TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 364, Number 5, May 2012, Pages 2339–2375
S 0002-9947(2011)05395-X
Article electronically published on December 15, 2011

NONSMOOTH HÖRMANDER VECTOR FIELDS AND THEIR CONTROL BALLS

ANNAMARIA MONTANARI AND DANIELE MORBIDELLI

Abstract. We prove a ball-box theorem for nonsmooth Hörmander vector fields of step $s \geq 2$.

Contents

1. Introduction

In this paper we give a self-contained proof of a ball-box theorem for a family $\{X_1, \dots, X_m\}$ of nonsmooth vector fields satisfying the Hörmander condition. This is the third paper, after [M] and [MM], where we investigate ideas of the classical article by Nagel Stein and Wainger [NSW].

Our purpose is to prove a ball-box theorem using only elementary analysis techniques and at the same time to relax as much as possible the regularity assumptions on the vector fields. Roughly speaking, our results hold as soon as the commutators

Received by the editors December 15, 2008 and, in revised form, March 5, 2010.
2010 *Mathematics Subject Classification.* Primary 53C17; Secondary 35R03.

involved in the Hörmander condition are Lipschitz continuous. Moreover, our proof does not rely on algebraic tools, such as formal series and the Campbell–Hausdorff formula.

To describe our work, we recall the basic ideas of [NSW]. Notation and language are more precisely described in Section 2. Any control ball $B(x,r)$ associated with a family $\{X_1,\dots,X_m\}$ of Hörmander vector fields in $\mathbb{R}^n$ satisfies, for x belonging to some compact set K and small radius $r<r_0$, the double inclusion

$$\Phi_x(Q(C^{-1}r))\subset B(x,r)\subset \Phi_x(Q(Cr)). \tag{1.1}$$

Here, the map Φ_x is an exponential of the form

$$\Phi_x(h)=\exp(h_1U_1+\cdots+h_nU_n)(x), \tag{1.2}$$

where the vector fields $U_1,\dots,U_n$ are suitable commutators of lengths $d_1,\dots,d_n$ and $Q(r)=\{h\in\mathbb{R}^n:\max_j|h_j|^{1/d_j}<r\}$. Usually, (1.1) is referred to as a *ball-box* inclusion. A control on the Jacobian matrix of Φ_x gives an estimate of the measure of the ball and ultimately provides the doubling property.

A remarkable achievement in [NSW] concerns the choice of the vector fields U_j which guarantee inclusions (1.1) for a given control ball $B(x,r)$; see also the discussion in [Ste, p. 440]. Enumerate as $Y_1,\dots,Y_q$ all commutators of length at most s and let ℓ_i be the length of Y_i. If the Hörmander condition of step s is fulfilled, then the vector fields Y_i span $\mathbb{R}^n$ at any point. Given a multi-index $I=(i_1,\dots,i_n)\in\{1,\dots,q\}^n=:\mathcal{S}$ and its corresponding $n-$tuple $Y_{i_1},\dots,Y_{i_n}$ of commutators, let

$$\lambda_I(x)=\det(Y_{i_1},\dots,Y_{i_n})(x)\quad\text{and}\quad \ell(I)=\ell_{i_1}+\cdots+\ell_{i_n}. \tag{1.3}$$

In [NSW], the authors prove the following fact: given a ball $B(x,r)$, inclusion (1.1) holds with $U_1=Y_{i_1},\dots,U_n=Y_{i_n}$ if the $n-$tuple $I\in\mathcal{S}$ satisfies the *η-maximality* condition

$$|\lambda_I(x)|r^{\ell(I)}>\eta\max_{K\in\mathcal{S}}|\lambda_K(x)|r^{\ell(K)}, \tag{1.4}$$

where $\eta\in(0,1)$ is greater than some absolute constant. Although the choice of the n-tuple I may depend on both the point and the radius, the constant C is uniform in $x\in K$ and $r\in(0,r_0)$.

In [M] the second author proved that (1.1) also holds if one changes the map Φ_x with the *almost exponential* map

$$E_x(h)=\exp_*(h_1U_1)\circ\cdots\circ\exp_*(h_nU_n)(x), \tag{1.5}$$

where $h_j\mapsto\exp_*(h_jU_j)$ is the *approximate exponential* of the commutator U_j, whose main feature is that it can be factorized as a suitable composition of exponentials of the original vector fields $X_1,\dots,X_m$. See (2.3) for the definition of $\exp_*$. Lanconelli and the second author in [LM] proved that, if inclusion (1.5), with pertinent estimates for the Jacobian of E_x are known, then the Poincaré inequality follows (see [J] for the original proof). It is worth observing now that all the results in [NSW] and [M] are proved for C^M vector fields, where M is much larger than the step s. This can be seen by carefully reading the proofs of Lemmas 2.10 and 2.13 in [NSW].

In [TW, Section 4], Tao and Wright gave a new proof of the ball-box theorem with a different approach, based on Gronwall's inequality. The authors in [TW] use scaling maps of the form $\Phi_{x,r}(t):=\exp(t_1r^{d_1}U_1+\cdots+t_nr^{d_n}U_n)x$, which are naturally defined on a box $|t| \le \varepsilon_0$, where $\varepsilon_0 > 0$ is a small constant independent of x and r; see the discussion in Subsection 5.2. The arguments in [TW] do not rely on the Campbell–Hausdorff formula.[1] Moreover, although the statement is phrased for C^∞ vector fields, one can see that their results hold under the assumption that the vector fields have a C^M smoothness, with $M = 2s$ for vector fields of step s. See Remark 5.10 for a more detailed discussion.

In [MM] we started to work in low regularity hypotheses and we obtained a ball-box theorem and the Poincaré inequality for Lipschitz continuous vector fields of step two with Lipschitz continuous commutators. We used the maps (1.5), but several aspects of the work [MM] are peculiar of the step two situation and until now it was not clear how to generalize those results to higher step vector fields.

Recently, Bramanti, Brandolini and Pedroni [BBP] have proved a doubling property and the Poincaré inequality for nonsmooth Hörmander vector fields with an algebraic method. Informally speaking, they truncate the Taylor series of the coefficients of the vector fields and then they apply to the polynomial approximations the results in [NSW, LM] and [M]. The paper [BBP] also involves a study of the almost exponential maps in (1.5). The results in [BBP] and in the present paper were obtained independently and simultaneously.

In this paper we complete the result in [MM], namely we prove a ball-box theorem for general vector fields of arbitrary step s, requiring basically that all the commutators involved in the Hörmander condition are Lipschitz continuous. Our precise hypotheses are stated in Definition 2.1. We improve all previous results in terms of regularity; see Remark 5.10. As in [MM], we use the almost exponential maps (1.5), but we need to provide a very detailed study of such functions in the higher step case.

The scheme of the proof of our theorem is basically the Nagel, Stein and Wainger one, but there are some new tools that should be emphasized. Namely, we obtain some noncommutative calculus formulas developed in order to show that, given a commutator Y, the derivative $\frac{d}{dt}\exp_*(tY)$ can be precisely written as a finite sum of higher order commutators plus an *integral remainder*. This is done in Section 3. These results are applied in Section 5 to the almost exponential maps E in (1.5). Our main structure theorem is Theorem 5.8. As in [MM], part of our computations will be given for smooth vector fields, namely the standard Euclidean regularization X_j^σ of the vector fields X_j. We will keep everywhere under control all constants in order to make sure that they are stable as σ goes to 0.

It is well known (see [LM, MM]) that the doubling property and the Poincaré inequality follow immediately from Theorem 5.8. Also observe that our ball-box theorem can be useful in all situations where integrals of the form $\int |f(x) - f(y)| w(x, y) dx dy$ need to be estimated, for some weight w. See for example [M] or [MoM]. As an application, in Proposition 6.2 we prove a subelliptic Hörmander–type estimate for nonsmooth vector fields. We believe that the results in Section 3 may be useful in other, related situations.

Concerning the machinary developed in Section 3, it is worth mentioning the papers [RaS, RaS2], where noncommutative calculus formulas are used in the proof of a nonsmooth version of Chow's Theorem for vector fields of step two.

Geometric analysis for nonsmooth vector fields started in the 1980s with the papers by Franchi and Lanconelli [FL1, FL2], who proved the Poincaré inequality

[1]The methods of [TW] have been further exploited in a very recent paper by Street [Str].

for diagonal vector fields in $\mathbb{R}^n$ of the form $X_j = \lambda_j(x)\partial_j$, $j = 1, \dots, n$. In the diagonal case completely different techniques are available. In a recent paper by Sawyer and Wheeden [SW], which probably contains the best results to date on diagonal vector fields, the reader can find a rich bibliography on the subject.

Plan of the paper. In Section 2 we introduce notation. In Section 3 we prove our noncommutative calculus formulas and in Section 4 we prove a stability property of the "almost-maximality" condition (1.4). These tools are applied in Subsection 5.1 to the maps E. In Subsection 5.2 we briefly discuss the "scaled version" of our maps E. Subsection 5.3 contains the ball-box theorem. In Section 6 we show some examples. Finally, Section 7 contains the smooth approximation result for the original vector fields.

2. Preliminaries and notation

We consider vector fields $X_1, \dots, X_m$ in $\mathbb{R}^n$. For any $\ell \in \mathbb{N}$ we define a *word* $w = j_1 \cdots j_\ell$ to be any finite ordered collection of ℓ letters, $j_k \in \{1, \dots, m\}$, and we introduce the notation $X_w = [X_{j_1}, \dots, [X_{j_{\ell-1}}, X_{j_\ell}]]$ for commutators. Let $|w| := \ell$ be the *length* of X_w. We assume the Hörmander condition of step s, i.e. that $\{X_w(x) : |w| \le s\}$ generate all $\mathbb{R}^n$ at any point $x \in \mathbb{R}^n$. Sometimes it will be useful to have a different notation between a vector field in $\mathbb{R}^n$ and its associated vector function. In these situations we will write $X_w = f_w \cdot \nabla = \sum_{k=1}^n f_w^k \partial_k$. We will also enumerate as $Y_1, \dots, Y_q$ all the commutators X_w with length $|w| \le s$ and denote by ℓ_i or $\ell(Y_i)$ the length of Y_i . We identify an ordered n–tuple of commutators $Y_{i_1}, \dots, Y_{i_n}$ by the index $I = (i_1, \dots, i_n) \in \mathcal{S} := \{1, \dots, q\}^n$.

For $x, y \in \mathbb{R}^n$, denote by $d(x, y)$ the control distance, that is, the infimum of the $r > 0$ such that there is a Lipschitz path $\gamma : [0, 1] \to \mathbb{R}^n$ with $\gamma(0) = x$, $\gamma(1) = y$ and $\dot\gamma = \sum_{j=1}^m b_j X_j(\gamma)$, for a.e. $t \in [0, 1]$. The measurable functions b_j must satisfy $|b_j(t)| \le r$ for almost any t. Corresponding balls will be indicated as $B(x, r)$.

Denote also by $\varrho(x, y)$ the infimum of the $r > 0$ such that there is a Lipschitz continuous path $\gamma : [0, 1] \to \mathbb{R}^n$ with $\gamma(0) = x$, $\gamma(1) = y$ and γ satisfies for a.e. $t \in [0, 1]$, $\dot\gamma = \sum_{i=1}^q c_i Y_i(\gamma)$ for suitable measurable functions c_j with $|c_j(t)| \le r^{\ell(Y_j)}$. Corresponding balls will be denoted by $B_\varrho(x, r)$. The definition of ϱ is meaningful as soon as the vector fields Y_j are at least continuous.

Definition 2.1 (Vector fields of class $\mathcal{A}_s$)**.** Let $X_1, \dots, X_m$ be vector fields in $\mathbb{R}^n$ and let $s \ge 2$. We say that the vector fields X_j are of class $\mathcal{A}_s$ if they are of class $C^{s-2,1}_{\mathrm{loc}}(\mathbb{R}^n)$ and for any word w with $|w| = s - 1$, and for every $j, k \in \{1, \dots, m\}$,

(1) the derivative $X_k f_w$ exists and is continuous;

(2) the distributional derivative $X_j(X_k f_w)$ exists and

$$X_j(X_k f_w) \in L^\infty_{\mathrm{loc}}(\mathbb{R}^n). \tag{2.1}$$

Recall that $X_j \in C^{s-2,1}_{\mathrm{loc}}$ means that all the Euclidean derivatives of order at most $s-2$ of the functions $f_1, \dots, f_m$ are locally Lipschitz continuous. In particular, all the commutators X_w with $|w| \le s - 1$ are locally Lipschitz continuous in the Euclidean sense and by item (1) all commutators X_w of length $|w| = s$ are pointwise defined. If we knew that d defined the Euclidean topology, condition (2) would be equivalent to the fact that X_w is locally d-Lipschitz if $|w| = s$; see [GN, FSSC].

Let $\{X_1, \dots, X_m\}$ be in the class $\mathcal{A}_s$ and assume that they satisfy the Hörmander condition of step s. Fix once for all a pair of bounded connected open sets $\Omega' \Subset \Omega$

and denote $K = \overline{\Omega'}$. We denote by D Euclidean derivatives. If $D = \partial_{j_1} \cdots \partial_{j_p}$ for some $j_1, \dots, j_p \in \{1, \dots, n\}$, then $|D| := p$ indicates the order of D. It is understood that a derivative of order 0 is the identity. Introduce the positive constant

$$
\begin{aligned}
L := \max_{\substack{1\le j\le m\\ 0\le |D|\le s-2}} \sup_{\Omega} |Df_j| + \max_{\substack{j=1,\dots,m\\ |D|=s-1}} \operatorname*{ess\,sup}_{\Omega} |Df_j| \\
+ \max_{\substack{k,j=1,\dots,m,\\ |w|=s-1}} \operatorname*{ess\,sup}_{\Omega} |X_k X_j f_w|.
\end{aligned}
\tag{2.2}
$$

Remark 2.2. We will prove in Section 5 a ball-box theorem for vector fields of step s in the class $\mathcal{A}_s$. This improves both the results in [TW] and [BBP] in terms of regularity. Indeed, in [TW] a C^M regularity with $M = 2s$ must be assumed (see Remark 5.10). In [BBP] the authors assume that the vector fields belong to the Euclidean Lipschitz space $C^{s-1,1}_{\mathrm{loc}}(\mathbb{R}^n)$, which requires the boundedness on the Euclidean gradient ∇f_w of any commutator f_w of length s, while we only need to control the "horizontal" gradient of f_w.

Approximate commutators. For vector fields $X_{j_1}, \dots, X_{j_\ell}$, and for $\tau > 0$, we define, as in [NSW], [M] and [MM],

$$
\begin{aligned}
C_\tau(X_{j_1}) &:= \exp(\tau X_{j_1}),\\
C_\tau(X_{j_1}, X_{j_2}) &:= \exp(-\tau X_{j_2}) \exp(-\tau X_{j_1}) \exp(\tau X_{j_2}) \exp(\tau X_{j_1}),\\
&\ \vdots\\
C_\tau(X_{j_1}, \dots, X_{j_\ell}) &:= C_\tau(X_{j_2}, \dots, X_{j_\ell})^{-1} \exp(-\tau X_{j_1}) C_\tau(X_{j_2}, \dots, X_{j_\ell}) \exp(\tau X_{j_1}).
\end{aligned}
$$

Then let

$$
e_*^{tX_{j_1 j_2\dots j_\ell}} := \exp_*(tX_{j_1 j_2\dots j_\ell}) := \begin{cases} C_{t^{1/\ell}}(X_{j_1}, \dots, X_{j_\ell}), & \text{if } t > 0,\\ C_{|t|^{1/\ell}}(X_{j_1}, \dots, X_{j_\ell})^{-1}, & \text{if } t < 0. \end{cases}
\tag{2.3}
$$

By standard ODE theory, there is t_0 depending on ℓ, K, Ω, $\sup |f_j|$ and $\operatorname{ess\,sup} |\nabla f_j|$ such that $\exp_*(tX_{j_1 j_2\dots j_\ell})x$ is well defined for any $x \in K$ and $|t| \le t_0$. The approximate commutators C_t are quite natural (indeed, they make an appearance in the original paper [NSW]). Assuming that the vector fields are smooth and using the Campbell–Hausdorff formula, we have the formal expansion

$$
C_\tau(X_{j_1}, \dots, X_{j_\ell}) = \exp\Big(\tau^\ell X_{j_1 j_2\dots j_\ell} + \sum_{k=\ell+1}^{\infty} \tau^k R_k\Big),
$$

where R_k denotes a linear combination of commutators of length k. See [NSW, Lemma 2.21]. A study of these maps in the smooth case based on this formula is carried out in [M].

Define, given $I = (i_1, \dots, i_n) \in \mathcal{S}$, $x \in K$ and $h \in \mathbb{R}^n$, with $|h| \le C^{-1}$,

$$
\begin{aligned}
&E_{I,x}(h) := E_I(x,h) := \exp_*(h_1 Y_{i_1}) \cdots \exp_*(h_n Y_{i_n})(x),\\
&\|h\|_I := \max_{j=1,\dots,n} |h_j|^{1/\ell_{i_j}}, \qquad Q_I(r) := \{h \in \mathbb{R}^n : \|h\|_I < r\},\\
&\Lambda(x,r) := \max_{K\in\mathcal{S}} |\lambda_K(x)|\, r^{\ell(K)},
\end{aligned}
\tag{2.4}
$$

where $\ell(K) = \ell_{k_1} + \dots + \ell_{k_n}$, the determinants λ_K are defined in (1.3), and we have

$$
\nu := \inf_{x\in\Omega} \Lambda(x,1) > 0.
\tag{2.5}
$$

The lower bound (2.5) will appear many times in the following sections. All the constants in our main theorem will depend on ν in (2.5) and on L in (2.2).

In order to refer to the crucial condition (1.4), we give the following definition.

Definition 2.3 ($\eta-$maximal triple)**.** Let $\eta \in]0,1[$, $I \in \mathcal{S}$, $x \in \mathbb{R}^n$ and $r > 0$. We say that (I, x, r) is $\eta-$maximal if we have $|\lambda_I(x)| r^{\ell(I)} > \eta \Lambda(x, r)$.

Regularized vector fields. Here we describe our procedure of the smoothing of the vector fields X_j of step s. For any function f, let $f^{(\sigma)}(x) = \int f(x - \sigma y)\varphi(y)dy$, where $\varphi \in C_0^\infty$ is a standard nonnegative averaging kernel supported in the unit ball. Define

$$
\begin{aligned}
&X_j^\sigma := \sum_{k=1}^n (f_j^k)^{(\sigma)} \partial_k \quad \text{and} \\
&X_{j_1 \dots j_\ell}^\sigma := [X_{j_1}^\sigma, \dots, [X_{j_{\ell-1}}^\sigma, X_{j_\ell}^\sigma]] =: \sum_{k=1}^n (f_{j_1 \dots j_\ell}^k)^\sigma \partial_k,
\end{aligned}
\tag{2.6}
$$

for any word $j_1 \dots j_\ell$, with $2 \le \ell \le s$. (Observe that $f_w^\sigma \ne f_w^{(\sigma)}$ if $|w| > 1$. See Section 7.) Then:

Proposition 2.4. *Let $X_1, \dots, X_m$ be vector fields in the class $\mathcal{A}_s$. Then the following hold:*

(1) *For any $\ell = 1, \dots, s$, for any word w of length $|w| \le \ell$,*

$$X_w^\sigma \to X_w, \tag{2.7}$$

as $\sigma \to 0$, uniformly on K. In particular, for any multi-index $I = (i_1, \dots, i_n) \in \mathcal{S}$, we have $\lambda_I^\sigma := \det(Y_{i_1}^\sigma, \dots, Y_{i_n}^\sigma) \longrightarrow \lambda_I$, uniformly on K, as $\sigma \to 0$.

