A FAMILY OF NONMINIMIZING ABNORMAL CURVES

ROBERTO MONTI

Abstract. In a four dimensional sub-Riemannian structure, we study a specific family of abnormal extremals and we show that they are not length minimizing, answering in the negative to a question that was recently asked. We extend the result to a class of 4-dimensional sub-Riemannian manifolds of step 5.

1. Introduction

One of the main open problems in sub-Riemannian geometry is the regularity of length minimizing curves. Minimizers may be abnormal extremals, in the language of Geometric Control Theory, and R. Montgomery showed in [5] an example of abnormal extremal that is length minimizing. Liu and Sussmann proved later that the existence of length minimizing abnormal extremals is typical of rank 2 distributions [3].

All known examples of length minimizing curves are smooth. On the other hand, there is no regularity theory of a general character for sub-Riemannian "geodesics", apart from the partial results of [2] and [6].

During the meeting *Geometric control and sub-Riemannian geometry* held in Cortona in May 2012, A. Agrachev and J. P. Gauthier suggested the following situation, in order to find a *nonsmooth* length-minimizing curve.

Consider in $\mathbb{R}^4$ the vector fields

$$X_1 = \frac{\partial}{\partial x_1} + 2x_2\frac{\partial}{\partial x_3} + x_3^2\frac{\partial}{\partial x_4}, \quad X_2 = \frac{\partial}{\partial x_2} - 2x_1\frac{\partial}{\partial x_3}, \tag{1.1}$$

and denote by $\mathcal{D}$ the distribution of 2-planes in $\mathbb{R}^4$ spanned pointwise by X_1 and X_2:

$$\mathcal{D}(x) = \mathrm{span}\{X_1(x), X_2(x)\}, \quad x \in \mathbb{R}^4. \tag{1.2}$$

A Lipschitz curve $\gamma : [0,1] \to \mathbb{R}^4$ is horizontal if $\dot\gamma(t) \in \mathcal{D}(\gamma(t))$ for a.e. $t \in [0,1]$. Namely, γ is horizontal if there exist two functions $h_1, h_1 \in L^\infty(0,1)$ such that

$$\dot\gamma = h_1X_1(\gamma) + h_2X_2(\gamma) \quad \text{a.e. on } [0,1]. \tag{1.3}$$

The length of γ is then defined as

$$L(\gamma) = \int_0^1 g_\gamma(\dot\gamma,\dot\gamma)^{1/2}dt, \tag{1.4}$$

where g_x is a metric on $\mathcal{D}(x)$, $x \in \mathbb{R}^4$.

Date: March 15, 2013.

Fix a parameter $\alpha > 0$ and consider the initial and final points $L = (-1, \alpha, 0, 0) \in \mathbb{R}^4$ and $R = (1, \alpha, 0, 0) \in \mathbb{R}^4$. Let $\bar\gamma : [-1, 1] \to \mathbb{R}^4$ be the curve

$$\bar\gamma_1(t) = t, \quad \bar\gamma_2(t) = \alpha|t|, \quad \bar\gamma_3(t) = 0, \quad \bar\gamma_4(t) = 0, \quad t \in [-1, 1]. \tag{1.5}$$

The curve $\bar\gamma$ is horizontal and joins L to R. Moreover, it can be easily checked that $\bar\gamma$ is an abnormal extremal in the sense of Geometric Control Theory. The question is whether the curve $\bar\gamma$ is length minimizing or not, especially for small $\alpha > 0$.

The interest of this question arises from the following consideration. Let M be an n-dimensional smooth manifold with $n \geq 3$, and let $\mathcal{D}$ be a completely nonintegrable (i.e., bracket generating) distribution on M. Let $\mathcal{D}_1 = \mathcal{D}$ and $\mathcal{D}_i = [\mathcal{D}_1, \mathcal{D}_{i-1}]$, i.e., $\mathcal{D}_i$ is the linear span of all commutators $[\mathcal{D}_1, \mathcal{D}_{i-1}]$. We let $\mathcal{L}_0 = \{0\}$ and $\mathcal{L}_i = \mathcal{D}_1 + \ldots + \mathcal{D}_i$, $i \geq 1$. By the nonintegrability condition, for any $p \in M$ there exists $r \in \mathbb{N}$ such that $\mathcal{L}_r(p) = T_pM$, the tangent space of M at p. Assume that $\mathcal{D}$ is equiregular, i.e., assume that for each $i = 1, \ldots, r$

$$\dim \big(\mathcal{L}_i(p)/\mathcal{L}_{i-1}(p)\big) \text{ is constant for } p \in M. \tag{1.6}$$

In [2], Leonardi and the author proved the following theorem.

Theorem 1.1. *Let $(M, \mathcal{D}, g)$ be a sub-Riemannian manifold, where g is a metric on $\mathcal{D}$, that satisfies* (1.6) *and*

$$[\mathcal{L}_i, \mathcal{L}_j] \subset \mathcal{L}_{i+j-1}, \qquad i, j \geq 2,\ i + j > 4. \tag{1.7}$$

Then any curve in M with a corner is not length minimizing in $(M, \mathcal{D}, g)$.

A "curve with a corner" is a $\mathcal{D}$-horizontal curve $\gamma : [0, 1] \to M$ such that at some point $t \in (0, 1)$ the left and right derivatives $\dot\gamma_L(t) \neq \dot\gamma_R(t)$ exist, are different and nonzero.

In view of Theorem 1.1, it is natural to look for a nonsmooth length minimizing curve in a sub-Riemannian manifold violating (1.7). The step r associated with $(M, \mathcal{D})$ has to be at least 5, because condition (1.7) is automatically satisfied if $r \leq 4$. A first attempt could be to consider the free Carnot group of step 5 and rank 2. This structure is diffeomorphic to $\mathbb{R}^8$ and, by the results in [1], all abnormal extremals are in principle computable.

If we drop the equiregularity condition (1.6), however, the search for structures violating (1.7) is easier. The manifold $M = \mathbb{R}^4$ with the distribution $\mathcal{D}$ spanned by the vector fields (1.1) is one such example.

For $n \in \mathbb{N}$, we define the set of multi-indexes $\mathcal{I}_n = \{1, 2\}^n = \big\{(\beta_1, \ldots, \beta_n) \in \mathbb{N}^n : \beta_1, \ldots, \beta_n = 1, 2\big\}$ and, for any $\beta \in \mathcal{I}_n$, we let

$$X_\beta = [X_{\beta_1}, [\ldots, [X_{\beta_{n-1}}, X_{\beta_n}] \ldots]]. \tag{1.8}$$

We define the length of the commutator X_β as $\mathrm{len}(X_\beta) = n$ if and only if $\beta \in \mathcal{I}_n$.

For any $\beta \in \mathcal{I}_3 \cup \mathcal{I}_4$, we have $X_\beta(0) = 0$. On the other hand, when $\beta \in \mathcal{I}_5$ we have the following situation:

$$\begin{aligned} &\frac{1}{16} X_\beta = \frac{\partial}{\partial x_4}, \quad \text{when } \beta = (1,1,2,2,1) \text{ or } \beta = (1,2,1,2,1) \\ &\frac{1}{32} X_\beta = -\frac{\partial}{\partial x_4}, \quad \text{when } \beta = (2,1,1,2,1) \\ &X_\beta = 0 \quad \text{otherwise.} \end{aligned} \tag{1.9}$$

In particular, there holds

$$\frac{1}{48}[[X_2, X_1], [[X_2, X_1], X_1]] = \frac{\partial}{\partial x_4},$$

and thus (1.7) is violated with $i = 2$ and $j = 3$.

We show that the curve $\bar{\gamma}$ in (1.5) is not length minimizing for any small α, thus answering in the negative to the question raised by Agrachev and Gauthier. Namely, we prove the following

Theorem 1.2. *For any $\alpha > 0$ with $\alpha \neq 1$, the curve $\bar{\gamma}$ in* (1.5) *is not length minimizing in $(\mathbb{R}^4, \mathcal{D}, g)$, for any choice of metric g on $\mathcal{D}$.*

The proof of the theorem is constructive. For $\alpha \neq 1$, we construct a horizontal curve joining the left and right end-points L and R that is strictly shorter than $\bar{\gamma}$.

When $\alpha = 1$ the construction does not work and our shorter curves for $\alpha \neq 1$ converge to the curve $\bar{\gamma}$ as $\alpha \to 1$ (see Remark 3.1). The problem of establishing whether $\bar{\gamma}$ for $\alpha = 1$ is length minimizing or not is open.

