EXISTENCE OF TANGENT LINES TO CARNOT-CARATHÉODORY GEODESICS

ROBERTO MONTI, ALESSANDRO PIGATI, AND DAVIDE VITTONE

Abstract. We prove that length minimizing curves in Carnot-Carathéodory spaces possess at any point at least one tangent curve (i.e., a blow-up in the nilpotent approximation) equal to a straight horizontal line. This is the first regularity result for length minimizers that holds with no assumption on either the space (e.g., its rank, step, or analyticity) or the curve.

1. Introduction

Let M be a connected n-dimensional C^∞-smooth manifold and $\mathscr{X} = \{X_1, \dots, X_r\}$, $r \geq 2$, a system of C^∞-smooth vector fields on M satisfying the Hörmander condition. We call the pair $(M, \mathscr{X})$ a *Carnot-Carathéodory (CC) structure* (see Section 2). Given an interval $I \subseteq \mathbb{R}$, a Lipschitz curve $\gamma : I \to M$ is said to be *horizontal* if there exist functions $h_1, \dots, h_r \in L^\infty(I)$ such that for a.e. $t \in I$ we have

$$\dot\gamma(t) = \sum_{i=1}^{r} h_i(t) X_i(\gamma(t)). \tag{1.1}$$

Letting $|h| := (h_1^2 + \dots + h_r^2)^{1/2}$, the length of γ is then defined as

$$L(\gamma) := \inf \left\{ \int_I |h(t)|\, dt \mid h \in L^\infty(I, \mathbb{R}^r) \text{ s.t. (1.1) holds} \right\}.$$

The infimum is attained by a unique $h \in L^\infty(I, \mathbb{R}^r)$: in the sequel, h denotes this minimal control. We will usually assume that curves are parameterized by arclength, i.e., $|h(t)| = 1$ for a.e. t and $\mathscr{L}^1(I) = L(\gamma)$.

Since M is connected, for any pair of points $x, y \in M$ there exists a horizontal curve joining x to y. We can therefore define a distance function $d : M \times M \to [0, \infty)$ letting

$$d(x, y) := \inf \{ L(\gamma) \mid \gamma : [0, T] \to M \text{ horizontal with } \gamma(0) = x \text{ and } \gamma(T) = y \}. \tag{1.2}$$

The resulting metric space (M, d) is a *Carnot-Carathéodory space*. Typical examples of Carnot-Carathéodory spaces are given by sub-Riemannian manifolds $(M, \mathscr{D}, g)$,

2010 *Mathematics Subject Classification.* 53C17, 49K30, 28A75.

Key words and phrases. Length minimizers, Carnot-Carathéodory spaces, sub-Riemannian geometry, Carnot groups.

where $\mathscr{D} \subset TM$ is a completely non-integrable distribution and g is a smooth metric on $\mathscr{D}$.

If the closure of any ball in (M, d) is compact, then the infimum in (1.2) is a minimum, i.e., any pair of points can be connected by a length-minimizing curve. A horizontal curve $\gamma : [0, T] \to M$ is a *length minimizer* if $L(\gamma) = d(\gamma(0), \gamma(T))$. In Carnot-Carathéodory spaces (or even in the model case of *Carnot groups*) it is not known whether constant-speed length minimizers are C^∞-smooth, or even C^1-smooth. The main obstacle is the presence of *abnormal* length minimizers, which are not captured by the natural Hamiltonian framework, see e.g. [13, 2]. In [12], Montgomery gave the first example of such a length minimizer. Contrary to the Riemannian case, stationarity conditions do not guarantee any smoothness of the curve: in [9] it is proved that no further regularity beyond the Lipschitz one can be obtained for abnormal extremals from the Pontryagin Maximum Principle and the Goh condition (which is a second-order necessary condition, see e.g. [2]).

However, some partial regularity results are known. If the *step* is at most 2 (i.e., for any x the tangent space T_xM is spanned by the $r + \binom{r}{2}$ vectors $X_i(x)$, $[X_i, X_j](x)$), then all constant-speed length minimizers are smooth. In the context of Carnot groups, the regularity problem was recently solved also when the step is at most 3 (independently by Tan-Yang in [18] and by Le Donne-Leonardi-Monti-Vittone in [8]). In [17] Sussmann proved that, in presence of analytic data (and in particular in Carnot groups), all length minimizers are *analytic* on a dense open set of times, although it is not known whether this set has full measure. Building on ideas contained in [11, 10], Hakavuori and Le Donne recently proved in [4] that length minimizers cannot have corner-type singularities. Other partial regularity results are contained in [14]. We also refer to [1, 15, 16, 19] for surveys about the known results on the problem.

At any point $x \in M$ the Carnot-Carathéodory structure $(M, \mathscr{X})$ has a *nilpotent approximation* $(M^\infty, \mathscr{X}^\infty)$, which is also a Carnot-Carathéodory structure. The construction, which is recalled in Section 2, uses a one-parameter group of anisotropic dilations $(\delta_\lambda)_{\lambda>0}$ associated with $\mathscr{X}$. In Definition 2.4, given a horizontal curve $\gamma : [-T, T] \to M$, we define the *tangent cone* $\mathrm{Tan}(\gamma; t)$ for $-T < t < T$. This cone contains the horizontal curves in $(M^\infty, \mathscr{X}^\infty)$ which are obtained as a blow-up limit of γ with respect to the dilations $(\delta_\lambda)_{\lambda>0}$ centered at $\gamma(t)$, along some infinitesimal sequence of scales. The manifold M^∞ is also a vector space and we call *horizontal line* a horizontal curve in $(M^\infty, \mathscr{X}^\infty)$ passing through $0 \in M^\infty$ and with *constant* minimal controls $h_1, \ldots, h_r$ (see (1.1)).

The following theorem is the main result of the paper.

Theorem 1.1. Let $\gamma : [-T, T] \to M$ be a length minimizer parametrized by arclength in a Carnot-Carathéodory space (M, d). Then, for any $t \in (-T, T)$, the tangent cone $\mathrm{Tan}(\gamma; t)$ contains a horizontal line.

Theorem 1.1 has a reformulation that does not depend on the construction of $(M^\infty, \mathscr{X}^\infty)$, see Remark 5.3. A version of Theorem 1.1 holds for the extremal points $t = 0$ and $t = T$ of a length minimizer $\gamma : [0, T] \to M$. In this case, the tangent cone contains a horizontal half-line; see Theorem 5.2. These results imply and improve the ones contained in [11, 10, 4]: while in these papers the existence of (linearly independent) left and right derivatives is assumed in order to construct a shorter competitor, Theorem 1.1 provides a mild form of pointwise differentiability which automatically excludes corner-type singularities.

Theorem 1.1 is deduced from a similar result for the case when $M = G$ is a Carnot group of rank $r \geq 2$ and $\mathscr{X} = \{X_1, \dots, X_r\}$ is a system of left-invariant vector fields forming a basis of the first layer of its Lie algebra $\mathfrak{g}$. The reduction to this case relies on the following facts:

(i) if $\kappa \in \mathrm{Tan}(\gamma; t)$ and $\widehat{\kappa} \in \mathrm{Tan}(\kappa; 0)$, then $\widehat{\kappa} \in \mathrm{Tan}(\gamma; t)$, see Proposition 2.8;
(ii) if γ is length-minimizing in $(M, \mathscr{X})$ and $\kappa \in \mathrm{Tan}(\gamma; t)$, then κ is length-minimizing in $(M^\infty, \mathscr{X}^\infty)$, see Proposition 2.7;
(iii) if G is a Carnot group lifting $(M^\infty, \mathscr{X}^\infty)$ and $\bar{\kappa}$ lifts to G a length-minimizing curve κ in $(M^\infty, \mathscr{X}^\infty)$, then $\bar{\kappa}$ is length-minimizing in G, see Proposition 2.11.

The proof of Theorem 1.1 in the case of a Carnot group, in turn, is a consequence of Theorem 1.2 below. Let $\mathfrak{g} = \mathfrak{g}_1 \oplus \cdots \oplus \mathfrak{g}_s$ be the stratification of $\mathfrak{g}$ and let $\langle \cdot, \cdot \rangle$ be the scalar product on $\mathfrak{g}_1$ making $X_1, \dots, X_r$ orthonormal. The integer $s \geq 2$ is the step of the group and $r = \dim \mathfrak{g}_1$ its rank. We denote by $S^{r-1} = \{v \in \mathfrak{g}_1 : \langle v, v \rangle = 1\}$ the unit sphere in $\mathfrak{g}_1$. We define the *excess* of a horizontal curve $\gamma : [-T, T] \to G$ over a Borel set $B \subseteq [-T, T]$ with positive measure as

$$\mathrm{Exc}(\gamma; B) := \inf_{v \in S^{r-1}} \left(\fint_B \langle v, \dot{\gamma}(t) \rangle^2 \, dt \right)^{1/2}.$$

The excess $\mathrm{Exc}(\gamma; B)$ measures how far $\dot{\gamma}_{|B}$ is from being contained in a single hyperplane of $\mathfrak{g}_1$, see Remark 3.2. For length-minimizing curves, the excess is infinitesimal at suitably small scales, as stated in our second main result.

Theorem 1.2. Let G be a Carnot group and let $\gamma : [-T, T] \to G$, $T > 0$, be a length-minimizing curve parametrized by arclength. Then there exists an infinitesimal sequence $\eta_i \downarrow 0$ such that

$$\lim_{i \to \infty} \mathrm{Exc}(\gamma; [-\eta_i, \eta_i]) = 0. \tag{1.3}$$

Again, this result has a version for extremal points: for a length minimizer $\gamma : [0, T] \to G$ the excess $\mathrm{Exc}(\gamma; [0, \eta_i])$ is infinitesimal, see Theorem 5.1. When $r = 2$, (1.3) implies that there exists $\kappa \in \mathrm{Tan}(\gamma; 0)$ of the form $\kappa(t) = \exp(tv)$ for some $v \in \mathfrak{g}_1$. This proves Theorem 1.1 for $M = G$ with $r = 2$. When $r > 2$, the situation can be reduced by induction to the case $r = 2$, using the facts (i) and (ii) above.

The proof of Theorem 1.2 goes by contradiction and uses a cut-and-adjust construction performed in s steps, see Section 5. If we had $\mathrm{Exc}(\gamma; [-\eta, \eta]) \geq \varepsilon$ for some $\varepsilon > 0$ and for all small $\eta > 0$, then we could find $t_1 < \cdots < t_r$ such that, roughly speaking, the vectors $\dot\gamma(t_1), \ldots, \dot\gamma(t_r) \in \mathfrak{g}_1$ are linearly independent in a quantitative way, see Lemma 3.6. We could replace the "horizontal projection" $\underline{\gamma}$ of γ on the interval $[-\eta, \eta]$ with the line segment joining $\underline{\gamma}(-\eta)$ to $\underline{\gamma}(\eta)$, whose gain of length would be estimated in terms of the excess, see Lemma 4.4, and we could lift the resulting "horizontal coordinates" to a horizontal curve in G. The end-point might have changed, but the vectors $\dot\gamma(t_1), \ldots, \dot\gamma(t_r)$ could then be used to build suitable correction devices restoring the end-point, taking care to keep a positive gain of length. This construction is detailed in Sections 3, 4 and 5 and is a refinement of the techniques introduced and developed in [11] and [4].

Acknowledgements. We thank L. Ambrosio for several discussions and for being an invaluable mentor and friend.

2. Nilpotent approximation and tangent cones to curves

In this section, we recall the construction of the nilpotent approximation of a Carnot-Carathéodory structure $(M, \mathscr{X})$, with $\mathscr{X} = \{X_1, \ldots, X_r\}$, and we define the tangent cone to a horizontal curve. We also recall the lifting to a Carnot group.

2.1. Nilpotent approximation of CC structures and horizontal lines.

We denote by $\mathrm{Lie}(X_1, \ldots, X_r)$ the real Lie algebra generated by $X_1, \ldots, X_r$ through iterated commutators. The evaluation of this Lie algebra at a point $x \in M$ is a subspace of the tangent space T_xM. If, for any $x \in M$, we have

$$\mathrm{Lie}(X_1, \ldots, X_r)(x) = T_xM,$$

we say that the system $\mathscr{X} = \{X_1, \ldots, X_r\}$ satisfies the *Hörmander condition* and we call the pair $(M, \mathscr{X})$ a *Carnot-Carathéodory (CC) structure.* If the Hörmander condition holds, then the topology induced by d is the standard one on M.

