\theta^t(\eta)$ evaluated at q_0; in other words, $F(t,\theta,\eta) = q_0 Q_\theta^t(\eta)$ (see appendix). We have (putting $Q = Q_\theta^t(\eta)$ for the sake of simplicity)

$$\begin{aligned} \tfrac{\partial}{\partial t} F(t,\theta,\eta) &= q_0 Q \circ ((\eta + \eta^2 c) f(u(\theta + t))) \\ &= q_0 Ad Q\, ((\eta + \eta^2 c) f(u(\theta + t))) \circ Q, \end{aligned}$$

$$\frac{\partial}{\partial\theta}F(t,\theta,\eta)=\frac{\partial}{\partial t}F(t,\theta,\eta)-(\eta+\eta^2q_0c)q_0f(u(\theta))\circ Q,$$

$$\frac{\partial}{\partial\eta}F(t,\theta,\eta)=\int\limits_0^t q_0AdQ_\theta^\tau(\eta)\left((1+2\eta c(\theta+\tau,\eta))f(u(\theta+\tau))\right)\,d\tau\circ Q.$$

Recall that $v\circ Q=Q_*v$ for any tangent vector v and

$$AdQ=id+\int\limits_0^t AdQ_\theta^\tau(\eta)\circ ad((\eta+\eta^2)f(u(\theta+\tau)))\,d\tau$$

(see appendix). Besides, $q_0f(u)=u_1q_0f(u)+u_2q_0f(u)$ belongs to the 2-dimensional subspace Δ_{q_0}, $\forall u$.

Denote by $\phi(t,\theta,\eta)$ the left–hand side of (3.9). A strightforward calculation shows that $\frac{\partial^i}{\partial t^i}\phi(t,\theta,\eta)|_{t=0}=0$ for $0\le i\le 3$. Hence $\phi(t,\theta,\eta)=t^4\hat\phi(t,\theta,\eta)$, where $\hat\phi$ is a smooth function. Then the Taylor–series expansion of ϕ as a function of η gives

$$\phi(t,\theta,\eta)=\eta^3\Bigg(q_0[\int\limits_\theta^{\theta+t}f(u(\tau))\,d\tau,f(u(\theta+t))]\wedge q_0f(u(\theta))\wedge\int\limits_\theta^{\theta+t}q_0f(u(\tau))\,d\tau$$
$$+q_0f(u(\theta+t))\wedge q_0f(u(\theta))\wedge\int\limits_\theta^{\theta+t}q_0[\int\limits_\theta^{\tau}f(u(\tau'))\,d\tau',f(u(\tau))]\,d\tau\Bigg)+O(\eta^4t^4).$$

Apply (3.3) and reduce (3.9) to the equation

$$0=\Big(\int\limits_\theta^{\theta+t}(u_2(\tau)u_1(\theta+t)-u_1(\tau)u_2(\theta+t))\,d\tau\int\limits_\theta^{\theta+t}(u_1(\theta)u_2(\tau)-u_2(\theta)u_1(\tau))\,d\tau$$
$$+\int\limits_\theta^{\theta+t}(\int\limits_\theta^{\tau}u_2(\tau')\,d\tau'u_1(\tau)-\int\limits_\theta^{\tau}u_1(\tau')\,d\tau'u_2(\tau))\,d\tau(u_1(\theta+t)u_2(\theta)$$
$$-u_2(\theta+t)u_1(\theta))+O(\eta t^4)\Big)q_0e\wedge q_0f^1\wedge q_0f^2.$$

Since $u_1(\tau)=\cos\tau, u_2(\tau)=\sin\tau$, and $q_0e\wedge q_0f_1\wedge q_0f_2\neq 0$, we obtain after the integrating

$$0=2\cos t-2+t\sin t+O(\eta t^4).$$

The statement of lemma 3.1 follows from the fact that the last equation restricted to $\eta=0$ has only simple positive roots while 0 is a root of order 4.

A simple analysis of the equation (3.8) shows that $\tau_{2k-1}=2\pi k$, $\frac{(4k+3)\pi}{2}<\tau_{2k}<2(k+1)\pi$, and $(2(k+1)\pi-\tau_{2k})\longrightarrow 0$ as $k\longrightarrow\infty$.

We have

$$\gamma_\nu(\ell_m(\nu)) = \int\limits_\theta^{\theta+\tau_m} \eta q_0 f(u(\tau))\, d\tau + O(\eta^2).$$

The strightforward integration gives

$$\gamma_\nu(\ell_m(\nu)) = \eta\sqrt{\tau_m \sin\tau_m} U_m q_0 f(u(\theta)) + O(\eta^2).$$

It remains to prove the last statement of the theorem concerning $\ell_{2k-1}(\nu)$ and
$\gamma_\nu(\ell_{2k-1}(\nu))$. We have $\int\limits_\theta^{\theta+2\pi k} q_0 f(u(t))\, dt = 0$. Hence

$$F(\theta+2\pi k, \theta, \eta) = q_0 Q_\theta^{\theta+2\pi k}(\eta) =$$
$$\eta^2 \int\limits_\theta^{\theta+2\pi k} \left(b(t) q_0 f(u(\tau)) + q_0 [\int\limits_\theta^t f(u(\tau))\, d\tau, f(u(t))] \right) dt + O(\eta^3).$$

Now apply (3.3) and obtain

$$F(2\pi k, \theta, \eta) = -\eta^2 e_{q_0} + O(\eta^3). \tag{3.11}$$

In particular, the right–hand side of (3.9) evaluated at $t = 2\pi k$ is $O(\eta^5)$. Let $t_m(\eta)$ be the m-th positive root of equation (3.9). Since (3.9) divided by η^3 and then restricted to $\eta = 0$ is a regular equation in positive t, we obtain

$$t_{2k-1}(\eta) = 2\pi k + O(\eta^2).$$

The last estimate together with (3.11) complete the proof of the theorem.
Remark 1. It follows from the proof that $\ell_m(\nu)$ are smooth in ν for $|\nu|$ big enough.
Remark 2. Conjugate lengths for left–invariant sub-Riemannian structures on three-dimensional Lie groups were computed in the paper [14]. We are forced to certify that all the lengths $\ell_{2k}(\nu),\ k = 1, 2, \dots,$ where lost there.

3.3 Principle invariants

We now focus on the first conjugate lengths and start with a useful symmetry property of F.

Proposition 3.1 *The following identity holds:*

$$F(t,\theta,\eta) \equiv F(t,\theta+\pi,-\eta). \tag{3.13}$$

Proof. Recall that $F(t,\theta,\eta) = q(t+\theta;\theta,\eta)$, where $t \mapsto q(t+\theta;\theta,\eta)$ is a solution to (3.6), $q(\theta;\theta,\eta) = q_0$. We have $u_i(t+\pi) = -u_i(t),\ i = 1,2,\ldots$. Applying the transformation $(t,q,w) \mapsto (t+\pi,q,-w)$ to (3.6) we realize that the right–hand side of the equation for $\dot q$ is preserved and the right–hand side of the equation for $\dot w$ changes the sign. Let $t \mapsto (q(t),w(t))$ be a solution to (3.6). It follows that $t \mapsto (q(t-\pi),-w(t-\pi))$ is again a solution to (3.6) and (3.13) is satisfied. $\square$

Suppose local coordinates of M are fixed in a neighborhood of q_0 and

$$F(2\pi,\theta,\eta) \approx \sum_{n=0}^{\infty} \eta^n q_n(\theta) \tag{3.14}$$

is the Taylor–series expansion of the vector-function $\eta \mapsto F(2\pi,\theta,\eta)$. In particular, $q_1(\theta) = 0$, $q_2(\theta) = -\pi e_{q_0}$, as it follows from theorem 3.1.

Corollary. $q_n(\theta+\pi) = (-1)^n q_n(\theta)$, $q_{2n+1}(\theta) = -\frac{1}{2}\int\limits_{\theta}^{\theta+\pi} \frac{dq_{2n+1}}{d\tau}(\tau)\,d\tau$, *for* $n = 0,1,\ldots$, $\theta \in S^1$.

We are now going to give an invariant expression for $\frac{dq_3}{d\theta}$. Before doing it, let us introduce a fundamental differential invariant of the contact sub-Riemannian structure.

By $\{h,e^*\}_{q_0}$ we denote the restriction of the function $\{h,e^*\}$ to the fiber at q_0. This restriction is a quadratic form and, according to (2.4), ω_q belongs to the kernel of this form. Hence $\{h,e^*\}_{q_0}$ is actually a well–defined quadratic form on the Euclidean plane $\Delta^*_{q_0} \cong \Delta_{q_0}$,

$$\{h,e^*\}_{q_0}(u) = \sum_{j,k=1}^{2} u_j u_k c^j_{0k}(q_0).$$

Quadratic form on the Euclidean space is nothing else but a symmetric operator on the same space. In particular, the trace and the determinant of $\{h,e^*\}_{q_0}$ are correctly defined. It follows from (3.4) that $tr\{h,e^*\} = 0$ and the form $\{h,e^*\}_{q_0}$ is a hyperbolic or null one. By $\chi(q_0)$ we denote the positive eigenvalue, $\chi(q_0) = \sqrt{-det\{h,e^*\}_{q_0}} = \max_{|u|=1}\{h,e^*\}_{q_0}(u)$; it is the only invariant of the hyperbolic form on the Euclidean plane.

The function χ is an important differential invariant of the sub-Riemannian structure. For instance, if $\chi \equiv 0$, then $\{h, e^*\} = 0$ and we are in the situation described in subsection **2.2**.