In the second part of the paper, we extend Theorem 1.2 to 4-dimensional sub-Riemannian manifolds having at the corner point of the involved curve the same infinitesimal structure as $(\mathbb{R}^4, \mathcal{D})$.

Let M be a 4-dimensional smooth manifold and let $\mathcal{D} \subset TM$ be a distribution of 2-planes on M. Locally, we have $\mathcal{D} = \text{span}\{Y_1, Y_2\}$, where Y_1 and Y_2 are linearly independent smooth vector fields on M. We denote by $\mathcal{L}_2 = \text{span}\{Y_i, [Y_j, Y_k] : i, j, k = 1, 2\}$ the pointwise linear span of $\mathcal{D}$ and of the brackets of $\mathcal{D}$.

Fix a point $p \in M$. We make the following three assumptions:

(H1) The vector fields

$$Y_1,\ Y_2,\ Y_3 = \frac{1}{4}[Y_2, Y_1], \text{ and } Y_4 = \frac{1}{16}[Y_1, [Y_1, [Y_2, [Y_2, Y_1]]]] \tag{1.10}$$

are linearly independent at the point p.

(H2) For all $\beta \in \mathcal{I}_3 \cup \mathcal{I}_4$, we have

$$Y_\beta(p) = 0 \bmod \mathcal{L}_2(p), \tag{1.11}$$

where $\mathcal{L}_2(p) = \{Y(p) : Y \in \mathcal{L}_2\}$.

(H3) For all $\beta \in \mathcal{I}_5$ with $\beta = (*, *, *, 2, 1)$, we have

$$\begin{aligned} &Y_\beta(p) = 16Y_4(p) \bmod \mathcal{L}_2(p), \text{ for } \beta = (1,2,1,2,1) \text{ or } \beta = (1,1,2,2,1),\\ &Y_\beta(p) = -32Y_4(p) \bmod \mathcal{L}_2(p), \text{ for } \beta = (2,1,1,2,1), \text{ and}\\ &Y_\beta(p) = 0 \bmod \mathcal{L}_2(p) \text{ otherwise.} \end{aligned} \tag{1.12}$$

Finally, let g be a metric on $\mathcal{D}$ making the vector fields Y_1 and Y_2 orthogonal at p,

$$g_p(Y_1, Y_2) = 0. \tag{1.13}$$

In Section 5, we prove the following result. Below, we denote by $\dot\gamma_L(t)$ and $\dot\gamma_R(t)$ the left and right derivative of a curve $\gamma : [0,1] \to M$ at the point $t \in (0,1)$. The curve γ has a corner at the point $p = \gamma(t)$, if the left and right derivatives at t exist, do not vanish, and $\dot\gamma_L(t) \neq \dot\gamma_R(t)$.

Theorem 1.3. *Let $(M, \mathcal{D}, g)$ be 4-dimensional sub-Riemannian manifold satisfying (H1)-(H3) and (1.13) at the point $p \in M$. Let $\gamma : [0,1] \to M$ be a horizontal curve in $(M, \mathcal{D})$ having a corner at the point $p = \gamma(t)$, $t \in (0,1)$, such that $g_p(\dot\gamma_L(t), \dot\gamma_R(t)) \neq 0$. Then γ is not length minimizing.*

The proof of Theorem 1.3 relies upon a blow-up argument reducing the situation to the one of Theorem 1.2. Theorem 1.3 is proved in Section 5 and the preliminary nilpotent approximation is explained in Section 4. In Section 2, we set up the construction of the shorter curve used to prove Theorem 1.2. In Section 3, we solve a system of end-point equations, concluding the proof of Theorem 1.2.

We thank D. Vittone for his careful reading of an early version of this paper. We also thank the referee for suggesting to us possible extensions of Theorem 1.2.

2. Construction of the competing curve

In this section, we construct the competing curve used to prove Theorem 1.2. In a first step, we cut the corner of $\bar\gamma$ at $t = 0$ in the x_1x_2-plane. Lifting the new curve in $\mathbb{R}^2$ to a horizontal curve in $\mathbb{R}^4$ produces an error on the third and fourth coordinates of the final point. To restore the final point, we use two devices. We describe the intuition behind the construction. The vector fields X_1 and X_2 in (1.1) have a Heisenberg group structure in the $x_1x_2x_3$ space. Thus, to restore the third coordinate we have to add an "area" in the x_1x_2-plane equaling the "area" cut at the corner. We do this by means of a "rectangle" having basis of fixed length and variable height ε_1, the first parameter.

Restoring the fourth coordinate is more delicate. This is due to the fact that the coefficient of $\partial/\partial x_4$ of X_1 in (1.1), the power x_1^2, is always non negative. This means that the cut produces a positive error, see (2.7). To correct it, the perturbation of $\bar\gamma$ must contain arcs where the first coordinate is decreasing. To do this, we add a square with side length ε_2, the second parameter, at the final point R in the x_1x_2 plane.

Our goal is to prove that the final end-point can be adjusted via this construction. Technically, this means that we have to find solutions $\varepsilon_1, \varepsilon_2$ to a system of two equations depending on the cut parameter. Moreover, we need to check that the cost of length of "rectangle" and square does not exceed the gain of length of the cut. This will be done in Section 3.

We begin the construction of the shorter curve.

2.1. **Cut.** Fix a cut-parameter $0 < \eta < 1/4$, and let

$$T_\eta = \big\{(x_1, x_2) \in \mathbb{R}^2 : \alpha|x_1| < x_2 < \alpha\eta\big\}.$$

The curve $\bar\gamma$ is deviated along the cut, i.e., along the side of the triangle T_η parallel to the x_1-axis. In other words, we cut the corner. The cut produces an error on the final point, namely on the third and fourth coordinates. To correct these errors we use two devices, a "rectangle" and a square.

2.2. **Rectangle.** We deviate $\bar\gamma$ along a "rectangle" put along $\bar\gamma$. The length of the basis is fixed. The height is a variable parameter. Namely, for $\varepsilon_1 > 0$ we let

$$R_{\varepsilon_1} = \Big\{(x_1, x_2) \in \mathbb{R}^2 : \frac{1}{4} < x_1 < \frac{1}{2},\, \alpha x_1 < x_2 < \alpha x_1 + \varepsilon_1\Big\}.$$

The curve $\bar\gamma$ is deviated following clockwise three sides of the rectangle. When $\varepsilon_1 < 0$ the construction is analogous, but the rectangle is below the curve and we follow its boundary counterclockwise.

2.3. **Square.** Next we use a square with bottom-left vertex at R, the final point. Namely, for any $\varepsilon_2 \in \mathbb{R}$ we let:

$$Q_{\varepsilon_2} = \big\{(x_1, x_2) \in \mathbb{R}^2 : 1 < x_1 < 1 + |\varepsilon_2|, \alpha < x_2 < \alpha + |\varepsilon_2|\big\}.$$

When $\varepsilon_2 > 0$ we follow the boundary of the square clockwise. When $\varepsilon_2 < 0$ we follow the boundary counterclockwise.

The devices R_{ε_1} and Q_{ε_2} produce effects in the coordinates 3 and 4 of $\bar\gamma$ after the cut. We call $\gamma : [-1, 1] \to \mathbb{R}^4$ the curve obtained after using T_η, R_{ε_1}, and Q_{ε_2}. To construct the coordinate γ_3, we use the formula

$$\gamma_3(t) = 2\int_{-1}^{t} (\gamma_2\dot\gamma_1 - \gamma_1\dot\gamma_2)ds, \quad t \in [0, 1]. \tag{2.1}$$

To construct γ_4, we use the formula

$$\gamma_4(t) = \int_{-1}^{t} \dot\gamma_1\gamma_3^2\, ds, \quad t \in [0, 1]. \tag{2.2}$$

On suitable subintervals, we shall use different parameterizations for γ.

We construct step by step the horizontal lift of γ.

