REPORT ON A SOLUTION TO A BIG OPEN PROBLEM IN SUBRIEMANNIAN GEOMETRY

RICHARD MONTGOMERY

1. Introduction

Are all subRiemannian geodesics smooth? I posed this question about 30 years ago after having shown that previous "proofs" asserting "yes" were fatally flawed. See [1] and [2]. The question was answered quite recently with a decisive "no". See [3]. This paper was written by Yacine Chitour, Fréderic Jean, Roberto Monti, Ludovic Rifford, Ludovic Sacchelli, Mario Sigalotti, and Alessandro Socionovo, whom I will be referring to as "the team". The team constructed an analytic subRiemannian structure on $\mathbb{R}^3$ having a geodesic segment which fails to be C^3 at one of its endpoint.

A few months before the team published their resounding 'no' I had advertised this question as 'Open Problem 1" for an article in the Open Problems section of the web site maintained by the Association for Mathematical Research. See [4]. The present article was written to set the record straight: the problem is now solved! I give the team's construction, say some words on their proof of the minimality of their non-smooth geodesic and state an important question which remains open about their example.

2. The construction

We can define a subRiemannian structure on $\mathbb{R}^3$ by specifying two everywhere linearly independent vector fields X and Y on $\mathbb{R}^3$. Declare X and Y to be an orthonormal frame for the 2-plane field D which they span. In this way we get a 2-plane field on $\mathbb{R}^3$ together with an inner product on the field – that is to say, a subRiemannian structure. A curve is called "horizontal" if it is tangent to D. We define the subRiemannian distance between two points to be the infimum of the lengths of the horizontal curves joining the two points. A subRiemannian geodesic is a shortest horizontal curve connecting its endpoints, i.e. a horizontal curve whose length realizes the distance between its endpoints. See [4] for more on these ideas and [6] or [7] for a detailed treatise on subRiemannian geometry.

The team took their vector fields to have the form

$$X = \frac{\partial}{\partial x}, Y = \frac{\partial}{\partial y} + A(x,y)\frac{\partial}{\partial z}. \tag{1}$$

I will call A the "vector potential". The specific A that the team used is recorded below as equation (5). The Lie bracket of our two vector fields is

$$[X,Y] = B(x,y)\frac{\partial}{\partial z} \text{ where } B = \frac{\partial A}{\partial x}. \tag{2}$$

I will call B the "magnetic field". The condition that D be a contact distribution in a neighborhood $U \subset \mathbb{R}^3$ is the condition that the triple of vector fields X, Y and

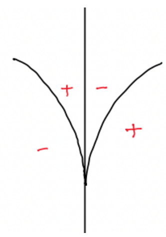

FIGURE 1. The zero locus of the magnetic field in the team's example is the A_k singularity $x^2 = y^{k+1}$ together with the y axis. The team needs $k \geq 4$ and even. The non-smooth geodesic is the horizontal lift of the right half of the cusp, starting from the cusp point, and parameterized by arclength. The $\pm$ signs indicate the sign taken by the magnetic field in the indicated regions.

$[X, Y]$ are linearly independent at every point of U. Thus our D is contact precisely on the open set $B \neq 0$. It is well-known that if a geodesic intersects the contact region then it must satisfy the "normal subRiemannian geodesic equations" stated momentarily. Solutions to these equations are analytic. It follows that the team's non-smooth geodesic must lie within the zero locus of the magnetic field, the locus where the distribution fails to be contact.

Write $c(s) = (x(s), y(s))$ for the planar projection of the space curve $\gamma(s) = (x(s), y(s), z(s))$ and $\kappa_c(s)$ for the signed curvature of the plane curve $c(s)$. The normal subRiemannian geodesic equations for γ are

$$\kappa_c(s) = \lambda B(c(s)) \tag{3}$$

for some constant λ (which arises as a Lagrange multiplier) together with the horizontality condition

$$\dot{z}(s) = A(c(s)). \tag{4}$$

We call these two equations, (3) and (4), the "normal geodesic equations". Being ODEs with analytic coefficients, all solutions to the normal geodesic equations are analytic.

Remark 2.1. *The ODEs* (3) *for* $c(s)$ *are precisely the Lorentz equations for a non-relativistic charged particle with charge* λ *and mass* 1 *travelling on the Euclidean xy plane under the influence of the magnetic field* $B(x, y)\frac{\partial}{\partial z}$ *orthogonal to that plane. For this and other reasons we refer to* B *as a "magnetic field".*

The team chose their "vector potential" $A(x, y)$ in equation (1) to be the square of the function whose zero locus defines the A_k singularity:

$$A(x, y) = (x^2 - y^{k+1})^2 \text{ for } k \geq 4 \text{ an even integer.} \tag{5}$$

