The Conjugate Locus for the Euler Top

I. The Axisymmetric Case

Larry Bates

Department of Mathematics and Statistics
University of Calgary
Calgary, Alberta, T2N 1N4, Canada

Francesco Fassò

Dipartimento di Matematica Pura e Applicata
Università di Padova
Via Trieste 63, 35121 Padova, Italy
(E-mail: fasso@math.unipd.it)

Abstract

The flow of the Euler top is a geodesic flow on $SO(3)$ with a left invariant metric. We determine the conjugate locus for this geodesic flow in the case that the metric has an S^1 invariance, which is the case when two of the three moments of inertia are equal. Depending on the ratios of these moments, the conjugate locus is either a segment or circle (if the body is oblate) or a non-injective mapping of an astroid of revolution (if the body is prolate).

Mathematics Subject Classification: 53C20

Keywords: Conjugate loci, Geodesic flow, Left-invariant metrics on $SO(3)$, Rigid body, Berger spheres.

1 Introduction

A. Motivations. If q is a point of a Riemannian manifold $\mathcal{Q}$ and γ an oriented geodesic through q , then the *cut point* q' of q along γ , if it exists, is that point of γ at which γ ceases to minimize the distance from q . The (first) *conjugate point* of q along γ (if it exists) is the point of γ at which the minimization property is lost among ‘infinitesimally close’ geodesics. Precisely, if $t \mapsto q^t(q, \dot{q})$ is a parameterization of the geodesic through q with initial

velocity $\dot{q} \in T_q\mathcal{Q}$ and $t_c(q, \dot{q})$ is the infimum of the positive t such that the Jacobian matrix

$$\frac{\partial q^t}{\partial \dot{q}}(q, \dot{q})$$

is singular, then the (first) conjugate point of q along the geodesic is $q^{t_c(q, \dot{q})}(q, \dot{q})$. The cut (conjugate) locus of q is the union of its cut (conjugate) points along all the geodesics through it.

For a general introduction to these notions see e.g. [17, 16, 5]. There is a connection between conjugate and cut points: if q_c is the (first) conjugate point of q along γ , then either q_c is the cut point or there is a cut point q' on γ between q and q_c which is joined to q by at least two minimizing geodesics. Hence, the study of the conjugate locus gives information on the cut locus, too. While the cut locus has probably a more clear geometric significance, the conjugate locus may be expected to have a richer structure and to help to understand the structure of the cut locus as well.

The study of conjugate loci in a Riemannian manifold has a long and renowned history, which goes back at least to Jacobi's study of the geodesic flow on ellipsoids: Jacobi's 'last geometric theorem' states that the conjugate locus of a non-umbilical point of a triaxial ellipsoid has exactly four cusps (see [14] for references and recent advances, including the proof of this fact). Notwithstanding a great number of studies (see e.g. [17, 8, 20, 13, 19, 15, 5, 21] and references therein), there are not yet many known examples of cut and conjugate loci, especially in dimension higher than two.

Thus, our aim in this article and in a forthcoming Part II [4] is that of determining the conjugate locus in a specific, though somehow special because of its symmetry, three-dimensional case—the geodesic flow on $\mathrm{SO}(3)$ endowed with a left invariant metric.

B. Left invariant metrics on $\mathrm{SO}(3)$ and the Euler top. This problem has both geometric and mechanical relevance. A left invariant metric on $\mathrm{SO}(3)$ is determined by the eigenvalues I_1, I_2, I_3 of its restriction to the group algebra. In Mechanics, this geodesic flow can be interpreted as the flow of an *Euler top*, namely a rigid body with a fixed point and no torques acting on it, and the eigenvalues I_1, I_2, I_3 play the role of moments of inertia.¹ This classical system is one of the prototypes of (super)integrable Hamiltonian systems and still plays an important role in Mechanics, see e.g. [1, 9, 12]. Even though our study will focus on the geometric problem, it will benefit of the corpus of techniques and ideas specific to the mechanical problem.

¹The moments of inertia of a rigid body satisfy in fact the three inequalities $I_1 + I_2 > I_3$, $I_2 + I_3 > I_1$, $I_3 + I_1 > I_2$. Nevertheless, the equations of motion of the rigid body retain their validity, and their properties, for all positive values of the parameters I_1, I_2, I_3 . So, we shall keep using the mechanical analogy even in unphysical cases.

The Euler top is a system with a large symmetry group, $\mathrm{SO}(3) \times S^1$, and quasi-periodic dynamics on invariant tori of (generically) dimension two [3, 11]. In the general case of three distinct moments of inertia, the flow can be expressed in terms of Jacobi elliptic functions and theta functions [2, 22, 18]. In the *axisymmetric* case in which two of the three eigenvalues are equal, say $I_1 = I_2 \neq I_3$, the flow can be expressed in terms of elementary functions of t and of the initial data. In the present paper we shall therefore consider the axisymmetric case and derive a detailed and essentially complete analytic description of the conjugate locus.

In Mechanics, left-invariant metrics with two equal eigenvalues on $\mathrm{SO}(3)$ correspond to axisymmetric Euler tops. In Geometry, they are covered by the so-called Berger metrics on S^3 , which have already been studied by Chavel [8] and Sakai [20] in the case $I_1 > I_3$. Chavel determined, among other things, the *tangential* conjugate locus and Sakai determined the cut locus (see below). We shall use their results on the tangential conjugate locus at an early stage of our construction.

The analytical determination of the conjugate locus in the general case of distinct eigenvalues appears to be significantly more complex. We shall discuss this problem at greater length in a companion article [4], using also numerical computations of the conjugate locus. The description of the axisymmetric case given here will be the starting point for the comprehension of the patterns emerging in the general case.

Remark: It should be stressed that the axisymmetric case does not have a larger symmetry group than the general case: the group is in both case $\mathrm{SO}(3) \times S^1$, but the factor S^1 acts in a particularly simple and recognizable way in the axisymmetric case. (A full account of this fact, which is implicit in Poinsot's classical 'geometric description' of the Euler top motions [3, 9], can be found in [6], chapter 12). Of course, the large symmetry group of the problem is at the same time its strength and its weakness: it allows (at least in the axisymmetric case) the determination of the conjugate locus but, at the same time, it certainly makes the structure of the conjugate locus non-generic.

C. The conjugate locus in the axisymmetric case. We describe now very informally our results for the axisymmetric case. Since the geodesic flow is just a shift in the group, the conjugate loci of different points are conjugate to each other. Thus, we study only the conjugate locus of the group identity and call it, accordingly, 'the' conjugate locus for the given metric. In order to visualize the conjugate locus we make use of the solid ball model of $\mathrm{SO}(3)$, that is, the unit ball of $\mathbb{R}^3$ with antipodal boundary points identified.

Elementary considerations based on the symmetry of the problem show that the conjugate locus, in the solid ball model, is symmetric with respect to the three coordinate planes and that, if $I_1 = I_2$, then it is also invariant

Figure 1: The conjugate locus in various prolate cases, cut so as to reveal the internal structure. All images have the same scale. The upper and lower circular edges lie on the boundary of the unit ball and must therefore be antipodally identified.

under rotations around the axis $(0, 0, 1)$. This implies that, if the three eigenvalues I_1, I_2, I_3 are all equal, then the conjugate locus reduces to a point. If $I_1 = I_2 \neq I_3$, instead, the conjugate loci are of two types:

- i.* If $I_1 = I_2 < I_3$ (the body is ‘oblate’) then the conjugate locus is one-dimensional. Specifically, in the solid ball model, it is a segment of length proportional to I_3/I_1 which, for $I_3/I_1 \geq 2$ touches the boundary of the unit ball and therefore closes to a circle.
- ii.* If $I_1 = I_2 > I_3$ (the body is ‘prolate’) then the conjugate locus is a surface of revolution with two conical cusp points on its symmetry axis, a cuspidal edge in the equatorial plane and a number of circles of transversal self-intersections on the equatorial plane and on the boundary of the unit ball, which grows without bound with I_1/I_3 .

Some examples of conjugate loci in the prolate case are shown in Figure 1. Their structure can be understood in simple terms:

- For $1 < I_1/I_3 < \frac{3}{2}$ the conjugate locus resembles an astroid of revolution, with height and equatorial radius both smaller than one and increasing with I_1/I_3 .
- When $I_1/I_3 = \frac{3}{2}$ the two cusps touch the boundary of the unit ball and must therefore be identified, while the equatorial radius remains smaller than one. Topologically, the conjugate locus is a pinched torus.
- As we further increase I_1/I_3 : the height of the astroid of revolution increases and its tips exit the unit ball—thus reentering it from the opposite side and creating two self-intersection circles. As I_1/I_3 keeps growing, each of the conical surfaces so created continues to grow in height and the process repeats indefinitely.

In other words, the increasingly complex structure of the conjugate locus for large values of I_1/I_3 is the result of a non-injective immersion with singularities into $\text{SO}(3)$ of a surface resembling an (increasingly elongated) astroid of revolution.

A precise description of this surface of revolution—specifically, of its generatrix—will be given in Section 3, and proved in Section 4. Section 2, which is not limited to the axisymmetric case, is devoted to some preliminary known facts about the geodesic flow on $\text{SO}(3)$ and the symmetries of its conjugate locus.

Remarks: (i) The cut locus for the prolate case has been determined in [20] and consists of the disk in the equatorial plane of the conjugate locus bounded by the cuspidal edge. The cut locus for the oblate case is instead, to our knowledge, still unknown. See also [5].

(ii) One may compute the conjugate locus on the universal covering S^3 once one knows it in $\text{SO}(3)$ (or conversely). We prefer the latter case because there is a cleaner picture of the geodesic flow on $\text{SO}(3)$.

2 The geodesic flow on $\mathrm{SO}(3)$

A. Left-trivialized equations of the geodesic flow. It is well known that the geodesic flow on a Riemannian manifold $\mathcal{Q}$ coincides (up to a time reparameterization, of course) with the flow of a Lagrangian system on $T\mathcal{Q}$ with Lagrangian $\frac{1}{2}G_q(\dot{q}, \dot{q})$, where G is the metric tensor. Equivalently, the system can be regarded as a Hamiltonian system on $T^*\mathcal{Q}$ with Hamiltonian $\frac{1}{2}G_q^{-1}(p, p)$.

In the case of a left-invariant metric on $\mathrm{SO}(3)$ the resulting equations of motions are the equations of motion of an Euler top with moments of inertia given by the eigenvalues I_1, I_2, I_3 of G_1 , where $\mathbf{1}$ is the identity matrix. These equations can be written in a variety of ways. We choose here the left-trivialized Hamiltonian formulation, which leads to the system of first order equations on $\mathrm{SO}(3) \times \mathbb{R}^3 \ni (\mathcal{R}, m)$ given by

$$\dot{\mathcal{R}} = \mathcal{R} \widehat{\mathcal{I}^{-1}m}, \quad \dot{m} = m \times \widehat{\mathcal{I}^{-1}m}. \quad (2.1)$$

Here $\mathcal{I} = \mathrm{diag}(I_1, I_2, I_3)$ and we have used the isomorphism $\widehat{\cdot} : \mathbb{R}^3 \rightarrow \mathfrak{so}(3)$ defined by $\widehat{u}v = u \times v$ for all $v \in \mathbb{R}^3$, where $\times$ denotes the cross product in $\mathbb{R}^3$. These equations are sometimes called Euler–Arnold equations, see e.g. [9] for details.

