Journal of Mathematical Sciences, Vol. 103, No. 6, 2001

SUB-RIEMANNIAN METRICS AND ISOPERIMETRIC PROBLEMS IN THE CONTACT CASE

A. A. Agrachev and J. P. A. Gauthier UDC 517.974; 517.972.9

1. Introduction

1. Invariants. A contact sub-Riemannian metric is a triple (X, Δ, g) consisting of a $(2n+1)$-dimensional manifold X, a distribution Δ on X, which is a contact structure, and a metric g on Δ. It defines a metric on X by measuring via g the length of smooth curves that are tangent to Δ. All the considerations in this paper are local: we consider only germs $(X, \Delta, g)_{q_0}$ at a fixed point q_0.

There are two canonical objects that are associated with (X, Δ, g), modulo their sign: the "defining" one-form ω and the "characteristic" vector field ν:

$$\begin{aligned} &\text{(i)} \quad \operatorname{Ker}\omega = \Delta, \quad (d\omega)^n_{|\Delta} = \text{Volume}, \\ &\text{(ii)} \quad \omega(\nu) = 1, \quad i_\nu d\omega = 0, \end{aligned} \tag{1.1}$$

or equivalently

$$\text{(ii}'\text{)} \quad i_\nu(\omega \wedge d\omega) = d\omega.$$

If n is odd and the signs are reversed, the form $\omega \wedge (d\omega)^n$ is unchanged; hence, it defines an orientation on X. In this case, the assignment of a vector field Z transversal to Δ is equivalent to the choice of an orientation on Δ; the volume form on Δ is defined by $\operatorname{sgn}(\omega(Z))(d\omega)^n_{|\Delta}$.

In the opposite case, where n is even, $(d\omega)^n_{|\Delta}$ defines an orientation on Δ, and the orientation on X is given by any Z transversal to Δ; we chose $\omega(Z) > 0$.

First, we will discuss normal forms and invariants for such a structure. We obtain a completely reduced normal form for any n. These results will be stated and proved in Sec. 2 of the paper.

In the three-dimensional case, a certain number of covariant symmetric tensor fields over Δ appear; they are invariants of the sub-Riemannian structure. Two of them, denoted by Q_2 and V_3, are very important, and they have covariance degree 2 and 3, respectively. They are irreducible under the action of $SO(2)$.

Generically, Q_2 is nonzero outside a smooth curve $\mathcal{C}$, and V_3 is nonzero on this curve. They are respectively called the *principal* and *second invariants* of the structure. They reflect the most important local properties of the sub-Riemannian metric.

2. Isoperimetric problems. We also consider general isoperimetric problems on a two-dimensional oriented Riemannian manifold (M, g) of the following type: we are given a two-form $\eta = \psi(\text{Volume})$, two points $q_0, q_1 \in M$ are fixed, together with a curve $\tilde{\gamma} : [0,1] \to M$, $\tilde{\gamma}(0) = q_0$, $\tilde{\gamma}(1) = q_1$, and we are looking for curves $\gamma : [0,1] \to M$, $\gamma(0) = q_1$, $\gamma(1) = q_0$, that have the minimum Riemannian length and are such that the value A of the integral $\int_\Omega \eta$ is prescribed, where Ω is the domain bounded by γ and $\tilde{\gamma}$.

The Dido problem is a particular case where $\psi = 1$, i.e., we want to minimize the perimeter for a prescribed area. (In fact, the classical Dido problem is the dual formulation of this problem.) Also, sometimes, it is called the Pappus problem in honor of the Greek mathematician who solved it in the particular case of a flat metric on a plane (see [9, pp. 366–370]).

We now reformulate the isoperimetric problems under consideration in the following two equivalent ways.

(1) Such an isoperimetric problem (M, g, η) can be reformulated locally in terms of sub-Riemannian geometry as follows: with a germ at q_0 $(M, g, \eta)_{q_0}$, one can associate a germ $(X, \Delta, g)_{p_0}$ of an *oriented*

Translated from Itogi Nauki i Tekhniki, Seriya Sovremennaya Matematika i Ee Prilozheniya. Tematicheskie Obzory. Vol. 64, Pontryagin Conference–3. Geometrical Control Theory, 1999.

1072–3374/01/1036–0639 $ 25.00 © 2001 Plenum Publishing Corporation

three-dimensional sub-Riemannian structure with a symmetry, where $X = M \times R = \{z, w\}, \pi : X \to M$, $p_0 = (q_0, 0)$, and if α is a given primitive of the form η and $\tilde{\gamma} : [0,1] \to X$, $\tilde{\gamma}(0) = (q_0, 0)$, $\tilde{\gamma}(1) = (q_1, w^*)$, is a smooth curve tangent to Δ, then the following conditions hold:

(a) the sub-Riemannian length of $\tilde{\gamma}$ is equal to the Riemannian length of its projection $\gamma = \pi\tilde{\gamma} : [0,1] \to M$,

(b) $w^* = \int_0^1 \alpha(\dot{\gamma}) d\tau$.

This sub-Riemannian metric $(X, g, \Delta)_{p_0}$ has the symmetry $\xi = \dfrac{\partial}{\partial w}$ and the form α, which is unique up to a closed 1-form dn and can be uniquely normalized as follows: we consider all sub-Riemannian geodesics passing through p_0 whose projections are Riemannian geodesics on M. If we require that all such curves be contained in the surface $\{w = 0\} \subset X$, then α is uniquely determined.

Therefore, starting from the isoperimetric problem $(M, g, \eta)_{q_0}$, we constructed a unique sub-Riemannian metric with symmetry $(X, \Delta, g, \xi)_{p_0}$.

This sub-Riemannian structure is contact iff ψ does not vanish; it corresponds to the Dido problem on (M, g) iff $\xi = \nu$, the characteristic vector field.

(2) Now, if we have a germ of a sub-Riemannian structure with symmetry $(X, \Delta, g, \xi)_{p_0}$, where the symmetry ξ is transversal to Δ, then the quotient space $M = X/\xi$ inherits the Riemannian structure g. Up to a certain trivial prolongation along the symmetry, X can be equipped with the structure of a (trivial) principal line bundle over the Riemannian manifold (M, g) with a connection whose horizontal space is Δ. Conversely, if we have (X, M, g, π, ∇), a (circle or line) principal bundle $\pi : X \to M$ with a connection ∇ over a Riemannian manifold (M, g), then we have a sub-Riemannian structure over X determined by the lift of the Riemannian metric to the horizontal space Δ of the connection. By construction, this sub-Riemannian structure has a symmetry transversal to Δ. Also, the curvature form of the connection defines an isoperimetric problem over M.

Since we are at the local level in the paper, it does not make any difference whether we consider a circle or a line bundle.

The sub-Riemannian structure is contact iff the curvature form of the connection does not vanish; it is Dido iff the curvature form is equal to the lift of the volume form.

Depending on the context, we will use these two equivalent reformulations of the isoperimetric problem and use both notations $(X, M, g, \xi)_{p_0}$ and $(X, M, g, \pi, \nabla)_{p_0}$.

Note that these new definitions of isoperimetric sub-Riemannian metrics and Dido metrics work not only for $n = 1$.

One of our goals is to find a (local) *optimal synthesis* for the isoperimetric problems, i.e., for a fixed q_0, for all sufficiently close q_1, and for all sufficiently small values w of the prescribed integral, find all minimum-length curves joining q_0 with q_1.

An equivalent problem is to compute the shape of all associated sub-Riemannian spheres of small radius (that are level sets of the sub-Riemannian distance function to q_0) in the space $\{(z, w)\}$.

This problem will be solved in the paper as a consequence of the treatment of the general case of small sub-Riemannian spheres:

(a) for a generic contact isoperimetric problem, the two main invariants Q_2 and V_3 mentioned above do not vanish simultaneously. Precisely as in the nonisoperimetric case, Q_2 is nonzero outside isolated smooth curves (that are fibers of the projection $(z, w) \to z$ in this case), and V_3 is nonzero on these fibers. As a consequence, there will be very small differences in all the results we show between isoperimetric sub-Riemannian metrics and general sub-Riemannian metrics. In particular, for the optimal synthesis, there will be no difference at all.

(b) On the contrary, for the Dido problem, the invariant Q_2 vanishes identically. Therefore, the results of the generic case are not applicable.

Just as an illustration of the results in this paper, let us state a theorem giving the shape of the (local) optimal synthesis in the Dido case.

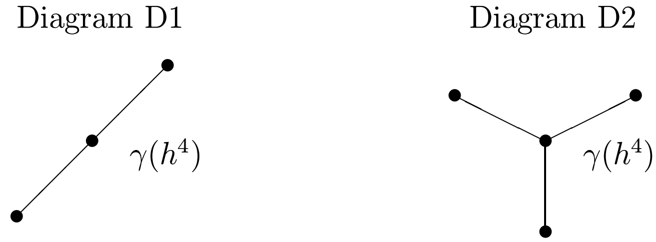

Fig. 1. Generic cut loci.

Again, $q_0 \in M$ is a pole; now ∇ is the Levi-Civita connection on M and $k(q)$ is the Gaussian curvature on M. A denotes a (small) prescribed value of the integral $\int_0^1 \alpha(\dot{\gamma}(\tau))\,d\tau$, where α is the primitive of the volume form described above. We will treat only the case $A > 0$. The case $A < 0$ is similar and is obtained by reversing orientation.

Let us set $h = \sqrt{\dfrac{A}{\pi}}$. The successive covariant derivatives $\nabla^j k$ are covariant symmetric tensor fields of degree j on M, and they can be decomposed under the action of the structural group $SO(2)$ of TM on the fibers of the corresponding vector bundles into isotypic components relative to successive powers $e^{li\varphi}$ of the basic character $e^{i\varphi}$, $i = \sqrt{-1}$:

$$\nabla^j k(q_0) = \sum_{l=0}^{j} (\nabla_l^j k(q_0)). \tag{2.1}$$

In particular, $\nabla^2 k(q_0)$ is a quadratic form, $\nabla^2 k(q_0) = \nabla_0^2 k(q_0) + \nabla_2^2 k(q_0)$, where

$$\nabla_0^2 k(q_0) = \frac{1}{2}\operatorname{trace}_g(\nabla^2 k(q_0)) g(q_0)$$

and $\nabla_2^2 k(q_0) = 0$ iff the discriminant $\operatorname{discr}_g(\nabla^2 k(q_0)) = 0$, $\nabla^3 k(q_0)$ is cubic, $\nabla^3 k(q_0) = \nabla_1^3 k(q_0) + \nabla_3^3 k(q_0)$.

Let us consider the following vectors: V_1, V_2^1, V_2^2, V_3^1, V_3^2, and V_3^3 in $T_{q_0}M$:

(1) V_1 is the vector that is normal to the gradient of k at q_0 with length $\dfrac{\pi}{4}|\operatorname{grad} k|_{q_0}$, and the frame $(\operatorname{grad} k, V_1)_{q_0}$ is direct;

(2) V_2^i are the vectors in the direction at which the quadratic form $\nabla_2^2 k(q_0)$ attain its maximum $\tilde{r}_2$ on the unit circle of length $\dfrac{\pi}{2}\tilde{r}_2$;

(3) V_3^j are the vectors that are normal to the directions I_1, I_2, and I_3 at which the cubic form $\nabla_3^3 k(q_0)$ attains its maximum $\tilde{r}_{32}$ on the unit circle of length $\dfrac{3\pi}{8}\tilde{r}_{32}$, and the frames (I_j, V_3^j) are direct, $j = 1, 2, 3$.

The "cut locus" $\operatorname{Cut} L(h)$ corresponding to the prescribed value h is defined as the subset of M formed by the points q_1 that are joined with q_0 by several (*not unique*) minimum-length trajectories.

Theorem 2.1. *For a germ of a Riemannian metric at q_0 and for h small enough, the following statements hold.*

There exists a germ of a smooth curve $\gamma(t)$ at q_0 such that $\dfrac{d\gamma(t)}{dt}(q_0) = V_1$, and

(1) *if $\nabla_2^2 k(q_0) \neq 0$, then $\operatorname{Cut} L(h)$ is a tree graph formed by two semi-open smooth curve segments emanating from the point $\gamma(h^4)$. The direction of these two segments is V_2^i and their length has the asymptotics $h^5|V_2^i|$;*

(2) *if $\nabla_2^2 k(q_0) = 0$ but $\nabla_3^3 k(q_0) \neq 0$, then $\operatorname{Cut} L(h)$ is a tree graph formed by three semi-open smooth curve segments emanating from the same point $\gamma(h^4)$. The direction of these segments is V_3^i and their length has asymptotics $h^6|V_3^i|$.*

Figure 1 shows two diagrams, D_1 and D_2, which depict the shape of the generic cut loci. Consequently the following assertion holds.

Corollary 2.2 (for generic Riemannian metrics g over M)**.** *There are the following two types of points $q_0 \in M$: $q_1 \in M$ denotes any point sufficiently close to q_0; $\gamma : [0,1] \to M$, $\gamma(0) = q_0$, $\gamma(1) = q_1$, denotes any curve emanating from q_0 with prescribed value A of the integral $\int_0^1 \alpha(\dot\gamma(\tau))\, d\tau$, $|A|$ sufficiently small. Then*

(1) *if $\nabla_2^2 k(q_0) \neq 0$ (generic points), then there are exactly one or two optimal curves γ;*
(2) *if $\nabla_2^2 k(q_0) = 0$ (isolated points), then there are points q_1 with three optimal curves γ from q_0 to q_1. (The optimal means the minimum length.)*

The triple point of case (2) is exactly the point $\gamma(h^4)$; all other points of the semi-open curve segments are double.

