SHARP REGULARITY OF SUB-RIEMANNIAN LENGTH-MINIMIZING CURVES

Alessandro Socionovo

SHARP REGULARITY OF SUB-RIEMANNIAN LENGTH-MINIMIZING CURVES

A. SOCIONOVO

Abstract. A longstanding open question in sub-Riemannian geometry is the smoothness of (the arc-length parameterization of) length-minimizing curves. In [19], this question is negative answered, with an example of a C^2 but not C^3 length-minimizer of a real-analytic (even polynomial) sub-Riemannian structure. In this paper, we study a class of examples of sub-Riemannian structures that generalizes that presented in [19], and we prove that length-minimizing curves must be at least of class C^2 within these examples. In particular, we prove that Theorem 1.1 in [19] is sharp.

Contents

1. Introduction

One of the most difficult and important open problems in sub-Riemannian geometry is the regularity of geodesics, i.e, of those curves that are locally length-minimizing (here and hereafter, by regularity of a curve we mean the regularity of its arc-length parameterization). The problem has been open since the work of Strichartz in [20, 21].

A sub-Riemannian manifold is a triplet $(M, \mathcal{D}, g)$, where M is a smooth manifold, $\mathcal{D} \subset TM$ is a smooth bracket-generating distribution (i.e., a distribution satisfying the Hörmander condition), and g is a smooth metric on $\mathcal{D}$. A *horizontal* (or, *admissible*) curve is an absolutely continuous trajectory in M which is tangent to $\mathcal{D}$ almost everywhere. The length of horizontal curves is defined in the standard way with respect to the metric g and the distance between two points $p, q \in M$ is the infimum of the length of horizontal curves joining p and q. The Chow-Rashewskii and the Hopf-Rinow theorems ensure the existence of geodesics between every pair of points sufficiently close.

Candidate curves to be length-minimizing are called *extremal curves* and they are individuated by some necessary conditions appearing in the Pontryagin maximum

principle. Extremal curves are divided in two non-disjoint classes, which are called *normal* and *abnormal* (or *singular*) curves. Normal curves, which are the only ones appearing in Riemannian geometry, are smooth and length-minimizing for short times. Instead, abnormal curves are a priori no more regular than being absolutely continuous, as well as their length-minimality properties are unknown. The difficulty of the problem of the regularity of length-minimizing curves lies in the presence of *strictly abnormal* (or *strictly singular*) extremals, i.e., those curves that are abnormal but not normal. We refer the reader to [2, 3, 11, 16, 18] for an exhaustive introduction to sub-Riemannian geometry and to the regularity problem of geodesics.

The first example of a strictly abnormal curve that is length-minimizing, was discovered by Montgomery in his ground-breaking work [15]. After that, many other examples have been found, see for instance [1, 14]. All these examples are smooth curves (by smooth we mean C^∞), indicating the profound complexity of the problem. Over the years, some partial results have been found. On the one hand, it was first proved in [9, 13] that abnormal length-minimizing curves cannot have corners, and later in [10, 17] that they must have a tangent line. In some special cases they are of class C^1, see [4, 6, 12], but in general the C^1 regularity remains an open question. On the other hand, necessary conditions for the minimality of abnormal curves can be derived by the differential analysis of the end-point map (see [2, Chapter 8] for a detailed introduction to the end-point map). This theory is well known up to the second order, see [3, 7], and it has been partially extended to the n-th order in [5]. Finally, an algorithm to produce abnormal curves in Carnot groups of sufficiently high dimension (without any new regularity or minimality result) is given in [8].

A recent and significant progress for the theory is the following result. On $\mathbb{R}^3 = (x_1, x_2, x_3)$, let X_1 and X_2 be the vector fields defined by

$$X_1(x) := \partial_{x_1}, \quad X_2(x) := \partial_{x_2} + P(x)^2 \partial_{x_3}, \tag{1.1}$$

where

$$P(x) = x_1^a - x_2^b, \quad a, b \in \mathbb{N} \setminus \{0\}, \tag{1.2}$$

and let $(\mathbb{R}^3, \mathcal{D}, g)$ be the sub-Riemannian manifold where $\mathcal{D} = \operatorname{span}\{X_1, X_2\}$ and g is the metric making X_1 and X_2 orthonormal. Let also consider, for $\varepsilon > 0$, the curve

$$\gamma_\varepsilon : [0, \varepsilon] \to \mathbb{R}^3, \quad \gamma_\varepsilon(t) = \begin{cases} (t^{\frac{b}{a}}, t, 0), & b \geq a, \\ (t, t^{\frac{a}{b}}, 0), & a \geq b. \end{cases} \tag{1.3}$$

Theorem 1.1. *If $a = 2$ and $b \geq 5$ is odd, then there exists $\varepsilon > 0$ such that the curve γ_ε is length-minimizing in $(\mathbb{R}^3, \mathcal{D}, g)$. Moreover, the curve γ_ε is non-smooth at 0.*

Theorem 1.1 is proved in [19] and provides the first ever examples of non-smooth sub-Riemannian length-minimizing curves, answering a longstanding open question in

sub-Riemannian geometry. The geodesic with the lowest regularity is of class $C^2 \setminus C^3$ and it is obtained for $a = 2$ and $b = 5$. The singularity is at a boundary point.

In this paper, we delve into the reasoning and methodology behind the construction of the examples in (1.1)-(1.2) and we prove that Theorem 1.1 is sharp, in the sense that it is not possible to find a geodesic of class $C^1 \setminus C^2$ within these examples. Our main result is the following.

Theorem 1.2. *If $b > a \geq 3$, or $a = 2$ and $b = 3$, or $a > b \geq 3$, or $b = 2$ and $a = 3$, then for every $\varepsilon > 0$ the curve γ_ε is not length-minimizing.*

The proof of Theorem 1.2 is constructive, as it provides an explicit example of a shorter curve connecting the same points. This kind of construction is, in our opinion, a general interesting tool to understand (as a first analysis) the minimality of abnormal curves, since in these examples they are a close approximation of normal curves, see [19, Proposition 2.2]. On the other hand, the proof makes explicit the analytical role of the parameters a and b, which are in fact the discriminant for both the minimality and the regularity of the curve γ_ε. Instead, the intrinsic and geometric nature of these parameters persists as an enigma, and we believe it is a deep and central question. Finding an abstract argument that unifies all these points –the explicit construction in the proof of Theorem 1.2, approximation of normal curves, and the interplay between analytical and geometric aspects of the parameters a and b– would represent, in our opinion, a significant breakthrough for the theory, which may leads either to find an example of a $C^1 \setminus C^2$ length-minimizing curve, or to prove the C^2 (or C^1) regularity, or to find an example of a non-smooth geodesic with internal singular point. Ultimately, Theorems 1.1 and 1.2 may open new frontiers of research in sub-Riemannian geometry.