Proposition 3.2 *The following asymptotic relation is valid:*

$$\frac{\partial F}{\partial \theta}(2\pi, \theta, \eta) = \eta^3 \pi [f(u(\theta)), e] + O(\eta^4).$$

Proof. We have

$$\frac{\partial F}{\partial \theta}(2\pi, \theta, \eta) = q_0 Q_\theta^{2\pi}(\eta) \circ \left(\frac{w(\theta + 2\pi)}{1 - w(\theta + 2\pi) b(\theta)} f(u(\theta)) \right) -$$

$$-q_0 \left(\frac{\eta}{1 - \eta b(\theta)} f(u(\theta)) \right) \circ Q_\theta^{2\pi}(\eta) =$$

$$= q_0 \left(Ad Q_\theta^{2\pi}(\eta) \left(\frac{w(\theta + 2\pi)}{1 - w(\theta + 2\pi) b(\theta)} f(u(\theta)) \right) - \frac{\eta}{1 - \eta b(\theta)} f(u(\theta)) \right) \circ Q_\theta^{2\pi}(\eta).$$

Then $Q_\theta^{2\pi}(\eta) = id - \eta^2 \pi e + O(\eta^3)$;

$$Ad Q_\theta^{2\pi}(\eta) = id - \eta^2 \pi ad\, e + O(\eta^3);$$

$$w(\theta + 2\pi) = \eta + \eta^3 \int_\theta^{\theta + 2\pi} \{h, e^*\}_{q_0}(u(\tau))\, d\tau + O(\eta^4) = \eta + O(\eta^4)$$

Hence

$$\frac{\partial F}{\partial \theta}(2\pi, \theta, \eta) = -\eta^3 \pi q_0 [e, f(u)] + O(\eta^4).$$

Our next goal is the cubic term in the asymptotic expansion for the conjugate locus and for the 1st conjugate length. Consider the geodesics

$$\gamma_{\theta,\nu} : \tau \mapsto F_{q_0}(\tau(\upsilon_\theta + \nu \omega_{q_0})), \quad \theta \in S^1,\ \nu \in R,$$

where $\upsilon_\theta = \sum u_j(\theta) \upsilon_{q_0}^j$. Thus $\dot\gamma_{\theta,\nu}(0) = f_{q_0}(u(\theta))$. By $\ell_1(\theta; \nu)$ we denote the first conjugate length for $\gamma_{\theta,\nu}$, appending an argument θ to the old notation. Set also $Con_{q_0}(\theta, \nu) = \gamma_{\theta,\nu}(\ell_1(\theta; \nu))$.

Theorem 3.2 *The conjugate locus and the 1st conjugate length admit the following asymptotic expansion as* $\nu \longrightarrow \pm\infty$:

$$Con_{q_0}(\theta, \nu) = q_0 \pm \frac{\pi}{\nu^2} e_{q_0} \pm \frac{3\pi}{2\nu^3} \int_{\theta}^{\theta+\pi} \{h, e^*\}_{q_0}(u(\tau)) f_{q_0}(u(\tau))\, d\tau + O(\frac{1}{\nu^4}),$$

$$\ell_1(\theta; \nu) = \frac{2\pi}{|\nu|} - \frac{\pi\kappa(q_0)}{|\nu|^3} + O(\frac{1}{\nu^4}),$$

where

$$\kappa = f^2 c^1_{12} - f^1 c^2_{12} - (c^1_{12})^2 - (c^2_{12})^2 + \frac{c^2_{01} - c^1_{02}}{2}$$

is a smooth function on M.

Remark. The function κ is the second, additional to χ, differential invariant of the sub-Riemannian structure. If $\chi \equiv 0$, then κ is constant along the trajectories of e and thus defines a function on $M/\mathcal{E}$ (see subsection **1.2**). In this case, it can be easily checked that κ is just the Gaussian curvature of the Riemannian surface $M/\mathcal{E}$. If both χ and κ vanish identically, then we obtain a "flat" contact sub-Riemannian structure locally isometric to a left–invariant structure on the Heisenberg group.

Proof. The point $Con(\theta, \nu)$ is a critical value of the mapping F, assigned to a critical point (t_1, η, ν), where $\eta = -\frac{1}{\nu}$ and interval $\{(\tau, \theta, \nu) : 0 < \tau < t_1\}$ contains only regular points of F. We know from the previous calculations that $t_1 = 2\pi + \eta^2 s(\theta) + O(\eta^3)$, where s is smooth in θ. The function $s(\theta)$ is the solution of the following equation evaluated at $\eta = 0$ (first solve, then evaluate):

$$0 = \frac{\partial F}{\partial s}(2\pi + \eta^2 s, \theta, \eta) \wedge \frac{\partial F}{\partial \theta}(2\pi + \eta^2 s, \theta, \eta) \wedge \frac{\partial F}{\partial \eta}(2\pi + \eta^2 s, \theta, \eta).$$

We have

$$\frac{\partial F}{\partial s}(2\pi + \eta^2 s, \theta, \eta) = \eta^3 f_{q_0}(u(\theta)) + O(\eta^4);$$

$$\frac{\partial F}{\partial \theta}(2\pi + \eta^2 s, \theta, \eta) = \eta^3(\pi[f_{q_0}(u(\theta)), e]_{q_0} + s(\theta) f(\frac{du}{d\theta})) + O(\eta^4);$$

$$\frac{\partial F}{\partial \eta}(2\pi + \eta^2 s, \theta, \eta) = \eta 2\pi e_{q_0} + O(\eta^2).$$

Set $\dot{u} = \frac{du}{d\theta},\ \dot{s} = \frac{ds}{d\theta}$. We obtain

$$0 = f(u) \wedge (sf(\dot{u}) + \pi[f(u), e]) \wedge e.$$

Fields $f(u), f(\dot{u})$ form an orthonormal frame in Δ and e is transversal to Δ. Hence $s = \pi\langle [e, f(u)] | f(\dot{u})\rangle$ and

$$\dot{s} = \pi\langle [e, f(\dot{u})] | f(\dot{u})\rangle - \pi\langle [e, f(u)] | f(u)\rangle,$$

since $f(\ddot{u}) = -f$. Then

$$F(2\pi + \eta^2 s, \theta, \eta) = F(2\pi, \theta, \eta) + \eta^3 s(\theta) f_{q_0}(u(\theta)) + O(\eta^4) = -\eta^2 \pi e_{q_0} -$$

$$-\frac{\eta^3}{2} \int\limits_{\theta}^{\theta+\pi} \left(\frac{\partial F}{\partial \tau}(2\pi, \tau, \eta) + \dot{s}(\tau) f_{q_0}(u(\tau)) + s(\tau) f_{q_0}(\dot{u}(\tau)) \right) d\tau = -\eta^2 \pi e_{q_0} -$$

$$-\frac{\eta^3 \pi}{2} \int\limits_{\theta}^{\theta+2\pi} \left(\langle [e, f(\dot{u}(\tau))] | f(\dot{u}(\tau))\rangle + 2\langle [f(u(\tau)), e] | f(u(\tau))\rangle \right) f_{q_0}(u(\tau))\, d\tau.$$

On the other hand,

$$\langle [e, f(\dot{u})] | f(\dot{u})\rangle = \langle [f(u), e] | f(u)\rangle = \sum_{j,k=1}^{2} c_{0k}^{j} u_k u_j = \{h, e^*\}$$

and we obtain the desired formula for $Con(\theta, \nu)$. It remains to compute the asymptotics of the conjugate length. In virtue of (3.6) (see also notations after that equation) we obtain

$$\ell_1(\theta; \eta) = \int\limits_{\theta}^{\theta+t_1} \frac{w(t)}{1 - w(t) Q_\theta^t(\eta) b(t)}\, dt = \eta 2\pi +$$

$$\eta^3 \left(\int\limits_{\theta}^{\theta+2\pi} \Big(\int\limits_{\theta}^{t} a(\tau)\, d\tau + \int\limits_{\theta}^{t} f(u(\tau))\, d\tau b(t) + b^2(t) \Big) dt + s(\theta) \right) + O(\eta^4) =$$

$$= \eta 2\pi + \eta^3 \pi \Big(f^1 c_{12}^2 - f^2 c_{12}^1 + \frac{c_{02}^1 - c_{01}^2}{2} + (c_{12}^1)^2 + (c_{12}^2)^2 \Big) + O(\eta^4).$$

4 Conjugate and cut loci

4.1 Conjugate locus

Assume that $\chi(q_0) \neq 0$; generic germs of contact sub-Riemannian structures enjoy this property. Until now, we were working with an arbitrary orthonormal frame f^1, f^2 in Δ. The assumption on χ permits to distinguish a privileged frame, which is defined up to a sign. Namely, the hyperbolic form $\{h, e^*\}_{q_0}$ has 2 orthogonal isotropic lines. We fix the frame up to a sign by the requirement that $\upsilon^1_{q_0}, \upsilon^2_{q_0}$ belong to isotropic lines and the form $\{h, e^*\}_{q_0}$ is positive at $\upsilon^1_{q_0} + \upsilon^2_{q_0}$; then

$$c^1_{01} = c^2_{02} = 0,\ \chi = \frac{c^2_{01} + c^1_{02}}{2},\ \{h, e^*\} = 2\chi u_1 u_2. \tag{4.1}$$

The asymptotic expansion for the conjugate locus from theorem 3.2 takes a form

$$Con_{q_0}(\theta, \nu) = q_0 \pm \frac{\pi}{\nu^2} e_{q_0} \pm \frac{2\pi\chi(q_0)}{\nu^3}(\cos^3(\theta) f^1_{q_0} - \sin^3(\theta) f^2_{q_0}) + O(\frac{1}{\nu^4}) \tag{4.2}$$

as $\nu \longrightarrow \pm\infty$. Note that quadratic term in the asymptotics does not depend on θ and cubic term is nothing else but a parametrization of an astroid in Δ_{q_0} with isotropic lines of $\{h, e^*\}_{q_0}$ as the cuspidal directions.