Remark 2.1. It follows from the theory of characteristic classes (see, e.g., [20, Chap. 40, p. 204]) that if M is a compact manifold and the Euler characteristic $\kappa(M)$ is nonzero, there are always isolated points of the second type: $\nabla_2^2 k$ defines a field of quadratic form with signature 1 on the complement of the set of these isolated points in M, and the sum of indices of the associated "field of line elements" is $2\kappa(M)$.

3. Equation of geodesics and the exponential mapping. Before going further in this introduction, we need to say a few words about the geodesics and the exponential mapping. Here, we will use the Hamiltonian formalism. There are several other possible ways of writing the equation of sub-Riemannian geodesics. In particular, in the case of isoperimetric problems, it is easily computed that the equation of the projections γ of the sub-Riemannian geodesics $\tilde\gamma$, $\gamma = \pi\tilde\gamma$ onto M is given by

$$k_g(\gamma(s)) = c\psi(\gamma(s)), \tag{3.1}$$

where $k_{g(\cdot)}$ is the geodesic curvature on M and c is a constant.

This equation is also the equation of motion of a charged particle in the magnetic field, and ψ is the value of the magnetic field on M. See our preprint [6], and [17] for more details on this point.

Also, there is a natural connection (called the Rumin connection, see [10, 19]) associated with a contact sub-Riemannian metric, and the equations of geodesics can be written in a very similar way in terms of this connection.

We will say a little more about this topic (the Rumin connection) in Sec. 6 of this paper.

Here, we will consider the Hamiltonian H:

$$H(p) = \frac{1}{2} \sup_{v \in \Delta\setminus\{0\}} \left(\frac{p(v)}{\|v\|}\right)^2. \tag{3.2}$$

$H(p)$ is a positive semi-definite quadratic form on the fibers of $\pi_X : T^*X \to X$ whose kernel is the annihilator of Δ.

Since we consider only the contact case, there is no abnormal geodesic, and all geodesics (i.e., locally C^0 minimizing curves) are the projections of trajectories of the Hamiltonian vector field $\vec{H}$ associated with H onto X. (For the study of several noncontact cases, see [3, 11].)

Let $H_{\frac{1}{2}} = H^{-1}(\frac{1}{2})$ be a level surface of H corresponding to geodesics that are parametrized by the arclength. Since Hamiltonian (3.2) is homogeneous w.r.t. p, the canonical contact structure of the projective cotangent bundle PT^*X corresponds to $H_{\frac{1}{2}}$. Let $\bar{C}_0 \subset T^*X$, $\bar{C}_0 = \pi_X^{-1}(q_0)$, $\pi_X : T^*X \to X$, and let C_0 be the cylinder $C_0 = \bar{C}_0 \cap H_{\frac{1}{2}}$. C_0 is a Legendre manifold for this contact structure, and the Hamiltonian flow preserves this contact structure.

The exponential mapping is the mapping

$$\varepsilon : C_0 \times R^+ \to X, \qquad (p, s) \to \pi_X \circ \exp(s\vec{H}(p)). \tag{3.3}$$

The *wave front* of radius s is $W_s = \varepsilon(C_0, s)$; the *sphere* of radius s is the set $S_s = \{q \in E | d(q, q_0) = s\}$.

Standard arguments (of Filippov's type for instance) show that for s small enough, if $d(q, q_0) = s$, there is at least a geodesic segment of length s joining q_0 with q. Hence $S_s \subset W_s$.

By the homogeneity of H, $\varepsilon(c, \lambda s) = \varepsilon(\lambda c, s)$. Hence we can also consider ε as a map $\varepsilon : \bar{C}_0 \to X, \varepsilon(p) = \pi_X \circ \exp \vec{H}(p)$. In that case, let us denote it by $\bar\varepsilon$.

At the end, we see that one can always consider the exponential mapping as the *suspension of a mapping between two-dimensional manifolds* (Sec. 3.10).

Thus, there are three ways to consider the exponential mapping: if s denotes the arclength parameter, then (1) we can consider $\bar{\varepsilon}$ as a Lagrangian mapping or (2) for s fixed, we can consider ε_s as a Legendre mapping. As we just said, it will also be possible and convenient (3) to consider ε as the suspension of an ordinary smooth mapping between 2-dimensional manifolds. In this case, let us denote it by ε^S.

4. Content of the paper. Section 2 is devoted to normal forms and invariants for contact sub-Riemannian metrics and isoperimetric problems.

Section 3 is devoted to the effective computation of the exponential mapping in all the considered three-dimensional cases, i.e., *contact* general sub-Riemannian metrics, isoperimetric metrics, and finally, the Dido case.

The computations in the general case have already been done in [2, 10], and in more detail in [4]. We merely state the results that we need. Even this is not easy, because the formulas are so long that it is not reasonable to show them completely. In practice, some of these computations in our series of papers have been done by using a formal computation program. In [2, 10], they have been done by hand.

For all details, we refer to these papers. For detailed computation in the Dido case, we refer to [6].

One of the goals of the paper is to study the singularities of ε_s, $\bar{\varepsilon}$, and ε^S, i.e., the singularities of ε as a Legendre or Lagrangian or (suspension of an) ordinary mapping.

As consequences of this study, we are able to describe spheres and wave fronts, together with their singularities, and caustics with their singularities. This is not new: it is mainly the results of [2, 4, 10].

But, what is new, as we said, is a similar study in the Dido case.

Section 4 is devoted to the study of sub-Riemannian caustics, i.e., the set of singular values of $\bar{\varepsilon}$ and ε^S. Stability of both mappings is studied in the Lagrangian and Thom–Mather sense, respectively. This was the purpose of [4]. We simply restate the results. We give complements for the isoperimetric problems.

Section 5 concerns the study of small spheres and wave fronts. It also concerns the Legendre stability of ε_s.

Theorem 2.1 and the Corollary 2.2 above are simply consequences of these results.

2. Normal Forms and Invariants

5. Normal coordinates. The task has already been done by us in the three-dimensional case in [4].

We start with a general $(2n+1)$-dimensional smooth contact sub-Riemannian metric $(X, \Delta, g)_{q_0}$. We assume that we are given a germ of a curve $\Gamma :]-\varepsilon, \varepsilon[\to X$, $\Gamma(0) = q_0$ that is transversal to the distribution. We will define special coordinates "adapted to this curve."

We set

$$A_\Gamma = \bigcup_{t \in]-\varepsilon,\varepsilon[} \left\{ p \in T^*_{\Gamma(t)}X \,\middle|\, p\left(\frac{d\Gamma(t)}{dt}\right) = 0 \right\}.$$

A_Γ is simply the union of all cotangent vectors along Γ that satisfy the Pontryagin transversality conditions w.r.t. Γ.

There are smooth coordinates (z, w) in a neighborhood of q_0 having the following properties:

(1) $\Gamma(t) = (0, t)$;

(2) geodesics passing through $\Gamma(t)$ that satisfy the Pontryagin transversality conditions w.r.t. Γ (i.e., geodesics of the form $\{\pi_X \circ \exp s\vec{H}(p)\}$, $p \in A_\Gamma$) are straight lines contained in the planes $\{w = t\}$;

(3) for s small enough, the cylinders $C_s = \left\{ \sum_{i=1}^{2n} z_i^2 = s^2 \right\}$ are simply the sets

$$C_s = \{q | d(q, \Gamma) = s\},$$

where d denotes the sub-Riemannian distance.

(These properties are easy consequences of the Pontryagin maximum principle.) These coordinates, called *adapted to* Γ *coordinates, are unique up to changes of coordinates of the form*

$$\tilde{z} = U(w)z, \qquad \tilde{w} = w,$$

where $U(w) \in O(2n)$. If these coordinates are compatible with the given orientation of Δ (or X), then they are unique up to $U(w) \in SO(2n)$.

Now we are able to define the "isoperimetric normal coordinates at q_0" for an (oriented) isoperimetric sub-Riemannian metric $(X, \Delta, g, \xi)_{q_0}$.

Definition 5.1. The (z, w) *isoperimetric normal coordinates*, or simply *isoperimetric coordinates* (resp., *Dido coordinates*) correspond to the choice $\Gamma(t) = \exp t\xi(q_0)$, and to the choice of $U(w)$ making $\xi = \dfrac{\partial}{\partial w}$ (resp. $\nu = \dfrac{\partial}{\partial w}$). The coordinates z are standard normal Riemannian coordinates over M (agreeing with the given orientation of M). This choice of $U(w)$ along Γ is unique up to the choice of $U(0) \in SO(2n)$.

In the general case, there is still the job of normalizing $U(w)$ along Γ, although the *canonical choice of* Γ is clear up to sign: we chose $\Gamma(t) = \exp t\nu(q_0)$.

5.1. Normalization of $U(w)$ in the general contact case. First, we work at the points of Γ, i.e., $q_w = (0, w)$.

We are given an orientation (on Δ if n is odd, and on X if n is even, i.e., in both cases, we have a vector field Z transversal to Δ). This completely defines ω and ν. By contactness, the 2-form $d\omega_{|\Delta}$ is nondegenerate. In the z coordinates just defined, $d\omega_{|\Delta}$ has the (skew symmetric) matrix A_w at q_w. Let us make the following assumption: the sub-Riemannian metric is *strongly nondegenerate* at q_0 in the sense that all the eigenvalues of A_w are distinct, $w \in \,]-\varepsilon, \varepsilon[$. For a generic contact sub-Riemannian metric, *this happens at points* q_0 *in an open dense set* of X. Also, *if* $n = 1$, *there is no assumption at all.* Note also that the product of the eigenvalues should be exactly $\left(\dfrac{1}{n!}\right)^2$, since $(d\omega)^n_{|\Delta} =$ Volume. Equivalently, the n positive imaginary parts α_i of these eigenvalues should satisfy $\prod\limits_{i=1}^{n} \alpha_i = \dfrac{1}{n!}$. These α_i are invariants of the sub-Riemannian structure at q_0. If $n = 1$, they are trivial. They are also the only invariants of the nilpotentization at q_0 of the metric. By construction, they are invariant under orientation changes.

The ordered sequence of 2-dimensional real eigenspaces V_i^w corresponding to the increasingly ordered set of these imaginary parts of eigenvalues $\alpha_i(w)$ are subspaces of $\Delta(q_w)$.

Now we go back to our given curve Γ. What we can first do is to reduce the action of $U(w)$ to the action of a maximal torus $T^n(w)$ in $SO(2n)(w)$ (a simultaneous normalization of the nilpotentizations along Γ). Therefore, we can split the coordinate z into (x, y) coordinates such that $V_i^w = V_i = \operatorname{Span}\left\{\dfrac{\partial}{\partial x_i}, \dfrac{\partial}{\partial y_i}\right\}$, $i = 1, \dots, n$, and the coordinates are now uniquely defined up to rotations $U(w)$ that live in a given maximal torus T^n of $SO(2n)$. In these coordinates, $d\omega = \sum\limits_{i=1}^{n} \alpha_i(w) dx_i \wedge dy_i$ at points of Γ.

Definition 5.2. Such coordinates are called Γ-*reduced coordinates.*

Now, let us consider any point $q = (x, y, w)$ in a sufficiently small neighborhood U of q_0. Let us consider the planes $S_{w_0} = \{q | w = w_0\}$. The projection $\Pi_{w_0} : U \to S_{w_0}$, $\Pi_{w_0}(x, y, w) = (x, y, w_0)$, has the differential $T\Pi_{w_0}$. For $q = (z, w)$, $T\Pi_w$ maps linearly and bijectively $\Delta(q)$ onto $T_q S_w$. The image of $g(q)$ under $T\Pi_w$ defines a Riemannian metric on S_w, which is denoted by g_w.

The space $V_i^w = V_i$ can be considered either as 2-dimensional subspaces of S_w or as 3-dimensional subspaces of R^{2n+1}.

We have the following theorem.

Theorem 5.1. (x, y) *are* (*Riemannian*) *normal coordinates at* $(0, 0, w)$ *for all* (S_w, g_w).

The proof is an easy verification, which is left to the reader.

In a neighborhood of q_0, we can consider any direct orthonormal frame field $\mathcal{F}$ for the metric, $\mathcal{F} = (F_1, G_1, \dots, F_n, G_n)$; in a trivial abbreviated way in the Γ-reduced coordinates $(x, y, w) = (z, w)$, it is written as $\mathcal{F} = Q\dfrac{\partial}{\partial z} + L\dfrac{\partial}{\partial w}$. We can choose the frame so that $Q_{i,j} = Q_{j,i}$: we begin with an arbitrary orthonormal frame field $\mathcal{F} = Q\dfrac{\partial}{\partial z} + L\dfrac{\partial}{\partial w}$, and by using a gauge transformation (i.e., a rotation of the frame, $U(z,w) \in SO(2n)$, depending on (z,w)), we simply take the polar decomposition of Q. $U(z,w)$ is unique. Thus, we can assume that Q is symmetric. In fact, there is more than that: the frame obtained is independent of the one from which we started for the following reasons:

(1) $Q_w(z) = Q(z,w)$, in our coordinates, is, by construction, the square root of the inverse of the matrix of the Riemannian metric g_w on S_w, which is well defined;

(2) if ω is the defining form, then ω is written as $\omega = f(z,w)(dw - \mu(z,w)dz)$, $f > 0$, in a unique way, and, since ω is defining, we have $L(z,w) = \mu Q$.

Therefore, this orthonormal frame is completely defined by the Γ-reduced coordinates.

Definition 5.3. This frame is called the *normal frame subordinated to the* Γ*-reduced coordinates.*

The only remaining choice to fix everything (the coordinates and the frame) is therefore the choice of a curve $\tilde{\Gamma} :]-\varepsilon,\varepsilon[\to T^n$, $w \to U(w)$.

We will prove the following theorem.