We conclude this introductory part with an outline of the paper. In Section 2, we introduce a class of sub-Riemannian structures on $\mathbb{R}^3$ that generalizes those defined in (1.1)-(1.2), see Definition 2.1, and we study their general properties. We also provide some concrete examples and we explain the ideas behind the construction of the examples in (1.1)-(1.2), which structure (in particular with the square of P) is inspired by the Liu-Sussmann example in [14].

In Section 3, we prove Theorem 1.2. The proof relies upon the construction of a competing curve which is shorter than γ_ε. Since every horizontal curve is implicitly defined by its first two coordinates, the construction of the competitor can be reduced to a constrained problem in $\mathbb{R}^2$, see Definition 3.1. The constraint we obtain is, due to the Stokes theorem, of isoperimetric type with a weighted area, see Lemma 3.2. The competing curve (actually, its projection onto the (x_1, x_2)-plane) is then built as follows: we cut γ_ε near the origin with a segment connecting 0 to the point $\gamma_\varepsilon(\rho)$, for some $0 < \rho < \varepsilon$; then we follow the curve γ_ε up to its end-point; finally, we add

the boundary of a square of length $0 < \delta < \varepsilon$ at the end-point of γ_ε. Thanks to Remark 2.4, the sub-Riemannian length of both γ_ε and the competitor is measured by the Euclidean length of their projection on the (x_1, x_2)-plane. Therefore, the cut produces a gain of length $\Delta L(\rho)$, but it modifies the third coordinate. The square is then needed to restore the third coordinate to 0, despite it wastes an amount of length equal to δ. The proof consists in showing that the parameters $\rho, \delta > 0$ can be chosen such that the length gained with the cut is greater than that lost with the boundary of the square, i.e., such that $\Delta L(\rho) > \delta$.

Research funding. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101034255.

2. Sub-Riemannian structures on $\mathbb{R}^3$

In this section, we study the sub-Riemannian structures on $\mathbb{R}^3$ defined in Definition 2.1 and their general properties, we provide some concrete examples, and we explain the ideas behind the construction of the structure defined in (1.1)-(1.2).

Definition 2.1. On $\mathbb{R}^3 = (x_1, x_2, x_3)$ we define the vector fields

$$X_1(x) = \frac{\partial}{\partial x_1}, \quad X_2(x) = \varphi(x)\frac{\partial}{\partial x_2} + \psi(x)\frac{\partial}{\partial x_3},$$

where $\varphi, \psi \in C^\infty(\mathbb{R}^3, \mathbb{R})$ are such that X_1 and X_2 satisfy the Hörmander condition. We also define the distribution $\mathcal{D} := \operatorname{span}\{X_1, X_2\}$ and the sub-Riemannian metric g that makes X_1 and X_2 orthonormal.

Definition 2.2. An absolutely continuous curve $\eta : [0, \tau] \to \mathbb{R}^3$ is *admissible* or *horizontal* in $(\mathbb{R}^3, \mathcal{D})$ if there exists a function $u = (u_1, u_2) \in L^2([0, \tau], \mathbb{R}^2)$, called the *control* of η, such that

$$\dot\eta(t) = u_1 X_1(\eta(t)) + u_2 X_2(\eta(t)), \quad \text{for a.e. } t \in [0, \tau]. \tag{2.1}$$

By (2.1) it follows that a horizontal curve $\eta : [0, \tau] \to \mathbb{R}^3$ and its control u satisfy, for almost every $t \in [0, \tau]$,

$$\dot\eta_1(t) = u_1(t), \quad \dot\eta_2(t) = u_2(t)\varphi(\eta(t)), \quad \dot\eta_3(t) = u_2(t)\psi(\eta(t)).$$

If $\varphi \circ \eta \neq 0$, then we have $u_2(t) = \dfrac{\dot\eta_2(t)}{\varphi(\eta(t))}$, for a.e. $t \in [0, \tau]$, and thus the third coordinate of η is uniquely determined by the first two coordinates of η via the integral constraint

$$\eta_3(t) = \eta_3(0) + \int_0^t \dot\eta_2(s)\frac{\psi(\eta(s))}{\varphi(\eta(s))}ds, \quad \text{for all } t \in [0, \tau]. \tag{2.2}$$

Definition 2.3. Let $\eta : [0,\tau] \to \mathbb{R}^3$ be a horizontal curve in $(\mathbb{R}^3, \mathcal{D}, g)$ and let $x, y \in \mathbb{R}^3$ be two points. We say that:

(i) the sub-Riemannian length of η is

$$L_{\mathrm{sR}}(\eta) := \int_0^\tau \sqrt{g\big(\dot\eta(t), \dot\eta(t)\big)}\, dt;$$

(ii) the sub-Riemannian distance between x and y is

$$d(x,y) := \inf\{L_{\mathrm{sR}}(\widetilde\eta) \mid \widetilde\eta : [0,\tau] \to \mathbb{R}^3 \text{ is horizontal, } \widetilde\eta(0) = x,\ \widetilde\eta(\tau) = y\};$$

(iii) η is length-minimizing if for every horizontal curve $\widetilde\eta : [0,\tau] to \mathbb{R}^3$ such that $\widetilde\eta(0) = \eta(0)$ and $\widetilde\eta(\tau) = \eta(\tau)$, we have $L_{\mathrm{sR}}(\eta) \le L_{\mathrm{sR}}(\widetilde\eta)$.

The fact that the distance in (ii) is well defined (i.e. $d(x,y) < +\infty$ for all $x, y \in \mathbb{R}^3$) is a direct consequence of the Hörmander condition and it is well-known as Chow-Rashewskii theorem, see for instance [2]. Consequently, also the definition of length-minimality in (iii) is well posed.