(2) *There is $\sigma_0 > 0$ such that if $|w| = s$ and $k = 1, \dots, m$, then*

$$\sup_{0<\sigma<\sigma_0} \sup_{x \in K} |X_k^\sigma f_w^\sigma| \le C, \tag{2.8}$$

with C depending on L in (2.2).

(3) *There is r_0 depending on K, Ω and the constant in (2.2) such that the following holds. Let $x \in K$, $r < r_0$ and $b \in L^\infty([0,1], \mathbb{R}^m)$ with $\|b_j\|_{L^\infty} \le r$ for all j. Then there is a unique $\varphi \in \mathrm{Lip}([0,1], \mathbb{R}^n)$, a.e. a solution of $\dot\varphi = \sum_j b_j X_j(\varphi)$, with $\varphi(0) = x$. Denote also by $\varphi^\sigma \in \mathrm{Lip}([0,1], \mathbb{R}^n)$ the a.e. solution of $\dot\varphi^\sigma = \sum_j b_j X_j^\sigma(\varphi^\sigma)$, with $\varphi^\sigma(0) = x$. Then*

$$\varphi^\sigma(1) \to \varphi(1), \tag{2.9}$$

as $\sigma \to 0$, uniformly in $x \in K$. As a consequence, for any $I \in \mathcal{S}$, uniformly in $x \in K$, $|h| \le C^{-1}$,

$$E_I^\sigma(x, h) := \exp_*(h_1 Y_{i_1}^\sigma) \cdots \exp_*(h_n Y_{i_n}^\sigma) \to E_I(x, h). \tag{2.10}$$

Proof. The proofs of items (1) and (2) are given in detail in Section 7. Item (3) follows from standard properties of ODE. $\square$

Remark 2.5. The approximation result contained in Proposition 2.4 is crucial for our subsequent arguments. Note that the class $\mathcal{A}_s$ requires a control on the Euclidean gradients of all commutators of length strictly less than s. However, it is natural to conjecture that a control only along the horizontal directions could be sufficient to ensure our main structure theorem in Section 5. Unfortunately, it seems quite difficult to get an approximation theorem such as Proposition 2.4 for a more general class than $\mathcal{A}_s$. On the other side, working without mollified vector fields seems to raise some nontrivial new issues which we plan to face in a further study.

Some more notation. Our notation for constants are the following: C, C_0 denote large absolute constants, and $\varepsilon_0, r_0, t_0, C^{-1}$ or C_0^{-1} denote positive small absolute constants. "Absolute constants" may depend on the dimension n, the number m of the fields, their step s, the constant L in (2.2) and possibly the constant ν in (2.5). We also use the notation ε_η (or C_η) to denote a small (or a large) constant depending also on η. The constants σ_0 or $\widetilde{\sigma}$ appearing in the regularizing parameter σ may also depend on the Euclidean continuity moduli of the vector fields f_w, with $|w| = s$, which are not included in L. Composition of functions are shortened as follows: fg stands for $f \circ g$. The notation u is always used for functions of the form $\exp(t_1 Z_1) \cdots \exp(t_\nu Z_\nu)$ for some $t_j \in \varrho$, $\nu \geq 1$, $Z_j \in \{X_1, \dots, X_m\}$.

3. Approximate exponentials of commutators

The main result of this section is Theorem 3.6 in Subsection 3.3, where we prove an exact formula for the derivative $\frac{d}{dt} u(e_*^{tX_w}(x))$, where X_w is a commutator of length $|w| \leq s$, while e_* is the approximate exponential defined in (2.3). This section is written for smooth vector fields, namely the mollified X_j^σ, but all constants appearing in our computations are stable as σ goes to 0. We drop everywhere in this section the superscript σ.

We will show that

$$\begin{aligned} \frac{d}{dt} e_*^{tX_w}(x) = X_w(e_*^{tX_w}(x)) + {} & \text{ higher order commutators} \\ + {} & \text{ integral remainder.} \end{aligned} \tag{3.1}$$

The integral remainder is rather complicated, but we do not need its exact form. In order to understand what we need to compute the derivative in (3.1), let us try to calculate for example the derivative $\frac{d}{dt} u(e^{tX} e^{tY} x)$, where $X, Y \in \{\pm X_1, \dots, \pm X_m\}$ and u denotes the identity function in $\mathbb{R}^n$. Since X and Y are C^1, we have

$$\frac{d}{dt} u(e^{tX} e^{tY} x) = (Xu)(e^{tX} e^{tY} x) + Y(ue^{tX})(e^{tY} x).$$

In order to compare the terms on the right-hand side, we may write

$$Y(ue^{tX})(e^{tY} x) = Yu(e^{tX} e^{tY} x) + \int_0^t \frac{d}{d\tau} Y(ue^{\tau X})(e^{-\tau X} e^{tX} e^{tY} x) d\tau.$$

Lemma 3.1 shows that the derivative inside the integral can be written in an exact form in terms of the commutator of X and Y. The purpose of the following Subsection 3.1 is to establish a formalism to study in a precise way more general, related, integral expressions.

Lemma 3.1. *Let Z, X be smooth vector fields. Then*

$$\frac{d}{dt} Z(ue^{-tX})(e^{tX} y) = [X, Z](ue^{-tX})(e^{tX} y). \tag{3.2}$$

Proof. The lemma is known, but we provide a proof for completeness. First observe that

$$\begin{aligned} \frac{d}{dt} Z(ue^{-tX})(e^{tX} x) &= \frac{d}{d\tau} Z(ue^{-tX})(e^{\tau X} x)\Big|_{\tau=t} + \frac{d}{d\tau} Z(ue^{-\tau X})(e^{tX} x)\Big|_{\tau=t} \\ &=: (1) + (2). \end{aligned}$$

Obviously, $(1) = XZ(ue^{-tX})(e^{tX}x)$. Now write (2) as follows:

$$\begin{aligned}\frac{d}{dt}Z(ue^{-tX})(\xi)\big|_{\xi=e^{tX}x} &= \frac{d}{dt}Z^j(\xi)\partial_{\xi_j}(ue^{-tX})(\xi)\big|_{\xi=e^{tX}x}\\ &= Z^j(\xi)\partial_{\xi_j}\frac{d}{dt}(ue^{-tX})(\xi)\big|_{\xi=e^{tX}x}.\end{aligned}$$

The proof of formula (3.2) will be concluded as soon as we prove that

$$\frac{d}{dt}(ue^{-tX})(\xi) = -X(ue^{-tX})(\xi). \tag{3.3}$$

To prove (3.3), start from the identity $u(\eta) = u(e^{-tX}e^{tX}\eta)$, for small t. Differentiating,

$$\begin{aligned}0 = \frac{d}{dt}(u(e^{-tX}e^{tX}\eta)) &= \frac{d}{d\tau}(u(e^{-tX}e^{\tau X}\eta))|_{\tau=t} + \frac{d}{d\tau}(u(e^{-\tau X}e^{tX}\eta))|_{\tau=t}\\ &= Z(ue^{-tX})(e^{tX}\eta) + \frac{d}{d\tau}(u(e^{-\tau X}e^{tX}\eta))|_{\tau=t}.\end{aligned}$$

Then, (3.3) is proved by letting $e^{tX}\eta = \xi$. □

3.1. **Notation for integral remainders.** Let $\lambda \in \mathbb{N}$, $p \in \{2, \dots, s+1\}$. We denote, for $y \in K$ and $t \in [0, t_0]$, t_0 small enough,

$$O_p(t^\lambda, u, y) = \sum_{i=1}^{N}\int_0^t \omega_i(t,\tau)X_{w_i}(u\varphi_i^{-1}e^{-\tau Z_i})(e^{\tau Z_i}\varphi_i y)d\tau, \tag{3.4}$$

where N is a suitable integer and u is the identity map or $u = \exp(tY_1)\cdots\exp(tY_\mu)$, for some integer μ and suitable vector fields $Y_j \in \{\pm X_1, \dots, \pm X_m\}$. Here X_{w_i} actually stands for a mollified $X_{w_i}^\sigma$, but we drop the superscript for simplicity. To describe the generic term of the sum above, we drop the dependence on i:

$$(R) := \int_0^t \omega(t,\tau)X_w(u\varphi^{-1}e^{-\tau X})(e^{\tau X}\varphi y)d\tau. \tag{3.5}$$

Here X_w is a commutator of length $|w| = p$ and $X \in \{\pm X_j\}$. Moreover, for any $t < t_0$, the function $\omega(t,\tau)$ is a polynomial, homogeneous of degree $\lambda - 1$ in all variables (t,τ), such that $\omega(t,\tau) > 0$ if $0 < \tau < t$. Thus

$$\int_0^t \omega(t,\tau)d\tau = bt^\lambda \quad \text{for any } \ t > 0, \tag{3.6}$$

for a suitable constant $b \in \varrho$. The map φ is the identity map or

$$\varphi = \exp(tZ_1)\cdots\exp(tZ_\nu)$$

for some $\nu \in \mathbb{N}$, where $Z_j \in \{\pm X_1, \dots, \pm X_m\}$.

Remark 3.2. All the numbers N, μ, ν, b, appearing in the computations of this paper will be bounded by absolute constants.

In order to explain how this formalism works, we give the main properties of our integral remainders.

Proposition 3.3. *A remainder of the form* (3.4) *satisfies for every* $\alpha \in \mathbb{N}$,

$$t^\alpha O_p(t^\lambda, u, y) = O_p(t^{\alpha+\lambda}, u, y) \quad \textit{for all } y \in K \quad t \in [0, t_0]. \tag{3.7}$$

Moreover, for $p \le s+1$,

$$|O_p(t^\lambda, u, y)| \le Ct^\lambda \qquad \textit{for all } y \in K \quad t \in [0, t_0], \tag{3.8}$$

where t_0 and C depend on the constant L in (2.2) *and on the numbers N, μ, ν, b appearing in the sum* (3.4). *Furthermore, if $\ell(Z) = 1$ and $p \le s+1$,*

$$O_p(t^\lambda, ue^{tZ}, y) = O_p(t^\lambda, u, e^{tZ}y). \tag{3.9}$$

Finally, if $p \le s$, we may write, for suitable constants c_w, $|w| = p$,

$$O_p(t^\lambda, u, y) = \sum_{|w|=p} c_w t^\lambda X_w u(y) + O_{p+1}(t^{\lambda+1}, u, y). \tag{3.10}$$

Proof. The proof of (3.7) and (3.9) are rather easy, and we leave them to the reader. So we start with the proof of (3.8). A typical term in $O_p(t^\lambda, u, y)$ has the form

$$\int_0^t \omega(t,\tau) Y(u\varphi^{-1}e^{-\tau Z})(e^{\tau Z}\varphi y)d\tau, \tag{3.11}$$

with $\ell(Y) = p \le s+1$. Thus, by Proposition 2.4, we have $\left|Y(u\varphi^{-1}e^{-\tau Z})(e^{\tau Z}\varphi y)\right| \le C$ (observe that we need (2.8), if $p = s+1$). Therefore, (3.8) follows from the property (3.6) of ω.

Finally, we establish the key property (3.10). Start from the generic term of $O_p(t^\lambda, u, y)$ in (3.11), where we introduce the notation $g_k := e^{tZ_k} \cdots e^{tZ_\nu}$ for $k = 1, \dots, \nu$ and where $g_{\nu+1}$ denotes the identity map. Recall also that $\ell(Y) \le s$. Therefore, we get

$$\begin{aligned}
&\int_0^t \omega(t,\tau) Y(ue^{-tZ_\nu} \cdots e^{-tZ_1}e^{-\tau X})\big(e^{\tau X}e^{tZ_1} \cdots e^{tZ_\nu}y\big)d\tau \\
&= \int_0^t \omega(t,\tau) Y(ug_1^{-1}e^{-\tau X})\big(e^{\tau X}g_1 y\big)d\tau \\
&\quad = \int_0^t \omega(t,\tau) Yu(y)d\tau + \int_0^t \omega(t,\tau)\Big\{Y(ug_1^{-1}e^{-\tau X})\big(e^{\tau X}g_1y\big) - Yu(y)\Big\}d\tau \\
&\quad = bYu(y)t^\lambda + \int_0^t \omega(t,\tau)\Big\{Y(ug_1^{-1}e^{-\tau X})\big(e^{\tau X}g_1y\big) - Y(ug_1^{-1})\big(g_1y\big)\Big\}d\tau \\
&\qquad\quad + \sum_{k=1}^{\nu} \int_0^t \omega(t,\tau)\Big\{Y(ug_k^{-1})\big(g_k y\big) - Y(ug_{k+1}^{-1})\big(g_{k+1}y\big)\Big\}d\tau.
\end{aligned}$$

Recall that Y has length $p \le s$. The penultimate term can be written as

$$\begin{aligned}
&\int_0^t \omega(t,\tau)\Big\{Y(ug_1^{-1}e^{-\tau X})\big(e^{\tau X}g_1y\big) - Yu(ug_1^{-1})\big(g_1y\big)\Big\}d\tau \\
&\quad = \int_0^t d\tau\ \omega(t,\tau) \int_0^\tau d\sigma \frac{d}{d\sigma} Y(ug_1^{-1}e^{-\sigma X})\big(e^{\sigma X}g_1y\big) \\
&\quad = \int_0^t d\sigma \Big\{\int_\sigma^t \omega(t,\tau)d\tau\Big\}[X,Y](ug_1^{-1}e^{-\sigma X})\big(e^{\sigma X}g_1y\big).
\end{aligned}$$

Observe that, as required, the function $\widetilde{\omega}(t,\sigma) := \int_\sigma^t \omega(t,\tau)d\tau$ satisfies

$$\int_0^t \widetilde{\omega}(t,\sigma)d\sigma = \int_0^t d\tau\omega(t,\tau)\int_0^\tau d\sigma = \int_0^t d\tau\ \tau\omega(t,\tau) = \widetilde{b}t^{\lambda+1},$$

because $(t,\tau) \mapsto \tau\omega(t,\tau)$ is homogeneous of degree λ.

The $k-$th term in the sum has the form

$$\begin{aligned}&\int_0^t d\tau\,\omega(t,\tau)\int_0^t d\sigma\frac{d}{d\sigma}Y(ug_{k+1}^{-1}e^{-\sigma Z_k})\big(e^{\sigma Z_k}g_{k+1}y\big)\\&\qquad=\int_0^t d\sigma\,\widetilde{\omega}(t,\sigma)[Z_k,Y](ug_{k+1}^{-1}e^{-\sigma Z_k})\big(e^{\sigma Z_k}g_{k+1}y\big),\end{aligned}$$

where $\widetilde{\omega}(t,\sigma):=\int_0^t\omega(t,\tau)d\tau=bt^\lambda$ has the correct form. The proof is concluded. $\square$

3.2. **Higher order noncommutative calculus formulas.** In order to prove Theorem 3.5, we first need to iterate formula (3.2). Start from smooth vector fields $X:=X_j^\sigma$ of length one and $Z:=X_w$ of length $\ell(Z):=|w|$. Differentiating identity (3.2) we get, by the Taylor formula,

$$Z(ue^{-tX})(e^{tX}y)=\sum_{k=0}^r\frac{t^k}{k!}\mathrm{ad}_X^kZu(y)+\int_0^t\frac{(t-\tau)^r}{r!}\mathrm{ad}_X^{r+1}Z(ue^{-\tau X})(e^{\tau X}y)d\tau,$$

where we introduced the notation $\mathrm{ad}_XZ=[X,Z]$, $\mathrm{ad}_X^2Z=[X,[X,Z]]$, etcetera. In other words,

$$\begin{aligned}&Z(ue^{tX})(y)-Zu(e^{tX}y)\\&\quad=\sum_{k=1}^r\frac{t^k}{k!}\mathrm{ad}_{-X}^kZu(e^{tX}y)+\int_0^t\frac{(t-\tau)^r}{r!}\mathrm{ad}_{-X}^{r+1}Z(ue^{\tau X})(e^{-\tau X}e^{tX}y)d\tau\\&\quad=\sum_{k=1}^r\frac{t^k}{k!}\mathrm{ad}_{-X}^kZu(e^{tX}y)+O_{r+1+\ell(Z)}(t^{r+1},u,e^{tX}y).\end{aligned}\tag{3.12}$$

If we take $r=s-\ell(Z)$, we may write

$$\begin{aligned}&Z(ue^{tX})(y)-Zu(e^{tX}y)\\&\qquad=\sum_{k=1}^{s-\ell(Z)}\frac{t^k}{k!}\mathrm{ad}_{-X}^kYu(e^{tX}y)+O_{s+1}(t^{s-\ell(Z)+1},u,e^{tX}y).\end{aligned}\tag{3.13}$$

In view of (3.8), this order of expansion is the highest that ensures the remainder can be estimated with $Ct^{s-\ell(Z)+1}$, with a control on C in terms of the constant in (2.2), as soon as $y\in K$ and $|t|\le C^{-1}$.

Next, we seek for a family of higher order formulas, in which we change e^{tX} with an approximate exponential $\exp_*(tX_w)$. The coefficients of the expansion (3.12) are all explicit, but we do not need such an accuracy in the higher order formulas. To explain what suffices for our purposes, start with the case of commutators of length two. Let $C_t=C_t(X,Y)=e^{-tY}e^{-tX}e^{tY}e^{tX}$, where $X:=X_j^\sigma$ and $Y:=X_k^\sigma$ are mollified vector fields with length one.

Let $Z:=X_v^\sigma$ be a smooth commutator with length $\ell(Z):=|v|$. Assume first that $\ell(Z)=s$. Then, iterating (3.13) we can write

$$Z(uC_t)(x)-Zu(C_tx)=O_{s+1}(t,u,C_tx).\tag{$F_{2,1}$}$$

If instead $\ell(Z) = s-1$, then some elementary computations based on (3.12) give

$$(F_{2,2})\qquad \begin{aligned} &Z(uC_t)(x) - Zu(C_tx) \\ &\quad = \sum_{k_1+k_2+k_3+k_4=1} \frac{t^{k_1+\cdots+k_4}}{k_1!\cdots k_4!}\mathrm{ad}_Y^{k_4}\mathrm{ad}_X^{k_3}\mathrm{ad}_{-Y}^{k_2}\mathrm{ad}_{-X}^{k_1} Zu(C_tx) \\ &\qquad\qquad + O_{2+\ell(Z)}(t^2, u, C_tx) \\ &\quad = O_{2+\ell(Z)}(t^2, u, C_tx) = O_{s+1}(t^2, u, C_tx). \end{aligned}$$

Next, if $\ell(Z) = s-2$ (this can happen only if $s \geq 3$), then we must expand more. Namely, we have

$$(F_{2,3})\qquad \begin{aligned} &Z(uC_t)(x) - Zu(C_tx) \\ &\quad = \sum_{k_1+k_2+k_3+k_4=1}^{2} \frac{t^{k_1+\cdots+k_4}}{k_1!\cdots k_4!}\mathrm{ad}_Y^{k_4}\mathrm{ad}_X^{k_3}\mathrm{ad}_{-Y}^{k_2}\mathrm{ad}_{-X}^{k_1} Zu(C_tx) \\ &\qquad\qquad + O_{3+\ell(Z)}(t^3, u, C_tx) \\ &\quad = t^2[Z,[X,Y]]u(C_tx) + O_{3+\ell(Z)}(t^3, u, C_tx) \\ &\quad = t^2[Z,[X,Y]]u(C_tx) + O_{s+1}(t^3, u, C_tx). \end{aligned}$$

Finally, if $\ell(Z) \leq s-3$ (this requires at least $s \geq 4$), we must expand even more:
$(F_{2,4})$

$$\begin{aligned} &Z(uC_t)(x) - Zu(C_tx) \\ &= \sum_{k_1+k_2+k_3+k_4=1}^{3} \frac{t^{k_1+\cdots+k_4}}{k_1!\cdots k_4!}\mathrm{ad}_Y^{k_4}\mathrm{ad}_X^{k_3}\mathrm{ad}_{-Y}^{k_2}\mathrm{ad}_{-X}^{k_1} Zu(C_tx) + O_{4+\ell(Z)}(t^4, u, C_tx) \\ &= t^2[Z,[X,Y]]u(C_tx) + t^3\Big\{\frac{1}{2}\mathrm{ad}_Y^2\mathrm{ad}_X Zu(C_tx) - \frac{1}{2}\mathrm{ad}_Y\mathrm{ad}_X^2 Zu(C_tx) \\ &\quad - \frac{1}{2}\mathrm{ad}_X^2\mathrm{ad}_Y Zu(C_tx) + \frac{1}{2}\mathrm{ad}_X\mathrm{ad}_Y^2 Zu(C_tx) \\ &\quad - \mathrm{ad}_Y\mathrm{ad}_X\mathrm{ad}_Y Zu(C_tx) + \mathrm{ad}_X\mathrm{ad}_Y\mathrm{ad}_X Zu(C_tx)\Big\} + O_{4+\ell(Z)}(t^4, u, C_tx). \end{aligned}$$

Observe that if $\ell(Z) = s-3$, then $O_{4+\ell(Z)}(t^4, u, C_tx) = O_{s+1}(t^4, u, C_tx)$. If instead $\ell(Z) < s-3$, then we can expand up to the order $O_{s+1}(t^{s+1-\ell(Z)}, u, C_tx)$ by means of (3.10).