Theorem 5.2. *Up to* $U(0) \in T^n$*, there is a unique choice of the curve* $\tilde{\Gamma}$ *for which the sectional curvatures relative to the subspaces* $V_i = V_i^w$ *at* $(0,w)$ *of all the metrics* (S_w, g_w) *are zero.*

This could seem surprising. However, there is not only a rotation of coordinates (which of course does not change the sectional curvatures), but there is also the effect of $\dfrac{\partial U}{\partial w}$ on the metrics g_w.

The proof of this theorem is postponed to the next section.

Definition 5.4. The coordinates (x,y,w) thus obtained, $(x,y) = (x_i,y_i)$, $1 \leq i \leq n$, ordered by the α_i, that are unique up to $U(0) \in T^n$, are called the *sub-Riemannian normal coordinates.* The subordinated normal frame is simply called the (*sub-Riemannian*) *normal form* at q_0 of the metric (which is assumed to be strongly nondegenerate).

6. (Sub-Riemannian) normal form. We choose any Γ-reduced coordinate system (x,y,w). Let $((F_i,G_i)$, $i = 1,\ldots,n)$ be an ordered frame field for the metric such that $(F_i,G_i)(0,0,w) = \left(\dfrac{\partial}{\partial x_i}, \dfrac{\partial}{\partial y_i}\right)$ (the order is given by the α_i). The $(2n+1) \times 2n$-matrix formed by this frame field splits into the matrices $Q(q)$ and $L(q)$ of dimension $2n \times 2n$ and $1 \times 2n$, respectively, and we can assume that Q is symmetric. (Abbreviated notations for the frame field are again $Q\dfrac{\partial}{\partial z} + L\dfrac{\partial}{\partial w}$.)

Theorem 6.1. (i) $Q(z,w)z = z$, $L(z,w)z = 0$;

(ii) *conversely, if* $Q\dfrac{\partial}{\partial z} + L\dfrac{\partial}{\partial w}$ *is an orthonormal frame field that satisfies* (i), *then, setting* $\Gamma(t) = (0,t)$, (z,w) *is an "adapted to* Γ *coordinate system."*

Proof. (i) Again, (p,r) denotes the dual coordinates in T^*X, and s is the arclength. Along our special sub-Riemannian geodesics, the adjoint vector has to satisfy $r(s) = 0$ and $p^t(s) = \dot{z}(s)$: by the Pontryagin maximum principle, $p(s)$ satisfies the transversality conditions w.r.t. the cylinders C_s; hence $r(s) = 0$, $p^t(s) = \mu(s)\dot{z}(s)$, and $\mu(s) \neq 0$. Also, $H = \frac{1}{2}$, hence $\|pQ\| = 1$. Otherwise, $\dot{z} = Q(pQ)^t$, $p^t = \mu Q(pQ)^t$, and $pp^t = \mu(pQ)(pQ)^t = \mu = \mu^2\|\dot{z}\|^2 = \mu^2$. Hence $\mu = 1$.

Now, $\dfrac{p^t}{\mu} = \dot{z} = Q^2p^t$ (Q is symmetric) and $p^t = Q^2p^t$. But, also, $\lambda(s)p^t(s) = z(s)$, $\lambda(0) = 0$, $\lambda(s) \neq 0$ for $s \neq 0$. Therefore, $z(s) = Q^2z(s)$ for all s and $(\mathrm{Id} + Q(s))(\mathrm{Id} - Q(s))z(s) = 0$. But, $\mathrm{Id} + Q(0)$ is invertible: $Q(0) = \mathrm{Id}$. Therefore, $(\mathrm{Id} + Q(s))$ is invertible for small s and $Q(s)z(s) = z(s)$. This is true along all our geodesics, so that $Q(z,w)z = z$.

Along these geodesics, $r(s)=0$, but also $\dot{w}(s)=0=-\dfrac{\partial H}{\partial r}=L(pQ)^t$ and $0=LQp^t=\dfrac{1}{\lambda(s)}LQz(s)$ for $s\neq 0$. By continuity again and by considering again all our geodesics, we have $L(z,w)Q(z,w)z=L(z,w)z=0$ for all (z,w).

(ii) It is easy to verify that the trajectories $z(s)=s\dfrac{z_0}{\|z_0\|}$, $p^t(s)=\dfrac{z_0}{\|z_0\|}$, $r(s)=0$, $w(s)=cte$ are solutions to the adjoint equations providing (i). Also, with (i), we find that along these trajectories, s is the arclength. (An obvious continuity argument at $s=0$ is needed.) $\square$

Proposition 6.2. *The conditions* (i) *of Theorem* 6.1 *are invariant under coordinate changes of the form* $\tilde{z}=U(w)z$, $\tilde{w}=w$ *for* $U(w)\in SO(2n)$.

Proof. First, let us show that if (i) is satisfied, then Q is invertible.

We write $Q(z,w)$ as a power series in $z : Q=Q^0(w)+Q^1(z,w)+\cdots+Q^n(z,w)+\cdots$; then $Q(z,w)z=z=Q^0(w)z+Q^1(z,w)z+\cdots$; this shows that $Q^0(w)=\mathrm{Id}$. In particular, for z sufficiently small, $Q(z,w)$ is invertible.

The formulas for these coordinate changes are easily computed. We obtain

$$\begin{aligned}\tilde{Q}&=U[(Q+AzL)(Q-L^tz^tA)]^{\frac{1}{2}}U^t,\\ \tilde{L}&=L(Q-L^tz^tA)U^t\tilde{Q}^{-1},\end{aligned}\tag{6.1}$$

where $A(w)=\dfrac{dU(w)}{dw}$ ($A(w)$ is skew symmetric).

Then $\tilde{Q}^2\tilde{z}=U(Q+AzL)(Q-L^tz^tA)z=U(Q+AzL)(z+L^tz^tAz)$. But $z^tAz=0$, since A is skew-symmetric; hence $\tilde{Q}^2\tilde{z}=Uz=\tilde{z}$. $Q(0)=\mathrm{Id}$, $(\mathrm{Id}+\tilde{Q})$ is invertible, and $\tilde{Q}\tilde{z}=\tilde{z}$.

For the second point, $\tilde{L}\tilde{z}=L(Q-L^tz^tA)U^t\tilde{Q}^{-1}\tilde{z}=L(Q-L^tz^tA)z=Lz=0$. $\square$

The following remark is also interesting and useful.

Proposition 6.3. *Let us again write* $Q(z,w)$ *as a power series in* z; $Q=\mathrm{Id}+Q^1(w,z)+\cdots$; *then,* $Q^1=0$.

Proof. First, for all $1\le i\le j\le 2n$, we have $Q^1_{i,j}(z,w)=\sum\limits_{k=1}^{n}Q^{1,k}_{i,j}(w)z_k$; the relation $Q(z)z=z$ yields $\sum\limits_{j,k=1}^{n}Q^{1,k}_{i,j}(w)z_kz_j=0$, which shows that $Q^{1,k}_{i,j}(w)=-Q^{1,j}_{i,k}(w)$. Now, Q^1 is symmetric; hence, $Q^{1,k}_{j,i}(w)=Q^{1,k}_{i,j}(w)=-Q^{1,j}_{i,k}(w)=-Q^{1,j}_{k,i}(w)=Q^{1,i}_{k,j}(w)=Q^{1,i}_{j,k}(w)=-Q^{1,k}_{j,i}(w)=0$.

Let us examine now how such a frame field should look when $n=1$. We take a Γ-adapted coordinate frame at q_0. In this case, it is automatically Γ-reduced. We know that $Q(0,w)=\mathrm{Id}$. Let us set $Q=\begin{pmatrix}Q_1 & Q_2\\ Q_2 & Q_3\end{pmatrix}$; let $\beta(z,w)=-\dfrac{1}{xy}Q_2(z,w)$ if $x,y\neq 0$. Then, using $Qz=z$, we find $Q_1=1+y^2\beta$ and $Q_3=1+x^2\beta$. Then βxy, $x^2\beta$, and $y^2\beta$ are smooth functions. Applying the following lemma twice, we see that, in fact, β is smooth even at $x=0$ or $y=0$. $\square$

Lemma 6.4 (desingularization)**.** *Let* $f(x,y,w)$ *be any function on* $\mathbf{R}^{p+2}$. *If* xf *and* yf *are smooth, then* f *is smooth.* (*Replace smooth either by* C^∞ *or by* C^ω).

Proof. Left to the reader. $\square$

Therefore, we see that

$$Q(z,w)=\begin{pmatrix}1+y^2\beta(x,y,w) & -xy\beta(x,y,w)\\ -xy\beta(x,y,w) & 1+x^2\beta(x,y,w)\end{pmatrix}.\tag{6.2}$$

Also, let us set $L(x,y,w)=(L^1,L^2)$ and $\gamma(x,y,w)=\dfrac{2}{y}L^1$. Then, using $Lz=0$, we get $L^2=-\dfrac{x}{2}\gamma$ outside $\{y=0\}$, and γx and γy are smooth functions. By the desingularization lemma, γ is smooth.

Also, since $n = 1$, we have $\alpha_1 = \alpha = 1$. Writing again $\omega = f(dw + \mu dz)$ for the defining form, we have $f > 0$; since $\omega(\nu) = 1$, we conclude that $f(0, 0, w) = 1$; this shows (after a small computation) that $\gamma(0, 0, w) = 1$.

Thus, in dimension 3, we have

$$\begin{aligned} F &= (1 + y^2\beta(x, y, w))\frac{\partial}{\partial x} - xy\beta(x, y, w)\frac{\partial}{\partial y} + \frac{y}{2}\gamma(x, y, w)\frac{\partial}{\partial w}, \\ G &= -xy\beta(x, y, w)\frac{\partial}{\partial x} + (1 + x^2\beta(x, y, w))\frac{\partial}{\partial y} - \frac{x}{2}\gamma(x, y, w)\frac{\partial}{\partial w}, \\ \gamma(0, 0, w) &= 1. \end{aligned} \tag{6.3}$$

By definition, the matrix $Q(x, y, w)$ is such that $Q\dfrac{\partial}{\partial z}$ is an orthonormal frame field for the metric g_w on the plane S_w. A direct computation shows that the Gaussian curvature (which is also the sectional curvature at $(0, 0, w)$ relative to the subspace $V_1 = V_1^w$ in that case) is

$$k_{g_w}(0, 0) = 6\beta(0, 0, w). \tag{6.4}$$

Therefore, by our normalization of the coordinates in Theorem 5.2, it should be zero.

It is now time to prove Theorem 5.2 of the previous section. To this end, we need a technical lemma.

Lemma 6.5. *The sectional curvature* $\mathrm{Sec}_0(V_i^w)$ *at* $(0, 0, w)$, *relative to the* 2*-dimensional subspaces* V_i^w *of the* (S^w, g_w) *satisfies*

$$\mathrm{Sec}_0(V_i^w) = \left(\frac{\partial^2 Q^{i,i}_{2,2}}{(\partial x_i)^2} + \frac{\partial^2 Q^{i,i}_{1,1}}{(\partial y_i)^2} - 2\frac{\partial^2 Q^{i,i}_{1,2}}{\partial y_i \partial x_i}\right) \quad (0, 0, w),$$

where Q *is* 2×2*-block,* $Q = \mathrm{Block}(Q^{i,j})$.

Proof. For a Riemannian metric g, in normal coordinates, we have $g = Q^{-2}$, where Q is symmetric matrix of an orthonormal frame field and $g(0) = \mathrm{Id}$. The proof of Theorem 6.1 also works for Riemannian metrics, that is, to show that $Q(z)z = z$. If $Q(z) = \mathrm{Id} + Q_1(z) + Q_2(z) + \cdots$, where Q_i has order i in z, then $Q_1(z) = 0$ (Proposition 6.3). Hence, $g(z) = \mathrm{Id} - 2Q_2(z) + O^3(z)$. Now, computations in normal coordinates show that

$$2\,\mathrm{Sec}_0\left(\mathrm{Span}\left\{\frac{\partial}{\partial z_k}, \frac{\partial}{\partial z_l}\right\}\right) = -\left(\frac{\partial^2 g_{l,l}}{(\partial z_k)^2} + \frac{\partial^2 g_{k,k}}{(\partial z_l)^2} - 2\frac{\partial^2 g_{k,l}}{\partial z_l \partial z_k}\right).$$

The result follows. $\square$

Notation. Below, for any function $\varphi(z)$, the notation $\varphi(\hat{z}_i)$ means $\varphi(0, \dots, 0, z_i, 0, \dots, 0)$, i.e., φ all of whose components (x_j, y_j) except (x_i, y_i) are zero.

Proof of Theorem 5.2. The relation $Q(z, w)z = z$ implies $Q^{i,i}(\hat{z}_i, w)z_i = z_i$, $1 \le i \le n$. As a consequence, $Q^{i,i}(\hat{z}_i, w)$ has the form (6.2). In the same way, if $L(z, w) = \mathrm{Block}(L^i)$, then L^i satisfies the following:

$$L^i(\hat{z}_i, w) = \left(\frac{y_i}{2}, -\frac{x_i}{2}\right)\gamma_i(\hat{z}_i, w), \qquad 1 \le i \le n. \tag{6.5}$$

Also, $\gamma_i(0, w) \neq 0 : \gamma_i(0, w)$ is exactly the eigenvalue $\alpha_i(w)$ considered above.

For the same reason, we have $Q^{j,i}(\hat{z}_i, w)z_i = 0$; hence, $Q^{i,j}(\hat{z}_i, w)Q^{j,i}(\hat{z}_i, w)z_i = 0$. With the same reasoning as before, this implies

$$Q^{i,j}(\hat{z}_i, w)Q^{j,i}(\hat{z}_i, w) = \begin{pmatrix} y_i^2\beta_{i,j}(z_i, w) & -x_i y_i\beta_{i,j}(z_i, w) \\ -x_i y_i\beta_{i,j}(z_i, w) & x_i^2\beta_{i,j}(z_i, w) \end{pmatrix}. \tag{6.6}$$

Otherwise, the coordinate changes $\tilde{z} = U(w)z$, $U(w) = \mathrm{Block}(e^{J\delta_i(w)})$, $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in T^n$, act as indicated in formula (6.1).