Fix local coordinates (x_0, x_1, x_2) in a neighborhood of q_0 in such a way that

$$(x_0, x_1, x_2)(q_0) = 0, \quad e = \partial_{x_0}, \quad f^i_{q_0} = \partial_{x_i},\ i = 1, 2.$$

In such coordinates, we have

$$\frac{1}{\pi} F(\pi(2 + \eta^2\tau), \theta, \eta) = (x_0(\tau, \theta, \eta), x_1(\tau, \theta, \eta), x_2(\tau, \theta, \eta)) =$$

$$= (-\eta^2, \eta^3(\tau - c^1_{02}) u_1(\theta), \eta^3(\tau + c^2_{01}) u_2(\theta)) + O(\eta^4), \tag{4.3}$$

where, as usually, $u_1(\theta) = \cos\theta, u_2(\theta) = \sin\theta$.

Set $\zeta = \sqrt{-x_0(\tau, \theta, \eta)} = \eta + O(\eta^3)$ and apply the smooth change of variables $(\tau, \theta, \eta) \mapsto (\tau, \theta, \zeta)$ to the mapping (4.3). We obtain

$$\frac{1}{\pi} F(\pi(2+\eta^2\tau), \theta, \eta) = \Big(-\zeta^2, \zeta^3(\tau - c^1_{02}) u_1(\theta) + O(\zeta^4), \zeta^3(\tau + c^2_{01}) u_2(\theta) + O(\zeta^4)\Big).$$

The first coordinate is constant when ζ is constant. The mapping (4.3) is thus reduced to the depending on ζ family of mappings of two variables,

$$\frac{1}{\pi}F(\pi(2+\eta^2\tau),\theta,\eta) = (-\zeta^2, \zeta^3\Phi_\zeta(\tau,\theta)), \tag{4.4}$$

where $\Phi_\zeta(\tau,\theta) = ((\tau - c_{02}^1)u_1(\theta), (\tau + c_{01}^2)u_2(\theta)) + O(\zeta)$.

The critical set of the mapping Φ_0 is a smooth closed curve in $R \times S^1$ defined by the equation $\tau = c_{02}^1 u_2^2(\theta) - c_{01}^2 u_1^2(\theta)$. Critical values of Φ_0 fill the astroid $2\chi(-u_1^3(\theta), u_2^3(\theta))$, $\theta \in S^1$, as we could predict because of asymptotic expansion (4.2). The restriction of Φ_0 to the critical set is a one-to-one mapping. Moreover, every critical point of Φ_0 is a fold or cusp (actually, there are 4 cusps). It means that Φ_0 is an Whitney mapping (see [15]); it is a stable mapping, according to the Thom–Mather theory. In other words, for any big enough $K \subset\subset R \times S^1$ there exist $\varepsilon > 0$ such that $\forall \zeta \in (0,\varepsilon)$, the mapping $\Phi_\zeta|_K$ is equivalent to $\Phi_0|_K$ under smooth transformations of variables in the domain and the range; moreover, the family of transformations of variables can be chosen smooth in ζ (see [10] for details).

Summing up, we come to the following essential supplement to theorem 3.2.

Theorem 4.1 *Suppose $\chi(q_0) \neq 0$ and a Riemannian structure is fixed in M; then $\exists \varepsilon > 0$ such that any containing q_0 open Riemannian ball B of radius less than ε enjoys the following property.*

There exists an open set U, $B \cap \mathsf{Con}_{q_0} \subset U \subset B \setminus q_0$ and a diffeomorphism $\Psi : U \to R^3 \times \{1,-1\}$ such that

$$\Psi(B \cap \mathsf{Con}_{q_0}) = \{(\zeta^2, (\zeta\cos\theta)^3, -(\zeta\sin\theta)^3) : \zeta > 0,\ \theta \in S^1\} \times \{1,-1\}.$$

In particular, each of 2 connected components of $B \cap \mathsf{Con}_{q_0}$ contains 4 cuspidal edges.

4.2 Cut locus and length

We have a satisfactory description of the conjugate locus near q_0, if $\chi_{q_0} \neq 0$. Let us now consider the cut locus. By $\ell_*(\theta;\nu)$ we denote the cut length for the geodesic $\gamma_{\theta,\nu}$ and set $Cut_{q_0}(\theta,\nu) = \gamma_{\theta,\nu}(\ell_*(\theta;\nu))$. We are continuing to use the privileged frame defined in the previous section.

Theorem 4.2 *Suppose* $\chi(q_0) \neq 0,\ \{h, e^*\} = 2\chi u_1 u_2$; *then the cut locus and length admit the following asymptotic expansion as* $\nu \longrightarrow \pm\infty$:

$$Cut_{q_0}(\theta, \nu) = q_0 \pm \frac{\pi}{\nu^2} e_{q_0} \pm \frac{2\pi\chi(q_0)\cos\theta}{\nu^3} f^1_{q_0} + O(\frac{1}{\nu^4}),$$

$$\ell_*(\theta; \nu) = \frac{2\pi}{|\nu|} - \frac{\pi(\kappa(q_0) + 2\chi(q_0)\sin^2\theta)}{|\nu|^3} + O(\frac{1}{\nu^4}).$$

Remark. Cubic terms in the asymptotic expansions for $Con(\theta, \nu)$ and for $\ell_*(\theta; \nu)$ are not surprising. We easily obtain them if reject $O(\eta^4)$ in the right–hand side of (4.3) and then compute the cut locus for the remaining mapping, which is cubic in η. The justification of such a procedure is not, however, trivial. We are going to formulate the "cut" analog of theorem 4.1 and shall prove it simultaneously with theorem 4.2.

Theorem 4.3 *Suppose* $\chi(q_0) \neq 0$ *and a Riemannian structure is fixed in* M; *then* $\exists \varepsilon > 0$ *such that any containing* q_0 *open Riemannian ball* B *of radius less than* ε *satisfies the following properties.*

- *There exists an open set* U, $B \cap \mathsf{Cut}_{q_0} \subset U \subset B \backslash q_0$ *and a diffeomorphism* $\Psi : U \to R^3 \times \{\pm 1\}$ *such that*

  $$\Psi(B \cap \mathsf{Cut}_{q_0}) = \{(\zeta^2, \xi, 0) : \zeta > 0, |\xi| \leq \zeta^3\} \times \{\pm 1\},$$

  where $\{\pm 1\}$ *is a two-point set;*

- $\partial(B \cap \mathsf{Cut}_{q_0}) = B \cap \mathsf{Cut}_{q_0} \cap \mathsf{Con}_{q_0}$ *is the union of 4 cuspidal edges of* $B \cap \mathsf{Con}_{q_0}$, *where simbol* ∂ *denotes the boundary relative to* $B \setminus q_0$.

Proof. We shall use notations of section 4.1, in particular, formula (4.4). By $\ell_\zeta(\tau, \theta)$ we denote the length of the segment of the geodesic $\gamma_{\theta,\nu}$ with the endpoints q_0 and $F(\pi(2+\eta^2\tau), \theta, \eta)$. We should describe the self-intersections of the mappings

$$(\tau, \theta) \mapsto (\ell_\zeta(\tau, \theta), \Phi_\zeta(\tau, \theta)) \tag{4.5}$$

for small $\zeta > 0$.

We have from the above calculations:

$$\ell_\zeta(\tau, \theta) = \eta 2\pi + \eta^3\pi(\tau + c^2_{01}u^2_1(\theta) - c^1_{02}u^2_2(\theta) - \kappa) + O(\eta^4)$$

and $\eta = \zeta + \frac{\alpha(\theta)}{2}\zeta^3 + O(\zeta^4)$, where $\alpha(\theta)$ comes from the expansion $x_0(0, \theta, \eta) = -\eta^2 + \alpha(\theta)\eta^4 + O(\eta^5)$ (see (4.3)). The strightforward calculation of $\alpha(\theta)$ is rather tedious, but we need this function only up to a constant summand; in other words, we need $\dot{\alpha}(\theta)$ and it can be find very easy. Indeed, let $\mathcal{L}$ be the Lagrangian submanifold in T^*M filled by the sub-Riemannian extremals started at $h_{q_0}^{-1}(\frac{1}{2})$; then $F|_{\eta\neq 0}$ is a composition of an immersion in $\mathcal{L}$ and the standard projection of T^*M on M. The length function ℓ is a primitive of the tautological 1-form ς restricted to $\mathcal{L}$. The form ς vanishes at the fibers of T^*M; hence any critical point of F is automatically a critical point of the mapping (ℓ, F) with values in $R \times M$. The mapping (ℓ, F) is a suspension over (4.5); it follows that any critical point of Φ_ζ must be a critical point of (4.5). The last property implies $\dot{\alpha}(\theta) = 6\chi u_1(\theta)u_2(\theta)$. Hence

$$\ell_\zeta(\tau, \theta) = C(\zeta) + \zeta^3\pi(\tau + \chi u_2^2(\theta)) + O(\zeta^4),$$

where $C(\zeta)$ does not depend on (τ, θ).