In this description, the left-invariant metric on $\mathrm{SO}(3)$ induces the scalar product

$$G_1^{-1}(m, m') = m \cdot \mathcal{I}^{-1}m' \quad (2.2)$$

on the factor $\mathbb{R}^3$. (Here and in the following, the dot denotes the Euclidean scalar product and $|\cdot|$ the Euclidean norm in $\mathbb{R}^n$). Note that, for the moment, we do not assume that two of the three eigenvalues are equal.

Remark: We recall very briefly, for the unfamiliar reader, some of the details about the rigid body interpretation of equations (2.1).

(i) First, the identification of the configuration space of a rigid body with the matrix group $\mathrm{SO}(3)$ is as follows. Consider two orthonormal bases $\{e_1^b, e_2^b, e_3^b\}$ and $\{e_1^s, e_2^s, e_3^s\}$ having the body fixed point as common origin and attached to the body and to the space, respectively. Choose the body base so that its axes e_j^b are principal axes of inertia, with moments of inertia I_j , and define $\mathcal{R} \in \mathrm{SO}(3)$ to be the matrix with entries $\mathcal{R}_{ij} = e_i^s \cdot e_j^b$.

(ii) Next, work on the Hamiltonian side, left-trivialize $T^*\mathrm{SO}(3)$ and identify $\mathfrak{so}(3)^*$ first with $\mathfrak{so}(3)$ via the Killing metric and then with $\mathbb{R}^3$ via the $\widehat{\cdot}$ isomorphism. The state of the system is thus a point $(\mathcal{R}, m)$ where the matrix $\mathcal{R}$ is as above and $m = (m_1, m_2, m_3) \in \mathbb{R}^3$ is the body angular momentum vector, that is, the representative in the body base $\{e_1^b, e_2^b, e_3^b\}$ of the angular momentum vector of the system.

B. The solid ball model of $\mathrm{SO}(3)$. Let B be the closed unit ball in $\mathbb{R}^3$, B_P the ball B with antipodal points on the boundary identified, and $p : B \rightarrow B_P$ the canonical projection. The solid ball model of $\mathrm{SO}(3)$ is based on the fact that, for any $m \in \mathbb{R}^3$, $\exp \hat{m}$ is the rotation matrix of angle $\pi|m|$ around m . Thus the map

$$B \rightarrow \mathrm{SO}(3), \quad m \mapsto \exp(\hat{m})$$

is surjective but not injective, and a diffeomorphism

$$\mathcal{F} : \mathrm{SO}(3) \rightarrow B_P$$

is obtained by regarding this association as (the inverse of) a map from $\mathrm{SO}(3)$ to B_P . For details, see e.g. [9].

Throughout the article we use the following terminology and notation. First of all, it will be understood that the coordinates in B , hereafter denoted x_1, x_2, x_3 , are the coordinates in which the scalar product induced by G on $\mathbb{R}^3$ is diagonal, see (2.2). The p -image in B_P of the coordinate line x_j in B will be called the x_j -axis of B_P or, when we want to stress the fact that it is topologically a circle, the x_j -coordinate circle or simply the x_j -circle of B_P . Similarly, the image of the disk $x_j = \text{const}$ of B will be called the coordinate plane x_j of B_P . The point $p(0)$ will be called the origin of B_P and $p(0, 0, \pm 1)$ its North-South pole.

If S is an orthogonal transformation of $\mathbb{R}^3$ and K is a subset of B_P , we define the image of K under S as $p(S(p^{-1}(K)))$. We then say that K is invariant under S if $S = p(S(p^{-1}(K)))$, that is, if $p^{-1}(K)$ is S -invariant. Finally note that, through pullback to B , we can define the distance from the origin of any point of B_P and the length of any arc of a coordinate circle.

Remark: In the rigid body analogy, the points of B on the x_j -axis correspond to rotations around the axis of inertia e_j^b of the body.

C. The conjugate locus and its symmetries. Denote by $t \mapsto (\mathcal{R}^t(\mathcal{R}, m), m^t(\mathcal{R}, m))$ the solution with initial datum $(\mathcal{R}, m)$ of the equations (2.1). For given m , let

$$t_c(\mathcal{R}, m) = \inf \left\{ t > 0 : \text{rank} \frac{\partial \mathcal{R}^t}{\partial m}(\mathcal{R}, m) < 3 \right\}. \quad (2.3)$$

Then,

$$\mathcal{L}(\mathcal{R}, m) := \mathcal{R}^{t_c(\mathcal{R}, m)}(\mathcal{R}, m)$$

is called the (first) conjugate point to $\mathcal{R}$ in the ‘direction’ m . The conjugate locus $\mathcal{L}_{\mathcal{R}}$ of $\mathcal{R} \in \mathrm{SO}(3)$ is the union of all its conjugate points $\mathcal{L}(\mathcal{R}, m)$, $m \in \mathbb{R}^3$. As for any geodesic flow, the norm of the (co)tangent vector m is of course

irrelevant to the definition of conjugate points.² Thus, the construction of the conjugate locus can be restricted to $m \in S^2$ (the unit sphere in $\mathbb{R}^3$; note that the norm we are using in $\mathbb{R}^3$ corresponds to the Killing metric on $\mathfrak{so}(3)$, not to the considered left-invariant metric on $\mathrm{SO}(3)$).

We now examine the consequences of the $\mathrm{SO}(3)$ symmetry of the problem on the conjugate locus. If $t \mapsto (\mathcal{R}^t, m^t)$ is a solution of equation (2.1) and $S \in \mathrm{SO}(3)$, then also $t \mapsto (S\mathcal{R}^t, m^t)$ is a solution. Note that these solutions reach their (first) conjugate points at the same time. Hence $\mathcal{L}_{S\mathcal{R}} = S\mathcal{L}_{\mathcal{R}}$: the conjugate loci of all points of $\mathrm{SO}(3)$ are isometric and we need to study only one of them, say that of the identity matrix. We shall therefore consider from now on only solutions with initial datum $\mathcal{R} = \mathbf{1}$ and we shall correspondingly drop everywhere the indication of the initial configuration $\mathcal{R}$. In particular, we call

$$\mathcal{L}(S^2) = \bigcup_{m \in S^2} \mathcal{L}(m) \subset \mathrm{SO}(3)$$

the conjugate locus for the given metric. However, we shall mainly work with the image of the conjugate locus in the solid ball model. Specifically, we define

$$L := \mathcal{F} \circ \mathcal{L} : S^2 \rightarrow B_P$$

and call $L(S^2)$ the conjugate locus in B_P . In the sequel, we shall commonly write $\mathcal{L}$ and L instead of $\mathcal{L}(S^2)$ and $L(S^2)$, respectively.

We now note some (rather obvious, but important for the sequel) symmetry property of the conjugate locus L , which are due to [8] in the prolate case.

Proposition 1

- i. L is invariant under reflections through the three coordinate planes.
- ii. If $I_1 = I_2$, then L is invariant under rotations around the axis e_3 . Moreover, $m \mapsto L(m)$ is a function of m_3 alone.

Proof. We first show that any orthogonal symmetry of the metric tensor $\mathcal{I} = \mathrm{diag}(I_1, I_2, I_3)$ produces a symmetry of the conjugate locus $\mathcal{L}$ in $\mathrm{SO}(3)$. Indeed, if $(\mathcal{R}^t, m^t)$ is a solution of the equations of motion and S is an orthogonal matrix such that $S\mathcal{I} = \mathcal{I}S$, then also $(S\mathcal{R}^t S^T, (\det S)m^t)$ is a solution. It follows that, for any m , the solutions with initial data $(\mathbf{1}, m)$ and $(\mathbf{1}, Sm)$ satisfy $\mathcal{R}^t(\mathbf{1}, (\det S)Sm) = S\mathcal{R}^t(\mathbf{1}, m)S^T$. Hence, they reach their (first) conjugate points $\mathcal{L}((\det S)Sm)$ and $\mathcal{L}(m)$ at the same time. Thus $\mathcal{L}((\det S)Sm) = S\mathcal{L}(m)S^T$ for all m . Thus, in the solid ball model,

$$L((\det S)Sm) = SL(m)$$

²Homogeneity of (2.1) implies in fact that if $t \mapsto (\mathcal{R}^t, m^t)$ is a solution, then so is $t \mapsto (\mathcal{R}^{\lambda t}, \lambda^{-1}m^{\lambda t})$ for any $\lambda \neq 0$. For $\lambda = -1$ this means that the equations are time-reversible; this is why we need not consider $t_c < 0$.

for all m and hence

$$SL = L. \quad (2.4)$$

This proves statement i and the rotational invariance of L in the axisymmetric case. But moreover, as discussed above, if $I_1 = I_2$ then the two conjugates points in the directions m and Sm are reached at the same time. Therefore, this time and the conjugate point $L(m)$ are functions of m_3 alone. ■

Remark: The homogeneity of the equations of motion implies also that a rescaling of the eigenvalues I_1, I_2, I_3 can be compensated by a corresponding rescaling of time. Hence, the conjugate locus depends in fact only on the ratios between the eigenvalues and one of them may be normalized to a fixed value.

3 The conjugate locus for the axisymmetric case

We consider now the case $I_1 = I_2$. Define the parameter η by

$$\eta := \frac{I_1}{I_3} - 1 > -1$$

Adopting a rigid body terminology, we say that the metric is *prolate* if $\eta > 0$ and *oblate* if $-1 < \eta < 0$.³ The case $\eta = 0$ ('spherical' bodies) is included in the subsequent treatment as a limit of both the oblate and the prolate cases.

If $I_1 = I_2$ then, as is easily checked, the solution $t \mapsto (\mathcal{R}^t, m^t)$ of the left-trivialized equations (2.1) with initial data $\mathcal{R} = \mathbf{1}$ and $m = (m_1, m_2, m_3)$ is

$$\mathcal{R}^t = \exp\left(\frac{t}{I_1}\widehat{m}\right) \exp\left(\frac{t}{I_1}\eta m_3 \widehat{e_3}\right), \quad m^t = \exp\left(-\frac{t}{I_1}\eta m_3 \widehat{e_3}\right)m \quad (3.5)$$

(see e.g. [10]). These formulas are the starting point for the study of the conjugate locus, which is performed in the next Section. The conclusions of this analysis are collected in the next three Propositions. First, we consider the oblate case:

Proposition 2 *In the solid ball model:*

- i. For $-1 < \eta < 0$, the conjugate locus L is the arc of the x_3 -coordinate circle centered at the origin and of length $4|\eta|$.*
- ii. If $\eta = 0$, then L is the origin.*

³In fact, $\eta \geq -\frac{1}{2}$ for 'real' rigid bodies: the limit case $\eta = -\frac{1}{2}$ corresponds to flat, i.e., disk-like, bodies.

Note that, when $\eta \leq -\frac{1}{2}$, the conjugate locus is a circle.

We now consider the case $\eta > 0$. Since, by Proposition 1, the conjugate locus is invariant under rotation around the axis x_3 , we can limit ourselves to describe its generatrix, that is, the curve in the plane $x_2 = \text{const}$ of B_P which produces the conjugate locus by rotation around the axis x_3 . We denote this plane by D_P and note that $D_P = p(D)$, where D is the section $x_2 = 0$ of B .