Remark 2.4. If $\varphi \equiv 1$ every horizontal curve η has control $u = (\dot\eta_1, \dot\eta_2)$. In this case, since g is the metric making X_1, X_2 orthonormal, we have that the sub-Riemannian length of η is the Euclidean length of the plane curve (η_1, η_2).

In the literature, necessary conditions (of the first order) for horizontal curves to be length-minimizing can be obtained in two different ways. The first possible approach is to study the first differential of the end-point map together with the Lagrange multiplier rules. The other way is to use the Pontryagin maximum principle provided by classical control theory. Both the methods are well illustrated in [2, 3].

Proposition 2.5. *Let $\eta : [0,\tau] \to \mathbb{R}^3$ be a length-minimizing curve in $(\mathbb{R}^3, \mathcal{D}, g)$. Then there exists a function of covectors $p : [0,\tau] \to (\mathbb{R}^3)^*$ such that one of the following conditions holds*

$$p(t)\big(X_i(\eta(t))\big) = g\big(\dot\eta(t), X_i(\eta(t))\big), \tag{2.3}$$

$$p(t)\big(X_i(\eta(t))\big) = 0, \quad \text{with } p(t) \neq 0, \tag{2.4}$$

for $i = 1, 2$ and for all $t \in [0,\tau]$.

If $\eta : [0,\tau] \to \mathbb{R}^3$ is horizontal and there exists a function of covectors $p : [0,\tau] \to (\mathbb{R}^3)^*$ such that (2.3) or (2.4) holds, then η is called an *extremal curve*. Extremal curves satisfying (2.3) are called *normal*, while extremal curves satisfying (2.4) are called *abnormal*. The conditions in (2.3)-(2.4) are mutually exclusive only if the function of covectors p is fixed, i.e., an extremal curve can be both normal and abnormal with different covectors. A curve that is only normal (resp., only abnormal) is called *strictly normal* (resp., *strictly abnormal*).

We briefly recall that normal curves are solution to the normal Hamiltonian equations. Here and hereafter, we use the short notation $\varphi_{x_i}, \psi_{x_i}$ to denote the partial derivatives of the functions φ and ψ (see Definition 2.1) with reaspect to x_i, $i = 1, 2, 3$. Defining the normal Hamiltonian of $(\mathbb{R}^3, \mathcal{D}, g)$ as

$$H : \mathbb{R}^3 \times (\mathbb{R}^3)^* \to \mathbb{R}, \quad H(x, p) := p_1^2 + \frac{1}{2} h(x, p)^2,$$

where $h(x, p) := p_2\varphi(x) + p_3\psi(x)$, we have that normal curves with their covectors are solution to the system of equations

$$\begin{cases} \dot{x}_1 = p_1 \\ \dot{x}_2 = h(x,p)\varphi(x) \\ \dot{x}_3 = h(x,p)\psi(x), \end{cases} \qquad \begin{cases} \dot{p}_1 = h(x,p)(p_2\varphi_{x_1}(x) + p_3\psi_{x_1}(x)) \\ \dot{p}_2 = h(x,p)(p_2\varphi_{x_2}(x) + p_3\psi_{x_2}(x)) \\ \dot{p}_3 = h(x,p)(p_2\varphi_{x_3}(x) + p_3\psi_{x_3}(x)), \end{cases} \tag{2.5}$$

We refer the reader to [2, 18] for all these facts, including a proof of (2.5).

We now focus on the study of the abnormal curves in $(\mathbb{R}^3, \mathcal{D})$, aiming to provide a geometric characterization. As a matter of fact, the abnormal condition (2.4) does not depend on the metric. We start by recalling that necessary conditions of the second order for abnormal curves to be length-minimizing, known as Goh conditions, can be obtained by the study of the second-order differential of the end-point map. Moreover, the optimality assumption is unnecessary when $\dim(\mathcal{D}_x) = 2$, for all $x \in \mathbb{R}^3$. Thus, every abnormal curve η in $(\mathbb{R}^3, \mathcal{D})$ with its function of covectors p satisfies

$$p(t)([X_1, X_2](\eta(t))) = 0, \quad \text{for all } t \in [0, \tau], \tag{2.6}$$

where $[\cdot, \cdot]$ denotes the Lie bracket of vector fields. For more details about Goh conditions and for a proof of (2.6) we refer the reader to [2].

Combining (2.4) and (2.6) we deduce the desired geometric characterization for abnormal curves in $(\mathbb{R}^3, \mathcal{D})$.

Lemma 2.6. *A horizontal curve $\eta : [0, \tau] \to \mathbb{R}^3$ is abnormal in $(\mathbb{R}^3, \mathcal{D})$ if and only if it is supported inside the surface*

$$\Sigma_{\mathcal{D}} := \{x \in \mathbb{R}^3 \mid \varphi(x)\psi_{x_1}(x) - \psi(x)\varphi_{x_1}(x) = 0\}. \tag{2.7}$$

Proof. Let $\eta : [0, \tau] \to \mathbb{R}^3$ be an abnormal curve and let $p = (p_1, p_2, p_3) : [0, \tau] \to (\mathbb{R}^3)^*$ be its covector. By (2.4) with $i = 1$ we get $p_1 = 0$. Then by (2.4) with $i = 2$ and by (2.6) we have that η satisfies, for all $t \in [0, \tau]$,

$$\begin{cases} p_2(t)\varphi(\eta(t)) + p_3(t)\psi(\eta(t)) = 0, \\ p_2(t)\varphi_{x_1}(\eta(t)) + p_3(t)\psi_{x_1}(\eta(t)) = 0. \end{cases} \tag{2.8}$$

Since $(p_2(t), p_3(t)) \neq 0$, we have that for all $t \in [0, \tau]$ the determinant of the matrix of the system (2.8) is forced to be 0, that is

$$(\varphi\psi_{x_1} - \psi\varphi_{x_1})(\eta(t)) = 0. \tag{2.9}$$

Conversely, if η is supported in $\Sigma_{\mathcal{D}}$ then it holds (2.9). Thus for every $t \in [0, \tau]$ there exists $(p_2(t), p_3(t)) \neq 0$ such that (2.8) holds. Consequently, the covector $p : [0, \tau] \ni t \mapsto (0, p_2(t), p_3(t)) \in (\mathbb{R}^3)^*$ ensures that η is abnormal. □

Definition 2.7. The set $\Sigma_{\mathcal{D}}$ defined in (2.7) is the *Martinet surface* of $\mathcal{D}$.

