Generalized Jacobi identities and ball-box theorem for horizontally regular vector fields*

Annamaria Montanari Daniele Morbidelli

Abstract

Consider a family $\mathcal{H} := \{X_j =: f_j \cdot \nabla : j = 1, \dots, m\}$ of C^1 vector fields in $\mathbb{R}^n$ and let $s \in \mathbb{N}$. We assume that for all $p \in \{1 \dots, s\}$ and $j_1, \dots, j_p \in \{1, \dots, m\}$ the *horizontal derivatives* $X_{j_1} X_{j_2} \cdots X_{j_{p-1}} f_{j_p}$ exist and are Lipschitz continuous with respect to the control distance defined by $\mathcal{H}$. Then we show that different notions of commutator agree. This involves an accurate analysis of some algebraic identities involving nested commutators which seem to have an independent interest.

Our principal applications are a ball-box theorem, the doubling property and the Poincaré inequality for Hörmander vector fields under an intrinsic "horizontal regularity" assumption on their coefficients.

Contents

1. Introduction and main results

In this paper we study the notion of higher order commutator for a given family $\mathcal{H} = \{X_1, \dots, X_m\}$ of vector fields in $\mathbb{R}^n$. Our main issue is to discuss to what extent the notion of higher order commutator can be extended to vector fields $X_j \in C^1_{\mathrm{Euc}}$ whose higher order derivatives are assumed to have regularity only along the "horizontal directions" provided by the family $\mathcal{H}$. The main application of such study consists of a discussion of a class of *almost exponential maps* under very low, intrinsic regularity assumptions which is carried out in [MM12a]. This enables us to prove a boll-box theorem, the doubling property and the Poincaré inequality for vector fields satisfying the Hörmander's bracket-generating condition of step $s \geq 1$ under very low regularity requirements.

*2010 Mathematics Subject Classification: 53C17. Key words and Phrases: Jacobi identities, Lie derivatives. Horizontal regularity, Ball-box theorem, Poincaré inequality

To understand the problem, which starts to appear for commutators of length three, assume that $X_1, \dots, X_m$ are vector fields of class $C^1_{\rm Euc}$, i.e. of class C^1 in the Euclidean sense. Write $X_i = f_i \cdot \nabla$ for $i = 1, \dots, m$. The definition of commutarors of length two is clear, namely we set

$$X_{jk} := (X_j^\sharp f_k - X_k^\sharp f_j) \cdot \nabla =: f_{jk} \cdot \nabla \quad \text{for all } j, k \in \{1, \dots, m\}$$

and $X_{jk}\psi := f_{jk} \cdot \nabla\psi$ for $\psi \in C^1_{\rm Euc}(\mathbb{R}^n)$. Here we denote by $X^\sharp f(x) := \lim_{t\to 0} \frac{1}{t}(f(e^{tX}x) - f(x))$ the Lie derivative along a vector field $X \in C^1_{\rm Euc}$ of a scalar funciton f (such unusual notation will be convenient for our purposes).

Passing to length three, we have two alternatives. For each i, j, k, we can define either

$$X_{ijk} := [X_i, [X_j, X_k]] := (X_i^\sharp X_j^\sharp f_k - X_i^\sharp X_k^\sharp f_j - X_j^\sharp X_k^\sharp f_i + X_k^\sharp X_j^\sharp f_i) \cdot \nabla, \tag{1.1}$$

or

$$\operatorname{ad}_{X_i} X_{jk} := (X_i^\sharp f_{jk} - X_{jk} f_i) \cdot \nabla. \tag{1.2}$$

Both operators act on $C^1_{\rm Euc}$ functions. The first one is the most natural and symmetric (for instance one gets the Jacobi identity for free). The second one appears in some useful non commutative calculus formulae which play a key role in our work, see Theorem 2.6. It is rather easy to see that $[X_i, [X_j, X_k]] = \operatorname{ad}_{X_i} X_{jk}$, if the involved vector fields are $C^2_{\rm Euc}$, so that Euclidean second order derivatives commute (here end hereafter by $C^k_{\rm Euc}$ we denote Euclidean C^k regularity). In this paper we are able to show that this regularity is not necessary. Indeed, if we denote by $C^{2,1}_{\mathcal{H},\rm loc}$ all functions f which have two horizontal derivatives and such that for all $i, j \in \{1, \dots, m\}$ the function $X_i^\sharp X_j^\sharp f$ is locally Lipschitz with respect to the distance associated to the vector fields in $\mathcal{H}$, then we have

Theorem 1.1 (see Theorem 3.1 for a higher order statement). *Let $\mathcal{H} = \{X_1, \dots, X_m\}$ be a family of $C^1_{\rm Euc}$ vector fields. Write $X_j = f_j \cdot \nabla$ and assume that $f_j \in C^{2,1}_{\mathcal{H},\rm loc}$ for all $j \in \{1, \dots, m\}$. Then we have*

$$[X_i, [X_j, X_k]] = \operatorname{ad}_{X_i} X_{jk} \quad \textit{for all } i, j, k \in \{1, \dots, m\}.$$

In Theorem 3.1 we give the general version of this statement which involves commutators of arbitrary length $s \geq 1$, where the vector fields f_j belong to the class $C^1_{\rm Euc} \cap C^{s-1,1}_{\mathcal{H},\rm loc}$ introduced in Definition 2.1.

Although the analysis of statements like Theorem 1.1—together with the techniques we develop in the proof—may have some independent interest, the strong motivation why we need to analyze operators like $\operatorname{ad}_{X_i} X_{jk}$ comes from their natural appearance in the noncommutative formulas of Theorem 2.6 and ultimately in the theory of differentiation of the almost exponential maps E introduced below. Let us mention that in [MM12a] we prove a higher order orbit theorem for families $\mathcal{H}$ in such class; [MM12a, Example 3.14] shows that our regularity classes capture examples which do not fall in the classical framework.

In order to show that (1.1) and (1.2) agree, we need to analyze carefully the algebraic properties of the coefficients appearing in the expansion of nested commutators as sums of higher order derivatives. In particular, we exploit some higher order algebraic identities which we denote as "generalized Jacobi identities"; see Proposition 3.3 and see also Proposition 3.7. It is interesting to observe that some of those identities, specialized to particular situations, give the proof of some old nested commutators identities going back to Baker and discussed in [Ote91]. This is discussed in Subsection 3.1. We believe that these algebraic features may have some independent interest.

From an historical point of view, let us mention that for commutators of length two, notions of nonsmooth Lie brackets have been studied deeply by Rampazzo and Sussmann [RS07]. The notion of set-valued commutator studied in [RS07] concerns vector fields which are quite less regular than ours, actually Lipschitz continuous only, but this approach does not provide a quantitative knowledge of control balls or Poincaré inequalities. Moreover, the notion of set-valued commutator is not clear, if the length exceeds two; see the counterexample in [RS07, Section 7.2]. We work here at a slightly more comfortable level of regularity, which ensures that commutators are pointwise defined and "horizontally" Lipschitz continuous, which will be sufficient to obtain some good information on control balls.

Next we discuss our applications to sub-Riemannian geometry. Let $\mathcal{H} = \{X_1, \dots, X_m\}$ be a family of vector fields and assume that $X_j \in C^1_{\mathrm{Euc}} \cap C^{s-1,1}_{\mathcal{H},\mathrm{loc}}$ for some $s \in \mathbb{N}$. Denote by $B_{\mathrm{cc}}(x,r)$ the Carnot–Carathéodory ball with center at x and radius r. Let $\mathcal{P} := \{Y_1, \dots, Y_q\}$ be the family of all nested commutators of length at most s. Let $\operatorname{length}(Y_j) =: \ell_j \le s$. Assume that $\mathcal{H}$ satisfies the Hörmander condition of step s, i.e. $\dim \operatorname{span}\{Y_j(x)\} = n$, for all $x \in \mathbb{R}^n$. To identify a family of n commutators, let us choose a multiindex $I = (i_1, \dots, i_n) \in \{1, \dots, q\}^n$. Given a radius $r > 0$, define the scaled commutators $\widetilde{Y}_{i_k} := r^{\ell_{i_k}} Y_{i_k}$ and the *almost exponential map*

$$E_{I,x,r}(h) := \exp_{\mathrm{ap}}(h_1 \widetilde{Y}_{i_1}) \cdots \exp_{\mathrm{ap}}(h_p \widetilde{Y}_{i_n}) x \tag{1.3}$$

for each h close to $0 \in \mathbb{R}^n$ (after passing to $\widetilde{Y}_{i_j}$, the variable h lives at a unit scale). See (4.8) for the definition of the *approximate exponential* $\exp_{\mathrm{ap}}$. Below, B_ϱ denotes the control ball defined by *all* commutators (with their degrees, see (4.5)), which trivially contains the Carnot–Carathéodory ball B_{cc} with same center and radius defined in (2.2). Then we have the following ball-box theorem and Poincaré inequality. A more detailed statement is contained in Section 4.

Theorem 1.2. *Let $\mathcal{H}$ be a family of vector fields in the class $C^1_{\mathrm{Euc}} \cap C^{s-1,1}_{\mathcal{H},\mathrm{loc}}$ for some s. Assume the Hörmander condition of step s and assume that $Y_j \in C^0_{\mathrm{Euc}}$ for all $Y_j \in \mathcal{P}$. Let $\Omega \subset \mathbb{R}^n$ be a bounded set. Then there is $C > 1$ such that the following holds. Let $x \in \Omega$ and take a positive radius $r < C^{-1}$. Then there is a subfamily $\{Y_{i_1}, \dots, Y_{i_n}\} \subset \mathcal{P}$ such that the map $E := E_{I,x,r}$ in (1.3) is C^1 in the Euclidean sense on the unit ball $B_{\mathrm{Euc}}(1) \subset \mathbb{R}^n$. Its Jacobian satisfies the estimate $C^{-1}|\det dE(0)| \le |\det dE(h)| \le C|\det dE(0)|$ and we have the ball-box inclusion*

$$E(B_{\mathrm{Euc}}(1)) \supseteq B_\varrho(x, C^{-1} r). \tag{1.4}$$

The map E is one-to-one on $B_{\mathrm{Euc}}(1)$ and we have

$$|B_{\mathrm{cc}}(x, 2r)| \le C |B_{\mathrm{cc}}(x,r)| \quad \text{for all } x \in \Omega \ r < C^{-1}. \tag{1.5}$$

Moreover, for any C^1 function f we have the Poincaré inequality

$$\int_{B_{\mathrm{cc}}(x,r)} |f(y) - f_{B_{\mathrm{cc}}(x,r)}| dy \le C \sum_{j=1}^m \int_{B_{\mathrm{cc}}(x,Cr)} |r X_j f(y)| dy. \tag{1.6}$$

It is well known that the doubling estimate and the Poincaré inequality are important tools in subelliptic PDEs, sub-Riemanninan geometry and analysis in metric spaces; see [FL83b, NSW85, Jer86, SC92, GN96, Che99, HK00]. Note that inequality (1.6) improves all previous versions of the Poincaré inequality from a regularity standpoint; compare [Jer86,LM00,BBP12a,MM12b,Man10]. Indeed, in such papers some higher order Euclidean regularity were assumed, whereas our higher order derivatives $X^\sharp_{j_1}\cdots X^\sharp_{j_{p-1}} f_{j_p}$ with $2\leq p\leq s$ are assumed to be horizontally Lipschitz continuous only. [1]

Let us mention that in [MM12a] and [MM11], relying on the results obtained here, we also prove an integrability result for orbits, a ball-box theorem and the Poincaré inequality in a setting where the Hörmander's condition is removed.

Our work in low regularity is also motivated by the appearance, in several recent papers, of subelliptic PDEs involving nonlinear first order operators. This happens for instance in several complex variables, while studying graphs with prescribed Levi curvature in $\mathbb{C}^n$, see [CLM02], or in the study of intrinsic regular hypersurfaces in Carnot groups, see [ASCV06], where vector fields with non Euclidean regularity naturally appear. These papers suggest that it would be desirable to remove even our assumption $X_j\in C^1_{\mathrm{Euc}}$ for the vector fields of the horizontal family $\mathcal{H}$. However, note that removing such assumption would destroy uniqueness of integral curves and dealing efficiently with our almost exponential maps would require nontrivial new ideas.

Before closing this introduction, we mention some recent papers where nonsmooth vector fields are discussed. In [SW06], the case of diagonal vector fields is discussed deeply. In the Hörmander case, in the model situation of equiregular families of vector fields, nonsmooth ball-box theorems have been studied by see [KV09, Gre10]. Finally, [BBP12b] contains a nonsmooth lifting theorem.

Acknowledgements. We thank Francesco Regonati, who helped us to formalize some of the questions we encountered in Section 3 in the language of polynomial identities.

2. Preliminary facts on horizontal regularity

2.1. Horizontal regularity classes

Vector fields and the control distance. Consider a family $\mathcal{H}=\{X_1,\dots,X_m\}$ of vector fields and assume that $X_j\in C^1_{\mathrm{Euc}}(\mathbb{R}^n)$ for all j. Here and later C^1_{Euc} means C^1 in the Euclidean sense. Write $X_j=: f_j\cdot\nabla$, where $f_j\colon\mathbb{R}^n\to\mathbb{R}^n$. The vector field X_j, evaluated at a point $x\in\mathbb{R}^n$, will be denoted by $X_{j,x}$ or $X_j(x)$. All the vector fields in this paper are always defined on the whole space $\mathbb{R}^n$.

Define the Franchi–Lanconelli distance [FL83a]

$$d(x,y):=\inf\Big\{r>0: y=e^{t_1Z_1}\cdots e^{t_\mu Z_\mu}x \text{ for some } \mu\in\mathbb{N} \text{ where } \sum|t_j|\leq 1 \text{ with } Z_j\in r\mathcal{H}\Big\}. \tag{2.1}$$

Here and hereafter we let $r\mathcal{H}:=\{rX_1,\dots,rX_m\}$ and $\pm r\mathcal{H}:=\{\pm rX_1,\dots,\pm rX_m\}$.

[1] Technically speaking, both the approaches adopted in in [BBP12a] and [MM12b]—via Euclidean Taylor approximation or Euclidean regularization—do not work in our situation, because one cannot prove that higher order commutators of mollified vector fields converge to mollified of the corresponding commutators.

Let also $d_{\rm cc}$ be the Fefferman–Phong and Nagel–Stein–Wainger distance [FP83,NSW85]

$$\begin{aligned} d_{\rm cc}(x,y) := \inf\big\{ r>0: \text{ there is } \gamma \in \mathrm{Lip_{Euc}}((0,1),\mathbb{R}^n) \text{ with } \gamma(0)=x \\ \gamma(1)=y \text{ and } \dot\gamma(t)\in \big\{ \textstyle\sum_{1\le j\le m} c_j X_{j,\gamma(t)} : |c|\le r\big\} \text{ for a.e. } t\in[0,1]\big\}. \end{aligned} \tag{2.2}$$

As usual, we call *Carnot–Carathéodory* or *control* distance the distance $d_{\rm cc}$. Note that in the definition of $d_{\rm cc}$ we may choose paths γ such that $\dot\gamma = \sum_j b_j(t) X_j(\gamma)$ where $b:(0,1)\to B_{\rm Euc}(0,r)$ is *measurable*, see Remark 4.2. In the present paper we shall make a prevalent use of the distance d. In the definition of both distances we agree that $d(x,y)=+\infty$ if there are no paths in the pertinent class which connect x and y.

Horizontal regularity classes. Here we define our notion of horizontal regularity in terms of the distance d. Note that we *do not* use the control distance $d_{\rm cc}$.

Definition 2.1. *Let $\mathcal{H} := \{X_1,,\dots,X_m\}$ be a family of vector fields, $X_j\in C^1_{\rm Euc}$. Let d be their distance (2.1) Let $g:\mathbb{R}^n\to\mathbb{R}$. We say that g is d-*continuous*, and we write $g\in C^0_{\mathcal{H}}(\mathbb{R}^n)$, if for all $x\in\mathbb{R}^n$, we have $|g(y)-g(x)|\to 0$, as $d(x,y)\to 0$. We say that $g:\mathbb{R}^n\to\mathbb{R}$ is $\mathcal{H}$-*Lipschitz* or d-*Lipschitz* in $A\subset\mathbb{R}^n$ if*

$$\mathrm{Lip}_{\mathcal{H}}(g;A) := \sup_{x,y\in A,\, x\neq y} \frac{|g(x)-g(y)|}{d(x,y)} < \infty.$$

We say that $g\in C^1_{\mathcal{H}}(\mathbb{R}^n)$ if the derivative $X_j^\sharp g(x) := \lim_{t\to 0}(f(e^{tX_j}x)-f(x))/t$ is a d-continuous function for any $j=1,\dots,m$. We say that $g\in C^k_{\mathcal{H}}(\mathbb{R}^n)$ if all the derivatives $X^\sharp_{j_1}\dots X^\sharp_{j_p} g$ are d-continuous for $p\le k$ and $j_1,\dots,j_p\in\{1,\dots,m\}$. If all the derivatives $X^\sharp_{j_1}\dots X^\sharp_{j_k} g$ are d-Lipschitz on each Ω bounded set in the Euclidean metric, then we say that $g\in C^{k,1}_{\mathcal{H},\rm loc}(\mathbb{R}^n)$. Finally, denote the usual Euclidean Lipschitz constant of g on $A\subset\mathbb{R}^n$ by $\mathrm{Lip_{Euc}}(g;A)$.