We have started to put tags of the form $(F_{k,\lambda})$ in our formulas. The number k indicates the length of the commutator we are approximating, while the number λ denotes the power of t which controls the remainder.

Note that in $(F_{2,4})$, the curly bracket changes sign if we exchange X with Y. Briefly, we can write

$$\begin{aligned} Z(uC_t)(x) = {} & Zu(C_tx) + t^2[Z,[X,Y]]u(C_tx) \\ & + t^3 \sum_{|w|=3+\ell(Z)} c_w X_w u(C_tx) + O_{4+\ell(Z)}(t^4, u, C_tx) \end{aligned}$$

for all $x \in K$, $t \in [0, C^{-1}]$, where the coefficients c_w are determined in $(F_{2,4})$. The corresponding formula for $C_t^{-1}(X,Y)$ is

$$Z(uC_t^{-1})(x) = Zu(C_t^{-1}x) - t^2[Z,[X,Y]]u(C_t^{-1}x) + t^3 \sum_{|w|=3+\ell(Z)} \widetilde{c}_w X_w u(C_t^{-1}x) + O_{4+\ell(Z)}(t^4, u, C_t^{-1}x),$$

where, since $C_t^{-1}(X,Y) = C_t(Y,X)$, the coefficients $\widetilde{c}_w$ are obtained again from $(F_{2,4})$ by changing X and Y. We are not interested in the explicit knowledge of all the coefficients c_w and $\widetilde{c}_w$. We only need to observe the following remarkable cancellation property:

$$\sum_{|w|=3+\ell(Z)} (c_w + \widetilde{c}_w) X_w(x) = 0 \quad \text{for any } x \in K.$$

Next we generalize formulas $(F_{2,\lambda})$ above. The general statement we prove tells us that this cancellation persists when the length of the commutator we are approximating with C_t is three or more.

Theorem 3.4. *For any $\ell \in \{2, \dots, s\}$, $x \in K$, and $t \in [0, C^{-1}]$, the following family $(F_{\ell,1}, F_{\ell,2}, \dots, F_{\ell,s})$ of formulas holds.*

Formula $F_{\ell,k}$. *For any $C_t = C_t(X_{w_1}, \dots, X_{w_\ell})$, $k = 1, \dots, \ell$, and for any commutator Z such that $\ell(Z) + k \le s + 1$, we have*

$$Z(uC_t)(x) - Zu(C_t x) = O_{k+\ell(Z)}(t^k, u, C_t x),$$
$$Z(uC_t^{-1})(y) - Zu(C_t^{-1} y) = O_{k+\ell(Z)}(t^k, u, C_t^{-1} x).$$

Formula $F_{\ell,\ell+1}$. *Let $\ell \ge 2$ be such that $\ell + 1 \le s$. Then, for any $C_t(X_{w_1}, \dots, X_{w_\ell})$ and Z such that $\ell + 1 + \ell(Z) \le s + 1$,*

$$Z(uC_t)(x) - Zu(C_t x) = t^\ell [Z, X_w] u(C_t x) + O_{\ell+1+\ell(Z)}(t^{\ell+1}, u, C_t x),$$
$$Z(uC_t^{-1})(y) - Zu(C_t^{-1} y) = -t^\ell [Z, X_w] u(C_t^{-1} x) + O_{\ell+1+\ell(Z)}(t^{\ell+1}, u, C_t^{-1} x).$$

Formula $F_{\ell,\ell+2}$. *If $s \ge 4$, let $\ell \ge 2$ be such that $\ell + 2 \le s$. Then, for any $C_t(X_{w_1}, \dots, X_{w_\ell})$ and Z such that $\ell + 2 + \ell(Z) \le s + 1$, there are numbers $c_v, \widetilde{c}_v$, with $|v| = \ell + \ell(Z) + 1$, such that*

(3.14)
$$Z(uC_t)(x) - Zu(C_t x) = t^\ell [Z, X_w] u(C_t x) + t^{\ell+1} \sum_{|v|=\ell+\ell(Z)+1} c_v X_v u(C_t x) + O_{\ell+2+\ell(Z)}(t^{\ell+2}, u, C_t x)$$
$$Z(uC_t^{-1})(x) - Zu(C_t^{-1} x) = -t^\ell [Z, X_w] u(C_t^{-1} x) + t^{\ell+1} \sum_{|v|=\ell+\ell(Z)+1} \widetilde{c}_v X_v u(C_t^{-1} x) + O_{\ell+2+\ell(Z)}(t^{\ell+2}, u, C_t^{-1} x).$$

Cancellation property. *If $s \ge 4$, let $\ell \ge 2$ be such that $\ell + 2 \le s$. If formulas $F_{\ell,1}$ through $F_{\ell,\ell+2}$ hold, then, for any $C_t(X_{w_1}, \dots, X_{w_\ell})$ and Z such that $\ell + 2 + \ell(Z) \le s + 1$, the coefficients $c_w, \widetilde{c}_w$ in (3.14) satisfy*

$$\sum_{|w|=\ell+\ell(Z)+1} (c_w + \widetilde{c}_w) X_w(x) = 0 \qquad \text{for any } x \in K. \tag{3.15}$$

Formula $F_{\ell,r}$, with $\ell+3 \le r \le s$. *Let $s \ge 5$ and assume that $\ell \ge 2$ and r are such that $\ell+3 \le r \le s$. Then, for any $C_t(X_{w_1},\dots,X_{w_\ell})$ and Z with $r+\ell(Z) \le s+1$, there are $c_v, \widetilde{c}_v$ such that*

$$\begin{aligned} Z(uC_t)(x) - Zu(C_t x) &= t^\ell [Z, X_w] u(C_t x) \\ &\quad + \sum_{|v|=\ell+\ell(Z)+1}^{r-1+\ell(Z)} t^{|v|-\ell(Z)} c_v X_v u(C_t x) \\ &\quad + O_{r+\ell(Z)}(t^r, u, C_t x), \\ Z(uC_t^{-1})(x) - Zu(C_t^{-1} x) &= -t^\ell [Z, X_w] u(C_t^{-1} x) \\ &\quad + \sum_{|v|=\ell+\ell(Z)+1}^{r-1+\ell(Z)} t^{|v|-\ell(Z)} \widetilde{c}_v X_v u(C_t^{-1} x) \\ &\quad + O_{r+\ell(Z)}(t^r, u, C_t^{-1} x). \end{aligned}$$

Observe again that in the formula $F_{\ell,k}$, ℓ is the length of the commutator which defines C_t, while k is the degree of the power of t which controls the remainder.

Proof of Theorem 3.4. If $\ell = 2$, we have already proved the statement. See formulas ($F_{2,1}$), ($F_{2,2}$), ($F_{2,3}$) and ($F_{2,4}$) and recall property (3.10) of the remainders. The proof will be accomplished in two steps.

Step 1. Let $s \ge 4$ and $\ell \ge 2$ be such that $\ell+2 \le s$. Assume that $F_{\ell,1}, F_{\ell,2}, \dots, F_{\ell,\ell+2}$ hold. Then the cancellation (3.15) holds for any $C_t(X_{j_1},\dots,X_{j_\ell})$ and W such that $\ell+2+\ell(W) \le s+1$.

Step 2. Assume that for some $\ell \ge 2$, all formulas $F_{\ell,k}$ hold, for $k = 1,\dots,s$. Then formula $F_{\ell+1,k}$ holds, for any $k = 1,\dots,s$.

Proof of Step 1. Let $C_t = C_t(X_{w_1},\dots,X_{w_\ell})$ and Z be such that $\ell(Z)+\ell+2 \le s+1$. Applying formula $F_{\ell,\ell+2}$ twice, we obtain

$$\begin{aligned} Zu(x) &= Z(uC_t^{-1}C_t)(x) \\ &= Z(uC_t^{-1})(C_t x) + t^\ell [Z, X_w](uC_t^{-1})(C_t x) \\ &\quad + t^{\ell+1} \sum_{|v|=\ell+\ell(Z)+1} c_v X_v (uC_t^{-1})(C_t x) \\ &\quad + O_{\ell+2+\ell(Z)}(t^{\ell+2}, uC_t^{-1}, C_t x) \\ &= Zu(x) - t^\ell [Z, X_w] u(x) + t^{\ell+1} \sum_{|v|=\ell+\ell(Z)+1} \widetilde{c}_v X_v u(x) \\ &\qquad + O_{\ell+\ell(Z)+2}(t^{\ell+2}, u, x) + t^\ell [Z, X_w](uC_t^{-1})(C_t x) \\ &\qquad + t^{\ell+1} \sum_{|v|=\ell+\ell(Z)+1} c_v X_v (uC_t^{-1})(C_t x) \\ &\qquad + O_{\ell+2+\ell(Z)}(t^{\ell+2}, uC_t^{-1}, C_t x). \end{aligned} \tag{3.16}$$

First observe that property (3.9) gives

$$O_{\ell+2+\ell(Z)}(t^{\ell+2}, uC_t^{-1}, C_t x) = O_{\ell+2+\ell(Z)}(t^{\ell+2}, u, x).$$

Later on, we will tacitly use such property many times. Recall that $\ell \geq 2$. By means of $F_{\ell,2}$ and $F_{\ell,1}$, respectively, we obtain

$$[Z, X_w](uC_t^{-1})(C_t x) = [Z, X_w]u(x) + O_{2+\ell(Z)+\ell}(t^2, u, x) \quad \text{and}$$
$$X_v(uC_t^{-1})(C_t x) = X_v u(x) + O_{2+\ell+\ell(Z)}(t, u, x).$$

Inserting this information into (3.16) gives, after algebraic simplifications,

$$\begin{aligned} 0 = t^{\ell+1} &\sum_{|v|=\ell+\ell(Z)+1} \widetilde{c}_v X_v u(x) + O_{\ell+2+\ell(Z)}(t^{\ell+2}, u, x) + t^\ell O_{2+\ell+\ell(Z)}(t^2, u, x) \\ &+ t^{\ell+1}\Big(\sum_{|v|=\ell+\ell(Z)+1} c_v X_v u(x) + O_{\ell+2+\ell(Z)}(t, u, x)\Big). \end{aligned}$$

To conclude the proof, recall (3.7), divide by $t^{\ell+1}$ and let $t \to 0$. ☐

Proof of Step 2. Let $\ell + 2 \leq s$. We prove formula $F_{\ell+1,\ell+2}$, which is the most significant among all formulas $F_{\ell+1,1}, \dots, F_{\ell+1,s}$. Indeed, once $F_{\ell+1,\ell+2}$ is proved, if $\ell + 3 \leq s$, then formulas $F_{\ell+1,\ell+3}, \dots, F_{\ell+1,s}$ follow easily from $F_{\ell+1,\ell+2}$ and from property (3.10). On the other side, the lower order formulas $F_{\ell+1,k}$ with $k < \ell + 2$ are easier (just truncate at the correct order all the expansions in the proof below). ☐

To start, recall that we are assuming that $F_{\ell,1}, \dots, F_{\ell,s}$ hold. Let, for $t > 0$

$$\begin{aligned} C_t &:= C_t(X_{w_1}, \dots, X_{w_\ell}) \quad \text{and} \\ C_t^0 &:= C_t(X, X_{w_1}, \dots, X_{w_\ell}) = C_t^{-1} e^{-tX} C_t e^{tX}, \end{aligned} \tag{3.17}$$

where $X = X_{w_0}$. Let Z be a commutator with $\ell(Z) + \ell + 2 \leq s + 1$. In the subsequent formulas, we expand everywhere up to a remainder of the form $O_{\ell+2+\ell(Z)}(t^{\ell+2}, u, C_t^0 x)$. By (3.12),

$$\begin{aligned} Z(uC_t^0)(x) &= Z(uC_t^{-1}e^{-tX}C_t)(e^{tX}x) + t[-X, Z](uC_t^{-1}e^{-tX}C_t)(e^{tX}x) \\ &\quad + \sum_{k=2}^{\ell+1} \frac{t^k}{k!} \mathrm{ad}_{-X}^k Z(uC_t^{-1}e^{-tX}C_t)(e^{tX}x) \\ &\quad + O_{\ell+2+\ell(Z)}(t^{\ell+2}, uC_t^{-1}e^{-tX}C_t, e^{tX}x) \\ &=: (A) + (B) + (C) + O_{\ell+2+\ell(Z)}(t^{\ell+2}, u, C_t^0 x), \end{aligned}$$

where we also used (3.9). Next we use $F_{\ell,\ell+2}$ in (A):

$$\begin{aligned} (A) &= Z(uC_t^{-1}e^{-tX})(C_t e^{tX}x) + t^\ell [Z, X_w](uC_t^{-1}e^{-tX})(C_t e^{tX}x) \\ &\quad + t^{\ell+1} \sum_{|v|=\ell+\ell(Z)+1} c_v X_v(uC_t^{-1}e^{-tX})(C_t e^{tX}x) + O_{\ell+2+\ell(Z)}(t^{\ell+2}, u, C_t^0 x) \\ &=: (A_1) + (A_2) + (A_3) + O_{\ell+2+\ell(Z)}(t^{\ell+2}, u, C_t^0 x). \end{aligned}$$

We first treat (A_1). By (3.12),

$$\begin{aligned} (A_1) &= Z(uC_t^{-1})(e^{-tX}C_t e^{tX}x) + t[X, Z](uC_t^{-1})(e^{-tX}C_t e^{tX}x) \\ &\quad + \sum_{k=2}^{\ell+1} \frac{t^k}{k!} \mathrm{ad}_X^k Z(uC_t^{-1})(e^{-tX}C_t e^{tX}x) + O_{\ell+2+\ell(Z)}(t^{\ell+2}, u, C_t^0 x). \end{aligned} \tag{3.18}$$

Now consider the various terms in (A_1). First use $F_{\ell,\ell+2}$ to get

$$\begin{aligned}Z(uC_t^{-1})(e^{-tX}C_te^{tX}x) &= Zu(C_t^0x) - t^\ell[Z,X_w]u(C_t^0x)\\ &\quad + t^{\ell+1}\sum_{|v|=\ell+\ell(Z)+1}\widetilde{c}_vX_vu(C_t^0x)\\ &\quad + O_{\ell+2+\ell(Z)}(t^{\ell+2},u,C_t^0x).\end{aligned}$$

Moreover, by $F_{\ell,\ell+1}$ we get

$$\begin{aligned}&t[X,Z](uC_t^{-1})(e^{-tX}C_te^{tX}x)\\ &\quad = t\Big\{[X,Z]u(C_t^0x) - t^\ell[[X,Z],X_w]u(C_t^0x) + O_{\ell+\ell(Z)+2}(t^{\ell+1},u,C_t^0x)\Big\}.\end{aligned}$$

Finally, we use $F_{\ell,\ell+2-k}$ in the $k-$th term of the sum in (3.18). Observe that $\ell+2-k\in\{1,\dots,\ell\}$ so that we use only remainders:

$$\frac{t^k}{k!}\mathrm{ad}_X^kZ(uC_t^{-1})(e^{-tX}C_te^{tX}x) = \frac{t^k}{k!}\Big\{\mathrm{ad}_X^kZu(C_t^0x) + O_{\ell+\ell(Z)+2}(t^{\ell+2-k},u,C_t^0x)\Big\}.$$

Therefore,

$$\begin{aligned}(A_1) &= Zu(C_t^0x) - t^\ell[Z,X_w]u(C_t^0x) + t^{\ell+1}\sum_{|v|=\ell+1+\ell(Z)}\widetilde{c}_vX_vu(C_t^0x)\\ &\quad + t[X,Z]u(C_t^0x) - t^{\ell+1}[[X,Z],X_w]u(C_t^0x)\\ &\quad + \sum_{k=2}^{\ell+1}\frac{t^k}{k!}\mathrm{ad}_X^kZu(C_t^0x) + O_{\ell+2+\ell(Z)}(t^{\ell+2},u,C_t^0x).\end{aligned}$$

Next we consider (A_2). Formula (3.12) gives

$$\begin{aligned}(A_2) &= t^\ell[Z,X_w](uC_t^{-1})(e^{-tX}C_te^{tX}x)\\ &\quad + t^{\ell+1}[X,[Z,X_w]](uC_t^{-1})(e^{-tX}C_te^{tX}x) + O_{\ell+\ell(Z)+2}(t^{\ell+2},u,C_t^0x).\end{aligned}$$

Since $\ell\ge 2$, formulas $F_{\ell,2}$ and $F_{\ell,1}$ give, respectively,

$$\begin{aligned}t^\ell[Z,X_w](uC_t^{-1})(e^{-tX}C_te^{tX}x) &= t^\ell[Z,X_w]u(C_t^0x) + O_{\ell+\ell(Z)+2}(t^{\ell+2},u,C_t^0x),\\ t^{\ell+1}[X,[Z,X_w]](uC_t^{-1})(e^{-tX}C_te^{tX}x) &= t^{\ell+1}[X,[Z,X_w]]u(C_t^0x)\\ &\quad + O_{\ell+\ell(Z)+2}(t^{\ell+2},u,C_t^0x),\end{aligned}$$

so that

$$(A_2) = t^\ell[Z,X_w]u(C_t^0x) + t^{\ell+1}[X,[Z,X_w]]u(C_t^0x) + O_{\ell+\ell(Z)+2}(t^{\ell+2},u,C_t^0x).$$

To handle (A_3), observe that a repeated application of (3.12) gives

$$(A_3) = t^{\ell+1}\sum_{|v|=\ell+\ell(Z)+1}c_vX_vu(C_t^0x) + O_{\ell+\ell(Z)+2}(t^{\ell+2},u,C_t^0x).$$

Next we study (B). Start with formula $F_{\ell,\ell+1}$:

$$\begin{aligned}(B) &= t[-X,Z](uC_t^{-1}e^{-tX})(C_te^{tX}x) + t^{\ell+1}[[-X,Z],X_w](uC_t^{-1}e^{-tX})(C_te^{tX}x)\\ &\quad + O_{\ell+\ell(Z)+2}(t^{\ell+2},u,C_t^0x)\\ &= t[-X,Z](uC_t^{-1}e^{-tX})(C_te^{tX}x)\\ &\quad + t^{\ell+1}[[-X,Z],X_w]u(C_t^0x) + O_{\ell+\ell(Z)+2}(t^{\ell+2},u,C_t^0x)\\ &=: (B_1) + (B_2) + O_{\ell+\ell(Z)+2}(t^{\ell+2},u,C_t^0x).\end{aligned}$$

We first consider (B_1). In view of (3.12), we obtain

$$(B_1) = t[-X,Z](uC_t^{-1})(e^{-tX}C_te^{tX}x) - t\sum_{k=1}^{\ell}\frac{t^k}{k!}\mathrm{ad}_X^k[X,Z](uC_t^{-1})(e^{-tX}C_te^{tX}x) + tO_{\ell+\ell(Z)+2}(t^{\ell+1},u,C_t^0x).$$

But by $F_{\ell,\ell+1}$ we get

$$t[-X,Z](uC_t^{-1})(e^{-tX}C_te^{tX}x) = t[-X,Z]u(C_t^0x) - t^{\ell+1}[[-X,Z],X_w]u(C_t^0x) + O_{\ell+\ell(Z)+2}(t^{\ell+2},u,C_t^0x),$$

while for any $k=1,\dots,\ell$, formula $F_{\ell,\ell+1-k}$ gives

$$t\frac{t^k}{k!}\mathrm{ad}_X^k([X,Z])(uC_t^{-1})(e^{-tX}C_te^{tX}x) = \frac{t^{k+1}}{k!}\mathrm{ad}_X^{k+1}Zu(C_t^0x) + O_{\ell+\ell(Z)+2}(t^{\ell+2},u,C_t^0x).$$

Therefore,

$$(B_1) = t^{\ell+1}[[X,Z],X_w]u(C_t^0x) - \sum_{k=0}^{\ell}\frac{t^{k+1}}{k!}\mathrm{ad}_X^{k+1}Zu(C_t^0x) + O_{\ell+\ell(Z)+2}(t^{\ell+2},u,C_t^0x).$$

Observe that $t^{\ell+1}[[X,Z],X_w]u(C_t^0x) = -(B_2)$.

