THE REGULARITY PROBLEM FOR SUB-RIEMANNIAN GEODESICS

DAVIDE VITTONE

Abstract. We study the regularity problem for sub-Riemannian geodesics, i.e., for those curves that minimize length among all curves joining two fixed endpoints and whose derivatives are tangent to a given, smooth distribution of planes with constant rank. We review necessary conditions for optimality and we introduce extremals and the Goh condition. The regularity problem is nontrivial due to the presence of the so-called abnormal extremals, i.e., of certain curves that satisfy the necessary conditions and that may develop singularities. We focus, in particular, on the case of Carnot groups and we present a characterization of abnormal extremals, that was recently obtained in collaboration with E. Le Donne, G. P. Leonardi and R. Monti, in terms of horizontal curves contained in certain algebraic varieties. Applications to the problem of geodesics' regularity are provided.

1. Introduction

A *sub-Riemannian manifold* is a smooth, connected n-dimensional manifold M endowed with a smooth, bracket-generating sub-bundle $\Delta \subset TM$ (called *horizontal*), having constant rank r, and with a smooth metric g on Δ. In these notes, we give a brief overview on the problem of the regularity of length minimizers, i.e., of the shortest (with respect to g) curves among all curves that join two fixed endpoints and are *horizontal*, i.e., tangent to Δ. We also present some results on the problem recently obtained, in the framework of *Carnot groups*, in collaboration with E. Le Donne, G. P. Leonardi and R. Monti [19, 20]. These notes are based on a course given by the author on the occasion of the ERC School *Geometric Measure Theory and Real Analysis* held at the Centro De Giorgi, Pisa, in October 2013.

It is well-known (see e.g. the basic references [3, 4, 27]) that length minimizers are *extremals*, i.e., satisfy certain necessary conditions given by the Pontryagin Maximum Principle of Optimal Control Theory. Extremals may be either *normal* or *abnormal*: while normal extremals are always smooth, abnormal ones may develop singularities. Hence, the regularity problem for length minimizers is reduced to the regularity of abnormal minimizers.

Let us spend a few words about the literature and the state of the art on the problem. We do not claim to be exhaustive and we refer to the beautiful introductions in [23, 30] for a more comprehensive account.

The author is supported by PRIN 2010-11 Project "Calculus of Variations" of MIUR (Italy), GNAMPA of INdAM (Italy), University of Padova, and Fondazione CaRiPaRo Project "Nonlinear Partial Differential Equations: models, analysis, and control-theoretic problems".

It was originally claimed in [35] that length minimizing curves are smooth, all of them being normal extremals. The wrong argument relied upon an incorrect application of Pontryagin Maximum Principle, ignoring the possibility of abnormal extremals; see also [13]. A correction to [35] appeared in [36], where it was proved that minimizers in *strong bracket-generating* distributions are always normal and, hence, smooth.

The first example of a strictly abnormal length minimizer was provided by R. Montgomery in [26]. Other examples in the same vein are studied in [22, 37]. Distributions of rank 2 are rich of abnormal geodesics: in [23], W. Liu and H. J. Sussmann introduced a class of abnormal extremals, called *regular abnormal* extremals, that are always locally length minimizing. Strictly abnormal length minimizers appear also in the setting of Carnot groups, see [11]. Notice, however, that all known examples of abnormal minimizers are smooth, so that the regularity problem is widely open.

As we said, abnormal extremals may have singularities. In the paper [16], G. P. Leonardi and R. Monti developed an elaborate *cutting-the-corner* technique (see also [2]) to show that, when the horizontal bundle satisfies a certain technical condition, length minimizers do not have corner-type singularities. In several interesting structures (among them, Carnot groups of rank 2 and nilpotency *step* at most 4), this is enough to conclude that length minimizers are smooth. More recently, R. Monti [29] was able to exclude certain singularities of higher order for length minimizers in structures satisfying the same condition introduced in [16].

Finally, a complete characterization of extremals in Carnot groups was recently obtained in [19, 20]. In particular, abnormal extremals in this setting are characterized as horizontal curves contained in certain algebraic varieties; the key tool here is represented by *extremal polynomials*. This allows for several applications; let us only mention the results discussed in these notes. First, one can give a very short proof of the regularity of length minimizers in Carnot groups of step not greater than 3 (a result first proved in [38]). Second, we describe a new technique for proving the negligibility of the endpoints of abnormal extremals; for the motivations behind this problem, which are only sketched in Remark 3.23, see [27, Section 10.2] and [2]. This technique cannot be applied to general Carnot groups; however, it is likely to work in many specific examples.

A few words about the organization of these notes. In Section 2, we introduce the sub-Riemannian (or *Carnot-Carathéodory*) distance. In Section 3, we derive the necessary conditions of extremality for length minimizers; the properties of normal and abnormal extremals are briefly described in Sections 3.3 and 3.4. In Section 4, we introduce Carnot groups and present the characterization of extremals obtained in [19, 20]. In Section 5, we apply our results to prove the smoothness of minimizers in Carnot groups of step at most 3. Finally, in Section 6 we describe the technique connected with the negligibility problem for the endpoints of abnormal extremals.

Acknowledgements. It is a pleasure to thank R. Monti for many invaluable comments, remarks and suggestions. We have to thank G. P. Leonardi for suggesting the characterization of normal extremals in Carnot groups contained in Theorem 4.12. We are indebted with E. Le Donne and E. Pasqualetto for their careful reading of a preliminary version of these notes. Finally, we want to thank the organizers G. Alberti, L. Ambrosio and C. De Lellis, as well as all the participants, for the nice time and the pleasant atmosphere during the ERC School on Geometric Measure Theory and Real Analysis.

2. The Carnot-Carathéodory distance

2.1. **Definition of Carnot-Carathéodory distance.** A sub-Riemannian manifold is a smooth, connected n-dimensional manifold M endowed with a smooth, bracket-generating sub-bundle $\Delta \subset TM$ (called *horizontal sub-bundle*) of constant rank r and with a smooth metric g on Δ. Without loss of generality, the regularity problem for length minimizers can be localized. Namely, we can assume that $M = \mathbb{R}^n$ and that the horizontal bundle Δ is generated by smooth, linearly independent vector fields $X_1, \dots, X_r$ which form an orthonormal system with respect to g.

A Lipschitz continuous curve $\gamma : [0,1] \to \mathbb{R}^n$ is said to be *horizontal* if $\dot\gamma(t) \in \Delta_{\gamma(t)}$ for a.e. $t \in [0,1]$, i.e., if

$$\dot\gamma(t) = \sum_{j=1}^{r} h_j(t) X_j(\gamma(t)) \quad \text{for a.e. } t \in [0,1] \tag{2.1}$$

for suitable functions $h = (h_1, \dots, h_r) \in L^\infty([0,1], \mathbb{R}^r)$. We will refer to the functions h_j as to the *controls* associated with γ. The length of γ is

$$L(\gamma) := \int_0^1 |h(t)|\, dt\,.$$

The fact that the length L is defined by integrating $|h(t)| := (h_1(t)^2 + \cdots + h_r(t)^2)^{1/2}$ corresponds to the fact that $X_1, \dots, X_r$ are orthonormal.

Definition 2.1. The *Carnot-Carathéodory (CC) distance* between $x, y \in \mathbb{R}^n$ is defined as

$$d(x,y) := \inf\left\{ L(\gamma) : \gamma \text{ is horizontal, } \gamma(0) = x \text{ and } \gamma(1) = y \right\}. \tag{2.2}$$

The structure induced by the Carnot-Carathéodory distance is often called *sub-Riemannian* because, intuitively speaking, the “allowed” directions form only a subspace of the whole tangent bundle.

Exercise 2.2. Given γ and h as above, define $L_2(\gamma) := \big(\int_0^1 |h(t)|^2\, dt \big)^{1/2}$; prove that, for any $x, y \in \mathbb{R}^n$, the CC distance $d(x,y)$ is equal to

$$d_2(x,y) := \inf\left\{ L_2(\gamma) : \gamma \text{ is horizontal, } \gamma(0) = x \text{ and } \gamma(1) = y \right\}.$$

2.2. **The Chow-Rashevski theorem.** The family of curves in the right hand side of (2.2) might be empty (i.e., no horizontal curve joins x and y), hence d is not necessarily a distance. Consider, for instance, $\mathbb{R}^3$ with horizontal distribution generated by the vector fields $X_1 := (1,0,0)$ and $X_2 := (0,1,0)$: clearly, in this case there is no horizontal curve joining the origin and the point $(0,0,1)$.

On the contrary, it is immediate to check that d is a distance whenever we can guarantee that any couple of points can be connected by horizontal curves. Sufficient conditions for connectivity are well-known; they are usually based on the following observation, which is a consequence of the Baker-Campbell-Hausdorff formula (see e.g. [40]). Here and in the sequel, we adopt the standard identification between vector fields and first-order derivations.

Fact. Given a point $p \in \mathbb{R}^n$, two vector fields X, Y and a positive real number $t \ll 1$, one has

$$e^{-tY}e^{-tX}e^{tY}e^{tX}(p) = e^{t^2[X,Y]}(p) + o(t^2)\,, \tag{2.3}$$

where we define $e^{tX}(p) := c(t)$ as the curve c solving the problem $\dot{c} = X(c),\ c(0) = p$, and where the *commutator* (or *bracket*) $[X, Y]$ is the vector field $XY - YX$.

Roughly speaking, if we are allowed to move along both X and Y, then we are also allowed to move in the direction of their commutator. This holds also for iterated brackets and suggests the following result, which we state without proof. Here and in the sequel, we denote by $\mathcal{L}(X_1, \dots, X_r)$ the Lie algebra of vector fields (with Lie product $[\cdot,\cdot]$) generated by $X_1, \dots, X_r$.

Theorem 2.3. *Assume that the* bracket-generating condition

$$\textit{rank}\ \mathcal{L}(X_1, \dots, X_r)(x) = n \quad \forall x \in \mathbb{R}^n \tag{2.4}$$

holds. Then, for any $x, y \in \mathbb{R}^n$ there exists a horizontal curve joining x and y; in particular, the Carnot-Carathéodory distance d is an actual distance.

Theorem 2.3 was proved independently by W. L. Chow [10] and P. K. Rashevski [33]; see also [8].

Condition (2.4) is also known as *Hörmander condition*, as it was used by L. Hörmander in the seminal paper [14] on hypoelliptic equations. In what follows, we will always assume that (2.4) is satisfied.

2.3. **The Ball-Box Theorem.** In this section, we state the classical Ball-Box Theorem by A. Nagel, E. M. Stein and S. Wainger [32], that allows to compare (small) CC balls $B(x, r)$ with suitable anisotropic boxes. See also [31].

If $\Omega \subset \mathbb{R}^n$ is an open bounded set, then there exists an integer κ such that condition (2.4) is verified at every $x \in \Omega$ by commutators of $X_1, \dots, X_r$ with length at most κ (the length of a commutator $[\cdots[X_{j_1}, X_{j_2}], X_{j_3}], \dots, X_{j_m}]$ is by definition m). Let $Y_1, \dots, Y_q$ be a fixed enumeration of all the commutators of length at most κ and let $d(Y_k) \in \{1, \dots, \kappa\}$ denote the length of Y_k.