We need to introduce some miscellaneous notations and terminology. For real numbers, besides the integer part $[x]$ and the ceiling $\lceil x \rceil$, we also use the function

$$\langle x \rangle := \lceil x \rceil - 1.$$

Moreover, we denote by $d(x)$ twice the distance of $x \in \mathbb{R}$ from its closest integer, which satisfies

$$d(x) = 2|x - [x + \tfrac{1}{2}]| = 2|x - \langle x + \tfrac{1}{2} \rangle| = \frac{2}{\pi} |\arcsin(\sin(\pi x))|.$$

Thus $d(x)$ lies between 0 and 1, equals 0 if x is an integer and 1 if x is a half-integer.

Recall now that a curve $u \mapsto (x(u), y(u)) \in \mathbb{R}^2$ has an (*ordinary*) *cusp tangent to the x -axis* at a point u_0 if

$$x'(u_0) = 0, \quad x''(u_0) \neq 0, \quad y(u_0) = y'(u_0) = y''(u_0) = 0, \quad y'''(u_0) \neq 0 \quad (3.6)$$

see e.g. [7]. The point $(x(u_0), y(u_0))$ in the trace of the curve is also called an (ordinary) cusp point, but the distinction is important in the sequel because it might happen that the images of two different cusp points u_0 and u_1 coincide. A cusp defines an obvious orientation on the tangent line through it, induced by the parameter u .

These notions extend in an obvious way, using local coordinates, to curves in D_P . The case we need to consider is that in which the curve has two cusps, in two points $P_1, P_2 \in D_P$ and tangent to the circle through P_1 and P_2 . If P_1, P_2 are distinct and do not belong to $p(\partial D)$, we say that the two cusps either *point towards the origin* or *point away from the origin* if so do their preimages in B . If $P_1 = P_2$ is either the origin or the North–South pole, then we use either expression to mean that the two cusps have opposite orientations.

Finally, given a point $P \in D_P$, we denote by $-P$ the point obtained by reflection through the origin, and by $+P$ the point P itself.

We can now describe the generatrix of the conjugate locus. First, we determine its singular and multiple points and its local structure at them:

Proposition 3 *If $\eta > 0$, the generatrix of the conjugate locus in D_P is a differentiable curve*

$$\tilde{\Gamma} : S^1 \rightarrow D_P$$

which is invariant under reflections through the coordinate circles x_1 and x_3 . All its points are regular, except:

Figure 2: The generatrix of the conjugate locus, for various values of I_1/I_3 .

- Two ordinary cusp points $\pm S_1$ on the x_1 -circle, whose distance from the origin is an increasing function of η which tends to 0 for $\eta \rightarrow 0$ and to 1 for $\eta \rightarrow \infty$. These cusps are tangent to the x_1 -circle and point away from the origin.
- Two (possibly coinciding) ordinary cusp points $\pm S_3$ on the x_3 -circle, at a distance $d(\eta)$ from the origin. These cusps are tangent to the x_3 -circle and point away from the origin if $\langle \eta + \frac{1}{2} \rangle = \langle \eta \rangle$, towards it if $\langle \eta + \frac{1}{2} \rangle > \langle \eta \rangle$.

Note that S_3 and $-S_3$ coincide only in two cases: if η is an integer then $\pm S_3$ is the origin, and if η is half-integer then it is the North-South pole. Furthermore:

- If $\eta > \frac{1}{2}$, the generatrix has $\langle \eta + \frac{1}{2} \rangle$ distinct pairs of double points $\pm P_1, \dots, \pm P_{\langle \eta + \frac{1}{2} \rangle} \in p(\partial D)$, none of which belong to the coordinate circles.
- If $\eta > 1$, the generatrix has $\langle \eta \rangle$ distinct pairs of double points $\pm A_1, \dots, \pm A_{\langle \eta \rangle}$ on the arc of the x_1 -circle comprised between $-S_1$ and S_1 and containing the origin, all different from $\pm S_1$ and from the origin.

The self-intersection of the generatrix at all the double points $P_1, \dots, A_{\langle \eta \rangle}$ is transversal. There are no other multiple points, except, in the case of integral or half-integral η , the double (cusp) point S_3 .⁴

Conditions defining the location of cusps and double points will be determined within the proof of Proposition 3, in the next section.

In order to describe the global structure of the generatrix, we now adopt a discrete point of view and specify the order in which the trace of the generatrix

⁴Note that $\pm S_3$ never coincide with some of the points $P_1, \dots, A_{\langle \eta \rangle}$. According to the Proposition, when $S_3 = -S_3$ is a double point, the two arcs of the generatrix which pass through it form two cusps with equal tangents and opposite orientations, not a transversal intersection.

connects these special points. This is more easily done by considering the (discontinuous) pull-back of $\tilde{\Gamma}$ to D .

First note that, since each double point $P_j \in p(D)$ is the image of a pair of antipodal points on ∂D , each of the pairs $\pm S_1, \pm S_3, \pm A_j, \pm P_j$ of points in D_P has exactly one preimage in the first quadrant of D , that is, a point with non-negative coordinates. Denote these points $S_1, S_3, A_1, \dots, A_{\langle \eta \rangle}, P_1, \dots, P_{\langle \eta + \frac{1}{2} \rangle}$ and note that, by Proposition 3, they are all distinct. Note also that the number of P -points equals that of A -points if $\langle \eta + \frac{1}{2} \rangle = \langle \eta \rangle$ (namely $\eta - [\eta] \leq \frac{1}{2}$) and exceeds it by one if $\langle \eta + \frac{1}{2} \rangle > \langle \eta \rangle$ (namely $\eta - [\eta] > \frac{1}{2}$). Consider now the sequence

$$S_1, P_1, A_1, P_2, A_2, \dots, P_{\langle \eta \rangle}, A_{\langle \eta \rangle}, (P_{\langle \eta + \frac{1}{2} \rangle}), S_3 \quad (3.7)$$

where the point between brackets is of course present only if $\langle \eta + \frac{1}{2} \rangle > \langle \eta \rangle$,⁵ ordered so that the sequences of the x_1 -coordinates of the two subsequences $S_1, A_1, \dots, A_{\langle \eta \rangle}$ and $P_1, P_2, \dots, P_{\langle \eta + \frac{1}{2} \rangle}, S_3$ are both strictly decreasing.

Proposition 4 *For $\eta > 0$, the pull-back to D of the trace of the generatrix is obtained by reflecting through the axes x_1 and x_3 a connected curve which is union of differentiable arcs which connect each point of the sequence (3.7) to the next one.*

Propositions 3 and 4, together, give an essentially complete account of the structure of the generatrix of the conjugate locus. The resulting structure is clearly visible in Figure 2, which shows the trace of the generatrix in a number of cases.

Remarks: (i) The images of Figure 2 have been constructed by numerical evaluation of the conditions (4.12) and (4.16) which define the generatrix. Note that there is perhaps only one feature of the generatrix that is implied by Figures 2 and is not accounted for by Propositions 3 and 4, namely, that each of the arcs of the generatrix between any pair of consecutive points of the sequence (3.7) has no inflection points. This fact is numerically evident, but the proof might be a little bit involved.

(ii) There is a simple heuristic way of describing the structure of the trace of the generatrix. Consider an astroid in $\mathbb{R}^2$ with cusps on the coordinate axes x_1, x_3 , (horizontal) width smaller than one and (vertical) height $2d(\eta)$. Cut it by pairs of horizontal lines symmetric to $x_3 = 0$ in such a way that each section, except possibly the two extreme ones, fits exactly into a circle of radius 1. Then, shifting each section back to the unit circle produces a curve which resembles the pull-back to D of the trace of the generatrix. There is

⁵That is, $S_1 S_3$ if $0 < \eta \leq \frac{1}{2}$, $S_1 P_1 S_3$ if $\frac{1}{2} < \eta \leq 1$, $S_1 P_1 A_1 S_3$ if $1 < \eta \leq \frac{3}{2}$, $S_1 P_1 A_1 P_2 S_3$ if $\frac{3}{2} < \eta \leq 2$, etc.

one aspect of this construction which requires some care: a look at Figure 2 shows indeed that the angles at which different arcs of the generatrix meet on ∂D is not compatible with shifting sections of an astroid: one should rather distort it to a Christmas tree-like continuous curve.

4 Proof of Propositions 2, 3 and 4

A. Quaternions. We work with (unit) quaternions instead of rotation matrices, that is, we work on the 2:1 covering S^3 of $\mathrm{SO}(3)$ (see e.g. [9]). The advantage is of course that quaternions easily allow the determination of the product of two rotations, which is necessary in view of (3.5).

We denote by $Q = (q_0, q_1, q_2, q_3) = (q_0, q)$, with $q = (q_1, q_2, q_3)$, the points of the unit sphere $S^3 \subset \mathbb{R}^4$ endowed with the quaternion group structure given by the product $Q \otimes P = (q_0 p^0 - q \cdot p, p^0 q + q^0 p + q \times p)$. The “Euler parameters” map

$$\mathcal{E} : S^3 \rightarrow \mathrm{SO}(3), \quad \mathcal{E}(Q) = \mathbf{1} + 2q_0 \hat{q} + 2\hat{q}^2$$

is a 2:1 covering map, with $\mathcal{E}(-Q) = \mathcal{E}(Q)$, as well as a group homomorphism, with $\mathcal{E}(Q)\mathcal{E}(P) = \mathcal{E}(Q \otimes P)$. If $\phi \in \mathbb{R}$ and $e \in \mathbb{R}^3$ has unit norm, then

$$\exp(\phi \hat{e}) = \mathcal{E}\left(\pm\left(\cos \frac{\phi}{2}, e \sin \frac{\phi}{2}\right)\right). \quad (4.8)$$

If $t \mapsto (\mathcal{R}^t(m), m^t(m))$ is the solution of the geodesic equations (2.1) with initial conditions $(\mathbf{1}, m)$, then the two covering motions $t \mapsto (Q^t, m^t) \in S^3 \times \mathbb{R}^3$, which satisfy $\mathcal{R}^t = \mathcal{E}(Q^t)$, have initial data $Q = (1, 0, 0, 0)$ and $Q = (-1, 0, 0, 0)$. We need of course to consider only one of them, say the first. Note that the condition $\mathrm{rank} \frac{\partial \mathcal{R}^t}{\partial m}(m) < 3$, which defines the time at which the solution reaches the conjugate point, see (2.3), is equivalent to

$$\mathrm{rank} \frac{\partial Q^t}{\partial m}(m) < 3.$$

Even though we work in S^3 , we describe the results in the solid ball model. Therefore, we use the 2 : 1 covering map

$$\mathcal{F} \circ \mathcal{E} : S^3 \rightarrow B_P$$

which, as one can check, can be written as the composition of the map

$$\mathcal{B} : S^3 \rightarrow B, \quad (q_0, q) \mapsto \begin{cases} \frac{2}{\pi} \mathrm{sign}(q_0) \arcsin(|q|) \frac{q}{|q|} & \text{if } q \neq 0 \text{ and } q_0 \neq 0 \\ 0 & \text{if } q = 0 \\ q & \text{if } q_0 = 0 \end{cases} \quad (4.9)$$

with the antipodal quotient $p : B \rightarrow B_P$.

B. The tangential conjugate locus. We begin the proof of Propositions 2, 3 and 4 by reviewing, here and in Subsection D, some known properties of the time at which the first conjugate point in a given direction is reached (that is, the ‘conjugate locus in the tangent space’). In the prolate case, this function has been determined in [8] and its properties studied in [20]. The extension to the oblate case does not present any difficulty.