We conclude this section providing some examples of distributions of the type appearing in Definition 2.1, which vary based on the functions φ and ψ. The aim is to study Martinet surfaces and abnormal curves within these examples. In particular, we mainly focus on the case where φ, ψ are chosen as in (1.1)-(1.2), and we explain through an heuristic argument why the curve γ_ε defined in (1.3) emerges as a natural candidate to be a non-smooth geodesic for every $b > a \geq 2$ and for small $\varepsilon > 0$.

The Heisenberg group. This represents the simplest and most famous example of a sub-Riemannian manifold. It is obtained by setting $\varphi = 1$ and $\psi = x_1$. In this example, the Martinet surface is empty, so there are no abnormal curves and it is not of our interest. The Heisenberg group has been extensively studied in the literature, and interested readers are directed for example to [16].

The Martinet example. This example involves a slight modification of the Heisenberg group by choosing $\varphi = 1$ and $\psi = x_1^2$. Consequently, the Martinet surface is no longer empty and is given by the plane $x_1 = 0$. Thus, for $\varepsilon > 0$, the curve

$$\alpha_\varepsilon : [0, \varepsilon] \to \mathbb{R}^3, \quad \alpha(t) = (0, t, 0), \tag{2.10}$$

is abnormal. Since α_ε is a segment in $\mathbb{R}^2$, by Remark 2.4 it is optimal in $(\mathbb{R}^3, \mathcal{D}, g)$ for all $\varepsilon > 0$, providing an example of an abnormal length-minimizing curve. However, the curve α_ε is not strictly abnormal since it is also normal in $(\mathbb{R}^3, \mathcal{D}, g)$ with the constant covector $p = (0, 1, 0)$.

The Liu-Sussmann example. This example, credited to Liu and Sussmann (see [14]), consists of a slight modification of the Martinet example to make the curve α in (2.10) strictly abnormal. The choice of φ and ψ here is $\varphi = 1 - x_1$ and $\psi = x_1^2$. Consequently, the defining function of $\Sigma_{\mathcal{D}}$ in (2.7) becomes

$$2x_1(1 - x_1) + x_1^2 = x_1(2 - x_1),$$

implying that the Martinet surface is the union of the two planes $x_1 = 0$ and $x_1 = 2$.

The curve α_ε defined in (2.10) remains abnormal and, since φ is no longer equal to 1, the sub-Riemannian length of horizontal curves here is no longer the Euclidean length of their projection onto $\mathbb{R}^2$. Therefore, the length-minimality of α_ε is no longer

obvious in this example. In [14] (see also [18, Section 2.5]) it is proved both that α_ε is strictly abnormal and that it is length-minimizing for small $\varepsilon > 0$.

Remark 2.8. The proof of Liu and Sussmann relies on the fact that the curve α_ε is C^1-isolated. Roughly speaking, a horizontal curve $\eta : [0, \tau] \to \mathbb{R}^3$ is said to be C^1-isolated if any other horizontal curve, parameterized on the same interval $[0, \tau]$ and having control close to that of η in the L^∞ topology, cannot have the same end-points of η. Here the curve α_ε is C^1-isolated, and thus length-minimizing, thanks to the squared power appearing on $\psi = x_1^2$, but, in general, the property of being C^1-isolated is not enough to be a minimizer, as explained at the end of this section.

The curve α_ε remains strictly abnormal by choosing $\psi = x_1^k$ for every integer $k \geq 2$. However, it remains C^1-isolated only for even k and, in fact, the proof of Liu and Sussmann showing the minimality of α_ε can be repeated only if k is even. Instead, for odd k the curve α_ε is not of minimal length, see [5, Theorem 10.1].

The examples defined in (1.1)**-**(1.2)**.** Our examples arise as a different modification of the Martinet example than the one as in the Liu-Sussmann case. The aim is to produce a singularity on the Martinet surface, then generating non-smooth and possibly lenght-minimizing abnormal curves.

We choose $\varphi = 1$ and $\psi = P(x)^2$, where $P(x) = (x_1^a - x_2^b)^2$ and $a, b \in \mathbb{N} \setminus \{0\}$. The choice of $\varphi = 1$ ensures, according to (2.7), that possible singularities on $\Sigma_{\mathcal{D}}$ only depend on ψ. With respect to the Liu-Sussmann example, we also recover the property that the sub-Riemannian length of horizontal curves is the Euclidean length of their projections onto $\mathbb{R}^2$, see Remark 2.4.

The choice of $\psi = P^2$ is then thought of as a perturbation of the square $(x_1^a)^2$, generating the desired singularity. The squared power on P aims to maintain a similar analytic structure as in the Liu-Sussmann example and its importance is better explained in Remarks 2.9 and 2.11 below.

By (2.7), the Martinet surface is given by the zero locus of the derivative of P^2 with respect to x_1. We define the abnormal polynomial

$$Q(x_1, x_2) := \partial_{x_1}\big(P(x_1, x_2)^2\big) = 2x_1^{a-1}P(x_1, x_2), \tag{2.11}$$

and so the Martinet surface is

$$\Sigma_{\mathcal{D}} = \{x \in \mathbb{R}^3 \mid Q(x_1, x_2) = 0\}.$$

If $a = 1$ we have $\{x \in \mathbb{R}^3 \mid Q = 0\} = \{x \in \mathbb{R}^3 \mid P = 0\}$, while for $a \geq 2$ we have

$$\Sigma_{\mathcal{D}} = \{x \in \mathbb{R}^3 \mid x_1 = 0\} \cup \{x \in \mathbb{R}^3 \mid P(x_1, x_2) = 0\}. \tag{2.12}$$

Remark 2.9. Thanks to the squared power on P, the zero locus of P is part of the Martinet surface for every $a \geq 1$. Therefore, the curve γ_ε defined in (1.3) is abnormal

in $(\mathbb{R}^3, \mathcal{D})$, for every $\varepsilon > 0$ and for every a and b. Moreover, for $a \neq b$ the curve γ_ε is strictly abnormal in $(\mathbb{R}^3, \mathcal{D}, g)$, as showed in the next lemma. Finally, when a and b are not multiples, γ_ε is not smooth at 0.