We will usually deal with vector fields which are of class at least $C^1_{\rm Euc}\cap C^{s-1,1}_{\mathcal{H},\rm loc}$, where $s\ge 1$ is a suitable integer. In this case it turns out that commutators up to the order s can be defined, see Definition 2.3 and Remark 2.5. It will take a quite hard work (the whole Section 3) to show that the different notions given in Definition 2.3 actually agree.

Definitions of commutator. Our purpose now is to show that, given a family $\mathcal{H}$ of vector fields with $X_j\in C^{s-1,1}_{\mathcal{H},\rm loc}\cap C^1_{\rm Euc}$, then commutators can be defined up to length s.

For any $\ell\in\mathbb{N}$, denote by $\mathcal{W}_\ell := \{w_1\cdots w_\ell : w_j\in\{1,\dots,m\}\}$ the words of length $|w|:=\ell$ in the alphabet $1,2,\dots,m$. Let also $\mathfrak{S}_\ell$ be the group of permutations of ℓ letters.

Definition 2.2 (Coefficients $\pi_\ell(\sigma)$). *Define $\pi_\ell:\mathfrak{S}_\ell\to\{-1,0,1\}$ as follows: let us agree that $\pi_1(\sigma)=1$ for the unique $\sigma\in\mathfrak{S}_1$. Then, for $\sigma\in\mathfrak{S}_2$, let $\pi_2(\sigma):=1$, if $\sigma(01)=01$ and $\pi_2(\sigma):=-1$, if $\sigma(01)=10$. Then, define inductively for $\ell\ge 2$,*

$$\begin{cases} \pi_{\ell+1}(\widetilde\sigma) := \pi_\ell(\sigma) & \text{if } \widetilde\sigma(01\cdots\ell) = 0\sigma(1\cdots\ell) \text{ and } \sigma\in\mathfrak{S}_\ell \\ \pi_{\ell+1}(\widetilde\sigma) := -\pi_\ell(\sigma) & \text{if } \widetilde\sigma(01\cdots\ell) = \sigma(1\cdots\ell)0 \text{ and } \sigma\in\mathfrak{S}_\ell \\ \pi_{\ell+1}(\widetilde\sigma) := 0 & \text{if } \widetilde\sigma_0(01\cdots\ell)\neq 0\neq \widetilde\sigma_\ell(01\cdots\ell). \end{cases} \tag{2.3}$$

Here we used the notation $\widetilde{\sigma}(01\cdots\ell) = \widetilde{\sigma}_0(01\cdots\ell)\widetilde{\sigma}_1(01\cdots\ell)\cdots\widetilde{\sigma}_\ell(01\cdots\ell)$. The coefficients π_ℓ are designed in order to write commutators in a convenient way. Indeed, if $A_1,\dots,A_m : V \to V$ are linear operators on a vector space V, then one can check inductively that, given a word $w = w_1\cdots w_\ell$, we have

$$[A_{w_1},[A_{w_2},\dots[A_{w_{\ell-1}},A_{w_\ell}]]\dots] = \sum_{\sigma\in\mathfrak{S}_\ell} \pi_\ell(\sigma)A_{\sigma_1(w)}A_{\sigma_2(w)}\cdots A_{\sigma_\ell(w)}. \tag{2.4}$$

We will benefit later of the property

$$\pi_\ell(\sigma_1\cdots\sigma_\ell) = (-1)^{\ell+1}\pi_\ell(\sigma_\ell\cdots\sigma_1) \quad \text{for all } \sigma\in\mathfrak{S}_\ell. \tag{2.5}$$

We are now ready to define commutators for vector fields in our regularity classes.

Definition 2.3 (Definitions of commutator)**.** *Given a family $\mathcal{H} = \{X_1,\dots X_m\}$ of vector fields of class $C^{s-1,1}_{\mathcal{H},\mathrm{loc}} \cap C^1_{\mathrm{Euc}}$, define, for $\psi \in C^1_{\mathcal{H}}$, $X_j^\sharp\psi(x) := \mathcal{L}_{X_j}\psi(x)$, the Lie derivative; let also $X_j\psi(x) := f_j(x)\cdot\nabla\psi(x)$ where $\psi\in C^1_{\mathrm{Euc}}$. Moreover, let*

$$\begin{aligned}
f_w &:= \sum_{\sigma\in\mathfrak{S}_\ell}\pi_\ell(\sigma)\big(X_{\sigma_1(w)}\cdots X_{\sigma_{\ell-1}(w)}f_{\sigma_\ell(w)}\big) \quad \text{for all } w \text{ with } |w|\le s,\\
X_w\psi &:= [X_{w_1},,\dots,[X_{w_{\ell-1}},X_{w_\ell}]]\psi := f_w\cdot\nabla\psi \quad \text{for all } \psi\in C^1_{\mathrm{Euc}} \quad |w|\le s,\\
X_w^\sharp\psi &:= \sum_{\sigma\in\mathfrak{S}_\ell}\pi_\ell(\sigma)X^\sharp_{\sigma_1(w)}\cdots X^\sharp_{\sigma_{\ell-1}(w)}X^\sharp_{\sigma_\ell(w)}\psi \quad \text{for all } \psi\in C^\ell_{\mathcal{H}} \quad |w|\le s-1.
\end{aligned}$$

Given words u and v, define the (possibly non-nested) commutators

$$\begin{aligned}
f_{[u]v} &:= X_u^\sharp f_v - X_v^\sharp f_u\\
&= \sum_{\alpha\in\mathfrak{S}_p,\beta\in\mathfrak{S}_q}\pi_p(\alpha)\pi_q(\beta)\big(X_{\alpha_1(u)}\cdots X_{\alpha_p(u)}X_{\beta_1(v)}\cdots X_{\beta_{q-1}(v)}f_{\beta_q(v)}\\
&\qquad - X_{\beta_1(v)}\cdots X_{\beta_q(v)}X_{\alpha_1(u)}\cdots X_{\alpha_{p-1}(u)}f_{\alpha_p(u)}\big),\\
X_{[u]v} &:= [X_u,X_v] := f_{[u]v}\cdot\nabla = (X_u^\sharp f_v - X_v^\sharp f_u)\cdot\nabla \qquad \text{if } |u|+|v|\le s,\\
[X_u,X_v]^\sharp &:= X^\sharp_{[u]v} := X_u^\sharp X_v^\sharp - X_v^\sharp X_u^\sharp \qquad \text{if } |u|+|v|\le s-1,
\end{aligned}$$

where $X_{[u]v}$ and $X^\sharp_{[u]v}$ act respectively on C^1_{Euc} and $C^{|u|+|v|}_{\mathcal{H}}$ functions. Finally, for any $j\in\{1,\dots,m\}$ and w with $1\le|w|\le s$, let

$$\mathrm{ad}_{X_j}X_w\psi := (X_j^\sharp f_w - f_w\cdot\nabla f_j)\cdot\nabla\psi = (X_j^\sharp f_w - X_w f_j)\cdot\nabla\psi \quad \text{for all } \psi\in C^1_{\mathrm{Euc}}. \tag{2.6}$$

Note that we will never need in this paper the commutators $X_w^\sharp\psi$ for $|w| = s$.

Remark 2.4. *Let $Z \in \pm\mathcal{H}$, where $\mathcal{H}$ is a family in $C^{s-1,1}_{\mathcal{H},\mathrm{loc}}\cap C^1_{\mathrm{Euc}}$. If $|w|\le s-1$, then there are no problems in defining $\mathrm{ad}_Z X_w$. More precisely, in Theorem 3.1 we will see that $\mathrm{ad}_Z X_w = [Z,X_w]$. If instead $|w| = s$, then the function $t\mapsto f_w(e^{tZ}x)$ is Lipschitz continuous. In particular it is differentiable for a.e. t. In other words, for any fixed $x\in\mathbb{R}^n$, the limit $\frac{d}{dt}f_w(e^{tZ}x) =: Z^\sharp f_w(e^{tZ}x)$ exists for a.e. t close to 0. Therefore the pointwise derivative $Z^\sharp f_w(y)$ exists for almost all $y\in\mathbb{R}^n$ and ultimately $\mathrm{ad}_Z X_w$ is defined almost everywhere. See the discussion in Theorem 2.6-(b) and Proposition 4.1.*

We will recognize that the first order operator X_w agrees with $X_w^\sharp$ against functions $\psi \in C^{s-1,1}_{\mathcal{H},\mathrm{loc}} \cap C^1_{\mathrm{Euc}}$ and for $|w| \leq s-1$. This is trivial if $|w| = 1$, because $X_k\psi := f_k \cdot \nabla\psi$ and $X_k^\sharp\psi := \mathcal{L}_{X_k}\psi$ are the same, if both X_k and ψ are C^1_{Euc}.

Remark 2.5. *Both our definitions of commutator, X_w and $X_w^\sharp$ are well posed from an algebraic point of view. Indeed, it is easy to check that $[X_u, X_v] = (X_u^\sharp f_v - X_v^\sharp f_u) \cdot \nabla = -[X_v, X_u]$. Moreover*

$$\begin{aligned}[X_w, [X_u, X_v]] &= (X_w^\sharp f_{[u]v} - [X_u, X_v]^\sharp f_w) \cdot \nabla \\ &= \{X_w^\sharp X_u^\sharp f_v - X_w^\sharp X_v^\sharp f_u - X_u^\sharp X_v^\sharp f_w + X_v^\sharp X_u^\sharp f_w\} \cdot \nabla,\end{aligned}$$

for any u, v, w with $|u| + |v| + |w| \leq s$. This immediatley implies the Jacobi identity

$$[X_u, [X_v, X_w]] + [X_v, [X_w.X_u]] + [X_w, [X_u, X_v]] = 0. \tag{2.7}$$

Antisymmetry and the Jacobi identity for the commutators $X_w^\sharp$ can be checked with the same argument.

Let $\Omega_0 \subset \mathbb{R}^n$ be a fixed open set, bounded in the Euclidean metric. Given a family $\mathcal{H}$ of vector fields of class $C^1_{\mathrm{Euc}} \cap C^{s-1,1}_{\mathcal{H},\mathrm{loc}}$, introduce the constant

$$\begin{aligned}L_0 := \sum_{j_1,\dots,j_s=1}^{m} \Big\{ \sup_{\Omega_0} \Big(|f_{j_1}| + |\nabla f_{j_1}| + \sum_{p\leq s} |X_{j_1}^\sharp \cdots X_{j_{p-1}}^\sharp f_{j_p}| \Big) \\ + \mathrm{Lip}_{\mathcal{H}}(X_{j_1}^\sharp \cdots X_{j_{s-1}}^\sharp f_{j_s}; \Omega_0) \Big\}.\end{aligned} \tag{2.8}$$

Fix also $\Omega \Subset \Omega_0$. We shall always choose points $x \in \Omega$ and we fix a constant $t_0 > 0$ small enough to ensure that

$$e^{\tau_1 Z_1} \cdots e^{\tau_N Z_N} x \in \Omega_0 \quad \text{if } x \in \Omega,\ Z_j \in \mathcal{H},\ |\tau_j| \leq t_0 \text{ and } N \leq N_0, \tag{2.9}$$

where N_0 is a suitable algebraic constant which depends on the data n, m and s associated with the family $\mathcal{H}$.

2.2. Non commutative formulas

In this section we discuss some preliminary tools on noncommutative calculus. Some of the objects we discuss here already appeared in [MM12b], for Hörmander vector fields, in a higher regularity setting. Observe that Theorem 2.6 has also a relevant role in [MM12a, Lemma 3.1 and Theorem 3.5].

Theorem 2.6. *Let $\mathcal{H}$ be a family of $C^1_{\mathrm{Euc}} \cap C^{s-1,1}_{\mathcal{H},\mathrm{loc}}$ smooth vector fields. Fix $Z \in \pm\mathcal{H}$ and X_w with $|w| \geq 1$. Then:*

(a) if $|w| \leq s-1$, then, for all $\psi \in C^1_{\mathrm{Euc}}$, $y \in \Omega$ and $|t| \leq t_0$ (see (2.9)) we have

$$\frac{d}{dt} X_w(\psi e^{-tZ})(e^{tZ}y) = \mathrm{ad}_Z\, X_w(\psi e^{-tZ})(e^{tZ}y); \tag{2.10}$$

(b) if $|w| = s$, then for any $\psi \in C^1_{\text{Euc}}$ and $y \in \Omega$ the function $\varphi(t) := X_w(\psi e^{-tZ})(e^{tZ}y)$ is Euclidean Lipschitz and satisfies

$$\frac{d}{dt}X_w(\psi e^{-tZ})(e^{tZ}x) = \mathrm{ad}_Z X_w(\psi e^{-tZ})(e^{tZ}x) \quad \textit{for a.e. } t \in (-t_0, t_0). \tag{2.11}$$

To comment on (2.10), assume that $Z = X_j$ for some $j \in \{1, \dots, m\}$. Note that in Theorem 3.1 we will show that $\mathrm{ad}_{X_j} X_w = [X_j, X_w] = X_{jw}$, if $|w| \leq s - 1$, $j \in \{1, \dots, m\}$. Looking instead at equation (2.11), the operator $\mathrm{ad}_Z X_w$ has been defined in (2.6). If we assume the Hörmander condition of step s, we shall see in Proposition 4.1 that we can write

$$\mathcal{L}_Z X_w(\psi e^{-tZ})(e^{tZ}x) := \frac{d}{dt}X_w(\psi e^{-tZ})(e^{tZ}x) = \sum_{1\leq|u|\leq s} b^u(t)X_u(\psi e^{-tZ})(e^{tZ}x), \tag{2.12}$$

where the functions b^u may depend on Z, w, x and are measurable and bounded.

Remark 2.7. *The proof of* (2.10) *is standard for smooth (say at least C^2) vector fields, see [KN96, Proposition 1.9], or, for a different argument, [Aub01, Proposition 3.5] and [MM12b, Lemma 3.1]. Note also that $Ue^{-tV}(e^{tV}x) = e^{-tV}_*(U_{e^{tV}x})$, by definition of tangent map. Thus, $(\mathcal{L}_V U)_x = \frac{d}{dt}Ue^{-tV}(e^{tV}x)\big|_{t=0}$, by definition of Lie derivative. Then, in Step 1 of the proof below, we are proving nothing but the probably known fact that $\mathcal{L}_V U = (V\xi - U\eta)\cdot\nabla$, if $V = \eta \cdot \nabla$ and $U = \xi \cdot \nabla \in C^1_{\text{Euc}}$.*

Proof of Theorem 2.6. We split the proof in three steps.

Step 1. We prove that for any $U = \xi \cdot \nabla \in C^1_{\text{Euc}}$ and $V = \eta \cdot \nabla \in C^1_{\text{Euc}}$, we have for all $x \in \mathbb{R}^n$

$$\frac{d}{dt}U(\psi e^{-tV})(e^{tV}x) = [V, U](\psi e^{-tV})(e^{tV}x) \quad \text{if } |t| \text{ is small enough.}$$

Here $[V, U] := (V\xi - U\eta) \cdot \nabla$ and $\psi \in C^1_{\text{Euc}}$.

Let $\psi \in C^1_{\text{Euc}}$. Take the usual smooth approximations $V^\sigma = \eta^\sigma \cdot \nabla$, $U^\sigma = \xi^\sigma \cdot \nabla$ and ψ^σ. Since V and U are C^1, elementary properties of Euclidean mollifiers show that $\eta^\sigma \to \eta$, $\xi^\sigma \to \xi$ and $V^\sigma\xi^\sigma - U^\sigma\eta^\sigma \to V\xi - U\eta$, uniformly on compact sets, as $\sigma \to 0$. Therefore, we have

$$\begin{aligned}\frac{d}{dt}U^\sigma(\psi^\sigma e^{-tV^\sigma})(e^{tV^\sigma}x) &= V^\sigma U^\sigma(\psi^\sigma e^{-tV^\sigma})(e^{tV^\sigma}x) - U^\sigma V^\sigma(\psi^\sigma e^{-tV^\sigma})(e^{tV^\sigma}x) \\ &= (V^\sigma\xi^\sigma - U^\sigma\eta^\sigma)(e^{tV^\sigma}x) \cdot \nabla(\psi^\sigma e^{-tV^\sigma})(e^{tV^\sigma}x) =: R(\sigma).\end{aligned}$$

First equality is provided in textbooks, see Remark 2.7. In our notation, the intermediate term here is $[V^\sigma, U^\sigma]^\sharp(\psi e^{-tV^\sigma})(e^{tV^\sigma}x)$. Both its addends may have a not clear behaviour, as $\sigma \to 0$. After the cancellation, second derivatives against $\psi^\sigma e^{-tV^\sigma}$ disappear and we may let $\sigma \to 0$ in the second line. By standard ODE theory, see for example [Har02, Chapter 5], $\nabla e^{tV^\sigma} \to \nabla e^{tV}$, uniformly on Ω and $|t| \leq t_0$, as $\sigma \to 0$. Then $\lim_{\sigma\to 0} R(\sigma) = [V, U](\psi e^{-tV})(e^{tV}x)$, uniformly on $t \in [-t_0, t_0]$ and $x \in \Omega$. Moreover,

$$\lim_{\sigma\to 0} U^\sigma(\psi^\sigma e^{-tV^\sigma})(e^{tV^\sigma}x) = U(\psi e^{-tV})(e^{tV}x) \quad \text{for all } t \in [-t_0, t_0] \quad x \in \Omega.$$

Therefore, Step 1 is accomplished.