Let $U \subset M$ be a neighborhood of a fixed point $x_0 \in M$ and let $\varphi \in C^\infty(U; \mathbb{R}^n)$ be a chart such that $\varphi(x_0) = 0$. Then $V := \varphi(U)$ is a neighborhood of $0 \in \mathbb{R}^n$. The system of vector fields $Y_i := \varphi_* X_i$, with $i = 1, \ldots, r$, satisfies the Hörmander condition.

For a multi-index $J = (j_1, \ldots, j_k)$ with $k \geq 1$ and $j_1, \ldots, j_k \in \{1, \ldots, r\}$, define the iterated commutator

$$Y_J := [Y_{j_1}, [\ldots, [Y_{j_{k-1}}, Y_{j_k}] \ldots]].$$

We say that Y_J is a commutator of length $\ell(J) := k$ and we denote by L^j the linear span of $\{Y_J(0) \mid \ell(J) \le j\}$, so that $\{0\} = L^0 \subseteq L^1 \subseteq \cdots \subseteq L^s = \mathbb{R}^n$ for some minimal $s \ge 1$. We select multi-indices $J_1, \dots, J_n$ such that, for each $1 \le j \le s$,

$$\ell(J_{\dim L^{(j-1)}+1}) = \cdots = \ell(J_{\dim L^j}) = j$$

and such that, setting $Z_i := Y_{J_i}$, $Z_1(0), \dots, Z_{\dim L^j}(0)$ form a basis of L^j.

Possibly composing φ with a diffeomorphism, we can assume that for any point $x = (x_1, \dots, x_n) \in V$ we have

$$x = \exp\Big(\sum_{i=1}^n x_i Z_i\Big)(0).$$

This means that $(x_1, \dots, x_n)$ are exponential coordinates of the first kind associated with the frame $Z_1, \dots, Z_n$. To each coordinate x_i we assign the weight $w_i := \ell(J_i)$ and we define the anisotropic dilations $\delta_\lambda : \mathbb{R}^n \to \mathbb{R}^n$

$$\delta_\lambda(x) := (\lambda^{w_1} x_1, \dots, \lambda^{w_n} x_n), \qquad \lambda > 0. \tag{2.4}$$

We will frequently use the homogeneous (pseudo-)norm

$$\|x\| := \sum_{i=1}^n |x_i|^{1/w_i}, \qquad x \in \mathbb{R}^n. \tag{2.5}$$

The following proposition is well-known. See [5, Theorem 2.1] for a proof of the analogous statement for exponential coordinates of the second kind. See also [6] for a general introduction to the nilpotent approximation.

Proposition 2.1. For any $i = 1, \dots, r$ we have $Y_i = a_{i1}\frac{\partial}{\partial x_1} + \ldots + a_{in}\frac{\partial}{\partial x_n}$ for functions $a_{ij} = p_{ij} + r_{ij}$, $j = 1, \dots, n$, such that:

(i) p_{ij} are homogeneous polynomials in $\mathbb{R}^n$ such that $p_{ij}(\delta_\lambda x) = \lambda^{w_j - 1} p_{ij}(x)$ for any $\lambda > 0$ and $x \in \mathbb{R}^n$;

(ii) $r_{ij} \in C^\infty(V)$ are functions such that $\lim\limits_{\lambda \to 0} \lambda^{1-w_j} r_{ij}(\delta_\lambda x) = 0$.

The vector fields $Y_1^\infty, \dots, Y_r^\infty$ in $\mathbb{R}^n$ defined by

$$Y_i^\infty := \sum_{j=1}^n p_{ij}(x) \frac{\partial}{\partial x_j}$$

are known as the *nilpotent approximation* of $Y_1, \dots, Y_r$ at the point 0 and satisfy the Hörmander condition in $\mathbb{R}^n$, see [6, Lemma 2.1]. The pair $(\mathbb{R}^n, \mathscr{X}^\infty)$ with $\mathscr{X}^\infty = \{Y_1^\infty, \dots, Y_r^\infty\}$ is a Carnot-Carathéodory structure. We set $M^\infty := \mathbb{R}^n$ and we call $(M^\infty, \mathscr{X}^\infty)$ a *tangent* Carnot-Carathéodory structure to $(M, \mathscr{X})$ at the point $x_0 \in M$.

The vector fields $Y_1^\infty, \dots, Y_r^\infty$ have polynomial coefficients. In fact, the coefficient $p_{ij}(x)$ depends only on the coordinates $x_1, \dots, x_{j-1}$. The real Lie algebra $\mathfrak{g} := \mathrm{Lie}(Y_1^\infty, \dots, Y_r^\infty)$ is stratified and nilpotent: all commutators with length at least $s+1$ vanish (see e.g. [3]). In particular, $\mathfrak{g}$ is a finite dimensional real vector space.

Let G be the abstract connected and simply connected Lie group having $\mathfrak{g}$ as its Lie algebra. By the nilpotency of $\mathfrak{g}$, the exponential map $\exp : \mathfrak{g} \to G$ is a global diffeomorphism. The group law $\cdot$ on G is related to the Lie bracket of $\mathfrak{g}$ via the Baker-Campbell-Hausdorff formula: for any $X, Y \in \mathfrak{g}$ we have $\exp(X) \cdot \exp(Y) = \exp(P(X,Y))$, where

$$P(X,Y) := X + Y + \sum_{p=1}^{s-1} \frac{(-1)^p}{p+1} \sum_{\substack{0 \le k_1,\dots,k_p < s \\ 0 \le \ell_1,\dots,\ell_p < s \\ k_i + \ell_i \ge 1}} \frac{(\operatorname{ad} X)^{k_1} (\operatorname{ad} Y)^{\ell_1} \cdots (\operatorname{ad} X)^{k_p} (\operatorname{ad} Y)^{\ell_p}}{(k_1 + \cdots + k_p + 1) k_1! \cdots k_p! \, \ell_1! \cdots \ell_p!} X. \tag{2.6}$$

Here, $\operatorname{ad} X : \mathfrak{g} \to \mathfrak{g}$ is the adjoint mapping $\operatorname{ad} X(Y) := [X,Y]$. The group G is a *Carnot group*, which means that it is a connected, simply connected and nilpotent Lie group whose Lie algebra $\mathfrak{g}$ is stratified, i.e., it has a decomposition $\mathfrak{g} = \mathfrak{g}_1 \oplus \cdots \oplus \mathfrak{g}_s$ satisfying $[\mathfrak{g}_1, \mathfrak{g}_{i-1}] = \mathfrak{g}_i$ and $[\mathfrak{g}, \mathfrak{g}_s] = \{0\}$. The number s is the *step* of the group.

For any $X \in \mathfrak{g}$, the flow $(x,t) \mapsto \Phi_t^X(x)$ is a polynomial function in $(x,t) \in M^\infty \times \mathbb{R}$ and X is therefore complete. For any $X, Y \in \mathfrak{g}$, $x \in M^\infty$ and $t \in \mathbb{R}$ we have the formula for the composition of flows

$$\Phi_t^{P(X,Y)}(x) = \Phi_t^Y \circ \Phi_t^X(x). \tag{2.7}$$

This identity follows from the fact that the left and right hand side are polynomial functions in t and have the same Taylor expansion in t by the Baker-Campbell-Hausdorff formula for vector fields, see [6, Lemma A.1].

If $X \in \mathfrak{g}_1$, its flow satisfies for any $\lambda > 0$, $t \in \mathbb{R}$ and $x \in M^\infty$ the following identity

$$\Phi_t^{\lambda X}(\delta_\lambda(x)) = \delta_\lambda(\Phi_t^X(x)), \tag{2.8}$$

which follows from $(\delta_\lambda)_* X = \lambda X$.

The group G acts on M^∞ on the right. The action $M^\infty \times G \to M^\infty$ is given by $(x,g) \mapsto x \cdot g := \Phi_1^X(x)$, where $g = \exp(X)$. In fact, by (2.7), for any $h = \exp(Y)$ we have

$$x \cdot (gh) = \Phi_1^{P(X,Y)}(x) = \Phi_1^Y \circ \Phi_1^X(x) = (x \cdot g) \cdot h. \tag{2.9}$$

The proof of the following proposition is elementary and is omitted.

Proposition 2.2. Let $\kappa : \mathbb{R} \to M^\infty$ be a horizontal curve in $(M^\infty, \mathscr{X}^\infty)$. The following statements are equivalent:

(i) The minimal control of κ is constant and $\kappa(0) = 0$.

(ii) There exist $c_1, \dots, c_r \in \mathbb{R}$ such that $\dot\kappa = \sum_{i=1}^r c_i Y_i^\infty(\kappa)$ and $\kappa(0) = 0$.

(iii) There exists $Y \in \mathfrak{g}_1$ such that $\kappa(t) = \Phi_t^Y(0)$.

(iv) There exists $x_0 \in M^\infty$ such that $\kappa(t) = \delta_t(x_0)$ (here δ_t is defined by (2.4) for any $t \in \mathbb{R}$).

Definition 2.3. We say that a horizontal curve κ in $(M^\infty, \mathscr{X}^\infty)$ is a *horizontal line* (through 0) if one of the conditions (i)-(iv) holds.

The definition of positive and negative half-line is similar, the formulas above being required to hold for $t \geq 0$ and $t \leq 0$, respectively.

2.2. **Tangent cone to a horizontal curve.** Let $(M, \mathscr{X})$ be a CC structure and let $\gamma : [-T, T] \to M$ be a horizontal curve. Let φ be a chart around the point $x = \gamma(t)$, for some $t \in (-T, T)$, such that $\varphi(x) = 0$. Finally, let δ_λ be the dilations introduced above and denote by $(M^\infty, \mathscr{X}^\infty)$ the tangent CC structure to $(M, \mathscr{X})$ at the point x.

Definition 2.4. The *tangent cone* $\mathrm{Tan}(\gamma; t)$ to γ at $t \in (-T, T)$ is the set of all horizontal curves $\kappa : \mathbb{R} \to M^\infty$ such that there exists an infinitesimal sequence $\eta_i \downarrow 0$ satisfying, for any $\tau \in \mathbb{R}$,

$$\lim_{i \to \infty} \delta_{1/\eta_i} \varphi(\gamma(t + \eta_i \tau)) = \kappa(\tau),$$

with uniform convergence on compact subsets of $\mathbb{R}$.

The definition of $\mathrm{Tan}(\gamma; t)$ depends on the chart φ and on the choice $Z_1, \ldots, Z_n$ of linearly independent iterated commutators. When $\gamma : [0, T] \to M$, the tangent cones $\mathrm{Tan}^+(\gamma; 0)$ and $\mathrm{Tan}^-(\gamma; T)$ can be defined in a similar way. $\mathrm{Tan}^+(\gamma; 0)$ contains curves in M^∞ defined on $[0, \infty)$, while $\mathrm{Tan}^-(\gamma; T)$ contains curves defined on $(-\infty, 0]$.

When $M = M^\infty$ or $M = G$ is a Carnot group, there is already a group of dilations on M itself. In such cases, when $\gamma(t) = 0$, we define the tangent cone $\mathrm{Tan}(\gamma; t)$ as the set of limiting curves of the form $\kappa(t) = \lim_{i \to \infty} \delta_{1/\eta_i} \gamma(t + \eta_i \tau)$.

The tangent cone is closed under uniform convergence of curves on compact sets. We need the following observation in order to initiate the proof of Theorem 1.1.

Proposition 2.5. For any horizontal curve $\gamma : [-T, T] \to M$ the tangent cone $\mathrm{Tan}(\gamma; t)$ is nonempty for any $t \in (-T, T)$. The same holds for $\mathrm{Tan}^+(\gamma; 0)$ and $\mathrm{Tan}^-(\gamma; T)$, for a horizontal curve $\gamma : [0, T] \to M$.

Proof. We prove that $\mathrm{Tan}^+(\gamma; 0) \neq \emptyset$. The other proofs are analogous.