Lemma 2.10. *Let $a, b \in \mathbb{N} \setminus \{0\}$ with $a \neq b$. Then every reparameterization of the curve γ_ε defined in (1.3) is not normal in $(\mathbb{R}^3, \mathcal{D}, g)$, for every $\varepsilon > 0$.*

Proof. Assume by contradiction that there exists a reparameterization $\bar{\gamma} : [0, 1] \to \mathbb{R}^3$ of γ_ε and a function of covectors $p : [0, 1] \to (\mathbb{R}^3)^*$ such that $\bar{\gamma}, p$ are solution of (2.5) with $\bar{\gamma} = x$. Since $P(\bar{\gamma}(t)) = 0$ for all $t \in [0, 1]$, the equations in (2.5) read

$$\begin{cases} \dot{\bar{\gamma}}_1 = p_1 \\ \dot{\bar{\gamma}}_2 = p_2 \\ \dot{\bar{\gamma}}_3 = 0, \end{cases} \qquad \begin{cases} \dot{p}_1 = 0 \\ \dot{p}_2 = 0 \\ \dot{p}_3 = 0. \end{cases}$$

Then both $p_1 = p_1(0)$ and $p_2 = p_2(0)$ are constants. Moreover, we have $\bar{\gamma}_1 = p_1(0)t$ and $\bar{\gamma}_2 = p_2(0)t$. Plugging these identities into $P(\bar{\gamma}) = 0$ we get

$$p_1(0)^a t^a - p_2(0)^b t^b = 0, \quad \text{for all } t \in [0, 1],$$

which is not possible for $a \neq b$. □

The case of $a = 1$ is not interesting since the arc-length reparameterization of γ is smooth at 0 for all $b \geq 1$. Therefore, our focus lies on cases where $a \geq 2$.

For $a \geq 2$, the Martinet surface is defined by (2.12). Therefore, the curve α_ε defined in (2.10) is abnormal, for every $\varepsilon > 0$, as in the Martinet and in the Liu-Sussmann examples. In particular, here the curve α_ε behaves as in the Martinet example, being both a normal curve with the covector $p = (0, 1, 0)$ and optimal by Remark 2.4.

Remark 2.11. We explain the second important role played by the squared power on P. On the one hand, for $b > a \geq 2$ and for $\varepsilon > 0$ small, the curve γ_ε and its control are close to the curve α_ε with its control in the Euclidean sense, as the two curves are tangent at the origin (when $a > b \geq 2$ the two curves are instead orthogonal at 0). On the other hand, we have an analytic structure similar to the Liu-Sussmann example, ensuring that γ_ε is C^1-isolated, see Remark 2.8. Since the Euclidean metric in $\mathbb{R}^2$ measures the length of horizontal curves (see Remark 2.4), the curve γ_ε is a good candidate to be a length minimizer.

The proof of Theorem 1.2 that we give in the next section shows that the parameters a and b partially destroy the heuristic intuition behind the construction of these examples. Their role seems to be much deeper than the analytic one they carry out in the computations. Also, Theorem 1.2 shows that the property of being C^1-isolated is not enough to deduce the length-minimality.

3. Proof of Theorem 1.2

The proof relies upon the construction of a shorter competitor. We consider the sub-Riemannian manifold $(\mathbb{R}^3, \mathcal{D}, g)$ with the strictly abnormal curve γ_ε, $\varepsilon > 0$, defined in (1.1), (1.2), and (1.3).

From now on, $\varepsilon > 0$ is fixed. We also introduce the useful notation

$$q := \frac{\max\{a, b\}}{\min\{a, b\}}.$$

We prove the theorem for $b > a$, thus we need to prove that γ_ε is not of minimal length for $b > a \geq 3$, or for $a = 2$ and $b = 3$, and for every $\varepsilon > 0$. Here we have $q = \frac{b}{a}$. Then, in Remark 3.6, we explain how to adapt the proof to the case $a > b$ (the two cases are essentially identical).

The starting point to prove Theorem 1.2 is to substitute competing curves for γ_ε, i.e., horizontal curves having the same end-points of γ_ε, with plane curves satisfying a certain constraint. The plane curve is the projection onto $\mathbb{R}^2$ of the competitor and the constraint to which it is subjected is the one in (2.2) defining the third coordinate. Thanks to Stokes theorem, this constraint is of isoperimetric type with some weighted area.

Definition 3.1. For an absolutely continuous curve $\omega : [0, \tau] \to \mathbb{R}^2$, we define the integral function $I_\omega : [0, \tau] \to \mathbb{R}$,

$$I_\omega(t) := \int_0^t \dot{\omega}_2(s) P(\omega(s))^2 ds. \tag{3.1}$$

We say that ω is a *competing curve for* γ_ε, or more simply a *competitor*, if $\omega(0) = \pi(\gamma(0)) = 0$, $\omega(\tau) = \pi(\gamma(\varepsilon)) = (\varepsilon^q, \varepsilon)$, and $I_\omega(\tau) = 0$.

In the following, we denote by $\mathcal{C}_\varepsilon([0, \tau], \mathbb{R}^2)$ the set of competitors parameterized in the interval $[0, \tau]$. For a competitor $\omega \in \mathcal{C}_\varepsilon([0, \tau], \mathbb{R}^2)$, we call the curve $(\omega_1, \omega_2, I_\omega) : [0, \tau] \to \mathbb{R}^3$ the horizontal lift of ω. Also, with abuse of notation, we denote by γ_ε the projection on $\mathbb{R}^2$ of γ_ε itself, according to the fact that γ_3 is constantly 0.

We next observe a consequence of Stokes theorem, providing that for all $\omega \in \mathcal{C}_\varepsilon([0, \tau], \mathbb{R}^2)$ the condition $I_\omega(\tau) = 0$ in Definition 3.1 is a constraint of isoperimetric nature, involving some weighted area.