The application of this formula allows us to compute $\tilde{Q}^{i,i}(\widehat{z}_i)$. We will compute it up to terms of order 3 only, since we want to apply Lemma 6.5 at the end:

$$\tilde{Q}^{i,i}(\widehat{z}_i) = e^{J\delta_i(w)} \left((Q^{i,i} + A_i z_i L^i)(Q^{i,i} - L^{it} z_i^t A_i) + \sum_{j \neq i} Q^{i,j} Q^{j,i} \right)^{\frac{1}{2}} e^{-J\delta_i(w)} + O^3(z_i) \tag{6.7}$$

with $A_i = J \dfrac{d\delta_i}{dw}$.

Using specific forms of $Q^{i,i}(\widehat{z}_i)$, $L^i(\hat{z}_i)$, A_i, and $Q^{i,i}(\widehat{z}_i)$ and computing a bit, we obtain

$$\tilde{Q}^{i,i}(\widehat{z}_i) = e^{J\delta_i(w)} \begin{pmatrix} 1 + y_i^2 \hat{\beta}(z_i, w) & -x_i y_i \hat{\beta}(z_i, w) \\ -x_i y_i \hat{\beta}(z_i, w) & 1 + x_i^2 \hat{\beta}(z_i, w) \end{pmatrix}^{\frac{1}{2}} e^{-J\delta_i(w)} + O^3(z_i)$$

with $\hat{\beta}(0, w) = \sum\limits_{j \neq i} \beta^{i,j}(0, w) + 2 \left(\beta_i(0, w) - \dfrac{d\delta_i}{dw}(0, w) \dfrac{\gamma_i}{2}(0, w) \right)$.

This shows that we can find $\delta_i(w)$ independently and in a unique way up to $\delta_i(0)$ for $\hat{\beta}(0, w) = 0$ and for all w small enough. Hence, after the change of variables $\tilde{z} = U(w)z$, we obtain

$$\tilde{Q}^{i,i}(\widehat{\tilde{z}}_i) = \begin{pmatrix} 1 + \tilde{y}_i^2 \hat{\beta}(\tilde{z}_i, w) & -\tilde{x}_i \tilde{y}_i \hat{\beta}(\tilde{z}_i, w) \\ -\tilde{x}_i \tilde{y}_i \hat{\beta}(\tilde{z}_i, w) & 1 + \tilde{x}_i^2 \hat{\beta}(\tilde{z}_i, w) \end{pmatrix} + O^3(\tilde{z}_i)$$

with $\hat{\beta}(0, w) = 0$.

By Lemma 6.5, this shows the result. $\square$

We can now again use the fact that the curve Γ is simply $\{\exp(t\nu)(q_0)\}$ in order to obtain additional conditions on L.

Taking the Taylor expansion of L w.r.t. $z = (x, y)$, we write $L(z, w) = L^1(w, z) + L^2(w, z) + \cdots$, where L^1 is linear and L^2 is quadratic in z. We already know that

$$L_i^1(w) = \alpha_i(w) \left(\frac{y_i}{2}, -\frac{x_i}{2} \right). \tag{6.8}$$

Setting $L^2 dz = \sum\limits_{i=1}^{2n} L_i^2 dz_i$, we see that $d(L^2 dz)$ is written as $d(L^2 dz) = \sum\limits_{j=1}^{n} \tilde{L}_j^2(x, y, w) dx_j \wedge dy_j +$ other terms, where $\tilde{L}_j^2(x, y, w) = \tilde{L}_j^2(z, w)$ is linear in z.

Computations show that the fact that $i_\nu(d\omega) = 0$ is equivalent to

$$\sum_{j=1}^{n} \frac{\partial}{\partial y_k} (\tilde{L}_j^2(z, w))_{|z=0} \prod_{i=1}^{n} \{\alpha_i(w)\}_{i \neq j} = 0, \tag{6.9}$$

$$\sum_{j=1}^{n} \frac{\partial}{\partial x_k} (\tilde{L}_j^2(z, w))_{|z=0} \prod_{i=1}^{n} \{\alpha_i(w)\}_{i \neq j} = 0, \quad 1 \leq k \leq n. \tag{6.10}$$

We have proved the following theorem.

Theorem 6.6 (normal form). *Assume that $(X, \Delta, g)_{q_0}$ is a strongly nondegenerate germ of a contact sub-Riemannian metric. Then, up to sign for the defining form ω and the characteristic vector field ν, there is a coordinate system (x, y, w), $z = (x, y)$, (unique up to an element of the maximal torus T^n of $SO(2n)$ defined by $d\omega$) and an orthonormal frame field $Q\dfrac{\partial}{\partial z} + L\dfrac{\partial}{\partial w}$ (unique once the coordinates are chosen) such that*

(0) *Q is symmetric,*

(1) *$Q(z, w)z = z$, $L(z, w)z = 0$, $Q = \mathrm{Block}(Q_{i,j})$, $L = \mathrm{Block}(L_i)$, $1 \leq i \leq j \leq n$,*

(2) (i) $Q_{ii}(\hat{z}_i, w) = \begin{pmatrix} 1 + y_i^2 \beta_i(z_i, w) & -x_i y_i \beta_i(z_i, w) \\ -x_i y_i \beta_i(z_i, w) & 1 + x_i^2 \beta_i(z_i, w) \end{pmatrix}$,

(ii) $\beta_i(0, w) = 0$,

(3) $L_i(z,w) = L_i^1(w,z) + L_i^2(w,z) + \text{higher order in } z$, $L_i^1 = \alpha_i(w)\left(\dfrac{y_i}{2}\dfrac{\partial}{\partial x_i} - \dfrac{x_i}{2}\dfrac{\partial}{\partial y_i}\right)$, *$L_i^2$ has order 2 in z,*
$d(\sum L_i^2 dz_i) = \sum \tilde{L}_i dx_i \wedge dy_i + \text{other terms}$,

(i) $L^i(\hat{z}_i, w) = \alpha_i(z_i, w)\left(\dfrac{y_i}{2}\dfrac{\partial}{\partial x_i} - \dfrac{x_i}{2}\dfrac{\partial}{\partial y_i}\right)$,

(ii) $\dfrac{1}{n!} = \prod_{i=1}^{n} \alpha_i(0,w)$,

(iii) $\sum_{j=1}^n \dfrac{\partial \tilde{L}_i}{\partial y_k}(0,w)\prod_{i=1}^{n}\{\alpha_i(0,w)\}_{i\neq j} = \sum_{j=1}^{n}\dfrac{\partial \tilde{L}_i}{\partial x_k}(0,w)\prod_{i=1}^{n}\{\alpha_i(0,w)\}_{i\neq j} = 0,\ 1 \le k \le n.$

If (x,y,w) and $Q\dfrac{\partial}{\partial z} + L\dfrac{\partial}{\partial w}$ are the coordinates and frame field such that (0)–(3) are satisfied, then (x,y,w) are normal coordinates and $Q\dfrac{\partial}{\partial z} + L\dfrac{\partial}{\partial w}$ is the corresponding normal frame field.

Also, it follows from our discussion that the following theorem holds.

Theorem 6.7 (invariance of the normal form)**.** *The normal form is invariant under the action of T^n on $\Delta(0)$. Precisely, if z is $U\tilde{z}$, then Q is replaced by $U^tQ(U\tilde{z},w)U$, and L is replaced by $L(U\tilde{z},w)U$.*

We restate these theorems in the special case $n = 1$ (this is not new; see [4]).

Theorem 6.8 (3-dimensional normal form)**.** *In the normal coordinates, the normal frame field is written as follows:*

$$\begin{aligned} F &= (1+y^2\beta(x,y,w))\frac{\partial}{\partial x} - xy\beta(x,y,w)\frac{\partial}{\partial y} + \frac{y}{2}\gamma(x,y,w)\frac{\partial}{\partial w}, \\ G &= -xy\beta(x,y,w)\frac{\partial}{\partial x} + (1+x^2\beta(x,y,w))\frac{\partial}{\partial y} - \frac{x}{2}\gamma(x,y,w)\frac{\partial}{\partial w}, \end{aligned} \tag{6.11}$$

with the "boundary conditions"

$$\begin{aligned} \beta(0,0,w) = 0, &\qquad \gamma(0,0,w) = 1, \\ \frac{\partial\gamma}{\partial x}(0,0,w) &= \frac{\partial\gamma}{\partial y}(0,0,w) = 0. \end{aligned} \tag{6.12}$$

These normal coordinates are unique up to a rotation of $SO(2)$ in $\Delta(0)$ (and changing orientation on Δ changes w to $-w$). Once the normal coordinates are chosen, the normal frame is unique, and the normal form is invariant under the action of $SO(2)$: if $U \in SO(2)$ is taken, and if we set $z = U\tilde{z} = z$, β and γ are replaced by $\beta \circ U$ and $\gamma \circ U$, respectively.

To conclude, let us restate our results about isoperimetric normal forms. Let us do this only in the 3-dimensional case, but note that it is possible to do the same for $n > 1$ (i.e., reducing to the action of a maximal torus in $SO(2n)$).

Remark 6.1. One can verify that in this way, we obtain normal forms and complete invariants for "Riemannian-symplectic" structures.

Theorem 6.9 (3-dimensional isoperimetric normal form)**.** *Let $(X,\Delta,g,\xi)_{q_0}$ be an isoperimetric structure. In isoperimetric coordinates (that are unique up to a rotation in $T_{p0}M$, $M = X\backslash\xi$, $\Pi : X \to M$, $\Pi(q_0) = p_0$), there is a unique orthonormal frame field of the form*

$$\begin{aligned} F &= (1+y^2\beta(x,y))\frac{\partial}{\partial x} - xy\beta(x,y)\frac{\partial}{\partial y} + \frac{y}{2}\gamma(x,y)\frac{\partial}{\partial w}, \\ G &= -xy\beta(x,y)\frac{\partial}{\partial x} + (1+x^2\beta(x,y))\frac{\partial}{\partial y} - \frac{x}{2}\gamma(x,y)\frac{\partial}{\partial w}, \end{aligned} \tag{6.13}$$

without any boundary condition, but with $\xi = \dfrac{\partial}{\partial w}$. This normal form is invariant under the action of $SO(2)$: $U \in SO(2)$ changes β and γ to $\beta \circ U$ and $\gamma \circ U$.

Theorem 6.10 (3-dimensional Dido normal form). *Let $(X,\Delta,g)_{p_0}$ be the Dido metric. Then, we have the same statement as the one of Theorem* 6.9, *and, moreover,*

$$\gamma(x,y)=(1+(x^2+y^2)\beta(x,y))\int_0^1\frac{2t\,dt}{1+t^2(x^2+y^2)\beta(tx,ty)}. \tag{6.14}$$

Also, $\beta(0,0)=\frac{1}{6}k(p_0), k(p_0)$ is the Gaussian curvature of M at p_0.

Proof. Left to the reader (a straightforward computation simply using the fact that $\xi=\nu$). $\square$

7. Invariants. All normal forms in the previous theorems allow us to define tensor fields that are invariants of the sub-Riemannian, or isoperimetric, or Dido structures (because of the invariance of the normal forms).

7.1. The general case. Let us begin with the general case, that is, with any n. We make the following identifications:

$$TX=\Delta\oplus_X R\nu,\qquad T^*X=\Delta^*\oplus_X\Delta_0,$$

where Δ^* is the dual of Δ and Δ_0 is the annihilator of Δ in T^*M ($\Delta_0=R\omega$).

If V is a vector space, we denote by S^kV^* the set of covariant symmetric tensors of degree k on V, which can be canonically identified with homogeneous polynomials of degree k on V^*.

Given an orientation (on Δ if n is odd and on X if n is even) together with a normal coordinate system at q_0, consider $Q\dfrac{\partial}{\partial z}+L\dfrac{\partial}{\partial w}$, the normal form of our metric. For any two vectors V_1 and V_2 in Δ_{q_0}, we can consider the values $A_1(z,w,V_1,V_2)=V_1^tQ(z,w)V_2$, and $A_2(z,w,V_1)=L(z,w)V_1$, (identifying V_1 and V_2 with their representants in normal coordinates). For each $k>0$, let us consider $A_k^i(z,V_1,V_2)=\dfrac{\partial^k A^i}{\partial w^k}_{|w=0}$ and $B_{k,l}^i(V_1)=D^lA_k^i(z)_{|z=0}$, the kth derivative with respect to z at 0 of $A_k^i(z)$. $B_{k,l}^1$ is an element of $S^k\Delta_{q_0}^*\otimes S^2\Delta_{q_0}^*$ and $B_{k,l}^2$ is an element of $S^k\Delta_{q_0}^*\otimes\Delta_{q_0}^*$. For a defining form ω, we set

$$\tilde{B}_{k,l}^i=\begin{cases}B_{k,l}^i & \text{if } k \text{ is even,}\\ B_{k,l}^i\otimes\omega & \text{if } k \text{ is odd.}\end{cases}$$

Then $\tilde{B}_{k,l}^i$, $i=1,2$, do not depend on the choice of U in T^n and on the orientation. The first fact comes from Theorem 6.7. Second, if we change orientation, w is changed to $-w$, but ω is changed to $-\omega$.

Now if we move the base point q_0 of this construction, we obtain tensor fields, still denoted by $\tilde{B}_{k,l}^i$, that are invariants of the sub-Riemannian structure.