We use exponential coordinates of the first kind centered at $\gamma(0)$. By (1.1), we have a.e.

$$\dot{\gamma} = \sum_{i=1}^{r} h_i Y_i(\gamma) = \sum_{j=1}^{n} \sum_{i=1}^{r} h_i a_{ij}(\gamma) \frac{\partial}{\partial x_j},$$

where $h_i \in L^\infty(-T, T)$ and $a_{ij} = p_{ij} + r_{ij}$ are functions satisfying the claims (i) and (ii) of Proposition 2.1. In fact, we can also assume that $\|h\|_\infty = 1$ and that the closure

$K := \overline{B(0,T)}$ of the CC ball $B(0,T)$ is compact: this is true for small enough T, which is enough for our purposes. Then $\gamma(t) \in K$ for all $t \in [0,T]$ and we have $|\dot\gamma(t)| \le C$ for some constant depending on $\|a_{ij}\|_{L^\infty(K)}$. This implies that $|\gamma(t)| \le Ct$ for all $t \in [0,T]$.

By induction on $k \ge 1$, we prove the following statement: for any j satisfying $w_j \ge k$ we have $|\gamma_j(t)| \le Ct^k$. The base case $k = 1$ has already been treated. Now assume that $w_j \ge k > 1$ and that the statement is true for $1, \dots, k-1$. Since r_{ij} is smooth, we have $r_{ij} = q_{ij,k} + r_{ij,k}$, where $q_{ij,k}$ is a polynomial containing only terms with homogeneous degree at least w_j and $|r_{ij,k}(x)| \le C|x|^{k-1}$ on K (here $|x|$ denotes the usual Euclidean norm).

Each monomial $c_\alpha x^\alpha$ of the polynomial $p_{ij} + q_{ij,k}$ has homogeneous degree $w_\alpha := \sum_{m=1}^n \alpha_m w_m \ge w_j - 1$. If $\alpha_m = 0$ whenever $w_m \ge k$, then we can estimate

$$|\gamma(t)^\alpha| = \prod_{m: w_m \le k-1} |\gamma_m(t)|^{\alpha_m} \le Ct^{w_\alpha} \le Ct^{k-1},$$

using the inductive hypothesis with k replaced by $w_m \le k-1$. Otherwise, there exists some index m with $w_m \ge k$ and $\alpha_m > 0$, in which case

$$|\gamma(t)^\alpha| \le C|\gamma_m(t)| \le Ct^{k-1},$$

using the inductive hypothesis with k replaced by $k-1$. Thus $|p_{ij}(\gamma(t)) + q_{ij,k}(\gamma(t))| \le Ct^{k-1}$. Combining this with the estimate $|r_{ij,k}(\gamma(t))| \le Ct^{k-1}$, we obtain $|a_{ij}(\gamma(t))| \le Ct^{k-1}$. So we finally have

$$|\gamma_j(t)| \le \sum_{i=1}^r \int_0^t |a_{ij}(\gamma(\tau))| \, d\tau \le Ct^k.$$

Applying the above statement with $k = w_j$, we obtain

$$|\gamma_j(t)| \le Ct^{w_j}, \tag{2.10}$$

for a suitable constant C depending only on K and T.

Now we prove that $\mathrm{Tan}^+(\gamma; 0)$ is nonempty. For $\eta > 0$ consider the family of curves $\gamma^\eta(t) := \delta_{1/\eta}(\gamma(\eta t))$, defined for $t \in [0, T/\eta]$. The derivative of γ^η is a.e.

$$\dot\gamma^\eta(t) = \sum_{j=1}^n \sum_{i=1}^r h_i(\eta t) \eta^{1-w_j} a_{ij}(\gamma(\eta t)) \frac{\partial}{\partial x_j},$$

where, by Proposition 2.1 and the estimates (2.10), we have

$$|a_{ij}(\gamma(\eta t))| \le C\|\gamma(\eta t)\|^{w_j - 1} \le C(\eta t)^{w_j - 1}.$$

We are using the homogeneous norm $\|\cdot\|$ defined in (2.5). This proves that the family of curves $(\gamma^\eta)_{\eta > 0}$ is Lipschitz equicontinuous. So it has a subsequence $(\gamma^{\eta_i})_i$ that is converging locally uniformly as $\eta_i \to 0$ to a curve $\kappa : [0, \infty) \to \mathbb{R}^n$. Now it is easy to see that κ is horizontal in $(M^\infty, \mathscr{X}^\infty)$: see also the proof of Lemma 3.5. $\square$

Remark 2.6. The following result was obtained along the proof of Proposition 2.5. Let $(M, \mathscr{X})$ be a Carnot-Carathéodory structure. Using exponential coordinates of the first kind, we (locally) identify M with $\mathbb{R}^n$ and we assign to the coordinate x_j the weight w_j, as above. Assume that $T > 0$ is such that $K := \overline{B(0,T)}$ is compact. Then there exists a positive constant $C = C(T)$ such that the following holds: for any horizontal curve $\gamma : [0,T] \to M = \mathbb{R}^n$ parametrized by arclength and such that $\gamma(0) = 0$, one has

$$|\gamma_j(t)| \leq C t^{w_j}, \quad \text{for any } j = 1, \dots, n \text{ and } t \in [0,T]. \tag{2.11}$$

We shall use this fact in the case of Carnot groups where, by homogeneity, the constant C does not depend on T.

If γ is a length-minimizing curve, then the curves in $\mathrm{Tan}(\gamma; t)$ are also locally length-minimizing.

Proposition 2.7. Let $\gamma : [-T,T] \to M$ be a length-minimizing curve in $(M, \mathscr{X})$, parametrized by arclength, and let $\kappa \in \mathrm{Tan}(\gamma; t)$ for some $t \in (-T,T)$. Then κ is parametrized by arclength and, when restricted to any compact interval, it is length-minimizing in the tangent Carnot-Carathéodory structure $(M^\infty, \mathscr{X}^\infty)$.

The proof of Proposition 2.7 is contained in [11, Proposition 2.4] (the quoted paper deals with equiregular CC spaces, but the proof of this proposition works also in our more general setting). The following fact is a special case of the general principle according to which the tangent to the tangent is (contained in the) tangent.

Proposition 2.8. Let $\gamma : [-T,T] \to M$ be a horizontal curve and $t \in (-T,T)$. If $\kappa \in \mathrm{Tan}(\gamma; t)$ and $\widehat{\kappa} \in \mathrm{Tan}(\kappa; 0)$, then $\widehat{\kappa} \in \mathrm{Tan}(\gamma; t)$.

Proof. We can without loss of generality assume that $t = 0$ and that γ takes values in $\mathbb{R}^n = M^\infty$. Let $N > 0$ be fixed. Since $\widehat{\kappa} \in \mathrm{Tan}(\kappa; 0)$, there exists an infinitesimal sequence $\xi_k \downarrow 0$ such that, for all $t \in [-N,N]$ and $k \in \mathbb{N}$, we have

$$\|\widehat{\kappa}(t) - \delta_{1/\xi_k}\kappa(\xi_k t)\| \leq \frac{1}{2^k}.$$

Since $\kappa \in \mathrm{Tan}(\gamma; 0)$, there exists an infinitesimal sequence $\eta_k \downarrow 0$ such that, for all $t \in [-N,N]$ and $k \in \mathbb{N}$, we have

$$\|\kappa(\xi_k t) - \delta_{1/\eta_k}\gamma(\eta_k \xi_k t)\| \leq \frac{\xi_k}{2^k}.$$

It follows that for the infinitesimal sequence $\sigma_k := \xi_k \eta_k$ we have, for all $t \in [-N,N]$,

$$\|\widehat{\kappa}(t) - \delta_{1/\sigma_k}\kappa(\sigma_k t)\| \leq \|\widehat{\kappa}(t) - \delta_{1/\xi_k}\kappa(\xi_k t)\| + \|\delta_{1/\xi_k}\kappa(\xi_k t) - \delta_{1/\sigma_k}\gamma(\sigma_k t)\| \leq \frac{1}{2^{k-1}}.$$

The thesis now follows by a diagonal argument. $\square$

When $\gamma : [0,T] \to M$, there are analogous versions of Propositions 2.7 and 2.8 for $\mathrm{Tan}^+(\gamma;0)$ and $\mathrm{Tan}^-(\gamma;T)$.

2.3. **Lifting of tangent structures to Carnot groups.** It might happen that, at some points, the vector fields $Y_1^\infty, \dots, Y_r^\infty$ are $\mathbb{R}$-linearly dependent and the group G could have rank strictly less than r. To avoid this situation, we lift the tangent CC structure $(M^\infty, \mathscr{X}^\infty)$ to a free Carnot group. We give some details of the construction, which is well-known (see [6]).

Let F be the free Carnot group of rank r and step s, and denote by $\mathfrak{f} = \mathfrak{f}_1 \oplus \cdots \oplus \mathfrak{f}_s$ its Lie algebra. Let $W_1, \dots, W_N$ be a basis of $\mathfrak{f}$ adapted to the stratification. In particular, $W_1, \dots, W_r$ is a basis of $\mathfrak{f}_1$ and we let $\mathscr{W} := \{W_1, \dots, W_r\}$. Via the exponential mapping $\exp : \mathfrak{f} \to F$, the one-parameter group of automorphisms of $\mathfrak{f}$ given by $W_k \mapsto \lambda^i W_k$ if and only if $W_k \in \mathfrak{f}_i$ induces a one-parameter group of automorphisms $(\widehat{\delta}_\lambda)_{\lambda>0}$ of F, called dilations. In the next sections, we will use the simpler notation $\delta_\lambda := \widehat{\delta}_\lambda$. In exponential coordinates for F these dilations coincide with the ones introduced in (2.4).

Let G be the abstract Carnot group constructed in the previous subsection starting from $(M^\infty, \mathscr{X}^\infty)$ with Lie algebra $\mathfrak{g}$. Since $\mathfrak{g}$ is a quotient of $\mathfrak{f}$, there exists a unique homomorphism $\psi : F \to G$ such that $\psi_*(W_i) = Y_i^\infty \in \mathfrak{g}$. We define the map $\pi^\infty : F \to M^\infty$ by $\pi^\infty(f) := 0 \cdot \psi(f)$, where the dot stands for the right action of G on M^∞.

Definition 2.9. We call the CC structure $(F, \mathscr{W})$ the *lifting* of $(M^\infty, \mathscr{X}^\infty)$ with projection $\pi^\infty : F \to M^\infty$.

Proposition 2.10. The lifting $(F, \mathscr{W})$ of $(M^\infty, \mathscr{X}^\infty)$ has the following properties.

(i) For any $x \in M^\infty$ and $i = 1, \dots, r$, we have $\pi^\infty_*(W_i)(x) = Y_i^\infty(x)$.

(ii) The dilations of F and M^∞ commute with the projection. Namely, for any $\lambda > 0$ we have

$$\pi^\infty \circ \widehat{\delta}_\lambda = \delta_\lambda \circ \pi^\infty.$$

Proof. (i) Let $x = \pi^\infty(f)$ for some $f \in F$. Using the homomorphism property for $\psi : F \to G$ and the action property (2.9), we find

$$\begin{aligned}\pi^\infty_*(W_i)(\pi^\infty(f)) &= \frac{d}{dt}\pi^\infty(f\exp(tW_i))\Big|_{t=0} = \frac{d}{dt} 0 \cdot (\psi(f)\psi(\exp(tW_i)))\Big|_{t=0} \\ &= \frac{d}{dt}\pi^\infty(f) \cdot \psi(\exp(tW_i))\Big|_{t=0} = \psi_*(W_i)(\pi^\infty(f)) = Y_i^\infty(\pi^\infty(f)).\end{aligned}$$

(ii) Let $\lambda > 0$ and $x \in M^\infty$. By (2.8), for any $X \in \mathfrak{g}_1$ we have

$$\delta_\lambda(x) \cdot \exp(\lambda X) = \Phi_1^{\lambda X}(\delta_\lambda(x)) = \delta_\lambda\Big(\Phi_1^X(x)\Big) = \delta_\lambda(x \cdot \exp(X)). \tag{2.12}$$

We prove the claim for $f = \exp(W)$ with $W \in \mathfrak{f}_1$:

$$\begin{aligned}\pi^\infty(\widehat{\delta}_\lambda(f)) &= \pi^\infty(\exp(\lambda W)) = 0 \cdot \exp(\lambda \psi_*(W)) \\ &= \Phi_1^{\lambda\psi_*(W)}(0) = \delta_\lambda(\Phi_1^{\psi_*(W)}(0)) = \delta_\lambda(\pi^\infty f).\end{aligned}$$

In general, any $f \in F$ is of the form $f = f_1 f_2 \dots f_k$ with each $f_i \in \exp(\mathfrak{f}_1)$. Assume by induction that the claim holds for $\widehat{f} = f_1 f_2 \dots f_{k-1}$. By (2.12), letting $f_k = \exp(W)$ we have

$$\begin{aligned}\pi^\infty(\widehat{\delta}_\lambda(f)) &= \pi^\infty(\widehat{\delta}_\lambda(\widehat{f})\exp(\lambda W)) = 0 \cdot \psi(\widehat{\delta}_\lambda(\widehat{f})\exp(\lambda W)) \\ &= (0 \cdot \psi(\widehat{\delta}_\lambda(\widehat{f}))) \cdot \psi(\exp(\lambda W)) = \pi^\infty(\widehat{\delta}_\lambda(\widehat{f})) \cdot \psi(\exp(\lambda W)) \\ &= \delta_\lambda(\pi^\infty(\widehat{f})) \cdot \psi(\exp(\lambda W)) = \delta_\lambda(\pi^\infty(\widehat{f}) \cdot \psi(\exp(W))) = \delta_\lambda(\pi^\infty(f)).\end{aligned}$$

□

Let $\kappa : I \to M^\infty$ be a horizontal curve in $(M^\infty, \mathscr{X}^\infty)$, with minimal control $h \in L^\infty(I, \mathbb{R}^r)$. A horizontal curve $\bar{\kappa} : I \to F$ such that

$$\dot{\bar{\kappa}}(t) = \sum_{i=1}^r h_i(t) W_i(\bar{\kappa}(t)) \qquad \text{for a.e. } t \in I$$

is called *lift* of κ to $(F, \mathscr{W})$.