Given $x \in \mathbb{R}^n$ and a multi-index $\mathcal{I} = (i_1, \dots, i_n) \in \{1, \dots, q\}^n$, define

$$\begin{aligned} d(\mathcal{I}) &:= d(Y_{i_1}) + \cdots + d(Y_{i_n}) \\ \lambda_{\mathcal{I}}(x) &:= \det \text{ col } \left[Y_{i_1}(x) \,|\, Y_{i_2}(x) \,|\, \cdots \,|\, Y_{i_n}(x) \right] \end{aligned}$$

and the map

$$E_{\mathcal{I}}(x, h) := e^{h_1 Y_{i_1} + h_2 Y_{i_2} + \cdots + h_n Y_{i_n}}(x), \quad h \in \mathbb{R}^n.$$

Let us define the *box*

$$\mathscr{B}_{\mathcal{I}}(x, r) := \{ E_{\mathcal{I}}(x, h) : h \in \mathbb{R}^n \text{ and } \max_{k=1,\dots,n} |h_k|^{1/d(Y_{i_k})} < r \}.$$

We can then state the following result.

Theorem 2.4. *Let $K \subset \Omega$ be a compact set; then, there exist positive numbers $\hat{r}, \alpha, \beta$, with $\beta < \alpha < 1$, such that the following holds. If $x \in K$, $r \in (0, \hat{r})$ and $\mathcal{I}$ are such that*

$$|\lambda_{\mathcal{I}}(x)| r^{d(\mathcal{I})} > \tfrac{1}{2} \max_{\mathcal{J}} |\lambda_{\mathcal{J}}(x)| r^{d(\mathcal{J})}, \tag{2.5}$$

then

$$B(x, \beta r) \subset \mathscr{B}_{\mathcal{I}}(x, \alpha r) \subset B(x, r).$$

In particular, there exists $C = C(K) > 0$ such that $d(x, y) \leqslant C|x - y|^{1/\kappa}$ for any $x, y \in K$.

Remark 2.5. As an important consequence, one can deduce from Theorem 2.4 that the topology induced by d is the standard one on $\mathbb{R}^n$.

3. Length minimizers and extremals

This section is devoted to the derivation of necessary conditions for length minimizing curves. Usually, such conditions are obtained by making use of the Pontryagin Maximum Principle of Optimal Control Theory; however, we will not directly refer to it. Our presentation is not meant to be exhaustive; the basic references [3, 4, 27] can be consulted for a more detailed account on these and related topics.

3.1. **Length minimizers, existence and non-uniqueness.**

Definition 3.1. A horizontal curve $\gamma : [0, 1] \to \mathbb{R}^n$ is a *length minimizer* if it realizes the distance between its endpoints, i.e., if $L(\gamma) = d(\gamma(0), \gamma(1))$.

As a preliminary result, we prove the local existence of minimizers.

Theorem 3.2. *For any $x \in \mathbb{R}^n$, there exists $\rho > 0$ with the following property: if $d(x, y) < \rho$, then there exists a length minimizer connecting x and y.*

Proof. Let $x \in \mathbb{R}^n$ be fixed and let $\rho > 0$ be such that the CC ball $B(x, \rho)$ is a bounded open subset of $\mathbb{R}^n$. The existence of such a ρ is guaranteed by Remark 2.5. We are going to prove that, for any point $y \in B(x, \rho)$, there exists a length minimizer from x to y.

Let then x, ρ, y be as above and consider a sequence of horizontal curves $\gamma^k : [0,1] \to \mathbb{R}^n, k \in \mathbb{N}$, such that

$$\gamma^k(0) = x, \quad \gamma^k(1) = y \quad \text{and} \quad L(\gamma^k) \to d(x,y) \text{ as } k \to \infty\,.$$

In particular, for large k we have Im $\gamma^k \subset B(x,\rho) \Subset \mathbb{R}^n$. Let $h^k : [0,1] \to \mathbb{R}^r$ be the controls associated with γ^k; we can assume that, for any k, $|h^k| \equiv L(\gamma^k)$ is constant on $[0,1]$. Thus, for large k, the Euclidean norm $\|\dot\gamma^k\|_{L^\infty}$ is bounded uniformly in k; by Ascoli-Arzelà's Theorem we deduce that, up to a subsequence, there exists a Lipschitz curve $\gamma : [0,1] \to \overline{B(x,\rho)}$ such that $\gamma^k \to \gamma$ uniformly on $[0,1]$. Now, by the Dunford-Pettis theorem, up to a further subsequence we have that $h^k \rightharpoonup h$ in $L^1([0,1],\mathbb{R}^r)$. For any $t \in [0,1]$ there holds

$$\gamma^k(t) = \int_0^t \sum_{j=1}^r h_j^k(s)\, X_j(\gamma^k(s))\, ds\,.$$

Taking into account the uniform convergence of γ^k and the weak convergence of h^k, on passing to the limit as $k \to \infty$ we get

$$\gamma(t) = \int_0^t \sum_{j=1}^r h_j(s)\, X_j(\gamma(s))\, ds\,,$$

i.e., the curve γ is horizontal with associated controls h. In particular we have

$$\gamma(0) = x, \quad \gamma(1) = y \quad \text{and} \quad L(\gamma) = \|h\|_{L^1} \leqslant \liminf_{k\to\infty} \|h^k\|_{L^1} = d(x,y)\,,$$

i.e., γ is a length minimizer connecting x and y. This concludes the proof. $\square$

Unlike Riemannian geodesics, sub-Riemannian length minimizers are not unique, even locally. To illustrate this situation, we consider the sub-Riemannian *Heisenberg group*, i.e., the space $\mathbb{R}^3$ with horizontal distribution generated by the linearly independent vector fields

$$X_1 := \partial_1 - \frac{x_2}{2}\partial_3, \quad X_2 := \partial_2 + \frac{x_1}{2}\partial_3\,.$$

Notice that the bracket-generating condition is trivially satisfied because $[X_1, X_2] = \partial_3$. Our aim it to describe length minimizers starting from the origin; a more detailed study can be found in [5].

It can be easily checked that a Lipschitz curve $\gamma = (\gamma_1, \gamma_2, \gamma_3) : [0,1] \to \mathbb{R}^3$ is horizontal if and only if

$$\dot\gamma_3 = -\frac{\gamma_2}{2}\dot\gamma_1 + \frac{\gamma_1}{2}\dot\gamma_2 \quad \text{a.e. on } [0,1].$$

In particular, if c denotes the planar curve $c(t) := (\gamma_1(t), \gamma_2(t))$ and Σ is the planar region bounded by c and by the (oriented) segment σ joining $c(1)$ to the origin, one can recover $\gamma_3(1)$ as

$$\gamma_3(1) = \int_c \Big(-\frac{x_2}{2}dx_1 + \frac{x_1}{2}dx_2\Big) = \int_{c\cup\sigma} \Big(-\frac{x_2}{2}dx_1 + \frac{x_1}{2}dx_2\Big) = \int_\Sigma dx_1 \wedge dx_2\,,$$

where we have used Stokes' theorem. Hence, the problem of connecting the origin $(0,0,0)$ to (x,y,t) with a length minimizing horizontal curve amounts to the problem of connecting $(0,0)$ to (x_1,x_2) with the shortest planar curve enclosing (algebraic) area x_3. This is (a version of) Dido's problem and it is well-known that, if $x_3 \neq 0$, its solutions are arcs of circles. The corresponding horizontal curves are spirals which can be parametrized by arclength by the formulae

$$\begin{cases} x_1(t) = \dfrac{A(1-\cos\varphi t) + B\sin\varphi t}{\varphi} \\ x_2(t) = \dfrac{-B(1-\cos\varphi t) + A\sin\varphi t}{\varphi} \\ x_3(t) = -\dfrac{\varphi t - \sin\varphi t}{2\varphi^2} \end{cases} \tag{3.1}$$

for suitable $(A,B) \in S^1 \subset \mathbb{R}^2$ and $\varphi \neq 0$. If $x_3 = 0$ we have instead the straight lines $\gamma(t) = (Bt, At, 0)$.

It can be proved that the spirals in (3.1) are length minimizing up to time $t = 2\pi/\varphi$, when they reach the point $(0,0,\pi/\varphi^2)$. In particular, for any $\varepsilon > 0$ there exists a family of length minimizers joining the origin and $(0,0,\varepsilon)$: this family is parametrized by $(A,B) \in S^1$ with the choice $\varphi = \sqrt{\pi/\varepsilon}$.

3.2. **First-order necessary conditions.** We want to derive necessary conditions for a horizontal curve to be length minimizing. To this end, we fix a length minimizer $\gamma : [0,1] \to \mathbb{R}^n$ with associated optimal controls h. Without loss of generality, we may assume that $\gamma(0) = 0$ and that γ is parametrized by constant speed, i.e., that $|h| = c$ a.e. on $[0,1]$. In particular, by Exercise 2.2, γ is also a minimizer for the problem

$$\inf \left\{ L_2(\tilde\gamma) : \tilde\gamma \text{ is horizontal, } \tilde\gamma(0) = \gamma(0) \text{ and } \tilde\gamma(1) = \gamma(1) \right\}.$$

For any fixed $x \in \mathbb{R}^n$, let $\gamma_x : [0,1] \to \mathbb{R}^n$ be the solution of

$$\begin{cases} \dot\gamma_x = h \cdot X(\gamma_x) \\ \gamma_x(0) = x\,, \end{cases}$$

where we write $h \cdot X(\gamma_x)$ to denote the function $\sum_{j=1}^r h_j X_j(\gamma_x)$ defined on $[0,1]$.

For any fixed $t \in [0,1]$, let us define the diffeomorphism $F_t : \mathbb{R}^n \to \mathbb{R}^n$ by

$$F_t(x) := \gamma_x(t)\,. \tag{3.2}$$

Given another control $k \in L^\infty([0,1],\mathbb{R}^r)$ we denote by q_k the horizontal curve solving

$$\begin{cases} \dot q_k = k \cdot X(q_k) \\ q_k(0) = 0\,, \end{cases} \tag{3.3}$$

Finally, given $v \in L^\infty([0,1],\mathbb{R}^r)$, we define the *(extended) endpoint map* $\varphi_v : \mathbb{R} \to \mathbb{R}^{n+1}$

$$\varphi_v(s) := \left(F_1^{-1}(q_{h+sv}(1)), \textstyle\int_0^1 (h+sv)^2 \right)\,. \tag{3.4}$$

The first component of $\varphi_v(s)$ is (up to a diffeomorphism) the endpoint of the horizontal curve q_{h+sv}, while the last component is (the square of) its 2-length $L_2(q_{h+sv})$.