Now, if J is any subbundle of the tensor bundle $\bigotimes^{p,q}\Delta$ (p times contravariant, q times covariant), the abelian group T^n acts on its typical fiber $J(0)$. $J(0)$ as a T^n-module can be decomposed into (1- or 2-dimensional) real irreducible components, $J(0)=\bigoplus\{J_k(0)|k\in Z^n,\ k$ not necessarily distinct$\}$, $J_k(0)$ corresponds to the characters $e^{\varepsilon i(k_1\theta_1+\cdots+k_n\theta_n)}$, $\varepsilon=+1,-1$.

In the case $n=1$, these considerations can be made more precise. This is the purpose of the following section.

7.2. The case $n=1$. In this case, it is more convenient to define these invariants in another (but equivalent) way: we can consider functions β and γ that appear in the normal form (6.11). By their invariance and by the same considerations for orientation, the following tensor fields are invariants of our sub-Riemannian structure: we define the tensors at the pole q_0 via the normal form at q_0. We set $\beta^k=\dfrac{\partial^k\beta}{\partial w^k}_{|w=0}$, $\gamma^k=\dfrac{\partial^k\gamma}{\partial w^k}_{|w=0}$ and we take the lth derivatives with respect to z at $z=0$, β_l^k, γ_l^k. If k is odd, then we tensorize with ω. (We still denote these tensors by β_l^k and γ_l^k in this case.)

If k is even, then β_l^k, $\gamma_l^k\in S^l\Delta^*$, and if k is odd, they belong to $J_l=S^l\Delta^*\otimes T^*X$.

We have $T^*X=\Delta^*\oplus_X R\omega$, and the structural group $SO(2)$ of $S^k\Delta^*$ acts on the typical fiber $J_k(q_0)$ of J_k in a trivial way. Hence all the tensor bundles considered above have decompositions into isotypic components. If N is an $SO(2)$-module, we denote by $(N)_k$ its isotypic component corresponding to the kth power of the basic character $\chi = e^{i\theta}$ (i.e., the direct sum of the real irreducible components corresponding to χ^k).

The following decompositions are the most important in the remaining part of the paper:

$$S^2\Delta^* = (S^2\Delta^*)_0 \oplus (S^2\Delta^*)_2, \qquad S^3\Delta^* = (S^3\Delta^*)_1 \oplus (S^3\Delta^*)_3.$$

$(S^2\Delta^*)_0$ can be identified with the line bundle Rg, $(S^3\Delta^*)_1$ endowed with the symmetric product $\Delta^* \odot Rg$. Then $(S^2\Delta^*)_2$ (resp. $(S^3\Delta^*)_3$) is simply the orthogonal complement of $(S^2\Delta^*)_0$ (resp. $(S^3\Delta^*)_1$) with respect to the metric induced by g on $S^2\Delta^*$ (resp. $S^3\Delta^*$).

For instance, if $q \in S^2\Delta^*$, then $q = (q)_0 + (q)_2$, $(q)_0 = \frac{1}{2}\,\mathrm{trace}_g(q)g$ and $\mathrm{trace}_g((q)_2) = 0$.

More generally, for $\varepsilon = 0, 1$, $S^{2p+\varepsilon}\Delta^* = \bigoplus_{k=0}^{p} (S^{2p+\varepsilon}\Delta^*)_{2k+\varepsilon}$, and if $q \in S^{2p+\varepsilon}\Delta^*$, then $q = (q)_\varepsilon + \cdots + (q)_{2p+\varepsilon}$. If one considers q as a homogeneous polynomial of degree p, then, being restricted to the unit circle in $\Delta^*_{q_0}$, it is a function $q(\theta)$ of the angle θ. The above decomposition is simply the (finite) Fourier series of $q(\theta)$, and $(q)_{2p+\varepsilon}$ is the "highest harmonic." In this study, the "highest harmonics" always play a crucial role as we will see.

Two of these tensors play a particularly important role in this paper; hence, we give them a special name: $Q = \gamma_2^0$ and $V = \gamma_3^0$. They belong to $S^2\Delta^*$ and $S^3\Delta^*$, respectively. Their highest harmonics are $Q_2 \in (S^2\Delta^*)_2$ and $V_3 \in (S^3\Delta^*)_3$ and were called the *principal and second invariant* in the introduction.

Theorem 7.1 (genericity)**.** *X is given. There is an open-dense set* (*in the Whitney topology*) *among the contact sub-Riemannian metrics on X such that the principal invariant Q_2 is nonzero outside a smooth* (*possibly empty*) *curve. On this smooth curve, the second invariant V_3 is nonzero.*

The proof is left to the reader. For more details, see [10].

7.3. The 3-dimensional isoperimetric case. Let us go back to the original statement of an isoperimetric problem in the 2-dimensional case. It is a triple $(M, g, \eta)_{q_0}$. $\eta = \psi$Volume. The sub-Riemannian formulation (X, Δ, g, ξ) is contact iff ψ does not vanish.

In the same way, we will consider S^kT^*M, the bundle of covariant symmetric tensors of degree k on TM. The structural group $SO(2)$ of TM acts on a typical fiber of this bundle, and we again have the following decompositions into isotypic components:

$$S^kT^*M = \bigoplus_M (S^kT^*M)_j.$$

If $\Pi_X : X \to M$ is the canonical projection, $q_0 \in X$, and $p_0 = \Pi_X(q_0) \in M$, then an element $T \in S^kT^*_{p_0}M$ is pulled back linearly to an element $\Pi_X^*(T) \in S^k\Delta^*_{q_0}$. Both bundles S^kT^*M and $S^k\Delta^*$ have the same fiber with the same action of $SO(2)$. Hence the typical fibers of both bundles have the same decomposition. The pull back mapping commutes with these decompositions, i.e., if $T \in (S^kT^*_{p_0}M)_j$, then $\Pi_X^*(T) \in (S^k\Delta^*_{q_0})_j$.

Fix $p_0 \in M$ and $q_0 \in (\Pi_X)^{-1}(p_0)$. Set $x_2 = \nabla^2 \log\psi$, $x_3 = \nabla^3 \log\psi$. $x_2 \in S^2T^*_{p_0}M$ and $x_3 \in S^3T^*_{p_0}M$. The decompositions of x_2 and x_3 (as elements of $S^2\Delta^*$ and $S^3\Delta^*$, in particular) yield $\Pi_X^*(x_2) = (x_2)_0 + (x_2)_2$ and $\Pi_X^*(x_3) = (x_3)_1 + (x_3)_3$.

Computations show the following.

Theorem 7.2. *In the contact isoperimetric case, Q_2 is a nonzero multiplier of $(x_2)_2$ and V_3 is a nonzero multiplier of $(x_3)_3$.*

Also, it is clear that for an open dense set of tuples (g, ψ) (among the open set of (g, ψ) with nonvanishing ψ) on a 2-dimensional manifold M, $(x_2)_2$ is nonzero, except for isolated points where $(x^3)_3$ is nonzero.

Therefore (except for small details due to the symmetry, which we will explain later on), the *study of generic isoperimetric problems is equivalent to the study of generic sub-Riemannian problems* (since, as we will see, most of the properties that we study are completely determined by Q_2 if it is nonzero, or V_3 if $Q_2 = 0$).

7.4. Dido case (3-dimensional). On the contrary, *for the Dido problem*, a computation with the formula (6.14) and the isoperimetric normal form (6.13) show that Q_2 is *identically zero.*

Of course, in the Dido case the only tensor fields of interest will be $\nabla^r k$, where k is the Gaussian curvature on M. For convenience in computations and coherence with the general sub-Riemannian case, we will prefer the invariants defined by β in the isoperimetric normal form (6.13). Clearly, β can be considered either as a function on the quotient $M = X\backslash\xi$ or as a function on X. The confusion will be not substantial because of the identification of $S^k T^*_{p_0} M$ with $S^k \Delta^*_{q_0}$ and their respective decompositions done in the previous section.

Again, because of the invariance of the normal form, the derivatives $D^k\beta$ with respect to z taken at $z = 0$ define elements $\beta^k \in S^k\Delta^*_{q_0}$ (or $S^k T^*_{p_0} M$). Moving the base points p_0, q_0, we obtain tensor fields that are invariants of the isoperimetric structure (they are also invariant w.r.t. to changes of orientation). The most important among β^r will be β^2 and β^3. They are related to the $\nabla^r k$ as follows:

Theorem 7.3 (invariants in the Dido case)**.** *The "highest harmonics" $(\nabla^r k)_r$ and $(\beta^r)_r$ are nonzero constant multipliers for $r = 2, 3$:*

$$(\nabla^2 k)_2 = 20(\beta^2)_2, \qquad (\nabla^3 k)_3 = 180(\beta^3)_3.$$

For more details, see [6].

3. Computation of the Exponential Mapping

In all of this section, we restrict ourselves to the 3*-dimensional case.*

We know that the sub-Riemannian geodesics are projections of the trajectories of the Hamiltonian vector field $\vec{H}$ onto X. Here, *let us present the computations in the general sub-Riemannian case and in the Dido case only.* As we said in Sec. 2.7.3 for the main results, it is not necessary to make the computations in the general isoperimetric case: they are direct consequences of the results in the general nonisoperimetric case. However, for some details that we will show in Sec. 4, it is necessary to do the computations. They have been done in [8].

We consider a sub-Riemannian metric that is either general or Dido. In the general case, we work in normal coordinates with the normal form (6.11) and with the boundary conditions (6.12). In the Dido case, we work in isoperimetric normal coordinates, with the isoperimetric normal form, and with relation (6.14).

In both cases, the dual coordinates in T^*X of the normal coordinates (x, y, w) (isoperimetric or Dido) are denoted by $(\tilde{p}, \tilde{q}, r)$.

8. Reparametrization of geodesics. Let us denote the arclength by s and reparametrize the time. The *new time* will be t, defined by $dt = r(s)ds$. In the Dido case, r is constant; hence, $t = rs$. We set

$$p = \frac{\tilde{p}}{r}, \qquad q = \frac{\tilde{q}}{r}, \qquad \rho = \frac{1}{r}. \tag{8.1}$$

Along a geodesic, let us write ρ for $\rho(t)$ at $t = 0 : \rho = \rho(0)$. When using ρ at the new time t, we always use explicitly the notation $\rho(t)$.

In these coordinates, the cylinder $C_0 = \{p = \rho\cos(\varphi),\ q = \rho\sin(\varphi)\}$, and ε is a mapping of the variables (ρ, φ, t), *which is smooth* even at $\rho = 0$. We denote it again by $\varepsilon(\rho, \varphi, t)$. For $z = (x, y)$, ε_z denotes the first two components of ε, and ε_w denotes the third one. $\varepsilon_z, \varepsilon_w$ have expansions in terms of ρ at $\rho = 0$ of the form

$$\varepsilon_z = \rho\varepsilon_1^z(t,\varphi) + \rho^3\varepsilon_3^z(t,\varphi) + \cdots + \rho^n\varepsilon_n^z(t,\varphi) + o(\rho^{n+1}) = \tilde{\varepsilon}_n^z + o(\rho^{n+1}). \tag{8.2}$$

Remark 8.1. Note the very important point that the *term of order* 2 *is missing.* This is due (i) to Proposition 6.3 and (ii) to the boundary conditions on γ in the general case and to the special form of γ in the Dido case. *This is not true in the general isoperimetric case for the isoperimetric coordinates* (*which are the natural ones*). However, we decided to forget these coordinates and to work in normal nonisoperimetric coordinates in this case. This results in the fact that the w component is not exactly the value of the prescribed integral, but this does not change the general shape of spheres, wave fronts, and conjugate loci:

$$\varepsilon_w = \rho^2\varepsilon_2^w(t,\varphi) + \rho^4\varepsilon_4^w(t,\varphi) + \cdots + \rho^n\varepsilon_n^w(t,\varphi) + o(\rho^{n+1}) = \tilde{\varepsilon}_n^w + o(\rho^{n+1}). \tag{8.3}$$

Remark 8.2. The term ε_3^w is missing. As in the remark above, this is not true for general isoperimetric problems in isoperimetric coordinates with the same consequences in the interpretation of the results on spheres, wave fronts, and optimal synthesis.

Also, let us set $\varepsilon_1 = (\rho\varepsilon_1^z(t,\varphi), \rho^2\varepsilon_2^w(t,\varphi)$, $\bar{\varepsilon}_n = (\tilde{\varepsilon}_n^z(t,\varphi), \tilde{\varepsilon}_{n+1}^w(t,\varphi))$, and $\hat{\varepsilon}_n = (\tilde{\varepsilon}_n^z(t,\varphi), \tilde{\varepsilon}_n^w(t,\varphi))$.

Remark 8.3. ε_1 is simply the exponential mapping of the "Heisenberg" right-invariant metric, which is the basic model. Everything will be computed by using formulas (8.2) and (8.3) as a perturbation of this basic (and totally degenerate) exponential mapping.

ε_1 can be computed easily:

$$\begin{aligned}\varepsilon_1(\rho,\varphi,t) &= (\rho\varepsilon_1^z(t,\varphi), \rho^2\varepsilon_2^w(t,\varphi)),\\ \varepsilon_1^z(t,\varphi) &= (2\cos(\varphi - t/2)\sin(t/2), 2\sin(\varphi - t/2)\sin(t/2)),\\ \varepsilon_2^w(t,\varphi) &= (t-\sin(t))/2.\end{aligned} \tag{8.4}$$

Let us now show how to compute the other terms that we need in formulas (8.2) and (8.3).