Proposition 2.11. Let $(F, \mathscr{W})$ be the lifting of $(M^\infty, \mathscr{X}^\infty)$ with projection $\pi^\infty : F \to M^\infty$. Then the following facts hold:

(i) If κ is length-minimizing in $(M^\infty, \mathscr{X}^\infty)$, then any horizontal lift $\bar{\kappa}$ of κ is length-minimizing in $(F, \mathscr{W})$.

(ii) If $\bar{\kappa}$ is a horizontal (half-)line in F, then $\pi^\infty \circ \bar{\kappa}$ is a horizontal (half-)line in $(M^\infty, \mathscr{X}^\infty)$.

Proof. Claim (i) follows from $L(\bar{\kappa}) = L(\kappa)$ and the inequality $L(\kappa') \geq L(\kappa)$, whenever κ' is horizontal and $\kappa = \pi^\infty \circ \kappa'$. We now turn to Claim (ii). Let $\bar{\kappa}(t) = \exp(tW)$ for some $W \in \mathfrak{f}_1$. The projection $\pi^\infty \circ \bar{\kappa}$ is horizontal by part (i) of Proposition 2.10. The thesis follows from characterization (ii) for horizontal lines, contained in Proposition 2.2.

□

3. Excess, compactness of length minimizers and first consequences

In this section we prove Lemma 3.6, which provides the correct position for the correction devices introduced in Section 4. We work in the setting of a Carnot group.

Let G be an n-dimensional Carnot group with Lie algebra $\mathfrak{g} = \mathfrak{g}_1 \oplus \dots \oplus \mathfrak{g}_s$, endowed with a positive definite scalar product $\langle \cdot, \cdot \rangle$ such that $\mathfrak{g}_i \perp \mathfrak{g}_j$ whenever $i \neq j$. We also let $|\cdot| := \langle \cdot, \cdot \rangle^{1/2}$. We fix an orthonormal basis $X_1, \dots, X_n$ of $\mathfrak{g}$ adapted to the stratification, i.e., such that $\mathfrak{g}_j = \mathrm{span}\{X_{r_{j-1}+1}, \dots, X_{r_j}\}$ for any $j = 1, \dots, s$, where

$r_j := \dim(\mathfrak{g}_1) + \cdots + \dim(\mathfrak{g}_j)$ and $r_0 := 0$. We identify $\mathfrak{g}_1$ with $\mathbb{R}^r$ through the fixed orthonormal basis $X_1, \ldots, X_r$ and denote by S^{r-1} and $G(r-1)$ the set of unit vectors and linear hyperplanes in $\mathfrak{g}_1$, respectively.

We denote by $\exp : \mathfrak{g} \to G$ the exponential mapping, by $\overline{\pi} : \mathfrak{g} \to \mathfrak{g}_1$ the projection onto the first layer and by $\pi : G \to \mathfrak{g}_1$ the mapping $\pi = \overline{\pi} \circ \exp^{-1}$. For any curve γ in G we use the short notation $\underline{\gamma} := \pi \circ \gamma$.

In this section, I denotes a compact interval of positive length.

Definition 3.1. Given a horizontal curve $\gamma : I \to G$ and a Borel subset $B \subseteq I$ with $\mathscr{L}^1(B) > 0$, we define the *excess* of γ on B as

$$\operatorname{Exc}(\gamma; B) := \inf_{v \in S^{r-1}} \left(\fint_B \langle v, \dot{\underline{\gamma}}(t) \rangle^2 \, dt \right)^{1/2}.$$

Remark 3.2. The excess can be equivalently defined as

$$\operatorname{Exc}(\gamma; B) := \inf_{\Pi \in G(r-1)} \left(\fint_B \left| \dot{\underline{\gamma}}(t) - \Pi \left(\dot{\underline{\gamma}}(t) \right) \right|^2 dt \right)^{1/2},$$

where we identify the hyperplane Π with the orthogonal projection $\mathfrak{g}_1 \to \Pi$.

Remark 3.3. Given a horizontal curve γ, $g \in G$ and $r > 0$, setting $\gamma_1(t) := g\,\gamma(t)$, $\gamma_2(t) := \delta_r(\gamma(t))$, we have

$$\operatorname{Exc}(\gamma_1; B) = \operatorname{Exc}(\gamma; B) \quad \text{and} \quad \operatorname{Exc}(\gamma_2; B) = r \operatorname{Exc}(\gamma; B).$$

Moreover, for $\gamma_3(t) := \delta_r(\gamma(t/r))$ we have $\operatorname{Exc}(\gamma_3; rB) = \operatorname{Exc}(\gamma; B)$.

Remark 3.4. The map from

$$S^{r-1} \times L^2(I, \mathfrak{g}_1) \ni (v, u) \mapsto \left(\fint_B \langle v, u(t) \rangle^2 \, dt \right)^{1/2} \in \mathbb{R}$$

is continuous. As a consequence, the infimum in Definition 3.1 is in fact a minimum and, by the compactness of S^{r-1}, we have

$$\operatorname{Exc}(\gamma_k; B) \to \operatorname{Exc}(\gamma; B)$$

whenever $\dot{\underline{\gamma_k}} \to \dot{\underline{\gamma}}$ in $L^2(I, \mathfrak{g}_1)$.

The following compactness result for length minimizers parametrized by arclength implies a certain uniform – though not explicit – estimate: see Lemma 3.6 below.

Lemma 3.5 (Compactness of minimizers)**.** Let I be a compact interval containing 0 and let $\gamma_k : I \to G$, $k \in \mathbb{N}$, be a sequence of length minimizers parametrized by arclength with $\gamma_k(0) = 0$. Then, there exist a subsequence γ_{k_p} and a length minimizer $\gamma_\infty : I \to G$, parametrized by arclength and with $\gamma_\infty(0) = 0$, such that $\gamma_{k_p} \to \gamma_\infty$ uniformly and $\dot{\gamma}_{k_p} \to \dot{\gamma}_\infty$ in $L^2(I)$.

Proof. By homogeneity, it is not restrictive to assume $I = [0,1]$. For any k we have $\gamma_k([0,1]) \subseteq \overline{B(0,1)}$, the closed unit ball, which is compact. Since all the curves γ_k are 1-Lipschitz with respect to the Carnot-Carathéodory distance d, we can find a subsequence γ_{k_p} converging uniformly to some curve γ_∞.

Since $\|\underline{\dot\gamma_{k_p}}\|_{L^2([0,1],\mathfrak{g}_1)}= 1$, up to selecting a further subsequence we can assume that $\underline{\dot\gamma_{k_p}} \rightharpoonup \overline{u}$ in $L^2([0,1],\mathfrak{g}_1)$. Thus, identifying G with $\mathbb{R}^n$ by exponential coordinates and passing to the limit as $p \to \infty$ in

$$\gamma_{k_p}(t) = \int_0^t \left(\sum_{i=1}^r \dot\gamma_{k_p,i}(\tau) X_i(\gamma_{k_p}(\tau)) \right) d\tau,$$

we obtain, for any $t \in [0,1]$,

$$\gamma_\infty(t) = \int_0^t \left(\sum_{i=1}^r \overline{u}_i(\tau) X_i(\gamma_\infty(\tau)) \right) d\tau.$$

This proves that γ_∞ is horizontal with $\underline{\dot\gamma_\infty} = \overline{u}$. Moreover,

$$\left\|\underline{\dot\gamma_\infty}\right\|_{L^2([0,1],\mathfrak{g}_1)} \geq L(\gamma_\infty) \geq d(\gamma_\infty(0),\gamma_\infty(1)) = \lim_{p\to\infty} d(\gamma_{k_p}(0),\gamma_{k_p}(1)) = 1. \tag{3.13}$$

We already know that $\left\|\underline{\dot\gamma_\infty}\right\|_{L^2([0,1],\mathfrak{g}_1)} \leq 1$ (because $\underline{\dot\gamma_{k_p}} \rightharpoonup \underline{\dot\gamma_\infty}$ and $\|\underline{\dot\gamma_{k_p}}\|_{L^2([0,1],\mathfrak{g}_1)}= 1$), so $\|\underline{\dot\gamma_{k_p}}\|_{L^2([0,1],\mathfrak{g}_1)}\to \left\|\underline{\dot\gamma_\infty}\right\|_{L^2([0,1],\mathfrak{g}_1)}$ and, since $L^2([0,1],\mathfrak{g}_1)$ is a Hilbert space, this gives $\underline{\dot\gamma_{k_p}} \to \underline{\dot\gamma_\infty}$ in $L^2([0,1],\mathfrak{g}_1)$ and $\dot\gamma_{k_p} \to \dot\gamma_\infty$ in $L^2([0,1])$. In particular, $\underline{\dot\gamma_\infty}(t)$ is for a.e. $t \in [0,1]$ a unit vector in $\mathfrak{g}_1$. As all inequalities in (3.13) must be equalities, we obtain $L(\gamma_\infty) = d(\gamma_\infty(0),\gamma_\infty(1))$, i.e., γ_∞ is a length minimizer parametrized by arclength. □

Lemma 3.6. For any $\varepsilon > 0$ there exists a constant $c = c(G,\varepsilon) > 0$ such that the following holds. For any length minimizer $\gamma : I \to G$ parametrized by arclength and such that $\mathrm{Exc}(\gamma; I) \geq \varepsilon$, there exist r subintervals $[a_1,b_1],\dots,[a_r,b_r] \subseteq I$, with $a_i < b_i \leq a_{i+1}$, such that

$$\left|\det\left(\underline{\gamma}(b_1) - \underline{\gamma}(a_1),\dots,\underline{\gamma}(b_r) - \underline{\gamma}(a_r)\right)\right| \geq c(\mathscr{L}^1(I))^r. \tag{3.14}$$

The determinant is defined by means of the identification of $\mathfrak{g}_1$ with $\mathbb{R}^r$ via the basis $X_1,\dots,X_r$.

Proof. By Remark 3.3 we can assume that $I = [0,1]$ and that $\gamma(0) = 0$. By contradiction, assume there exist length minimizers $\gamma_k : [0,1] \to G$ parametrized by arclength, with $\gamma_k(0) = 0$ and $\mathrm{Exc}(\gamma_k; [0,1]) \geq \varepsilon$, such that

$$\left|\det\left(\underline{\gamma_k}(b_1) - \underline{\gamma_k}(a_1),\dots,\underline{\gamma_k}(b_r) - \underline{\gamma_k}(a_r)\right)\right| \leq 2^{-k}, \tag{3.15}$$

for any $0 \leq a_1 < b_1 \leq \dots \leq a_r < b_r \leq 1$. By Lemma 3.5, there exists a subsequence $(\gamma_{k_p})_{p\in\mathbb{N}}$ such that $\gamma_{k_p} \to \gamma_\infty$ uniformly and $\dot\gamma_{k_p} \to \dot\gamma_\infty$ in $L^2([0,1])$ for some length minimizer γ_∞ parametrized by arclength. Passing to the limit as $p \to \infty$ in (3.15) we deduce that

$$\det\big(\underline{\gamma_\infty}(b_1) - \underline{\gamma_\infty}(a_1), \ldots, \underline{\gamma_\infty}(b_r) - \underline{\gamma_\infty}(a_r)\big) = 0, \tag{3.16}$$

for any $0 \le a_1 < b_1 \le \cdots \le a_r < b_r \le 1$.