2.4. **Effect of T_η.** We parameterize on the interval $[-\eta,\eta]$ the piece of γ that is on the horizontal part of ∂T_η. We have, by (2.1),

$$\gamma_3(t) = \int_{-\eta}^{t} 2\alpha\eta ds = 2\alpha\eta(t+\eta), \quad t\in[-\eta,\eta].$$

The final value of γ_3 is $\gamma_3(\eta) = 4\alpha\eta^2$. The final value of γ_4 is, by (2.2),

$$\gamma_4(\eta) = \int_{-\eta}^{\eta} 4\alpha^2\eta^2(t+\eta)^2 dt = \frac{32}{3}\alpha^2\eta^5.$$

2.5. **First segment.** Now we have to follow for a time $t\in[\eta,1/4]$ the piece of the original curve $\bar\gamma$ connecting the cut to the rectangle R_{ε_1}. Here γ_3 is constant: $\gamma_3 \equiv 4\alpha\eta^2$. The final value of γ_4 is

$$\gamma_4(1/4) = \frac{32}{3}\alpha^2\eta^5 + \int_{\eta}^{1/4} 16\alpha^2\eta^4\, dt = 16\alpha^2\eta^4\Big(\frac{1}{4}-\frac{\eta}{3}\Big).$$

2.6. **Rectangle. 1st side.** We compute γ_3 along the first vertical side $\big\{(1/4,\alpha/4+t)\in\mathbb{R}^2 : t\in[0,\varepsilon_1]\big\}$ of R_{ε_1}:

$$\gamma_3(t) = 4\alpha\eta^2 - \frac{t}{2}, \quad t\in[0,\varepsilon_1].$$

The final value is $\gamma_3(\varepsilon_1) = 4\alpha\eta^2 - \frac{\varepsilon_1}{2}$. This value is correct also when $\varepsilon_1 < 0$. Along this piece of curve, γ_4 stays constant because $\dot\gamma_1 = 0$.

2nd side. Along the second side of the rectangle we have

$$\gamma_1(t) = \frac{1}{4}+t, \quad \gamma_2(t) = \frac{\alpha}{4}+\varepsilon_1+\alpha t, \quad t\in[0,1/4].$$

Then the third coordinate is

$$\begin{aligned}\gamma_3(t) &= 4\alpha\eta^2 - \frac{\varepsilon_1}{2} + 2\varepsilon_1 t, \quad t\in[0,1/4],\\ \gamma_3(1/4) &= 4\alpha\eta^2.\end{aligned}$$

The final value of γ_4 at the end of this side is

$$\begin{aligned}\gamma_4(1/4) &= 16\alpha^2\eta^4\Big(\frac{1}{4}-\frac{\eta}{3}\Big) + \int_0^{1/4}\Big(4\alpha\eta^2 - \frac{\varepsilon_1}{2} + 2\varepsilon_1 t\Big)^2 dt\\ &= 16\alpha^2\eta^4\Big(\frac{1}{4}-\frac{\eta}{3}\Big) + 4\alpha^2\eta^4 - \frac{1}{2}\alpha\eta^2\varepsilon_1 + \frac{1}{48}\varepsilon_1^2\\ &= \varrho.\end{aligned} \tag{2.3}$$

This is the final value of γ_4 after the rectangle R_{ε_1}. The number ϱ is defined via the last identity in (2.3).

3rd side. We compute γ_3 along the second vertical side $\big\{(1/2,\alpha/2+t)\in\mathbb{R}^2 : t\in[0,\varepsilon_1]\big\}$ of the rectangle R_{ε_1}. We have to follow $-X_2$ and we get

$$\begin{aligned}\gamma_3(t) &= 4\alpha\eta^2 + t, \quad t\in[0,\varepsilon_1],\\ \gamma_3(\varepsilon_1) &= 4\alpha\eta^2 + \varepsilon_1.\end{aligned}$$

Here, γ_4 is constant because $\dot\gamma_1 = 0$ on this side of the rectangle.

2.7. **Second diagonal segment.** Now we have to follow for a time $t \in [1/2, 1]$ the piece of the original curve $\bar\gamma$ connecting R_{ε_1} to the final point R. Here γ_3 is constant: $\gamma_3 \equiv 4\alpha\eta^2 + \varepsilon_1$. We call $\Delta_0(\varepsilon_1; \eta)$ the final value of γ_4 before the square Q_{ε_2}, and namely (recall the definition of ϱ in (2.3)),

$$\begin{aligned}\Delta_0(\varepsilon_1;\eta) &= \varrho + \frac{1}{2}(4\alpha\eta^2 + \varepsilon_1)^2 \\ &= 16\alpha^2\eta^4\Big(1 - \frac{\eta}{3}\Big) + \frac{7}{2}\alpha\eta^2\varepsilon_1 + \frac{25}{48}\varepsilon_1^2.\end{aligned} \tag{2.4}$$

2.8. **Square. Case $\varepsilon_2 > 0$.** We compute the effect on the third and fourth coordinates of the square Q_{ε_2}. Here, we do the computations for the case $\varepsilon_2 > 0$. In this case, we follow the boundary $\partial Q_{\varepsilon_2}$ clockwise.

1st vertical side. We follow the segment $\big\{(1, \alpha + t) \in \mathbb{R}^2 : t \in [0, \varepsilon_2]\big\}$. The coordinate γ_3 is

$$\begin{aligned}\gamma_3(t) &= 4\alpha\eta^2 + \varepsilon_1 - 2t, \quad t \in [0, \varepsilon_2],\\ \gamma_3(\varepsilon_2) &= 4\alpha\eta^2 + \varepsilon_1 - 2\varepsilon_2.\end{aligned}$$

The coordinate γ_4 stays constant along this side, $\gamma_4 \equiv \Delta_0(\varepsilon_1; \eta)$.

2nd (horizontal) side. We follow the line segment $\big\{(1 + t, \alpha + \varepsilon_2) \in \mathbb{R}^2 : t \in [0, \varepsilon_2]\big\}$. The coordinate γ_3 is

$$\begin{aligned}\gamma_3(t) &= 4\alpha\eta^2 + \varepsilon_1 - 2\varepsilon_2 + 2(\alpha + \varepsilon_2)t, \quad t \in [0, \varepsilon_2]\\ \gamma_3(\varepsilon_2) &= 4\alpha\eta^2 + \varepsilon_1 - 2\varepsilon_2 + 2(\alpha + \varepsilon_2)\varepsilon_2.\end{aligned}$$

Here, there is an important contribution to the fourth coordinate:

$$\begin{aligned}\gamma_4(\varepsilon_2) &= \Delta_0(\varepsilon_1;\eta) + \int_0^{\varepsilon_2} \big(4\alpha\eta^2 + \varepsilon_1 - 2\varepsilon_2 + 2(\alpha+\varepsilon_2)t\big)^2 dt \\ &= \Delta_0(\varepsilon_1;\eta) + (4\alpha\eta^2 + \varepsilon_1 - 2\varepsilon_2)^2\varepsilon_2 \\ &\qquad + 2(4\alpha\eta^2 + \varepsilon_1 - 2\varepsilon_2)(\alpha+\varepsilon_2)\varepsilon_2^2 + \frac{4}{3}(\alpha+\varepsilon_2)^2\varepsilon_2^3 \\ &= \Delta_1(\varepsilon_2, \varepsilon_2; \eta).\end{aligned}$$

Above, $\Delta_1(\varepsilon_2, \varepsilon_2; \eta)$ is defined via the last identity.

3rd (vertical) side. We have to follow the vertical side $\big\{(1 + \varepsilon_2, \alpha + t) \in \mathbb{R}^2 : t \in [0, \varepsilon_2]\big\}$ along the vector field $-X_2$. The final coordinate γ_3 is:

$$\gamma_3(\varepsilon_2) = 4\alpha\eta^2 + \varepsilon_1 + 2\alpha\varepsilon_2 + 4\varepsilon_2^2.$$

The coordinate γ_4 does not change.