From (4.8) and (3.5) it follows that Q -component of the covering motion $t \mapsto (Q^t, m^t) \in S^3 \times \mathbb{R}^3$ with initial datum $Q = (1, 0, 0, 0)$ is

$$Q^t = (c_\tau, \mu s_a) \otimes (c_\rho, \mu s_\rho) \quad (4.10)$$

where $\mu = m/|m|$ and we have written

$$c_\tau = \cos \tau, \quad s_\tau = \sin \tau, \quad c_\rho = \cos \rho, \quad s_\rho = \sin \rho$$

with

$$\tau = \frac{t}{2I_1} |m|, \quad \rho = \frac{t}{2I_1} \eta m_3.$$

In order to simplify the notation we parameterize the curve (4.10) with the rescaled time $\tau = \frac{t}{2I_1}$, thus denoting it $Q^\tau(m)$. (Equivalently, choose $I_1 = 2$). Its four components are

$$\begin{aligned} q_0^\tau(m) &= c_\tau c_\rho - s_\tau s_\rho \mu_3 \\ \begin{pmatrix} q_1^\tau(m) \\ q_2^\tau(m) \end{pmatrix} &= s_\tau \begin{pmatrix} c_\rho & s_\rho \\ -s_\rho & c_\rho \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \\ q_3^\tau(m) &= c_\tau s_\rho + s_\tau c_\rho \mu_3. \end{aligned} \quad (4.11)$$

Let $\tau_\eta(m)$ be the value of τ at which this solution reaches its first conjugate point. In the prolate case, $\tau_\eta(m)$ has been determined in [8]:

Lemma 5 *Assume $|m| = 1$. Then:*

- i. *If $-1 < \eta \leq 0$ then $\tau_\eta(m) = \pi$.*
- ii. *If $\eta > 0$, then $\tau_\eta(m)$ is the smallest positive root τ of the equation*

$$\tan \tau = -\eta \frac{1 - m_3^2}{1 + \eta m_3^2} \tau \quad (4.12)$$

and satisfies $\frac{\pi}{2} < \tau_\eta(m) \leq \pi$, with equality if and only if $m_3 = \pm 1$.

Proof. An elementary though somewhat lengthy computation shows that, evaluated at a point m of unit norm, the four 3×3 minors of the matrix $\frac{\partial Q^T}{\partial m}$ are the four components of the vector

$$\tau s_\tau f(\tau, m) Q^\tau(m)$$

Figure 3: (a) Solutions of equation (4.12). (b) Plot of $m_3 \rightarrow \hat{\tau}_\eta(m_3)$ ($\eta = 1$.)

where

$$f(\tau, m) = (1 + \eta m_3^2) s_\tau + \eta(1 - m_3^2) \tau c_\tau.$$

Since Q^τ has norm one and $m_3^2 \leq 1$, the four minors vanish simultaneously at $\tau > 0$ if and only if either $s_\tau = 0$ or $f(\tau, m) = 0$. Clearly, s_τ first vanishes at $\tau = \pi$. We thus check whether $f(\tau, m)$ vanishes for some $0 < \tau < \pi$. Note that c_τ vanishes in this interval only at $\tau = \frac{\pi}{2}$ but that $f(\frac{\pi}{2}, m) = (1 + \eta m_3^2) > 0$. Hence, the vanishing of $f(\tau, m)$ for some $\tau \in (0, \pi)$ is equivalent to condition (4.12) for some $\tau \in (0, \pi)$. Since $\frac{1-m_3^2}{1+\eta m_3^2} < 1$ for all $\eta > -1$ and $m_3^2 \leq 1$, equation (4.12) does not have any solution τ in $(0, \pi)$ if $\eta < 0$ while it has exactly one if $\eta > 0$, see Figure 3.a. In the latter case, moreover, the solution belongs to the interval $(\frac{\pi}{2}, \pi]$, and equals π if and only if $m_3 = \pm 1$. ■

C. Proof of Proposition 2. Lemma 1 shows that in the oblate case, $-1 < \eta < 0$, all conjugate points are reached after the time $t = 2\pi I_1$. Hence, according to (3.5),

$$\mathcal{L}(m) = \exp(2\pi\eta m_3 \widehat{e}_3) \quad \forall m \in S^2.$$

In the solid model ball this corresponds to the point $(0, 0, 2\eta m_3)$. For $\eta = 0$, this is the origin.

D. Properties of τ_η for $\eta > 0$. We consider now the case $\eta > 0$ and review some elementary properties of the function $\tau_\eta : S^2 \rightarrow \mathbb{R}$ already remarked in [20]. Equation (4.12) implies that $\tau_\eta(m)$ depends on m only through the component m_3 . It is therefore sufficient to study the restriction of τ_η to a meridian of the unit sphere, say the circle $m_2 = 0$, $m_1^2 + m_3^2 = 1$. In order to simplify the notation we identify this meridian with the unit circle

$$C := \{(m_1, m_3) \in \mathbb{R}^2 : m_1^2 + m_3^2 = 1\}$$

and, with a small abuse, we regard τ_η as a function on C . We anticipate that we shall often use the parameterization

$$\mathcal{C} : S^1 \rightarrow C, \quad u \mapsto (\cos u, \sin u) \quad (4.13)$$

of C . We also need to consider the dependency of τ_η on m_3 alone. To this end, we consider the function $\hat{\tau}_\eta : (-1, 1) \rightarrow \mathbb{R}$ defined by

$$\hat{\tau}_\eta(m_3) := \tau_\eta\left(\sqrt{1 - m_3^2}, m_3\right) \quad (4.14)$$

which, by (4.12), is an even function.

Lemma 6 [20] *The function $(\eta, m) \rightarrow \tau_\eta(m)$ is a smooth real function on $\mathbb{R}_+ \times C$. Moreover, for any $\eta > 0$:*

- i. $\tau_\eta : C \rightarrow \mathbb{R}$ has exactly four critical points: $(0, \pm 1)$ where it attains its maximum and $(\pm 1, 0)$ where it attains its minimum. The maximum equals π . The minimum is an increasing function of η which tends to $\frac{\pi}{2}$ for $\eta \rightarrow 0^+$ and to π for $\eta \rightarrow \infty$.*
- ii. The function $m_3 \rightarrow \hat{\tau}_\eta(m_3)$ is strictly increasing in the interval $(0, 1)$.*

The graph of $\hat{\tau}_\eta$, for $\eta > 0$, is shown in Figure 3.b.

Proof. We know from the proof of Lemma 1 that, for any $u \in S^1$ and $\eta > 0$, $\tau_\eta(\mathcal{C}(u))$ is the unique solution in $(\frac{\pi}{2}, \pi]$ of the equation $g(\tau, \eta, u) = 0$, where

$$g(\tau, \eta, u) = \tan \tau + \frac{\cos^2 u}{1 + \eta \sin^2 u} \eta \tau.$$

Since g is smooth in $(\frac{\pi}{2}, \frac{3\pi}{2}) \times \mathbb{R}_+ \times \mathcal{S}^1$ and $\frac{\partial g}{\partial \tau}$ is there everywhere nonzero, the implicit function theorem ensures that $(\eta, u) \mapsto \tau_\eta(\mathcal{C}(u))$ is smooth. A routine computation of the derivative of $u \mapsto \tau_\eta(\mathcal{C}(u))$ gives

$$(\tau_\eta \circ \mathcal{C})' = \frac{2\eta(1 + \eta)\tau_\eta m_1 m_3}{(1 + \eta)(1 + \eta m_3^2) + (\eta \tilde{\tau}_\eta m_1^2)^2}, \quad (4.15)$$

where it is understood that, on the right hand side, m_1 and m_3 stand for $\sin u$ and $\cos u$, respectively, and that τ_η stands for $\tau_\eta \circ \mathcal{C}$. Thus, the only critical points of τ_η on C are those where either $m_1 = 0$ or $m_3 = 0$. The statements about the values of τ_η at these points follow immediately from Figure 3.a. Together with the fact that $m_3 \rightarrow \frac{1 - m_3^2}{1 + \eta m_3^2}$ is a decreasing function of $m_3 \in (0, 1)$, Figure 3.a also shows that $m_3 \mapsto \hat{\tau}_\eta(m_3)$ is increasing in that interval. ■

Remark: The differentiability of $m \mapsto \tau_\eta(m)$ implies that, in the case at hand, the conjugate locus is the image of a differentiable function $S^2 \rightarrow \text{SO}(3)$.

E. The generatrix: singular and double points. Since the conjugate locus $L \subset B_P$ is invariant under rotations around the axis x_3 , we can limit ourselves to determining its generatrix.

First note that the equations (4.11) show that the considered $\mathcal{E}$ -preimage in S^3 of the conjugate locus can be constructed by rotating about the axis $(0, 0, 1)$ the (q_1, q_2, q_3) -component of the curve

$$\Gamma : C \rightarrow S^3, \quad \Gamma(m_1, m_3) := Q^{\tau_\eta(m_1, 0, m_3)}(m_1, 0, m_3).$$

This curve has components

$$\gamma_0 = c_\tau c_\rho - s_\tau s_\rho m_3, \quad \gamma_1 = s_\tau m_1, \quad \gamma_2 = 0, \quad \gamma_3 = c_\tau s_\rho + s_\tau c_\rho m_3 \quad (4.16)$$

and belongs to the section $q_2 = 0$ of S^3 . Here, and everywhere in the following, τ stands for τ_η and ρ for $\rho_\eta := \eta \tau_\eta m_3$. Correspondingly, s_τ and c_τ stand for $\sin \tau_\eta$ and $\cos \tau_\eta$, respectively, and s_ρ and c_ρ for $\sin \rho_\eta$ and $\cos \rho_\eta$. We also note, once for all, that

$$1 > s_\tau \geq 0 > c_\tau \geq -1 \quad \forall \eta > 0, m \in S^2, \quad (4.17)$$

with equalities if and only if $m_3 = \pm 1$.

On the other hand, formulas (4.9) show that the restriction of $\mathcal{F} \circ \mathcal{E}$ to this section of S^3 is onto the section D_P of B_P given by $x_2 = 0$, and that this restriction is in fact a 2 : 1 covering map. Therefore, the conjugate locus in the solid ball model is obtained by rotating about the x_3 -axis the trace of the curve

$$\tilde{\Gamma} := \mathcal{F} \circ \mathcal{E} \circ \Gamma : C \rightarrow D_P.$$

Note that the trace of this curve, that we will call the *generatrix* of the conjugate locus, is symmetric with respect to the axes x_1 and x_3 (as also ensured by Proposition 1).

We begin the study of the generatrix by determining and locating its singular and multiple points.

Lemma 7 Assume $\eta > 0$. Then:

- i. $\tilde{\Gamma} : C \rightarrow D_P$ is regular at all points $(m_1, m_3) \in C$ except:
 - $(\pm 1, 0)$, the images of which belong to the x_1 -axis
 - $(0, \pm 1)$, the images of which belong to the x_3 -axis.
- ii. $\tilde{\Gamma}$ is injective in $C \setminus K$ and 2 : 1 in K , where $K \subset C$ consists:
 - Of all the pairs of points $\pm(m_1, m_3)$ with $\gamma_0(m_1, m_3) = 0$, that is, the points whose image belongs to $p(\partial D)$.