Let S be the set of differentiability points $t \in (0,1)$ of γ_∞ and let

$$\mathfrak{h}_1 := \mathrm{span}\{\dot\gamma_\infty(t) \mid t \in S\} \subseteq \mathfrak{g}_1$$

be the linear subspace of $\mathfrak{g}_1$ spanned by the derivatives $\dot\gamma_\infty(t)$. We claim that $\dim \mathfrak{h}_1 < r$. If this were not the case, we could find $0 < t_1 < \cdots < t_r < 1, t_i \in S$, such that $\dot\gamma_\infty(t_1), \ldots, \dot\gamma_\infty(t_r)$ are linearly independent. Setting

$$a_i := t_i,\ b_i := t_i + \delta, \quad i = 1, \ldots, r$$

and letting $\delta \downarrow 0$ in (3.16), we would deduce that $\det(\dot\gamma_\infty(t_1), \ldots, \dot\gamma_\infty(t_r)) = 0$, which is a contradiction.

As a consequence, there exists a unit vector $v \in \mathfrak{g}_1$ orthogonal to $\mathfrak{h}_1$ and we obtain

$$\mathrm{Exc}(\gamma_\infty; [0,1]) \le \left(\int_0^1 \langle v, \dot\gamma_\infty(t)\rangle^2\, dt\right)^{1/2} = 0.$$

But from $\mathrm{Exc}(\gamma_{k_p}; [0,1]) \ge \varepsilon$ and Remark 3.4 we also have $\mathrm{Exc}(\gamma_\infty; [0,1]) \ge \varepsilon$. This is a contradiction and the proof is accomplished. □

Remark 3.7. Under the same assumptions and notation of Lemma 3.6, we also have

$$|\underline{\gamma}(b_i) - \underline{\gamma}(a_i)| \ge c\mathscr{L}^1(I) \qquad \text{for any } i = 1, \ldots, r. \tag{3.17}$$

Indeed, one has $|\underline{\gamma}(b_i) - \underline{\gamma}(a_i)| \le \mathscr{L}^1(I)$ by arclength parametrization and (3.14) could not hold in case (3.17) were false for some index i.

4. Cut and correction devices

In this section we introduce the cut and the iterated correction of a horizontal curve. In Lemma 4.4 we compute the gain of length in terms of the excess. In the formula (4.21), we establish an algebraic identity for the displacement of the endpoint produced by an iterated correction. We keep on working in a Carnot group G.

The *concatenation* of two curves $\alpha : [a, a + a'] \to G$ and $\beta : [b, b + b'] \to G$ is the curve $\alpha * \beta : [a, a + (a' + b')] \to G$ defined by the formula

$$\alpha * \beta(t) := \begin{cases} \alpha(t) & \text{if } t \in [a, a + a'] \\ \alpha(a + a')\beta(b)^{-1}\beta(t + b - (a + a')) & \text{if } t \in [a + a', a + a' + b']. \end{cases}$$

The concatenation $\alpha * \beta$ is continuous if α and β are continuous and it is horizontal if and only if α and β are horizontal. The operation $*$ is associative.

Definition 4.1 (Cut curve)**.** Let $\gamma : [a,b] \to G$ be a curve. For any subinterval $[s,s'] \subseteq [a,b]$ with $\underline{\gamma}(s') \neq \underline{\gamma}(s)$ we define the *cut curve* $\mathrm{Cut}(\gamma;[s,s']) : [a,b''] \to G$, with $b'' := b - (s'-s) + \left|\underline{\gamma}(s') - \underline{\gamma}(s)\right|$, by the formula

$$\mathrm{Cut}(\gamma;[s,s']) := \gamma|_{[a,s]} * \exp(\,\cdot\, w)|_{\left[0,\left|\underline{\gamma}(s')-\underline{\gamma}(s)\right|\right]} * \gamma|_{[s',b]},$$

where

$$w = \frac{\underline{\gamma}(s') - \underline{\gamma}(s)}{\left|\underline{\gamma}(s') - \underline{\gamma}(s)\right|}.$$

When $\underline{\gamma}(s') = \underline{\gamma}(s)$, the cut curve is defined by

$$\mathrm{Cut}(\gamma;[s,s']) = \gamma|_{[a,s]} * \gamma|_{[s',b]}.$$

Remark 4.2. If γ is parametrized by arclength and horizontal, then the cut curve $\mathrm{Cut}(\gamma;[s,s'])$ is parametrized by arclength and horizontal. Moreover, its length is

$$L(\mathrm{Cut}(\gamma;[s,s'])) = L(\gamma) - (s'-s) + |\underline{\gamma}(s') - \underline{\gamma}(s)|. \tag{4.18}$$

Remark 4.3. The final point of the cut curve has the same projection on $\mathfrak{g}_1$ as the final point of γ, i.e., $\pi\left(\mathrm{Cut}(\gamma;[s,s'])(b'')\right) = \pi(\gamma(b))$. Indeed, by Lemma 4.6 below we have

$$\begin{aligned}\pi\left(\mathrm{Cut}(\gamma;[s,s'])(b'')\right) &= \pi(\gamma(s)\exp\left(\left|\underline{\gamma}(s') - \underline{\gamma}(s)\right| w\right)\gamma(s')^{-1}\gamma(b)) \\ &= \underline{\gamma}(s) + \left|\underline{\gamma}(s') - \underline{\gamma}(s)\right| w + \left(\underline{\gamma}(b) - \underline{\gamma}(s')\right) \\ &= \underline{\gamma}(b).\end{aligned}$$

Lemma 4.4. Let $\gamma : I \to G$ be a horizontal curve parametrized by arclength on a compact interval I and let $J \subseteq I$ be a subinterval with $\mathscr{L}^1(J) > 0$. Then we have

$$L(\gamma) - L(\mathrm{Cut}(\gamma;J)) \geq \frac{\mathscr{L}^1(J)}{2}\,\mathrm{Exc}(\gamma;J)^2.$$

Proof. Let $J = [s,s']$ for some $s < s'$. As in Definition 4.1, let $w \in \mathfrak{g}_1$ be a unit vector such that $\left\langle w, \underline{\gamma}(s') - \underline{\gamma}(s)\right\rangle = \left|\underline{\gamma}(s') - \underline{\gamma}(s)\right|$, i.e.,

$$\left\langle w, \fint_s^{s'} \dot{\underline{\gamma}}(t)\,dt\right\rangle = \frac{|\underline{\gamma}(s') - \underline{\gamma}(s)|}{s'-s}.$$

Since $\left|\dot{\underline{\gamma}}\right| = 1$ a.e., we have $\left|\dot{\underline{\gamma}} - w\right|^2 = 2\left(1 - \left\langle w, \dot{\underline{\gamma}}\right\rangle\right)$, and since $r \geq 2$ there exists a unit vector $v \in \mathfrak{g}_1$ with $\langle v, w\rangle = 0$. Thus, for all t such that $\dot{\underline{\gamma}}(t)$ is defined we have

$$\left|\left\langle v, \dot{\underline{\gamma}}(t)\right\rangle\right| = \left|\left\langle v, \dot{\underline{\gamma}}(t) - w\right\rangle\right| \leq \left|\dot{\underline{\gamma}}(t) - w\right|.$$

We deduce that

$$\begin{aligned}\mathrm{Exc}(\gamma;[s,s'])^2 &\leq \fint_s^{s'} \left\langle v, \dot{\underline{\gamma}}(t)\right\rangle^2 dt \leq \fint_s^{s'} \left|\dot{\underline{\gamma}}(t) - w\right|^2 dt \\ &= 2\left(1 - \left\langle w, \fint_s^{s'} \dot{\underline{\gamma}}(t)\,dt\right\rangle\right) = 2\left(1 - \frac{|\underline{\gamma}(s') - \underline{\gamma}(s)|}{s'-s}\right).\end{aligned}$$

Multiplying by $\mathscr{L}^1(J) = s' - s$ and using (4.18), we obtain the claim:

$$\mathscr{L}^1(J)\,\mathrm{Exc}(\gamma;J)^2 \le 2\left((s'-s) - \left|\underline{\gamma}(s') - \underline{\gamma}(s)\right|\right) = 2\left(L(\gamma) - L(\mathrm{Cut}(\gamma;J))\right).$$

$\square$

Given $Y \in \mathfrak{g}$, we hereafter denote by $\delta_Y : [0,\ell_Y] \to G$ a geodesic from 0 to $\exp(Y)$ parametrized by arclength; in particular, $\ell_Y = d(0,\exp(Y))$. We denote by $\delta_Y(\ell_Y - \cdot\,)$ the curve δ_Y traveled backwards from $\exp(Y)$ to 0.

Definition 4.5 (Corrected curve and displacement)**.** Let $\gamma : [a,b] \to G$ be a horizontal curve parametrized by arclength. For any subinterval $[s,s'] \subseteq [a,b]$ and $Y \in \mathfrak{g}$, we define the *corrected curve* $\mathrm{Cor}(\gamma;[s,s'],Y) : [a,b'''] \to G$, with $b''' := b + 2\ell_Y$, by

$$\mathrm{Cor}(\gamma;[s,s'],Y) := \gamma|_{[a,s]} * \delta_Y * \gamma|_{[s,s']} * \delta_Y(\ell_Y - \cdot\,) * \gamma|_{[s',b]}.$$

We refer to the process of transforming γ into $\mathrm{Cor}(\gamma;[s,s'],Y)$ as to the application of the *correction device* associated with $[s,s']$ and Y. The *displacement* of the final point produced by the correction device associated with $[s,s']$ and Y is

$$\mathrm{Dis}(\gamma;[s,s'],Y) := \gamma(b)^{-1}\,\mathrm{Cor}(\gamma;[s,s'],Y)(b''').$$

We will later express the displacement in terms of suitable conjugations $C_g(h) := ghg^{-1}$ and commutators $[g,h] := ghg^{-1}h^{-1}$ in G.

For any $1 \le j \le s$, we denote by $\overline{\pi}_j : \mathfrak{g} \to \mathfrak{g}_j$ the canonical projection with respect to the direct sum. The mappings $\pi_j : G \to \mathfrak{g}$ are defined as $\pi_j := \overline{\pi}_j \circ \exp^{-1}$. Clearly, one has $\overline{\pi}_1 = \overline{\pi}$ and $\pi_1 = \pi$. We let $\mathfrak{w}_j := \mathfrak{g}_j \oplus \cdots \oplus \mathfrak{g}_s$ and $G_j := \exp(\mathfrak{w}_j)$. We also agree that $G_{s+1} := \{0\}$, the identity element of G, and $\mathfrak{w}_{s+1} := \{0\}$.

Lemma 4.6. The mapping $\pi : G \to (\mathfrak{g}_1,+)$ is a group homomorphism. For any $1 \le j \le s$, G_j is a subgroup of G and $\pi_j : G_j \to (\mathfrak{g}_j,+)$ is a group homomorphism.

Proof. Given points $g = \exp(x_1X_1 + \cdots + x_nX_n)$ and $g' = \exp(x_1'X_1 + \cdots + x_n'X_n)$ in G, by (2.6) we have $\exp^{-1}(g\,g') = (x_1+x_1')X_1 + \cdots + (x_r+x_r')X_r + R$ with $R \in \mathfrak{w}_2$ and hence

$$\pi(gg') = \overline{\pi}(\exp^{-1}(gg')) = (x_1+x_1')X_1 + \cdots + (x_r+x_r')X_r = \pi(g) + \pi(g').$$

The fact that G_j is a subgroup follows from the Baker-Campbell-Hausdorff formula, and the assertion that $\pi_j : G_j \to \mathfrak{g}_j$ is a homomorphism can be obtained as above. $\square$

The following lemmas describe how the homomorphisms π_j interact with conjugations, commutators and Lie brackets. We denote by $\mathrm{Ad}(g)$ the differential of the conjugation C_g at the identity $0 \in G$. This is an automorphism of $\mathfrak{g}$. For $X, Y \in \mathfrak{g}$ and $g \in G$, we have the formulas $\mathrm{Ad}(\exp(X)) = \mathrm{e}^{\mathrm{ad}(X)}$ and $C_g(\exp(Y)) = \exp(\mathrm{Ad}(g)Y)$, see e.g. [7].