Lemma 3.3. *If γ is length minimizing and parametrized by constant speed, then there exists $\overline{\xi} \in \mathbb{R}^{n+1} \setminus \{0\}$ such that*

$$\langle \overline{\xi}, \varphi'_v(0) \rangle = 0 \quad \forall v \in L^\infty([0,1], \mathbb{R}^r)\,. \tag{3.5}$$

Proof. Assume not: then, there exist $v_1, \dots, v_{n+1} \in L^\infty([0,1], \mathbb{R}^r)$ such that the vectors $\varphi'_{v_1}(0), \dots, \varphi'_{v_{n+1}}(0) \in \mathbb{R}^{n+1}$ are linearly independent. Writing $s \cdot v := s_1 v_1 + \dots + s_{n+1} v_{n+1}$, it follows that the map

$$\begin{aligned} \Phi : &\mathbb{R}^{n+1} \to \mathbb{R}^{n+1} \\ &\Phi(s_1, \dots, s_{n+1}) := \big(F_1^{-1}(q_{h+s\cdot v}(1)), L_2(q_{h+s\cdot v})^2\big) \end{aligned}$$

is such that $\nabla \Phi(0)$ is invertible because $\frac{\partial \Phi}{\partial s_i}(0) = \varphi'_{v_i}(0)$. In particular, Φ is an open map (in a neighbourhood of 0), hence one can find $\bar{s} \in \mathbb{R}^{n+1}$ such that the control $\bar{h} := h + \bar{s}_1 v_1 + \dots + \bar{s}_{n+1} v_{n+1}$ satisfies

$$\begin{aligned} &F_1^{-1}(q_{\bar{h}}(1)) = F_1^{-1}(q_h(1)) \\ &L_2(q_{\bar{h}})^2 < L_2(q_h)^2\,. \end{aligned}$$

Since F_1 is a diffeomorphism, if $\bar{s}$ is close enough to 0, the first equality above implies that $q_{\bar{h}}(1) = q_h(1)$. This contradicts the minimality of $\gamma = q_h$. $\square$

Remark 3.4. An important role in the derivation of the necessary conditions in Theorem 3.6 will be played by the previous lemma. A key point in its proof is the fact that the extended endpoint map cannot be an open map in any neighbourhood of length minimizers. This suggests a sort of recipe to produce necessary conditions for optimality: in principle, any open mapping theorem might be used to derive necessary conditions. Also the Goh condition in the subsequent Theorem 3.20 is obtained by exploiting a suitable open mapping theorem.

Lemma 3.5. *If $v \in L^\infty([0,1], \mathbb{R}^r)$ is fixed and φ_v is as in (3.4), then*

$$\varphi'_v(0) = \left(\int_0^1 JF_t(0)^{-1}(v \cdot X(\gamma(t)))\, dt, 2\int_0^1 \langle h(t), v(t) \rangle\, dt \right) \in \mathbb{R}^n \times \mathbb{R}\,, \tag{3.6}$$

where JF_t is the Jacobian matrix of F_t and, again, $v \cdot X = v_1 X_1 + \dots + v_r X_r$.

Proof. Let $s \in \mathbb{R}$ be fixed and, for any $t \in [0,1]$, define $x_{h+sv}(t) := F_t^{-1}(q_{h+sv}(t))$; equivalently,

$$q_{h+sv}(t) = F_t(x_{h+sv}(t))\,. \tag{3.7}$$

In particular, the first n components in the definition of $\varphi_v(s)$ are equal to $x_{h+sv}(1)$.

We can differentiate (3.7) with respect to t to obtain

$$(h + sv) \cdot X(q_{h+sv}) = h \cdot X(q_{h+sv}) + JF_t(x_{h+sv})\dot{x}_{h+sv}\,,$$

hence

$$\dot{x}_{h+sv} = s\; JF_t(x_{h+sv})^{-1}[v \cdot X(q_{h+sv})]\,.$$

It follows that

$$x_{h+sv}(t) = s\int_0^t JF_\tau(x_{h+sv}(\tau))^{-1}[v\cdot X(F_\tau(x_{h+sv}(\tau)))]\,d\tau\,,$$

i.e.,

$$\begin{aligned}\varphi_v'(0) &= \left(\left.\frac{\partial x_{h+sv}(1)}{\partial s}\right|_{s=0}, 2\int_0^1 \langle h(\tau), v(\tau)\rangle\,d\tau\right)\\ &= \left(\int_0^1 JF_\tau(x_h(\tau))^{-1}[v\cdot X(F_\tau(x_h(\tau)))]\,d\tau, 2\int_0^1 \langle h(\tau), v(\tau)\rangle\,d\tau\right).\end{aligned}$$

The desired equality (3.6) easily follows on noticing that $x_h(\tau) = F_\tau^{-1}(q_h(\tau)) = F_\tau^{-1}(\gamma(\tau)) = 0$. ☐

We can now pass to the main result of this section.

Theorem 3.6 (First-order necessary conditions)**.** *Let $\gamma : [0,1]\to\mathbb{R}^n$ be a length minimizer with $\gamma(0)=0$ and with associated controls h; assume that γ is parametrized by constant speed, i.e., $|h|\equiv c$. Then, there exist $\xi_0\in\{0,1\}$ and $\xi\in \mathrm{Lip}([0,1],\mathbb{R}^n)$ such that*

(i) $(\xi(t),\xi_0)\neq 0$ *for any* $t\in[0,1]$*;*
(ii) *for any* $j=1,\dots,r$*, the equality* $\xi_0 h_j + \langle \xi, X_j(\gamma)\rangle = 0$ *holds a.e. on* $[0,1]$*;*
(iii) $\dot\xi = -(\sum_{j=1}^r h_j\, JX_j(\gamma))^T\xi$ *a.e. on* $[0,1]$*,*

where JX_j denotes the $n\times n$ Jacobian matrix of $X_j:\mathbb{R}^n\to\mathbb{R}^n$ and the superscript ^{T} denotes matrix transposition.

Proof. Let $\overline{\xi}\in\mathbb{R}^{n+1}\setminus\{0\}$ be as in Lemma 3.3; write $\overline{\xi} =: (\xi(0),\xi_0/2)\in\mathbb{R}^n\times\mathbb{R}$. Using Lemma 3.5, we deduce from (3.5) the following necessary condition:

$$\begin{aligned}0 &= \int_0^1 \sum_{j=1}^r v_j(t)\left\{\langle \xi(0), JF_t(0)^{-1}(X_j(\gamma(t)))\rangle + \xi_0 h_j\right\}\,dt\\ &= \int_0^1 \sum_{j=1}^r v_j(t)\left\{\langle [JF_t(0)^{-1}]^T\xi(0), X_j(\gamma(t))\rangle + \xi_0 h_j\right\}\,dt \quad \forall v\in L^\infty([0,1],\mathbb{R}^r)\,.\end{aligned}$$

Upon defining $\xi(t) := [JF_t(0)^{-1}]^T\xi(0)$, the Fundamental lemma of the Calculus of Variations immediately implies statement (ii).

Statement (i) is clearly true if $\xi_0\neq 0$ (notice that, in this case, one can also normalize $\overline{\xi}$ to have $\xi_0=1$); on the contrary, if $\xi_0=0$ we have $\xi(0)\neq 0$, hence $\xi(t)\neq 0$ for all $t\in[0,1]$ because $JF_t(0)^{-1}$ is invertible. Hence, also (i) is proved.

We are left with statement (iii). By definition of $\xi(t)$, we have $\xi(0) = JF_t(0)^T\xi(t)$ and, on differentiating with respect to t,

$$0 = \left(\frac{d}{dt}JF_t(0)^T\right)\xi(t) + JF_t(0)^T\dot\xi(t) \quad \text{a.e. on } [0,1]\,. \tag{3.8}$$

Let us compute

$$\begin{aligned}\frac{d}{dt}JF_t(0) &= J\,\frac{d}{dt}F_t(x)\Big|_{x=0} = J\left(\sum_{j=1}^r h_j(t)X_j(F_t(x))\right)\Big|_{x=0}\\ &= \sum_{j=1}^r h_j(t)\,JX_j(F_t(0))\,JF_t(0)\\ &= \left(\sum_{j=1}^r h_j(t)\,JX_j(\gamma(t))\right)JF_t(0) \quad \text{a.e. on } [0,1]\,.\end{aligned}$$

Recalling (3.8) and the fact that $JF_t(0)^T$ is invertible, we obtain

$$\dot\xi(t) = -\left(\sum_{j=1}^r h_j(t)\,JX_j(\gamma(t))\right)^T \xi(t) \quad \text{for a.e. } t\in[0,1]\,,$$

as desired. □

Definition 3.7. A horizontal curve $\gamma:[0,1]\to\mathbb{R}^n$ with $\gamma(0)=0$ and with associated controls h is said to be an *extremal* if there exist $\xi_0\in\{0,1\}$ and $\xi\in\mathrm{Lip}([0,1],\mathbb{R}^n)$ such that statements (i), (ii) and (iii) in Theorem 3.6 hold. The function ξ is called *dual curve* (or *dual variable*).
If $\xi_0=1$, we say that γ is a *normal extremal*.
If $\xi_0=0$, we say that γ is an *abnormal extremal*.

Theorem 3.6 states that length minimizers parametrized by constant speed are also extremals; on the contrary, there exist extremals that are not minimizers, see Section 3.5. We do not require extremals to be parametrized by constant speed because this is automatically satisfied for normal extremals (see Exercise 3.11), while for abnormal extremals the parametrization plays essentially no role (see Exercise 3.17).

We will review the main properties of normal and abnormal extremals in Sections 3.3 and 3.4; now, a few observations are in order.

Remark 3.8. An extremal γ might be normal and abnormal at the same time, in the sense that it could possess two different dual curves that make γ normal and abnormal. An example of this phenomenon is given in Exercise 4.5. An extremal which is normal but not abnormal is called *strictly normal*; on the contrary, we call *strictly abnormal* an extremal which is abnormal but not also normal.

Exercise 3.9. Prove that, if γ is strictly normal, then it possesses a unique dual curve $\xi(t)$.

Hint: assume that $\xi_1(t),\xi_2(t)$ are dual curves making γ normal; prove that γ is abnormal with associated dual curve $\xi_1-\xi_2$.

Theorem 3.6 possesses also an Hamiltonian formulation. Define the Hamiltonian

$$H(x,\xi) := \sum_{j=1}^r \langle X_j(x),\xi\rangle^2 \quad x,\xi\in\mathbb{R}^n\,.$$

Then, the following result holds.

Exercise 3.10. If γ is a normal extremal with dual variable ξ, then the couple (γ, ξ) solves the system of Hamiltonian equations

$$\begin{cases} \dot{\gamma} = -\dfrac{1}{2}\dfrac{\partial H}{\partial \xi}(\gamma, \xi) \\ \dot{\xi} = \dfrac{1}{2}\dfrac{\partial H}{\partial x}(\gamma, \xi) \,. \end{cases}$$

If γ is an abnormal extremal with dual variable ξ, then $H(\gamma, \xi) \equiv 0$.

3.3. **Normal extremals.** In this section we deal with basic properties and facts about normal extremals. We begin with the following exercise.

Exercise 3.11. Let γ be a normal extremal; then, γ is parametrized by constant speed.

Hint: use Exercise 3.10 and the fact that, if h denotes the controls associated with γ, then $|h(t)|^2 = H(\gamma(t), \xi(t))$.

The most important result in this subsection is the following one.

Proposition 3.12. *Normal extremals are C^∞ smooth.*

Proof. Let $\gamma : [0, 1] \to \mathbb{R}^n$ be a normal extremal with associated controls h and dual curve ξ. Using (2.1) and (ii), (iii) in Theorem 3.6 we easily obtain the following chain of implications

$$\begin{array}{rlll} & \gamma, \xi \in C^0([0, 1]) & \overset{\text{(ii)}}{\Longrightarrow} & h_j \in C^0([0, 1]) \ \forall j = 1, \dots, r \\ \overset{(2.1),\text{(iii)}}{\Longrightarrow} & \gamma, \xi \in C^1([0, 1]) & \overset{\text{(ii)}}{\Longrightarrow} & h_j \in C^1([0, 1]) \ \forall j = 1, \dots, r \\ \overset{(2.1),\text{(iii)}}{\Longrightarrow} & \gamma, \xi \in C^2([0, 1]) & \overset{\text{(ii)}}{\Longrightarrow} & h_j \in C^2([0, 1]) \ \forall j = 1, \dots, r \\ \Longrightarrow & \dots & & \end{array}$$

□

Exercise 3.13. Prove that, if γ is a normal extremal with dual curve ξ, then condition (ii) in Theorem 3.6 holds on the whole interval $[0, 1]$ (and not only almost everywhere).