The resulting magnetic field

$$B(x, y) = 4x(x^2 - y^{k+1}) \tag{6}$$

has zero locus the two branches $x = 0$ and $x^2 - y^{k+1} = 0$, the latter being the equation of the A_k singularity, or "cusp". This zero locus is depicted in figure 1. The team's putative non-smooth geodesic Γ is half of the A_k cusp, one endpoint being the cusp point which is the origin. Specifically, parametrize the A_k branch by $x = t^{k+1}, y = t^2$. Take the right branch $t \geq 0$. Then the team's main assertion is that the horizontal lift of this right branch of the cusp is a geodesic (in the metric sense) through the origin when $k \geq 4$ is even. Here, the horizontal lift is obtained by setting $z = 0$ since $A(x, y) = 0$ along the cusp. When reparameterized by arc length s the right branch of the cusp is given by $\bar{\omega}(s) = (x(s), y(s))$ were $x = s^{(k+1)/2}[1+O(s)]$ and $y = s[1+O(s)]$ and so its $(k/2)+1$-st derivative blows up as $s^{-1/2}$ when $s \to 0$. Hence the arc-length parameterized curve $\Gamma(s) = (\bar{\omega}(s), 0)$ is not smooth at the origin. When $k = 4$ this curve is C^2 but not C^3 at the origin.

The hard work of the paper is to show that the putative geodesic Γ just described, beginning from the origin which is the cusp point, is actually a geodesic. The team argues by contradiction. Suppose Γ is not a geodesic upon leaving the origin. Then for all $\epsilon > 0$ there must be shorter horizontal curves connecting the origin to $\Gamma(\epsilon)$. For fixed small ϵ take γ_ϵ to be the shortest such curve and write $\omega = \omega_\epsilon$ for its plane projection. Of necessity this shortest curve must leave the zero-locus and hence, being minimal, and travelling into the contact region $B \neq 0$, must satisfy the normal subRiemannian geodesic equations. See figure 2.

At this point the argument becomes a detailed intricate proof, one might even call it a slog, which reminded me of my own old proof in [2]. The team must understand in detail *all solutions* to the normal subRiemannian geodesic equations (3) starting at the origin and satisfying the required endpoint conditions. The curve ω has been set up so that the x and y values of $\gamma = \gamma_\epsilon$ and Γ match up at 0 and $\Gamma(\epsilon)$. The interesting and useful endpoint condition is the z-endpoint condition on ω which we can write as

$$0 = \int_\omega A dy - \int_{\bar{\omega}} A dy = \int\int B dx \wedge dy \tag{7}$$

where the last equality comes from Stoke's equations using $d(Ady) = Bdx \wedge dy$ and that last integral is over the immersed disc bounded by ω and $\bar{\omega}$. This integral equality is repeatedly used in combination with (3) in their analysis. The team rightly refers to their study of properties of all solutions to (3) satisfying the endpoint conditions (7) as a study of the "anatomy of ω". They summarize their anatomical findings as Proposition 2.2 on p. 4 of the article.

First, they observe that the "charge" λ appearing (3) cannot be zero, for otherwise ω_ϵ would be a straight line, and would fail to satisfy the vertical endpoint condition (7). Equation (3) shows that every time the curve ω crosses the zero locus the sign of its curvature must switch. See figure 1. Being analytic, ω can only cross the zero locus a finite number of times. ANATOMICAL FINDING ONE: ω does not cross the zero locus at all except at its endpoints where it lies on the zero locus. ANATOMICAL FINDING TWO: ω must suffer exactly one loop. This loop is required to cancel weighted area (the weight being the magnetic field) accumulated during travel so that the vertical endpoint condition (7) will hold. In addition they show that the charge λ in (3) is negative and ω lies in the region $B > 0$ except at its endpoints. Thus ω lies entirely in the + region in figure 1 and has the shape indicated in figure 2.

The anatomical findings stated on p. 4 within Proposition 2.2 require at least 22 pages to prove. These pages include detailed analysis regarding lengths and B-weighted areas (see (7)) of closed curves cut out by ω and $\bar{\omega}$. Many non-trivial estimates and theorems are invoked along the way. A detailed analysis regarding the asymptotics of the length and weighted area of the loop within ω_ϵ versus the