Finally we consider (C). In the $k-$th term of the sum use formula $F_{\ell,\ell+2-k}$. Then

$$\begin{aligned}(C) &= \sum_{k=2}^{\ell+1}\frac{t^k}{k!}\mathrm{ad}_{-X}^kZ(uC_t^{-1}e^{-tX})(C_te^{tX}x) + O_{\ell+2+\ell(Z)}(t^{\ell+2},u,C_t^0x) \text{ by (3.12)}\\ &= \sum_{k=2}^{\ell+1}\frac{t^k}{k!}\Big\{\sum_{h=0}^{\ell+1-k}\frac{t^h}{h!}\mathrm{ad}_X^h\mathrm{ad}_{-X}^kZ(uC_t^{-1})(e^{-tX}C_te^{tX}x)\\ &\qquad + O_{\ell+\ell(Z)+2}(t^{\ell+2-k},u,C_t^0x)\Big\} + O_{\ell+\ell(Z)+2}(t^{\ell+2},u,C_t^0x)\\ &= \sum_{k=2}^{\ell+1}\sum_{h=0}^{\ell+1-k}\frac{t^{k+h}}{k!h!}(-1)^k\mathrm{ad}_X^{k+h}Zu(C_t^0x) + O_{\ell+\ell(Z)+2}(t^{\ell+2},u,C_t^0x).\end{aligned}$$

Collecting together all the previous computations and making some simplifications (in particular, here we need the cancellation property (3.15)), we get

$$\begin{aligned}Z(uC_t^0)(x) &= (A_1)+(A_2)+(A_3)+(B_1)+(B_2)+(C)\\ &= Zu(C_t^0x) + t^{\ell+1}\big\{-[[X,Z],X_w]u(C_t^0x) + [X,[Z,X_w]]u(C_t^0x)\big\}\\ &\quad + \sum_{k=1}^{\ell+1}\frac{t^k}{k!}\mathrm{ad}_X^kZu(C_t^0x) - \sum_{k=0}^{\ell}\frac{t^{k+1}}{k!}\mathrm{ad}_X^{k+1}Zu(C_t^0x)\\ &\quad + \sum_{k=2}^{\ell+1}\sum_{h=0}^{\ell+1-k}\frac{t^{k+h}}{k!h!}(-1)^k\mathrm{ad}_X^{k+h}Zu(C_t^0x) + O_{\ell+\ell(Z)+2}(t^{\ell+2},u,C_t^0x)\\ &=: Zu(C_t^0x) + t^{\ell+1}\{\cdots\} + (1) + (2) + (3) + O_{\ell+\ell(Z)+2}(t^{\ell+2},u,C_t^0x).\end{aligned}$$

The Jacobi identity gives $t^{\ell+1}\{\cdots\} = t^{\ell+1}[Z,[X,X_w]]$, which is the desired term.

Ultimately we need to consider all the terms with sums. Changing k and h in (2), we may write

$$(2)+(3)=\sum_{k=1}^{\ell+1}\sum_{h=0}^{\ell+1-k}(-1)^k\frac{t^{k+h}}{k!h!}\mathrm{ad}_X^{k+h}Zu(C_t^0x) \qquad \text{and}$$

$$(1)+Zu(C_t^0x)=\sum_{h=0}^{\ell+1}\frac{t^h}{h!}\mathrm{ad}_X^h Zu(C_t^0x).$$

Therefore,

$$\begin{aligned}(1)+(2)+(3)+Zu(C_t^0x)&=\sum_{k=0}^{\ell+1}\sum_{h=0}^{\ell+1-k}(-1)^k\frac{t^{k+h}}{k!h!}\mathrm{ad}_X^{k+h}Zu(C_t^0x)\\ &=\sum_{s=0}^{\ell+1}\Big(\sum_{\substack{k+h=s\\ k,h\geq 0}}\frac{(-1)^k}{k!h!}\Big)t^s\mathrm{ad}_X^s Zu(C_t^0x)=Zu(C_t^0x),\end{aligned}$$

because $\displaystyle\sum_{\substack{k+h=s\\ k,h\geq 0}}\frac{(-1)^k}{k!h!}=0$ for all $s\geq 1$. The proof of Step 2 and of Theorem 3.4 is concluded. □

3.3. **Derivatives of approximate exponentials.** Here we give the formula for the derivative of an approximate exponential. All of this subsection is written for the mollified vector fields X_j^σ, but we drop the superscript everywhere.

Theorem 3.5. *There is $t_0>0$ such that, for any $\ell\in\{2,\dots,s\}$, $w=(w_1,\dots,w_\ell)$, letting $C_t=C_t(X_{w_1},\dots,X_{w_\ell})$, there are constants $a_w,\widetilde{a}_w$ such that, for any $x\in K$ and $t\in[0,t_0]$,*

$$\frac{d}{dt}u(C_tx)=\ell t^{\ell-1}X_wu(C_tx)+\sum_{|v|=\ell+1}^{s}a_vt^{|v|-1}X_vu(C_tx)+O_{s+1}(t^s,u,C_tx), \tag{3.19}$$

$$\begin{aligned}\frac{d}{dt}u(C_t^{-1}x)=-\ell t^{\ell-1}X_wu(C_t^{-1}x)+\sum_{|v|=\ell+1}^{s}\widetilde{a}_vt^{|v|-1}X_vu(C_t^{-1}x)\\ +O_{s+1}(t^s,u,C_t^{-1}x),\end{aligned} \tag{3.20}$$

where, if $\ell=s$, the sum is empty, while, if $2\leq\ell<s$, we have the cancellation

$$\sum_{|w|=\ell+1}\big\{a_w+\widetilde{a}_w\big\}X_w(x)=0 \quad \text{for all } x\in K. \tag{3.21}$$

From Theorem 3.5 it is very easy to obtain the following result:

Theorem 3.6. *For any commutator X_w with length $|w|=\ell\leq s$, we have, for $x\in K$ and $t\in[-t_0,t_0]$,*

$$\begin{aligned}\frac{d}{dt}u(e_*^{tX_w}(x))=X_wu(e_*^{tX_w}(x))+\sum_{|v|=\ell+1}^{s}\alpha_v(t)X_vu\big(e_*^{tX_w}(x)\big)\\ +O_{s+1}\big(|t|^{(s+1-\ell)/\ell},u,e_*^{tX_w}(x)\big),\end{aligned} \tag{3.22}$$

where the sum is empty if $\ell = s$, $\alpha_v(t) = \ell^{-1} a_v t^{(|v|/\ell)-1)}$ *if* $t > 0$*, and* $\alpha_v(t) = -\ell^{-1}\widetilde{a}_v |t|^{(|v|/\ell)-1)}$ *if* $t < 0$*. In particular, the map* $(t, x) \longmapsto e_*^{tX_w}(x)$ *is of class* C^1 *on* $(-t_0, t_0) \times \Omega'$.

Example 5.7 shows that, even if the vector fields are smooth, then the map $\exp_*(tX_w)$ is at most $C^{1,\alpha}$ for some $\alpha < 1$.

Proof of Theorem 3.6. Formula (3.22) follows immediately from (3.19), (3.20) and the definition (2.3) of e_*. Now we only need to show that the map is C^1 in both variables t, x.

Recall that the vector fields X_j^σ are smooth and in particular C^1. By classical ODE theory (see [Ha, Chap. 5]), any map of the form

$$(\tau_1, \dots, \tau_\nu, x) \mapsto e^{\tau_1 X_{i_1}} \cdots e^{\tau_\nu X_{i_\nu}} x$$

is C^1 if the τ_j's belong to some neighborhood of the origin and $x \in \Omega'$. This implies that for any commutator X_w, the map $\nabla_x \exp_*(tX_w)x$ is continuous on $(t, x) \in I \times \Omega'$, while $\frac{d}{dt}\exp_*(tX_w)x$ is continuous on $(t, x) \in \big(I \setminus \{0\}\big) \times \Omega'$.

Next we prove that $\frac{d}{dt}\exp_*(tX_w)x$ exists and that it is also continuous at all points of the form $(0, x)$. First observe that formula (3.22) gives

$$\lim_{t\to 0} \frac{d}{dt}\exp_*(tX_w)x = X_w(x) \qquad \text{uniformly in } x \in \Omega'. \tag{3.23}$$

Now, (3.23) and l'Hôpital's rule imply that $\frac{d}{dt}\exp_*(tX_w)x\big|_{t=0} = X_w(x)$, for all $x \in \Omega'$. Finally, the uniformity of the limit ensures that the map $(t, x) \mapsto \frac{d}{dt}\exp_*(tX_w)x$ is actually continuous in $I \times \Omega'$. □

Proof of Theorem 3.5. We divide the proof in two steps.

Step 1. We first prove that, if (3.19) and (3.20) hold for some w with $\ell := |w| \in \{2, \dots, s-1\}$, then the cancellation formula (3.21) must hold. Fix such a w and start from the identity $\frac{d}{dt}u(C_t^{-1}C_t x) = 0$:

(3.24)
$$\begin{aligned}
0 &= \frac{d}{ds}u(C_s^{-1}C_t x)\Big|_{s=t} + \frac{d}{ds}(uC_t^{-1})(C_s x)\Big|_{s=t} \\
&= -\ell t^{\ell-1}X_w u(x) + \sum_{|v|=\ell+1} \widetilde{a}_v t^\ell X_v u(x) + O_{\ell+2}(t^{\ell+1}, u, x) + \ell t^{\ell-1} X_w(uC_t^{-1})(C_t x) \\
&\quad + \sum_{|v|=\ell+1} a_v t^\ell X_v(uC_t^{-1})(C_t x) + O_{\ell+2}(t^{\ell+1}, uC_t^{-1}, C_t x).
\end{aligned}$$

But, since $\ell \geq 2$, formula $F_{\ell,2}$ shows that

$$t^{\ell-1}\big\{X_w(uC_t^{-1})(C_t x) - X_w u(x)\big\} = t^{\ell-1}O_{2+|w|}(t^2, u, x) = O_{\ell+2}(t^{\ell+1}, u, x),$$

while $F_{\ell,1}$ gives for any v with $|v| = \ell+1$,

$$t^\ell\{X_v(uC_t^{-1})(C_t x) - X_v u(x)\} = t^\ell O_{1+|v|}(t, u, x) = O_{\ell+2}(t^{\ell+1}, u, x).$$

Divide (3.24) by t^ℓ and let $t \to 0$ to get (3.21). Step 1 is concluded.

Step 2. We prove by an induction argument that, if Theorem 3.5 holds for some $\ell \in \{2, \dots, s-1\}$, then it holds for $\ell + 1$. To show the result for $\ell = 2$, it suffices to follow the proof below, taking into account that formulas (3.19) and (3.20) are trivial if $\ell = 1$. We use the notation in (3.17) for C_t and C_t^0. In view of (3.10) and of the already accomplished Step 1, it suffices to prove that

$$
\begin{aligned}
\frac{d}{dt}u(C_t^0x) &= (\ell+1)t^\ell[X,X_w]u(C_t^0x) + O_{\ell+2}(t^{\ell+1},u,C_t^0x) \quad \text{and} \\
\frac{d}{dt}u((C_t^0)^{-1}x) &= -(\ell+1)t^\ell[X,X_w]u((C_t^0)^{-1}x) + O_{\ell+2}(t^{\ell+1},u,(C_t^0)^{-1}x).
\end{aligned} \tag{3.25}
$$

We prove only the first line of (3.25). The latter is similar. We know that

$$
\begin{aligned}
\frac{d}{dt}\left(u(C_tx)\right) =& \ell t^{\ell-1}X_wu(C_tx) + t^\ell\sum_{|v|=\ell+1}a_vX_vu(C_tx) + O_{\ell+2}(t^{\ell+1},u,C_tx) \\
\frac{d}{dt}u(C_t^{-1}x) =& -\ell t^{\ell-1}X_wu(C_t^{-1}x) + t^\ell\sum_{|v|=\ell+1}\widetilde{a}_vX_vu(C_t^{-1}x) \\
&+ O_{\ell+2}(t^{\ell+1},u,C_t^{-1}x),
\end{aligned} \tag{3.26}
$$

with the remarkable cancellation (3.21). Observe that $a_v = \widetilde{a}_v = 0$ if $\ell = 1$. Next,

$$
\begin{aligned}
\frac{d}{dt}\left(u(C_t^0x)\right) =& \frac{d}{dt}\left(u(C_t^{-1}e^{-tX}C_te^{tX}x)\right) \\
=& X\left(uC_t^{-1}e^{-tX}C_t\right)(e^{tX}x) + \frac{d}{ds}(uC_t^{-1}e^{-tX})(C_se^{tX}x)\Big|_{s=t} \\
&- X\left(uC_t^{-1}\right)(e^{-tX}C_te^{tX}x) + \frac{d}{ds}u(C_s^{-1}e^{-tX}C_te^{tX}x)\Big|_{s=t} \\
=:& A_1 + A_2 + A_3 + A_4.
\end{aligned}
$$

First we study $A_1 + A_3$, by (3.12) and $F_{\ell,\ell+1}$:

$$
\begin{aligned}
A_1 + A_3 =& X\left(uC_t^{-1}e^{-tX}C_t\right)(e^{tX}x) - X\left(uC_t^{-1}e^{-tX}\right)(C_te^{tX}x) = \ (\text{by } F_{\ell,\ell+1}) \\
=& t^\ell[X,X_w](uC_t^{-1}e^{-tX})(C_te^{tX}x) + O_{\ell+2}(t^{\ell+1},uC_t^{-1}e^{tX},C_te^{tX}x) \\
=& t^\ell[X,X_w]u(C_t^0x) + O_{\ell+2}(t^{\ell+1},u,C_t^0x).
\end{aligned}
$$

Next we study $A_2 + A_4$, by means of (3.26):

$$
\begin{aligned}
A_2 + A_4 =& \ell t^{\ell-1}X_w(uC_t^{-1}e^{-tX})(C_te^{tX}x) + t^\ell\sum_{|v|=\ell+1}a_vX_v(uC_t^{-1}e^{-tX})(C_te^{tX}x) \\
&- \ell t^{\ell-1}X_wu(C_t^0x) + t^\ell\sum_{|v|=\ell+1}\tilde{a}_vX_vu(C_t^0x) + O_{\ell+2}(t^{\ell+1},u,C_t^0x) \\
=& \ell t^{\ell-1}\Big\{X_w(uC_t^{-1})(e^{-tX}C_te^{tX}x) + t[X,X_w](uC_t^{-1})(e^{-tX}C_te^{tX}x) \\
&+ O_{\ell+2}(t^2,u,C_t^0x)\Big\} \\
&+ t^\ell\sum_{|v|=\ell+1}a_v\left\{X_vu(C_t^0x) + O_{\ell+2}(t,u,C_t^0x)\right\} \\
&- \ell t^{\ell-1}X_wu(C_t^0x) + t^\ell\sum_{|v|=\ell+1}\tilde{a}_vX_vu(C_t^0x) + O_{\ell+2}(t^{\ell+1},u,C_t^0x).
\end{aligned}
$$

Now observe that by formula $F_{\ell,2}$ we have, if $\ell \geq 2$,

$$
t^{\ell-1}\Big\{X_w(uC_t^{-1})(e^{-tX}C_te^{tX}x) - X_wu(C_t^0x)\Big\} = O_{\ell+2}(t^{\ell+1},u,C_t^0x),
$$

while, if $\ell = 1$ the left-hand side vanishes identically. Thus, cancellation (3.21) gives $A_2+A_4 = \ell t^\ell [X, X_w]u(C_t^0 x)+O_{\ell+2}(t^{\ell+1}, u, C_t^0 x)$ and ultimately $A_1+A_2+A_3+A_4 = (\ell+1)t^\ell [X, X_w]u(C_t^0 x) + O_{\ell+2}(t^{\ell+1}, u, C_t^0 x)$. The proof is concluded. □

4. Persistence of maximality conditions on balls

Here we establish a key property of stability of the $\eta-$maximality condition. The argument, as in [TW], is based on Gronwall's inequality.

Theorem 4.1. *Let $X_1, \dots, X_m$ be vector fields in $\mathcal{A}_s$. Then, there are $r_0 > 0$ and $\varepsilon_0 > 0$ depending on the constants L and ν in* (2.5) *and* (2.2) *such that if for some $\eta \in \,]0, 1[\,$, $x \in K$ and $r < r_0$, the triple (I, x, r) is $\eta-$maximal, then for any $y \in B(x, \eta\varepsilon_0 r)$, we have the estimates*

$$|\lambda_I(y) - \lambda_I(x)| \le \frac{1}{2}|\lambda_I(x)|, \tag{4.1}$$

$$|\lambda_I(y)|r^{\ell(I)} > C^{-1}\eta\Lambda(y, r). \tag{4.2}$$

To prove Theorem 4.1 we need the following easy lemma.

Lemma 4.2. *There is $C > 0$ depending on L and ν such that, given $y \in \Omega$ and $z \in \mathbb{R}^n$, the linear system $\sum_{i=1}^q Y_i(y)\xi^i = z$ has a solution $\xi \in \mathbb{R}^q$ such that $|\xi| \le C|z|$.*

Proof. Take $y \in \Omega$ and choose $(k_1, \dots, k_n) \in \mathcal{S}$ such that $\det(Y_{k_1}(y), \dots, Y_{k_n}(y)) \ge \nu$. Let $A := (Y_{k_1}(y), \dots, Y_{k_n}(y))$. Thus, $\left|A^{-1}\right| \le C/|\det(A)| \le C/\nu$, where C depends on L. The lemma is easily proved by studying the system $A\xi = z$ with $\xi = (\xi_{k_1}, \dots, \xi_{k_n}) \in \mathbb{R}^n$. □

Proof of Theorem 4.1. Observe that if (I, x, r) is η-maximal, then there is $\widetilde{\sigma} > 0$ which may also depend on I, x, r, such that (I, x, r) is η-maximal for the mollified X_j^σ for all $\sigma \le \widetilde{\sigma}$. Therefore, we will give the proof for smooth vector fields (without writing any superscript). The nonsmooth case will follow by passing to the limit as $\sigma \to 0$ and taking into account that all constants are stable.