The following results, as well as the Proposition 3.12, show that normal minimizers/extremals share several common features with Riemannian geodesics.

Remark 3.14. When $r = n$ (i.e., the CC structure is indeed Riemannian), any length minimizer/extremal γ is strictly normal. Otherwise, there would exist a dual curve ξ such that $\langle \xi, X_j(\gamma) \rangle = 0$ for any $j = 1, \dots, n$. Since $X_1, \dots, X_n$ now form a basis of $\mathbb{R}^n$, we obtain that $\xi \equiv 0$, which contradicts (i) in Theorem 3.6.

Exercise 3.15. Assume again that we are in the Riemannian case $r = n$. Then, by (2.1) and (ii) in Theorem 3.6, there is a natural way of identifying $\dot{\gamma}, h$ and ξ, in the sense that any of the three uniquely determines the others. Prove that equation (iii) in Theorem 3.6 corresponds to the ODE of Riemannian geodesics.

The following important result is a special case of more general results in Optimal Control Theory, see for instance [7, 15, 13] and [23, Appendix C].

Theorem 3.16. *Every normal extremal is locally length minimizing.*

On the contrary, strictly abnormal extremals might not be length minimizers, see Section 3.5.

3.4. **Abnormal extremals.** By Theorem 3.6 (ii), an abnormal extremal γ and its dual variable ξ satisfy

$$\langle \xi, X_j(\gamma)\rangle = 0 \text{ on } [0,1] \quad \forall j = 1,\dots,r\,. \tag{3.9}$$

The compact notation $\xi \perp \Delta_\gamma$ will often be used to abbreviate the previous formula.

When dealing with abnormal extremals, it is not necessary to require that they are parametrized by constant speed; this is justified by the following fact.

Exercise 3.17. Assume that $\tilde\gamma : [0,1] \to \mathbb{R}^n$ is an abnormal extremal parametrized by constant speed and with dual curve $\tilde\xi$. Let γ be a different parametrization of the same curve; namely, let $\gamma := \tilde\gamma \circ f$ for an increasing, Lipschitz continuous homeomorphism $f : [0,1] \to [0,1]$. Then, γ satisfies (i), (ii) and (iii) in Theorem 3.6 with $\xi := \tilde\xi \circ f$.

Exercise 3.18. Prove that, if γ is an abnormal extremal with dual curve ξ, then condition (ii) in Theorem 3.6 holds on the whole interval $[0,1]$ (and not only almost everywhere).

Abnormal extremals are often introduced in the literature as singular points of the endpoint map; a few comments on this viewpoint are in order.

Going back to Section 3.2, let $\gamma : [0,1] \to \mathbb{R}^n$ be an extremal with $\gamma(0) = 0$ and associated optimal controls $h \in L^\infty([0,1],\mathbb{R}^r)$. Define the *endpoint map* $\mathrm{End} : L^\infty([0,1],\mathbb{R}^r) \to \mathbb{R}^n$ by

$$\mathrm{End}(k) := q_k(1), \quad k \in L^\infty([0,1],\mathbb{R}^r)\,,$$

the curve q_k being defined as in (3.3). For any $v \in L^\infty([0,1],\mathbb{R}^r)$, the map $\varphi_v(s)$ in (3.4) can then be rewritten as

$$\varphi_v(s) = \big(F_1^{-1} \circ \mathrm{End}(h+sv), L_2(q_{h+sv})^2\big)\,,$$

where the diffeomorphism F_1 is defined as in (3.2).
Now, if γ is an abnormal extremal, then the vector $\overline{\xi} = (\xi(0), \xi_0/2) \in \mathbb{R}^n \times \mathbb{R}$ provided by Lemma 3.3 is such that $\xi_0 = 0$. Hence (again by Lemma 3.3), the vector $\xi(0) \neq 0$ is such that

$$\xi(0) \perp \frac{d}{ds}\big(F_1^{-1} \circ \mathrm{End}(h+sv)\big)\Big|_{s=0} \qquad \forall v \in L^\infty([0,1],\mathbb{R}^r)\,.$$

Since F_1^{-1} is a diffeomorphism, there exists also a vector $\eta \neq 0$ such that

$$\eta \perp \frac{d}{ds}\big(\mathrm{End}(h+sv)\big)\Big|_{s=0} = d\mathrm{End}(h)[v] \quad \forall v \in L^\infty([0,1],\mathbb{R}^r)\,,$$

where $d\mathrm{End}(h)[v]$ denotes the differential of End at h in direction v. In particular, the image of $d\mathrm{End}(h)$ does not contain the vector η; equivalently, h is a point where the differential of the endpoint map is not surjective.

We have proved that (the controls associated with) abnormal extremals are singular points of End; the following exercise shows that the converse is also true.

Exercise 3.19. Prove that, if the differential of the endpoint map End is not surjective at some controls h associated with an horizontal curve γ, then γ is an abnormal extremal.

As already pointed out in Remark 3.8, an extremal might be normal and abnormal at the same time. By Proposition 3.12 any minimizer/extremal is C^∞ smooth unless it is strictly abnormal; hence, the relevant curves in the regularity problem for length minimizers are precisely the strictly abnormal ones. For such minimizers, a further necessary condition, the so-called *Goh condition*, can be proved.

Theorem 3.20 (Goh condition)**.** *Let $\gamma : [0,1] \to \mathbb{R}^n$ be a strictly abnormal length minimizer. Then, there exists an associated dual curve ξ that satisfies*

$$\langle \xi, [X_i, X_j](\gamma)\rangle = 0 \text{ on } [0,1] \quad \text{for any } i,j = 1,\dots,r. \tag{3.10}$$

We refer to [4, Chapter 20] for the proof of Theorem 3.20. The proof is in the spirit of Remark 3.4: if (3.10) does not hold for any dual curve ξ, then a suitable open mapping theorem allows to conclude that a certain mapping of endpoint-type is open at γ, contradicting its minimality.

Corollary 3.21. *If the horizontal distribution $X_1,\dots,X_r$ is of step 2, i.e., if*

$$\text{dim span}\,\{X_i, [X_i, X_j] : i,j \in \{1,\dots,r\}\}(x) = n \quad \forall x \in \mathbb{R}^n\,,$$

then any length minimizer is C^∞ smooth.

Proof. Assume by contradiction that there exists a length minimizer γ that is not of class C^∞; then, by Proposition 3.12, γ is strictly abnormal. By (3.9) and Theorem 3.20, there exists a dual variable ξ that is orthogonal (at points of γ) to X_i and $[X_i, X_j]$ for any $i,j \in \{1,\dots,r\}$. Since, by assumption, these elements generates all the tangent space at any point, then we have necessarily $\xi \equiv 0$, which contradicts Theorem 3.6 (i). □

We stress the fact that the minimality assumption is crucial in Theorem 3.20. In general, (3.10) might not hold for strictly abnormal extremals, with the following remarkable exception concerning general abnormal extremals in structures with rank 2.

Remark 3.22. If the horizontal distribution has rank $r = 2$, then any abnormal extremal $\gamma : [0,1] \to \mathbb{R}^n$ and any associated dual curve ξ satisfy

$$\langle \xi(t), [X_1, X_2](\gamma(t))\rangle = 0 \quad \forall t \in [0,1] \tag{3.11}$$

Let us prove (3.11); we claim that it is enough to show that

$$\langle \xi(t), [X_1, X_2](\gamma(t))\rangle = 0 \quad \text{for a.e. } t \in [0,1] \text{ such that } \dot\gamma(t) \text{ exists and } \dot\gamma(t) \neq 0\,. \tag{3.12}$$

Indeed, (3.12) and the continuity of ξ imply (3.11) for any γ such that $\dot\gamma \neq 0$ a.e. on $[0,1]$; for instance, whenever γ is parametrized by constant speed. For different parametrizations of γ, it is enough to reason as in Exercise 3.17.

Let us prove (3.12). By equation (iii) in Theorem 3.6 and the abnormality of γ, we get

$$\begin{aligned}
0 &= \frac{d}{dt}\langle \xi, X_1(\gamma)\rangle \\
&= -\langle (h_1\, JX_1(\gamma) + h_2\, JX_2(\gamma))^T \xi, X_1(\gamma)\rangle + \langle \xi, JX_1(\gamma)[\dot\gamma]\rangle \\
&= -\langle \xi, (h_1\, JX_1(\gamma) + h_2\, JX_2(\gamma))[X_1(\gamma)]\rangle + \langle \xi, JX_1(\gamma)[h_1 X_1(\gamma) + h_2 X_2(\gamma)]\rangle \\
&= h_2\langle \xi, -JX_2(\gamma)[X_1(\gamma)] + JX_1(\gamma)[X_2(\gamma)]\rangle \\
&= -h_2\langle \xi, [X_1, X_2](\gamma)\rangle \qquad \text{a.e. on } [0,1].
\end{aligned}$$

With similar computations one gets

$$0 = \frac{d}{dt}\langle \xi, X_2(\gamma)\rangle = h_1\langle \xi, [X_1, X_2](\gamma)\rangle \quad \text{a.e. on } [0,1].$$

In particular, if $t \in [0,1]$ is such that $\dot\gamma(t) \neq 0$, then either $h_1(t) \neq 0$ or $h_2(t) \neq 0$, and this is enough to obtain (3.12).

As done before, for notational convenience we write $\xi \perp (\Delta \cup [\Delta, \Delta])_\gamma$ whenever the Goh condition holds for the couple (γ, ξ), to mean that the dual variable ξ is orthogonal to both horizontal vectors and brackets of horizontal vector fields. With a slight change of notation, we could also introduce the time-dependent 1-form

$$\xi^*(t) := \xi_1(t)dx_1 + \cdots + \xi_n(t)dx_n \tag{3.13}$$

and write $\xi^* \in \Delta_\gamma^\perp$ (for abnormal extremals) or $\xi^* \in \Delta_\gamma^\perp \cap [\Delta, \Delta]_\gamma^\perp$ (when the Goh condition is in force). The 1-form ξ^* is going to appear again later in these notes.

Remark 3.23. Another important fact about abnormal minimizers has been proved in [1] (see also [34]) in connection with the smoothness problem for the CC distance d: if the horizontal vectors $X_1, \dots, X_r$ are analytic, then the set Σ of point in $\mathbb{R}^n$ that can be connected to the origin (or to any other base point) with abnormal length minimizers is a closed set with empty interior. An important open question is the following Morse-Sard problem for the endpoint map: does Σ have measure zero? We refer to [27, Section 10.2] and to the recent preprint [2] for more detailed discussions on this and other topics.