Lemma 4.7. For any $g \in G$ and $h \in G_j$ we have $ghg^{-1} \in G_j$ (i.e., G_j is normal in G) and $\pi_j(ghg^{-1}) = \pi_j(h)$.

Proof. With $g = \exp(X)$ and $h = \exp(Y)$, we have

$$\exp^{-1}(ghg^{-1}) = \mathrm{Ad}(g)Y = \mathrm{e}^{\mathrm{ad}\,X}Y = \sum_{k=0}^{\infty} \frac{(\mathrm{ad}\,X)^k}{k!}Y = Y + R,$$

with $R \in \mathfrak{w}_{j+1}$, because in the previous sum all the terms with $k \geq 1$ belong to $\mathfrak{w}_{j+1}$. Hence, we have $ghg^{-1} \in G_j$ and

$$\pi_j(ghg^{-1}) = \overline{\pi}_j \circ \exp^{-1}(ghg^{-1}) = \overline{\pi}_j(Y + R) = \overline{\pi}_j(Y) = \pi_j(h).$$

□

Lemma 4.8. For any $g \in G$ and $h \in G_j$ with $1 \leq j < s$ we have

$$[g, h] \in G_{j+1} \quad \text{and} \quad \pi_{j+1}([g, h]) = [\pi(g), \pi_j(h)].$$

A similar statement holds if $g \in G_j$ and $h \in G$.

Proof. We prove only the first part of the statement, the second one following from the first one and the identity $[g, h] = [h, g]^{-1}$. Combining Lemma 4.7 with Lemma 4.6, we obtain $[g, h] = (ghg^{-1})h^{-1} \in G_j$ and

$$\pi_j([g, h]) = \pi_j(ghg^{-1}) + \pi_j(h^{-1}) = \pi_j(h) - \pi_j(h) = 0,$$

so that $[g, h] \in G_{j+1}$. Now, writing $g = \exp(X)$, $h = \exp(Y)$ and using the formula $\exp^{-1}(ghg^{-1}) = \mathrm{e}^{\mathrm{ad}\,X}Y$ as in the previous proof, we obtain

$$\exp^{-1}(ghg^{-1}) = \sum_{k=0}^{\infty} \frac{(\mathrm{ad}\,X)^k}{k!}Y = Y + [X, Y] + R',$$

where the remainder R' is the sum of all terms with $k \geq 2$ and thus belongs to $\mathfrak{w}_{j+2}$. As $h^{-1} = \exp(-Y)$, the Baker-Campbell-Hausdorff formula gives

$$\exp^{-1}([g, h]) = P(Y + [X, Y] + R', -Y) = [X, Y] + R' + R'',$$

where R'' is given by the double sum in (2.6). Now, thinking each term of this double sum as a $(k_1 + \ell_1 + \cdots + k_p + \ell_p + 1)$-multilinear function (and expanding each factor containing $Y + [X, Y] + R'$ accordingly), we obtain that R'' is a linear combination of elements of the form

$$(\mathrm{ad}\,Z_1) \cdots (\mathrm{ad}\,Z_k)Z_{k+1},$$

where $k \geq 1$ and $Z_i \in \{Y, [X, Y], R'\}$. Those elements where only Y appears vanish, while the other terms belong to $\mathfrak{w}_{j+2}$, since $[X, Y], R' \in \mathfrak{w}_{j+1}$ and $k \geq 1$. We deduce that $R'' \in \mathfrak{w}_{j+2}$. Finally,

$$\pi_{j+1}([g, h]) = \overline{\pi}_{j+1}([X, Y] + R' + R'') = \overline{\pi}_{j+1}([X, Y]) = [\overline{\pi}(X), \overline{\pi}_j(Y)],$$

since $X = \overline{\pi}(X) + R_X$ and $Y = \overline{\pi}_j(Y) + R_Y$, with $R_X \in \mathfrak{w}_2$ and $R_Y \in \mathfrak{w}_{j+1}$. □

Hereafter, we adopt the short notation $\gamma|_a^b := \gamma(a)^{-1}\gamma(b)$.

Lemma 4.9. Under the assumptions and notation of Definition 4.5, the displacement is given by the formula

$$\mathrm{Dis}(\gamma; [s, s'], Y) = C_{\gamma|_b^s}\left(\left[\exp(Y), \gamma|_s^{s'}\right]\right). \tag{4.19}$$

In particular, if $Y \in \mathfrak{g}_j$ and $1 \le j < s$, then $\mathrm{Dis}(\gamma; [s, s'], Y) \in G_{j+1}$ and

$$\pi_{j+1}\left(\mathrm{Dis}(\gamma; [s, s'], Y)\right) = [Y, \underline{\gamma}(s') - \underline{\gamma}(s)].$$

Proof. We have

$$\begin{aligned} \mathrm{Cor}(\gamma; [s, s'], Y)(b''') &= \gamma(s) \exp(Y)\, \gamma|_s^{s'} \exp(-Y)\, \gamma|_{s'}^b \\ &= \gamma(s)\left[\exp(Y), \gamma|_s^{s'}\right] \gamma|_s^{s'}\, \gamma|_{s'}^b \\ &= \gamma(s)\left[\exp(Y), \gamma|_s^{s'}\right] \gamma|_s^b, \end{aligned}$$

hence

$$\mathrm{Dis}(\gamma; [s, s'], Y)) = \gamma|_b^s\left[\exp(Y), \gamma|_s^{s'}\right](\gamma|_b^s)^{-1} = C_{\gamma|_b^s}\left(\left[\exp(Y), \gamma|_s^{s'}\right]\right).$$

By Lemma 4.6, we have $\pi(\gamma|_s^{s'}) = \underline{\gamma}(s') - \underline{\gamma}(s)$; moreover, $\pi_j(\exp(Y)) = Y$. Hence, using Lemma 4.8, we obtain

$$\left[\exp(Y), \gamma|_s^{s'}\right] \in G_{j+1} \quad \text{and} \quad \pi_{j+1}\left(\left[\exp(Y), \gamma|_s^{s'}\right]\right) = [Y, \underline{\gamma}(s') - \underline{\gamma}(s)].$$

The lemma now follows from equation (4.19) and Lemma 4.7. $\square$

Definition 4.10 (Iterated correction)**.** Let $\gamma : I \to G$ be a horizontal curve parametrized by arclength on the interval I and let $I_1 := [s_1, t_1], \dots, I_k := [s_k, t_k] \subseteq I$ be subintervals with $t_i \le s_{i+1}$. For any $Y_1, \dots, Y_k \in \mathfrak{g}$ we define by induction on $k \ge 2$ the iterated correction

$$\mathrm{Cor}(\gamma; I_1, Y_1; \dots; I_k, Y_k) := \mathrm{Cor}(\mathrm{Cor}(\gamma; I_1, Y_1; \dots; I_{k-1}, Y_{k-1}); I_k + 2\textstyle\sum_{i<k}\ell_{Y_i}, Y_k).$$

The iterated correction is a curve defined on the interval $[a, \widehat{b}]$, with $\widehat{b} := b + 2\sum_{i=1}^k \ell_{Y_i}$. The displacement of the final point produced by this iterated correction is

$$\mathrm{Dis}(\gamma; I_1, Y_1; \dots; I_k, Y_k) := \gamma(b)^{-1}\,\mathrm{Cor}(\gamma; I_1, Y_1; \dots; I_k, Y_k)(\widehat{b}).$$

Corollary 4.11. For any $I_i = [s_i, t_i] \subseteq I$ and $Y_i \in \mathfrak{g}_j$, with $i = 1, \dots, k$ and $j < s$, we have

$$\mathrm{Dis}(\gamma; I_1, Y_1; \dots; I_k, Y_k) \in G_{j+1} \tag{4.20}$$

and

$$\pi_{j+1}\left(\mathrm{Dis}(\gamma; I_1, Y_1; \dots; I_k, Y_k)\right) = \sum_{i=1}^k [Y_i, \underline{\gamma}(t_i) - \underline{\gamma}(s_i)]. \tag{4.21}$$

Proof. We prove (4.21) by induction on k. The case $k=1$ is in Lemma 4.9. Assume the formula holds for $k-1$. Letting $\widehat{\gamma} := \mathrm{Cor}(\gamma; I_1, Y_1; \dots; I_{k-1}, Y_{k-1})$, which is defined on the interval $[a, \widehat{b}]$ (where $\widehat{b} := b + 2\sum_{i<k} \ell_{Y_i}$), we have

$$\begin{aligned}
\mathrm{Dis}(\gamma; I_1, Y_1; \dots; I_k, Y_k) &= \gamma(b)^{-1} \mathrm{Cor}(\widehat{\gamma}; I_k + (\widehat{b} - b), Y_k) \\
&= \gamma(b)^{-1}\, \widehat{\gamma}(\widehat{b})\ \mathrm{Dis}(\widehat{\gamma}; I_k + (\widehat{b} - b), Y_k) \\
&= \mathrm{Dis}(\gamma; I_1, Y_1; \dots; I_{k-1}, Y_{k-1})\ \mathrm{Dis}(\widehat{\gamma}; I_k + (\widehat{b} - b), Y_k).
\end{aligned}$$

Then, by Lemma 4.6, by the inductive assumption and by Lemma 4.9 applied to $\widehat{\gamma}$ we have

$$\begin{aligned}
&\pi_{j+1}(\mathrm{Dis}(\gamma; I_1, Y_1; \dots; I_k, Y_k)) \\
&= \sum_{i=1}^{k-1} [Y_i, \underline{\gamma}(t_i) - \underline{\gamma}(s_i)] + [Y_k, \underline{\widehat{\gamma}}(t_k + (\widehat{b} - b)) - \underline{\widehat{\gamma}}(s_k + (\widehat{b} - b))] \\
&= \sum_{i=1}^{k} [Y_i, \underline{\gamma}(t_i) - \underline{\gamma}(s_i)],
\end{aligned}$$

because $\underline{\widehat{\gamma}}(t_k + (\widehat{b} - b)) - \underline{\widehat{\gamma}}(s_k + (\widehat{b} - b)) = \underline{\gamma}(t_k) - \underline{\gamma}(s_k)$. □

When dealing with curves γ defined on symmetric intervals, it is convenient to use modified versions of Cut and Cor, which we will denote by $\mathrm{Cut}'(\gamma; [s, s'])$ and $\mathrm{Cor}'(\gamma; [s, s'], Y)$. They are obtained from $\mathrm{Cut}(\gamma; [s, s'])$ and $\mathrm{Cor}(\gamma; [s, s'], Y)$ by composition with the time translation such that the new domain is a symmetric interval. The iterated correction is then defined in the following way:

$$\mathrm{Cor}'(\gamma; I_1, Y_1, \dots; I_k, Y_k) := \mathrm{Cor}'(\mathrm{Cor}'(\gamma; I_1, Y_1; \dots; I_{k-1}, Y_{k-1}); I_k + \textstyle\sum_{i<k} \ell_{Y_i}, Y_k).$$

The related displacement satisfies the properties (4.20) and (4.21) of Corollary 4.11 with Cor' replacing Cor.

5. Proof of the main results

Let G be a Carnot group with rank $r \geq 2$ and step s, and let $\mathscr{X} = \{X_1, \dots, X_r\}$ be an orthonormal basis for $\mathfrak{g}_1$ (recall that $\mathfrak{g}$ is endowed with a scalar product such that $\mathfrak{g}_i \perp \mathfrak{g}_j$). We first prove the one-sided version of Theorem 1.2; we will illustrate later how to adapt the proof in order to obtain Theorem 1.2.