The mapping (ℓ, F) is a "wave front" or "Legendre projection" (see [7]) and mappings (4.5) are also like that. More precisely, an extension

$$(\tau, \theta) \mapsto (\ell_\zeta(\tau, \theta), \Phi_\zeta(\tau, \theta), u(\theta) + O(\zeta))$$

of (4.5) to the family of mappings with values in R^5 annihilates the contact form $d\ell - u_1 dx_1 - u_2 dx_2$. The mapping

$$(\tau, \theta) \mapsto (\tau + \chi u_2^2(\theta), \Phi_0(\tau, \theta)) \tag{4.6}$$

has only stable Legendre singularities (do not confuse with stable singularities of general smooth mappings treated in [10]!): folds of Φ_0 are cuspidal edges for (4.6) and cusps of Φ_0 are "swallow tails" for (4.6). The self-intersections and swallow tails fill 2 circles defined by the equations $\tau = c_{02}^1$ and $\tau = -c_{01}^2$ in the domain and they are transversal self-intersections. In the range, the self-intersections fill two nonintersecting pieces of parabolas. Hence the mapping (4.6) is stable: the family of mappings (4.5) (restricted to a big enough compact in the domain)can be transformed to (4.6) by a smooth family of smooth transformations of variables in the domain and in the range for small enough ζ.

Not all the self-intersections are related to the cut locus. Recall that the pre-images of geodesics in our model are lines $\{(\tau, \theta) : \tau \in R,\ \theta = const\}$

and the lengths of started at q_0 geodesic pieces grow with τ. Hence only self-intersections coming from the circle $\{(-c_{01}^2, \theta) : \theta \in S^1\}$ can define the cut locus.

$fig.1$

Figure 1 shows the singularities of the mapping Φ_0 and the self-intersections of mapping (4.6) in the domain and in the range. The domain is presented in the coordinates $(e^\tau \cos\theta, e^\tau \sin\theta)$. The restriction of Φ_0 to the domain $\{(\tau, \theta) : \tau < -c_{01}^2, \ \theta \in S^1\}$ is a one-to-one mapping. Take an arbitrary segment $[\tau_0, \tau_1] \subset (-\infty, -c_{01}^2)$; it follows from the stability of (4.6) that $\exists \eta_0 > 0$ such that the mapping

$$(\tau, \theta, \eta) \mapsto F(\pi(2 + \eta^2\tau), \theta, \eta), \quad \tau \in [\tau_0, \tau_1], \ \theta \in S^1, \ \eta \in (0, \eta_0],$$

is a regular one-to-one mapping. Besides that, the mapping
$(t, \theta, \eta) \mapsto F(\tau, \theta, \eta)$ resticted to the domain

$$\mathcal{D}(\tau_1, \eta_0) = \{(t, \theta, \eta) : 0 < t \leq \pi(2 + \eta^2\tau_1), \ \theta \in S^1, \ 0 < \eta \leq \eta_0\}$$

is a submersion (no critical points!)

Lemma 4.1 *There exists a constant $a > 0$ such that*

$$|\Pi(F(t, \theta, \eta))| \geq a\eta t, \quad |\Pi(F(2\pi \pm t, \theta, \eta))| \geq a\eta t - \frac{\eta^3}{a}$$

for $0 < t < a$, $\theta \in S^1$ and all small enough η, where $\Pi : (x_0, x_1, x_2) \mapsto (x_1, x_2)$ is the standard projection.

Proof. Let us identify vectors (x_1, x_2) with the complex numbers: $(x_1, x_2) = x_1 + ix_2$. We have

$$F(t, \theta, \eta) = \eta \int\limits_{\theta}^{\theta+t} e^{i\tau}\, d\tau + O(t\eta^2),$$

$$F(2\pi \pm t, \theta, \eta) = \eta \int\limits_{\theta}^{\theta\pm t} e^{i\tau}\, d\tau + O(t\eta^2) + O(\eta^3),$$

and $\left|\int\limits_{\theta}^{\theta\pm t} e^{i\tau}\, d\tau\right| = |(1 - e^{it})| \geq \frac{t}{2}$ for small enough t. $\square$

Choosing $(-\tau_0)$ much greater then the constant a from lemma 4.1 we obtain that the sets

$$F^{-1}\Big(F(\pi(2+\eta^2\tau_1),\theta,\eta)\Big)\cap\mathcal{D}(\tau_1,\eta_0)\quad \theta\in S^1,\ 0<\eta\leq\eta_0,$$

are actually one-point sets, if η_0 is small enough. We know that $F|_{\mathcal{D}(\tau_1,\eta_0)}$ is a submersion. If the sets

$$F^{-1}\left(F(t,\theta,\eta)\right)\cap\mathcal{D}(\tau_1,\eta_0)\quad \theta\in S^1,\ 0<\eta\leq\eta_0 \tag{4.7}$$

are finite $\forall(t,\theta,\eta)\in\mathcal{D}(\tau_1,\eta_0)$, then $F|_{\mathcal{D}(\tau_1,\eta_0)}$ is a covering, all the sets (4.7) have equal capacity, and this capacity must be 1. It is exactly what we need to complete the proof of theorems 4.2 and 4.3. Suppose a point $q\in M$ has an infinite number of pre-images in $\mathcal{D}(\tau_1,\eta_0)$; then there exists a sequence $\{(t^n,\theta^n,\eta^n): n=1,\ldots\}\in\mathcal{D}(\tau_1,\eta_0)$ such that $F(t^n,\theta^n,\eta^n)=q$ $\forall n$ and t^n or η^n tends to zero as $n\longrightarrow\infty$. Hence $q=q_0$. It follows, however, from lemma 4.1 that $q_0\notin F\left(\mathcal{D}(\tau_1,\eta_0)\right)$.

5 A degenerate case

5.1 Extra invariants

It was shown in the previous section that the asymptotic expansion of F_{q_0} till the third order contains all the essential information on the structure of small spheres and on local behavior of the distance function, if $\chi(q_0)\neq 0$. We need extra terms of the expansion, if $\chi(q_0)=0$. The privileged frame does not defined in the last case and we are free to select a frame that could simplify calculations.

Lemma 5.1 *Suppose a frame $\hat f_q^1,\hat f_q^2$ is the result of the rotation of the frame f_q^1,f_q^2 on the angle $\varphi(q)$ in the plane Δ_q, where φ is a smooth function. Then*

$$[\hat f_q^2,\hat f_q^1]=[f_q^2,f_q^1]+(f^1\varphi)f^1+(f^2\varphi)f^2.$$

Proof is the strightforward calculation. □

So we can assume without loss of generality that $c_{12}^1(q_0)=c_{12}^2(q_0)=0$; if it is not true for the original frame, then it is true for a rotated one. We suppose that $\{h, e^*\}_{q_0} = 0,\ c^1_{12}(q_0) = c^2_{12}(q_0) = 0$ in all the calculations in this subsection, unless otherwise stated. We also use simplified notations

$$\alpha = c^1_{01} = -c^2_{02},\ \beta = \frac{c^2_{01} + c^1_{02}}{2},\ c = \frac{c^2_{01} - c^1_{02}}{2};$$

then $a(\theta) \overset{\text{(def)}}{=} \{h, e^*\}|_{u(\theta)} = \alpha \cos 2\theta + \beta \sin 2\theta,\ \alpha(q_0) = \beta(q_0) = 0$. Structural equations $d^2 \upsilon^j = 0,\ j = 1, 2$ take the form

$$\begin{array}{rcl} ec^1_{12} + f^1 c + f^2 \alpha - f^1 \beta & = & 0 \\ ec^2_{12} + f^2 c + f^1 \alpha + f^2 \beta & = & 0. \end{array}$$

We obtain a more elegant version of these equations using notations actively exploited in section 3:

$$eb(\theta) + f(u(\theta))c + f(\dot{u}(\theta))a(\theta) - \frac{1}{2} f(u(\theta))\dot{a}(\theta) = 0. \tag{5.1}$$

Theorem 5.1 *Suppose* $\chi(q_0) = 0$; *then the following asymptotic expansions hold as* $\nu \longrightarrow \pm\infty$:

$$\frac{\partial}{\partial \theta} Con_{q_0}(\theta, \nu) =$$

$$= \mp \frac{2\pi}{\nu^4} \Big(2\partial_\theta \{h, \{h, e^*\}\}_{q_0}(u(\theta)) - \{h, \partial_\theta \{h, e^*\}\}_{q_0}(u(\theta)) \Big) f(u(\theta)) + O(\frac{1}{\nu^5}),$$

$$\frac{\partial}{\partial \theta} \ell_1(\theta; \nu) = \mp \frac{2\pi}{\nu^4} \Big(3\partial_\theta \{h, \{h, e^*\}\}_{q_0}(u(\theta)) - 2\{h, \partial_\theta \{h, e^*\}\}_{q_0}(u(\theta)) +$$

$$+ \{h, \kappa\}_{q_0}(u(\theta)) \Big) + O(\frac{1}{\nu^5}),$$

$$\ell_1(\theta; \nu) = \frac{2\pi}{\nu} - \frac{\pi \kappa(q_0)}{\nu^2} - \frac{1}{2} \int_\theta^{\theta+\pi} \frac{\partial \ell_1}{\partial t}(t; \nu)\, dt + O(\frac{1}{\nu^5}).$$

Remark. ∂_θ is a well-defined vertical vector field on T^*M, hence above expressions are absolutely invariant; the field ∂_θ annihilates h and e^* but not the symplectic structure! κ is a function on M, we automatically identify such functions with constant on fibers functions on T^*M.