Step 2. We prove statement (a). By uniqueness of the flow of Z, we may work with $t = 0$.

$$\frac{d}{dt}X_w(\psi e^{-tZ})(e^{tZ}x)\Big|_{t=0} = \lim_{t\to 0}\frac{1}{t}\Big\{[f_w(e^{tZ}x) - f_w(x)]\cdot\nabla(\psi e^{-tZ})(e^{tZ}x) + f_w(x)\cdot[\nabla(\psi e^{-tZ})(e^{tZ}x) - \nabla\psi(x)]\Big\}.$$

But $\lim_{t\to 0}\frac{1}{t}[f_w(e^{tZ}x) - f_w(x)] =: Z^\sharp f_w(x)$ exists, because $f_w \in C^1_{\mathcal{H}}$. Moreover, since $\partial_1, \dots, \partial_n, Z \in C^1_{\text{Euc}}$, Step 1 gives for all $\alpha \in \{1, \dots, n\}$,

$$\lim_{t\to 0}\frac{1}{t}[\partial_\alpha(\psi e^{-tZ})(e^{tZ}x) - \partial_\alpha\psi(x)] = [Z, \partial_\alpha]\psi(x) = -\partial_\alpha f(x)\cdot\nabla\psi(x), \tag{2.13}$$

where $Z = f\cdot\nabla$. This concludes the proof of Step 2.

Step 3. Proof of (b). We will show that $t \mapsto X_w(\psi e^{-tZ})(e^{tZ}x) =: \varphi(t)$ is Lipschitz continuous on $[-t_0, t_0]$ for all $x \in \Omega$.

$$\varphi(\tau) - \varphi(t) = f_w(e^{\tau Z}x)\cdot\big[\nabla(\psi e^{-\tau Z})(e^{\tau Z}x) - \nabla(\psi e^{-tZ})(e^{tZ}x)\big] + [f_w(e^{\tau Z}x) - f_w(e^{tZ}x)]\cdot\nabla(\psi e^{-tZ})(e^{tZ}x).$$

But, since $f_w \in \text{Lip}_{\mathcal{H}}$, we have $|f_w(e^{\tau Z}x) - f_w(e^{tZ}x)| \le C|\tau - t|$. Moreover, if $t \in (-t_0, t_0)$ is a differentiability point for $t \mapsto f_w(e^{tZ}x)$, we have $\lim_{\tau\to t}(f_w(e^{\tau Z}x) - f_w(e^{tZ}x))/(\tau - t) = Z^\sharp f_w(e^{tZ}x)$, by definition of derivative along Z. Finally, for any $\alpha \in \{1, \dots, n\}$, (2.13) shows that the function $t \mapsto \partial_\alpha(\psi e^{-tZ})(e^{tZ}x) \in C^1_{\text{Euc}}$ and that

$$\frac{d}{dt}\partial_\alpha(\psi e^{-tZ})(e^{tZ}x) = -\partial_\alpha f(e^{tZ}x)\cdot\nabla(\psi e^{-tZ})(e^{tZ}x).$$

Then the proof of (b) is easily concluded. □

2.3. Integral remainders

Here we introduce a class of integral remainders $O_p(t^\lambda, \psi, y)$. There is a reason why we use a different notation from the seemingly similar remainders $R_p(t^\lambda, \psi, y)$ appearing in the Taylor formula below, see (2.21). Namely, the remainders $O_p(\dots)$ have a much more "balanced" structure. Under suitable involutivity conditions and using such balanced structure, in [MM12a] we will be able to show that they can be given a *pointwise* form (a consequence of this fact is the pointwise form of the remainders in expansion (4.9)).

Let $\mathcal{H}$ be a family in the regularity class $C^{s-1,1}_{\mathcal{H}} \cap C^1_{\text{Euc}}$. Let $\lambda \in \mathbb{N}$, $p \in \{2, \dots, s+1\}$. We denote, for $y \in \Omega$ and $t \in [0, t_0]$, $\psi \in C^1_{\text{Euc}}(\Omega_0)$

$$O_p(t^\lambda, \psi, y) = \sum_{i=1}^{N}\int_0^t \omega_i(t,\tau)\frac{d}{d\tau}X_{w_i}(\psi\varphi_i^{-1}e^{-\tau Z_i})(e^{\tau Z_i}\varphi_i y)d\tau, \tag{2.14}$$

where N is a suitable integer and ψ is the identity map or $\psi = \exp(tZ_1)\cdots\exp(tZ_\mu)$, for some integer μ and suitable vector fields $Z_j \in \pm\mathcal{H}$. The "balanced structure" we mentioned above, follows from identity $(\psi\varphi_i^{-1}e^{-\tau Z_i})(e^{\tau Z_i}\varphi_i y) = \psi(y)$.

To describe the generic term of the sum above, we drop the dependence on i:

$$(*) := \int_0^t \omega(t,\tau)\frac{d}{d\tau}X_w(\psi\varphi^{-1}e^{-\tau X})(e^{\tau X}\varphi y)d\tau. \tag{2.15}$$

Here X_w is a commutator of length $|w| = p-1$ and $X \in \pm\mathcal{H}$. Moreover, for any $t < t_0$, the function $\omega(t,\tau)$ is a polynomial, homogeneous of degree $\lambda - 1$ in all variables (t,τ), so that

$$\int_0^t \omega(t,\tau)d\tau = bt^\lambda \quad \text{for any} \quad t > 0 \tag{2.16}$$

for a suitable constant $b \in \mathbb{R}$. The map φ is the identity map or otherwise it has the form $\varphi = \exp(tZ_1)\cdots\exp(tZ_\nu)$ for some $\nu \in \mathbb{N}$, where $Z_j \in \pm\mathcal{H}$. Observe that, if $p \le s$, i.e. $|w| \le s-1$, Lemma 2.6 (a) gives

$$(*) = \int_0^t \omega(t,\tau)\,\mathrm{ad}_X\, X_w(\psi\varphi^{-1}e^{-\tau X})(e^{\tau X}\varphi y)d\tau.$$

Therefore the remainder has the same form of the analogous term in [MM12b, Eq. (3.5)], provided that we are able to show that $\mathrm{ad}_X\, X_w = [X, X_w]$ (this will be achieved in Theorem 3.1). If instead $p = s+1$, i.e. $|w| = s$, we need to use part Theorem 2.6-(b) to get some information on the remainder. See also Proposition 4.1 and see the paper [MM12a] for a detailed discussion of remainders of higher order $O_{s+1}(\cdots)$.

A remainder of the form (2.14) satisfies for every $\alpha, \lambda \in \mathbb{N}$ and $p \le s+1$ estimate

$$t^\alpha O_p(t^\lambda, \psi, y) = O_p(t^{\alpha+\lambda}, \psi, y) \quad \text{for all} \quad y \in \Omega \quad t \in [0, t_0]. \tag{2.17}$$

Let us also recall estimate

$$|O_p(t^\lambda, \psi, y)| \le Ct^\lambda, \tag{2.18}$$

which holds for $p \le s+1$, $\lambda \in \mathbb{N}$. To see (2.18), just observe that, at any t and for $j \in \{1,\dots,m\}$ and $|w| \le s$, we have

$$\Big|\frac{d}{d\tau} X_w(\psi\varphi^{-1}e^{-\tau X_j})(e^{\tau X_j}\varphi x)\Big| = \big|\mathrm{ad}_{X_j}\, X_w(\psi\varphi^{-1}e^{-\tau X_j})(e^{\tau X_j}\varphi x)\big| \le C,$$

at any τ such that $X_j^\sharp f_w(e^{\tau X_j}\varphi x)$ exists. Here we use the trivial estimate $|\mathrm{ad}_{X_j}\, X_w| \le |X_j^\sharp f_w| + |X_w f_j| \le C$, because $f_w \in \mathrm{Lip}_\mathcal{H}$ and $f_j \in C^1_{\mathrm{Euc}}$.

Note finally that, if $j \in \{1,\dots,m\}$, $p \le s+1$ and $Z \in \pm\mathcal{H}$, we have

$$O_p(t^\lambda, \psi e^{tZ}, y) = O_p(t^\lambda, \psi, e^{tZ}y).$$

Proposition 2.8. *Assume that $p \le s$ and assume that $\mathrm{ad}_{X_j}\, X_w = X_{jw}$ for all word w with length $|w| \le p-1$ and $j \in \{1,\dots,m\}$. Then there are constants c_w, $|w| = p$, such that*

$$O_p(t^\lambda, \psi, y) = \sum_{|w|=p} c_w t^\lambda X_w\psi(y) + O_{p+1}(t^{\lambda+1}, \psi, y). \tag{2.19}$$

The proof of Proposition 2.8 has been given in [MM12b, Proposition 3.3] for smooth vector fields. The argument is the same in our case. One just needs to check that all the computations we made there work perfectely in our regularity setting. We omit the details.

Remark 2.9.
- *The statement of Proposition 2.8, and in particular the assumption $\mathrm{ad}_{X_j}\, X_w = X_{jw}$, is designed in order to be a part of the induction machinery we shall implement to prove Theorem 3.1 in the following section.*

- *The generalization of* (2.19) *to the case* $p = s + 1$ *is discussed under the Hörmander condition in Section 4, see* (4.2). *In the companion paper [MM12a] we deal with a more general situation.*

Remark 2.10. *Let for a while* $\mathcal{H} = \{X_1, \dots, X_m\}$ *be a family of smooth vector fields. Iterating Theorem 2.6, we have for* $x \in \Omega$ *and* $|t|$ *sufficiently small*

$$\begin{aligned} X_w(\psi e^{-tZ_\mu} \cdots e^{-tZ_1})x = \sum_{|\alpha|=0}^{\ell} \operatorname{ad}_{Z_\mu}^{\alpha_\mu} \cdots \operatorname{ad}_{Z_1}^{\alpha_1} X_w \psi(e^{-tZ_\mu} \cdots e^{-tZ_1} x) \frac{t^{|\alpha|}}{\alpha!} \\ + O_{\ell+|w|}(t^{\ell+1}, \psi, x). \end{aligned} \tag{2.20}$$

Formula (2.20) *will be referred to later.*

Taylor formula with integral remainder. Here we show that functions of class $C^{s-1,1}_{\mathcal{H},\mathrm{loc}}$ enjoy an elementary Taylor expansion with integral remainder. Let $p, \lambda \in \mathbb{N}$. Denote by $R_p(t^\lambda, \psi, x)$ a sum of a finite number of terms of the form

$$\int_0^t \omega(t,\tau) \frac{d}{d\tau} (X_{j_1}^\sharp)^{k_1} \cdots (X_{j_\mu}^\sharp)^{k_\mu} \psi(e^{\tau X_i} \varphi x) d\tau, \tag{2.21}$$

where the polynomial $\omega(t,\tau)$ is homogeneous of degree $\lambda - 1$ in all variables (t,τ). This ensures that $\int_0^t \omega(t,\tau) d\tau = Ct^\lambda$, for any $t > 0$. Moreover, $i, j_1, \dots, j_\mu \in \{1, \dots, m\}$, $k_1 + \cdots + k_\mu = p-1$. The map φ is the identity map or it has the form $\varphi = \exp(tZ_1) \cdots \exp(tZ_\nu)$ for some $\nu \in \mathbb{N}$, where $Z_j \in \pm\mathcal{H}$. If $\Omega \subset \mathbb{R}^n$ is bounded, then we have, for all $x \in \Omega$, $|t| \le t_0$,

$$|R_p(t^\lambda, \psi, x)| \le C \operatorname{Lip}_{\mathcal{H}} \big((X_{j_1}^\sharp)^{k_1} \cdots (X_{j_\mu}^\sharp)^{k_\mu} \psi, B_d(x, C|t|) \big) t^\lambda,$$

where t_0 is positive, small enough, see (2.9).

Denote $\Delta^{j_1 \cdots j_q} x := \Delta^{j_1} \Delta^{j_2} \cdots \Delta^{j_q} x := e^{tX_{j_1}} \cdots e^{tX_{j_q}} x$, where $j_1, \dots, j_q \in \{1, \dots, m\}$.

Lemma 2.11. *Let* $\psi \in C^{\ell-1,1}_{\mathcal{H},\mathrm{loc}}$, *for some* $\ell \le s$. *Then for any* $q \ge 1$ *and* $j_1, \dots, j_q \in \{1, \dots, m\}$, *we have in standard multi-index notation*

$$\psi(\Delta^{j_1 \cdots j_q} x) = \sum_{\substack{k_1, \dots, k_q \ge 0 \\ k_1 + \cdots + k_q \le \ell - 1}} (X_{j_q}^\sharp)^{k_q} \cdots (X_{j_1}^\sharp)^{k_1} \psi(x) \frac{t^{|k|}}{k!} + R_\ell(t^\ell, \psi, x). \tag{2.22}$$

Proof. We prove formula (2.22) by induction on $q \ge 1$. Fix $j \in \{1, \dots, m\}$. Let $\psi \in C^{\ell-1,1}_{\mathcal{H},\mathrm{loc}}$, Then $\left(\frac{d}{dt}\right)^k \psi(e^{tX_j} x) = (X_j^\sharp)^k \psi(e^{tX_j} x)$, for $k = 0, 1, \dots, \ell - 1$. Moreover, the function $t \mapsto (X_j^\sharp)^{\ell-1} \psi(e^{tX_j} x)$ is Euclidean Lipschitz and, for a.e. $t \in (-t_0, t_0)$, its derivative can be estimated by $\operatorname{Lip}_{\mathcal{H}}((X_j^\sharp)^{\ell-1} \psi; B_d(x, t_0))$. Therefore, the Taylor formula gives

$$\begin{aligned} \psi(e^{tX_j} x) &= \sum_{k=0}^{\ell-1} (X_j^\sharp)^k \psi(x) \frac{t^k}{k!} + \int_0^t \frac{(t-\tau)^{\ell-1}}{(\ell-1)!} \frac{d}{d\tau} (X_j^\sharp)^{\ell-1} \psi(e^{\tau X_j} x) d\tau \\ &= \sum_{k=0}^{\ell-1} (X_j^\sharp)^k \psi(x) \frac{t^k}{k!} + R_\ell(t^\ell, \psi, x). \end{aligned}$$

Next we give the induction step. Let $q \geq 1$. Then,

$$\begin{aligned}\psi(\Delta^{j_0}\Delta^{j_1\cdots j_q}x) &= \sum_{k_0=0}^{\ell-1}(X_{j_0}^\sharp)^{k_0}\psi(\Delta^{j_1\cdots j_q}x)\frac{t^{k_0}}{k_0!} + R_\ell(t^\ell,\psi,\Delta^{j_1\cdots j_q}x)\\ &= \sum_{k_0=0}^{\ell-1}\frac{t^{k_0}}{k_0!}\Big\{\sum_{\substack{k_1,\dots,k_q\geq 0\\ k_1+\cdots+k_q\leq \ell-1-k_0}}(X_{j_q}^\sharp)^{k_q}\cdots(X_{j_1}^\sharp)^{k_1}(X_{j_0}^\sharp)^{k_0}\psi(x)\frac{t^{k_1+\cdots+k_q}}{k_1!\cdots k_q!}\\ &\qquad + R_{\ell-k_0}(t^{\ell-k_0},(X_{j_0}^\sharp)^{k_0}\psi,x)\Big\} + R_\ell(t^\ell,\psi,x).\end{aligned}$$

The proof is concluded by property $t^{k_0}R_{\ell-k_0}(t^{\ell-k_0},(X_{j_0}^\sharp)^{k_0}\psi,x) = R_\ell(t^\ell,\psi,x)$. □

3. Commutator identities

In this section we show that, if the vector fields of the family $\mathcal{H}$ belong to $C^{s-1,1}_{\mathcal{H},\mathrm{loc}} \cap C^1_{\mathrm{Euc}}$, then the various notions of commutators introduced in Definition 2.3 agree. This requires a quite elaborate algebraic work which will be performed in the first subsection. Later on, we will show that the machinary we construct, in particular the generalized Jacobi identities in Proposition 3.3 can be useful to detect nested commutators identities.