Let $\vartheta_i \in AC([0, \tau_i], \mathbb{R}^2)$, $\tau_i > 0$, $i = 1, 2$, be two plane curves. If $\vartheta_1(\tau_1) = \vartheta_2(0)$, we say that ϑ_1 and ϑ_2 are *concatenable*, and we define their concatenation as

$$(\vartheta_1 * \vartheta_2)(t) := \begin{cases} \vartheta(t), & t \in [0, \tau_1], \\ \vartheta_2(t - \tau_1), & t \in [\tau_1, \tau_1 + \tau_2]. \end{cases}$$

We also define the *inverse parameterization* of ϑ_1 as

$$-\vartheta_1(t) := \vartheta_1(\tau_1 - t).$$

If $\omega \in \mathcal{C}_\varepsilon([0,\tau],\mathbb{R}^2)$, we denote by $\Omega = \Omega(\omega,\varepsilon) \subset \mathbb{R}^2$ the bounded open subset of $\mathbb{R}^2$ enclosed by $\omega * (-\gamma_\varepsilon)$. For each connected component $U \subset \Omega$, we denote by $\mathrm{ind}(U) \in \mathbb{Z}$ its winding number with respect to the curve $\omega * (-\gamma_\varepsilon)$. Then we define the weighted area of U as

$$\mathscr{L}_Q(U) := \mathrm{ind}(U) \iint_U Q(x_1,x_2)dx_1dx_2,$$

where Q is the abnormal polynomial in (2.11).

Lemma 3.2. *For any $\omega \in \mathcal{C}_\varepsilon([0,\tau],\mathbb{R}^2)$ we have*

$$\sum_U \mathscr{L}_Q(U) = 0, \tag{3.2}$$

where the sum is over all connected components $U \subset \Omega$.

Proof. Since $\omega \in \mathcal{C}_\varepsilon([0,\tau],\mathbb{R}^2)$ we have $I_\omega(\tau) = 0$ (see Definition 3.1). By (3.1) this condition reads

$$0 = \int_0^\tau \dot\omega_2(t)P(\omega(t))^2dt = \int_\omega P(x_1,x_2)^2dx_2.$$

Since $P(\gamma_\varepsilon)$ is identically 0, from the last equation we deduce

$$0 = \int_\omega P(x_1,x_2)^2dx_2 + \int_{-\gamma_\varepsilon} P(x_1,x_2)^2dx_2 = \int_{\omega*(-\gamma_\varepsilon)} P(x_1,x_2)^2dx_2.$$

The statement then follows by Stokes Theorem. □

We are now ready to construct the competitor and to prove that it is shorter than γ_ε if $b > a \geq 3$, or if $a = 2$ and $b = 3$ (we recall that we are proving the theorem when $b > a$ and that we explain how to adapt the proof to the case $a > b$ in Remark 3.6).

The competitor is built in the following way:

i) we cut γ_ε with a segment near the origin. For $0 < \rho < \varepsilon$, we define the cutting segment $\kappa_\rho(t) := (t\rho^{q-1}, t)$, $t \in [0,\rho]$, where $q = \frac{b}{a}$ and the parameter ρ will be chosen later;

ii) we continue from the point (ρ^q,ρ) by following γ_ε until its end point, i.e., we run across the curve $\gamma_\varepsilon|_{[\rho,\varepsilon]}$;

iii) finally, we add a small square at the end-point and we run across its boundary, namely, for $0 < \delta < \varepsilon$, we let $E_\delta := (\varepsilon^q,\varepsilon) + [0,\varepsilon^q\delta] \times [-\varepsilon\delta, 0]$ and we let

$\sigma = \sigma_\delta : [0, 2(\varepsilon + \varepsilon^q)\delta] \to \mathbb{R}^2$ to be the parameterization of ∂E_δ given by

$$\sigma_\delta(t) = (\varepsilon^q, \varepsilon) + \begin{cases} \sigma_\delta(t) = (t, 0), & t \in [0, \varepsilon^q\delta], \\ \sigma_\delta(t) = (\varepsilon^q\delta, \varepsilon^q\delta - t), & t \in [\varepsilon^q\delta, (\varepsilon + \varepsilon^q)\delta], \\ \sigma_\delta(t) = ((\varepsilon + 2\varepsilon^q)\delta - t, -\delta\varepsilon), & t \in [(\varepsilon + \varepsilon^q)\delta, (\varepsilon + 2\varepsilon^q)\delta], \\ \sigma_\delta(t) = (0, t - 2(\varepsilon + \varepsilon^q)\delta), & t \in [(\varepsilon + 2\varepsilon^q)\delta, 2(\varepsilon + \varepsilon^q)\delta]. \end{cases}$$

Definition 3.3. For $0 < \rho, \delta < \varepsilon$ we define the family of absolutely continuous plane curves $\omega_{\rho,\delta}$ as the concatenation

$$\omega_{\rho,\delta} := \kappa_\rho * \gamma|_{[\rho,\varepsilon]} * \sigma_\delta. \tag{3.3}$$

The curves defined in (3.3) are not competitors a priori: the cut made with the curve κ_ρ modifies the third coordinate of the horizontal lift of $\omega_{\rho,\delta}$, then we add the square to correct this error, restoring the end-point of the third coordinate to 0.

Remark 3.4. The region enclosed by the curve $\omega_{\rho,\delta} * (-\gamma_\varepsilon)$ has, according to Lemma 3.2, two connected components, which are the square E_δ and the region enclosed by $\kappa_\rho * (-\gamma_\rho)$, which we call U_ρ. The winding numbers of U_ρ and E_δ have absolute value equal to 1, and we also have $Q|_{U_\rho}, Q|_{E_\delta} \geq 0$. Then the parameterization σ_δ of ∂E_δ has been chosen in such a way that $\mathrm{ind}(E_\delta) = -\mathrm{ind}(U_\rho)$, so that (3.2) can be satisfied.

We have to chose suitable $\rho = \rho(\varepsilon) < \varepsilon$ and $\delta = \delta(\varepsilon) < \varepsilon$ (if it is possible) such that the curve $\omega_{\rho,\delta}$ both satisfies $I_{\omega_{\rho,\delta}}(\tau) = 0$ and is shorter than γ_ε.

In the following, we omit the dependence on ε in ρ, δ, and we omit the dependence on ρ, δ in $\omega, \kappa, \sigma, E, U$.