3.5. **An interesting family of extremals.** An interesting sub-Riemannian structure was proposed by A. Agrachev and J. P. Gauthier during the meeting “Geometric control and sub-Riemannian geometry" held in Cortona in May 2012. Consider the CC structure of rank 2 induced on $\mathbb{R}^4$ by the vector fields

$$X_1(x) := \partial_1 + 2x_2\partial_3 + x_3^2\partial_4, \quad X_2(x) := \partial_2 - 2x_1\partial_3 .$$

It can be easily checked that the bracket-generating condition holds and that, for any $\alpha \in \mathbb{R}$, the curve $\gamma^\alpha(t) := (t, \alpha|t|, 0, 0)$, $t \in \mathbb{R}$, is a strictly abnormal extremal with dual curve $\xi(t) = (0, 0, 0, 1)$. It is fairly easy to show that γ^0 is a minimizer; R. Monti has proved with a cutting-the-corner technique that γ^α is not a minimizer when $\alpha \notin \{0, 1, -1\}$. Using a different and much simpler argument, the remaining case $\alpha = \pm 1$ was recently settled in [21], where it is also proved that all length minimizers in the CC structure under consideration are smooth.

Exercise 3.24. Prove that γ^0 is uniquely length minimizing.

4. Carnot groups

4.1. **Stratified groups.** In this section we are going to describe a few basic facts on stratified groups. Recall that the Lie algebra $\mathfrak{g}$ associated with a Lie group $\mathbb{G}$ is defined as the Lie algebra of left-invariant vector fields on $\mathbb{G}$. A vector field X on $\mathbb{G}$ is said to be *left-invariant* if

$$X(p) = d\ell_p(X(0)) \quad \forall p \in \mathbb{G} ,$$

where $d\ell_p$ denotes the differential of the left-translation $\ell_p(z) = p \cdot z$ by p, $\cdot$ denotes the group product and 0 denotes the identity of $\mathbb{G}$. Equivalently, X is left-invariant if $(Xf)(\ell_p(x)) = X(f \circ \ell_p)(x)$ for any $p, x \in \mathbb{G}$ and any $f \in C^\infty(\mathbb{G})$.

Definition 4.1. A *stratified group* $\mathbb{G}$ is a connected, simply connected and nilpotent Lie group whose Lie algebra $\mathfrak{g}$ admits a *stratification*, i.e., a decomposition

$$\mathfrak{g} = V_1 \oplus V_2 \oplus \cdots \oplus V_s$$

with the properties that $V_i = [V_1, V_{i-1}]$ for any $i = 2, \ldots, s$ and $[V_1, V_s] = \{0\}$.

A few comments are in order:

- the Lie algebra $\mathfrak{g}$ is nilpotent of *step* s;
- one can easily see that $[V_i, V_j] \subset V_{i+j}$ for any $i, j \geqslant 1$ such that $i + j \leqslant s$;
- if $i + j \geqslant s + 1$, then $[V_i, V_j] = \{0\}$.

Moreover, the exponential map $\exp : \mathfrak{g} \to \mathbb{G}$ induces a diffeomorphism between $\mathbb{G}$ and $\mathbb{R}^n \equiv \mathfrak{g}$, n being the dimension of $\mathfrak{g}$. However, in the sequel we will identify $\mathbb{G}$ with $\mathbb{R}^n$ by means of a different set of coordinates, the so-called exponential coordinates of the second-type (see (4.1) and (4.2) below).

Let us fix an *adapted basis* of $\mathfrak{g}$, i.e., a basis $X_1, \ldots, X_n$ whose order is coherent with the stratification:

$$\underbrace{X_1, \ldots, X_r}_{\text{basis of } V_1}, \underbrace{X_{r+1}, \ldots, X_{r_2}}_{\text{basis of } V_2}, \underbrace{X_{r_2+1}, \ldots}_{\text{basis of } V_3} \cdots\cdots \underbrace{\ldots, X_n}_{\text{basis of } V_s} .$$

The integer $r_2 := \dim V_1 + \dim V_2$ will be used also in the sequel. We can then identify $\mathbb{G}$ with $\mathbb{R}^n$ by introducing *exponential coordinates of the second type*

$$\mathbb{R}^n \ni (x_1, \dots, x_n) \longleftrightarrow \exp(x_n X_n) \cdot \exp(x_{n-1} X_{n-1}) \cdots \exp(x_1 X_1) \in \mathbb{G} \tag{4.1}$$

or, equivalently, by using flows of vector fields

$$\mathbb{R}^n \ni (x_1, \dots, x_n) \longleftrightarrow e^{x_1 X_1} \circ \cdots \circ e^{x_{n-1} X_{n-1}} \circ e^{x_n X_n}(0) \in \mathbb{G}\,. \tag{4.2}$$

As a matter of fact (see e.g. [20]), one can prove that in these coordinates

$$\begin{aligned} &X_1 = \partial_1 \\ &X_i(x) = \partial_i + \textstyle\sum_{j=r+1}^n f_{ij}(x)\partial_j \quad \forall i = 2, \dots, r \end{aligned} \tag{4.3}$$

for suitable analytic functions $f_{ij} : \mathbb{R}^n \to \mathbb{R}$.

The stratification of $\mathfrak{g}$ allows to define a family of intrinsic dilations on $\mathbb{G}$. For any $i = 1, \dots, r$, let us define its *degree* $d(i) \in \{1, \dots, s\}$ by

$$d(i) = k \iff X_i \in V_k\,.$$

One can define a one-parameter family of dilations on $\mathfrak{g}$ in the following way. For any $r > 0$, let $\delta_r : \mathfrak{g} \to \mathfrak{g}$ be the unique linear map such that

$$\delta_r(X_i) = r^{d(i)} X_i.$$

Then, by the stratification assumption, δ_r is a Lie algebra isomorphism. One can also define dilations on the group (in coordinates) by

$$\delta_r(x_1, \dots, x_n) := (rx_1, \dots, r^{d(i)} x_i, \dots, r^s x_n)\,.$$

Clearly, $\delta_r : \mathbb{G} \to \mathbb{G}$ defines a one-parameter family of group isomorphisms.

Example 4.2. The Heisenberg group (see also Section 3.1, where it is presented in a different set of coordinates) is the stratified group associated with the Lie algebra of step 2 $\mathfrak{g} := V_1 \oplus V_2$, where $V_1 = \operatorname{span}\{X_1, X_2\}$, $V_2 = \operatorname{span}\{X_3\}$ and

$$[X_2, X_1] = X_3, \quad [X_3, X_1] = [X_3, X_2] = 0\,.$$

The Heisenberg group can be represented in exponential coordinates of the second type as $\mathbb{R}^3$ with

$$X_1 = \partial_1, \quad X_2 = \partial_2 - x_1\partial_3, \quad X_3 = \partial_3\,.$$

Group dilations read as $\delta_r(x_1, x_2, x_3) = (rx_1, rx_2, r^2 x_3)$.

Example 4.3. The *Engel group* is the stratified group associated with the Lie algebra of step 3 $\mathfrak{g} := V_1 \oplus V_2 \oplus V_3$, where $V_1 = \operatorname{span}\{X_1, X_2\}$, $V_2 = \operatorname{span}\{X_3\}$, $V_3 = \operatorname{span}\{X_4\}$ and

$$[X_2, X_1] = X_3, \quad [X_3, X_1] = X_4, \quad [X_3, X_2] = [X_4, X_1] = [X_4, X_2] = [X_4, X_3] = 0\,.$$

The Engel group can be represented in exponential coordinates of the second type as $\mathbb{R}^4$ with

$$X_1 = \partial_1, \quad X_2 = \partial_2 - x_1\partial_3 + \frac{x_1^2}{2}\partial_4, \quad X_3 = \partial_3 - x_1\partial_4, \quad X_4 = \partial_4\,.$$

Group dilations read as $\delta_r(x_1,x_2,x_3,x_4)=(rx_1,rx_2,r^2x_3,r^3x_4)$.

4.2. **Carnot groups.** Stratified groups can be endowed with a canonical CC structure induced by a basis $X_1,\dots,X_r$ of the first layer V_1. Notice that the horizontal sub-bundle $\Delta:=V_1$ is left-invariant and bracket-generating (by the stratification assumption), hence the CC distance d is well defined. We refer to [17] for a metric characterization of Carnot groups and to [18] for an introduction to sub-Riemannian geometry on groups.

Exercise 4.4. Prove that, for any $p,x,y\in\mathbb{G}$ and any $r>0$, there holds

$$d(p\cdot x,p\cdot y)=d(x,y)\quad\text{and}\quad d(\delta_r x,\delta_r y)=rd(x,y)\,.$$

Exercise 4.5. Prove that the horizontal curve $\gamma(t)=(0,t,0,0)$ in the Engel group (represented in the coordinates of Example 4.3) is an extremal that is normal and abnormal at the same time.

Our interest in Carnot groups is motivated by the well-known fact that the tangent metric space (in the Gromov-Hausdorff sense) to a CC space at a "generic" point is a Carnot group: roughly speaking, Carnot groups are the infinitesimal models of CC spaces. See e.g. [25, 24, 6].

4.3. **The dual curve and extremal polynomials.** Let $\gamma:[0,1]\to\mathbb{G}$ be an extremal with associated controls $h\in L^\infty([0,1],\mathbb{R}^r)$ and dual curve $\xi\in\mathrm{Lip}([0,1],\mathbb{R}^n)$; assume that $\gamma(0)=0$. Recall that ξ induces a time-dependent 1-form ξ^* as in (3.13); we are going to write ξ^* in a different system of coordinates for 1-forms.

The group structure allows to define a frame $\theta_1,\dots,\theta_n$ of left-invariant 1-forms, dual to the adapted basis $X_1,\dots,X_n$, by imposing that

$$\theta_i(X_j)=\delta_{ij}\quad\text{on }\mathbb{G}\,,\tag{4.4}$$

δ_{ij} denoting the Kronecker delta. We can therefore define $\lambda\in\mathrm{Lip}([0,1],\mathbb{R}^n)$ by imposing that

$$\xi^*(t)=\xi_1(t)dx_1+\dots+\xi_n(t)dx_n=\big(\lambda_1(t)\theta_1+\dots+\lambda_n(t)\theta_n\big)(\gamma(t))\quad\forall\,t\in[0,1]\,.\tag{4.5}$$

We use the term *dual curve* also for the function λ. One can immediately notice that, by (4.4), statement (iii) in Theorem 3.6 is equivalent to

$$\xi_0h_i+\lambda_i=0\quad\text{a.e. on }[0,1],\qquad\forall\,i=1,\dots,r\,.\tag{4.6}$$

Moreover, the differential equation (iii) of Theorem 3.6 is equivalent to the following ODE for λ (we refer to [19, Theorem 2.6] for details). For any $i=1,\dots,n$, there holds

$$\dot\lambda_i=-\sum_{j=1}^r\sum_{k=1}^n c_{ij}^k h_j\lambda_k\quad\text{a.e. on }[0,1]\,,\tag{4.7}$$

where the constants c_{ij}^k are the *structure constants* of the Lie algebra $\mathfrak{g}$ defined by

$$[X_i,X_j]=\sum_{k=1}^n c_{ij}^k X_k\quad\forall\,i,j=1,\dots,n\,.$$

Exercise 4.6. Prove the implication

$$d(k) \neq d(i) + d(j) \Rightarrow c_{ij}^k = 0 \quad \forall i, j, k = 1, \dots, n\,. \tag{4.8}$$

Hint: recall that $[X_i, X_j] \in V_{d(i)+d(j)}$.

Deduce, as a consequence, that (4.7) is equivalent to

$$\dot\lambda_i = -\sum_{j=1}^{r} \sum_{\substack{k=1,\dots,n\\ d(k)=d(i)+1}} c_{ij}^k h_j \lambda_k \quad \text{a.e. on } [0,1]\,. \tag{4.9}$$

From the technical viewpoint, the main achievement of [19, 20] is an explicit formula for the dual curve λ as a function of γ, see Theorem 4.11 below. This is obtained through the integration of the ODE (4.7), which is in turn based on the following result.