Proof. We follow the way of proof and notations used for theorem 3.2 and just compute extra terms in asymptotic expansions. We have $\frac{\partial}{\partial\theta}F(2\pi,\theta,\eta) = O(\eta^4)$ and have to find $s_\prime$ such that

$$\langle\frac{\partial}{\partial\theta}F(2\pi+\eta^3 s_\prime(\theta,\eta)|f(\dot{u}(\theta))\rangle = O(\eta^5); \tag{5.2}$$

then $\frac{\partial}{\partial\theta}Con_{q_0}(\theta,\nu) = \eta^4(\frac{\partial}{\partial\theta}F(2\pi+\eta^3 s_\prime,\theta,\eta) + \frac{ds_\prime}{d\theta}f(u(\theta))) + O(\eta^5)$. Now compute:

$$\frac{\partial}{\partial\theta}F(2\pi+\eta^3 s_\prime,\theta,\eta) = \frac{\partial}{\partial\theta}q_0 Q_\theta^{2\pi+\eta^3 s_\prime} =$$

$$q_0 Q_\theta^{2\pi}\circ\left(\frac{w(\theta+2\pi)}{1-\eta b(\theta)}f(u(\theta)) + \eta^4 s_\prime f(\dot{u}(\theta))\right) +$$

$$+q_0\left(\int\limits_\theta^{\theta+2\pi} AdQ_\theta^{t-\theta}\frac{\partial}{\partial\theta}\Big(\frac{w(t)}{1-w(t)b(t)}\Big)f(u(t))\,dt - \eta f\right)\circ Q_\theta^{2\pi} + O(\eta^5).$$

Equation (3.6) implies

$$w(t) = \eta + \eta^3\int\limits_\theta^t \frac{Q_\theta^{\tau-\theta}a(\tau)}{1-\eta b(\tau)}\,d\tau + O(\eta^5) =$$

$$= \eta^3\int\limits_\theta^t\Big(a(\tau)(1+\eta b(\tau)) + \eta(\int\limits_\theta^\tau f(u(\tau'))\,d\tau' a(\tau)\Big)\,d\tau + O(\eta^5) =$$

$$= \eta^3\int\limits_\theta^t\Big(a(\tau)(1+\eta b(\tau)) + \eta(f(\dot{u}(\theta)) - f(\dot{u}(\tau))\Big)\,d\tau + O(\eta^5).$$

In particular, $w(\theta+2\pi) = \eta + O(\eta^5)$. It follows from proposition 3.2 that

$$Q_\theta^{2\pi} = \eta^2\pi e - \frac{\eta^3\pi}{2}\int\limits_\theta^{\theta+\pi}[f(u(t)),e]\,dt + O(\eta^4) =$$

$$= -\eta^2\pi e - \eta^3\pi[f(\dot{u}(\theta)),e] + O(\eta^4).$$

Collecting all the terms we reduce (5.2) to the equation

$$s_\prime - \pi f(u)a - \pi\langle[[f(\dot{u}),e],f]|f(\dot{u})\rangle = 0.$$

We have

$$[f(\dot u), e] = (\frac{\dot a}{2} - c) f(u) - a f(\dot u),$$

$$[[f(\dot u), e], f] = (f(u)c - f(u)\frac{\dot a}{2}) f(u) + (fa) f(\dot u).$$

Hence $s_\prime = 2\pi f(u)a$; collecting terms again and applying (5.1) we obtain

$$\frac{\partial}{\partial\theta} Con(\theta,\nu) = \eta^4 \pi (4 f(\dot u(\theta)) a(\theta) + 2 f(u(\theta)) \dot a(\theta)) + O(\eta^5). \tag{5.3}$$

The first statement of theorem 5.1 is just equality (5.3) rewritten in invariant terms. Now compute the length:

$$\ell_1(\theta;\nu) = \int\limits_{\theta}^{\theta+2\pi+\eta^3 s_\prime} \frac{w(t)}{w(t) Q_\theta^{t-\theta} b(t)}\, dt + O(\eta^5) =$$

$$= \int\limits_{\theta}^{\theta+2\pi} w(t)\, dt + \eta^4 s_\prime(\theta) + \eta^2 \int\limits_{\theta}^{\theta+2\pi} Q_\theta^{t-\theta} b(t)\, dt + O(\eta^5) =$$

$$= \eta 2\pi - \eta^3 \pi\kappa + \eta^4 \pi \psi(\theta) + O(\eta^5),$$

where $\psi(\theta)$ is a cubic trigonometric polynomial such that

$$\dot\psi = 2 f(\dot u) a + \frac{3}{2} f\dot a + f\kappa - fc - eb \overset{(5.1)}{=} 3 f(\dot u) a + f(u)\dot a + f\kappa.$$

Cubic trigonometric polynomials are odd functions on S^1; hence $\psi(\theta) = -\psi(\theta+\pi) = -\frac{1}{2} \int\limits_{\theta}^{\theta+\pi} \dot\psi(t)\, dt$. $\square$

We have

$$\begin{array}{c} 2\partial_\theta \{h, \{h, e^*\}\}_{q_0}(u(\theta)) - \{h, \partial_\theta\{h, e^*\}\}_{q_0}(u(\theta)) = \\ = 2(f^2_{q_0}\alpha + f^1_{q_0}\beta)\cos 3\theta + 2(f^2_{q_0}\beta - f^1_{q_0}\alpha)\sin 3\theta. \end{array} \tag{5.4}$$

Moreover, at an arbitrary point of T^*M, without any assumptions on structural constants, we have

$$\begin{array}{c} 2\partial_\theta\{h, \{h, e^*\}\} - \{h, \partial_\theta\{h, e^*\}\} = 2(\nu\{h, e^*\} + \\ +(f^2\alpha + f^1\beta + 2c^1_{12}\alpha - 2c^2_{12}\beta)\cos 3\theta + \\ +(f^2\beta - f^1\alpha + 2c^2_{12}\alpha + 2c^1_{12}\beta)\sin 3\theta). \end{array} \tag{5.5}$$

Let R_q be the restriction of function (5.5) to the plane $e_q^{\perp}$; then R_q is a cubic form on the Euclidean plane with the following symmetry property: rotation of the plane on the angle $\frac{\pi}{3}$ implies multiplication of the form by(-1). The only invariant of such a form is its maximum on the unite circle, a "nonlinear eigenvalue". Set $\rho(q) = \max_{|u|=1} \mathsf{R}_q$; then

$$\rho(q_0) = 2\sqrt{(f_{q_0}^2\alpha + f_{q_0}^1\beta)^2 + (f_{q_0}^2\beta - f_{q_0}^1\alpha)^2}.$$

So conjugate locus possesses an asymptotic $\frac{\pi}{2}$-rotation symmetry for generic germs and $\frac{\pi}{3}$-rotation symmetry for germs with a simplest degeneration. This symmetry is, in my opinion, the most misterеous fact in all the theory. It appears as a result of calculations, but I cannot recognize it in the original problem. Perhaps we observe here a fragment of a deep hidden symmetry.

5.2 Conjugate and cut loci

Suppose $\chi(q_0) = 0, \rho(q_0) \neq 0$. Let us fix an appropriate frame f^1, f^2. Without loss of generality, we can suppose that R_{q_0} vanishes at $v_{q_0}^1$ and is negative at $v_{q_0}^2$; then $\mathsf{R}_{q_0}(u) = \rho(q_0)u_2(3u_1^2 - u_2^2)$, $\mathsf{R}_{q_0}(u(\theta)) = \rho(q_0)\sin 3\theta$. Fix local coordinates (x_0, x_1, x_2) in a neighborhood of q_0 in such a way that

$$(x_0, x_1, x_2)(q_0) = 0, \quad e = \partial_{x_0}, \quad f_{q_0}^i = \partial_{x_i}, \ i = 1, 2,$$

and identify (x_1, x_2) with a complex number $y = x_1 + ix_2$. In these coordinates, we have

$$\frac{1}{\pi}F(\pi(2 + \eta^3(s_\prime(\theta) + \rho\tau)), \theta, \eta) = (x_0(\tau, \theta, \eta), y(\tau, \theta, \eta)) =$$

$$= \left(-\eta^2, \eta^4\rho(\delta + \tau e^{\theta i} - \frac{1}{2}e^{-2\theta i} - \frac{1}{4}e^{4\theta i})\right) + O(\eta^5), \tag{5.6}$$

where δ is a constant vector.

Set $\zeta = \sqrt{-x_0(\tau, \theta, \eta)} = \eta + O(\eta^4)$ and apply the smooth change of variables $(\tau, \theta, \eta) \mapsto (\tau, \theta, \zeta)$ to the mapping (4.3). We obtain

$$\frac{1}{\pi}F(\pi(2 + \eta^3(s_\prime(\theta) + \rho\tau), \theta, \eta) = \left(-\zeta^2, \zeta^4\rho\Phi_\zeta(\tau, \theta)\right),$$

where $\Phi_\zeta(\tau, \theta)) = \delta + \tau e^{\theta i} - \frac{1}{2}e^{-2\theta i} - \frac{1}{4}e^{4\theta i} + O(\zeta)$.

The mapping (5.6) is thus reduced to the family of mappings Φ_ζ. The critical set of the mapping Φ_0 is the circle $\{(0,\theta) : \theta \in S^1\}$. There are 6 cusps $(0, \frac{k\pi}{3})$, $k = 0, 1, \ldots$, and all other critical points are folds. Standard stability arguments imply the following

Proposition 5.1 *Suppose* $\chi(q_0) = 0, \rho(q_0) \neq 0$; *then the curves*

$$\theta \mapsto Con_{q_0}(\theta, \nu), \quad \theta \in S^1,$$

contain exactly 6 cuspidal points in S^1, *for all small enough* $|\nu|$.

Well, cuspidal points in S^1 are stable, but the mapping Φ_0 is not. Indeed, the restriction of Φ_0 to the critical curve is π-periodic, it is a double covering! In particular, only 3 cusps of Φ_0 are visible in the range. There is no reason to expect that it is the same for Φ_ζ with nonzero ζ and calculations show that it is not the same. We shall prove in the next subsection that the function $\chi^2 + \rho^2$ is strictly positive for generic contact sub-Riemannian structure on a three- dimensional manifold. There are rather convicing arguments in favor of the following

Conjecture. For generic contact sub-Riemannian structure on M, the equality $\chi(q_0) = 0$ implies the following properties of the conjugate locus: each of 2 connected components of the intersection of Con_{q_0} with any centered at q_0 small Riemannian ball contains exactly 6 cuspidal edges and 3 or 5 self-intersection lines diffeomorphic to intervals.