- Of all the pairs of points $(m_1, \pm m_3)$ with $m_3 \neq 0$ and $\gamma_3(m_1, m_3) = 0$, that is, the regular points whose image belongs to the x_1 -axis.

Proof. (i) Since $\mathcal{F} \circ \mathcal{E}$ is a covering map it suffices to show that the statement is true for $\Gamma : C \rightarrow S^3$. Let $\mathcal{C}$ be the parameterization of C given by (4.13) and ' denote differentiation with respect to u . Using (4.12) and $\cos \tau_\eta \neq 0$, a computation gives

$$(\gamma_1 \circ \mathcal{C})' = \left[(\tau_\eta \circ \mathcal{C})' + \frac{\eta m_1 m_3}{1 + \eta m_3^2} \tau_\eta \right] m_1 \cos \tau_\eta.$$

Using (4.15), one then sees that the vanishing of $(\gamma_1 \circ \mathcal{C})'$ is equivalent to $m_1 m_3 = 0$. Thus, the only possible critical points of Γ in C are the four points where $m_1 m_3 = 0$. Equation (4.15) shows that the derivative of τ_η vanishes at these points, as well. It easily follows that the derivatives of s_τ , c_τ , m_1 , m_3 , s_ρ and c_ρ , and hence those of γ_0 and γ_3 , vanish at these four points. To conclude the proof note that $m_3 = 0$ implies $s_\rho = 0$ and hence, by (4.16), $\gamma_3 = 0$. By (4.9), this in turn implies that the component $\tilde{\gamma}_3$ of $\tilde{\Gamma}$ also vanishes. Similarly, if $m_1 = 0$ then $\gamma_1 = 0$ and hence $\tilde{\gamma}_1 = 0$.

(ii) We must check whether there exist different points (m_1, m_3) and (m_1^*, m_3^*) in C such that either

$$\Gamma(m_1, m_3) = \Gamma(m_1^*, m_3^*) \quad (4.18)$$

or

$$\Gamma(m_1, m_3) = -\Gamma(m_1^*, m_3^*). \quad (4.19)$$

In order to simplify the notation, let us denote by $\Gamma = (\gamma_0, \gamma_1, 0, \gamma_3)$ and $\Gamma^* = (\gamma_0^*, \gamma_1^*, 0, \gamma_3^*)$ the values of Γ at these two points and by τ and τ^* the times at which the conjugate points in the directions of $m = (m_1, 0, m_3)$ and of $m^* = (m_1^*, 0, m_3^*)$, respectively, are reached. Moreover, we write ρ for $\eta \tau m_3$ and ρ^* for $\eta \tau^* m_3^*$. Let us distinguish a few mutually incompatible cases.

1. $m_3 = m_3^*$, $m_1^* = -m_1$. Note that $m \neq m^*$ implies $m_1 \neq 0$ and $m_3 \neq \pm 1$. In this case $\frac{\pi}{2} < \tau = \tau^* < \pi$ and therefore $s_\tau = s_{\tau^*} > 0$, $c_\tau = c_{\tau^*}$, $s_\rho = s_{\rho^*}$, $c_\rho = c_{\rho^*}$. It follows from (4.16) that $\gamma_0^* = \gamma_0$, $\gamma_3^* = \gamma_3$, and $\gamma_1^* = -\gamma_1 \neq 0$. Condition (4.18) is therefore impossible. In turn, condition (4.19) is fulfilled if and only if $\gamma_0 = \gamma_3 = 0$, that is, given that the quaternions have unit norm, $\gamma_1 = s_\tau m_1 = \pm 1$. However, this never happens because $|s_\tau| < 1$.
2. $m_3 = -m_3^* \neq 0$, $m_1^* = \pm m_1$. In this case $\tau = \tau^*$ and hence $s_\tau = s_{\tau^*}$, $c_\tau = c_{\tau^*}$, $s_\rho = -s_{\rho^*}$, $c_\rho = c_{\rho^*}$. This implies $\gamma_0^* = \gamma_0$ and $\gamma_3^* = -\gamma_3$. This also implies that $\gamma_1 = \gamma_1^*$ if and only if $m_1 = m_1^*$. (Indeed, if $m_3 = \pm 1$ then $s_\tau = s_{\tau^*} > 0$ while if $m_3 = \pm 1$ then $m_1 = m_1^* = 0$). Thus, condition

(4.18) is satisfied if and only if $m_1^* = m_1$ and $\gamma_3 = 0$; the latter condition means exactly that $\gamma_0^* = \gamma_0$ and $\gamma_3^* = -\gamma_3$. Similarly, $\gamma_1 = -\gamma_1^*$ if and only if $m_1 = -m_1^*$. Hence, condition (4.19) is satisfied if and only if $m_1^* = -m_1$ and $\gamma_0 = 0$; the latter condition means precisely that $|\tilde{\Gamma}| = |\tilde{\Gamma}^*| = 1$.

3. If $|m_3| < |m_3^*|$ then $|m_1| > |m_1^*|$ and, by Lemma 2 and by the fact that $\hat{\tau}_\eta$ is an even function of m_3 , $\tau < \tau^*$ and hence $s_\tau > s_{\tau^*}$; thus $|\gamma_1| > |\gamma_1^*|$ and neither condition (4.18) nor condition (4.19) are satisfied. The case $|m_3| > |m_3^*|$ is analogous.

Thus, two different points (m_1, m_3) and (m_1^*, m_3^*) of C produce the same value of $\tilde{\Gamma}$ if and only if $m_3^* = -m_3$ and either $m_1^* = m_1$ (in which case $m_3 \neq 0$) or $m_1^* = -m_1$. In the first case this value belongs to the x_1 axis, in the second case it is at unit distance from the origin. In both cases, the value has exactly two preimages. Conversely, it follows from formulas (4.16) that if $\gamma_3(m_1, m_3) = 0$ then $\Gamma(m_1, m_3) = \Gamma(m_1, -m_3)$ and that, if $\gamma_0(m_1, m_3) = 0$ then $\tilde{\Gamma}(-m_1, -m_3) = -\tilde{\Gamma}(m_1, m_3)$. ■

Remark: Statements (i) of Lemmas 2 and 3 show that the singular points of the generatrix coincide with the critical points of the time τ_η at which the conjugate point is reached. In two-dimensional Riemannian manifolds, this property is always true for the conjugate locus and follows in that case by Gauss Lemma, as claimed by [14]. This is in general not true for the conjugate locus in the three-dimensional case.

We now show that the singular and double points of the generatrix are the only points at which it intersects the coordinate axes and the unit circle, and we count and locate them.

Lemma 8 *For $\eta > 0$ the trace of the curve $\tilde{\Gamma} : C \rightarrow D_P$ has the following properties:*

- i. *It intersects the axis x_3*
 - *at the origin, if η is integral*
 - *in the North–South pole, if η is half–integral*
 - *in two distinct points at a distance $1 > d(\eta) > 0$ from the origin, in all other cases.*

The intersection points are the images of $(0, \pm 1)$ and are singular points of $\tilde{\Gamma}$.

- ii. *Besides the origin in case of integral η , it intersects the x_1 –axis in $2\langle\eta\rangle + 2$ distinct points, symmetrically disposed with respect to the origin. The two farthest of these points from the origin are a distance $\tau_\eta(1, 0)/\pi$ from it and are the singular points $\tilde{\Gamma}(\pm 1, 0)$. All others are double points.*
- iii. *Besides the North–South pole in case of half–integral η , it has $2\langle\eta + \frac{1}{2}\rangle$ distinct points on $p(\partial D)$, which are all double points.*

Proof. (i) By (4.9), $\tilde{\Gamma}(m_1, m_3)$ belongs to the x_3 -axis if and only if $\gamma_1(m_1, m_3) = 0$ that is, by (4.16), if and only if $s_\tau m_1 = 0$. Since both s_τ and m_1 vanish if and only if $|m_3| = 1$, we conclude that $(m_1, m_3) = (0, \pm 1)$. At these points $\tau_\eta = \pi$, $c_\tau = -1$, $s_\tau = 0$, $c_\rho = \cos(\pi\eta)$ and $s_\rho = \pm \sin(\pi\eta)$. Thus $\Gamma(0, \pm 1) = (-\cos \pi\eta, 0, 0, \mp \sin \pi\eta)$. Therefore, the two points $\tilde{\Gamma}(0, \pm 1)$ lie on the x_3 -axis, at a distance $\frac{2}{\pi} |\arcsin(\sin(\eta\pi))| = d(\eta)$ from the origin and symmetric with respect to it.

(ii) Intersections with the x_1 -axis occur at those (m_1, m_3) such that $\gamma_3(m_1, m_3) = 0$. Let us distinguish two cases.

- If $m_3 = 0$, then $m_1 = \pm 1$. Thus $s_\rho = 0$, $c_\rho = 1$ and hence $\Gamma(\pm 1, 0) = (c_\tau, \pm s_\tau, 0, 0)$ are the points on the x_1 -axis at distance $d(\frac{1}{\pi}\tau_\eta(1, 0))$ from the origin. Since $\frac{\pi}{2} < \tau_\eta(1, 0) < \pi$ (see Lemma 2) this distance equals $\frac{1}{\pi}\tau_\eta(1, 0)$. As we know from Lemma 2, these two intersection points are singular points of $\tilde{\Gamma}$.
- Assume now $m_3 \neq 0$. We have already seen that the two points $\tilde{\Gamma}(0, \pm 1)$ belong to the x_3 -axis. Therefore, they belong to the x_1 -axis only if they coincide with the origin, which happens for integral η . Let us thus assume $m_3 \neq \pm 1$. The condition $\gamma_3(m_1, m_3) = 0$ is $c_\tau s_\rho + s_\tau c_\rho m_3 = 0$ and, if $0 < |m_3| < 1$, implies

$$s_\rho c_\rho \neq 0. \quad (4.20)$$

Indeed, $c_\rho \neq 0$ since otherwise $0 \neq |c_\tau| = |c_\tau s_\rho| = |s_\tau c_\rho m_3| = 0$; hence $s_\rho = -c_\rho m_3 \tan \tau \neq \eta c_\rho m_3 \frac{1-m_3^2}{1+\eta m_3^2} \neq 0$. Therefore, the condition $\gamma_3 = 0$ can be written as $m_3 \tan \tau_\eta = -\tan \rho$, or equivalently, and using a more precise notation, as

$$\frac{1 - m_3^2}{1 + \eta m_3^2} \eta m_3 \tau_\eta(m_1, m_3) = \tan(\eta m_3 \tau_\eta(m_1, m_3)).$$

If (m_1, m_3) is a solution to this equation, then also $(\pm m_1, \pm m_3)$ are solutions and $\tau_\eta(\pm m_1, \pm m_3) = \hat{\tau}_\eta(m_3)$, see (4.14). It follows that it suffices to find the solutions $m_3 \in (0, 1)$ of

$$k(m_3) \hat{\rho}(m_3) = \tan(\hat{\rho}(m_3)) \quad (4.21)$$

where of course $\hat{\rho}(m_3) = \eta m_3 \hat{\tau}_\eta(m_3)$ and

$$k(m_3) := \frac{1 - m_3^2}{1 + \eta m_3^2}.$$

Note that, since $m_3 \mapsto \hat{\tau}_\eta(m_3)$ is increasing,

$$\frac{d\hat{\rho}}{dm_3}(m_3) = \frac{d}{dm_3}(m_3 \hat{\tau}_\eta(m_3)) = \hat{\tau}_\eta(m_3) + m_3 \frac{d\hat{\tau}_\eta}{dm_3}(m_3) \geq \tau_\eta(m_3) > 0$$