Lemma 4.7. *Let $\gamma : [0,1] \to \mathbb{G}$ be an extremal with $\gamma(0) = 0$; let $h \in L^\infty([0,1], \mathbb{R}^r)$ be the associated controls and $\lambda \in \mathrm{Lip}([0,1], \mathbb{R}^n)$ be its dual curve. Suppose that there exist functions $P_i \in C^1(\mathbb{G}), i = 1, \dots, n$, such that*

$$P_i(0) = \lambda_i(0) \quad \text{and} \quad X_j P_i = \sum_{k=1}^{n} c_{ji}^k P_k \text{ on } \mathbb{G} \tag{4.10}$$

for any $i, j = 1, \dots, n$. Then, for any $i = 1, \dots, n$, there holds

$$\lambda_i(t) = P_i(\gamma(t)) \quad \forall\, t \in [0,1]. \tag{4.11}$$

Proof. The proof is based on a reverse-order inductive argument on i; we start by proving (4.11) for $i = n$. Since $X_n \in V_s$ is in the kernel of $\mathfrak{g}$, we have $[X_j, X_n] = 0$ for any $j = 1, \dots, n$, i.e., $c_{jn}^k = -c_{nj}^k = 0$. In particular, by (4.7) and (4.10)

- $\dot\lambda_n = -\sum_{j=1}^r \sum_{k=1}^n c_{nj}^k h_j \lambda_k = 0$, hence λ_n is constant on $[0,1]$;
- for any $j = 1, \dots, n$, $X_j P_n = \sum_{k=1}^n c_{jn}^k P_k = 0$, hence P_n is constant on $\mathbb{G}$.

Since, by assumption, $P_n(0) = \lambda_n(0)$, we obtain that $\lambda_n(t) = P_n(\gamma(t))$ for any $t \in [0,1]$.

Assume now that $\lambda_k = P_k(\gamma)$ for any $k \geqslant i+1$; recalling that $\dot\gamma = \sum_{j=1}^r h_j X_j(\gamma)$, we have

$$\begin{aligned}\frac{d}{dt}(P_i \circ \gamma) &= \sum_{j=1}^{r} h_j\, X_j P_i(\gamma) = \sum_{j=1}^{r}\sum_{k=1}^{n} h_j c_{ji}^k P_k(\gamma)\\ &\overset{(4.8)}{=} \sum_{j=1}^{r} \sum_{\substack{k=1,\dots,n\\ d(k)=d(i)+1}} h_j c_{ji}^k P_k(\gamma)\,.\end{aligned}$$

We can now use the inductive assumption together with the equality $c_{ji}^k = -c_{ij}^k$ to get

$$\frac{d}{dt}(P_i \circ \gamma) = -\sum_{j=1}^{r} \sum_{\substack{k=1,\dots,n\\ d(k)=d(i)+1}} h_j c_{ij}^k \lambda_k(\gamma) \overset{(4.9)}{=} \dot\lambda_i\,.$$

In particular, the Lipschitz functions λ_i and $P_i \circ \gamma$ have the same derivative and, by assumption, they coincide at time $t = 0$. This is sufficient to conclude the validity of (4.11). □

The integration of the dual variable λ is thus reduced to the search for functions P_i satisfying (4.10); these functions are provided by the *extremal polynomials* introduced below in Definition 4.8. Let us introduce some preliminary notation. Given a multi-index $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ and $x \in \mathbb{R}^n \equiv \mathbb{G}$, we write

$$\begin{aligned} x^\alpha &= x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \\ |\alpha| &= \alpha_1 + \cdots + \alpha_n \\ \alpha! &= \alpha_1!\, \alpha_2! \,\cdots\, \alpha_n!\,. \end{aligned}$$

For the sake of precision: we agree that $0 \in \mathbb{N}$, hence the null multi-index $\alpha = 0$ is admissible. If $x = 0$ and $\alpha = 0$, we agree that $x^\alpha = 1$.

Definition 4.8. For any $v \in \mathbb{R}^n$ and $i = 1, \dots, n$, we define the *extremal polynomial* $P_i^v : \mathbb{G} \to \mathbb{R}$ by

$$P_i^v(x) = \sum_{\alpha \in \mathbb{N}^n} \sum_{k=1}^n \frac{(-1)^{|\alpha|}}{\alpha!}\, c_{i\alpha}^k\, v_k\, x^\alpha\,, \tag{4.12}$$

where the symbols $c_{i\alpha}^k$ denote the *generalized structure constants* of $\mathfrak{g}$ defined by

$$[\cdots[X_i, \underbrace{X_1], X_1], \dots X_1]}_{\alpha_1 \text{ times}}, \underbrace{X_2], \dots X_2]}_{\alpha_2 \text{ times}}, X_3], \dots] \dots] = \sum_{k=1}^n c_{i\alpha}^k X_k\,.$$

Exercise 4.9. Prove that the summation in (4.12) is finite and, more precisely, that P_i^v is a polynomial of both degree and *homogeneous degree* (see e.g. [19, Remark 4.2]) at most $s - d(i)$.

Hint: define $d(\alpha) := \sum_{j=1}^n \alpha_i d(i)$ and prove the implication

$$d(k) \neq d(i) + d(\alpha) \Rightarrow c_{i\alpha}^k = 0\,.$$

As already mentioned, extremal polynomials satisfy (4.10) in Lemma 4.7.

Theorem 4.10. *For any $v \in \mathbb{R}^n$ and $i = 1, \dots, n$, the extremal polynomials satisfy*

$$P_i^v(0) = v_i \quad \textit{and} \quad X_j P_i^v = \sum_{k=1}^n c_{ji}^k P_k^v \ \textit{on } \mathbb{G}. \tag{4.13}$$

While the first equality in (4.13) can be easily checked, the formulae for the derivatives of the P_i^v's are not trivial at all. Their proof is however beyond the scopes of these notes. In the framework of *free* Carnot groups, the second equality in (4.13) was first proved in [19] as a consequence of certain algebraic identities obtained along the proof of [19, Theorem 4.6]. The latter is nothing but Theorem 4.11 below (in the special case of free groups), but its proof follows a completely different line from the one presented here, being based on explicit formulae for the horizontal vector fields (see [12]) rather than on Lemma 4.7. For general Carnot groups, the proof of (4.13) was achieved in [20] with an argument of a differential-geometric flavour involving also non-trivial algebraic identities.

Lemma 4.7 and Theorem 4.10 have the following, immediate consequence.

Theorem 4.11. *Let $\gamma : [0,1] \to \mathbb{G}$ be an extremal with $\gamma(0) = 0$; let $\lambda \in \mathrm{Lip}([0,1], \mathbb{R}^n)$ be an associated dual curve and set $v := \lambda(0) \in \mathbb{R}^n$. Then,*

$$\lambda_i(t) = P_i^v(\gamma(t)) \quad \textit{for any } t \in [0,1] .$$

4.4. **Extremals in Carnot groups.** Theorem 4.11 is our main result from a technical viewpoint. Its consequences, however, are probably even more interesting; let us start by discussing its implications in the case of normal extremals.

Theorem 4.12 (Characterization of normal extremals in Carnot groups)**.** *Let $\gamma : [0,1] \to \mathbb{G}$ be an horizontal curve with $\gamma(0) = 0$. Then, the following conditions are equivalent:*

(a) *γ is a normal extremal;*
(b) *there exists $v \in \mathbb{R}^n$ such that $\dot\gamma = -\sum_{i=1}^r P_i^v(\gamma) X_i(\gamma)$.*

In particular, the sum $P_1^v(\gamma)^2 + \cdots + P_r^v(\gamma)^2$ is constant on $[0,1]$.

Proof. Let us begin with the implication (a)⇒(b). Let h be the controls associated with γ and λ be the dual variable; set $v := \lambda(0) \in \mathbb{R}^n$. By (4.6) and Theorem 4.11, we have

$$h_i = -\lambda_i = -P_i^v(\gamma) \quad \text{on } [0,1], \qquad \forall i = 1, \dots, n ,$$

and (b) immediately follows. The fact that $P_1^v(\gamma)^2 + \cdots + P_r^v(\gamma)^2$ is constant is equivalent to $|h|$ being constant.

Concerning the implication (b)⇒(a), notice that the controls $h_i = -P_i^v(\gamma)$, together with the functions $\lambda_i := P_i^v(\gamma)$, satisfy (4.6) (with $\xi_0 = 1$) and (4.7), because

$$\begin{aligned} \dot\lambda_i &= \frac{d}{dt}\left(P_i^v(\gamma)\right) \overset{\text{(b)}}{=} -\sum_{j=1}^r P_j^v(\gamma)\, X_j P_i^v(\gamma) \\ &\overset{(4.13)}{=} \sum_{j=1}^r \sum_{k=1}^n (-P_j^v(\gamma)) c_{ji}^k P_k^v(\gamma) = -\sum_{j=1}^r \sum_{k=1}^n c_{ij}^k\, h_j\, \lambda_k . \end{aligned}$$

This proves that γ is a normal extremal with dual curve λ, as desired. □

Theorem 4.12 characterizes normal extremals as solutions to a certain ODE: notice that we have reduced the $2n$-variables Hamiltonian system of Exercise 3.10 to a system of ODEs in n variables.

Recalling that left-invariant vector fields are analytic, by Theorem 4.12 one can improve Proposition 3.12 on the regularity of normal extremals.

Corollary 4.13. *Let $\gamma : [0,1] \to \mathbb{G}$ be a normal extremal; then, γ is analytic regular.*

Let us now examine the case of abnormal extremals. If λ is the dual curve associated with an abnormal extremal γ, then (4.6) and Theorem 4.11 imply that

$$\lambda_i = P_i^v(\gamma) = 0 \quad \text{on } [0,1] \qquad \forall i = 1, \dots, r,$$

provided $v := \lambda(0) \in \mathbb{R}^n$. Moreover, by (4.4), (4.5) and the fact that the basis $X_1, \dots, X_n$ is adapted to the stratification, the Goh condition (3.10) is equivalent to

$$\lambda_i = P_i^v(\gamma) = 0 \text{ on } [0,1] \quad \text{for any } i = r+1, \dots, r_2\,.$$

Recall that the integer r_2 has been defined as $\dim V_1 + \dim V_2$. We have therefore

Theorem 4.14 (Characterization of abnormal extremals in Carnot groups)**.** *Let $\gamma : [0,1] \to \mathbb{G}$ be an horizontal curve with $\gamma(0) = 0$. Then, the following conditions are equivalent:*

(a) *γ is an abnormal extremal;*
(b) *there exists $v \in \mathbb{R}^n \setminus \{0\}$ such that $P_1^v(\gamma) = \cdots = P_r^v(\gamma) = 0$.*

Moreover, the Goh condition (3.10) *holds if and only if $P_{r+1}^v(\gamma) = \cdots = P_{r_2}^v(\gamma) = 0$.*

The proof is left as an exercise to the reader, who will also notice that the parameter $v \in \mathbb{R}^n$ is equal to $\lambda(0)$, which is not zero due to Theorem 3.6 (i).

Remark 4.15. The fact that $v \neq 0$ implies that there exist at least one index $i \in \{1, \dots, r\}$ and another index $j \in \{r+1, \dots, r_2\}$ such that neither P_i^v nor P_j^v are the null polynomial; see [20, Proposition 2.6] for more details. In particular, any abnormal extremal γ belongs to an algebraic variety (the one defined by the equalities in Theorem 4.14 (b)) that is not trivial.