4th (horizontal) side. We have to follow the horizontal side $\{(1+t,\alpha)\in\mathbb{R}^2 : t\in[0,\varepsilon_2]\}$ along the vector field $-X_1$. The coordinate γ_3 is :

$$\begin{aligned}\gamma_3(t) &= 4\alpha\eta^2+\varepsilon_1+2\alpha\varepsilon_2+4\varepsilon_2^2-2\alpha t,\quad t\in[0,\varepsilon_2]\\ \gamma_3(\varepsilon_2) &= 4\alpha\eta^2+\varepsilon_1+4\varepsilon_2^2.\end{aligned}$$

This is the final value of γ_3 after the entire construction, when $\varepsilon_2>0$. When $\varepsilon_2<0$ we have to replace ε_2^2 above with $\mathrm{sgn}(\varepsilon_2)|\varepsilon_2|^2$. There is a change of sign. We call $\Gamma_3(\varepsilon_1,\varepsilon_2;\eta)$ the final value of γ_3 after the entire construction, and namely,

$$\Gamma_3(\varepsilon_1,\varepsilon_2;\eta)=4\alpha\eta^2+\varepsilon_1+4\mathrm{sgn}(\varepsilon_2)|\varepsilon_2|^2. \tag{2.5}$$

The final value of γ_4 is:

$$\begin{aligned}\gamma_4(\varepsilon_2) &= \Delta_1(\varepsilon_2,\varepsilon_2;\eta)-\int_0^{\varepsilon_2}\big(4\alpha\eta^2+\varepsilon_1+2\alpha\varepsilon_2+4\varepsilon_2^2-2\alpha t\big)^2dt\\ &= \Delta_1(\varepsilon_2,\varepsilon_2;\eta)-\Big[(4\alpha\eta^2+\varepsilon_1+4\varepsilon_2^2)^2\varepsilon_2+2(4\alpha\eta^2+\varepsilon_1+4\varepsilon_2^2)\alpha\varepsilon_2^2+\frac{4}{3}\alpha^2\varepsilon_2^3\Big]\\ &= \Delta_0(\varepsilon_1;\eta)+4\varepsilon_2^2(1+2\varepsilon_2)\big[\varepsilon_2(1-2\varepsilon_2)-4\alpha\eta^2-\varepsilon_1\big]\\ &\qquad+2\varepsilon_2^3\big[4\alpha\eta^2+\varepsilon_1-2(\alpha+\varepsilon_2)-4\alpha\varepsilon_2\big]+\frac{4}{3}\varepsilon_2^4(2\alpha+\varepsilon_2).\end{aligned}$$

Recalling (2.4), we let for $\varepsilon_2>0$

$$\begin{aligned}\Gamma_4(\varepsilon_1,\varepsilon_2;\eta) &= 16\alpha^2\eta^4\Big(1-\frac{\eta}{3}\Big)+\frac{7}{2}\alpha\eta^2\varepsilon_1+\frac{25}{48}\varepsilon_1^2\\ &\quad+4\varepsilon_2^2(1+2\varepsilon_2)\big[\varepsilon_2(1-2\varepsilon_2)-4\alpha\eta^2-\varepsilon_1\big]\\ &\quad+2\varepsilon_2^3\big[4\alpha\eta^2+\varepsilon_1-2(\alpha+\varepsilon_2)-4\alpha\varepsilon_2\big]+\frac{4}{3}\varepsilon_2^4(2\alpha+\varepsilon_2).\end{aligned} \tag{2.6}$$

2.9. **Square. Case $\varepsilon_2<0$.** We compute the effect of the square Q_{ε_2} on the fourth coordinate when $\varepsilon_2<0$. In this case, we follow $\partial Q_{\varepsilon_2}$ counterclockwise.

1st horizontal side. We follow the segment $\{(1+t,\alpha)\in\mathbb{R}^2 : t\in[0,|\varepsilon_2|]\}$. The coordinate γ_3 is

$$\begin{aligned}\gamma_3(t) &= 4\alpha\eta^2+\varepsilon_1+2\alpha t,\quad t\in[0,|\varepsilon_2|],\\ \gamma_3(|\varepsilon_2|) &= 4\alpha\eta^2+\varepsilon_1+2\alpha|\varepsilon_2|.\end{aligned}$$

The final value of the coordinate γ_4 is

$$\begin{aligned}\gamma_4(|\varepsilon_2|) &= \Delta_0(\varepsilon_1;\eta)+\int_0^{|\varepsilon_2|}\big(4\alpha\eta^2+\varepsilon_1+2\alpha t\big)^2dt\\ &= \Delta_0(\varepsilon_1;\eta)+|\varepsilon_2|(4\alpha\eta^2+\varepsilon_1)^2+2\alpha|\varepsilon_2|^2(4\alpha\eta^2+\varepsilon_1)+\frac{4}{3}\alpha^2|\varepsilon_2|^3\\ &= \Delta_1(\varepsilon_1,\varepsilon_2;\eta).\end{aligned}$$

Here, $\Delta_1(\varepsilon_1,\varepsilon_2)$ for $\varepsilon_2<0$ is defined via the last identity.

2nd (vertical) side. We follow the line segment $\{(1+|\varepsilon_2|, \alpha+t) \in \mathbb{R}^2 : t \in [0, |\varepsilon_2|]\}$. The coordinate γ_3 is

$$\gamma_3(t) = 4\alpha\eta^2 + \varepsilon_1 + 2\alpha|\varepsilon_2| - 2(1 + |\varepsilon_2|)t, \quad t \in [0, |\varepsilon_2|]$$
$$\gamma_3(|\varepsilon_2|) = 4\alpha\eta^2 + \varepsilon_1 + 2\alpha|\varepsilon_2| - 2(1 + |\varepsilon_2|)|\varepsilon_2|.$$

The fourth coordinate does not change.

3rd (horizontal) side. We have to follow the horizontal side $\{(1 + t, \alpha + |\varepsilon_2|) \in \mathbb{R}^2 : t \in [0, |\varepsilon_2|]\}$ along the vector field $-X_1$. The coordinate γ_3 is:

$$\gamma_3(t) = 4\alpha\eta^2 + \varepsilon_1 + 2\alpha|\varepsilon_2| - 2(1 + |\varepsilon_2|)|\varepsilon_2| - 2(\alpha + |\varepsilon_2|)t,$$
$$\gamma_3(|\varepsilon_2|) = 4\alpha\eta^2 + \varepsilon_1 + 2\alpha|\varepsilon_2| - 2(1 + |\varepsilon_2|)|\varepsilon_2| - 2(\alpha + |\varepsilon_2|)|\varepsilon_2|,$$

The final value of the coordinate γ_4 is

$$\begin{aligned}\gamma_4(|\varepsilon_2|) &= \Delta_1(\varepsilon_1, \varepsilon_2; \eta) - \int_0^{|\varepsilon_2|} \Big(4\alpha\eta^2 + \varepsilon_1 + 2\alpha|\varepsilon_2| - 2(1 + |\varepsilon_2|)|\varepsilon_2| - 2(\alpha + |\varepsilon_2|)t\Big)^2 dt \\ &= \Delta_1(\varepsilon_1, \varepsilon_2; \eta) - |\varepsilon_2|\Big(4\alpha\eta^2 + \varepsilon_1 + 2\alpha|\varepsilon_2| - 2(1 + |\varepsilon_2|)|\varepsilon_2|\Big)^2 \\ &\quad + 2(\alpha + |\varepsilon_2|)|\varepsilon_2|^2\Big(4\alpha\eta^2 + \varepsilon_1 + 2\alpha|\varepsilon_2| - 2(1 + |\varepsilon_2|)|\varepsilon_2|\Big) - \frac{4}{3}(\alpha + |\varepsilon_2|)^2|\varepsilon_2|^3.\end{aligned}$$

This is the final value of γ_4 after the entire construction when $\varepsilon_2 < 0$. Then for $\varepsilon_2 < 0$ we let:

$$\begin{aligned}\Gamma_4(\varepsilon_1, \varepsilon_2; \eta) &= 16\alpha^2\eta^4\Big(1 - \frac{\eta}{3}\Big) + \frac{7}{2}\alpha\eta^2\varepsilon_1 + \frac{25}{48}\varepsilon_1^2 \\ &\quad + |\varepsilon_2|(4\alpha\eta^2 + \varepsilon_1)^2 + 2\alpha|\varepsilon_2|^2(4\alpha\eta^2 + \varepsilon_1) + \frac{4}{3}\alpha^2|\varepsilon_2|^3 \\ &\quad - |\varepsilon_2|\Big(4\alpha\eta^2 + \varepsilon_1 + 2\alpha|\varepsilon_2| - 2(1 + |\varepsilon_2|)|\varepsilon_2|\Big)^2 \\ &\quad + 2(\alpha + |\varepsilon_2|)|\varepsilon_2|^2\Big(4\alpha\eta^2 + \varepsilon_1 + 2\alpha|\varepsilon_2| - 2(1 + |\varepsilon_2|)|\varepsilon_2|\Big) \\ &\quad - \frac{4}{3}(\alpha + |\varepsilon_2|)^2|\varepsilon_2|^3.\end{aligned} \tag{2.7}$$

3. Proof of Theorem 1.2

In this section, we complete the proof of Theorem 1.2. We show that the system of end-point equations has a solution, and we prove that, for a small $\eta > 0$, this solution provides a curve that is shorter than $\bar{\gamma}$.