Let $\mathcal{H} = \{X_1,\dots,X_m\}$ be family of vector fields of class $C^1_{\mathrm{Euc}} \cap C^{s-1,1}_{\mathcal{H},\mathrm{loc}}$. We use the notation $\mathcal{W}_\ell$ to indicate the set of words $w = w_1\cdots w_\ell$ of length ℓ.

The main result of this section is the following theorem, which has a key role in the proof of [MM12a, Theorems 3.5 and 3.8] and ultimately of Theorem 4.3 here; see the discussion at the beginning of [MM12a, Subsection 3.2].

Theorem 3.1. *Let $\mathcal{H}$ be a family of vector fields of class $C^1_{\mathrm{Euc}} \cap C^{s-1,1}_{\mathcal{H},\mathrm{loc}}$. Then, if $1 \leq \ell \leq s-1$, the following statements are equivalent and true.*

(i) For any $w \in \mathcal{W}_\ell$ and for all $\psi \in C^1_{\mathrm{Euc}} \cap C^{\ell,1}_{\mathcal{H},\mathrm{loc}}$ we have

$$X_w\psi = X_w^\sharp\psi. \tag{3.1}$$

(ii) For any $Z = \psi\cdot\nabla \in C^1_{\mathrm{Euc}} \cap C^{\ell,1}_{\mathcal{H},\mathrm{loc}}$ and for all $w \in \mathcal{W}_\ell$, we have

$$\mathrm{ad}_Z X_w\varphi = [Z, X_w]\varphi \qquad \text{for all } \varphi \in C^1_{\mathrm{Euc}}. \tag{3.2}$$

Remark 3.2. *In view of Theorem 3.1, formula* (2.10) *in Theorem 2.6 becomes,*

$$\frac{d}{dt}X_w(\psi e^{-tZ})(e^{tZ}y) = [Z, X_w](\psi e^{-tZ})(e^{tZ}y) \quad \text{if } |w| \leq s-1 \quad t \in (-t_0,t_0). \tag{3.3}$$

The case $|w| = s$ will be discussed in Section 4, see e.g. Proposition 4.1.

To prove Theorem 3.1, we need the following proposition which may have some independent interest.

Proposition 3.3 (Generalized Jacobi identities)**.** *Let $\mathcal{H}$ be a family in the regularity class $C^{s-1,1}_{\mathcal{H},\mathrm{loc}} \cap C^1_{\mathrm{Euc}}$. For any $v \in \mathcal{W}_p$, $w \in \mathcal{W}_q$, $p, q \geq 1$, $p + q \leq s$, we have*

$$X_{[v]w} = \sum_{\sigma \in \mathfrak{S}_p} \pi_p(\sigma) X_{\sigma_1(v) \dots \sigma_p(v) w}. \tag{3.4}$$

If $|w| = 0$, then (3.4) *fails, but for any $v = v_1 \cdots v_\ell \in \mathcal{W}_\ell$, $\ell \leq s$, we have*

$$X_v = \frac{1}{\ell} \sum_{\sigma \in \mathfrak{S}_\ell} \pi_\ell(\sigma) X_{\sigma_1(v) \dots \sigma_\ell(v)}. \tag{3.5}$$

Before proving the proposition, to explain the reason of our terminology, we give a couple of examples to show that our identities, suitably specialized, give back some familiar identities. See also Subsection 3.1.

Example 3.4. *Let X_1, X_2 and X_3 be sufficiently smooth vector fields. Then*

$$\begin{aligned}
[X_1, [X_2, X_3]] =: X_{123} &= \frac{1}{3} \sum_{\sigma \in \mathfrak{S}_3} \pi_3(\sigma) X_{\sigma_1(123)\sigma_2(123)\sigma_3(123)} \\
&= \frac{1}{3}\big\{X_{123} - X_{132} - X_{231} + X_{321}\big\} \\
&= \frac{1}{3}\big\{[X_1, [X_2, X_3]] - [X_1, [X_3, X_2]] - [X_2, [X_3, X_1]] + [X_3, [X_2, X_1]]\big\} \\
&= \frac{2}{3}[X_1, [X_2, X_3]] - \frac{1}{3}[X_2, [X_3, X_1]] - \frac{1}{3}[X_3, [X_1, X_2]].
\end{aligned}$$

Comparing the first and the list line one can recognize the familiar Jacobi identity.

Example 3.5. *Here, looking at the fourth order identity* (3.5) *with $\ell = 4$ and taking $w = 1212$, we check the nested commutators identity*

$$X_{1212} = X_{2112} = -X_{1221} \tag{3.6}$$

discussed in [Ote91, eq. (4.3)]. To get (3.6)*, start from the 4-th order formula*

$$X_{1234} = \frac{1}{4}\{X_{1234} - X_{1243} - X_{1342} + X_{1432} - X_{2341} + X_{2431} + X_{3421} - X_{4321}\}.$$

Letting 1 instead of 3 and 2 instead of 4, we get

$$4X_{1212} = X_{1212} - X_{1221} - X_{1122} + X_{1212} - X_{2121} + X_{2211} + X_{1221} - X_{2121},$$

which is equivalent to $2X_{1212} = -2X_{2121}$, and gives immediately (3.6).

Proof of Proposition 3.3. To prove (3.4), we argue by induction on $|v|$. The property is trivial if $|v| = 1$ and $1 \leq |w| \leq s - 1$. Assume that for a given $p \in \{1, \dots, s-2\}$, formula (3.4) holds for all v, w with $|v| = p$ and $1 \leq |w| \leq s - p$ and we will prove that it holds for any v, w with $|v| = p + 1$ and $1 \leq |w| \leq s - p - 1$.

Write $\widetilde{v} = v_0 v \in \mathcal{W}_{p+1}$ and $\widetilde{\sigma}(\widetilde{v}) = \widetilde{\sigma}_0(v_0 v)\dots\widetilde{\sigma}_p(v_0 v)$. Then the defining property (2.3) of the coefficients $\pi(\sigma)$ gives

$$\begin{aligned}
&\sum_{\widetilde{\sigma}\in\mathfrak{S}_{p+1}} \pi_{p+1}(\widetilde{\sigma}) X_{\widetilde{\sigma}_0(\widetilde{v})\widetilde{\sigma}_1(\widetilde{v})\cdots\widetilde{\sigma}_p(\widetilde{v})w} \\
&= \sum_{\sigma\in\mathfrak{S}_p} \pi_p(\sigma)\big(X_{v_0\sigma_1(v)\cdots\sigma_p(v)w} - X_{\sigma_1(v)\cdots\sigma_p(v)v_0 w}\big) \\
&= \Big[X_{v_0}, \sum_{\sigma\in\mathfrak{S}_p} \pi_p(\sigma) X_{\sigma_1(v)\cdots\sigma_p(v)w}\Big] - \sum_{\sigma\in\mathfrak{S}_p} \pi_p(\sigma) X_{\sigma_1(v)\cdots\sigma_p(v)v_0 w} \quad \text{(inductive assumption)} \\
&= [X_{v_0}, X_{[v]w}] - X_{[v]v_0 w} = X_{[v_0 v]w},
\end{aligned}$$

by the Jacobi identity (2.7) and the antisymmetry. Thus (3.4) is proved.

To prove (3.5), we work by induction. The statement for $\ell = 2$ is obvious. Assume that (3.5) holds for some $\ell \in \{2, \dots, s-1\}$. We need to show that

$$X_{v_0 v} = \frac{1}{\ell+1} \sum_{\sigma\in\mathfrak{S}_{\ell+1}} \pi_{\ell+1}(\widetilde{\sigma}) X_{\widetilde{\sigma}_0\widetilde{\sigma}_1\dots\widetilde{\sigma}_\ell} \quad \text{for all } v_0 v = v_0 v_1 \cdots v_\ell \in \mathcal{W}_{\ell+1},$$

where for all j we denoted $\widetilde{\sigma}_j = \widetilde{\sigma}_j(v_0 v)$. But the definition of $\pi_{\ell+1}$, the induction assumption and (3.4) show that

$$\begin{aligned}
\sum_{\widetilde{\sigma}\in\mathfrak{S}_{\ell+1}} \pi(\widetilde{\sigma}) X_{\widetilde{\sigma}_0\widetilde{\sigma}_1\cdots\widetilde{\sigma}_\ell} &= \sum_{\sigma\in\mathfrak{S}_\ell} \pi_\ell(\sigma)(X_{v_0\sigma_1\cdots\sigma_\ell} - X_{\sigma_1\cdots\sigma_\ell v_0}) \\
&= [X_{v_0}, \sum_{\sigma\in\mathfrak{S}_\ell} \pi_\ell(\sigma) X_{\sigma_1\dots\sigma_\ell}] - \sum_{\sigma\in\mathfrak{S}_\ell} \pi_\ell(\sigma) X_{\sigma_1\cdots\sigma_\ell v_0} \\
&= \ell X_{v_0 v} - X_{[v]v_0} = (\ell+1) X_{v_0 v},
\end{aligned}$$

as desired. ☐

Recall the notation $\Delta^{k_1\cdots k_\ell}x := e^{tX_{k_1}}\cdots e^{tX_{k_\ell}}x$, where $\ell\in\mathbb{N}$ and $k_j\in\{1,\dots,m\}$.

Lemma 3.6. *For any $\ell \in \{2, \dots, s-1\}$, for each $w \in \mathcal{W}_\ell$ and for each $\psi \in C^{\ell,1}_{\mathcal{H},\mathrm{loc}}$, we have*

$$X^\sharp_w \psi(x) = \lim_{t\to 0} \frac{1}{t^\ell} \sum_{\sigma\in\mathfrak{S}_\ell} \pi_\ell(\sigma) \psi(\Delta^{\sigma_\ell(w)\cdots\sigma_1(w)}x) \quad \textit{for all } x \in \mathbb{R}^n.$$

Proof. To prove the statement, we shall show the Taylor expansion

$$\sum_{\sigma\in\mathfrak{S}_\ell} \pi_\ell(\sigma)\psi(\Delta^{\sigma_\ell\cdots\sigma_1}x) = t^\ell X^\sharp_w\psi(x) + R_{\ell+1}(t^{\ell+1},\psi,x) \quad \text{for all } \psi \in C^{\ell,1}_{\mathcal{H},\mathrm{loc}}. \tag{3.7}$$

We will work by induction. The statement for $\ell = 2$ follows immediately from the Taylor formula (2.22). Indeed

$$\begin{aligned}
\psi(\Delta^{kj}x) &= \psi(x) + t(X^\sharp_j\psi + X^\sharp_k\psi)(x) \\
&\quad + \frac{t^2}{2}\big((X^\sharp_k)^2\psi + (X^\sharp_j)^2\psi + 2X^\sharp_j X^\sharp_k\psi\big)(x) + R_3(t^3,\psi,x),
\end{aligned}$$

where $j,k\in\{1,\dots,m\}$. Thus $\psi(\Delta^k\Delta^j x)-\psi(\Delta^j\Delta^k x)=t^2X^\sharp_{jk}\psi(x)+R_3(t^3,\psi,x)$.

Let us assume that (3.7) holds for some $\ell\in\{2,\dots,s-2\}$. Looking at the Taylor expansion (2.22), this means that

$$\sum_{\sigma\in\mathfrak{S}_\ell}\sum_{|\alpha|=0}^{\ell}\pi_\ell(\sigma)\frac{t^{|\alpha|}}{\alpha!}(X^\sharp_{\sigma_1})^{\alpha_1}\cdots(X^\sharp_{\sigma_\ell})^{\alpha_\ell}\psi(x)=t^\ell X^\sharp_w\psi(x)\quad\text{for all } t,x \text{ and } \psi\in C^{\ell,1}_{\mathcal{H},\mathrm{loc}}.$$

In particular, if $k\le\ell-1$, we have

$$\sum_{|\alpha|=0}^{k}\frac{t^{|\alpha|}}{\alpha!}\sum_{\sigma\in\mathfrak{S}_\ell}\pi_\ell(\sigma)(X^\sharp_{\sigma_1})^{\alpha_1}\cdots(X^\sharp_{\sigma_\ell})^{\alpha_\ell}\psi(x)=0\quad\text{for all } t,x. \tag{3.8}$$

In order to prove the induction step, let $\psi\in C^{\ell+1,1}_{\mathcal{H},\mathrm{loc}}$. Then, omitting all the $\sharp$ symbols

$$\begin{aligned}
&\sum_{\widetilde\sigma\in\mathfrak{S}_{\ell+1}}\pi_{\ell+1}(\widetilde\sigma)\psi(\Delta^{\widetilde\sigma_\ell\cdots\widetilde\sigma_1\widetilde\sigma_0}x)\\
&=\sum_{\sigma\in\mathfrak{S}_\ell}\pi_\ell(\sigma)\Big(\psi(\Delta^{\sigma_\ell\cdots\sigma_1 w_0}x)-\psi(\Delta^{w_0\sigma_\ell\cdots\sigma_1}x)\Big)\\
&=\sum_{|\alpha|+\beta=0}^{\ell+1}\frac{t^{|\alpha|+\beta}}{\alpha!\beta!}\sum_{\sigma\in\mathfrak{S}_\ell}\pi_\ell(\sigma)\big(X^\beta_{w_0}X^{\alpha_1}_{\sigma_1}\cdots X^{\alpha_\ell}_{\sigma_\ell}\psi(x)-X^{\alpha_1}_{\sigma_1}\cdots X^{\alpha_\ell}_{\sigma_\ell}X^\beta_{w_0}\psi(x)\big)\\
&\quad+R_{\ell+2}(t^{\ell+2},\psi,x)\\
&=\sum_{|\alpha|=0}^{\ell+1}\frac{t^{|\alpha|}}{\alpha!}\sum_{\sigma\in\mathfrak{S}_\ell}\pi_\ell(\sigma)\big(X^{\alpha_1}_{\sigma_1}\cdots X^{\alpha_\ell}_{\sigma_\ell}\psi(x)-X^{\alpha_1}_{\sigma_1}\cdots X^{\alpha_\ell}_{\sigma_\ell}\psi(x)\big)\\
&\quad+t\sum_{|\alpha|=0}^{\ell}\frac{t^{|\alpha|}}{\alpha!}\sum_{\sigma\in\mathfrak{S}_\ell}\big(X_{w_0}X^{\alpha_1}_{\sigma_1}\cdots X^{\alpha_\ell}_{\sigma_\ell}\psi(x)-X^{\alpha_1}_{\sigma_1}\cdots X^{\alpha_\ell}_{\sigma_\ell}X_{w_0}\psi(x)\big)\\
&\quad+\sum_{\beta=2}^{\ell+1}\frac{t^\beta}{\beta!}\sum_{|\alpha|=0}^{\ell+1-\beta}\frac{t^{|\alpha|}}{\alpha!}\sum_{\sigma\in\mathfrak{S}_\ell}\pi_\ell(\sigma)\big(X^\beta_{w_0}X^{\alpha_1}_{\sigma_1}\cdots X^{\alpha_\ell}_{\sigma_\ell}\psi(x)-X^{\alpha_1}_{\sigma_1}\cdots X^{\alpha_\ell}_{\sigma_\ell}X^\beta_{w_0}\psi(x)\big)\\
&\quad+R_{\ell+2}(t^{\ell+2},\psi,x).
\end{aligned}$$

Now note that the first line (case $\beta=0$) vanishes trivially. The third line, where $\beta\ge2$, vanishes by virtue of (3.8) (note that $X^\beta_{w_0}\psi\in C^{\ell+1-\beta,1}_{\mathcal{H},\mathrm{loc}}$). It remains the term with $\beta=1$ which gives

$$\begin{aligned}
\sum_{\widetilde\sigma\in\mathfrak{S}_{\ell+1}}\pi_{\ell+1}(\widetilde\sigma)\psi(\Delta^{\widetilde\sigma_\ell\cdots\widetilde\sigma_1\widetilde\sigma_0}x)&=t^{\ell+1}(X^\sharp_{w_0}X^\sharp_w\psi(x)-X^\sharp_wX^\sharp_{w_0}\psi(x))+R_{\ell+2}(t^{\ell+2},\psi,x)\\
&=t^{\ell+1}X^\sharp_{w_0w}\psi(x)+R_{\ell+2}(t^{\ell+2},\psi,x),
\end{aligned}$$

by definition of $X^\sharp_{w_0w}$. □

Proof of Theorem 3.1. We first show that (i) and (ii) are equivalent for $\ell=2,\dots,s-1$. The statement is obvious if $\ell=1$. Let now $\ell\in\{2,\dots,s-1\}$ and take $Z=\psi\cdot\nabla\in$

$C^{\ell,1}_{\mathcal{H}} \cap C^1_{\mathrm{Euc}}$. Fix also $w \in \mathcal{W}_\ell$. Comparing the definitions $\mathrm{ad}_Z X_w := (Z^\sharp f_w - X_w\psi)\cdot\nabla$ and $[Z, X_w] := (Z^\sharp f_w - X^\sharp_w \psi)\cdot\nabla$, we immediately recognize that (i) and (ii) are equivalent.