We set $\sigma = (x,y,p,q)$. Then our geodesics have the expansion

$$\begin{aligned}\sigma(\rho,\varphi,t) &= \sigma^n(\rho,\varphi,t) + o(\rho^{n+1}),\\ \sigma^n(\rho,\varphi,t) &= \rho\sigma_1(\varphi,t) + \rho^3\sigma_3(\varphi,t) + \cdots + \rho^n\sigma_n(\varphi,t);\end{aligned} \tag{8.5}$$

$\sigma_1(\rho,\varphi,t)$ is the "Heisenberg term," the first two components of which are given in (8.4). Let $\vec{B}^1$ denote the $(x,y,\tilde{p},\tilde{q})$ components of our vector field $\vec{H}$ in which, moreover, we set $r=1$, $\tilde{p}=p$, and $\tilde{q}=q$. Let $\vec{B}^2 = \left(0,0,p\dfrac{\partial H}{\partial w}, q\dfrac{\partial H}{\partial w}\right)$ in which also $r=1$, $\tilde{p}=p$, and $\tilde{q}=q$, and let $\vec{B} = \vec{B}^1 + \vec{B}^2$. Let $\vec{B}_n$ denote the nth jet of $\vec{B}$ with respect to (x,y,p,q,w) and with respect to the gradation $\deg(x) = \deg(y) = \deg(p) = \deg(q) = 1$, $\deg(w) = 2$. Let $\vec{C}$ be the w-component of $\vec{H}$ (in which $r=1$, $\tilde{p}=p$, $\tilde{q}=q$), $\vec{C} = \dfrac{\partial H}{\partial r}(\sigma,w)_{|r=1}$, and let $\vec{C}_n$ be the nth jet of $\vec{C}$ in the same gradation. Then we can compute $(\sigma_n(\varphi,t), \varepsilon_n^w(\varphi,t)$ by induction, since

$$\sigma^{n+1} = \rho\ \sigma_1(\varphi,t) + \int_0^t e^{A(t-s)}(\vec{B}_{n+1} - \vec{B}_1)(\sigma^n(\rho,\varphi,s), \tilde{\varepsilon}_n^w(\rho,\varphi,t))\ ds + O(\rho^{n+2}), \tag{8.6}$$

$$\tilde{\varepsilon}_{n+1}^w = \int_0^t \vec{C}_{n+1}(\sigma^n(\rho,\varphi,s), \tilde{\varepsilon}_n^w(\rho,\varphi,t))ds + O(\rho^{n+2}), \tag{8.7}$$

where $A(x,y,p,q) = \vec{B}_1(x,y,p,q)$ is the Heisenberg linear operator with the matrix

$$A = \begin{pmatrix} 0 & 1/2 & 1 & 0\\ -1/2 & 0 & 0 & 1\\ -1/4 & 0 & 0 & 1/2\\ 0 & -1/4 & -1/2 & 0\end{pmatrix}.$$

These computations are rather long. They were done by hand in [10] and for higher order using a formal program in [4]. This higher order was necessary for the purposes of Sec. 4. We show here the expressions of ε_3^z and ε_4^w *in the Dido case*. The other expressions are too long. In the following formulas, $\alpha_0 = 1/6k(p_0)$,

$$\begin{aligned}\varepsilon_3^z =& \alpha_0/2(6t\cos(\varphi - t) - 6\sin(\varphi) + 2\sin(\varphi - 2t) + 3\sin(\varphi - t) + \sin(\varphi + t),\\ & 6t\sin(\varphi - t) + 6\cos(\varphi) - 2\cos(\varphi - 2t) - 3\cos(\varphi - t) - \cos(\varphi + t)),\\ \varepsilon_4^w =& 3\alpha_0/8(-2t - 4t\cos(t) + 4\sin(t) + \sin(2t)).\end{aligned} \tag{8.8}$$

Similar expressions for the general sub-Riemannian case are given in [4, 10].

Note also that in the Dido case the computations are simpler because the two steps computing σ_n and ε_n^w are completely decoupled. The counterpart is that we will need to compute higher-order jets.

We will also need an expression for $\rho(t)$; by setting $\mu(t,\rho,\varphi) = \frac{\partial H}{\partial w}_{|r=1}(\sigma(t,\rho,\rho), \varepsilon_w(t,\rho,\varphi))$, we obtain

$$\rho(t) = \rho \exp\left(\int_0^t \mu(\tau,\rho,\varphi)d\tau\right), \quad \text{which is of the form } \rho(t) = \rho + O(\rho^5). \tag{8.9}$$

9. Back to wave fronts, spheres, conjugate loci, and cut loci. Any geodesic is optimal on small pieces of itself. For such a geodesic $\varepsilon(c,\cdot)$, $c \in C_0$, we define the conjugate-time (of the pole) $s_{\text{conj}}(c)$ (resp. the cut-time $s_{\text{cut}}(c)$) as the first instant of time at which the geodesic tail is locally optimal, i.e., optimal among admissible curves having the same endpoints and lying in a certain C^0 neighborhood of the geodesic segment (resp. globally optimal).

It is possible to verify that $s_{\text{conj}}(c)$ is also the first strictly positive instant of time at which the exponential mapping does not have full rank.

The *conjugate locus* CL is the union $\bigcup_{c\in C_0} \varepsilon(c, s_{\text{conj}}(c))$, i.e., the set of (*first*) *singular values of the exponential mapping* ε. The cut locus $\text{Cut}\,L$ is the union $\bigcup_{c\in C_0} \varepsilon(c, s_{\text{cut}}(c))$.

The conjugate locus is also a part of the set of singular values of $\bar{\varepsilon}$ ($\bar{\varepsilon}$ is defined to be equal to ε, but taken as a Lagrangian mapping $\bar{\varepsilon} : \bar{C}_0 \to X$). $\bar{C}_0$ is a Lagrangian submanifold of T^*E, $\bar{\varepsilon}$ is a Lagrangian mapping (in the sense of [7]), and the conjugate locus is part of the associated *"caustic."*

In the same way, $\varepsilon_s = \varepsilon(s,\cdot) : C_0 \to E$ is a Legendre mapping (in the same sense), and its image is the wave front W_s.

Therefore, all the objects under study (spheres, conjugate loci, cut loci) are, in a local level, elementary objects of the theory of Lagrangian and Legendre singularities in dimension 3. Of course, we will show nothing new in this level.

A simple computation shows that the conjugate time for the Heisenberg approximation $\bar{\varepsilon}_1(\rho,\varphi,t)$ is $t_H = 2\pi$. Hence all the phenomena of importance occur for t close to 2π.

Our study is local: first, we restrict ourselves to a neighborhood of the base point q_0 in X; second, we restrict ourselves to instants that are at most close to 2π, i.e., *we restrict ourselves to what happens in a neighborhood of the first conjugate points.* The reason for this is that only this restriction is important from the point of view of optimality.

However, *our study has a "global character"*: we consider the exponential mapping as a mapping restricted to a neighborhood U^S of the first conjugate points, and we study this restriction *globally.*

What will be new is that the elementary singularities appearing in our study consist of *nontrivial arrangements of classical Lagrangian and Legendre singularities* (generic and hence stable, since the dimension is 3). These collections of elementary stable singularities *are not organized in an arbitrary way.*

10. The exponential mapping as a suspension. Once we have computed the expansion of ε, it is possible and very convenient for our purposes to choose coordinates at the source of the exponential mapping so that it is in suspended form. For $w > 0$ and $\rho > 0$, we set

$$h = \sqrt{w/\pi}, \quad \sigma = (s - 2\pi h)/h, \tag{10.1}$$

and we will use the coordinates (x,y,h) at the image and (φ,σ,h) in the source.

One can verify that these coordinate changes are valid in a certain neighborhood of $\{t = 2\pi\}$ at the source (merely by using the expression (8.4) of ε_2^w, and the expression of s, $s = \int_0^t (\rho(\tau)d\tau$ and $\rho(t) = \rho + O(\rho^5))$.

Remark 10.1. The case $w < 0$, $\rho < 0$ is absolutely similar and leads to completely parallel results by setting $h = \sqrt{-w/\pi}$. Also, in the Dido case, the results for $w < 0$ can be obtained from the results for $w > 0$ simply by reversing the orientation a posteriori. *From now on, we will consider the case* $w > 0$ *only.*

Remark 10.2. This coordinate system for the suspended mapping ε^S goes back to the arclength. This is not the case for the coordinate systems (at the image and at the source) that we used in previous papers, (x,y,h) and (φ,t,h), which forget the arclength. For this reason, this new coordinate system is much better for computations of cut loci, as we will see. Also, the next lemma shows that it is very convenient for the computation of the conjugate loci.

10.1. Basic properties of the conjugate locus and the cut locus. Let us fix a geodesic $\varepsilon_{(\varphi,h)}(\sigma)$. Then, in these suspended coordinates, the conjugate time and the cut time are functions $\sigma_{\mathrm{conj}}(\varphi, h)$ and $\sigma_{\mathrm{cut}}(\varphi, h)$. As we will see, *these two functions are smooth*, and the conjugate locus and the cut locus are (in the generic case) 2-dimensional surfaces with singularities. If we cut them by the planes $\{h = ct\}$, we obtain one-dimensional objects. The conjugate locus (cut by $h = ct$) will be a closed curve that is smooth, except for a finite number of cusp points. (In fact, the conjugate locus is a surface that is smooth except for cuspidal lines transversal to the planes $\{h = ct\}$.)

Lemma 10.1. *In the coordinates (x, y, h) of the image and (φ, σ, h) of the source, the following properties hold*:

(i) *the conjugate time is given by the equation* $\dfrac{\partial \varepsilon}{\partial \varphi} = 0$;

(ii) *the conjugate time $\sigma_{\mathrm{conj}}(\varphi, h)$, for h constant, is a function of φ having its extrema at the position φ of cusp points of the conjugate locus.*

Simple general arguments also show in our contact case that $\mathrm{Cut}\, L \backslash CL$ is formed by points that are joined to the origin by several (not unique) geodesics at the same arclength time.

As a consequence, $\mathrm{Cut}\, L \backslash CL$ is simply *"the optimal part of the union of the self-intersections of all wave fronts."*

The complement of $\mathrm{Cut}\, L \backslash CL$ in $\mathrm{Cut}\, L$ is called *"the boundary of the cut locus."* We also have the following lemma.

Lemma 10.2. *The boundary of the cut locus coincides with the cuspidal lines of the conjugate locus.*

The proof of Lemmas 10.1 and 10.2 are given in [6]. Both proofs deeply use Liouville's theorem.

10.2. Sufficient jets for the exponential mapping. One of the conclusions of this study is that all singularities that appear for the exponential mapping are stable singularities: for the mapping as an (suspension of a) ordinary mapping between 2-dimensional manifolds (ε^S) or for the mapping as a Legendre mapping (ε_s), or for the mapping as a Lagrangian mapping ($\bar{\varepsilon}$).

As a suspension of an ordinary mapping, ε will have sufficient asymptotics in terms of h (or in terms of ρ), at least in restriction to a certain neighborhood of the first conjugate locus at the source. Hence ε will be determined by certain jets of the metric at the pole, q_0, in the general case, and by certain jets of the Riemannian metric at p_0 in the Dido case.

Also, as we will see, not all Lagrangian singularities appear. This will be shown in the next section of the paper.

4. Caustics

Here, we study the conjugate locus. We simply study ε as a suspension of an ordinary mapping. We observe that ε is stable in Thom–Mather's sense (in restriction to a neighborhood of its singular set at the source). Hence all singularities of the conjugate locus that appear are simply cuspidal lines and transversal self-intersections. Thus, as Lagrangian singularities, swallow tails do not appear, and the $D^4_{+,-}$ singularities do not appear. It is explained in [10] that, in fact, the swallow tail A^4 appears in the transition between the nondegenerate situation (points at which the principal invariant is nonzero) and the degenerate situation (points at which it is zero). All the results of the general case are valid for the isoperimetric case but are different in the Dido case: the sufficient asymptotics have higher order, and the size of the conjugate locus is different.

From now on, the *nondegenerate case* is the case of base points q_0 where the principal invariant $Q_2 \neq 0$. In the isoperimetric case, this is equivalent to $(\nabla^2 \log \psi)_2 \neq 0$. In the Dido case, Q_2 is always zero, and we say that the basepoint p_0 is nondegenerate if $(\nabla^2 k)_2 \neq 0$.

Also, in the Dido case, which we want to present in detail, we will need extra notations:

$$\begin{aligned}\beta^1(p_0) &= r_1\cos(t_1)dx - r_1\sin(t_1)dy = 1/12\nabla k(p_0),\\ \beta^2_2 &= r_2 R_e(e^{it_2}(dx+idy)^2), \qquad \beta^2_0 = \tau_2(dx^2+dy^2),\\ \beta^3_1 &= (r_{31}\cos(t_{31})dx - r_{31}\sin(t_{31})dy)\odot(dx^2+dy^2),\\ \beta^3_3 &= r_{32}R_e(e^{it_{32}}(dx+idy)^3), \alpha_0 = 1/6k(p_0),\end{aligned} \tag{10.2}$$

where $\beta^i_j = (\beta^i)_j$. $\nabla^2_2 k = 0$ iff $r_2 = 0$ and $\nabla^3_3 k = 0$ iff $r_{32} = 0$.

11. Nondegenerate case. Using Lemma 10.1, we see first that the conjugate time is a smooth function $\sigma_{conj}(\varphi, h)$ and obtain its expansion:

$$\sigma_{\text{conj}}(\varphi,h) = h^2\pi/48(-72\alpha_0 + 197(\alpha_0)^2h^2 - 320h^2\tau_2 - 720h^2r_2\cos(2\varphi)) + o^5(h). \tag{11.1}$$

Replacing the expression of the conjugate time in the expansion of the exponential mapping, we obtain the asymptotics of the conjugate locus.