To correct the coordinates γ_3 and γ_4, we have to find solutions ε_1 and ε_2, depending on the (small) parameter $\eta > 0$, to the following system of equations

$$\begin{cases} \Gamma_3(\varepsilon_1, \varepsilon_2; \eta) = 0 \\ \Gamma_4(\varepsilon_1, \varepsilon_2; \eta) = 0. \end{cases} \tag{3.1}$$

Here, $\Gamma_4(\varepsilon_1, \varepsilon_2; \eta)$ is defined in (2.6) and (2.7), for $\varepsilon_2 > 0$ and $\varepsilon_2 < 0$, respectively.

By (2.5), the first equation $\Gamma_3(\varepsilon_1, \varepsilon_2; \eta) = 0$ is

$$4\alpha\eta^2 + \varepsilon_1 + 4\mathrm{sgn}(\varepsilon_2)|\varepsilon_2|^2 = 0. \tag{3.2}$$

We compute ε_1 using equation (3.2) and replace this value into the second equation $\Gamma_4(\varepsilon_1, \varepsilon_2; \eta) = 0$.

3.1. **Case** $\alpha > 1$**.** In this case, we look for a solution $\varepsilon_2 > 0$ and we use formula (2.6) for $\Gamma_4(\varepsilon_1, \varepsilon_2; \eta)$. Letting $\varepsilon_2 = \sigma > 0$, we obtain the following equation in σ

$$\begin{aligned} &\frac{1}{3}\Big(31\alpha^2\eta^4 - 16\alpha^2\eta^5 + 8\alpha\eta^2\sigma^2 + 25\sigma^4\Big)+ \\ &+4\sigma^3\big(1-\alpha+\sigma(3-2\alpha)+2\sigma^2\big) + \frac{4}{3}\sigma^4(2\alpha+\sigma) = 0. \end{aligned} \tag{3.3}$$

When $\alpha > 1$ equation (3.3) reads

$$\phi(\sigma) = \alpha^2\eta^4(31 - 16\eta) + 8\alpha\eta^2\sigma^2 + 12(1-\alpha)\sigma^3 + o(\sigma^3) = 0, \tag{3.4}$$

where $o(\sigma^3)$ contains terms with σ^4 and σ^5. Notice that the coefficient of σ^3 is negative, because $\alpha > 1$. The function ϕ introduced in (3.4) is continuous, and moreover

$$\phi(0) = \alpha^2\eta^4(31 - 16\eta) > 0,$$

as soon as $\eta < 31/16$. In fact, we have $\eta < 1/4$. On the other hand, we have

$$\phi\Big(C\sqrt[3]{\frac{\alpha^2}{\alpha-1}}\eta^{4/3}\Big) = \alpha^2\eta^4(31 - 12C^3) + o(\eta^4) < 0,$$

provided that we choose a constant C such that $12C^3 > 31$ and $\eta > 0$ is small enough.

By the theorem of zeros for continuous functions, for any $\eta > 0$ that is small enough, equation (3.4) has a solution $\varepsilon_2 = \sigma > 0$ such that

$$\varepsilon_2 = \sigma \le C\sqrt[3]{\frac{\alpha^2}{\alpha-1}}\eta^{4/3}. \tag{3.5}$$

Moreover, by (3.2),

$$\varepsilon_1 = -4\alpha\eta^2 - 4\mathrm{sgn}(\varepsilon_2)|\varepsilon_2|^2 = -4\alpha\eta^2 - 4\varepsilon_2^2.$$

In particular, ε_1 is asymptotic to $-4\alpha\eta^2$, for small $\eta > 0$, and thus we have

$$|\varepsilon_1| \le 5\alpha\eta^2, \tag{3.6}$$

for all $\eta > 0$ that are sufficiently small.

3.2. **Case** $0 < \alpha < 1$**.** In this case, we look for solutions $\varepsilon_1, \varepsilon_2$ to the system (3.1) such that $\varepsilon_2 < 0$. The unknown ε_1 is determined by (3.2). We use formula (2.7) for

$\Gamma_4(\varepsilon_1, \varepsilon_2; \eta)$. Letting $\sigma = |\varepsilon_2| > 0$, equation $\Gamma_4(\varepsilon_1, \varepsilon_2; \eta) = 0$ reads

$$\begin{aligned} &16\alpha^2\eta^4\Big(1 - \frac{\eta}{3}\Big) + 14\alpha\eta^2(\sigma^2 - \alpha\eta^2) + \frac{25}{3}(\sigma^2 - \alpha\eta^2)^2 \\ &+ 16\sigma^5 + 8\alpha\sigma^4 + \frac{4}{3}\alpha^2\sigma^3 \\ &- 4\sigma^3(\sigma + \alpha - 1)^2 \\ &+ 4(\alpha + \sigma)\sigma^3(\sigma + \alpha - 1) \\ &- \frac{4}{3}(\alpha + \sigma)^2\sigma^3 = 0. \end{aligned} \tag{3.7}$$

We may shorten equation (3.7) in the following way:

$$\alpha^2\eta^4(31 - 16\eta) - 8\alpha\eta^2\sigma^2 + 12(\alpha - 1)\sigma^3 + o(\sigma^3) = 0. \tag{3.8}$$

Compare equations (3.8) and (3.4). Notice that the coefficient of σ^3 in (3.8) is negative. The same argument used in the case $\alpha > 1$ proves that there exists a solution $|\varepsilon_2| = \sigma$ to equation (3.8) such that

$$0 < |\varepsilon_2| = \sigma \le C\sqrt[3]{\frac{\alpha^2}{1-\alpha}}\eta^{4/3}, \tag{3.9}$$

where C is the same constant as in the case $\alpha > 1$. As above, we deduce that

$$|\varepsilon_1| \le 5\alpha\eta^2. \tag{3.10}$$

3.3. **Gain of length.** We compute the difference between the length of the original curve $\bar{\gamma}$ and the curve γ obtained after applying the cut T_η, the "rectangle" R_{ε_1}, and the square Q_{ε_2}

Without loss of generality, we can assume that the metric g on $\mathcal{D}$ is the one making X_1 and X_1 orthonormal. In this case, the length $L(\gamma)$ of a horizontal curve γ as in (1.4) is

$$L(\gamma) = \int_0^1 |h(t)|dt, \quad h = (h_1, h_2). \tag{3.11}$$

By (3.11), the gain of length $G(\eta)$ obtained through the cut T_η is

$$G(\eta) = 2\eta(\sqrt{1+\alpha^2} - 1). \tag{3.12}$$

By (3.6) and (3.9), the cost of length $C_R(\varepsilon_1)$ of the correction made through the "rectangle" R_{ε_1} is

$$C_R(\varepsilon_1) = 2|\varepsilon_1| \le 10\alpha\eta^2.$$

By (3.5) and (3.9), the cost of length $C_Q(\varepsilon_2)$ of the correction made through the square Q_{ε_2} is

$$C_Q(\varepsilon_2) = 4|\varepsilon_2| \le 4C\sqrt[3]{\frac{\alpha^2}{\alpha-1}}\eta^{4/3}.$$

We conclude that

$$G(\eta) - C_R(\varepsilon_1) - C_Q(\varepsilon_2) \ge 2\eta(\sqrt{1+\alpha^2} - 1) - 10\alpha\eta^2 - 4C\sqrt[3]{\frac{\alpha^2}{\alpha-1}}\eta^{4/3} > 0,$$

for all $\eta > 0$ that are sufficiently small.

Thus, the construction provides a horizontal curve γ shorter than $\bar\gamma$ and joining the same end-points L and R. This ends the proof of Theorem 1.2.

Remark 3.1. When $\alpha \to 1$, our solutions $\varepsilon_1, \varepsilon_2$ to the system (3.1), along with the parameter η, converge to 0 and the shorter curve γ converges to $\bar\gamma$.

We may look for solutions to (3.1) in the case $\alpha = 1$. However, when $\alpha = 1$, equation (3.3) for the case $\sigma = \varepsilon_2 > 0$, reads

$$\eta^4(31 - 16\eta) + 8\eta^2\sigma^2 + 45\sigma^4 + o(\sigma^4) = 0.$$

This equation has no small solution $\sigma > 0$ for small η.