Next we prove that (i) holds for all $\ell \in \{2, \dots, s-1\}$. In view of Lemma 3.6, it suffices to prove that for all $w = w_1 \cdots w_\ell \in \mathcal{W}_\ell$, we have

$$\lim_{t\to 0} \sum_{\sigma\in\mathfrak{S}_\ell} \frac{1}{t^\ell}\pi_\ell(\sigma)\psi(\Delta^{\sigma_\ell(w)\cdots\sigma_1(w)}x) = \sum_{\sigma\in\mathfrak{S}_\ell} \pi_\ell(\sigma) X_{\sigma_1(w)}\cdots X_{\sigma_{\ell-1}(w)} f_{\sigma_\ell(w)}(x)\cdot\nabla\psi(x), \tag{3.9}$$

for any $\psi \in C^{\ell,1}_{\mathcal{H},\mathrm{loc}} \cap C^1_{\mathrm{Euc}}$ and for all $\ell = 2, 3, \dots, s-1$.

We first prove the statement for $\ell = 2$. Fix $X_{w_1}, X_{w_2} \in \{X_1, \dots, X_m\}$ and $\psi \in C^{2,1}_{\mathcal{H},\mathrm{loc}} \cap C^1_{\mathrm{Euc}}$. We need to show that

$$\lim_{t\to 0}\frac{1}{t^2}\sum_{\sigma\in\mathfrak{S}_2} \pi_2(\sigma)\psi(\Delta^{\sigma_2(w)}\Delta^{\sigma_1(w)}x) = X_{w_1w_2}\psi(x) \quad \text{for all } x\in\mathbb{R}^n, \tag{3.10}$$

where $X_{w_1w_2} := (X_{w_1}f_{w_2} - X_{w_2}f_{w_1})\cdot\nabla$. Observe that since trivially $X_k\psi = X^\sharp_k\psi$ for all $\psi \in C^1_{\mathrm{Euc}}$ and $k = 1, \dots, m$, we already have

$$\mathrm{ad}_{X_k} X_i = X_{ki} \quad \text{for all } k, i \in \{1, \dots, m\}. \tag{3.11}$$

For each fixed x, let $g(t) := \sum_\sigma \pi_2(\sigma)\psi(\Delta^{\sigma_2}\Delta^{\sigma_1}x)$. Here we take the abridged notation $\sigma_i = \sigma_i(w)$. We will prove (3.10) by calculating the limit in the left-hand side with de l'Hôpital's rule.

$$\begin{aligned} g'(t) &= \sum_{\sigma\in\mathfrak{S}_2}\pi_2(\sigma)\big\{X_{\sigma_1}(\psi\Delta^{\sigma_2})(\Delta^{\sigma_1}x) + X_{\sigma_2}\psi(\Delta^{\sigma_2}\Delta^{\sigma_1}x)\big\} \\ &= \sum_{\sigma\in\mathfrak{S}_2}\pi_2(\sigma)\big\{X_{\sigma_1}\psi(\Delta^{\sigma_2}\Delta^{\sigma_1}x) + X_{\sigma_2}\psi(\Delta^{\sigma_2}\Delta^{\sigma_1}x) \\ &\qquad + X_{\sigma_2\sigma_1}\psi(\Delta^{\sigma_2}\Delta^{\sigma_1}x)(-t) + O_3(t^2,\psi,\Delta^{\sigma_2\sigma_1}x)\big\}. \end{aligned} \tag{3.12}$$

Here we already used Theorem 2.6 and we also invoked (3.11) to claim that $\mathrm{ad}_{X_{\sigma_2}} X_{\sigma_1} = X_{\sigma_2\sigma_1}$. To accomplish the proof for $\ell = 2$, observe first that

$$\lim_{t\to 0}\frac{1}{2t}\sum_{\sigma\in\mathfrak{S}_2}\pi_2(\sigma)\Big(X_{\sigma_2\sigma_1}\psi(\Delta^{\sigma_2}\Delta^{\sigma_1}x)(-t) + O_3(t^2,\psi,\Delta^{\sigma_2\sigma_1}x)\Big) = X_{w_1w_2}\psi(x).$$

Here we used estimate (2.18) and the definition of $\pi_2(\sigma)$. Therefore the last line of (3.12) has the expected behaviour. It remains to show that the second one behaves as $O(t^2)$, as $t \to 0$. To prove this claim, introduce $\varphi := X_{w_1}\psi + X_{w_2}\psi \in C^{1,1}_{\mathcal{H},\mathrm{loc}}$. The Taylor formula (2.22) gives

$$\begin{aligned} \varphi(\Delta^{w_2}\Delta^{w_1}x) - \varphi(\Delta^{w_1}\Delta^{w_2}x) &= \varphi(x) + (X_{w_2}\varphi(x) + X_{w_1}\varphi(x))t + R_2(t^2,\varphi,x) \\ &\quad - \big\{\varphi(x) + (X_{w_1}\varphi(x) + X_{w_2}\varphi(x))t + R_2(t^2,\varphi,x)\big\} = R_3(t^2,\psi,x), \end{aligned}$$

as $t \to 0$. Therefore,

$$\lim_{t\to 0}\frac{1}{t}\sum_{\sigma\in\mathfrak{S}_2}\pi_2(\sigma)\Big(X_{\sigma_1}\psi(\Delta^{\sigma_2\sigma_1}x) + X_{\sigma_2}\psi(\Delta^{\sigma_2\sigma_1}x)\Big) = \lim_{t\to 0}\frac{1}{t}\sum_{\sigma\in\mathfrak{S}_2}\pi_2(\sigma)\varphi(\Delta^{\sigma_2\sigma_1}x) = 0,$$

as desired.

Next we show the induction step (which is not needed if $s \leq 3$). Assume that $s \geq 4$ and that for some $\ell \in \{3, \dots, s-1\}$ we have

$$X_v \varphi = X_v^\sharp \varphi \quad \text{for all } \varphi \in C^1_{\mathrm{Euc}} \cap C^{\ell-1,1}_{\mathcal{H},\mathrm{loc}} \quad |v| \leq \ell - 1. \tag{3.13}$$

We want to show (3.9) for all $\psi \in C^{\ell,1}_{\mathcal{H},\mathrm{loc}} \cap C^1_{\mathrm{Euc}}$ and $|w| = \ell$.

Fix x and let $g(t) := \sum_{\sigma \in \mathfrak{S}_\ell} \pi_\ell(\sigma) \psi(\Delta^{\sigma_\ell} \cdots \Delta^{\sigma_1} x)$, where $\sigma_j = \sigma_j(w)$. It suffices to show that $\lim_{t\to 0} g(t)/t^\ell = X_w \psi(x)$. This will follow by de l'Hôpital's rule, as soon as we prove that

$$\lim_{t\to 0} \frac{g'(t)}{\ell t^{\ell-1}} = X_w \psi(x). \tag{3.14}$$

To show (3.14), observe first that by the induction assumption we have

$$\mathrm{ad}_{X_j} X_v = X_{jv} \qquad \text{for all } j \in \{1, \dots, m\} \quad |v| \leq \ell - 1. \tag{3.15}$$

Now we calculate $g'(t)$ keeping (2.20) into account.

$$\begin{aligned} g'(t) &= \sum_{\sigma \in \mathfrak{S}_\ell} \pi_\ell(\sigma) \sum_{j=1}^{\ell} X_{\sigma_j} \big(\psi \Delta^{\sigma_\ell \cdots \sigma_{j+1}}\big)\big(\Delta^{\sigma_j \cdots \sigma_1} x\big) \\ &= \sum_{\sigma \in \mathfrak{S}_\ell} \pi_\ell(\sigma) \sum_{j=1}^{\ell} \sum_{0 \leq k_{j+1} + \cdots + k_\ell \leq \ell - 1} X_{\sigma_\ell^{k_\ell} \cdots \sigma_{j+1}^{k_{j+1}} \sigma_j} \psi(\Delta^{\sigma_\ell \cdots \sigma_1} x) \frac{(-t)^{k_{j+1} + \cdots + k_\ell}}{k_{j+1}! \cdots k_\ell!} \\ &\quad + O_{\ell+1}(t^\ell, \psi, \Delta^{\sigma_\ell \cdots \sigma_1} x). \end{aligned}$$

In view of (3.15), we can expand as in (2.20) and use identity $\mathrm{ad}^{k_\ell}_{X_{\sigma_\ell}} \cdots \mathrm{ad}^{k_{j+1}}_{X_{\sigma_{j+1}}} X_{\sigma_j} \psi = X_{\sigma_\ell^{k_\ell} \cdots \sigma_{j+1}^{k_{j+1}} \sigma_j} \psi$, which is legitimate because $k_\ell + \cdots + k_{j+1} \leq \ell - 1$, see (3.15). We may rearrange as

$$\begin{aligned} g'(t) &= \sum_{\sigma \in \mathfrak{S}_\ell} \pi_\ell(\sigma) \sum_{1 \leq i_1 \leq \ell} X_{\sigma_{i_1}} \psi(\Delta^{\sigma_\ell \cdots \sigma_1} x) \\ &+ \sum_{\mu=2}^{\ell} (-t)^{\mu-1} \sum_{p=2}^{\mu} \sum_{\substack{1 + b_2 + \cdots + b_p = \mu \\ b_2, \dots, b_p \geq 1}} \frac{1}{b_2! \cdots b_p!} \sum_{\sigma \in \mathfrak{S}_\ell} \pi_\ell(\sigma) \sum_{1 \leq i_1 < \cdots < i_p \leq \ell} X_{\sigma_{i_p}^{b_p} \cdots \sigma_{i_2}^{b_2} \sigma_{i_1}} \psi(\Delta^{\sigma_\ell \cdots \sigma_1} x) \\ &+ O_{\ell+1}(t^\ell, \psi, \Delta^{\sigma_\ell \cdots \sigma_1} x) \\ &=: H_1(t) + \sum_{\mu=2}^{\ell-1} (-t)^{\mu-1} H_\mu(t) + (-t)^{\ell-1} \sum_{p=2}^{\ell-1} h_{\ell,p}(t) + (-t)^{\ell-1} h_{\ell,\ell}(t) \\ &+ O_{\ell+1}(t^\ell, \psi, \Delta^{\sigma_\ell \cdots \sigma_1} x). \end{aligned}$$

Everywhere σ_j stands for $\sigma_j(w)$.

The proof of (3.14) will be a consequence of the following three facts.

Fact 1. We have

$$\lim_{t \to 0} \frac{(-t)^{\ell-1} h_{\ell,\ell}(t)}{\ell t^{\ell-1}} = \frac{(-1)^\ell}{\ell} h_{\ell,\ell}(0) = X_w \psi(x).$$

Fact 2. For any $p \in \{1, \dots, \ell - 1\}$, we have

$$\lim_{t\to 0} \frac{(-t)^{\ell-1} h_{\ell,p}(t)}{t^{\ell-1}} = (-1)^{\ell-1} h_{\ell,p}(0) = 0.$$

Fact 3. We have

$$\lim_{t\to 0} \sum_{\mu=1}^{\ell-1} \frac{(-t)^{\mu-1} H_\mu(t)}{t^{\ell-1}} = 0. \tag{3.16}$$

Facts 1,2, and 3 give easily the proof of (3.14) and of the theorem.

To check *Fact 1*, just observe that property (2.5) and the generalized Jacobi identity (3.5) give

$$\lim_{t\to 0} \frac{(-t)^{\ell-1}}{\ell t^{\ell-1}} h_{\ell,\ell}(t) = \frac{(-1)^{\ell-1}}{\ell} h_{\ell,\ell}(0) = \frac{(-1)^{\ell-1}}{\ell} \sum_{\sigma\in\mathfrak{S}_\ell} \pi_\ell(\sigma) X_{\sigma_\ell\cdots\sigma_1}\psi(x) = X_w\psi(x),$$

as desired. Note that we used $\lim_{t\to 0} h_{\ell,\ell}(t) = h_{\ell,\ell}(0)$.

To verify *Fact 2*, note first that $\lim_{t\to 0} h_{\ell,p}(t) = h_{\ell,p}(0)$. Thus

$$\begin{aligned} h_{\ell,p}(0) &= \sum_{\substack{1+b_2+\cdots+b_p=\ell \\ b_2,\dots,b_p\geq 1}} \frac{1}{b_2!\cdots b_p!} \Big\{ \sum_{\sigma\in\mathfrak{S}_\ell} \pi_\ell(\sigma) \sum_{i\leq i_1<\cdots<i_p\leq \ell} X_{\sigma_{i_p}^{b_p}\cdots\sigma_{i_2}^{b_2}\sigma_{i_1}} \psi(x) \Big\} \\ &= \sum_{\substack{1+b_2+\cdots+b_p=\ell \\ b_2,\dots,b_p\geq 1}} \frac{1}{b_2!\cdots b_p!} 0, \end{aligned}$$

because for any $p \leq \ell - 1$ and $b_2, \dots, b_p \geq 1$, the term $\{\cdots\}$ vanishes by Proposition 3.7 below.

Finally we discuss *Fact 3*. Here it does not suffice to know that $H_\mu(t) \to H_\mu(0)$, as $t \to 0$. We need instead a more refined expansion, whose explicit analysis is of considerable algebraic difficulty. Therefore, we use a slightly more implicit argument. First of all we expand all the terms by means of the Taylor formula, taking into account that $\psi \in C^{\ell,1}_{\mathcal{H},\mathrm{loc}}$. By inductive assumption we may claim that $X_{\sigma_{i_p}^{b_p}\cdots\sigma_{i_2}^{b_2}\sigma_{i_1}}\psi = X^\sharp_{\sigma_{i_p}^{b_p}\cdots\sigma_{i_2}^{b_2}\sigma_{i_1}}\psi$. Thus we express the latter as a sum of horizontal derivatives of order μ with suitable coefficients. This gives for $\mu \in \{2, \dots, , \ell - 1\}$,

$$\begin{aligned} H_\mu(t) &= \sum_{p=1}^{\mu} \sum_{\substack{1+b_2+\cdots+b_p=\mu \\ b_2,\dots,b_p\geq 1}} \frac{1}{b_2!\cdots b_p!} \sum_\sigma \pi_\ell(\sigma) \sum_{1\leq i_1<\cdots<i_p\leq \ell} X^\sharp_{\sigma_{i_p}^{b_p}\cdots\sigma_{i_2}^{b_2}\sigma_{i_1}} \psi(\Delta^{\sigma_\ell\cdots\sigma_1} x) \\ &=: \sum_{p,b,\sigma,i} \sum_{(k_1,\dots,k_\mu)\in\{w_1,\dots,w_\ell\}^\mu} c^k_{p,b,\sigma,i} X^\sharp_{k_1}\cdots X^\sharp_{k_\mu}\psi(\Delta^{\sigma_\ell\cdots\sigma_1}x) \\ &= \sum_{|\alpha|=0}^{\ell-\mu} \sum_{p,b,\sigma,i} \sum_{(k_1,\dots,k_\mu)\in\{w_1,\dots,w_\ell\}^\mu} \frac{c^k_{p,b,\sigma,i}}{\alpha!} t^{|\alpha|} (X^\sharp_{\sigma_1})^{\alpha_1}\cdots(X^\sharp_{\sigma_\ell})^{\alpha_\ell} X^\sharp_{k_1}\cdots X^\sharp_{k_\mu}\psi(x) \\ &\qquad + R_{\ell+1}(t^{\ell+1-\mu}, \psi), \end{aligned}$$

where we also used the Taylor expansion. A similar expansion holds for $\mu = 1$. Algebra of such coefficients is quite complicated, and it seems rather difficult to show *Fact 3* directly. We are instead able to prove what we need indirectly. What we actually have is a polynomial expansion of the form

$$\sum_{\mu=1}^{\ell-1}(-t)^{\mu-1}H_\mu(t) = \sum_{\lambda=1}^{\ell} t^{\lambda-1}P_\lambda(X^\sharp_{w_1},\dots,X^\sharp_{w_\ell})\psi(x) + R_{\ell+1}(t^\ell,\psi,x), \tag{3.17}$$

where P_λ is an homogeneous polynomial of degree λ involving the coefficents $c^k_{p,b,\sigma,i}/\alpha!$ above. Taking *Fact 1* and *Fact 2* for granted, this gives

$$\begin{aligned} X^\sharp_w\psi(x) &= \lim_{t\to 0}\frac{g(t)}{t^\ell} \overset{\text{(H)}}{=} \lim_{t\to 0}\frac{g'(t)}{\ell t^{\ell-1}} \\ &= X_w\psi(x) + \lim_{t\to 0}\frac{1}{\ell t^{\ell-1}}\Big(\sum_{\mu=1}^{\ell-1}(-t)^{\mu-1}H_\mu(t) + R_{\ell+1}(t^\ell,\psi)\Big) \\ &= X_w\psi(x) + \lim_{t\to 0}\frac{1}{\ell t^{\ell-1}}\sum_{\lambda=1}^{\ell} t^{\lambda-1}P_\lambda(X^\sharp_{w_1},\dots,X^\sharp_{w_\ell})\psi(x) \end{aligned} \tag{3.18}$$

Equality $\overset{\text{(H)}}{=}$ should be iterpreted in the usual conditional sense provided by de l'Hôpital's rule (the limit in the left-hand side exists and takes a value L if the limit in the right-hand side exists and takes the same value L). We do not know at this stage the value of the limit in the right-hand side. Our purpose is to show that it vanishes.