Let $J \in \mathcal{S}$ and let $\lambda_J(x) := \det[Y_{j_1}(x), \dots, Y_{j_n}(x)]$. Let X be a vector field of length one. Recall the following formula (see [NSW, Lemma 2.6]):

$$\begin{aligned}
X\lambda_J &= (\mathrm{div}X)\lambda_J + \sum_{k=1}^n \det(\dots, Y_{j_{k-1}}, [X, Y_{j_k}], Y_{j_{k+1}}, \dots) \\
&= (\mathrm{div}X)\lambda_J + \sum_{k\le n,\ \ell_{j_k}<s} \det(\dots, Y_{j_{k-1}}, [X, Y_{j_k}], Y_{j_{k+1}}, \dots) \\
&\quad + \sum_{k\le n,\ \ell_{j_k}=s} \det(\dots, Y_{j_{k-1}}, [X, Y_{j_k}], Y_{j_{k+1}}, \dots) \\
&=: (A) + \sum_{k\le n,\ \ell_{j_k}<s} (B)_k + \sum_{k\le n,\ \ell_{j_k}=s} (C)_k.
\end{aligned}$$

We claim that

$$r^{\ell(J)}\,|X\lambda_J(y)| \le \frac{C}{r}\Lambda(y, r) \qquad \text{for all} \quad y \in \Omega \quad J \in \mathcal{S} \quad r \le r_0. \tag{4.3}$$

To prove (4.3), observe first that if $y \in \Omega$, then

$$|(A)(y)| \le C|\lambda_J(y)| \le C r^{-\ell(J)} \Lambda(y,r),$$

by definition of Λ. This immediately gives the correct estimate for (A). Next we look at $(B)_k$. Since $\ell(Y_{j_k}) \le s-1$, we get for $y \in \Omega$

$$|(B)_k(y)| \le \big| \det\big(\dots, Y_{j_{k-1}}, [X, Y_{j_k}], Y_{j_{k+1}}, \dots\big)(y)\big| \le C\Lambda(y,r)\, r^{-\ell(J)-1}.$$

Finally we consider $(C)_k$. In view of Lemma 4.2, we may write $[X, Y_{j_k}](y) = \sum_{i=1}^q \xi_{j_k,i} Y_i(y)$, where $|\xi_{j_k,i}| \le C\big|[X, Y_{j_k}](y)\big| \le C$ for any $i = 1, \dots, q$. Therefore,

$$\begin{aligned} |(C)_k(y)| &= \Big| \sum_{i=1}^q \xi_{j_k,i} \det\big(\dots, Y_{j_{k-1}}, Y_i, Y_{j_{k+1}}, \dots\big)(y)\Big| \\ &\le C \sum_{i=1}^q \Big| \det\big(\dots, Y_{j_{k-1}}, Y_i, Y_{j_{k+1}}, \dots\big)(y)\Big| \\ &\le C \sum_{i=1}^q r^{-\ell(J)+\ell_{j_k}-\ell_i} \Lambda(y,r) \le C\Lambda(y,r) r^{-\ell(J)-1}, \end{aligned}$$

because $\ell_{j_k} = s \ge \ell_i$, for any $i = 1, \dots, q$. This finishes the proof of (4.3).

Let $\gamma : [0,r] \to \mathbb{R}^n$ be a subunit path with $\gamma(0) = x$, $\gamma(r) = y$. Assume that $x \in K$ and r is small enough to ensure that $B(x,r) \subset \Omega$. Then, by (4.3),

$$r^{\ell(J)}\, |\lambda_J(y) - \lambda_J(x)| \le \frac{C}{r} \int_0^r \Lambda(\gamma(s), r) ds. \tag{4.4}$$

To have differentiability, define $\Lambda_2(x,r) := \Big\{ \sum_I \Big(\lambda_I(x)\, r^{\ell(I)}\Big)^2 \Big\}^{1/2}$, which is equivalent to $\Lambda(x,r)$, through absolute constants. Therefore, (4.3) gives

$$\left|\frac{d}{ds}\Lambda_2(\gamma(s),r)\right| = \Big| \frac{1}{\Lambda_2(\gamma(s),r)} \sum_J r^{2\ell(J)} \lambda_J(\gamma(s)) \partial_s \lambda_J(\gamma(s))\Big| \le \frac{C}{r} \Lambda_2(\gamma(s),r).$$

Integrating the inequality we get

$$\big|\Lambda_2(x,r) - \Lambda_2(\gamma(s),r)\big| \le \Lambda_2(x,r)\Big(\exp\Big(\frac{C}{r}s\Big) - 1\Big). \tag{4.5}$$

Moreover, integrating (4.4) for $J = I$, we get for $0 \le s \le r$,

$$\begin{aligned} \big|r^{\ell(I)}\lambda_I(\gamma(s)) - r^{\ell(I)}\lambda_I(x)\big| &\le \frac{C}{r} \int_0^s \Lambda_2(\gamma(\tau), r) d\tau \\ &\le \frac{C}{r} \int_0^s \Lambda_2(x,r) e^{C\tau/r} d\tau = \Lambda_2(x,r)\big(e^{Cs/r} - 1\big) \\ &\le \frac{Cs}{r} \Lambda(x,r) \le \frac{Cs}{r\eta} r^{\ell(I)}\, |\lambda_I(x)|, \end{aligned} \tag{4.6}$$

because (I,x,r) is η-maximal. Then (4.1) and (4.2) follow from (4.6) and (4.5). $\square$

At this point we can prove the following statement.

Corollary 4.3. *Assume that (I,x,r) is $\eta-$maximal for the vector fields $X_1, \dots, X_m$ in $\mathcal{A}_s$ and for some $x \in K$ and $r \le r_0$. Then for any $y \in B(x, \varepsilon_0 \eta r)$, $i = 1, \dots, q$, we may write $Y_j(y) = \sum_{k=1}^n a_j^k Y_{i_k}(y)$, where $|a_j^k| \le \frac{C}{\eta} r^{\ell_{i_k} - \ell_j}$.*

Proof. Write Y_i instead of $Y_i(y)$. Look at the linear system $Y_j = \sum_{k=1}^n a_j^k Y_{i_k}$. The Cramer's rule furnishes

$$a_j^k = \frac{\det[Y_{i_1}, \dots, Y_{i_{k-1}}, Y_j, Y_{i_{k+1}}, \dots, Y_{i_n}]}{\det[Y_{i_1}, \dots, Y_{i_{k-1}}, Y_{i_k}, Y_{i_{k+1}}, \dots, Y_{i_n}]} \le \frac{C}{\eta} r^{\ell_{i_k} - \ell_j},$$

by (4.2), and the proof is concluded. □

5. Ball-box theorem

5.1. **Derivatives of almost exponential maps.** Here we take Hörmander vector fields $X_1, \dots, X_m$ in $\mathcal{A}_s$. When we choose an $n-$tuple $I = (i_1, \dots, i_n) \in \mathcal{S}$ and the $n-$tuple is understood, we write $Y_{i_j} = U_j$ and $\ell(Y_{i_j}) = \ell(U_j) = d_j$, for $j = 1, \dots, n$. Our first result is:

Theorem 5.1. *There are σ_0, r_0, σ_0 and $C > 0$ such that, given $I \in \mathcal{S}$, then, for any $j = 1, \dots, n$, $\sigma \le \sigma_0$, $x \in K$ and $h \in Q_I(r_0)$, the C^1 map $E^\sigma_{I,x}$ satisfies*

$$\frac{\partial}{\partial h_j} E^\sigma_{I,x}(h) = U^\sigma_j(E^\sigma_{I,x}(h)) + \sum_{|w|=d_j+1}^{s} a^w_j(h) X^\sigma_w(E^\sigma_{I,x}(h)) + \omega^\sigma_j(x,h), \tag{5.1}$$

where the sum is empty if $d_j = \ell(U_j) = s$ and the following estimates hold:

$$|\omega^\sigma_j(x,h)| \le C\|h\|_I^{s+1-d_j} \qquad \text{for any } x \in K, \quad h \in Q_I(r_0), \quad \sigma \le \sigma_0, \tag{5.2}$$

$$|a^w_j(h)| \le C\|h\|_I^{|w|-d_j} \qquad \text{for all } h \in Q_I(r_0), \qquad |w| = d_j + 1, \dots, s. \tag{5.3}$$

Theorem 5.1 holds without assuming η-maximality. If the triple (I, x, r) is $\eta-$maximal, we have more. To state the result, fix once for all a dimensional constant $\chi > 0$ such that

$$\det(I_n + A) \in \left[\frac{1}{2}, 2\right] \qquad \text{for all } A \in \mathbb{R}^{n\times n} \quad \text{with norm } |A| \le \chi. \tag{5.4}$$

Theorem 5.2. *Let $r_0, \sigma_0 > 0$ as in Theorem* 5.1*. Given an $\eta-$maximal triple (I, x, r) for the vector fields X_i, with $x \in K$, $r < r_0$ and $\sigma \le \sigma_0$, then, for any $h \in Q_I(\varepsilon_0 \eta r)$, $j = 1, \dots, n$, we may write*

$$\frac{\partial}{\partial h_j} E^\sigma_{I,x}(h) = U^\sigma_j(E^\sigma_{I,x}(h)) + \sum_{k=1}^{n} (b^k_j)^\sigma(x,h) U^\sigma_k(E^\sigma_{I,x}(h)), \tag{5.5}$$

where

$$|(b^k_j)^\sigma(x,h)| \le \frac{C}{\eta}\frac{\|h\|_I}{r} r^{d_k - d_j} \le \chi r^{d_k - d_j} \quad \text{for all } \ h \in Q_I(\eta\varepsilon_0 r). \tag{5.6}$$

Remark 5.3. Estimate (5.6) and the results in Section 4 imply that, under the hypotheses of Theorem 5.2, we have

$$|\lambda^\sigma_I(x)| \le C_1 |\lambda^\sigma_I(E^\sigma_{I,x}(h))| \le C_2 \left|\det \frac{\partial}{\partial h} E^\sigma_{I,x}(h)\right| \le C_3 |\lambda^\sigma_I(x)| \quad \text{for all } h \in Q_I(\varepsilon_0 \eta r).$$

Proof of Theorem 5.1*.* Without loss of generality we may work in $\mathbb{R}^2$. We drop the superscript σ everywhere. Then $E_I(x,h) = e_*^{h_1 U_1} e_*^{h_2 U_2} x$. Denote by u the identity function in $\mathbb{R}^n$.

We first look at $\partial/\partial h_1$. Theorem 3.6 with $X_w = U_1$ and $t = h_1$ gives

$$\frac{\partial}{\partial h_1} u\big(e_*^{h_1U_1} e_*^{h_2U_2} x\big) = U_1 u(E_{I,x}(h)) + \sum_{|v|=d_1+1}^{s} \alpha_v(h_1) X_v u(E_{I,x}(h)) + O_{s+1}\big(|h_1|^{(s+1-d_1)/d_1}, u, E_{I,x}(h)\big),$$

where we know that $|\alpha_v(h_1)| \le C|h_1|^{(|v|-d_1)/d_1}$ and

$$\big|O_{s+1}\big(|h_1|^{(s+1-d_1)/d_1}, u, E_{I,x}(h)\big)\big| \le C|h_1|^{(s+1-d_1)/d_1}.$$

Thus, since $|h_1|^{1/d_1} \le \|h\|_I$, we have proved (5.10) and (5.11) for $j = 1$.

Next we look at the variable h_2. Theorem 3.6 gives

$$\begin{aligned}\frac{\partial}{\partial h_2} u\big(e_*^{h_1U_1} e_*^{h_2U_2} x\big) &= U_2\big(ue_*^{h_1U_1}\big)\big(e_*^{h_2U_2}x\big)\\ &\quad + \sum_{|v|=d_2+1}^{s} \alpha_v(h_2) X_v\big(ue_*^{h_1U_1}\big)\big(e_*^{h_2U_2}x\big)\\ &\quad + O_{s+1}\big(h_2^{(s+1-d_2)/d_2}, ue_*^{h_1U_1}, e_*^{h_2U_2}x\big),\end{aligned} \tag{5.7}$$

where we know that $|\alpha_v(h_2)| \le C|h_2|^{(|v|-d_2)/d_2} \le C\|h\|_I^{|v|-d_2}$ and

$$\Big|O_{s+1}\big(h_2^{(s+1-d_2)/d_2}, ue_*^{h_1U_1}, e_*^{h_2U_2}x\big)\Big| \le C|h_2|^{(s+1-d_2)/d_2} \le C\|h\|_I^{s+1-d_2}.$$

Now, a repeated application of formula (3.12) gives

$$\begin{aligned}U_2\big(ue_*^{h_1U_1}\big)\big(e_*^{h_2U_2}x\big) &= U_2 u\big(E_{I,x}(h)\big)\\ &+ \sum_{\alpha_1+\cdots+\alpha_\nu=1}^{s-d_2} C_\alpha h_1^{(\alpha_1+\cdots+\alpha_\nu)/d_1} \operatorname{ad}_{Z_1}^{\alpha_1}\cdots\operatorname{ad}_{Z_\nu}^{\alpha_\nu} U_2 u\big(E_{I,x}(h)\big)\\ &+ O_{s+1}\big(h_1^{(s+1-d_2)/d_1}, u, E_{I,x}(h)\big),\end{aligned} \tag{5.8}$$

where we denoted briefly $e_*^{h_1U_1} = e^{-h_1^{1/d_1}Z_1}\cdots e^{-h_1^{1/d_1}Z_\nu}$, where ν is suitable, $h_1 > 0$ and $Z_j \in \{\pm X_1, \dots, \pm X_m\}$. If $h_1 < 0$ the computation is analogous.

To conclude the proof it suffices to write all the terms $X_v(ue_*^{h_1U_1})(e_*^{h_2U_2}x)$ in (5.7) in the form $X_v u\big(E_{I,x}(h)\big)$ plus an appropriate remainder. The argument is the same as that used in equation (5.8), and we leave it to the reader. □

Proof of Theorem 5.2. The proof relies on Corollary 4.3. We drop the superscript σ everywhere. If (I, x, r) is $\eta-$maximal, then (4.2) gives $|\lambda_I(E_{I,x}(h))|r^{\ell(I)} \ge C^{-1}\eta\Lambda(E_{I,x}(h), r)$, as soon as $h \in Q_I(\varepsilon_0\eta r)$.

Write briefly E instead of $E_{I,x}(h)$. Looking at the right-hand side of (5.1), we need to study, for any word w of length $|w| = \ell$, with $\ell = d_j + 1, \dots, s$, the linear system $a_j^w(h)X_w(E) = \sum_{k=1}^n b_j^k U_k(E)$, and we must show that the solution b_j^k satisfies (5.6) if $\|h\|_I \le \varepsilon_0\eta r$. By Corollary 4.3 write $X_w(E) = \sum p_w^k U_k(E)$, where $|p_w^k| \le \frac{C}{\eta} r^{d_k-|w|}$. Thus

$$|b_j^k| = |a_j^w p_w^k| \le C\|h\|_I^{|w|-d_j} \frac{C}{\eta} r^{d_k-|w|} \le \Big(\frac{\|h\|_I}{r}\Big)^{|w|-d_j} \frac{C}{\eta} r^{d_k-d_j}.$$

Here we also used (5.3). This gives the estimate of the terms in the sum in (5.1).

Next we look at the the remainder ω_j. Fix $j = 1, \dots, n$. We know that $|\omega_j| \le C\|h\|_I^{s+1-d_j}$, and we want to write $\omega_j = \sum_k b_j^k U_k(E)$ with estimate (5.6). It is convenient to multiply by r^{d_j}. Let $r^{d_j}\omega_j =: \theta \in \mathbb{R}^n$ and $\xi^k = r^{d_j} b_j^k$. Thus it suffices to show that we can write $\theta = \sum_k \xi^k U_k(E)$, where ξ^k satisfies the estimate $|\xi^k| \leq \frac{C}{\eta}\frac{\|h\|}{r} r^{d_k}$. We know that

$$|\theta| = |r^{d_j}\omega_j| \leq C\|h\|_I^{s+1-d_j} r^{d_j} = C\Big(\frac{\|h\|_I}{r}\Big)^{s+1-d_j} r^{s+1}.$$

To estimate ξ^k, we follow a two-step argument:

Step 1. Write, by Lemma 4.2, $\theta = \sum_{i=1}^q \mu^i Y_i(E)$, for some $\mu \in \mathbb{R}^q$ satisfying $|\mu| \leq C|\theta| \leq C\Big(\frac{\|h\|}{r}\Big)^{s+1-d_j} r^{s+1}$.

Step 2. For any $i = 1, \dots, q$ write $Y_i(E) = \sum_{k=1}^n \lambda_i^k U_k(E)$. This can be done in a unique way and estimate $|\lambda_i^k| \leq \frac{C}{\eta} r^{d_k - \ell(Y_i)}$ holds, by Corollary 4.3.

Collecting Step 1 and Step 2, we conclude that

$$|\xi^k| = \Big|\sum_{i=1}^q \mu^i \lambda_i^k\Big| \leq C\Big(\frac{\|h\|}{r}\Big)^{s+1-d_j} r^{s+1} \cdot \frac{C}{\eta} r^{d_k-\ell(Y_i)} \leq \frac{C}{\eta}\frac{\|h\|}{r} r^{d_k},$$

as required. This ends the proof. □

Next we pass to the limit as $\sigma \to 0$ in both Theorems 5.1 and 5.2.

Theorem 5.4. *If (I, x, r) is $\eta-$maximal for some $x \in K$, $r \leq r_0$, then the map $E_{I,x}\big|_{Q_I(\varepsilon_0\eta r)}$ is locally biLipschitz in the Euclidean sense and satisfies for a.e. h,*

$$\begin{aligned} \frac{\partial}{\partial h_j} E_{I,x}(h) &= U_j(E_{I,x}(h)) + \sum_{|w|=d_j+1}^{s} a_j^w(h) X_w(E_{I,x}(h)) + \omega_j(x,h) \\ &= U_j(E_{I,x}(h)) + \sum_{k=1}^n b_j^k(x,h) U_k(E_{I,x}(h)), \end{aligned} \tag{5.9}$$

where the sum is empty if $d_j = \ell(U_j) = s$ and otherwise the following estimates hold:

$$|\omega_j(x,h)| \leq C\|h\|_I^{s+1-d_j} \qquad \textit{for all } x \in K, \quad h \in Q_I(r_0), \tag{5.10}$$

$$|a_j^w(h)| \leq C\|h\|_I^{|w|-d_j} \qquad \textit{if} \quad |w| = d_j+1, \dots, s \quad \textit{and} \quad h \in Q_I(r_0), \tag{5.11}$$

and

$$|b_j^k(x,h)| \leq \frac{C}{\eta}\frac{\|h\|_I}{r} r^{d_k-d_j} \leq \chi r^{d_k-d_j} \qquad \textit{for all } h \in Q_I(\varepsilon_0\eta r). \tag{5.12}$$

Remark 5.5. If $s \geq 3$, then vector fields of the class $\mathcal{A}_s$ are C^1. Then, as discussed at the beginning of the proof of Theorem 3.6, the map $E_{I,x}$ is actually C^1 smooth. This is not ensured if $s = 2$.

Proof of Theorem 5.4. Look first at the C^1 map $E^\sigma = E^\sigma_{I,x}$ defined on $Q_I(r_0)$. Denote by E its pointwise limit as $\sigma \to 0$. By Theorem 5.1, the map E^σ satisfies for any $\sigma < \sigma_0$, $\|h\|_I \leq r_0$,

$$\frac{\partial}{\partial h_j} E^\sigma(h) = U_j^\sigma(E^\sigma(h)) + \sum_{|w|=d_j+1}^{s} a_j^w(h) X_w^\sigma(E^\sigma(h)) + \omega_j^\sigma(h), \tag{5.13}$$

where a_j^w does not depend on σ, while $|\omega_j^\sigma(h)| \leq C\|h\|_I^{s+1-d_j}$, uniformly in $\sigma \leq \sigma_0$.