Figure 4: Solutions of $\rho\hat{k}(\rho) = \tan \rho$ (dotted line) and of $\rho\hat{k}(\rho) = -\cot \rho$ (dashed). The solid line represents the graph of $\rho\hat{k}(\rho)$.

for all $m_3 \in (0, 1)$. By Lemma 1, we see that $m_3 \mapsto \hat{\rho}(m_3)$ maps the interval $(0, 1)$ diffeomorphically onto the interval $(0, \eta\pi)$. Thus, instead of (4.21) we can equivalently solve the equation

$$\rho\hat{k}(\rho) = \tan \rho, \quad \rho \in (0, \eta\pi) \quad (4.22)$$

where $k = \hat{k} \circ \hat{\rho}$. Note that $\hat{k}(0) = 1$, $\hat{k}(\eta\pi) = 0$ and $0 < \hat{k}(y) < 1$ for all $y \in (0, \eta\pi)$. Thus $\rho\hat{k}(\rho)$ vanishes at $\rho = 0$ and at $\rho = \eta\pi$ and its graph lies between the ρ -axis and the diagonal. Therefore, its graph intersects all branches of $\tan \rho$ in $0 < \rho < \eta\pi$ in at least one point, except the first one, see Figure 4. In fact, there is exactly one intersection with each such branch, because the derivative of $\tan \rho$ is larger than 1 while, as we now show,

$$\left| \frac{d}{d\rho}(\rho\hat{k}(\rho)) \right| \leq 1, \quad 0 < \rho < \eta\pi.$$

Indeed, denoting by a prime the derivative with respect to m_3 , we have $\frac{d}{d\rho}(\rho\hat{k}) = (k\hat{\rho}' + \hat{\rho}k')/\hat{\rho}'$; this implies $|\frac{d}{d\rho}(\rho\hat{k})| \leq |k| \leq 1$ because k , $\hat{\rho}$ and $\hat{\rho}'$ are everywhere positive in $(0, 1)$ while k' is there everywhere negative.

We thus conclude that equation (4.22) has $\langle \eta \rangle$ distinct solutions $m_3^{(1)}(\eta), \dots, m_3^{(\langle \eta \rangle)}(\eta)$ in the interval $(0, 1)$. (If η is an integer then there is an intersection also at $\rho = \eta\pi$, see Figure 4; since $\rho = \eta\pi$ corresponds to $m_3 = 1$, the intersection point is the origin). Correspondingly, there are $4\langle \eta \rangle$ points $(\pm m_1^{(j)}, \pm m_3^{(j)})$ in C whose images under $\tilde{\Gamma}$ belong to the x_1 -axis, and none of them coincides with the origin. By Lemma 3, all these images are double points of $\tilde{\Gamma}$, so that only $2\langle \eta \rangle$ of them are distinct. Note now that the distance from the origin of $\tilde{\Gamma}(\pm m_1^{(j)}, \pm m_3^{(j)})$ is $\frac{2}{\pi} \arcsin(|s_\tau m_1^{(j)}|)$, which is a strictly increasing function of $|s_\tau m_1^{(j)}|$. Since $|m_1^{(j)}| \leq 1$ and $s_\tau \leq \sin(\tau_\eta(0, 1))$, these points are all at a distance $< \tau_\eta(1, 0)/\pi$ from the origin.

(iii) We already know from (i) that if η is half-integral then the images of $(0, \pm 1)$ coincide with the North–South pole of D_P . We also know from (ii) that the images of points of C with $m_3 = 0$ lie on the x_1 -axis at a distance < 1 from the origin. We thus restrict our considerations to $0 < m_3 < 1$. The condition for $|\tilde{\Gamma}(m_1, m_3)| = 1$ is $\gamma_0(m_1, m_3) = 0$, namely $c_\tau c_\rho = s_\tau s_\rho m_3$. If this condition is satisfied, then necessarily (4.20) is satisfied too, that is $s_\rho c_\rho \neq 0$ ($s_\rho \neq 0$ since otherwise $0 \neq |c_\tau| = |c_\tau c_\rho| = |s_\tau s_\rho m_3| = 0$; in turn, $c_\rho = s_\rho m_3 \tan \tau \neq 0$). Hence, using also (4.12), this condition can be written as

$$-\cot(\eta m_3 \tau_\eta) = k(m_3) \eta m_3 \tau_\eta.$$

An argument analogous to that of the proof of (ii) shows now that this equation has $\langle \eta + \frac{1}{2} \rangle$ distinct solutions $0 < m_3^j < 1$, see Figure 4. Correspondingly, there are $4\langle \eta + \frac{1}{2} \rangle$ points $(\pm m_1^j, \pm m_3^j)$ whose images belong to the boundary of D . Identifying antipodal points gives $2\langle \eta + \frac{1}{2} \rangle$ distinct double points in D_P . ■

F. Local structure near the singular and double points. We study now the local structure of the generatrix near its singular and double points. We shall show that the singular points are ordinary cusps and that all self-intersections are transversal, thus completing the proof of Proposition 3.

We say that an ordinary cusp tangent to the x -axis of a curve $u \mapsto (x(u), y(u)) \in \mathbb{R}^2$, see (3.6), is *positively oriented* if $x''(u_0) < 0$ and that it is *negatively oriented* if $x''(u_0) > 0$. If $x(u_0) \neq 0$, then we also say that the cusp *points away from the origin* if $x(u_0)x''(u_0) < 0$ and that it *points towards the origin* if $x(u_0)x''(u_0) > 0$.

Lemma 9 *For any $\eta > 0$:*

- i. $\tilde{\Gamma}$ has ordinary cusps at $(\pm 1, 0)$ tangent to the x_1 -axis and facing away from the origin.
- ii. $\tilde{\Gamma}$ has ordinary cusps at $(0, \pm 1)$ tangent to the x_3 -axis. If not at the origin, both cusps point away from the origin if $\langle \eta + \frac{1}{2} \rangle = \langle \eta \rangle$, towards it otherwise. If at the origin, the cup at $(0, 1)$ is positively oriented and that at $(0, -1)$ is negatively oriented.

Proof. We begin by linking the cusp conditions for the generatrix $\tilde{\Gamma}$ in D_P to conditions on its pull-back Γ to S^3 , which will be easier to verify. This is the content of the next Lemma:

Lemma 10 *Assume $\eta > 0$. Consider the parameterization $\mathcal{C} : u \mapsto (\cos u, \sin u)$ of C and regard the components of Γ and $\tilde{\Gamma}$ as functions of u instead of $(\cos u, \sin u)$. Let $u_0 \in S^1$ be a singular point of Γ . Assume that*

$$\gamma_0(u_0) \neq 0, \pm 1$$

(that is, $\tilde{\Gamma}(u_0) \neq p(0)$ and $\notin p(\partial D)$) and that

$$\gamma_h(u_0) = \gamma_h''(u_0) = 0, \quad \gamma_h'''(u_0) \neq 0, \quad \gamma_k''(u_0) \neq 0 \quad (4.23)$$

where either $h = 1, k = 3$ or $h = 3, k = 1$. Then, $\tilde{\Gamma}$ has a an ordinary cusp in u_0 which is tangent to the axis x_k . Moreover, at u_0 :

$$\begin{aligned} \text{sign}(\tilde{\gamma}_k \tilde{\gamma}_k'') &= \text{sign}(\gamma_k \gamma_k'') & \text{if } \tilde{\gamma}_k \neq 0 \\ \text{sign}(\tilde{\gamma}_k'') &= \text{sign}(\gamma_0 \gamma_k'') & \text{if } \tilde{\gamma}_k = 0. \end{aligned}$$

Proof. Since Γ and $\tilde{\Gamma}$ have the same singular points, $\tilde{\gamma}_1'(u_0) = \tilde{\gamma}_3'(u_0) = 0$. In order to show that $\tilde{\Gamma}$ has a cusp at u_0 tangent to the axis x_k we show that also $\tilde{\gamma}_h$ and $\tilde{\gamma}_k$ satisfy conditions (4.23). According to (4.9),

$$\tilde{\gamma}_j(u) = f(u) \gamma_j(u), \quad j = 1, 3,$$

with

$$f(u) = \frac{2}{\pi} \text{sign}(\gamma_0(u)) \frac{\arcsin |\gamma(u)|}{|\gamma(u)|}.$$

Hence $\tilde{\gamma}_j'' = f'' \gamma_j + 2f' \gamma_j' + f \gamma_j''$ etc. and so, in u_0 :

$$\tilde{\gamma}_h'' = 0, \quad \tilde{\gamma}_h''' = f \gamma_h''' \neq 0, \quad \tilde{\gamma}_k'' = f'' \gamma_k + f \gamma_k'',$$

where we have also used that $f(u_0) \neq 0$. So, we need only to show that $f \gamma_k'' + f'' \gamma_k$ does not vanish at u_0 . Write

$$\sigma(u) = \frac{2}{\pi} \text{sign}(\gamma_0(u)), \quad g(x) = \frac{\arcsin \sqrt{x}}{\sqrt{x}} \quad (x > 0),$$

so that

$$f(u) = \sigma(u) g(|\gamma(u)|^2)$$

for $\gamma = (\gamma_1, \gamma_2, \gamma_3)$. Observing that σ is constant in a neighbourhood of u_0 (since $\gamma_0(u_0) \neq 0$) we deduce

$$f''(u_0) = 2\sigma(u_0) g'(|\gamma(u_0)|^2) \gamma(u_0) \cdot \gamma''(u_0),$$

where g' is the derivative of $x \mapsto g(x)$, that is,

$$g'(x) = \frac{1}{2x} \left(\frac{1}{\sqrt{1-x}} - g(x) \right).$$

Hence, at u_0 , where $|\gamma|^2 = \gamma_k^2$,

$$f'' = 2\sigma g'(\gamma_k^2) \gamma_k \gamma_k'' = \frac{\sigma \gamma_k''}{\gamma_k \sqrt{1-\gamma_k^2}} - \frac{f \gamma_k''}{\gamma_k}$$

and so, finally,

$$\tilde{\gamma}_k'' = f''\gamma_k + f\gamma_k'' = \frac{\sigma\gamma_k''}{\sqrt{1-\gamma_k^2}} \neq 0. \quad (4.24)$$

From this it also follows that, at u_0 :

$$\tilde{\gamma}_k\tilde{\gamma}_k'' = \frac{\sigma^2 g(\gamma_k^2)}{\sqrt{1-\gamma_k^2}} \gamma_k\gamma_k''$$

which shows that, at u_0 , $\tilde{\gamma}_k\tilde{\gamma}_k''$ has the same sign of $\gamma_k\gamma_k''$. ■

We now go back to the proof of Lemma 5:

(i) Because of the symmetry of $\tilde{\Gamma}$ with respect to the two axes x_1 and x_3 , we need only consider the point $(m_1, m_3) = (1, 0)$. We know from Lemma 4 that for no value of η does the point $\tilde{\Gamma}(1, 0)$ belong to either the origin or $p(\partial D)$, and so we can apply Lemma 6. First compute, thinking of τ_η as a function of u and using (4.15),

$$\tau_\eta' = 0, \quad \tau_\eta'' = \frac{2\eta(1+\eta)\tau_\eta}{1+\eta+\eta^2\tau_\eta^2} > 0.$$

It is then not difficult to verify using (4.16) that, at $(1, 0)$, where $s_\rho = 0$ and $c_\rho = 1$,

$$\gamma_1 = s_\tau, \quad \gamma_1' = 0, \quad \gamma_1'' = c_\tau\tau_\eta'' - s_\tau$$

and

$$\gamma_3 = 0, \quad \gamma_3' = 0, \quad \gamma_3'' = 0, \quad \gamma_3''' = [2\eta^3\tau_\eta^3 + 3(1+\eta+\eta^2\tau_\eta^2)\tau_\eta'']c_\tau.$$

Since $c_\tau < 0$, this proves that $\gamma_3''' \neq 0$. Moreover, since $s_\tau > 0$, using (4.12) we also see that

$$\gamma_1'' = -s_\tau \frac{3+3\eta+\eta^2\tau_\eta^2}{1+\eta+\eta^2\tau_\eta^2} < 0.$$

(ii) Here too, we can consider only one of the two points, say $(m_1, m_3) = (0, 1)$. Note that then $s_\tau = 0$, $c_\tau = -1$, $\gamma_0 = -c_\rho$, $\gamma_3 = -s_\rho$.