On the other hand, when $\alpha = 1$, equation (3.7) for the case $\sigma = |\varepsilon_2|$ with $\varepsilon_2 < 0$ reads

$$\eta^4(31 - 16\eta) - 8\eta^2\sigma^2 + 53\sigma^4 + o(\sigma^4) = 0.$$

And also this equation has for small $\eta > 0$ no small solution $\sigma > 0$.

Remark 3.2. The shorter curve γ constructed above has a "curl" at the end-point R. When $\alpha > 1$ this curl is oriented clockwise (i.e., $\varepsilon_2 > 0$). When $0 < \alpha < 1$ it is oriented counterclockwise (i.e., $\varepsilon_2 < 0$). This suggests that the geodesic joining L to R displays a similar behavior. One may wonder what is the behavior at the point R of the length minimizing curve joining L to R when $\alpha = 1$.

Remark 3.3. The proof of Theorem 1.2 shows why the problem of extending Theorem 1.1 to situations where (1.6) fails is technically complicated. The presence of coefficients depending on nonhorizontal coordinates (as the coefficient x_3^2 in the vector field X_1 in (1.1)) makes the construction of horizontal competitors complicated.

4. Nilpotent approximation

Let $(M, \mathcal{D})$ be a 4-dimensional sub-Riemannian manifold satisfying (H1)-(H3) at the point $p \in M$, where $\mathcal{D} = \mathrm{span}\{Y_1, Y_2\}$, and let $Y_1, \dots, Y_4$ be linearly independent vector fields at the point p, as in (1.10). In this section, we construct the homogeneous nilpotent approximation of $(M, \mathcal{D})$ at p.

In a neighborhood of $p \in M$, we fix the exponential coordinates of the first type induced by the frame $Y_1, \dots, Y_4$ starting from p. Then we can identify M with $\mathbb{R}^4$, $Y_1, \dots, Y_4$ with vector fields on $\mathbb{R}^4$, and p with $0 \in \mathbb{R}^4$. We have

$$x = \exp\Big(\sum_{i=1}^4 x_i Y_i\Big)(0), \quad x = (x_1, \dots, x_4) \in \mathbb{R}^4. \tag{4.1}$$

Here, the exponential mapping is defined by $\exp(Y)(0) = \gamma(1)$ where γ is the solution of $\dot\gamma = Y(\gamma)$ and $\gamma(0) = 0$.

We assign to the coordinate x_1 the weight $w_1 = 1$, to x_2 the weight $w_2 = 1$, to x_3 the weight $w_3 = 2$, and to x_4 the weight $w_4 = 5$. In fact, the length of Y_3 is

$\text{len}(Y_3) = 2$ and the length of Y_4 is $\text{len}(Y_4) = 5$. The natural dilations adapted to the frame $Y_1, \dots, Y_4$ are

$$\delta_\lambda(x) = (\lambda x_1, \lambda x_2, \lambda^2 x_3, \lambda^5 x_4), \quad x \in \mathbb{R}^4,\ \lambda > 0.$$

Let $Y = Y_\beta$ be a commutator of $Y_1, \dots, Y_4$, with the notation (1.8) for iterated commutators. Then we have

$$Y = \sum_{i=1}^{4} a_i(x) \frac{\partial}{\partial x_i}, \tag{4.2}$$

where $a_i \in C^\infty(\mathbb{R}^4)$, $i = 1, \dots, 4$, are smooth functions that have the structure described in the following proposition.

Proposition 4.1. *Assume that (H1) and (H2) hold. There are polynomials* $p_i : \mathbb{R}^4 \to \mathbb{R}$ *and functions* $r_i : \mathbb{R}^4 \to \mathbb{R}$, $i = 1, \dots, 4$, *such that:*

i) $a_i(x) = p_i(x) + r_i(x)$, $x \in \mathbb{R}^4$;
ii) $p_i(\delta_\lambda(x)) = \lambda^{w_i - \text{len}(Y)} p_i(x)$, *where* $\text{len}(Y)$ *is the length of* Y;
iii) $\lim_{\lambda \to \infty} \lambda^{w_i - \text{len}(Y)} r_i(\delta_{1/\lambda}(x)) = 0$, $x \in \mathbb{R}^4$.

Proposition 4.1 can be proved as in [4] on page 306. We omit the details, here.

Let $Y = Y_\beta$ be a vector field as in (4.2). For $\lambda > 0$, we let

$$Y^\lambda(x) = \sum_{i=1}^{4} \lambda^{w_i - \text{len}(Y)} a_i(\delta_{1/\lambda}(x)) \frac{\partial}{\partial x_i}, \quad x \in \mathbb{R}^4. \tag{4.3}$$

The mapping $Y \mapsto Y^\lambda$ is bracket-preserving. Namely, for any multi-index β and for $i = 1, 2$ we have

$$[Y_i, Y_\beta]^\lambda = [Y_i^\lambda, Y_\beta^\lambda], \quad \lambda > 0. \tag{4.4}$$

The vector fields $Y_1^\lambda, \dots, Y_4^\lambda$ induce on $\mathbb{R}^4$ exponential coordinates of the first type starting from 0:

$$x = \exp\Big(\sum_{i=1}^{4} x_i Y_i^\lambda\Big)(0), \quad x \in \mathbb{R}^4. \tag{4.5}$$

The proof of (4.5) relies upon the fact that if a curve γ solves $\dot\gamma = \sum_{i=1}^4 x_i Y_i(\gamma)$ then the curve $\gamma_\lambda = \delta_\lambda(\gamma)$ solves $\dot\gamma_\lambda = \sum_{i=1}^4 \lambda^{w_i} x_i Y_i^\lambda(\gamma_\lambda)$.

By Proposition 4.1, for any $Y = Y_\beta$ as in (4.2), we can define the vector field Y^∞ in $\mathbb{R}^4$

$$Y^\infty(x) = \lim_{\lambda \to \infty} Y^\lambda(x) = \sum_{i=1}^{4} p_i(x) \frac{\partial}{\partial x_i}, \quad x \in \mathbb{R}^4,$$

where p_i, $i = 1, \dots, 4$, are polynomials such that $p_i \circ \delta_\lambda = \lambda^{w_i - \text{len}(Y)} p_i$. In particular, if $w_i < \text{len}(Y)$ then $p_i = 0$. Passing to the limit as $\lambda \to \infty$ in (4.4), we see that also the mapping $Y \mapsto Y^\infty$ is bracket-preserving

$$[Y_i, Y_\beta]^\infty = [Y_i^\infty, Y_\beta^\infty], \quad \lambda > 0. \tag{4.6}$$

Moreover, passing to the limit as $\lambda \to \infty$ in (4.5) we see that $Y_1^\infty, \dots, Y_4^\infty$ induce exponential coordinates of the first type:

$$x = \exp\Big(\sum_{i=1}^4 x_i Y_i^\infty\Big)(0), \quad x \in \mathbb{R}^4. \tag{4.7}$$

Then the vector fields $Y_1^\infty, \dots, Y_4^\infty$ are

$$\begin{aligned} Y_1^\infty &= \frac{\partial}{\partial x_1} + q_1 \frac{\partial}{\partial x_3} + p_1 \frac{\partial}{\partial x_4}, \\ Y_2^\infty &= \frac{\partial}{\partial x_2} + q_2 \frac{\partial}{\partial x_3} + p_2 \frac{\partial}{\partial x_4}, \\ Y_3^\infty &= \frac{1}{4}[Y_2^\infty, Y_1^\infty] = \frac{\partial}{\partial x_3} + p_3 \frac{\partial}{\partial x_4}, \\ Y_4^\infty &= \frac{1}{16}[Y_1^\infty, [Y_1^\infty, [Y_2^\infty, [Y_2^\infty, Y_1^\infty]]]] = \frac{\partial}{\partial x_4}. \end{aligned} \tag{4.8}$$

Above, q_i and p_i are polynomials such that $q_i \circ \delta_\lambda = \lambda q_i$ and $p_i \circ \delta_\lambda = \lambda^4 p_i$ for $i = 1, 2$ and $\lambda > 0$. In particular, for $i = 1, 2$, we have

$$q_i = a_{i1} x_1 + a_{i2} x_2 \quad \text{and} \quad p_i = p_{i1} + x_3 p_{i2} + d_i x_3^2,$$

where $a_{ij}, d_i \in \mathbb{R}$ are constants, p_{i1} are homogeneous polynomials of degree 4 in the variables x_1, x_2, and p_{i2} are homogeneous polynomials of degree 2 in the variables x_1, x_2. By the relation $Y_3^\infty = \frac{1}{4}[Y_2^\infty, Y_2^\infty]$, we deduce that the polynomial p_3 is

$$p_3 = \frac{1}{4}\Big\{ \Big(\frac{\partial}{\partial x_2} + q_2 \frac{\partial}{\partial x_3} \Big) p_1 - \Big(\frac{\partial}{\partial x_1} + q_1 \frac{\partial}{\partial x_3} \Big) p_2 \Big\}. \tag{4.9}$$