The characterization of abnormal extremals in Carnot groups allows for several applications; here, we are going to recall a few of those presented in [19] and [20].

It is possible to construct very irregular abnormal extremals satisfying also the Goh condition. For instance, there exists a 32-dimensional Carnot group $\mathbb{G}$ such that, for *any* Lipschitz function $\phi : [0,1] \to \mathbb{G}$, there exists a Goh abnormal extremal of the form

$$\gamma(t) = (t^2, t, \phi(t), *, \dots, *)\,.$$

See [19, Section 6.4] for more details. In the same spirit, a "spiral-like" abnormal Goh extremal has been provided in [20, Section 5]. These examples somehow suggest that a finer analysis of necessary conditions is needed if one aims at proving smoothness of minimizers, since even second-order necessary conditions (the Goh one) are not enough to ensure regularity.

W. Liu and H. J. Sussman have proved in [23] that, if γ is an abnormal extremal in a CC structure of rank $r = 2$ with dual curve ξ satisfying

$$\xi(t) \not\perp [[\Delta, \Delta], \Delta]_{\gamma(t)} \quad \text{for any } t \in [0,1]\,,$$

then γ is smooth. Abnormal extremals satisfying the previous condition are called *regular abnormal* and are somehow "generic"; let us recall that the Goh condition $\xi \perp (\Delta \cup [\Delta, \Delta])_\gamma$ holds for abnormal extremals because of Remark 3.22. When working in Carnot groups of rank 2, the regularity of such extremals can be proved in a plain way by using extremal polynomials, see [19, Section 6.2]. The results in [23] are anyway much finer, as they show (in a more general framework) that regular abnormal extremals are also locally minimizing.

In the following sections we analyze with more details two further applications of our machinery.

5. Minimizers in step 3 Carnot group

In this section, we review the proof given in [19, Section 6.1] of the following result, that was first proved by K. Tan and X. Yang in [38].

Theorem 5.1. *Any minimizer in a Carnot group of step 3 is C^∞ smooth.*

Proof. By contradiction, assume that there exists a length minimizing curve $\gamma : [0,1] \to \mathbb{G}$ that is not of class C^∞; then, γ is a strictly abnormal minimizer and satisfies the Goh condition. By left invariance, we can assume that $\gamma(0) = 0$. By Theorem 4.14 and Remark 4.15, there exist $v \in \mathbb{R}^n \setminus \{0\}$ and $j \in \{r+1, \dots, r_2\}$ such that P^v_j is not the null polynomial and

$$P^v_j(\gamma) = 0 \quad \text{on } [0,1]. \tag{5.1}$$

By Exercise 4.9, P^v_j has homogeneous degree at most 1, hence there exists $(a_1, \dots, a_r) \in \mathbb{R}^r \setminus \{0\}$ such that

$$P^v_j(x) = a_1 x_1 + \dots + a_r x_r \,, \tag{5.2}$$

where we have also used the fact that $P^v_j(0) = 0$. Define the left-invariant horizontal vector field $Y_1 := a_1 X_1 + \dots + a_r X_r$ and complete it to a basis $Y_1, \dots, Y_r$ of V_1. Using (4.3), (5.1) and (5.2), we obtain that $\dot\gamma$ is of the form

$$\dot\gamma = h_2 Y_2(\gamma) + \dots + h_r Y_r(\gamma)\,.$$

Hence, γ is contained in the subgroup of $\mathbb{G}$ associated with the Lie subalgebra of $\mathfrak{g}$ generated by $Y_2, \dots, Y_r$ and, in particular, it is contained in a Carnot group of rank $r-1$ and step (at most) 3. An easy argument by induction on the rank of the group allows to conclude. □

6. On the negligibility of the abnormal set

In this Section we review the results contained in [20, Section 4]; to this end, we have to introduce some preliminary notions.

The *Tanaka prolongation* $\operatorname{Prol}\mathfrak{g}$ of a stratified Lie algebra $\mathfrak{g} = V_1 \oplus \dots \oplus V_s$ is the largest stratified Lie algebra which can be written in the form

$$\operatorname{Prol}\mathfrak{g} = \dots \oplus V_{-1} \oplus V_0 \oplus V_1 \oplus \dots \oplus V_s$$

and with the property that $[V_i, V_j] \subset V_{i+j}$ for any $i \leqslant s, j \leqslant s$. Here, “largest” means that any other extension of $\mathfrak{g}$ with these properties is (isomorphic to) a sub-algebra of $\operatorname{Prol}\mathfrak{g}$. The explicit construction of $\operatorname{Prol}\mathfrak{g}$ was provided by N. Tanaka in [39]. The prolongation is never trivial, in the sense that $\operatorname{Prol}\mathfrak{g} \neq \mathfrak{g}$; indeed, it can be proved that $\dim V_0 \geqslant 1$. Notice that the number of layers in $\operatorname{Prol}\mathfrak{g}$ in not necessarily finite; when $\operatorname{Prol}\mathfrak{g}$ is infinite dimensional we say that $\mathbb{G}$ is *nonrigid*.

Let $X_1, \ldots, X_n$ be an adapted basis of $\mathfrak{g}$; let us extend it to an adapted basis of $\operatorname{Prol}\mathfrak{g}$

$$\cdots, \underbrace{\ldots, X_{-j}, \ldots}_{\text{basis of } V_{-1}}, \underbrace{\ldots, X_{-1}, X_0}_{\text{basis of } V_0}, \underbrace{X_1, \ldots, X_r}_{\text{basis of } V_1}, \ldots\ldots, \underbrace{\ldots, X_n}_{\text{basis of } V_s}.$$

With a slight abuse of notation, we denote this basis by $(X_i)_{i \leqslant n}$, where the notation "$i \leqslant n$" means

- either $-\infty < i \leqslant n$, if $\dim \operatorname{Prol}\mathfrak{g} = \infty$;
- or $-m \leqslant i \leqslant n$, if $m \in \mathbb{N}$ is such that $\dim \operatorname{Prol}\mathfrak{g} = m + n + 1$.

We will adopt a similar convention for notations like "$i \leqslant r$" and "$i \leqslant 0$".

The Lie algebra $\operatorname{Prol}\mathfrak{g}$ possesses its own structure constants and generalized structure constants. We still denote these constants (that are defined for $i, j, k \leqslant n$ and $\alpha \in \mathbb{N}^n$) by c_{ij}^k and $c_{i\alpha}^k$ because they clearly coincide with those of $\mathfrak{g}$ when $1 \leqslant i, j, k \leqslant n$. Hence, as in Definition 4.8, for any fixed $v \in \mathbb{R}^n$ and any $i \leqslant n$ (i.e., also for $i \leqslant 0$) one can define the extremal polynomial

$$P_i^v(x) := \sum_{\alpha \in \mathbb{N}^n} \sum_{k=1}^{n} \frac{(-1)^{|\alpha|}}{\alpha!} c_{i\alpha}^k \, v_k \, x^\alpha, \quad x \in \mathbb{G}, \tag{6.1}$$

where we agree that $v_k = 0$ whenever $k \leqslant 0$. As in Theorem 4.10, one can prove that

$$\begin{aligned} &P_i^v(0) = v_i \text{ for any } i \leqslant n \\ &X_j P_i^v = \sum_{k \leqslant n} c_{ji}^k P_k^v \quad \text{for any } i \leqslant n, 1 \leqslant j \leqslant n. \end{aligned} \tag{6.2}$$

These formulae are key tools in the proof of the following result; see [20] for more details.

Theorem 6.1. *Let $\gamma : [0,1] \to \mathbb{G}$ be an abnormal extremal with $\gamma(0) = 0$. Then, there exists $v \in \mathbb{R}^n$ such that*

$$P_i^v(\gamma) = 0 \text{ on } [0,1] \qquad \text{for any } i \leqslant r\,. \tag{6.3}$$

If the Goh condition holds, then the previous formula holds for any $i \leqslant r_2$.

Proof. Given the formulae (6.2), the proof is quite elementary and similar to that of Lemma 4.7. We prove (6.3) by reverse induction on the homogeneous degree[1] $d(i)$ of i. We set again $v := \lambda(0)$, λ being the dual curve associated with γ.

The base of the induction is the case $d(i) = 1$, where (6.3) holds by Theorem 4.14.

[1]Clearly, the homogeneous degree is defined by $d(i) = k \Leftrightarrow X_i \in V_k$ also for $i \leqslant 0$.

Assume then that $P^v_k(\gamma) \equiv 0$ for any k such that $d(i) < d(k) \leqslant 1$. Let $h \in L^\infty([0,1],\mathbb{R}^r)$ be the controls associated with γ, so that $\dot\gamma = \sum_{j=1}^r h_j X_j(\gamma)$. Then

$$\begin{aligned}\frac{d}{dt} P^v_i \circ \gamma &= \sum_{j=1}^r h_j\, X_j P^v_i(\gamma) \overset{(6.2)}{=} \sum_{j=1}^r \sum_{k\leqslant n} h_j c^k_{ji} P^v_k(\gamma)\\ &= \sum_{j=1}^r \sum_{\substack{k\leqslant n\\ d(k)=d(i)+1}} h_j c^k_{ji} P^v_k(\gamma) = 0\,,\end{aligned}$$

i.e., $P^v_i(\gamma)$ is constant and equal to $P^v_i(\gamma(0)) = 0$. □

Theorem 4.14 states that abnormal extremals in Carnot groups are contained in certain algebraic varieties (of a very specific type). Theorem 6.1 improves it because it states that these algebraic varieties can be made smaller, as there are more polynomials (than in Theorem 4.14) that vanish along γ.

We show an application of our techniques to the Morse-Sard problem for abnormal extremals. In our opinion, the strategy we follow has chances to be adapted to many Carnot groups; however, we present it only in a specific group.

Let us consider the free[2] Carnot group $\mathbb{G}$ of rank 2 and step 4, i.e., the group associated with the stratified Lie algebra

$$\mathfrak{g} = V_1 \oplus V_2 \oplus V_3 \oplus V_4$$

with

$$\begin{aligned} V_1 &= \mathrm{span}\{X_1, X_2\}, \quad & V_2 &= \mathrm{span}\{X_3\},\\ V_3 &= \mathrm{span}\{X_4, X_5\}, \quad & V_4 &= \mathrm{span}\{X_6, X_7, X_8\}\end{aligned}$$

and commutation relations

$$\begin{aligned} &[X_2, X_1] = X_3\\ &[X_3, X_1] = X_4,\ [X_3, X_2] = X_5\\ &[X_4, X_1] = X_6,\ [X_4, X_2] = [X_5, X_1] = X_7,\ [X_5, X_2] = X_8\,.\end{aligned}$$

Using exponential coordinates of the second type (see [12]), $\mathbb{G}$ can be identified with $\mathbb{R}^8$ in such a way that

$$\begin{aligned} X_1 &= \partial_1\\ X_2 &= \partial_2 - x_1\partial_3 + \frac{x_1^2}{2}\partial_4 + x_1x_2\partial_5 - \frac{x_1^3}{6}\partial_6 - \frac{x_1^2x_2}{2}\partial_7 - \frac{x_1x_2^2}{2}\partial_8\,.\end{aligned}$$

We are going to prove the following result.