Assume first that η is neither integral nor half-integral, so that $\tilde{\Gamma}(0, 1)$ is not the origin and is not in $p(\partial D)$ and therefore $c_\rho s_\rho \neq 0$. We now find

$$\tau = \pi, \quad \tau' = 0, \quad \tau'' = -\frac{2\eta}{1+\eta}\pi \neq 0$$

so that

$$\gamma_1 = \gamma_1' = \gamma_1'' = 0, \quad \gamma_1''' = 3\tau'' \neq 0$$

and

$$\gamma_3 = -s_\rho, \quad \gamma'_3 = 0, \quad \gamma''_3 = (\eta\tau - (1 + \eta)\tau'')c_\rho = 3\eta\pi c_\rho \neq 0.$$

From this it also follows that $\text{sign}(\gamma_3\gamma''_3) = -\text{sign}(\sin(2\pi\eta))$, which equals $+1$ if $\langle\eta + \frac{1}{2}\rangle > \langle\eta\rangle$ and -1 if $\langle\eta + \frac{1}{2}\rangle = \langle\eta\rangle$.

Consider now an integral or half-integral value $\bar{\eta}$ of the parameter η , so that $\tilde{\Gamma}(0, 1)$ is either the origin or the North–South pole and we cannot use Lemma 6. By Lemma 2, however, the time at which the conjugate point is reached is a smooth function of η and m . Therefore, the values of the generatrix and of its derivatives at $(0, 1)$ are continuous functions of η and we can compute their value at $\bar{\eta}$ by taking the limit as $\eta \rightarrow \bar{\eta}$ of the values found above, that is,

$$\tilde{\gamma}''_1 = 0, \quad \tilde{\gamma}'''_1 = f\gamma'''_1, \quad \tilde{\gamma}''_3 = \frac{\sigma\gamma''_3}{\sqrt{1-\gamma_3^2}} \quad \text{for } \eta \neq \bar{\eta}.$$

Thus $\tilde{\gamma}''_1 \rightarrow 0$. On the other hand, since $\gamma_0 = -c_\rho$ and $\gamma_3 = -\sin\rho$,

$$f = \frac{2}{\pi} \text{sign}(-c_\rho) \frac{|\rho(\bmod \pi)|}{|s_\rho|}, \quad \gamma''_3 = 3\eta\pi c_\rho \quad \text{for } \eta \neq \bar{\eta}$$

and therefore,

$$\tilde{\gamma}''_3 = \frac{2}{\pi} \text{sign}(-c_\rho) \frac{3\eta\pi c_\rho}{|c_\rho|} \rightarrow -6\bar{\eta} < 0.$$

In order to prove that the singular point is an ordinary cusp it only remains to show that $\tilde{\gamma}'''_1 \neq 0$. Since

$$\gamma'''_1 = 3\tau'' \rightarrow -\frac{6\bar{\eta}}{1+\bar{\eta}} \neq 0$$

this amounts to the fact that $f(0, 1) \neq 0$ as $\eta \rightarrow \bar{\eta}$. This is obvious if $\bar{\eta}$ is half-integral because then $\tilde{\gamma}_3 = f\gamma_3 \rightarrow \pm 1$, while if $\bar{\eta}$ is integral then $s_\rho \rightarrow \sin(\bar{\eta}\pi) = 0$, $c_\rho \rightarrow \pm 1$ and so $f \rightarrow -\frac{2}{\pi}\text{sign}(c_\rho) \neq 0$. It only remains to prove the statements about the cusp orientation. If $\bar{\eta}$ is integral then the cusp is at the origin and, since $\tilde{\gamma}''_3 = -6\bar{\eta} < 0$, it is positively oriented. If instead $\bar{\eta}$ is half-integral then $\text{sign}(\tilde{\gamma}_3\tilde{\gamma}''_3) = -\text{sign}(\sin(2\pi\eta))$ for any $\eta \rightarrow \bar{\eta}$, which in the limit equals $-\text{sign}(\sin(2\pi\bar{\eta}))$. ■

Next, we consider the structure of the generatrix at its double points:

Lemma 11 *If $\eta > 0$, then the self-intersections of the generatrix at its double points are all transversal.*

Proof. We consider the usual parameterization (4.13) of the circle C , regard all functions as functions of u and denote by a prime the derivative with respect to u . Let u and $u^* \neq u$ be such that $\tilde{\Gamma}(u) = \tilde{\Gamma}(u^*)$. We must prove that $\tilde{\Gamma}'(u)$ and $\tilde{\Gamma}'(u^*)$ are not parallel. Since $\mathcal{F} \circ \mathcal{E}$ is a local diffeomorphism, this is equivalent to the fact that $\Gamma'(u)$ and $\Gamma'(u^*)$ are not parallel, that is, that at least one of the following three inequalities is satisfied:

$$\begin{aligned}\gamma'_0(u)\gamma'_1(u^*) &\neq \gamma'_0(u^*)\gamma'_1(u) \\ \gamma'_0(u)\gamma'_3(u^*) &\neq \gamma'_0(u^*)\gamma'_3(u) \\ \gamma'_1(u)\gamma'_3(u^*) &\neq \gamma'_1(u^*)\gamma'_3(u).\end{aligned}\tag{4.25}$$

As we know from Lemmas 3 and 4 the double points of the generatrix are of two types.

(i) The double point is in $p(\partial D)$. In this case $m(u^*) = -m(u)$ with $0 < |m_3(u)| < 1$ and $\gamma_0(u) = 0$. Note now that, since γ_0 is an even function of m while γ_1 and γ_3 are odd, we have $\gamma'_0(u^*) = \gamma'_0(u)$, $\gamma'_1(u^*) = -\gamma'_1(u)$ and $\gamma'_3(u^*) = -\gamma'_3(u)$. Therefore, inequalities (4.25) are equivalent to the fact that one at least of the two inequalities

$$\gamma'_0(u)\gamma'_1(u) \neq -\gamma'_0(u)\gamma'_1(u), \quad \gamma'_0(u)\gamma'_3(u) \neq -\gamma'_0(u)\gamma'_3(u)$$

is satisfied. This happens if and only if $\gamma'_0(u) \neq 0$ and at least one between γ'_1 and γ'_3 does not vanish at u . But as we have shown in the proof of statement i. of Lemma 3, γ'_1 vanishes only at the singular points. Therefore, we must show that $\gamma'_0(u) \neq 0$. From (4.16), using $\rho' = \eta\tau'm_3 + \eta\tau m_1$, we obtain

$$\gamma'_0(u) = -[(1 + \eta m_3^2)s_\tau c_\rho + (1 + \eta)m_3 c_\tau s_\rho]\tau' - [s_\tau s_\rho + \eta\tau(c_\tau s_\rho + s_\tau c_\rho)]m_1.$$

The two conditions (4.12) and $\gamma_0 = 0$ give $s_\tau = -\eta\frac{1-m_3^2}{1+\eta m_3^2}\tau c_\tau$ and $c_\rho = -\eta\frac{1-m_3^2}{1+\eta m_3^2}\tau m_3 s_\rho$. Using the expression (4.15) of τ' we eventually obtain

$$\begin{aligned}\gamma'_0(u) &= -\eta\tau c_\tau s_\rho m_1 m_3^2 \frac{3(1 + \eta)^2 m_3^2 + \eta^2 \tau^2 m_1^4}{1 + \eta m_3^2} \\ &\quad \times \frac{(1 + \eta)(1 + \eta m_3^2) + \eta^2 \tau^2 m_1^2 + \eta^2 \tau^2 (1 - m_3)^2}{(1 + \eta)(1 + \eta m_3^2) + \eta^2 \tau^2 m_1^4}\end{aligned}$$

which shows that $\gamma'_0(u) < 0$.

(ii) The case of a double point on the x_1 -axis may be treated in a very similar way. Now $m_1(u^*) = m_1(u)$ and $m_3(u^*) = -m_3(u)$ with $0 < |m_3(u)| < 1$ and $\gamma_3(u) = 0$. At variance from the previous case we have $\gamma'_0(u^*) = \gamma'_0(u)$, $\gamma'_1(u^*) = \gamma'_1(u)$, $\gamma'_3(u^*) = -\gamma'_3(u)$ and inequalities (4.25) reduce to $\gamma'_3(u) \neq 0$. Using (4.12) and $\gamma_3 = 0$ yields $s_\rho = \eta\frac{1-m_3^2}{1+\eta m_3^2}\tau m_3 c_\rho$. Eventually, one finds an

expression for $\gamma'_3(u)$ which coincides with the one above for γ'_0 , except for the sign and for the fact that the factor s_ρ is replaced by c_ρ . Thus $\gamma'_3(u) > 0$. ■

Note that at this point we have proved all statements of Proposition 3. We now describe the global structure of the generatrix, thus proving Proposition 4.

G. Proof of Proposition 4. We study now the pull-back of the trace of $\tilde{\Gamma}$ to D . Because of the symmetry of $\tilde{\Gamma}$ with respect to the two axes x_1 and x_3 , we can restrict ourselves to values of (m_1, m_3) with $m_1 \geq 0$ and $m_3 \geq 0$: the images of points with $m_1 < 0$ (respectively $m_3 < 0$) will then be obtained by reflecting through the x_1 -axis (respectively, the x_3 -axis).

Let C_+ be the subset of C where $m_1 \geq 0$, $m_3 \geq 0$, ordered from $s^1 := (1, 0)$ to $s^3 := (0, 1)$. According to Lemma 4 and the symmetry properties of the components of Γ (see (4.16)), C_+ contains:

- the two singular points s^1 and s^3 .
- $\langle \eta + \frac{1}{2} \rangle$ regular points $b^j = (b_1^j, b_3^j)$ whose images belong to ∂D
- $\langle \eta \rangle$ regular points $a^j = (a_1^j, 0)$ whose images belong to the x_1 -axis.

Figure 4 shows that these points are reached in the following order:

$$s^1, b^1, a^1, \dots, b^{\langle \eta \rangle}, a^{\langle \eta \rangle}, \left(b^{\langle \eta + \frac{1}{2} \rangle}\right), s^3 \quad (4.26)$$

where the point between brackets is present only if $\langle \eta + \frac{1}{2} \rangle > \langle \eta \rangle$.