Let E^{σ_k} be a sequence weakly converging to E in $W^{1,2}$. Therefore, by (5.13), the remainder j^{σ_k} has a weak limit in L^2. Denote it by ω_j. Standard properties of weak convergence ensure that $|\omega_j(h)| \leq C_0\|h\|_I^{s+1-d_j}$ for a.e. h. Therefore, we have proved the first line of (5.9) and estimates (5.10) and (5.11). To prove the second line and (5.12), it suffices to repeat the argument of Theorem 5.2, taking into account that the main ingredient there, namely Corollary 4.3, holds for nonsmooth vector fields in $\mathcal{A}_s$.

Now we have to prove the local injectivity of E. Let σ be small enough to ensure that (I, x, r) is η-maximal for the vector fields X_j^σ. In view of Theorem 5.2, we can write $dE^\sigma(h) = U^\sigma(E^\sigma(h))(I_n + B^\sigma(h))$, where $U^\sigma = [U_1^\sigma, \dots, U_n^\sigma]$, and the entries of the matrix B satisfy $|(b_j^k)^\sigma| \leq Cr^{d_k-d_j}$, by (5.5). Now fix $h_0 \in Q_I(\varepsilon_0\eta r)$, where $\varepsilon_0\eta$ comes from Theorem 5.2. We will show that E^σ is locally one-to-one around h_0, with a stable coercivity estimate as $\sigma \to 0$. By Proposition 2.4 and by the continuity of the vector fields U_j, we may claim that for any $\delta > 0$ there is $\varrho > 0$ such that $|U_j^\sigma(\xi) - U_j^\sigma(\xi')| < \delta$ as soon as $\xi, \xi' \in K$, $|\xi - \xi'| < \varrho$ and $\sigma < \varrho$. Recall also that E^σ is Lipschitz continuous, uniformly in σ; see (5.13). Then, for any $\delta > 0$ there is $\varrho > 0$ such that $B_{\mathrm{Eucl}}(h_0, \varrho) \subset Q_I(\varepsilon_0\eta r)$, and, if $|h - h_0| \leq \varrho$ and $\sigma < \varrho$, then $|U^\sigma(E^\sigma(h)) - U^\sigma(E^\sigma(h_0))| \leq \delta$.

Take $h, h' \in B_{\mathrm{Eucl}}(h_0, \delta)$. By integrating on the path $\gamma(t) = h' + t(h - h')$, we have

$$\begin{aligned}
|E^\sigma(h) - E^\sigma(h')| &= \Big| \int_0^1 U^\sigma(E^\sigma(\gamma))(I + B^\sigma(\gamma))(h - h')dt \Big| \\
&\geq \Big| \int_0^1 U^\sigma(E^\sigma(h_0))(I + B^\sigma(\gamma))(h - h')dt \Big| \\
&\quad - \Big| \int_0^1 \big(U^\sigma(E^\sigma(\gamma)) - U^\sigma(E^\sigma(h_0))\big)(I + B^\sigma(\gamma))(h - h')dt \Big|.
\end{aligned}$$

To estimate from below the first line, recall the easy inequality $|Ax| \geq C^{-1}\frac{|\det A|}{|A|^{n-1}}|x|$, for all $A \in \mathbb{R}^{n\times n}$. The pointwise estimate $|(b_j^k)^\sigma| \leq \chi r^{d_k-d_j}$ gives $|\int_0^1 (b_j^k)^\sigma(\gamma))dt| \leq \chi r^{d_k-d_j}$. Thus (5.4) gives

$$\Big|\det \int_0^1 (I + B^\sigma(\gamma))dt\Big| = \Big|\det\Big(I + \int_0^1 B^\sigma(\gamma)dt\Big)\Big| \geq \frac{1}{2}.$$

Observe also that $|I + B^\sigma(\gamma)| \leq Cr^{1-s}$. Moreover, in view of Remark 5.3, it must be $|\det U^\sigma(E^\sigma(h_0))| \geq C^{-1}|\lambda_I(x)|$, for small σ. This suffices to estimate from below the first line. To get an estimate of the second line we again need the inequality $|I + B^\sigma(\gamma)| \leq Cr^{1-s}$. Eventually we get

$$|E^\sigma(h) - E^\sigma(h')| \geq \{C_0^{-1}|\lambda_I(x)|\, r^{(n-1)(s-1)} - C_0 r^{1-s}\delta\}|h - h'|,$$

for any $\sigma < \varrho$ and $|h - h'| < \varrho$. The proof is concluded as soon as we choose $\delta = \delta(I, x, r)$ small enough and let $\sigma \to 0$.

This argument shows that the map is locally biLipschitz, as desired. □

5.2. **Pullback of vector fields through scaling maps.** Given an η-maximal triple (I, x, r), for vector fields of the class $\mathcal{A}_s$ we can define, as in [TW], the "scaling map"

$$\Phi_{I,x,r}(t) = \exp\Big(\sum_{j=1}^{n} t_j r^{\ell_{i_j}} Y_{i_j}\Big)x, \tag{5.14}$$

for small $|t|$. The dilation $\delta_r^I(t) := (t_1 r^{\ell_{i_1}}, \dots, t_n r^{\ell_{i_n}})$ makes the natural domain of $\Phi_{I,x,r}$ independent of r. Observe the property $\|\delta_r^I t\|_I = r\|t\|_I$. It turns out that if $\widehat{X}_k$ $(k = 1, \dots, m)$ denotes the pullback of rX_k under $\Phi_{I,x,r}$, then $\widehat{X}_1, \dots, \widehat{X}_m$ satisfy the Hörmander condition in a uniform way. This fact enables the authors in [TW] to give several simplifications to the arguments in [NSW].

We can also consider the scaling map associated with our exponentials. Namely,

$$S_{I,x,r}(t) := \exp_*(t_1 r^{\ell_{i_1}} Y_{i_1}) \cdots \exp_*(t_n r^{\ell_{i_n}} Y_{i_n}) = E_{I,x}(\delta_r^I t). \tag{5.15}$$

It will be proved in Subsection 5.3 that if (I, x, r) is η-maximal, then S is one-to-one on the set $\{\|t\|_I \le \varepsilon_0 \eta\}$. If we assume that the original vector fields are of class C^1 (see Remark 5.5), thus we may define, for all $i \in \{1, \dots, q\}$, the vector fields $\widehat{Y}_j := S_*^{-1}(r^{\ell_i} Y_i)$.

Theorem 5.4 thus becomes

Proposition 5.6. *Let $X_1, \dots, X_m$ be vector fields in $\mathcal{A}_s$. Let (I, x, r) be an $\eta-$maximal triple and let $S := S_{I,x,r}$ be the associated scaling map. Then $S\big|_{Q_I(\varepsilon_0\eta)}$ is a locally biLipschitz map and for a.e. $t \in Q_I(\varepsilon_0\eta)$ we may write*

$$S_*(\partial_{t_j}) = r^{\ell_{i_j}} Y_{i_j}(S(t)) + \sum_{k=1}^{n} \widehat{b}_j^k r^{\ell_{i_k}} Y_{i_k}(S(t)), \tag{5.16}$$

where the functions $\widehat{b}_j^k$ satisfy

$$|\widehat{b}_j^k| \le \frac{C}{\eta}\|t\|_I \qquad \text{for a.e. } t \in Q_I(\varepsilon_0\eta). \tag{5.17}$$

Moreover, if S is C^1 and we write $\widehat{Y}_{i_j} = \partial_{t_j} + \sum_{k=1}^n a_j^k(t)\partial_{t_k}$, then

$$|a_j^k(t)| \le \frac{C}{\eta}\|t\|_I \qquad \text{for all } t \in Q_I(\varepsilon_0\eta). \tag{5.18}$$

Proof. Formula (5.16) is just Theorem 5.4. The proof of (5.18) is a consequence of (5.17) and of the following elementary fact: given a square matrix $B \in \mathbb{R}^{n\times n}$ with norm $|B| \le \frac{1}{2}$, we may write $(I_n + B)^{-1} = I_n + A$ and $|A| = \big|\sum_{k\ge1}(-B)^k\big| \le 2|B|$. $\square$

In the framework of our almost exponential maps, estimate (5.18) is sharp, even for smooth vector fields. The better estimate $Y_{i_j}(t) = \partial_j + \sum_k a_j^k(t)\partial_k$ with $|a_j^k(t)| \le C|t|$, obtained in [TW] for maps of the form (5.14), generically fails for S, as the following example shows.

Example 5.7. Let $X_1 = \partial_1$, $X_2 = a(x_1)\partial_2$ with $a(s) = s + s^2$, or any smooth function with $a(0) = 0$ and $a'(0) \ne 0 \ne a''(0)$. A computation shows that

$$\exp_*(h[X_1, X_2])(x_1, x_2) = \big(x_1, x_2 + \big\{a(x_1 + |h|^{1/2}) - a(x_1)\big\}|h|^{-1/2}h\big).$$

Therefore, at $(x_1, x_2) = (0, 0)$, for small r, we must choose the maximal pair of commutators $X_1, [X_1, X_2]$ and we have

$$\begin{aligned}S(t_1, t_2) &= \exp_*(t_1 r X_1) \exp_*(t_2 r^2 [X_1, X_2])(0, 0) = \big(t_1 r, a(r|t_2|^{1/2})|t_2|^{-1/2} t_2 r\big) \\ &= \big(t_1 r, t_2 r^2 + |t_2|^{1/2} t_2 r^3\big).\end{aligned}$$

Therefore,

$$\widehat{X}_1 = \partial_{t_1}, \qquad \widehat{X}_2 = \frac{t_1 + r t_1^2}{1 + \frac{3}{2} r |t_2|^{1/2}} \partial_{t_2}, \qquad \widehat{[X_1, X_2]} = \frac{1 + 2 r t_1}{1 + \frac{3}{2} r |t_2|^{1/2}} \partial_{t_2}.$$

Clearly the formula $\widehat{[X_1, X_2]} = \partial_{t_2} + O(|t|)$ cannot hold, but (5.18) holds. Observe also that, writing $\widehat{[X_1, X_2]} = \widehat{f}_{(1,2)} \cdot \nabla$, we have $\sup_{t \in U} \big|\widehat{X}_2 \widehat{f}_{(1,2)}\big| \simeq \sup_{t \in U} |t_2|^{-1/2} = +\infty$ for any neighborhood U of the origin. Therefore, the vector fields $\widehat{X}_1, \widehat{X}_2$ do not even belong to the class $\mathcal{A}_2$.

5.3. **Ball-box theorem.** Here we give our main result. We keep the notation from Subsection 5.1.

Theorem 5.8. *Let $X_1, \dots, X_m$ be Hörmander vector fields of step s in the class $\mathcal{A}_s$. There are $r_0, \widetilde{r}_0, C_0 > 0$, and for all $\eta \in (0, 1)$ there are $\varepsilon_\eta, C_\eta > 0$ such that:*

(A) *if (I, x, r) is $\eta-$maximal for some $x \in K$, $r \le r_0$ then, for any $\varepsilon \le \varepsilon_\eta$, we have*

$$E_{I,x}(Q_I(\varepsilon r)) \supset B_\varrho(x, C_\eta^{-1} \varepsilon^s r); \tag{5.19}$$

(B) *if (I, x, r) is $\eta-$maximal for some $x \in K$, $r \le \widetilde{r}_0$, then the map $E_{I,x}$ is one-to-one on the set $Q_I(\varepsilon_\eta r)$.*

Remark 5.9. Observe that on the right-hand side of inclusion (5.19) we use the distance d_ϱ. Therefore, a standard consequence of (5.19) is the well-known property $B(x, r) \supset B_E(x, C^{-1} r^s)$, for any $x \in K$, $r < r_0$. See [FP].

Remark 5.10. In the paper [TW] the authors use the exponential maps in (1.2). If the vector fields have step s, then their method requires that the commutators of length $2s$ are at least continuous. (Here, we specialize [TW] to the case $\varepsilon = 1$ and we do not discuss the higher regularity estimate [TW, Eq. (2.1)].) This appears in the proof of (22) and (23) of [TW, Proposition 4.1]. Indeed in equation (29), the commutator $[X_w, X_{w_j}]$ must be written as a linear combination of commutators $X_{w'}$, where for algebraic reasons it must be $|w'| = |w| + |w_k|$. If $|w| = |w_k| = s$, then commutators of degree $2s$ appear. A similar issue appears for $[Y_{w_i}, Y_{w_j}]$ at the beginning of p. 619.

Remark 5.11. The reason why we introduce two different constants r_0 and $\widetilde{r}_0$ is that C_0, ε_0 and r_0 depend only on L and ν in (2.2) and (2.5) (together with universal constants, such as m, n and s). The constants ε_η and C_η also depend on ν, L and η. We do not have control of $\widetilde{r}_0$ (which appears only in the injectivity statement) in terms of L and ν. This is a delicate question because of the covering argument implicitly contained in [NSW, p. 132] and described in [M, p. 230]. Below we provide a constructive procedure to provide a lower bound for $\widetilde{r}_0$ in terms of the functions λ_I. This can be of some interest in view of the applications of our results to nonlinear problems.

Remark 5.12. The proof of the injectivity result would be considerably simplified if we could prove (uniformly in $x \in K, r < r_0$) an equivalence between the balls and their convex hulls, i.e. $\mathrm{co}B(x,r) \subset B(x,Cr)$, which is reasonable for diagonal vector fields (see [SW, Remark 5]) or a "contractability" property of the ball $B(x,r)$ inside $B(x,Cr)$. See [Sem, Definition 1.7]. Unfortunately, in spite of their reasonable aspect, both these conditions seem quite difficult to prove in our situation. It also seems that the clever argument in [TW, p. 622] can not be adapted to our almost exponential maps.

In the proof of inclusion (5.19), we follow the argument in [NSW, M]. Before giving the proof, we need to show that some constants in the proof actually depend only on L and ν in (2.2) and (2.5). Basically, what we need is contained in Corollary 4.3 and in the following lemma. See [NSW, p. 129].

Lemma 5.13. *Assume that (I,x,r) is $\eta-$maximal for vector fields X_j in $\mathcal{A}_s$, $x \in K$, and $r \le r_0$. Let $\widetilde{\sigma} > 0$ be such that (I,x,r) is η-maximal for the mollified X_j^σ for all $\sigma \le \widetilde{\sigma}$. Let $U \subset Q_I(\varepsilon_\eta r)$, where ε_η comes from Theorem* 5.4*, and assume that a C^1 diffeomorphism $\psi = (\psi^1, \dots, \psi^n) : E^\sigma(U) \to U$ satisfies $\psi(E^\sigma(h)) = h$, for any $h \in U$. Then we have the estimate $|U_j^\sigma \psi^k(E^\sigma(h))| \le Cr^{d_k - d_j}$, for all $h \in U$, where C is independent of σ.*

Proof. It is convenient to work with the map $S^\sigma(t) := E^\sigma(\delta_r t)$ so that $\varphi := \delta_{1/r}\psi$ satisfies $t = \varphi(S^\sigma(t))$, for all $t \in V := \delta_{1/r}U$. The chain rule gives $d\varphi(S^\sigma(t))dS^\sigma(t) = I$, for all $t \in V$. But, by (5.16) we have

$$dS^\sigma(t) = [r^{d_1}U_1^\sigma(S^\sigma), \dots, r^{d_n}U_n^\sigma(S^\sigma)](I + \widehat{B}^\sigma(t)),$$

where $|\widehat{B}^\sigma(t)| \le \frac{C}{\eta}\|t\|_I$ if $\|t\|_I \le \varepsilon_0\eta$. Therefore we may write

$$d\varphi(S^\sigma)[r^{d_1}U_1(S^\sigma), \dots, r^{d_n}U_n(S^\sigma)] = I + A^\sigma,$$

where, as in the proof of Proposition 5.6, $|A^\sigma(t)| \le 2|\widehat{B}^\sigma(t)|$. This implies that $|r^{d_j}U_j^\sigma\varphi^k(S^\sigma(t))| \le C$ and ultimately that $|r^{d_j-d_k}U_j^\sigma\psi^k| \le C$, as desired. □

Proof of Theorem 5.8 (A). Since the vector fields Y_j are not Euclidean Lipschitz continuous, if $\ell_j = s$, we do not know whether or not any point in a ϱ-ball of the Y_j can be approximated by points in the analogous ball of the mollified Y_j^σ. In order to avoid this problem, observe the inclusion $B_\varrho(x,r) \subset B_{\widetilde{\varrho}}(x,Cr)$, where C is absolute and the distance $\widetilde{\varrho}$ is defined using the family $\{Y_j : \ell_j \le s-1, \partial_k : k = 1, \dots, n\}$, where we assign to the vector fields ∂_k a maximal weight s. Therefore, we will prove the inclusion using the distance $\widetilde{\varrho}$, which is defined by Lipschitz vector fields.

Let (I,x,r) be an $\eta-$maximal triple for the original vector fields X_j and let $\widetilde{\sigma}$ be as in Lemma 5.13. Let $y \in B_{\widetilde{\varrho}}(x, C_\eta^{-1}\varepsilon^s r)$, where $\varepsilon \le \varepsilon_\eta$ and where ε_η comes from statement (A), while C_η will be discussed below. Thus, $y = \gamma(1)$, where $\dot\gamma = \sum_{\ell_j \le s-1} b_j Y_j(\gamma) + \sum_{i=1}^n \tilde{b}_j \partial_i(\gamma)$ a.e. on $[0,1]$, with $|b_j(t)| \le (C_\eta^{-1}\varepsilon^s r)^{\ell(Y_j)}$ and $|\tilde{b}_i(t)| \le (C_\eta^{-1}\varepsilon^s r)^s$ for a.e. $t \in [0,1]$. Also let $y^\sigma \in B_{\widetilde{\varrho}}(x, C_\eta^{-1}\varepsilon^s r)$ be an approximating family, $y^\sigma = \gamma^\sigma(1)$, where $\dot\gamma^\sigma = \sum_{\ell_j \le s-1} b_j Y_j^\sigma(\gamma^\sigma) + \sum_{i=1}^n \tilde{b}_j \partial_i(\gamma^\sigma)$ a.e. on $[0,1]$. Observe that $y^\sigma \to y$, as $\sigma \to 0$.

Claim. If C_η is large enough, then for any $\sigma \le \widetilde{\sigma}$ there is a lifting map $\theta^\sigma(t)$, $t \in [0,1]$, with $\theta^\sigma(0) = 0$ and such that

$$E^\sigma(\theta^\sigma(t)) = \gamma^\sigma(t) \quad \text{and} \quad \|\theta^\sigma(t)\|_I < \varepsilon r \quad \text{for all } t \in [0,1]. \tag{5.20}$$

Once the claim is proved, the surjectivity statement follows.

To prove the Claim the key estimate we need is the following. Let $U \subset Q_I(\varepsilon_\eta r)$, $\sigma \le \widetilde{\sigma}$ and assume that a C^1-diffeomorphism $\psi = (\psi^1, \dots, \psi^n)$ satisfies locally $\psi(E^\sigma(h)) = h$, for all $h \in U$, where, for some $t \in [0,1]$, $E^\sigma(U)$ is a neighborhood of $\gamma^\sigma(t)$. Then, for $\mu = 1, \dots, n$ and for all τ close to t,

$$\begin{aligned}
&\Big|\frac{d}{d\tau}\psi^\mu(\gamma^\sigma(\tau))\Big| \\
&= \Big|\sum_{\ell_j \le s-1} b_j(\tau) Y_j^\sigma \psi^\mu(\gamma^\sigma(\tau)) + \sum_{i=1}^n \tilde{b}_i(\tau)\partial_j \psi^\mu(\gamma^\sigma(\tau))\Big| \\
&= \Big|\sum_{\ell_j \le s} b_j(\tau) \sum_{k=1}^n a_j^k(\gamma^\sigma(\tau)) U_k^\sigma \psi^\mu(\gamma^\sigma(\tau)) \\
&\qquad + \sum_{i=1}^n \tilde{b}_i(\tau) \sum_{k=1}^n \tilde{a}_i^k(\gamma^\sigma(\tau)) U_k^\sigma \psi^\mu(\gamma^\sigma(\tau))\Big| \\
&\le \sum_{j,k} C(C_\eta^{-1}\varepsilon^s r)^{\ell(Y_j)} \cdot \frac{C}{\eta} r^{d_k - \ell(Y_j)} \cdot C r^{d_\mu - d_k} \\
&\qquad + \sum_{i,k} C(C_\eta^{-1}\varepsilon^s r)^s \cdot \frac{C}{\eta} r^{d_k - s} \cdot C r^{d_\mu - d_k} \\
&\le \frac{CC_\eta^{-1}}{\eta}\varepsilon^s r^{d_\mu} \le \frac{CC_\eta^{-1}}{\eta}(\varepsilon r)^{d_\mu}.
\end{aligned} \tag{5.21}$$

The constant C_η will be chosen below, while C depends on L, ν, in force of Corollary 4.3 and Lemma 5.13. We used the estimate $\partial_i = \tilde{a}_i^k U_k^\sigma$ with $\left|\tilde{a}_i^k\right| \le \frac{C}{\eta} r^{d_k - s}$, which follows from Lemma 4.2 and Corollary 4.3.