Figure 2 shows a probable form of the set of critical values of Φ_ζ for small nonzero ζ.

$$fig.2$$

Fortunately, we need no the detailed structure of the conjugate locus in order to describe the cut locus and thus the singularities of small sub-Riemannian spheres.

Theorem 5.2 *Suppose* $\chi(q_0) = 0, \rho(q_0) \neq 0$, $\mathsf{R}_{q_0}(u) = \rho(q_0) u_2 (3u_1^2 - u_2^2)$; *then the cut locus and length admit the following asymptotic expansion as* $\nu \longrightarrow \pm\infty$:

$$\mathsf{Cut}_{q_0}(\theta,\nu) = q_0 \pm \frac{\pi}{\nu^2} e_{q_0} \mp \frac{\pi\rho(q_0)}{\nu^4}\Big(\delta + (\frac{1}{4} - \cos^2(\theta - \frac{2k\pi}{3}))f_{q_0}(u(\frac{2k\pi}{3}))\Big) + O(\frac{1}{\eta^5})$$

$$\ell_*(\theta;\nu)=\ell_1(\theta;\nu)+\frac{\pi\rho(q_0)}{2\nu^4}\Big(\cos 3\theta-\cos(\theta-\frac{2k\pi}{3})\Big)+O(\frac{1}{\eta^5}),\ |\theta-\frac{2k\pi}{3}|\leq\frac{\pi}{3},$$

$k=0,1,2,$ *where* $\delta\in T^*_{q_0}$ *is a constant vector.*

This theorem will be proved simultaneously with the following

Theorem 5.3 *Suppose* $\chi(q_0)=0, \rho(q_0)\neq 0$ *and a Riemannian structure is fixed in* M; *then* $\exists\varepsilon>0$ *such that any containing* q_0 *open Riemannian ball* B *of radius less than* ε *satisfies the following properties.*

- *There exists an open set* U, $B\cap\mathsf{Cut}_{q_0}\subset U\subset B\backslash q_0$ *and a diffeomorphism* $\Psi: U\to R\times C\times\{\pm1\}$ *such that*

  $$\Psi(B\cap\mathsf{Cut}_{q_0})=\{(\zeta^2,\xi e\frac{i2k\pi}{3}):\zeta>0,|\xi|\leq\zeta^4\}\times\{\pm1\},$$

  where C *is the complex plane;*

- $\partial(B\cap\mathsf{Cut}_{q_0})=B\cap\mathsf{Cut}_{q_0}\cap\mathsf{Con}_{q_0}$ *is the union of 6 cuspidal edges of* $B\cap\mathsf{Con}_{q_0}$, *where simbol* ∂ *denotes the boundary relative to* $B\setminus q_0$.

Proof. We follow the same way as for the proof of theorems 4.2-4.3. Let $\ell_\zeta(\tau,\theta)$ be the length of the segment of the geodesic $\gamma_{\theta,\nu}$ with the endpoints q_0 and $\gamma_{\theta,\nu}(\ell_1(\theta;\nu)+\eta^4\pi\rho\tau)$. Then

$$\ell_\zeta(\tau,\theta)=\ell_1(\theta;\nu)+\eta^4\pi\rho\tau+O(\eta^5)$$

and $\eta=\zeta+\frac{\alpha(\theta)}{2}\zeta^4+O(\eta^5)$, where $\alpha(\theta)$ comes from the expansion $x_0(0,\theta,\eta)=-\eta^2+\alpha(\theta)\eta^5+O(\eta^6)$. We should study the "wave fronts"

$$(\tau,\theta)\mapsto(\ell_\zeta(\tau,\theta),\Phi_\zeta(\tau,\theta)) \tag{5.7}$$

for small $\zeta>0$. Any critical point of Φ_ζ is automatically a critical point for $\ell_\zeta(\tau,\theta)$; this fact permits us to obtain the asymptotics of $\frac{\partial\ell}{\partial\theta}$ without strightforward calculation of $\alpha(\theta)$. We obtain

$$\ell_\zeta(\tau,\theta)=C(\zeta)+\zeta^4\pi\rho(\tau-\frac{2}{3}\cos 3\theta)+O(\zeta^5),$$

where $C(\zeta)$ does not depend on (τ,θ).

let us consider the wave front

$$(\tau,\theta)\mapsto(\tau-\frac{2}{3}\cos 3\theta,\Phi_0(\tau,\theta)). \tag{5.8}$$

I do not know is it stable as a wave front or not. Doubtful points are $\tau=0,\theta=\frac{(4k+3)\pi}{6}$, $k=0,1,2$, where cuspidal edges meet self-intersections. Fortunately, these points are far from the related to cut locus part of the wave front and we obtain a stable wave front if cut out small neighborhoods of the bad points. The remaining part contains only cuspidal edges (folds of Φ_0), swallow tails (cusps of Φ_0), and transversal self-intersections defined by the equations

$$2\tau=\cos 3\theta-\cos(\theta-\frac{2k\pi}{3}),\quad k=0,1,2; \tag{5.9}$$

in particular, there are 2 triple points in the range: $(\pm\frac{1}{12},0)$.

The pre-images of geodesics in our model are lines $\{(\tau,\theta):\tau\in R,\theta=const\}$ and the length of started at q_0 geodesic pieces grow with τ. We derive that only segments of self-intersections (5.9) satisfying the inequality $|\theta-\frac{2k\pi}{3}|\leq\frac{\pi}{3}$ can define the cut locus.

$$fig.3$$

Figure 3 shows the singularities of Φ_0 and the related to cut locus part of the self-inter sections of the mapping (5.8) (in the domain and in the range). The domain is presented in the coordinates $(e^\tau\cos\theta,e^\tau\sin\theta)$. The restriction of Φ_0 to the domain$\{(\tau,\theta):\tau<-\frac{3}{4},\theta\in S^1\}$ is a one-to-one mapping.

To complete the proof of theorems 5.2, 5.3 it remains to repeate arguments used in proof of theorems 4.2, 4.3, with obvious changes.

5.3 Isoperimetric problems

Plane isoperimetric problems are in fact special cases of three-dimensional sub-Riemannian problems; associated to these problems geodesics have also a natural physical interpretation as trajectories of charged particles in magnetic fields (see [12] for details).

We consider a smooth 2-form $\varphi(x_1,x_2)dx_1\wedge dx_2$ on R^2. Let $\varphi dx_1\wedge dx_2=d\vartheta$, where ϑ is a 1-form. The problem is to describe plane curves that have minimal Euclidean length among all Lipschitzian curves $\xi : [0,1] \to R^2$ with fix endpoints $\xi(0), \xi(1)$ and fix integral $\int\limits_{\xi} \vartheta$. It is clear that the desired curves depend only on φ. Set $M = R^3$, $q = (x_0, x_1, x_2)$, $\Delta = (dx_0 - \vartheta)^{\perp}$, and consider the sub-Riemannian structure $(dx_1)^2 + (dx_2)^2$ on Δ, induced by the Euclidean structure on the plane. Admissible trajectories for Δ are curves of the form $t \mapsto \Big(\int\limits_{\xi|_{[0,t]}} \vartheta,\ \xi(t) \Big)$, where ξ is an arbitrary Lipschitzian curve in R^2; the sub-Riemannian length coincides with the length of ξ. Thus plane projections of the sub-Riemannian length minimizers are actually solutions to the isoperimetric problem.

The distribution Δ is a contact one iff $\varphi(x) \neq 0\ \forall x = (x_1, x_2) \in R^2$. Suppose $\varphi > 0$; then $\omega = \frac{1}{\varphi}(\vartheta - dx_0)$. Let $\vartheta = \delta_1 dx_1 + \delta_2 dx_2$; we have

$$e = \frac{\partial \ln \varphi}{\partial x_1}(\partial_{x_2} + \delta_2 \partial_{x_0}) - \frac{\partial \ln \varphi}{\partial x_2}(\partial_{x_1} + \delta_1 \partial_{x_0}) - \varphi \partial_{x_0}.$$

We may set $f_i = \partial_{x_i} + \delta_i \partial_{x_0}$, $i = 1, 2$; then structural constants take a form:

$$\alpha = -\frac{\partial^2 \ln \varphi}{\partial x_1 \partial x_2}, \quad \beta = \frac{1}{2}\left(\frac{\partial^2 \ln \varphi}{\partial x_1^2} - \frac{\partial^2 \ln \varphi}{\partial x_2^2} \right).$$

Recall that $\chi^2 = \alpha^2 + \beta^2$ is an important differential invariant. It follows from the standard transversality theorem that χ can vanish only in isolated points in R^2 for generic φ. An extra invariant ρ has a form

$$\rho^2 = \left(\frac{\partial^3 \ln \varphi}{\partial x_1^3} - 3\frac{\partial^3 \ln \varphi}{\partial x_1 \partial x_2^2} \right)^2 + \left(\frac{\partial^3 \ln \varphi}{\partial x_2^3} - 3\frac{\partial^3 \ln \varphi}{\partial x_1^2 \partial x_2} \right)^2$$

at any point, where $\gamma = 0$. Hence γ and ρ do not vanish simultaneously for generic φ. We also give an expression for the invariant κ :

$$\kappa = \frac{3}{2}\left(\frac{\partial^2 \ln \varphi}{\partial x_1^2} + \frac{\partial^2 \ln \varphi}{\partial x_2^2} \right) - \left(\frac{\partial \ln \varphi}{\partial x_1} \right)^2 - \left(\frac{\partial \ln \varphi}{\partial x_2} \right)^2 = \frac{3}{2} \triangle \ln \varphi - |\nabla \ln \varphi|^2.$$

We now come back to a general contact sub-Riemannian structure on an arbitrary three dimensional manifold M and use derived formulas to compute the codimension of degenerations.