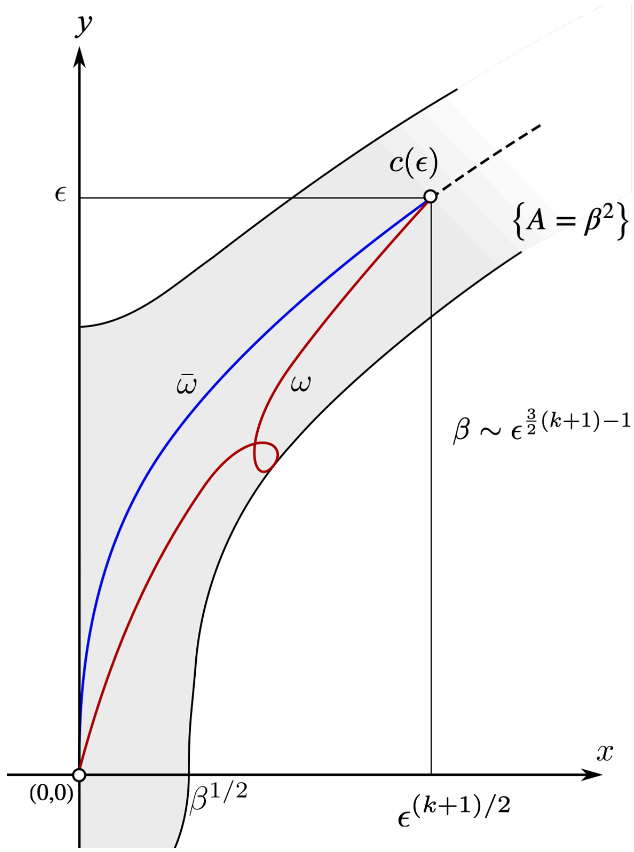

FIGURE 2. The putative geodesic $\bar{\omega}$ drawn in comparison with nearby solutions ω to the normal geodesic equations which share its endpoints. A key part of the proof reduces to analyzing those solutions which lie in a region $A(x, y) \leq \beta^2$ where $\beta \sim \epsilon^{\frac{3}{2}(k+1)-1}$, a region which is a strip of thickness $\beta^{1/2}$ about the zero locus. The figure is modified from one found in [3].

rest of ω_ϵ as $\epsilon \to 0$ yield contradictions which show that a solution $\omega = \omega_\epsilon$ with the required properties cannot exist. These contradictions conclude the proof.

Remark 2.2. *That Γ is a geodesic away from the cusp point at $s = 0$ follows from my own work in* [2].

Remark 2.3. *Team member Socionovo showed in* [5] *that using the A_2 curve, which is to say the vector potential $A(x, y) = (x^2 - y^3)^2$, will not work. What Socionovo showed is that the the lifted half-cusp $\Gamma([0, \epsilon])$ in this case fails to be a minimizing geodesic between its endpoints. His result means that one cannot obtain curves that fail to be C^2 using the team's construction.*

3. But the question remains ...

Can the team's geodesic be extended beyond the cusp and remain geodesic?

The obvious extension of the team's geodesic is to follow the lift of the negative y-axis, which is to say, follow the curve $s \mapsto (0, s, 0)$ for $s \leq 0$, up to the cusp point and then continue along their geodesic $\Gamma(s)$ for $s \geq 0$. In other words, concatenate their lifted axis with their lifted cusp Γ. To date the team has been unable to say whether or not this extension is a geodesic. Call this extension $\tilde{\Gamma}$. What would be required is showing that for sufficiently small $\epsilon > 0$ the curve $\tilde{\Gamma}([-\epsilon, \epsilon])$ beats out all competing normal subRiemannian geodesics sharing its endpoints (which are $(0, -\epsilon, 0)$ and $(\bar{\omega}(\epsilon), 0)$)

Either answer would be interesting. If 'no' then subRiemannian geometries enjoy yet one more property not enjoyed by Riemannian geometries: they can support the existence of inextendible geodesics. If 'yes', well then, we get a non-smooth point inside of a geodesic.

References

[1] R. Montgomery, [1996], *Survey of Singular Geodesics*, in **Subriemannian Geometry**, Progress in Math., vol. 144, 325-339, Birkhauser.

[2] R. Montgomery, [1994], *Abnormal minimizers*, SIAM J. Control Optim. 32 (1994), no. 6, 1605- 1620.

[3] Yacine Chitour, Frédéric Jean, Roberto Monti, Ludovic Rifford, Ludovic Sacchelli, Mario Sigalotti and Alessandro Socionovo, [2025], *Not all sub-Riemannian minimizing geodesics are smooth*, https://arxiv.org/abs/2501.18920

[4] R. Montgomery, [2024], *SubRiemannian Geometry: Two Open Problems and One Falling Cat* in https://amathr.org/subriemanniangeometry/.

[5] A. Socionovo, [2025], *Sharp regularity of sub-Riemannian length-minimizing curves*, in https://arxiv.org/abs/2502.00403

[6] R. Montgomery, [2002], **A tour of sub-Riemannian geometries, their geodesics and applications**, Mathematical Surveys and Monographs, Vol. 91. American Mathematical Society, Providence, RI.

[7] A. Agrachev, D. Barilari and U. Boscain, [2020], **A comprehensive introduction to sub-Riemannian geometry from the Hamiltonian viewpoint** Cambridge Studies in Advanced Mathematics, Vol. 181, Cambridge University Press.

Mathematics Department, University of California, Santa Cruz, Santa Cruz CA 95064

Email address: rmont@ucsc.edu