To prove such claim, note that equality (3.18) has an algebraic feature. Namely, all the coefficients $c^k_{p,b,\sigma,i}$ appearing implicitely in the polynomials P_λ do not change if we take different vector fields Z_j instead ov X_{w_j} in some $\mathbb{R}^N$ with possibly $N \neq n$, provided that we do not change the number ℓ of vector fields.

If we choose analytic vector fields Z_j in $\mathbb{R}^N$, we clearly have $Z_w\psi = Z^\sharp_w\psi$ for all w and for any $\psi \in C^\omega$. Moreover, the conditional equality (3.18) becomes a true equality, because all functions depend analytically on t and x. Therefore we have found a family of polynomial identities of the form

$$0 = P_\lambda(Z_1,\dots,Z_\ell)\psi(x) =: \sum_{(k_1,\dots,k_\lambda)\in\{1,\dots,\ell\}^\lambda} C(k_1,\dots,k_\lambda)Z_{k_1}\cdots Z_{k_\lambda}\psi(x)$$

which holds for any family $Z_1,\dots,Z_\ell$ of analytic vector fields in $\mathbb{R}^N$, for each $N\in\mathbb{N}$, for all analytic $\psi:\mathbb{R}^N\to\mathbb{R}$ and any $x\in\mathbb{R}^N$. Theorem 3.8 shows that the polynomial should be trivial, i.e. $C(k_1,\dots,k_\lambda) = 0$ for all $(k_1,\dots,k_\lambda)$. This concludes the proof of *Fact 3* and of the theorem. □

Next we state and prove the relevant results needed to accomplish the proof of *Fact 2* and *Fact 3*, that we took for granted in the argument above.

The following family of nested commutators identities is relevant for the proof of *Fact 2*.

Proposition 3.7. *Let $X_1,\dots,X_m$ be vector fields in the regularity class $C^{s-1,1}_{\mathcal{H},\mathrm{loc}}\cap C^1_{\mathrm{Euc}}$. For any $\ell\in\{2,\dots,s\}$ and $1\le p\le \ell-1$, we have the following statement:*

$$\begin{aligned}\sum_{\sigma\in\mathfrak{S}_\ell}\pi_\ell(\sigma)\sum_{1\le i_1<\dots<i_p\le\ell} X_{\sigma_{i_p}^{b_p}(v)\cdots\sigma_{i_2}^{b_2}(v)\sigma_{i_1}^{b_1}(v)w}=0\quad &\text{for all } b_1,\dots,b_p\in\mathbb{N}\cup\{0\}\\ |w|\ge 0\quad |v|=\ell\quad 1\le b_1+\dots+b_p\le s-|w|.&\end{aligned}\tag{$F_{\ell,p}$}$$

We agree that if $|w|=0$, then $X_{vw}=X_v$ for any word v with $|v|\ge 1$. To prove *Fact 2* we need the case $|w|=0$ and $b_1=1$ of the proposition, but the case $|w|=0$ is included for convenience in the proof. Observe also that

- if $|w|=0$ and $b_1\ge 2$, then the statement is trivial;
- if $\ell=1$, then the statement is empty;
- Proposition 3.7 fails for $\ell=p$, as (3.5) shows.

Since the statement of Proposition 3.7 is quite intricated, we first check its correctness in the already significant case $\ell=3$ and $p=2$, $|w|=0$ and $b\in\mathbb{N}$. The general case is based on the same cancellation mechanism. In this model case, identity ($F_{3,2}$) becomes

$$\sum_{\sigma\in\mathfrak{S}_3}\pi_3(\sigma)\big\{X_{\sigma_3^b\sigma_2}+X_{\sigma_3^b\sigma_1}+X_{\sigma_2^b\sigma_1}\big\}=0,$$

which can be checked by writing explicitly the twelve terms (in the notation $[j^bk]:=X_{j^bk}$):

$$\begin{aligned}&[3^b2]+[3^b1]+[2^b1]-\{[2^b3]+[2^b1]+[3^b1]\}\\&\qquad-\{[1^b3]+[1^b2]+[3^b2]\}+\{[1^b2]+[1^b3]+[2^b3]\}=0.\end{aligned}$$

Proof of Propoosition 3.7. We first prove by induction that ($F_{\ell,1}$) holds for any $\ell\in\{2,\dots,s\}$. Introduce the abridged notation $[i_1^{b_1}i_2^{b_2}]$ instead of $X_{i_1^{b_1}i_2^{b_2}}$ and so on. For convenience of notation, we prove $F_{\ell+1,1}$ for all $\ell\in\{1,\dots,s-1\}$. Let $\widetilde{v}=v_0v\in\mathcal{W}_{\ell+1}$, w and $b_1\ge 1$ be such that $b_1+|w|\le s$. Then

$$\begin{aligned}\sum_{\widetilde\sigma\in\mathfrak{S}_{\ell+1}}\pi_\ell(\widetilde\sigma)\sum_{0\le i_1\le\ell}[\widetilde\sigma_{i_1}^{b_1}(\widetilde v)w]&=\sum_{\widetilde\sigma\in\mathfrak{S}_{\ell+1}}\pi_\ell(\widetilde\sigma)\big([\widetilde\sigma_0^{b_1}(\widetilde v)w]+[\widetilde\sigma_1^{b_1}(\widetilde v)w]+\dots+[\widetilde\sigma_\ell^{b_1}(\widetilde v)w]\big)\\&=\sum_{\sigma\in\mathfrak{S}_\ell}\pi_\ell(\sigma)\big([v_0^{b_1}w]+[\sigma_1^{b_1}(v)w]+\dots+[\sigma_\ell^{b_1}(v)w]\\&\qquad-\big([\sigma_1^{b_1}(v)w]+\dots+[\sigma_\ell^{b_1}(v)w]+[v_0^{b_1}w]\big)\big)=0,\end{aligned}$$

as we claimed.

To fill up the triangle, we prove that if ($F_{\ell,p-1}$) holds for some $\ell\in\{2,\dots,s-1\}$ and $p\in\{2,\dots,\ell\}$, then ($F_{\ell+1,p}$) holds. This will imply that ($F_{\ell,p}$) holds for all the required couples (p,ℓ). We argue as usual by the defining property (2.3). Denote below

$\widetilde{v} = v_0 v \in \mathcal{W}_{\ell+1}$.

$$\begin{aligned}
&\sum_{\widetilde{\sigma}\in\mathfrak{S}_{\ell+1}} \pi_{\ell+1}(\widetilde{\sigma}) \sum_{0\le i_1<\cdots<i_p\le\ell} [\widetilde{\sigma}_{i_p}^{b_p}(\widetilde{v})\cdots\widetilde{\sigma}_{i_1}^{b_1}(\widetilde{v})w] \\
&= \sum_{\sigma\in\mathfrak{S}_\ell} \pi_\ell(\sigma)\Big(\sum_{0\le i_1<\cdots<i_p\le\ell} [\widetilde{\sigma}_{i_p}^{b_p}(\widetilde{v})\cdots\widetilde{\sigma}_{i_1}^{b_1}(\widetilde{v})w]\Big)\Big|_{\widetilde{\sigma}(v_0v)=v_0\sigma(v)} \\
&\qquad - \sum \pi_\ell(\sigma)\Big(\sum_{0\le i_1<\cdots<i_p\le\ell} [\widetilde{\sigma}_{i_p}^{b_p}(\widetilde{v})\cdots\widetilde{\sigma}_{i_1}^{b_1}(\widetilde{v})w]\Big)\Big|_{\widetilde{\sigma}(v_0v)=\sigma(v)v_0} \\
&= \sum_{\sigma\in\mathfrak{S}_\ell} \pi_\ell(\sigma)\Big(\sum_{1\le i_1<\cdots<i_p\le\ell} [\widetilde{\sigma}_{i_p}^{b_p}(\widetilde{v})\cdots\widetilde{\sigma}_{i_1}^{b_1}(\widetilde{v})w] + \sum_{\substack{1\le i_2<\cdots<i_p\le\ell\\ i_1=0}} [\widetilde{\sigma}_{i_p}^{b_p}(\widetilde{v})\cdots\widetilde{\sigma}_{i_1}^{b_1}(\widetilde{v})w]\Big)\Big|_{\widetilde{\sigma}(\widetilde{v})=v_0\sigma(v)} \\
&- \sum_{\sigma\in\mathfrak{S}_\ell} \pi_\ell(\sigma)\Big(\sum_{0\le i_1<\cdots<i_p\le\ell-1} [\widetilde{\sigma}_{i_p}^{b_p}(\widetilde{v})\cdots\widetilde{\sigma}_{i_1}^{b_1}(\widetilde{v})w] + \sum_{\substack{0\le i_1<\cdots<i_{p-1}\le\ell-1\\ i_p=\ell}} [\widetilde{\sigma}_{i_p}^{b_p}(\widetilde{v})\cdots\widetilde{\sigma}_{i_1}^{b_1}(\widetilde{v})w]\Big)\Big|_{\widetilde{\sigma}(\widetilde{v})=\sigma(v)v_0} \\
&= \sum_{\sigma\in\mathfrak{S}_\ell} \pi_\ell(\sigma)\Big\{\sum_{1\le i_1<\cdots<i_p\le\ell} [\sigma_{i_p}^{b_p}(v)\cdots\sigma_{i_1}^{b_1}(v)w] + \sum_{1\le i_2<\cdots<i_p\le\ell} [\sigma_{i_p}^{b_p}(v)\cdots\sigma_{i_2}^{b_2}(v)v_0^{b_1}w]\Big\} \\
&- \Big\{\sum_{0\le i_1<\cdots<i_p\le\ell-1} [\sigma_{i_p+1}^{b_p}(v)\cdots\sigma_{i_1+1}^{b_1}(v)w] + \sum_{0\le i_1<\cdots<i_{p-1}\le\ell-1} [v_0^{b_p}\sigma_{i_{p-1}+1}^{b_{p-1}}(v)\cdots\sigma_{i_1+1}^{b_1}(v)w]\Big\} \\
&= 0,
\end{aligned}$$

because the first and the third term cancel, while both the second and the fourth vanish by inductive assumption. ☐

The following theorem has been used to check *Fact 3* in the proof of Theorem 3.1. See (3.16).

Theorem 3.8. *Let m and p be natural numbers. Let $C : \{1,\dots,m\}^p \to \mathbb{R}$ be given coefficients. Consider the polynomial*

$$P(X_1,\dots,X_m) := \sum_{(k_1,\dots,k_p)\in\{1,\dots,m\}^p} C(k_1,\dots,k_p)X_{k_1}X_{k_2}\cdots X_{k_p}. \tag{3.19}$$

Assume that for all $N\in\mathbb{N}$, for any $X_1,\dots,X_m$ analytic vector fields in $\mathbb{R}^N$ and for each analytic $\psi:\mathbb{R}^N\to\mathbb{R}$ we have

$$P(X_1,\dots,X_m)\psi(x) = 0 \quad \text{for all } x\in\mathbb{R}^N. \tag{3.20}$$

Then P is the trivial polynomial, i.e. $C(k_1,\dots,k_p)=0$ for all $(k_1,\dots,k_p)\in\{1,\dots,m\}^p$.

Proof. The argument is inspired to some ideas contained in the proof the Amitsur–Levitzki theorem [Lev50, AL50]. We start by separating homogeneous parts in each variable. Let $N\in\mathbb{N}$ and take $X_1,\dots,X_m$ analytic vector fields in $\mathbb{R}^N$ and ψ analytic in $\mathbb{R}^N$. Consider the function

$$\begin{aligned}
f(t_1,\dots,t_m) :=& P(t_1X_1,\dots,t_mX_m)\psi(x) \\
=:& \sum_{q=1}^{\min\{p,m\}} \sum_{1\le i_1<\cdots<i_q\le m} \sum_{\substack{d_1,\dots,d_q\ge1\\ d_1+\cdots+d_q=p}} t_{i_1}^{d_1}\cdots t_{i_q}^{d_q} P_{d_1\cdots d_q}^{i_1\cdots i_q}(X_{i_1},\dots,X_{i_q})\psi(x),
\end{aligned}$$

where x is fixed. [2] The function f should vanish identically in $t_1, \dots, t_m$. Therefore it is clear that it must be for each fixed $q, i_1, \dots, i_q, d_1, \dots, d_q$

$$P^{i_1 \cdots i_q}_{d_1 \cdots d_q}(X_{i_1}, \dots, X_{i_q})\psi(x) = 0 \quad \text{for all } X_{i_1}, \dots, X_{i_q}, \psi \in C^\omega(\mathbb{R}^N) \quad N \in \mathbb{N} \quad x \in \mathbb{R}^N.$$

In other words we can work with homogeneous polynomials in each variable. Renaming variables, it suffices to prove the theorem for a polynomial P in q variables, where $1 \le q \le p$ and such that

$$P(\lambda_1 X_1, \dots, \lambda_q X_q) = \lambda_1^{d_1} \cdots \lambda_q^{d_q} P(X_1, \dots, X_q) \qquad \text{for all } \lambda_1, \dots, \lambda_q \in \mathbb{R},$$

where $d_1, \dots, d_q \ge 1$.

Next we show by a standard multilinearization argument that, possibly adding new variables, we can assume that $d_j = 1$ for all $j = 1, \dots, q$. Indeed, assume that $d_1 \ge 2$. Define

$$\widetilde{P}(U, T, X_2, \dots, X_q) := P(U + T, X_2, \dots, X_q) - P(U, X_2, \dots, X_q) - P(T, X_2, \dots, X_q).$$

It turns out that $\widetilde{P}$ is a homogeneous polynomial in $q + 1$ variables, but the degrees in the new variables U and T are both strictly less that the original degree d_1. Note that if $P(X_1, X_2, \dots, X_q)\psi \equiv 0$ for all $\psi, X_1, \dots, X_q \in C^\omega$, then $\widetilde{P}(U, T, X_2, \dots, X_q)\psi \equiv 0$ for all $\psi, U, T, X_2, \dots, X_q \in C^\omega$. On the other side, if $\widetilde{P}$ is the trivial polynomial (all its coefficients vanish), then also the polynomial P must be trivial. Clearly, the polynomial $\widetilde{P}$ can be decomposed in a sum of homogeneous polynomials, where each of them is homogeneous in each variable separately, as above.

Iterating this argument we may assume that we have a polynomial of the form

$$Q(X_1, \dots, X_p) = \sum_{\sigma \in \mathfrak{S}_p} B(\sigma) X_{\sigma_1} \cdots X_{\sigma_p}$$

in p variables, where p is the original degree of the polynomial P in (3.19). Here $12 \cdots p \mapsto \sigma_1 \sigma_2 \cdots \sigma_p$ are permutations and $\sigma_j = \sigma_j(12 \cdots p)$. We know that

$$Q(X_1, \dots, X_p)\psi(x) = 0 \quad \text{for all } X_1, \dots, X_p, \psi \in C^\omega \quad N \in \mathbb{N} \quad x \in \mathbb{R}^N \tag{3.21}$$

and we want to show that $B(\sigma) = 0$ for all σ. Since we are free to increase the dimension N of the underlying space, take $N \ge p+1$, let $X_j = x_j \partial_{j+1}$ for any $j = 1, \dots, p$. Therefore, it turns out that, if we let $\psi(x) = x_{p+1}$, we have

$$X_{\sigma_1} \cdots X_{\sigma_p} \psi = \begin{cases} 1 & \text{if } \sigma_1 \cdots \sigma_p = 1 \cdots p; \\ 0 & \text{if } \sigma_1 \cdots \sigma_p \neq 1 \cdots p. \end{cases}$$

Therefore, if we make use of (3.21), we discover that it must be $B(1, 2, \dots, p) = 0$. Letting then $X_{\sigma_j} = x_j \partial_{j+1}$ we see that $B(\sigma) = 0$ for all $\sigma \in \mathfrak{S}_p$. Therefore Q is the trivial polynomial and the proof is concluded. □

[2] An informal example to understand quickly this splitting could be:

$$\begin{aligned} P(X_1, X_2, X_3) &= X_1^2 X_2^2 + X_3^4 + (X_1 X_2 X_3 X_1 + X_1^2 X_3 X_2) \\ &= P^{1,2}_{2,2}(X_1, X_2) + P^3_4(X_3) + P^{1,2,3}_{2,1,1}(X_1, X_2, X_3). \end{aligned}$$

3.1. An old nested commutators identity due to Baker

Here we show as an application that some very old nested commutator identities going back to Baker (see the discussion in [Ote91]) can be found as a particular case of our Proposition 3.3. All vector fields in this subsection are smooth.