If $Y_1, \dots, Y_4$ satisfy (1.12), then the vector fields $Y_1^\infty, \dots, Y_4^\infty$ satisfy, for all $\beta \in \mathcal{I}_5$ with $\beta = (*, *, *, 2, 1)$,

$$\begin{aligned} &Y_\beta^\infty(0) = 16 Y_4^\infty(0) \text{ for } \beta = (1,2,1,2,1) \text{ or } \beta = (1,1,2,2,1), \\ &Y_\beta^\infty(0) = -32 Y_4^\infty(0) \text{ for } \beta = (2,1,1,2,1), \text{ and} \\ &Y_\beta^\infty(0) = 0 \text{ otherwise.} \end{aligned} \tag{4.10}$$

We claim that

$$Y_1^\infty = X_1 \quad \text{and} \quad Y_2^\infty = X_2, \tag{4.11}$$

where X_1 and X_2 are the vector fields (1.1).

For any $x \in \mathbb{R}^4$, let $\gamma : [0,1] \to \mathbb{R}^4$ be the curve such that $\dot\gamma = \exp\big(\sum_{i=1}^4 x_i Y_i^\infty(\gamma)\big)(0)$ and $\gamma(0) = 0$. Condition (4.7) implies that $\gamma(t) = tx$ for all $t \in [0,1]$. Differentiating this identity at $t = 1$, we obtain

$$\sum_{i=1}^4 x_i \frac{\partial}{\partial x_i} = \sum_{i=1}^4 x_i Y_i^\infty(x), \quad x \in \mathbb{R}^4. \tag{4.12}$$

The equation for the coordinate $i = 3$ in (4.12) is $x_1q_1 + x_2q_2 = 0$, i.e.,

$$\sum_{i,j=1}^{2} a_{ij}x_ix_j = 0.$$

This equation along with the relation $\frac{1}{4}[Y_2^\infty, Y_1^\infty](0) = \partial/\partial x_3$ gives

$$q_1 = 2x_1 \quad \text{and} \quad q_2 = -2x_1. \tag{4.13}$$

The equation for the coordinate $i = 4$ in (4.12) is $x_1p_1 + x_2p_2 + x_3p_3 = 0$. By (4.13) and (4.9), this equation reads

$$x_1p_1 + x_2p_2 + \frac{x_3}{4}\Big(\partial_2 p_1 - 2x_1\partial_3 p_1 - \partial_1 p_2 - 2x_2\partial_3 p_2\Big) = 0,$$

or, equivalently,

$$x_1p_{11} + x_2p_{21} + \frac{x_3}{4}\Big(2x_1p_{12} + 2x_2p_{22} + \partial_2 p_{11} - \partial_1 p_{21}\Big) + \frac{x_3^2}{4}\Big(\partial_2 p_{12} - \partial_1 p_{22}\Big) = 0. \tag{4.14}$$

Setting to 0 the coefficients of the powers of x_3 in (4.14), we obtain the equations:

$$x_1p_{11} + x_2p_{21} = 0, \tag{4.15}$$

$$2x_1p_{12} + 2x_2p_{22} + \partial_2 p_{11} - \partial_1 p_{21} = 0, \tag{4.16}$$

$$\partial_2 p_{12} - \partial_1 p_{22} = 0. \tag{4.17}$$

Equation (4.15) implies that

$$p_{11} = x_2 p \quad \text{and} \quad p_{21} = -x_1 p, \tag{4.18}$$

where

(4.19) p is a homogeneous polynomial of degree 3 in the variables x_1, x_2.

Now equation (4.16) reads

$$5p + 2x_1p_{12} + 2x_2p_{22} = 0. \tag{4.20}$$

The vector fields Y_1^∞ and Y_2^∞ are

$$Y_1^\infty = X_1 + w_1\frac{\partial}{\partial x_4} \quad \text{and} \quad Y_2^\infty = X_2 + w_2\frac{\partial}{\partial x_4},$$

where X_1, X_2 are the vector fields (1.1), and

$$\begin{aligned} w_1 &= x_2 p + x_3 p_{12} + (d_1 - 1)x_3^2, \\ w_2 &= -x_1 p + x_3 p_{22} + d_2 x_3^2. \end{aligned} \tag{4.21}$$

We claim that $w_1 = w_2 = 0$. Once this claim is proved, the main claim (4.11) will follow.

We compute commutators of length 5 of Y_1^∞ and Y_2^∞, as functions of X_1, X_1, w_1, and w_2. First of all we have

$$[Y_2^\infty, Y_1^\infty] = [X_2, X_1] + (X_2w_1 - X_1w_2)\frac{\partial}{\partial x_4}.$$

Then, for any $i,j,k=1,2$ we have

$$
\begin{aligned}
[Y_i^\infty,[Y_j^\infty,[Y_k^\infty,[Y_2^\infty,Y_1^\infty]]]] &= [X_i,[X_j,[X_k,[X_2,X_1]]]]+ \\
&\quad + \big(X_iX_jX_kX_2w_1 - X_iX_jX_kX_1w_2 - 4X_iX_j\partial_3 w_k\big)\frac{\partial}{\partial x_4}.
\end{aligned}
\tag{4.22}
$$

By (1.9) and (4.10), commutators (of length 5) of Y_1^∞ and Y_2^∞, and commutators of X_1 and X_2 satisfy the same relations at the point $x=0$. From (4.22), we deduce that we have for all $i,j,k=1,2$

$$
X_iX_jX_kX_2w_1 - X_iX_jX_kX_1w_2 - 4X_iX_j\partial_3 w_k = 0, \quad \text{at } x=0. \tag{4.23}
$$

Notice that $X_iX_jX_kX_2w_1 - X_iX_jX_kX_1w_2 - 4X_iX_j\partial_3 w_k$ is a homogeneous polynomial of degree 0 for the dilations $(x_1,x_2,x_3)\mapsto(\lambda x_1,\lambda x_2,\lambda^2 x_3)$, $\lambda>0$. Thus, equation (4.23) holds for all $x\in\mathbb{R}^4$. By integration, we deduce that $X_jX_kX_2w_1 - X_jX_kX_1w_2 - 4X_j\partial_3 w_k$ is constant in $\mathbb{R}^3$. For $X_jX_kX_2w_1 - X_jX_kX_1w_2 - 4X_j\partial_3 w_k$ is a homogeneous polynomial of degree 1, we deduce that it is identically zero. Repeating the same argument, we conclude that for $k=1,2$

$$
X_kX_2w_1 - X_kX_1w_2 - 4\partial_3 w_k = 0 \quad \text{in } \mathbb{R}^3. \tag{4.24}
$$

In order to analyze equation (4.24), we preliminarily compute

$$
\begin{aligned}
X_2w_1 &= p + x_2\partial_2 p + x_3\partial_2 p_{12} - 2x_1p_{12} - 4(d_1-1)x_1x_3,\\
X_1w_2 &= -p - x_1\partial_1 p + x_3\partial_1 p_{22} + 2x_2p_{22} + 4d_2x_2x_3.
\end{aligned}
$$

When $k=1$, equation (4.24) reads

$$
\begin{aligned}
&3\partial_1 p + x_1\partial_1\partial_1 p + x_2\partial_1\partial_2 p - 6p_{12} - 2x_1\partial_1 p_{12} - 2x_2\partial_1 p_{22} + 2x_2(\partial_2 p_{12} - \partial_1 p_{22})+\\
&\qquad -8(d_1-1)x_1x_2 - 8d_2x_2^2 + x_3(\partial_1\partial_2 p_{12} - \partial_1\partial_1 p_{22} - 12(d_1-1)) = 0.
\end{aligned}
$$