We now note that the absolute values of the x_1 -coordinates of the two subsequences

$$\tilde{\Gamma}(s^1), \tilde{\Gamma}(a^1), \dots, \tilde{\Gamma}(a^{\langle \eta \rangle})$$

and

$$\tilde{\Gamma}(b^1), \tilde{\Gamma}(b^2), \dots, \left(\tilde{\Gamma}(b^{\langle \eta + \frac{1}{2} \rangle})\right), \tilde{\Gamma}(s^3)$$

form two decreasing sequences. In fact, we already know from Lemma 4 that the absolute values of the x_1 -coordinates of $\tilde{\Gamma}(a^j)$ are smaller than that of $\tilde{\Gamma}(s^1)$. On the other hand, the distance from the origin of a point $\tilde{\Gamma}(m_1, m_3)$ on the x_1 -axis is $\frac{2}{\pi} \arcsin(s_\tau |m_1|)$, which is an increasing function of m_1 . The same happens if $\Gamma(m_1, m_3) = (\gamma_0, \gamma_1, \gamma_3)$ is such that $\gamma_0 = 0$, so that $\tilde{\Gamma}(m_1, m_3)$ lies on the boundary of D , because then the absolute value of the x_1 -coordinate of the latter point is $s_\tau |m_1|$.

We now locate the p -preimages of the $\tilde{\Gamma}$ -images of the points (4.26) in the four quadrants of $D \subset \mathbb{R}^2$, by determining the signs of the coordinates of these points, that we denote $x_1(s^1)$, $x_3(s^1)$ etc. To this end, in view of (4.9) we need only study the signs of the components of Γ , using (4.16). Clearly, $\gamma_1 = s_\tau m_1$ is positive for positive m_1 . It is also immediately checked that

- $x_1(s^1) < 0$. (If $m_3 = 0$ then $c_\rho = 1$ and $\gamma_0 = c_\tau c_\rho < 0$).
- $x_3(s^3)$, if nonzero, has the same sign as $s_\rho c_\rho$. (If $m_3 = 1$ then $s_\tau = 0$ and $c_\tau = -1$ and so $\gamma_0 = -c_\rho$ and $\gamma_3 = -s_\rho$.)

At all other points of (4.26), $0 < m_3 < 1$; hence, from the proof of Lemma 4 we know that $s_\rho c_\rho \neq 0$. Therefore

$$\begin{aligned}\operatorname{sign}(\gamma_0) &= \operatorname{sign}(s_\rho) \operatorname{sign}((-\cot \rho) - \rho \tilde{k}(\rho)) \\ \operatorname{sign}(\gamma_3) &= \operatorname{sign}(c_\rho) \operatorname{sign}(\rho \tilde{k}(\rho) - \tan \rho).\end{aligned}$$

Based on this, and on Figure 4, one then easily checks that

- If (m_1, m_3) lies between s^1 and the second point of (4.26), then $x_1(m_1, m_3) < 0$ and $x_3(m_1, m_3) > 0$. (A glance at Figure 4 shows that γ_0 and γ_3 are both negative).
- When (m_1, m_3) passes through a point b^j , both coordinates x_1 and x_3 change sign. This means that, in D , the curve reaches ∂D and jumps to the antipodal point.
- When (m_1, m_3) passes through a point a^j , the x_3 -coordinate changes sign while the sign of x_1 remains unchanged.
- When (m_1, m_3) passes through s_1 , x_1 changes sign while x_3 does not (γ_1 changes sign, while γ_0 and γ_3 do not).

Taking into account that, by Lemma 4, $\tilde{\Gamma}$ intersects the coordinate axes or $p(\partial D)$ only at the points of the sequence (4.26), the previous information allows us to describe the pullback to D of the trace of $\tilde{\Gamma}|_{C_+}$.

Let us use the following notation: the two p -preimages of $\tilde{\Gamma}(\pm s^1)$ and of $\tilde{\Gamma}(\pm a^j)$, which belong to the x_1 -axis, will be denoted S_1^+ and A_j^+ if their x_1 -coordinate is positive, S_1^- and A_j^- if their x_1 -coordinate is negative. Similarly, the p -preimages of $\tilde{\Gamma}(\pm s^3)$ will be denoted S_3^+ and S_3^- , depending on whether their x_3 -coordinate is positive or negative (and any of the two symbols if it is zero). For the four p -preimages of the points $\tilde{\Gamma}(\pm b_1^j, \pm b_3^j)$ we write $P_j^{\pm\pm}$ where the first sign specifies the sign of the x_1 -coordinate and the second that of their x_3 -coordinate.

Move now (m_1, m_3) in an anticlockwise direction along C , starting from s^1 . In D , $p^{-1}(\tilde{\Gamma}(m_1, m_3))$ starts from S_1^- and reaches P_1^{-+} , where p is $2 : 1$. However, at this point both coordinates change sign and so $p^{-1}(\tilde{\Gamma}(m_1, m_3))$ ‘antipodally’ jumps to the $(+, -)$ -quadrant. The next point that it reaches is therefore A_1^+ . Continuing, one sees that when m has reached s^3 , $p^{-1}(\tilde{\Gamma}(m_1, m_3))$ has passed through the following sequence of points of D :

$$S_1^-, P_1^{-+} \equiv P_1^{+-}, A_1^+, P_2^{++} \equiv P_2^{--}, A_2^-, \dots, S_3^\pm$$

where the last point is S^+ if $\langle \eta + \frac{1}{2} \rangle = \langle \eta \rangle$ (so that the previous point is either $A_{\langle \eta \rangle}^+$ or $A_{\langle \eta \rangle}^-$, depending on whether $\langle \eta \rangle$ is even or odd) and S^- if $\langle \eta + \frac{1}{2} \rangle > \langle \eta \rangle$ (so that the previous point is either $P_{\langle \eta \rangle}^{++} = P_{\langle \eta \rangle}^{--}$ or $P_{\langle \eta \rangle}^{+-} = P_{\langle \eta \rangle}^{-+}$, depending on whether $\langle \eta \rangle$ is even or odd).

Figure 5: Local structure of the trace of the pull-back to D of the trace of the generatrix $\tilde{\Gamma}$. (a), (b) and (c): structure near $S_1^\pm$ in the three cases $\eta < 1/2$, $\eta = 1/2$ and $\eta > 1/2$. (d): structure of the subsequence b^j, a^j, b^{j+1} . (e)–(h) structure near $S_3^\pm$, in the various possible cases. All pictures also show the mirror images of the portions of the curve.

If we continue moving (m_1, m_3) beyond s^3 along C , $p^{-1}(\tilde{\Gamma}(m_1, m_3))$ passes through the m_1 -mirrored images of the points of the sequence (4.26), but in the reverse order. Hence, its trace runs through the points

$$, \dots, A_2^+ P_2^{+-} \equiv P_2^{-+}, A_1^-, P_1^{-+} \equiv P_1^{+-}, S_1^+.$$

Continuing, m_3 becomes negative while m_1 is unchanged, so $p^{-1}(\tilde{\Gamma}(m_1, m_3))$ passes through the sequence of points

$$P_1^{+-} \equiv P_1^{-+}, A_1^-, P_2^{--} \equiv P_2^{++}, A_2^+, \dots, S_3^\mp$$

and finally through

$$, \dots, A_2^- P_2^{+-} \equiv P_2^{-+}, A_1^+, P_1^{-+} \equiv P_1^{+-}, S_1^-.$$

Note that the generatrix closes itself after passing exactly twice through each of these points.

With some bookkeeping one sees that the curve is built out of a number of simple blocks, which are shown in Figure 5.

The statement of Proposition 4 now follows by taking into account the reflectional symmetry of the generatrix and the fact that it does not have any other self-intersection points. ■

ACKNOWLEDGMENTS. This research has been partially supported by NSERC and by MIUR.

References

- [1] R. Abraham and J. E. Marsden, *Foundations of Mechanics*, Benjamin, Reading, 1978.
- [2] P. Appell, *Traité de Mécanique Rationnelle*, Gauthier–Villars, Paris, 1909.
- [3] V.I. Arnold, *Mathematical Methods of Classical Mechanics*. Graduate Texts in Mathematics **60**, Springer-Verlag, New York, 1989.
- [4] L. Bates and F. Fassò, The conjugate locus for the Euler top. II: Numerical investigation of the general case. In preparation.
- [5] M. Berger, *A panoramic view of Riemannian geometry*, Springer-Verlag, Berlin, 2003.
- [6] A.D. Blaom, A geometric setting for Hamiltonian perturbation theory. *Mem. Amer. Math. Soc.*, **153** (2001), 1-112.
- [7] J.W. Bruce and P.J. Giblin, *Curves and singularities. A geometrical introduction to singularity theory*, Cambridge University Press, Cambridge, 1992.
- [8] I. Chavel, A class of Riemannian homogeneous spaces. *J. Differential Geometry*, **4** (1970), 13–20.
- [9] R.H. Cushman and L. Bates, *Global Aspects of Classical Integrable Systems*, Birkhäuser, Basel, 1997.
- [10] A. Dullweber, B. Leimkuhler and R. McLachlan, Symplectic splitting methods for rigid body molecular dynamics, *Journal Chemical Physics*, **107** (1997), 5840–5851.
- [11] F. Fassò, The Euler-Poinsot top: a non-commutatively integrable system without global action-angle coordinates, *Zeitschrift für Angewandte Mathematics und Physics*, **47** (1996), 953–976.
- [12] F. Fassò, Superintegrable Hamiltonian systems: geometry and perturbations. *Acta Appl. Math.* **87** (2005), 93–121.
- [13] J. Itoh and K. Kiyohara, The Lipschitz continuity of the distance function to the cut locus. *Trans. Amer. Math. Soc.* **353** (2001), 21–40.
- [14] J. Itoh and K. Kiyohara, The cut loci and the conjugate loci on ellipsoids. *Manuscripta Mathematica* **114**, 247–264 (2004).

- [15] C. Jang and P.E. Parker, Conjugate loci of pseudo-Riemannian 2-step nilpotent Lie groups with nondegenerate center. *Ann. Global Anal. Geom.* **28** (2005), 1–18.
- [16] W. Klingenberg, *Riemannian Geometry* (de Gruyter, Berlin, 1981).
- [17] S. Kobayashi, *On conjugate and cut loci*. In: S.S. Chern (ed) *Global Geometry and Analysis*, MAA Studies in Mathematic **27**, 96–122 (Mathematical Association of America, Washington, 1989).
- [18] D.F. Lawden, *Elliptic functions and applications*. Applied Mathematical Sciences **80**. Springer-Verlag, New York, 1989.
- [19] G. Misiolek, Conjugate points in $D_\mu(T^2)$. *Proceeding of the American Mathematical Society* **124**, 977–982 (1996).
- [20] T. Sakai, Cut loci of Berger's spheres. *Hokkaido Math. J.* **10** (1981), 143–155.
- [21] R. Sinclair, Cut and conjugate loci on compact Liouville surfaces. Preprint (2005).
- [22] E.T. Whittaker, *A treatise on the analytical dynamics of particles and rigid bodies. With an introduction to the problem of three bodies*. Reprint of the 1937 edition. Cambridge University Press, Cambridge, 1988.

Received: October 31, 2006