In the general case, after normalization we have

$$\begin{aligned}x_{\text{conj}}(\varphi,h) &= |Q_2|h^3\cos(\varphi)^3 + O^4(h),\\ y_{\text{conj}}(\varphi,h) &= |Q_2|h^3\sin(\varphi)^3 + O^4(h),\end{aligned} \tag{11.2}$$

and in the Dido case,

$$\begin{aligned}z_{\text{conj}}(\varphi,h) &= z^5_{\text{conj}}(\varphi,h) + o^6(h)\\ &= h^4\pi(3r_1\sin(t_1) - 15h\cdot r_2\cos(\varphi) - 5h\cdot r_2\cos(3\varphi), 3r_1\cos(t_1) + 20h\ r_2\sin(\varphi)^3) + o^6(h).\end{aligned} \tag{11.3}$$

Moreover, for both cases, the following assertion holds.

Theorem 11.1. *For $h > 0$ small enough, there is a neighborhood U of the singular set at the source such that the exponential mapping restricted to U is left-right equivalent (via h-preserving diffeomorphisms) to its 3rd jet in h (resp. 5th jet in the Dido case), which is determined by the 3rd jet of the sub-Riemannian metric at q_0 (resp. 2nd jet of the curvature at p_0 in the Dido case).*

Sections $CL(h)$ of the conjugate locus CL by the planes $\{h = ct\}$ are *closed curves with 4 cusp points and without self-intersection.*

12. Degenerate case. The situation is much more complicated. Mainly, this is due to the very strange asymptotic symmetry of π: a symmetry at the level of the 4th jet in h of the exponential mapping (resp. at the level of the 6th jet in the Dido case).

Using again Lemma 10.1, we obtain the following asymptotics for the conjugate time.

In the Dido case, we obtain

$$\sigma_{\text{conj}}(\varphi,h) = -h^2\pi/48(72\alpha_0 - 197(\alpha_0)^2h^2 + 320h^2\tau_2 + 2880h^3r_{32}\sin(3\varphi)) + O^6(h). \tag{12.1}$$

For the expression in the general case, see [10].

Replacing the expansion of the exponential mapping by the suspended form, we obtain the asymptotics of the conjugate locus:

(a) *In the general case*:

$$\begin{aligned}z_{\text{conj}}(\varphi,h) &= z^4_{\text{conj}}(\varphi,h) + O^5(h)\\ &= |V_3|h^4\{2\cos(2\varphi)+\cos(4\varphi), -2\sin(2\varphi)+\sin(4\varphi)\} + O^5(h).\end{aligned} \tag{12.2}$$

(b) *In the Dido case*:

$$\begin{aligned}z_{\text{conj}}(\varphi,h) &= z^6_{\text{conj}}(\varphi,h) + o^7(h)\\ &= -h^4\pi/2(-6r_1\sin(t_1) + h^2(31\alpha_0 r_1\sin(t_1) + 25r_{31}\sin(t_{31}))\\ &+ 45h^2r_{32}(2\sin(2\varphi)+\sin(4\varphi)), -6r_1\cos(t_1) - h^2(25r_{31}\cos(t_{31}) - 31\alpha_0r_1\cos(t_1))\\ &+ 45h^2r_{32}(2\cos(2\varphi)-\cos(4\varphi))) + o^7(h).\end{aligned} \tag{12.3}$$

One can see that, at the level of these approximations, in both cases,

$$z_{\text{conj}}(\varphi, h) = z_{\text{conj}}(\varphi + \pi, h). \tag{12.4}$$

This is the announced asymptotic symmetry.

Hence the approximations at order 6 w.r.t. h of the sections $CL(h)$ are *double closed curves with* 3 *cusp points* (6 if counted with multiplicity).

Because of this symmetry, we have a theorem stating that these approximations are sufficient, locally only, along the singular locus at the source S.

For both cases, the following theorem holds.

Theorem 12.1. *For $h > 0$ small enough, there is a neighborhood U_x of each point of the singular set at the source S such that the exponential mapping restricted to U_x is left-right equivalent (via h-preserving diffeomorphisms) to its 4th jet in h (resp. 6th jet in the Dido case), which is determined by the 4th jet of the sub-Riemannian metric at q_0 (resp. 3rd jet of the curvature at p_0 in the Dido case).*

Singularities that appear at the local level on the caustic are the same: cuspidal lines only.

If we want to go further, general arguments from singularity theory say that we have to analyze the self-intersections of the conjugate locus at the level of higher-order jets: if the *conjugate locus is in general position*, then the suspended exponential mapping will be (Thom–Mather) stable.

13. Symbols for conjugate loci. For all the details (which are very complicated), see [4] in the general case and [8] in Dido case.

We will conclude that generic conjugate loci $CL(h)$ (cut by $\{h = ct\}$) are closed curves with 6 cusp points, which are in general position.

To analyze their self-intersections, we are led to consider reciprocal polynomials $P(z)$ of the form

$$P_{\mu,\nu}(z) = \mu z^4 + \nu z^3 + \bar{\nu} z + \bar{\mu}. \tag{13.1}$$

Theorem 13.1 (general case). *We are given a generic sub-Riemannian metric $(X, \Delta, g)_{q_0}$; q_0 is a degenerate point. The normal coordinates at q_0 are chosen for the asymptotic position $\varphi = \dfrac{2k\pi}{3}$ of the cusp points of the conjugate locus relative to q_0. There is a surjective mapping $\mathcal{M}$, $\mathcal{M} : \mathcal{S}^{\triangle}\Delta^* \times \mathcal{S}^{\triangle}\Delta^* \to \mathcal{C}^{\ni}$, which associates (ν^+, ν^-, μ) with the "invariants" γ_4^1, γ_4^0 of the sub-Riemannian metric at the basepoint q_0. The asymptotic positions on the unit circle of Δ_{q_0} as $h \to 0$ of the self-intersections of the conjugate locus* $\mathrm{CL}(h)$ *fall into two classes:*

(a) *those that coincide with the position of the cusp points, and*

(b) *the others.*

Points of class (b) *are the roots on the unit circle of the polynomial $P\mu, \nu^+$ if $h > 0$ (resp. $P\mu, \nu^-$ if $h < 0$).*

Note that γ_4^1 is invariant w.r.t. the action of $SO(2)$, but covariant w.r.t orientation changes.

Theorem 13.2 (Dido case). *We are given a generic Riemannian metric (M, g); p_0 is a degenerate point, $p_0 = \Pi(q_0)$. The Dido coordinates at q_0 are chosen for the asymptotic position $\varphi = \dfrac{2k\pi}{3}$ of the cusp points of the conjugate locus relative to q_0. There is a surjective mapping $\mathcal{N}$, $\mathcal{N} : \mathcal{S}^{\triangle}\mathcal{T}^*\mathcal{M} \to \mathcal{C}^{\in}$, which associates (ν, μ) with $\nabla^4 k(p_0)$. The asymptotic positions on the unit circle of TM_{p_0}, as $h \to 0$, of the self-intersections of the conjugate locus $CL(h)$ fall into two classes:*

(a) *those that coincide with the position of the cusp points, and*

(b) *the others.*

Points of class (b) *are the roots on the unit circle of the polynomial $P\mu, \nu$.*

A similar statement to the one of Theorem 13.2 holds in the general isoperimetric case. Here we see the only difference between the general sub-Riemannian case and the isoperimetric case: due to the symmetry, in the isoperimetric case conjugate loci are the same if we reverse the orientation. This is not true in the general case.

We define a semi-conjugate locus CL^+, or CL^-, as the restriction of CL to $\{w > 0\}$ (resp. $\{w < 0\}$).

Definition 13.1. We select any cusp point and any orientation. Following the curve $CL(h)$, we count the number of self-intersections between the ith and $(i+1)$th cusp point and we divide by 2. The result is a sequence of 6 rational numbers that we take modulo reflection and cyclic permutation. This sequence is the "symbol" of $CL^{+,-}$.

Recall that, generically, the degenerate points live on a curve $\mathcal{C}$ in X. The following (difficult) theorem is proved in [4]. This theorem says that *there are* 3 *types of points on* $\mathcal{C}$. At all of them, the exponential mapping ε^S restricted to a neighborhood of the semi-singular locus at the source is still stable and finitely determined.

Theorem 13.3. (1) *At generic points of the curve* $\mathcal{C}$, *the possible symbols for generic "semi-conjugate loci" are as follows:*

$$S_1 = (2,1,1,2,1,0), \qquad S_2 = (2,1,1,1,1,1), \qquad S_3 = (0,1,1,1,1,1).$$

(2) *At isolated points of the first type of* $\mathcal{C}$,

$$S_1^* = S_1 = (2,1,1,2,1,0), \qquad S_2^* = S_2 = (2,1,1,1,1,1).$$

But in this case, the exponential mapping is determined by a higher-order jet of the metric (7*th jet*) *than at generic points of* $\mathcal{C}$ (5*th jet*).

(3) *At isolated points of the second type, the possible symbols are*

$$S_4 = \left(\frac{1}{2},\frac{1}{2},1,0,0,1\right), \qquad S_5 = \left(1,\frac{1}{2},\frac{1}{2},1,1,1\right), \qquad S_6 = \left(\frac{3}{2},\frac{1}{2},1,1,0,1\right), \qquad S_7 = \left(2,\frac{1}{2},\frac{1}{2},2,0,0\right).$$

Also, *the semi-conjugate locus is recovered from the symbol* (this is not obvious at all). Hence, the above list of symbols provides a complete classification of semi-conjugate loci with respect to germs of origin-preserving homeomorphisms of R^3 that are smooth, together with their inverse, outside the origin.

Problem. Find all possible couples of symbols, modulo cyclic permutation and reversing orientation, for generic metrics. This should give a complete classification up to left-right equivalence of exponential mappings restricted to neighborhoods of their conjugate locus. Quick estimations show that there should be hundreds; however, this problem should not be difficult to solve using the results of our paper.

In the general isoperimetric and Dido cases, the situation is simpler. It follows from this study that the possible full conjugate loci are recovered from the symbols of the semi-conjugate loci, and the following assertion holds.

Theorem 13.4. *For generic Dido structures* (*or Riemannian metrics*), *at the isolated degenerate points, the possible symbols for conjugate loci are as follows:*

$$S_1 = (2,1,1,2,1,0), \qquad S_2 = (2,1,1,1,1,1), \qquad S_3 = (0,1,1,1,1,1).$$

These symbols give a complete classification of conjugate loci of degenerate Dido structures under the action of (*origin-preserving*) *homeomorphisms that are smooth, together with their inverse, outside the origin.*

At these points, the 8*th jet in* h *is sufficient for the exponential mapping, w.r.t. left-right equivalence when restricted to a neighborhood of the conjugate locus at the source. It is determined by the* 5*th jet of the curvature at* p_0.

There is a statement similar to that of Theorem 13.4 in the general isoperimetric case. The sufficient jet of the sub-Riemannian metric is the 5th one.

5. Wave Fronts and Spheres

Using Lemmas 10.2 and 10.1 and also the fact that the cut locus is simply the optimal part of the self-intersection of all wave fronts, we can draw a conclusion. This is not an easy job. All details can be found in [6].

To compute the self-intersections of the wave fronts, we have to solve the following equation in suspended coordinates:

$$\varepsilon^S(\sigma, \varphi', h) - \varepsilon^S(\sigma, \varphi, h) = 0. \tag{13.2}$$

It is obvious that in Eq. (13.2), the term $\sin\left(\dfrac{\varphi' - \varphi}{2}\right) = \sin\left(\dfrac{\delta\varphi}{2}\right)$ factors out.

Once this is done, we obtain the cut time $s_{\text{cut}}(\varphi, \delta\varphi)$ simply by applying the implicit function theorem (which shows that it is a smooth function).

Let us show the results in *the Dido case only*:

(a) *nondegenerate case*

$$\begin{aligned} \sigma_c = \frac{\pi}{48} h^2 (&-72\alpha_0 + 197\alpha_0^2 h^2 - 320\tau_2 h^2 \\ &- 120h^2 r_2(\cos(2\varphi) + 4\cos(d\varphi + 2\varphi)) \\ &- 120h^2 r_2 \cos(2(d\varphi + \varphi))) + o(h^5); \end{aligned} \tag{13.3}$$

(b) *degenerate case*

$$\begin{aligned} \sigma_c = \frac{\pi}{48} h^2 (&-72\alpha_0 + 197\alpha_0^2 h^2 - 320\tau_2 h^2 \\ &- 360h^3 r_{32}(\sin(3(d\varphi + \varphi)) + 3\sin(d\varphi + 3\varphi) \\ &+ 3\sin(2d\varphi + 3\varphi) + \sin(3\varphi)) + o(h^6). \end{aligned} \tag{13.4}$$

After that, we can replace this expression in one of the nontrivial equations of (13.2) (there are two). We obtain the following:

(a) *nondegenerate case*

$$h^4(\sin(d\varphi/2))^2 \sin(d\varphi + 2\varphi) + o(h^5) = 0; \tag{13.5}$$

(b) *degenerate case*

$$h^5(\sin(d\varphi/2))^2 \cos(d\varphi/2) \cos(3/2(d\varphi + 2\varphi)) + o(h^6) = 0. \tag{13.6}$$

The following lemma is crucial.

Lemma 13.5. *In formulas* (13.5) *and* (13.6), *the term* $(\sin(d\varphi/2))^2$ *factors out.*

At this level, let us forget the proof in the degenerate case: due partly to the asymptotic symmetry, it is really much more complicated. In the nondegenerate case, we obtain

$$d\varphi_1 = -2\varphi + o(h), d\varphi_2 = -2\varphi + \pi + o(h). \tag{13.7}$$

Only one of these solutions is optimal (it is $d\varphi_1$).