With estimate (5.21) in hand we can prove the claim along the lines of [M, p. 228]. Here is a sketch of the argument.

Step 1. If C_η is large enough, then if $\theta^\sigma(t)$ satisfies $E(\theta^\sigma(t)) = \gamma^\sigma(t)$ on $[0, \bar{t}]$, for some $\bar{t} \le 1$, then $\|\theta^\sigma(t)\|_I < \frac{1}{2}\varepsilon r$, for any $t \le \bar{t}$.

To prove Step 1, assume by contradiction that the statement is false. There is $\widetilde{t} \le \bar{t}$ such that $\|\theta^\sigma(t)\|_I < \frac{1}{2}\varepsilon r$ for all $t < \widetilde{t}$ and $\|\theta^\sigma(\widetilde{t})\|_I = \frac{1}{2}\varepsilon r$. Then for some $\mu \in \{1, \dots, n\}$, we have

$$\Big(\frac{1}{2}\varepsilon r\Big)^{d_\mu} = |\theta_\mu^\sigma(\widetilde{t})| \le \frac{CC_\eta^{-1}}{\eta}(\varepsilon r)^{d_\mu}. \tag{5.22}$$

This estimate can be obtained writing locally $\theta^\sigma(t) = \psi(\gamma^\sigma(t))$ and using (5.21). If we choose C_η large enough to ensure that $\frac{CC_\eta^{-1}}{\eta} < (\frac{1}{2})^s$, then (5.22) cannot hold and we have a contradiction. This ends the proof of Step 1.

Step 2. There exists a path θ^σ on $[0,1]$ satisfying (5.20).

The proof of Step 2 can be done as in [M, p. 229] by a very classical argument involving an upper bound "of Hadamard type" $\|dE^\sigma(\theta^\sigma(t))^{-1}\| \le C$, which holds uniformly in t.

The proof of the statement (A) is concluded. $\square$

Before proving part (B) of Theorem 5.8, we need the following rough injectivity statement.

Lemma 5.14. *Let $x \in K$ and I such that $\lambda_I(x) \neq 0$. Then the function $E_{I,x}$ is one-to-one on the set $Q_I(C^{-1}|\lambda_I(x)|)$.*

Proof. First observe that for all $j = 1, \dots, n$ and small σ, we have

$$\begin{aligned}&\Big|\frac{\partial}{\partial h_j}E^\sigma(h) - U_j^\sigma(x)\Big| \\ &\quad\le \Big|\frac{\partial}{\partial h_j}E^\sigma(h) - U_j^\sigma(E^\sigma(h))\Big| + |U_j^\sigma(E(h)) - U_j^\sigma(x)| \le C\|h\|_I\end{aligned} \tag{5.23}$$

by estimates (5.2), (5.3) and the d-Lipschitz continuity of U_j^σ.

Fix $h, h' \in Q_I(C^{-1}|\lambda_I(x)|)$ and let $\gamma(t) = h' + t(h - h')$. Then

$$\begin{aligned}|E^\sigma(h) - E^\sigma(h')| &= \Big|\int_0^1 dE^\sigma(\gamma)(h-h')dt\Big| \\ &\ge |dE^\sigma(0)(h-h')| - \Big|\int_0^1 \big\{dE^\sigma(\gamma) - dE^\sigma(0)\big\}dt(h-h')\Big| \\ &\ge \big\{C^{-1}|\lambda_I^\sigma(x)| - C\max\{\|h\|_I, \|h'\|_I\}\big\}|h-h'|\end{aligned}$$

by (5.23) and because $dE^\sigma(0) = U^\sigma(x) = [U_1^\sigma(x), \dots, U_n^\sigma(x)]$ has determinant $\lambda_I^\sigma(x)$. The proof is concluded by letting $\sigma \to 0$. $\square$

As announced in Remark 5.11, we provide a constructive procedure for the "injectivity radius" $\widetilde{r}_0$ in Theorem 5.8 in terms of the functions λ_I. Compare [M, pp. 229-230].

Denote by $D_1, \dots, D_p$ all the values attained by $\ell(I)$, as $I \in \mathcal{S}$. Assume that $D_1 < \dots < D_p$ and introduce the notation

$$\sum_I |\lambda_I(x)| r^{\ell(I)} = \sum_{j=1}^p r^{D_j} \sum_{\ell(I)=D_j} |\lambda_I(x)| =: \sum_{j=1}^p r^{D_j}\mu_j(x), \tag{5.24}$$

where μ_j is defined by (5.24). Let $\Sigma_1 := K$ and, for all $k = 2, \dots, p$,

$$\Sigma_k := \{x \in K : \mu_j(x) = 0 \text{ for any } j = 1, \dots, k-1\}.$$

Observe that $\Sigma_1 = K \supseteq \Sigma_2 \supseteq \cdots \supseteq \Sigma_p$. Let $x \in K$. Take $j(x) = \min\{j \in \{1, \dots, p\} : \mu_j(x) \neq 0\}$. Then choose $I_x \in \mathcal{S}$ such that $|\lambda_{I_x}(x)| = \max_{\ell(J)=D_{j(x)}} |\lambda_J(x)|$ is maximal. Therefore, we have $|\lambda_{I_x}(x)| \simeq \mu_{j(x)}(x)$, through absolute constants.

From the construction above we get the following proposition.

Proposition 5.15. *There is $C > 1$ such that, letting $r_x := C^{-1}|\lambda_{I_x}(x)|$ for all $x \in K$*

(1) *we have*

$$|\lambda_{I_x}(y)| r_x^{\ell(I_x)} > C^{-1}\Lambda(y, r_x) \qquad \text{for all } y \in B(x, \varepsilon_0 r_x); \tag{5.25}$$

(2) *the map* $h \mapsto E_{I_x}(y,h)$ *is one-to-one on the set* $Q_{I_x}(r_x)$*, for any* $y \in B(x,\varepsilon_0 r_x)$.

Observe that Proposition 5.15 is far from what we need because it may be $\inf_K r_x = 0$ (for example, this happens in the elementary situation $X_1 = \partial_1, X_2 = x_1\partial_2$).

Proof. We first prove (1) for $y = x$. Namely, we show that

$$|\lambda_{I_x}(x)|r^{\ell(I_x)} \geq |\lambda_J(x)|r^{\ell(J)} \qquad \text{for all } J \in \mathcal{S} \quad r \in [0,r_x], \tag{5.26}$$

where $r_x = C^{-1}|\lambda_{I_x}(x)|$, as required. Let $J \in \mathcal{S}$. If $\lambda_J(x) = 0$, then (5.26) holds for all $r > 0$. If instead $\lambda_J(x) \neq 0$, by the choice of I_x it must be $\ell(J) = \ell(I_x)$ or $\ell(J) > \ell(I_x)$. If $\ell(J) = \ell(I_x)$, then (5.26) holds for any $r > 0$, because $|\lambda_{I_x}(x)|$ is maximal by the construction above. If $\ell(J) > \ell(I_x)$, then

$$|\lambda_J(x)|r^{\ell(J)} \leq |\lambda_{I_x}(x)|r^{\ell(I_x)} \Leftarrow Cr^{\ell(J)-\ell(I_x)} \leq |\lambda_{I_x}(x)| \Leftarrow r \leq C^{-1}|\lambda_{I_x}(x)|.$$

Thus (5.26) holds for any $r \leq r_x$, where r_x has the required form.

The proof of (1) for $y \neq x$ follows from Theorem 4.1.

Finally, to prove (2) observe that, in view of Lemma 5.14, the map $h \mapsto E_{I_x}(y,h)$ is one-to-one on $Q_{I_x}(C^{-1}|\lambda_{I_x}(y)|)$. But Theorem 4.1 and in particular (4.1) show that, if $d(x,y) \leq \varepsilon_0 r_x$, then $|\lambda_{I_x}(y)|$ and $|\lambda_{I_x}(x)|$ are comparable. This concludes the proof. $\square$

Proof of Theorem 5.8 (C). Let $p_1 \leq p$ be the largest integer such that $\Sigma_{p_1} \neq \varnothing$. Then define the "injectivity radius"

$$r_{(p_1)} := \min_{x\in\Sigma_{p_1}} r_x = \min_{x\in\Sigma_{p_1}} C^{-1}|\lambda_{I_x}(x)| \geq C^{-1} \min_{x\in\Sigma_{p_1}} \mu_{p_1}(x) > 0. \tag{5.27}$$

Also denote

$$\Omega_{p_1} = \bigcup_{x\in\Sigma_{p_1}} \Omega' \cap B(x, r_{(p_1)}),$$

where the open set Ω' was introduced before (2.2). Recall that all metric balls $B(x,r)$ are open, by the already accomplished Theorem 5.8, part (A). Then, by Proposition 5.15, for any $y \in \Omega_{p_1}$ there is $x \in \Sigma_{p_1}$ such that the map $h \mapsto E_{I_x}(y,h)$ is one-to-one on $Q_{I_x}(\varepsilon_0 r_x)$ and (I_x, y, r_x) is C^{-1}-maximal. Recall that $r_x \geq r_{(p_1)}$ on Σ_{p_1}.

Next let $p_2 < p_1$ be the largest number such that $K_{p_2} := \Sigma_{p_2} \setminus \Omega_{p_1} \neq \varnothing$. Then, let

$$r_{(p_2)} := \min_{x\in\Sigma_{p_2}\setminus\Omega_{p_1}} r_x \geq C^{-1} \min_{x\in\Sigma_{p_2}\setminus\Omega_{p_1}} \mu_{p_2}(x) > 0.$$

We may claim that for any $y \in \Omega_{p_2} := \bigcup_{x\in K_{p_2}} \Omega' \cap B(x, r_{(p_2)})$, there is $x \in K_{p_2}$ such that the map $h \mapsto E_{I_x}(y,h)$ is one-to-one on the set $Q_{I_x}(\varepsilon_0 r_x)$ and (I_x, y, r_x) is C^{-1}-maximal.

Iterating the argument and letting $\widetilde{r}_0 = \min\{r_{(p_k)}\}$, we conclude that for any $x \in K$ there is an n-tuple $I_0 = I_0(x)$ and $\varrho_0 = \varrho_0(x) \geq \widetilde{r}_0$ such that $E_{I_0}(x,\cdot)$ is one-to-one on the set $Q_{I_0}(\varepsilon_0\varrho_0)$ and (I_0, x, ϱ_0) is C^{-1}-maximal. Clearly, I_0 can be different from I_x. This is the starting point for the proof of the injectivity statement, Theorem 5.8, item (B).

From now on I, $x \in K$ and $r < \widetilde{r}_0$ are fixed and (I,x,r) is $\eta-$maximal, as in the hypothesis of (B). Let I_0 and ϱ_0 be the n-tuple and the injectivity radius associated with x by the argument above. Recall that $\varrho_0 \geq \widetilde{r}_0$. Arguing as in [M, p. 230], see also [NSW, p. 133], we may find a sequence of $n-$tuples $I = I_N, I_{N-1}, \dots, I_1, I_0$

and correspoding numbers $0 \le \varrho_{N+1} < \varrho_N < \cdots < \varrho_0$, with $\varrho_0 \ge \widetilde{r}_0$, $r \in [\varrho_{N+1}, \varrho_N]$ such that for any $j = 0, 1, \dots, N-1$,

$$|\lambda_{I_j}(x)|\varrho^{\ell(I_j)} \ge \eta\Lambda(x, \varrho), \quad \forall\, \varrho \in [\varrho_{j+1}, \varrho_j]. \tag{5.28}$$

In order to show that $E_I = E_{I_N}$ is one-to-one on the set $Q_I(\varepsilon_\eta r)$, for some $\varepsilon_\eta > 0$, we start by showing that E_{I_1} is one-to-one on the set $Q_{I_1}(\varepsilon'_\eta\varrho_1)$, for a suitable ε'_η. What we know is that E_{I_0} is one-to-one on the set $Q_{I_0}(\varrho_0)$. We also know that (5.28) holds for $j = 0, 1$ and $\varrho = \varrho_1$. Therefore, applying (5.19) twice, we have

$$E_{I_1}(Q_{I_1}(\varepsilon_\eta\varrho_1)) \supseteq E_{I_0}(Q_{I_0}(C_\eta^{-1}\varrho_1)) \supseteq E_{I_1}\big(Q_{I_1}(\varepsilon'_\eta\varrho_1)\big). \tag{5.29}$$

Assume by contradiction that $E_{I_1}(h) = E_{I_1}(h') = y$ for some $h, h' \in Q_{I_1}(\varepsilon'_\eta\varrho_1)$. Let $r(t) = h' + t(h - h')$, $t \in [0, 1]$, be the line segment connecting h and h'. Also let $\gamma(t) = E_{I_1}(r(t))$. Since E_{I_0} is one-to-one (actually a C^1 diffeomorphism on its image), we may contract γ to a point just by letting

$$q(\lambda, t) = E_{I_0}\big(\lambda E_{I_0}^{-1}(y) + (1-\lambda)E_{I_0}^{-1}(\gamma(t))\big),$$

where $(\lambda, t) \in [0, 1] \times [0, 1]$. Observe that q is continuous on $[0, 1]^2$, and $q(\lambda, t) \in Q_{I_1}(\varepsilon_\eta\varrho_1)$, by (5.29). Moreover, $q(0, t) = \gamma(t) = E_{I_1}(r(t))$ and $q(1, t) = y$, for any $t \in [0, 1]$. By standard properties of local diffeomorphisms we may claim that there is a continuous lifting $p : [0, 1]^2 \to Q_{I_1}(\varepsilon_\eta\varrho_1)$ such that $E_{I_1}(p(\lambda, t)) = q(\lambda, t)$ and $p(0, t) = r(t)$ for all λ and $t \in [0, 1]$. Next observe that both the maps $\lambda \mapsto E_{I_1}(p(\lambda, 1))$ and $\lambda \mapsto E_{I_1}(p(\lambda, 0))$ are constants on $[0, 1]$. Therefore, since E_{I_1} is a local diffeomorphism, both $\lambda \mapsto p(\lambda, 0)$ and $\lambda \mapsto p(\lambda, 1)$ must be constant. In particular, $p(1, 1) = p(0, 1) = h'$ and $p(1, 0) = p(0, 0) = h$. Finally observe that the path $t \mapsto p(1, t)$ must be constant, because $E_{I_1}(p(1, t)) = y$ for all $t \in [0, 1]$. Therefore we conclude that $h = h'$.

Then we have proved that E_{I_1} is one-to-one on $Q_{I_1}(\varepsilon'_\eta\varrho_1)$. Iterating the argument at most N times, we get the proof of statement (B) of Theorem 5.8. □

6. Examples

Example 6.1 (Levi vector fields)**.** In order to illustrate the previous procedure to find $\tilde{r}_0$, we exhibit the following three-step example. In $\mathbb{R}^3$ consider the vector fields $X_1 = \partial_{x_1} + a_1\partial_{x_3}$ and $X_2 = \partial_{x_2} + a_2\partial_{x_3}$. Assume that the vector fields belong to the class $\mathcal{A}_3$. Let us define $f = X_1a_2 - X_2a_1$. Moreover, assume that $|f| + |X_1f| + |X_2f| \ne 0$ at every point of the closure of a bounded set $\Omega \supset K = \overline{\Omega'}$. Assume also that f has some zero inside K. This condition naturally arises in the regularity theory for graphs of the form $\{(z_1, z_2) \in \mathbb{C}^2 : \mathrm{Im}(z_2) = \varphi(z_1, \bar{z}_1, \mathrm{Re}(z_2))\}$ having some first order zeros. See [CM], where the smoothness of $C^{2,\alpha}$ graphs with prescribed smooth Levi curvature is proved.

In this situation we have $n = 3, m = 2, s = 3$ and $Y_1 = X_1, Y_2 = X_2, Y_3 = [X_1, X_2] = f\partial_{x_3}, Y_4 = [X_1, [X_1, X_2]] = (X_1f - f\partial_{x_3}a_1)\partial_{x_3}, Y_5 = [X_2, [X_1, X_2]] = (X_2f - f\partial_{x_3}a_2)\partial_{x_3}$. Thus, $q = 5$ and

$$\begin{aligned}
\lambda_{(1,2,3)} &= f, & d(1,2,3) &= 4,\\
\lambda_{(1,2,4)} &= X_1f - f\partial_{x_3}a_1, & d(1,2,4) &= 5,\\
\lambda_{(1,2,5)} &= X_2f - f\partial_{x_3}a_2, & d(1,2,5) &= 5.
\end{aligned}$$

Let us put $D_1 = 4, D_2 = 5$ and, by (5.24), $\mu_1 = |f|, \mu_2 = |X_1 f - f\partial_{x_3} a_1| + |X_2 f - f\partial_{x_3} a_2|$. In this situation $\Sigma_1 = K$, $\Sigma_2 = \{x \in K : \mu_1(x) = 0\} = \{x \in K : f(x) = 0\}$. Hence, $r_{(2)} = \min_{x\in\Sigma_2} r_x = \min_{x\in\Sigma_2} \max\{|X_1 f(x)|, |X_2 f(x)|\} > 0$. Let $\Omega_2 = \bigcup_{x\in\Sigma_2} \Omega' \cap B(x, r_{(2)})$, with $\bar{\Omega}' = K$, and let $K_1 = \Sigma_1 \setminus \Omega_2$. Since $K_1 \subseteq \{x \in K : f(x) \neq 0\}$, if $K_1 \neq \varnothing$, then $r_{(1)} = \min_{x\in K_1} r_x = \min_{x\in K_1} |f(x)| > 0$. Finally, if $K_1 \neq \varnothing$, then $\tilde{r}_0 = \min\{r_{(1)}, r_{(2)}\}$, while if $K_1 = \varnothing$, then $\tilde{r}_0 = r_{(2)}$.

In the next example we show a subelliptic-type estimate for nonsmooth vector fields. The argument of the proof below is due to Ermanno Lanconelli (unpublished).