[2] *Free* means, roughly speaking, that it is the Carnot group with largest dimension among those with rank 4 and step 2; equivalently, that any other such group is (isomorphic to) a quotient of the free one.

Theorem 6.2. *Let $\mathbb{G} \equiv \mathbb{R}^8$ be the free Carnot group of rank 2 and step 4. Then, there exists a non-zero polynomial in 8 variables $Q : \mathbb{R}^8 \to \mathbb{R}$ such that the following holds: if $p \in \mathbb{G}$ is the endpoint of an abnormal extremal starting from 0, then $Q(p) = 0$.*

In particular, the set of points in $\mathbb{G}$ that can be connected to the origin with abnormal extremals is contained in the algebraic variety $\{x \in \mathbb{R}^8 : Q(x) = 0\}$ and has measure zero.

Remark 6.3. Theorems 4.14 and 6.1 show that any abnormal extremal is contained in an algebraic variety whose definition depends on a parameter v, i.e., on the extremal itself. On the contrary, by Theorem 6.2 there exists a *universal* algebraic variety containing *all* abnormal extremals.

Proof. As proved in [41], the Tanaka prolongation of $\mathbb{G}$ is of the form $\operatorname{Prol}\mathfrak{g} = V_0 \oplus \mathfrak{g}$ with $\dim V_0 = 4$. Let us extend $X_1, \dots, X_8$ to an adapted basis $\{X_i\}_{-3 \leqslant i \leqslant 8}$ of $\operatorname{Prol}\mathfrak{g}$. By Theorem 6.1 we know that, for any abnormal extremal $\gamma : [0,1] \to \mathbb{G}$ with $\gamma(0) = 0$, there exists $v \in \mathbb{R}^8$ such that

$$P_i^v(\gamma) = 0 \text{ on } [0,1] \quad \text{for any } i = -3, \dots, 3\,. \tag{6.4}$$

We have also used Remark 3.22, i.e., the fact that the Goh condition holds. In particular,

$$v_i = P_i^v(0) = 0 \qquad \text{for } i = 1, 2, 3.$$

Therefore, recalling (6.1), any P_i^v can be written in the form

$$P_i^v(x) = \sum_{k=4}^{8} v_k Q_{ik}(x), \quad i = -3, \dots, 8$$

for suitable polynomials $Q_{ik}(x)$ that are independent from v. For any $i = -3, \dots, 3$, let us define the map $Q_i : \mathbb{G} \to \mathbb{R}^5$ by

$$Q_i(x) = \big(Q_{i4}(x), Q_{i5}(x), Q_{i6}(x), Q_{i7}(x), Q_{i8}(x)\big),$$

so that $P_i^v(x) = \langle (v_4, \dots, v_8), Q_i(x) \rangle$. Hence, (6.4) can be rewritten as

$$Q_i(\gamma(t)) \perp (v_4, \dots, v_8) \quad \forall t \in [0,1],\ \forall\, i = -3, \dots, 3\,.$$

In particular, for any $t \in [0,1]$, the seven 5-dimensional vectors $Q_i(\gamma(t)), -3 \leqslant i \leqslant 3$, belong to the vector space $(v_4, \dots, v_8)^\perp \subset \mathbb{R}^5$; this vector space has dimension 4 because $(v_4, \dots, v_8) \neq 0$ due to Theorem 3.6 (i). Hence, any 5 of these 7 vectors are linearly dependent, i.e., any 5×5 minor of the 5×7 matrix

$$(Q_{ik}(x))_{\substack{-3 \leqslant i \leqslant 3 \\ 4 \leqslant k \leqslant 8}} = \operatorname{col}\big[Q_{-3}|Q_{-2}| \cdots |Q_3\big](x) \tag{6.5}$$

has determinant 0 at any point x on γ. In particular, the determinant of the minor

$$\operatorname{col}\big[Q_{-1}|Q_0|Q_1|Q_2|Q_3\big](x)$$

is a polynomial $Q(x)$ (independent from v) which vanish along γ.

It is now a boring task to prove that Q is not the null polynomial; we refer to the proof of [20, Theorem 4.1] for details. This concludes the proof. ☐

Remark 6.4. The determinant of *any* 5×5 minor of the matrix in (6.5) has to vanish along abnormal extremals; hence, in principle, one could produce $\binom{7}{5} = 21$ polynomials as in the statement of Theorem 6.2. See [20, Remark 4.2.] for a more detailed discussion on these and other considerations.

References

[1] A. A. Agrachev, *Any sub-Riemannian metric has points of smoothness* (Russian). Dokl. Akad. Nauk 424 (2009), no. 3, 295–298; translation in Dokl. Math. 79 (2009), no. 1, 45–47.

[2] A. A. Agrachev, *Some open problems.* In *Geometric Control Theory and Sub-Riemannian Geometry*, Springer INdAM Series 5 (2014), 1–14.

[3] A. A. Agrachev, D. Barilari & U. Boscain, *Introduction to Riemannian and Sub-Riemannian geometry.* Preprint available at www.math.jussieu.fr/ barilari/

[4] A. A. Agrachev & Y. L. Sachkov, *Control theory from the geometric viewpoint.* Encyclopaedia of Mathematical Sciences, 87. Control Theory and Optimization, II. Springer-Verlag, Berlin (2004), xiv+412.

[5] L. Ambrosio & S. Rigot, *Optimal mass transportation in the Heisenberg group.* J. Funct. Anal. 208 (2004), no. 2, 261–301.

[6] A. Bellaïche, *The tangent space in subriemannian geometry.* In *Subriemannian Geometry*, Progress in Mathematics 144, ed. by A. Bellaiche and J. Risler, Birkhauser Verlag, Basel, 1996.

[7] V. G. Boltyanskii, *Sufficient conditions for optimality and the justification of the dynamic programming method.* SIAM J. Control 4 (1966), 326–361.

[8] C. Carathéodory, *Untersuchungen über die Grundlagen der Thermodynamik.* Math. Ann. 67 (1909), 355–386.

[9] Y. Chitour, F. Jean & E. Trélat, *Genericity results for singular curves.* J. Differential Geom. 73 (2006), no. 1, 45–73.

[10] W. L. Chow, *Über Systeme von linearen partiellen Differentialgleichungen erster Ordnung.* Math. Ann. 117 (1939), 98–105.

[11] C. Golé & R. Karidi, *A note on Carnot geodesics in nilpotent Lie groups.* J. Dynam. Control Systems 1 (1995), no. 4, 535–549.

[12] M. Grayson & R. Grossman, *Models for free nilpotent Lie algebras.* J. Algebra 135 (1990), no. 1, 177–191.

[13] U. Hamenstädt, *Some regularity theorems for Carnot-Carathéodory metrics.* J. Differential Geom. 32 (1990), no. 3, 819–850.

[14] L. Hörmander, *Hypoelliptic second-order differential equations.* Acta Math. 121 (1968), 147–171.

[15] E. B. Lee & L. Markus, *Foundations of optimal control theory*, second edition. Robert E. Krieger Publishing Co., Inc., Melbourne, FL, 1986. xii+576 pp.

[16] G. P. Leonardi & R. Monti, *End-point equations and regularity of sub-Riemannian geodesics.* Geom. Funct. Anal. 18 (2008), no. 2, 552–582.

[17] E. Le Donne, *A metric characterization of Carnot groups.* To appear in Proc. Amer. Math. Soc. Preprint available at arxiv.org/abs/1304.7493.

[18] E. Le Donne, *Lecture notes on sub-Riemannian geometry.* Preprint available at sites.google.com/site/enricoledonne/.

[19] E. Le Donne, G. P. Leonardi, R. Monti & D. Vittone, *Extremal curves in nilpotent Lie groups.* Geom. Funct. Anal. 23 (2013), 1371–1401.

[20] E. Le Donne, G. P. Leonardi, R. Monti & D. Vittone, *Extremal polynomials in stratified groups.* Preprint available at cvgmt.sns.it.

[21] E. Le Donne, G. P. Leonardi, R. Monti & D. Vittone, *Corners in non-equiregular sub-Riemannian manifolds*. Preprint available at cvgmt.sns.it.

[22] W. Liu & H. J. Sussman, *Abnormal sub-Riemannian minimizers*. Differential equations, dynamical systems, and control science, 705–716, Lecture Notes in Pure and Appl. Math., 152, Dekker, New York, 1994.

[23] W. Liu & H. J. Sussman, *Shortest paths for sub-Riemannian metrics on rank-two distributions*. Mem. Amer. Math. Soc. 118 (1995), no. 564, x+104 pp.

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

[25] J. Mitchell, *On Carnot-Caratheodory metrics*. J. Diff. Geom. 21 (1985), 35–45.

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

[27] R. Montgomery, *A tour of subriemannian geometries, their geodesics and applications*. Mathematical Surveys and Monographs, 91. American Mathematical Society, Providence, RI, 2002. xx+259.

[28] R. Monti, *A family of nonminimizing abnormal curves*. Ann. Mat. Pura Appl (online). Preprint available at cvgmt.sns.it.

[29] R. Monti, *Regularity results for sub-Riemannian geodesics*. Calc. Var. Partial Differential Equations 49 (2014), no. 1-2, 549–582.

[30] R. Monti, *The regularity problem for sub-Riemannian geodesics*. In *Geometric Control Theory and Sub-Riemannian Geometry*, Springer INdAM Series 5 (2014), 313–332.

[31] D. Morbidelli, *Fractional Sobolev norms and structure of Carnot-Caratheodory balls for Hörmander vector fields*, Studia Math. 139 (2000), no. 3, 213–244.

[32] A. Nagel, E. M. Stein & S. Wainger, *Balls and metrics defined by vector fields I: basic properties*, Acta Math. 155 (1985), 103–147.

[33] P. K. Rashevsky, *Any two points of a totally nonholonomic space may be connected by an admissible line*. Uch. Zap. Ped. Inst. im. Liebknechta, Ser. Phys.Math., 2 (1938) 83–94.

[34] L. Rifford & E. Trélat, *Morse-Sard type results in sub-Riemannian geometry*. Math. Ann. 332 (2005), no. 1, 145–159.

[35] R. Strichartz, *Sub-Riemannian geometry*. J. Differential Geom. 24 (1986), no. 2, 221–263.

[36] R. Strichartz, *Corrections to: "Sub-Riemannian geometry"*. J. Differential Geom. 30 (1989), no. 2, 595–596.

[37] H. J. Sussmann, *A cornucopia of four-dimensional abnormal sub-Riemannian minimizers*. In *Subriemannian Geometry*, Progress in Mathematics 144, ed. by A. Bellaiche and J. Risler, Birkhauser Verlag, Basel, 1996.

[38] K. Tan & X. Yang, *Subriemannian geodesics of Carnot groups of step 3*. ESAIM Control Optim. Calc. Var. 19 (2013), no. 1, 274–287.

[39] N. Tanaka, *On differential systems, graded Lie algebras and pseudogroups*. J. Math. Kyoto Univ. 10 (1970), 1–82.

[40] V. S. Varadarajan, *Lie Groups, Lie Algebras, and Their Representations*. Graduate Texts in Mathematics, Springer-Verlag, New York, 1974.

[41] B. Warhurst, *Tanaka prolongation of free Lie algebras*. Geom. Dedicata, 130 (2007), 59–69.

Davide Vittone: Dipartimento di Matematica, Università di Padova, Via Trieste, 63, 35121 Padova, Italy

E-mail address: vittone@math.unipd.it