Theorem 5.1. Let $\gamma : [0, T] \to G$, $T > 0$, be a length-minimizing curve parametrized by arclength. Then there exists an infinitesimal sequence $\eta_i \downarrow 0$ such that

$$\lim_{i \to \infty} \mathrm{Exc}(\gamma; [0, \eta_i]) = 0. \tag{5.22}$$

Proof. Step 1. We can assume that $\gamma(0) = 0$. Suppose by contradiction that there exists $\varepsilon > 0$ such that $\mathrm{Exc}(\gamma; [0, t]) \geq \varepsilon$ for any sufficiently small $t > 0$. For $k =$

$1, \dots, s$, we inductively define horizontal curves $\gamma^{(k)} : [0, T_k] \to G$ parametrized by arclength such that:

(i) $\gamma^{(k)}(0) = \gamma(0) = 0$;
(ii) $\gamma(T)^{-1}\gamma^{(k)}(T_k) \in G_{k+1}$, where $G_{s+1} = \{0\}$;
(iii) $L\left(\gamma^{(k)}\right) < L(\gamma)$, i.e., $T_k < T$.

In particular, $\gamma^{(s)}$ is a horizontal curve with the same endpoints as γ, but with smaller length: this contradicts the minimality of γ.

We define $\gamma^{(1)} := \mathrm{Cut}(\gamma; [0, \eta])$, where the parameter $\eta > 0$ will be chosen later; in fact, any sufficiently small η will work. In this proof, the notation $O(\cdot)$ and $o(\cdot)$ is used for asymptotic estimates which hold as $\eta \to 0$. By Remarks 4.2 and 4.3, $\gamma^{(1)}$ satisfies (i), (ii) and (iii) with $k = 1$.

Step 2. Let us fix parameters $\beta > 0$ and $\varrho_1 := 1 > \varrho_2 > \dots > \varrho_s > 0$ such that for all $k = 1, \dots, s-1$

$$\frac{(k+1)\varrho_k - \varrho_{k+1}}{k} > 1 + \beta. \tag{5.23}$$

This is possible if β is small enough: indeed, the inequality (5.23) is equivalent to

$$\varrho_k > \frac{\varrho_{k+1} + k}{k+1} + \frac{k}{k+1}\beta,$$

and we can choose any $\varrho_s \in (0,1)$ and then $\varrho_{s-1} < 1$ so as to verify the (strict) inequality when $\beta = 0$ and $k = s-1$, then ϱ_{s-2} similarly and so on. By continuity, the inequalities will still hold for a small enough $\beta > 0$.

For any $k = 1, \dots, s-1$, we define $I_k := [0, \eta^{\varrho_k}]$; the curve $\gamma^{(k+1)}$ is defined from $\gamma^{(k)}$ by applying several correction devices within I_{k+1}, see (5.26). As soon as $\eta \le 1$, there holds $[0, \eta] = I_1 \subseteq I_2 \subseteq \dots \subseteq I_{s-1}$.

By Lemma 4.4, since $\mathrm{Exc}(\gamma; [0, \eta]) \ge \varepsilon$, the gain of length obtained by performing the cut is

$$L(\gamma) - L(\gamma^{(1)}) \ge \frac{\eta\varepsilon^2}{2} \ge \eta^{1+\beta},$$

provided η is small enough.

The curves $\gamma^{(k)} : [0, T_k] \to G$ will be constructed inductively so as to satisfy (i), (ii) and (iii), as well as the following additional technical properties, which hold for $\gamma^{(1)}$:

(iv) $T_k \ge T_{k-1}$ if $k \ge 2$;
(v) $L\left(\gamma^{(k)}\right) \le L(\gamma) - (1 + o(1))\eta^{1+\beta}$;
(vi) $\underline{\gamma^{(k)}}(t) = \underline{\gamma}(t + (T - T_k))$ for any $t \in [2\eta^{\varrho_k}, T_k]$, i.e., on $[2\eta^{\varrho_k}, T_k]$ the curve $\gamma^{(k)}$ has the same projection on $\mathfrak{g}_1$ as the corresponding final piece of γ;
(vii) $\left\| \underline{\gamma^{(k)}} - \underline{\gamma}\big|_{[0,T_k]} \right\|_{L^\infty} = O(\eta)$.

Notice that (v) implies (iii) for small enough η.

Step 3. Assume that $\gamma^{(k)}$ has been constructed, for some $1 \leq k \leq s-1$, in such a way that (i)–(vii) hold. By (ii), there exists $E_k \in \mathfrak{g}_{k+1} \oplus \cdots \oplus \mathfrak{g}_s$ such that

$$\gamma(T)^{-1}\gamma^{(k)}(T_k) = \exp(E_k).$$

Let us estimate $\overline{\pi}_{k+1}(E_k)$. First, by (vi) and the uniqueness of horizontal lifts, we have

$$\gamma^{(k)}|_{2\eta^{\varrho_k}}^{T_k} = \gamma|_{\tau_k}^{T}, \qquad \text{where } \tau_k := 2\eta^{\varrho_k} + (T - T_k).$$

Hence, defining $g_k := \gamma(\tau_k)^{-1}\gamma^{(k)}(2\eta^{\varrho_k})$, we have

$$\begin{aligned}\gamma^{(k)}(T_k) =& \gamma^{(k)}(2\eta^{\varrho_k})\ \gamma^{(k)}\big|_{2\eta^{\varrho_k}}^{T_k} = \gamma(\tau_k) g_k\ \gamma|_{\tau_k}^{T} \\ =& \gamma(\tau_k)\ \gamma|_{\tau_k}^{T}\, C_{\gamma|_T^{\tau_k}}(g_k) = \gamma(T) C_{\gamma|_T^{\tau_k}}(g_k),\end{aligned}$$

i.e., $g_k = C_{\gamma|_{\tau_k}^{T}}(\gamma(T)^{-1}\gamma^{(k)}(T_k))$. By (ii), from Lemma 4.7 we obtain $g_k \in G_{k+1}$ and

$$\overline{\pi}_{k+1}(E_k) = \pi_{k+1}\left(\gamma(T)^{-1}\gamma^{(k)}(T_k)\right) = \pi_{k+1}(g_k) = O\left(\eta^{(k+1)\varrho_k}\right). \tag{5.24}$$

The last estimate follows from Remark 2.6 applied to the curve

$$\gamma(\tau_k)^{-1}\ \gamma|_{[0,\tau_k]}(\tau_k - \cdot) * \gamma^{(k)}\big|_{[0,2\eta^{\varrho_k}]},$$

which connects 0 to $\gamma(\tau_k)^{-1}\gamma^{(k)}(2\eta^{\varrho_k})$. Its length is $\tau_k + 2\eta^{\varrho_k}$ and is controlled by $5\eta^{\varrho_k}$ because, by (iv),

$$T - T_k \leq T - T_1 = L(\gamma) - L(\gamma^{(1)}) \leq \eta \leq \eta^{\varrho_k}.$$

Step 4. We now define $\gamma^{(k+1)}$. As $\mathfrak{g}_{k+1} = [\mathfrak{g}_k, \mathfrak{g}_1]$, using estimate (5.24) for $\overline{\pi}_{k+1}(E_k)$, there exist $Y_1, \ldots, Y_r \in \mathfrak{g}_k$ such that

$$\overline{\pi}_{k+1}(E_k) = \sum_{i=1}^{r}[Y_i, X_i] \quad \text{and} \quad |Y_1|, \ldots, |Y_r| = O\left(\eta^{(k+1)\varrho_k}\right). \tag{5.25}$$

Furthermore, we have $\mathrm{Exc}(\gamma; I_{k+1}) \geq \varepsilon$ whenever η is small enough. We can then apply Lemma 3.6 to I_{k+1} and find $[a_1, b_1], \ldots, [a_r, b_r] \subseteq I_{k+1}$ (with $b_i \leq a_{i+1}$) such that

$$\left|\det\left(\underline{\gamma}(b_1) - \underline{\gamma}(a_1), \ldots, \underline{\gamma}(b_r) - \underline{\gamma}(a_r)\right)\right| \geq c\eta^{r\varrho_{k+1}}.$$

Using (vii) we obtain, for small η,

$$\begin{aligned}\left|\det\left(\underline{\gamma^{(k)}}(b_1) - \underline{\gamma^{(k)}}(a_1), \ldots, \underline{\gamma^{(k)}}(b_r) - \underline{\gamma^{(k)}}(a_r)\right)\right| &\geq c\eta^{r\varrho_{k+1}} - O(\eta^{1+(r-1)\varrho_{k+1}}) \\ &\geq \frac{c}{2}\eta^{r\varrho_{k+1}}.\end{aligned}$$

This implies that for $i = 1, \ldots, r$ we have

$$X_i = \sum_{j=1}^{r} c_{ij}\left(\underline{\gamma^{(k)}}(b_j) - \underline{\gamma^{(k)}}(a_j)\right),$$

where $|c_{ij}| = O\left(\eta^{-\varrho_{k+1}}\right)$. This estimate depends on c and thus on ε. So, defining $Z_i := \sum_{j=1}^r c_{ji} Y_j$, from (5.25) we obtain

$$\overline{\pi}_{k+1}(E_k) = \sum_{i=1}^r [Z_i, \underline{\gamma^{(k)}}(b_i) - \underline{\gamma^{(k)}}(a_i)],$$

with $|Z_i| = O\left(\eta^{(k+1)\varrho_k - \varrho_{k+1}}\right)$. Finally, we let

$$\gamma^{(k+1)} := \mathrm{Cor}(\gamma^{(k)}; [a_1, b_1], -Z_1; \dots; [a_r, b_r], -Z_r). \tag{5.26}$$

Since $d(0, \exp(Z)) = O(|Z|^{1/k})$ for $Z \in \mathfrak{g}_k$, the extra length $T_{k+1} - T_k$ needed for the application of these r correction devices is

$$T_{k+1} - T_k = \sum_{i=1}^r O\left(|Z_i|^{1/k}\right) = O\left(\eta^{\frac{(k+1)\varrho_k - \varrho_{k+1}}{k}}\right) = o\left(\eta^{1+\beta}\right),$$

thanks to the inequalities (5.23) on the parameters ϱ_k. Thus, we obtain

$$L(\gamma^{(k+1)}) \le L(\gamma^{(k)}) + o(\eta^{1+\beta}).$$

Step 5. We check that $\gamma^{(k+1)}$ satisfies properties (i)-(vii). We have just verified (iii) and (v), while (i) and (iv) are trivial. The property (vii) follows from the fact that $\gamma^{(k+1)}$ (as well as $\underline{\gamma^{(k+1)}}$) is obtained from $\gamma^{(k)}$ (from $\underline{\gamma^{(k)}}$) by the application of correction devices of total length $o(\eta^{1+\beta})$.

In order to check (vi), we remark that

$$\underline{\gamma^{(k+1)}} = \underline{\gamma^{(k+1)}}\Big|_{[0, \eta^{\varrho_{k+1}} + (T_{k+1} - T_k)]} * \underline{\gamma^{(k)}}\Big|_{[\eta^{\varrho_{k+1}}, T_k]}$$

and that the final point of the first curve in this concatenation coincides with the starting point of the second one. Since $T_{k+1} - T_k = O\left(\eta^{\frac{(k+1)\varrho_k - \varrho_{k+1}}{k}}\right) = o\left(\eta^{\varrho_{k+1}}\right)$, if η is small enough we obtain

$$\begin{aligned}
\underline{\gamma^{(k+1)}}\Big|_{[2\eta^{\varrho_{k+1}}, T_{k+1}]} &= \underline{\gamma^{(k)}}\Big|_{[2\eta^{\varrho_{k+1}} - (T_{k+1} - T_k), T_k]} (\,\cdot\, - (T_{k+1} - T_k)) \\
&= \underline{\gamma}\Big|_{[2\eta^{\varrho_{k+1}} + (T - T_{k+1}), T]} (\,\cdot\, + (T - T_{k+1})),
\end{aligned}$$

the last equality holding because $2\eta^{\varrho_{k+1}} - (T_{k+1} - T_k) \ge 2\eta^{\varrho_k}$ when η is small. Thus, $\gamma^{(k+1)}$ satisfies (vi).