Let $v = v_1 \cdots v_\ell$ be a word of length ℓ in the alphabet $1, \cdots, m$. Let us adopt the notation

$$X_a^k X_b := \begin{cases} X_a X_b & \text{if } k = 1 \\ X_b X_a & \text{if } k = -1 \end{cases} \quad \text{for all } a, b \in \{1, \dots, m\},$$

$$X_a^k X_b^h X_c := \begin{cases} X_a X_b^h X_c & \text{if } k = 1 \text{ and } h \in \{-1, 1\} \\ X_b^h X_c X_a & \text{if } k = -1 \text{ and } h \in \{-1, 1\} \end{cases} \quad \text{for all } a, b, c \in \{1, \dots, m\}$$

and analogous notation for higher order derivatives. Then it is rather easy to check that we may write for all $v = v_1 \cdots v_\ell$

$$X_{v_1 \cdots v_\ell} = \sum_{k_1, \dots, k_{\ell-1} \in \{-1,1\}} (-1)^{k_1 + \cdots + k_{\ell-1}} X_{v_1}^{k_1} \cdots X_{v_{\ell-1}}^{k_{\ell-1}} X_{v_\ell}. \tag{3.22}$$

This is an alternative way to write commutators, less focused on the inductive point of view than the form (2.4). Let now $n \in \mathbb{N}$, $v \in \mathcal{W}_{n+1}$ and $w \in \mathcal{W}_1$. Thus,

$$X_{wv} + \sum_{\sigma \in \mathfrak{S}_{n+1}} \pi_{n+1}(\sigma) X_{\sigma_1(v) \cdots \sigma_{n+1}(v) w} = -X_{[v]w} + \sum_{\sigma \in \mathfrak{S}_{n+1}} \pi_{n+1}(\sigma) X_{\sigma_1(v) \cdots \sigma_{n+1}(v) w} = 0,$$

by (3.4). Comparing (3.22) and (2.4), this is equivalent to

$$X_{wv_1 \cdots v_n v_{n+1}} + \sum_{k_1, \dots, k_n \in \{-1,1\}} (-1)^{k_1 + \cdots + k_n} X_{v_1^{k_1} \cdots v_n^{k_n} v_{n+1}, w} = 0,$$

or in the typographically better, self-explanatory notation

$$[wv_1 \cdots v_n v_{n+1}] + \sum_{k_1, \dots, k_n \in \{-1,1\}} (-1)^{k_1 + \cdots + k_n} [v_1^{k_1} \cdots v_n^{k_n} v_{n+1}, w] = 0. \tag{3.23}$$

Note that we introduced a comma before w to avoid confusion. Namely, when some of the v_j has power -1, then it goes on the right side of the previous $v_{j+1}, \dots, v_n, v_{n+1}$ but *not* of w. For instance, we have $[v_1^{-1} v_2^{-1} v_3 v_4, w] := [v_3 v_4 v_2 v_1 w]$ and so on (a precise definition can be given by induction).

Now we show that the following Baker's identity of order six

$$[ab^4 a] - 2[bab^3 a] + [b^2 ab^2 a] = 0 \quad \text{for all } a, b \in \{1, \dots, m\}, \tag{3.24}$$

see [Ote91, eq. (4.4)], can be easily obtained specializing (3.23). Let $n = 4$ and choose $v_1 \cdots v_n v_{n+1} = v_1 \cdots v_4 v_5 = b \cdots ba = b^4 a$ and $w = a$. Thus (3.23) becomes

$$[ab^4 a] + \sum_{k_1, \dots, k_4 \in \{-1,1\}} (-1)^{k_1 + \cdots + k_4} [b^{k_1} \cdots b^{k_4} a, a] = 0.$$

Note that at least one among the numbers k_j must be -1, otherwise we get $[b^4aa] = 0$. Therefore we get

$$[ab^4a] + \Big\{-\binom{4}{1}[b^3aba] + \binom{4}{2}[b^2ab^2a] - \binom{4}{3}[bab^3a] + \binom{4}{4}[ab^4a]\Big\} = 0$$

which gives $[ab^4a] - 2[b^3aba] + 3[b^2ab^2a] - 2[bab^3a] = 0$. But the fourth order identity (3.6) gives $[b^3aba] = [b^2ab^2a]$. Thus (3.24) follows.

4. Applications to ball-box theorems

In this section we describe some applications of our results to ball-box theorems. We shall use some results from [MM12a] on the expansion of almost exponential maps.

Assume that a family $\mathcal{H}$ of vector fields belongs to the regularity class $C^{s-1,1}_{\mathcal{H},\mathrm{loc}} \cap C^1_{\mathrm{Euc}}$ and assume that the vector fields satisfy the Hörmander condition of step s, namely $\dim\operatorname{span}\{X_w(x)\colon 1 \le |w| \le s\} = n$, at any $x \in \mathbb{R}^n$. Following a standard notation, denote by $\mathcal{P} := \{Y_1, \dots, Y_q\} = \{X_w : 1 \le |w| \le s\}$ the family of commutators of length at most s. Let $\ell_j \le s$ be the length of Y_j and write $Y_j =: g_j \cdot \nabla$. For each $I = (i_1, \dots, i_n) \in \{1, \dots, q\}^n$, let $\ell(I) = \ell_{i_1} + \cdots + \ell_{i_n}$ $\lambda_I(x) := \det[Y_{i_1}(x), \dots, Y_{i_n}(x)]$ and $\ell(I) := \ell_{i_1} + \cdots + \ell_{i_n}$. Define also the vector valued function $\Lambda(x, r) := (\lambda_I(x) r^{\ell(I)})_{I \in \{1,\dots,q\}^n}$. Finally, for all $A \subset \mathbb{R}^n$, put

$$\nu(A) := \inf_{x \in A} |\Lambda(x, 1)|. \tag{4.1}$$

Assume that each commutator Y_j is continuous in the Euclidean topology. Then, on the open set $\Omega_0 \subset \mathbb{R}^n$ fixed before (2.8), we have $\nu(\Omega_0) > 0$. Moreover, take $j \in \{1, \dots, m\}$ and any word w with $|w| = s$. For any $x \in \Omega_0$ where the derivative $X_j^\sharp f_w(x)$ exists, we have the obvious bound $|X_j^\sharp f_w(x)| \le L_0$, the constant in (2.8). Furthermore we also have $|X_w f_j(x)| \le L_0$ for all x. Therefore we can write

$$\operatorname{ad}_{X_j} X_w(x) = \sum_{1 \le |u| \le s} b^u X_u(x) \quad \text{where} \tag{4.2}$$

$$|b^u| \le C_0 \qquad \text{for all } u \text{ with } 1 \le |u| \le s. \tag{4.3}$$

Here the constant C_0 can be estimated in terms of the constant L_0 in (2.8) and of the infimum $\nu(\Omega_0)$; see [MM12b, Lemma 4.2].

Therefore, the vector fields are in the class $\mathcal{A}_s$ introduced in in [MM12a] (actually in a subclass, because here we assume the Hörmander condition, while in [MM12a] we did not). Moreover we have the following measurability property:

Proposition 4.1 (measurability). *Let $\mathcal{H}$ be a family of vector fields in the regularity class $C^1_{\mathrm{Euc}} \cap C^{s-1,1}_{\mathcal{H},\mathrm{loc}}$. Assume the Hörmander condition at step s and assume that $f_w \in C^0_{\mathrm{Euc}}$, if $1 \le |w| \le s$. Let w be a word with $|w| = s$ and let $Z = f \cdot \nabla \in \pm\mathcal{H}$. Then for any $x \in \Omega$ we can write*

$$\operatorname{ad}_Z X_w(e^{tZ}x) = \sum_{1 \le |v| \le s} b^v(t) X_v(e^{tZ}x) \quad \textit{for a.e. } t \in (-t_0, t_0), \tag{4.4}$$

where the functions $t \mapsto b^v(t)$ are measurable *and $|b^v(t)| \le C_0$, the constant in (4.3).*

Proof. Denote $\gamma(t) := e^{tZ}x$. Since $t \mapsto f_w(\gamma(t))$ is Lipschitz and $x \mapsto X_w f(x)$ is continuous, the function $t \mapsto \operatorname{ad}_Z X_w(\gamma(t)) := \frac{d}{dt} f_w(\gamma(t)) - X_w f(\gamma(t))$ is measurable and bounded, as observed above. Let for any x the matrix $Y_x = [Y_{1,x}, \dots, Y_{q,x}] \in \mathbb{R}^{n\times q}$. Then let $Y_x^\dagger$ be the Moore–Penrose pseudoinverse of Y_x. Therefore, choose $b(t) = Y^\dagger_{\gamma(t)}(\operatorname{ad}_Z X_w)_{\gamma(t)}$ at any differentiability point t. Note that $b(t)$ is the least-norm solution of the system $\sum_{j=1}^q Y_{j,\gamma(t)}\xi^j = (\operatorname{ad}_Z X_w)_{\gamma(t)}$, where $\xi \in \mathbb{R}^q$. The Tychonoff approximation $Y^\dagger = \lim_{\delta\downarrow 0}(Y^T Y + \delta I_q)^{-1} Y^T$ (see the appendix) shows measurability. ☐

Remark 4.2.

- *One can prove Proposition 4.1 in a less elegant but more analytic way, without using the Moore–Penrose inverse, looking instead for "almost least-squares" solutions.*
- *The argument above can be used to see that in the definition of subunit distance we may work with paths γ such that for a.e. t we have $\dot\gamma(t) = \sum_j b^j(t) X_j(\gamma(t))$, where the function $t \mapsto b(t)$ is* measurable. *Indeed, let γ be a subunit path as in the definition of d_{cc} in (2.2). Given a differentiability point t of γ, let $b(t) := \lim_{\delta\downarrow 0}\big(X^T_{\gamma(t)} X_{\gamma(t)} + \delta I_p\big)^{-1} X^T_{\gamma(t)}\dot\gamma(t)$, where $X_x := [X_{1,x}, \dots, X_{m,x}]$ for all x. The function b is measurable and at any differentiability point t of γ, the vector $b(t)$ is the least-norm solution of the system $X_{\gamma(t)}\xi = \dot\gamma(t)$, with $\xi \in \mathbb{R}^m$. See [JSC87] for a related discussion.*

The distance associated with $\mathcal{P}$ where each Y_j has degree ℓ_j will be denoted by ϱ:

$$\begin{aligned}\varrho(x,y) := \inf\big\{ & r \ge 0 : \text{there is } \gamma \in \mathrm{Lip}_{\mathrm{Euc}}((0,1),\mathbb{R}^n) \text{ with } \gamma(0) = x \\ & \gamma(1) = y \text{ and } \dot\gamma(t) = \textstyle\sum_{j=1}^q b_j r^{\ell_j} Y_j(\gamma(t)) \text{ with } |b| \le 1 \text{ for a.e. } t \in [0,1]\big\}.\end{aligned} \tag{4.5}$$

Next we recall the definition of approximate exponential. Let $w_1, \dots, w_\ell \in \{1, \dots, m\}$. Given $\tau > 0$, we define, as in [NSW85, Mor00] and [MM12b],

$$\begin{aligned} C_\tau(X_{w_1}) &:= \exp(\tau X_{w_1}), \\ C_\tau(X_{w_1}, X_{w_2}) &:= \exp(-\tau X_{w_2})\exp(-\tau X_{w_1})\exp(\tau X_{w_2})\exp(\tau X_{w_1}), \\ &\ \vdots \\ C_\tau(X_{w_1}, \dots, X_{w_\ell}) &:= C_\tau(X_{w_2}, \dots, X_{w_\ell})^{-1}\exp(-\tau X_{w_1}) C_\tau(X_{w_2}, \dots, X_{w_\ell})\exp(\tau X_{w_1}). \end{aligned} \tag{4.6}$$

Then let

$$\mathrm{e}_{\mathrm{ap}}^{tX_{w_1 w_2\dots w_\ell}} := \exp_{\mathrm{ap}}(tX_{w_1 w_2 \dots w_\ell}) := \begin{cases} C_{t^{1/\ell}}(X_{w_1}, \dots, X_{w_\ell}), & \text{if } t \ge 0, \\ C_{|t|^{1/\ell}}(X_{w_1}, \dots, X_{w_\ell})^{-1}, & \text{if } t < 0. \end{cases} \tag{4.7}$$

Let Ω_0 be the open bounded set fixed before (2.8). By standard ODE theory, there is t_0 depending on ℓ, Ω, Ω_0, $\sup|f_j|$ and $\sup|\nabla f_j|$ such that $\exp_*(tX_{w_1 w_2 \dots w_\ell})x \in \Omega_0$ for any $x \in \Omega$ and $|t| \le t_0$. Given $r > 0$, define $\widetilde{Y}_j = r^{\ell_j} Y_j$ for $j = 1, \dots, q$. Moreover, if $I = (i_1, \dots, i_n) \in \{1, \dots, q\}^n$, $x \in \Omega$, $r \in (0,1]$ and $h \in \mathbb{R}^n$ is sufficiently close to the origin, define

$$\begin{aligned} E_{I,x,r}(h) &:= \exp_{\mathrm{ap}}(h_1 \widetilde{Y}_{i_1}) \cdots \exp_{\mathrm{ap}}(h_n \widetilde{Y}_{i_n})(x) \\ \|h\|_I &:= \max_{j=1,\dots,n} |h_j|^{1/\ell_{i_j}} \qquad Q_I(r) := \{h \in \mathbb{R}^n : \|h\|_I < r\}. \end{aligned} \tag{4.8}$$

Recall that, given $\eta \in (0,1)$, $x \in K$, $r < r_0$ and $I \in \{1,\dots,q\}^n$, the triple (I,x,r) is said to be η-maximal if $|\lambda_I(x)| r^{\ell(I)} > \eta \max_{J\in\mathcal{I}(p_x,q)} |\lambda_J(x)| r^{\ell(J)}$.

Theorem 4.3. *Let $\mathcal{H}$ be a family of vector fields of class $C^{s-1,1}_{\mathcal{H},\mathrm{loc}} \cap C^1_{\mathrm{Euc}}$ satisfying the Hörmander condition of step s. Assume that all nested commutators up to length s are continuous in the Euclidean sense. Then there is $C > 1$ such that the following properties hold. Let $I \in \{1,\dots,q\}^n$, $x \in \Omega$ and $r < C^{-1}$. Let also $E := E_{I,x,r}$ be the map in (4.8). Then*

(a) $E \in C^1_{\mathrm{Euc}}(Q_I(C^{-1}))$.
(b) *We have the expansion*

$$\frac{\partial}{\partial h_k} E(h) = \widetilde{Y}_{i_k}(E(h)) + \sum_{\ell_j=\ell_{i_k}+1}^{s} a_k^j(h)\widetilde{Y}_j(E(h)) + \sum_{i=1}^{q} \omega_k^i(x,h)\widetilde{Y}_i(E(h)). \tag{4.9}$$

where $\widetilde{Y}_k := r^{\ell_k} Y_k$ and the functions a_k^j and ω_k^j satisfy

$$|a_k^j(h)| \le C\|h\|_I^{\ell_j-\ell_{i_k}} \quad \textit{for all } h \in Q_I(C^{-1}) \tag{4.10}$$

$$|\omega_k^j(x,h)| \le C\|h\|_I^{s+1-\ell_{i_k}} \quad \textit{for all } h \in Q_I(C^{-1}) \quad x \in \Omega. \tag{4.11}$$

(c) *If moreover (I,x,r) is $\frac{1}{2}$-maximal with $I \in \{1,\dots,q\}^n$, $x \in \Omega$ and $r < r_0$, then, for all $\varepsilon \le C^{-1}$ we have*

$$E_{I,x,r}(Q_I(\varepsilon)) \supset B_\rho(x, C^{-1}\varepsilon^s r). \tag{4.12}$$

Note that constants in Theorem 4.3 depend quantitatively on C_0 and L_0. Inclusion (4.12) ensures the Fefferman–Phong type estimate $d(x,y) \le C|x-y|^{1/s}$; see [FP83].
Moreover, we have

Theorem 4.4. *Assume that the hypotheses of Theorem 4.3 hold. Then there is is a constant $C > 0$ such that the following holds. Let $x \in \Omega \Subset \Omega_0$. Then, for any $\frac{1}{2}$-maximal triple (I,x,r) with $I \in \{1,\dots,q\}^n$, $x \in \Omega$ and $r < C^{-1}$, the map $E_{I,x,r}$ is one-to-one on the set $Q_I(C^{-1})$.*

The constant C in Theorem 4.4 does not depend quantitatively on C_0 and L_0, because (vi) below involves a qualitative covering argument. A more precise control on such constant can be obtained assuming more regularity (for instance if the vector fields belong to the class $\mathcal{B}_s$ of [MM11]).