Theorem 5.4 *Given a three-dimensional manifold M, the following conditions hold for any structure from an open dense subset in the space of contact sub-Riemannian structures on M endowed with the Whitney topology:*

- *The equation $\chi(q) = 0$ defines a smooth 1-dimensional or empty submanifold in M;*

- $\chi^2(q) + \rho^2(q) > 0 \quad \forall q \in M.$

Proof. A pair f^1, f^2 of germs at q of smooth vector fields defines a germ of a contact sub-Riemannian structure iff

$$f^1_q \wedge f^2_q \wedge [f^1, f^2]_q \neq 0. \tag{5.10}$$

Let $J^k_q VectM$ be the space of k-jets at q of smooth vector fields. Inequality (5.10) defines a Zariski-open subset $\mathcal{C}^k_q$ in $J^k_q VectM \times J^k_q VectM$ for $k \geq 1$; moreover, $\alpha(q), \beta(q)$ are actually regular rational functions on $\mathcal{C}^3_q$ and hence on $\mathcal{C}^k_q$ for $k \geq 3$. We have to prove that the equations $\alpha(q) = \beta(q) = 0$ define a codimension 2 subset in $\mathcal{C}^k_q$. It is true, if the equations are algebraically independent. They are certainly independent, if their restrictions to a linear subspace in the jet space are independent; the jet space of isoperimetric problems is a required subspace.

Further, the equation $\chi^2(q) + \rho^2(q) = 0$ is equivalent to the following system of equations:

$$\alpha(q) = \beta(q) = f^2_q\alpha + f^1_q\beta = f^2_q\beta - f^1_q\alpha = 0. \tag{5.11}$$

System (5.11) is actually a system of 4 rational equations on $\mathcal{C}^4_q$ and hence on $\mathcal{C}^k_q$ for $k \geq 4$. We have to prove that (5.11) defines a codimension 4 subset in $\mathcal{C}^k_q$. It is enough to show that the equations are algebraically independent. Their restrictions to the jets space of isoperimetric problems are independent and we are done.

6 A metric interpretation of invariants

Let $\mathsf{d}(q_0, q_1)$ denote the Carnot–Caratheodory distance between points $q_0, q_1 \in M$, i.e. the infimum of the length of admissible curves connecting q_0 and q_1. In his paper [11] Michael Gromov discusses the problem of a reconstructing of infinitesimal invariants for sub-Riemannian structures in terms of the metric d, without explicit using of the smooth structure on M. In this section we give metric definitions for the invariants χ, κ, ρ, and for the trajectories of the Stokes field e.

The following simple proposition is quit general; it is true for all sub-Riemannian structures, not only for the contact ones.

Proposition 6.1 *A curve $\xi : [0,1] \to M$ is a Lipschitzian curve with respect to the Carnot–Caratheodory metric d, if and only if ξ is an admissible curve. The sub-Riemannian length of ξ is equal to*

$$\sup \left\{ \sum_{i=1}^{k} \mathsf{d}(\xi(t_{i-1}), \xi(t_i)) : k > 0,\ 0 = t_0 < t_1 < \ldots < t_k = 1 \right\}.$$

□

It follows from proposition 6.1 that the sub-Riemannian geodesics are correctly defined in terms of the metric d. Moreover, the metric d defines the standard topology in M, hence C^0-local minimizers are also correctly defined in terms of d. In particular, for given geodesic γ its cut length $\ell_{cut}(\gamma)$ and the first conjugate length $\ell_{con}(\gamma)$ are metric invariants, if these lengths exist. Set

$$\mathsf{L}_{q_0} = \{(\ell_{cut}(\gamma), \ell_{con}(\gamma)) : \gamma(0) = q_0\} \subset \{(l_1, l_2) : 0 < l_1 \leq l_2\} \subset R^2.$$

The germ at 0 of the plane set L_{q_0} is a metric invariant of the germ at q_0 of the Carnot–Caratheodory space (M, d). It is true for any sub-Riemannian structure. Coming back to a contact sub-Riemannian structure on a 3-dimensional manifold we obtain the following

Theorem 6.1 *The germ at 0 of the set L_{q_0} has the following properties:*

- $\overline{\lim\limits_{\substack{l_i \to 0 \\ (l_1,l_2) \in \mathsf{L}_{q_0}}}} \frac{l_2 - l_1}{l_i^3} = \frac{\chi(q_0)}{4\pi^2}, \quad i = 1, 2;$

- *if* $\chi(q_0) = 0$, *then* $\overline{\lim\limits_{\substack{l_i \to 0 \\ (l_1,l_2) \in \mathsf{L}_{q_0}}}} \frac{l_2 - l_1}{l_i^4} = \frac{\rho(q_0)}{12\sqrt{3}\pi^3}, \quad i = 1, 2;$

- *if* $\chi^2(q_0) + \rho^2(q_0) \neq 0$, *then* $(l, l) \in \mathsf{L}_{q_0}$ *for any small enough* $l > 0$.

Proof. The desired results are strightforward corollaries of theorems 3.2, 4.2, 4.3, 5.1–5.3. □

The sets Cut_q, Con_q are, of course, defined by the metric d. Consider a piecewise constant curve in M :

$$t \mapsto q_i, \quad t_{i-1} < t \leq t_i, \ i = 1, \ldots, k, \tag{6.1}$$

where $q_i \in M,\ 0 = t_0 < t_1 < \cdots < t_k,\ k > 0$. We say that the curve (6.1) is a started at q_0 *(2nd order) pre-indicatrix* for the metric d, if $\forall i$

$$q_i \in \mathsf{Cut}_{q_{i-1}} \quad \mathsf{d}(q_i, q_{i-1}) = (t_i - t_{i-1})^{\frac{1}{2}} < \mathsf{d}(q_{i+1}, q_{i-1}).$$

A continuous curve $\varrho : [0, \tau] \to M$ is a started at q_0 *(2nd order) indicatrix* for the metric d, if ϱ is the uniform limit of a sequence of started at q_0 pre-indicatrixes

$$\{t \mapsto q_i^j, \quad t_{i-1}^j < t \leq t_i^j, \ i = 1, \ldots, k^j\}_{j=1}^{\infty},$$

where $0 = t_0 < \cdots < t_{k^j} = \tau, \quad \max\limits_i (t_i^j - t_{i-1}^j) \longrightarrow 0$ as $j \longrightarrow \infty$. The following result is a corollary of theorem 4.2.

Theorem 6.2 *There are exactly 2 started at q_0 germs of indicatrixes: the germs of trajectories $\varrho_\pm$ of the vector fields $\pm\frac{1}{4\pi}e$. The distance $\mathsf{d}(\varrho_\pm(t), q_0)$ admits the following asymptotic expansion as $t \longrightarrow 0$:*

$$\mathsf{d}(\varrho_\pm(t), q_0) = t^{\frac{1}{2}} - \frac{t^{\frac{3}{2}}}{8\pi^2}(\kappa(q_0) + 2\chi(q_0)) + O(t^2).$$

□

Remark. The statement of theorem 6.2 remains true after the replacing $\mathsf{Cut}_{q_{i-1}}$ by $\mathsf{Con}_{q_{i-1}}$ in the definition of pre-indicatrixes.

Appendix

We list notations and formulas of chronological calculus that are used in the paper (see [3, 5]) for details). We identify C^∞-smooth diffeomorphisms $P : M \to M$ with automorphisms $\varphi(\cdot) \mapsto \varphi(P(\cdot))$ of the algebra $C^\infty(M)$ of smooth functions on M. A point $q \in M$ is identified with the homomorphism $\varphi \mapsto \varphi(q)$ of $C^\infty(M)$ in R. These identifications of nonlinear mappings with dual linear ones justify "operator" notations $q\varphi$ and qP for the values at q of the function $\varphi(\cdot) \in C^\infty(M)$ and the diffeomorphism P. Smooth vector fields on M are derivations of the algebra $C^\infty(M)$, i.e. R-linear mappings $X : C^\infty(M) \to C^\infty(M)$, meeting Leibnitz rule: $X(\alpha\beta) = (X\alpha)\beta + \alpha(X\beta)$. The value $X_q \in T_qM$ of a vector field X at a point $q \in M$ can also be denoted qX . By $[X^1, X^2]$ we denote the Lie bracket $X_1 \circ X_2 - X_2 \circ X_1$ of vector fields X^1, X^2. In local coordinates on M it is calculated as

$$[X^1, X^2] = \left[\sum_{i=1}^n X_i^1 \partial_{x_i}, \sum_{i=1}^n X_i^2 \partial_{x_i}\right] = \sum_{i=1}^n \left(\frac{\partial X_i^2}{\partial x} X^1 - \frac{\partial X_i^1}{\partial x} X^2\right) \partial_{x_i}.$$

This operation introduces in the space of vector fields the structure of a Lie algebra $VectM$. For $X \in VectM$ the notation adX stands for inner derivation of $VectM$: $(adX)X' = [X, X']$, $\forall X' \in VectM$.

For a diffeomorphism P we put the notation AdP for the following inner automorphism of the Lie algebra $VectM$: $AdPX = P \circ X \circ P^{-1} = P_*^{-1}X$. The last notation stands for the result of translation of the vector field X by the differential of the diffeomorphism P^{-1}.