Proposition 6.2 (Hörmander–type estimate [H])**.** *Let $X_1, \dots, X_m$ be a family of vector fields of step s and in the class $\mathcal{A}_s$. Then, given $\Omega' \Subset \Omega$ and $\varepsilon \in]0, 1/s[$, there is $\widetilde{r}_0$ and $C > 0$ such that, for any $f \in C^1(\Omega)$,*

$$[f]_\varepsilon^2 := \int_{\substack{\Omega'\times\Omega',\\ d(x,y)\le\widetilde{r}_0}} \frac{|f(x) - f(y)|^2}{|x-y|^{n+2\varepsilon}} dxdy \le C \int_\Omega \sum_j |X_j f(y)|^2 dy. \tag{6.1}$$

Proof. We just sketch the proof, leaving some details to the reader. For any $I \in \mathcal{S}$, let $\Omega_I := \{x \in \Omega : I_0(x) = I\}$, where $I_0(x)$ comes from the proof of Theorem 5.8, together with $\varrho_0 = \varrho_0(x) \ge \widetilde{r}_0$; see the discussion before equation (5.28). If $x \in \Omega_I$, we have $B(x, \varrho_0) \subset E_I(x, Q_I(C\varrho_0))$, where the biLipschitz map E_I satisfies $C^{-1} \le |\det dE_I(x,h)| \le C$, for a.e. $h \in Q_I(C\varrho_0)$. Thus,

$$\begin{aligned}
[f]_\varepsilon^2 = \int_{\substack{\Omega'\times\Omega'\\ d(x,y)\le\widetilde{r}_0}} \frac{|f(x) - f(y)|^2}{|x-y|^{n+2\varepsilon}} dxdy &\le \sum_{I\in\mathcal{S}} \int_{\Omega_I} dx \int_{d(x,y)\le\varrho_0} dy \frac{|f(x) - f(y)|^2}{|x-y|^{n+2\varepsilon}} \\
&\le C \sum_I \int_{\Omega_I} dx \int_{Q_I(C\varrho_0)} dh \frac{|f(x) - f(E_I(x,h))|^2}{|x - E_I(x,h)|^{n+2\varepsilon}}.
\end{aligned}$$

Now observe that, arguing as in the proof of Lemma 5.14, we have $|E_I(x,h) - x| \ge C^{-1}|h|$ if $\|h\|_I \le C\varrho_0$. Let $\delta_0 = \max_{x\in K} \varrho_0(x)$. Next we follow the argument in [LM]. Write $E_I(x,h) = \gamma_I(x,h,T(h))$, where $\gamma_I(x,h,t)$, $t \in [0, T(h)]$ is a *control function* with the properties described in [LM]. Therefore

$$\begin{aligned}
[f]_\varepsilon^2 &= \sum_I \int_{\Omega_I} dx \int_{Q_I(C\delta_0)} dh \frac{|f(x) - f(E_I(x,h))|^2}{|h|^{n+2\varepsilon}} \\
&\le C \int_{Q_I(C\delta_0)} \frac{dh}{|h|^{n+2\varepsilon}} \int_{\Omega_I} dx \Big| \int_0^{T(h)} dt |Xf(\gamma_I(x,h,t))| \Big|^2 \\
&\le C \int_{Q_I(C\delta_0)} \frac{dh}{|h|^{n+2\varepsilon}} \Big\{ \int_0^{T(h)} dt \Big(\int_{\Omega_I} dx |Xf(\gamma_I(x,h,t))|^2 \Big)^{1/2} \Big\}^2 \\
&\le C \int_{Q_I(C\delta_0)} \frac{dh}{|h|^{n+2\varepsilon}} T(h)^2 \|Xf\|_{L^2(\Omega)}^2 \le C \|Xf\|_{L^2(\Omega)}^2,
\end{aligned}$$

because $x \mapsto \gamma_I(x,h,t)$ is a change of variable, by estimate $T(h) \le \|h\|_I \le |h|^{1/s}$ and the strict inequality $\varepsilon < 1/s$. $\square$

The borderline inequality $\|f\|_{1/s} \le C\|Xf\|_{L^2}$, which cannot be obtained with the argument above, was proved in the smooth case by Rothschild and Stein [RoS].

7. Proof of Proposition 2.4

Here we prove Proposition 2.4. By definition, (2.1) means that for all $j,k \in \{1,\dots,m\}$ and $|w| \le s-1$ there is a bounded function $X_j(X_k f_w)$ such that for any test function $\psi \in C_c^\infty(\varrho^n)$,

$$\int (X_k f_w)(X_j\psi) = -\int \{X_j(X_k f_w) + \operatorname{div}(X_j) X_k f_w\}\psi. \tag{7.1}$$

If $D = \partial_{j_1}\cdots\partial_{j_p}$ for some $j_1,\dots,j_p \in \{1,\dots,n\}$ is an Euclidean derivative, denote by $|D| = p$ its order. It is understood that a derivative of order 0 is the identity.

The first item of Proposition 2.4 is a consequence of the following lemma:

Lemma 7.1. *Let $X_1,\dots,X_m$ be vector fields in $\mathcal{A}_s$. Then for any word w with $|w| \le s$ and for any Euclidean derivative D of order $|D| = p \in \{0,\dots,s-|w|\}$, we have*

$$\sup_K \Big| D f_w^\sigma - (D f_w)^{(\sigma)} \Big| \le C\sigma. \tag{7.2}$$

Note that the case $p=0$ of (7.2) provides the proof of item (1) of Proposition 2.4.

Observe also that if $|w| = s$, then we have $|f_w - f_w^\sigma| \le |f_w - (f_w)^\sigma| + |f_w^\sigma - (f_w)^\sigma|$. Lemma 7.1 gives the estimates of the second term. The first one is estimated by means of the continuity modulus of f_w, which is not included in L in (2.2).

Proof of Lemma 7.1. We argue by induction on $|w|$. If $|w|=1$, then the left-hand side of (7.2) vanishes. Assume that for some $\ell \in \{1,\dots,s-1\}$, (7.2) holds for any word w of length ℓ and for each D with $|D| \le s-\ell$. Let $v = kw$ be a word of length $|kw| = \ell+1$. We must show that for any Euclidean derivative D of order $0 \le |D| \le s-|v|$, (7.2) holds. We have $f_v = X_k f_w - X_w f_k$ and $f_v^\sigma = X_k^\sigma f_w^\sigma - X_w^\sigma f_k^\sigma$. We first prove (7.2) when the order of D satisfies $1 \le |D| \le s-|v| = s-\ell-1$, which can occur only if $\ell \le s-2$ (in particular, this implies $s \ge 3$). The easier case occurs when D is the identity operator, and it will be proven below:

$$\begin{aligned} D f_v^\sigma - (D f_v)^{(\sigma)} &= D\big(X_k^\sigma f_w^\sigma - X_w^\sigma f_k^\sigma\big) - \big(D(X_k f_w - X_w f_k)\big)^{(\sigma)} \\ &= D\big(X_k^\sigma f_w^\sigma\big) - (D X_k f_w)^{(\sigma)} - \{D(X_w^\sigma f_k^\sigma) - \big(D X_w f_k\big)^{(\sigma)}\} \\ &=: (A) - (B). \end{aligned}$$

Omitting the summation sign on $\alpha = 1,\dots,n$, we may write

$$\begin{aligned} (A) &= (D f_k^\alpha)^{(\sigma)}\big\{\partial_\alpha f_w^\sigma - (\partial_\alpha f_w)^{(\sigma)}\big\} + \big\{(D f_k^\alpha)^{(\sigma)}(\partial_\alpha f_w)^{(\sigma)} - \big((D f_k^\alpha)\partial_\alpha f_w\big)^{(\sigma)}\big\} \\ &\quad + (f_k^\alpha)^{(\sigma)}\big\{D\partial_\alpha f_w^\sigma - (D\partial_\alpha f_w)^{(\sigma)}\big\} + \big\{(f_k^\alpha)^{(\sigma)}(D\partial_\alpha f_w)^{(\sigma)} - \big(f_k^\alpha D\partial_\alpha f_w\big)^{(\sigma)}\big\} \\ &=: (A_1) + (A_2) + (A_3) + (A_4). \end{aligned}$$

The estimate $|(A_1)| + |(A_3)| \le C\sigma$ follows from the induction assumption. To estimate (A_4) observe that

$$|(A_4)| = \Big| \int \big\{(f_k^\alpha)^{(\sigma)}(x) - f_k^\alpha(x-\sigma y)\big\} D\partial_\alpha f_w(x+\sigma y)\varphi(y)dy \Big| \le C\sigma,$$

because f_k is Lipschitz, while $D\partial_\alpha f_w \in L^\infty_{\mathrm{loc}}$. Indeed, since $|w| = \ell$, $f_w \in W^{s-\ell,\infty}$. Moreover, D has length at most $s-\ell-1$ so that $D\partial_\alpha$ has length at most $s-\ell$. The estimate of (A_2) is analogous to that of A_4. Just recall that $D f_k^\alpha$ is Lipschitz and $\partial_\alpha f_w$ is bounded.

Next we estimate (B):

$$\begin{aligned}(B) &= D((f_w^\alpha)^\sigma \partial_\alpha f_k^\sigma) - (D(f_w^\alpha \partial_\alpha f_k))^{(\sigma)} \\ &= \{D(f_w^\alpha)^\sigma - (Df_w^\alpha)^{(\sigma)}\}\partial_\alpha f_k^\sigma + \big\{(Df_w^\alpha)^{(\sigma)}\partial_\alpha f_k^\sigma - ((Df_w^\alpha)\partial_\alpha f_k)^{(\sigma)}\big\} \\ &\quad + (f_w^\alpha)^\sigma D\partial_\alpha f_k^\sigma - (f_w^\alpha(D\partial_\alpha f_k))^{(\sigma)} \\ &=: (B_1) + (B_2) + (B_3).\end{aligned}$$

The term (B_1) can be estimated by the inductive assumption. Moreover,

$$|(B_2)| = \Big| \int \big\{\partial_\alpha f_k^\sigma(x) - \partial_\alpha f_k(x-\sigma y)\big\} Df_w^\alpha(x-\sigma y)\varphi(y)dy\Big| \le C\sigma,$$

because $\partial_\alpha f_k$ is Lipschitz and $Df_w^\alpha \in L^\infty_{\mathrm{loc}}$. Finally,

$$|(B_3)| = \Big| \int \big\{(f_w^\alpha)^{(\sigma)}(x) - f_w^\alpha(x-\sigma y)\big\} D\partial_\alpha f_k(x-\sigma y)\varphi(y)dy\Big| \le C\sigma.$$

Indeed, since $|w| \le s-2$, f_w^α is locally Lipschitz. Moreover, since the length of the derivative $D\partial_\alpha$ is at most $s-1$ and $f_k \in W^{s-1,\infty}_{\mathrm{loc}}$, we have $D\partial_\alpha f_k \in L^\infty$.

Next we look at the case where D has length zero, i.e. D is the identity operator. We have to estimate, for v with $|v| \le s$, the difference $f_v^\sigma - (f_v)^{(\sigma)}$. Write $v = kw$, where $k \in \{1, \dots, m\}$. Thus

$$\begin{aligned}f_v^\sigma - (f_v)^{(\sigma)} &= X_k^\sigma f_w^\sigma - X_w^\sigma f_k^\sigma - (X_k f_w)^{(\sigma)} + (X_w f_k)^{(\sigma)} \\ &= (f_k^\alpha)^{(\sigma)}\{\partial_\alpha f_w^\sigma - (\partial_\alpha f_w)^{(\sigma)}\} + \big\{(f_k^\alpha)^{(\sigma)}(\partial_\alpha f_w)^{(\sigma)} - (f_k^\alpha \partial_\alpha f_w)^{(\sigma)}\big\} \\ &\quad - \{(f_w^\alpha)^\sigma \partial_\alpha f_k^\sigma - (f_w^\alpha \partial_\alpha f_k)^{(\sigma)}\} \\ &= (S_1) + (S_2) + (S_3).\end{aligned}$$

Now (S_1) can be estimated by the inductive assumption. Moreover,

$$|(S_2)| = \Big| \int \{(f_k^\alpha)^{(\sigma)}(x) - f_k^\alpha(x-\sigma y)\}\partial_\alpha f_w(x-\sigma y)\varphi(y)dy\Big| \le C\sigma,$$

because f_k^α is locally Lipschitz continuous and $\partial_\alpha f_w$ is locally bounded, since $|w| \le s-1$. Finally,

$$|(S_3)| = \Big| \int \big\{(f_w^\alpha)^\sigma(x) - f_w^\alpha(x-\sigma y)\big\}\partial_\alpha f_k(x-\sigma y)\varphi(y)dy\Big| \le C\sigma,$$

because f_w is Lipschitz and $\partial_\alpha f_k$ is bounded. This concludes the proof of the first item of Proposition 2.4. ∎

Proof of Proposition 2.4, *item* (2). We need to show that, for any $j, k \in \{1, \dots, m\}$, $|w| = s-1$, we have the estimate $|X_j^\sigma X_k^\sigma f_w^\sigma| \le C$, uniformly in $x \in K$ and $\sigma \le \sigma_0$. Write

$$X_j^\sigma X_k^\sigma f_w^\sigma = X_j^\sigma (X_k f_w)^{(\sigma)} + X_j^\sigma\big(X_k^\sigma f_w^\sigma - (X_k f_w)^{(\sigma)}\big) =: M + N.$$

Now, letting $\varphi_\sigma(\xi) = \sigma^{-n}\varphi(\xi/\sigma)$, we have

$$\begin{aligned}
M(x) &= (f_j^\alpha)^{(\sigma)}(x)\partial_{x_\alpha}\int X_k f_w(x-\sigma y)\varphi(y)dy \\
&= -\int (f_j^\alpha)^\sigma(x) X_k f_w(z)\partial_{z_\alpha}(\varphi_\sigma(x-z))dz \\
&= -\int f_j^\alpha(z) X_k f_w(z)\partial_{z_\alpha}(\varphi_\sigma(x-z))dz \\
&\quad + \int \left\{(f_j^\alpha)^\sigma(x) - f_j^\alpha(z)\right\} X_k f_w(z)\frac{(\partial_\alpha\varphi)_\sigma(x-z)}{\sigma}dz.
\end{aligned}$$

The first line can be estimated by integrating by parts by means of (7.1). The estimate of the second line follows from the Lipschitz continuity of the functions f_i.

Next we control N:

$$\begin{aligned}
N(x) &= (f_j^\alpha)^{(\sigma)}(x)\partial_{x_\alpha}\Big\{X_k^\sigma\int f_w(x-\sigma y)\varphi(y)dy - \int (X_k f_w)(x-\sigma y)\varphi(y)dy\Big\} \\
&= (f_j^\alpha)^{(\sigma)}(x)\partial_{x_\alpha}\Big\{\int (f_k^\beta)^\sigma(x)\partial_\beta f_w(z)\varphi_\sigma(x-z)dz \\
&\qquad\qquad - \int f_k^\beta(z)\partial_\beta f_w(z)\varphi_\sigma(x-z)dz\Big\} \\
&= (f_j^\alpha)^{(\sigma)}(x)\int\Big\{(\partial_\alpha f_k^\beta)^\sigma(x)\partial_\beta f_w(z)\varphi_\sigma(x-z) \\
&\qquad + [(f_k^\beta)^\sigma(x) - f_k^\beta(z)]\partial_\beta f_w(z)\frac{1}{\sigma}(\partial_\alpha\varphi)_\sigma(x-z)\Big\}dz.
\end{aligned}$$

The estimate is concluded, because $\partial_\beta f_w$ is bounded, while $|(f_k^\beta)^\sigma(x) - f_k^\beta(z)| \le C\sigma$. ☐

Acknowledgment

The authors wish to express their gratitude to Ermanno Lanconelli for his continuous advice, encouragement and interest in our work, past and present. We dedicate this paper to him with admiration.

References

[BBP] M. Bramanti, L. Brandolini, M. Pedroni, Basic properties of nonsmooth Hörmander's vector fields and Poincaré's inequality, preprint arXiv:0809.2872v2.

[CM] G. Citti, A. Montanari, Regularity properties of solutions of a class of elliptic-parabolic nonlinear Levi type equations, Trans. Amer. Math. Soc. **354** (2002), no. 7, 2819–2848. MR1895205 (2004k:35276)

[FP] C. Fefferman, D. H. Phong, Subelliptic eigenvalue problems. Conference on harmonic analysis in honor of Antoni Zygmund, Vol. I, II (Chicago, Ill., 1981), 590–606, Wadsworth Math. Ser., Wadsworth, Belmont, CA, 1983. MR730094 (86c:35112)

[FL1] B. Franchi, E. Lanconelli, Une métrique associée à une classe d'opérateurs elliptiques dégénérés (French) [A metric associated with a class of degenerate elliptic operators]. Conference on linear partial and pseudodifferential operators (Torino, 1982). Rend. Sem. Mat. Univ. Politec. Torino 1983, Special Issue, 105–114 (1984). MR745979 (86d:35057)

[FL2] B. Franchi, E. 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), 523–541. MR753153 (85k:35094)

[FSSC] B. Franchi, R. Serapioni, F. Serra Cassano, Approximation and imbedding theorems for weighted Sobolev spaces associated with Lipschitz continuous vector fields, Boll. Un. Mat. Ital. B (7) **11** (1997), 83–117. MR1448000 (98c:46062)

[GN] N Garofalo, D. M. Nhieu, Lipschitz continuity, global smooth approximations and extension theorems for Sobolev functions in Carnot-Carathéodory spaces, J. Anal. Math. **74** (1998), 67–97. MR1631642 (2000i:46025)

[Ha] P. Hartman, Ordinary differential equations, John Wiley & Sons, Inc., New York-London-Sydney 1964. MR0171038 (30:1270)

[H] L. Hörmander, Hypoelliptic second order differential equations, Acta Math. **119** (1967), 147–171. MR0222474 (36:5526)

[J] D. Jerison, The Poincaré inequality for vector fields satisfying the Hörmander condition, Duke Math. J. **53** (1986), No. 2, 503–523. MR850547 (87i:35027)

[LM] E. Lanconelli, D. Morbidelli, On the Poincaré inequality for vector fields, Ark. Mat. **38** (2000), 327–342. MR1785405 (2002a:46037)

[MM] A. Montanari, D. Morbidelli, Balls defined by nonsmooth vector fields and the Poincaré inequality, Ann. Inst. Fourier (Grenoble) **54** (2004), 431–452. MR2073841 (2005e:46053)

[MoM] R. Monti, D. Morbidelli, Trace theorems for vector fields, Math. Z. **239** (2002), 747–776. MR1902060 (2003c:46046)

[M] D. Morbidelli, Fractional Sobolev norms and structure of Carnot–Carathéodory balls for Hörmander vector fields, Studia Math. **139** (2000), 213–244. MR1762582 (2002a:46039)

[NSW] A. Nagel, E. M. Stein, S. Wainger, Balls and metrics defined by vector fields. I. Basic properties, Acta Math **155** (1985), 103–147. MR793239 (86k:46049)

[RaS] F. Rampazzo, H. J. Sussmann, Set–valued differential and a nonsmooth version of Chow's theorem, Proceedings of the 40th IEEE Conference on Decision and Control, Orlando, Florida, 2001.

[RaS2] F. Rampazzo, H. J. Sussmann, Commutators of flow maps of nonsmooth vector fields, J. Differential Equations **232** (2007), 134–175. MR2281192 (2007j:49021)

[RoS] L. P. Rothschild, E. M. Stein, Hypoelliptic differential operators and nilpotent groups, Acta Math. **137** (1976), 247–320. MR0436223 (55:9171)

[SW] E. T. Sawyer, R. L. Wheeden, Hölder continuity of weak solutions to subelliptic equations with rough coefficients, Mem. Amer. Math Soc. **180** (2006). MR2204824 (2007f:35037)

[Sem] S. Semmes, Finding curves on general spaces through quantitative topology, with applications to Sobolev and Poincaré inequalities, Selecta Math. (N.S.) **2** (1996), 155–295. MR1414889 (97j:46033)

[Ste] E. M. Stein, Some geometrical concepts arising in harmonic analysis, GAFA 2000 (Tel Aviv, 1999). Geom. Funct. Anal. 2000, Special Volume, Part I, 434–453. MR1826263 (2002f:42014)

[Str] B. Street, Multi-parameter Carnot-Carathéodory balls and the theorem of Frobenius, preprint arXiv:0901.2910v3.

[TW] T. Tao, J. Wright, L^p improving bounds for averages along curves, J. Amer. Math. Soc. **16** (2003), 605–638. MR1969206 (2004j:42005)

Dipartimento di Matematica, Università di Bologna, Piazza di Porta S Donato 5, 40127 Bologna, Italy

E-mail address: `montanar@dm.unibo.it`

Dipartimento di Matematica, Università di Bologna, Piazza di Porta S Donato 5, 40127 Bologna, Italy

E-mail address: `morbidel@dm.unibo.it`