Lemma 3.5. *For every $\varepsilon > 0$ and $\rho, \delta \in (0, \varepsilon)$ small enough there exist $C_1 = C_1(a, b, \varepsilon) > 0$ and $0 < C_2 = C_2(a, b, \varepsilon, \delta) \leq 1/2$ such that the curve $\omega_{\rho,\delta}$ is a competitor as soon as*

$$\rho^{2b+1} = C_1\delta^3(1 + C_2). \tag{3.4}$$

Proof. We have to compute and then compare the effects made by the cut and by the square on the third coordinate, as $\rho, \delta \to 0$.

For the cut we have

$$\begin{aligned} \Delta_3^{\mathrm{cut}}(\rho) &:= \int_0^\rho P(\kappa(t))^2 \dot{\kappa}_2(t) dt = \int_0^\rho (t^a \rho^{b-a} - t^b)^2 dt \\ &\overset{(t=\rho s)}{=} \rho^{2b+1} \int_0^1 (s^a - s^b)^2 ds. \end{aligned} \tag{3.5}$$

For the square we use the Stokes theorem, see Lemma 3.2. We compute

$$\begin{aligned}\Delta_3^{\mathrm{cor}}(\delta) &:= \iint_{E_\delta} Q(x_1,x_2)dx_1dx_2\\ &= \int_{\varepsilon(1-\delta)}^{\varepsilon}\int_{\varepsilon^q}^{\varepsilon^q(1+\delta)} 2ax_1^{a-1}(x_1^a-x_2^b)dx_1dx_2 \qquad (x_1=\varepsilon^q x,\ x_2=\varepsilon y)\\ &= 2a\varepsilon^{2b+1}\left(\int_{1-\delta}^{1}\int_{1}^{1+\delta} x^{a-1}(x^a-y^b)dxdy\right)\\ &= a(b+a)\varepsilon^{2b+1}\delta^3(1+o(1)),\end{aligned} \tag{3.6}$$

where in the last line we omitted some basic integral calculations and Taylor developments as $\delta\to 0$. Here we need δ to be small.

By Remark 3.4, the effects on the third coordinate produced by the cut and by the square have opposite sign. Therefore the proof of (3.4) follows by (3.5) and (3.6). Since by (3.6) $\delta>0$ is small, then by (3.4) also $\rho>0$ is small a posteriori. □

We are now ready to prove Theorem 1.2. We denote by $L(\cdot)$ the Euclidean length of plane curves: by Remark 2.4 the Euclidean lengths of ω and γ_ε are the sub-Riemannian lengths of their horizontal lift. The aim of the proof is to show that it is possible to choose small $\rho,\delta>0$ such that both (3.4) and $L(\omega)<L(\gamma_\varepsilon)$ are satisfied.

Proof of Theorem 1.2. Let $\rho,\delta>0$ be small enough and such that (3.4) is satisfied. We fix these parameters later. Along this proof, $C=C(a,b,\varepsilon)>0$ is a constant that may change from line to line. We also recall the notation $q=\frac{b}{a}$.

We start by computing the gain of length obtained with the cut, as $\rho\to 0$. On the one hand, we have $\dot\gamma(t)=(qt^{q-1},1)$ and so

$$\begin{aligned}L(\gamma_\rho) &= \int_0^\rho \sqrt{1+q^2t^{2q-2}}dt = \int_0^\rho\left(1+\frac{1}{2}q^2t^{2q-2}(1+o(1))\right)dt\\ &= \rho+\frac{q^2}{2(2q-1)}\rho^{2q-1}(1+o(1)).\end{aligned} \tag{3.7}$$

On the other hand, we have

$$L(\kappa)=\sqrt{\rho^2+\rho^{2q}}=\rho\left(1+\frac{1}{2}\rho^{2q-2}(1+o(1))\right). \tag{3.8}$$

Then, for small $\rho>0$, we get

$$\Delta L(\rho) := L(\gamma_\rho)-L(\kappa)\geq\frac{1}{4}\left(\frac{q^2}{2q-1}-1\right)\rho^{2q-1}. \tag{3.9}$$

Since $L(\partial E)=L(\sigma)=2\delta(\varepsilon+\varepsilon^q)$, then by (3.9) the competitor ω is shorter than γ_ε as soon as

$$2\delta(\varepsilon+\varepsilon^q)\leq\frac{1}{4}\left(\frac{q^2}{2q-1}-1\right)\rho^{2q-1}. \tag{3.10}$$

Moreover, by Lemma 3.5 we have

$$\delta \geq C\rho^{\frac{2b+1}{3}}. \tag{3.11}$$

By inserting (3.11) into (3.10) we obtain the following sufficient condition on ρ for the competitor ω to be shorter than γ_ε:

$$\rho^{\frac{2b+1}{3}} \leq C\rho^{2q-1}. \tag{3.12}$$

The latter equation is satisfied, when $\rho > 0$ is small enough, if and only if

$$\frac{2b+1}{3} > 2q - 1,$$

which in fact reads

$$b + 2 > 3\,\frac{b}{a}. \tag{3.13}$$

If the latter equation is satisfied, then we can chose small $\rho, \delta > 0$ such that γ_ε is not length-minimizing. Indeed, it is enough to choose $\rho > 0$ small enough such that (3.12) is satisfied. This choice implicitly fixes $\delta > 0$ by (3.4) and ensures $L(\omega) < L(\gamma_\varepsilon)$.

It is very easy to see that (3.13) is satisfied if and only if $a \geq 3$, or $a = 2$ and $b = 3$. Since $\varepsilon > 0$ was arbitrary, the proof is complete. □

Remark 3.6. We explain how to adapt the proof of Theorem 1.2 in the case $a > b$, exploiting the symmetry within the computations between a and b. If $a > b$ then $q = \frac{a}{b}$, and the competing curve ω is essentially the same: we cut from 0 to (ρ, ρ^q) and we add the square $E_\delta = (\varepsilon, \varepsilon^q) + [-\varepsilon\delta, 0] \times [0, \varepsilon^q\delta]$. Again, we orient ∂E_δ such that $\mathrm{ind}(E_\delta) = -\mathrm{ind}(U_\rho)$ (see Remark 3.4), and $\rho, \delta > 0$ are small parameters to be fixed at the end.