As $\partial_1 p$ is a homogeneous polynomial of degree 2, we have $x_1\partial_1\partial_1 p + x_2\partial_2\partial_1 p = 2\partial_1 p$. Also using the fact that p_{12} is a homogeneous polynomial of degree 2, and using identity (4.17), the previous equation reads

$$
5\partial_1 p - 10p_{12} - 8(d_1-1)x_1x_2 - 8d_2x_2^2 - 12(d_1-1)x_3 = 0.
$$

This implies

$$
d_1 - 1 = 0 \quad \text{and} \quad \partial_1 p - 2p_{12} - \frac{8}{5}d_2x_2^2 = 0. \tag{4.25}
$$

When $k=2$, equation (4.24) reads

$$
\begin{aligned}
&3\partial_2 p + x_1\partial_2\partial_1 p + x_2\partial_2\partial_2 p - 6p_{22} - 2x_1\partial_2 p_{12} - 2x_2\partial_2 p_{22} + 2x_1(\partial_1 p_{22} - \partial_2 p_{12})+\\
&\qquad +8(d_1-1)x_1^2 + 8d_2x_1x_2 + x_3(\partial_2\partial_2 p_{12} - \partial_2\partial_1 p_{22} - 12d_2) = 0.
\end{aligned}
$$

Using the fact that $\partial_2 p$ is homogeneous of degree 2, identity (4.17), the fact that p_{22} is homogeneous of degree 2, and $d_1=1$, the previous equation reads

$$
5\partial_2 p - 10p_{22} + 8d_2x_1x_2 - 12d_2x_3 = 0.
$$

This implies $d_2=0$ and $\partial_2 p - 2p_{22} = 0$. We multiply the latter equation by x_2, we multiply by x_1 the second equation in (4.25) (with $d_2=0$), we sum the two resulting equations and we obtain $x_1\partial_1 p + x_2\partial_2 p - 2x_1p_{12} - 2x_2p_{22} = 0$. Using the fact that p is homogeneous of degree 3, we finally obtain

$$3p - 2x_1p_{12} - 2x_2p_{22} = 0. \tag{4.26}$$

From (4.26) and (4.20), we deduce that $p = 0$ and thus $p_{12} = p_{22} = 0$. We conclude that $w_1 = w_2 = 0$ and this proves the main claim (4.11).

5. Proof of Theorem 1.3

In this section, we prove Theorem 1.3. As in Section 4, we assume that $M = \mathbb{R}^4$, $p = 0$, $\mathcal{D} = \mathrm{span}\{Y_1, Y_2\}$, and that the vector fields $Y_1, \dots, Y_4$ satisfy (H1)-(H3) and induce exponential coordinates of the first type in $\mathbb{R}^4$, as in (4.1). On $\mathcal{D}$, we fix a metric g satisfying (1.13).

For $\lambda > 0$, let Y_1^λ and Y_2^λ be the vector fields in (4.3) and let $\mathcal{D}^\lambda = \mathrm{span}\{Y_1^\lambda, Y_2^\lambda\}$. Let g^λ be the metric on $\mathcal{D}^\lambda$ defined by

$$g_x^\lambda(Y_i^\lambda, Y_j^\lambda) = g_{\delta_{1/\lambda}(x)}(Y_i, Y_j), \quad x \in \mathbb{R}^4, \quad i, j = 1, 2.$$

When $\lambda = \infty$, we let $\mathcal{D}^\infty = \mathrm{span}\{Y_1^\infty, Y_2^\infty\}$, where, by the discussion of Section 4, we have $Y_1^\infty = X_1$ and $Y_2^\infty = X_2$. On $\mathcal{D}^\infty$, we define the metric g^∞

$$g_x^\infty(Y_i^\infty, Y_j^\infty) = \lim_{\lambda\to\infty} g_x^\lambda(Y_i^\lambda, Y_j^\lambda) = \lim_{\lambda\to\infty} g_{\delta_{1/\lambda}(x)}(Y_i, Y_j) = g_0(Y_i, Y_j), \quad x \in \mathbb{R}^4.$$

Remark 5.1. From (1.13), it follows that

$$g_x^\infty(Y_1^\infty, Y_2^\infty) = g_0(Y_1, Y_2) = 0, \quad x \in \mathbb{R}^4.$$

Without loss of generality, we also assume that $g_x^\infty(Y_i^\infty, Y_i^\infty) = 1$ for all $x \in \mathbb{R}^4$ and $i = 1, 2$. In other words, the metric g^∞ makes the vector fields Y_1 and Y_2 orthonormal. It follows that linear mappings $T : \mathbb{R}^4 \to \mathbb{R}^4$ of the form

$$T(x) = \big(U(x_1, x_2), \det(U)x_3, x_4\big), \quad x = (x_1, \dots, x_4) \in \mathbb{R}^4,$$

where $U \in O(2)$ is an orthogonal mapping in $\mathbb{R}^2$, are isometries of $(\mathbb{R}^4, \mathcal{D}^\infty, g^\infty)$.

Let $\gamma : [-1, 1] \to M = \mathbb{R}^4$ be a $\mathcal{D}$-horizontal curve such that $\gamma(0) = 0$ and

$$\dot\gamma_L(0) \neq \dot\gamma_R(0), \tag{5.1}$$

where $\dot\gamma_L(0) \neq 0$ and $\dot\gamma_R(0) \neq 0$ are the left and right derivatives of γ at $t = 0$.

For $\lambda > 0$, the curves $\gamma^\lambda(t) = \delta_\lambda\gamma(t/\lambda)$, $t \in [-\lambda, \lambda]$, are $\mathcal{D}^\lambda$-horizontal. Moreover, the limit curve $\gamma^\infty : (-\infty, \infty) \to \mathbb{R}^4$

$$\gamma^\infty(t) = \begin{cases} \lim_{\lambda\to\infty} \gamma^\lambda(t) = \dot\gamma_R(0)t & t \geq 0, \\ \lim_{\lambda\to\infty} \gamma^\lambda(t) = \dot\gamma_L(0)t & t < 0, \end{cases} \tag{5.2}$$

is $\mathcal{D}^\infty$-horizontal. Here, the vectors $\dot\gamma_L(0)$ and $\dot\gamma_R(0)$ are identified with vectors of $\mathbb{R}^4$, and have the form $(*, *, 0, 0)$. The curve γ^∞ lies therefore in the x_1x_2-plane.

Proposition 5.2. *Assume that the curve γ is length minimizing in $(\mathbb{R}^4, \mathcal{D}, g)$. Then the curve γ^∞ is length minimizing in $(\mathbb{R}^4, \mathcal{D}^\infty, g^\infty)$.*

Proposition 5.2 is proved in [2], Proposition 2.4.

We conclude the proof of Theorem 1.3. Assume by contradiction that there exists a curve γ as in Theorem 1.3 that is length minimizing in $(\mathbb{R}^4, \mathcal{D}, g)$. By Proposition 5.2, the curve γ^∞ in (5.2) is length minimizing in $(\mathbb{R}^4, \mathcal{D}^\infty, g^\infty)$. By assumption, we have $g_0(\dot\gamma_L(0), \dot\gamma_R(0)) \neq 0$. It follows that

$$g_0^\infty(\dot\gamma_L^\infty(0), \dot\gamma_R^\infty(0)) = \lim_{\lambda\to\infty} g_0^\lambda(\dot\gamma_L^\lambda(0), \dot\gamma_R^\lambda(0)) = g_0(\dot\gamma_L(0), \dot\gamma_R(0)) \neq 0. \tag{5.3}$$

By Remark 5.1 and (5.1), we can assume that γ^∞ has the form (1.5) with $\alpha > 0$ (up to re-parameterization), and by (5.3) we have $\alpha \neq 1$. By Theorem 1.2, the curve γ^∞ is not length minimizing. This is a contradiction.

References

[1] E. Le Donne, G. P. Leonardi, R. Monti, D. Vittone, Extremal curves in nilpotent Lie groups, Geom. Funct. Anal. 2013 (to appear)

[2] G. P. Leonardi, R. Monti, End-point equations and regularity of sub-Riemannian geodesics, Geom. Funct. Anal. 18 (2008), 552–582

[3] W. Liu, H. Sussmann, Shortest paths for sub-Riemannian metrics on rank-two distributions, Mem. Amer. Math. Soc. 118 (1995), x+104

[4] G. A. Margulis, G. D. Mostow, Some remarks on the definition of tangent cones in a Carnot-Carathèodory space. J. Anal. Math. 80 (2000), 299–317

[5] R. Montgomery, Abnormal minimizers, SIAM J. Control Optim., **32** (1994), 1605–1620

[6] R. Monti, Regularity results for sub-Riemannian geodesics, Calc. Var. 2013 (to appear)