Replacing in (13.3), we obtain the following final expression of the cut time:

$$\sigma_{\text{cut}} = h^2\pi(-72\alpha_0 + 197\alpha_0^2 h^2 - 160h^2(3r_2 + 2\tau_2) - 240h^2 r_2 \cos(2\varphi)))/48 + o(h^5). \tag{13.8}$$

This is smooth, as we claimed.

Plugging the cut time in the exponential mapping, we obtain the following asymptotics of the cut locus:

$$z_{\text{cut}} = h^4\pi(3r_1 \sin(t1) - 20hr_2 \cos(\varphi), 3r_1 \cos(t1)) + o(h^6). \tag{13.9}$$

Putting everything together, we obtain the following theorem.

Theorem 13.6 (cut locus and sphere in the nondegenerate Dido case)**.** *The cut time and the cut locus have the smooth asymptotics* (13.8) *and* (13.9). *In particular, the cut locus has the size* h^5 (*in place of* h^3 *for generic sub-Riemannian metrics or generic isoperimetric problems*).

Sections by $h = cst$ *of the cut locus are curve segments joining two cusp points among those of the conjugate locus.*

The small spheres are homeomorphic to Euclidean spheres, but they are not smooth: the subset where they are not differentiable consists of two curve segments.

The singularities of the typical small wave front fall into two parts (*upper and lower hemisphere*)*. Each of them consists of a closed cuspidal curve with 4 cusp points. At the cusp points, swallow tails appear. On each part, there are also two curve segments* (*each of them joining two of the cusp points*), *and the*

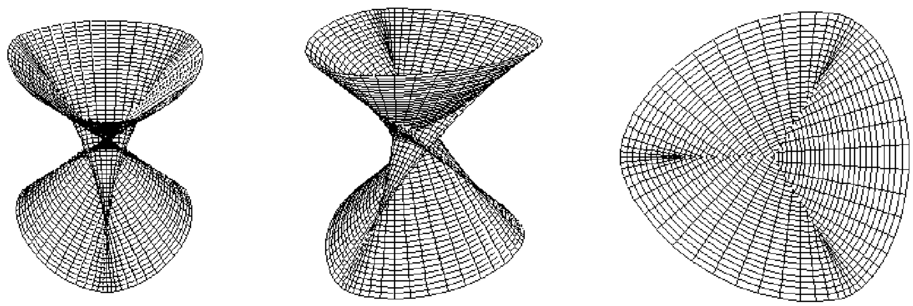

Fig. 2. Degenerate case, wave front.

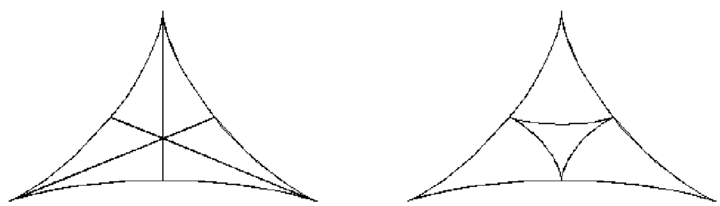

Fig. 3. Degenerate case. Parts 1 and 2 of the self-intersection of the wave front with the approximation at order 6 of the conjugate locus.

nonboundary points of these segments correspond to transversal self-intersections of the wave front. One of these segments is the (*upper or lower*) *singular set of the sphere.*

Thus, *contrary to what happens for caustics, all the stable elementary Legendre singularities occur.*

Let us now describe what happens in the degenerate case.

Theorem 13.7 (cut locus, sphere, and wave front in the degenerate case)**.** *The cut locus has the size* h^6 *in the Dido case* (*in place of* h^4 *in the case of generic sub-Riemannian metrics or generic isoperimetric problems*)*.*

Sections of the cut locus by $h = cst$ *consist of* 3 *curve segments with a common endpoint. The other endpoints of the* 3 *segments coincide with cusps of the conjugate locus.*

The small spheres are homeomorphic to Euclidean spheres but they are not differentiable. The singular set of the sphere falls into two pieces (*in both hemispheres*)*, each of them consisting of* 3 *curve segments with one common endpoint.*

The singularities of the wave fronts that are close to the corresponding spheres are shown in Fig. 2*. They consist of certain complicated self-intersections* (*described below*) *plus a closed cuspidal line with* 6 *cusp points that are endpoints of* 6 *swallow tails.*

Now let us completely describe this singularity of the wave front, which is close to the singularity of the sub-Riemannian sphere. It is shown in Fig. 2. The rightmost object in Fig. 2 is the picture of the sub-Riemannian sphere from above.

To analyze the singularity, it is easier to deal only with the 6th jet in the Dido case, 4th jet in the general case) of the exponential mapping. There is the asymptotic symmetry, but everything is easier to understand, and what will happen after perturbation is very clear.

Moreover, this analysis, in agreement with computations, shows that, in fact, *for the purposes of determining the sub-Riemannian sphere, these jets generically are sufficient.*

This is not the case for the wave front and for the caustic for which breaking the symmetry is necessary: the caustic cut by $\{h = ct\}$ has to be in general position, as we said.

The generic caustic is described by one of the symbols shown in Sec. 4.

Since we are looking at the nonsufficient 4th jet for which the asymptotic symmetry is present, $CL(h)$ is a double closed curve with 6 cusp points.

The self-intersection of the wave front (the part of it that is close to the corresponding sphere) has two pieces (adherent to the upper and lower hemisphere). Only one of them is shown in Fig. 2.

Each of these pieces falls into two parts, called Part 1 and Part 2.

If we cut the union of these self-intersection loci by $h = ct$, we obtain what is shown in Fig. 3. In this figure, these two parts are drawn at $h = ct$, together with the conjugate locus.

The first part only contains points of the cut locus, easily recognizable, compared with Fig. 2, right, which shows the sphere from above. The triple point appearing is perfectly stable: it corresponds to the transversal intersection of 3 surfaces.

All these pieces of the self-intersection are *quadruple curves* (in Fig. 3): they are double, since any cut locus should be double, and they are quadruple because of the asymptotic symmetry.

The most unstable phenomena at the level of this approximation are the cusps of Part 2: they are stable cusps also, but they are also 6, and looking at the middle of Fig. 2, one can see how they will separate: at these points, in Fig. 2 (middle), there is an unstable intersection of twofold lines on the wave front. After perturbation, they will separate, giving two cusps.

On a given wave front, there is a cuspidal closed curve. This curve is without self-intersections (at this level of approximation, it has 3) and with 6 cusps. Each of these cusps is the base point of a swallow tail on the wave front.

6. Complements

14. Rumin's Connection. For details, see [10, 19].

Again, we deal with the 3-dimensional case only. We are given a contact sub-Riemannian metric (X, Δ, g). $\underline{TX}$ and $\underline{\Delta}^*$ denote the sheaves of sections of TX and Δ, respectively.

Theorem 14.1. *There is a unique linear connection such that if* $\nabla : TM \times_M \underline{TM} \to TM$ *is the associated covariant derivative, then*

(i) $\nabla : TM \times_M \underline{\Delta} \to \Delta$ (∇ *has a restriction to* Δ);
(ii) $\nabla g = 0, \nabla\omega = 0$;
(iii) $\nabla_\nu \nu = 0$;
(iv) *the torsion of* ∇ *is* $\nu \otimes d\omega$;
(v) *the form* $II : \Delta \times_M \Delta \to R$, $II(X, Y) = \langle \nabla_X \nu, Y \rangle_g$, *is symmetric.*

The proof of this theorem is given in [10].

All the invariants appearing in this paper, in particular the principal and the second invariant, can be computed in terms of this canonical connection as follows.

In the 3-dimensional case, there is a natural complex structure on Δ (that we already used implicitly in the paper) $j : \Delta \to \Delta$, defined by $d\omega(X, Y) = \langle j(X), Y \rangle_g$. If the orientation is reversed, j is changed to $-j$.

The restriction $\nabla II_{|\Delta}$ is a section of $\Delta^* \otimes S^2\Delta^* = \bigoplus_{n=1,3} (\Delta^* \otimes S^2\Delta^*)_n$. Let $III = \nabla II_{|\Delta}$. $(III)_3$ is a component of III in $(\Delta^* \otimes S^2\Delta^*)_3$.

Theorem 14.2. *For* $u \in \Delta$, *one has*

$$Q_2(u) = \frac{1}{4} II(u, j(u)), \qquad V_3(u) = \frac{1}{15} (III)_3(j(u), u, u).$$

Also, the Gaussian curvature of ∇ *is* $\mathcal{K} = \not{t}\mathrm{race}_\}(\mathcal{Q}) = \not{t}\mathrm{race}_\}(\gamma'_\in)$.

In terms of this connection, the geodesics $\lambda(s)$ are solutions of the following equations:

$$\nabla_{\dot\lambda} \dot\lambda = r j(\dot\lambda), \qquad \dot r + II(\dot\lambda) = 0, \qquad r(0) = r_0.$$

7. A Large Number of Cusps

We focus on the Dido case. In the generic situation, the number of cusps of the conjugate locus n_{cusp} is equal to double the number of free endpoints of the singular segments on the hemisphere, n_e, and is also equal to the number n_{sw} of swallow tails adherent to the (hemi)sphere on the (hemi)wave front of same radius:

$$n_{\text{cusp}} = 2\, n_e = n_{sw}. \tag{15.1}$$

Also, $\frac{n_{\text{cusp}}}{2}$ *is equal to the first integer* $j > 1$ *such that* $(\nabla^j k(q_0))_j \neq 0$, where k is the Gaussian curvature and q_0 is the pole. It seems that there is something general beyond these facts.

Open problem. Given any (nongeneric) germ at q_0 of a Riemannian metric that is "nonflat" in the sense that $\nabla_i^i k(q_0) \neq 0$ for some integer $i > 1$. Set j= first integer >1 such that $\nabla_j^j k(q_0) \neq 0$. Is it still true that:

(a) $n_{\text{cusp}} = 2j$?

(b) formula (15.1) holds?

(c) if j is odd, then the conjugate locus is asymptotically double?

The same problem also makes sense in the general case of (isoperimetric or not) sub-Riemannian metrics.

Perhaps, accordingly to the generic cases, this term $\nabla_j^j k(q_0)$ *dominates* all other terms. We are unable to prove this in general. Nevertheless, what we can prove is the following theorem.

Theorem 14.3. *Assume that all tensors* $\nabla_k^l k(q_0)$ *are zero, except for* $\nabla_j^j k(q_0)$*, for some* j*. Then,* $n_{\text{cusp}} = 2j$*.*

REFERENCES

1. A. A. Agrachev, "Methods of control theory in nonholonomic geometry," In: *Proc. ICM-94*, Zürich, Birkhauser (1995).
2. A. A. Agrachev, "Exponential mappings for contact sub-Riemannian structures," *J. Dyn. Contr. Syst.*, **2**, No. 3, 321–358 (1996).
3. A. A. Agrachev, B. Bonnard, M. Chyba, and I. Kupka, "Sub-Riemannian sphere in Martinet flat case" (in press).
4. A. A. Agrachev, H. Chakir, and J. P. Gauthier, "Sub-Riemannian metrics on R^3," *Proc. Can. Math. Soc.*, **25**, 29–78 (1998).
5. A. A. Agrachev, H. Chakir, J. P. Gauthier, and I. Kupka, "Generic singularities of sub-Riemannian metrics on R^3," *C. R. Acad. Sci. Sér.* 1, **322**, 377–384 (1996).
6. A. A. Agrachev and J. P. Gauthier, "On the Dido problem and plane isoperimetric problems," *Prépublication du Laboratoire de Topologie de l'Université de Dijon, Mai* (1998).
7. V. Arnold, A. Varchenko, and S. Goussein-Zadé, *Singularities of Differentiable Mappings* [in Russian], Mir, Moscow (1986).
8. M. Baldé, "Remarks on some isoperimetric problems" (in press).
9. C. Carathéodory, *Calculus of Variations and Partial Differential Equations of the First Order*, New York (1989).
10. H. Chakir, J. P. Gauthier, and I. Kupka, "Small sub-Riemannian balls on R^3," *J. Dyn. Contr. Syst.*, **2**, No. 3, 359–421 (1996).
11. M. Chyba, "Exponential mapping, spheres and wave fronts in sub-Riemannian geometry: The Martinet case," Ph.D. thesis, University of Burgandy (1997).
12. M. Golubitskii and V. Guillemin, *Stable Mappings and Their Singularities*, Springer-Verlag, New York (1973).
13. M. Gromov, "Carnot–Carathéodory metrics seen from within," Preprint IHES M/94/6 (1994).
14. S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, John Wiley, New York (1969).
15. I. Kupka, "Géométrie sous-Riemannienne," *Séminaire Bourbaki*, 48ème année, No. 817 (1995–1996).
16. J. Mather, "Stability of C^∞-mappings, I–VI," *Ann. Math.*, **87**, 89–104 (1968); **89**, 254–291 (1969); *Publ. Sci. IHES*, **35**, 127–156 (1969); **37**, 223–248 (1970); *Adv. Math.*, **4**, 301–335 (1970); *Lect. Notes Math.*, **192**, 207–253 (1971).

17. R. Montgommery, “Isholonomic problems and some applications,” *Commun. Math. Phys.*, **128**, 565–592 (1990).
18. L. Pontryagin, V. Boltyanskii, R. Gamkrelidzé, and E. Mitchenko, *Mathematical Theory of Optimal Processes* [in Russian], Nauka, Moscow (1976).
19. M. Rumin, “Formes différentielles sur les variétés de contact,” Ph.D. Thésis, Univ. of Paris sud (1992).
20. N. Steenrod, *The Topology of Fiber Bundles*, Princeton University Press, New Jersey (1951).
21. H. Whitney, “On singularities of mappings of Euclidean spaces, I–II,” *Symp. Intern. de Topologia Algebraica*, Mexico City (1958); *Ann. Math.*, **62**, No. 3 (1955).