To introduce topology in the space of vector fields and diffeomorphisms we start with a family of seminorms $\|\cdot\|_{s,K}$ in $C^\infty(R^N)$, where s is a nonegative integer and $K \subset R^N$ is a compact. This family defines in $C^\infty(R^N)$ the *topology of convergence of all derivatives on compacts.* We call a family of functions $t \mapsto \varphi_t$, $(t \in R)$ *measurable* if $\forall x \in R^N$ $t \mapsto \varphi_t(x)$ is measurable. A measurable family is called *locally integrable* if $\int\limits_{t_1}^{t_2} \|\varphi_t\|_{s,K}\, dt < \infty$ $\forall s \geq 0$, $\forall K$, $\forall t_1, t_2 \in R$. A family ω_t is called *absolutely continuous with respect to* t if $\omega_t = \omega_{t_0} + \int\limits_{t_0}^{t} \varphi_\tau\, d\tau$ for some locally inegrable family φ_τ. Since any manifold can be properly embedded into the Euclidean space of a sufficiently big dimension N, one can introduce such a topology (independent on the embedding) in the space $C^\infty(M)$ of smooth functions on M.

As far as we treat the vector fields and the diffeomorphisms as operators on the $C^\infty(M)$ we may introduce the properties of local integrability or absolute continuity for parametrized by t families of the operators in a weak sense (see [3] for details). Thus we call time dependent vector field $t \mapsto X_t$ locally integrable if $t \mapsto X_t\varphi$ is locally integrable for any $\varphi \in C^\infty(M)$. From now on we assume all time-dependent vector fields to be locally integrable. A flow on M is an absolutely continuous family $t \mapsto P_\tau$ of diffeomorphisms, satisfying the condition $P_0 = I$ (where I is the identity diffeomorphism). This means that $\forall\varphi \in C^\infty(M) : \ (P_t\varphi)(q) = \varphi(P_t(q))$ is abosolutely continuous family of functions; $P_0\varphi = \varphi$.

A time-dependent vector field X_τ defines an ordinary differential equation $\dot{q} = X_\tau(q(\tau))$, $q(0) = q^0$ on the manifold M; if solutions of this differential equation exist for all $q^0 \in M, \tau \in R$, then the vector field X_τ is called *complete* and defines a flow on M, being the unique absolutaly continuous solution of the (operator) differential equation:

$$\frac{dP_\tau}{d\tau} = P_\tau \circ X_\tau, \quad P_0 = I. \tag{A.1}$$

This solution will be denoted by $P_t = \overrightarrow{\exp}\int\limits_0^t X_\tau \, d\tau$, and is called (see [3, 5]) a *right chronological exponential* of X_τ. If the vector field $X_\tau \equiv X$ is time-independent, then the corresponding flow is denoted by $P_t = e^{tX}$.

We introduce also Volterra expansion (or Volterra series) for the chronological exponential. It is (see [3, 5]):

$$\overrightarrow{\exp}\int\limits_0^t X_\tau \, d\tau \approx I + \sum_{i=1}^{\infty} \int\limits_0^t d\tau_1 \int\limits_0^{\tau_1} d\tau_2 \ldots \int\limits_0^{\tau_{i-1}} d\tau_i (X_{\tau_i} \circ \cdots \circ X_{\tau_1}).$$

We essentially use the terms of zero-, first- and second-order in this expansion, which are

$$\overrightarrow{\exp}\int\limits_0^t X_\tau \, d\tau \approx I + \int\limits_0^t X_\tau \, d\tau + \int\limits_0^t d\tau_1 \int\limits_0^{\tau_1} d\tau_2 (X_{\tau_2} \circ X_{\tau_1}) + \cdots$$

For time-independent X one obtains

$$e^{tX} \approx I + tX + (t^2/2) X \circ X + \cdots$$

One more tool from chronological calculus is a "generalized variational formula"(see [3, 5] for its drawing):

$$\overrightarrow{\exp}\int\limits_0^t (\hat{X}_\tau + X_\tau) \, d\tau = \overrightarrow{\exp}\int\limits_0^t \hat{X}_\tau \, d\tau \circ \overrightarrow{\exp}\int\limits_0^t Ad(\overrightarrow{\exp}\int\limits_t^\tau \hat{X}_\theta \, d\theta) X_\tau \, d\tau. \tag{A.2}$$

Applying the operator $Ad(\overrightarrow{\exp}\int\limits_0^\tau \hat{X}_\theta\, d\theta)$ to a vector field Y and differentiating $Ad(\overrightarrow{\exp}\int\limits_0^\tau \hat{X}_\theta\, d\theta)Y = (\overrightarrow{\exp}\int\limits_0^\tau \hat{X}_\theta\, d\theta) \circ Y \circ (\overrightarrow{\exp}\int\limits_0^\tau \hat{X}_\theta\, d\theta)^{-1}$ with respect to τ one comes to the equality (see [3, 5]):

$$\frac{d}{d\tau} Ad(\overrightarrow{\exp}\int\limits_0^\tau \hat{X}_\theta\, d\theta Y) = Ad(\overrightarrow{\exp}\int\limits_0^\tau \hat{X}_\theta\, d\theta) ad\hat{X}_\tau Y,$$

which is of the same form as (A.1). Therefore $Ad(\overrightarrow{\exp}\int\limits_0^\tau \hat{X}_\theta\, d\theta)$ can be presented as an operator chronologocal exponential $\overrightarrow{\exp}\int\limits_0^t ad\hat{X}_\theta\, d\theta$ which for a time-independent vector field $\hat{X}_\tau \equiv \hat{X}$ is written as $e^{tad\hat{X}}$. These exponentials also admit Volterra expansions:

$$\overrightarrow{\exp}\int\limits_0^t adX_\tau\, d\tau \approx I + \sum_{i=1}^{\infty} \int\limits_0^t d\tau_1 \int\limits_0^{\tau_1} d\tau_2 \ldots \int\limits_0^{\tau_{i-1}} d\tau_i (adX_{\tau_i} \circ \cdots adX_{\tau_1)} \approx$$

$$\approx I + \int\limits_0^t adX_\tau\, d\tau + \int\limits_0^t d\tau_1 \int\limits_0^{\tau_1} d\tau_2 (adX_{\tau_2} \circ adX_{\tau_1}) + \cdots,$$

and

$$e^{tadX} \approx I + tadX + (t^2/2) adX \circ adX + \cdots.$$

In this new notation the generalized variational formula (A.2) can be rerepresented as:

$$\overrightarrow{\exp}\int\limits_0^t (\hat{X}_\tau + X_\tau)\, d\tau = \overrightarrow{\exp}\int\limits_0^t \hat{X}_\tau\, d\tau \circ \overrightarrow{\exp}\int\limits_0^t (\overrightarrow{\exp}\int\limits_t^\tau ad\hat{X}_\theta\, d\theta) X_\tau\, d\tau =$$

$$= \overrightarrow{\exp}\int\limits_0^t (\overrightarrow{\exp}\int\limits_0^\tau ad\hat{X}_\theta\, d\theta) X_\tau\, d\tau \circ \overrightarrow{\exp}\int\limits_0^t \hat{X}_\tau\, d\tau.$$

References

[1] A. A. Agrachev, *Methods of control theory in nonholonomic geometry,* Proceed. ICM-94, Zürich. Birkhäuser, 1995, pp.1473-1483.

[2] A. A. Agrachev, El-H. C. El Alaoui, J.-P. Gauthier, I. Kupka, *Generic Sub-Riemannian metrics on R^3*, Compt. Rend. Acad. Sci.

[3] A. A. Agrachev, R. V. Gamkrelidze, *Exponential representation of flows and chronological calculus,* Mat. sb. vol.107, 1978, pp.467-532. English translation: Math. USSR Sb. vol.29, 1979, pp.727-785.

[4] A. A. Agrachev, R. V. Gamkrelidze, *Symplectic Methods for optimization and control,* to appear in: "Geometry of Feedback and Optimal Control", Marcel Dekker.

[5] A. A. Agrachev, R. V. Gamkrelidze, A. V. Sarychev, *Local invariants of smooth control systems,* Acta Appl. Math. vol.14, 1989, pp.191-237.

[6] A. A. Agrachev, S. Stefani, P. L. Zezza, *Strong minima in optimal control,* in preparation.

[7] V. I. Arnold, *Mathematical methods in classical mechanics,* 3d ed., Nauka, Moscow 1989.

[8] R. W. Brockett, *Control theory and singular Riemannian geometry,* New Directions in Applied Mathematics, P. J. Hilton and G. S. Young, Eds. Springer Verlag, 1981

[9] Ge Zhong, *Horizontal path space and Carnot–Caratheodory metrics,* Pac. J. Math. vol.161, 1993, pp.255-286.

[10] M. Golubitsky, V. Guillemin, *Stable mappings and their singularities,* Springer Verlag, New York 1973.

[11] M. Gromov, *Carnot–Caratheodory spaces seen from within,* Preprint IHES /M/94/6, 1994, 221p.

[12] R. Montgomery, *The isoholonomic problem and some applications,* Commun. Math. Phys. vol.128, 1990, pp.565-592.

[13] A. V. Sarychev, *The index of the second variation of a control system,* Mat. Sb. vol.113, 1980, pp.464-486. English translation: Math. USSR Sb. vol.41, 1982, pp.383-401.

[14] A. M. Vershik, V. Y. Gershkovich, *Nonholonomic dynamical systems. Geometry of disributions and variational problems,* Itogi nauki: Sovremennye problemy matematiki. Fundamentalnye napravleniya, vol. 16, VINITI, Moscow, 1987, pp.5-85; English translation in Encyclopedia of Mathematical Sciences, vol. 16 (Dynamical Systems, VII), Springer Verlag.

[15] H. Whitney, *On singularities of mappings of Euclidean spaces,* Annals of math. vol.62, 1955.