Proof of Theorems 4.3 and 4.4. All arguments of the proofs are contained in the papers [NSW85, Mor00, MM12b, MM12a] and [MM11]. Let us recapitulate the skeleton of the proof with precise references to the mentioned papers.

(i) Specializing [MM12a, Remark 3.3] to our setting, we may claim that if (I,x,r) is η-maximal, then (I,y,r) is $C^{-1}\eta$-maximal for all $y \in B_d(x, C^{-1}\eta r)$.
(ii) The proof of Theorem 4.3, items (a) and (b) are contained in [MM12a, Theorem 3.11]. Note that the mentioned result holds even in a more general setting where the Hörmander's rank condition is not assumed.

(iii) In view of (i), (ii) and expansion (4.9) we can follow the proof of [MM12b, Lemma 5.14] (just letting $\sigma = 0$). Thus, we may claim that if $\xi \in \Omega$ and $|\lambda_I(\xi)| \neq 0$, then $E_{I,\xi,r}$ is one-to-one on $Q_I(C^{-1} r^{\ell(I)} |\lambda_I(\xi)|)$.

(iv) For all $\eta \in (0,1)$ there is $C_\eta > 0$ such that given an η-maximal triple (I, x, r), then the map $E_{I,x,r}$ satisfies for all $j \in \{1, \dots, n\}$ the expansion

$$\frac{\partial}{\partial h_j} E(h) = \widetilde{Y}_{i_j}(E(h)) + \sum_{1\le k\le n} \chi_j^k(h) \widetilde{Y}_{i_k}(E(h)) \quad \text{for all } h \in Q_I(C_\eta^{-1}), \tag{4.13}$$

where $\chi \in C^0_{\mathrm{Euc}}(Q_I(C_\eta^{-1}), \mathbb{R}^{n\times n})$ satisfies

$$|\chi(h)| \le C_\eta \, \|h\|_I \qquad \text{if } \|h\|_I \le C_\eta^{-1}. \tag{4.14}$$

Therefore, for a suitable $\widetilde{C}_\eta$ possibly larger that C_η, the map $E_{I,x,r}\big|_{Q_I(\widetilde{C}_\eta^{-1})}$ is a local C^1 diffeomorphism and in particular it is open. This ensures that the topologies of the distances ϱ, d_{cc} and d are all locally equivalent to the Euclidean one. Expansion (4.13) with estimate (4.14) has been proved in [MM11, Theorem 3.1]. As observed after the statement in [MM11], such result holds in the broader class $\mathcal{A}_s$. Note that in [MM11] we discuss the case $\eta = \frac{1}{2}$. The case with $\eta < \frac{1}{2}$ can be treated with minor modifications.

(v) To prove Theorem 4.3-(c), it suffices to follow the proof of [MM11, Lemma 3.7]. This is explained in [MM11, Remark 3.8].

(vi) Finally, keeping all previous items into account, to prove the injectivity result Theorem 4.4, it suffices to follow [NSW85, pp. 132–133] or [Mor00, Lemma 3.6]. In the proof of the latter lemma, note that in third line of [Mor00, Eq.(30)], which reads

$$|\lambda_{I_{0,k}}(x)| \delta_{0,k}^{d(I_{0,k})} > \frac{1}{2} \max_J |\lambda_J(x)| \delta_{0,k}^{d(J)} \quad \text{for all } x \in U_k,$$

by (i) we may choose $U_k = B_d(x_k, C^{-1}\delta_{0,k})$, which is open by (iv); moreover, by (iii) we may assume that $E_{I,x,\delta_{0,k}}\big|_{Q_I(\delta_{0,k})}$ is one-to-one for each $x \in U_k$. The remaining part of the proof in [Mor00] can be applied verbatim to our setting.

□

Remark 4.5. *Theorem 4.4 implies the* doubling property *for vector fields satisfying the hypotheses of Theorem 4.3. Let $\Omega \subset \mathbb{R}^n$ be a bounded open set. Then there are C and $r_0 > 0$ so that*

$$|B_{\mathrm{cc}}(x, 2r)| \le C |B_{\mathrm{cc}}(x, r)| \quad \textit{for all } x \in \Omega \ r < r_0.$$

Moreover, following [LM00], one gets for all $f \in C^1(B_{\mathrm{cc}}(x, Cr))$, the Poincaré inequality

$$\int_{B_{\mathrm{cc}}(x,r)} |f(y) - f_B| dy \le Cr \int_{B_{\mathrm{cc}}(x,Cr)} \sum_j |X_j f(y)| dy \quad \textit{for all } x \in \Omega \ r < r_0.$$

Finally, as in [MM12b, Proposition 6.2], given $\Omega' \subset\subset \Omega$, and $\varepsilon \in]0, 1/s[$, there is r_0 and $C > 0$ such that, for any $f \in C^1(\Omega)$,

$$\int_{\substack{\Omega'\times\Omega' \\ d(x,y)\le r_0}} \frac{|f(x) - f(y)|^2}{|x-y|^{n+2\varepsilon}} dx dy \le C \int_\Omega \sum_j |X_j f(y)|^2 dy. \tag{4.15}$$

A. Appendix

Tychonoff regularization for the Moore–Penrose pseudoinverse. Here we discuss an approximation formula for the Moore–Penrose inverse of a matrix which has been used in Proposition 4.1. This result is proposed as an exercise in some matrix-analysis textbooks (see [GVL89, Problem 5.5.2]). We include here a short discussion for completeness.

Let $a_1, \dots, a_q \in \mathbb{R}^n$ and let $A = [a_1, \dots, a_q] \in \mathbb{R}^{n\times q}$. Take $b \in \operatorname{span}\{a_1, \dots, a_q\}$ and look at the system $Ax = b$ where $x \in \mathbb{R}^q$. We do not assume that the vectors a_j are independent. Let x_{LS} be the solution of minimal norm. We claim that

$$x_{\mathrm{LS}} = \lim_{\lambda\to 0}(A^TA + \lambda^2 I_q)^{-1}A^T b. \tag{A.1}$$

In other words, the family if matrices $(A^TA + \lambda^2 I_q)^{-1}A^T$ gives an approximation of the Moore–Penrose inverse $A^\dagger$, as $\lambda \to 0$. Note that, if $a_1, \dots, a_q$ are independent, then it is well known that $A^\dagger = (A^TA)^{-1}A^T$. If they are dependent, then A^TA is singular, but still we have $\lim_{\lambda\to 0}(A^TA + \lambda^2 I_q)^{-1}A^T = A^\dagger$.

To show (A.1), write $A = U\Sigma V^T$ as a singular value decomposition, i.e. $U \in O(n)$, $V \in O(q)$, while $\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0) \in \mathbb{R}^{n\times q}$, where $\sigma_1 \geq \dots \geq \sigma_r > 0$ are the singular values of A and $r \leq \min\{q, n\}$ is its rank. Note that $U^T[a_1, \dots, a_q] = \Sigma V^T$. Therefore, $U^Ta_j \in \mathbb{R}^r \times \{0_{n-r}\}$ and $U^Tb \in \mathbb{R}^r \times \{0_{n-r}\}$, too.

By definition, the vector $x \in \mathbb{R}^q$ is a (not unique) least-square solution of the system $Ax = b$ if and only if it solves $A^TAx = A^Tb$, which is equivalent to $\Sigma^T\Sigma V^Tx = \Sigma^TU^Tb$, or, letting $V^Tx =: \xi$ and $U^Tb =: \beta \in \mathbb{R}^n$, to the system

$$\Sigma^T\Sigma\xi = \Sigma^T\beta. \tag{A.2}$$

Since $\Sigma^T\Sigma = \operatorname{diag}(\sigma_1^2, \dots, \sigma_r^2, 0, \dots, 0) \in \mathbb{R}^{q\times q}$ and since the system (A.2) has solutions by assumptions on the data b, it must be $\Sigma^T\beta = (\sigma_1\beta_1, \dots, \sigma_r\beta_r, 0, \dots)^T \in \mathbb{R}^q$ and the solutions of (A.2) are $\xi = (\beta_1/\sigma_1, \dots, \beta_r/\sigma_r, \xi_{r+1}, \dots, \xi_q)^T$, with $\xi_{r+1}, \dots, \xi_q$ free parameters. Clearly, the minimal-norm one is $\xi_{\mathrm{LS}} = (\beta_1/\sigma_1, \cdots, \beta_r/\sigma_r, 0, \dots)^T \in \mathbb{R}^q$.

Define now the vector $x_\lambda := (A^TA + \lambda^2 I_q)^{-1}A^Tb = V(\Sigma^T\Sigma + \lambda^2 I_q)^{-1}\Sigma^TU^Tb$. Since $\Sigma^TU^Tb = \Sigma^T\beta = (\sigma_1\beta_1, \dots, \sigma_r\beta_r, 0, \dots)^T$, we have

$$V^Tx_\lambda =: \xi_\lambda = (\sigma_1\beta_1/(\sigma_1^2 + \lambda^2), \dots, \sigma_r\beta_r/(\sigma_r^2 + \lambda^2), 0, \dots)^T \in \mathbb{R}^q.$$

Thus, as $\lambda \to 0$,

$$|x_{\mathrm{LS}} - x_\lambda| = |\xi_{\mathrm{LS}} - \xi_\lambda| = \Big|\Big(\frac{\lambda^2\beta_1}{\sigma_1(\sigma_1^2 + \lambda^2)}, \dots, \frac{\lambda^2\beta_r}{\sigma_r(\sigma_r^2 + \lambda^2)}\Big)\Big| \longrightarrow 0.$$

This concludes the proof of (A.1).

References

[AL50] A. S. Amitsur and J. Levitzki, *Minimal identities for algebras*, Proc. Amer. Math. Soc. **1** (1950), 449–463. MR 0036751 (12,155d)

[ASCV06] Luigi Ambrosio, Francesco Serra Cassano, and Davide Vittone, *Intrinsic regular hypersurfaces in Heisenberg groups*, J. Geom. Anal. **16** (2006), no. 2, 187–232. MR 2223801 (2007g:49072)

[Aub01] Thierry Aubin, *A course in differential geometry*, Graduate Studies in Mathematics, vol. 27, American Mathematical Society, Providence, RI, 2001. MR 1799532 (2001m:53001)

[BBP12a] M. Bramanti, L. Brandolini, and M. Pedroni, *Basic properties of nonsmooth Hormander's vector fields and Poincare's inequality*, ArXiv e-prints, `http://arxiv.org/abs/0809.2872`. Forum Math. (to appear) (2012).

[BBP12b] ———, *On the lifting and approximation theorem for nonsmooth vector fields*, ArXiv e-prints; Indiana Univ. Math. J. (to appear) (2012).

[Che99] J. Cheeger, *Differentiability of Lipschitz functions on metric measure spaces*, Geom. Funct. Anal. **9** (1999), no. 3, 428–517. MR 1708448 (2000g:53043)

[CLM02] G. Citti, E. Lanconelli, and A. Montanari, *Smoothness of Lipschitz-continuous graphs with nonvanishing Levi curvature*, Acta Math. **188** (2002), no. 1, 87–128. MR 1947459 (2003m:35035)

[FL83a] B. Franchi and E. Lanconelli, *Une métrique associée à une classe d'opérateurs elliptiques dégénérés*, Rend. Sem. Mat. Univ. Politec. Torino (1983), no. Special Issue, 105–114 (1984), Conference on linear partial and pseudodifferential operators (Torino, 1982).

[FL83b] Bruno Franchi and Ermanno Lanconelli, *Hölder regularity theorem for a class of linear nonuniformly elliptic operators with measurable coefficients*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **10** (1983), no. 4, 523–541. MR 753153 (85k:35094)

[FP83] C. Fefferman and D. H. Phong, *Subelliptic eigenvalue problems*, Conference on harmonic analysis in honor of Antoni Zygmund, Vol. I, II (Chicago, Ill., 1981), Wadsworth Math. Ser., Wadsworth, Belmont, CA, 1983, pp. 590–606. MR 730094 (86c:35112)

[GN96] Nicola Garofalo and Duy-Minh Nhieu, *Isoperimetric and Sobolev inequalities for Carnot-Carathéodory spaces and the existence of minimal surfaces*, Comm. Pure Appl. Math. **49** (1996), no. 10, 1081–1144. MR 1404326 (97i:58032)

[Gre10] A. V. Greshnov, *On a class of Lipschitz vector fields in* $\mathbb{R}^3$, Sibirsk. Mat. Zh. **51** (2010), no. 3, 517–527. MR 2683094 (2011g:53059)

[GVL89] Gene H. Golub and Charles F. Van Loan, *Matrix computations*, second ed., Johns Hopkins Series in the Mathematical Sciences, vol. 3, Johns Hopkins University Press, Baltimore, MD, 1989. MR 1002570 (90d:65055)

[Har02] Philip Hartman, *Ordinary differential equations*, Classics in Applied Mathematics, vol. 38, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002, Corrected reprint of the second (1982) edition [Birkhäuser, Boston, MA; MR0658490 (83e:34002)], With a foreword by Peter Bates. MR 1929104 (2003h:34001)

[HK00] Piotr Hajłasz and Pekka Koskela, *Sobolev met Poincaré*, Mem. Amer. Math. Soc. **145** (2000), no. 688, x+101. MR 1683160 (2000j:46063)

[Jer86] David Jerison, *The Poincaré inequality for vector fields satisfying Hörmander's condition*, Duke Math. J. **53** (1986), no. 2, 503–523. MR 850547 (87i:35027)

[JSC87] David Jerison and Antonio Sánchez-Calle, *Subelliptic, second order differential operators*, Complex analysis, III (College Park, Md., 1985–86), Lecture Notes in Math., vol. 1277, Springer, Berlin, 1987, pp. 46–77. MR 922334 (89b:35021)

[KN96] Shoshichi Kobayashi and Katsumi Nomizu, *Foundations of differential geometry. Vol. I*, Wiley Classics Library, John Wiley & Sons Inc., New York, 1996, Reprint of the 1963 original, A Wiley-Interscience Publication. MR 1393940 (97c:53001a)

[KV09] Maria Karmanova and Sergey Vodop'yanov, *Geometry of Carnot-Carathéodory spaces, differentiability, coarea and area formulas*, Analysis and mathematical physics, Trends Math., Birkhäuser, Basel, 2009, pp. 233–335. MR 2724617

[Lev50] J. Levitzki, *A theorem on polynomial identities*, Proc. Amer. Math. Soc. **1** (1950), 334–341. MR 0035758 (12,6c)

[LM00] Ermanno Lanconelli and Daniele Morbidelli, *On the Poincaré inequality for vector fields*, Ark. Mat. **38** (2000), no. 2, 327–342.

[Man10] Maria Manfredini, *A note on the Poincaré inequality for Lipschitz vector fields of step two*, Proc. Amer. Math. Soc. **138** (2010), no. 2, 567–575. MR 2557174 (2010j:35011)

[MM11] A. Montanari and D. Morbidelli, *Step-s involutive families of vector fields, their orbits and the Poincaré inequality*, ArXiv e-prints (2011).

[MM12a] ______, *Almost exponential maps and integrability results for a class of horizontally regular vector fields*, ArXiv e-prints (2012).

[MM12b] ______, *Nonsmooth Hörmander vector fields and their control balls*, Trans. Amer. Math. Soc (to appear). (2012).

[Mor00] Daniele Morbidelli, *Fractional Sobolev norms and structure of Carnot-Carathéodory balls for Hörmander vector fields*, Studia Math. **139** (2000), no. 3, 213–244. MR 1762582 (2002a:46039)

[NSW85] Alexander Nagel, Elias M. Stein, and Stephen Wainger, *Balls and metrics defined by vector fields. I. Basic properties*, Acta Math. **155** (1985), no. 1-2, 103–147. MR 793239 (86k:46049)

[Ote91] J. A. Oteo, *The Baker-Campbell-Hausdorff formula and nested commutator identities*, J. Math. Phys. **32** (1991), no. 2, 419–424. MR 1088363 (92b:17055)

[RS07] Franco Rampazzo and Héctor J. Sussmann, *Commutators of flow maps of nonsmooth vector fields*, J. Differential Equations **232** (2007), 134–175.

[SC92] L. Saloff-Coste, *A note on Poincaré, Sobolev, and Harnack inequalities*, Internat. Math. Res. Notices (1992), no. 2, 27–38. MR 1150597 (93d:58158)

[SW06] Eric T. Sawyer and Richard L. Wheeden, *Hölder continuity of weak solutions to subelliptic equations with rough coefficients*, Mem. Amer. Math. Soc. **180** (2006), no. 847, x+157. MR 2204824 (2007f:35037)

Annamaria Montanari, Daniele Morbidelli
Dipartimento di Matematica, Università di Bologna (Italy)
Email: annamaria.montanari@unibo.it, daniele.morbidelli@unibo.it