With the same computations as in (3.5)-(3.6) we get

$$\Delta_3^{\mathrm{cut}}(\rho) = \rho^{2a+1}\int_0^1 (s^a - s^b)^2 ds \quad \text{and} \quad \Delta_3^{\mathrm{cor}}(\delta) = a(b+a)\varepsilon^{2a+q}\delta^3(1 + o(1)),$$

where, as before, Δ_3^{cut} is the error produced by the cut and Δ_3^{cor} is the error produced by the square, both in absolute value. The power of ρ in $\Delta_3(\rho)$ replaces b with a (as expected due to the symmetry in a, b), while the slight modification in the power of ε into $\Delta_3(\delta)$ is irrelevant ($\varepsilon > 0$ is fixed). Condition (3.4) then implies that ω is a competitor as soon as

$$\rho^{2a+1} = C_1\delta^3(1 + C_2),$$

for some $C_1 = C_1(a, b, \varepsilon) > 0$ and $0 < C_2 = C_2(a, b, \varepsilon, \delta) < 1/2$.

The computations for the gain of length in (3.7)-(3.8) are identical, then (3.9) remains the same.

Finally, with similar estimates as in (3.10)-(3.11)-(3.12), we deduce that ω is shorter than γ_ε as soon as

$$a + 2 > 3\frac{a}{b}, \tag{3.14}$$

for a suitable choice of $\rho > 0$ small enough. Once a small $\rho > 0$ is fixed, then also the small $\delta > 0$ is automatically determined. The conclusion is symmetric in a, b with respect to the one obtained in the proof of Theorem 1.2.

Remark 3.7. The statement of Theorem 1.2 is stable with respect to arbitrary increases in the power of the polynomial P defining ψ. That is, the statement remains true for every choice of ψ of the form $\psi = (x_1^a - x_2^b)^c$, for every natural numbers $a, b, c \geq 2$. Indeed, by repeating the same computations as in the proof Theorem 1.2 and in Remark 3.6, we obtain, in place of (3.13)-(3.14), that ω is shorter than γ_ε as soon as

$$\frac{c}{2}\max\{a, b\} + 2 > 3q.$$

Since $c \geq 2$ the latter equation is implied by (3.13) when $b > a$, and by (3.14) if $a > b$.

References

[1] A. A. Agrachev and A. V. Sarychev. Abnormal sub-Riemannian geodesics: Morse index and rigidity. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 13(6):635–690, 1996.

[2] Andrei Agrachev, Davide Barilari, and Ugo Boscain. *A comprehensive introduction to sub-Riemannian geometry*, volume 181 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2020. With an appendix by Igor Zelenko.

[3] Andrei A. Agrachev and Yuri L. Sachkov. *Control theory from the geometric viewpoint*, volume 87 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 2004. Control Theory and Optimization, II.

[4] D. Barilari, Y. Chitour, F. Jean, D. Prandi, and M. Sigalotti. On the regularity of abnormal minimizers for rank 2 sub-Riemannian structures. *J. Math. Pures Appl. (9)*, 133:118–138, 2020.

[5] Francesco Boarotto, Roberto Monti, and Alessandro Socionovo. Higher order Goh conditions for singular extremals of corank 1. *Arch. Rational Mech. Anal.*, 248(23), 2024.

[6] A. Belotto da Silva, A. Figalli, A. Parusiński, and L. Rifford. Strong Sard conjecture and regularity of singular minimizing geodesics for analytic sub-Riemannian structures in dimension 3. *Invent. Math.*, 229(1):395–448, 2022.

[7] B. S. Goh. Necessary conditions for singular extremals involving multiple control variables. *SIAM J. Control*, 4:716–731, 1966.

[8] Eero Hakavuori. ODE trajectories as abnormal curves in Carnot groups. *J. Differential Equations*, 300:458–486, 2021.

[9] Eero Hakavuori and Enrico Le Donne. Non-minimality of corners in subriemannian geometry. *Invent. Math.*, 206(3):693–704, 2016.

[10] Eero Hakavuori and Enrico Le Donne. Blowups and blowdowns of geodesics in Carnot groups. *J. Differential Geom.*, 123(2):267–310, 2023.

[11] Frédéric Jean. *Control of nonholonomic systems: from sub-Riemannian geometry to motion planning*. SpringerBriefs in Mathematics. Springer, Cham, 2014.

[12] Enrico Le Donne, Nicola Paddeu, and Alessandro Socionovo. Metabelian distributions and C^1 regularity of sub-riemannian geodesics. *Preprint arXiv*, id. arXiv:2405.14997, 2024.

[13] Gian Paolo Leonardi and Roberto Monti. End-point equations and regularity of sub-Riemannian geodesics. *Geom. Funct. Anal.*, 18(2):552–582, 2008.

[14] Wensheng Liu and Héctor J. Sussman. Shortest paths for sub-Riemannian metrics on rank-two distributions. *Mem. Amer. Math. Soc.*, 118(564):x+104, 1995.

[15] Richard Montgomery. Abnormal minimizers. *SIAM J. Control Optim.*, 32(6):1605–1620, 1994.

[16] Richard Montgomery. *A tour of subriemannian geometries, their geodesics and applications*, volume 91 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2002.

[17] Roberto Monti, Alessandro Pigati, and Davide Vittone. Existence of tangent lines to Carnot-Carathéodory geodesics. *Calc. Var. Partial Differential Equations*, 57(3):Paper No. 75, 18, 2018.

[18] Ludovic Rifford. *Sub-Riemannian geometry and optimal transport*. SpringerBriefs in Mathematics. Springer, Cham, 2014.

[19] Alessandro Socionovo, Yacine Chitour, Frédéric Jean, Roberto Monti, Ludovic Rifford, Ludovic Sacchelli, and Mario Sigalotti. Not all sub-Riemannian minimizing geodesics are smooth. *Preprint HAL*, id. hal-04885315, 2025.

[20] Robert S. Strichartz. Sub-Riemannian geometry. *J. Differential Geom.*, 24(2):221–263, 1986.

[21] Robert S. Strichartz. Corrections to: Sub-Riemannian geometry. *J. Differential Geom.*, 30(2):595–596, 1989.