Finally, let us check (ii). By Lemma 4.6 and Corollary 4.11, we have

$$\gamma(T)^{-1}\gamma^{(k+1)}(T_{k+1}) = \left(\gamma(T)^{-1}\gamma^{(k)}(T_k)\right)\left(\gamma^{(k)}(T_k)^{-1}\gamma^{(k+1)}(T_{k+1})\right) \in G_{k+1}$$

and

$$\begin{aligned}
\pi_{k+1}\left(\gamma(T)^{-1}\gamma^{(k+1)}(T_{k+1})\right) &= \pi_{k+1}(\exp(E_k)) + \pi_{k+1}\left(\gamma^{(k)}(T_k)^{-1}\gamma^{(k+1)}(T_{k+1})\right) \\
&= \overline{\pi}_{k+1}(E_k) + \sum_{i=1}^r [-Z_i, \underline{\gamma}(b_i) - \underline{\gamma}(a_i)] \\
&= 0.
\end{aligned}$$

This concludes the proof. □

We now prove Theorem 1.2. The proof is basically the same as that of Theorem 5.1 and we just list the required minor modifications below.

Proof of Theorem 1.2. The constraints imposed on the curves $\gamma^{(k)}$, as well as the cut and correction operations, have to be replaced by their symmetric counterparts. For $k = 1, \ldots, s$ we inductively construct horizontal curves $\gamma^{(k)} : [-T_k, T_k] \to G$ parametrized by arclength satisfying:

(i') $\gamma^{(k)}(-T_k) = \gamma(-T)$;
(ii') $\gamma(T)^{-1}\gamma^{(k)}(T_k) \in G_{k+1}$ (in particular, $\gamma^{(s)}(T_s) = \gamma(T)$);
(iii') $L\left(\gamma^{(k)}\right) < L(\gamma)$, i.e., $T_k < T$;
(iv') $T_k \geq T_{k-1}$ if $k \geq 2$;
(v') $L\left(\gamma^{(k)}\right) \leq L(\gamma) - (1 + o(1))\eta^{1+\beta}$;
(vi') $\underline{\gamma^{(k)}}\big|_{[2\eta^{\varrho_k}, T_k]} = \underline{\gamma}\big|_{[2\eta^{\varrho_k}+(T-T_k), T]}(\,\cdot\, + (T - T_k))$ and
$\underline{\gamma^{(k)}}\big|_{[-T_k, -2\eta^{\varrho_k}]} = \underline{\gamma}\big|_{[-T, -2\eta^{\varrho_k}-(T-T_k)]}(\,\cdot\, - (T - T_k))$;
(vii') $\left\|\underline{\gamma^{(k)}} - \underline{\gamma}\big|_{[-T_k, T_k]}\right\|_\infty = O(\eta)$.

We list the necessary modifications in the various steps.

Step 1. The first curve is $\gamma^{(1)} := \mathrm{Cut}'(\gamma; [-\eta, \eta])$, which satisfies (i')–(vii') for $k = 1$.

Step 2. The interval I_k is now $[-\eta^{\varrho_k}, \eta^{\varrho_k}]$.

Step 3. Let E_k, τ_k, g_k be defined as in the proof of Theorem 5.1. The estimate $\pi_{k+1}(g_k) = O\left(\eta^{(k+1)\varrho_k}\right)$ follows by applying Remark 2.6 to the curve

$$\gamma(\tau_k)^{-1}\ \gamma|_{[-\tau_k, \tau_k]}(\tau_k - \cdot) * \gamma^{(k)}\big|_{[-2\eta^{\varrho_k}, 2\eta^{\varrho_k}]} \tag{5.27}$$

and observing that $\gamma(-\tau_k) = \gamma^{(k)}(-2\eta^{\varrho_k})$. The length of the curve in (5.27) is $2\tau_k + 4\eta^{\varrho_k} \leq 10\eta^{\varrho_k}$.

Step 4. In the definition (5.26) of $\gamma^{(k+1)}$, Cor is replaced by Cor$'$.

Step 5. The fact that $\gamma^{(k+1)}$ satisfies (vi') follows from the identity

$$\underline{\gamma^{(k+1)}} = \underline{\gamma^{(k)}}\big|_{[-T_k, -\eta^{\varrho_{k+1}}]} * \underline{\gamma^{(k+1)}}\big|_{[-\eta^{\varrho_{k+1}}-(T_{k+1}-T_k), \eta^{\varrho_{k+1}}+(T_{k+1}-T_k)]} * \underline{\gamma^{(k)}}\big|_{[\eta^{\varrho_{k+1}}, T_k]},$$

where the final point of each curve in the concatenation coincides with the starting point of the next one. □

We finally prove Theorem 1.1 and then state its one-sided version.

Proof of Theorem 1.1. As explained in the introduction, by Propositions 2.5, 2.7, 2.8 and 2.11 it is not restrictive to assume that the Carnot-Carathéodory structure $(M, \mathscr{X})$ is that of a Carnot group G; moreover, we can assume that $t = 0$, $\gamma(0) = 0$.

Let $\eta_i \downarrow 0$ be the sequence provided by Theorem 1.2. Since $\mathrm{Exc}(\gamma; [-\eta_i, \eta_i]) \to 0$, we can find a sequence $\zeta_i \downarrow 0$ satisfying

$$\zeta_i^{-1/2}\,\mathrm{Exc}(\gamma; [-\eta_i, \eta_i]) \to 0.$$

Let us set $\lambda_i := \zeta_i\eta_i \downarrow 0$. Up to subsequences, using Lemma 3.5 and a diagonal argument, we can assume that there exists a length minimizer $\gamma_\infty : \mathbb{R} \to G$ parametrized by arclength such that

$$\gamma_i(t) := \delta_{\lambda_i^{-1}}(\gamma(\lambda_i t)) \to \gamma_\infty(t),$$

uniformly on compact subsets of $\mathbb{R}$, and that $\dot\gamma_i \to \dot\gamma_\infty$ in $L^2_{loc}(\mathbb{R})$. For any fixed $N > 0$ we have by Remark 3.3

$$\mathrm{Exc}\,(\gamma_i; [-N, N]) = \mathrm{Exc}(\gamma; [-N\zeta_i\eta_i, N\zeta_i\eta_i]) \le (N\zeta_i)^{-1/2}\,\mathrm{Exc}(\gamma; [-\eta_i, \eta_i]) \to 0,$$

the last inequality being true for any i such that $N\zeta_i \le 1$. So, by Remark 3.4, we deduce that $\mathrm{Exc}(\gamma_\infty; [-N, N]) = 0$, which means that $\underline{\dot\gamma_\infty}(t)$ is contained in a hyperplane $\mathfrak{h}_1$ of $\mathfrak{g}_1$ for a.e. $t \in [-N, N]$. Since this is true for any N, we deduce that there exists a hyperplane $\mathfrak{h}_1$ of $\mathfrak{g}_1$ such that $\underline{\dot\gamma_\infty}(t) \in \mathfrak{h}_1$ for a.e. $t \in \mathbb{R}$; in particular, γ_∞ is contained in the Carnot subgroup H associated with the Lie algebra generated by $\mathfrak{h}_1$.

If the rank of G is $r = 2$, we conclude that γ_∞ is contained in a one-parameter subgroup of G. Since $\gamma_\infty \in \mathrm{Tan}(\gamma; 0)$ is a length minimizer parametrized by arclength, we deduce that γ_∞ is a line in G.

Otherwise, we can reason by induction on $r > 2$: since H has rank $r - 1$ and γ_∞ is a length minimizer in H parametrized by arclength, there exists $\widehat\gamma \in \mathrm{Tan}(\gamma_\infty; 0)$ such that $\widehat\gamma$ is a line in $H \subset G$. By Proposition 2.8 we have $\widehat\gamma \in \mathrm{Tan}(\gamma; 0)$ and the proof is accomplished. □

We state without proof the following version of Theorem 1.1, which holds for extremal points of length-minimizers. The proof uses the same arguments as the previous one and can be easily deduced from Theorem 5.1.

Theorem 5.2. *Let $\gamma : [0, T] \to M$ be a length minimizer parametrized by arclength in a Carnot-Carathéodory space (M, d). Then, each of the tangent cones $\mathrm{Tan}^+(\gamma; 0)$ and $\mathrm{Tan}^-(\gamma; T)$ contains a horizontal half-line.*

Remark 5.3. Theorem 1.1 admits the following reformulation in terms of the minimal control $h = (h_1, \dots, h_r)$ of a length-minimizer $\gamma : [-T, T] \to M$ (parametrized by arclength): for any $t \in (-T, T)$ there exist an infinitesimal sequence $\eta_i \downarrow 0$ and a constant unit vector $v \in S^{r-1}$ such that

$$h(t + \eta_i\,\cdot) \to v \qquad \text{in } L^2_{loc}(\mathbb{R}).$$

Of course, an analogous version holds for extremal points.

Let us prove this fact. Let the chart φ and the vector fields Y_i, Y_i^∞ be as in Section 2. We can assume $t = 0$. Theorem 1.1 provides a sequence $\eta_i \downarrow 0$ such that the curves $\gamma^i(\tau) := \delta_{1/\eta_i}\varphi(\gamma(\eta_i\tau))$ converge locally uniformly to a horizontal line κ, in some tangent CC structure $(M^\infty, \mathscr{X}^\infty)$. We have

$$\gamma^i(\tau) = \int_0^\tau \sum_{j=1}^r h_j(\eta_i s) Y_j^{1/\eta_i}(\gamma^i(s))\, ds,$$

where $Y_j^\lambda(x) := \lambda^{-1}(\delta_\lambda)_* Y_j(\delta_{1/\lambda}(x))$. Up to subsequences we have $h(\eta_i\,\cdot) \rightharpoonup h_\infty$ in $L^2_{loc}(\mathbb{R})$, with $\|h_\infty\|_\infty \leq 1$. Since $Y_j^{1/\eta_i} \to Y_j^\infty$ locally uniformly, we obtain

$$\kappa(\tau) = \int_0^\tau \sum_{j=1}^r h_\infty(s) Y_j^\infty(\kappa(s))\, ds.$$

By Proposition 2.7, κ is parametrized by arclength. So $|h_\infty| = 1$ a.e. and h_∞ is the minimal control of κ. Since κ is a line, h_∞ is constant. Finally, for any compact set $K \subset \mathbb{R}$, we trivially have $\|h(\eta_i\,\cdot)\|_{L^2(K)} \to \|h_\infty\|_{L^2(K)}$. This gives $h(\eta_i\,\cdot) \to h_\infty$ in $L^2(K)$.

References

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

[2] 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.

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

[4] E. Hakavuori & E. Le Donne, *Non-minimality of corners in subRiemannian geometry.* Invent. Math. (2016), online.

[5] H. Hermes, *Nilpotent and High-Order Approximations of Vector Field Systems.* SIAM Rev. 33 (1991), no. 2, 238–264.

[6] F. Jean, *Control of Nonholonomic Systems: from Sub-Riemannian Geometry to Motion Planning.* Springer Briefs in Mathematics. Springer, Cham, 2014.

[7] A. W. Knapp, *Lie groups beyond an introduction.* Progress in Mathematics, 140. Birkhäuser 2002.

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

[9] E. Le Donne, G. P. Leonardi, R. Monti & D. Vittone, *Extremal polynomials in stratified groups.* Comm. Anal. Geom., to appear. Preprint available at cvgmt.sns.it.

[10] E. Le Donne, G. P. Leonardi, R. Monti & D. Vittone, *Corners in non-equiregular sub-Riemannian manifolds.* ESAIM Control Optim. Calc. Var. 21 (2015), 625–634.

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

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

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

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

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

[16] L. Rifford, *Singulières minimisantes en géometrié sous-riemannienne*, Séminaire Bourbaki, Mars 2016 68ème année, 2015–2016, no. 1113.

[17] H. J. Sussmann, *A regularity theorem for minimizers of real-analytic subriemannian metrics.* Proceedings of the 53rd IEEE Conference on Decision and Control (2015), 4801–4806.

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

[19] D. Vittone, *The regularity problem for sub-Riemannian geodesics.* In: Geometric Measure Theory and Real Analysis. Ed. by Luigi Ambrosio. Pisa: Scuola Normale Superiore, 2014, pp. 193–226.

E-mail address: `monti@math.unipd.it`

E-mail address: `alessandro.pigati@math.ethz.ch`

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

(Monti and Vittone) Università di Padova, Dipartimento di Matematica, via Trieste 63, 35121 Padova, Italy

(Pigati) Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy

(Pigati) ETH Zürich, Department of Mathematics, Rämistrasse 101, 8092 Zürich, Switzerland